GENETICS AND EVOLUTION OF INFECTIOUS DISEASES



Edited by **MICHEL TIBAYRENC**

ELSEVIER

THIRD EDITION

Genetics and Evolution of Infectious Diseases

This page intentionally left blank

Genetics and Evolution of Infectious Diseases

Third Edition

Edited by

Michel Tibayrenc
Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle, MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD Center, Montpellier, France



Elsevier
Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands
125 London Wall, London EC2Y 5AS, United Kingdom
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Publisher's note: Elsevier takes a neutral position with respect to territorial disputes or jurisdictional claims in its published content, including in maps and institutional affiliations.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices
Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-443-28818-0

For information on all Elsevier publications visit our website at
https://www.elsevier.com/books-and-journals

Publisher: Stacy Masucci
Acquisitions Editor: Kattie Washington
Editorial Project Manager: Ashi Jain
Production Project Manager: Selvaraj Raviraj
Cover Designer: Greg Harris
Cover image: Courtesy of Dr Jenny Telleria
(IRD, Montpellier, France)

Typeset by TNQ Technologies



Contents

Contributors xxv

PART 1 Methodological/Generalist Chapters

CHAPTER 1 A New Official Definition of Virus Species Has Led to a Controversial Linnaean Latinized Binomial Format of Virus Classification 3

Marc H.V. Van Regenmortel and Michel Tibayrenc

1. The Nature of Virus Classification 3
2. The Late Acceptance of the Species Concept in Virus Classification 5
3. Properties Used for Demarcating New Virus Species Are Not the Same as the Diagnostic Properties Used for Identifying the Members of an Established Virus Species 6
4. Popular Anglicized Non-Latinized Virus Species Names Were Subsequently Abandoned 8
5. The ICTV Introduced Names of Virus Species that Correspond to the Italicized Version of the Virus Name, Thereby Creating Unnecessary Confusion 8
6. The New ICTV Definition of Virus Species Is at Odds with the Definition of Species in Other Biological Classifications 9
7. Adrian Gibbs Questions the Merits of a New ICTV Latinized Binomial Nomenclature for Virus Species Associated with Metagenomic Sequences 9
8. Ontology and Epistemology of Viruses and Living Organisms 12

References 12

CHAPTER 2 A Completionist Approach to Discovering and Characterizing Bacterial Diversity 17

Frederick M. Cohan and Sarah Kopac

1. Introduction—Toward a Completionist Bacterial Systematics 17
2. Genomes Enable Completionist Systematics at the Species Level 19
3. The Ecological Breadth of Recognized Species Taxa 24
4. Sexual Isolation and Ecological Divergence in the Origins of Species-like Lineages That can Coexist Indefinitely 27
5. How Modes of Speciation Affect the Correspondence Between Sequence Clusters and Ecotypes 31
6. Genomes Enable Completionist Systematics at the Ecotype Level 38
7. Rapprochement Between Theory-based and Tradition-based Systematics of Bacteria 44

Acknowledgments 46

References 46

CHAPTER 3 Population Structure of Pathogenic Bacteria 57
Xueting Qiu, Lavanya Challagundla, Madikay Senghore, William P. Hanage and D. Ashley Robinson
1. Introduction 57
2. Recombination and Selection in Bacterial Populations 58
2.1 Emergence and Persistence of Sequence Clusters 58
2.2 Heterogeneity in Recombination 59
2.3 The Structure of the Pan-genome of Species and Populations 61
2.4 Negative Frequency−Dependent Selection 62
3. Bacterial Population Structure Within and Between Hosts 64
3.1 Within-host Diversity and Bottlenecks Shape Transmission Dynamics 64
3.2 Between-host Evolution and Impact on Population Structure 65
4. Geography and Bacterial Population Structure 67
4.1 Bacterial Phylogeography and Migration Between Populations 67
4.2 Range Expansion as a Model of Geographic Spread 68
5. Conclusions 69
Acknowledgments 69
References 69

CHAPTER 4 Diverse Strategies and Evolutionary Histories of Fungal Pathogens 79
Cene Gostinčar and Nina Gunde-Cimerman
1. Introduction 79
2. A Brief Overview of Fungal Pathogens 80
2.1 Plant Pathogens 80
2.2 Animal Pathogens 81
3. Selected Examples of Human Pathogens 82
3.1 Major Pathogens: *Candida*, *Cryptococcus, Aspergillus* 83
3.2 Primary Pathogens: *Blastomyces*, *Coccidioidomycosis*, *Paracoccidioides*, *Histoplasma*, *Pneumocystis* 85
3.3 Emerging Fungal Pathogens 86
4. Facultative Pathogenesis and Adaptation Versus Exaptation 87
5. Virulence Factors 89
5.1 Conventional Virulence Factors 89
5.2 Thermotolerance 90
5.3 Extremotolerance 90
5.4 Resistance to Antifungals 91
6. Host and Environment Populations, and Emergence of New Fungal Pathogens 92
6.1 Population Genetics and Genomics 92
6.2 Recombination and Hybridization 95
6.3 Emergence of New Fungal Pathogens 96

7. Conclusions 97
Acknowledgments 98
References 98

CHAPTER 5 Molecular Epidemiology of Pathogenic Microorganisms and the Predominant Clonal Evolution Model 111
Michel Tibayrenc

1. Introduction: What Is Molecular Epidemiology? 111
2. Brief Recall on the PCE Model 113
2.1 What Is, and what Is Not the PCE Model 114
2.2 Deep Phylogenies/Multigene Bifurcating Trees 114
2.3 Other PCE Traits 115
2.4 Evolutionary Scales; The Russian Doll Model 115
2.5 Other Models of Pathogen Population Structure 116
2.6 The Problem of the Wahlund Effect*: The "Starving Sex Hypothesis" 116
3. Bacterial Species 116
3.1 Escherichia coli 116
3.2 Neisseria meningitidis 118
4. Parasite Species 123
4.1 *Leishmania donovani/infantum* 123
4.2 Trypanosoma cruzi 125
5. Yeasts 128
5.1 The *Cryptococcus neoformans/gattii* Complex 128
6. Concluding Remarks 130
Glossary of Specialized Terms 132
Acknowledgments 134
References 135

CHAPTER 6 Parasite–Host Coevolution 141
Brenyn Kelly, Katie Izenour and Sarah Zohdy

1. Introduction 141
2. Common Parasites and Their Definition/Description 143
2.1 Macroparasites 143
2.2 Microparasites 143
2.3 Endoparasites 143
2.4 Fungi 143
2.5 Protozoa 143
2.6 Bacteria 143
2.7 Ectoparasites 144
3. Viruses 144

4. Red Queen Hypothesis and the Evolutionary Arms Race 144
5. Rapidly Evolving Pathogens—RNA Viruses as a Model 145
6. Where Did Viruses Come From? Viral Evolution 146
7. How Do Viruses Evolve and Selection Pressures Not Just on Individual Pathogens With Mutations, but on Groups of Pathogens? Mutant Clouds/ Quasispecies 147
8. Phylogenetics Reveals Coevolutionary Patterns 148
9. Parasite–Host Relationships in Human Infectious Diseases 149
10. Parasite–Host Relationships in Animal Infectious Diseases 150
11. Host Immunogenetics Adaptations 151
12. Parasite Manipulation of Host Behavior 152
13. A Parasite-absent World 153
14. Conclusions 154
References 155

CHAPTER 7 Human Genetic Variability and Transmissible Diseases 163
Michel Tibayrenc

1. Introduction 163
2. A Population Genetics* Framework for the Study of Transmissible Diseases 164
2.1 Overall Genetic Variation Seems to be Greater Within Major Continental Groups Than Between Them 165
2.2 Clines* or Trees? 165
2.3 Additional Variability and Stratifications 166
2.4 Low-Frequency and Rare Variants 166
2.5 "Russian Doll" Patterns 166
2.6 Genetic Stratifications and Polygenic Scores* 167
2.7 We are All Half-Breed: Mixtures are Ubiquitous 167
2.8 If We are All Mestizos, Don't Human Populations Tend Toward Homogenization? 167
2.9 Ancestry Informative Markers and Self-Reported Ancestry 168
2.10 General Conclusions 168
3. Communicable Diseases: A Phenomenon of Co-Evolution Involving Two or Three Actors 168
3.1 The Considerable Role of Environment 169
3.2 Genetic Inequality and Transmissible Diseases 169
3.3 How do Humans Protect Themselves Against Pathogens? 171
3.4 Other Diseases 171
3.5 Pathogen Strains as Markers of Human Populations 172
3.6 Influence of Human Behavior on Transmissible Disease Epidemiology: The Baldwin Effect and the Niche Construction Theory 172
3.7 Transmissible Diseases and History 173

3.8 Archaic Adaptive Introgression 173
4. Conclusions on Human Genetic Variability and Susceptibility to Communicable Diseases 173
Glossary of Specialized Terms 174
Acknowledgments 176
References 176

CHAPTER 8 Pathogens as Tracers of Past Human Demography and Migrations 181
Charlotte J. Houldcroft
1. Introduction 181
2. Pathogen Genomics to Infer Host History 182
3. Bacteria 185
3.1 Helicobacter pylori 185
3.2 Mycobacterium tuberculosis 187
3.3 Mycobacterium leprae 189
4. Viruses 193
4.1 Human Polyomaviruses 193
4.2 Human Herpesviruses 195
5. Conclusion 197
Acknowledgments 199
References 199

CHAPTER 9 Evolutionary Responses to Infectious Disease 213
G. Cochran and H. Harpending
1. Introduction 213
2. Parasites as Our Friends 215
3. Demography and Parasites 215
4. Agriculture 216
5. Some Lessons from Malaria 217
6. The Columbian Exchange 219
7. Disease and Standard of Living in Preindustrial Societies: A Simple Model 220
8. Population Limitation 221
9. Disease, Mating, and Reproductive Strategy 223
10. Prosperity and the Postindustrial Era Mortality Decline 224
References 225

CHAPTER 10 Infectious Disease Genomics 229
Yu-Tsueng Liu
1. Introduction 229
2. Vaccine Target 230
3. New Drug Discovery 231

4. Drug Target 231
5. Therapeutic Response and Drug Resistance 232
6. Vector Control 233
7. Clinical Application 234
8. Conclusion 237
References 237

CHAPTER 11 Proteomics and Host–Pathogen Interactions: Still a Bright Future? **243**
David G. Biron, Dorothée Missé, Dobrin Nedelkov and Philippe Holzmuller

1. Introduction 243
2. Interest of Proteomics to Study Host–(Vector)–Pathogen Interactions 244
3. Retrospective of Pioneering Proteomics Studies 245
3.1 Deciphering of the Molecular Strategies Involved in Pathogen Immune Evasion 246
3.2 Host Proteome Responses to Parasite Infection 249
3.3 Biomarkers Linked to Infection Process by a Pathogen Using SELDI-TOF-MS Technology 251
4. Evolution of Conceptual Approaches to Decipher the Host–Pathogen Interactions for Microorganisms With Simple or Complex Life Cycle 252
4.1 From "Parasito-Proteomics" to "Pathogeno-Proteomics": An Holistic View to Disentangle the Hosts and Pathogens Genome Responses During Their Interactions 252
4.2 Proteomics of Arthropod Vectors: Understanding the Role of Sialomes in the Host–Vector Interactions 256
4.3 Pitfalls of the Current Proteomics Approaches 257
4.4 Proteo-Genomics: Next-Generation Proteomics to Deepen the Characterization of Proteins, Their Biological Functions, and Their Interactions in Complex Environments 258
4.5 Host–Pathogen Interactomes 260
4.6 Population Proteomics, a Growing Discipline, to Study Host–Pathogen Interactions 261
5. Five-Year View 270
5.1 Metabolomics and Fluxomics, Key "Omics" Tools to Complete Deciphering of Host–Pathogen Crosstalks 270
5.2 Applied Outputs: New Diagnostic Tools and Identification of New Therapeutic Target 271
5.3 Bioterrorism and Proteomics 272
5.4 Environment and HostMicroorganisms Interactions 272
6. Conclusions 274

Glossary 274
Short Biographical Lines for Contributors 275
References 276

CHAPTER 12 Genomic Epidemiology and Surveillance of Antimicrobial Resistance 291
Neris García-González, Irving Cancino-Muñoz, Leonor Sánchez-Busó and Fernando González-Candelas

1. Antimicrobial Resistance: A Current Major Threat 291
2. Genomic Epidemiology and Surveillance in the Fight Against AMR 292
3. Genotypic Prediction of Antimicrobial Resistance 293
3.1 Methods Based on Raw Reads 294
3.2 Assembly-based Methods 295
3.3 Databases 295
3.4 New AMR Determinants 296
4. Case Studies: Genomics and AMR in Clinically Relevant Bacteria 296
4.1 *Mycobacterium tuberculosis* 296
4.2 *Neisseria gonorrhoeae* 298
4.3 *Klebsiella pneumoniae* 302
5. Future Perspectives: A Gold Standard Becoming Routine? 305
6. Conclusions 306
Acknowledgments 307
References 307

CHAPTER 13 Evolution of Resistance to Insecticides in Disease Vectors 317
Philippe Nwane, Achille Binyang, Lili Ranaise Mbakop and Josiane Etang

1. Introduction 317
2. Insecticide Classification and Modes of Action 319
2.1 Natural Insecticides 319
2.2 Synthetic Insecticides 320
2.3 Alternative Insecticides 322
2.4 New Molecules Used in Vector Control 322
3. Mechanisms of Resistance 324
3.1 Metabolic Resistance 324
3.2 Target-Site Modification 327
3.3 Other Resistance Mechanisms 331
4. Methods for Testing and Monitoring Insecticide Resistance in Mosquitoes 332
4.1 Bioassays 332
4.2 Identification of Resistance Mechanisms 334
5. Features of Insecticide Resistance 335

5.1 Evolvement of Resistance 335
5.2 Cross-Resistance 335
5.3 Multiple Resistance 336
6. Factors Governing the Evolution of Insecticide Resistance 336
6.1 Biological Factors 337
6.2 Genetic Factors 337
6.3 Operational Factors 339
7. Insecticide Resistance in Disease Vectors 339
7.1 Malaria Vectors 340
7.2 Arbovirus Vectors 340
7.3 Other Vectors 340
8. Approaches to Insecticide Resistance Management 341
9. Conclusion 342
References 343
Further reading 355

CHAPTER 14 Genetics of Tsetse Flies (Diptera: Glossinidae), Vectors of Human and Animal Trypanosomoses 357
Philippe Solano
1. Introduction 357
2. Reminder on the Taxonomy of Tsetse Flies 358
3. Some Important Life Traits of Tsetse regarding Population Genetics 359
3.1 Reproduction 359
3.2 Lifespan and Population Dynamics 360
3.3 Natural Habitats 361
3.4 Dispersal, Diet, Activity Cycle 361
3.5 Tsetse Bacteriome 362
4. Populations Genetics of Tsetse Flies 362
4.1 Species Complexes, Subspecies and Ecotypes 363
4.2 Use of Genetics for Eradication/Control 364
5. Future Work 365
Acknowledgments 365
References 365

CHAPTER 15 Genetics of the Triatominae (Hemiptera: Reduviidae), Vectors of Chagas Disease 369
S.F. Brenière, E. Waleckx, S. Pita, F. Panzera, C. Barnabé and P.L. Dorn
1. Introduction 369
2. The Triatominae, Vectors of Chagas Disease 369
3. Evolution and Taxonomy of the Triatominae 371
4. Cytogenetics of the Triatominae 371

5. Genetic Diversity and Population Structure of Triatominae Species 377
6. Conclusions and Future Directions 378
Acknowledgments 380
References 380

CHAPTER 16 Modern Morphometrics of Arthropods: A Phenotypic Approach to Species Recognition and Population Structure 385
Jean-Pierre Dujardin

1. Introduction 385
1.1 Modern and Traditional Morphometrics 385
2. Landmark-Based Geometric Morphometry 386
2.1 The Landmark 386
2.2 Landmark-Based Size 386
2.3 Landmark-Based Shape 387
2.4 Semilandmarks for Open Curves 389
3. Pseudolandmark-Based Geometric Morphometry (Outlines) 389
3.1 Pseudolandmark-Based Shape 390
3.2 Pseudolandmark-Based Size 390
4. Allometry 390
5. Some Considerations About the Genetics of Metric Change 391
5.1 Shape as a Polygenic Character 391
5.2 Genetic Drift 392
5.3 Heritability 392
5.4 Hidden Genetic Variability 393
5.5 Hybridism 393
6. Some Considerations About Phenotypic Variation 395
6.1 Phenotypic Plasticity 395
6.2 The Regulation of Phenotype 396
7. Taxonomic Questions 398
7.1 The Taxonomic Signal of Shape 399
7.2 The Taxonomic Use of Morphometrics 399
7.3 Examples 403
8. Population Structure Studies 406
8.1 Gene Flow and the Flow of Migrants 406
8.2 The Need for a Heuristic 408
8.3 Reinfestation Studies 409
8.4 Species Richness 410
9. Conclusion 412
Glossary 412
Acknowledgments 412
References 412

CHAPTER 17 Microbial Sequence Typing Strategies for Genetic and Genomic Data 427
Marcos Pérez-Losada, Eduardo Castro-Nallar and Miguel Arenas
1. Introduction 427
2. DNA-Based Typing Approaches 428
2.1 Standard Typing Approaches: MLST and MLVA 428
2.2 NGS-Based Approaches: WGS, SNP, rMLST, cgMLST, HiMLST, NGMLST, and MetaMLST 431
3. DNA-Based Typing Analyses 432
3.1 Allele-Based Analytical Methods 434
3.2 Nucleotide-Based Analytical Methods 435
3.3 SNPs-Based Analyses 439
3.4 GWS-Based Analyses 441
4. Microbial Sequence Typing Applications 443
4.1 Molecular Epidemiology 444
4.2 Genotyping and Phylogenetic Inference 446
5. Conclusions and Prospects 448
Acknowledgments 449
References 449

CHAPTER 18 Proteomics, Bioinformatics, and Infectious Diseases 465
Jennifer Geddes-McAlister, Florence Roux-Dalvai and Arnaud Droit
1. Introduction 465
1.1 Mass Spectrometry—Based Proteomics 466
1.2 Bioinformatics Tools in Proteomics 470
2. Proteomics of Infectious Diseases 471
2.1 Disease Applications 471
2.2 Clinical Applications 474
2.3 Technological Advances 475
3. Combining Proteomics and Bioinformatics Towards Fighting Infectious Diseases 476
3.1 Diagnostic and Prognostic Markers 477
3.2 Target and Drug Discovery 478
4. Concluding Remarks and Future Directions 479
Acknowledgments 479
References 480

CHAPTER 19 Genomics of Infectious Diseases and Private Industry 493
Alex van Belkum and Guy Vernet
1. Introduction 493
2. Customers and Applications 496

3. Commercial Sequencing Instruments, Reagents, and Software 498
4. Library Preparation 505
5. Genomics Service Companies and Infectious Diseases 506
6. Conclusion 506
References 507

CHAPTER 20 Emerging Zoonoses, Biodiversity, and Ecosystems 511
Serge Morand
1. Introduction 511
2. A Long History 512
2.1 Environment as a Health Factor 512
2.2 The Rise of Hygienism 513
2.3 Environment and Health in Colonial Sciences 513
2.4 The Unesco Conference on the Biosphere 514
2.5 Ecosystem Services and Health 515
3. One Health 516
3.1 The Origin of One Health 516
3.2 A New One Health Definition 516
3.3 Preventing Emerging Zoonoses 518
4. Conclusion 519
Acknowledgments 520
References 520

CHAPTER 21 Epigenetics and Transmissible Diseases 523
Mohamed Kamel, Josiah Levi Davidson, Rachel A. Munds, Jiangshan Wang, Simerdeep Kaur, Virendra Kumar and Mohit S. Verma
1. Introduction and Definition of Epigenetics 523
1.1 Definition of Epigenetics 523
1.2 The Impact of Epigenetic Modifications on Gene Expression 524
1.3 The Role of Epigenetics in Disease Development 525
2. Epigenetics and Its Role in Infectious Diseases 525
2.1 The Impact of Epigenetic Modifications on Infectious Disease Transmission and Host Susceptibility 525
2.2 Epigenetics Role in Resistance to Infectious Pathogens 526
2.3 Epigenetics Role in Greater Susceptibility to Infectious Pathogens 528
3. The Role of Infectious Diseases in Epigenetic Alterations 529
3.1 Mechanisms by Which Infectious Diseases Induce Epigenetic Change 530
3.2 Epigenetic Alterations Associated With Different Infectious Diseases 531
3.3 Long-term Effects of Epigenetic Changes Induced by Infectious Diseases 532
4. Potential Role of Laboratory Tests in Detecting Infectious Diseases Based on Epigenetic Signatures 533

4.1 Epigenetic Biomarkers and Laboratory Tests in Detecting Infectious Diseases 533
4.2 Epigenetic Lab Tests for Detecting Gene Expressions Related to Infectious Diseases 534
4.3 Diagnostic Use of Gene Expression Profiling in Epigenetics for Infectious Diseases 534
5. Strategies to Limit the Impact of Infectious Diseases on Epigenetics 535
5.1 Strategies to Reduce the Impact of Infectious Diseases on Epigenetics 535
5.2 Preventative Measures for Minimizing Epigenetic Changes Caused by Infectious Diseases 536
5.3 Reducing the Impact of Infectious Diseases on Epigenetics 537
6. Potential Applications of Epigenetic Factors in Preventive Medicine 537
6.1 Use of Epigenetic Modifications for Prevention and Treatment of Infectious Diseases 538
6.2 Potential Therapies Based on Epigenetic Factors Against Infectious Diseases 539
6.3 Informing Public Health Policy Through Epigenetic Research 540
7. Conclusion and Future Perspectives 540
References 541

CHAPTER 22 The Evolution of SARS-CoV-2 and Influenza A Virus at the Human–Animal Interface 549
Tavis K. Anderson, Rafael A. Medina and Martha I. Nelson
1. Introduction: New Perspectives on the Bidirectional Human–Animal Interface 549
2. A Tangled History of Influenza A Virus Evolution in Humans and Swine 551
3. Viral Adaptations Required for Influenza A Virus Host Switches 554
4. SARS-CoV-2 Evolution at the Human–Animal Interface 557
5. Future Perspectives 559
Acknowledgments 560
References 560

CHAPTER 23 Emerging Viruses 573
Rafael Sanjuán
1. Introduction: Emergent Viruses and the Virosphere 573
2. Ecological and Societal Drivers of Viral Emergence 575
2.1 Viral Emergence as an Ecological Process 575
2.2 Role of Societal Changes in Viral Emergence and Evolution 577
3. The Sources of Viral Emergence 578
3.1 Zoonotic Viruses Originate Preferentially From Some Host Taxa 578
3.2 The Importance of Viral Diversification in Reservoir Hosts 579

3.3 The Role of Intermediate Hosts in Viral Zoonosis 580
4. Genetic and Molecular Determinants of Viral Emergence 582
4.1 Virological Features 582
4.2 Receptor Specificity as a Determinant of Cross-Species Transmission 583
4.3 Additional Molecular Determinants of Cross-Species Transmission: Examples from Influenza Virus 585
5. Viral Establishment Following Cross-Species Transmission 588
5.1 Unique Versus Multiple and Overt Versus Cryptic Zoonoses 588
5.2 Selection for Increased Transmissibility in Emerging Viruses 589
5.3 Longer-Term Evolution of Emerging Viruses: Virulence and Immune Escape 591
6. Closing Remarks and Future Perspectives 592
Acknowledgments 593
References 593

PART 2 Specialized Chapters

CHAPTER 24 Genetic Exchange in Trypanosomatids and its Relevance to Epidemiology 607

Wendy Gibson, Michael D. Lewis, Matthew Yeo and Michael A. Miles

1. Introduction 607
2. Trypanosoma brucei 608
2.1 Genetic Crosses 608
2.2 Location of Genetic Exchange 609
2.3 Mendelian Inheritance and Meiosis 610
2.4 Inheritance of Kinetoplast DNA 613
2.5 Implications for Epidemiology 613
3. Trypanosoma cruzi 614
3.1 *T. cruzi* Diversity 614
3.2 Genetic Crosses 616
3.3 Location of Genetic Exchange 617
3.4 Behavior of Experimental Hybrids 618
3.5 Mechanism of Genetic Exchange in Experimental and Natural Hybrids 619
3.6 Implications for Epidemiology 621
4. Leishmania 621
4.1 Evidence of Genetic Exchange in Natural Populations 621
4.2 Genetic Crosses in Sand Flies 623
4.3 Genetic Crosses In Vitro 623
4.4 Implications for Epidemiology 624
Abbreviations 624

Acknowledgments 624
References 625

CHAPTER 25 Genomic Insights Into the Past, Current, and Future Evolution of Human Parasites of the Genus *Plasmodium* 635
Colin J. Sutherland, Khalid B. Beshir and Spencer D. Polley

1. Introduction 635
1.1 Overview of Plasmodium Phylogeny and Description of Species Infecting *Homo sapiens* 635
1.2 Population Genetics and Design of Public Health Interventions 637
1.3 Genomic Signals of Selection due to Host Immunity 637
1.4 The Impact of New Technologies on Genomic Studies of the Genus 638
2. Evolution of *Plasmodium*: The Last 10 Million Years 639
2.1 Role of Host Transitions in Speciation Events Within the Genus 639
2.2 *P. falciparum* and *P. reichenowi*: Divergent Host Specificities? 640
2.3 Speciation Between *P. ovale curtisi* and *P. ovale wallikeri*: Separate Host Transitions? 641
2.4 Importance of Host Specificity 643
2.5 Contrasting Drug Susceptibilities: Divergent Biochemical Adaptation Within the Genus 644
3. Evolution of *Plasmodium*: The 20th Century 645
3.1 Habitat Destruction and Ecological Transitions 645
3.2 Emergence of *P. knowlesi* as an Important Zoonosis 646
3.3 Chemotherapy and the Pharmaceutical Selection of New Parasite Gene Variants 647
4. Future Challenges From Continuing *Plasmodium* Genome Adaptation 648
References 649
Further reading 656

CHAPTER 26 Host Adaptation in the *Mycobacterium* Genus: An Evolutionary and Genomic Perspective 657
Juan Calvet Seral, Elena Campos Pardos and Jesús Gonzalo-Asensio

1. Overview: Tuberculosis Disease and Historic Events in Mycobacterial Genetics 657
2. Host–Pathogen Co-evolution of the Tubercle Bacillus in Humans and Other Mammals 659
2.1 *M. tuberculosis* 662
2.2 *M. africanum* 663
2.3 *M. canettii* 663
2.4 Animal-Adapted Ecotypes: The Four Animal Clades 664
3. Evolution of the MTBC From a Genomic Perspective 666

4. Evolution in the Laboratory Environment and In Vitro Attenuation of Bacteria From the MTBC....668
5. Short-term Evolution of *M. tuberculosis* During Infection, Drug Treatment, and Disease....670
6. Adaptative Cues of the MTBC as the Most Successful Pathogens....672
7. Pending Questions and Concluding Remarks....675
List of Authors or Contributors....676
References....676

CHAPTER 27 Evolution, Phylogenetics, and Phylogeography of *Mycobacterium tuberculosis* complex....683
Christophe Sola, Igor Mokrousov, Muhammed Rabiu Sahal, Kevin La, Gaetan Senelle, Christophe Guyeux, Guislaine Refrégier and Emmanuelle Cambau

1. Introduction to *Mycobacterium tuberculosis* Complex Genetic Diversity Studies....683
1.1 Tuberculosis as an Animal or Human Disease: Historical, Epidemiological, Ecological, and Archeological Considerations....683
1.2 The Origin of TB Molecular Epidemiology....684
1.3 Back to the Present Epidemiological Situation and Link With Science and Politics....685
1.4 The Discovery of BCG....685
1.5 Milk Pastorization and Its Impact on TB....686
1.6 A Change of Paradigm....686
1.7 Tuberculosis Archeology....688
1.8 Evolutionary History of TB: Theoretical Issues....688
1.9 Development of Tuberculosis Genetics and Links to Clinics....690
1.10 Genetic Diversification Mechanisms and Phage Typing....690
1.11 Main Lineages Identified so Far....691
2. An Introduction to Genomics and Computational Analysis Methods of Whole-Genome Sequencing Data....692
2.1 Basic Methodology of WGS....692
2.2 WGS Applications: Drug-Resistance Studies and Genomic Epidemiology....694
2.3 WGS Allows the Most Accurate Classification of MTBC....694
2.4 MTBC Genomes Classification Across Time....695
2.5 Technological Evolution of Molecular Epidemiology....695
3. Lineage 1: The "East-African Indian" (EAI) Lineage According to Sola et al. (2001) or "Indo-Oceanic" According to Gagneux et al. (2006)....696
3.1 An History of L1 Discovery and on Issues Left Unsolved....696
3.2 Genomic Characteristics of L1....698

3.3 Geographical Distribution of L1 698
3.4 Importance of South-East and South Asia in Early History of TB 698
3.5 Specific Data on the Manila Lineage 699
4. Lineage 2: The "Beijing" or East-Asian Lineage 699
4.1 History of the Beijing Lineage Naming 699
4.2 Nomenclature Inside L2 Suggests a Complex History 700
4.3 Diversification History of L2 700
4.4 Recent Expansion of the L2 Phylogenetical Tree 701
5. Lineage 3: The Delhi or Central-Asian (CAS) Lineage According to Filliol et al. (2002) or "East-African Indian" According to Gagneux et al. (2006) 701
5.1 Lineage 3 Discovery History and Specific Characteristics 701
5.2 Description of Main L3 Sublineages 702
5.3 Phylogeography of L3 and Comparison to L1 702
6. Lineage 4: The By-default "T" Lineage (Sola et al. 1999) Renamed as "Euro-American" Lineage (Gagneux et al. 2006) 703
6.1 Lineage 4 Characterization and Discovery History 703
6.2 Description of the L4 Sublineages Characteristics (SNPs Variant and RD Content), With an Attempt to Define Their Phylogeographical Specificity 708
7. Lineages 5 and 6: *Mycobacterium africanum* (Maf) West African 1 and 2 728
7.1 Focus on Lineage 5: *Mycobacterium africanum*, West African I 728
7.2 Focus on Lineage 6: *Mycobacterium africanum*, West-African II 733
8. Focus on Lineage 7 734
8.1 Chronology of Discovery of Lineage 7 734
9. The Lineage 8 (Uganda, Rwanda) 735
9.1 A Sister Lineage of MTBC Found in the Great Lakes Region 735
10. The Lineage 9 and 10: Newly Discovered *M. africanum* Sublineages 735
10.1 Discovery of New *Mycobacterium africanum* Lineages 735
10.2 Geographical Origin of L9 Is Likely to Be Found in East Africa 736
11. Animal Lineages and *Mycobacterium canettii* 736
11.1 Animal Lineages 736
11.2 *Mycobacterium canettii* 740
12. Variability of the Clinical Expression of the Tuberculosis Disease and of Virulence of MTBC as a Function of the Lineage 741
12.1 The Issue of Adaptation of the Pathogen to Its Host 741
12.2 Differential Virulence Inside MTBC 742
12.3 Importance of Polyketide Synthases 743
12.4 Differential drug resistance according to lineages and sublineages: 743
12.5 Quantitative Genetics and GWAS 743

13. Conclusion 744
References 745
Further Reading 772

CHAPTER 28 The Evolution of *Staphylococcus aureus* **773**
JoAnn McClure and Kunyan Zhang
1. Introduction 773
2. Evolution of *Staphylococcus aureus* 774
3. Methicillin Resistance and the Staphylococcal Cassette Chromosome *mec* 778
3.1 *mec* Gene Complex 779
3.2 *ccr* Gene Complex 780
3.3 J Regions 782
3.4 SCC*mec* Types 782
4. Typing Methods for *S. aureus* Characterization 784
4.1 Pulsed Field Gel Electrophoresis (PFGE) 784
4.2 Multilocus Sequence Typing (MLST) 784
4.3 Staphylococcal Protein A (*spa*) Typing 785
4.4 SCC*mec* Typing 785
4.5 Whole Genome Sequence (WGS) Typing 785
5. Molecular Dynamics in MRSA Evolution 786
5.1 The Contribution of Mobile Genetic Elements to MRSA Evolution 787
6. Global MRSA Lineages 789
6.1 HA-MRSA 789
6.2 CA-MRSA 793
6.3 LA-MRSA 794
7. Concluding Remarks 795
References 795

CHAPTER 29 Origin and Emergence of HIV/AIDS **811**
Martine Peeters, Mirela D'arc, Lucie Etienne, Eric Delaporte and Ahidjo Ayouba
1. History of AIDS 811
2. Human Immunodeficiency Viruses are Closely Related to Simian Immunodeficiency Virus From Nonhuman Primates 813
2.1 Discovery of the First SIV 813
2.2 SIVs in African NHP 813
2.3 Pathogenicity of SIVs in Their Natural Hosts 816
2.4 Evolution and Phylogeny of SIVs 818
3. HIV-1 Is Derived From SIVs Circulating Among African Apes 819
3.1 SIVs From Chimpanzees and Gorillas are the Ancestors of HIV-1 in Humans 819

3.2 The Cross-Species Transmissions Resulting in HIV-1 Viruses in Humans Occurred in West Central Africa ... 822
3.3 HIV-1 Started to Diverge in the Human Population at the Beginning of the 20th Century ... 822
3.4 SIVs are Transmitted to Humans by Exposure to Infected Primates ... 823
4. Origin of HIV-2: An Other Emergence, an Other Epidemic ... 823
5. Ongoing Exposure of Humans to a Large Diversity of SIVs: Risk for a Novel HIV? ... 824
5.1 Exposure to SIV Infected NHP is Still Ongoing ... 824
5.2 SIV Prevalences and Cross-Species Transmissions ... 825
5.3 Host Restriction Factors ... 826
5.4 Viral Adaptation ... 826
5.5 Human Factors ... 827
5.6 Ongoing Cross-Species Transmissions From Other Retroviruses From Primates to Humans ... 828
6. Conclusion ... 828
References ... 829

CHAPTER 30 Zoonotic Origin and Evolution of SARS Coronavirus ... 841
Zhengli Shi, Ben Hu and Lin-Fa Wang
1. A Brief History of SARS ... 841
2. SARS Coronavirus (SARS-CoV) ... 843
3. The Animal Link of SARS-CoV ... 843
4. Natural Reservoirs of SARS-CoV ... 845
5. Molecular Evolution of SARS-CoV in Humans and Animals ... 849
5.1 Rapid Adaptation of SARS-CoVs in Humans ... 849
5.2 Receptor Usage and Evolutionary Selection ... 849
5.3 Potential Recombination Origin of SARS-CoV ... 851
6. Preparedness of Future SARS-like Disease Emergence ... 852
6.1 Virus Surveillance in Wildlife ... 852
6.2 Development of Pan-Coronavirus Therapeutics and Vaccines ... 853
7. Concluding Remarks ... 853
References ... 854

CHAPTER 31 Ecology and Evolution of Avian Influenza A Viruses in Wild Birds ... 863
Michelle Wille
1. Role of Wild Birds as a Reservoir for a Diversity of Viruses ... 863
2. Avian Influenza A Virus ... 865
2.1 Introduction ... 865
2.2 Subtype Diversity in Birds ... 870

3. AIV Ecology in Waterfowl 870
3.1 Mallards as Model Species for Studies of LPAIV Ecology 870
4. Beyond the Duck-Centric, Temperate Northern Hemisphere Model of LPAIV Ecology 876
4.1 AIV Ecology in the Southern Hemisphere 876
4.2 AIV in Shorebirds 877
4.3 AIV in Gulls and Other Seabirds 879
5. Effect of Infection on Birds 881
6. Evolution and Genetics 882
7. The Role of Wild Birds in the HPAIV H5N1 Panzootic 884
References 888

CHAPTER 32 Evolution of SARS-COV-2 899
Manuela Sironi and Fernando González-Candelas
1. Introduction 899
2. Coronaviruses 899
2.1 Classification and Phylogenetic Relationships 899
2.2 Genomic Features 900
2.3 Host Range and Ecology 902
3. Origin of SARS-CoV-2 903
3.1 Phylogenetic Relationships of SARS-CoV-2 to Other CoVs 903
3.2 Initial Emergence and Spread 905
3.3 Hypotheses About the Origin of SARS-CoV-2 906
4. Genetic Diversity of SARS-CoV-2 907
4.1 Mutation and Substitution Rates 907
4.2 Recombination 909
4.3 Nomenclature of SARS-CoV-2 Lineages 913
5. Evolution of SARS-CoV-2 915
5.1 VOC/VOI Emergence 915
5.2 Evolution of Transmissibility 917
5.3 Evasion of Host Immunity 918
5.4 Ongoing Evolution 920
6. Conclusion 921
Acknowledgments 922
References 922

CHAPTER 33 Molecular Tools for Systematic and Epidemiological Investigations of Socioeconomically Important Parasitic Nematodes of Animals and Humans 933
Robin B. Gasser, Ian Beveridge and Neil B. Chilton
1. Background and Historic Account 933
2. Critical Evaluation of Internal Transcribed Spacers of Nuclear Ribosomal DNA as Specific Markers for Parasitic Nematodes 934
3. Taxonomic Investigations 936
4. Phylogenetic Investigations 938
5. Diagnostic and Epidemiological Investigations 939
6. Nemabiome Investigations 940
7. Considerations and Extensions 941
8. Mitochondrial Genetic Markers 942
9. Technological Advances and Prospects 943
10. Concluding Remarks 944
Acknowledgments 944
References 945
Further reading 958

Index 959

Contributors

Tavis K. Anderson
Virus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames, IA, United States

Miguel Arenas
CINBIO, Universidade de Vigo, Vigo, Spain; Department of Biochemistry, Genetics and Immunology, Universidade de Vigo, Vigo, Spain

Ahidjo Ayouba
TransVIHMI, University of Montpellier, Institut de Recherche pour le Développement (IRD), INSERM, Montpellier, France

C. Barnabé
Institut de Recherche pour le Développement, UMR INTERTRYP IRD, CIRAD, Université de Montpellier, Montpellier, France

Khalid B. Beshir
Department of Infection Biology, London School of Hygiene & Tropical Medicine, London, United Kingdom

Ian Beveridge
Department of Veterinary Biosciences, Melbourne Veterinary School, The University of Melbourne, Melbourne, VIC, Australia

Achille Binyang
Laboratory of Animal Biology and Physiology, Faculty of Sciences, University of Yaoundé I, Yaoundé, Cameroon

David G. Biron
Laboratoire Microorganismes: Génome et Environnement, UMR 6023 CNRS/UCA, Campus Universitaire des Cézeaux, Aubière, France

S.F. Brenière
Institut de Recherche pour le Développement, UMR INTERTRYP IRD, CIRAD, Université de Montpellier, Montpellier, France

Emmanuelle Cambau
Associate laboratory of the National Reference Center for Mycobacteria and Antimycobacterial Resistance, AP-HP GHU Nord, Mycobacteriology, Inserm, IAME UMR 1137, University of Paris Cité, Paris, France

Irving Cancino-Muñoz
Joint Research Unit "Infection and Public Health" FISABIO-University of Valencia, Institute for Integrative Systems Biology (I2SysBio, UV-CSIC), Valencia, Spain

Eduardo Castro-Nallar
Departamento de Microbiología, Facultad de Ciencias de la Salud, Universidad de Talca, Talca, Chile; Centro de Ecología Integrativa, Universidad de Talca, Talca, Chile

Lavanya Challagundla
Department of Pharmacology and Toxicology, University of Mississippi Medical Center, Jackson, MS, United States

Neil B. Chilton
Department of Biology, University of Saskatchewan, Saskatoon, SK, Canada

G. Cochran
University of Utah, Salt Lake City, UT, United States

Frederick M. Cohan
Department of Biology, Wesleyan University, Middletown, CT, United States

Josiah Levi Davidson
Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, United States; Birck Nanotechnology Center, Purdue University, West Lafayette, IN, United States

Eric Delaporte
TransVIHMI, University of Montpellier, Institut de Recherche pour le Développement (IRD), INSERM, Montpellier, France

P.L. Dorn
Loyola University New Orleans, New Orleans, LA, United States

Arnaud Droit
Proteomics Platform, CHU de Québec - Université Laval Research Center, Québec City, QC, Canada; Computational Biology Laboratory, CHU de Québec - Université Laval Research Center, Québec City, QC, Canada; Canadian Proteomics and Artificial Intelligence Consortium, Guelph, ON, Canada

Mirela D'arc
TransVIHMI, University of Montpellier, Institut de Recherche pour le Développement (IRD), INSERM, Montpellier, France; Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

Jean-Pierre Dujardin
Institut de Recherche pour le Développement (IRD), UMR INTERTRYP IRD-CIRAD, University of Montpellier, Montpellier, France

Josiane Etang
Malaria Research Laboratory, Yaoundé Research Institute (IRY), Organization for the Coordination of Endemic Diseases' control in Central Africa (OCEAC), Yaoundé, Cameroon; Department of Biological Sciences, Faculty of Medicine and Pharmaceutical Sciences, University of Douala, Douala, Cameroon; Institute for Insect Biotechnology, Justus Liebig University Gießen, Gießen, Germany

Lucie Etienne
TransVIHMI, University of Montpellier, Institut de Recherche pour le Développement (IRD), INSERM, Montpellier, France; International Center for Infectiology Research, INSERM U1111 and Ecole Normale Supérieure de Lyon and Université Lyon 1 and CNRS UMR Lyon, France

Neris García-González
Joint Research Unit "Infection and Public Health" FISABIO-University of Valencia, Institute for Integrative Systems Biology (I2SysBio, UV-CSIC), Valencia, Spain

Robin B. Gasser
Department of Veterinary Biosciences, Melbourne Veterinary School, The University of Melbourne, Melbourne, VIC, Australia

Jennifer Geddes-McAlister
Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada; Canadian Proteomics and Artificial Intelligence Consortium, Guelph, ON, Canada

Wendy Gibson
University of Bristol, School of Biological Sciences, Bristol, United Kingdom

Jesús Gonzalo-Asensio
Mycobacterial Genetics Group, Department of Microbiology, Faculty of Medicine, University of Zaragoza, Zaragoza, Spain

Fernando González-Candelas
University of Valencia, Institute for Integrative Systems Biology (I2SysBio, UV-CSIC), Valencia, Spain; CIBER in Epidemiology and Public Health, Valencia, Spain; FISABIO Foundation, Joint Research Unit Infection and Public Health, Valencia, Spain; Joint Research Unit "Infection and Public Health" FISABIO-University of Valencia, Institute for Integrative Systems Biology (I2SysBio, UV-CSIC), Valencia, Spain

Cene Gostinčar
University of Ljubljana, Biotechnical Faculty, Department of Biology, Ljubljana, Slovenia

Nina Gunde-Cimerman
University of Ljubljana, Biotechnical Faculty, Department of Biology, Ljubljana, Slovenia

Christophe Guyeux
FEMTO-ST Institute, UMR 6174 CNRS, DISC Computer Science Department, University Franche-Comté (UFC), Besançon, France

William P. Hanage
Department of Epidemiology, Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, MA, United States

H. Harpending
University of Utah, Salt Lake City, UT, United States

Philippe Holzmuller
ASTRE (Animal, santé, territoires, risques et écosystèmes), UMR 117 CIRAD/INRAE/UM, TA A-117/E, Campus International de Baillarguet, Montpellier, France

Charlotte J. Houldcroft
Department of Genetics, University of Cambridge, Cambridge, United Kingdom; McDonald Institute for Archaeological Research, Cambridge, United Kingdom

Ben Hu
CAS Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China

Katie Izenour
College of Veterinary Medicine, Auburn University, Auburn, AL, United States

Mohamed Kamel
Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, United States; Department of Medicine and Infectious Diseases, Faculty of Veterinary Medicine, Cairo University, Giza, Egypt; Birck Nanotechnology Center, Purdue University, West Lafayette, IN, United States

Simerdeep Kaur
Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, United States; Birck Nanotechnology Center, Purdue University, West Lafayette, IN, United States

Brenyn Kelly
College of Forestry, Wildlife, and the Environment, Auburn University, Auburn, AL, United States

Sarah Kopac
Department of Biology, Wesleyan University, Middletown, CT, United States

Virendra Kumar
Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, United States; Birck Nanotechnology Center, Purdue University, West Lafayette, IN, United States

Kevin La
INSERM-Université Paris-Cité, Paris, France

Michael D. Lewis
Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom

Yu-Tsueng Liu
Division of Infectious Diseases, Department of Medicine, University of California San Diego, La Jolla, CA, United States

Lili Ranaise Mbakop
Laboratory of Animal Biology and Physiology, Faculty of Sciences, University of Yaoundé I, Yaoundé, Cameroon

JoAnn McClure
Centre for Antimicrobial Resistance, Alberta Health Services/Alberta Precision Laboratories /University of Calgary, Calgary, AB, Canada

Rafael A. Medina
Department of Pathology and Experimental Medicine, School of Medicine, Emory University, Atlanta, GA, United States; Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, United States; Departamento de Enfermedades Infecciosas e Inmunología Pediátrica, Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile

Michael A. Miles
Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom

Dorothée Missé
MIVEGEC (Maladies Infectieuses et Vecteurs: Écologie, Génétique, Évolution et Contrôle), UMR 5290 CNRS/IRD/UM, Montpellier, France

Igor Mokrousov
Laboratory of Molecular Epidemiology and Evolutionary Genetics, St. Petersburg Pasteur Institute, St. Petersburg, Russia

Serge Morand
IRL HealthDEEP (Health, Disease Ecology, Environment and Policy) CNRS — Kasetsart University — Mahidol University, Bangkok, Thailand; Faculty of Veterinary Technology, University of Kasetsart, Bangkok, Thailand; Faculty of Tropical Medicine, University of Mahidol, Bangkok, Thailand

Rachel A. Munds
Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, United States; Krishi Inc., West Lafayette, IN, United States

Dobrin Nedelkov
Isoformix Inc., Phoenix, AZ, United States

Martha I. Nelson
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, United States

Philippe Nwane
Malaria Research Laboratory, Yaoundé Research Institute (IRY), Organization for the Coordination of Endemic Diseases' control in Central Africa (OCEAC), Yaoundé, Cameroon; Laboratory of Animal Biology and Physiology, Faculty of Sciences, University of Yaoundé I, Yaoundé, Cameroon; Centre de Recherche sur les Filarioses et autres Maladies Tropicales (CRFilMT), Yaoundé, Cameroon

F. Panzera
Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

Elena Campos Pardos
Mycobacterial Genetics Group, Department of Microbiology, Faculty of Medicine, University of Zaragoza, Zaragoza, Spain

Martine Peeters
TransVIHMI, University of Montpellier, Institut de Recherche pour le Développement (IRD), INSERM, Montpellier, France

S. Pita
Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay

Spencer D. Polley
Department of Clinical Parasitology, Hospital for Tropical Diseases, London, United Kingdom

Marcos Pérez-Losada
Computational Biology Institute, Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, George Washington University, Washington DC, United States

Xueting Qiu
Department of Epidemiology, Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, MA, United States

Guislaine Refrégier
Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, Gif-Sur-Yvette, France

D. Ashley Robinson
Department of Cell and Molecular Biology, Center for Immunology and Microbial Research, University of Mississippi Medical Center, Jackson, MS, United States

Florence Roux-Dalvai
Proteomics Platform, CHU de Québec - Université Laval Research Center, Québec City, QC, Canada; Computational Biology Laboratory, CHU de Québec - Université Laval Research Center, Québec City, QC, Canada; Canadian Proteomics and Artificial Intelligence Consortium, Guelph, ON, Canada

Muhammed Rabiu Sahal
Université Paris-Saclay and IAME, INSERM-Université Paris-Cité, Paris, France

Rafael Sanjuán
Institute for Integrative Systems Biology (I^2SysBio), Universitat de València — CSIC, Paterna, Spain

Gaetan Senelle
FEMTO-ST Institute, UMR 6174 CNRS, DISC Computer Science Department, University Franche-Comté (UFC), Besançon, France

Madikay Senghore
Department of Epidemiology, Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, MA, United States

Juan Calvet Seral
Mycobacterial Genetics Group, Department of Microbiology, Faculty of Medicine, University of Zaragoza, Zaragoza, Spain

Zhengli Shi
CAS Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China

Manuela Sironi
Scientific Institute IRCCS E.MEDEA, Bioinformatics, Bosisio Parini, Italy

Christophe Sola
Université Paris-Saclay and IAME, INSERM-Université Paris-Cité, Paris, France

Philippe Solano
Intertryp, University of Montpellier, IRD Cirad, Montpellier, France

Leonor Sánchez-Busó
Joint Research Unit "Infection and Public Health" FISABIO-University of Valencia, Institute for Integrative Systems Biology (I2SysBio, UV-CSIC), Valencia, Spain; CIBER in Epidemiology and Public Health, Valencia, Spain

Colin J. Sutherland
Department of Infection Biology, London School of Hygiene & Tropical Medicine, London, United Kingdom

Michel Tibayrenc
Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle, MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD Center, Montpellier, France

Alex van Belkum
BaseClear, Leiden, the Netherlands

Marc H.V. Van Regenmortel
Medical University of Vienna, Vienna, Austria

Mohit S. Verma
Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, United States; Birck Nanotechnology Center, Purdue University, West Lafayette, IN, United States; Krishi Inc., West Lafayette, IN, United States; Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, United States

Guy Vernet
Institut Pasteur de Bangui, Bangui, Central African Republic

E. Waleckx
Institut de Recherche pour le Développement, UMR INTERTRYP IRD, CIRAD, Université de Montpellier, Montpellier, France; Laboratorio de Parasitología, Centro de Investigaciones Regionales "Dr Hideyo Noguchi", Universidad Autónoma de Yucatán, Mérida, Mexico

Lin-Fa Wang
Program in Emerging Infectious Disease, Duke-NUS Medical School, Singapore, Singapore

Jiangshan Wang
Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, United States; Birck Nanotechnology Center, Purdue University, West Lafayette, IN, United States

Michelle Wille
Centre for Pathogen Genomics, Department of Microbiology and Immunology, The University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia; WHO Collaborating Centre for Reference and Research on Influenza, at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia

Matthew Yeo
Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom

Kunyan Zhang
Centre for Antimicrobial Resistance, Alberta Health Services/Alberta Precision Laboratories /University of Calgary, Calgary, AB, Canada; University of Calgary, Pathology and Laboratory Medicine, Calgary, AB, Canada

Sarah Zohdy
College of Forestry, Wildlife, and the Environment, Auburn University, Auburn, AL, United States; College of Veterinary Medicine, Auburn University, Auburn, AL, United States

PART

Methodological/ Generalist Chapters

1

This page intentionally left blank

CHAPTER

1

A New Official Definition of Virus Species Has Led to a Controversial Linnaean Latinized Binomial Format of Virus Classification

Marc H.V. Van Regenmortel[1] **and Michel Tibayrenc**[2]

[1]*Medical University of Vienna, Vienna, Austria;* [2]*Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle, MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD Center, Montpellier, France*

1. The Nature of Virus Classification

The need for a reliable virus nomenclature and classification led the Virology Division of the International Union of Microbiological Societies in 1966 to establish the International Committee on Nomenclature of Viruses (ICNV). In 1975, its name was changed to the International Committee on Taxonomy of Viruses (ICTV) because this Committee had become responsible for regulating both the creation and the naming of virus taxa (Adams et al., 2017). The ICTV developed an independent taxonomy that differed considerably from the codes of nomenclature used in botany, zoology, and microbiology. Extensive ICTV Reports were regularly published every 4–5 years (Fauquet et al., 2005; Fenner, 1976; Francki, 1991; King et al., 2012; Matthews, 1979, 1982; Murphy et al., 1995; Van Regenmortel et al., 2000; Wildy, 1971) which described thousands of viral taxa using the hierarchical classes of species, genera, families, and orders used in biological classification. Although the first ICTV Reports advocated a Latinized viral nomenclature for virus species, the use of Latin was never implemented and in the fifth, sixth, and seventh Reports, the use of Latin was increasingly abolished altogether. In his book on the History of Viral Taxonomy, Matthews (1983) described the continued opposition to the use of Latin in virus taxonomy which was actually responsible for the late adoption of the species concept in virology only in 1991 because many virologists were convinced that the introduction of species would lead to the use of Latin species names. Only when the use of Latin in viral taxonomy was completely abolished (Pringle, 1991) were species names accepted in virology (Van Regenmortel, 1990; Van Regenmortel et al., 1991).

All biological classifications deal with abstract classes that are conceptual constructions of the mind. A virus classification is built up with conceptual classes of viruses called taxa, and these classes have members that are the concrete viral objects studied by virologists. Although viruses are usually considered not to be alive (Van Regenmortel, 2010, 2016a), they are viewed as biological entities and are therefore classified using the hierarchical categories universally used in biology. In the case of genera and families, virologists readily accept that these categories are indeed conceptual

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00017-3
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

constructions of the mind and not physical objects that they could purify, sequence, or visualize with an electron microscope. On the other hand, virologists often believe that species are more "real" than genera and families because they think of a species as being an individual type of virus capable of infecting a particular host. The resulting confusion between a species as a conceptual category in a classification and as a concrete kind of object is actually very common in the whole of biology because the vast majority of living organisms do not have common names that differ from the Latin name of the species to which the organism belongs (Ereshefsky, 2009). As a result, it is commonly said that a person can be infected by the bacterial species *Escherichia coli* as if a conceptual category could infect a physical organism. It is worth mentioning that virologists are the only biologists who can easily avoid such a logical confusion since every pathogenic virus does have a common vernacular name that differs from the species name to which the virus belongs and that name should therefore be used to refer to the infectious agent.

The root of the word classification is "class" and all biological classifications follow the logic of the hierarchical Linnaean format of classes from the lowest species class to the higher classes of genera, families, and orders (Buck and Hull, 1966). The viruses that are members of a species are also members of the taxa above it, and the relation between a lower taxon and higher ones is called class inclusion. Class inclusion obviates the need to repeat the properties used for defining higher taxa in the description of the lower taxa that are included in them. The properties of a viral genus are thus always present in the members of the species that are included in the genus, although these properties should not be confused with the species-defining properties that allow individual species to be differentiated. Because of class inclusion, higher taxa such as genera and families have more members than species, which means that they require fewer properties (for instance, virion morphology or the type of genome replication) to meet the qualification for membership in higher taxa. The logical principle (Buck and Hull, 1966) which requires that it is necessary to increase the number of qualifications for defining a species class does actually invalidate the common belief among virologists that it is possible to define a polythetic species by the presence of a single genetic metric, for instance, a single short nucleotide sequence in the viral genome (Gibbs and Gibbs, 2006; Adams et al., 2013; Simmonds, 2015). Since a virus species always has fewer members than genera and families, it is actually imperative to use several different properties of viruses for demarcating a new species (Van Regenmortel, 1990, 2018).

As is the case for all concepts, the concept of virus species cannot be "described" and can only be "defined" by listing a number of species—defining properties of its members that are not all necessarily present in every member of the species. Class membership is the logical relation that allows a link to be established between physical objects and conceptual constructs (Quine, 1990).

Since an infectious viral agent is a subcellular genetic parasite (Lwoff, 1957) that can be incorporated into the genome of an infected living cell, it should be obvious that a virus does not replicate itself since it is the living host cell that does replicate the virus. Viruses should therefore not be considered to be subcellular living organisms possessing a two-part heterogeneous so-called "life" cycle that would consist of both nonliving virions and of living virus−infected host cells as was argued by Forterre (2013) and Gibbs (2020). Viruses do not possess the functional autonomy of living cells that would enable them to actively reproduce themselves since they are actually being passively reproduced by the cells they have infected (Moreira and Lopez-Garcia, 2009). A virus present in a host cell is no more alive than other cellular constituents such as genes or macromolecules and it makes little sense to claim that a nonliving virion can be transformed by spontaneous generation into a new type of living cell metaphorically labeled a virocell (Forterre, 2013; Moreira and Lopez-Garcia, 2009).

2. The Late Acceptance of the Species Concept in Virus Classification

Species in biology are commonly considered to be populations of organisms that can breed among themselves and are reproductively isolated from other populations (Mayr, 1970). This led virologists to accept that the concept of species could not be applied to asexual viruses that replicate as clones (Milne, 1984). Many virologists were also opposed to using the species concept because they thought that it would lead to the use of Latin species names which they strongly opposed (Matthews, 1983; Van Regenmortel, 2018).

The International Committee on Taxonomy of Viruses (ICTV) is the body officially empowered to make decisions on matters of virus classification and nomenclature. Although the first ICTV Reports (Wildy, 1971; Fenner, 1976) had advocated Latinized virus species names, these names were never officially introduced. The ICTV also remained opposed to the introduction of virus species in viral taxonomy because there was no agreement on the definition of a virus species.

An earlier definition had stated that "Virus species are strains whose properties are so similar that there seems little value in giving them separate names" (Gibbs and Harrison, 1976). However, strains were not defined and the species definition implied that attributing names to virus species was equivalent to developing a viral taxonomy. The concept of species in biology had also remained controversial as illustrated by the fact that 22 different species definitions were discussed at an international conference on species (Mayden, 1997).

In 1989, the following definition of virus species was proposed "A virus species is a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche" (Van Regenmortel, 1990). This definition which indicated that the members of a virus species were not simply objects devoid of a common origin but were collections of viral objects related by common descent, was approved by the ICTV in 1991 (Pringle, 1991). The term "polythetic" is commonly used in taxonomy (Beckner, 1959), and it refers to a class whose members possess a variable combination of species properties, none of which is necessarily present in every member of the species. It must be stressed that the term polythetic refers to a particular distribution of properties characteristic of a conceptual class and that the viral members of this class do not themselves possess polythetic properties. Unfortunately, this was misunderstood by many virologists who thought that polythetic meant "variable" which led them to believe that the individual viral members of a species class did possess polythetic properties, and Gibbs and Gibbs (2006), for instance, suggested that the term polythetic should be removed from the species definition because they argued that a monothetic species could constitute a distinct lineage only if all its members shared a common property inherited from its ancestors. Since this is incompatible with the definition of a polythetic species class whose members do not share a single common defining property, the term polythetic was removed from the initial 1991 definition of virus species. Gibbs (2003) had previously argued that virus species had never been defined in practice by using the "philosophically elegant polythetic principle" and that they should rather be recognized pragmatically as strains!

Although monothetic classes are defined by one or a few properties that are both necessary and sufficient for membership in the class, the members of a polythetic class share a variable combination of properties although they do not possess a property that must be present in each member and absent in the members of other classes. The properties of monothetic and polythetic classes are illustrated in Fig. 1.1.

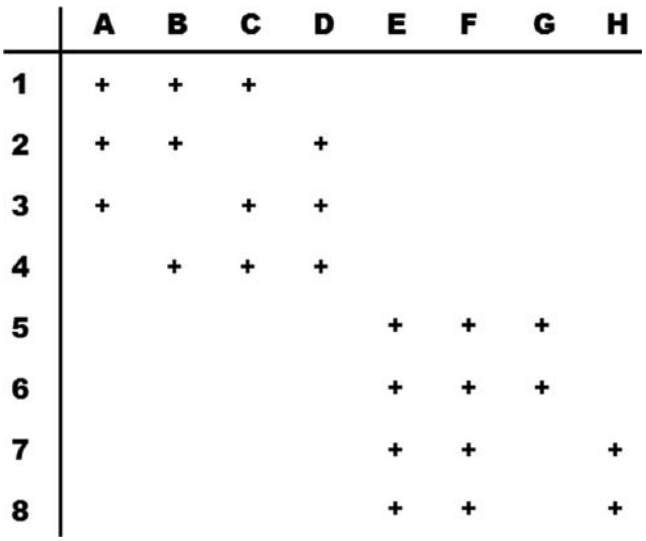


FIGURE 1.1

Distinction between polythetic and monothetic classes in the case of 8 individuals (1–8) and 8 properties (A–H). The possession of a property is indicated by a plus sign. Individuals 1–4 constitute a polythetic class, each member possessing 3 out of 4 properties, while in addition no common property is present in all the members of a polythetic class. Individuals 5–6 and 7–8 form two monothetic classes with 3 properties present in all the members (Van Rijsbergen, 1979) (see also: http://www.iva.dk/bh/lifeboat_ko/CONCEPTS/monothetic.htm). It should be pointed out that Gibbs and Gibbs (2006) did not clearly distinguish monothetic from polythetic classes since they defined both classes simply as sharing three out of four common properties.

3. Properties Used for Demarcating New Virus Species Are Not the Same as the Diagnostic Properties Used for Identifying the Members of an Established Virus Species

It is important to differentiate between properties useful for defining a virus species and properties used for identifying individual viruses (Van Regenmortel, 2016b).

Properties are possessed by things and objects and cannot be detached from them, a property like a color not being a physical part of an object (Quine, 1990). Properties like the diameter of a virion or the number of nucleotides in a viral genome are also properties of a virus although these numerical values are also not a physical part of the virus. Relational properties of viruses, also called emergent properties, are possessed by virtue of a relation between the virus and other objects in the environment such as a vector, a host, or an ecological niche which emerge only during viral transmission and infection.

A property is also different from a predicate or attribute which is the conceptual, mental representation of the property. This distinction is important because not all predicates represent properties of real things. Although tapeworms do not think, it makes no sense to say that tapeworms possess the

property of "not-thinking" because there are no such things as negative properties (Mahner and Bunge, 1997, p. 10).

The term character is often used in the sense of both property and part, although a part of a thing is a thing and not a property (Fristrup, 1992). The notion of character has been called the central mystery of taxonomy (Inglis, 1991) since the possession of a certain part may be viewed as a property although it is unclear whether a complete genome sequence, a particular nucleotide motif or the presence of a certain nucleotide somewhere in a viral genome should be considered as a single character.

Species taxa are defined intentionally by what is called the intention of the taxon which refers to the properties that provide the qualification for membership in the species. The so-called extension of a species class, on the other hand, refers to the set of all the members that are the concrete referents of the class. Since the intention of a class determines its extension, the extension of a class can only be determined if it is possible to distinguish members from nonmembers, which means that the intention must precede the extension (Mahner and Bunge, 1997, p. 227). A species taxon must therefore be established by taxonomists before it becomes possible to ascertain if a sufficient number of species-defining properties are present in an individual virus to make it a member of the species. Since monothetic species classes are defined by one or very few properties that are both necessary and sufficient for membership in the class, the claim by Gibbs and Gibbs (2006) that it is possible to rely on the presence of a single nucleotide motif for demarcating a new monothetic species is unrealistic because it would be necessary to know beforehand that this motif is present in all the members of the species and absent in other species; this implies that the extension would need to precede the intension which is impossible.

The tradition of bionominalism in taxonomy views species as individuals and historical entities (Hull, 1976) that form cohesive holes and it accepts that a species lineage is a concrete object that actually is a case of logical reification (i.e., viewing an abstract concept as if it were an object). However, the relational concepts of ancestry and lineage are not real objects and they cannot act upon each other unless they exist at the same time (Mahner and Bunge, 1997, p. 238). Species also cannot descend from each other in a literal sense since only concrete organisms and viruses can do this.

A species must thus first be established and defined by a taxonomist before it becomes possible to allocate a virus to that species by using so-called diagnostic properties. Virologists can obtain such specific diagnostic tools, for example, by producing polyclonal or monoclonal antibodies against viruses that are members of the species. Such antibodies, however, are obviously not species—defining properties that could have been used for demarcating and defining the species taxon initially.

The term species has two meanings since it can refer (1) to the many individual species classes created by virologists that have viruses as their members, or (2) to the lowest category in a virus classification which is the class of all the virus species taxa that virologists have demarcated. The 1989 polythetic species definition actually referred to the species category which was of little use to virologists when they attempted to allocate viruses to a new species taxon. The members of a polythetic class always share a number of so-called relational phenotypic properties of the virus that arise by virtue of relations between the virus and its hosts and vectors, which become actualized only during transmission and infectious processes. These species-defining properties are easily altered by a few mutations which can modify the host range, the cell and tissue tropism, the pathogenicity, as well as small parts of the genome of the virus. As a result, these properties may not be identical in every member of the polythetic class and taxonomists may have to create species by drawing boundaries across a continuum of phenotypic and genetic variability. This biological variation may sometimes

lead to an absence of clear-cut discontinuities between closely related species which could then be considered as fuzzy classes with blurred boundaries (Van Regenmortel, 2016b). This of course would not justify abandoning the species concept since the continuous nature of rainbow colors still allows different colors to be distinguished.

4. Popular Anglicized Non-Latinized Virus Species Names Were Subsequently Abandoned

Anglicized non-Latinized binomial names (NLBNs) for species were introduced by Fenner (1976) by replacing the terminal word "virus" that occurs in all common English virus names with the name of the genus to which the virus belongs which also ends in virus. Measles virus, for instance, became a member of the italicized species *Measles morbillivirus* and thousands of such names became very popular because genus names and the English names of viruses are well-known to all virologists. This is due to the fact that the major journals and reference books in virology as well as all the ICTV Reports are written in English, which is undeniably the predominant communication language used by scientists and virologists. In 2016, the ICTV initiated a so-called thought exercise in which they converted currently existing 175 NLBNs into an inverse Latinized Linnaean binomial format (LLBNs) that consists of the genus name followed by a Latinized epithet (Postler et al., 2017). Adelaide River virus, for instance, became the NLBN *Adelaide River Ephemerovirus*, while the LLBN was *Ephemerovirus fiumeadelaidense*. The nonofficial NLBN was easily recognizable by virologists and by quarantine officials or medical/veterinary professionals, whereas thousands of unpronounceable LLBNs that would need to be coined and remembered (Siddell et al., 2020) are less likely to become popular. For more than 45 years (Fenner, 1976; Francki, 1991), virologists had always used their own taxonomic rules and code of nomenclature that differ from the Linnaean format used with living organisms but unfortunately the benefits of changing the well-known NLBN names of thousands of virus species were never clearly explained. Replacing the NLBN names and imposing Latinized epithets that mimic the Linnaean format used for organisms was supposed by some to facilitate the discussion between virologists in case they would use their computers for simultaneously comparing Latinized binomial names of organisms and of virus species that are based on different binomial formats (Van Regenmortel, 2020). However, LLBNs are more likely to create unnecessary confusion in the minds of virologists who may wonder what is the species that is being referred to when newly created Latin epithets coined for LLBNs are used. Reintroducing Latin terminology to a scientific community that is quite familiar with known English scientific virus names may be difficult to justify since it seems to be an unnecessary hurdle with no obvious advantage.

5. The ICTV Introduced Names of Virus Species that Correspond to the Italicized Version of the Virus Name, Thereby Creating Unnecessary Confusion

In 1998, the ICTV EC was asked to approve two proposals submitted by committee members. The first proposal was to accept the popular Anglicized NLBNs of virus species as the official terminology for species. This could have been immediately implemented for 90% of the species currently in use but the proposal was rejected (Van Regenmortel, 2019).

The second proposal was to introduce as official species names an Italicized version of the corresponding virus name, i.e., the measles virus name would give rise to the species name *Measles virus*. Since distinguishing between viruses and species classes had always been somewhat problematic for virologists, this new terminology which abolished the characteristic binomial format used for taxonomic species names made it easier to confuse the two names that only differed by typography. After a few years, the ICTV once more allowed binomial species names (*Measles morbillivirus*) to be used if they had been approved by the relevant ICTV Study Group. However, another attempt in 2010 to approve such NLBNs as official species names (Van Regenmortel et al., 2010) was again rejected by the ICTV in spite of the fact that when virologists were asked for their opinion regarding the use of NLBNs, a significant majority (80%−85%) of the 250 virologists who voted had been in favor of NLBNs for species (Van Regenmortel et al., 2013).

6. The New ICTV Definition of Virus Species Is at Odds with the Definition of Species in Other Biological Classifications

The ICTV EC ratified in March 2021 (Walker et al., 2021) numerous changes to the International Code of Virus Classification and Nomenclature (ICVCN). A notable change was a new rule for naming virus species which followed the binomial Linnaean tradition of putting the genus name first followed by the species epithet which is the pattern used in all biological classifications, with the only difference that either a Latinized or a "free form" could be followed. The ICTV Study Groups were given the task of converting thousands of existing species names into the new Linnaean binomial format, which is expected to be completed in 2023. Unfortunately, the ICTV presented few arguments to justify the removal of thousands of Anglicized NLBN names of species based on well-known virus names and genus names. Following the Linnaean format used for living organisms and applying it to nonliving viruses seems an arbitrary decision (see Section 6) and the new official definition of a virus species as a group of physical viral isolates is also not in line with the definition of all biological species of organisms as conceptual classes. The justification for using the Linnaean Latinized format which had always been rejected by virologists is indeed strange since viral nomenclature for 45 years had always followed a different typography for taxa (italics with a capital initial) compared to that used by the Biological Code of Nomenclature.

7. Adrian Gibbs Questions the Merits of a New ICTV Latinized Binomial Nomenclature for Virus Species Associated with Metagenomic Sequences

Adrian Gibbs has for years been a regular critic of ICTV proposals and decisions (Gibbs, 2003, 2013; Gibbs and Gibbs, 2006) and in his recent review (Gibbs, 2020) he analyzed two recent controversial proposals that the ICTV had presented for discussion (Gibbs, 2013; Siddell et al., 2020; Simmonds et al., 2017; Van Regenmortel, 2003, 2019, 2020; Van Regenmortel et al., 2010; Walker et al., 2021). With the rapid development of high-throughput sequencing methods for viral genomes, large numbers of virus-like gene sequences called metagenomes had been obtained from a variety of living materials. The ICTV Executive Committee reacted to this avalanche of sequences by organizing a workshop attended by expert viral taxonomists who quickly produced a so-called consensus statement that

accepted that these virus-like metagenomes did correspond to viral genomes that should be incorporated in the existing ICTV taxonomy, in spite of the absence of any known biological properties of what were nevertheless referred to as being sequence-viruses (Hull and Rima, 2020; Simmonds, 2018). Since the hosts and vectors linked to these sequences had not been identified, these so-called sequence viruses are indeed only sequences. Gibbs (2020) reminded virologists that the ICTV taxonomy should be a taxonomy of viruses, not virions, not gene sequences but viruses. Gibbs also endorsed the view that viruses are subcellular organisms with a two-part life cycle (i.e., nonliving virions and living virus−infected host cells called virocells), a terminology which had been proposed earlier by Forterre (2013), in spite of the fact that the majority of virologists still consider viruses to be nonliving subcellular genetic parasites (Lwoff, 1957) devoid of any metabolic activity (Van Regenmortel, 2010). Gibbs also accepted the new 2013 definition of virus species, as a group of real material viral objects called isolates although he conceded that this definition was not helpful for demarcating individual virus species. He also approved the species definition proposed by Gibbs and Harrison (1976) according to which a species is an isolate or a group of virus isolates that are material objects (instead of classes) and which experts, using multiple criteria, considered to be so distinct that they could be given a single species name. Once more, taxonomy was confused with nomenclature. He also accepted that a portion of the genomic sequence of an isolate provided sufficient evidence for identifying a genomic type as a characteristic of an individual material species. Gibbs remains convinced that species and other taxonomic classes used in all biological classifications are not constructs of the human mind and he refuses to admit that conceptual taxonomic classes may have tangible, material viral objects or organisms as their members (Buck and Hull, 1966). He approved the Latinized Linnaean binomial format for virus species introduced by the ICTV which discarded the 45-year old tradition of Anglicized species names and he did endorse the new 2013 ICTV species definition that was based on the removal of the term polythetic which he had proposed in 2006 (Gibbs and Gibbs, 2006, fig. 1.1). In 2013, the ICTV accepted the removal of the term polythetic from the new definition of species which then made it possible for all the members of a virus species to possess a common nucleotide motif that was both necessary and sufficient for demarcating a novel species class.

The ICTV had ratified the following species definition: A species is a monophyletic group of viruses whose properties can be distinguished from those of other species by multiple criteria (Adams et al., 2013). This new definition made it possible to demarcate a new virus species only on the basis of the presence of a single common nucleotide motif considered to be necessary and sufficient for creating a new species. Such a definition was also incompatible with the logic of classes based on class membership and class inclusion present in all hierarchical biological classifications and has given rise to continued heated debate (Ball, 2005; Kuhn and Jahrling, 2010; Van Regenmortel, 2016a,b, 2020; Van Regenmortel et al., 2013).

This made it possible, for instance, to create more than 300 new species of begomovirus species on the basis of an arbitrarily chosen percentage (89%) of pairwise sequence identity in the viral DNA-A sequences of these viruses (Brown et al., 2015). All these so-called species infected the same hosts, produced similar disease symptoms, and possessed no specific properties that would have justified the label of species instead of strain. It is also well-known that recombination and reassortment in the genomes of geminiviruses give rise to chimeric viruses with polyphyletic genomes that cannot be correctly represented by single pairwise sequence comparisons.

Gibbs (2020) did not accept that it was possible to incorporate metagenomic sequences, sometimes called sequence-viruses (Hull and Rima, 2020) in the list of official ICTV species since there was no

information regarding the biological properties of the hypothetical viruses presumed to be hiding within metagenomic sequences. However, members of the ICTV and of a recently created metagenomics think-thank had agreed that it should be feasible to identify most of the species that were hiding within metagenomes (Simmonds, 2015, 2018).

Since viruses are genetic parasites that do not self-multiply but are reproduced by the living cells of the host that the virus had infected, it is of course not possible to create additional new virus species in the existing ICTV classification based on the viral genome since the phenotype and its relational properties derived from hosts and vectors remain unknown, when only the viral genome is known. When the viral host remains unknown epigenetic factors together with alternative splicing and the discarding of introns lead to unpredictable RNA transcripts that interact with the host but cannot reveal the causal connections between viral genes and the development of viral phenotypes (Van Regenmortel, 2016a,b). A further limitation is that single reference genomes are not adequate for representing the total amount of genetic variation in a species since the pangenome of a virus species may contain as much as 10% more genes that are present in an individual species (Bao et al., 2010; Brito et al., 2015; Wikipedia contributors, 2022).

It should be stressed that the variable distribution of properties in the members of a polythetic species class as well as the absence of a single common property in all its members are both characteristics of the class rather than a property of all its members. Every membership condition determines a class although not being a member is not a defining property since there are no such things as negative properties. Membership of a polythetic class can thus be determined by using some of many species membership conditions except one which is nonself-membership (Quine and Ullian, 1978). The absence of a single common property in all the members of a polythetic class does not make it possible to recognize a polythetic species class, since one would land with the well-known Russell paradox of the barber (Baldwin and Lessmann, 1998). This paradox arises if one assumes that when a village barber shaves all and only those men in the village who do not shave themselves, this implies that the barber will need to shave himself only if he does not do so! This paradox is resolved only when it is realized that there is no such barber (Quine, 1990). It is equally impossible to identify a polythetic virus species on the basis of a single unknown property that is absent in all its members!

Gibbs (2020) also recommended that the ballot system used by the ICTV for approving its new rules should not involve the participation of too many national "honorary" members who often have a limited knowledge of viral taxonomy, since this could lead to the generalized acceptance of every new proposal formulated by the ICTV EC, which in the past has sometimes been counterproductive.

In his amusing discussion of polythetic classes, Calisher (2018) clarified a number of philosophical problems that viral taxonomists have to face when they classify virus species. Defining what is an individual can be as problematic as defining a species (Hull, 1976), and it would be equally problematic to give a list of all the properties that must necessarily be absent in every member of a viral polythetic species in order to satisfy the definition of a polythetic class as a class that must not share a single common property in all its members (Van Regenmortel et al., 1991). Philosophy of science is rarely taught in virology curricula and a light-hearted approach to the subject as done by Calisher is probably more helpful than a more rigorous academic introduction to the philosophy of science itself. Interested readers could consult two books that provide a good introduction to logic and philosophy of science (Quine, 1990; Mahner and Bunge, 1997).

8. Ontology and Epistemology of Viruses and Living Organisms

The word ontology is derived from the Greek words "*ontos*" (which means being) and "*logos*" (which means study) and it refers to the branch of philosophy that analyzes the nature of what exists in the world in the form of substances, objects, living organisms, and abstract entities. The term epistemology is derived from the Greek words "*episteme*" (which means knowledge) and "*logos*," and it refers to the branch of philosophy that analyses how humans acquire knowledge and determine its extent and validity for understanding the reality of the world we live in.

When viruses were discovered about a century ago, it was accepted that they were minute microbial organisms too small to be seen with optical microscopes but able to infect animals, plants, and microbes and to kill the hosts they had infected. Subsequently, it was established that viruses were actually genetic parasites that invaded the genomes of their hosts and that they did not reproduce themselves but were reproduced by the living cells of their hosts (Lwoff, 1957). Since a virus present in a host cell is no more alive than other cellular constituents such as genes and macromolecules, it makes little sense to claim that a nonliving virion can be transformed by spontaneous generation into a new type of living cell metaphorically called a virocell (Gibbs, 2020). The ontology of living cells is easily differentiated from that of dead organisms which lack the intrinsic characteristics of life such as metabolism and the functional autonomy to grow, reproduce, and repair themselves (Van Regenmortel, 2010). Descartes' statement: *Cogito, ergo sum* (I think, therefore I am) did provide the irrefutable truth that humans are alive because they are able to think.

Epistemology views knowledge as a mental state existing in human minds acquired by objective personal experience and by sound reasoning, but it also analyzes the methods used for acquiring human knowledge as well as for analyzing its scope and limits.

In recent years, a new ontology and epistemology of living organisms has been developed which argues that the living world actually consists of a hierarchy of processes rather than of things and substances as previously thought. Such a processual view of organisms is often compared to the ceaseless current and flow of a river, which gave rise to the *stream of life conception.* Although organisms sometimes may give the impression of being static, processes lead to perpetual changes that extend in time, because living organisms undergo continuous self-generating processes of development, metabolism, and evolution (Dupré, 2012; Nicholson and Dupré, 2018; Rescher, 1996).

References

Adams, M.J., et al., 2013. Recently agreed changes to the international code of virus classification and nomenclature. Arch. Virol. 158, 2633e9.

Adams, M.J., et al., 2017. 50 years of the international committee on taxonomy of viruses: progress and prospects. Arch. Virol. 162, 1441−1446. https://doi.org/10.1007/s00705-016-3215-y.

Baldwin, J.T., Lessmann, O., August 17, 1998. What is Russell's paradox? Sci. Am. http://www.scientificamerican.com/article/what-is-russells-paradox/.

Ball, L.A., 2005. The universal taxonomy of viruses in theory and practice. In: Fauquet, C.M., Mayo, M.A., Maniloff, J., Desselberger, U., Ball, L.A. (Eds.), Virus Taxonomy. Eight Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Amsterdam, pp. 3−8. https://doi.org/10.1016/b978-0-12-249951-7.50003-1.

Bao, Y., Kapustin, Y., Tatusova, T., 2010. Virus classification by pairwise sequence comparison (PASC) in Van Regenmortel MHV. In: Mahy, B. (Ed.), Desk Encyclopedia of General Virology. Academic Press, Elsevier, Oxford, pp. 95−100.
Beckner, M., 1959. The Biological Way of Thought. Columbia University Press, New York.
Brito, A.F., et al., 2015. The Pangenome of the Anticarsia gemmatalis multiple nucleopolyhedrovirus (AgMNPV). Genome Biol. Evol. 8, 94.
Brown, J.K., et al., 2015. Revision of Begomovirus taxonomy based on pairwise sequence comparisons. Arch. Virol. 160, 1593−1619.
Buck, R.C., Hull, D.L., 1966. The logical structure of the Linnaean hierarchy. Syst. Zool. 15, 97e111.
Calisher, C.H., 2018. Viruses do not have polythetic properties; species are polythetic classes and do not have any properties. Arch. Virol. 163, 2033−2036. https://doi.org/10.1007/s00705-018-3937-0.
Dupré, J., 2012. Processes of Life: Essays in the Philosophy of Biology. Oxford University Press, Oxford.
Ereshefsky, M., 2009. Darwin's solution to the species problem. Synthese 175, 405−425. https://doi.org/10.1007/s11229-009-9538-4.
Fauquet, C., et al., 2005. Virus Taxonomy. Eighth Report of the ICTV. Elsevier, Academic Press, London, San Diego.
Fenner, F., 1976. Classification and nomenclature of viruses. Second Report of the ICTV. Intervirology 7, 1−115.
Forterre, P., 2013. The virocell concept and environmental microbiology. ISME J. 7, 233−236.
Francki, R.I.B., et al., 1991. Classification and nomenclature of viruses fifth report of the ICTV. Arch. Virol. Suppl. 2 (Springer, Vienna, New York).
Fristrup, K., 1992. Character: current usages. In: Keller, E.F., Lloyd, E.A. (Eds.), Keywords in Evolutionary Biology. Harvard University Press, Cambridge, MA, pp. 45−51.
Gibbs, A.J., 2003. Virus nomenclature: where next. Arch. Virol. 148, 1645e53.
Gibbs, A.J., 2013. Viral taxonomy needs a spring clean; its exploration era is over. Virol. J. 10, 254. https://doi.org/10.1186/1743-422X-10-254.
Gibbs, A.J., 2020. Binomial nomenclature for virus species: a long view. Arch. Virol. 165, 3079−3083. https://doi.org/10.1007/s00705-020-04828-8.
Gibbs, A.J., Gibbs, M.J., 2006. A broader definition of the 'virus species'. Arch. Virol. 151, 1419e22.
Gibbs, A.J., Harrison, B., 1976. Plant Virology. The Principles. Edward Arnold, London.
Hull, D.L., 1976. Are species really individuals? Syst. Zool. 25, 174e91.
Hull, D.L., Rima, B., 2020. Virus taxonomy and classification: naming of virus species. Arch. Virol. 165, 2733−2736. https://doi.org/10.1007/s00705-020-04748-7.
Inglis, W.G., 1991. Characters: the central mystery of taxonomy and systematics. Biol. J. Linn. Soc. 44, 121−139.
King, A.M.K., et al., 2012. Virus taxonomy. In: Ninth Report of the ICTV. Elsevier, Academic Press, London, San Diego.
Kuhn, J.H., Jahrling, P.B., 2010. Clarification and guidance on the proper usage of virus and virus species names. Arch. Virol. 155, 445−453. https://doi.org/10.1007/s00705-010-0600-9.
Lwoff, A., 1957. The concept of virus. J. Gen. Microbiol. 2, 239−253. https://doi.org/10.1099/00221287-17-2-239.
Mahner, M., Bunge, M., 1997. Foundations of Biophilosophy. Springer-Verlag, Berlin.
Matthews, R.E.F., 1979. Classification and nomenclature of viruses. Third report of the ICTV. Intervirology 12, 129−296.
Matthews, R.E.F., 1982. Classification and nomenclature of viruses. Fourth report of the ICTV. Intervirology 17, 1−199.
Matthews, R.E.F., 1983. The history of viral taxonomy. In: Matthews, R.E.F. (Ed.), A Critical Appraisal of Virus Taxonomy.

Mayden, R.L., 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. In: Claridge, M.F., Dawah, H.A., Wilson, M.R. (Eds.), Species: The Units of Biodiversity. Chapman and Hall, London, pp. 381–424.

Mayr, E., 1970. Populations, Species and Evolution. Harvard University Press, Cambridge, MA.

Milne, R.G., 1984. The species problem in plant virology. Microbiol. Sci. 1, 113e22.

Moreira, D., Lopez-Garcia, P., 2009. Ten reasons to exclude viruses from the tree of life. Nat. Rev. Microbiol. 4, 306–311. https://doi.org/10.1038/nrmicro2108.

Murphy, F.A., et al., 1995. Virus Taxonomy Sixth Report of the ICTV. Springer, Vienna, New York.

Nicholson, D.J., Dupré, J., 2018. Everything Flows. Towards a Processual Philosophy of Biology. Oxford University Press, Oxford.

Postler, T.S., et al., 2017. Possibility and challenges of conversion of current virus species names to Linnaean binomials. Syst. Biol. 66, 463–473.

Pringle, C.R., 1991. The 20th Meeting of the Executive Committee of the ICTV. Virus species, higher taxa and other matters. Arch. Virol. 119, 303–304.

Quine, W.V., 1990. Classes versus properties. In: Quiddities: An Intermittently Philosophical Dictionary. Penguin Books, London, p. 22.

Quine, W.V., Ullian, J.S., 1978. The Web of Belief. McGraw-Hill, New York.

Rescher, N., 1996. Process Metaphysics: An Introduction to Process Philosophy. State University of New York.

Siddell, S.G., et al., 2020. Correction to: binomial nomenclature for virus species: a consultation. Arch. Virol. 165, 1263–1264. https://doi.org/10.1007/s00705-020-04555-0. Erratum for: Arch. Virol. 2020; 165, 519–525. PMID: 32065315; PMCID: PMC7160066.

Simmonds, P., 2015. Methods for virus classification and the challenge of incorporating metagenomic sequence data. J. Gen. Virol. 96, 1193–1206.

Simmonds, P., 2018. A clash of ideas - the varying uses of the 'species' term in virology and their utility for classifying viruses in metagenomic datasets. J. Gen. Virol. 99, 277–287.

Simmonds, P., et al., 2017. Virus taxonomy in the age of metagenomics. Nat. Rev. Microbiol. 15, 161–168.

Van Regenmortel, M.H.V., 1990. Virus species, a much overlooked but essential concept in virus classification. Intervirology 31, 241–254.

Van Regenmortel, M.H.V., 2003. Viruses are real, virus species are man-made, taxonomic constructions. Arch. Virol. 148, 2481–2488. https://doi.org/10.1007/s00705-003-0246-y.

Van Regenmortel, M.H.V., 2010. Logical puzzles and scientific controversies. The nature of species, viruses and living organisms. Syst. Appl. Microbiol. 33, 1–6. https://doi.org/10.1016/j.syapm.2009.11.001.

Van Regenmortel, M.H.V., 2016a. Classes, taxa and categories in hierarchical virus classification: a review of current debates on definitions and names of virus species. Bionomina 10, 1–21.

Van Regenmortel, M.H.V., 2016b. Only viruses, but not their genome sequences can be classified into hierarchical species and genus classes. Curr. Top. Virol. 13, 59–68.

Van Regenmortel, M.H.V., 2018. Viral species, viral genomes and HIV vaccine design: is the rational design of biological complexity a utopia? Arch. Virol. 163, 2047–2054. https://doi.org/10.1007/s00705-018-3955-y.

Van Regenmortel, M.H.V., 2019. Solving the species problem in viral taxonomy: recommendations on non-Latinized binomial species names and on abandoning attempts to assign metagenomic viral sequences to species taxa. Arch. Virol. 164, 2223–2229. https://doi.org/10.1007/s00705-019-04320-y.

Van Regenmortel, M.H.V., 2020. A brief history of the species concept in virology and an opinion on the proposal to introduce Linnaean binomial species names. Arch. Virol. 165, 3073–3077. https://doi.org/10.1007/s00705-020-0495-0.

Van Regenmortel, M.H.V., et al., 2000. Virus taxonomy. In: Seventh Report of the ICTV. Academic Press, New York.

Van Regenmortel, M.H.V., et al., 2010. A proposal to change existing virus species names to non-latinized binomials. Arch. Virol. 155, 1909e19.

Van Regenmortel, M.H.V., et al., 2013. Virus species polemics: 14 senior virologists oppose a proposed change to the ICTV definition of virus species. Arch. Virol. 158, 1115−1119. https://doi.org/10.1007/s00705-012-1583-5.

Van Regenmortel, M.H.V., Maniloff, J., Calisher, C.H., 1991. The concept of virus species. Arch. Virol. 120, 313−314.

Van Rijsbergen, K., 1979. Information Retrieval, second ed. Butterworths, London.

Walker, P.J., et al., 2021. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses. Arch. Virol. 166, 2633−2648. https://doi.org/10.1007/s00705-021-05156-1.

Wikipedia contributors, January 30, 2022. Pan-genome. In: Wikipedia, The Free Encyclopedia. Retrieved 13:53, February 17, 2022 from. https://en.wikipedia.org/w/index.php?title=Pan-genome&oldid=1068820129.

Wildy, P., 1971. Classification and nomenclature of viruses. First report of the ICTV. Monogr. Virol. 5. ISBN 3-8055-1196-5.

This page intentionally left blank

CHAPTER

A Completionist Approach to Discovering and Characterizing Bacterial Diversity

2

Frederick M. Cohan and Sarah Kopac
Department of Biology, Wesleyan University, Middletown, CT, United States

1. Introduction—Toward a Completionist Bacterial Systematics

How odd that video gamers have enriched the vocabulary and zeitgeist of science with their concept of "completionism"! A completionist is a gamer who aims "to complete every challenge and earn every achievement or trophy" in a video game (dictionary.com, 2022). In video games, *one is not required* to try for every achievement to finish a game, and most gamers do not, but there is the temptation to accomplish everything offered by a game because that is possible. With inspiration from the genomic and bioinformatic revolutions, we scientists can now aim to understand biological processes and diversity to a completionist level. For example, a pharmaceutical developer can characterize *every possible drug* within a class of compounds through in silico approaches, as prelude to wet-lab testing of the most promising compounds (Jabalia et al., 2021). When analyzing the genetic basis of an evolutionary adaptation, we can go beyond testing the individual genes suspected a priori to play a role in the adaptation; we are now driven to enumerate the entire genome's contributions to any evolutionary response (Toll-Riera et al., 2016). Aiming to prevent a future pandemic, virologists are now motivated to characterize every virus on the planet that could possibly infect and harm our species (Carroll et al., 2018).

Similarly, bacterial systematists are driven toward a completionist approach to characterizing the diversity of bacteria—to describe all the species of bacteria (Parks et al., 2020) and to explore whatever significant diversity may exist within any species we describe (Cohan, 2021). Scientists in many disciplines rely on a reasonably complete accounting of the world's bacterial species—epidemiologists, biotechnologists, agriculturists, microbial ecologists, evolutionary biologists, and of course, systematists. But what is a reasonably complete accounting of diversity? We are now finding that rare bacterial species play substantial roles in ecosystems (Lynch and Neufeld, 2015; Wang et al., 2022). Moreover, ecologically distinct variants within pathogenic bacterial species can play dramatically different roles in public health (Proctor et al., 2006). It is therefore increasingly important to adopt a completionist accounting of all species and ecologically distinct variants within species, as well as their ecological functions. Unfortunately, bacterial systematics is far behind in describing the trillions of species we suspect exist (Garrity, 2016; Locey and Lennon, 2016; Yarza et al., 2014), but genomic approaches allow us to conceive the tasks of discovering, characterizing, and naming every species of bacteria, whether or not the species are cultivable.

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00009-4
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Genomes have granted us an exciting time for the discovery of bacterial diversity. However, we cannot reach a completionist understanding of bacterial diversity through genomic technology alone. Sequencing the genome of every organism on the planet would not organize the diversity into meaningful and useful groups for us. We first need a rubric for demarcating species so that we can recognize novel species when we find them. In a theory-based approach, we may demarcate species as groups that hold certain species-like, dynamic properties (Bobay and Ochman, 2017; Cohan, 2021; de Queiroz, 2005). Alternatively, species demarcation may defer to theory-free traditions on the quantity of diversity that we should include within a species (Rosselló-Móra and Amann, 2001). To choose between these alternatives, systematists need to know whether there are real species in nature bearing some set of species-like dynamics, or whether species are just the invention of taxonomists.

Bacterial systematics has chosen the path of tradition rather than theory to demarcate species as closely related groups that differ in their disease-causing properties, their ecological roles in biological communities, or their physiological capacities (Rosselló-Móra and Amann, 2001, 2015). Systematists have defined species as groups (or clusters) of close relatives separated by large gaps in phenotypic and molecular characters (Rosselló-Móra and Amann, 2001; Vandamme et al., 1996). This practical approach has the cachet of approval from no less an evolutionary biologist than Charles Darwin (Darwin, 1859; Mallet, 2008b). Darwin proposed that animal and plant species should be defined as closely related groups that can coexist as phenotypically distinct clusters (Darwin, 1859; Mallet, 2008b; Stamos, 2007), and this is largely the approach taken by bacterial systematists. This cluster-based approach has proven to be remarkably robust, even as the criteria for defining bacterial species have changed over the decades, from being based on phenotype (usually metabolism) (Rosselló-Móra and Amann, 2001; Vandamme et al., 1996), to whole-genome similarity (as measured through genome hybridization) (Wayne et al., 1987), to in silico approaches measuring the overall similarity of genomes (Chun and Rainey, 2014; Jain et al., 2018; Meier-Kolthoff et al., 2021; Olm et al., 2020; Parks et al., 2020). Bacterial systematists have argued about whether the species they recognize are too narrowly or too broadly defined, and whether they are using the best criteria for demarcating species, but they have agreed that species should hold the essential property of being clusters of close relatives with gaps between them (Gevers et al., 2005; Jain et al., 2018). Our first aim here is to discuss how we can reach a completionist accounting of the clusters separated by gaps that are recognized as species by the traditions of systematics.

Nevertheless, many microbiologists and most systematists outside of microbiology have understood species to be more than closely related groups separated by gaps (de Queiroz, 2005; Mayr, 1982). They note that demarcating species as clusters with intervening gaps applies equally well to *any* taxonomic rank, from species to phyla. A pattern of clustering with gaps merely reflects the genealogical continuity among all organisms, where some lineages have disappeared and others have thrived (Cohan, 2021). Indeed, higher ranks such as genera and families are marked as phenotypic clusters with gaps between them, simply resulting from uneven successes among different lineages.

Is there something special about the rank of species? Largely under the influence of Ernst Mayr, the property of cohesion has become understood as a quintessential aspect of species (Cohan, 1994, 2017; de Queiroz, 2005; Mayr, 1963, 1982; Templeton, 1989). In this view, species are real because they are phylogenetic groups whose diversity is constrained by a force of cohesion. In Mayr's Biological Species Concept, the force constraining diversity within animal and plant species is understood to be genetic exchange. Here, speciation requires certain unusual circumstances that allow newly divergent populations to break free of cohesion by recurrent, high-frequency genetic exchange (Mayr, 1963).

Some theory-oriented bacteriologists have sought to apply the Biological Species Concept to bacteria. They have found evidence for genetic exchange among the most closely related bacteria, with

recombination rates diminishing with greater genetic distance (Bobay and Ochman, 2017; Cadillo-Quiroz et al., 2012; Dykhuizen and Green, 1991; Majewski and Cohan, 1999b). Some have argued that recombination is therefore a force of cohesion only among the homogeneous close relatives within a species (Bobay and Ochman, 2017; Cadillo-Quiroz et al., 2012), but here we will challenge that view (Melendrez et al., 2016).

Many concepts of species have been developed since Mayr's Biological Species Concept, and most have in common certain quintessential features, most related to cohesion (Cohan, 2011a; de Queiroz, 2005). These concepts hold that the diversity within a species is limited by a force of cohesion, different species are ecologically distinct and irreversibly separate, and species are invented only once. In what we might consider Mayrian concepts of species, these essential properties have been extended to other groups where genetic exchange is rare or absent, such as the bacteria (Cohan, 1994; Templeton, 1989).

With our colleagues, we have developed an alternative Mayrian concept of bacterial speciation that focuses on ecology. This is the "ecotype" theory of bacterial species and speciation, where cohesion within species-like ecotypes is based on the ecological homogeneity of a species' members, and speciation occurs through ecological divergence (Kopac et al., 2014). We have shown how the ecological homogeneity within an ecotype limits diversification within the ecotype through a process called "periodic selection" (Fig. 2.1). Then, when the members of an ecotype somehow diverge in their ecology, they break free of cohesive forces and can diverge without bound. In the ecotype theory, the splitting of one ecotype into two ecologically distinct lineages that can coexist is the fundamental, diversity-creating process of speciation in bacteria (Cohan, 1994, 2011a; Cohan and Perry, 2007; Koeppel et al., 2008; Kopac et al., 2014; Ward, 1998).

Here we will show that the recognized species taxa of bacterial systematics generally contain numerous species-like, infraspecific ecotypes, each with the species-like property of being ecologically distinct and able to coexist with the other ecotypes. We will show how genomics can potentially enrich bacterial systematics through completionist discovery of all the species-like ecotypes within each species taxon.

Finally, we note a similarity in outcome between the theory-based systematics that enumerates diversity at the ecotype level and the tradition-based systematics that demarcates diversity at the species taxon level. Remarkably, it turns out that species demarcations based on theory-free, pre-genomic criteria may universally be demarcated by a genomic criterion of similarity, the Average Nucleotide Identity of core genomes (Jain et al., 2018; Konstantinidis et al., 2017; Olm et al., 2020), although there is some variation among species for how much ANI they each contain (Murray et al., 2021; Rodriguez-R et al., 2021). Moreover, we will show that the species taxa demarcated in this way appear to each have the property of cohesion (Cohan, 2019). We propose a rapprochement between theory-based and cluster-based approaches, acknowledging that they have each led to discovery of cohesive phylogenetic groups, the former at the ecotype level and the latter at the level of recognized species taxa.

2. Genomes Enable Completionist Systematics at the Species Level

How to meet the completionist agenda when there are so many bacterial species? The Global Prokaryotic Census has identified bacterial species taxa from sequences of 16S rRNA genes amplified from environmental DNA from disparate habitats. Assuming a cutoff of 99% identity for species members, the GPC has directly enumerated an enormous number of species (5.6 million) (Gilbert et al., 2014). Statistically accounting for the number of rare species that inevitably are missing in finite

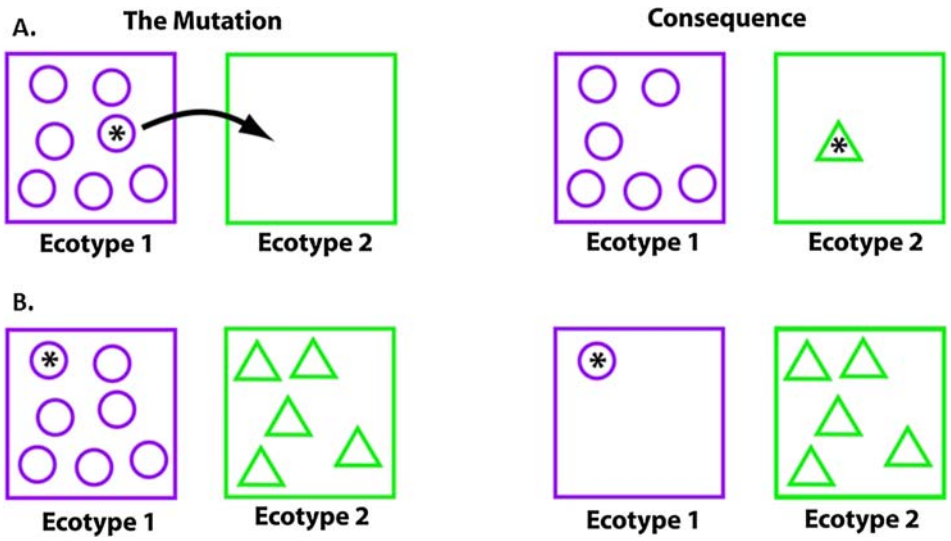


FIGURE 2.1 The dynamics of ecotype formation and periodic selection within an ecotype.

Circles represent different genotypes, and asterisks represent adaptive mutations. (A) Ecotype formation event. A mutation or a recombination event allows the cell to occupy a new ecological niche, founding a new ecotype. A new ecotype can be formed only if the founding organism has undergone a fitness trade-off, whereby it cannot compete successfully with the parental ecotype in the old niche. (B) Periodic selection event. A periodic selection mutation improves the fitness of an individual such that the mutant and its descendants outcompete all other cells within the ecotype; these mutations do not affect the diversity within other ecotypes because ecological differences between ecotypes prevent direct competition. Periodic selection leads to the distinctness of ecotypes by purging the divergence within, but not between ecotypes (Cohan, 2005).

Used with permission from Landes Publishers.

samples, Lennon and Locey have estimated the number of species at one trillion (Lennon and Locey, 2020; Locey and Lennon, 2016)! We may contrast this to the paltry 19,000 species that have been described, according to the All-Species Living Tree Project https://imedea.uib-csic.es/mmg/ltp/. Considering that only 1000 new species' descriptions are published per annum (Trujillo and Oren, 2018), one might conclude that the completionist agenda would overwhelm even the most hopeful of researchers. However, genomic and computational approaches can help in every aspect of describing and naming new species. Let us begin with the challenge of archiving a living type specimen for each species.

Using genome sequences as the type material. The International Code for Nomenclature of Prokaryotes rules that any new taxon (from species through phylum) can be named only after a living type specimen is archived in culture (Parker et al., 2019). Most directly, this rule eliminates from consideration the description and naming of any organism that cannot be cultivated. For decades, we have assumed the number of noncultivable species to be extremely high (at least 99%), although massive efforts in culturomics, which aims to culture novel species in a

high-throughput manner, has mitigated this limitation (Martiny, 2019). Nevertheless, there are phyla that remain entirely uncultivated, for example, the phyla of the Candidate Phyla Radiation (Hug et al., 2016). Even when we further improve our abilities for high-throughput culturing of unknown bacteria, many have argued that our rate of describing the enormity of bacterial diversity should not be limited by the time, effort, and expense required to place organisms in culture (Garrity, 2016; Whitman, 2016).

The urgency for ramping up the speed of species description has led to debate on the need for cultivating new species. Whitman and other systematists recently proposed to relax Rules 27 and 30 of the taxonomic code, such that a "type" genome sequence can stand in for a type strain (Whitman, 2016). Beyond speeding up species-level systematics, supporters of reform argue that this change would further democratize systematics, such that anyone who can sequence and analyze genomes would be able to demarcate novel species and characterize their metabolic features (Garrity, 2016). It is unfortunate, to our minds, that the proposal failed (Sutcliffe et al., 2020).

In response to this decision, a new approach to systematics called the SeqCode has now developed in parallel to the International Code for Nomenclature of Prokaryotes (Whitman et al., 2022). We believe that in all the most important ways, the SeqCode offers the same information content as the ICNP. As we will describe, through the SeqCode we can assure that a novel species is separate from other, established species; we can confidently place the novel species in the proper place in the bacterial phylogeny; and we can comprehensively (if tentatively) describe the phenotype of the new species.

Incorporating genomes of uncultivated organisms into systematics. One option for generating a genome from an uncultivated bacterium is by the Single Amplified Genome (SAG) approach, which involves isolating a single cell, lysing it, amplifying the whole genome, and then sequencing the genome (Alneberg et al., 2018). The SAG method has the advantage of describing one real organism, albeit uncultivated. In this sense, a species represented by a type sequence can be treated identically to one represented by a type organism.

The SeqCode also allows for a metagenome-assembled genome sequence (MAG) to stand in for a type specimen. MAGs are based on an amalgam of sequence reads from closely related organisms that are assembled into contigs on the basis of a certain level of sequence identity (e.g., 99.5%). Then the partial genomes represented by these contig assemblies are binned together according to similarities in their nucleotide composition and relative abundance, preferably across multiple environmental samples (Alneberg et al., 2018).

The MAGs so produced differ from the SAGs of single cells in that the MAG contains genes from multiple organisms, and thus potentially contains the sum of metabolic qualities of ecologically different close relatives (Cohan, 2021; Nelson et al., 2016; Sangwan et al., 2016). We should note a weakness of the MAG approach that was revealed by comparing MAGs with the genomes of isolates from *Escherichia coli*—the MAG was likely to miss a substantial fraction of both the core and variable genes in the species' genome (Meziti et al., 2021). Another downside of the MAG type sequence is that it can confound multiple ecotypes into one organism (Cohan, 2016a). On the other hand, it may be argued that the MAG type sequence is beneficial for approximating the metabolic diversity included within a species taxon (Rosselló-Mora, Pers. Comm). Either way, allowing type sequences to stand in for type specimens allows progress toward a completionist systematics, and will encourage systematists to eventually find a way to culture what is presently described by a type sequence (Whitman et al., 2022).

How to name the enormous number of species described through genomics. The term "Candidatus" has been used for a quarter century to name uncultivable taxa that have been uniquely described with sequence, morphological, and ecological data (Murray and Stackebrandt, 1995; Pallen, 2021b). The SeqCode proposes to extend Candidatus status to new species described only by their genomes and the phenotypes inferred from genomes (Whitman et al., 2022).

The prospect of providing names for the trillions of species could easily overwhelm our imagination. Traditionally, naming species has been limited by finding Latin and Greek words that describe the new taxa, or by latinizing names that honor a person or describe a geographic source. Clearly, we just do not have the imagination or time to apply this tradition to the volume of species waiting for our attention. Moreover, many systematists lack a working knowledge of Latin, as required by systematics tradition. Recently, Pallen and colleagues have addressed this challenge in their "The next million names for Archaea and Bacteria," where they have introduced the Great Automatic Nomenclator (Pallen et al., 2021). This is an engaging and playful argument for giving ourselves license to make up novel and meaningless names, foregoing any ambition to connect an organism's name to its qualities. Pallen's algorithm produces random but euphonious combinations and concatenations of Latin and Greek roots to be used as taxon names. An alternative (but more limited) means to extend the number of possible taxon names is to continue the centuries-old, Linnaean tradition of naming in accordance with a taxon's properties, but using words from other languages. Pallen gives examples of chicken-infecting microbes that might use the word for chicken from other languages, for example *Ofimonas* (from Hebrew), *Kazabacterium* (from Hausa), or *Jicola* from Chinese (Pallen, 2021a). Recently, the species *Subdoligranulum didolesgii*, causing rheumatoid arthritis, was named after the Cherokee word for arthritis to celebrate indigenous people's contributions to health science (Chriswell, 2022).

Determining whether a genome sequence should be placed in a novel species. Genomics has already fully reached the standard ICNP systematics for classifying novel organisms. Whether based on live type organisms or type sequences, novel organisms may be classified by various Overall Genome Relatedness Indices (OGRIs) (Chun and Rainey, 2014), including Average Nucleotide Identity of shared genes (ANI) (Jain et al., 2018; Olm et al., 2020) and digital DNA-DNA Hybridization (dDDH), which measures the percentage of genes that are shared (Meier-Kolthoff et al., 2021). That is, if an unclassified organism is found to have a low enough ANI ($<$95%) or dDDH score (70%) to the type strain of any described species, it can potentially be classified as a novel species (Chun et al., 2018; Parks et al., 2019, 2020). Thus, once the type genome for a novel species is chosen, other genomes would be added to the species based on having an ANI value $>$ 95% or dDDH value $>$ 70% to the type genome (Chun et al., 2018). This would capture the central importance of type strains in taxonomy.

A genomes-based classification must consider that nearly 20% of described species have not yet been genome sequenced (Whitman et al., 2022), while nearly all species have been sequenced for at least 16S rRNA. Therefore, we will need a two-step process in deciding whether a novel organism is different enough from other established species, starting with 98.7% identity in 16S rRNA as a cutoff criterion (Chun et al., 2018).

Genome-based analyses can readily place any new species on the tree of life, as a minimum of 31 universal genes yields a reliable tree (Chun et al., 2018).

Characterizing a novel species' phenotype from genomes. While the traditions of systematics have encouraged experiment-based description of a new species' phenotype, microbiologists are increasingly demanding a genome-based route to characterizing new species (Garrity and Lyons,

2011; Garrity, 2016; Konstantinidis et al., 2017; Thompson et al., 2014; Vandamme and Peeters, 2014). Like the ICNP, the SeqCode requires that any new species taxon be described for its phenotype and for diagnostic phenotypic characteristics. The ICNP now accepts phenotypic diagnoses and descriptions inferred entirely from genome content (Sutcliffe et al., 2021), so this aspect of the SeqCode is not novel (Whitman et al., 2022).

Genome content analyses can potentially yield the metabolic, chemical, and ecological properties of a novel organism. Most straightforwardly, the standard phenotypic traits used for decades can be estimated by analyzing gene content. For example, the capacity for the Voges-Proskauer reaction, indole production, and utilization of any number of carbon sources can be tentatively determined through gene content analyses (Thompson et al., 2014). Algorithms (e.g., Traitar and microTrait) have been developed to estimate a broad diversity of phenotypic characters from genome content (Karaoz and Brodie, 2022; Karlsen et al., 2023; Weimann et al., 2016).

Using genomes to characterize novel bacteria for their tolerances of physical and chemical conditions is challenging. Changes in temperature preference do not seem to involve gene content but changes in amino acid content (Dore et al., 2020). Various approaches to identifying genes responsible for complex traits such as salt and pH tolerance in laboratory studies promise future predictions of phenotype (Barberán et al., 2017; Hahne et al., 2010; Mirete et al., 2015). Genome-Wide Association Studies indirectly identify genes responsible for phenotypic differences by correlating, among close relatives, the genome content with phenotypes and habitats (Lees et al., 2020).

Future microbiologists can perform GWAS analyses on the organisms we describe, but only if microbiologists are careful to publish detailed accounts of the environments from which they isolate novel organisms. Following the MIxS protocols for describing habitats (e.g., including salinity, pH, texture, and many other features for soil habitats) will become especially important as we try to characterize ecological abilities from genome-wide association studies (Cohan, 2011b; Jorge et al., 2022; Yilmaz et al., 2011).

Lacking a cultured organism, confirming genome-based phenotypes with certainty will be difficult. One can glean that the phenotypes inferred from gene content are correct by a transcriptomic approach, where the genes of a novel species are shown to be upregulated in the appropriate environments (Pombo et al., 2019).

Genomics can best infer the ecology of a novel organism when the organism is related to known, described, and cultured bacteria. However, there are so many bacteria that have no characterized close relatives, the most extreme case being a cluster of uncultivated phyla, the Candidate Phyla Radiation (Hug et al., 2016). To make progress toward describing the phenotypes of these bacteria, it will be difficult to rely entirely on culture-free methods. Genome-based phenotyping will benefit from a surge of culturomics (Lewis et al., 2021).

We will need to accommodate and tolerate the intrinsic instability of a species' description based on metabolism we cannot observe. George Garrity has anticipated that descriptions of species based entirely on genomes will be in flux, and he and his coworkers have proposed and patented the Names for Life database, which allows continual updating the data resources of a taxon. This includes a record of the original description as well as further taxonomic and nomenclatural events relating to the taxon (Garrity and Lyons, 2011).

Now that we are aware of the enormous diversity of bacterial species, an easier, genome-based path to describing species is most welcome. Genome-based systematics will allow us to spend less time describing new species, and more time on synthesizing the taxa we have (Sutcliffe et al., 2021).

Genome-based deposition, description, and demarcation of new species will give us time to employ phylogenomics to revise genera and higher levels (Xu et al., 2021). Also, we could more efficiently employ genomics toward systematics at the ecotype level.

In this spirit, we next delve below the species level to propose a completionist accounting of ecological diversity at the level of infraspecific ecotypes. We first motivate the need for understanding the diversity contained within the recognized species taxa.

3. The Ecological Breadth of Recognized Species Taxa

The animal ecologist G. Evelyn Hutchinson provided a pragmatic motivation for demarcating species as homogeneous groups—as based on their physiological, biochemical, morphological, and ecological characteristics (Hutchinson, 1968). He noted that species so defined have the useful property that the characteristics of any individual classified to a species could be easily predicted. While we believe Hutchinson was overly optimistic about the homogeneity of animal and plant species, microbiologists could probably agree that a taxonomy based on homogeneity, if possible, would be extremely beneficial (Cohan and Perry, 2007). For example, under this approach, the total membership of a pathogenic species would have the same disease-causing properties, the same tissue tropisms, the same transmission properties, and the same host range, while organisms with significantly different properties would be recognized as different species. It is widely understood that the species recognized by bacterial systematics are far from satisfying the Hutchinsonian property of homogeneity.

The classification scheme of bacterial systematics focuses on finding species that are significantly different from one another in DNA sequence identity, genome content, and physiology (Rosselló-Móra and Amann, 2001, 2015; Vandamme et al., 1996), but places almost no emphasis on ensuring that each individual species is homogeneous in any characteristic (Cohan, 2002; Staley, 2006; Ward et al., 2006). Under this system, two individuals are understood to be in the same species if they show a critical level of sequence identity, in 16S rRNA genes (Stackebrandt and Ebers, 2006) or more recently in various Overall Genome Relatedness Indices (OGRIs) (Chun and Rainey, 2014), including Average Nucleotide Identity of shared genes (ANI) (Jain et al., 2018; Olm et al., 2020) and digital DNA-DNA Hybridization (dDDH) (Meier-Kolthoff et al., 2021).

The degree of genetic diversity included within these species limits allows for enormous ecological (and in some cases, disease-causing) differences within a species, as illustrated by *Escherichia coli*. Some members of *E. coli* are specialized as pathogens, while others are specialized as commensals; some *E. coli* lineages are specialized to colonize the large intestine or other parts of the body (Touchon et al., 2009; Walk et al., 2009). These are vastly different environments where the bacteria encounter different extracellular secretions, pH, and notable differences in the extracellular matrix, which they must attach to. Moreover, different *E. coli* populations may be specialized to different hosts (Gordon and Lee, 1999) and different outside environments (Walk et al., 2007). The profound ecological and physiological differences among *E. coli* populations are reflected by huge genomic differences, with some divergent populations sharing fewer than half of their genes (Welch et al., 2002). Other named species have also been found to contain a high diversity of ecologically, physiologically, and genomically distinct members (Aboal et al., 2016; Arevalo et al., 2019; Dugat et al., 2015; Kettler et al., 2007; Kopac et al., 2014; Lefebure and Stanhope, 2007; Marri et al., 2006; Mukherjee et al., 2022; Paul et al., 2010; Vernikos et al., 2007).

How did bacterial systematists agree to house such a huge amalgam of diversity within the species they recognize? In the case of the animals and plants, humans long ago developed an "umwelt," a foundation for demarcating natural groups of consequence for our survival, through natural selection and cultural evolution (Yoon, 2009). The umwelt we have inherited from our hunter-gatherer past has been useful for identifying dangerous and useful plants and animals, and from that beginning, systematists could build their intuition about the diversity of macroorganisms. However, until the late 19th century the bacteria escaped the attention of human interest in biodiversity, and so systematists of bacteria had to develop a way of seeing and classifying the diversity of bacteria from scratch (Cohan, 2011a). Moreover, until recently, bacterial systematists have not had the advantage of being able to anticipate either the ecological differences between close relatives of bacteria or the physiological differences underlying their ecological divergence.

Bacteriologists were successful from the middle of the 20th century on in developing an objectively based umwelt for species demarcation. While limited at the time to metabolic and other phenotypic characteristics, "numerical taxonomy" allowed bacterial systematists to develop standard levels of phenotypic diversity within and between species (Sneath and Sokal, 1973; Yoon, 2009). In principle, species could have been defined narrowly, even on metabolic grounds, by focusing on subtle, quantitative differences in metabolic expression (Cohan, 2002; Feldgarden et al., 2003). However, systematists made a pragmatic, but fateful decision early on to include strains within a species that were heterogeneous in the presence versus absence of many metabolic capabilities (Cohan, 2011a; Rosselló-Móra and Amann, 2001; Vandamme et al., 1996). Thus, bacterial species were from the start defined to be extremely diverse in their physiological and hence ecological characteristics.

Molecular technology has not brought about a refinement in the breadth of diversity subsumed within a recognized taxon. Rather, as each new technology has been embraced, including DNA-DNA hybridization (Wayne et al., 1987), 16S rRNA sequence (Stackebrandt and Ebers, 2006), multilocus sequence analysis (Gevers et al., 2005), genome-wide average nucleotide identity (ANI) (Goris et al., 2007; Konstantinidis and Tiedje, 2005), and in silico DNA-DNA hybridization based on genome sequences (Meier-Kolthoff et al., 2021), systematists have attempted to calibrate every newest method to yield the species taxa originally demarcated decades ago on metabolic grounds (Cohan, 2002, 2021).

Thus, while the approaches of systematics have brought pragmatic solutions for the practice of systematists, we might ask whether these approaches have been pragmatic for microbiologists outside of systematics. The problem is that when systematists reify an amalgam of ecological and functional heterogeneity into a species taxon, other microbiologists tend to assume that each such species constitutes a natural and fundamental unit of biodiversity. This has led to numerous unfortunate consequences for microbiologists outside of systematics.

One such consequence is the classification of genes within a recognized species as essential, "core" genes, shared by all "species" members, versus the nonessential, "dispensable" (Tettelin et al., 2005) or "flexible" (Kettler et al., 2007) genes that are shared only by a subset of species members. This dichotomy is false because it is based on the reification of the named species. A more finely demarcated species taxon would yield more core genes and fewer dispensable genes. Also, this gives the unintended impression that those genes held only by one subclade are somehow dispensable and not essential to the ecology or physiology of that group.

This reification of the core genome may have a negative impact on vaccine development. Vaccine development can be based on choosing a target protein that is a member of the core genome of a

bacterial species (Tettelin et al., 2005). However, if the pathogenic strains of concern constitute only a single ecotype within the species diversity, the choice of vaccine target is unnecessarily restricted to the small core genome of the entire named species, rather than the larger set of genes shared among members of the pathogen ecotype.

The broad definition of recognized species has led to innumerable errors in estimating the critical population genetic parameters of evolution (Cohan, 2021). For example, attributing the name *Escherichia coli* to the huge diversity of ecological specialists within the species taxon gives population geneticists the impression that they are dealing with a group of ecologically interchangeable organisms, such as the members of the fruit fly species *Drosophila pseudoobscura* within a particular oak woodland habitat. This has led to incorrect application of various algorithms for estimating effective population size and recombination rates (Gordon and Lee, 1999), which assume that the organisms sampled are interchangeable members of a single population (McVean et al., 2002). (More specifically, the algorithms assume that any one individual alive today could become the ancestor of the entire population in the future.) In the case of estimating effective population size from sequence diversity, ecological heterogeneity artificially increases the sequence diversity, and thereby the estimate of population size, by conflating the divergence between populations (which is not affected by population size) with divergence within them (Wielgoss, 2022). In our own work, sequence-based estimations of migration rates have also erred by pooling within a species taxon various ecologically distinct groups (Roberts and Cohan, 1995).

In addition to the errors caused by species reification, the broad brush of systematics has incurred an opportunity cost for different subfields of microbiology, starting with systematics itself. When a systematist discovers a new species and sees that its diversity can be placed within one new species taxon, there is no further motivation from systematics to further explore the ecologically distinct clades within the species. As the novelist Alex Michaelides noted, "Once you name something, it stops you seeing the whole of it, or why it matters" (Michaelides, 2019). There is so much diversity that gets subsumed within each of our species taxa. Research in systematics is impoverished by a standard of detail that leaves much of a clade's diversity uncharacterized.

The broad brush may also incur an opportunity cost on epidemiology whenever we force closely related but ecologically disparate pathogens into the same species taxon. The most consequential and lethal example comes to us recently from the world of viral taxonomy. Virus systematics placed the pandemic Coronavirus of 2019 (SARS-CoV-2) into the same species as the original SARS Coronavirus of 2003 because of their high amino acid identity (Gorbalenya et al., 2020; Zhou et al., 2020), although these viruses differ significantly in their epidemiological features. For example, SARS-CoV-2 transmits before symptoms appear (Rothe et al., 2020), while the original SARS virus was not transmissible until *after* severe headache and myalgia symptoms appeared (Mazzulli et al., 2004). The viruses also differ dramatically in their case mortality rates, 2% for the novel coronavirus versus 9.6% for SARS (World Health Organization, 2020). Demarcating SARS-CoV-1 and SARS-CoV-2 into the same species encouraged scientists to think that the viruses would have certain similarities, in particular that asymptomatic or presymptomatic transmission would be impossible in COVID-19, on the basis of the original SARS (Gorbalenya et al., 2020). As a consequence, it took months for the World Health Organization to accept mounting data that the virus could be transmitted from people with no symptoms (Apuzzo et al., 2020). The reluctance to consider that SARS-CoV-2 could differ profoundly in its transmission from SARS-CoV-1 may have cost tens of thousands of lives.

Biotechnologists could also take advantage of a more fine-grained systematics of species. After discovering a strain with a valuable enzyme, one could then search for homologs of the enzyme across closely related, ecologically distinct populations, if they were highlighted by taxonomic recognition (Cohan and Perry, 2007; Jensen, 2010).

The molecular revolution has taken us far beyond the early days of systematics, when species demarcation was based entirely on metabolism and other phenotypic traits. Sequencing has now revealed ecologically distinct populations within the recognized species, yet we do not take advantage of this information to refine the demarcations of species. This would be an auspicious time to incorporate the high resolution of genomic technology into our taxonomy, so that the physiological and ecological diversity we know to exist within the named species can be officially recognized. An important challenge is to develop universal algorithms to analyze sequence and genome content data to identify populations that are each ecologically homogeneous but are ecologically distinct from one another.

4. Sexual Isolation and Ecological Divergence in the Origins of Species-like Lineages That can Coexist Indefinitely

How can we integrate the species-like ecotypes into taxonomic classification? To this end, we must consider how cohesive forces constrain diversity within ecotypes and how one ecotype can break through cohesion to split into two ecotypes that can coexist indefinitely. Some bacteriologists have focused on the Biological Species Concept, where sexual isolation is required before populations can diverge in their ecology. The allure of the Biological Species Concept stems from the midcentury work of Ernst Mayr and Theodosius Dobzhansky. They defined animal species as populations that could exchange genes at a high frequency, whereas populations from different species could not (Dobzhansky, 1951; Mayr, 1942). Thus, the high recombination rate within species could limit the divergence among populations of the same species but not the divergence between different species. The model states that populations could diverge without bound only when they break free of their recurrent recombination, through evolving sexual isolation (Bobay and Ochman, 2017; Cohan, 2017). This concept has held hegemony in the species-level systematics of animals and plants since the mid-20th century.

The appeal for a systematics based on patterns of sexual isolation has long run strong in microbiology (Bobay and Ochman, 2017; Cadillo-Quiroz et al., 2012; Dykhuizen and Green, 1991; Kashtan et al., 2014; Polz et al., 2013). Recently, Bobay and Ochman provided evidence that the Biological Species Concept can apply across all of life, from viruses to bacteria and archaea and beyond (Bobay and Ochman, 2017). Through inferring recombination rates from thousands of viral and prokaryotic genomes, they found that recombination is universally much greater within the recognized species taxa than between them.

Several workers have presented evidence that ecological divergence cannot occur between bacterial lineages until sexual isolation has evolved (Cadillo-Quiroz et al., 2012; Kashtan et al., 2014; Polz et al., 2013). They have found that extremely closely related, ecologically distinct clades showed reduced recombination between them, with greater recombination within clades, and many others have found similar results (Arevalo et al., 2019; Eppley et al., 2007; Stanojković et al., 2022). These results are consistent with the hypothesis that sexual isolation is a prerequisite for evolution of ecological

divergence. However, one should consider, alternatively, that the clades studied may have diverged first, with no prerequisite of reduced recombination, and that sexual isolation evolved later (Mallet, 2008a; Melendrez et al., 2016).

There is abundant empirical data that one bacterial population can diverge into two ecologically distinct lineages without prior reduction of their recombination rate. In laboratory evolution experiments, a single clone is apt to diversify into multiple populations within a culture flask, owing to specialization either on different soluble resources (Blount et al., 2012; Treves et al., 1998) or on different microhabitats within the same flask (Koeppel et al., 2013; Rainey and Travisano, 1998). Likewise, surveys of ecological diversity within one natural habitat have demonstrated ecological divergence among extremely closely related strains (Shapiro and Polz, 2014). Indeed, our work with hot spring *Synechococcus* has demonstrated ecological divergence among extremely close relatives, even those recombining at the highest levels (Melendrez et al., 2016).

Nearly a century ago, the population geneticist J. B. S. Haldane gave a theoretical argument for why we should not fuss about recombination in systematics (Haldane, 1932). He showed mathematically that recurrent recombination between populations adapted to different circumstances would likely have only a negligible effect on the populations' abilities to maintain their unique adaptations. That is, if c_b is the rate of recombination between populations (Haldane called it m, for migration) and s is the selection intensity against migrant alleles, then the equilibrium frequency of the maladaptive, migrant allele in a population is c_b/s, which would be tiny for any set of populations with limited recombination between them (Cohan, 2011a; Vos and Didelot, 2009).

Haldane was clearly thinking about the role of recombination in animal and plant divergence, but recombination is even less likely to limit adaptive divergence in bacteria. This is because even the most closely related bacteria from the same population recombine at extremely low rates, usually at not much more than the rate of mutation (Vos and Didelot, 2009). Even if different populations were to recombine at the same rate as cells of the same population, the equilibrium frequency of a foreign allele would be negligible. Thus, two populations could diverge even if they live in exactly the same place (e.g., two populations living on different soluble compounds in the same aquatic environment) and recombine at the same rate as members of the same population. We have previously concluded that the evolution of sexual isolation is not a milestone in the ecological divergence of bacterial lineages (Cohan, 1994).

If our aim is to find bacterial lineages with the species-like properties that they are ecologically distinct and are poised to diverge indefinitely, we should not focus on the property of sexual isolation. Instead, we should aim directly to find lineages that are ecologically distinct and have shown a history of coexisting with one another.

In the ecotype theory of bacterial species and speciation, ecological diversification does not require sexual isolation. Ecotypes are the most newly divergent populations that are ecologically distinct from one another and are each ecologically homogeneous; as a result of their ecological distinctness, ecotypes are able to coexist indefinitely (Kopac et al., 2014) (Fig. 2.2). Nevertheless, the ecotype theory recognizes that every population will have *some* ecological diversity: ecotypes are defined such that the ecological divergence among lineages within an ecotype is not sufficient to allow them to coexist indefinitely; we may refer to ecologically distinct lineages within an ecotype as "ephemeral ecotypes."

In the ecotype model, the ecological near-homogeneity of an ecotype's members can grant cohesion to the ecotype. Here, natural selection favoring one competitively superior adaptive mutation

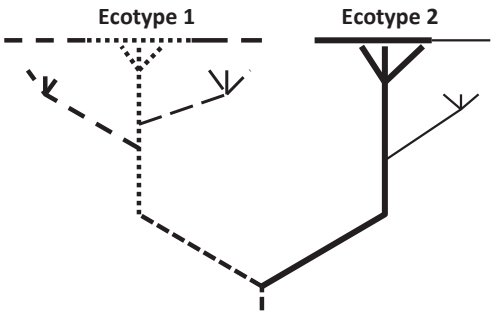


FIGURE 2.2 Ecological divergence between ecotypes and ecological near-homogeneity within ecotypes.

The ecological divergence between ecotypes is sufficient for them to coexist into the indefinite future. Ecotypes are defined so that any ecological differences among lineages within ecotypes are not sufficient to allow them to coexist indefinitely. We thus refer to ecologically distinct lineages within ecotypes as "ephemeral ecotypes." The different styles of dashed lines within ecotype 1 refer to different ephemeral ecotypes; note that only one of these lineages persists to the present. The different weights of solid lines represent different ephemeral ecotypes within ecotype 2.

within an ecotype causes the mutation to reach 100% frequency within the ecotype. Because recombination is so rare in bacteria (Vos and Didelot, 2009), nearly the entire genome of the adaptive mutant can reach 100% frequency (Cohan, 1994; Wiedenbeck and Cohan, 2011). Thus, the ecological near-homogeneity within an ecotype can result in recurrent, genome-wide purges of diversity called periodic selection (Fig. 2.2) (Atwood et al., 1951; Cohan, 1994; Templeton, 1989).

Periodic selection is likely the oldest concept in the population biology of bacteria, discovered in the early 1950s by Atwood and colleagues (Atwood et al., 1951). They explored the population dynamics within an extremely simple community in minimal medium, containing one population of *Escherichia coli* with four selectable genotypes that were nevertheless ecologically interchangeable. That is, the genotypes had the same fitness and utilized the same nutrients in the context of the minimal medium. Atwood and colleagues found that the four genotypes would remain at the relative frequencies at which they were introduced, but then after some random time, one of the genotypes swiftly rose to 100% at the expense of the others. Atwood et al. showed that the sweep of diversity occurred as a result of selection favoring an adaptive mutation, and they named this phenomenon "periodic selection."

While periodic selection is the foundation for ecotype theory, some have doubted whether the phenomenon extends to natural populations of bacteria (Shapiro and Polz, 2015). One reason is that no examples of a genome-wide sweep were discovered from nature for many decades after the seminal periodic selection paper. However, the hypothesis of periodic selection was supported recently by a metagenomic study of diversity at various depths within a lake in Wisconsin, USA (Bendall et al., 2016; Cohan, 2016a). In a longitudinal survey, Bendall et al. assembled metagenomic sequences into clusters of genetically similar organisms, usually with up to 2% sequence divergence, and one such cluster was shown to lose its diversity in a genome-wide sweep over a span of eight years—the first periodic selection event observed in nature. It is not surprising that metagenomics would document such an incident, as the approach allows us to monitor a whole community of ecotypes simultaneously

for a genome-wide purge. In contrast, finding such an event in one researcher's favorite organism over a short window of observation would be extremely unlikely. (We should note that metagenomic evidence for periodic selection would be strengthened by experimental tests of the competitive superiority of the winning genome sequence.)

Another promising approach for finding periodic selection events in *nearly* natural communities was developed by Frazão and colleagues (Frazão et al., 2022). Here, one lineage of *E. coli* with selectable markers was introduced into the gut microbiomes of several mice (with all their native microbial diversity intact). Genome sequencing of the introduced lineage over time revealed the spontaneous appearance of sequence variants within the lineage. The authors found that within some of the mice, one variant of the introduced lineage swept to nearly 100% of that lineage, which they interpreted as periodic selection. Thus, metagenomic and genomic techniques now allow us to detect periodic selection events in nature.

Nevertheless, many genomic surveys of diversity of close relatives (e.g., within a recognized species taxon) appear to contradict the model of periodic selection. Here, a species taxon or a smaller lineage shows a sweep of just one short chromosomal segment rather than the whole genome (Shapiro and Polz, 2015). We might call this a "whisk-broom sweep" (Cohan, 2016a), as occurs in a highly sexual population of eukaryotes when selection favors an adaptive mutation (Pritchard et al., 2010). How can these whisk-broom sweeps be reconciled with periodic selection, which predicts "garage-broom sweeps" that will clean the diversity of the whole genome (Cohan, 2016a)?

Some have interpreted these cases as evidence that perhaps whole-genome sweeps predicted by periodic selection theory are prevented by abundant rates of recombination within a population (Papke et al., 2007). This interpretation, though, contradicts what we know about the rarity of recombination in bacteria—that it occurs at about the same rate as mutation (Vos and Didelot, 2009), which is too infrequent to prevent selection from pushing the entire genome of an adaptive mutation to near 100% (Cohan, 1994, 2016a).

We have argued that, while short-segment sweeps do not support the model of genome-wide sweeps by periodic selection, they do not contradict the model, either (Cohan, 2016a; Majewski and Cohan, 1999a). Consider, for example, the case of two ecotypes of hot spring *Synechococcus* that are adapted to extremely different temperatures but are nearly identical in their nitrogen fixation operons (Bhaya et al., 2007), or a similar case where two pathogen species yielding different diseases are identical in an antibiotic resistance gene (Shapiro and Polz, 2015). These cases demonstrate that a small-segment sweep can occur when a generally adaptive gene transfers between two ecologically distinct populations. Here, the transfer of an adaptation can homogenize two ecologically different populations over just a small chromosomal region while the populations maintain their divergence over the rest of the chromosome. Inspired by such cases, our "Adapt Globally Adapt Locally" model shows how the transfer of a globally adaptive mutation throughout a set of ecotypes can make each ecotype better adapted to its own local environment (Majewski and Cohan, 1999a). We have argued that the short-segment sweeps found within bacterial species taxa represent the spread of an adaptive gene across ecotypes while the ecotypes remain ecologically distinct after the transfer (Cohan, 2016a).

In summary, we now have evidence that periodic selection sweeps are occurring in nature, provided that we look for them with the appropriate tools, and moreover, that short-segment sweeps appearing over a set of ecologically distinct ecotypes within a taxon are consistent with periodic selection.

We will next explore variations of the ecotype model. We shall see that in some models, the dynamics of periodic selection make ecotypes easy to discover as sequence clusters, but that in other

models the discovery of ecotypes is more challenging. Moreover, we shall see that in one ecotype model, periodic selection does not act as a cohesive force.

5. How Modes of Speciation Affect the Correspondence Between Sequence Clusters and Ecotypes

Stable Ecotype model. In the Stable Ecotype model, ecotypes are long-lived; different ecotypes coexist with one another indefinitely; and speciation is rare (i.e., new ecotypes are formed infrequently) (Fig. 2.3A). The long-term coexistence of different ecotypes may be fostered by a *qualitative* ecological divergence, where each ecotype utilizes some unique resource not shared with others (Kopac et al., 2014). Any resource that a particular ecotype does not share with closely related ecotypes creates an ecological haven from competition, such that adaptive mutations in other ecotypes will not drive the focus ecotype to extinction.

The longevity of each ecotype provides ample opportunity for the ecotype to acquire a unique set of neutral mutations in each gene in the genome; the longevity also provides opportunities for many periodic selection events to occur during the long lifetime of an ecotype. Thus, in the Stable Ecotype model, each ecotype is cohesive by virtue of periodic selection events that recurrently purge the ecotype of sequence diversity (Cohan and Perry, 2007).

New ecotypes are founded when a single individual acquires a mutation (or a recombination event) that changes its ecology. The ecological change may involve utilizing a new set of resources, thriving under a new set of environmental conditions, or adopting some other change in lifestyle. Since new ecotypes are each founded by a single individual, they start out with zero diversity. A new ecotype is not in direct competition with the members of the parental ecotype because it lives in a different microhabitat or uses at least somewhat different resources. For example, a member of a primarily impetigo-causing (skin-infecting) ecotype of *Streptococcus pyogenes* might mutate or acquire a gene that allows it to primarily infect the throat (Bessen, 2009), thus founding a new ecotype. Although the new ecotype may share the same host species as the parental ecotype, it is utilizing different host resources, and so the two ecotypes may not experience the same periodic selection events.

In the Stable Ecotype model, ecotypes have all the fundamental characteristics broadly attributed to species (Kopac et al., 2014). Like species, ecotypes are ecologically distinct. They are cohesive because an ecotype's diversity is recurrently purged genome-wide by periodic selection events. Ecotypes are irreversibly separate because their ecological distinctness prevents periodic selection within one from extinguishing the other and because recombination in bacteria is insufficient to reverse ecological diversification (Wiedenbeck and Cohan, 2011).

In principle, ecotypes that abide by the slow speciation of the Stable Ecotype model can be easily discovered as sequence clusters, even with the low resolution of one or a few genes. Each ecotype accumulates a unique set of neutral mutations, while the sequence diversity within each ecotype is recurrently purged by its periodic selection events. Thus, the diversity within each ecotype is represented only by the diversity that has accumulated since its last periodic selection event (a short time), while the ecotypes are separated by the long time since the last speciation event.

We will next consider alternative models of ecological diversification where sequence clusters fail to predict ecotypes.

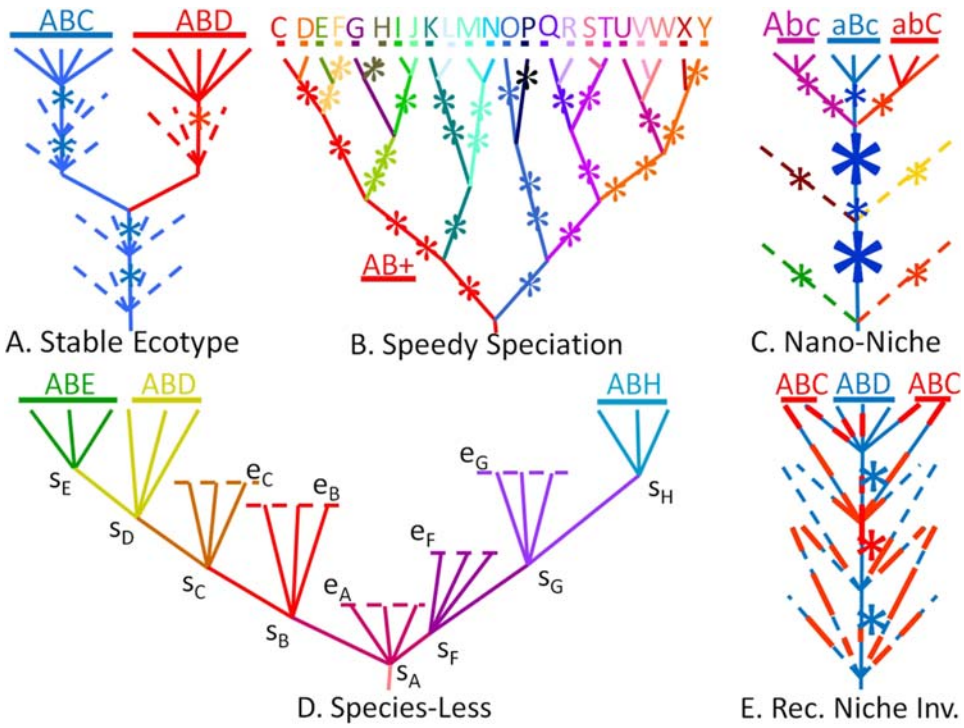


FIGURE 2.3 Models of bacterial speciation.

Ecotypes are represented by different colors; periodic selection events are indicated by asterisks, and extinct lineages are represented by dashed lines. The letters at the top represent the resources that each group of organisms can utilize. In cases where ecotypes utilize the same set of resources but in different proportions, the predominant resource of each ecotype is noted by a capital letter. (A) The Stable Ecotype model. In the Stable Ecotype model, each ecotype endures many periodic selection events during its long lifetime. The Stable Ecotype model generally yields a one-to-one correspondence between ecotypes and sequence clusters because ecotypes are formed at a low rate. The ecotypes are able to coexist indefinitely because each has a resource not shared with the others. (B) The Speedy Speciation model. This model is much like the Stable Ecotype model, except that speciation occurs so rapidly that most newly divergent ecotypes cannot be detected as sequence clusters in multilocus analyses. (C) The Nano-Niche model of bacterial speciation. In the figure, there are three Nano-Niche ecotypes that use the same set of resources but in different proportions (noted by Abc, aBc, and abC). Each Nano-Niche ecotype can coexist with the other two because they have partitioned their resources, at least quantitatively. However, because the ecotypes share all their resources, each is vulnerable to a possible speciation-quashing mutation that may arise in the other ecotypes. (D) The Species-Less model. Here the diversity within an ecotype is not limited by periodic selection but instead by the short time from the ecotype's invention as a single mutant until its extinction. The origination and extinction of each ecotype i is indicated by s_i and e_i, respectively. In the absence of periodic selection, each extant ecotype that has given rise to another ecotype is a paraphyletic group, and each recent ecotype that has not yet given rise to another ecotype is monophyletic (Kopac et al., 2014). (E) Recurrent Niche Invasion model. Here a lineage may move, frequently and recurrently, from one population to another, usually by acquisition and loss of niche-determining plasmids.

Re-used with permission Wiedenbeck J and Cohan FM. 2011. Origins of bacterial diversity through horizontal gene transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. 35:957–976. Cohan FM and Perry EB. 2007. A systematics for discovering the fundamental units of bacterial diversity. Curr. Biol. 17:R373–R386 and Cohan FM. 2011. Are species cohesive? A view from bacteriology, p 43–65. In Walk S and Feng P (ed).

Speedy Speciation model. In the Speedy Speciation model, cohesion occurs through periodic selection, just as in the Stable Ecotype model (Cohan and Perry, 2007) (Fig. 2.3B). The difference is that speciation is greatly accelerated, as occurs in an adaptive radiation. The practical consequence of rapid speciation is in reducing our ability to discover ecotypes easily as sequence clusters.

When ecological diversification is slow (as in the Stable Ecotype model), the phylogenetic resolution of even a small number of genes (e.g., 1−3 genes) may distinguish ecotypes as sequence clusters, as we have seen with ecotypes of hot spring *Synechococcus* (Becraft et al., 2015). On the other hand, if ecotypes are recurrently and rapidly splitting into new ecotypes, then more phylogenetic resolution will be required to find the clusters that correspond to the most newly divergent ecotypes (Kopac et al., 2014; Melendrez et al., 2011). When phylogenies of both low and high molecular resolution predict the same ecotypes, we may conclude that ecotypes are forming slowly. However, when a high-resolution phylogeny (e.g., based on a core genome) predicts many more ecotypes than a low-resolution phylogeny, this is consistent with rapid diversification into new ecotypes.

A likely example of Speedy Speciation comes from the diversification of *Pseudomonas aeruginosa* within the lungs of cystic fibrosis patients (Jorth et al., 2015). Jorth et al. found that in different regions of a CF lung of a single patient, multiple populations had diversified in situ and were divergent in numerous features that apparently adapted them to their local environments, including differences in nutrient requirements, antibiotic resistance, and virulence. These ecotypes have certainly evolved quickly (within one short-lived human), and the accumulated adaptive differences over time probably represent periodic selection events within each; so the Speedy Speciation model appears to apply. Alternatively, the situation would be different if *P. aeruginosa* ecotypes adapted to CF lungs were to spread frequently from one CF patient to another (Saiman and Siegel, 2004). In this case, the Speedy Speciation model would probably not apply, as a given lung niche could be colonized by an already-adapted ecotype rather than requiring evolution in situ.

A similar case of Speedy Speciation appears to occur in a chronic *Mycobacterium tuberculosis* infection within a single human. It appears that, owing to bottlenecks in transmission of the tuberculosis pathogen between humans, individual hosts are often infected by a single lineage, which can then diversify rapidly within a new host. Thus, a given tuberculosis patient is likely to host a diversity of ecologically divergent ecotypes that have evolved in situ (Warner et al., 2015). In both *P. aeruginosa* and *M. tuberculosis*, many ecotypes originating within one human's lungs may not be transmitted and so are short-lived.

Species-Less model. The Species-Less model is profoundly different from the Stable Ecotype model and all models that assume cohesion within species (Cohan and Perry, 2007) (Fig. 2.3D). The Species-Less model assumes both rapid speciation and rapid extinction, leading to a high turnover of species. In this case, a species may not persist long enough, from its time of origin to its extinction, to undergo even a single periodic selection event. In the Species-Less model, each ephemeral ecotype, while ecologically homogeneous, could not be considered a cohesive unit. Like the case of the Speedy Speciation model, where species are cohesive, the Species-Less model will lead to a diversity of ecotypes (in this case, ephemeral ecotypes) that cannot be easily distinguished as sequence clusters.

In the Species-Less model, ecotypes evolve not by becoming more efficient in utilizing their current ecological niche, but instead by evolving to invade a new ecological niche. The Species-Less model may apply to the case for pathogens, where immune-escape mutations may each constitute a new ephemeral ecotype (Achtman and Wagner, 2008; Cohan, 2011a). Also, the Species-Less model may apply in cases where an environment undergoes a succession process, where organisms at a site

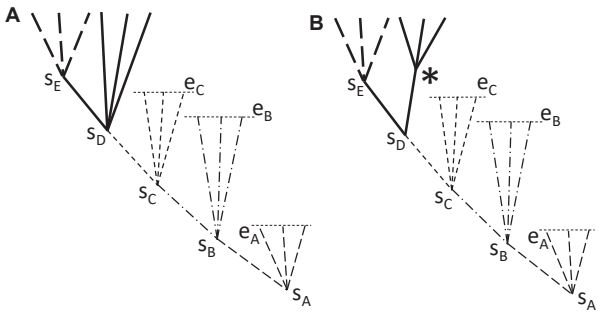


FIGURE 2.4 The Species-Less model of bacterial diversification.

In the Species-Less model, the diversity within an ecotype is not limited by periodic selection but instead by the short time from the ecotype's invention until its extinction. Each ephemeral ecotype in the figure is indicated by a unique line style; the origination and extinction of each ephemeral ecotype *i* is represented by s_i and e_i, respectively. (A) In the absence of periodic selection, each ephemeral ecotype that has given rise to another ephemeral ecotype is a paraphyletic group (e.g., ecotype D), and each recent ecotype that has not yet given rise to another ecotype is a monophyletic group (e.g., ecotype E). (B) If instead a periodic selection event has occurred in the parental ecotype since it has given rise to a daughter ecotype, the mother-daughter pair will be seen as a pair of sister monophyletic clades. Observing that pairs of most-closely related ecotypes are usually observed as paraphyletic-monophyletic pairs would indicate that the origin of new ecotypes is more common than periodic selection in existing ecotypes, giving support to the Species-Less model (Cohan, 2011a).

Used with permission from the American Society for Microbiology.

must adapt to rapidly changing conditions, for example, the successions that occur on mine tailings, with pH and oxidation levels changing predictably and quickly (Remonsellez et al., 2009).

Because periodic selection is not operating in the Species-Less model to purge the diversity within ecotypes, closely related ecotypes will not be discernible as separate, monophyletic sequence clusters. Rather, most-closely-related ecotypes will appear as a mother-daughter pair, where the mother ecotype is paraphyletic and the daughter ecotype is monophyletic (Fig. 2.4). Gene content analyses may be required to identify ecotypes in this model.

Nano-Niche model. Like the Species-Less model, the Nano-Niche model also assumes a high rate of speciation, but here cohesion occurs over a set of different ephemeral ecotypes (Cohan and Perry, 2007) (Fig. 2.3C). In the Nano-Niche model, closely related, ephemeral ecotypes are subtly and only quantitatively different in their ecology. These "Nano-Niche ecotypes" use the same set of resources and conditions, but they coexist much like closely related animal species by using their shared resources and conditions *in different proportions*. Not having any unique resources that might constitute a haven from competition from other ecotypes, each ecotype is ephemeral and vulnerable to extinction from competition with other ecotypes. For a time, the various Nano-Niche, ephemeral ecotypes may coexist while each has its own private periodic selection events. At some point, however, an extremely competitive adaptive mutant (which we call a speciation-quashing mutation) from one ephemeral ecotype may extinguish not only the other members of its own population, but also other closely related, ephemeral ecotypes (Cohan, 2005). In the Nano-Niche model, divergence among very closely

related ephemeral ecotypes is limited by these speciation-quashing mutations, and many ephemeral ecotypes might not last long enough to appear as separate sequence clusters.

We recently found evidence for the Nano-Niche model in *Bacillus spizizenii* within one extremely closely related cluster of sequences that was initially interpreted as an ecotype (Connor et al., 2010; Kopac et al., 2014). A survey of genomic diversity showed that each of five isolates sampled had a unique history of positive selection (adaptive evolution revealed by an excess of nonsynonymous mutations in some genes). This indicated that no two members of the putative ecotype were ecologically identical (Kopac et al., 2014; Vos, 2011). A comparison of genome content indicated no differences in genome content that would indicate unshared resources. Thus, the strains appeared to be ecologically distinct only in the extent to which they utilize the same set of resources. For example, all strains sampled had the capacity to utilize maltose, but one strain had additional genes for maltose utilization and was able to grow faster on maltose than the others. These strains may have a limited future of coexistence and likely constitute ephemeral ecotypes, since a speciation-quashing event originating in one lineage could extinguish the others.

The Nano-Niche model may also apply to bacterial ecotypes that adapt over a long course of infection within a host individual (e.g., a chronically infecting pathogen or commensal within one person). The course of evolutionary adaptation to one human body may bring about multiple periodic selection events within that population, and different individual humans could support different ephemeral ecotypes. However, the individual hosts might not be different enough to support indefinite coexistence of lineages that are specialized to different human individuals. Any speciation-quashing mutation, which makes an individual bacterium not just superior in its own host but also in other hosts, would put an end to the speciation among the various ephemeral ecotypes adapting to different host individuals. In our quest to identify ecologically homogeneous groups, we should perhaps be satisfied with finding *sets* of Nano-Niche, ephemeral ecotypes whose diversity will be purged with a speciation-quashing mutation, rather than identifying every ephemeral ecotype.

Recurrent Niche Invasion model. In the Recurrent Niche Invasion model, mobile genetic elements such as plasmids or phage may determine bacterial niches (Cohan and Perry, 2007) (Fig. 2.3E). For example, in the case of *Rhizobium*, a bacterial lineage may acquire a symbiosis-encoding plasmid that adapts it as an endosymbiotic mutualist for a particular set of legume hosts; the lineage may then lose that plasmid and gain another, which adapts it to another set of legume hosts. Ecotypes in this model are not irreversibly separate, as a lineage can recurrently leave one plasmid-determined ecotype and then join another (by gaining and losing different symbiotic plasmids).

In the case of *Rhizobium leguminosarum*, there are five sequence clusters that can each be infected with various symbiosis plasmids that adapt the bacteria to either vetch or clover (Kumar et al., 2015). While the five bacterial clusters have no known diagnostic features that would explain their ecological coexistence, one possibility is that these populations are distinguished quantitatively in their propensities to be infected by different symbiosis plasmids, a result suggested by a contingency test of association between the clusters and plasmid types (vetch-adapting vs. clover-adapting, a 5X2 Fisher's exact test, $P = .011$, based on data in (Kumar et al., 2015)). So, perhaps one dimension of ecological distinctness among the five clusters is a quantitative difference in their tendencies to infect vetch versus clover.

A similar situation arises within a set of longstanding sequence clusters that include the pathogens *Bacillus cereus* (a gut pathogen of mammals), *B. anthracis* (causing anthrax and systemic infection in mammals), and *B. thuringiensis* (an insect pathogen). These taxa are problematic in that they are

classified by these phenotypes, which are provided by niche-determining plasmids. Because the plasmids can move between sequence clusters, the species taxa *B. cereus* and *B. thuringiensis* are both polyphyletic (Liu et al., 2015). As in the case of *R. leguminosarum*, a contingency test of four plasmid categories versus five clusters yielded a high degree of association ($P \approx 0$, based on data in Liu et al. (2015)). Thus, part of what is contributing to the coexistence of the various longstanding sequence clusters appears to be their propensities toward infection by the different niche-determining plasmids.

Cohesive Recombination model. The Cohesive Recombination model provides another mechanism by which ecologically distinct populations will fail to be recognizable as sequence clusters (Cohan and Perry, 2007). As analyzed quantitatively by Hanage et al. (2006), bacteria with the highest rates of recombination may exchange genes so frequently that ecotypes do not accumulate sequence divergence in niche-neutral genes (Doroghazi and Buckley, 2014), and so we will not be able to discern the ecotypes as distinct sequence clusters. We note, however, that the rates of genetic exchange in bacteria are never sufficient to hinder or reverse adaptive divergence in niche-specifying genes, as we have previously discussed (Cohan and Koeppel, 2008; Melendrez et al., 2016; Wiedenbeck and Cohan, 2011). Thus, while genetic exchange in rapidly recombining bacteria will not prevent ecotype formation, it may prevent our ability to discover ecotypes using niche-neutral sequence diversity.

Geotype plus Boeing model. In some cases, a single ecotype from a given community may fall into several distinct sequence clusters, as seen in the Geotype plus Boeing model (Cohan and Perry, 2007). Provided that a given taxon has not dispersed frequently, geographically isolated members of a single ecotype may diverge into different clusters (geotypes), even while remaining ecologically interchangeable (Papke and Ward, 2004). This would yield a different sequence cluster in each geographical region, a common phenomenon in the systematics of all organisms that do not readily disperse (Whitaker et al., 2003). In the case of bacteria, geotypes may be a source of confusion for systematists if geotypes have a history of being geographically isolated, while modern human travel and transport has recently accelerated their dispersal. In this case (the Geotype plus Boeing model), members of one ecotype isolated from a single site may contain multiple clusters representing the ecotype's various, formerly isolated geotypes from all over the world. Because sequence clusters, even within a single location, may represent geotypes rather than ecotypes, we will need to confirm the ecological distinctness of sequence clusters we identify.

How ecology determines the rate of ecotype formation. Of all these models, only the Stable Ecotype model is likely to yield a one-to-one correspondence between sequence clusters and ecologically distinct lineages. Let us next consider a general hypothesis about what taxa are likely to consistently undergo the slow diversification required for the Stable Ecotype model.

Our interest in the rates of ecotype formation emerged from two contrasting model systems of diversification that we have studied: the generalist heterotroph *Bacillus* in desert soils (Kopac et al., 2014) versus the photoautotrophic *Synechococcus* in hot springs (Becraft et al., 2015). We have discussed the rapid rate of diversification within one putative ecotype of *Bacillus* in the context of the Nano-Niche model. Apparently, there is much room for diversification in this generalist heterotroph simply by becoming better at utilizing resources that all closely related strains are already utilizing.

We found a very different pattern of diversification in the hot spring cyanobacterial genus *Synechococcus*. Here, each putative ecotype identified as a cluster with about 0.5% diversity genome-wide appeared to be ecologically homogeneous. The evidence was that each putative ecotype consisted of various sequence types that maintained the same relative frequencies across a great diversity of natural and experimentally perturbed habitats (Becraft et al., 2015). Thus, in the time that these putative

ecotypes accumulated 0.5% sequence diversity across their genomes, they had not diversified ecologically, and so diversification in *Synechococcus* appears to abide by the Stable Ecotype model.

The contrast between *Bacillus* and *Synechococcus* in their rates of speciation suggested a hypothesis to predict which organisms follow the slow speciation of the Stable Ecotype model. Among free-living bacteria, we have predicted that generalist heterotrophs like *Bacillus*, with many options for metabolic diversification, will speciate rapidly (Cohan, 2016a). On the other hand, photoautotrophs like *Synechococcus* minimally utilize organic compounds, and so they may be much more limited in their ecological opportunities for speciation.

We may hypothesize more generally that greater metabolic plasticity in a taxon should lead to rapid diversification into new ecotypes that can diverge from one another by specializing on alternative resources. A metagenomic approach offers promise to test this hypothesis systematically, by comparing many more systems than is possible for a small group of investigators focusing on their favorite organisms. Here we will return to Bendall and collaborators' metagenomic survey of evolution within a Wisconsin lake (Bendall et al., 2016; Cohan, 2016a). We saw earlier that when these researchers assembled metagenomic sequences into clusters of genetically similar organisms, usually with up to 2% sequence divergence, one such cluster lost its diversity in a genome-wide sweep over a span of 8 years. The authors also found evidence of some genome-wide sweeps occurring before the study began. Each cluster of organisms that they found swept of diversity, genome-wide, was interpreted as evidence that the entire cluster was ecologically homogeneous, such that one adaptive mutant could outcompete the entire diversity of the cluster. Because these clusters failed to diversify in the time that they accumulated as much as 2% sequence divergence, we may conclude that diversification within these clusters occurred at a slow rate consistent with the Stable Ecotype model. Interestingly, these clusters could be predicted a priori to have little opportunity to diversify—some of these clusters were photoautotrophs of the phylum Chlorobi; others were heterotrophs of single-carbon molecules of various phyla (with few options for changing their organic resources) (Cohan, 2016a).

The metagenomic study also found various taxa undergoing less-profound sweeps, where only a single chromosome region was swept within a cluster (Bendall et al., 2016). These clusters were interpreted to contain not a single, ecologically homogeneous population, but rather an amalgam of many newly divergent, ecologically distinct ecotypes (Cohan, 2016a), following the Adapt Globally Act Locally model (Cohan, 2016a; Majewski and Cohan, 1999a; Takeuchi et al., 2015; Wiedenbeck and Cohan, 2011). That is, an adaptive mutation that was generally useful across the diversity of ecotypes within a cluster could cause a genome-wide sweep within one ecotype, then transfer as a small recombinant segment to another ecotype, causing a genome-wide sweep there, and so on. The result is that the entire cluster would be swept of diversity only in the gene region around the adaptive mutation and would be heterogeneous everywhere else on the genome. So, the clusters undergoing sweeps over just a small chromosomal region were interpreted as undergoing rapid speciation (Cohan, 2016a), where a great diversity of ecotypes could be found within 2% sequence divergence. These taxa were for the most part generalist heterotrophs like *Bacillus*, supporting the hypothesis that a highly plastic metabolism is the key to rapid speciation (Bendall et al., 2016; Cohan, 2016a).

To sum up the data from free-living organisms, the Stable Ecotype model of slow speciation appears to apply to organisms with little opportunity to diversify through utilization of different organic resources, while more rapid speciation occurs in taxa that utilize a vast diversity of organic resources (Cohan, 2016a).

Where do bacterial pathogens fit in between the extremes of metabolic plasticity from *Bacillus* to *Synechococcus*? Some pathogens may follow the slow diversification of the Stable Ecotype model, especially if they diversify primarily by adapting to new host species and new host tissues. That is, if susceptible host species and tissues are not numerous, speciation might be infrequent. For example, the *Mycobacterium tuberculosis* complex (which includes several recognized, taxonomic species) includes sequence clusters that are each ecologically homogeneous and associated with different clades of hosts (humans, artiodactyls, pinnipeds) (Smith et al., 2006b). Also, within *Anaplasma phagocytophilum*, a tick-borne pathogen, sequence clusters have diversified into associations with different host species and are each ecologically homogeneous (Dugat et al., 2015).

Another pathogen group likely following the Stable Ecotype model is the complex of species related to *Borrelia burgdorferi* sensu stricto, the spirochete responsible for Lyme Disease in humans. These pathogens are tick-borne and are hosted and vectored by a variety of vertebrate hosts, and some are more specialized to host and vector than others. A genomic analysis of diversity in 14 species of the complex revealed multiple sequence clusters that were associated with different sets of hosts and tick vectors (Becker et al., 2016), and the authors concluded that diversification followed the Stable Ecotype model (which the authors referred to as the "cohesion ecotype" model).

The Stable Ecotype model may reasonably apply to two closely related pathogens within *Yersinia* that infect different tissues. *Y. pseudotuberculosis* is transmitted by the fecal−oral route; however, it has the capacity to be lethal if it should invade the lungs or the blood (Fisher et al., 2007). *Y. pestis*, the Plague pathogen, emerged from *Y. pseudotuberculosis* and has developed a lifestyle of systemic infection and transmission by fleas (Sun et al., 2014). Because each taxon appears to be homogeneous in its ecology, these two taxa may be considered ecotypes that are distinguished by host tissue and mode of transmission, while both are generalist with respect to host species.

For pathogens where diversification events are infrequent (for a lack of either susceptible, novel host species, or possible tissues to infect), we hypothesize that the Stable Ecotype model is likely to apply, with many periodic selection events occurring within the long lifetime of any given ecotype.

Given the diversity of models and rates of ecotype origination, how can we achieve a completionist accounting of the world's ecotypes? Following progress in applying genomic approaches to classifying organisms at the species level, we will propose a genome-based systematics at the infraspecific, ecotype level.

6. Genomes Enable Completionist Systematics at the Ecotype Level

A completionist accounting of the ecological diversity within bacterial species must overcome our ignorance about the physiological and ecological dimensions along which one ecotype splits into two. Unlike the case for animal systematists, we cannot readily anticipate the physiological and other ecological dimensions by which ecotypes might differ. In contrast to songbirds, for example, we do not have features like the shapes and sizes of beaks that could point us toward the ecological distinctions among close relatives. Also, our explorations into ecotype diversity are stifled by not knowing the microhabitat dimensions along which closely related ecotypes may diverge.

We can accommodate our ignorance with an ecology-blind, "reverse ecology" approach to discovering the ecologically distinct lineages within a bacterial species taxon, assuming no prior ecological knowledge about the lineages, either through habitat or genomic data (Arevalo et al., 2019;

Koeppel et al., 2008). Here, we identify ecotypes as sequence clusters, based on any set of genes, whether or not they are involved in the ecotypes' ecological divergence. This is possible because, while each gene in the genome is recurrently purged of diversity within an ecotype, different ecotypes are each accumulating unique neutral mutations (of no fitness consequence) in every gene (Palys et al., 1997). Basing our phylogenetic analysis on many genes or even the entire core genome will help us accommodate rapid speciation.

The first step in ecology-blind identification of ecotypes is to hypothesize and demarcate putative ecotypes as clusters based on sequence data, using one of several algorithms: Ecotype Simulation (Francisco et al., 2014; Koeppel et al., 2008; Wood et al., 2021), BAPS (Corander et al., 2008), GYMC (Barraclough et al., 2009), and Minimum Entropy Decomposition (Eren et al., 2015). In contrast to the approaches of bacterial systematics at the species level, none of these algorithms assumes a universal criterion of sequence divergence for demarcation. Rather, each algorithm uses sequence data from the taxon of focus to identify the appropriate sequence divergence criterion for distinguishing ecotypes. For example, Ecotype Simulation identifies sequence clusters that are most consistent with ecotypes after estimating the rates of ecotype formation and periodic selection from the sequence data; each ecotype is hypothesized to represent the largest phylogenetic group subject to periodic selection events (Francisco et al., 2014; Wood et al., 2021).

The putative ecotypes hypothesized by these algorithms have consistently been confirmed to be ecologically distinct from one another, based on differences in their habitat associations (Becraft et al., 2015; Koeppel et al., 2008; Ward et al., 2006). For example, we have confirmed the ecological distinctness of putative ecotypes in soil *Bacillus* by differences in their habitat associations, including differences in solar exposure, soil texture, rhizospheres, and elevation (Connor et al., 2010; DelGaudio, 2019; Koeppel et al., 2008). Putative ecotypes of *Synechococcus* from hot spring mats have differed in their temperature and depth associations (Becraft et al., 2015; Ward et al., 2006), and putative ecotypes of *Legionella* have differed in their host ranges (Cohan et al., 2006). In addition, many ecologists have noted that very closely related sequence clusters (demarcated by eye, rather than by a computer algorithm) are different in their habitat associations (Brisson and Dykhuizen, 2004; Koch et al., 2020; Manning et al., 2008; Normand et al., 2007; Smith et al., 2006a; Walk et al., 2007, 2009). To sum up, putative ecotypes that were identified as sequence clusters have generally been shown to be ecologically distinct, over a great diversity of phyla (Bacillota, Pseudomonadota, Spirochaetota, Actinomycetota, and Cyanobacteria).

It is important to note that ecotypes need not be *absolutely* specialized to different habitats: closely related ecotypes frequently are only somewhat ecologically specialized and coexist by *quantitative* differences in habitat distinctness (Hunt et al., 2008; Thompson and Kouba, 2019). For example, we have found that ecotypes of *Bacillus spizizenii* that are adapted to the hotter temperatures of south-facing slopes are also occasionally isolated from the shadier, north-facing slopes in a desert canyon (Connor et al., 2010). Perhaps we should not be surprised that alternatively specialized ecotypes are generalists to some extent, given that many bacteria can disperse broadly across habitats and even across continents (Roberts and Cohan, 1995; Walters et al., 2022).

Once we have identified ecotypes as sequence clusters that are differentially adapted to different habitats, we can aim to find physiological and genomic differences that underlie their habitat associations. For example, putative ecotypes of *Bacillus subtilis* associated with more direct solar exposure have membranes yielding greater thermal tolerances (Connor et al., 2010). Putative ecotypes of *Synechococcus* from downstream communities of a hot spring are less tolerant of extreme

temperatures than ecotypes from upstream communities (Allewalt et al., 2006); also, the downstream ecotypes have genes for utilizing nutrient ions not consumed by upstream communities (Bhaya et al., 2007; Olsen et al., 2015).

Habitat associations can take us only so far in confirming the ecological distinctness of ecotypes that we hypothesize as sequence clusters. One limitation is that some ecotypes may live in exactly the same microhabitat yet differ ecologically, by utilizing different soluble resources in a shared aquatic habitat (Blount et al., 2012), or through differing in their rates of metabolism and replication (Thompson and Kouba, 2019). In other cases, ecotypes may differ in their preferred habitats, but we may be unaware of the resources and conditions by which they differ.

Genome content comparisons may help us confirm the ecological distinctness of putative ecotypes when they are not clearly associated with different habitats. That is, one can investigate whether different sequence clusters stemming from a single habitat differ in their gene content; such differences could then suggest ecological differences. We see this as a future development for discovering and characterizing ecotypes, but to our knowledge no one has yet confirmed the ecological distinctness of closely related sequence clusters (falling within a single species taxon) from a single habitat based on their genome content. If we aspire toward a completionist characterization of ecotypes, genome content analyses will allow us to confirm ecotypes whose ecological differences are beyond what we can currently imagine.

Genome content has played a role in confirming that sequence clusters from *different* habitats are ecologically distinct in their adaptations. In a metagenomic approach, Mukherjee and colleagues sampled DNA from two abutting and chemically contrasting soil environments at the "Evolution Plateau" in Israel (Mukherjee et al., 2022). The sequence reads from each environment were assembled into metagenome-assembled genomes (MAGs), and within 20 species (some previously described, and others not yet named or described), assemblies were shown to differ for numerous single-nucleotide variants. The authors interpreted these cases as different ecotypes within species. We are encouraged by this metagenomic approach because it does not focus on any particular taxon of interest to the investigator; the approach can thereby address diversity within many species simultaneously.

Phil Arevalo and colleagues have developed a promising method to discover ecotypes based on the full genome sequences and the habitats of isolation of close relatives (Arevalo et al., 2019). They began by identifying genes that have transferred through homologous recombination among a set of closely related isolates, using their algorithm PopCOGenT. They then searched for correlations between recombined alleles at a given gene locus and habitat type. This approach identifies the ecotype lineages that are adapted to each habitat type as well as the genes underlying adaptations to the different habitat types. Using this approach, Arevalo and colleagues identified ecotypes within the human commensal *Ruminococcus gnavus* that differ in their habitat associations, in this case within the guts of humans with and without Crohn's disease.

In their *Ruminococcus* study, Arevalo et al. used habitat associations to confirm that their putative ecotypes were differentially adapted. However, we note that their search for recombined segments can potentially yield demarcations of putative ecotypes even within a single habitat type. Therefore, we see a role for this approach in discovering ecotypes when we do not know ahead of time the ecotypes' physiological or habitat preferences.

Beyond gene content comparisons, we can also confirm the ecological distinctness of putative ecotypes from analysis of genes that they share. Michiel Vos has argued for using genome-wide

analyses of positive selection to identify ecotypes (Vos, 2011). That is, when two closely related lineages show different histories of adaptive evolution in their shared genes, we may conclude that they are ecologically distinct (Kopac et al., 2014).

To sum up, we have laid out a sequence- and genome-based approach that will help us discover putative ecotypes, the ecologically distinct lineages within a recognized species taxon. Systematics cannot succeed in reaching a completionist outline of bacterial diversity without discovering these infraspecific lineages.

We consider next whether a completionist systematics may be possible at the ecotype level. This depends critically on how we define ecotypes. On one hand, we might simply define an ecotype as an ecologically distinct lineage with its own unique history of natural selection, for example, following Vos's approach (Vos, 2011). However, this approach could lead to an infinitude of ecotypes within a species taxon. For example, in the case of *Bacillus spizizenii*, all five extremely closely related isolates that we studied each had its own unique history of natural selection on shared genes (Kopac et al., 2014), and following Vos (2011) these would be considered different ecotypes. Apparently, using ecological distinctness as our criterion for demarcating ecotypes yields an unlimited number of ecotypes, at least within a *Bacillus* species. An ecotype-level systematics would then be utterly useless, as argued by Doolittle (2009), and would be endlessly distracted by an algorithm that identifies all the ecologically distinct lineages within a species.

On the other hand, we might demarcate ecotypes by a more stringent criterion—that members of an ecotype must have a history of cohesion through periodic selection, as discerned by forming sequence clusters with a history of coexistence. This criterion allows for limited ecological heterogeneity within an ecotype, provided that some periodic selection event will eventually purge the diversity among an ecotype's members (Fig. 2.2). In the Nano-Niche model, for example, "ephemeral ecotypes" may be subtly and quantitatively different in the resources they use. Here, one periodic selection event may purge the diversity across many ephemeral ecotypes.

The Ecotype Simulation algorithm avoids ephemeral ecotypes by attempting to find the largest sequence clusters that are each subject to the diversity-purging effects of periodic selection (Koeppel et al., 2008). In our studies of *Bacillus*, we have found that Ecotype Simulation does not find an infinitude of ecotypes (at least when applied with a low resolution of one to three genes); it has typically yielded a modest number of ecotypes within a species taxon (e.g., 5–20 ecotypes within each *Bacillus* species taxon) (Connor et al., 2010; Koeppel et al., 2008).

We must consider, however, whether the small number of ecotypes discovered within species may be an artifact of the low molecular resolution (using one to three genes) analyzed. Previous studies have revealed that as resolution has increased from 16S rRNA sequences to 16–23S intervening transcribed spacer sequences to protein-coding genes, more ecotypes are revealed (Melendrez et al., 2011). Therefore, it is reasonable to hypothesize that if we were to utilize the enormous molecular resolution of a species' entire core genome, we might discover an infinitude of ecotypes (Fig. 2.5).

We recently investigated in a rarefaction analysis how the number of shared genes used determines the number of ecotypes demarcated (Wang, 2024). We found that, whether 1 or 700 shared genes were used, the number of ecotypes remained the same within a species taxon (e.g., 5 ecotypes in the case of *B. spizizenii*). That is, within the rapidly diversifying, generalist heterotroph *B. spizizenii*, the number of ecotypes is not the infinitude suggested by the number of ecologically distinct lineages, where every isolate appears to be distinct (Kopac et al., 2014). Rather, it is the limited (and manageable) number of lineages that form sequence clusters at any level of molecular resolution. We conclude that the number

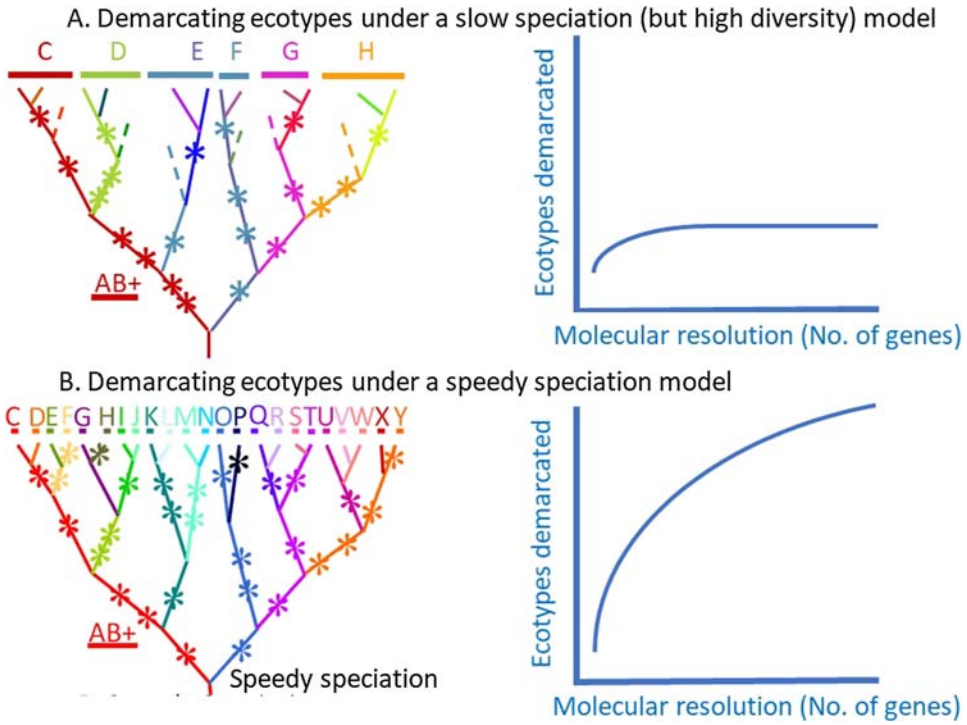


FIGURE 2.5 Demarcating ecotypes under models of slow and speedy speciation.

As in Fig. 2.3, each ecotype's unique resource set is indicated by the colored line above the ecotype in the phylogeny. That is, all ecotypes utilize resources A and B, plus each ecotype utilizes one additional resource (C—Y) that it does not share with other ecotypes. The ecotypes can coexist indefinitely because each ecotype utilizes a unique set of resources. Each asterisk represents a periodic selection event that purges diversity within an ecotype. (A) Under slow speciation, even low molecular resolution based on a small number of genes can identify all the ecotypes. Finding that higher resolution yields no more ecotypes indicates that only a small number of ecotypes exist. Illustrated here is the Nano-Niche model, where there are various ecologically distinct lineages forming ephemeral ecotypes within each ecotype, but there are few ecotypes. In the Nano-Niche model, a periodic selection event within an ecotype can extinguish the ephemeral ecotypes (each a distinct color) within the ecotype; periodic selection can thus purge the diversity within an entire ecotype. The dashed lines illustrate extinctions of ephemeral ecotypes caused by periodic selection. (B) Under speedy speciation, higher resolutions yield more ecotypes. Finding that each increase in resolution yields more ecotypes (up through the resolution of the entire core genome) indicates speedy speciation, and that we do not have the resolution to discover all the ecotypes. Our results with *Bacillus spizizenii* indicate that this species has only five ecotypes because molecular resolutions based on 1—700 genes unanimously yielded five ecotypes; two other *Bacillus* species yielded similar results (Wang, 2024).

of cohesive ecotypes within a species taxon that can coexist into the indefinite future is extremely small. We suspect that the Bacilli have among the most ecotypes within species taxa because these bacteria are generalist heterotrophs and are therefore metabolically plastic (Cohan, 2016a).

Protocol for classifying ecotypes. We suggest that systematics should recognize the infraspecific ecotypes in addition to the broadly defined, ecologically heterogeneous amalgams currently recognized as species-level taxa. To this end, we lay out guidelines for selecting and classifying infraspecific ecotypes that systematists might have the motivation and confidence to recognize.

First, we should employ sequence data to identify the most closely related, ecologically distinct lineages that have coexisted owing to their ecological distinctness. These infraspecific ecotypes could be hypothesized by any of various universal, sequence-based methods, including AdaptML (Hunt et al., 2008), Ecotype Simulation (Francisco et al., 2014; Wood et al., 2021), GMYC (Barraclough et al., 2009), BAPS (Corander et al., 2008), and Minimum Entropy Decomposition (Eren et al., 2015). We suggest that when possible, such analyses should be based on many genes, ideally the entire set of shared genes in the genome, to resolve some of the most newly divergent ecotypes.

Second, to insure that we have used enough molecular resolution to discover all the ecotypes, we recommend employing a rarefaction approach (Wang, 2024). Provided that the number of ecotypes does not increase as more genes are included, one can be confident that all ecotypes have been identified.

Third, in keeping with an important tradition of bacterial systematics, ecotypes should be confirmed to be phenotypically distinct (Rosselló-Móra and Amann, 2001), and we add that ideally the phenotypic differences should confer the ecological niche specificity of the ecotypes.

Fourth (and related to phenotypic differences, above), the most-closely related ecotypes should be confirmed to be ecologically distinct from one another. This is important because in the Geotype plus Boeing model, sequence clusters do not necessarily reflect ecological diversification. Ecological distinctness could be confirmed by differences in habitat association, differences in physiology, or differences in history of positive selection. Inferring phenotype from genomes should help streamline the completionist agenda at the ecotype level.

Fifth, following the traditions of bacterial systematics and the newly developed SeqCode (Whitman et al., 2022), each newly described ecotype should have a designated type strain or sequence.

Sixth, we abide by the preferences of systematists to describe only those infraspecific taxa for which there is a "sound rationale" for taxonomic recognition (Chun et al., 2018), but we suggest relaxing what taxa might reach this criterion. Most systematists might limit descriptions to those infraspecific lineages with properties that affect human well-being in some way or are otherwise notable. However, in the interest of a completionist description of the bacteria's ecological diversity, we urge description of any ecotype at the point that we have identified what is ecologically distinct about it. We are encouraged by our rarefaction analysis on *Bacillus spizizenii* that the number of ecotypes with a history of long coexistence may be small within many species taxa (Wang, 2024).

Let us next consider how we might follow these guidelines to incorporate ecotypes into the systematics of bacteria. When a recognized, legacy species is found to contain multiple ecotypes, such as is the case for *Bacillus simplex* (Koeppel et al., 2008), *Vibrio splendidus* (Hunt et al., 2008), and probably many cases where sequence clusters within a pathogenic species are known to differ in host range and/or tissue tropism (Gordon and Cowling, 2003; Smith et al., 2006a; Walk et al., 2007, 2009), we suggest adding a trinomial "ecovar" epithet to describe the ecotype taxon. For example, an

oak-forest-associated and a grassland-associated ecotype within *Bacillus simplex*, from a canyon near Haifa, Israel (Koeppel et al., 2008), might be named *B. simplex* ecovar Alon and *B. simplex* ecovar Esev (based on the Hebrew words for oak and grass).

We believe that the approach we have laid out is pragmatic both for systematists and for those whose work would benefit from a full accounting of the ecological diversity among close relatives. The proposed system is pragmatic because it identifies the likely ecotypes through universally available and applicable techniques of genomics and DNA sequencing, as well as computer algorithms to recognize the ecotypes from sequence diversity patterns. Moreover, it does not reify heterogeneous groups by attempting to apply a universal molecular criterion to all bacterial species. Microbiologists outside of systematists would benefit from a systematics that would recognize the most recent products of bacterial diversification. Perhaps most importantly, we will eventually reach a completionist knowledge of the unique ecological roles played by each member of a vast and diverse microbial community.

7. Rapprochement Between Theory-based and Tradition-based Systematics of Bacteria

Here we have laid out a plan for discovering, characterizing, and classifying ecotypes within species taxa. For those ecotypes whose diversity is recurrently purged of diversity by periodic selection (i.e., excluding the Species-Less model), we have argued that the ecotypes abide by the quintessential property of species. That is, the ecotypes are each a cohesive unit that can diverge indefinitely from other ecotypes (Cohan, 2013). Moreover, with genomic tools, we can aim to automate discovery and classification of the whole of ecotype diversification within species to a completionist level.

Surprisingly, the property of cohesion appears to apply also to the bacterial species taxa. This is a remarkable prospect, given that each species taxon is heterogeneous in its physiology, ecology, and genome content (Cohan, 2016b), and that each species taxon appears to abound with ecotypes, each cohesive (Becraft et al., 2015; Connor et al., 2010).

Nevertheless, early evidence of cohesion within but not between species taxa emerged from metagenomic, "tiling" studies. In tiling, the genome sequence of one isolate from an environment is compared to random sequence reads from the same environment. For example, when my colleagues and I compared the genome sequences of each of two *Synechococcus* isolates from a Yellowstone hot spring to the metagenome reads from the same hot spring, the environmental reads were either nearly identical to the isolate (>95% identical) or they were much more divergent (usually <80%) (Bhaya et al., 2007). Other tiling experiments have yielded similar results (Caro-Quintero and Konstantinidis, 2012; Konstantinidis and DeLong, 2008).

Pairwise genome sequence comparisons have yielded the same pattern. In the first ANI analysis, Konstantinidis and colleagues found that the recognized species taxa of bacterial systematics generally contained organisms within 95% identity of one another, while the most closely related species were generally divergent at less than 83% identity (Konstantinidis and Tiedje, 2005). A much more extensive set of 90,000 genomes revealed the same pattern (Jain et al., 2018). A more recent study, using both cultivated and uncultivated genomes, has confirmed this universal gap in ANI, where many close relatives span values in the range of 95%−100%, and extremely few pairs of organisms are found between 83% and 95% ANI (Olm et al., 2020).

Matt Olm and colleagues rated several molecules used in species demarcation (various single genes vs. whole genomes) for their ability to demarcate the recognized species taxa, and they found the gap between species in ANI scores the most reliable (Olm et al., 2020). Over all species pairs tested, an ANI value of 94.5% yielded the highest delineation of species of any molecular marker, far greater than for any single gene.

Nevertheless, Murray and colleagues have argued that a single value of divergence over all bacteria was too much to expect, given the disparity of the bacterial kingdom in ecology, recombination, and biochemistry (Murray et al., 2021). Indeed, systematists have acknowledged an uncertainty around the ANI cutoff around 93%–96% (Murray et al., 2021; Rosselló-Móra and Amann, 2015). Murray et al. reevaluated the 90K genomes analyzed by Jain, and found that species vary considerably in their levels of diversity, as measured by ANI scores. They found that, while many species are limited to ANI scores over 95%, there are also many species whose divergence extends below that (Murray et al., 2021). In some cases, there are clusters of ANI divergence in a range far below 95%. For example, *Campylobacter jejuni* ANI values extend down to 95%, but then there is an additional cluster showing 91% and another at 85% divergence from other species members. ANI is likely suggesting that, if we were to follow the 95% guideline, we could split this and similar species into taxa that cluster like most species.

While the species taxa vary by the level of divergence at which we obtain ANI clusters, most species taxa do show a gap in ANI between species (Rodriguez-R et al., 2021). This suggests that the species taxa are each subject to cohesion, with some unidentified force limiting divergence within but not (or to a much lesser extent) between them (Cohan, 2019).

One such universal force may be recombination among species for their homologous, functionally interchangeable genes. In genomic analyses of recombination rates in thousands of fully sequenced genomes, Bobay and Ochman found that recombination declines to near zero for genomes from different named species (Bobay and Ochman, 2017), and Olm and colleagues found a similar result from metagenomic data (Olm et al., 2020). The dynamics of homologous recombination yield an exponential decay in the rate of recombination (Costechareyre et al., 2009; Majewski, 2001; Majewski and Cohan, 1999b), and this may contribute to the ANI gap at 95% (Cohan, 2021).

Alternatively, natural selection might contribute to homogenization of genomes at ANI values greater than 95% (within a species taxon). Among such close relatives, generally adaptive genes might pass from one ecotype to another and could thereby provide their adaptation to all the ecotypes within a species taxon (Cohan, 2016a, 2019). Following the Adapt Globally Act Locally model, these small-segment sweeps could tend to homogenize the genomes among close relatives. However, this model does not seem to explain homogeneity at more than a few tens of thousands of base pairs (Arevalo et al., 2019); moreover, recombination tends to occur in so-called "hot spots," reducing the likelihood of genome-wide cohesion by transfer of adaptations (Bobay and Ochman, 2017).

We see a rapprochement between systematists who are focused on the infraspecific ecotypes and those focused on the broader species taxa. Ecotype-level systematists are deliberately aiming to find the most newly divergent, ecologically distinct lineages where periodic selection is imposing genome-wide cohesion. In contrast, species-level systematists have traditionally not focused on the evolutionary dynamics of their charges. Fortuitously, however, the species sought by species-level systematists appear to correspond to groups abiding by a different kind of cohesion. Recombination, whether passively homogenizing most genes, or whether facilitating the spread of generally adaptive genes across a species' ecotypes, appears to limit the ANI divergence within but not between

species. How unexpected and auspicious that systematists from these two traditions share a pursuit of taxa with the property of cohesion!

To conclude, we have the rules and technology in place to fully yield a completionist account of diversity of cohesive units, at and below the species taxon level. Systematics has been derided as a postage-stamp collecting activity (Sutcliffe et al., 2021), but we see an appealing postage stamp metaphor if species share properties outside our esthetics. If species are real, cohesive units, we could liken the process of systematics to the work of stamp collectors. The philatelists start with a preprinted stamp collection album, and they aim to fill in all the blanks with stamps that they collect. If species share a reality of cohesion, then our charge is to fill in species descriptions into a preexisting book with empty slots that represent the species that we have not yet discovered. The slots are predetermined as the world's cohesive units before any systematist enters the arena.

Acknowledgments

We thank William Cohan for introducing us to the video gaming concept of completionism. This work was sponsored by research funds from the Huffington Foundation, the John and Rosemarie Dooley Foundation, and Wesleyan University.

References

Aboal, M., Werner, O., García-Fernández, M.E., Palazón, J.A., Cristóbal, J.C., Williams, W., 2016. Should ecomorphs be conserved? The case of *Nostoc flagelliforme*, an endangered extremophile cyanobacteria. J. Nat. Conserv. 30, 52−64.

Achtman, M., Wagner, M., 2008. Microbial diversity and the genetic nature of microbial species. Nat. Rev. Microbiol. 6, 431−440.

Allewalt, J.P., Bateson, M.M., Revsbech, N.P., Slack, K., Ward, D.M., 2006. Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the octopus spring microbial mat community of Yellowstone National Park. Appl. Environ. Microbiol. 72, 544−550.

Alneberg, J., Karlsson, C.M.G., Divne, A.-M., Bergin, C., Homa, F., Lindh, M.V., Hugerth, L.W., Ettema, T.J.G., Bertilsson, S., Andersson, A.F., Pinhassi, J., 2018. Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. Microbiome 6, 173.

Apuzzo, M., Gebrekidan, S., Kirkpatrick, D.K., 2020. How the World Missed Covid-19's Silent Spread. New York Times.

Arevalo, P., Vanlnsberghe, D., Elsherbini, J., Gore, J., Polz, M.F., 2019. A reverse ecology approach based on a biological definition of microbial populations. Cell 178, 820.

Atwood, K.C., Schneider, L.K., Ryan, F.J., 1951. Periodic selection in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 37, 146−155.

Barberán, A., Caceres Velazquez, H., Jones, S., Fierer, N., 2017. Hiding in plain sight: mining bacterial species records for phenotypic trait information. mSphere 2, e00237-17.

Barraclough, T.G., Hughes, M., Ashford-Hodges, N., Fujisawa, T., 2009. Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data. Biol. Lett. 5, 425−428.

Becker, N.S., Margos, G., Blum, H., Krebs, S., Graf, A., Lane, R.S., Castillo-Ramírez, S., Sing, A., Fingerle, V., 2016. Recurrent evolution of host and vector association in bacteria of the *Borrelia burgdorferi* sensu lato species complex. BMC Genom. 17, 734.

Becraft, E.D., Wood, J.M., Rusch, D.B., Kühl, M., Jensen, S.I., Bryant, D.A., Roberts, D.W., Cohan, F.M., Ward, D.M., 2015. The molecular dimension of microbial species: 1. Ecological distinctions among, and homogeneity within, putative ecotypes of *Synechococcus* inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. Front. Microbiol. 6, 590.

Bendall, M.L., Stevens, S.L., Chan, L.K., Malfatti, S., Schwientek, P., Tremblay, J., Schackwitz, W., Martin, J., Pati, A., Bushnell, B., Froula, J., Kang, D., Tringe, S.G., Bertilsson, S., Moran, M.A., Shade, A., Newton, R.J., McMahon, K.D., Malmstrom, R.R., 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J. 10, 1589−1601.

Bessen, D.E., 2009. Population biology of the human restricted pathogen, *Streptococcus pyogenes*. Infect. Genet. Evol. 9, 581−593.

Bhaya, D., Grossman, A.R., Steunou, A.S., Khuri, N., Cohan, F.M., Hamamura, N., Melendrez, M.C., Bateson, M.M., Ward, D.M., Heidelberg, J.F., 2007. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J. 1, 703−713.

Blount, Z.D., Barrick, J.E., Davidson, C.J., Lenski, R.E., 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489, 513−518.

Bobay, L.-M., Ochman, H., 2017. Biological species are universal across life's domains. Genom. Biol. Evol. 9, 491−501.

Brisson, D., Dykhuizen, D.E., 2004. *ospC* diversity in *Borrelia burgdorferi*: different hosts are different niches. Genetics 168, 713−722.

Cadillo-Quiroz, H., Didelot, X., Held, N.L., Herrera, A., Darling, A., Reno, M.L., Krause, D.J., Whitaker, R.J., 2012. Patterns of gene flow define species of thermophilic Archaea. PLoS Biol. 10, e1001265.

Caro-Quintero, A., Konstantinidis, K.T., 2012. Bacterial species may exist, metagenomics reveal. Environ. Microbiol. 14, 347−355.

Carroll, D., Daszak, P., Wolfe, N.D., Gao, G.F., Morel, C.M., Morzaria, S., Pablos-Mendez, A., Tomori, O., Mazet, J.A.K., 2018. The global virome project. Science 359, 872−874.

Chriswell, M., 2022. Newly Discovered Species of Bacteria in the Microbiome May Be a Culprit Behind Rheumatoid Arthritis. The Conversation.

Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D.R., da Costa, M.S., Rooney, A.P., Yi, H., Xu, X.-W., De Meyer, S., Trujillo, M.E., 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int. J. Syst. Evol. Microbiol. 68, 461−466.

Chun, J., Rainey, F.A., 2014. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int. J. Syst. Evol. Microbiol. 64, 316−324.

Cohan, F.M., 1994. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. Am. Nat. 143, 965−986.

Cohan, F.M., 2002. What are bacterial species? Annu. Rev. Microbiol. 56, 457−487.

Cohan, F.M., 2005. Periodic selection and ecological diversity in bacteria. In: Nurminsky, D. (Ed.), Selective Sweep. Landes Bioscience, Georgetown, Texas, pp. 78−93.

Cohan, F.M., 2011a. Are species cohesive?–A view from bacteriology. In: Walk, S., Feng, P. (Eds.), Bacterial Population Genetics: A Tribute to Thomas S. Whittam. American Society for Microbiology Press, Washington, DC, pp. 43−65.

Cohan, F.M., 2011b. Koufax's Perfect Game—The Tale of the Data. Los Angeles Times. Tribune Newspapers, Los Angeles.

Cohan, F.M., 2013. Species. In: Maloy, S., Hughes, K. (Eds.), Brenner's Encyclopedia of Genetics, second ed. Elsevier, Amsterdam, pp. 506−511.

Cohan, F.M., 2016a. Bacterial speciation: genetic sweeps in bacterial species. Curr. Biol. 26, R112−R115.

Cohan, F.M., 2016b. Bacterial species concepts. In: Kliman, R.M. (Ed.), Encyclopedia of Evolutionary Biology. Academic Press, Oxford, pp. 119−129.

Cohan, F.M., 2017. Species. In: Osmanaj, B., Escalante Santos, L. (Eds.), Reference Module in Life Sciences. Elsevier.

Cohan, F.M., 2019. Systematics: the cohesive nature of bacterial species taxa. Curr. Biol. 29, R169−R172.

Cohan, F.M., 2021. Genomes reveal the cohesiveness of bacterial species taxa and provide a path towards describing all of bacterial diversity. In: Bridge, P., Smith, D., Stackebrandt, E. (Eds.), Trends in the Systematics of Bacteria and Fungi.

Cohan, F.M., Koeppel, A., Krizanc, D., 2006. Sequence-based discovery of ecological diversity within *Legionella*. In: Cianciotto, N.P., Abu Kwaik, Y., Edelstein, P.H., Fields, B.S., Geary, D.F., Harrison, T.G., Joseph, C., Ratcliff, R.M., Stout, J.E., Swanson, M.S. (Eds.), *Legionella*: State of the Art 30 Years after its Recognition. ASM Press, Washington, D.C., pp. 367−376

Cohan, F.M., Koeppel, A.F., 2008. The origins of ecological diversity in prokaryotes. Curr. Biol. 18, R1024−R1034.

Cohan, F.M., Perry, E.B., 2007. A systematics for discovering the fundamental units of bacterial diversity. Curr. Biol. 17, R373−R386.

Connor, N., Sikorski, J., Rooney, A.P., Kopac, S., Koeppel, A.F., Burger, A., Cole, S.G., Perry, E.B., Krizanc, D., Field, N.C., Slaton, M., Cohan, F.M., 2010. The ecology of speciation in *Bacillus*. Appl. Environ. Microbiol. 76, 1349−1358.

Corander, J., Marttinen, P., Siren, J., Tang, J., 2008. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinf. 9, 539.

Costechareyre, D., Bertolla, F., Nesme, X., 2009. Homologous recombination in Agrobacterium: potential implications for the genomic species concept in bacteria. Mol. Biol. Evol. 26, 167−176.

Darwin, C., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London.

de Queiroz, K., 2005. Ernst Mayr and the modern concept of species. Proc. Natl. Acad. Sci. U.S.A. 102 (Suppl. 1), 6600−6607.

DelGaudio, N., 2019. A Genomic Perspective on Ecological Diversification in *Bacillus* in Death Valley National Park Biology. Wesleyan University, Middletown.

dictionary.com, 2022. Completionist, dictionary.com Unabridged. Random House.

Dobzhansky, T., 1951. Genetics and the Origin of Species, third ed. Columbia Univ. Press, New York.

Doolittle, W.F., 2009. Eradicating typological thinking in prokaryotic systematics and evolution. Cold Spring Harbor Symp. Quant. Biol. 74, 197−204.

Dore, H., Farrant, G.K., Guyet, U., Haguait, J., Humily, F., Ratin, M., Pitt, F.D., Ostrowski, M., Six, C., Brillet-Gueguen, L., Hoebeke, M., Bisch, A., Le Corguille, G., Corre, E., Labadie, K., Aury, J.M., Wincker, P., Choi, D.H., Noh, J.H., Eveillard, D., Scanlan, D.J., Partensky, F., Garczarek, L., 2020. Evolutionary mechanisms of long-term genome diversification associated with niche partitioning in marine picocyanobacteria. Front. Microbiol. 11, 567431.

Doroghazi, J.R., Buckley, D.H., 2014. Intraspecies comparison of *Streptomyces pratensis* genomes reveals high levels of recombination and gene conservation between strains of disparate geographic origin. BMC Genom. 15, 1−14.

Dugat, T., Lagrée, A.-C., Maillard, R., Boulouis, H.-J., Haddad, N., 2015. Opening the black box of *Anaplasma phagocytophilum* diversity: current situation and future perspectives. Front. Cell. Infect. Microbiol. 5.

Dykhuizen, D.E., Green, L., 1991. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. 173, 7257−7268.

Eppley, J.M., Tyson, G.W., Getz, W.M., Banfield, J.F., 2007. Genetic exchange across a species boundary in the archaeal genus ferroplasma. Genetics 177, 407−416.

Eren, A.M., Morrison, H.G., Lescault, P.J., Reveillaud, J., Vineis, J.H., Sogin, M.L., 2015. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. ISME J. 9, 968−979.

Feldgarden, M., Byrd, N., Cohan, F.M., 2003. Gradual evolution in bacteria: evidence from *Bacillus* systematics. Microbiology 149, 3565−3573.

Fisher, M.L., Castillo, C., Mecsas, J., 2007. Intranasal inoculation of mice with *Yersinia pseudotuberculosis* causes a lethal lung infection that is dependent on Yersinia outer proteins and PhoP. Infect. Immun. 75, 429−442.

Francisco, J.C., Cohan, F.M., Krizanc, D., 2014. Accuracy and efficiency of algorithms for the demarcation of bacterial ecotypes from DNA sequence data. Int. J. Bioinf. Res. Appl. 10, 409−425.

Frazão, N., Konrad, A., Amicone, M., Seixas, E., Güleresi, D., Lässig, M., Gordo, I., 2022. Two modes of evolution shape bacterial strain diversity in the mammalian gut for thousands of generations. Nat. Commun. 13, 5604.

Garrity, G., Lyons, C., 2011. Systems and Methods for Resolving Ambiguity Between Names and Entities. Board of Trustees of Michigan State University.

Garrity, G.M., 2016. A new genomics-driven taxonomy of bacteria and archaea: are we there yet? J. Clin. Microbiol. 54, 1956−1963.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., Swings, J., 2005. Opinion: re-evaluating prokaryotic species. Nat. Rev. Microbiol. 3, 733−739.

Gilbert, J.A., Jansson, J.K., Knight, R., 2014. The Earth Microbiome project: successes and aspirations. BMC Biol. 12, 69.

Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W., Penzar, D., Perlman, S., Poon, L.L.M., Samborskiy, D.V., Sidorov, I.A., Sola, I., Ziebuhr, J., Coronaviridae Study Group of the International Committee on Taxonomy of, V., 2020. The species *Severe acute respiratory syndrome-related coronavirus*: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. 39, 2324.

Gordon, D.M., Cowling, A., 2003. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. Microbiology 149, 3575−3586.

Gordon, D.M., Lee, J., 1999. The genetic structure of enteric bacteria from Australian mammals. Microbiology 145 (Pt 10), 2673−2682.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M., 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81−91.

Hahne, H., Mader, U., Otto, A., Bonn, F., Steil, L., Bremer, E., Hecker, M., Becher, D., 2010. A comprehensive proteomics and transcriptomics analysis of *Bacillus subtilis* salt stress adaptation. J. Bacteriol. 192, 870−882.

Haldane, J.B.S., 1932. The Causes of Evolution. Longmans, Green, and Co., London.

Hanage, W.P., Fraser, C., Spratt, B.G., 2006. Sequences, sequence clusters and bacterial species. Phil. Trans. Roy. Soc. Ser. B. 361, 1917−1927.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., Butterfield, C.N., Hernsdorf, A.W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D.A., Finstad, K.M., Amundson, R., Thomas, B.C., Banfield, J.F., 2016. A new view of the tree of life. Nat. Microbiol. 1, 16048.

Hunt, D.E., David, L.A., Gevers, D., Preheim, S.P., Alm, E.J., Polz, M.F., 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science 320, 1081−1085.

Hutchinson, G.E., 1968. When are species necessary? In: Lewontin, R.C. (Ed.), Population Biology and Evolution. Syracuse University Press, Syracuse, pp. 177−186.

Jabalia, N., Kumar, A., Kumar, V., Rani, R., 2021. In silico approach in drug design and drug discovery: an update. In: Singh, S.K. (Ed.), Innovations and Implementations of Computer Aided Drug Discovery Strategies in Rational Drug Design. Springer Singapore, Singapore, pp. 245–271.

Jain, C., Rodriguez, R.L., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9, 5114.

Jensen, P.R., 2010. Linking species concepts to natural product discovery in the post-genomic era. J. Ind. Microbiol. Biotechnol. 37, 219–224.

Jorge, F., Brealey, J.C., Brindley, P.J., Buysse, M., Cantacessi, C., Duron, O., Fichorova, R., Fitzpatrick, C.R., Hahn, M., Hunter, C., Hervé, V., Knoll, L.J., Kohl, K.D., Lalle, M., Lukeš, J., Martínez, J.M., Perkins, S.L., Poulin, R., Rosario, K., Schneider, A.C., Schriml, L.M., Thompson, L.R., Walls, R.L., Dheilly, N.M., 2022. MIxS-SA: a MIxS extension defining the minimum information standard for sequence data from symbiont-associated micro-organisms. ISME Commun. 2, 9.

Jorth, P., Staudinger, B.J., Wu, X., Hisert, K.B., Hayden, H., Garudathri, J., Harding, C.L., Radey, M.C., Rezayat, A., Bautista, G., Berrington, W.R., Goddard, A.F., Zheng, C., Angermeyer, A., Brittnacher, M.J., Kitzman, J., Shendure, J., Fligner, C.L., Mittler, J., Aitken, M.L., Manoil, C., Bruce, J.E., Yahr, T.L., Singh, P.K., 2015. Regional isolation drives bacterial diversification within cystic fibrosis lungs. Cell Host Microbe 18, 307–319.

Karaoz, U., Brodie, E.L., 2022. microTrait: a toolset for a trait-based representation of microbial genomes. Front. Bioinform. 2.

Karlsen, S.T., Rau, M.H., Sánchez, B.J., Jensen, K., Zeidan, A.A., 2023. From genotype to phenotype: computational approaches for inferring microbial traits relevant to the food industry. FEMS Microbiol. Rev. 47.

Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., Ding, H., Marttinen, P., Malmstrom, R.R., Stocker, R., Follows, M.J., Stepanauskas, R., Chisholm, S.W., 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild *Prochlorococcus*. Science 344, 416–420.

Kettler, G.C., Martiny, A.C., Huang, K., Zucker, J., Coleman, M.L., Rodrigue, S., Chen, F., Lapidus, A., Ferriera, S., Johnson, J., Steglich, C., Church, G.M., Richardson, P., Chisholm, S.W., 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. PLoS Genet. 3, e231.

Koch, H., Germscheid, N., Freese, H.M., Noriega-Ortega, B., Lücking, D., Berger, M., Qiu, G., Marzinelli, E.M., Campbell, A.H., Steinberg, P.D., Overmann, J., Dittmar, T., Simon, M., Wietz, M., 2020. Genomic, metabolic and phenotypic variability shapes ecological differentiation and intraspecies interactions of Alteromonas macleodii. Sci. Rep. 10, 809.

Koeppel, A., Perry, E.B., Sikorski, J., Krizanc, D., Warner, W.A., Ward, D.M., Rooney, A.P., Brambilla, E., Connor, N., Ratcliff, R.M., Nevo, E., Cohan, F.M., 2008. Identifying the fundamental units of bacterial diversity: a paradigm shift to incorporate ecology into bacterial systematics. Proc. Natl. Acad. Sci. USA 105, 2504–2509.

Koeppel, A.F., Wertheim, J.O., Barone, L., Gentile, N., Krizanc, D., Cohan, F.M., 2013. Speedy speciation in a bacterial microcosm: new species can arise as frequently as adaptations within a species. ISME J. 7, 1080–1091.

Konstantinidis, K.T., DeLong, E.F., 2008. Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. ISME J. 2, 1052–1065.

Konstantinidis, K.T., Rosselló-Móra, R., Amann, R., 2017. Uncultivated microbes in need of their own taxonomy. ISME J. 11, 2399–2406.

Konstantinidis, K.T., Tiedje, J.M., 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U.S.A. 102, 2567–2572.

Kopac, S., Wang, Z., Wiedenbeck, J., Sherry, J., Wu, M., Cohan, F.M., 2014. Genomic heterogeneity and ecological speciation within one subspecies of *Bacillus subtilis*. Appl. Environ. Microbiol. 80, 4842–4853.

Kumar, N., Lad, G., Giuntini, E., Kaye, M.E., Udomwong, P., Shamsani, N.,J., Young, J.P.W., Bailly, X., 2015. Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. Open Biol. 5, 140133.

Lees, J.A., Mai, T.T., Galardini, M., Wheeler, N.E., Horsfield, S.T., Parkhill, J., Corander, J., 2020. Improved prediction of bacterial genotype-phenotype associations using interpretable pangenome-spanning regressions. mBio 11, e01344.

Lefebure, T., Stanhope, M.J., 2007. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol. 8, R71.

Lennon, J.T., Locey, K.J., 2020. More support for Earth's massive microbiome. Biol. Direct 15, 5.

Lewis, W.H., Tahon, G., Geesink, P., Sousa, D.Z., Ettema, T.J.G., 2021. Innovations to culturing the uncultured microbial majority. Nat. Rev. Microbiol. 19, 225−240.

Liu, Y., Lai, Q., Goker, M., Meier-Kolthoff, J.P., Wang, M., Sun, Y., Wang, L., Shao, Z., 2015. Genomic insights into the taxonomic status of the *Bacillus cereus* group. Sci. Rep. 5, 14082.

Locey, K.J., Lennon, J.T., 2016. Scaling laws predict global microbial diversity. Proc. Natl. Acad. Sci. USA 113, 5970−5975.

Lynch, M.D.J., Neufeld, J.D., 2015. Ecology and exploration of the rare biosphere. Nat. Rev. Microbiol. 13, 217−229.

Majewski, J., 2001. Sexual isolation in bacteria. FEMS Microbiol. Lett. 199, 161−169.

Majewski, J., Cohan, F.M., 1999a. Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. Genetics 152, 1459−1474.

Majewski, J., Cohan, F.M., 1999b. DNA sequence similarity requirements for interspecific recombination in Bacillus. Genetics 153, 1525−1533.

Mallet, J., 2008a. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363, 2971−2986.

Mallet, J., 2008b. Mayr's view of Darwin: was Darwin wrong about speciation? Biol. J. Linn. Soc. 95, 3−16.

Manning, S.D., Motiwala, A.S., Springman, A.C., Qi, W., Lacher, D.W., Ouellette, L.M., Mladonicky, J.M., Somsel, P., Rudrik, J.T., Dietrich, S.E., Zhang, W., Swaminathan, B., Alland, D., Whittam, T.S., 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. U.S.A. 105, 4868−4873.

Marri, P.R., Hao, W., Golding, G.B., 2006. Gene gain and gene loss in *Streptococcus*: is it driven by habitat? Mol. Biol. Evol. 23, 2379−2391.

Martiny, A.C., 2019. High proportions of bacteria are culturable across major biomes. ISME J. 13, 2125−2128.

Mayr, E., 1942. Systematics and the Origin of Species From the Viewpoint of a Zoologist. Columbia University, New York.

Mayr, E., 1963. Animal Species and Evolution. Belknap Press of Harvard Univ Press, Cambridge.

Mayr, E., 1982. The Growth of Biological Thought: Diversity, Evolution, and Inheritance. Harvard Univ Press, Cambridge.

Mazzulli, T., Kain, K., Butany, J., 2004. Severe acute respiratory syndrome: overview with an emphasis on the Toronto experience. Arch. Pathol. Lab Med. 128, 1346−1350.

McVean, G., Awadalla, P., Fearnhead, P., 2002. A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160, 1231−1241.

Meier-Kolthoff, J.P., Carbasse, J.S., Peinado-Olarte, R.L., Göker, M., 2021. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res. 50, D801−D807.

Melendrez, M.C., Becraft, E.D., Wood, J.M., Olsen, M.T., Bryant, D.A., Heidelberg, J.F., Rusch, D.B., Cohan, F.M., Ward, D.M., 2016. Recombination does not hinder formation or detection of ecological species of *Synechococcus* inhabiting a hot spring cyanobacterial mat. Front. Microbiol. 6, 1540.

Melendrez, M.C., Lange, R.K., Cohan, F.M., Ward, D.M., 2011. Influence of molecular resolution on sequence-based discovery of ecological diversity among *Synechococcus* populations in an alkaline siliceous hot spring microbial mat. Appl. Environ. Microbiol. 77, 1359−1367.

Meziti, A., Rodriguez-R, L.M., Hatt, J.K., Peña-Gonzalez, A., Levy, K., Konstantinidis, K.T., 2021. The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample. Appl. Environ. Microbiol. 87, e02593.

Michaelides, A., 2019. The Silent Patient. Celadon.

Mirete, S., Mora-Ruiz, M.R., Lamprecht-Grandío, M., de Figueras, C.G., Rosselló-Móra, R., González-Pastor, J.E., 2015. Salt resistance genes revealed by functional metagenomics from brines and moderate-salinity rhizosphere within a hypersaline environment. Front. Microbiol. 6, 1121-1121.

Mukherjee, S., Kuang, Z., Ghosh, S., Detroja, R., Carmi, G., Tripathy, S., Barash, D., Frenkel-Morgenstern, M., Nevo, E., Li, K., 2022. Incipient sympatric speciation and evolution of soil bacteria revealed by metagenomic and structured non-coding RNAs analysis. Biology 11.

Murray, C.S., Gao, Y., Wu, M., 2021. Re-evaluating the evidence for a universal genetic boundary among microbial species. Nat. Commun. 12, 4059.

Murray, R.G., Stackebrandt, E., 1995. Taxonomic note: implementation of the provisional status Candidatus for incompletely described procaryotes. Int. J. Syst. Bacteriol. 45, 186−187.

Nelson, W.C., Maezato, Y., Wu, Y.-W., Romine, M.F., Lindemann, S.R., 2016. Identification and resolution of microdiversity through metagenomic sequencing of parallel consortia. Appl. Environ. Microbiol. 82, 255−267.

Normand, P., Lapierre, P., Tisa, L.S., Gogarten, J.P., Alloisio, N., Bagnarol, E., Bassi, C.A., Berry, A.M., Bickhart, D.M., Choisne, N., Couloux, A., Cournoyer, B., Cruveiller, S., Daubin, V., Demange, N., Francino, M.P., Goltsman, E., Huang, Y., Kopp, O.R., Labarre, L., Lapidus, A., Lavire, C., Marechal, J., Martinez, M., Mastronunzio, J.E., Mullin, B.C., Niemann, J., Pujic, P., Rawnsley, T., Rouy, Z., Schenowitz, C., Sellstedt, A., Tavares, F., Tomkins, J.P., Vallenet, D., Valverde, C., Wall, L.G., Wang, Y., Medigue, C., Benson, D.R., 2007. Genome characteristics of facultatively symbiotic Frankia sp. strains reflect host range and host plant biogeography. Genome Res. 17, 7−15.

Olm, M.R., Crits-Christoph, A., Diamond, S., Lavy, A., Matheus Carnevali, P.B., Banfield, J.F., 2020. Consistent metagenome-derived metrics verify and delineate bacterial species boundaries. mSystems 5, e00731.

Olsen, M.T., Nowack, S., Wood, J.M., Becraft, E.D., LaButti, K., Lipzen, A., Martin, J., Schackwitz, W.S., Rusch, D.B., Cohan, F.M., Bryant, D.A., Ward, D.M., 2015. The molecular dimension of microbial species: 3. Comparative genomics of *Synechococcus* isolates with different light responses and in situ diel transcription patterns of associated putative ecotypes in the Mushroom Spring microbial mat. Front. Microbiol. 6, 604.

Pallen, M.J., 2021a. Bacterial nomenclature in the era of genomics. New Microb. New Inf. 44, 100942.

Pallen, M.J., 2021b. The status Candidatus for uncultured taxa of Bacteria and Archaea: SWOT analysis. Int. J. Syst. Evol. Microbiol. 71.

Pallen, M.J., Telatin, A., Oren, A., 2021. The next million names for archaea and bacteria. Trends Microbiol. 29, 289−298.

Palys, T., Nakamura, L.K., Cohan, F.M., 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. Int. J. Syst. Bacteriol. 47, 1145−1156.

Papke, P.T., Ward, D.M., 2004. The importance of physical isolation in microbial evolution. FEMS Microbiol. Ecol. 48, 293−303.

Papke, R.T., Zhaxybayeva, O., Feil, E.J., Sommerfeld, K., Muise, D., Doolittle, W.F., 2007. Searching for species in haloarchaea. Proc. Natl. Acad. Sci. U.S.A. 104, 14092−14097.

Parker, C.T., Tindall, B.J., Garrity, G., 2019. International code of nomenclature of prokaryotes. Int. J. Syst. Evol. Microbiol. 69, S1−S111.

Parks, D.H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A.J., Hugenholtz, P., 2019. Selection of Representative Genomes for 24,706 Bacterial and Archaeal Species Clusters Provide a Complete Genome-Based Taxonomy. bioRxiv, 771964.

Parks, D.H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A.J., Hugenholtz, P., 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat. Biotechnol. 38, 1079−1086.

Paul, S., Dutta, A., Bag, S.K., Das, S., Dutta, C., 2010. Distinct, ecotype-specific genome and proteome signatures in the marine cyanobacteria *Prochlorococcus*. BMC Genom. 11, 103.

Polz, M.F., Alm, E.J., Hanage, W.P., 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet. : TIG 29, 170−175.

Pombo, M.A., Ramos, R.N., Zheng, Y., Fei, Z., Martin, G.B., Rosli, H.G., 2019. Transcriptome-based identification and validation of reference genes for plant-bacteria interaction studies using Nicotiana benthamiana. Sci. Rep. 9, 1632.

Pritchard, J.K., Pickrell, J.K., Coop, G., 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. 20, R208−R215.

Proctor, R.A., von Eiff, C., Kahl, B.C., Becker, K., McNamara, P., Herrmann, M., Peters, G., 2006. Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. Nat. Rev. Microbiol. 4, 295−305.

Rainey, P.B., Travisano, M., 1998. Adaptive radiation in a heterogeneous environment. Nature 394, 69−72.

Remonsellez, F., Galleguillos, F., Moreno-Paz, M., Parro, V., Acosta, M., Demergasso, C., 2009. Dynamic of active microorganisms inhabiting a bioleaching industrial heap of low-grade copper sulfide ore monitored by real-time PCR and oligonucleotide prokaryotic acidophile microarray. Microb. Biotechnol. 2, 613−624.

Roberts, M.S., Cohan, F.M., 1995. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. Evolution 49, 1081−1094.

Rodriguez-R, L.M., Jain, C., Conrad, R.E., Aluru, S., Konstantinidis, K.T., 2021. Reply to: "Re-evaluating the evidence for a universal genetic boundary among microbial species". Nat. Commun. 12, 4060.

Rosselló-Móra, R., Amann, R., 2001. The species concept for prokaryotes. FEMS Microbiol. Rev. 25, 39−67.

Rosselló-Móra, R., Amann, R., 2015. Past and future species definitions for Bacteria and Archaea. Syst. Appl. Microbiol. 38, 209−216.

Rothe, C., Schunk, M., Sothmann, P., Bretzel, G., Froeschl, G., Wallrauch, C., Zimmer, T., Thiel, V., Janke, C., Guggemos, W., Seilmaier, M., Drosten, C., Vollmar, P., Zwirglmaier, K., Zange, S., Wölfel, R., Hoelscher, M., 2020. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. N. Engl. J. Med. 382, 970−971.

Saiman, L., Siegel, J., 2004. Infection control in cystic fibrosis. Clin. Microbiol. Rev. 17, 57−71.

Sangwan, N., Xia, F., Gilbert, J.A., 2016. Recovering complete and draft population genomes from metagenome datasets. Microbiome 4, 8.

Shapiro, B.J., Polz, M.F., 2014. Ordering microbial diversity into ecologically and genetically cohesive units. Trends Microbiol. 22, 235−247.

Shapiro, B.J., Polz, M.F., 2015. Microbial speciation. Cold Spring Harbor Perspect. Biol. 7 a018143-a018143.

Smith, N.H., Gordon, S.V., de la Rua-Domenech, R., Clifton-Hadley, R.S., Hewinson, R.G., 2006a. Bottlenecks and broomsticks: the molecular evolution of *Mycobacterium bovis*. Nat. Rev. Microbiol. 4, 670−681.

Smith, N.H., Kremer, K., Inwald, J., Dale, J., Driscoll, J.R., Gordon, S.V., van Soolingen, D., Hewinson, R.G., Smith, J.M., 2006b. Ecotypes of the *Mycobacterium tuberculosis* complex. J. Theor. Biol. 239, 220−225.

Sneath, P., Sokal, R., 1973. Numerical Taxonomy: The Principles and Practice of Numerical Classification. W.H. Freeman, San Francisco.

Stackebrandt, E., Ebers, J., 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiol. Today 33, 152−155.

Staley, J.T., 2006. The bacterial species dilemma and the genomic-phylogenetic species concept. Philos. Trans. R. Soc. Lond. B Biol. Sci. 361, 1899−1909.

Stamos, D.N., 2007. Darwin and the Nature of Species. State University of New York Press, Albany.

Stanojković, A., Skoupý, S., Škaloud, P., Dvořák, P., 2022. High genomic differentiation and limited gene flow indicate recent cryptic speciation within the genus Laspinema (cyanobacteria). Front. Microbiol. 13.

Sun, Y.-C., Jarrett, C.O., Bosio, C.F., Hinnebusch, B.J., 2014. Retracing the evolutionary path that led to flea-borne transmission of *Yersinia pestis*. Cell Host Microbe 15, 578−586.

Sutcliffe, I.C., Dijkshoorn, L., Whitman, W.B., Executive Board, O., 2020. Minutes of the International Committee on Systematics of Prokaryotes online discussion on the proposed use of gene sequences as type for naming of prokaryotes, and outcome of vote. Int. J. Syst. Evol. Microbiol. 70, 4416−4417.

Sutcliffe, I.C., Rosselló-Móra, R., Trujillo, M.E., 2021. Addressing the sublime scale of the microbial world: reconciling an appreciation of microbial diversity with the need to describe species. New Microb. New Inf. 43, 100931.

Takeuchi, N., Cordero, O.X., Koonin, E.V., Kaneko, K., 2015. Gene-specific selective sweeps in bacteria and archaea caused by negative frequency-dependent selection. BMC Biol. 13, 20.

Templeton, A., 1989. The meaning of species and speciation: a genetic perspective. In: Otte, D., Endler, J. (Eds.), Speciation and its Consequences. Sinauer Assoc, Sunderland MA, pp. 3−27.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., Deboy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., Fraser, C.M., 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". Proc. Natl. Acad. Sci. U.S.A. 102, 13950−13955.

Thompson, A.W., Kouba, K., 2019. Differential activity of coexisting prochlorococcus ecotypes. Front. Mar. Sci. 6.

Thompson, C.C., Amaral, G.R., Campeao, M., Edwards, R.A., Polz, M.F., Dutilh, B.E., Ussery, D.W., Sawabe, T., Swings, J., Thompson, F.L., 2014. Microbial taxonomy in the post-genomic era: rebuilding from scratch? Arch. Microbiol. 197, 359−370.

Toll-Riera, M., San Millan, A., Wagner, A., MacLean, R.C., 2016. The genomic basis of evolutionary innovation in *Pseudomonas aeruginosa*. PLoS Genet. 12, e1006005.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi, S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A., Diard, M., Dossat, C., Karoui, M.E., Frapy, E., Garry, L., Ghigo, J.M., Gilles, A.M., Johnson, J., Le Bouguenec, C., Lescat, M., Mangenot, S., Martinez-Jéhanne, V., Matic, I., Nassif, X., Oztas, S., Petit, M.A., Pichon, C., Rouy, Z., Ruf, C.S., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D., Médigue, C., Rocha, E.P., Denamur, E., 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. PLoS Genet. 5, e1000344.

Treves, D.S., Manning, S., Adams, J., 1998. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. Mol. Biol. Evol. 15, 789−797.

Trujillo, M.E., Oren, A., 2018. Avoiding 'salami slicing' in publications describing new prokaryotic taxa. Int. J. Syst. Evol. Microbiol. 68, 977−978.

Vandamme, P., Peeters, C., 2014. Time to revisit polyphasic taxonomy. Antonie Leeuwenhoek 106, 57−65.

Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., Swings, J., 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol. Rev. 60, 407−438.

Vernikos, G.S., Thomson, N.R., Parkhill, J., 2007. Genetic flux over time in the Salmonella lineage. Genome Biol. 8, R100.

Vos, M., 2011. A species concept for bacteria based on adaptive divergence. Trends Microbiol. 19, 1−7.

Vos, M., Didelot, X., 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 3, 199−208.

Walk, S.T., Alm, E.W., Calhoun, L.M., Mladonicky, J.M., Whittam, T.S., 2007. Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. Environ. Microbiol. 9, 2274−2288.

Walk, S.T., Alm, E.W., Gordon, D.M., Ram, J.L., Toranzos, G.A., Tiedje, J.M., Whittam, T.S., 2009. Cryptic lineages of the genus *Escherichia*. Appl. Environ. Microbiol. 75, 6534−6544.

Walters, K.E., Capocchi, J.K., Albright, M.B.N., Hao, Z., Brodie, E.L., Martiny, J.B.H., 2022. Routes and rates of bacterial dispersal impact surface soil microbiome composition and functioning. ISME J. 16, 2295−2304.

Wang, J., 2024. Only a Few Lineages Within Bacterial Species Can Coexist Into the Future. College of Integrative Science. Wesleyan, Middletown, Connecticut.

Wang, P., Ding, L., Zou, C., Zhang, Y., Wang, M., 2022. Rhizosphere element circling, multifunctionality, aboveground productivity and trade-offs are better predicted by rhizosphere rare taxa. Front. Plant Sci. 13.

Ward, D.M., 1998. A natural species concept for prokaryotes. Curr. Opin. Microbiol. 1, 271−277.

Ward, D.M., Bateson, M.M., Ferris, M.J., Kühl, M., Wieland, A., Koeppel, A., Cohan, F.M., 2006. Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. Phil. Trans. Roy. Soc. Ser. B. 361, 1997−2008.

Warner, D.F., Koch, A., Mizrahi, V., 2015. Diversity and disease pathogenesis in *Mycobacterium tuberculosis*. Trends Microbiol. 23, 14−21.

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O., Krichevsky, M.I., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., Trüper, H.G., 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463−464.

Weimann, A., Mooren, K., Frank, J., Pope, P.B., Bremges, A., McHardy, A.C., 2016. From genomes to phenotypes: traitar, the microbial trait analyzer. mSystems 1.

Welch, R.A., Burland, V., Plunkett 3rd, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L., Donnenberg, M.S., Blattner, F.R., 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. 99, 17020−17024.

Whitaker, R.J., Grogan, D.W., Taylor, J.W., 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science 301, 976−978.

Whitman, W.B., 2016. Modest proposals to expand the type material for naming of prokaryotes. Int. J. Syst. Evol. Microbiol. 66, 2108−2112.

Whitman, W.B., Chuvochina, M., Hedlund, B.P., Hugenholtz, P., Konstantinidis, K.T., Murray, A.E., Palmer, M., Parks, D.H., Probst, A.J., Reysenbach, A.L., Rodriguez, R.L., Rosselló-Móra, R., Sutcliffe, I., Venter, S.N., 2022. Development of the SeqCode: a proposed nomenclatural code for uncultivated prokaryotes with DNA sequences as type. Syst. Appl. Microbiol. 45, 126305.

Wiedenbeck, J., Cohan, F.M., 2011. Origins of bacterial diversity through horizontal gene transfer and adaptation to new ecological niches. FEMS Microbiol. Rev. 35, 957−976.

Wielgoss, S., 2022. When maths meets phytoplankton ecology. Nat. Ecol. Evol. 6, 138−139.

Wood, J.M., Becraft, E.M., Cohan, F., Krizanc, D., Ward, D.M., 2021. Ecotype Simulation 2: An Improved Algorithm for Efficiently Demarcating Microbial Species from Large Sequence Datasets. bioRxiv, 2020.2002.2010.940734.

World Health Organization, 2020. Emergency Preparedness, Response: Severe Acute Respiratory Syndrome (SARS).

Xu, S., Li, Z., Huang, Y., Han, L., Che, Y., Hou, X., Li, D., Fan, S., Li, Z., 2021. Whole genome sequencing reveals the genomic diversity, taxonomic classification, and evolutionary relationships of the genus Nocardia. PLoS Neglected Trop. Dis. 15, e0009665.

Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzeby, J., Amann, R., Rosselló-Móra, R., 2014. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat. Rev. Microbiol. 12, 635−645.

Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J.R., Amaral-Zettler, L., Gilbert, J.A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., Birren, B.W., Blaser, M.J., Bonazzi, V., Booth, T., Bork, P., Bushman, F.D., Buttigieg, P.L., Chain, P.S., Charlson, E., Costello, E.K., Huot-Creasy, H., Dawyndt, P., DeSantis, T., Fierer, N., Fuhrman, J.A., Gallery, R.E., Gevers, D., Gibbs, R.A., San Gil, I., Gonzalez, A., Gordon, J.I., Guralnick, R., Hankeln, W., Highlander, S., Hugenholtz, P., Jansson, J., Kau, A.L., Kelley, S.T., Kennedy, J., Knights, D., Koren, O., Kuczynski, J., Kyrpides, N., Larsen, R., Lauber, C.L., Legg, T., Ley, R.E., Lozupone, C.A., Ludwig, W., Lyons, D., Maguire, E., Methe, B.A., Meyer, F., Muegge, B., Nakielny, S., Nelson, K.E., Nemergut, D., Neufeld, J.D., Newbold, L.K., Oliver, A.E., Pace, N.R., Palanisamy, G., Peplies, J., Petrosino, J., Proctor, L., Pruesse, E., Quast, C., Raes, J., Ratnasingham, S., Ravel, J., Relman, D.A., Assunta-Sansone, S., Schloss, P.D., Schriml, L., Sinha, R., Smith, M.I., Sodergren, E., Spo, A., Stombaugh, J., Tiedje, J.M., Ward, D.V., Weinstock, G.M., Wendel, D., White, O., Whiteley, A., Wilke, A., Wortman, J.R., Yatsunenko, T., Glockner, F.O., 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat. Biotechnol. 29, 415−420.

Yoon, C.K., 2009. Naming Nature: The Clash Between Instinct and Science. Norton, New York.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., Shi, Z.-L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270−273.

CHAPTER

Population Structure of Pathogenic Bacteria

3

Xueting Qiu[1], Lavanya Challagundla[2], Madikay Senghore[1], William P. Hanage[1] and D. Ashley Robinson[3]

[1]*Department of Epidemiology, Center for Communicable Disease Dynamics, Harvard T. H. Chan School of Public Health, Boston, MA, United States;* [2]*Department of Pharmacology and Toxicology, University of Mississippi Medical Center, Jackson, MS, United States;* [3]*Department of Cell and Molecular Biology, Center for Immunology and Microbial Research, University of Mississippi Medical Center, Jackson, MS, United States*

1. Introduction

There is no universally accepted definition for a bacterial population. This deficiency might be viewed as a consequence of the long-standing, and sometimes contentious, discussion on the nature of species in the bacterial world (Bapteste and Boucher, 2009; Cohan and Perry, 2007; Doolittle and Papke, 2006; Fraser et al., 2009; Godreuil et al., 2005; Shapiro and Polz, 2015). At present, existing bacterial species labels are more often a reflection of practical necessity and arbitrary thresholds, many of which are rooted in historical methods of phenotypic characterization and without reference to any ecological or evolutionary theory. This is particularly noticeable for many bacterial pathogens that were named after the disease they cause (e.g., *Burkholderia pseudomallei* and *Burkholderia mallei*, *Neisseria meningitidis* and *Neisseria gonorrhoeae*) (Gevers et al., 2005). Resolution of the species "dilemma" and study of population structure are therefore essential as they have practical application to human and veterinary health, agriculture, and biosafety. However, the species concept also embraces more fundamental biological questions that remain unanswered today: What are the characteristics shared between all members of a species? How does the standard taxonomic demarcation of species reflect ecology and niche boundaries? What are the relative contributions of the evolutionary processes that affect a bacterial species? What is the unit of selection during bacterial evolution?

Although what exactly constitutes a bacterial population is debatable, there are some common concepts used to refer to units of population structure. Lineages or sequence clusters, which are isolates grouped together on the basis of genetic similarity, can be defined by phylogenomic approaches reconstructing the tree-like history of isolates or by population genomic approaches assigning isolates to sequence clusters (Lees et al., 2019; Tonkin-Hill et al., 2018, 2019). Both approaches define groups of isolates that provide the units for population structure studies. These groups of isolates tend to replace the earlier term of "strain" Ecotype is another concept distinguishing the ecological niches of bacterial strains. Notably, ecotypes imply a specific biological mechanism for the existence of strains. Ecological adaptation can certainly produce strains, but so can strictly neutral

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00011-2
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

processes (Jeraldo et al., 2012). While ecotype can be an accurate concept for strain behavior, lineage or sequence cluster is easier to define with genomic data alone.

Over the past decade, ample genomic data and efficient analytical tools have become available for bacterial pathogens that enable the identification of population structure and the dissection of evolutionary processes that shape this structure. Historically, algorithms such as BAPS and STRUCTURE have been widely used for sequence cluster assignments on relatively small amounts of genetic data (Cheng et al., 2013), but have struggled with the huge amounts of genomic data now available. In response, approaches like FastBAPS (Tonkin-Hill et al., 2019) and PopPUNK (Population Partitioning Using Nucleotide K-mers) (Lees et al., 2019) have been developed. PopPUNK is of particular interest because it can produce consistent and compatible results from different datasets. In contrast, BAPS, STRUCTURE, and similar approaches produce a locally unique result for each dataset—the exact number of sequence clusters and isolates in each sequence cluster depend on the genetic variation present in the dataset. This algorithmic behavior has been a problem in studies of bacterial population structure reaching back to earlier work on clonal complexes defined by eBURST (Feil et al., 2004). Recently, visualization tools, including Microreact (Argimón et al., 2016) and NextStrain (Hadfield et al., 2018), have also been developed or applied to display bacterial population structure.

We know that the population structure of pathogenic bacteria is a combined effect resulting from many different evolutionary processes, including those that produce genetic variation (i.e., polymorphisms) and those that modulate the frequency of polymorphisms in a population. In this chapter, we elaborate on the different evolutionary processes that shape the population structure of bacterial pathogens, with emphasis on recent findings from large-scale genomic studies. Our discussion attempts to demonstrate how an understanding of bacterial population structure can be translated into practical applications.

2. Recombination and Selection in Bacterial Populations

2.1 Emergence and Persistence of Sequence Clusters

A fundamental question in bacterial speciation is how sequence clusters of closely related strains emerge and persist. These clusters arise despite significant variations in colonization, virulence, transmissibility, and other clinically important phenotypes. For relatively clonal species (i.e., those with low rates of homologous recombination), sequence clusters are characterized by sharp and unambiguous boundaries. Some of the most epidemiologically important pathogens such as *Mycobacterium tuberculosis* and *Yersinia pestis* evolve in such a fashion, with little or no recombination occurring between strains (Godreuil et al., 2010). For highly recombining species, the mechanisms of how sequence clusters pull away from the force imposed by recombination between clusters and subsequently remain stable is not so straightforward. If the recombination rate, measured as the rate at which polymorphisms accumulate through recombination relative to mutation (r/m), varies over time or under spatially and temporally fluctuating selective pressures, this might allow some lineages to form genetically cohesive clusters (Cox et al., 2013).

The probability of recombination is governed by multiple ecological and genetic factors (Mourkas et al., 2022). In some cases, strains coexisting within the same host species remain isolated despite having ample opportunity for genetic exchange. This situation has been observed in *Campylobacter*

jejuni, a gut colonizer of many animal species and a causal agent of gastroenteritis in humans (Sheppard et al., 2014). Despite having a high degree of niche overlap and the ability to readily recombine with each other in vitro, two generalist lineages possess separate gene pools. A cryptic ecological barrier within the host appears to exist between the two lineages, which likely explains the lack of recombination between them. The exact nature of this barrier, however, remains unclear. In other cases, genetic barriers can arise that prevent recombination between strain lineages; this has been reported for *Burkholderia pseudomallei* where lineage-specific restriction−modification systems on mobile genetic elements allow recombination within but not between clades (Nandi et al., 2015).

The ability of some bacterial pathogens to obtain genetic material beyond the bounds of named species suggests that their gene pool is extremely large. This is particularly worrisome because the constant barrage of clinical interventions imposed on pathogenic bacteria is an opportunity for these bacteria to rapidly adapt to the interventions. The pneumococcus, *Streptococcus pneumoniae*, is an excellent example of this adaptability. The pneumococcus exchanges genes with closely related streptococcal species, such as *S. mitis, S. infantis*, and *S. oralis* (Donati et al., 2010). Genomic comparisons between these species indicate that ∼30% of biallelic polymorphisms in the pneumococcus are also polymorphic in *S. mitis*. These polymorphisms include those in pathogenicity genes, and their existence supports the notion that *S. mitis* is a genetic reservoir for the pneumococcus (Sanguinetti et al., 2012). On the other hand, some bacteria appear to be more discriminating in exchanging genes with other species. Homologous recombination between *S. aureus* and *S. epidermidis* is generally limited, but transfer of mobile genetic elements between these species is known to occur (Méric et al., 2015; Smyth et al., 2011).

Hybrid or mosaic sequence clusters (i.e., those intermediate forms resulting from recombination between two species or strains) are not neatly partitioned from other clusters. This phenomenon has been widely observed in *Neisseria* spp. whereby some isolates cannot be unambiguously classified into either *Neisseria meningitidis* or *Neisseria lactamica* because they contain DNA from both species, reflecting a history of recombination (Corander et al., 2012; Hanage et al., 2005). On the other hand, new lineages within species can originate by large-scale recombination, as was first reported in methicillin-resistant *Staphylococcus aureus* (MRSA) where the pandemic ST239 lineage formed from a single large chromosomal recombination between distinct ST8-like and ST30-like parent lineages (Robinson and Enright, 2004). Hence, for species and lineages to remain distinct, barriers to genetic exchange must exist between them.

Bacterial pathogens form and maintain distinct species and sequence clusters when barriers to recombination exist. In many instances, these barriers are leaky or the mechanisms underlying cluster formation may not be apparent. Frequent transfers between sequence clusters (be they species or populations) can be a homogenizing force, whereas rare transfers between clusters can be a diversifying force. The question as to whether these clusters are biologically meaningful must be addressed on a case-by-case basis.

2.2 Heterogeneity in Recombination

Recent population genomic studies of hundreds and even thousands of bacterial pathogens reveal a remarkable level of variation in recombination rates and patterns across very closely related lineages. Hence, this variation adds another layer of complexity that may conflict with blanket definitions of bacterial species and populations. Moreover, this variation may be associated with certain clinically

relevant phenotypes, such as antibiotic resistance. In the pneumococcus, for example, populations with a history of more recombination at unlinked housekeeping loci have been reported to exhibit higher levels of resistance to multiple antibiotics (Hanage et al., 2009).

Recombination rates can vary significantly within a species. The genetically promiscuous pneumococcus exhibits a dramatic range in recombination rates. Early studies using multilocus sequence typing (MLST) detected differences in r/m between capsular serotypes of the same serogroup (Robinson et al., 2002). Recent surveys using genome sequences from numerous serotypes of pneumococci have estimated a range of r/m values from 0.06 to 34.06 (Croucher et al., 2013). Other bacterial pathogens also exhibit such variation. In *Escherichia coli*, different recombination rates emerge at different clustering levels such as phylogenetic group, lineage, and clonal complex (González-González et al., 2013). The causative agent of listeriosis, a serious infection caused by eating food contaminated with *Listeria monocytogenes*, consists of at least four evolutionary lineages. Higher recombination rates are found in lineage II strains, which are widespread in natural and farm environments, and are also commonly isolated from animal listeriosis cases. The higher recombination rates in lineage II may contribute to its adaptation to diverse environments and hosts. In contrast, lineage I, the predominant cause of human listeriosis outbreaks, is largely clonal (den Bakker et al., 2008; Orsi et al., 2011).

Different strains also vary in terms of how often they donate or receive recombined DNA, which may greatly influence the population structure and dynamics of pathogens. These biases, wherein some lineages act as frequent donors while others receive DNA more often, can create highways of gene exchange. In the pneumococcus, certain sporadically occurring, nonencapsulated lineages (e.g., ST1106) appear to recombine more often than others, potentially forming a hub for gene flow that is an important source of genetic diversity for the wider population (Chewapreecha et al., 2014). This characteristic is not simply due to the absence of a polysaccharide capsule, which may act as a physical barrier to entry of exogenous DNA, because other nonencapsulated lineages (e.g., ST344 and ST448) show no significant difference in r/m compared to encapsulated lineages (Croucher et al., 2013; Hilty et al., 2014). In the emerging opportunistic pathogen *Mycobacterium abscessus* species complex, which causes lung diseases in immune-compromised individuals, asymmetrical gene flow occurs among three subspecies; subspecies *M. bolletii* appears to donate more often to *M. massiliense* than the other way around (Sapriel et al., 2016).

Recombination does not occur at the same rates across the genome as "recombination hotspots" are frequently found. In the pneumococcus, the common hotspots contain genes encoding cell surface antigens, such as pneumococcal surface protein A (*pspA*) and pneumococcal surface protein C (*pspC*), and antibiotic resistance, such as penicillin-binding proteins (*pbp2x*, *pbp1a, pbp2b*) and dihydrofolate reductase (*folA*) (Chewapreecha et al., 2014). These hotspots are likely driven by selective pressure due to host immunity and clinical intervention (Chewapreecha et al., 2014). In *C. jejuni* genomes, three recombination hotspots have been identified (Yahara et al., 2014). More than half of the genes in these hotspots related to membrane proteins are crucial for host interaction, colonization, and adhesion to intestinal epithelial cells, and are likely under diversifying selection as a response to the host immune system.

The length of recombined DNA is also observed to differ within a species. Large-scale chromosomal recombination has been detected in several important antimicrobial-resistant pathogens including *S. aureus* ST239 (Gill et al., 2021), *Klebsiella pneumoniae* ST258 (Chen et al., 2014), *Campylobacter coli* ST1150 (Sheppard et al., 2013), and *Streptococcus agalactiae* (Brochet et al., 2008). Although very

large recombination events can occur, the majority are of a much smaller size. Where such asymmetry in length of recombined DNA exists, a given accessory gene or indeed a mobile genetic element is more likely to be lost than gained. This asymmetric recombination has been proposed as an adaptation to preserve the genome as a whole from invading prophage that diminishes the fitness of the organism. Such asymmetric recombination, favoring deletion of accessory loci over insertion, is observed in *S. pneumoniae* (Harrow et al., 2021).

K. pneumoniae (Kp), one of the most important nosocomial pathogens causing multiorgan infections and hospital outbreaks worldwide, has demonstrated that these heterogeneities in recombination can exist simultaneously. A recent study (Comandatore et al., 2019) showed that large recombination events occur with higher frequency in specific Kp lineage pairs and that the prevalence of these events is not evenly distributed across the Kp genome. Moreover, although very large recombination events (>100 kb) occur frequently in certain lineages of Kp, the donor of a successful large recombination must possess genes essential to the recipient strain within the exchanged genomic portion (i.e., genes already present within the replaced portion of the recipient genome) (Comandatore et al., 2019). This pattern limits the genomic regions that can have successful genetic exchange to some hotspots, which indicates that certain lineages are more likely to act as donors. Therefore, the success of a large recombination event depends on gene composition in the exchanged genomic portion. In other words, the gene composition in a length of recombined DNA can be a potential barrier to large recombination events in some pathogenic bacteria.

2.3 The Structure of the Pan-genome of Species and Populations

In general, studies involving whole-genome sequencing (WGS) reinforce the concept that a single strain does not fully represent the diversity of its species. A more profound understanding of the total genetic makeup of a species is essential, and the concept of the pan-genome is helpful in this regard. The full complement of genes in a given species or population under consideration is referred to as the pan-genome, including genes present in all strains (core genome) and only in some strains of a species (accessory, flexible, or dispensable genome). In the following paragraphs, we provide some examples of how the pan-genome may relate to bacterial population structure.

In a classic study of three *E. coli* genomes, each one having a different pathogenic characteristic (nonpathogenic strain K12, uropathogenic strain CFT073, enterohemorrhagic strain EDL933), only 39.2% of the combined set of genes is common to all three strains (Welch et al., 2002). While this backbone of common genes has largely evolved by vertical inheritance, the remaining genes unique to each strain have been acquired by frequent gene gain and loss. More recent genomic studies of *E. coli* indicate a more remarkable level of intraspecific diversity. A study reported that in 186 *E. coli* genomes, the core genome had only 1072 gene families but the pan-genome contained 16,373 gene families (Kaas et al., 2012). Despite the relatively small size of the core genome, the same study also reported that the core genome tree shows a clear demarcation between seven *E. coli* phylogroups, which are mostly recapitulated in the accessory gene presence−absence tree. Pan-genome analysis has also clarified the relationship of *E. coli* and its close relative *Shigella*. The genetic repertoire of the two species is very similar, indicating a common gene pool with no evidence for *Shigella*-specific genes (Gordienko et al., 2013). Hence, *Shigella*, which cannot be justified as being its own genus based on its unique gene content or its paraphyletic location within *E. coli,* is another example of a historical anomaly in bacterial taxonomy.

Many species of pathogenic bacteria can be isolated from environmental sources. Pan-genome analysis is an excellent approach to uncover the genetic differences between clinical and environmental populations, and to understand the evolution of pathogenicity in a species. One example is *Pseudomonas aeruginosa*, an opportunistic pathogen able to thrive in diverse ecological niches and one of the leading causes of nosocomial infections and an infectious agent for cystic fibrosis patients. *P. aeruginosa* has a large core genome, which in one study was estimated at 5233 gene families or ~88% of the average genome (Valot et al., 2015). Remarkably, the gene content is extremely well conserved between environmental (plant, animal, water) and clinical sources. Thus, *P. aeruginosa* may constitute a single population with high phenotypic diversity and the potential to become pathogenic to humans (Grosso-Becerra et al., 2014). This contrasts with the population structure of *Vibrio vulnificus*, a halophilic bacterium typically found in marine environments and also an important cause of fatal bloodstream and skin infections. Phylogenies based on the pan-, core, and accessory genomes showed that *V. vulnificus* has genetically distinct clusters between clinical and environmental sources. The biotype 3 lineage, which caused human outbreaks as early as 1996, evolved as a genetically distinct cluster within the environmental population (Koton et al., 2015).

Pan-genome analysis brings a deeper recognition of the enormous genetic diversity in bacterial pathogens. Simultaneously, it has also indicated that species and population boundaries may not be resolved as quickly and unambiguously as we might have hoped. While the terms pan-, core, and accessory genomes may be useful in population genomic studies, we have to keep in mind that their gene contents are partially a result of the sampling criteria used and may not necessarily reflect the ecology of the population under study. One concept that has been put forward is the existence of a gene pool comprising niche-adaptive genes from the accessory genome (Polz et al., 2013). The dynamics of this gene pool drive population structuring through an initial ecological separation, which reduces the rate of gene flow between nascent populations. Population-specific mutations will subsequently accumulate, thus further reducing gene flow and leading to the formation of distinct, ecologically differentiated sequence clusters (Polz et al., 2013).

2.4 Negative Frequency–Dependent Selection

Natural selection can determine the fate of a genetic variant. In some situations, the direction and intensity of natural selection in bacterial populations will depend on the relative frequencies of the variants (Levin, 1988). A form of natural selection that has been found to structure pneumococcal populations, which may also be important for other bacteria with large accessory genomes, is negative frequency–dependent selection (NFDS). NFDS occurs when the selective value of a given allele or gene (relative to others) is a function of its frequency in the population such that its relative fitness increases up to a point along with its frequency, after which it decreases. That is, rare variants have a selective advantage due to their rarity while common variants are disadvantaged (Ayala and Campbell, 2003). A familiar example of NFDS is the balanced polymorphism at the outer surface protein C (*ospC*) locus in populations of *Borrelia burgdorferi*, the cause of human Lyme disease (Qiu et al., 1997). The population diversity at this locus reveals the hallmark of NFDS, where if it becomes common, the significant host population immunity is stimulated to decrease its proportion (Brisson and Dykhuizen, 2004). NFDS has the potential to maintain selected and linked polymorphisms within populations. In this regard, NFDS is a specific form of balancing selection that has only rarely been studied across genomes of bacterial pathogens (Thomas et al., 2012).

To study NFDS, it is important to have an unbiased sample of the naturally transmitting population, not one biased toward particularly resistant or virulent strains. We are fortunate that several such samples exist for the pneumococcus (Azarian et al., 2018, 2020; Croucher et al., 2015; Mitchell et al., 2019). *S. pneumoniae* is diverse in population structures with more than 100 serotypes scattered across many sequence clusters. *S. pneumoniae* also has a large accessory genome that is enriched for genes with functions such as surface antigens (e.g., virulence-related hair-like structure "pili"), receptors for bacteriophage, and toxin−antitoxin pairs. NFDS was first proposed to have structured pneumococcal populations based on the observation that frequencies of accessory genes were highly correlated in population samples from distant locations on three continents (Corander et al., 2017). This correlation can also be explained by a lack of phylogeographic structure due to high migration rates whereby the same strains were present everywhere in similar proportions. However, the proportions of specific strains in different geographic locations were not correlated. Therefore, the NFDS mechanism retained these highly correlated accessory gene frequencies even in populations with different strains. Further evidence of NFDS in *S. pneumoniae* comes from the comparison of samples collected before and after vaccination (Azarian et al., 2018). After the introduction of the seven-valent pneumococcal conjugate vaccine (PCV7), the carriage prevalence of the pathogen did not change. Instead, serotype replacement occurred, whereby the prevalence of most vaccine serotypes markedly decreased while nonvaccine serotypes increased to take their place. In a 14-year longitudinal sample from one carriage population from the southwest United States, accessory gene frequencies were highly correlated between pre-PCV7 and post-PCV7 time periods. Moreover, after approximately 10 years of PCV7 introduction, in 2010−12, the accessory gene frequencies presented a "bouncing back" trend closer to pre-PCV7 frequencies in 1998 than those in 2006 (Azarian et al., 2018).

This aspect of pneumococcal population structure potentially has an important application in that the composition of the accessory genome can predict which nonvaccine type strains would be most successful following vaccination. A simple prediction model was developed to study how the pneumococcal population evolved via NFDS after PCV7 perturbation (Azarian et al., 2020; Corander et al., 2017). This prediction model posited that each accessory gene was under NFDS, and the pre-PCV7 frequency of each gene was an estimate of the "ideal" or equilibrium frequency at which the presence of the gene was neutral. When the population frequency of this gene increased above equilibrium due to vaccine perturbation, this gene contributed negatively to the fitness of strains with this gene, and when the gene frequency was below equilibrium, this gene contributed positively. This model could accurately predict the dynamics of individual strains post-PCV7 as a selective process in which NFDS changed strain frequencies to produce accessory gene frequencies similar to those observed at the prevaccine equilibrium. This NFDS model made a better prediction of the population structure following vaccination compared with a null model which only accounted for proportional increases or decreases in the strain prevalence after the vaccine removed the vaccine serotypes in the strains (Azarian et al., 2020).

Many consequences of the NFDS process remain unclear; its ability to determine exactly what strains are present in the population and at what proportions has not been systematically assessed. Building simple population genetic models has shown that NFDS, when combined with asymmetric recombination, can produce stable population structures similar to those seen in nature (Harrow et al., 2021). While this phenomenon has been most studied in *S. pneumoniae*, one study also finds that NFDS acting on accessory genes plays a role in the evolution of *E. coli* (McNally et al., 2019). It will be important to focus future work on collecting appropriate samples to test for NFDS in pathogenic

bacteria. Prediction models built upon this evolutionary process may contribute to vaccine designs and public health interventions.

3. Bacterial Population Structure Within and Between Hosts

3.1 Within-host Diversity and Bottlenecks Shape Transmission Dynamics

As with epidemiological study samples, bacterial populations may be defined by the scale of observation. Prior to the availability of low-cost WGS, the scale of observation was most often between-host and used molecular typing methods with poorer resolution. However, the fine-grained typing resolution provided by WGS has allowed investigators to narrow the scale of observation to the level of an individual host. As a result, we have gained a better understanding of population dynamics of bacterial pathogens within and between hosts.

Within-host diversity is a proven phenomenon that has been demonstrated in several bacterial species. Earlier studies using low resolution genotyping techniques like pulsed-field gel electrophoresis and *spa* typing confirmed that a host can be colonized by multiple genotypes of the same species at a given body site; this was seen in multiple species such as *Haemophilus influenzae*, *S. pneumoniae*, and *S. aureus* (Leung et al., 2015; Sauver et al., 2000). However, genotyping techniques were only able to identify co-colonization if the strains belonged to phylogenetically distinct genotypes. While genomic studies overcame limitations in the resolution of genotypes, they remained hindered by the practice of isolating a single bacterial colony and thus a single predominant strain from a specimen. The extent of within-host diversity became apparent through WGS of multiple colonies. Studies of *S. aureus* colonization, for example, have found cohesive populations of bacteria separated by up to ~40 single nucleotide polymorphisms (SNPs) with no apparent subpopulation structure among body sites sampled (Golubchik et al., 2013; Popovich et al., 2016; Young et al., 2012). Variations ranging from 1−40 SNPs of the same strain can coexist among carriage and invasive disease isolates (Tong et al., 2015b). The observed level of within-host diversity can vary because of multiple factors starting with the amount of diversity acquired during a transmission, which is modulated by a genetic bottleneck, followed by microevolution within the host, which is shaped by genetic drift and selective pressures such as functional constraints, host-immunity, and treatment with antibiotics (Didelot et al., 2016).

Longitudinal cohorts that capture cross-sectional snapshots of within-host bacterial populations have been instrumental in showing the range of nucleotide diversity and the mechanisms that generate this diversity. Recent work on *S. pneumoniae* showed rapid adaptation during natural colonization in infants, where neutral evolution through accumulation of mutations was up to 40 times faster compared with what was observed over longer time scales. Moreover, divergent strains emerged through homologous recombination, and parallel evolution was observed in antibiotic resistance genes and immune evasion genes (Chaguza et al., 2020). These observations built on existing knowledge from longitudinal studies of *S. aureus* carriage, which showed significant within-host fluctuations in effective population size (N_e) (Golubchik et al., 2013), and heterogeneity in evolutionary rate and accumulation of mutations preceding the transition from colonization to disease (Young et al., 2012).

The role of within-host diversity in shaping the wider population structure of bacteria remains unclear. However, to fully characterize bacterial transmission, within-host diversity must be considered. Simulation studies have demonstrated that ignoring within-host diversity during epidemic

investigations hinders accurate reconstruction of between-host transmission networks (Worby et al., 2014). Deep sequencing to identify sub-consensus diversity provides added resolution by identifying within-host variants at genomic loci where a fraction of the reads detected an alternative allele. When within-host variants are shared by closely related isolates and are otherwise rare in the population they serve as good indicators of transmission. However, this approach requires the identification of haplotypes and ignores their relative abundance in the host. New deep sequencing approaches for studying transmission chains should devise innovative ways to quantify and measure differences in diversity within-host as a way of discriminating transmission pairs.

Contributions to within-host diversity must take several factors into consideration. Primarily, the duration of colonization and transmission bottleneck size is often unknown in these studies. The size of the bacterial population acquired during a transmission event (i.e., bottleneck size) affects the diversity of the founding within-host population, and assuming constant evolutionary rate, diversity would increase over time. Secondly, bioinformatics pipelines for identifying SNPs in bacterial populations are not standardized and can vary significantly between studies. Lastly, multiple transmission events of a closely related strain are likely within highly connected transmission chains or high transmission settings (e.g., within a household), potentially resulting in inflated within-host diversity estimates. When also considering strain characteristics (e.g., hypermutators and highly recombinant strains), it becomes clear that a single snapshot of the within-host population may not accurately reflect the demographic history of the pathogen. For this reason, a single cut-off value of SNP difference should not be used to distinguish the within-host diversity from the between-host transmission events. Sampling multiple bacterial isolates over time better captures the history of the bacterial population within a host.

Studies of bacterial populations often neglect the individual scale. However, as we have illustrated, recent technological advances in WGS combined with epidemiological studies sampling multiple bacterial isolates from individuals over time have redefined our concept of a bacterial population. Within a single patient that is harmlessly colonized by a bacterial strain or that is being treated for a bacterial infection, we can observe extensive polymorphisms among the bacterial isolates and can infer these to be the result of the gamut of evolutionary processes. In essence, within-host studies serve as a model for gaining a better understanding of population structure on a larger scale.

3.2 Between-host Evolution and Impact on Population Structure

Bacterial populations may possess temporal, geographic, environmental, or host-specific structure. The evolutionary processes shaping the population structure of isolates from different hosts are indeed similar to those acting on within-host populations, and the way in which they do so has long been investigated by researchers in microbiology, epidemiology, phylogenetics, and population genetics. Now, with the widespread application of WGS and an assortment of phylogenetic and population genetic analysis tools, the demographic history and transmission dynamics of bacteria can be studied at an unprecedented level of resolution. For example, recent studies have used these methods to investigate clonal emergence and intercontinental spread (Grad et al., 2014; Harris et al., 2010), transmission within the community and in high-risk contact networks (Alam et al., 2015; Gardy et al., 2011), and to track geographic origin (Frerichs et al., 2012). The challenge is that in high burden endemic settings, most bacterial pathogens are under-sampled, resulting in datasets that fail to capture the full extent of diversity in the population for a sufficient time period; thus, many datasets are unable

to reliably estimate relevant parameters such as the population growth rate and time to the most recent common ancestor.

Bacteria evolve at different rates, and this affects the choice of tools and approaches that can appropriately define and keep track of the population structure over time. Genetically monomorphic bacteria like *M. tuberculosis,* are easily stratified based on lineage-defining mutations because strains typically do not recombine with each other, and their populations have low rates of convergent evolution (Coll et al., 2014; Walker et al., 2013). Mapping the population structure of many other bacterial pathogens is more challenging because they evolve at a faster pace (Didelot et al., 2013; Mathers et al., 2015; Young et al., 2012) and are more prone to recombination and convergent evolution (Croucher et al., 2014; Stott and Bobay, 2020).

Bacteria may evolve differently in controlled environments such as hospitals that house vulnerable populations, and this can be reflected in their population structure. Hospital environments can be rich in bacterial diversity, and selection pressures from the routine use of antibiotics can lead to the emergence of resistant bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA) and other species of major public health concern (Jernigan et al., 2020). It is not surprising that MRSA is a leading cause of hospital infection, given its ability to colonize humans as well as environmental surfaces and a range of animals (Tong et al., 2015a). Hospital-acquired MRSA are genetically distinct and typically spread less efficiently in the community and are less virulent than community-acquired MRSA (Prosperi et al., 2013). The root cause of this difference in transmissibility is unclear, and it is further complicated by the constant introduction of community-associated MRSA into the hospital environment. One hypothesis is that over time hospital-acquired MRSA become better at causing nosocomial infection by adapting to ease of transmission and intense selection pressure from antibiotics in hospitals compared to the community (Ellis et al., 2009; Vysakh and Jeya, 2013). A study of a network of hospitals in California showed that MRSA populations were genetically more similar between hospitals that had a high traffic of patients between them, highlighting the unique transmission dynamics that can exist in hospitals (Chang et al., 2016).

Studies of *Vibrio cholerae* provide an excellent example of what can be learned about between-host evolution and population structure. *V. cholerae* serogroup O1, consisting of "classical" and "El Tor" biotypes, causes epidemic cholera and is the most epidemiologically important lineage in this species (Haley et al., 2014). Molecular clock analysis using time-dated genomes found that the current seventh pandemic of O1 cholera originated in the 1950s in the Bay of Bengal and caused at least three overlapping epidemic waves (Mutreja et al., 2011). It remains unclear why the first of these waves spread globally, while the second and third waves were geographically isolated. However, recombination played a significant role in the diversification of each wave, which were also preceded by considerable population bottlenecks (Mutreja et al., 2011).

More recently in 2010, a cholera epidemic struck the already earthquake-torn country of Haiti, which had been cholera-free for over 100 years (Ivers and Walton, 2012). While this epidemic caused significant morbidity and mortality, it also provided a unique opportunity to study the emergence of *V. cholerae* in a new setting. Phylogenetic analysis was used to complement epidemiological data in order to trace the introduction of cholera to a Nepalese garrison on a United Nations peacekeeping duty in Haiti (Frerichs et al., 2012; Hendriksen et al., 2011). The introduction to Haiti was an example of a substantial population bottleneck, given that Nepalese cholera epidemics were divided into four distinct clades (Hendriksen et al., 2011) and the Haiti isolates formed a monophyletic clade and differed from several Nepalese strains by only one to two SNPs (Azarian et al., 2014; Katz et al., 2013). The epidemic

quickly spread to all regions of the country as well as to the neighboring Dominican Republic, and environmental surveillance subsequently isolated the epidemic clone from several Haitian rivers and estuaries (Alam et al., 2014). In all, these examples demonstrate how WGS and phylogenetic analysis can facilitate the understanding of historical patterns of between-host transmission, the role of recombinations and mutations, and the spatiotemporal origin and spread of a pathogen.

4. Geography and Bacterial Population Structure

4.1 Bacterial Phylogeography and Migration Between Populations

Bacterial genomic data with accompanying spatial and temporal meta-data can be used to understand geographic population structure and to infer origin and spread (Bentley and Parkhill, 2015). Advances in phylogeographic models have provided powerful tools to make such inferences. However, as with other observational studies, phylogeographic studies are impacted by embedded issues of the sample collection process. For example, biased inclusion of geographically related samples can lead to underestimates of migration events between populations. Biased inclusion of genetically related samples can lead to a poorly resolved and "shallow" tree, which may suggest a rapid expansion of the pathogen (Colijn and Gardy, 2014).

Historically, human migration has led to the spread and subsequent establishment of some human-associated pathogens with previously limited geographic distributions such as *M. tuberculosis* (Barbier and Wirth, 2016; Comas et al., 2013; Pepperell et al., 2010) and *Helicobacter pylori* (Falush et al., 2003; Linz et al., 2007). These studies demonstrated that the population structure of the pathogens mirrors the phylogeographical structure of their hosts. However, this association between bacterial population structure and host migration may not always be as straightforward as expected and there may be other ecological barriers that may contribute to the observed population structure. In multiple *Borrelia* species, the migration of both host and vector appears to greatly determine the geographic distribution of the pathogen. In a study of *Borrelia burgdorferi* across the United States it was found that populations of the pathogen clustered geographically, and loosely corresponded to the primary vector ranges, which could suggest adaptive evolution in *B. burgdorferi* to the vector species or neutral evolution as a result of vicariance and genetic drift (Margos et al., 2008, 2012). In another study, limited geographic structuring between countries in populations of *Borrelia* spp. associated with birds (*B. garinii* and *B. valaisiana*) was observed, likely as a result of higher rates of migration of the host, but there existed a strong signal in geographic structure in the bacterium associated with small mammals (*B. afzelii*) (Vollmer et al., 2011). *B. burgdorferi* populations from Europe and the United States isolated from human patients exhibited an overlap of sequence types, which was in contrast to populations found in tick vectors (Jungnick et al., 2015). These findings were reinforced in a more recent study (Norte et al., 2020) of *B. garinii* in ticks associated with birds, where there was no detectable geographic structuring within the populations from Europe or between ticks associated with birds, humans, or questing ticks.

For some bacteria, the spatial isolation of lineages or allopatry can give rise to local variants. For *S. aureus* strain ST239, there is a strong geographic clustering at continental, national, and city scales. At the continental scale, genomic comparisons indicate that three monophyletic clades, which mostly represent European, Asian, and South American populations, reflect multiple independent exports from Europe over only a few decades (Harris et al., 2010; Smyth et al., 2010). On the other hand,

countrywide geographic patterns are consistent with patterns of human movement between cities (Castillo-ramirez et al., 2012). On a finer spatial scale, the populations of nosocomial multidrug-resistant *Enterococcus faecium* from different hospitals have emerged and diversified independently of each other, and adaptation to a specific hospital niche has resulted in genetic isolation and limited gene flow between them (Willems et al., 2012). However, a similar degree of genetic isolation in hospital-adapted populations has not been observed within *Staphylococcus epidermidis* (Thomas et al., 2014). Patient sharing between hospitals is likely to erode the separation of populations, as previously mentioned for MRSA whereby the genetic similarity of isolates from different hospitals correlated with the degree of patient sharing between hospitals (Chang et al., 2016; Ke et al., 2012).

4.2 Range Expansion as a Model of Geographic Spread

The spatial spread of a species by a geographic "range expansion" can profoundly alter genetic diversity and adaptive potential (Excoffier et al., 2009). In nature, these expansions can result from environmental changes (Hewitt, 1999; Schmitt, 2007), from introducing a species into a favorable niche (Lawson Handley et al., 2011), or from evolutionary adaptations that enable a species to expand into a previously unoccupied niche. This process leaves distinct signatures across neutral mutations that are typically associated with selection and distinguishable from those expected in an expansion without spatial structure (Excoffier et al., 2009).

Some of the population genetic signatures of geographic range expansion include a decrease in genetic diversity (Excoffier, 2004) and an increase in derived allele frequencies (Edmonds et al., 2004; Klopfstein et al., 2006; Peter and Slatkin, 2013) between the origin population and the edge populations at the boundaries of the expanding range. These population genetic signatures are the result of sequential bottlenecks that occur during spread from an origin population, which highlights the role of genetic drift in spatially expanding populations. Geographic range expansions have been characterized for numerous species including humans (Cavalli-Sforza and Feldman, 2003; Li et al., 2008; Ramachandran et al., 2005; Tishkoff et al., 2009) and their associated pathogens such as *H. pylori* (Falush et al., 2003; Linz et al., 2007). More recently, the use of WGS data allowed range expansions to be detected at the level of individual bacterial strains. Population genetic signatures were able to identify a geographic origin for the USA300 North American Epidemic MRSA strain in the United States and the LAM4/KZN strain of extensively drug-resistant *M. tuberculosis* in KwaZulu-Natal, South Africa (Brown et al., 2019; Challagundla et al., 2018). Of note, both of these strains had poorly resolved phylogenies potentially due to their recent origin and/or rapid spread, which demonstrates that population genetic study of the geographic spread of bacteria can complement a phylogenetic analysis. Moreover, the USA300 MRSA study found evidence of both neutral (synonymous mutation in *ssa*-1) and potentially non-neutral (*gyrA* mutations conferring high-level fluoroquinolone resistance) mutations that "surfed" to high frequency during the range expansion (Challagundla et al., 2018). Many aspects of geographic range expansion remain to be tested in bacterial populations. This includes predictions about the effects of sequential bottlenecks and the resulting low effective population sizes on bacterial fitness (expansion load; Peischl and Excoffier, 2015), and the potential for genome-level surfing with transient increases in imported DNA (Choudoir et al., 2017).

Due to the dependence of some pathogenic bacteria upon their hosts for spatial dissemination, bacterial populations can be dispersed over long distances creating founding populations well beyond locations proximate to the origin (Becheler et al., 2016). This movement may not be as simple as a

smooth diffusion process but rather can be by "leaps and bounds" involving long distance dispersal (LLD) events (Ibrahim et al., 1996; Slatkin and Excoffier, 2012). These LDD events can create a "spatially incoherent" population structure in which geographic distances poorly correlate with genetic diversity (Brown et al., 2022; Excoffier et al., 2009; Paulose and Hallatschek, 2020). Brown et al. (2022) suggests exploring alternate measures of social connectivity that best tracks host movements between genetically distinct bacterial populations. This measure of social connectivity may vary across bacterial species. For example, *Neisseria gonorrhea* can be spread within certain host networks in cities (Williamson et al., 2019). The spread of gonococcus in a sexual network of men having sex with men suggested multiple transmissions from the west coast to the east coast and was consistent with the introduction/emergence of resistance to Cefixime on the west coast followed by eastward travel in the same sexual network (Grad et al., 2014).

5. Conclusions

Analyses of genome sequences from samples of hundreds and even thousands of bacterial isolates have allowed the precise identification of sequence clusters that sometimes correspond to lineages within species and sometimes may deserve their own species labels. Moreover, these analyses have provided precise estimates of recombination and mutation rates occurring within and between these clusters, the understanding of adaptations driving the associations of bacterial pathogens to their hosts and ecological niches, and the migration and range expansion impacting bacterial population structure. However, evolutionary processes do not act separately when shaping bacterial population structure. What we observe is the results of joint processes. Some processes may be easier to study and estimate relevant parameters; for example, mutation rates that quantify a baseline rate of genetic variations from point mutations. The complexity of horizontal gene transfer, recombination, drift, and selection, occurring in various demographic contexts, makes it difficult to disentangle the relative contributions of these processes to the genetic diversity that we observe. Deciphering complex evolutionary processes that shape bacterial population structure is a difficult task. Nevertheless, by building upon prior findings, excellent data availability, and advanced technical methodologies, we will draw a more complete map to define and explain the population structure of bacterial pathogens. Ultimately, this area of research can provide answers to what underlies the heterogeneity in antimicrobial resistance and invasive disease potential, and can aid in the design of vaccines and improve public health interventions.

Acknowledgments

We thank the prior coauthors of this chapter, including Jonathan C. Thomas, Cheryl P. Andam, and Taj Azarian, for their contributions that remain in this current version of the chapter.

References

Alam, M.T., Read, T.D., Petit, R.A., Boyle-Vavra, S., Miller, L.G., Eells, S.J., Daum, R.S., David, M.Z., 2015. Transmission and microevolution of USA300 MRSA in U.S. households: evidence from whole-genome sequencing. mBio 6. https://doi.org/10.1128/MBIO.00054-15.

Alam, M.T., Weppelmann, T.A., Weber, C.D., Johnson, J.A., Rashid, M.H., Birch, C.S., Brumback, B.A., de Rochars, V.E.M.B., Glenn Morris, J., Ali, A., 2014. Monitoring water sources for environmental reservoirs of toxigenic *Vibrio cholerae* O1, Haiti. Emerg. Infect. Dis. 20, 356. https://doi.org/10.3201/EID2003.131293.

Argimón, S., Abudahab, K., Goater, R.J.E., Fedosejev, A., Bhai, J., Glasner, C., Feil, E.J., Holden, M.T.G., Yeats, C.A., Grundmann, H., Spratt, B.G., Aanensen, D.M., 2016. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb. Genom. 2, e000093. https://doi.org/10.1099/MGEN.0.000093.

Ayala, F.J., Campbell, C.A., 2003. Frequency-dependent selection. Ann. Rev. Ecol. Syst. 5, 115−138. https://doi.org/10.1146/ANNUREV.ES.05.110174.000555.

Azarian, T., Ali, A., Johnson, J.A., Mohr, D., Prosperi, M., Veras, N.M., Jubair, M., Strickland, S.L., Rashid, M.H., Alam, M.T., Weppelmann, T.A., Katz, L.S., Tarr, C.L., Colwell, R.R., Morris, J.G., Salemi, M., 2014. Phylodynamic analysis of clinical and environmental *Vibrio cholerae* isolates from Haiti reveals diversification driven by positive selection. mBio 5. https://doi.org/10.1128/MBIO.01824-14.

Azarian, T., Grant, L.R., Arnold, B.J., Hammitt, L.L., Reid, R., Santosham, M., Weatherholtz, R., Goklish, N., Thompson, C.M., Bentley, S.D., O'Brien, K.L., Hanage, W.P., Lipsitch, M., 2018. The impact of serotype-specific vaccination on phylodynamic parameters of Streptococcus pneumoniae and the pneumococcal pan-genome. PLoS Pathog. 14, e1006966. https://doi.org/10.1371/journal.ppat.1006966.

Azarian, T., Martinez, P.P., Arnold, B.J., Qiu, X., Grant, L.R., Corander, J., Fraser, C., Croucher, N.J., Hammitt, L.L., Reid, R., Santosham, M., Weatherholtz, R.C., Bentley, S.D., O'Brien, K.L., Lipsitch, M., Hanage, W.P., 2020. Frequency-dependent selection can forecast evolution in Streptococcus pneumoniae. PLoS Biol. 18, e3000878. https://doi.org/10.1371/JOURNAL.PBIO.3000878.

Bapteste, E., Boucher, Y., 2009. Epistemological impacts of horizontal gene transfer on classification in microbiology. Methods Mol. Biol. https://doi.org/10.1007/978-1-60327-853-9_4.

Barbier, M., Wirth, T., 2016. The evolutionary history, demography, and spread of the *Mycobacterium tuberculosis* complex. Microbiol. Spectr. 4. https://doi.org/10.1128/MICROBIOLSPEC.TBTB2-0008-2016.

Becheler, R., Xhaard, C., Klein, E.K., Hayden, K.J., Frey, P., de Mita, S., Halkett, F., 2016. Genetic signatures of a range expansion in natura: when clones play leapfrog. Ecol. Evol. 6, 6625−6632. https://doi.org/10.1002/ECE3.2392.

Bentley, S.D., Parkhill, J., 2015. Genomic perspectives on the evolution and spread of bacterial pathogens. Proc. Biol. Sci. 282. https://doi.org/10.1098/RSPB.2015.0488.

Brisson, D., Dykhuizen, D.E., 2004. ospC diversity in Borrelia burgdorferi: different hosts are different niches. Genetics 168, 713. https://doi.org/10.1534/GENETICS.104.028738.

Brochet, M., Rusniok, C., Couvé, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., Kunst, F., Glaser, P., 2008. Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of Streptococcus agalactiae. Proc. Natl. Acad. Sci. U. S. A. 105, 15961−15966. https://doi.org/10.1073/PNAS.0803654105.

Brown, T.S., Challagundla, L., Baugh, E.H., Omar, S.V., Mustaev, A., Auld, S.C., Shah, N.S., Kreiswirth, B.N., Brust, J.C.M., Nelson, K.N., Narechania, A., Kurepina, N., Mlisana, K., Bonneau, R., Eldholm, V., Ismail, N., Kolokotronis, S.O., Robinson, D.A., Gandhi, N.R., Mathema, B., 2019. Pre-detection history of extensively drug-resistant tuberculosis in KwaZulu-Natal, South Africa. Proc. Natl. Acad. Sci. U. S. A. 116, 23284−23291. https://doi.org/10.1073/PNAS.1906636116.

Brown, T.S., Robinson, D.A., Buckee, C.O., Mathema, B., 2022. Connecting the dots: understanding how human mobility shapes TB epidemics. Trends Microbiol. https://doi.org/10.1016/J.TIM.2022.04.005.

Castillo-ramirez, S., Corander, J., Marttinen, P., Aldeljawi, M., Hanage, W.P., Westh, H., Boye, K., Gulay, Z., Bentley, S.D., Parkhill, J., Holden, M.T., Feil, E.J., 2012. Phylogeographic variation in recombination rates within a global clone of Methicillin-Resistant *Staphylococcus aureus* (MRSA). Genome Biol. https://doi.org/10.1186/gb-2012-13-12-r126.

Cavalli-Sforza, L.L., Feldman, M.W., 2003. The application of molecular genetic approaches to the study of human evolution. Nat. Genet. 33 (Suppl. l), 266−275. https://doi.org/10.1038/NG1113.

Chaguza, C., Senghore, M., Bojang, E., Gladstone, R.A., Lo, S.W., Tientcheu, P.E., Bancroft, R.E., Worwui, A., Foster-Nyarko, E., Ceesay, F., Okoi, C., McGee, L., Klugman, K.P., Breiman, R.F., Barer, M.R., Adegbola, R.A., Antonio, M., Bentley, S.D., Kwambana-Adams, B.A., 2020. Within-host microevolution of Streptococcus pneumoniae is rapid and adaptive during natural colonisation. Nat. Commun. 11. https://doi.org/10.1038/S41467-020-17327-W.

Challagundla, L., Luo, X., Tickler, I.A., Didelot, X., Coleman, D.C., Shore, A.C., Coombs, G.W., Sordelli, D.O., Brown, E.L., Skov, R., Larsen, A.R., Reyes, J., Robledo, I.E., Vazquez, G.J., Rivera, R., Fey, P.D., Stevenson, K., Wang, S.H., Kreiswirth, B.N., Mediavilla, J.R., Arias, C.A., Planet, P.J., Nolan, R.L., Tenover, F.C., Goering, R.v., Robinson, D.A., 2018. Range expansion and the origin of USA300 north American epidemic methicillin-resistant *Staphylococcus aureus*. mBio 9. https://doi.org/10.1128/MBIO.02016-17.

Chang, H.H., Dordel, J., Donker, T., Worby, C.J., Feil, E.J., Hanage, W.P., Bentley, S.D., Huang, S.S., Lipsitch, M., 2016. Identifying the effect of patient sharing on between-hospital genetic differentiation of methicillin-resistant staphylococcus aureus. Genome Med. 8, 1−10. https://doi.org/10.1186/S13073-016-0274-3/FIGURES/3.

Chen, L., Mathema, B., Pitout, J.D.D., DeLeo, F.R., Kreiswirth, B.N., 2014. Epidemic *Klebsiella pneumoniae* ST258 is a hybrid strain. mBio 5. https://doi.org/10.1128/MBIO.01355-14.

Cheng, L., Connor, T.R., Sirén, J., Aanensen, D.M., Corander, J., 2013. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. Mol. Biol. Evol. 30, 1224−1228. https://doi.org/10.1093/MOLBEV/MST028.

Chewapreecha, C., Harris, S.R., Croucher, N.J., Turner, C., Marttinen, P., Cheng, L., Pessia, A., Aanensen, D.M., Mather, A.E., Page, A.J., Salter, S.J., Harris, D., Nosten, F., Goldblatt, D., Corander, J., Parkhill, J., Turner, P., Bentley, S.D., 2014. Dense genomic sampling identifies highways of pneumococcal recombination. Nat. Genet. https://doi.org/10.1038/ng.2895.

Choudoir, M.J., Panke-Buisse, K., Andam, C.P., Buckley, D.H., 2017. Genome surfing as driver of microbial genomic diversity. Trends Microbiol. 25, 624−636. https://doi.org/10.1016/j.tim.2017.02.006.

Cohan, F.M., Perry, E.B., 2007. A systematics for discovering the fundamental units of bacterial diversity. Curr. Biol. https://doi.org/10.1016/j.cub.2007.03.032.

Colijn, C., Gardy, J., 2014. Phylogenetic tree shapes resolve disease transmission patterns. Evol. Med. Public Health 2014, 96−108. https://doi.org/10.1093/EMPH/EOU018.

Coll, F., McNerney, R., Guerra-Assunção, J.A., Glynn, J.R., Perdigão, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., Clark, T.G., 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nat. Commun. 5 (1 5), 1−5. https://doi.org/10.1038/ncomms5812.

Comandatore, F., Sassera, D., Bayliss, S.C., Scaltriti, E., Gaiarsa, S., Cao, X., Gales, A.C., Saito, R., Pongolini, S., Brisse, S., Feil, E.J., Bandi, C., 2019. Gene composition as a potential barrier to large recombinations in the bacterial pathogen *Klebsiella pneumoniae*. Genome Biol. Evol. 11, 3240−3251. https://doi.org/10.1093/GBE/EVZ236.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S., 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat. Genet. 45, 1176−1182. https://doi.org/10.1038/NG.2744.

Corander, J., Connor, T.R., O'Dwyer, C.A., Kroll, J.S., Hanage, W.P., 2012. Population structure in the Neisseria, and the biological significance of fuzzy species. J. R. Soc. Interface. https://doi.org/10.1098/rsif.2011.0601.

Corander, J., Fraser, C., Gutmann, M.U., Arnold, B., Hanage, W.P., Bentley, S.D., Lipsitch, M., Croucher, N.J., 2017. Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. Nat. Ecol. Evol. 1, 1950−1960. https://doi.org/10.1038/s41559-017-0337-x.

Cox, M.P., Holland, B.R., Wilkins, M.C., Schmid, J., 2013. Reconstructing past changes in locus-specific recombination rates. BMC Genet. https://doi.org/10.1186/1471-2156-14-11.

Croucher, N.J., Coupland, P.G., Stevenson, A.E., Callendrello, A., Bentley, S.D., Hanage, W.P., 2014. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nat. Commun. 5 (1 5), 1−12. https://doi.org/10.1038/ncomms6471.

Croucher, N.J., Finkelstein, J.A., Pelton, S.I., Mitchell, P.K., Lee, G.M., Parkhill, J., Bentley, S.D., Hanage, W.P., Lipsitch, M., 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat. Genet. https://doi.org/10.1038/ng.2625.

Croucher, N.J., Finkelstein, J.A., Pelton, S.I., Parkhill, J., Bentley, S.D., Lipsitch, M., Hanage, W.P., 2015. Population genomic datasets describing the post-vaccine evolutionary epidemiology of Streptococcus pneumoniae. Sci. Data 2. https://doi.org/10.1038/sdata.2015.58.

den Bakker, H.C., Didelot, X., Fortes, E.D., Nightingale, K.K., Wiedmann, M., 2008. Lineage specific recombination rates and microevolution in Listeria monocytogenes. BMC Evol. Biol. https://doi.org/10.1186/1471-2148-8-277.

Didelot, X., Nell, S., Yang, I., Woltemate, S., van der Merwe, S., Suerbaum, S., 2013. Genomic evolution and transmission of *Helicobacter pylori* in two South African families. Proc. Natl. Acad. Sci. U. S. A. 110, 13880−13885. https://doi.org/10.1073/PNAS.1304681110/SUPPL_FILE/ST05.DOC.

Didelot, X., Walker, A.S., Peto, T.E., Crook, D.W., Wilson, D.J., 2016. Within-host evolution of bacterial pathogens. Nat. Rev. Microbiol. 14 (3 14), 150−162. https://doi.org/10.1038/nrmicro.2015.13.

Donati, C., Hiller, N.L., Tettelin, H., Muzzi, A., Croucher, N., Angiuoli, S., Oggioni, M., Dunning Hotopp, J., Hu, F., Riley, D., Covacci, A., Mitchell, T., Bentley, S., Kilian, M., Ehrlich, G., Rappuoli, R., Moxon, E.R., Masignani, V., 2010. Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. Genome Biol. https://doi.org/10.1186/gb-2010-11-10-r107.

Doolittle, W.F., Papke, R.T., 2006. Genomics and the bacterial species problem. Genome Biol. https://doi.org/10.1186/gb-2006-7-9-116.

Edmonds, C.A., Lillie, A.S., Cavalli-Sforza, L.L., 2004. Mutations arising in the wave front of an expanding population. Proc. Natl. Acad. Sci. USA 101, 975−979. https://doi.org/10.1073/PNAS.0308064100.

Ellis, M.W., Griffith, M.E., Jorgensen, J.H., Hospenthal, D.R., Mende, K., Patterson, J.E., 2009. Presence and molecular epidemiology of virulence factors in methicillin-resistant *Staphylococcus aureus* strains colonizing and infecting soldiers. J. Clin. Microbiol. 47, 940. https://doi.org/10.1128/JCM.02352-08.

Excoffier, L., 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Mol. Ecol. 13, 853−864. https://doi.org/10.1046/J.1365-294X.2003.02004.X.

Excoffier, L., Foll, M., Petit, R.J., 2009. Genetic consequences of range expansions. Ann. Rev. Ecol. Evol. Syst. 40, 481−501. https://doi.org/10.1146/ANNUREV.ECOLSYS.39.110707.173414.

Falush, D., Wirth, T., Linz, B., Pritchard, J.K., Stephens, M., Kidd, M., Blaser, M.J., Graham, D.Y., Vacher, S., Perez-Perez, G.I., Yamaoka, Y., Mégraud, F., Otto, K., Reichard, U., Katzowitsch, E., Wang, X., Achtman, M., Suerbaum, S., 2003. Traces of human migrations in *Helicobacter pylori* populations. Science 299, 1582−1585. https://doi.org/10.1126/SCIENCE.1080857.

Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., Spratt, B.G., 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 186, 1518−1530. https://doi.org/10.1128/JB.186.5.1518-1530.2004/ASSET/28BAF22D-0A29-46AA-A47B-089F2590A526/ASSETS/GRAPHIC/ZJB0050410900010.JPEG.

Fraser, C., Alm, E.J., Polz, M.F., Spratt, B.G., Hanage, W.P., 2009. The bacterial species challenge: making sense of genetic and ecological diversity. Science. https://doi.org/10.1126/science.1159388.

Frerichs, R.R., Keim, P.S., Barrais, R., Piarroux, R., 2012. Nepalese origin of cholera epidemic in Haiti. Clin. Microbiol. Infect. 18. https://doi.org/10.1111/J.1469-0691.2012.03841.X.

Gardy, J.L., Johnston, J.C., Ho Sui, S.J., Cook, V.J., Shah, L., Brodkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., Varhol, R., Birol, I., Lem, M., Sharma, M.K., Elwood, K., Jones, S.J., Brinkman, F.S., Brunham, R.C., Tang, P., 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N. Engl. J. Med. 364. https://doi.org/10.1056/NEJMoa1003176.

Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., van de Peer, Y., Vandamme, P., Thompson, F.L., Swings, J., 2005. Re-evaluating prokaryotic species. Nat. Rev. Microbiol. https://doi.org/10.1038/nrmicro1236.

Gill, J.L., Hedge, J., Wilson, D.J., MacLean, R.C., 2021. Evolutionary processes driving the rise and fall of *Staphylococcus aureus* ST239, a dominant hybrid pathogen. mBio 12. https://doi.org/10.1128/MBIO.02168-21.

Godreuil, S., Cohan, F., Shah, H., Tibayrenc, M., 2005. Which species concept for pathogenic bacteria? An E-Debate. Infect. Genet. Evol. https://doi.org/10.1016/j.meegid.2004.03.004.

Godreuil, S., Renaud, F., Choisy, M., Depina, J.J., Garnotel, E., Morillon, M., van de Perre, P., Bañuls, A.L., 2010. Highly structured genetic diversity of the *Mycobacterium tuberculosis* population in Djibouti. Clin. Microbiol. Infection. https://doi.org/10.1111/j.1469-0691.2009.03025.x.

Golubchik, T., Batty, E.M., Miller, R.R., Farr, H., Young, B.C., Larner-Svensson, H., Fung, R., Godwin, H., Knox, K., Votintseva, A., Everitt, R.G., Street, T., Cule, M., Ip, C.L.C., Didelot, X., Peto, T.E.A., Harding, R.M., Wilson, D.J., Crook, D.W., Bowden, R., 2013. Within-host evolution of *Staphylococcus aureus* during asymptomatic carriage. PLoS One 8, e61319. https://doi.org/10.1371/JOURNAL.PONE.0061319.

González-González, A., Sánchez-Reyes, L.L., Delgado Sapien, G., Eguiarte, L.E., Souza, V., 2013. Hierarchical clustering of genetic diversity associated to different levels of mutation and recombination in *Escherichia coli*: a study based on Mexican isolates. Infect. Genet. Evol. 13, 187−197. https://doi.org/10.1016/J.MEEGID.2012.09.003.

Gordienko, E.N., Kazanov, M.D., Gelfand, M.S., 2013. Evolution of pan-genomes of *Escherichia coli*, Shigella spp., and *Salmonella enterica*. J. Bacteriol. https://doi.org/10.1128/JB.02285-12.

Grad, Y.H., Kirkcaldy, R.D., Trees, D., Dordel, J., Harris, S.R., Goldstein, E., Weinstock, H., Parkhill, J., Hanage, W.P., Bentley, S., Lipsitch, M., 2014. Genomic epidemiology of Neisseria gonorrhoeae with reduced susceptibility to cefixime in the USA: a retrospective observational study. Lancet Infect. Dis. 14, 220−226. https://doi.org/10.1016/S1473-3099(13)70693-5.

Grosso-Becerra, M.-V., Santos-Medellín, C., González-Valdez, A., Méndez, J.-L., Delgado, G., Morales-Espinosa, R., Servín-González, L., Alcaraz, L.-D., Soberón-Chávez, G., 2014. *Pseudomonas aeruginosa* clinical and environmental isolates constitute a single population with high phenotypic diversity. BMC Genom. https://doi.org/10.1186/1471-2164-15-318.

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34, 4121−4123. https://doi.org/10.1093/BIOINFORMATICS/BTY407.

Haley, B.J., Choi, S.Y., Grim, C.J., Onifade, T.J., Cinar, H.N., Tall, B.D., Taviani, E., Hasan, N.A., Abdullah, A.H., Carter, L., Sahu, S.N., Kothary, M.H., Chen, A., Baker, R., Hutchinson, R., Blackmore, C., Cebula, T.A., Huq, A., Colwell, R.R., 2014. Genomic and phenotypic characterization of *Vibrio cholerae* non-O1 isolates from a US Gulf Coast cholera outbreak. PLoS One 9. https://doi.org/10.1371/JOURNAL.PONE.0086264.

Hanage, W.P., Fraser, C., Spratt, B.G., 2005. Fuzzy species among recombinogenic bacteria. BMC Biol. https://doi.org/10.1186/1741-7007-3-6.

Hanage, W.P., Fraser, C., Tang, J., Connor, T.R., Corander, J., 2009. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. Science. https://doi.org/10.1126/science.1171908.

Harris, S.R., Feil, E.J., Holden, M.T.G., Quail, M.A., Nickerson, E.K., Chantratita, N., Gardete, S., Tavares, A., Day, N., Lindsay, J.A., Edgeworth, J.D., de Lencastre, H., Parkhill, J., Peacock, S.J., Bentley, S.D., 2010.

Evolution of MRSA during hospital transmission and intercontinental spread. Science 327, 469−474. https://doi.org/10.1126/SCIENCE.1182395.

Harrow, G.L., Lees, J.A., Hanage, W.P., Lipsitch, M., Corander, J., Colijn, C., Croucher, N.J., 2021. Negative frequency-dependent selection and asymmetrical transformation stabilise multi-strain bacterial population structures. ISME J. 15 (5), 1523−1538. https://doi.org/10.1038/s41396-020-00867-w.

Hendriksen, R.S., Price, L.B., Schupp, J.M., Gillece, J.D., Kaas, R.S., Engelthaler, D.M., Bortolaia, V., Pearson, T., Waters, A.E., Upadhyay, B.P., Shrestha, S.D., Adhikari, S., Shaky, G., Keim, P.S., Aarestrupa, F.M., 2011. Population genetics of vibrio cholerae from Nepal in 2010: evidence on the origin of the haitian outbreak. mBio 2, 157−168. https://doi.org/10.1128/MBIO.00157-11/SUPPL_FILE/MBO004111153SF02.PDF.

Hewitt, G.M., 1999. Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. 68, 87−112.

Hilty, M., Wüthrich, D., Salter, S.J., Engel, H., Campbell, S., Sá-Leão, R., de Lencastre, H., Hermans, P., Sadowy, E., Turner, P., Chewapreecha, C., Diggle, M., Pluschke, G., McGee, L., Eser, Ö.K., Low, D.E., Smith-Vaughan, H., Endimiani, A., Ffer, M.K., Dupasquier, M., Beaudoing, E., Weber, J., Bruggmann, R., Hanage, W.P., Parkhill, J., Hathaway, L.J., Hlemann, K.M., Bentley, S.D., 2014. Global phylogenomic analysis of nonencapsulated Streptococcus pneumoniae reveals a deep-branching classic lineage that is distinct from multiple sporadic lineages. Genome Biol. Evol. 6, 3281−3294. https://doi.org/10.1093/GBE/EVU263.

Ibrahim, K.M., Nichols, R.A., Hewitt, G.M., 1996. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity 77 (3 77), 282−291. https://doi.org/10.1038/hdy.1996.142.

Ivers, L.C., Walton, D.A., 2012. The "first" case of cholera in Haiti: lessons for global health. Am. J. Trop. Med. Hyg. 86, 36. https://doi.org/10.4269/AJTMH.2012.11-0435.

Jeraldo, P., Sipos, M., Chia, N., Brulc, J.M., Dhillon, A.S., Konkel, M.E., Larson, C.L., Nelson, K.E., Qu, A., Schook, L.B., Yang, F., White, B.A., Goldenfeld, N., 2012. Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. Proc. Natl. Acad. Sci. U. S. A. 109, 9692−9698. https://doi.org/10.1073/PNAS.1206721109/SUPPL_FILE/PNAS.201206721SI.PDF.

Jernigan, J.A., Hatfield, K.M., Wolford, H., Nelson, R.E., Olubajo, B., Reddy, S.C., McCarthy, N., Paul, P., McDonald, L.C., Kallen, A., Fiore, A., Craig, M., Baggs, J., 2020. Multidrug-resistant bacterial infections in U.S. Hospitalized patients, 2012−2017. N. Engl. J. Med. 382, 1309−1319. https://doi.org/10.1056/NEJMOA1914433/SUPPL_FILE/NEJMOA1914433_DISCLOSURES.PDF.

Jungnick, S., Margos, G., Rieger, M., Dzaferovic, E., Bent, S.J., Overzier, E., Silaghi, C., Walder, G., Wex, F., Koloczek, J., Sing, A., Fingerle, V., 2015. Borrelia burgdorferi sensu stricto and Borrelia afzelii: population structure and differential pathogenicity. Int. J. Med. Microbiol. https://doi.org/10.1016/j.ijmm.2015.08.017.

Kaas, R.S., Friis, C., Ussery, D.W., Aarestrup, F.M., 2012. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. BMC Genom. https://doi.org/10.1186/1471-2164-13-577.

Katz, L.S., Petkau, A., Beaulaurier, J., Tyler, S., Antonova, E.S., Turnsek, M.A., Guo, Y., Wang, S., Paxinos, E.E., Orata, F., Gladney, L.M., Stroika, S., Folster, J.P., Rowe, L., Freeman, M.M., Knox, N., Frace, M., Boncy, J., Graham, M., Hammer, B.K., Boucher, Y., Bashir, A., Hanage, W.P., van Domselaar, G., Tarr, C.L., 2013. Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti. mBio 4. https://doi.org/10.1128/MBIO.00398-13/SUPPL_FILE/MBO003131547SF03.PDF.

Ke, W., Huang, S.S., Hudson, L.O., Elkins, K.R., Nguyen, C.C., Spratt, B.G., Murphy, C.R., Avery, T.R., Lipsitch, M., 2012. Patient sharing and population genetic structure of methicillin-resistant *Staphylococcus aureus*. Proc. Natl. Acad. Sci. U. S. A. 109, 6763−6768. https://doi.org/10.1073/PNAS.1113578109/SUPPL_FILE/PNAS.201113578SI.PDF.

Klopfstein, S., Currat, M., Excoffier, L., 2006. The fate of mutations surfing on the wave of a range expansion. Mol. Biol. Evol. 23, 482−490. https://doi.org/10.1093/MOLBEV/MSJ057.

Koton, Y., Gordon, M., Chalifa-Caspi, V., Bisharat, N., 2015. Comparative genomic analysis of clinical and environmental Vibrio vulnificus isolates revealed biotype 3 evolutionary relationships. Front. Microbiol. https://doi.org/10.3389/fmicb.2014.00803.

Lawson Handley, L.J., Estoup, A., Evans, D.M., Thomas, C.E., Lombaert, E., Facon, B., Aebi, A., Roy, H.E., 2011. Ecological genetics of invasive alien species. BioControl 56 (4 56), 409−428. https://doi.org/10.1007/S10526-011-9386-2.

Lees, J.A., Harris, S.R., Tonkin-Hill, G., Gladstone, R.A., Lo, S.W., Weiser, J.N., Corander, J., Bentley, S.D., Croucher, N.J., 2019. Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Res. 29, 304−316. https://doi.org/10.1101/GR.241455.118/-/DC1.

Leung, N.S., Padgett, P., Robinson, D.A., Brown, E.L., 2015. Prevalence and behavioural risk factors of *Staphylococcus aureus* nasal colonization in community-based injection drug users. Epidemiol. Infect. 143, 2430. https://doi.org/10.1017/S0950268814003227.

Levin, B.R., 1988. Frequency-dependent selection in bacterial populations. Philos. Trans. R. Soc. Lond. B Biol. Sci. 319, 459−472. https://doi.org/10.1098/RSTB.1988.0059.

Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., Myers, R.M., 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319, 1100−1104. https://doi.org/10.1126/SCIENCE.1153717.

Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S.W., Yamaoka, Y., Graham, D.Y., Perez-Trallero, E., Wadstrom, T., Suerbaum, S., Achtman, M., 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature 445, 915−918. https://doi.org/10.1038/NATURE05562.

Margos, G., Gatewood, A.G., Aanensen, D.M., Hanincová, K., Terekhova, D., Vollmer, S.A., Cornet, M., Piesman, J., Donaghy, M., Bormane, A., Hurn, M.A., Feil, E.J., Fish, D., Casjens, S., Wormser, G.P., Schwartz, I., Kurtenbach, K., 2008. MLST of housekeeping genes captures geographic population structure and suggests a European origin of Borrelia burgdorferi. Proc. Natl. Acad. Sci. U. S. A. 105, 8730−8735. https://doi.org/10.1073/PNAS.0800323105.

Margos, G., Tsao, J.I., Castillo-Ramírez, S., Girard, Y.A., Hamer, S.A., Hoen, A.G., Lane, R.S., Raper, S.L., Ogden, N.H., 2012. Two boundaries separate borrelia burgdorferi populations in north America. Appl. Environ. Microbiol. 78, 6059−6067. https://doi.org/10.1128/AEM.00231-12/SUPPL_FILE/ZAM999103580SO2.PDF.

Mathers, A.J., Stoesser, N., Sheppard, A.E., Pankhurst, L., Giess, A., Yeh, A.J., Didelot, X., Turner, S.D., Sebra, R., Kasarskis, A., Peto, T., Crook, D., Sifri, C.D., 2015. *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* at a single institution: insights into endemicity from whole-genome sequencing. Antimicrob. Agents Chemother. 59, 1656−1663. https://doi.org/10.1128/AAC.04292-14.

McNally, A., Kallonen, T., Connor, C., Abudahab, K., Aanensen, D.M., Horner, C., Peacock, S.J., Parkhill, J., Croucher, N.J., Corander, J., 2019. Diversification of colonization factors in a multidrug-resistant *Escherichia coli* lineage evolving under negative frequency-dependent selection. mBio 10. https://doi.org/10.1128/MBIO.00644-19.

Méric, G., Miragaia, M., de Been, M., Yahara, K., Pascoe, B., Mageiros, L., Mikhail, J., Harris, L.G., Wilkinson, T.S., Rolo, J., Lamble, S., Bray, J.E., Jolley, K.A., Hanage, W.P., Bowden, R., Maiden, M.C.J., Mack, D., de Lencastre, H., Feil, E.J., Corander, J., Sheppard, S.K., 2015. Ecological overlap and horizontal gene transfer in *Staphylococcus aureus* and Staphylococcus epidermidis. Genome Biol. Evol. 7, 1313−1328. https://doi.org/10.1093/GBE/EVV066.

Mitchell, P.K., Azarian, T., Croucher, N.J., Callendrello, A., Thompson, C.M., Pelton, S.I., Lipsitch, M., Hanage, W.P., 2019. Population genomics of pneumococcal carriage in Massachusetts children following introduction of PCV-13. Microb. Genom. 5. https://doi.org/10.1099/mgen.0.000252.

Mourkas, E., Yahara, K., Bayliss, S.C., Calland, J.K., Johansson, H., Mageiros, L., Muñoz-Ramirez, Z.Y., Futcher, G., Méric, G., Hitchings, M.D., Sandoval-Motta, S., Torres, J., Jolley, K.A., Maiden, M.C.J., Ellström, P., Waldenström, J., Pascoe, B., Sheppard, S.K., 2022. Host ecology regulates interspecies recombination in bacteria of the genus Campylobacter. Elife 11. https://doi.org/10.7554/ELIFE.73552.

Mutreja, A., Kim, D.W., Thomson, N.R., Connor, T.R., Lee, J.H., Kariuki, S., Croucher, N.J., Choi, S.Y., Harris, S.R., Lebens, M., Niyogi, S.K., Kim, E.J., Ramamurthy, T., Chun, J., Wood, J.L.N., Clemens, J.D., Czerkinsky, C., Nair, G.B., Holmgren, J., Parkhill, J., Dougan, G., 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. Nature 477 (7365), 462−465. https://doi.org/10.1038/nature10392.

Nandi, T., Holden, M.T.G., Didelot, X., Mehershahi, K., Boddey, J.A., Beacham, I., Peak, I., Harting, J., Baybayan, P., Guo, Y., Wang, S., How, L.C., Sim, B., Essex-Lopresti, A., Sarkar-Tyson, M., Nelson, M., Smither, S., Ong, C., Aw, L.T., Hoon, C.H., Michell, S., Studholme, D.J., Titball, R., Chen, S.L., Parkhill, J., Tan, P., 2015. Burkholderia pseudomallei sequencing identifies genomic clades with distinct recombination, accessory, and epigenetic profiles. Genome Res. 25, 129. https://doi.org/10.1101/GR.177543.114.

Norte, A.C., Margos, G., Becker, N.S., Albino Ramos, J., Núncio, M.S., Fingerle, V., Araújo, P.M., Adamík, P., Alivizatos, H., Barba, E., Barrientos, R., Cauchard, L., Csörgő, T., Diakou, A., Dingemanse, N.J., Doligez, B., Dubiec, A., Eeva, T., Flaisz, B., Grim, T., Hau, M., Heylen, D., Hornok, S., Kazantzidis, S., Kováts, D., Krause, F., Literak, I., Mänd, R., Mentesana, L., Morinay, J., Mutanen, M., Neto, J.M., Nováková, M., Sanz, J.J., Pascoal da Silva, L., Sprong, H., Tirri, I.S., Török, J., Trilar, T., Tyller, Z., Visser, M.E., Lopes de Carvalho, I., 2020. Host dispersal shapes the population structure of a tick-borne bacterial pathogen. Mol. Ecol. 29, 485−501. https://doi.org/10.1111/MEC.15336.

Orsi, R.H., Bakker, H.C.D., Wiedmann, M., 2011. Listeria monocytogenes lineages: genomics, evolution, ecology, and phenotypic characteristics. Int. J. Med. Microbiol. https://doi.org/10.1016/j.ijmm.2010.05.002.

Paulose, J., Hallatschek, O., 2020. The impact of long-range dispersal on gene surfing. Proc. Natl. Acad. Sci. U. S. A. 117, 7584−7593. https://doi.org/10.1073/PNAS.1919485117/SUPPL_FILE/PNAS.1919485117.SAPP.PDF.

Peischl, S., Excoffier, L., 2015. Expansion load: recessive mutations and the role of standing genetic variation. Mol. Ecol. 24, 2084−2094. https://doi.org/10.1111/MEC.13154.

Pepperell, C., Hoeppner, V.H., Lipatov, M., Wobeser, W., Schoolnik, G.K., Feldman, M.W., 2010. Bacterial genetic signatures of human social phenomena among *M. tuberculosis* from an Aboriginal Canadian population. Mol. Biol. Evol. 27, 427−440. https://doi.org/10.1093/MOLBEV/MSP261.

Peter, B.M., Slatkin, M., 2013. Detecting range expansions from genetic data. Evolution 67, 3274−3289. https://doi.org/10.1111/EVO.12202.

Polz, M.F., Alm, E.J., Hanage, W.P., 2013. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet. 29, 170−175. https://doi.org/10.1016/J.TIG.2012.12.006.

Popovich, K.J., Snitkin, E., Green, S.J., Aroutcheva, A., Hayden, M.K., Hota, B., Weinstein, R.A., 2016. Genomic epidemiology of USA300 methicillin-resistant *Staphylococcus aureus* in an urban community. Clin. Infect. Dis. 62, 37−44. https://doi.org/10.1093/CID/CIV794.

Prosperi, M., Veras, N., Azarian, T., Rathore, M., Nolan, D., Rand, K., Cook, R.L., Johnson, J., Morris, J.G., Salemi, M., 2013. Molecular epidemiology of community-associated methicillin-resistant *Staphylococcus aureus* in the genomic era: a cross-sectional study. Sci. Rep. 3. https://doi.org/10.1038/SREP01902.

Qiu, W.G., Bosler, E.M., Campbell, J.R., Ugine, G.D., Wang, I.N., Luft, B.J., Dykhuizen, D.E., 1997. A population genetic study of Borrelia burgdorferi sensu stricto from eastern long island, New York, suggested frequency-dependent selection, gene flow and host adaptation. Hereditas 127, 203−216. https://doi.org/10.1111/J.1601-5223.1997.00203.X.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., Cavalli-Sforza, L.L., 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. U. S. A. 102, 15942−15947. https://doi.org/10.1073/PNAS.0507611102.

Robinson, D.A., Briles, D.E., Crain, M.J., Hollingshead, S.K., 2002. Evolution and virulence of serogroup 6 pneumococci on a global scale. J. Bacteriol. 184, 6367. https://doi.org/10.1128/JB.184.22.6367-6375.2002.

Robinson, D.A., Enright, M.C., 2004. Evolution of *Staphylococcus aureus* by large chromosomal replacements. J. Bacteriol. 186, 1060. https://doi.org/10.1128/JB.186.4.1060-1064.2004.

Sanguinetti, L., Toti, S., Reguzzi, V., Bagnoli, F., Donati, C., 2012. A novel computational method identifies intra- and inter-species recombination events in italicStaphylococcus aureus/italic and italicStreptococcus pneumoniae/italic. PLoS Comput. Biol. 8, 1−11. https://doi.org/10.1371/journal.pcbi.1002668.

Sapriel, G., Konjek, J., Orgeur, M., Bouri, L., Frézal, L., Roux, A.-L., Dumas, E., Brosch, R., Bouchier, C., Brisse, S., Vandenbogaert, M., Thiberge, J.-M., Caro, V., Ngeow, Y.F., Tan, J.L., Herrmann, J.-L., Gaillard, J.-L., Heym, B., Wirth, T., 2016. Genome-wide mosaicism within Mycobacterium abscessus: evolutionary and epidemiological implications. BMC Genom. 17, 1−16. https://doi.org/10.1186/s12864-016-2448-1.

Sauver, J.S., Marrs, C.F., Foxman, B., Somsel, P., Madera, R., Gilsdorf, J.R., 2000. Risk factors for otitis media and carriage of multiple strains of Haemophilus influenzae and Streptococcus pneumoniae. Emerg. Infect. Dis. 6, 622. https://doi.org/10.3201/EID0606.000611.

Schmitt, T., 2007. Molecular biogeography of Europe: pleistocene cycles and postglacial trends. Front. Zool. 4. https://doi.org/10.1186/1742-9994-4-11.

Shapiro, B.J., Polz, M.F., 2015. Microbial speciation. Cold Spring Harbor Perspect. Biol. https://doi.org/10.1101/cshperspect.a018143.

Sheppard, S.K., Cheng, L., Méric, G., de Haan, C.P.A., Llarena, A.K., Marttinen, P., Vidal, A., Ridley, A., Clifton-Hadley, F., Connor, T.R., Strachan, N.J.C., Forbes, K., Colles, F.M., Jolley, K.A., Bentley, S.D., Maiden, M.C.J., Hänninen, M.L., Parkhill, J., Hanage, W.P., Corander, J., 2014. Cryptic ecology among host generalist Campylobacter jejuni in domestic animals. Mol. Ecol. 23, 2442−2451. https://doi.org/10.1111/MEC.12742.

Sheppard, S.K., Didelot, X., Jolley, K.A., Darling, A.E., Pascoe, B., Meric, G., Kelly, D.J., Cody, A., Colles, F.M., Strachan, N.J.C., Ogden, I.D., Forbes, K., French, N.P., Carter, P., Miller, W.G., McCarthy, N.D., Owen, R., Litrup, E., Egholm, M., Affourtit, J.P., Bentley, S.D., Parkhill, J., Maiden, M.C.J., Falush, D., 2013. Progressive genome-wide introgression in agricultural Campylobacter coli. Mol. Ecol. 22, 1051−1064. https://doi.org/10.1111/MEC.12162.

Slatkin, M., Excoffier, L., 2012. Serial founder effects during range expansion: a spatial analog of genetic drift. Genetics 191, 171. https://doi.org/10.1534/GENETICS.112.139022.

Smyth, D.S., McDougal, L.K., Gran, F.W., Manoharan, A., Enright, M.C., Song, J.H., de Lencastre, H., Robinson, D.A., 2010. Population structure of a hybrid clonal group of methicillin-resistant *Staphylococcus aureus*, ST239-MRSA-III. PLoS One 5, e8582. https://doi.org/10.1371/JOURNAL.PONE.0008582.

Smyth, D.S., Wong, A., Robinson, D.A., 2011. Cross-species spread of SCCmec IV subtypes in staphylococci. Infect. Genet. Evol. 11, 446−453. https://doi.org/10.1016/J.MEEGID.2010.12.005.

Stott, C.M., Bobay, L.M., 2020. Impact of homologous recombination on core genome phylogenies. BMC Genom. 21, 1−10. https://doi.org/10.1186/S12864-020-07262-X/FIGURES/6.

Thomas, J.C., Godfrey, P.A., Feldgarden, M., Robinson, D.A., 2012. Candidate targets of balancing selection in the genome of *Staphylococcus aureus*. Mol. Biol. Evol. 29, 1175. https://doi.org/10.1093/MOLBEV/MSR286.

Thomas, J.C., Zhang, L., Robinson, D.A., 2014. Differing lifestyles of Staphylococcus epidermidis as revealed through Bayesian clustering of multilocus sequence types. Infect. Genet. Evol. 22, 257. https://doi.org/10.1016/J.MEEGID.2013.06.020.

Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.M., Doumbo, O., Ibrahim, M., Juma, A.T., Kotze, M.J., Lema, G., Moore, J.H., Mortensen, H., Nyambo, T.B., Omar, S.A., Powell, K., Pretorius, G.S., Smith, M.W., Thera, M.A., Wambebe, C., Weber, J.L., Williams, S.M., 2009. The genetic structure and history of Africans and African Americans. Science 324, 1035−1044. https://doi.org/10.1126/SCIENCE.1172257.

Tong, S.Y.C., Davis, J.S., Eichenberger, E., Holland, T.L., Fowler, V.G., 2015a. *Staphylococcus aureus* infections: epidemiology, pathophysiology, clinical manifestations, and management. Clin. Microbiol. Rev. 28, 603−661. https://doi.org/10.1128/CMR.00134-14.

Tong, S.Y.C., Holden, M.T.G., Nickerson, E.K., Cooper, B.S., Koser, C.U., Cori, A., Jombart, T., Cauchemez, S., Fraser, C., Wuthiekanun, V., Thaipadungpanit, J., Hongsuwan, M., Day, N.P., Limmathurotsakul, D., Parkhill, J., Peacock, S.J., 2015b. Genome sequencing defines phylogeny and spread of methicillin-resistant *Staphylococcus aureus* in a high transmission setting. Genome Res. 25, 111−118. https://doi.org/10.1101/GR.174730.114.

Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W., Corander, J., 2019. Fast hierarchical Bayesian analysis of population structure. Nucleic Acids Res. 47, 5539. https://doi.org/10.1093/NAR/GKZ361.

Tonkin-Hill, G., Lees, J.A., Bentley, S.D., Frost, S.D.W., Corander, J., 2018. RhierBAPS: an R implementation of the population clustering algorithm hierBAPS. Wellcome Open Res. 3. https://doi.org/10.12688/WELLCOMEOPENRES.14694.1.

Valot, B., Guyeux, C., Rolland, J.Y., Mazouzi, K., Bertrand, X., Hocquet, D., 2015. What it takes to be a *Pseudomonas aeruginosa*? The core genome of the opportunistic pathogen updated. PLoS One. https://doi.org/10.1371/journal.pone.0126468.

Vollmer, S.A., Bormane, A., Dinnis, R.E., Seelig, F., Dobson, A.D.M., Aanensen, D.M., James, M.C., Donaghy, M., Randolph, S.E., Feil, E.J., Kurtenbach, K., Margos, G., 2011. Host migration impacts on the phylogeography of Lyme Borreliosis spirochaete species in Europe. Environ. Microbiol. https://doi.org/10.1111/j.1462-2920.2010.02319.x.

Vysakh, P.R., Jeya, M., 2013. A comparative analysis of community acquired and hospital acquired methicillin resistant Staphylococcus aureus. J. Clin. Diagn. Res. 7, 1339. https://doi.org/10.7860/JCDR/2013/5302.3139.

Walker, T.M., Ip, C.L.C., Harrell, R.H., Evans, J.T., Kapatai, G., Dedicoat, M.J., Eyre, D.W., Wilson, D.J., Hawkey, P.M., Crook, D.W., Parkhill, J., Harris, D., Walker, A.S., Bowden, R., Monk, P., Smith, E.G., Peto, T.E.A., 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect. Dis. 13, 137−146. https://doi.org/10.1016/S1473-3099(12)70277-3.

Welch, R.A., Burland, V., Plunkett, G., Redford, P., Roesch, P., Rasko, D., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J., Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C., Perna, N.T., Mobley, H.L.T., Donnenberg, M.S., Blattner, F.R., 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. https://doi.org/10.1073/pnas.252529799.

Willems, R.J.L., Top, J., van Schaik, W., Leavis, H., Bonten, M., Sirén, J., Hanage, W.P., Corander, J., 2012. Restricted gene flow among hospital subpopulations of Enterococcus faecium. mBio. https://doi.org/10.1128/mBio.00151-12.

Williamson, D.A., Chow, E.P.F., Gorrie, C.L., Seemann, T., Ingle, D.J., Higgins, N., Easton, M., Taiaroa, G., Grad, Y.H., Kwong, J.C., Fairley, C.K., Chen, M.Y., Howden, B.P., 2019. Bridging of Neisseria gonorrhoeae lineages across sexual networks in the HIV pre-exposure prophylaxis era. Nat. Commun. 10. https://doi.org/10.1038/S41467-019-12053-4.

Worby, C.J., Lipsitch, M., Hanage, W.P., 2014. Within-host bacterial diversity hinders accurate reconstruction of transmission networks from genomic distance data. PLoS Comput. Biol. 10, e1003549. https://doi.org/10.1371/JOURNAL.PCBI.1003549.

Yahara, K., Didelot, X., Ansari, M.A., Sheppard, S.K., Falush, D., 2014. Efficient inference of recombination hot regions in bacterial genomes. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msu082.

Young, B.C., Golubchik, T., Batty, E.M., Fung, R., Larner-Svensson, H., Votintseva, A.A., Miller, R.R., Godwin, H., Knox, K., Everitt, R.G., Iqbal, Z., Rimmer, A.J., Cule, M., Ip, C.L.C., Didelot, X., Harding, R.M., Donnelly, P., Peto, T.E., Crook, D.W., Bowden, R., Wilsona, D.J., 2012. Evolutionary dynamics of *Staphylococcus aureus* during progression from carriage to disease. Proc. Natl. Acad. Sci. U. S. A. 109, 4550−4555. https://doi.org/10.1073/PNAS.1113219109/SUPPL_FILE/ST05.DOCX.

CHAPTER

Diverse Strategies and Evolutionary Histories of Fungal Pathogens

4

Cene Gostinčar and Nina Gunde-Cimerman

University of Ljubljana, Biotechnical Faculty, Department of Biology, Ljubljana, Slovenia

1. Introduction

The number of mycoses in humans is growing at an alarming rate. It is estimated that more than 150 million people suffer from severe fungal infections and more than 1.5 million of them die each year (Bongomin et al., 2017; Brown et al., 2012; Rayens and Norris, 2022; World Health Organization, 2022). Changes in the environment and human population lead to the emergence of new pathogens (Friedman and Schwartz, 2019). The spread of fungal diseases poses a threat to biodiversity and food security (Fisher et al., 2020). With limited tools to combat this, a better understanding of the evolution of fungal pathogens is critical to predicting the nature and limiting the burden of fungal threats in the future.

Most fungal pathogens of animals are opportunists and even more are facultative pathogens—they do not require the host to survive. They are increasingly numerous and diverse and occur in all major branches of the fungal tree of life (Gostinčar et al., 2018). For opportunistic pathogens, infection is typically a rare event with negligible impact on the survival—and evolution—of the pathogen species. Our understanding of such opportunistic pathogenesis is that it is random and arises from the accidental inoculation of the host and the at least temporary survival of the relatively hostile conditions encountered by the pathogen in an (otherwise nutrient-rich) animal body (Casadevall and Pirofski, 2007; Gostinčar et al., 2018). To do this, pathogens use the cellular toolbox that has evolved in the environment outside the host rather than in response to previous infections.

While the number of primary fungal pathogens of mammals is extremely small, especially compared to the growing diversity of fungal opportunists, the small number of species poses a disproportionate medical problem (Bongomin et al., 2017; World Health Organization, 2022). In these species, the incidence of infection is much higher and may support (in some cases along with commensalism) coevolution between host and pathogen and specialization for pathogenesis (Alves et al., 2020). The same is true for many plant pathogens, some of which are highly specialized to their hosts (Fones et al., 2020).

The modes of pathogenicity and evolutionary trajectories of opportunistic and primary fungal pathogens may have less overlap than is often assumed (Gostinčar et al., 2018). To fully understand the role of fungi in animal and plant disease, we need to study not one but two distinct phenomena that have important implications for the future epidemiology of fungal infections.

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00020-3

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

2. A Brief Overview of Fungal Pathogens

2.1 Plant Pathogens

From saprophytic to mycorrhizal species, fungal interactions with plants shape the functions of terrestrial ecosystems. Fungi have also evolved a spectrum of parasitic relationships, ranging from biotrophic (using specialized structures to acquire nutrients from host tissues while downmodulating host immunity) to necrotrophic (killing host tissues in order to feed on them) (Köhler et al., 2017). Since approximately 8000 fungal and oomycete species are associated with plant diseases (Fisher et al., 2020), we present here only some of the most striking examples of emerging infectious diseases (EIDs). In the 20th century, Dutch elm blight, caused by *Ophiostoma novo-ulmi*, caused the loss of about 100 million elms in the United Kingdom and the United States, and 3.5 billion chestnut trees have succumbed to chestnut blight in the United States, caused by *Cryphonectria parasitica* (Fisher et al., 2012). In the 1990s, the oomycete *Phytophthora ramorum* emerged simultaneously in the United States on oaks and in Europe on rhododendrons and later spread globally (Grünwald et al., 2019). More recent global pathogens are myrtle rust (*Austropuccinia psidii*) and ash dieback (*Hymenoscyphus fraxineus)*, which can affect various plant species (Fisher et al., 2020).

Fungal plant pathogens are not only responsible for endangering plant species in nature, but also for threatening global food security through crop losses (Fisher et al., 2020). In the nineteenth century, the potato blight *Phytophthora infestans* led to a catastrophic shortage of food and economic ruin during the Irish Famine. Today, fungal diseases of agronomically important plants include old and new devastating epidemics. Wheat and soybean, two of the world's top five calorie crops, are currently threatened by emerging fungal and oomycete pathogens: the wheat blast fungus *Magnaporthe oryzae*, which causes 10%−35% crop losses; the soybean rust fungus *Phakopsora pachyrhizi*; and the soybean oomycete *Phytophthora sojae* (Fantozzi et al., 2021; Fisher et al., 2020; McCoy et al., 2023). Some regional EIDs also pose a growing threat to food security. The most important include wheat rust caused by *Puccinia graminis* f. *tritici* (virulent Ug99 strain), causing 5%−10% losses of wheat yields in the EU every year, *Fusarium oxysporum* f. sp. *cubense* (tropical Race 4), which infects Cavendish bananas, maize smut caused by *Ustilago maydis*, and potato blight caused by *P. infestans* (Edel-Hermann and Lecomte, 2019; Fantozzi et al., 2021; Fisher et al., 2020).

Although many fungal pathogens of plants exhibit extreme host specialization, exemplified by the gene-for-gene interactions between biotrophic fungi and their plant hosts, some combine a broad host range with high virulence. The oomycete *Phytophthora ramorum* infects 109 host species, *Puccinia graminis* f. *tritici* infects 365 cereal or grass species in addition to wheat. As many of these other hosts survive infection, they themselves produce inoculum for new infections (Fisher et al., 2012; Grünwald et al., 2019). Pathogenic fungi spread in response to environmental change, human-mediated intercontinental dispersal and trade in fungus-infected produce and food. Examples include the above-mentioned Irish Famine (a consequence of the introduction of *P. infestans* from South America), the destruction of North American chestnuts (caused by the importation of *C. parasitica* from Asia) and the introduction of *Heterobasidion annosum* from the United States into Italy (through untreated wood) (Chaloner et al., 2021; Fisher et al., 2012; Rigling and Prospero, 2018).

In addition to the introduction of pathogens into new areas, several other factors contribute to the challenges posed by EIDs. These include modern agricultural intensification practices, large areas of genetically uniform crops, the use of single-target antifungals that accelerate the emergence of new

virulent and fungicide-resistant strains, and the changed spread and impacts of plant pathogenic fungi due to climate change (Bebber et al., 2019; Fisher et al., 2012). A combination of these factors can lead to a rapid deterioration of a precariously balanced situation. For example, the impact of wheat pathogens is currently limited by the use of resistant host cultivars and the application of chemical fungicides. However, both of these methods are prone to failure due to the rapid evolution of pathogen populations, as illustrated by the example of *Zymoseptoria tritici*. This pathogen has evolved multiple independent populations that are resistant to fungicides and virulent on resistant cultivars (Croll and McDonald, 2017; McDonald et al., 2019; Yang et al., 2022). The breakdown of host resistance is devastating because it takes breeders many years to develop replacement cultivars and, in the meantime, it increases the already heavy reliance on fungicides. The intensive use of fungicides accelerates the evolution of resistant pathogen populations (McDonald et al., 2019), threatening food security worldwide.

2.2 Animal Pathogens

The animal kingdom is widely infested with fungi. Resistance to fungal infections in vertebrates depends on innate and acquired immunity supported by high body temperatures in endothermic mammals and birds (Casadevall, 2022; Fisher et al., 2020; Köhler et al., 2017). Of particular concern are fungi infecting nonthermoregulating amphibians (frogs, salamanders), reptiles (e.g., snakes) and hibernating mammalian hosts (e.g., bats). These emerging fungal diseases spread globally due to complex interactions between biotic and abiotic factors: international introduction of novel pathogens by humans or livestock into naïve wildlife populations (Bernard et al., 2017), followed by invasion across landscapes and host communities (Fisher et al., 2020; Köhler et al., 2017). Two fungal genera have recently caused overly aggressive infections associated with die-offs, population declines and, in some cases, species extinctions. These are the aquatic chytrid fungi *Batrachochytrium dendrobatidis* and *B. salamandrivorans*, which cause chytridiomycosis in amphibians (Fisher et al., 2020; Köhler et al., 2017; Taylor, 2015), and the ascomycete fungus *Pseudogymnoascus destructans*, which causes white-nose syndrome (WNS) in North American bats in hibernation (Bernard et al., 2017; Campbell et al., 2022; Fisher et al., 2012, 2020; Garzoli et al., 2021; Meteyer and Verant, 2018).

The first reports of infections with *B. dendrobatidis* date back to 1998. In the following decades, this fungus was found infecting species on all continents except Antarctica. Phylogenomic analyses identified the Korean peninsula as the place of origin of the panzootic. One of the lineages that escaped from Asia caused the outbreak of *B. salamandrivorans*, which was first described in the Netherlands in 2013 (Fisher et al., 2020). Other intercontinental and transnational introductions occurred for both batrachochytrids through the global amphibian trade and translocated species (Fisher et al., 2020). *B. dendrobatidis* has been shown to infect over 50% of all amphibian species, threaten 30% of amphibians worldwide and was capable of killing up to 80% of a population within 4−5 months of its introduction. Chytridiomycosis led to the extinction of 100 different species and represents the largest documented loss of biodiversity attributable to an invading microorganism (Bernard et al., 2017; Fisher et al., 2020).

The outbreak of white-nose syndrome (WNS) in North American bats can be similarly explained by the recent introduction of this pathogen from Eurasia (Bernard et al., 2017). While WNS has not been associated with mass mortality in Europe, the pathogen *P. destructans* has killed millions of

hibernating bats in North America. By 2020, the pathogen had spread to 38 US states and seven Canadian provinces (Bernard et al., 2017). Globally, *P. destructans* has been found in at least 62 bat species. In some cases, declines have exceeded 95% and entire populations have been extirpated (Bernard et al., 2017; Campbell et al., 2022). The disease may represent the largest mammalian die-off in recorded history, with significant ecological and economic consequences due to expected increases in populations of plant and forest pests (Garzoli et al., 2021; Meteyer and Verant, 2018). *P. destructans* is an obligate psychrophile (with the upper growth limit of about 20°C) found in cave soil. Since metabolically active bats have a body temperature of about 37°C, the fungus infects susceptible bat hosts during hibernation, when their body temperature drops to 7°C (Fisher et al., 2020).

Both batrachochytrids and *P. destructans* are cutaneous pathogens. Because the skin of amphibians and the wing skin of bats are mediators of complex physiological functions, damage to the skin has far-reaching physiological consequences, such as electrolyte imbalance leading to death by cardiac arrest in amphibians and chronic respiratory acidosis and hyperkalaemia in bats. These skin infections are causing catastrophic population declines through mechanisms previously unknown in human or veterinary medicine (Fisher et al., 2020).

3. Selected Examples of Human Pathogens

Among the expansive fungal kingdom, numbering over six million species, a modest 625 fungi species are recognized for their ability to infect vertebrates, with approximately 200−300 directly associated with humans either as commensals and members of our microbiome or as agents triggering infectious diseases (Fisher et al., 2020). Despite this limited subset, fungi provoke billions of infections globally, spanning from life-threatening invasive infections like fungemia and meningitis to chronic conditions such as pulmonary aspergillosis, asthma, oral and vaginal candidiasis, as well as skin, hair, and nail infections, impacting more than a billion individuals (Fisher et al., 2020; Ksiezopolska et al., 2021; Robbins et al., 2017; Strickland and Shi, 2021; World Health Organization, 2022). While a handful of fungi can afflict healthy individuals, the majority of invasive fungal infections manifest in patients with compromised innate or acquired immunity (Casadevall, 2022). These invasive fungal diseases impose a significant public health burden, with mortality rates ranging from 30% to 90%, depending on the specific pathogen and patient demographic (Fisher et al., 2020; Strickland and Shi, 2021).

For fungi to establish infection in humans, they must meet specific criteria: evasion or penetration of surface barriers, tissue breakdown, and resistance to immune defenses, including adaptation to elevated body temperatures (Casadevall, 2022; Fisher et al., 2020; Köhler et al., 2017). The dissemination of fungi constitutes a pivotal phase in the progression of invasive diseases. Fungi adeptly utilize various routes—hijacking phagocytes from initial sites of infection, direct entry into the bloodstream via transcytosis, paracellular crossing, or exploiting damaged barriers (Strickland and Shi, 2021). Such invasive behaviors are well-described in some *Aspergillus*, *Candida*, and *Cryptococcus* species, which account for over 90% of fatal mycotic infections in humans (Robbins et al., 2017). Four species from these genera, *Aspergillus fumigatus*, *Candida albicans*, *Candida auris*, and *Cryptococcus neoformans*, form the Critical Priority Group of fungal pathogens proposed by the World Health Organization (WHO) (World Health Organization, 2022).

3.1 Major Pathogens: *Candida*, *Cryptococcus*, *Aspergillus*

Candida species are among the most common opportunistic fungal pathogens in humans and are responsible for both superficial (mucosal and cutaneous) and systemic infections. Population-based studies have estimated the incidence rate of candidiasis to be 2−14 cases per 100,000 people, and candidemia (*Candida* bloodstream infection) affects more than 250,000 people worldwide each year and results in more than 50,000 deaths (Ksiezopolska et al., 2021; World Health Organization, 2022). *Candida albicans* remains the leading cause of these infections in North America and Europe, with a mortality from 20% to 50% (Robbins et al., 2017; World Health Organization, 2022). However, a rapid increase in infections has been observed in non-albicans species, some of which are intrinsically resistant to azole antifungals, including *Nakaseomyces glabrata* (*C. glabrata)*, *C. krusei*, *C. auris, C. tropicalis*, *and C. parapsilosis*. In the United States, over 30% of cases of candidemia are now caused by *N. glabrata.* Infections with *C. parapsilosis* and *C. tropicalis* are also increasingly common (Chen et al., 2018; Fisher et al., 2020; Friedman and Schwartz, 2019; Ksiezopolska et al., 2021; Nett, 2019; Robbins et al., 2017; Spivak and Hanson, 2018; Turner and Butler, 2014; World Health Organization, 2022). The complete list of *Candida* species reported as causative agents of human infections includes up to 30 different species belonging to different clades within Saccharomycotina (Gabaldón and Fairhead, 2019; Taylor, 2015).

Candida species primarily reside within mammalian hosts, thriving as commensals in the oral cavity, gastrointestinal tract, skin, and/or vagina, and are kept in check until the host's immune defenses are compromised. As commensals, the yeast form facilitates swift movement across mucosal surfaces using bodily fluids. Upon systemic infection, dissemination to target organs occurs rapidly through the bloodstream. Transitioning into its hyphal form, *Candida* can adhere to epithelial cell layers and various tissues, forming biofilms and triggering the lysis of macrophages and neutrophils (Köhler et al., 2017).

Stress resistance, thermotolerance, adherence to host cells, and evasion of the immune system are important virulence traits across *Candida* species. However, the species differ in their virulence and epidemiology. *C. parapsilosis* tends to infect infants younger than 1 year, while other *Candida* species, in particularly *C. glabrata*, are associated with infections in older patients, particularly those older than 64 years (Turner and Butler, 2014). Despite its previous name, *C. glabrata* (now *N. glabrata*) is more closely related to *Saccharomyces cerevisiae* than to other pathogens of the genus, suggesting that *N. glabrata* and *C. albicans* have acquired the ability to infect humans independently (Gabaldón and Carreté, 2016). The incidence of *N. glabrata* infections has steadily increased since the 1980s. The species can also be isolated from nonhuman sources and is spreading with global trade and climate change (Gabaldón and Fairhead, 2019; Roy and Thompson, 2015).

A recent emerging pathogen, *C. auris*, was first documented in Japan in 2009 and has since triggered outbreaks within healthcare settings around the world (Friedman and Schwartz, 2019; World Health Organization, 2022). This species is closely related to the *C. haemulonii* species complex, a group displaying occasional pathogenicity in humans and animals and exhibiting substantial resistance to antifungal treatments (Chybowska et al., 2020). In contrast to other *Candida* spp. that often originate from the microbiome of the patient, particularly from the gastrointestinal tract, *C. auris* has a propensity for skin colonization but is not typically found in the gastrointestinal environment. Remarkably, *C. auris* presently represents nearly one-fifth of candida isolates obtained from blood samples, surpassing the prevalence of *C. albicans*. *C. auris* strains are often resistant to multiple classes of

drugs, they can remain viable on plastic surfaces for more than 2 weeks, efficiently transmit from person to person, and incur an alarmingly high mortality rate ranging from 29% to 53% (Chybowska et al., 2020; Forsberg et al., 2019; Nett, 2019; Rhodes and Fisher, 2019; Spivak and Hanson, 2018; World Health Organization, 2022).

Cryptococcosis, caused by *Cryptococcus neoformans* and *Cr. gattii*, infects over one million people annually, with mortality rates as high as 70% in developing countries (Fisher et al., 2020; Robbins et al., 2017). These fungi primarily affect immunocompromised individuals, particularly those with HIV/AIDS. In these patients *Cr. neoformans* causes mortality ranging from 41% to 61%. Overall *Cr. gattii* accounts for 11%−33% cryptococcal infections, with mortality from 10% to 25% worldwide, three-quarters of them in sub-Saharan Africa. In comparison, the Asia and Pacific region has the second highest burden of disease, while Europe has only 4400 new cases and 1800 deaths per year (Edwards et al., 2021; Köhler et al., 2017; World Health Organization, 2022).

Cryptococcus spp. are cosmopolitan basidiomycetous yeasts. Only two species, *Cr. neoformans* and *Cr. gattii*, can cause the disease cryptococcosis when aerosolized spores and/or desiccated yeast cells (smaller than 5 μm) are inhaled. In immunocompetent individuals, they are either successfully cleared from the lungs or establish long-term latent infections. In immunocompromised individuals, the infection can spread to other organs, including the skin, often presenting as acute pneumonia or fatal meningitis (Köhler et al., 2017). *Cr. neoformans,* which causes the vast majority of cryptococcoses worldwide, inhabits avian excrement and soil, while *Cr. gattii* is found in the soil and on certain trees.

Among immunocompromised individuals, *Cr. neoformans* can cause symptomatic pulmonary infections leading to pneumonia, acute respiratory distress syndrome, and subsequent spread beyond the lungs (Strickland and Shi, 2021). This dissemination is thought to occur via several proposed pathways: intracellularly within phagocytes, transcellularly through epithelial cells, paracellularly between epithelial cells, and extracellularly through compromised epithelial barriers (Strickland and Shi, 2021). Once disseminated, *Cr. neoformans* enters the bloodstream, targeting distant organs, primarily the central nervous system (CNS), resulting in meningoencephalitis. *Cr. neoformans* are classified into five major molecular types: VNI, VNII, and VNB, collectively referred to as *Cr. neoformans* var. *grubii*; VNIV, also known as *Cr. neoformans* var. *neoformans*; and VNIII, a hybrid of the two varieties. Occasional interspecies hybrids have been documented. The predominant clinical burden worldwide stems from infections with *Cr. neoformans* var. *grubii*, accounting for nearly 95% of all cryptococcosis cases (Strickland and Shi, 2021). *Cr. neoformans* also infects a broad spectrum of animal hosts besides humans (Cuomo et al., 2018; Edwards et al., 2021; Köhler et al., 2017; May et al., 2016).

The first report of infection with *Cr. gattii* was in 1970 in Congo. Initially, the species was only found in tropical and subtropical climates and was mainly associated with eucalyptus. Nowadays, it is associated with more than 50 species of trees and has spread worldwide in temperate and developed areas. Infections that were originally reported from immunocompetent individuals also affect immunocompromised patients (Espinel-Ingroff and Kidd, 2015; May et al., 2016). Within *Cr. gattii*, several lineages are recognized: VGI, VGII, VGIII, VGIV, and the recently discovered and described lineage VGV, which is spreading worldwide (Edwards et al., 2021). These lineages have also been described as *Cr. gattii* s. str., *Cr. bacillisporus*, *Cr. deuterogattii, Cr. tetragattii*, and *Cr. decagattii* (Friedman and Schwartz, 2019).

Annually, over a quarter of a million instances of invasive aspergillosis are documented, leading to a 100% mortality rate in cases left untreated (Fisher et al., 2020; Robbins et al., 2017; Strickland and Shi, 2021). Recently, a rise in aspergilloses has been noted in people who are not immunocompromised but are recovering from severe viral respiratory infections such as influenza or SARS-CoV-2 (Drissi, 2021; Friedman and Schwartz, 2019).

Aspergilli, filamentous saprophytic fungi, are widespread, predominantly thriving in soil and decaying organic matter. Their impact on human health encompasses a spectrum of diseases collectively termed aspergillosis. These diseases exhibit diverse pathogenic mechanisms and clinical manifestations, ranging from aspergilloma (the development of a fungus ball) to a chronic inflammatory and fibrotic condition now categorized as chronic pulmonary aspergillosis (CPA). Among the causative agents, *A. fumigatus* stands as the primary culprit, responsible for the majority of fatalities and ranking second only to *Candida albicans* in infection rates.

Characteristic ecological features important for the pathogenicity of *A. fumigatus* are the ability to grow at 37°C, a large production of small, buoyant and hydrophobic conidia, the ability to sequester iron, a wide range of proteolytic enzymes and secondary metabolites, including gliotoxin, fumagillin, helvolic acid, verruculogen, trypacidin, and gliotoxin. Transmission of aspergillosis typically occurs via inhalation of minute airborne conidia that penetrate deep into the lungs (Latgé and Chamilos, 2019; Strickland and Shi, 2021). In healthy individuals they are quickly eliminated, whereas in immunocompromised patients they can germinate and grow as hyphae. The resulting invasive aspergillosis is characterized by tissue damage, necrosis, and hypoxia. From the lungs, the target organ, the hyphae can spread angiotropically and cause sepsis and dissemination to distant organs (Strickland and Shi, 2021).

Until the late 1990s, *A. fumigatus* was almost universally susceptible to triazole antifungals. In the following decade, multiple azole resistance was observed in clinical isolates. Azole resistance is now reported from six continents and includes *A. flavus,* the second most common cause of aspergillosis, with resistance rates as high as 5% (Friedman and Schwartz, 2019).

The pervasive emergence of antifungal resistance has evolved into a widespread concern across various fungal species, posing a significant challenge in combating fungal infections effectively. Each of the briefly presented major opportunistic invaders has become a dominant threat to human health, and the emergence of drug-resistant strains has been reported in all species (Robbins et al., 2017; World Health Organization, 2022).

3.2 Primary Pathogens: *Blastomyces*, *Coccidioidomycosis*, *Paracoccidioides*, *Histoplasma*, *Pneumocystis*

The pathogenic species of Onygenales comprises saprotrophic soil inhabitants, which can cause systemic infections in otherwise healthy individuals. Predominantly affiliated with the Ajellomycetaceae family, notable members include *Blastomyces dermatitidis*, *Histoplasma capsulatum, Coccidioides immitis, C. posadasii*, and *Paracoccidioides brasiliensis*. Recently, several other clinically relevant species have been described as causative agents of blastomycosis and histoplasmosis (Friedman and Schwartz, 2019). In the United States, these pathogens cause over 650,000 new infections each year (Van Dyke et al., 2019).

Phylogenomic studies reveal a shift in lifestyle from plant-based to carnivorous tendencies among these organisms, marked by the loss of cellulases and pectinases while gaining proteinases and

keratinases. While typically confined to specific geographic regions, recent findings suggest an expansion of their geographic range (Deepe, 2018; Friedman and Schwartz, 2019). They prefer particular soil consistency and specific host mammals, e.g., rodents for *Coccidioides*, bats and birds for *Histoplasma*, and armadillos for *Paracoccidioides* (Köhler et al., 2017; Vite-Garín et al., 2014). *Histoplasma* and *Blastomyces* pathogens generate airborne microconidia from specialized hyphae and *Coccidioides* and *Paracoccidioides* produce arthrospores, which transform into parasitic yeasts (or spherules in the case of *Coccidioides* spp.) in the human body and thus initiate an infection (Köhler et al., 2017; Kollath et al., 2019). Infections in immunocompromised hosts tend to be more clinically conspicuous, severe, and prone to multiorgan spread. The typical locations of the pathogens in the body include the mucous membranes of the oral cavity and respiratory tract for *Paracoccidioides*, bones, joints, and skin for *Blastomyces*, multiple organs (including the gastrointestinal tract, bones, and skin) for *Histoplasma*, and primarily the lungs, skin, deep organs, and infrequently the CNS for *Coccidioides* spp. (Deepe, 2018; Guerra et al., 2019; Köhler et al., 2017; Lockhart et al., 2021; Van Dyke et al., 2019; Vite-Garín et al., 2014).

The genus *Pneumocystis* belongs to the subphylum Taphrinomycotina, a distinct clade of fungi that diverged upon the separation of Ascomycota from Basidiomycota. *Pneumocystis* was first described in 1909 and until the late 1980s it was unclear whether it belonged to protozoa or fungi. *Pneumocystis* comprises uncultured, obligately biotrophic, highly diversified microfungal organisms that grow in the lungs of mammals as abundant ameboid cells with thin cell walls ("trophic forms") and as oval forms with thick walls ("cysts," now recognized as asci). *Pneumocystis* depends on the host for a broad range of metabolic functions, resulting in the loss of numerous metabolic and transporter-encoding genes (Köhler et al., 2017). The human-associated *Pneumocystis jirovecii* successfully adapted to colonize the lungs of both immunocompromised and healthy hosts, which serve as reservoirs for further infections. The incidence of *Pneumocystis* pneumonia has increased dramatically since the emergence of AIDS. Mortality ranges from 0% to 100% overall (highly variable) (Alanio and Bretagne, 2017; Chen et al., 2018; Cissé and Hauser, 2018; Fisher et al., 2020; Taylor, 2015; World Health Organization, 2022).

3.3 Emerging Fungal Pathogens

An array of phylogenetically diverse emerging fungal pathogenesis are undergoing rapid expansion in their geographic and host ranges, as well as in their incidence and virulence. So far, we have highlighted the increased incidence of *Nakaseomyces glabrata,* the recent global emergence of *C. auris* and the spread of *Cr. gattii* into nonendemic areas. Furthermore, an increasing number of reports describe atypical and unusually severe fungal infections related to the use of novel immunomodulators and viral disease (Friedman and Schwartz, 2019).

The implementation of antifungal prophylaxis among immunosuppressed individuals has contributed to the upsurge in invasive infections involving species such as *Apophysomyces* spp., *Lomentospora prolificans*, and *Fusarium* spp., particularly *F. proliferatum* and *F. solani* species complex (World Health Organization, 2022). Novel dimorphic fungal pathogens have surfaced, encompassing various species within the new genus *Emergomyces*, causing disseminated mycoses worldwide, particularly prevalent among HIV-infected patients. Additionally, *Blastomyces helicus* and *B. percursus* have emerged as causative agents of atypical blastomycoses in North America and Africa. In Brazil, zoonotic sporotrichosis attributed to *Sporothrix brasiliensis* has gained significance as a

disease affecting both cats and humans. The widespread utilization of topical antifungals containing corticosteroids in India has led to cases of persistent dermatophytosis caused by *Trichophyton mentagrophytes* (Friedman and Schwartz, 2019). In the future, indoor-related infections with black and red yeasts, such as *Exophiala dermatitidis* and *Rhodotorula mucilaginosa,* may be expected (Gostinčar et al., 2011; Novak Babič et al., 2018; Villar et al., 2012; Wirth and Goldani, 2012).

Of all the fungal infections reported in recent times, the most dramatic incidence was associated with COVID-19. The deadly second wave of the global COVID-19 pandemic has resulted in an unprecedented surge in mortality associated with mucormycosis, also referred to as the "black fungus." By May 2021, more than 40,000 cases of mucormycosis had been reported in India (Bhatia, 2022; World Health Organization, 2022). Infection begins with the entry of airborne sporangiospores into air-filled spaces, i.e., the sinuses and lungs. In a susceptible person, the spores germinate into hyphae, which then invade the surrounding tissue, including blood vessels, causing a hemorrhagic infarction. The name "black fungus" refers to the dark, necrotic tissue seen in affected individuals. From the nose and lungs, the infection can spread to the eyes and, in some cases, the brain. Poorly controlled diabetes mellitus, immunosuppression, prolonged and indiscriminate use of steroids and unhygienic environments are some of the commonly recognized risk factors. Sinus, pulmonary, and disseminated infections have been observed to have mortality rates of 46%, 76%, and 96%, respectively, with an overall mortality rate of 54% (Sharma et al., 2022). The causative agents of mucormycosis are various ubiquitous environmental Zygomycota molds found in soil and decaying organic matter. The main genera involved are *Mucor*, *Rhizopus*, and *Thamnostylum*, with *Rhizopus arrhizus* being the most important pathogen, followed by *Mucor circinelloides* (Bhatia, 2022; Drissi, 2021; Mahalaxmi et al., 2021; Pattnaik et al., 2021; Selarka et al., 2021; Sharma et al., 2022; Singh et al., 2021).

4. Facultative Pathogenesis and Adaptation Versus Exaptation

While several plant-infecting fungi are obligate pathogens that require a host to complete their life cycle (Lorrain et al., 2019), this is not true for fungal pathogens of animals, with few exceptions (Microsporidia, *Pneumocystis* spp.). For other fungi, the population outside the host is typically large and self-sufficient (Rokas, 2022). This has significant implications for the evolution of the pathogen and its virulence.

Time spent in the host is, by definition, important for evolutionary success of obligate pathogens. It can also play an important role in the survival of those facultative pathogens that cause frequent infections, or even for those that spend considerable time in nonpathogenic interactions with the host, e.g., as commensals (as is the case in some *Candida* spp.) or during persistence in healthy individuals (D'Enfert, 2009). Passage through a mammalian host has been shown to cause genomic and phenotypic changes in a pathogen (Sephton-Clark et al., 2023). This applies either to entire species or to phylogenetic lineages within these species. Some examples of such pathogens are described above. However, many of the opportunistic infections reported today involve causative agents that are very rarely recognized as clinically relevant. Numerous fungal species are implicated as pathogens in only a handful of case reports (De Pauw, 2011).

Host defense mechanisms exert selective pressure on the invading pathogen. Less well-adapted pathogens that cause only occasional infections are subject to higher stress and selection pressure, but their overall host populations are small, so they are less likely to undergo mutations that would

increase their virulence (Gostinčar et al., 2018). However, the major evolutionary difference between common and occasional pathogens probably lies in the fate of such mutations. An important factor in deciding this fate is the ability of the pathogen to escape from the host. In many cases, this ability remains unknown.

But even if cells carrying an adaptive mutation manage to escape into the environment, their long-term fate is uncertain. In the case of common pathogens, which only stay in the environment for a limited time before infecting the next host, the adaptive mutation may have sufficient opportunity to spread within the infecting population. In contrast, for occasional pathogens, the fate of the mutation is mainly determined by its performance in the population outside the host (Gostinčar et al., 2018). If the mutation has adaptive value outside the host, it would likely have evolved there in the first place. If the mutation is neutral outside the host, there is little chance that it will spread in a large pool of ancestral genotypes. And if the mutation is actually detrimental to survival outside the host, it will be rapidly selected out of the population. The suggestion that a passage in animal hosts might increase the fitness of environmental microbes in their out-of-host environment (Casadevall and Pirofski, 2007) therefore seems unlikely (Gostinčar et al., 2018).

It logically follows that the rarer the pathogen–host interactions are and the smaller the number of cells that escape the host (relative to the population of the pathogen outside the host), the less likely it is that virulence will develop as a response to interactions with the host. Therefore, many, if not most, animal fungal pathogens are unlikely to have evolved their virulence during infection (Casadevall and Pirofski, 2007; Gostinčar et al., 2018; Rokas, 2022). But how then can we explain the ability of these fungi to cause infection in the first place?

In recent years, several authors have suggested that the features that allow fungi to survive in an animal body are in most cases not adaptations but exaptations, i.e., they evolved for a different purpose and then proved useful during infection (Casadevall, 2007; Song et al., 2017). Several examples have been given, of which resistance to predation by wild amebae is perhaps the most intriguing, as it has been shown to select for several putative virulence factors (Albuquerque et al., 2019; Casadevall, 2012; Hillmann et al., 2015; Steenbergen et al., 2001) and can easily be argued to share similarities with phagocytosis, one of the pillars of the animal immune system. Environmental stress has been proposed as a selection pressure that increases the virulence of *Cryptococcus neoformans* (Brown et al., 2007; van Burik and Magee, 2001; Zaragoza, 2019) and the role of stress tolerance in pathogenesis has been extended to other fungal pathogens as well (Gostinčar et al., 2018), as discussed in more detail later in this chapter. The evolution of some traits predisposing the fungus for pathogenesis may also occur during nonpathogenic interactions between the fungus and its host, e.g., in commensals such as *Candida albicans* and *Malassezia* spp. (Kumamoto et al., 2020; Vijaya Chandra et al., 2021).

In all the cases described above, the ability of the pathogen to survive in the hostile environment created by the host's immune defenses may suggest a story of co-evolution, but in reality it is an example of ecological fitting. Casadevall and Pirofski (2007) proposed the term "accidental virulence" to describe a concept where the fungus finds itself in an animal body and is able to survive long enough to cause disease due to a combination of traits that have evolved outside the host.

The phenotype that enables survival in an animal body has evolved multiple times in different parts of the fungal tree of life (Rokas, 2022). However, this does not mean that every fungal species is equally likely to become pathogenic. The transition to a pathogenic lifestyle has only been reported in a minority of fungal orders (Gostinčar et al., 2018), in some of them occurring independently in several related species (Rokas, 2022). The genomic toolbox supporting the transition is thought to have been

shaped by the specific evolutionary history of some phylogenetic lineages and is correlated with thermotolerance (Bergman and Casadevall, 2010) and general stress tolerance (Gostinčar et al., 2018)—both discussed in more detail below.

5. Virulence Factors

5.1 Conventional Virulence Factors

Researchers have struggled to identify conventional virulent factors in fungi that infect animals. Even in notorious human pathogens, only a handful of traits have been found that are clearly linked to virulence (Casadevall and Pirofski, 2014)—a stark contrast to bacterial pathogens. This is even more true for opportunistic pathogens, which rarely cause infections (Gostinčar et al., 2018). Many of these traits can be easily explained as adaptations to an environment outside the host, which is in line with the concept of "accidental virulence" described above. For example, hydrolytic enzymes are often associated with pathogenesis, but as a broad repertoire of extracellular enzymatic activities is typical of many fungi, it is difficult to interpret these as adaptations to pathogenesis. The production of the pigment melanin protects against various types of stress and is characteristic of a large group of "black fungi" which contains (opportunistic) pathogenic species, but many more nonpathogenic species. Survival of iron limitation is crucial for growth in natural oligotrophic environments. Selected examples of well-studied pathogens are listed below.

Virulence factors of *Candida* spp. include membrane and cell wall barriers disruption, dimorphism, biofilm formation, stress tolerance, hydrolytic enzymes, but also one of the few true "bacteria-like" virulence factors: candidalysin, a toxin that facilitates tissue invasion (reviewed by Staniszewska (2020)). Biofilm formation, adherence, and hydrophobicity have been proposed as virulence factors of *Malassezia furfur* (Angiolella et al., 2018).

Cryptococcus neoformans produces various hydrolases (including a urease), a polysaccharide capsule (to protect against phagocytosis, desiccation, and free radicals) and melanin (to protect cells from various types of environmental stress). Both the polysaccharides of the capsule and melanin are thought to modulate the host immune response (reviewed by Zaragoza (2019)). They help the fungus to survive in and escape from phagocytes, but they also help it to survive predation by amebae (Casadevall, 2012). Melanin production additionally protects against ultraviolet light, oxidative stress, mechanical stress, and temperature fluctuations, and increases resistance to high temperature and drugs in the host (Edwards et al., 2021; Mattoon et al., 2023). Laccase production supports lignin degradation in the environment and protects against oxidative bursts in the host (Edwards et al., 2021). Demonstrating the potential of high-throughput sequencing and comparative genomics in studying virulence factors, the genome-wide association study of *Cr. neoformans* strains identified several genetic variations associated with specific clinical outcomes (Sephton-Clark et al., 2022).

In *Aspergillus fumigatus*, genes under positive selection have been reported to be involved in signal transduction, metabolism, mitochondrial activity, regulation of transcription, but also in siderophore biosynthesis, stress tolerance and thermotolerance. No positive selection was observed for gliotoxin, one of the most important virulence factors of *A. fumigatus*. This was explained by the amebicidal activity of this metabolite, a property that is also important for nonpathogenic members of the genus (e.g., it is also produced by *Aspergillus fischeri*) (Hillmann et al., 2015; Knowles et al., 2020).

In comparison to the few true virulence factors in fungal pathogens of animals, many virulence factors of phytopathogens are known, including toxins that often show a striking host specificity (Pontes et al., 2020). Enzymes for the degradation of plant compounds are also crucial for pathogenesis, and although decomposition of plant biomass is a common feature of fungi, plant pathogens show characteristic enrichment of some enzymatic activities (Doehlemann et al., 2017; Ohm et al., 2012). Many plant pathogens (but also some nonpathogens) can form appressoria, specialized structures used to invade the plant (Chethana et al., 2021).

5.2 Thermotolerance

In the great variety of fungi that infect warm-blooded animals, the one characteristic that unites them—and at the same time distinguishes them from most other fungi—is thermotolerance. Thermotolerance in fungi has evolved repeatedly, but remains limited to a relatively small number of species (Robert et al., 2015). This may explain why the number of fungal pathogens of plants and insects is orders of magnitude greater than the number of fungal pathogens of mammals (Robert and Casadevall, 2009). In other words, endothermy creates a thermal restriction zone for about 95% of fungi (Robert and Casadevall, 2009). At the level of fungal orders, there is an almost complete correlation between thermotolerance and clinical occurrence (Gostinčar et al., 2018). Concerns have been raised that ongoing global climate change will act as a selection pressure, leading to more fungi surviving higher temperatures and the emergence of new diseases (Casadevall, 2020; Robert et al., 2015; Steffen et al., 2022). However, at the taxonomic level, fungal opportunism is also correlated with tolerance to low temperatures (Gostinčar et al., 2018). While this may seem unexpected, it supports the suggestion that opportunistic potential is not only associated with thermotolerance, but with good stress tolerance in general. Polyextremotolerant opportunistic pathogens are therefore sometimes isolated from unexpected environments, such as Arctic glaciers (Perini et al., 2019).

5.3 Extremotolerance

High temperature may be a major barrier to fungal infections in mammals, but it is only one of many defense mechanisms employed by the animal body. Low water activity and low pH inhibit skin colonization (Elias, 2007), reactive oxygen and reactive nitrogen species result in oxidative stress (Warris and Ballou, 2019), limited availability of iron, zinc, and manganese lead to nutritional immunity (Crawford and Wilson, 2015), to name just a few barriers to infection.

Immune defenses thus create an inhospitable environment in which any surviving microorganism is subjected to significant stress. Efficient stress responses are thus critical for fungal pathogens (Brown et al., 2017; Day and Quinn, 2019). Of course, many types of stress to which the pathogen is exposed in the host do not only occur during infection. In their environment outside the host, fungi are repeatedly exposed to low water activity, oxidative stress, oligotrophic conditions, and other stressors. Since for the vast majority of fungi the population outside the host is incomparably more important for the evolution of the species, the stress tolerance of fungal pathogens has most likely evolved independently of the host (Rokas, 2022).

Do fungi with more efficient mechanisms for coping with stress have a higher pathogenic potential? There are several lines of evidence to support this conjecture. A relationship between osmotolerance and pathogenicity has been suggested (de Hoog et al., 2005) and later confirmed as

statistically significant (Gostinčar et al., 2018). Like thermotolerance, osmotolerance appears to have evolved independently in various parts of the fungal tree of life (de Hoog et al., 2005; Gostinčar et al., 2018). The authors of the second study (Gostinčar et al., 2018) distinguished between two types of stress-tolerant fungi. Monoextremotolerance, a specialization to a specific type of stress and with low potential for habitat shifts, was not associated with pathogenic potential. Polyextremotolerance, i.e., tolerance to a wide range of stress types combined with high adaptability and high potential for habitat shifts, was significantly associated with opportunism. Polyextremotolerant fungi excel at colonizing novel habitats, especially those that are generally suboptimal for microbial growth, and infection of a mammalian body can be viewed simply as survival in another new and inhospitable habitat (Gostinčar et al., 2011, 2018). Other phenotypic traits, such as the ability to assimilate monoaromatic hydrocarbons, have been reported to further increase the pathogenic potential of fungi (Lavrin et al., 2020; Prenafeta-Boldu et al., 2006).

Interestingly, the polyextremotolerance of fungi overlaps strongly not only with the phenotype that enables mammalian pathogenicity but also with the phenotype that is desired in biocontrol agents used in agriculture, justifying some caution in the selection of fungal strains used in the fast-growing biocontrol industry (Zajc et al., 2019).

5.4 Resistance to Antifungals

The evolution of drug-resistant fungal pathogens limits the already small arsenal of antifungal drugs available for use in clinical practice. The mechanisms of resistance are diverse and often similar to those that confer antibiotic resistance to bacteria, from mutation or overexpression of the target to increased expression of multidrug transporters or activation of stress response mechanisms (Fisher et al., 2022; Lee et al., 2023; Revie et al., 2018; Shapiro et al., 2011).

Azole resistance is most commonly acquired by mutation or overexpression of the gene encoding the target of the drug, P450-dependent lanosterol 14-α-demethylase, while accumulation of sterols other than ergosterol can confer resistance to both azoles and polyenes (Robbins et al., 2017). Several ATP-binding cassette (ABC) transporters have also been linked to azole resistance, either through upregulation of gene transcription or through increased gene dosage (reviewed by Revie et al. (2018)). The extensive use of azoles in agriculture and cross-resistance between agricultural and medicinal azoles is a cause of concern (Jørgensen and Heick, 2021). Resistance to echinocandins, which inhibit the (1,3)-β-D-glucan synthase, is achieved by mutation of the enzyme or by changes in other pathways of cell wall synthesis or remodeling, e.g., upregulation of chitin synthases (reviewed by Revie et al. (2018)).

The mutations conferring resistance can be diverse, as shown by numerous cases of experimental evolution, but target a limited number of genes (Ksiezopolska et al., 2021). Evolution of multiple antifungal resistance can involve hypermutator strains (Healey et al., 2016). Genomic plasticity and epigenomic changes also play a role in the development of resistance (Lee et al., 2023). Some pathogens are innately resistant to some of the antifungal agents (Revie et al., 2018). Finally, the buildup of resistance as a response to selection pressures imposed by the use of antimicrobials in nonclinical settings has been recognized as especially important in case of fungal opportunists, which are widespread in the environment and thus fit well into the framework of One Health (Fisher et al., 2022; Rhodes et al., 2022).

Antifungal resistance is often accompanied by reduced fitness in a drug-free medium (and possibly but not necessarily by reduced virulence (Duxbury et al., 2020; Lee et al., 2023)), a situation sometimes ameliorated by compensatory mutations and sometimes leading to loss of resistance without the ongoing selective pressure of the drug (Lee et al., 2023; Robbins et al., 2017).

The evolution of antifungal resistance is a complex and dynamic area of research, which can only be summarized here in broad strokes. However, this brief description should sufficiently illustrate that while this evolution is of critical importance to clinical practice, it is also very different from the evolution of fungal pathogenesis. Antifungal resistance can develop rapidly and with few mutations, sometimes within the time frame of a single infection (Beekman and Ene, 2020). While hostile conditions within the host can also cause adaptation in relatively short time frames, as discussed below, the basis of pathogenesis is much more complex and usually cannot be reduced to a few causative genes.

6. Host and Environment Populations, and Emergence of New Fungal Pathogens

6.1 Population Genetics and Genomics

Population genetics has been a valuable tool in epidemiology for decades. The increasing ease of whole genome sequencing and the extension of population genetics methods to whole genomes greatly enhance the power of these methods and give us unprecedented insight into the introduction, spread and evolution of pathogens. Pathogens have received a disproportionate amount of attention in fungal population genomics, with whole genomes of dozens, sometimes even hundreds, of fungal isolates being compared.

However, despite the increasing research on population structure and comparative genomics of fungal pathogens, population genomics has revealed evidence of pathogen−host co-evolution in only a handful of fungal species. This is perhaps most clearly seen in *Candida albicans*, a species with a highly plastic genome that adapts to its environment through a combination of mitotic recombination, changes in ploidy, including aneuploidy, and loss of heterozygosity (Hirakawa et al., 2015; Mixão et al., 2021). When this yeast was exposed to oxidative stress in vitro, this resulted in recombination that produced short regions of lost heterozygosity (Forche et al., 2011), and the same was observed during growth in the host (Forche, 2014). Adaptation at the level of genetic variation has also been reported for several virulence factors (reviewed by Sitterlé et al. (2019)).

Candida auris appears to have emerged independently in different geographical locations, forming five known clades. Particularly worrying is its unusually good resistance to antifungal drugs. The species is characterized by high thermotolerance and relatively good halotolerance and can persist for long periods on inanimate surfaces (reviewed by Du et al. (2020)). Many isolates from patients from the same and sometimes even from distant institutions were shown to be of clonal origin, indicating efficient transmission of the strains (Chybowska et al., 2020). The tropical coast has been suggested as the original habitat of the species (Arora et al., 2021).

In contrast, *Nakaseomyces glabrata* may be an opportunistic colonizer rather than a specialized commensal and its population is diverse both genotypically and phenotypically, but with no discernible geographical structure (Gabaldón and Fairhead, 2019). Even the clinical isolates from a geographically limited sampling are genetically diverse (Helmstetter et al., 2022). It has been suggested that

diversifying lineages can coexist in a single individual even when the yeast is a commensal (Sitterlé et al., 2019). Interestingly, a study of in-host microevolution identified polymorphisms in cell surface proteins, adhesins, and genes involved in antifungal resistance (Helmstetter et al., 2022) but it is not clear whether these adaptations are evolutionarily relevant.

While some *Candida* species show host adaptation at the genomic level and also show evidence of host-to-host transmission, there is little in most other pathogenic fungi to suggest a similar strategy. Most population genomic studies of fungal pathogens have failed to demonstrate separation of clinical and environmental strains, supporting the long-standing hypothesis that environmental strains are the source of infection. For example, genome comparison confirmed that the clinical *Candida krusei* and *Pichia kudriavzevii* used in biotechnology are actually a single species and that the clinical strains do not segregate into separate clades (Douglass et al., 2018).

Population genomics of *Aspergillus fumigatus* detected three primary populations that rarely recombine, differ in gene groups possibly linked to environmental niche specialization and differ in antifungal resistance genes and alleles (Lofgren et al., 2022). Strains isolated from the UK and Ireland cluster into two clades, with the majority of azole resistant strains confined to one clade (Rhodes et al., 2022). However, no genetic differentiation was observed between clinical and environmental isolates, and they were found in both clades. The extent of recombination between clades was limited. Resistance to azoles was likely acquired in the environment before infections, presumably due to the widespread use of azole fungicides in agriculture (Rhodes et al., 2022).

Analysis of global and African populations of *Cr. neoformans* revealed two African subpopulations, one enriched in clinical strains and the other showing greater melanization and resistance to oxidative stress—an unexpected finding, explained by the authors of the study with greater amount of stress in the environment than in the human host. *Cr. neoformans* var. *grubii* was divided into four lineages, with evidence of recombination within, but not between, groups (Desjardins et al., 2017). A genome-wide association analysis of differences between lineages revealed the location of several polymorphisms in the loci of known virulence factors, and some lineages showed a loss of diversity consistent with repeated population bottlenecks (Desjardins et al., 2017). Sequencing of African and Asian isolates showed that the species recently underwent a population expansion with low recombination rates and high levels of long-distance geographic transmission, suggesting efficient out-of-host dispersal. In this study, three subclades of one lineage accounted for the vast majority of clinical isolates (Ashton et al., 2019). During an investigation of four cases of cryptococcosis in a single hospital in Glasgow, no environmental strains could be isolated, but none of the four clinical isolates of *Cr. neoformans* were clonally related (Farrer et al., 2021). Similarly, *Cr. gattii* is composed of multiple clades containing a mixture of clinical, environmental and veterinary isolates (Firacative et al., 2016). A new clade was found by sampling environments associated with a small mammal in Zambia (Farrer et al., 2019), demonstrating the importance of environmental sampling to maximize our understanding of pathogen origin and evolution.

Sequencing of *Pneumocystis* species revealed a low level of recombination, but also considerable diversity with only hints of geographic structure. This suggests that species diverged before the divergence of their hosts (Cissé et al., 2018). Comparative genomics suggested several genomic changes accompanying the shift to mammal-associated lifestyle and later to specific hosts (Cissé et al., 2021).

If most relatively common fungal pathogens, such as those discussed so far, appear to have an environmental source of infection and follow the concept of accidental virulence, according to which

their virulence factors have evolved in response to environmental challenges and not due to interaction with the host, there are even stronger reasons to believe that this is also true for more exotic and emerging fungal pathogens. The latter cause infections under very specific circumstances, their environmental populations are many orders of magnitude larger than those within the host, and there is little evidence that the infecting cells can ever escape from the host. It is therefore highly unlikely that their evolution will lead to true pathogenesis, at least not through co-evolution with their host. However, adaptation to novel anthropogenic conditions, for example through exposure to agricultural antifungals (Rhodes et al., 2022), increased global temperatures (Casadevall, 2020) or novel conditions in our homes (Gostinčar et al., 2011), may lead to an increase in their pathogenicity potential.

While humans and other mammals, and even more so birds, are relatively well protected against fungal infections by the high body temperature, this is not the case for other animals and also not for plants. Perhaps the most striking example of the danger posed by fungi to cold-blooded animals after an introduction of a pathogen to a new geographical area is *Batrachochytrium dendrobatidis*. Early population genetic studies of this fungus, which caused a global decline in amphibian populations, suggested that the pathogen had low diversity, no host specificity or geographic structure, but some local recombining populations (Morgan et al., 2007). However, a comparison at the genome level revealed three lineages, only one of which was globally distributed (Farrer et al., 2011). The following study reported a deeper phylogenetic structure that predates the modern amphibian decline (Rosenblum et al., 2013). None of these studies clarified the geographic origin of *B. dendrobatidis* until finally O'Hanlon et al. (2018) discovered a fourth lineage within the species and located the origin of the pathogen in Asia, where the characteristic lack of lethal outbreaks of amphibian disease indicates a long history of pathogen-host co-evolution (Fisher and Garner, 2020). The onset of the outbreak has been roughly estimated to have occurred in the first half of the twentieth century and with several introductions of the pathogen, possibly through global trade in amphibian species (O'Hanlon et al., 2018).

The European epidemic of *Batrachochytrium salamandrivorans* is caused by unexpectedly diverse strains that diverged at least 50,000 years ago, also suggesting multiple introductions of the pathogen. The species has also been shown to have (isolate-specific) saprotrophic potential enabling it to survive in the environment, which explains its ability to cause host extinction, typical of pathogens where species survival is decoupled from population density of the host (Kelly et al., 2021).

In plants, there are numerous cases of devastating damage caused by newly introduced pathogens or examples of pathogen-host co-evolution that reflect the difference between opportunistic fungal pathogens of humans and primary pathogens of plants. Only a few examples are given here. *Zymoseptoria tritici*, an important pathogen of wheat, has emerged in parallel with the domestication of its host (Stukenbrock et al., 2011). Analysis of isolates from a single field revealed a highly diverse population on a small area with evidence of extensive recombination (Singh et al., 2021). Two clades of the maze pathogen *Ustilago maydis* were found in Mexico, both showing signs of bottleneck overlapping with the time of maize domestication, with a high rate of adaptive mutations in virulence-related genomic regions (Schweizer et al., 2021). Three distinct clades were observed in *Magnaporthe oryzae*, one of which was diverse and the other two originated as clonal lineages no more than 400 years ago, while isolates infecting rice overall split from strains infecting wild relatives of rice at a time corresponding to the period of rice domestication (Zhong et al., 2018). *Hymenoscyphus fraxineus* has been introduced from Asia to Europe twice. The host European ash trees are susceptible, resulting in massive ash dieback (McMullan et al., 2018). The strains of the chestnut blight fungus *Cryphonectria*

parasitica that caused an outbreak of the disease in Europe originated from the diverse European population of the fungus and are primarily clonal and dominated by a single mating type. This asexuality has been suggested as a factor in the evolution of the highly invasive lineage of the pathogen, but it also increases the susceptibility of *C. parasitica* to mycoviral infection (Stauber et al., 2021). Finally, Dutch elm disease is caused by three groups of strains of *Ophiostoma ulmi* and two subspecies of *Ophiostoma novo-ulmi*, which may occasionally hybridize. The resulting introgression is important for the diversification of pathogens (Hessenauer et al., 2020).

6.2 Recombination and Hybridization

The understanding of fungal sexuality has evolved from the assumption that many fungi are strictly asexual to the realization that most species recombine at least occasionally (Taylor et al., 2015). Rather than dividing them into two clear categories, sexual and strictly clonal, it is more accurate to view fungi as a continuous spectrum that lies between the two categories. In addition to the conventional sexual cycle of meiosis and cell fusion, fungi can use several other strategies to exchange genetic material, as will be explained later.

It is unclear what level of sexual reproduction (if any) is ideal for fungal pathogens. The general advantages and disadvantages of recombination are well-known. On the one hand, sexual cycles generate diversity and are promoted under stressful conditions (Nielsen and Heitman, 2007). Increased genetic diversity provides a substrate for natural selection, increasing the chance of better-adapted offspring and the elimination of deleterious mutations. The heterogeneity of pathogenic strains can be clinically relevant and influence the outcome of infection (Rokas, 2022). On the other hand, clonal reproduction preserves advantageous combinations of interacting alleles, the break-up of which is a problem known as recombination load (Otto, 2009; Sun and Heitman, 2011). By foregoing a mating system and sexual structures, energy can be conserved (Sun and Heitman, 2011; Xu, 2005). Asexual reproduction allows for rapid proliferation and may be unavoidable when strains of one mating type are absent or there is an imbalance of mating types. This may occur, for example, after the introduction of a single mating type of the pathogen into a new geographical area where mate search becomes impossible or too costly. This has been observed particularly in some plant pathogens (Paoletti et al., 2005; Stauber et al., 2021), but is unlikely to be significant in occasional human pathogens with a large environmental reservoir. However, strictly clonal propagation carries the risk of accumulating deleterious mutations, a process known as Muller's Ratchet. Furthermore, clonal lineages may be limited in their adaptability, although diversity can also be generated by increased mutation rates, mitotic recombination, genome restructuring or changes in transposon activity (reviewed by Gostinčar et al. (2018)). While successful clonal lineages can prevail temporarily, their long-term survival is doubtful (Drenth et al., 2019).

Population genomic analyses revealed evidence of recombination in many presumably clonal species, including those that can infect animals. *N. glabrata* was thought to be asexual, but recent genomic studies suggest that this may not be the case (Gabaldón and Fairhead, 2019). In *B. dendrobatidis*, evidence of recombination has been found (Fisher and Garner, 2020), despite previous reports of widespread mitotic reproduction (Rosenblum et al., 2013). Even in *Pneumocystis* species, which are mostly restricted to living in the lungs of their hosts and show low levels of recombination, genetic diversity is surprisingly high, possibly due to interactions during multiclonal infections (Cissé et al., 2018).

As mentioned earlier, sexual reproduction is not the only way for fungi to exchange genetic material. *C. albicans* can undergo a well-studied parasexual cycle in which diploid cells fuse to form a tetraploid that reverts to a diploid state through random chromosome loss (as opposed to carefully orchestrated and high-fidelity meiosis). During this process, mitotic recombination often occurs, resulting in further diversity (Mishra et al., 2021). Parasexuality appears to facilitate adaptation in the survival of *A. fumigatus* in the lungs of cystic fibrosis patients (Engel et al., 2020), but the population genomics analysis of this species detected unexpectedly high levels of recombination although the fungus disperses primarily via asexual conidia (Lofgren et al., 2022). While parasexuality is usually followed by a return to the original ploidy through a series of aneuploid generations, in some cases cell fusion leads to the formation of a stable increased ploidy. Such hybridization has received increased recognition in recent years. *C. albicans* itself has a hybrid ancestor with unknown parental lineages, which then diverged through extensive loss of heterozygosity (Mixão and Gabaldón, 2020), mitotic recombination events, and changes in ploidy including aneuploidy (Hickman et al., 2015). *Candida inconspicua*, *Candida orthopsilosis*, and *Candida metapsilosis* are all hybrids of medical relevance, in some cases resulting from multiple hybridization events (Mixão et al., 2019; Pryszcz et al., 2015; Schröder et al., 2016). Several hybrids have been described in the *Cr. gattii/Cr. neoformans* species complex (Hagen et al., 2015). Hybrids have also been found in *Candida tropicalis*, but these mostly originate from the environment and are therefore probably not associated with increased pathogenicity (O'Brien et al., 2021).

Evidence of hybridization can easily be overlooked, even after genome sequencing. It is likely that careful population genomic analyses will reveal more hybridization events in fungi in general and in fungal pathogens in particular. As shown in the example of *Hortaea werneckii* and *Aureobasidium melanogenum*, two rare opportunistic pathogens, hybrids can occur repeatedly even in otherwise strictly clonal species, forming stable hybrid lineages that are isolated more frequently than their nonhybrid progenitor lineages (Gostinčar et al., 2022). The role of such a reproductive strategy in fungal pathogenicity is not yet fully understood, but hybridization (and aneuploidy) has been shown to increase under environmental stress, in response to new environmental conditions (Gerstein and Sharp, 2021; Gilchrist and Stelkens, 2019; Hill et al., 2021; Naranjo−Ortiz and Gabaldón, 2020). The diversity generated by hybridization has been shown to be of clinical relevance (reviewed by Naranjo−Ortiz and Gabaldón (2020)). In addition to combining genetic material from otherwise separate phylogenetic lineages, hybridization by definition also leads to altered ploidy. This has effects that are typical of any change in ploidy: changes in cell size and shape, changes in gene dosage, masking of deleterious mutations and an improved ability to repair DNA damage (Chen et al., 2012; Gerstein and Sharp, 2021; Naranjo−Ortiz and Gabaldón, 2020; Todd et al., 2017). There is growing evidence that ploidy variation is a widespread transient adaptation of fungi to novel conditions (reviewed by Naranjo−Ortiz and Gabaldón (2020)).

6.3 Emergence of New Fungal Pathogens

The number of fungal infections in humans is increasing (Vallabhaneni et al., 2016). This is not the case only for true pathogens that may co-evolve with the host and develop increased pathogenicity, but also for rare opportunistic pathogens where such co-evolution is unlikely, as discussed above. What, apart from the evolution of increased pathogenicity and improved diagnostics, is leading to a global explosion of mycoses? The answers fall into three categories: (i) changes in the host population; (ii)

changes in the environment outside the host; (iii) changes in the nature and frequency of fungus−host interactions.

Advances in modern medicine may extend people's lifespans, but they do not always improve their health. Organ transplants and cancer therapies have greatly improved survival for many patients but are also leading to a growing cohort of immunosuppressed individuals, either intentionally (to prevent rejection of transplanted organs) or as a side effect (e.g., of chemotherapy or radiotherapy), increasing susceptibility to infection (Papon et al., 2021; Ruhnke et al., 2020). The aging population itself provides more opportunities for fungal infections, as people live many more years with deteriorating health than in the past. The spread of human immunodeficiency virus (HIV) has increased the incidence of fatal fungal infections, which account for nearly half of HIV-related deaths worldwide (Fisher et al., 2016). Secondary fungal infections are also important in patients with COVID-19 and affect a significant proportion of mechanically ventilated patients (Verweij and van de Veerdonk, 2022).

While it is unlikely that most opportunistic pathogens co-evolve with humans, this does not mean that they do not evolve toward greater pathogenicity at all. For example, Casadevall (2020) suggested that global warming could lead to the selection of more thermotolerant microorganisms that could overcome the thermal restriction zone, which is one of the main defense mechanisms of mammals against fungal infections. Localized changes in our environment could also be important. Selection pressures introduced to our habitats may lead to the selection of traits relevant to virulence. Several thermotolerant opportunistic fungi have been consistently found in household dishwashers around the world (Gümral et al., 2016; Zalar et al., 2011; Zupančič et al., 2016) and in other indoor environments (Hamada and Abe, 2010; Moat et al., 2016; Novak Babič et al., 2020; Wang et al., 2018). Growth on aromatic hydrocarbons has been associated with certain patterns of infection in mammals (Prenafeta-Boldú et al., 2018). Hygiene measures used to prevent microbial overgrowth in our homes, by definition, expose the surviving indoor microbiome to high levels of stress, and as described above, adaptation to stressful environments can lead to exaptations promoting opportunistic infections (Gostinčar et al., 2011, 2018).

Even if hostile conditions in our immediate environment do not actively shape the evolution of microorganisms, they do select for species and strains that are particularly capable of surviving stress. By trying to make our homes inhospitable to microorganisms, we have opened up new opportunities for the most adaptable and resilient species. These are adept at habitat switches and often have the potential to cause opportunistic infections in humans (Gostinčar et al., 2011, 2015). Some studies even suggest that cleaning and urbanization are associated with increased fungal diversity (McCall et al., 2020). As opportunistic fungal pathogens number in the hundreds of species, even though many of them are extremely rare pathogens (De Pauw, 2011), their accumulation and enrichment in our immediate environment may contribute to the growing number of mycoses simply by increasing the frequency of their interactions with their potential hosts.

7. Conclusions

The animal pathogenesis of most fungi differs substantially from primary pathogenesis, which evolves through frequent or even obligatory interaction of the pathogen with the host. While some species, particularly commensals such as *Candida* spp. show evidence of co-evolution between pathogen and host, most fungi infecting animals invade the host due to their thermotolerance, high stress tolerance

and ability of habitat shifts. This phenotype is generally thought to have evolved in response to selection in the environment rather than selection in the host. It is therefore unlikely that opportunistic fungal pathogens gradually evolve higher virulence and progress towards true, primary pathogenicity over a series of infections. This is in stark contrast to fungal pathogens of plants, which often follow more conventional scenarios of epidemiology and evolution, evolving their virulence through repeated interactions with the host and often causing outbreaks after invasion of a new area.

In the great diversity of fungi, the ability to infect an animal body, especially a warm-blooded one, remains rare. However, anthropogenic changes such as increases in global temperature and changes in the indoor microbiota may be driving fungal evolution toward increased pathogenesis or simply enriching our immediate environment with potentially harmful species. The opportunistic nature of most fungal infections and the lack of a clear evolutionary progression toward greater pathogenicity should therefore not be confused with a lack of medical relevance. The growing burden of mycoses clearly proves otherwise. To limit the dangers posed by opportunistic fungal infections in the future, we need to apply epidemiological approaches that differ substantially from those used to combat conventional pathogens.

Acknowledgments

This study was supported by funding from the Slovenian Research and Innovation Agency to Infrastructural Center Mycosmo (MRIC UL, I0-0022), and programmes P4-0432 and P1-0198.

Declaration of Competing Interest

The authors declare no conflict of interest.

References

Alanio, A., Bretagne, S., 2017. *Pneumocystis jirovecii* detection in asymptomatic patients: what does its natural history tell us? F1000Research 6, 739. https://doi.org/10.12688/f1000research.10619.1.

Albuquerque, P., Nicola, A.M., Magnabosco, D.A.G., Derengowski, L. da S., Crisóstomo, L.S., Xavier, L.C.G., Frazão, S. de O., Guilhelmelli, F., de Oliveira, M.A., Dias, J. do N., Hurtado, F.A., Teixeira, M. de M., Guimarães, A.J., Paes, H.C., Bagagli, E., Felipe, M.S.S., Casadevall, A., Silva-Pereira, I., 2019. A hidden battle in the dirt: soil amoebae interactions with *Paracoccidioides* spp. PLoS Neglected Trop. Dis. 13, e0007742. https://doi.org/10.1371/journal.pntd.0007742.

Alves, R., Barata-Antunes, C., Casal, M., Brown, A.J.P., Van Dijck, P., Paiva, S., 2020. Adapting to survive: how *Candida* overcomes host-imposed constraints during human colonization. PLoS Pathog. 16, e1008478. https://doi.org/10.1371/journal.ppat.1008478.

Angiolella, L., Leone, C., Rojas, F., Mussin, J., de los Angeles Sosa, M., Giusiano, G., 2018. Biofilm, adherence, and hydrophobicity as virulence factors in *Malassezia furfur*. Med. Mycol. 56, 110−116. https://doi.org/10.1093/mmy/myx014.

Arora, P., Singh, P., Wang, Y., Yadav, A., Pawar, K., Singh, A., Padmavati, G., Xu, J., Chowdhary, A., 2021. Environmental isolation of Candida auris from the coastal wetlands of Andaman Islands, India. mBio 12. https://doi.org/10.1128/mBio.03181-20.

Ashton, P.M., Thanh, L.T., Trieu, P.H., Van Anh, D., Trinh, N.M., Beardsley, J., Kibengo, F., Chierakul, W., Dance, D.A.B., Rattanavong, S., Davong, V., Hung, L.Q., Chau, N.V.V., Tung, N.L.N., Chan, A.K., Thwaites, G.E., Lalloo, D.G., Anscombe, C., Nhat, L.T.H., Perfect, J., Dougan, G., Baker, S., Harris, S., Day, J.N., 2019. Three phylogenetic groups have driven the recent population expansion of *Cryptococcus neoformans*. Nat. Commun. 10, 2035. https://doi.org/10.1038/s41467-019-10092-5.

Bebber, D.P., Field, E., Gui, H., Mortimer, P., Holmes, T., Gurr, S.J., 2019. Many unreported crop pests and pathogens are probably already present. Global Change Biol. 25, 2703−2713. https://doi.org/10.1111/gcb.14698.

Beekman, C.N., Ene, I.V., 2020. Short-term evolution strategies for host adaptation and drug escape in human fungal pathogens. PLoS Pathog. 16, e1008519. https://doi.org/10.1371/journal.ppat.1008519.

Bergman, A., Casadevall, A., 2010. Mammalian endothermy optimally restricts fungi and metabolic costs. mBio 1. https://doi.org/10.1128/mBio.00212-10.

Bernard, R.F., Willcox, E.V., Parise, K.L., Foster, J.T., McCracken, G.F., 2017. White-nose syndrome fungus, *Pseudogymnoascus destructans*, on bats captured emerging from caves during winter in the southeastern United States. BMC Zool. 2, 12. https://doi.org/10.1186/s40850-017-0021-2.

Bhatia, M., 2022. The rise of mucormycosis in Covid-19 patients in India. Expert Rev. Anti-infect. Ther. 20, 137−138. https://doi.org/10.1080/14787210.2021.1960822.

Bongomin, F., Gago, S., Oladele, R., Denning, D., 2017. Global and multi-national prevalence of fungal diseases—estimate precision. J. Fungi 3, 57. https://doi.org/10.3390/jof3040057.

Brown, A.J.P., Cowen, L.E., di Pietro, A., Quinn, J., 2017. Stress adaptation. Microbiol. Spectr. 5. https://doi.org/10.1128/microbiolspec.FUNK-0048-2016.

Brown, G.D., Denning, D.W., Gow, N.A.R., Levitz, S.M., Netea, M.G., White, T.C., 2012. Hidden killers: human fungal infections. Sci. Transl. Med. 4. https://doi.org/10.1126/scitranslmed.3004404.

Brown, S.M., Campbell, L.T., Lodge, J.K., 2007. *Cryptococcus neoformans*, a fungus under stress. Curr. Opin. Microbiol. 10, 320−325. https://doi.org/10.1016/j.mib.2007.05.014.

Campbell, C.J., Nelson, D.M., Gates, J.E., Gibbs, H.L., Stevenson, E.R., Johnson, B., Nagel, J., Trott, R., Wieringa, J.G., Vander Zanden, H.B., 2022. White-nose syndrome pathogen *Pseudogymnoascus destructans* detected in migratory tree-roosting bats. J. Wildl. Dis. https://doi.org/10.7589/JWD-D-21-00160.

Casadevall, A., 2022. Immunity to invasive fungal diseases. Annu. Rev. Immunol. 40, 121−141. https://doi.org/10.1146/annurev-immunol-101220-034306.

Casadevall, A., 2020. Climate change brings the specter of new infectious diseases. J. Clin. Invest. 130, 553−555. https://doi.org/10.1172/JCI135003.

Casadevall, A., 2012. Amoeba provide insight into the origin of virulence in pathogenic fungi. Adv. Exp. Med. Biol. 710, 1−10. https://doi.org/10.1007/978-1-4419-5638-5_1.

Casadevall, A., 2007. Determinants of virulence in the pathogenic fungi. Fungal Biol. Rev. 21, 130−132. https://doi.org/10.1016/j.fbr.2007.02.007.

Casadevall, A., Pirofski, L.A., 2014. Microbiology: ditch the term pathogen. Nature 516, 165−166. https://doi.org/10.1038/516165a.

Casadevall, A., Pirofski, L.A., 2007. Accidental virulence, cryptic pathogenesis, Martians, lost hosts, and the pathogenicity of environmental microbes. Eukaryot. Cell 6, 2169−2174. https://doi.org/10.1128/Ec.00308-07.

Chaloner, T.M., Gurr, S.J., Bebber, D.P., 2021. Plant pathogen infection risk tracks global crop yields under climate change. Nat. Clim. Change 11, 710−715. https://doi.org/10.1038/s41558-021-01104-8.

Chen, G., Rubinstein, B., Li, R., 2012. Whole chromosome aneuploidy: big mutations drive adaptation by phenotypic leap. Bioessays 34, 893−900. https://doi.org/10.1002/bies.201200069.

Chen, M., Xu, Y., Hong, N., Yang, Y., Lei, W., Du, L., Zhao, J., Lei, X., Xiong, L., Cai, L., Xu, H., Pan, W., Liao, W., 2018. Epidemiology of fungal infections in China. Front. Med. 12, 58−75. https://doi.org/10.1007/s11684-017-0601-0.

Chethana, K.W.T., Jayawardena, R.S., Chen, Y.-J., Konta, S., Tibpromma, S., Abeywickrama, P.D., Gomdola, D., Balasuriya, A., Xu, J., Lumyong, S., Hyde, K.D., 2021. Diversity and function of appressoria. Pathogens 10, 746. https://doi.org/10.3390/pathogens10060746.

Chybowska, A.D., Childers, D.S., Farrer, R.A., 2020. Nine things genomics can tell us about *Candida auris*. Front. Genet. 11. https://doi.org/10.3389/fgene.2020.00351.

Cissé, O.H., Hauser, P.M., 2018. Genomics and evolution of *Pneumocystis species*. Infect. Genet. Evol. 65, 308−320. https://doi.org/10.1016/j.meegid.2018.08.015.

Cissé, O.H., Ma, L., Dekker, J.P., Khil, P.P., Youn, J.-H., Brenchley, J.M., Blair, R., Pahar, B., Chabé, M., Van Rompay, K.K.A., Keesler, R., Sukura, A., Hirsch, V., Kutty, G., Liu, Y., Peng, L., Chen, J., Song, J., Weissenbacher-Lang, C., Xu, J., Upham, N.S., Stajich, J.E., Cuomo, C.A., Cushion, M.T., Kovacs, J.A., 2021. Genomic insights into the host specific adaptation of the *Pneumocystis* genus. Commun. Biol. 4, 1−14. https://doi.org/10.1038/s42003-021-01799-7.

Cissé, O.H., Ma, L., Wei Huang, D., Khil, P.P., Dekker, J.P., Kutty, G., Bishop, L., Liu, Y., Deng, X., Hauser, P.M., Pagni, M., Hirsch, V., Lempicki, R.A., Stajich, J.E., Cuomo, C.A., Kovacs, J.A., 2018. Comparative population genomics analysis of the mammalian fungal pathogen *Pneumocystis*. mBio 9, e00381-18. https://doi.org/10.1128/mBio.00381-18.

Crawford, A., Wilson, D., 2015. Essential metals at the host-pathogen interface: nutritional immunity and micronutrient assimilation by human fungal pathogens. FEMS Yeast Res. fov071. https://doi.org/10.1093/femsyr/fov071.

Croll, D., McDonald, B.A., 2017. The genetic basis of local adaptation for pathogenic fungi in agricultural ecosystems. Mol. Ecol. 26, 2027−2040. https://doi.org/10.1111/mec.13870.

Cuomo, C.A., Rhodes, J., Desjardins, C.A., 2018. Advances in Cryptococcus genomics: insights into the evolution of pathogenesis. Memórias do Instituto Oswaldo Cruz 113. https://doi.org/10.1590/0074-02760170473.

Day, A.M., Quinn, J., 2019. Stress-activated protein kinases in human fungal pathogens. Front. Cell. Infect. Microbiol. 9. https://doi.org/10.3389/fcimb.2019.00261.

de Hoog, G.S., Zalar, P., Gerrits van den Ende, A.H.G., Gunde-Cimerman, N., van den Ende, B.G., Gunde-Cimerman, N., 2005. Relation of halotolerance to human pathogenicity in the fungal tree of life: an overview of ecology and evolution under stress. In: Gunde-Cimerman, N., Oren, A., Plemenitas, A. (Eds.), Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya. Springer, Dordrecht, The Netherlands, pp. 373−395.

De Pauw, B.E., 2011. What are fungal infections? Mediterr. J. Hematol. Infect. Dis. 3, e2011001. https://doi.org/10.4084/mjhid.2011.001.

Deepe, G.S., 2018. Outbreaks of histoplasmosis: the spores set sail. PLoS Pathog. 14, e1007213. https://doi.org/10.1371/journal.ppat.1007213.

D'Enfert, C., 2009. Hidden killers: persistence of opportunistic fungal pathogens in the human host. Curr. Opin. Microbiol. 12, 358−364. https://doi.org/10.1016/j.mib.2009.05.008.

Desjardins, C.A., Giamberardino, C., Sykes, S.M., Yu, C.-H., Tenor, J.L., Chen, Y., Yang, T., Jones, A.M., Sun, S., Haverkamp, M.R., Heitman, J., Litvintseva, A.P., Perfect, J.R., Cuomo, C.A., 2017. Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. Genome Res. 27, 1207−1219. https://doi.org/10.1101/gr.218727.116.

Doehlemann, G., Ökmen, B., Zhu, W., Sharon, A., 2017. Plant pathogenic fungi. Microbiol. Spectr. 5. https://doi.org/10.1128/microbiolspec.FUNK-0023-2016.

Douglass, A.P., Offei, B., Braun-Galleani, S., Coughlan, A.Y., Martos, A.A.R., Ortiz-Merino, R.A., Byrne, K.P., Wolfe, K.H., 2018. Population genomics shows no distinction between pathogenic *Candida krusei* and environmental *Pichia kudriavzevii*: one species, four names. PLoS Pathog. 14, e1007138. https://doi.org/10.1371/journal.ppat.1007138.

Drenth, A., McTaggart, A.R., Wingfield, B.D., 2019. Fungal clones win the battle, but recombination wins the war. IMA Fungus 10, 1−6. https://doi.org/10.1186/s43008-019-0020-8.

Drissi, C., 2021. Black fungus, the darker side of COVID-19. J. Neuroradiol. 48, 317−318. https://doi.org/10.1016/j.neurad.2021.07.003.

Du, H., Bing, J., Hu, T., Ennis, C.L., Nobile, C.J., Huang, G., 2020. *Candida auris*: epidemiology, biology, antifungal resistance, and virulence. PLoS Pathog. 16, e1008921. https://doi.org/10.1371/journal.ppat.1008921.

Duxbury, S.J.N., Bates, S., Beardmore, R.E., Gudelj, I., 2020. Evolution of drug-resistant and virulent small colonies in phenotypically diverse populations of the human fungal pathogen *Candida glabrata*. Proc. Biol. Sci. 287, 20200761. https://doi.org/10.1098/rspb.2020.0761.

Edel-Hermann, V., Lecomte, C., 2019. Current status of *Fusarium oxysporum formae speciales* and races. Phytopathology 109, 512−530. https://doi.org/10.1094/PHYTO-08-18-0320-RVW.

Edwards, H.M., Cogliati, M., Kwenda, G., Fisher, M.C., 2021. The need for environmental surveillance to understand the ecology, epidemiology and impact of *Cryptococcus* infection in Africa. FEMS Microbiol. Ecol. 97. https://doi.org/10.1093/femsec/fiab093.

Elias, P.M., 2007. The skin barrier as an innate immune element. Semin. Immunopathol. 29, 3−14. https://doi.org/10.1007/s00281-007-0060-9.

Engel, T., Verweij, P.E., van den Heuvel, J., Wangmo, D., Zhang, J., Debets, A.J.M., Snelders, E., 2020. Parasexual recombination enables *Aspergillus fumigatus* to persist in cystic fibrosis. ERJ Open Res. 6, 1−10. https://doi.org/10.1183/23120541.00020-2020.

Espinel-Ingroff, A., Kidd, S., 2015. Current trends in the prevalence of *Cryptococcus gattii* in the United States and Canada. Infect. Drug Resist. 89. https://doi.org/10.2147/IDR.S57686.

Fantozzi, E., Kilaru, S., Gurr, S.J., Steinberg, G., 2021. Asynchronous development of *Zymoseptoria tritici* infection in wheat. Fungal Genet. Biol. 146, 103504. https://doi.org/10.1016/j.fgb.2020.103504.

Farrer, R.A., Borman, A.M., Inkster, T., Fisher, M.C., Johnson, E.M., Cuomo, C.A., 2021. Genomic epidemiology of a Cryptococcus neoformans case cluster in Glasgow, Scotland, 2018. Microb. Genom. 7. https://doi.org/10.1099/mgen.0.000537.

Farrer, R.A., Chang, M., Davis, M.J., van Dorp, L., Yang, D.-H., Shea, T., Sewell, T.R., Meyer, W., Balloux, F., Edwards, H.M., Chanda, D., Kwenda, G., Vanhove, M., Chang, Y.C., Cuomo, C.A., Fisher, M.C., Kwon-Chung, K.J., 2019. A new lineage of Cryptococcus gattii (VGV) discovered in the central Zambezian Miombo woodlands. mBio 10. https://doi.org/10.1128/mBio.02306-19.

Farrer, R.A., Weinert, L.A., Bielby, J., Garner, T.W.J., Balloux, F., Clare, F., Bosch, J., Cunningham, A.A., Weldon, C., du Preez, L.H., Anderson, L., Pond, S.L.K., Shahar-Golan, R., Henk, D.A., Fisher, M.C., 2011. Multiple emergences of genetically diverse amphibian-infecting chytrids include a globalized hypervirulent recombinant lineage. Proc. Natl. Acad. Sci. USA 108, 18732−18736. https://doi.org/10.1073/pnas.1111915108.

Firacative, C., Roe, C.C., Malik, R., Ferreira-Paim, K., Escandón, P., Sykes, J.E., Castañón-Olivares, L.R., Contreras-Peres, C., Samayoa, B., Sorrell, T.C., Castañeda, E., Lockhart, S.R., Engelthaler, D.M., Meyer, W., 2016. MLST and whole-genome-based population analysis of *Cryptococcus gattii* VGIII links clinical, veterinary and environmental strains, and reveals divergent serotype specific sub-populations and distant ancestors. PLoS Neglected Trop. Dis. 10, e0004861. https://doi.org/10.1371/journal.pntd.0004861.

Fisher, M.C., Alastruey-Izquierdo, A., Berman, J., Bicanic, T., Bignell, E.M., Bowyer, P., Bromley, M., Brüggemann, R., Garber, G., Cornely, O.A., Gurr, S.J., Harrison, T.S., Kuijper, E., Rhodes, J., Sheppard, D.C., Warris, A., White, P.L., Xu, J., Zwaan, B., Verweij, P.E., 2022. Tackling the emerging threat of antifungal resistance to human health. Nat. Rev. Microbiol. 20, 557−571. https://doi.org/10.1038/s41579-022-00720-1.

Fisher, M.C., Garner, T.W.J., 2020. Chytrid fungi and global amphibian declines. Nat. Rev. Microbiol. 18, 332−343. https://doi.org/10.1038/s41579-020-0335-x.

Fisher, M.C., Gow, N.A.R., Gurr, S.J., 2016. Tackling emerging fungal threats to animal health, food security and ecosystem resilience. Phil. Trans. Biol. Sci. 371, 20160332. https://doi.org/10.1098/rstb.2016.0332.

Fisher, M.C., Gurr, S.J., Cuomo, C.A., Blehert, D.S., Jin, H., Stukenbrock, E.H., Stajich, J.E., Kahmann, R., Boone, C., Denning, D.W., Gow, N.A.R., Klein, B.S., Kronstad, J.W., Sheppard, D.C., Taylor, J.W., Wright, G.D., Heitman, J., Casadevall, A., Cowen, L.E., 2020. Threats posed by the fungal kingdom to humans, wildlife, and agriculture. mBio 11. https://doi.org/10.1128/mBio.00449-20.

Fisher, M.C., Henk, D.A., Briggs, C.J., Brownstein, J.S., Madoff, L.C., McCraw, S.L., Gurr, S.J., 2012. Emerging fungal threats to animal, plant and ecosystem health. Nature 484, 186−194. https://doi.org/10.1038/nature10947.

Fones, H.N., Bebber, D.P., Chaloner, T.M., Kay, W.T., Steinberg, G., Gurr, S.J., 2020. Threats to global food security from emerging fungal and oomycete crop pathogens. Nat. Food 1, 332−342. https://doi.org/10.1038/s43016-020-0075-0.

Forche, A., 2014. Large-scale chromosomal changes and associated fitness consequences in pathogenic fungi. Curr. Fungal Infect. Rep. 8, 163−170. https://doi.org/10.1007/s12281-014-0181-2.

Forche, A., Abbey, D., Pisithkul, T., Weinzierl, M.A., Ringstrom, T., Bruck, D., Petersen, K., Berman, J., 2011. Stress alters rates and types of loss of heterozygosity in *Candida albicans*. mBio 2. https://doi.org/10.1128/mBio.00129-11.

Forsberg, K., Woodworth, K., Walters, M., Berkow, E.L., Jackson, B., Chiller, T., Vallabhaneni, S., 2019. *Candida auris*: the recent emergence of a multidrug-resistant fungal pathogen. Med. Mycol. 57, 1−12. https://doi.org/10.1093/mmy/myy054.

Friedman, Schwartz, 2019. Emerging fungal infections: new patients, new patterns, and new pathogens. J. Fungi 5, 67. https://doi.org/10.3390/jof5030067.

Gabaldón, T., Carreté, L., 2016. The birth of a deadly yeast: tracing the evolutionary emergence of virulence traits in *Candida glabrata*. FEMS Yeast Res. 16, fov110. https://doi.org/10.1093/femsyr/fov110.

Gabaldón, T., Fairhead, C., 2019. Genomes shed light on the secret life of *Candida glabrata*: not so asexual, not so commensal. Curr. Genet. 65, 93−98. https://doi.org/10.1007/s00294-018-0867-z.

Garzoli, L., Bozzetta, E., Varello, K., Cappelleri, A., Patriarca, E., Debernardi, P., Riccucci, M., Boggero, A., Girometta, C., Picco, A.M., 2021. White-nose syndrome confirmed in Italy: a preliminary assessment of its occurrence in bat species. J. Fungi 7, 192. https://doi.org/10.3390/jof7030192.

Gerstein, A.C., Sharp, N.P., 2021. The population genetics of ploidy change in unicellular fungi. FEMS Microbiol. Rev. 45, 1−20. https://doi.org/10.1093/femsre/fuab006.

Gilchrist, C., Stelkens, R., 2019. Aneuploidy in yeast: segregation error or adaptation mechanism? Yeast 3427. https://doi.org/10.1002/yea.3427.

Gostinčar, C., Grube, M., Gunde-Cimerman, N., 2011. Evolution of fungal pathogens in domestic environments? Fungal Biol. 115, 1008−1018. https://doi.org/10.1016/j.funbio.2011.03.004.

Gostinčar, C., Gunde-Cimerman, N., Grube, M., 2015. Polyextremotolerance as the fungal answer to changing environments. In: Bakermans, C. (Ed.), Microbial Evolution Under Extreme Conditions. De Gruyter, Berlin, München, Boston, pp. 185−208. https://doi.org/10.1515/9783110340716.185.

Gostinčar, C., Sun, X., Černoša, A., Fang, C., Gunde-Cimerman, N., Song, Z., 2022. Clonality, inbreeding, and hybridization in two extremotolerant black yeasts. GigaScience 11, 1−16. https://doi.org/10.1093/gigascience/giac095.

Gostinčar, C., Zajc, J., Lenassi, M., Plemenitaš, A., de Hoog, S., Al-Hatmi, A.M.S., Gunde-Cimerman, N., 2018. Fungi between extremotolerance and opportunistic pathogenicity on humans. Fungal Divers. 93, 195−213. https://doi.org/10.1007/s13225-018-0414-8.

Grünwald, N.J., LeBoldus, J.M., Hamelin, R.C., 2019. Ecology and evolution of the sudden oak death pathogen *Phytophthora ramorum*. Annu. Rev. Phytopathol. 57, 301−321. https://doi.org/10.1146/annurev-phyto-082718-100117.

Guerra, B.T., Almeida-Silva, F., Almeida-Paes, R., Basso, R.P., Bernardes, J.P.R.A., Almeida, M.A., Damasceno, L.S., Xavier, M.O., Wanke, B., Zancopé-Oliveira, R.M., de Melo Teixeira, M., 2019. Histoplasmosis outbreaks in Brazil: lessons to learn about preventing exposure. Mycopathologia. https://doi.org/10.1007/s11046-019-00389-w.

Gümral, R., Özhak-Baysan, B., Tümgör, A., Saraçlı, M.A., Yıldıran, Ş.T., Ilkit, M., Zupančič, J., Novak-Babič, M., Gunde-Cimerman, N., Zalar, P., de Hoog, G.S., 2016. Dishwashers provide a selective extreme environment for human-opportunistic yeast-like fungi. Fungal Divers. 76, 1–9. https://doi.org/10.1007/s13225-015-0327-8.

Hagen, F., Khayhan, K., Theelen, B., Kolecka, A., Polacheck, I., Sionov, E., Falk, R., Parnmen, S., Lumbsch, H.T., Boekhout, T., 2015. Recognition of seven species in the *Cryptococcus gattii/Cryptococcus neoformans* species complex. Fungal Genet. Biol. 78, 16–48. https://doi.org/10.1016/j.fgb.2015.02.009.

Hamada, N., Abe, N., 2010. Comparison of fungi found in bathrooms and sinks. Biocontrol Sci. 15, 51–56. https://doi.org/10.4265/bio.15.51.

Healey, K.R., Zhao, Y., Perez, W.B., Lockhart, S.R., Sobel, J.D., Farmakiotis, D., Kontoyiannis, D.P., Sanglard, D., Taj-Aldeen, S.J., Alexander, B.D., Jimenez-Ortigosa, C., Shor, E., Perlin, D.S., 2016. Prevalent mutator genotype identified in fungal pathogen *Candida glabrata* promotes multi-drug resistance. Nat. Commun. 7, 11128. https://doi.org/10.1038/ncomms11128.

Helmstetter, N., Chybowska, A.D., Delaney, C., Da Silva Dantas, A., Gifford, H., Wacker, T., Munro, C., Warris, A., Jones, B., Cuomo, C.A., Wilson, D., Ramage, G., Farrer, R.A., 2022. Population genetics and microevolution of clinical *Candida glabrata* reveals recombinant sequence types and hyper-variation within mitochondrial genomes, virulence genes, and drug targets. Genetics 221. https://doi.org/10.1093/genetics/iyac031.

Hessenauer, P., Fijarczyk, A., Martin, H., Prunier, J., Charron, G., Chapuis, J., Bernier, L., Tanguay, P., Hamelin, R.C., Landry, C.R., 2020. Hybridization and introgression drive genome evolution of Dutch elm disease pathogens. Nat. Ecol. Evol. 4, 626–638. https://doi.org/10.1038/s41559-020-1133-6.

Hickman, M.A., Paulson, C., Dudley, A., Berman, J., 2015. Parasexual ploidy reduction drives population heterogeneity through random and transient aneuploidy in *Candida albicans*. Genetics 200, 781–794. https://doi.org/10.1534/genetics.115.178020.

Hill, R., Leitch, I.J., Gaya, E., 2021. Targeting Ascomycota genomes: what and how big? Fungal Biol. Rev. 36, 52–59. https://doi.org/10.1016/j.fbr.2021.03.003.

Hillmann, F., Novohradská, S., Mattern, D.J., Forberger, T., Heinekamp, T., Westermann, M., Winckler, T., Brakhage, A.A., 2015. Virulence determinants of the human pathogenic fungus *Aspergillus fumigatus* protect against soil amoeba predation. Environ. Microbiol. 17, 2858–2869. https://doi.org/10.1111/1462-2920.12808.

Hirakawa, M.P., Martinez, D.A., Sakthikumar, S., Anderson, M.Z., Berlin, A., Gujja, S., Zeng, Q., Zisson, E., Wang, J.M., Greenberg, J.M., Berman, J., Bennett, R.J., Cuomo, C.A., 2015. Genetic and phenotypic intraspecies variation in *Candida albicans*. Genome Res. 25, 413–425. https://doi.org/10.1101/gr.174623.114.

Jørgensen, L.N., Heick, T.M., 2021. Azole use in agriculture, horticulture, and wood preservation – is it indispensable? Front. Cell. Infect. Microbiol. 11. https://doi.org/10.3389/fcimb.2021.730297.

Kelly, M., Pasmans, F., Muñoz, J.F., Shea, T.P., Carranza, S., Cuomo, C.A., Martel, A., 2021. Diversity, multifaceted evolution, and facultative saprotrophism in the European *Batrachochytrium salamandrivorans* epidemic. Nat. Commun. 12, 6688. https://doi.org/10.1038/s41467-021-27005-0.

Knowles, S.L., Mead, M.E., Silva, L.P., Raja, H.A., Steenwyk, J.L., Goldman, G.H., Oberlies, N.H., Rokas, A., 2020. Gliotoxin, a known virulence factor in the major human pathogen *Aspergillus fumigatus*, is also biosynthesized by its nonpathogenic relative *Aspergillus fischeri*. mBio 11. https://doi.org/10.1128/mBio.03361-19.

Köhler, J.R., Hube, B., Puccia, R., Casadevall, A., Perfect, J.R., 2017. Fungi that infect humans. Microbiol. Spectr. 5. https://doi.org/10.1128/microbiolspec.FUNK-0014-2016.

Kollath, D.R., Miller, K.J., Barker, B.M., 2019. The mysterious desert dwellers: *Coccidioides immitis* and *Coccidioides posadasii*, causative fungal agents of coccidioidomycosis. Virulence 10, 222−233. https://doi.org/10.1080/21505594.2019.1589363.

Ksiezopolska, E., Schikora-Tamarit, M.À., Beyer, R., Nunez-Rodriguez, J.C., Schüller, C., Gabaldón, T., 2021. Narrow mutational signatures drive acquisition of multidrug resistance in the fungal pathogen *Candida glabrata*. Curr. Biol. 31, 5314−5326.e10. https://doi.org/10.1016/j.cub.2021.09.084.

Kumamoto, C.A., Gresnigt, M.S., Hube, B., 2020. The gut, the bad and the harmless: *Candida albicans* as a commensal and opportunistic pathogen in the intestine. Curr. Opin. Microbiol. 56, 7−15. https://doi.org/10.1016/j.mib.2020.05.006.

Latgé, J.-P., Chamilos, G., 2019. *Aspergillus fumigatus* and aspergillosis in 2019. Clin. Microbiol. Rev. 33. https://doi.org/10.1128/CMR.00140-18.

Lavrin, T., Konte, T., Kostanjšek, R., Sitar, S., Sepčič, K., Prpar Mihevc, S., Žagar, E., Županski, V., Lenassi, M., Rogelj, B., Gunde Cimerman, N., 2020. The neurotropic black yeast *Exophiala dermatitidis* induces neurocytotoxicity in neuroblastoma cells and progressive cell death. Cells 9, 963. https://doi.org/10.3390/cells9040963.

Lee, Y., Robbins, N., Cowen, L.E., 2023. Molecular mechanisms governing antifungal drug resistance. NPJ Antimicrob Res. 1, 1−9. https://doi.org/10.1038/s44259-023-00007-2.

Lockhart, S.R., Toda, M., Benedict, K., Caceres, D.H., Litvintseva, A.P., 2021. Endemic and other dimorphic mycoses in the Americas. J. Fungi 7, 151. https://doi.org/10.3390/jof7020151.

Lofgren, L.A., Ross, B.S., Cramer, R.A., Stajich, J.E., 2022. The pan-genome of *Aspergillus fumigatus* provides a high-resolution view of its population structure revealing high levels of lineage-specific diversity driven by recombination. PLoS Biol. 20, e3001890. https://doi.org/10.1371/journal.pbio.3001890.

Lorrain, C., Gonçalves dos Santos, K.C., Germain, H., Hecker, A., Duplessis, S., 2019. Advances in understanding obligate biotrophy in rust fungi. New Phytol. 222, 1190−1206. https://doi.org/10.1111/nph.15641.

Mahalaxmi, I., Jayaramayya, K., Venkatesan, D., Subramaniam, M.D., Renu, K., Vijayakumar, P., Narayanasamy, A., Gopalakrishnan, A.V., Kumar, N.S., Sivaprakash, P., Sambasiva Rao, K.R.S., Vellingiri, B., 2021. Mucormycosis: an opportunistic pathogen during COVID-19. Environ. Res. 201, 111643. https://doi.org/10.1016/j.envres.2021.111643.

Mattoon, E.R., Cordero, R.J.B., Casadevall, A., 2023. Melaninization reduces *Cryptococcus neoformans* susceptibility to mechanical stress. mSphere 8. https://doi.org/10.1128/msphere.00591-22.

May, R.C., Stone, N.R.H., Wiesner, D.L., Bicanic, T., Nielsen, K., 2016. *Cryptococcus*: from environmental saprophyte to global pathogen. Nat. Rev. Microbiol. 14, 106−117. https://doi.org/10.1038/nrmicro.2015.6.

McCall, L.-I., Callewaert, C., Zhu, Q., Song, S.J., Bouslimani, A., Minich, J.J., Ernst, M., Ruiz-Calderon, J.F., Cavallin, H., Pereira, H.S., Novoselac, A., Hernandez, J., Rios, R., Branch, O.H., Blaser, M.J., Paulino, L.C., Dorrestein, P.C., Knight, R., Dominguez-Bello, M.G., 2020. Home chemical and microbial transitions across urbanization. Nat. Microbiol. 5, 108−115. https://doi.org/10.1038/s41564-019-0593-4.

McCoy, A.G., Belanger, R.R., Bradley, C.A., Cerritos-Garcia, D.G., Garnica, V.C., Giesler, L.J., Grijalba, P.E., Guillin, E., Henriquez, M.A., Kim, Y.M., Malvick, D.K., Matthiesen, R.L., Mideros, S.X., Noel, Z.A., Robertson, A.E., Roth, M.G., Schmidt, C.L., Smith, D.L., Sparks, A.H., Telenko, D.E.P., Tremblay, V., Wally, O., Chilvers, M.I., 2023. A global-temporal analysis on *Phytophthora sojae* resistance-gene efficacy. Nat. Commun. 14, 6043. https://doi.org/10.1038/s41467-023-41321-7.

McDonald, M.C., Renkin, M., Spackman, M., Orchard, B., Croll, D., Solomon, P.S., Milgate, A., 2019. Rapid parallel evolution of azole fungicide resistance in Australian populations of the wheat pathogen *Zymoseptoria tritici*. Appl. Environ. Microbiol. 85. https://doi.org/10.1128/AEM.01908-18.

McMullan, M., Rafiqi, M., Kaithakottil, G., Clavijo, B.J., Bilham, L., Orton, E., Percival-Alwyn, L., Ward, B.J., Edwards, A., Saunders, D.G.O., Garcia Accinelli, G., Wright, J., Verweij, W., Koutsovoulos, G., Yoshida, K., Hosoya, T., Williamson, L., Jennings, P., Ioos, R., Husson, C., Hietala, A.M., Vivian-Smith, A., Solheim, H.,

MaClean, D., Fosker, C., Hall, N., Brown, J.K.M., Swarbreck, D., Blaxter, M., Downie, J.A., Clark, M.D., 2018. The ash dieback invasion of Europe was founded by two genetically divergent individuals. Nat. Ecol. Evol. 2, 1000−1008. https://doi.org/10.1038/s41559-018-0548-9.

Meteyer, C., Verant, M., 2018. White-nose syndrome: cutaneous invasive ascomycosis in hibernating bats. In: Miller, E.R., Lamberski, N., Calle, P. (Eds.), Fowler's Zoo and Wild Animal Medicine Current Therapy. Elsevier, pp. 508−513.

Mishra, A., Forche, A., Anderson, M.Z., 2021. Parasexuality of *Candida* species. Front. Cell. Infect. Microbiol. 11. https://doi.org/10.3389/fcimb.2021.796929.

Mixão, V., Gabaldón, T., 2020. Genomic evidence for a hybrid origin of the yeast opportunistic pathogen *Candida albicans*. BMC Biol. 18, 48. https://doi.org/10.1186/s12915-020-00776-6.

Mixão, V., Hansen, A.P., Saus, E., Boekhout, T., Lass-Florl, C., Gabaldón, T., 2019. Whole-genome sequencing of the opportunistic yeast pathogen *Candida inconspicua* uncovers its hybrid origin. Front. Genet. 10. https://doi.org/10.3389/fgene.2019.00383.

Mixão, V., Saus, E., Boekhout, T., Gabaldón, T., 2021. Extreme diversification driven by parallel events of massive loss of heterozygosity in the hybrid lineage of *Candida albicans*. Genetics 217. https://doi.org/10.1093/genetics/iyaa004.

Moat, J., Rizoulis, A., Fox, G., Upton, M., 2016. Domestic shower hose biofilms contain fungal species capable of causing opportunistic infection. J. Water Health 14, 727−737. https://doi.org/10.2166/wh.2016.297.

Morgan, J.A.T., Vredenburg, V.T., Rachowicz, L.J., Knapp, R.A., Stice, M.J., Tunstall, T., Bingham, R.E., Parker, J.M., Longcore, J.E., Moritz, C., Briggs, C.J., Taylor, J.W., 2007. Population genetics of the frog-killing fungus *Batrachochytrium dendrobatidis*. Proc. Natl. Acad. Sci. USA 104, 13845−13850. https://doi.org/10.1073/pnas.0701838104.

Naranjo−Ortiz, M.A., Gabaldón, T., 2020. Fungal evolution: cellular, genomic and metabolic complexity. Biol. Rev. 95, 1198−1232. https://doi.org/10.1111/brv.12605.

Nett, J.E., 2019. *Candida auris*: an emerging pathogen "incognito"? PLoS Pathog. 15, e1007638. https://doi.org/10.1371/journal.ppat.1007638.

Nielsen, K., Heitman, J., 2007. Sex and virulence of human pathogenic fungi. In: Advances in Genetics. Academic Press, pp. 143−173. https://doi.org/10.1016/S0065-2660(06)57004-X.

Novak Babič, M., Gostinčar, C., Gunde-Cimerman, N., 2020. Microorganisms populating the water-related indoor biome. Appl. Microbiol. Biotechnol. 104, 6443−6462. https://doi.org/10.1007/s00253-020-10719-4.

Novak Babič, M., Zupančič, J., Gunde-Cimerman, N., de Hoog, S., Zalar, P., Babič, M.N., Zupančič, J., Gunde-Cimerman, N., de Hoog, S., Zalar, P., 2018. Ecology of the human opportunistic black yeast *Exophiala dermatitidis* indicates preference for human-made habitats. Mycopathologia 183, 201−212. https://doi.org/10.1007/s11046-017-0134-8.

O'Brien, C.E., Oliveira-Pacheco, J., Ó Cinnéide, E., Haase, M.A.B., Hittinger, C.T., Rogers, T.R., Zaragoza, O., Bond, U., Butler, G., 2021. Population genomics of the pathogenic yeast *Candida tropicalis* identifies hybrid isolates in environmental samples. PLoS Pathog. 17, e1009138. https://doi.org/10.1371/journal.ppat.1009138.

O'Hanlon, S.J., Rieux, A., Farrer, R.A., Rosa, G.M., Waldman, B., Bataille, A., Kosch, T.A., Murray, K.A., Brankovics, B., Fumagalli, M., Martin, M.D., Wales, N., Alvarado-Rybak, M., Bates, K.A., Berger, L., Böll, S., Brookes, L., Clare, F., Courtois, E.A., Cunningham, A.A., Doherty-Bone, T.M., Ghosh, P., Gower, D.J., Hintz, W.E., Höglund, J., Jenkinson, T.S., Lin, C.-F., Laurila, A., Loyau, A., Martel, A., Meurling, S., Miaud, C., Minting, P., Pasmans, F., Schmeller, D.S., Schmidt, B.R., Shelton, J.M.G., Skerratt, L.F., Smith, F., Soto-Azat, C., Spagnoletti, M., Tessa, G., Toledo, L.F., Valenzuela-Sánchez, A., Verster, R., Vörös, J., Webb, R.J., Wierzbicki, C., Wombwell, E., Zamudio, K.R., Aanensen, D.M., James, T.Y., Gilbert, M.T.P., Weldon, C., Bosch, J., Balloux, F., Garner, T.W.J., Fisher, M.C., 2018. Recent Asian origin of chytrid fungi causing global amphibian declines. Science 360, 621−627. https://doi.org/10.1126/science.aar1965.

Ohm, R.A., Feau, N., Henrissat, B., Schoch, C.L., Horwitz, B.A., Barry, K.W., Condon, B.J., Copeland, A.C., Dhillon, B., Glaser, F., Hesse, C.N., Kosti, I., LaButti, K., Lindquist, E.A., Lucas, S., Salamov, A.A., Bradshaw, R.E., Ciuffetti, L., Hamelin, R.C., Kema, G.H.J., Lawrence, C., Scott, J.A., Spatafora, J.W., Turgeon, B.G., de Wit, P.J.G.M., Zhong, S.B., Goodwin, S.B., Grigoriev, I.V., 2012. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog. 8, e1003037. https://doi.org/10.1371/journal.ppat.1003037.

Otto, S.P., 2009. The evolutionary enigma of sex. Am. Nat. 174, 1−14. https://doi.org/10.1086/599084.

Paoletti, M., Buck, K.W., Brasier, C.M., 2005. Selective acquisition of novel mating type and vegetative incompatibility genes via interspecies gene transfer in the globally invading eukaryote *Ophiostoma novo-ulmi*. Mol. Ecol. 15, 249−262. https://doi.org/10.1111/j.1365-294X.2005.02728.x.

Papon, N., Nevez, G., Le Gal, S., Vigneau, C., Robert-Gangneux, F., Bouchara, J.-P., Cornely, O.A., Denning, D.W., Gangneux, J.-P., 2021. Fungal infections in transplant recipients: pros and cons of immunosuppressive and antimicrobial treatment. Lancet Microbe 2, e6−e8. https://doi.org/10.1016/S2666-5247(20)30199-3.

Pattnaik, B., Patil, S.S., Vaidya, G., Yadav, M.P., Prasad, S.K., Dharmashekar, C., Shreevatsa, B., Srinivasa, C., Prasad, K.S., Ashwini, P., Shivamallu, C., 2021. Emergence of mucormycosis during COVID-19 pandemic in India. J. Pharmaceut. Res. Int. 98−103. https://doi.org/10.9734/jpri/2021/v33i50B33432.

Perini, L., Mogrovejo, D.C., Tomazin, R., Gostinčar, C., Brill, F.H.H., Gunde-Cimerman, N., 2019. Phenotypes associated with pathogenicity: their expression in Arctic fungal isolates. Microorganisms 7, 1−15. https://doi.org/10.3390/microorganisms7120600.

Pontes, J.G. de M., Fernandes, L.S., dos Santos, R.V., Tasic, L., Fill, T.P., 2020. Virulence factors in the phytopathogen−host interactions: an overview. J. Agric. Food Chem. 68, 7555−7570. https://doi.org/10.1021/acs.jafc.0c02389.

Prenafeta-Boldú, F.X., Hoog, G.S. de, Summerbell, R.C., 2018. Fungal communities in hydrocarbon degradation. In: McGenity, T.J. (Ed.), Microbial Communities Utilizing Hydrocarbons and Lipids: Members, Metagenomics and Ecophysiology. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-60063-5.

Prenafeta-Boldu, F.X., Summerbell, R., de Hoog, G.S., 2006. Fungi growing on aromatic hydrocarbons: biotechnology's unexpected encounter with biohazard? FEMSMicrobiol. Rev. 30, 109−130. https://doi.org/10.1111/j.1574-6976.2005.00007.x.

Pryszcz, L.P., Németh, T., Saus, E., Ksiezopolska, E., Hegedűsová, E., Nosek, J., Wolfe, K.H., Gacser, A., Gabaldón, T., 2015. The genomic aftermath of hybridization in the opportunistic pathogen *Candida metapsilosis*. PLoS Genet. 11, e1005626. https://doi.org/10.1371/journal.pgen.1005626.

Rayens, E., Norris, K., 2022. Prevalence and healthcare burden of fungal infections in the United States, 2018. Open Forum Infect. Dis. 9. https://doi.org/10.1093/ofid/ofab593.

Revie, N.M., Iyer, K.R., Robbins, N., Cowen, L.E., 2018. Antifungal drug resistance: evolution, mechanisms and impact. Curr. Opin. Microbiol. 45, 70−76. https://doi.org/10.1016/j.mib.2018.02.005.

Rhodes, J., Abdolrasouli, A., Dunne, K., Sewell, T.R., Zhang, Y., Ballard, E., Brackin, A.P., van Rhijn, N., Chown, H., Tsitsopoulou, A., Posso, R.B., Chotirmall, S.H., McElvaney, N.G., Murphy, P.G., Talento, A.F., Renwick, J., Dyer, P.S., Szekely, A., Bowyer, P., Bromley, M.J., Johnson, E.M., Lewis White, P., Warris, A., Barton, R.C., Schelenz, S., Rogers, T.R., Armstrong-James, D., Fisher, M.C., 2022. Population genomics confirms acquisition of drug-resistant *Aspergillus fumigatus* infection by humans from the environment. Nat. Microbiol. 7, 663−674. https://doi.org/10.1038/s41564-022-01091-2.

Rhodes, J., Fisher, M.C., 2019. Global epidemiology of emerging *Candida auris*. Curr. Opin. Microbiol. 52, 84−89. https://doi.org/10.1016/j.mib.2019.05.008.

Rigling, D., Prospero, S., 2018. *Cryphonectria parasitica*, the causal agent of chestnut blight: invasion history, population biology and disease control. Mol. Plant Pathol. 19, 7−20. https://doi.org/10.1111/mpp.12542.

Robbins, N., Caplan, T., Cowen, L.E., 2017. Molecular evolution of antifungal drug resistance. Annu. Rev. Microbiol. 71, 753−775. https://doi.org/10.1146/annurev-micro-030117-020345.

Robert, V., Cardinali, G., Casadevall, A., 2015. Distribution and impact of yeast thermal tolerance permissive for mammalian infection. BMC Biol. 13, 18. https://doi.org/10.1186/s12915-015-0127-3.

Robert, V.A., Casadevall, A., 2009. Vertebrate endothermy restricts most fungi as potential pathogens. J. Infect. Dis. 200, 1623−1626. https://doi.org/10.1086/644642.

Rokas, A., 2022. Evolution of the human pathogenic lifestyle in fungi. Nat. Microbiol. 7, 607−619. https://doi.org/10.1038/s41564-022-01112-0.

Rosenblum, E.B., James, T.Y., Zamudio, K.R., Poorten, T.J., Ilut, D., Rodriguez, D., Eastman, J.M., Richards-Hrdlicka, K., Joneson, S., Jenkinson, T.S., Longcore, J.E., Parra Olea, G., Toledo, L.F., Arellano, M.L., Medina, E.M., Restrepo, S., Flechas, S.V., Berger, L., Briggs, C.J., Stajich, J.E., 2013. Complex history of the amphibian-killing chytrid fungus revealed with genome resequencing data. Proc. Natl. Acad. Sci. USA 110, 9385−9390. https://doi.org/10.1073/pnas.1300130110.

Roy, S., Thompson, D., 2015. Evolution of regulatory networks in *Candida glabrata*: learning to live with the human host. FEMS Yeast Res. 15, 1−87. https://doi.org/10.1093/femsyr/fov087.

Ruhnke, M., Cornely, O.A., Schmidt-Hieber, M., Alakel, N., Boell, B., Buchheidt, D., Christopeit, M., Hasenkamp, J., Heinz, W.J., Hentrich, M., Karthaus, M., Koldehoff, M., Maschmeyer, G., Panse, J., Penack, O., Schleicher, J., Teschner, D., Ullmann, A.J., Vehreschild, M., Lilienfeld-Toal, M., Weissinger, F., Schwartz, S., 2020. Treatment of invasive fungal diseases in cancer patients — revised 2019 recommendations of the infectious diseases working party (AGIHO) of the German society of hematology and oncology (DGHO). Mycoses 63, 653−682. https://doi.org/10.1111/myc.13082.

Schröder, M.S., Martinez de San Vicente, K., Prandini, T.H.R., Hammel, S., Higgins, D.G., Bagagli, E., Wolfe, K.H., Butler, G., 2016. Multiple origins of the pathogenic yeast *Candida orthopsilosis* by separate hybridizations between two parental species. PLoS Genet. 12, e1006404. https://doi.org/10.1371/journal.pgen.1006404.

Schweizer, G., Haider, M.B., Barroso, G.V., Rössel, N., Münch, K., Kahmann, R., Dutheil, J.Y., 2021. Population genomics of the maize pathogen *Ustilago maydis*: demographic history and role of virulence clusters in adaptation. Genome Biol. Evol. 13. https://doi.org/10.1093/gbe/evab073.

Selarka, L., Sharma, S., Saini, D., Sharma, S., Batra, A., Waghmare, V.T., Dileep, P., Patel, S., Shah, M., Parikh, T., Darji, P., Patel, A., Goswami, G., Shah, A., Shah, S., Lathiya, H., Shah, M., Sharma, P., Chopra, S., Gupta, A., Jain, N., Khan, E., Sharma, V.K., Sharma, A.K., Chan, A.C.Y., Ong, J.J.Y., 2021. Mucormycosis and COVID-19: an epidemic within a pandemic in India. Mycoses 64, 1253−1260. https://doi.org/10.1111/myc.13353.

Sephton-Clark, P., McConnell, S.A., Grossman, N., Baker, R.P., Dragotakes, Q., Fan, Y., Fu, M.S., Gerbig, G., Greengo, S., Hardwick, J.M., Kulkarni, M., Levitz, S.M., Nosanchuk, J.D., Shoham, S., Smith, D.F.Q., Stempinski, P., Timp, W., Wear, M.P., Cuomo, C.A., Casadevall, A., 2023. Similar evolutionary trajectories in an environmental *Cryptococcus neoformans* isolate after human and murine infection. Proc. Natl. Acad. Sci. USA 120, e2217111120. https://doi.org/10.1073/pnas.2217111120.

Sephton-Clark, P., Tenor, J.L., Toffaletti, D.L., Meyers, N., Giamberardino, C., Molloy, S.F., Palmucci, J.R., Chan, A., Chikaonda, T., Heyderman, R., Hosseinipour, M., Kalata, N., Kanyama, C., Kukacha, C., Lupiya, D., Mwandumba, H.C., Harrison, T., Bicanic, T., Perfect, J.R., Cuomo, C.A., 2022. Genomic variation across a clinical *Cryptococcus* population linked to disease outcome. mBio 13, e02626-22. https://doi.org/10.1128/mbio.02626-22.

Shapiro, R.S., Robbins, N., Cowen, L.E., 2011. Regulatory circuitry governing fungal development, drug resistance, and disease. Microbiol. Mol. Biol. Rev. 75, 213−267. https://doi.org/10.1128/MMBR.00045-10.

Sharma, A., Bedi, M., Patel, P., Singh Arora, A., 2022. Black fungus outbreak in India − a direct consequence of COVID-19 surge: a myth or reality. Gondwana Res. https://doi.org/10.1016/j.gr.2021.12.016.

Singh, A.K., Singh, R., Joshi, S.R., Misra, A., 2021a. Mucormycosis in COVID-19: a systematic review of cases reported worldwide and in India. Diabetes Metabol. Syndr.: Clin. Res. Rev. 15, 102146. https://doi.org/10.1016/j.dsx.2021.05.019.

Singh, N.K., Karisto, P., Croll, D., 2021. Population-level deep sequencing reveals the interplay of clonal and sexual reproduction in the fungal wheat pathogen *Zymoseptoria tritici*. Microb. Genom. 7. https://doi.org/10.1099/mgen.0.000678.

Sitterlé, E., Maufrais, C., Sertour, N., Palayret, M., D'Enfert, C., Bougnoux, M.-E., 2019. Within-host genomic diversity of *Candida albicans* in healthy carriers. Sci. Rep. 9, 2563. https://doi.org/10.1038/s41598-019-38768-4.

Song, Y., Laureijssen-van de Sande, W.W.J., Moreno, L.F., Gerrits van den Ende, B., Li, R., de Hoog, S., 2017. Comparative ecology of capsular *Exophiala* species causing disseminated infection in humans. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.02514.

Spivak, E.S., Hanson, K.E., 2018. *Candida auris*: an emerging fungal pathogen. J. Clin. Microbiol. 56. https://doi.org/10.1128/JCM.01588-17.

Staniszewska, M., 2020. Virulence factors in *Candida* species. Curr. Protein Pept. Sci. 21, 313−323. https://doi.org/10.2174/1389203720666190722152415.

Stauber, L., Badet, T., Feurtey, A., Prospero, S., Croll, D., 2021. Emergence and diversification of a highly invasive chestnut pathogen lineage across southeastern Europe. Elife 10. https://doi.org/10.7554/eLife.56279.

Steenbergen, J.N., Shuman, H.A., Casadevall, A., 2001. *Cryptococcus neoformans* interactions with amoebae suggest an explanation for its virulence and intracellular pathogenic strategy in macrophages. Proc. Natl. Acad. Sci. USA 98, 15245−15250. https://doi.org/10.1073/pnas.261418798.

Steffen, H., Bosch, C., Wolfaardt, G., Botha, A., 2022. Rising environmental temperatures and polluted surface waters: the prelude to the rise of mycoses in South Africa. Water SA 48. https://doi.org/10.17159/wsa/2022.v48.i2.3918.

Strickland, A.B., Shi, M., 2021. Mechanisms of fungal dissemination. Cell. Mol. Life Sci. 78, 3219−3238. https://doi.org/10.1007/s00018-020-03736-z.

Stukenbrock, E.H., Bataillon, T., Dutheil, J.Y., Hansen, T.T., Li, R., Zala, M., McDonald, B.A., Wang, J., Schierup, M.H., 2011. The making of a new pathogen: insights from comparative population genomics of the domesticated wheat pathogen *Mycosphaerella graminicola* and its wild sister species. Genome Res. 21, 2157−2166. https://doi.org/10.1101/gr.118851.110.

Sun, S., Heitman, J., 2011. Is sex necessary? BMC Biol. 9, 1−4. https://doi.org/10.1186/1741-7007-9-56.

Taylor, J.W., 2015. Evolutionary perspectives on human fungal pathogens. Cold Spring Harb. Perspec. Med. 5, a019588. https://doi.org/10.1101/cshperspect.a019588.

Taylor, J.W., Hann-Soden, C., Branco, S., Sylvain, I., Ellison, C.E., 2015. Clonal reproduction in fungi. Proc. Natl. Acad. Sci. USA 112, 8901−8908. https://doi.org/10.1073/pnas.1503159112.

Todd, R.T., Forche, A., Selmecki, A., 2017. Ploidy variation in fungi: polyploidy, aneuploidy, and genome evolution. Microbiol. Spectr. 5, 1−31. https://doi.org/10.1128/microbiolspec.FUNK-0051-2016.

Turner, S.A., Butler, G., 2014. The *Candida* pathogenic species complex. Cold Spring Harb. Perspec. Med. 4. https://doi.org/10.1101/cshperspect.a019778 a019778−a019778.

Vallabhaneni, S., Mody, R.K., Walker, T., Chiller, T., 2016. The global burden of fungal diseases. Infect. Dis. Clin. 30, 1−11. https://doi.org/10.1016/j.idc.2015.10.004.

van Burik, J.A.H., Magee, P.T., 2001. Aspects of fungal pathogenesis in humans. Annu. Rev. Microbiol. 55, 743−772. https://doi.org/10.1146/annurev.micro.55.1.743.

Van Dyke, M.C.C., Teixeira, M.M., Barker, B.M., 2019. Fantastic yeasts and where to find them: the hidden diversity of dimorphic fungal pathogens. Curr. Opin. Microbiol. 52, 55−63. https://doi.org/10.1016/j.mib.2019.05.002.

Verweij, P.E., van de Veerdonk, F.L., 2022. Managing secondary fungal infections in severe COVID-19: how to move forward? Lancet Respir. Med. 10, 127−128. https://doi.org/10.1016/S2213-2600(21)00500-2.

Vijaya Chandra, S.H., Srinivas, R., Dawson, T.L., Common, J.E., 2021. Cutaneous *Malassezia*: commensal, pathogen, or protector? Front. Cell. Infect. Microbiol. 10. https://doi.org/10.3389/fcimb.2020.614446.

Villar, J.M., Delgado, J.D.J., Velasco, C.G., 2012. Fungemia due to *Rhodotorula mucilaginosa* in an immunocompetent, critically ill patient. J. Infect. Chemother. 18, 581−583. https://doi.org/10.1007/s10156-011-0347-6.

Vite-Garín, T., Estrada-Bárcenas, D.A., Cifuentes, J., Taylor, M.L., 2014. The importance of molecular analyses for understanding the genetic diversity of *Histoplasma capsulatum*: an overview. Rev. Iberoam. De. Micol. 31, 11−15. https://doi.org/10.1016/j.riam.2013.09.013.

Wang, X., Cai, W., van den Ende, A.H.G.G., Zhang, J., Xie, T., Xi, L., Li, X., Sun, J., de Hoog, S., 2018. Indoor wet cells as a habitat for melanized fungi, opportunistic pathogens on humans and other vertebrates. Sci. Rep. 8, 7685. https://doi.org/10.1038/s41598-018-26071-7.

Warris, A., Ballou, E.R., 2019. Oxidative responses and fungal infection biology. Semin. Cell Dev. Biol. 89, 34−46. https://doi.org/10.1016/j.semcdb.2018.03.004.

Wirth, F., Goldani, L.Z., 2012. Epidemiology of *Rhodotorula*: an emerging pathogen. Interdiscip. Perspect. Infect. Dis. 2012, 1−7. https://doi.org/10.1155/2012/465717.

World Health Organization, 2022. WHO Fungal Priority Pathogens List to Guide Research, Development and Public Health Action. Geneva. https://www.who.int/publications/i/item/9789240060241.

Xu, J., 2005. Cost of interacting with sexual partners in a facultative sexual microbe. Genetics 171, 1597−1604. https://doi.org/10.1534/genetics.105.045302.

Yang, N., Ovenden, B., Baxter, B., McDonald, M.C., Solomon, P.S., Milgate, A., 2022. Multi-stage resistance to *Zymoseptoria tritici* revealed by GWAS in an Australian bread wheat diversity panel. Front. Plant Sci. 13. https://doi.org/10.3389/fpls.2022.990915.

Zajc, J., Gostinčar, C., Černoša, A., Gunde-Cimerman, N., 2019. Stress-tolerant yeasts: opportunistic pathogenicity versus biocontrol potential. Genes 10. https://doi.org/10.3390/genes10010042.

Zalar, P., Novak, M., De Hoog, G.S., Gunde-Cimerman, N., 2011. Dishwashers - a man-made ecological niche accommodating human opportunistic fungal pathogens. Fungal Biol. 115, 997−1007. https://doi.org/10.1016/j.funbio.2011.04.007.

Zaragoza, O., 2019. Basic principles of the virulence of *Cryptococcus*. Virulence 10, 490−501. https://doi.org/10.1080/21505594.2019.1614383.

Zhong, Z., Chen, M., Lin, L., Han, Y., Bao, J., Tang, W., Lin, L., Lin, Y., Somai, R., Lu, L., Zhang, W., Chen, J., Hong, Y., Chen, X., Wang, B., Shen, W.-C., Lu, G., Norvienyeku, J., Ebbole, D.J., Wang, Z., 2018. Population genomic analysis of the rice blast fungus reveals specific events associated with expansion of three main clades. ISME J. 12, 1867−1878. https://doi.org/10.1038/s41396-018-0100-6.

Zupančič, J., Novak Babič, M., Zalar, P., Gunde-Cimerman, N., 2016. The black yeast *Exophiala dermatitidis* and other selected opportunistic human fungal pathogens spread from dishwashers to kitchens. PLoS One 11, e0148166. https://doi.org/10.1371/journal.pone.0148166.

This page intentionally left blank

CHAPTER

5 Molecular Epidemiology of Pathogenic Microorganisms and the Predominant Clonal Evolution Model

Michel Tibayrenc

Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle, MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD Center, Montpellier, France

*See abbreviations and definitions in the glossary of specialized terms.

1. Introduction: What Is Molecular Epidemiology?

Molecular epidemiology is the discipline that aims at following the epidemiological routes of pathogens by characterizing their genetic make-up. In common words, it is often called: "strain typing."

Early attempts of molecular epidemiology relied on antigen characterization of pathogens by antibody−antigen reactions (immunocharacterization). The approach got a real start in the late 1970s with the advent of isoenzyme* markers. As a matter of fact, although strictly speaking, isoenzymes correspond to a phenotypic* variation, their genetic control is fairly known and then, isoenzyme polymorphism does reflect genetic variation (Fig. 5.1). Pioneer studies dealing with isoenzyme analysis permitted to elucidate several informative traits of parasite genetic polymorphism (Gibson et al., 1980; Miles et al., 1978; Maazoun et al., 1981). The technology of isoenzyme typing is termed multilocus enzyme electrophoresis* (MLEE*).

A new step has been taken when isoenzyme polymorphism was analyzed in terms of population genetics (Tibayrenc et al., 1981a).

Later, in the early 1980s, molecular epidemiology turned to true genetic markers with direct survey of DNA variation. These new markers concerned restriction fragment length polymorphism* or RFLP* (Tibayrenc and Ayala, 1987, Fig. 5.2), amplified fragment length polymorphism* or AFLP*, random primed amplified polymorphic DNA or RAPD (Tibayrenc et al., 1993, Fig. 5.1), and pulse field gel electrophoresis* or PFGE*. AFLP and RFLP rely on specific polymerase chain reaction* (PCR*) protocols.

The advent of Multilocus Sequence Typing* (MLST*; Maiden et al., 1998; Maiden, 2006) introduced DNA sequencing in the field of molecular epidemiology. Classic MLST protocol involves the sequencing of 470 bp fragments of 8−11 housekeeping genes. The considerable advantages of MLST in comparison with other DNA typing methods is that it is highly reproducible and is portable. As a matter of fact, DNA sequences can be easily sent by email. Moreover, MLST can be applied to any

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00021-5
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

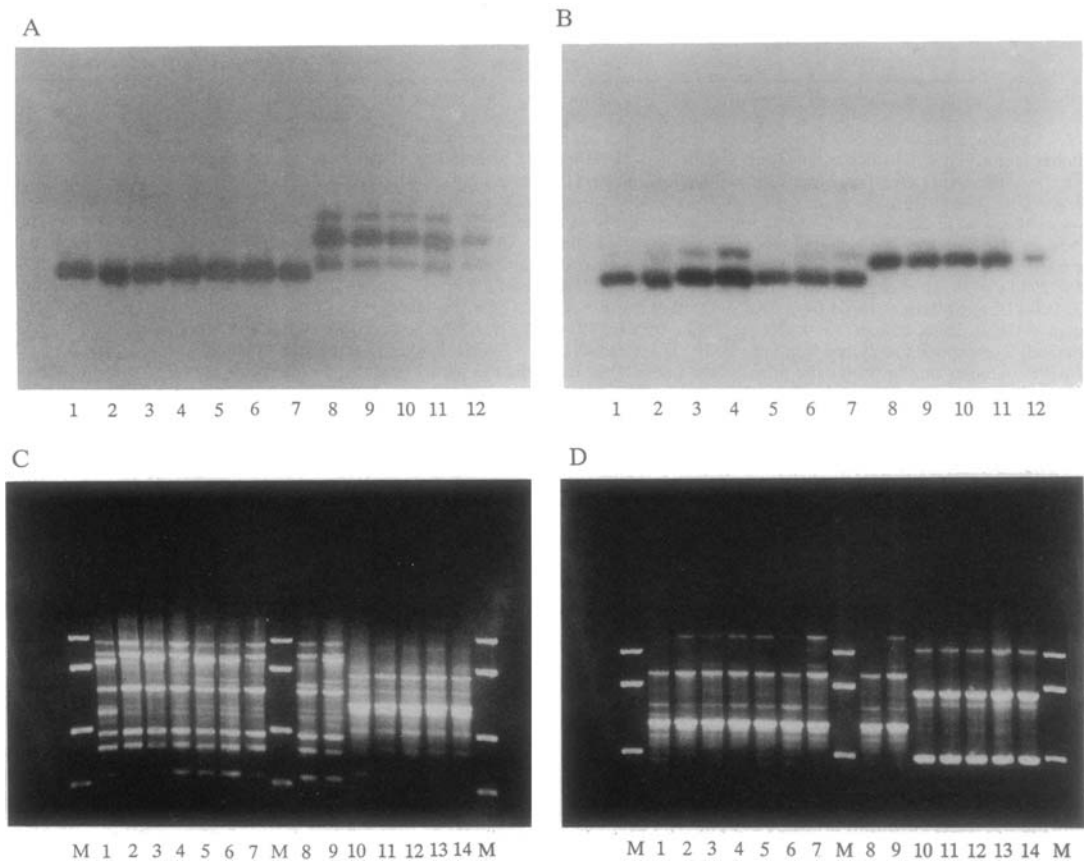


FIGURE 5.1

Top: isoenzyme polymorphism; bottom: random amplified polymorphic DNA* (RAPD*) polymorphism. The samples correspond to various strains of *Trypanosoma cruzi*, the parasite responsible for Chagas disease. M = molecular weight ladders. The patterns observed show a striking example of linkage disequilibrium* (LD*). Isoenzyme genotypes A 1–7 are linked to isoenzyme genotypes B 1–7 and to RAPD genotypes C 1–9 and D 1–9. It is the same for isoenzyme genotypes A 8–12 and B 8–12 and RAPD genotypes C 10–14 and D 10–14. Recombinant genotypes such as A 1 + B 8 or A 8 + C 1 have never been observed on more than 500 *T. cruzi* strains.

After Tibayrenc, M. et al., 1993. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proc. Natl. Acad. Sci. USA 90 (4), 1335–1339. https://doi.org/10.1073/pnas.90.4.1335

kind of pathogen. It has been massively used to characterize various bacterial species. Specialized databases have been designed for each species (Maiden, 2006). More recently, MLST has been applied to several species of parasitic protozoa (Ankarklev et al., 2018; Boité et al., 2012; Lauthier et al., 2012) and fungi (Ferreira-Paim et al., 2017). Multilocus genotypes* (MLGs*) that appear identical for a given MLST protocol are termed "sequence types"* (STs*).

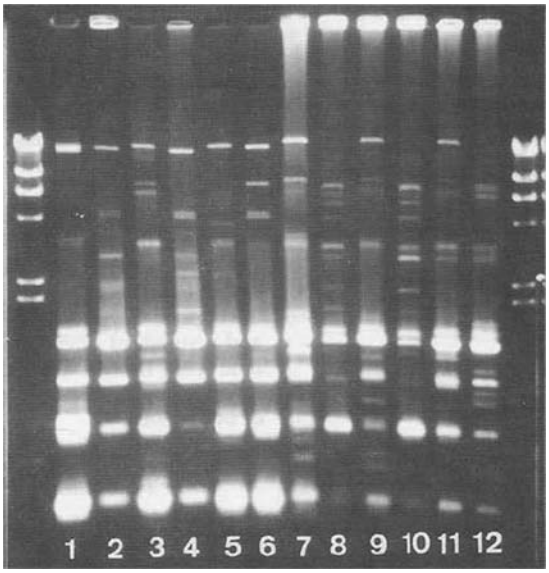


FIGURE 5.2

RFLP polymorphism of the kinetoplast* DNA of various *Trypanosoma cruzi* strains.

After Tibayrenc, M., Ayala, F.J., 1988. Isozyme variability in Trypanosoma cruzi, the agent of Chagas'disease: genetical, taxonomical and epidemiological significance. Evolution 42 (2), 277–292. https://doi.org/10.1111/j.1558-5646.1988.tb04132.x.

Molecular epidemiology is now being revolutionized by the spread of genomic* approaches. As a matter of fact, the cost of sequencing protocols has dramatically dropped in the last 10 years. The sequencing of whole viral and bacterial genomes (whole genome sequencing; WGS) tends to become routine, and allows for a level of resolution much higher than the resolution power of previous methods, including MLST. This permits the survey of "measurably evolving pathogens," that is to say: epidemiological analyses down to the scales of kilometers and years (Biek et al., 2015). Routine WGS permits to identify specific single nucleotide polymorphisms* or SNPs*. The typing of pathogen strains by several thousand SNPs is now common-place (McNally et al., 2016), allowing an unparalleled level of resolution. Genomic approaches are no longer the prerogative of viruses and bacteria. Although their much larger genomes makes it more difficult, genomic analysis of eukaryotic pathogens becomes more and more common (Imamura et al., 2016; Schwabl et al., 2019, 2021; Van den Broeck et al., 2020). These advances provide us with an abundant material for exploring more in depth both macro* and microevolutionary* levels, which has fueled recent developments of the predominant clonal evolution* (PCE*) approach (see below).

2. Brief Recall on the PCE Model

The PCE model has been exposed in details in several recent papers (Tibayrenc and Ayala, 2021a, 2021b, 2022). Only the key points will be recalled here.

2.1 What Is, and what Is Not the PCE Model

The PCE model deals only with the population structure of the organism under survey. It does not consider precise cytological mechanisms involved in the reproductive mode of the concerned species. It concerns all cases in which genetic recombination* is severely restricted, which generates a typical population structure exposed below. It therefore includes not only mitotic propagation but also selfing*/strong homogamy* and several cases of parthenogenesis*.

It does not state that genetic exchange is absent or epidemiologically/evolutionary meaningless, as wrongly stated many times (Gutiérrez-Corbo et al., 2021; Lewis et al., 2017; Miles et al., 2009; Ortega-Pierres et al., 2009; Ramírez and Llewellyn, 2014; Weedall and Hall, 2015), but rather that its impact is not sufficient to erase the predominant clonal evolution pattern.

2.2 Deep Phylogenies/Multigene Bifurcating Trees

The most typical manifestation of this predominant clonal evolution pattern is the generation and maintenance in the long run of deep phylogenies/bifurcating evolutionary trees at all evolutionary levels and in the whole ecogeographical range of the species. The identification of these bifurcating trees should not be based on only one gene, or on a limited range of genes, but rather, on many different genes ("multilocus bifurcating trees* or MLBTs*). As a matter of fact, evolutionary trees based on only one gene or a limited set of genes may not reflect the evolution of the whole species: "gene trees and species trees are not the same" (Nichols, 2001). It is therefore misleading to consider that "the evolutionary history of each gene tree is identical to the species tree" (Gómez-Hernández et al., 2019). This presence of stable bifurcating trees ("clonal backbone"; "tree-like phylogenies"), attesting for vertical, rather than horizontal inheritance, is considered by many authors as the clearest manifestation of clonality (Hanage, 2016; Maiden, 2006; Pearson et al., 2009; Pérez-Losada et al., 2006; Posada et al., 2002). Robustness of the MLBTs can be evaluated either by classic bootstrap* analysis, with the threshold of 0.70 as recommended by Hillis and Bull (1993) or by a Bayesian* procedure (Pybus and Rambaut, 2009).

If PCE obtains, deep phylogenies/MLBTs should get more and more phylogenetic support, as more data (samples of genetic markers and of strains = MLGs) are added, according to the congruence criterion (Tibayrenc and Ayala, 2017a). This is what we have called the "clonality threshold" (Tibayrenc and Ayala, 2017a), beyond which recombination is efficiently countered by PCE, and evolutionary lines within the species (near-clades* or NCs*) are bound to irreversibly diverge. This strict criterion of the presence or absence of deep phylogenies/MLBTs provides a clear-cut definition of predominant clonality, that is accepted by many authors working in this field (see many references in Tibayrenc and Ayala, 2021a). It is by far more relevant and falsifiable than the abundance of vague, subjective, poorly informative terms that litter this field of research: "relative clonality" (Dunne et al., 2018); "high levels of recombination" (Pérez-Losada et al., 2013); "widespread genetic exchange," "regular outcrossing," "frequent genetic exchange" (Ramírez and Llewellyn, 2014); "higher impact of genetic recombination" (Llanes et al., 2022); "recombination may be quite frequent" (Gutiérrez-Corbo et al., 2021); "genetic exchange appears common" (Schwabl et al., 2021); "substantial recombination" (de Paula Baptista et al., 2014); "high frequency of recombination," "recombination at high rate" (Ramírez et al., 2013).

2.3 Other PCE Traits

Other key manifestations of PCE are: (i) The presence of a statistically significant linkage disequilibrium (LD): see Fig. 5.1. LD is a remarkable property of the PCE model that is highly relevant for molecular epidemiology. As a matter of fact, it allows the procedure of "indirect typing." If the LD is strong, as it is in the case exposed at Fig. 5.1, knowing one genotype at one locus makes it possible to predict with a high probability the genotypes that will occur at all other loci. As an example, LD in the bacterium *Escherichia coli* (*E. coli*) makes it possible to design specific PCR tests for identifying the different evolutionary lines that subdivide this species (Denamur et al., 2021). (ii) The occurrence of widespread MLGs over large spans of time and vast geographical distances, with frequencies incompatible with panmictic* expectations. Such widespread MLGs are the result of clonal propagation. They should not be considered as true genetic clones, but rather, as "clonets*" (Tibayrenc et al., 1991).

2.4 Evolutionary Scales; The Russian Doll Model

The PCE model stipulates that these key features are observed at all evolutionary scales. In particular, they should be verified at a microevolutionary level. The population structure recorded within each of the main genetic subdivisions ("near clades" or NCs) that are observed in the species under study should look like a miniature picture of the whole species' population structure (Russian doll pattern* or RDP* (Tibayrenc and Ayala, 2013); see Fig. 5.3).

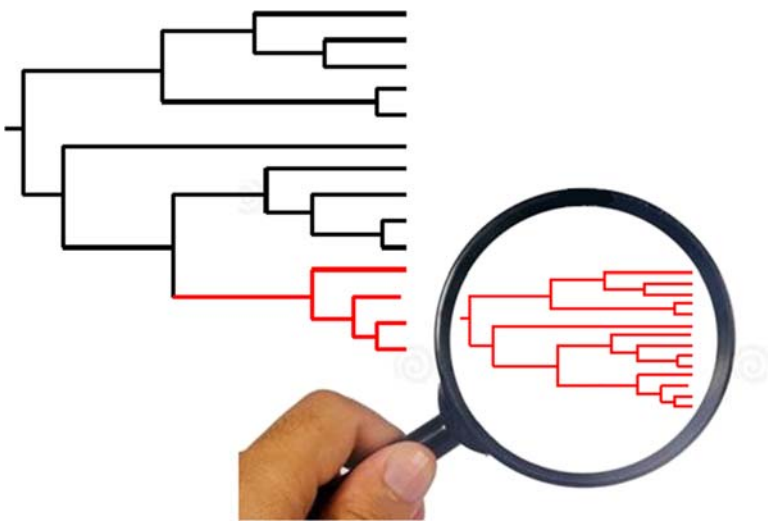


FIGURE 5.3

Russian doll pattern (RDP). One of the NCs that subdivide the species under study is analyzed with markers with a high resolution. This analysis shows that within this NCs, a miniature picture of the whole species' population structure is observed.

After Tibayrenc, M., Ayala, F.J., 2017b. Relevant units of analysis for applied and basic research dealing with neglected transmissible diseases: the predominant clonal evolution model of pathogenic microorganisms. PLoS Neglected Trop. Dis. 11 (4), e0005293. https://doi.org/10.1371/journal.pntd.0005293.

2.5 Other Models of Pathogen Population Structure

According to the PCE model, together with lesser NCs, LD and widespread MLGs should be observed within each of the main NCs of the species under study. Now it has been proposed that the PCE model is artifactual, since lack of genetic exchange is observed between the main NCs of the species considered, whereas genetic exchange is abundant within the NCs (Ramírez and Llewellyn, 2014; "pseudospeciation model"; see Fig. 5.4B). The Russian doll model makes it possible to refute this hypothesis by showing that recombination is identically restrained within each of the NCs that subdivide the species considered.

The presence of deep phylogenies/MLBTs makes it possible to refute another model of population structure, namely the epidemic clonality/semiclonality* model (Fig. 5.4C; Maiden, 2006; Maynard Smith et al., 1993). As a matter of fact, the epidemic/semiclonal model concerns a population structure where occasional bouts of clonal propagation occur in an otherwise recombining species. In the framework of this model, genetic recombination should erase deep phylogenies/MLBTs in the long run.

2.6 The Problem of the Wahlund Effect*: The "Starving Sex Hypothesis"

It is important to verify that the restrictions to genetic exchange are not due to trivial physical obstacles (Wahlund effect). Time and space always have an impact on the population structure of any species ("isolation by distance and/or time"; IBDT). In all the species analyzed by us (Tibayrenc and Ayala, 2021a), the Wahlund effect has not been a global explanation on the PCE features recorded by us. We have observed a very constant pattern in the cases surveyed by us: generally, IBD plays a role in the results observed but it can explain them only partially.

A special case of the Wahlund effect is the situation where different genotypes of the same species, although they occur at the same geographical place and at the same time, do not have opportunity for mating: for example, they may feature host specificity, or they may be hosted by different organs in a given host. We have called this case "starving sex hypothesis" or "passive clonality" (Tibayrenc and Ayala, 2014a): in this case, mating between different genotypes is not hampered by in-built biological properties of the species (true PCE) but rather, by lack of physical opportunities for mating. None of the species surveyed by us (Tibayrenc and Ayala, 2021a) appears to undergo such a "passive clonality." For all of them, in-built PCE traits are the most parsimonious hypothesis to account for their population structure.

After this brief recall, let us illustrate the PCE model with five illustrative species, two bacteria, two parasitic protozoa and one yeast. A broader sampling has been exposed in Tibayrenc and Ayala (2021a).

3. Bacterial Species

3.1 Escherichia coli

E. coli is generally a nonharmful commensal of the digestive tract of humans and many other species. However, some strains may be highly pathogenic (Denamur et al., 2021).

A seminal article by Ochman and Selander (1984), based on isoenzyme analysis, has inferred that this species has a clonal population structure. The main argument for supporting this hypothesis was

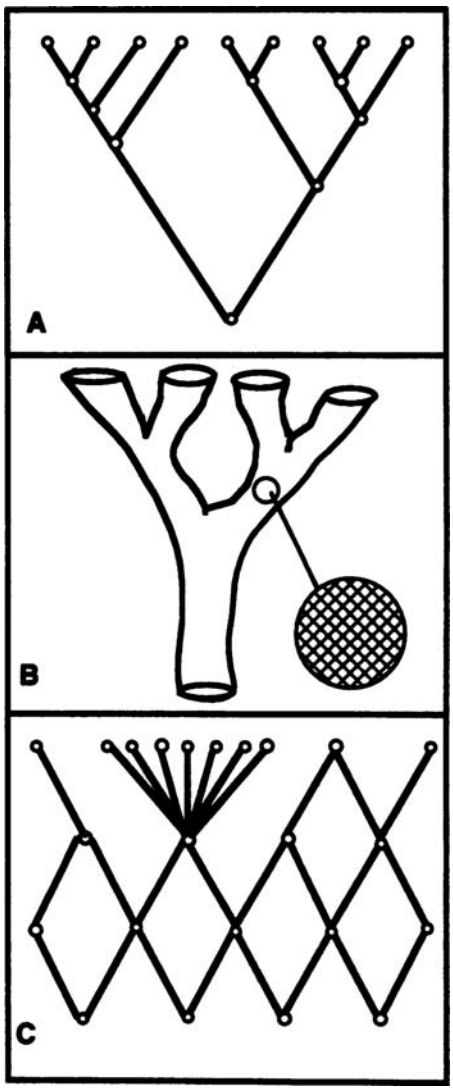


FIGURE 5.4

Three kinds of pathogen population structure. Top: "True" clonality; middle: Pseudospeciation; bottom: Epidemic clonality*.

After Maynard Smith, J.M. et al., 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. U. S. A. 90 (10), 4384–4388. https://doi.org/10.1073/pnas.90.10.4384.

the presence of a considerable LD, and the fact that most MLGs that would have been expected in case of panmixia were lacking.

Maynard Smith et al. (1991) have noted the paradox of an apparently clonal population structure in *E. coli* and the presence of mosaic genes, resulting from genetic recombination, in this species. The inferred explanation was that these recombining DNA sequences are short, "localized." Bobay et al. (2015) note that genetic recombination "does not destroy the clonal state of *E. coli*." Similarly, Denamur et al. (2021) note that recombination "does not 'blur' the phylogenetic signal in *E. coli*."

In the light of the most recent high-resolution genetic and genomic data, *E. coli* appears as a paradigmatic case of PCE. MLBTs and deep phylogenies are recorded at all evolutionary levels, above the species, within the species and within each of the NCs that subdivide it (Fig. 5.5).

At an upper evolutionary scale, the family Enterobacteriaceae comprises six "clades" (=NCs), of which the *Escherichia* "clade." The species *E. coli* itself is only a part of this NC (Fig. 5.5 left). When addressing the species *E. coli*, a clear phylogenetic signal is detectable within it (Chaudhuri and Henderson, 2012).

At the level of the species itself, *E. coli* exhibits five to six inner genetic subdivisions (=NCs), individualized by various, mutually corroborative, genetic markers (Chaudhuri and Henderson, 2012; Clermont et al., 2011, Fig. 5.5 middle). These NCs had been already recorded in the historical ECOR collection of strains as early as the year 1984 (Chaudhuri and Henderson, 2012) and still are present at nowadays, which shows their stability in space and time. This is one of the remarkable properties of the NCs: they show permanency in the whole ecogeographical range of the species under survey, over space and time. Historical collections of strains and ancient studies are a fine resource for evidencing this trait.

At a lower evolutionary scale, the 6 *E. coli* NCs are clearly structured into various lesser NCs supported by significant bootstrap values (Fig. 5.5 right) in a typical Russian doll pattern. Within the B2 NC, the subdivision I appears close to monomorphic with classical markers. It corresponds to the widespread "clone" ST131, a very recently appeared genotype, around the year 2000 (McNally et al., 2016). However, despite this very recent origin, when high-resolution genomic markers are used, ST131 exhibits considerable genetic variability and is clustered into three "clades" and various well-supported, lesser NCs Fig. 5.5, right (Fig. 5.2 in Ben Zakour et al., 2016; Fig. 5.1 in McNally et al. (2016), with several clusters, with a bootstrap >99%; Fig. 5.1 in Petty et al., 2014; Fig. 5.1 in Price et al., 2013; Fig. 5.1 in Stoesser et al., 2016).

This MLST MLG ST 131, which is worldwide distributed, is a typical example of a widespread clonal genotype. In terms of evolution, it should be equated to a clonet, rather than a true clone, since it is genetically heterogeneous and exhibits a typical Russian doll pattern. It is an illustration that the PCE model is valid also at a microevolutionary level, at a scale of time of a few years ("measurably evolving pathogen"; Biek et al., 2015).

ST393 is another typical example of a widespread MLST *E. coli* MLG that is worldwide distributed and exhibits a clear Russian doll pattern (Fig. 5.2 in Vianello et al., 2021).

3.2 Neisseria meningitidis

The presence of *Neisseria meningitidis* (*N. meningitidis*) in the framework of the PCE model could appear as somewhat "iconoclastic." As a matter of fact, this species, which is transformable*, has long been considered as "highly recombining" (Caugant, 2008), with "extensive genetic exchange"

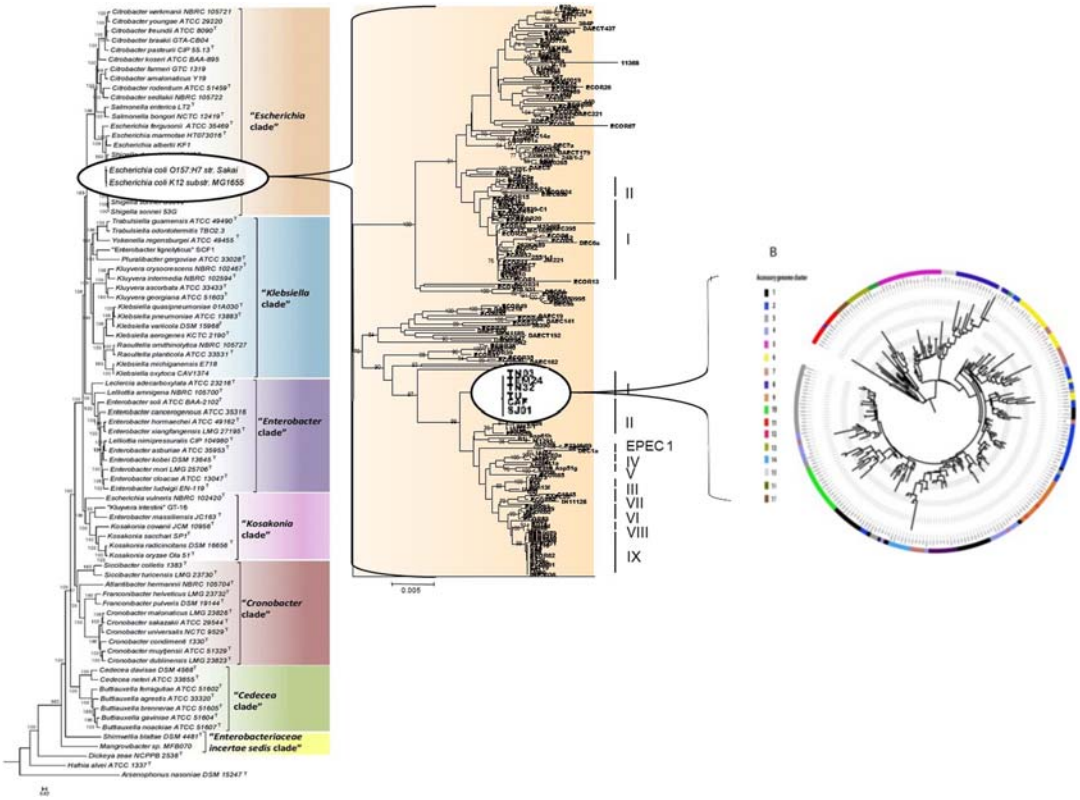


FIGURE 5.5

Russian doll pattern within the family Enterobacteriaceae and the genus *Escherichia*. From left to right: (1) the whole family Enterobacteriaceae, subdivided into six main clades. The species *Escherichia coli* constitutes only a small part of the "*Escherichia* clade" (top); (2) the species *E. coli*, according to many congruent studies, is clustered into seven near-clades, A, B1, B2, C, D, E, and F. Within the B2 near-clade, the "sequence type" ST131 corresponds to a highly virulent and widespread clone that is almost monomorphic with classic markers; (3) with high resolution markers (multi-SNP typing), the "clone" ST131 proves to be highly variable and strongly structured into several lesser near-clades that are only partially linked to geographical distance.

(1) After Alnajar, S., Gupta, R.S., 2017. Phylogenomics and comparative genomic studies delineate six main clades within the family Enterobacteriaceae and support the reclassification of several polyphyletic members of the family. Infect. Genet. Evol. 54, 108–127. https://doi.org/10.1016/j.meegid.2017.06.024; their Figure 1. (2) After Clermont, O., et al., 2011. Animal and human pathogenic Escherichia coli strains share common genetic backgrounds. Infect. Genet. Evol. 11 (3), 654–662. https://doi.org/10.1016/j.meegid.2011.02.005; their Figure S1. (3) After McNally, A., et al., 2016. Accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. PLoS Genet. 12 (9). https://doi.org/10.1371/journaLpgeN1006280; their Figure 1.

(Caugant and Maiden, 2009) and "extensive recombination" (Hill et al., 2015). Actually, *N. meningitidis* was not considered by these authors as truly panmictic, but rather, as a paradigmatic case of epidemic clonality/semiclonality population structure (Maynard Smith et al., 1993; Maiden, 2006), a situation in which ephemeral clones propagate themselves in an otherwise recombining species. Such a pattern is consistent with the presence of some LD, but should be incompatible with the presence of MLBTs and deep phylogenies, since such phylogenies should be erased in the long term by prevalent genetic recombination.

The *N. meningitidis* group is a monophyletic* subset of the *Neisseria* genus (Fig. 5.6, left).

We have evidenced (Bart et al., 2001) the presence, within the *N. meningitidis* serogroup A, of deep phylogenies corroborated by isoenzymes, RAPD and MLST. Such a congruence between very different kinds of genetic markers is a strong indication of the robustness of the phylogeny considered (test g; Tibayrenc et al., 1990).

Since then, indications against the epidemic clonality/semiclonality model have accumulated (Tibayrenc and Ayala, 2015). According to this model, clonality should be limited to the propagation of ephemeral clones in an otherwise recombining species (Fig. 5.4, bottom). However, against this statement, genetic recombination does not prevent the possibility to observe discrete, widespread and stable lineages and "clonal complexes* (CCs*) in *N. meningitidis* (Caugant and Brynildsrud, 2020).

Deep phylogenies and MLBTs, fairly corroborated by phylogenetic analysis, have been identified in *N. meningitidis* under the name of "phylogenetic clades" (=NCs) (Budroni et al., 2011), which contradicts the semiclonal model. Each of the NCs recorded by Budroni et al. (2011) is subdivided into CCs, following an RDP (Fig. 5.2 in Budroni et al., 2011). Various other examples of RDPs can be found in this species: The worldwide-widespread hyperinvasive "lineage ET5-ST32," a tiny part of *N. meningitidis*, that can be equated to an NC, is divided into three sublineages or lesser NCs (Harrison et al., 2015). The MLG ST2881 in Ghana is partitioned into 5 NCs (Lamelas et al., 2017) that are fairly supported by phylogenetic analysis, even if SNPs involved in recombination are considered (G. Pluschke, personal communication). In Hao et al. (2011), a complex clustering pattern into many well-supported CCs can be observed (see their Fig. 5.1), as well as additional clustering within each of these CCs (RDP). Upper branches are weaklier supported. However, several of them fit the criterion of bootstrap at 0.70 (Hillis and Bull, 1993). Within each CC, many STs have strong bootstrap support when the whole genome sequences are used to design the tree (W. Hao, personal communication). Four major phylogenetic groups A to D (=NCs) have been recorded in *N. meningitidis* by Bratcher et al. (2014) (their Fig. 5.3). Group A includes lineages 1, 4, and 10 (lesser NCs); group B includes lineages 8 and 11, group C includes lineages 2 and 5. The CC 5, which can be equated as a tiny subdivision of *N. meningitidis*, fits the definition of an NC and itself includes various lesser NCs, most of them being fairly supported (Diallo et al., 2017). MacAlasdair et al. (2022) have surveyed with genomic data a vast sample of *N. meningitidis* strains from Burkina Faso, from 2009 to 2012. They have identified nine very persistent clusters (=NCs) that exhibit broad links with phenotypes (serogroups). The congruence between molecular markers is a strong indication that the NCs are robust. The link between genetic/genomic data and phenotypes has been already noted for *E. coli* (Miller and Hartl, 1986) and is one of the main PCE traits. According to MacAlasdair et al. (2022), this *N. meningitidis* sample exhibits a "deeply stratified population structure, with clusters (NCs) evolving largely independently." Although the authors note different recombination rates among the lineages, the overall pattern does not fit the semiclonal model and much better supports PCE.

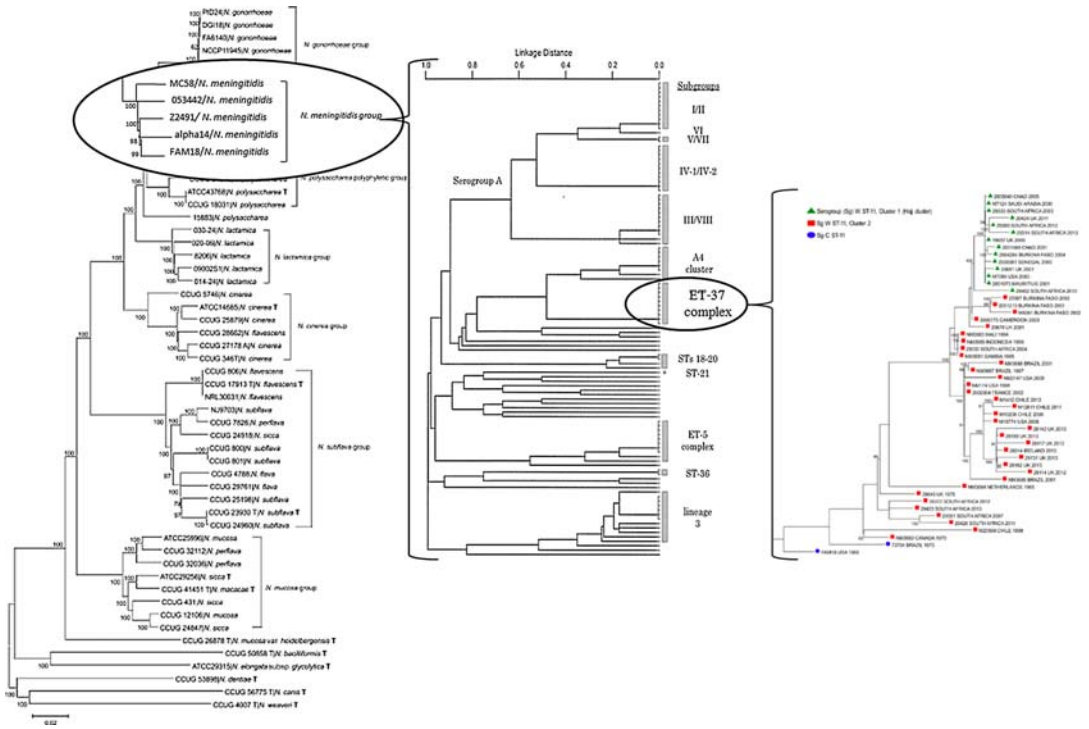


FIGURE 5.6

Russian doll pattern within the genus *Neisseria*. From left to right: (1) various species within the genus *Neisseria*. The species *N. meningitidis* constitutes only a tiny part of this set (second group, top); (2) various clusters (=near-clades) of the species *N. meningitidis* individualized by multilocus sequence typing (MLST). The "ET*-37 complex" (ET = electrophoretic type*) is a tiny, monomorphic, subdivision (lesser near-clade) of the whole species; (3) with high-resolution genomic markers (multi-SNP typing), the "clone" ST11, which corresponds to ET-37, appears to be highly variable and strongly structured into various well-supported lesser near-clades that are only partially linked to geographical distance.

(1) After Figure 2 in Bennett, J.S. et al., 2012. A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus Neisseria. Microbiology 158 (6), 1570–1580. https://doi.org/10.1099/mic.0.056077-0. (2) After Figure 2 in Maiden, M.C.J., et al., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95 (6), 3140–3145. https://doi.org/10.1073/pnas.95.6.3140. (3) after Figure 3 in Mustapha, M.M., et al., 2015. Genomic epidemiology of hypervirulent serogroup W, ST-11 Neisseria meningitidis. EBioMedicine 2 (10), 1447–1455. https://doi.org/10.1016/j.ebiom.2015.09.007.

Restriction-modification systems and DNA uptake sequences could be specific of lineages and therefore explain why genetic recombination is inhibited among NCs in this species (Caugant and Brynildsrud, 2020).

Some *N. meningitidis* STs, that can be equated to clonal MLST MLGs, persist over broad geographical and time scales (Buckee et al., 2008), which does not fit the predictions of the epidemic/semiclonal model. Table 5.1 in Caugant (2008) shows the strong persistence and vast geographical spread of many STs. As an example, the ST1 complex has been recorded from 1930 to 2007 in China

Table 5.1 The Many Different Terms Recorded in the Literature to Designate Pathogens' Genetic Clusters

Viruses	Bacteria	Parasitic Protozoa	Fungi
Clades	Clades	Assemblages	AFLP groups
Clusters	Clonal complexes	Clades	Clades
Genogroups	Clonal lineages	Clonal haplogroups	Clonal groups
Genotypes	Clonal subgroups	Clonal haplotypes	Clonal lineages
Groups	Clusters	Clonal lineages	Clusters
Lineages	Family strains	Clonal types	Clonal groups
Major genotypes	Genetic groups	Clones	Genetically distinct subgroups
Major lineages	Genoclouds	Clonotypes	Genotypes
Phylogenetic groups	Genogroups	Clusters	Genotypic groups
Phylogroups	Genome groups	Core subgroups	Groups
Subclades	Genospecies	Discrete typing units (DTUs)	Lineages
Subgenotypes	Groups	Divergent entities	Major clades
Subgenotype clusters	Lineages	Genetic groups	Minor clades
Subgroups	Major branches	Genetic types	Molecular genotypes
Sublineages	Major clusters	Genotypes	Molecular types
Substrains	Main/major lineages	Groups	Phylogenetic species
Subtypes	Major phylogenetic groups	Haplogroups	Subclusters
Subvariants	Phylogenetic clades	Haplotypes	Subgenotypes
Types	Phylogenetic groupings	Lesser subgroups	Subgroups
Variants	Populations	Lineages	Subpopulations
	Primary clusters	Main haplogroups	Varieties
	Principal genetic groups	Major clades	
	Secondary clusters	Major clonal lineages	
	Semi discrete lineages	Major groups	
	Subclones	Major monophyletic groups	
	Subclusters	Phylogenetic lineages	
	Subgroups	Populations	
	Sublineages	Subassemblages	
	Subpopulations	Subclades	
	Subspecies	Subclusters	
	Subspecies groups	Subgroups	
	Subtypes	Sublineages	
		Subpopulations	
		Subgenotypes	
		Subgroups	
		Subspecies	
		Subtypes	
		Subtype groups	
		Types	

and Russia. The ST4 complex has been sampled in Gambia from 1917 to 1992. The worldwide ST11 complex (=ET-37; see Fig. 5.6) has been recorded from 1960 to 2006. These traits do not fit the term of "ephemeral clones."

4. Parasite Species

The clonality/sexuality debate has been and still is, remarkably lively in the case of parasitic protozoa.

4.1 *Leishmania donovani/infantum*

The *Leishmania donovani* (*L. donovani/infantum*) "complex" includes two very closely related species. We have proposed (Tibayrenc et al., 1990) that these species, together with other *Leishmania* species, have a clonal population structure. As recalled many times, this is compatible with occasional bouts of genetic exchange, including in *Leishmania* parasites (Serafim et al., 2023).

L. infantum exhibits a remarkable case of a widespread MLEE genotype: it is the MLEE MLG termed "MON1." MON refers to the faculty of medicine in Montpellier, France, where a highly standardized MLEE technology has been designed since the 1980s. This technology relies on the survey of 15 isoenzyme loci on starch gel electrophoresis. A "MON" with a given numbering therefore corresponds to all MLGs that appear identical in the MON system (Pratlong et al., 2013). By definition, MON1 then is monomorphic with this MLEE technology. MON1 is a very frequent *L. infantum* MLEE MLG. It proves to have a widespread distribution. It has been sampled in the Near and Middle-East, Africa, North Africa, Southern Europe, and South America (used to be termed: "*L. chagasi*" in that part of the world). MON1 is not a true clone, but rather, a very typical "clonet." As a matter of fact, when performing microsatellite* analysis, which has a much higher resolution power than MLEE, MON1, which is only a tiny part of *L. infantum* (Fig. 5.7, left), proves to be highly heterogeneous. This diversity is distributed into many different, fairly supported lesser NCs, in a typical RDP (Fig. 5.7, right).

Another microsatellite study (Rugna et al., 2018; see their Fig. 5.4) confirms the monophyletism of MON1 and its subdivision into various lesser NCs.

Again, through microsatellite analysis (Amro et al., 2013), in Moroccan *L. infantum* strains, MON1 is subdivided into two clusters (lesser NCs). One includes Moroccan strains, together with European strains. The other one is composed of other Moroccan strains, with Algerian and Tunisian strains. This phylogenetic pattern is corroborated by the STRUCTURE software, which supports its phylogenetic robustness. Each of the two MON1 inner clusters exhibits many lesser NCs (Figs. 5.2 and 5.3 in Amro et al., 2013). IBDT certainly influences this phylogenetic pattern, however, not at 100%.

In Tunisia, Chargui et al. (2009) have similarly evidenced many, well-supported lesser NCs withing MON1 by microsatellite analysis (their Fig. 5.3).

Castelli et al. (2020) have analyzed 78 *L. infantum* strains from Sicilia with microsatellites. They have recorded two main clusters, one being linked to human hosts, the other one, to dogs. Within each of these two clusters, many lesser NCs with fair bootstrap values are observed (Fig. 5.5 in Castelli et al. (2020)). 91% of the isolates were attributed to MON1. Geographical separation can only partially explain this overall pattern.

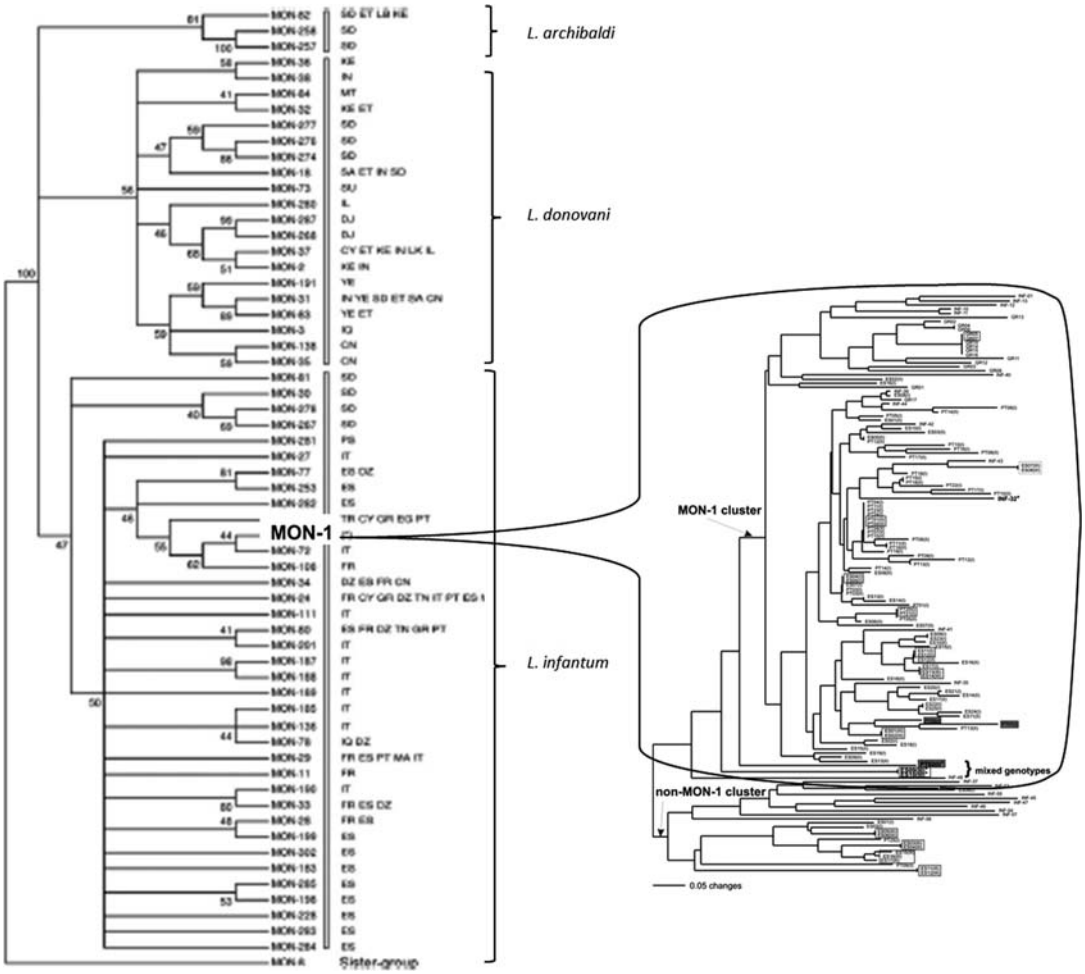


FIGURE 5.7

An illustrative case of RDP in the species *Leishmania infantum*. Left: a phylogenetic tree of the whole "*Leishmania donovani/infantum* complex," designed from MLEE analysis. Right: The MLEE MLG MON1, which is a tiny part of the *L. infantum* species (left) appears considerably heterogeneous and structured through microsatellite analysis.

Left: After Pratlong, F., et al., 2013. Geographical distribution and epidemiological features of Old World Leishmania infantum and Leishmania donovani foci, based on the isoenzyme analysis of 2277 strains. Parasitology 140 (4), 423–434. https://doi.org/10.1017/S0031182012001825. Right: After Kuhls, K., et al., 2008. Differentiation and gene flow among European populations of Leishmania infantum Mon-1. PLoS Neglected Trop. Dis. 2 (7), e261. https://doi.org/10.1371/journal.pntd.0000261.

Ochsenreither et al. (2006) have evidenced the presence of three microsatellite subclusters within MON1, with many lesser NCs. Phylogenetic robustness is supported by several high levels of bootstrap. This pattern cannot be accounted for by geographical distance (Fig. 5.2 in Ochsenreither et al. (2006)).

More recently, Kuhls et al. (2021) have evidenced a complex clustering pattern with many lesser NCs within MON1 *L. infantum* strains from Armenia (their Figs. 5.2 and 5.3).

The monophyletism of MON1, a then small part of the taxon *L. infantum* (Pratlong et al., 2013; see Fig. 5.7), is corroborated by both microsatellites (Kuhls et al., 2008; Rugna et al., 2018 and genomics* (Franssen et al., 2020). This parity between three different categories of genetic markers relying on quite different parts of the genome, namely MLEE, microsatellites and genomics, is a strong support for the robustness of the inferred phylogenies, and for PCE ("g test" of clonality: Tibayrenc et al., 1990; congruence criterion: Tibayrenc and Ayala, 2017a). As a matter of fact, this evidences an overall LD that involves the whole genome of the species under study.

Whole genome sequencing and the use of 395,684 SNPs of 151 isolates of *L. donovani/infantum* fully confirm that MON1 (the "MON1 clade") is quite heterogeneous and structured into various lesser NCs (see Fig. 5.1 suppl. 2, A in Franssen et al., 2020).

The MLEE newly identified genotype MON309, a small part of the *L. infantum* species, appears monophyletic through WGS and PCR-RFLP typing (Rogers et al., 2014). Within MON309, "mainly clonal reproduction in the parasite population" and strong LD are observed (Rogers et al., 2014). This shows that PCE traits are recorded within this small subdivision of the species, which is definitely not panmictic (RDP).

Although microsatellites have a high-resolution power, they do not evidence any polymorphism in Nepal and India within the *L. donovani* microsatellite Core 191 genotype, a small subdivision of the species (Imamura et al., 2016). Now this Core 191 genotype appears to be highly polymorphic and strongly structured into six monophyletic groups by genomic analysis: see figure suppl. 2 in Imamura et al. (2016). This study definitely deals with microevolution, since the age of the common ancestor of the clusters ISC2 and ISC 4−6 is around the year 1960 and the Core 191 genotype originates in mid-19th century. *L. donovani* therefore appears as a true "measurably evolving pathogen" (Biek et al., 2015).

Schwabl et al. (2021) surveyed the genome of 207 *L. infantum* strains from Brazil through WGS and multi-SNP typing. They inferred that, "similar to observations in *Trypanosoma cruzi* (*T. cruzi*,), different rates of sex and clonality may occur in *L. infantum*, depending on demographic or ecological variation within landscapes or between the parasite's evolutionarily native (Old World) and introduced (New World) range." Now the phylogenetic tree of their Fig. 5.4 clearly shows that the Brazilian strains are clustered into various lesser NCs, and that geographical separation cannot account for this overall pattern. According to the authors (Boité, personal communication), all these Brazilian strains pertain to the MLEE MLG MON1, again showing a remarkable RDP within this MLEE MLG.

4.2 Trypanosoma cruzi

T. cruzi, the agent of Chagas disease, has become one of the most studied pathogens by evolutionists, to the point that it could have become a classic evolutionary model like *Drosophila melanogaster*, *Mus musculus*, *Caenorhabditis elegans*, or *Escherichia coli*. Only the fact that it is highly pathogenic has hampered its status as a star of basic biology. We have recently exposed in details the status of *T. cruzi*

regarding the PCE model (Tibayrenc and Ayala, 2022). Only the main features and new data will be treated in this chapter.

T. cruzi has been one of the favorite "bugs" of parasitologists interested in the "clonality/sexuality" debate. Early MLEE studies have shown that sexuality was hampered in *T. cruzi* at the level of the whole species (Tibayrenc et al., 1981b). Later, still according to MLEE data dealing with a much more diversified strain sample and more MLEE loci, it has been inferred that *T. cruzi* featured a "complex multiclonal structure" (Tibayrenc et al., 1986). The arguments for this proposal were as follows: at the level of the whole species, a strong LD is observed, and several MLGs are widespread over a vast ecogeographical range, with frequencies that are totally incompatible with panmictic expectations. The introduction of DNA data (RAPD) made it possible to identify six within-species "discrete typing units*" or DTUs* (Tibayrenc, 1998), namely TC I to VI (Brisse et al., 2000; Zingales et al., 2012). Evolutionarily speaking, *T. cruzi* DTUs perfectly fit the definition of NCs. Indications for occasional hybridization are recorded among some of them (Brisse et al., 2003). However, these hybridization events are unable to erase the prevalent PCE pattern, namely the presence of MLBTs and the persistent structuring into six NCs. More than 20 years later, the NCs still persist unchanged, and "the frequency of natural genetic exchange events between lineages/clades/DTUs appears to be too low to disrupt the overall stability of the main six clades" (Machado Matos et al., 2022). The robustness of these six NCs has been recently supported by the evidence that different genetic markers show an overall congruent phylogenetic pattern (Barnabé et al., 2023).

An additional NC, specific of bats, has been more recently identified (Pinto et al., 2012).

At the intraspecies level, *T. cruzi* exhibits various cases of RDPs within each of the main NCs, especially within TcI.

A very illustrative example is given by a genomic analysis of some Ecuadorian strains (Schwabl et al., 2019). This study enlightens the fact that, despite occasional mating events, the main PCE traits are conserved at a microevolutionary level. Similarly, as we have seen it, at the whole *T. cruzi* species level, hybridization events do not break up the prevalent PCE pattern attested by deep phylogenies and stable MLBTs. The study in Ecuador has evidenced within the NC TcI the presence of two distinct genetic clusters (= lesser NCs), hence an RDP. In one of these lesser NCs, the authors have found fair evidence for meiosis events. No such events have been detected in the other lesser NC. However, even the cluster harboring meiosis is far from panmixia and still shows the PCE traits: the phylogenetic signal remains strong: "support is unambiguous for main clusters and high within subclusters, except where last branch lengths are quite short in Cluster 2" (P. Schwabl, personal communication) (see Fig. 5.8).

The claim that the population corresponding to this first cluster is "almost panmictic*" (Rusman et al., 2020) is therefore misleading.

Many other studies involving various countries evidence within the NC TcI the presence of typical RDPs with statistically significant LD and MLBTs supported by fair bootstrap values.

In Ecuador, within TcI, two different populations, corresponding partly to peridomestic and sylvatic cycles, have been individualized by 23 microsatellite loci. A strongly significant LD ($P < 10^{-3}$) has been recorded in the total population as well as in each of the two subpopulations (Costales et al., 2015).

In Brazil, three bootstrap-supported clusters with many subclusters (lesser NCs) have been evidenced within TcI (Fig. 5.2 in Roman et al., 2018).

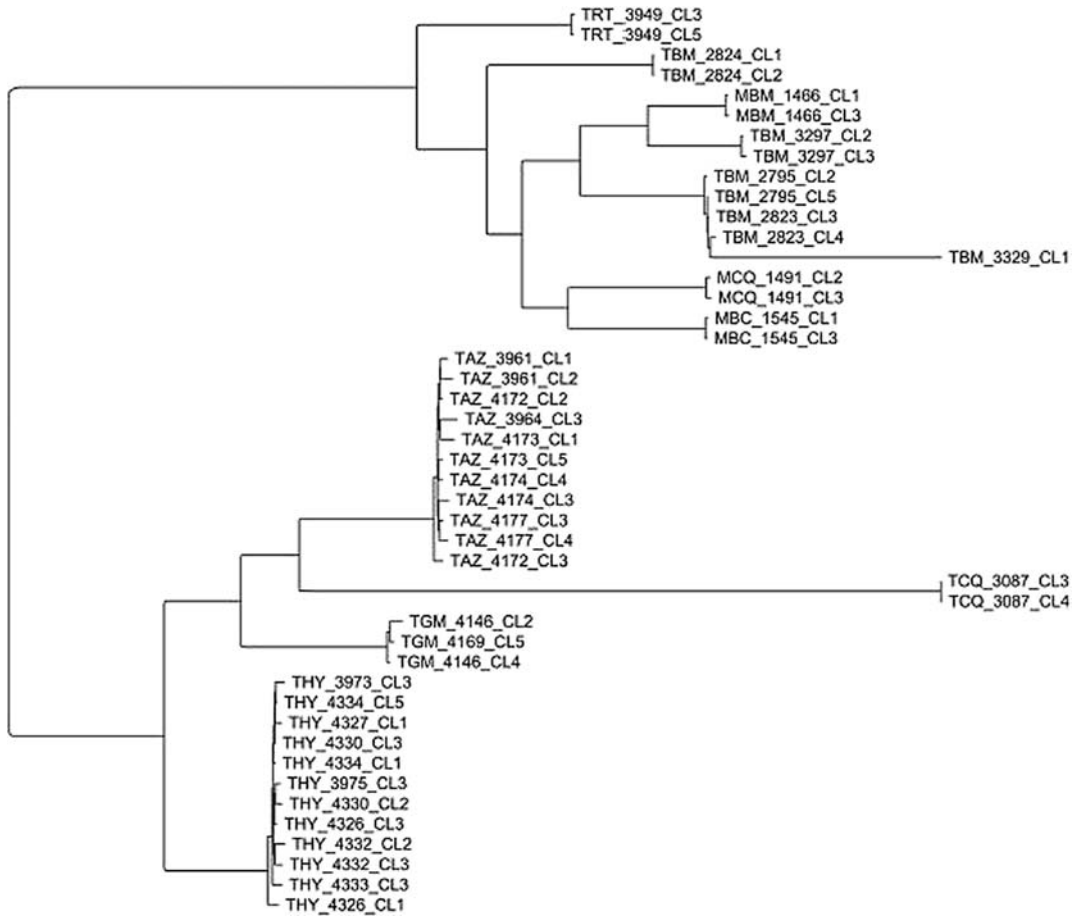


FIGURE 5.8

Two lesser near-clades within the *Trypanosoma cruzi* TcI near-clade in Ecuador. Despite convincing indications of meiosis in the top cluster, a strong phylogenetic signal is observed at the level of the whole sample and within each of the two lesser near-clades.

After Schwabl, P., et al., 2019. Meiotic sex in Chagas disease parasite Trypanosoma cruzi. Nat. Commun. 10 (1). https://doi.org/10.1038/s41467-019-11771-z; their suppl. Figure 1.

Within the NC TcII, a phylogenetic signal has been detected by a genomic analysis (Reis-Cunha et al., 2018).

Lastly, each of the six *T. cruzi* NCs (Brisse et al., 2000; Zingales et al., 2012) are clustered into many lesser NCs (Fig. 5.9), with highly significant bootstrap values (Ramírez et al., 2017) (J.C. Ramírez, personal communication).

It remains to know whether genetic clustering and lesser NCs within the *T. cruzi* NCs show constant patterns over space and time, as it is the case in the pathogenic yeast *Cryptococcus neoformans/gattii*

for example (see further). This has been hypothesized for the subdivisions observed within TcI (Guhl and Ramírez, 2011). If this is verified, lesser *T. cruzi* NCs would behave like "simili-taxa" (Tibayrenc and Ayala, 2022) and could deserve taxonomic nomenclature. The following results seem to support this hypothesis:

The lesser NC "TCIDom" (Zumaya-Estrada et al., 2012) seems to correspond to Chaco-1 (Tomasini et al., 2014), to VENdom (Llewellyn et al., 2009), and to Haplotype A (Herrera et al., 2007). This lesser NC would have been therefore identified by several studies relying on different genetic markers (mini-exon gene intergenic region, microsatellites, MLST) in several countries (Argentina, Colombia, Venezuela) and ecosystems. It therefore appears as a reliable and stable lesser NC of the NC TcI. It seems to be specifically linked to domestic cycles.

Similarly, the lesser TcI NC Chaco-4 (Tomasini et al., 2014) seems to be identical to the lesser TcI NC SL-IR group E characterized by the intergenic region of spliced leader (Cura et al., 2010). It has been recorded in Argentina, Bolivia, and Chile.

Lastly, the lesser NC TcIV-USA has been sampled in Arizona, California, New Mexico, and Texas (Flores-López et al., 2022).

According to the above-summarized data, *T. cruzi* remains a paradigmatic case for the PCE model. Hybridization events at the species level (Brisse et al., 2003) and meiotic processes at a lesser evolutionary level (Schwabl et al., 2019) most certainly have a high evolutionary and medical relevance. However, they are unable to disrupt the "clonality frame" with deep phylogenies and MLBTs, neither at the species level, nor down to microevolutionary levels.

5. Yeasts

5.1 The *Cryptococcus neoformans/gattii* Complex

Two species compose this complex, namely *Cryptococcus neoformans* and *C. gattii.* Several "molecular types" have been identified within each of these two species (Desjardins et al., 2017; Samarasinghe and Xu, 2018). As we will see it, these "molecular types" perfectly fit the evolutionary definition of NCs (Tibayrenc and Ayala, 2014b). New species names have been proposed for these molecular types (Table 5.1 in Samarasinghe and Xu, 2018). Now, the boundaries between *C. neoformans* and *C. gattii* and among the molecular types are not strict, with hybridization phenomena and occasional genetic transfer (Samarasinghe and Xu, 2018). This perfectly fits the NC definition of NC and closely mimics the comparable phenomena observed in *T. cruzi*.

Authors working on these yeasts emphasize the specific mating strategies of them: "same-sex," "unisex" (Samarasinghe and Xu, 2018). These phenomena certainly have a high evolutionary and epidemiological relevance. However, when exclusively focusing on them, there is a risk of hiding important traits that *C. neoformans* and *C. gattii* share with other pathogens. This is the same with hybridization and meiosis events recorded in *T. cruzi* (Brisse et al., 2003; Schwabl et al., 2019). These phenomena do not break up the prevalent PCE pattern, with LD, widespread MLGS, deep phylogenies and MLBTs at all evolutionary scales.

In the VNI genotype of *C. neoformans* "variety *grubii*," many sequence types (STs), that can be equated to clonal MLGs, are widely distributed (Ferreira-Paim et al., 2017). Another clonal MLG, the VNB genotype of the *C. neoformans* "variety *grubii*," has been recorded both in South America and in Africa (Rhodes et al., 2017).

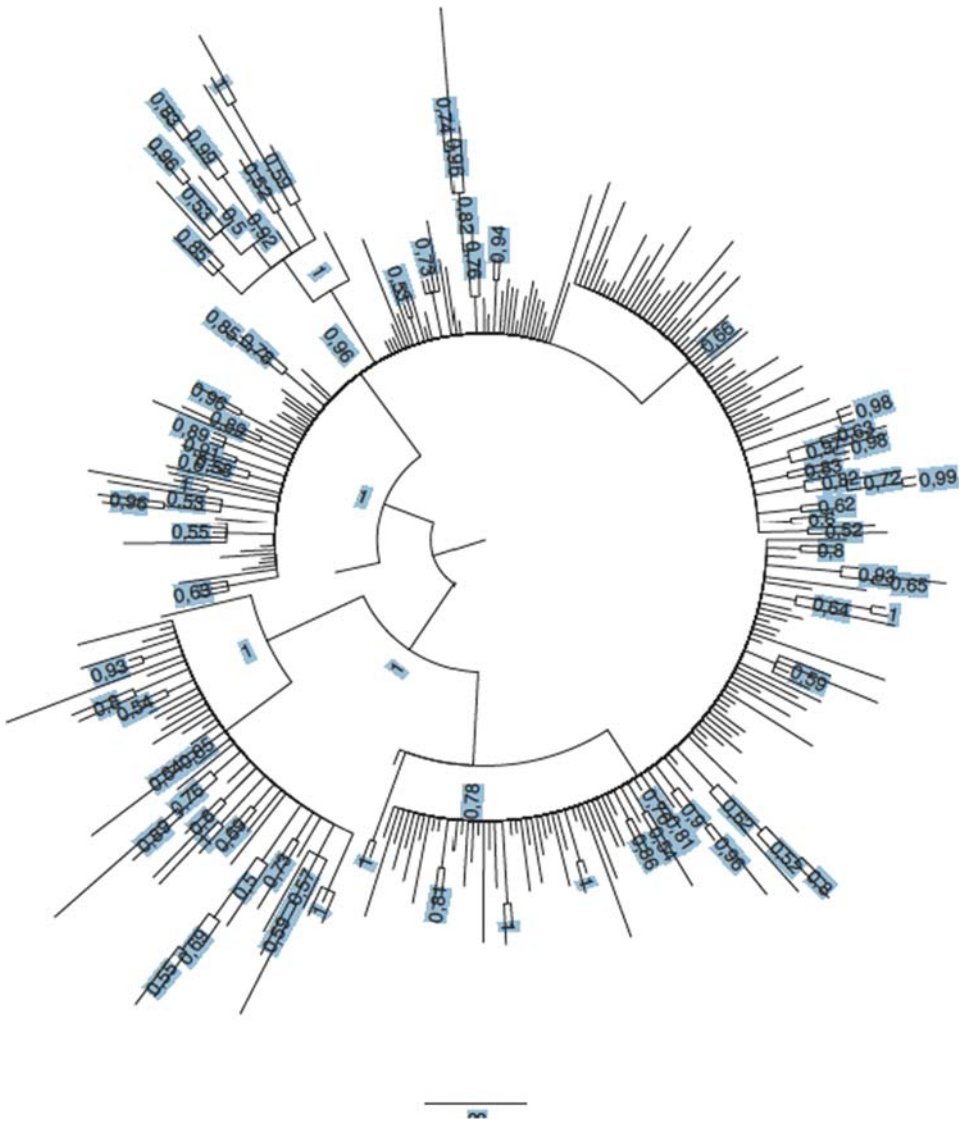


FIGURE 5.9

The analysis by 335 independent satellite DNA sequences of 19 *T. cruzi* isolates shows various lesser NCs within each of the six *T. cruzi* main NCs. These lesser NCs are supported by strongly significant bootstrap values (J.C. Ramírez, personal communication).

After Ramírez, J.C. et al., 2017. New insights into Trypanosoma cruzi evolution, genotyping and molecular diagnostics from satellite DNA sequence analysis. PLoS Negl. Trop. Dis. 11 (12), e0006139. https://doi.org/10.1371/journal.pntd.0006139; original figure communicated by J.C. Ramírez.

This "variety *grubii*" is a genetic subdivision of *C. neoformans*. It is actually an NC of this species. It includes three molecular types, VNI, VNII, and VNB (Table 5.1 in Samarasinghe and Xu, 2018) that correspond to lesser NCs (RDP). The genotype (NC) VNI is structured into many lesser NCs (Fig. 5.2 in Ferreira-Paim et al., 2017). Many lesser NCs are evidenced within the B, VNI, and VNII genotypes (NCs) (Fig. 5.10). Most of the branchings in the SNP MLBT (Fig. 5.10) are strongly supported (C.A. Cuomo, personal communication). These results parallel those by Rhodes et al. (2017) (see their Fig. 5.1), and Vanhove et al. (2017) with different sampling procedures. Similar branching patterns are observed in these three studies, which corroborates the stability of these subdivisions over space and time.

Such a stability of the lesser *C. neoformans* NCs makes that they behave like simili-taxa and could deserve a nomenclature procedure.

6. Concluding Remarks

The results above recalled demonstrate a striking similarity of PCE traits in the surveyed organisms, that include bacteria, parasitic protozoa, and a yeast. Other examples, including parasite protozoa (*Giardia intestinalis*, *Toxoplasma gondii*, *Trypanosoma congolense*) and bacteria (*Bacillus cereus*, *B. anthracis*, *Neisseria gonorrhoeae*, *Salmonella enterica*, *Shigella flexneri*, *S. sonnei*, *Staphylococcus aureus*) have been recently exposed (Tibayrenc and Ayala, 2021a).

To our knowledge, this is the first time that such a similarity of relevant evolutionary traits has been evidenced in pathogenic microorganisms that feature a considerable evolutionary and taxonomical disparity. This is to the point that a blind lecture of evolutionary data could make it impossible to guess which species is involved. We have called this pattern "evolutionary twins" (Tibayrenc and Ayala, 2021a).

The broad, global hypothesis of PCE should not lead to underemphasize the specificities of each model, for example, selfing/homogamy versus mitotic clonality, hybridization, occasional meiotic events in *T. cruzi*, "same-sex" and "unisex" in *C. neoformans*. Such specificities certainly have a considerable epidemiological and evolutionary relevance. However, focusing exclusively on these specific cases makes that "the trees will hide the forest." One of the goals of science is to design very general hypotheses and models that are able to harbor different cases sharing relevant similarities. This is the case for example of horizontal gene transfer (HGT) in bacteria, that include conjugation, transduction, and transformation. This is also the case of PCE, that includes mitotic reproduction, selfing/homogamy and several cases of parthenogenesis, and is compatible with some degree of genetic exchange, provided that it does not erase the prevalent PCE pattern, with deep phylogenies and MLBTs.

When terminology is concerned, since the evolutionary status of the genotypes observed within the species level was quite unclear, an incredibly confusing semantics has been generated (Table 5.1).

The near-clade concept, which has a clear evolutionary definition, makes it possible to discard this confusing terminology and to replace all these terms with a unique one.

The PCE pattern, by definition, implies the presence, in the species featuring it, of MLBTs, deep phylogenies, NCs, and widespread clonal MLGs at all evolutionary scales, down to microevolutionary levels (RDPs). All these entities are stable in space and time. They therefore constitute relevant units of analysis (Tibayrenc and Ayala, 2017b) for epidemiological tracking. Actually, it can be said that a PCE

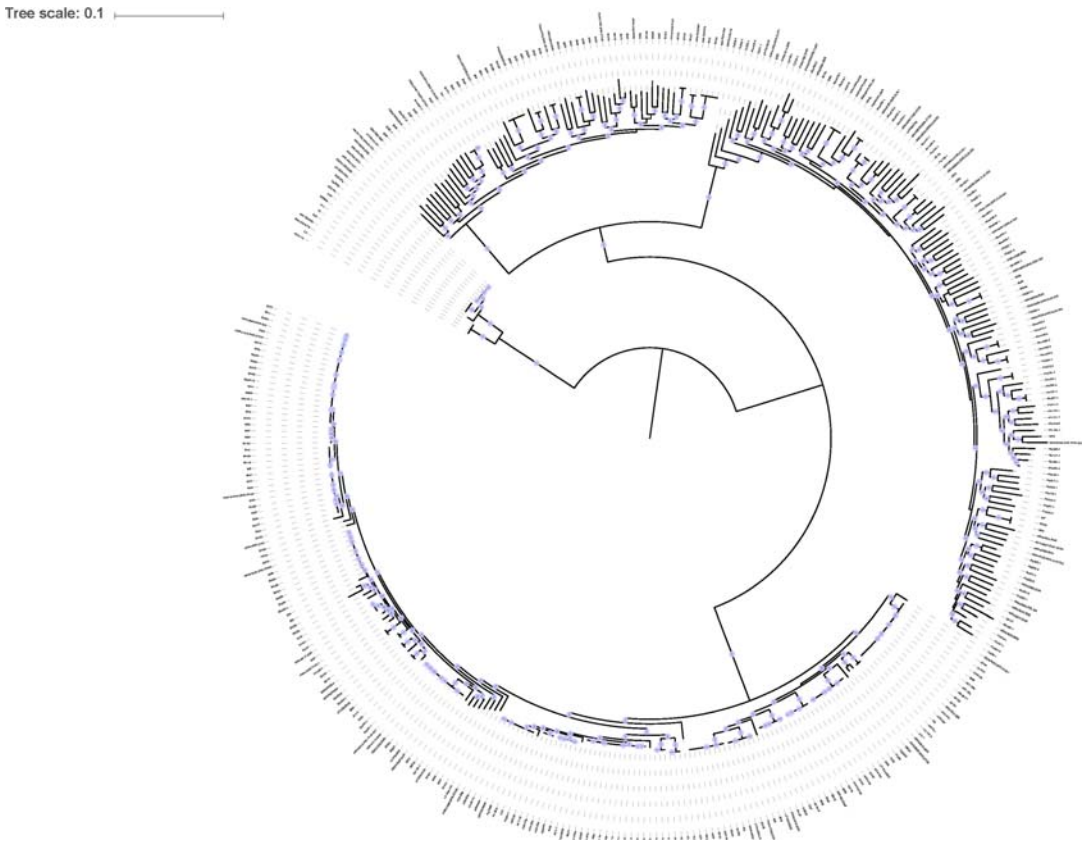


FIGURE 5.10

Many lesser NCs are recorded within the VNI, VNII, and VNB genotypes (=NCs) of the yeast *Cryptococcus neoformans* variety *grubii*. "Most of the nodes in the whole genome SNP tree are highly supported" (C.A. Cuomo, personal communication).

After Figure 1 in Desjardins, C.A. et al., 2017. Population genomics and the evolution of virulence in the fungal pathogen Cryptococcus neoformans. Genome Res. 27 (7), 1207–1219. doi:10.1101/gr.218727.116.

pattern is a prerequisite for any attempt to molecular epidemiology. If PCE does not obtain, MLGs vanish soon in the common gene pool due to frequent recombination. They cannot therefore be traced in the long term.

Moreover, since NCs are discrete and behave like "simili-taxa" (Tibayrenc and Ayala, 2021a), they can be taken as bases for the description of new species by using a flexible approach of the phylogenetic species concept* (Cracraft, 1983). However, the presence of RDPs at all evolutionary scales means that the potential number of phylogenetic species is virtually unlimited. It will be the task of the specialists of the concerned species to decide whether the NCs observed in these species are worthwhile being described as new species. Phylogenetic discreteness should not be the only criterion for

that. Epidemiological and clinical specificities should also be considered. This has been the case for example for the "assemblages" (=NCs) that subdivide the species *Giardia intestinalis*. Since these "assemblages" feature some host and epidemiological specificities, at least statistically, it has been proposed (Cacciò and Sprong, 2010) to call assemblage A: *Giardia duodenalis* and assemblage B, *G. enterica*.

Clonal framework/deep phylogenies/MLBTs/NCs open the possibility to design a "phylogenetic character mapping" approach (Avise, 2004). When exploring the evolution of a given pathogen species, it is recommendable to first establish its phylogenetic framework. Only in a second time, relevant phenotypic traits (pathogenicity, drug resistance, epidemiological traits) will be placed onto this overall phylogenetic framework.

Glossary of Specialized Terms

Amplification fragment length polymorphism (AFLP) Selective amplification of genomic restriction fragments (obtained by RFLP) by PCR

Bayesian analysis Uses a likelihood function to create a quantity called the posterior probability of trees using a model of evolution, based on some prior probabilities, yielding the most likely phylogenetic tree for the data under study

Bootstrap analysis In phylogenetic analysis, the generation of pseudoreplicate data sets by sampling at random the original character matrix to create new matrices that are the same size as the original. The frequency by which a given branch is reproduced through this randomization procedure is recorded as the bootstrap proportion. These proportions can be used as a measure of the robustness of individual branches in the tree. A bootstrap value of 70 for a given branch means that this branch has been found 70 times out of 100 by the procedure (Hillis and Bull, 1993)

Clonal complex (CC) A set of genetically similar genotypes identified by MLST

Clonet[§] A set of MLGs that are identical for a given genetic marker in a species that undergoes predominant clonal evolution (Tibayrenc et al., 1991)

Discrete typing unit[§] **(DTU)** Sets of strains that are genetically more similar to each other than to any other stock, are stable in time and space, and are identifiable by common immunological, molecular or genetic markers called tags (Tibayrenc, 1998). Should be replaced by the term "near-clade" (Tibayrenc and Ayala, 2012), which has a clearer evolutionary meaning

Electrophoretic type (ET) MLG identified by isoenzyme electrophoresis

Epidemic clonality Semiclonality (see this term)

Genetic recombination Exchange of genetic material occurring at two or more different genetic loci between different individuals, which leads to the production of offspring with combinations of genetic characters that differ from those observed in either parent

Genomics Field of research that explores the whole genome, while genetics considers individual genes or limited sets of genes

Homogamy The tendency of an organism to mate with individuals that are genetically very similar or identical to itself. It concerns also phenotypic similarity

Isoenzyme Protein extracts from given samples, for example, various pathogen isolates, are separated by electrophoresis. The gel is then subjected to a histochemical reaction involving the specific substrate of a given enzyme, and the zone of activity of this enzyme is specifically stained. The same enzyme from different samples may migrate at different rates that are the manifestation of genetic differences in the genes that code for these enzymes. Isoenzyme typing is therefore an indirect gene typing. The different electrophoretic forms

of the same enzyme are referred to as isoenzymes or isozymes. If a genetic analysis of this polymorphism is performed, and allelic variation is inferred, isoenzymes/isozymes are termed allozymes

Kinetoplast, kinetoplast DNA An organelle specific of the protozoa of the class Kinetoplastida (*Trypanosoma* and *Leishmania*). The kinetoplast is composed of two categories of molecules: (1) The maxicircles are between 20 and 40 kb in size. There are a few dozen of them per kinetoplast, while there are several thousand minicircles per kinetoplast. The size of the minicircles ranges from 0.5 to 1 kb

Linkage disequilibrium (LD) Nonrandom association of genotypes observed at different loci: these genotypes are not shuffled by genetic recombination, which makes it possible to predict with a high probability a given genotype at locus B when one knows the genotype at locus A, whereas this is impossible in sexual organisms. Several factors can generate linkage disequilibrium in sexual organisms also (for example: natural selection of proximity of two loci on the same chromosome). However, especially when high levels of significance are observed, and linkage disequilibrium is observed at a large number of loci, it is considered by specialists of the field (see a number of references in Tibayrenc and Ayala, 2012, 2017a, 2017b) as strong circumstantial evidence that genetic recombination is restrained and the organism undergoes PCE

Macroevolution Related to a scale of time that could go up to geological periods

Microevolution Classically: evolution at the level of a given population. More specifically in this chapter: a scale of time that could go down to a few kilometers and to recent years or at least historical times, a definition that is better adapted to the epidemiology of microbial pathogens and to the testing of the Russian doll model (see text)

Microsatellite A short DNA sequence, usually 1−4 bp long, that is repeated together in a row along the DNA sequence. In humans, as in many other species, there is great variation in the number of repeats from one individual to another and among different populations. Numbers of repeats for a given locus define microsatellite alleles

Monophyletic An evolutionary lineage that originates from a unique common ancestor

Multilocus bifurcating tree (MLBT) Phylogenetic tree that exhibits a hierarchized, bifurcating pattern (see Fig. 5.3) whose robustness is verified by various phylogenetic procedures, for example, bootstrap analysis (Hillis and Bull, 1993)

Multilocus Enzyme Electrophoresis (MLEE) Electrophoretic variants of a set of isoenzyme loci reflecting the variability of the DNA sequence of the genes that code for these enzymes in the species under survey

Multilocus genotype (MLG) A combined genotype identified by several genetic loci

Multilocus sequence typing (MLST) Method of pathogen characterization based on the sequencing of several (generally 8−11) housekeeping genes

Near-Clade[§] (NC[§]) Widespread, discrete and stable phylogenetic lines whose discreteness can be somewhat blurred by occasional genetic exchange. The term has been coined by us (Tibayrenc and Ayala, 2012), since "clade" corresponds to different phylogenetic lines among which no genetic exchange at all happens, which is virtually never observed in the case of pathogenic microorganisms

Panmixia, panmictic A situation in which genetic exchange occurs at random in a given population. Panmictic expectations are the verification of this situation by various population genetics tests

Parthenogenesis Mode of reproduction of some metazoa (insects, amphibians, reptiles) performed without the genetic contribution of a mating partner

Phenotype, phenotypic All observable characteristics of a given individual or a given population distinct from the genome. The phenotype is not limited to morphological characteristics and includes, for example, physiological parameters (pathogenicity, etc.) or biochemical polymorphism

Phylogenetic species concept (Cracraft, 1983) A taxonomical nomenclature system in which species are equated to distinct phylogenetic lineages

Polymerase chain reaction (PCR) A technique that copies the complementary strands of a target DNA chain through a set of cycles until the needed DNA amount is produced. PCR uses synthesized DNA primers whose nucleotide sequences are complementary to the DNA flanking the target region

Predominant clonal evolution[§] **(PCE**[§]**)** A population structure in which the impact of genetic exchange is not enough to disrupt a pattern of predominant clonality, the main manifestation of it being a stable and persistent phylogenetic signal observed at all evolutionary scales in the whole ecogeographical range of the species under survey

Pulse field gel electrophoresis (PFGE) Separation of large DNA fragments by a particular electrophoresis technique using alternately pulsed, perpendicularly oriented electrical fields. Strains that share the same PFGE are referred to as pulse types. In the case of bacteria, the large DNA fragments result from the action of a low-frequency cutter (a bacterial endonuclease whose restriction action has a low frequency) on the bacterial chromosome. It is therefore a specific case of RFLP. With parasitic protozoa (*Trypanosoma* and *Leishmania*) and yeasts, the large DNA fragments correspond to entire chromosomes (molecular karyotype)

Random amplified polymorphic DNA (RAPD) Method of DNA characterization, also called Arbitrarily Primed Polymerase Chain Reaction or AP-PCR. In the classical Polymerase Chain Reaction (PCR) method, the primers used are identified DNA sequences, while the RAPD protocol is based on primers whose sequence is randomly designed

Restriction fragment length polymorphism (RFLP) Variability in the DNA sequenced of the organism under study revealed by the use of restriction endonucleases. The endonuclease cuts the DNA at specific restriction sites, and the polymorphism of the DNA fragments so obtained can be visualized on gels

Russian doll pattern[§] **(RDP**[§]**) (Tibayrenc and Ayala, 2013)** When the population structure of the species under study at lower evolutionary scales (under the level of the NC) looks like a miniature of the whole species' population structure, with widespread clonal MLGs, LD and lesser NCs. This indicates that PCE is also verified at these microevolutionary levels

Selfing, self-fertilization Fertilization of an organism by itself, hence by a genotype that is identical to itself

Semiclonality (= Epidemic clonality) a population structure in which ephemeral clones propagate themselves in an otherwise recombining species (Maiden, 2006; Maynard Smith et al., 1993)

Sequence type (ST) Individual MLG identified by MLST

Single-nucleotide polymorphism (SNP) Polymorphisms of one-letter variations in the DNA sequence

Starving sex hypothesis[§] A situation where genetic recombination does not occur because different genotypes of the species under survey do not have the physical opportunity to mate, although they are potentially capable of interbreeding (Tibayrenc and Ayala, 2014a)

Transformation, transformable bacterium the property of a bacterium to modify its genetic make-up by catching extracellular ("naked") DNA from its environment through a process called transformation

Wahlund effect Classically, in population genetics, the Wahlund effect refers to a situation in which heterozygote deficit is explained by the fact that two distinct populations with different allelic frequencies, between which physical obstacles (time and/or space) obtain, have been wrongly considered as a unique population. Here, the Wahlund effect is taken in a broad sense, when any apparent departures from panmictic expectations are due only to physical obstacles (time and/or space) to genetic exchange

[§] Concepts specifically designed by us in the framework of the PCE model

Acknowledgments

We thank Dr Jenny Telleria (IRD, Montpellier, France) for designing Fig. 5.3.

References

Alnajar, S., Gupta, R.S., 2017. Phylogenomics and comparative genomic studies delineate six main clades within the family Enterobacteriaceae and support the reclassification of several polyphyletic members of the family. Infect. Genet. Evol. 54, 108−127. https://doi.org/10.1016/j.meegid.2017.06.024.

Amro, A., et al., 2013. Moroccan *Leishmania infantum*: genetic diversity and population structure as revealed by multi-locus microsatellite typing. PLoS ONE 8 (10), e77778. https://doi.org/10.1371/journal.pone.0077778.

Ankarklev, J., et al., 2018. A novel high-resolution multilocus sequence typing of *Giardia intestinalis* Assemblage A isolates reveals zoonotic transmission, clonal outbreaks and recombination. Infect. Genet. Evol. 60, 7−16. https://doi.org/10.1016/j.meegid.2018.02.012.

Avise, J.C., 2004. Molecular Markers, Natural History and Evolution. Chapman and Hall.

Barnabé, C., et al., 2023. Revisiting gene typing and phylogeny of *Trypanosoma cruzi* reference strains: comparison of the relevance of mitochondrial DNA, single-copy nuclear DNA, and the intergenic region of mini-exon gene. Infect. Genet. Evol. 115, 105504.

Bart, A., et al., 2001. The population structure of *Neisseria meningitidis* serogroup. A fits the predictions for clonality. Infect. Genet. Evol. 1 (2), 117−122. https://doi.org/10.1016/S1567-1348(01)00011-9.

Ben Zakour, N.L., et al., 2016. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. mBio 7 (2). https://doi.org/10.1128/mBio.00347-16.

Bennett, J.S., et al., 2012. A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. Microbiology 158 (6), 1570−1580. https://doi.org/10.1099/mic.0.056077-0.

Biek, R., et al., 2015. Measurably evolving pathogens in the genomic era. Trends Ecol. Evol. 30 (6), 306−313. https://doi.org/10.1016/j.tree.2015.03.009.

Bratcher, H.B., et al., 2014. A gene-by-gene population genomics platform: de novo assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. BMC Genom. 15 (1). https://doi.org/10.1186/1471-2164-15-1138.

Bobay, L.M., Traverse, C.C., Ochman, H., 2015. Impermanence of bacterial clones. Proc. Natl. Acad. Sci. U. S. A. 112 (29), 8893−8900. https://doi.org/10.1073/pnas.1501724112.

Boité, M.C., et al., 2012. New insights on taxonomy, phylogeny and population genetics of *Leishmania* (*Viannia*) parasites based on multilocus sequence analysis. PLoS Negl. Trop. Dis. 6 (11). https://doi.org/10.1371/journal.pntd.0001888.

Brisse, S., Barnabé, C., Tibayrenc, M., 2000. Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. Int. J. Parasitol. 30 (1), 35−44. https://doi.org/10.1016/S0020-7519(99)00168-X.

Brisse, S., et al., 2003. Evidence for genetic exchange and hybridization in *Trypanosoma cruzi b*ased on nucleotide sequences and molecular karyotype. Infect. Genet. Evol. 2 (3), 173−183. https://doi.org/10.1016/S1567-1348(02)00097-7.

Buckee, C.O., et al., 2008. Role of selection in the emergence of lineages and the evolution of virulence in *Neisseria meningitidis*. Proc. Natl. Acad. Sci. U. S. A. 105 (39), 15082−15087. https://doi.org/10.1073/pnas.0712019105.

Budroni, S., et al., 2011. *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. Proc. Natl. Acad. Sci. U. S. A. 108 (11), 4494−4499. https://doi.org/10.1073/pnas.1019751108.

Cacciò, S.M., Sprong, H., 2010. *Giardia duodenalis*: genetic recombination and its implications for taxonomy and molecular epidemiology. Exp. Parasitol. 124 (1), 107−112. https://doi.org/10.1016/j.exppara.2009.02.007.

Castelli, G., et al., 2020. Genetic tools discriminate strains of *Leishmania infantum* isolated from humans and dogs in Sicily, Italy. PLOS Negl. Trop. Dis. 14 (7), e0008465. https://doi.org/10.1371/journal.pntd.0008465.

Caugant, D.A., 2008. Genetics and evolution of *Neisseria meningitidis*: importance for the epidemiology of meningococcal disease. Infect. Genet. Evol. 8 (5), 558−565. https://doi.org/10.1016/j.meegid.2008.04.002.

Caugant, D.A., Brynildsrud, O.B., 2020. *Neisseria meningitidis*: using genomics to understand diversity, evolution and pathogenesis. Nat. Rev. Microbiol. 18 (2), 84−96. https://doi.org/10.1038/s41579-019-0282-6.

Caugant, D.A., Maiden, M.C.J., 2009. Meningococcal carriage and disease-population biology and evolution. Vaccine 27 (2), B64−B70. https://doi.org/10.1016/j.vaccine.2009.04.061.

Chargui, N., et al., 2009. Population structure of Tunisian *Leishmania infantum* and evidence for the existence of hybrids and gene flow between genetically different populations. Int. J. Parasitol. 39 (7), 801−811. https://doi.org/10.1016/j.ijpara.2008.11.016.

Chaudhuri, R.R., Henderson, I.R., 2012. The evolution of the *Escherichia coli* phylogeny. Infect. Genet. Evol. 12 (2), 214−226. https://doi.org/10.1016/j.meegid.2012.01.005.

Clermont, O., et al., 2011. Animal and human pathogenic *Escherichia coli* strains share common genetic backgrounds. Infect. Genet. Evol. 11 (3), 654−662. https://doi.org/10.1016/j.meegid.2011.02.005.

Costales, J.A., et al., 2015. *Trypanosoma cruzi* population dynamics in the Central Ecuadorian Coast. Acta Trop. 151 (1), 88−93. https://doi.org/10.1016/j.actatropica.2015.07.017.

Cracraft, J., 1983. Species concepts and speciation analysis. Curr. Ornithol. 1, 159−187. https://doi.org/10.1007/978-1-4615-6781-3_6.

Cura, C.I., et al., 2010. *Trypanosoma cruzi* I genotypes in different geographical regions and transmission cycles based on a microsatellite motif of the intergenic spacer of spliced-leader genes. Int. J. Parasitol. 40 (14), 1599−1607. https://doi.org/10.1016/j.ijpara.2010.06.006.

de Paula Baptista, R.d.P., et al., 2014. Evidence of substantial recombination among *Trypanosoma cruzi* II strains from Minas Gerais. Infect. Genet. Evol. 22, 183−191. https://doi.org/10.1016/j.meegid.2013.11.021.

Denamur, E., et al., 2021. The population genetics of pathogenic *Escherichia coli*. Nat. Rev. Microbiol. 19 (1), 37−54. https://doi.org/10.1038/s41579-020-0416-x.

Desjardins, C.A., et al., 2017. Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. Genome Res. 27 (7), 1207−1219. https://doi.org/10.1101/gr.218727.116.

Diallo, K., et al., 2017. Hierarchical genomic analysis of carried and invasive serogroup A *Neisseria meningitidis* during the 2011 epidemic in Chad. BMC Genom. 18 (1). https://doi.org/10.1186/s12864-017-3789-0.

Dunne, M.W., et al., 2018. Epidemiology of transmissible diseases: array hybridization and next generation sequencing as universal nucleic acid-mediated typing tools. Infect. Genet. Evol. 63, 332−345. https://doi.org/10.1016/j.meegid.2017.09.019.

Ferreira-Paim, K., et al., 2017. MLST-based population genetic analysis in a global context reveals clonality amongst *Cryptococcus neoformans* var. *grubii* VNI isolates from HIV patients in southeastern Brazil. PLoS Negl. Trop. Dis. 11 (1). https://doi.org/10.1371/journal.pntd.0005223.

Flores-López, C.A., et al., 2022. Phylogenetic diversity of two common *Trypanosoma cruzi* lineages in the Southwestern United States. Infect. Genet. Evol. 99. https://doi.org/10.1016/j.meegid.2022.105251.

Franssen, S.U., et al., 2020. Global genome diversity of the Leishmania donovani complex. Elife 9. https://doi.org/10.7554/eLife.51243.

Gibson, W.C., Marshall, T. F. de, Godfrey, D.G., 1980. Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus trypanozoon. Adv. Parasitol. 18 (C), 175−246. https://doi.org/10.1016/S0065-308X(08)60400-5.

Guhl, F., Ramírez, J.D., 2011. *Trypanosoma cruzi* I diversity: towards the need of genetic subdivision? Acta Trop. 119 (1), 1−4. https://doi.org/10.1016/j.actatropica.2011.04.002.

Gutiérrez-Corbo, C., et al., 2021. Reproduction in trypanosomatids: past and present. Biology 10 (6), 471. https://doi.org/10.3390/biology10060471.

Gómez-Hernández, C., et al., 2019. Evaluation of the multispecies coalescent method to explore intra-*Trypanosoma cruzi* i relationships and genetic diversity. Parasitology 146 (8), 1063−1074. https://doi.org/10.1017/S0031182019000428.

Hanage, W.P., 2016. Not so simple after all: bacteria, their population genetics, and recombination. Cold Spring Harbor Perspect. Biol. 8 (7). https://doi.org/10.1101/cshperspect.a018069.

Hao, W., et al., 2011. Extensive genomic variation within clonal complexes of *Neisseria meningitidis*. Genome Biol. Evol. 3 (1), 1406−1418. https://doi.org/10.1093/gbe/evr119.

Harrison, O.B., et al., 2015. Genomic analysis of the evolution and global spread of hyper-invasive meningococcal lineage 5. EBioMedicine 2 (3), 234−243. https://doi.org/10.1016/j.ebiom.2015.01.004.

Herrera, C., et al., 2007. Identifying four Trypanosoma cruzi I isolate haplotypes from different geographic regions in Colombia. Infect. Genet. Evol. 7 (4), 535−539. https://doi.org/10.1016/j.meegid.2006.12.003.

Hill, D.M.C., et al., 2015. Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study. Lancet Infect. Dis. 15 (12), 1420−1428. https://doi.org/10.1016/S1473-3099(15)00267-4.

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42 (2), 182−192. https://doi.org/10.1093/sysbio/42.2.182.

Imamura, H., et al., 2016. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. Elife 5.

Kuhls, K., et al., 2008. Differentiation and gene flow among European populations of Leishmania infantum Mon-1. PLoS Negl. Trop. Dis. 2 (7), e261. https://doi.org/10.1371/journal.pntd.0000261.

Kuhls, K., et al., 2021. Microsatellite based molecular epidemiology of *Leishmania infantum* from re-emerging foci of visceral leishmaniasis in Armenia and pilot risk assessment by ecological niche modeling. PLoS Negl. Trop. Dis. 15 (4), e0009288. https://doi.org/10.1371/journal.pntd.0009288.

Lamelas, A., et al., 2017. Emergence and genomic diversification of a virulent serogroup W:ST-2881(CC175) *Neisseria meningitidis* clone in the African meningitis belt. Microb. Genom. 3 (8). https://doi.org/10.1099/mgen.0.000120.

Lauthier, J.J., et al., 2012. Candidate targets for Multilocus Sequence Typing of *Trypanosoma cruzi*: validation using parasite stocks from the Chaco Region and a set of reference strains. Infect. Genet. Evol. 12 (2), 350−358. https://doi.org/10.1016/j.meegid.2011.12.008.

Lewis, M.D., et al., 2017. Experimental and natural recombination in Trypanosoma cruzi. In: American Trypanosomiasis Chagas Disease: One Hundred Years of Research, second ed. Elsevier Inc., pp. 455−473. https://doi.org/10.1016/B978-0-12-801029-7.00020-4

Llanes, A., et al., 2022. Genomic diversity and genetic variation of *Leishmania panamensis* within its endemic range. Infect. Genet. Evol. 103. https://doi.org/10.1016/j.meegid.2022.105342.

Llewellyn, M.S., et al., 2009. Genome-scale multilocus microsatellite typing of *Trypanosoma cruzi* discrete typing unit I reveals phylogeographic structure and specific genotypes linked to human infection. PLoS Pathog. 5 (5). https://doi.org/10.1371/journal.ppat.1000410.

Maazoun, R., et al., 1981. Significant du polymorphisme enzymatique chez les leishmanies. Ann. Parasitol. Hum. Comp. 56 (5), 467−475. https://doi.org/10.1051/parasite/1981565467.

MacAlasdair, N., et al., 2022. The effect of recombination on the evolution of a population of *Neisseria meningitidis*. Genome Res. 31, 1258−1268.

Machado Matos, G.M., et al., 2022. Microevolution of Trypanosoma cruzi reveals hybridization and clonal mechanisms driving rapid genome diversification. Elife 11. https://doi.org/10.7554/eLife.75237.

Maiden, M.C.J., 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. 60, 561−588. https://doi.org/10.1146/annurev.micro.59.030804.121325.

Maiden, M.C.J., et al., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95 (6), 3140−3145. https://doi.org/10.1073/pnas.95.6.3140.

Maynard Smith, J.M., Dowson, C.G., Spratt, B.G., 1991. Localized sex in bacteria. Nature 349 (6304), 29−31. https://doi.org/10.1038/349029a0.

Maynard Smith, J.M., et al., 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. U. S. A. 90 (10), 4384−4388. https://doi.org/10.1073/pnas.90.10.4384.

McNally, A., et al., 2016. Accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. PLoS Genet. 12 (9). https://doi.org/10.1371/journaLpgeN1006280.

Miles, M.A., et al., 1978. Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochthonous patients with Chagas' disease in Amazonian Brazil. Nature 272 (5656), 819−821. https://doi.org/10.1038/272819a0.

Miles, M.A., et al., 2009. The molecular epidemiology and phylogeography of *Trypanosoma cruzi* and parallel research on *Leishmania*: looking back and to the future. Parasitology 136, 1509−1528.

Miller, R.D., Hartl, D.L., 1986. Biotyping confirms a nearly clonal population structure in *Escherichia coli*. Evolution 40 (1), 1−12. https://doi.org/10.2307/2408598.

Mustapha, M.M., et al., 2015. Genomic epidemiology of hypervirulent serogroup W, ST-11 *Neisseria meningitidis*. EBioMedicine 2 (10), 1447−1455. https://doi.org/10.1016/j.ebiom.2015.09.007.

Nichols, R., 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16 (7), 358−364. https://doi.org/10.1016/S0169-5347(01)02203-0.

Ochman, H., Selander, R.K., 1984. Evidence for clonal population structure in *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 81 (1 I), 198−201. https://doi.org/10.1073/pnas.81.1.198.

Ochsenreither, S., et al., 2006. Multilocus microsatellite typing as a new tool for discrimination of *Leishmania infantum* Mon-1 strains. J. Clin. Microbiol. 44 (2), 495−503. https://doi.org/10.1128/jcm.44.2.495-503.2006.

Ortega-Pierres, G., et al., 2009. New tools provide further insights into *Giardia* and *Cryptosporidium* biology. Trends Parasitol. 25 (9), 410−416. https://doi.org/10.1016/j.pt.2009.06.002.

Pearson, T., et al., 2009. Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. BMC Biol. 7. https://doi.org/10.1186/1741-7007-7-78.

Pérez-Losada, M., et al., 2006. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect. Genet. Evol. 6 (2), 97−112. https://doi.org/10.1016/j.meegid.2005.02.003.

Pérez-Losada, M., et al., 2013. Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. Infect. Genet. Evol. 16, 38−53. https://doi.org/10.1016/j.meegid.2013.01.009.

Petty, N.K., et al., 2014. Global dissemination of a multidrug resistant *Escherichia coli* clone. Proc. Natl. Acad. Sci. U. S. A. 111 (15), 5694−5699. https://doi.org/10.1073/pnas.1322678111.

Pinto, C.M., et al., 2012. TcBat a bat-exclusive lineage of *Trypanosoma cruzi* in the Panama Canal Zone, with comments on its classification and the use of the 18S rRNA gene for lineage identification. Infect. Genet. Evol. 12, 1328−1332.

Posada, D., Crandall, K.A., Holmes, E.C., 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. 36, 75−97. https://doi.org/10.1146/annurev.genet.36.040202.111115.

Pratlong, F., et al., 2013. Geographical distribution and epidemiological features of Old World *Leishmania infantum* and *Leishmania donovani* foci, based on the isoenzyme analysis of 2277 strains. Parasitology 140 (4), 423−434. https://doi.org/10.1017/S0031182012001825.

Price, L.B., et al., 2013. The epidemic of extended-spectrum-β-lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. mBio 4 (6). https://doi.org/10.1128/mBio.00377-13.

Pybus, O.G., Rambaut, A., 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nat. Rev. Genet. 10 (8), 540−550. https://doi.org/10.1038/nrg2583.

Ramírez, J.D., Llewellyn, M.S., 2014. Reproductive clonality in protozoan pathogens - truth or artefact? Mol. Ecol. 23 (17), 4195−4202. https://doi.org/10.1111/mec.12872.

Ramírez, J.D., Tapia-Calle, G., Guhl, F., 2013. Genetic structure of *Trypanosoma cruzi* in Colombia revealed by a high-throughput nuclear multilocus sequence typing (nMLST) approach. BMC Genet. 14. https://doi.org/10.1186/1471-2156-14-96.

Ramírez, J.C., et al., 2017. New insights into *Trypanosoma cruzi* evolution, genotyping and molecular diagnostics from satellite DNA sequence analysis. PLoS Negl. Trop. Dis. 11 (12), e0006139. https://doi.org/10.1371/journal.pntd.0006139.

Reis-Cunha, J.L., et al., 2018. Whole genome sequencing of Trypanosoma cruzi field isolates reveals extensive genomic variability and complex aneuploidy patterns within TcII DTU. BMC Genom. 19 (1). https://doi.org/10.1186/s12864-018-5198-4.

Rhodes, J., et al., 2017. Tracing genetic exchange and biogeography of *Cryptococcus neoformans* var. *Grubii* at the global population level. Genetics 207 (1), 327−346. https://doi.org/10.1534/genetics.117.203836.

Rogers, M.B., et al., 2014. Genomic confirmation of hybridisation and recent inbreeding in a vector-isolated *Leishmania* population. PLoS Genet. 10 (1). https://doi.org/10.1371/journal.pgen.1004092.

Roman, F., et al., 2018. Dissecting the phyloepidemiology of *Trypanosoma cruzi* I (TcI) in Brazil by the use of high resolution genetic markers. PLoS Negl. Trop. Dis. 12 (5). https://doi.org/10.1371/journal.pntd.0006466.

Rugna, G., et al., 2018. Multilocus microsatellite typing (MLMT) reveals host-related population structure in *Leishmania infantum* from northeastern Italy. PLoS Negl. Trop. Dis. 12 (7), e0006595. https://doi.org/10.1371/journal.pntd.0006595.

Rusman, F., et al., 2020. Evidence of hybridization, mitochondrial introgression and biparental inheritance of the kdna minicircles in *Trypanosoma cruzi* I. PLoS Negl. Trop. Dis. 14 (1), 1−16. https://doi.org/10.1371/JOURNAL.PNTD.0007770.

Samarasinghe, H., Xu, J., 2018. Hybrids and hybridization in the *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes. Infect. Genet. Evol. 66, 245−255. https://doi.org/10.1016/j.meegid.2018.10.011.

Schwabl, P., et al., 2021. Colonization and genetic diversification processes of *Leishmania infantum* in the Americas. Nat. Commun. 4 (1). https://doi.org/10.1038/s42003-021-01658-5.

Schwabl, P., et al., 2019. Meiotic sex in Chagas disease parasite *Trypanosoma cruzi*. Nat. Commun. 10 (1). https://doi.org/10.1038/s41467-019-11771-z.

Serafim, T.D., et al., 2023. *Leishmania* genetic exchange is mediated by IgM natural antibodies. Nature 623, 149−156. https://doi.org/10.1038/s41586-023-06655-8.

Stoesser, N., et al., 2016. Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. mBio 7.

Tibayrenc, M., 1998. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. Int. J. Parasitol. 28, 85−104.

Tibayrenc, M., Ayala, F.J., 1987. Forte corrélation entre classification isoenzymatique et variabilité de l'ADN kinétoplastique chez *Trypanosoma cruzi*. C. R. Acad. Sci. Hebd. Seances Acad. Sci. D 304, 89−93.

Tibayrenc, M., Ayala, F.J., 1988. Isozyme variability in *Trypanosoma cruzi*, the agent of Chagas'disease: genetical, taxonomical and epidemiological significance. Evolution 42 (2), 277−292. https://doi.org/10.1111/j.1558-5646.1988.tb04132.x.

Tibayrenc, M., Ayala, F.J., 2012. Reproductive clonality of pathogens: a perspective on pathogenic viruses, bacteria, fungi, and parasitic protozoa. Proc. Natl. Acad. Sci. U. S. A. 109 (48), E3305−E3313.

Tibayrenc, M., Ayala, F.J., 2013. How clonal are *Trypanosoma* and *Leishmania*? Trends Parasitol. 29 (6), 264−269. https://doi.org/10.1016/j.pt.2013.03.007.

Tibayrenc, M., Ayala, F.J., 2014a. New insights into clonality and panmixia in *Plasmodium* and *Toxoplasma*. In: Advances in Parasitology. Academic Press, pp. 253−268. https://doi.org/10.1016/B978-0-12-800099-1.00005-3.

Tibayrenc, M., Ayala, F.J., 2014b. *Cryptosporidium, Giardia, Cryptococcus, Pneumocystis* genetic variability: cryptic biological species or clonal near-clades? PLoS Pathog. 10 (4), e1003908. https://doi.org/10.1371/journal.ppat.1003908.

Tibayrenc, M., Ayala, F.J., 2015. How clonal are *Neisseria* species? The epidemic clonality model revisited. Proc. Natl. Acad. Sci. U. S. A. 112 (29), 8909−8913. https://doi.org/10.1073/pnas.1502900112.

Tibayrenc, M., Ayala, F.J., 2017a. Is predominant clonal evolution a common evolutionary adaptation to parasitisms in parasitic protozoa, fungi, bacteria and viruses? Adv. Parasitol. 96, 243−325.

Tibayrenc, M., Ayala, F.J., 2017b. Relevant units of analysis for applied and basic research dealing with neglected transmissible diseases: the predominant clonal evolution model of pathogenic microorganisms. PLoS Negl. Trop. Dis. 11 (4), e0005293. https://doi.org/10.1371/journal.pntd.0005293.

Tibayrenc, M., Ayala, F.J., 2021a. Models in parasite and pathogen evolution: genomic analysis reveals predominant clonality and progressive evolution at all evolutionary scales in parasitic protozoa, yeasts and bacteria. Adv. Parasitol. 75−117. https://doi.org/10.1016/bs.apar.2020.12.001.

Tibayrenc, M., Ayala, F.J., 2021b. *Leishmania* and the model of predominant clonal evolution. Microorganisms 9 (11). https://doi.org/10.3390/microorganisms9112409.

Tibayrenc, M., Ayala, F.J., 2022. Microevolution and subspecific taxonomy of *Trypanosoma cruzi*. Infect. Genet. Evol. 103, 105344. https://doi.org/10.1016/j.meegid.2022.105344.

Tibayrenc, M., Cariou, M.L., Solignac, M., 1981a. Interprétation génétique des zymogrammes de flagellés des genres *Trypanosoma* and *Leishmania*. C. R. Acad. Sci. Hebd. Seances Acad. Sci. D Series III 292 (9), 623−625.

Tibayrenc, M., et al., 1981b. Arguments génétiques contre l'existence d'une sexualité actuelle chez *Trypanosoma cruzi*. Implications taxinomiques. C. R. Acad. Sci. Hebd. Seances Acad. Sci. D Series III 293 (3), 207−209.

Tibayrenc, M., Kjellberg, F., Ayala, F.J., 1990. A clonal theory of parasitic protozoa: the population structures of *Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas*, and *Trypanosoma* and their medical and taxonomical consequences. Proc. Natl. Acad. Sci. U. S. A. 87 (7), 2414−2418. https://doi.org/10.1073/pnas.87.7.2414.

Tibayrenc, M., Kjellberg, F., Ayala, F.J., 1991. The clonal theory of parasitic protozoa: a taxonomic proposal applicable to other clonal organisms. Bioscience 41 (11), 767−774. https://doi.org/10.2307/1311726.

Tibayrenc, M., et al., 1993. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proc. Natl. Acad. Sci. U. S. A 90 (4), 1335−1339. https://doi.org/10.1073/pnas.90.4.1335.

Tibayrenc, M., et al., 1986. Natural populations of *Trypanosoma cruzi*, the agent of Chagas disease, have a complex multiclonal structure. Proc. Natl. Acad. Sci. U. S. A. 83 (1), 115−119. https://doi.org/10.1073/pnas.83.1.115.

Tomasini, N., et al., 2014. Preponderant clonal evolution of *Trypanosoma cruzi* I from Argentinean Chaco revealed by multilocus sequence typing (MLST). Infect. Genet. Evol. 27, 348−354. https://doi.org/10.1016/j.meegid.2014.08.003.

Van den Broeck, F., et al., 2020. Ecological divergence and hybridization of Neotropical *Leishmania* parasites. Proc. Natl. Acad. Sci. U. S. A. 117 (40), 25159−25168. https://doi.org/10.1073/pnas.1920136117.

Vanhove, M., et al., 2017. Genomic epidemiology of *Cryptococcus* yeasts identifies adaptation to environmental niches underpinning infection across an African HIV/AIDS cohort. Mol. Ecol. 26 (7), 1991−2005. https://doi.org/10.1111/mec.13891.

Vianello, M.A., et al., 2021. International high-risk clone of fluoroquinolone-resistant *Escherichia coli* O15:H1-D-ST393 in remote communities of Brazilian Amazon. Infect. Genet. Evol. 91. https://doi.org/10.1016/j.meegid.2021.104808.

Weedall, G.D., Hall, N., 2015. Sexual reproduction and genetic exchange in parasitic protists. Parasitology 142, S120−S127. https://doi.org/10.1017/S0031182014001693.

Zingales, B., et al., 2012. The revised *Trypanosoma cruzi* subspecific nomenclature: rationale, epidemiological relevance and research applications. Infect. Genet. Evol. 12 (2), 240−253. https://doi.org/10.1016/j.meegid.2011.12.009.

Zumaya-Estrada, F.A., et al., 2012. North American import? Charting the origins of an enigmatic *Trypanosoma cruzi* domestic genotype. Parasites Vectors 5 (1). https://doi.org/10.1186/1756-3305-5-226.

CHAPTER

Parasite–Host Coevolution

6

Brenyn Kelly[1], Katie Izenour[2] and Sarah Zohdy[1,2]

[1]*College of Forestry, Wildlife, and the Environment, Auburn University, Auburn, AL, United States;* [2]*College of Veterinary Medicine, Auburn University, Auburn, AL, United States*

1. Introduction

"*People are either born hosts or born guests*" is Max Beerbohm's famous saying attributed to human hosting and hospitality, but it can be well-applied to the dynamic relationship between host and parasite species. All living organisms have parasites that are adapted to their systems for survival. In the same way hosts adapt to environmental and ecological conditions, parasites must evade predators and adapt to their environment, which in most cases, is the host. Living organisms all serve as a host to various pathogens or parasites that depend on them for survival, proliferation, and transmission. This symbiotic host–parasite relationship is a result of adaptation. Parasites can adapt to evade predators, as well as adjust to changes in the environment and changes within their preferred host (Schulte et al., 2011). The sole purpose of a parasite is to survive, and when survival is dependent on a host, the parasite must adapt to changing host conditions to ensure survival.

The dependence of parasites on their hosts for survival creates a constant cycle of adaptation, feeding into each other's evolutionary processes. Hosts adapt both to changes in their environment, as well as changes in the parasite in order to ensure the parasite does not become detrimental or lethal to the host (Ebert, 2005; Quigley et al., 2012). Additionally, parasites also adapt to changes within the host to ensure they are optimally fit to the conditions of the host to ensure survival and reproduction (Ebert, 2005). This symbiotic host–parasite relationship requiring both organisms to be "best fit" for each other is known as coevolution (Quigley et al., 2012). Although never explicitly using the word "coevolution," the theory is ascribed to Charles Darwin for his works examining orchids and predicting that a species of moth must possess a long proboscis to access the nectar (Crilly, 2021). Darwin believed that both species' unique adaptations influenced the other's survival (Crilly, 2021). Paul Ehrlich and Peter Raven were the first to use the term coevolution while studying specific butterflies and plant species (Crilly, 2021).

The degree to which a host and parasite are able to survive and thrive in their environment, thus coevolving, is mainly determined by genetics. Genetic mutations can either be beneficial or detrimental to an organism and result in a change in fitness (Brajesh et al., 2019). Beneficial mutations will improve the ability of an organism, host, or parasite to survive under new environmental pressures. Organisms with mutations that enhance their fitness and ability to survive are able to reproduce and

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00008-2

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

spawn offspring that will be similarly phenotypically similar and thus, well-suited to survive in the present environment (Imhof and Schlötterer, 2001). These genetic mutations happen gradually and result in a change in the proportion of traits in a population, changing an organism's ability to survive and favoring those with beneficial mutations that reproduce (Gregory, 2009). Over time, these small enhancements for adaptation become the dominant phenotype in the population (Clayton et al., 1999).

For the parasite to be sustained, it requires the host to survive and thrive in order to keep the host system properly functioning. If an adaptation or change in the parasite results in the death of the host, this co-dependence will no longer benefit the parasite as it will no longer have a host system in which to reside. But adaptation and evolution do not occur in a vacuum. The parasite adapts and responds to changes in the host, while the host adapts and responds to changes in the pathogen. There is a cyclical synergy in how hosts and parasites coevolve, both ensuring they are closely aligned to ensure each other's mutual survival. If the parasite cannot complete its life cycle in the host, it results in a dead-end relationship for the parasite. A parasite's goal is to survive and reproduce, thus, it needs a host that is optimally positioned to facilitate that successfully.

Parasites fall into several categories based on the structure, behavior, and environment they require. For example, ectoparasites are *on* the skin of the host (Meredith and Ulrich, 2013) and endoparasites are *in* a host (Scheifler et al., 2019). Microparasites multiply *in* a host, macroparasites multiply *outside* of the host. Microparasites include viruses, bacteria, and fungi (Anderson and May 1981) while macroparasites consist of helminths and arthropods (Møller et al., 1993).

In many scientific fields, the words disease, parasite, and pathogen are often used interchangeably to describe how an organism interacts with a host and whether the relationship is beneficial or detrimental. However, these three terms are very different and should be delineated:

A *disease* is a condition, or a host's state of being that can be caused by a parasite or a pathogen. A key concept of disease is that a disease is not transmissible or contagious. Hosts do not "catch" a disease; *pathogens* are a type of disease-causing agent that can be transmitted or passed between hosts (Balloux and van Dorp, 2017). Viruses, bacteria, fungi, and protozoa are all organisms that can be pathogens but are not always. *Parasites* are considered to be organisms that live in or on another organism, often referred to as its host, and derive some benefit from that host. Some of these parasites are pathogenic, meaning they cause disease, and some pathogens are parasitic, meaning they depend on another organism for survival. Parasites are categorized into two groups: microparasites and macroparasites. Microparasites, like viruses, bacteria, and fungi, reproduce inside their host and are dependent on that host for the completion of their lifecycle. Macroparasites, like helminths or lice, are usually larger than microparasites and rarely require a single host to complete their life cycle, typically completing some part of their life cycle outside of their host (Møller et al., 1993).

This chapter will examine the many complexities of parasite—host coevolution. We will describe the Red Queen hypothesis and the evolutionary arms race, view rapidly evolving pathogens with RNA viruses as a model, and examine viral evolution. We will also introduce the concept of quasispecies and mutant clouds with modern, real-world examples and survey phylogenetics to learn more about coevolution. Lastly, we will go over parasite—host relationships in both animal and human infectious diseases, host immunogenetics, parasite manipulation, and the concept and consequences of a parasite-free world. Understanding more about the dynamics of parasite—host coevolution is vitally important for the future, especially as we live in an interconnected world with parasites that often share our bodies.

2. Common Parasites and Their Definition/Description

2.1 Macroparasites

Definition: "macro" refers to pathogens that can be visualized with the naked eye.

Macroparasitic parasites can live in or on the host organism. Examples of internal pathogenic macroparasites include intestinal helminths like *Toxocara canis* and *Trichuris vulpi.* External microparasites like fleas, ticks, or lice may or may not be pathogens. Whether internal or external, pathogenic or not, macroparasites can be seen without assistance from the naked eye.

2.2 Microparasites

Definition: "micro-" refers to pathogens that are too small to be visualized with the naked eye (Guzik and McCoy-Simandle, 2017) and organisms that require assistance like a microscope to be visualized.

Examples of microparasites are bacteria, protozoa, and some fungi like yeast.

2.3 Endoparasites

Definition: a pathogen that lives inside the host.

Endoparasitic pathogens live inside the host organism and can be any number of organisms. Examples are intestinal parasites like hookworms and roundworms that live in the intestines of mammals. The pathogen that causes malaria is a protozoan of the genus *Plasmodium*, this pathogen is an endoparasite, living inside red blood cells.

2.4 Fungi

Definition: single or multicellular organisms, including yeasts, mushrooms, and mold.

Fungi can be endo- or ecto-parasitic. *Candida* colonization of the gastrointestinal tract is an example of a pathogenic, endoparasitic yeast infection. Ringworm is caused by a fungus and is a pathogenic ectoparasitic fungus.

2.5 Protozoa

Definition: single-celled free-living or parasitic eukaryotic microorganisms.

Pathogenic protozoa include the waterborne, endoparasitic pathogen *Giardia lamblia,* which causes gastrointestinal illness in mammals. The mosquito-borne pathogen *Plasmodium* spp. that causes malaria in humans is an intracellular endoparasitic microorganism that lives in the red blood cells.

2.6 Bacteria

Definition: unicellular microorganism with a cell wall and nucleus. Bacteria can be pathogenic but not always.

Examples of pathogenic bacteria include *Escherichia coli*, which causes gastrointestinal upset, often associated with food poisoning. In severe cases, infection with *E. coli* can cause life-threatening

hemolytic urea syndrome. *Staphylococcus aureus* is often associated with skin and soft tissue infections but can cause serious infections of the bloodstream and joints.

2.7 Ectoparasites

Definition: A pathogen that lives outside or on the exterior of its host.

Ectoparasitic pathogens live outside or on the surface of the host organism. Examples of ectoparasites are fleas, ticks, and lice. There are many others, but fleas, ticks, and lice all live on the exterior of mammals on the skin, hair, or fur.

3. Viruses

Definition: an endoparasitic microorganism that survives by entering the cells of the host and capitalizing on the functionality and metabolism of the host's cells.

Examples of viruses include the human immunodeficiency virus (HIV), which causes acquired immunodeficiency syndrome (AIDS) in humans.

4. Red Queen Hypothesis and the Evolutionary Arms Race

> Now here, you see, it takes all the running you can do to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!
>
> **The Red Queen, Lewis Carroll, Through the Looking Glass (1871)**

Coevolution may sound like a beautiful push and pull, a poetic give and take between a host and pathogen as they both work to become the most biologically fit versions of themselves, while also surviving extrinsic factors beyond their control. It is easy to overlook one critical detail of this perfectly timed and executed dance of nature—each one needs to evade the natural tendency of the other to kill or damage the other. Hosts are trying to stay healthy and pathogen-free, while parasites are determined to survive and multiply. The "Red Queen Hypothesis" is often used to describe what is essentially an arms race between host and parasite as they each evolve and adapt to changes in each other and evade any changes that may be detrimental.

The Red Queen is a character from Lewis Carroll's books *Alice's Adventures in Wonderland* (Carroll, 1865) and *Through the Looking Glass* (Carroll, 1871) who is made to resemble and represent the Queen in the game of chess. In Chapter 2 of *Through the Looking Glass* (Carroll, 1871), Alice, the main character in *Alice's Adventures in Wonderland,* and *Through the Looking Glass*, is described as running as fast as she can but remaining in the same physical location. In reference to Alice's lack of movement in spite of her frantic effort, the Red Queen remarks, "Now here, you see, it takes all the running you can do to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!" (Carroll, 1871). This quote from Alice and Wonderland was first applied to coevolution by Van Valen in 1973, who proposed "The Red Queen Hypothesis" (Strotz et al., 2018) stating a constant need for adaptation to selective pressure to allow for evolutionary changes to move forward (Papkou et al., 2019). The hypothesis was later refined to highlight the characteristic negative impacts a parasite has on its host and also when examining the survival of the host—parasite relationship where there are repeated cycles of adaptation and counteradaptation (Strotz et al., 2018). In a

sense, the faster one of them runs, the other one must run twice as fast to stay ahead of any compensatory detrimental or deleterious mutations (Chakrabarti and Panchenko, 2009).

With a rapid life history, parasites are inclined to possess faster evolution in comparison to host species (Kochin et al., 2010). These parasites must be able to adapt quickly to their host environment if they want to survive and proliferate. An identical goal exists for the host species, and they must be able to evolve with the parasite when the pressures placed on them are strong (Foitzik, 2003). The term "arms race" is traditionally used to describe competitive weapon accumulation and in accretions (Gray, 1971), such as the race for nuclear weaponry acquisition in the Cold War (McGuire, 1977), but has recently been attributed to biology, taking on a similar meaning when attributed to the coevolutionary phenomenon. The evolutionary arms race is a component of coevolution and is represented by antagonistic interactions between species, with increasing amounts of defensive mechanisms (Medel et al., 2010). Unfortunately, there are often no real winners in these never-ceasing "Red Queen" scenarios (Rabajante et al., 2015) and rely, many times, on negative frequency-dependent selection (Anzia and Rabajante, 2018). This sometimes-negative effect of the Red Queen Hypothesis can be especially prominent when viewing the coevolution of numerous RNA viruses with humans.

5. Rapidly Evolving Pathogens—RNA Viruses as a Model

When discussing adaptation and evolution, it may be easy to overlook that every genetic mutation that leads to a favorable phenotype for survival is actually an incremental change in the diversity of the organism or pathogen. Mutation rates vary greatly between pathogenic species with deleterious mutations being more common than beneficial ones (Murray et al., 2021). An increase in mutation, results in a higher diversity among viruses (Sanjuán R, Domingo-Calap, 2016). Viewing a one-hundred-year period, hosts, obligate ectoparasites, and parasites change very little from time = 0 years to time = 100 years (Fig. 6.1). If mutations occurred and the virus failed to survive or killed the host,

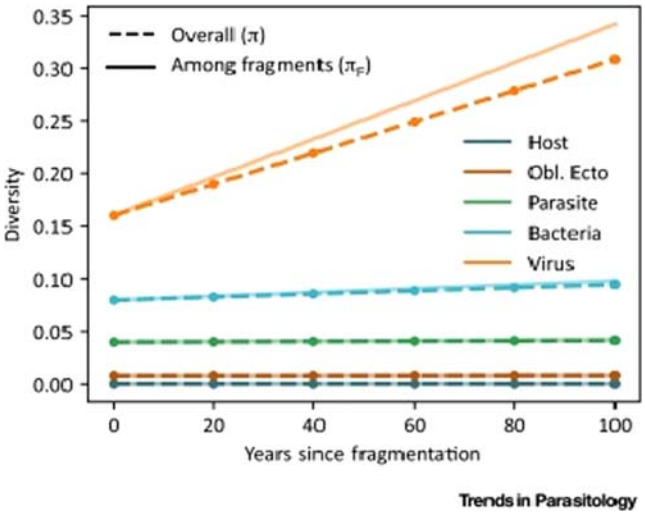


FIGURE 6.1

Overview of evolution of different pathogen types over time and diversification with habitat fragmentation.

Adapted [reprinted] from Zohdy, S.,Schwartz, T.S., Oaks, J., 2019. The Coevolution Effect as a Driver of Spillover. Trends in Parasitology 35 (6), 399–408.

there would be a negative trend seen (Fig. 6.1), but the trend line shows that over time, viruses grow more proficient at adapting and, ultimately, become more successful survivalists. There are minute changes overall and among fragments with the other organisms; however, viral pathogens demonstrate an almost exponential increase in diversity in this period. So, why do viral pathogens evolve more frequently than other organisms?

RNA viruses can have mutation rates up to one million times greater than their host, and this rapid mutation is often associated with increased virulence and survival (Duffy, 2018). And the rapid increase in diversity that viral pathogens can achieve means they have the ability to quickly adapt to change, evolving more swiftly than their host (Seal et al., 2021). For instance, viruses have been becoming more resistant to antiviral drugs that are specifically intended to stop viral replication (Irwin et al., 2016). However, if some viral genomes survive the medication and replicate, those viruses will pass on the most-fit, drug-resistant mutations (Irwin et al., 2016). Rapidly evolving pathogens may have the upper hand when it comes to their ability to mutate, evolve, and ultimately, pass on their genetics to ensure the best outcomes for future generations.

6. Where Did Viruses Come From? Viral Evolution

The cunning adaptability and survivability of viruses are not surprising when considering their origin. Viruses have been with humans since the dawn of our history, often making appearances in recorded historical texts centuries before they were accurately discovered (Taylor, 2014). Viruses are obligate, intracellular parasites consisting of genomic DNA or RNA material contained within a protein shell called a capsid (Louten, 2016). Their simple structure means they must enter the cell of another living organism, like a human, to access the cellular and metabolic machinery they need for replication (Birch et al., 2012). Since viruses cannot live or replicate outside of host cells, high levels of connection exist between humans and these organisms (Durzyńska and Goździcka-Józefiak, 2015). Like all existing things, survival is the goal. The complex dynamic process that exists between a host and virus leads to constant changes and coevolution on both sides of the host—virus dynamic. Viruses have short generation times, meaning they can adapt to changes in the environment rapidly, which can potentially introduce mutations (Stern and Andino, 2016). The cornerstones of viral evolution are:

(1) Mutation rate
(2) Population size
(3) Selection
(4) The multiplicity of infection

When viewing mutation rates through the lens of viral evolution, mutation rates are the rate at which errors or mutations occur during viral genomic replication and are often used to determine genetic diversity (Peck and Lauring, 2018). High mutation rates are often associated with high evolutionary rates and control the amount of genetic variation in populations (Peck and Lauring, 2018). Another component of viral evolution is population size. Population size is the size of a population with no generations that overlap in an area (Rousseau et al., 2017). RNA viruses possess large population sizes, but these can change due to bottlenecks and coevolution between hosts and viruses drive alterations in population size (Moya et al., 2004). Similarly to population size, selection also plays a key role in coevolution. Selection is a significant driver of viral evolution (Kustin and

Stern, 2021) and functions as a constraint on deleterious mutations that negatively impact the population (Dolan et al., 2018). Lastly, the multiplicity of infection is considered to be the number of genomes of a virus that are entering and replicating in host cells (Gutiérrez et al., 2010, 2015). The multiplicity of infection can largely affect evolutionary outcomes, especially when examining this factor in the context of viruses (Wodarz et al., 2018).

7. How Do Viruses Evolve and Selection Pressures Not Just on Individual Pathogens With Mutations, but on Groups of Pathogens? Mutant Clouds/Quasispecies

Viruses are a great model of the rapid rate of evolution to view how environmental and host factors can influence change. When viruses evolve and mutate, quasispecies form composed of mutant clouds that often hold abundant genetic heterogeneity (Nowak, 1992; Rachmadi et al., 2018; Domingo and Perales, 2016). Mutant clouds, often called mutant swarms, or mutant spectra (Domingo et al., 2012), are RNA viruses that are an aggregation of genetically diverse, quickly replicating communities (Agarwal et al., 2018). Mutant clouds make up a plethora of closely intertwined, dynamic variants (Domingo et al., 2012). The quasispecies that form from these mutant spectra are typically dependent on the population size of the virus and possess high mutation rates, as well as fluctuations in frequency (Domingo and Holland, 1997; Domingo and Perales, 2016). A great illustration of quasispecies in today's world is the SARS-CoV-2 virus which has released a number of rapidly evolving variants across countries.

The COVID-19 pandemic is an excellent example of the adaptability of a virus. The world watched as mutations continually arose, guidelines about mask-wearing in public spaces shifted, recommendations to get vaccinated or boosted were released, and alterations to normal everyday activities such as students attending class in person precipitated. The fact is, the SARS-CoV-2 virus that caused the COVID-19 pandemic did extensively mutate (Fig. 6.2, Nextstrain Team), and this pandemic is a testament to how new variants caused by viral evolution influenced health on a global scale.

Genomes and genetic mutations are inherently difficult to visualize or see. Computers are used to analyze genetic sequence data, and it is represented in phylogenetic trees (Young and Gillung, 2019), with each circle on the phylogenetic tree showing an individual report of a specific sequence. The lines or branches work like a family tree and show the generations and relatedness of each sequence and their mutations (de Queiroz, 2013). This figure shows that some of the early mutations, far left in the image, continued circulating throughout the entire pandemic. Over time, strains of COVID-19 quickly shifted in dominance in response to changes in human behavior, such as mask-wearing and the introduction of the vaccine (Konishi, 2022; Malik et al., 2022). Remember, pathogens and parasites are in a race for survival and that survival depends on their ability to respond to pressure, like a vaccine, and evade that pressure. As we, the human hosts, are evolving ways of opposition to this virus, SARS-CoV-2 is evolving with us. A new cause for concern is the risk of SARS-CoV-2's vaccine resistance due to resistant mutations, as we have seen with many viruses that have evolved in the past (Vitiello, 2021). This figure's data is based on specimens that are sequenced and reported, it is possible that there are mutations that haven't been reported. This phylogenetic tree of SARS-CoV-2 over a 2.5-year period of time depicts extensive mutations.

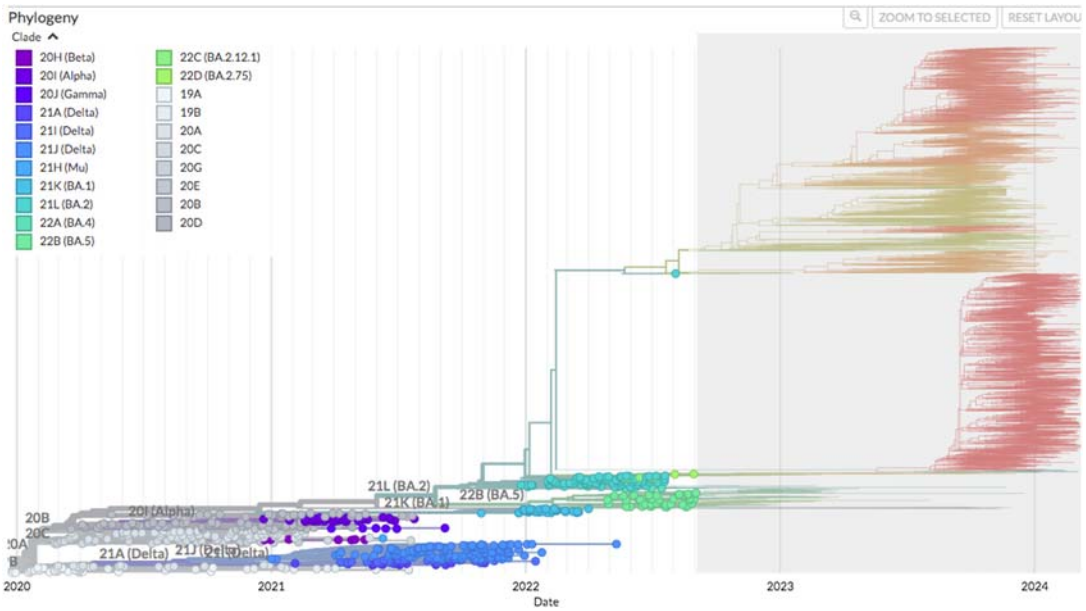


FIGURE 6.2

This phylogenetic tree shows the clades and reported mutations in SARS-CoV-2 since the start of the 2019 COVID-19 pandemic through September 2022; each dot represents one reported sequence from a specimen anywhere in the world.

Adapted [reprinted] from Nextstrain Team, 2022. "Genomic epidemiology of SARS-CoV-2 with subsampling focused globally over the past 6 months" is licensed by CC-BY 4.0. Auspice, https://nextstrain.org/ncov/gisaid/global/6m. (Accessed 16 December, 2022).

8. Phylogenetics Reveals Coevolutionary Patterns

Phylogenetics is used to visualize the relationship of genetic sequences to one another. Traditionally, the genetic sequences of a host and parasite would be analyzed separately from each other, but this method treats the host genome as one entity in isolation from the parasitic genome, completely overlooking the interconnectedness of the host and parasite through coevolution. Light and Hafner (2007) examined the cophylogeny and coevolution of chewing lice, an obligate ectoparasite, with their host, pocket gophers. This study tests the hypothesis that closely related lineages of parasites will respond similarly to the same host. This study collected chewing lice from the pelage of pocket gophers and performed molecular sequencing on the collected lice and the gophers. The findings of this study show that two clades of lice recovered from the pocket gophers developed mutations specific to the pocket gopher host that are distinct from mutations found in lice, not on pocket gophers.

Similarly, specific lice have also been found to be specific to certain birds, providing the basis for Harrison's rule, which states that host body size is linked to parasite size (Villa et al., 2018). More specifically, it applies to large-bodied hosts possessing large-bodied parasites specific to size (Harnos

et al., 2017). When examining the phylogenies of these species, a connection between host and parasite body size coevolution is shown (Harnos et al., 2017). Villa et al. (2019) examined this idea further by researching the evolutionary relationship between feather-feeding lice and pigeons. The results found that when lice are removed from their host pigeon and transferred to a pigeon of a different size, they can evolve their size in 4 years—roughly 60 of their generations—to correlate with host size (Villa et al., 2019). Although feather lice, like many other parasites, are somewhat host-specific and often possess phylogenies fairly consistent with their host species, they can exhibit host-switching to a degree (de Moya et al., 2019; Villa et al., 2019). For example, when winged hippoboscid flies disperse feather-feeding lice to new hosts other than their host pigeons, they are able to switch and proliferate more competently than body lice from the same pigeon (Harbison and Clayton, 2011, Fig. 6.3), thus, showing how phylogenies of body lice are more identical to pigeons than feather-feeding lice. These incredible forms of adaptation by lice, along with their phylogenetic patterns, reveal how parasites have coevolved with their host over time.

9. Parasite–Host Relationships in Human Infectious Diseases

Throughout time, parasites and humans have had an extremely complex, dynamic relationship. Parasites are one of the most prevalent organisms on earth and infect more than one billion humans worldwide (Naveed and Abdullah, 2021). Parasites' wide range, infectious nature, reliance on humans,

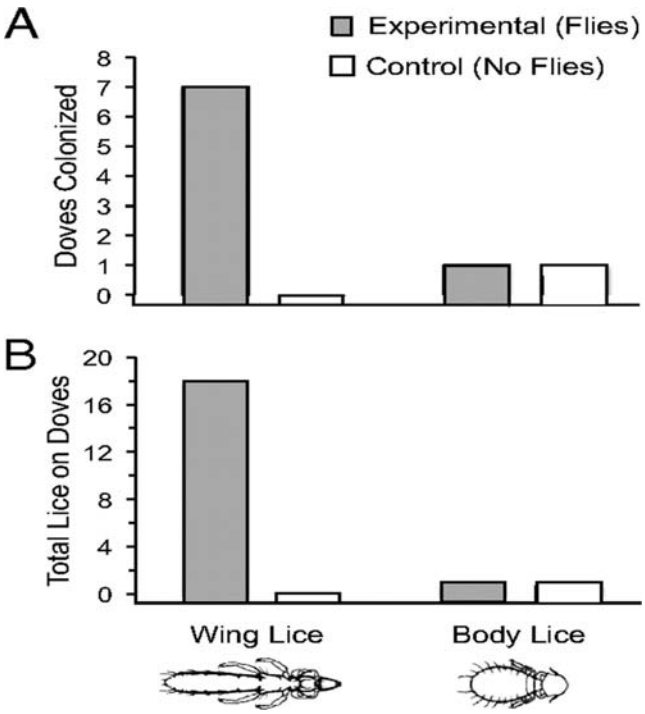


FIGURE 6.3

Comparison of host-switching between feather-feeding and body lice.

Adapted [reprinted] from Harbison, C.W., Clayton, D.H., 2011. Community interactions govern host-switching with implications for host–parasite coevolutionary history. PNAS 108 (23), 9525–9529.

and proclivity for causing diseases have led to a closely intertwined, yet adverse association with humans. Since parasites and hosts evolve in response to each other to increase their own fitness, they often form a coevolutionary feedback loop (Buckingham and Ashby, 2022). A significant problem that has become apparent over recent decades is emerging infectious diseases (EIDs), with humans becoming drivers of the spread of these parasitic species due to global travel (Rogalski et al., 2017). Habitat fragmentation is also shown to be correlated with EIDs due to zoonotic pathogen spillover (Zohdy et al., 2019). In addition, areas with a tightly packed host population increase the likelihood of parasite transmission and infection among individuals (Rogalski et al., 2017). With parasites being an ever-present threat in our lives, understanding more about their role in human infectious diseases is critical.

Perhaps one of the best examples of parasite—human infectious disease coevolution is malaria, which affects over 200 million people annually (Talapko et al., 2019) and exhibits some of the most intriguing instances in nature of coevolving species working against each other. Malaria is caused by *Plasmodium* spp. and is transmitted by the *Anopheles* mosquito (Deroost et al., 2016) to a human host. Host immunity is one of the biggest factors that drive the evolution of parasites, while humans must continually evolve to avoid illness (Su et al., 2020). Although the human immune system is well-equipped to deal with infections, *Plasmodium* has evolved ways of circumventing a host's immune system responses to avoid detection (Deroost et al., 2016). While inside its host, this parasite will avoid antibody-mediated clearance by altering cell-surface proteins through a process called antigenic variation (Siao et al., 2020; Deroost et al., 2016) this manner of immune evasion and deception successfully allows *Plasmodium* to proliferate and spread. Another method of this parasite evolving against its host is seen during human infections, when *Plasmodium* increases skin attractiveness to *Anopheles* vectors, thereby resulting in a greater probability of malaria transmission (Robinson et al., 2018). Individuals infected with this parasite produce more types of specific aldehydes attractive to mosquito vectors (Robinson et al., 2018). *Plasmodium* has evolved ways to increase its chance of transmission in an age of modern antimalarial medicine which has resulted in decreased malaria case numbers (Talapko et al., 2019); however, the issue of antimalarial drug resistance has recently emerged. The introduction of antimalarial drugs placed a pressure on *Plasmodium* not seen before, causing a reduction in cases (Su et al., 2020). Despite this, a number of drug-resistant parasites survived and are passing their mutation on to future generations (Siao et al., 2020; Su et al., 2020). These ways of parasitic evolution to increase fitness are troublesome for human populations plagued with these infectious diseases.

10. Parasite—Host Relationships in Animal Infectious Diseases

Like the human—parasite relationship, the animal—parasite relationship is equally complex and dynamic. Parasites are a pervasive threat in nature, and many animals develop avoidance strategies for parasites (Hart and Hart, 2018). Although parasites can have functional roles in their animal host, they also cause a wide variety of deadly, infectious diseases (Poulin, 1999). Brucellosis is caused by the microparasite, the gram-negative bacteria, *Bartonella* spp. (Portillo et al., 2020). *Bartonella* has been found to coevolve with its hosts, especially bats and rodents (Lei and Olival, 2014). Bats and rodents consist of the two largest percentages of mammals (Williams et al., 2021) and are reservoirs for many emerging infectious diseases (EIDs) (Lei and Olival, 2014). A study by Lei and Olival (2014) found

that *Bartonella* significantly coevolved with both bats and rodents. One reason for this species-specific coevolution entails the presence of a vector in brucellosis disease transmission (Lei and Olival, 2014). Since *Bartonella* is transmitted through a vector (Portillo et al., 2020), there is an extra evolutionary step to drive speciation that involves the adaptation of the arthropod vectors (Lei and Olival, 2014).

Another example of parasite–host relationship coevolution is *Dirofilaria immitis,* the parasitic nematode causing deadly heartworm disease, and its resistance to heartworm drugs. Like in human infections with *Plasmodium,* host species must fight against parasite infections, and one way to accomplish this is to increase host immunity with medicine and drugs. Although medicine for heartworm has been largely beneficial, *Dirofilaria immitis* has begun to evolve resistance to the macrocyclic lactones (Bowman, 2012). The issue arises when the prevention of heartworm by the use of medicine no longer works. With the rise of heartworm cases across the country, *Dirofilaria immitis*'s evolved resistance to increase fitness is a significant concern (Drake and Wiseman, 2018). When examining the evolved mechanisms of heartworm's initial host, the mosquito, this arthropod vector has evolved ways of controlling parasitic burden. *Dirofilaria immitis* requires a mosquito vector for the early stages of development and transmission to a canine host (Noack et al., 2021). The mosquito has biological and physical mechanisms that can offer a more resistant state to infections (Bartholomay, 2014). Mosquitoes can impede the development and hinder the transmission of *Dirofilaria immitis* through blood coagulation in the digestive tract (Bartholomay, 2014). This evolved means of resistance to infection inhibits the transmissibility of heartworm disease to dogs. In response, *Dirofilaria immitis* has adapted; the parasites can manipulate elements of the mosquito host's immune system that typically function in parasite recognition and response (Bartholomay, 2014). The parasite and hosts have all coevolved ways to increase their own fitness through a variety of different methods.

11. Host Immunogenetics Adaptations

In order to counteract the negative effects of parasites, host species need to mate to optimize immunity for their offspring. Many secondary sex traits and mating strategies have evolved to combat parasitism (Ashby and Boots, 2015). Host species have adopted a more choosy mating strategy in order to ascertain the health of potential mates, including investing in those uninfected with parasites with better immune systems (Ashby and Boots, 2015). More specifically, the insurance of genetic diversity in a mate is vitally important for resistance to parasites and pathogens, specifically when viewing the major histocompatibility complex (MHC) (Sommer, 2005). MHC possessing a high diversity is thought to be because of the selection of quickly evolving pathogens (Phillips et al., 2018). Variation in MHC is essential for pathogen and parasite resistance, and this type of genetic diversity should be favored (Eizaguirre et al., 2012; Sommer, 2005). Hosts adapting mating strategies to ensure genetic and immune diversity is paramount for survival due to the threat of virulent parasite evolution.

When viewing the relationships between host immunogenetics and parasites, bananaquit bird and avian malaria parasites are prime examples of a host species adapting genetic responses to a parasitic threat. The bananaquit is a species of bird that can harbor parasites that cause harmful avian malaria (Antonides et al., 2019). The study by Antonides et al. (2019) found that certain bananaquit individuals possess specific alleles that confer resistance to the malaria parasites. With these certain birds having genetic resilience to the parasites that cause avian malaria, these genes may be passed to future generations in order to counteract this disease. But although hosts have adapted and passed along

immune responses to counter infections, parasites have also evolved in response. When viewing *Dirofilaria immitis,* the parasite that causes heartworm, it has been found that this parasite can retaliate in reaction to their mosquito host's adapted immune defenses. These parasites can camouflage themselves to avoid recognition in the mosquito's system, instigate immune system suppression, and migrate to mosquito host tissues for seclusion (Bartholomay, 2014; Castillo et al., 2011). Host immunogenetic adaptations show the versatility and flexibility of our biological systems to avoid pathogenic and parasitic threats; however, due to our direct relationship with parasites and the premise of coevolution, this selection placed on parasites causes parasites to mount their own adaptations in response due to selection pressures.

12. Parasite Manipulation of Host Behavior

Aspects of parasite and host relationships have been researched for decades, but perhaps none are more perplexing than the manipulation of host behavior by a parasite. Many parasites have an unparalleled ability to alter host phenotypes in order to increase their own fitness (Poulin, 2013). Here, the infected host is often placed at a disadvantage as they begin to behave in ways that enhance the transmission of the parasite while simultaneously reducing their own fitness (Poulin, 2013; Hafer-Hahmann, 2019). There is no "one size fits all" method of host behavioral modification by parasites; each parasite has its own unique ability to modify host behavior that is designed specifically to stimulate proliferation and transmission for its own species. They may occupy the nervous and muscular systems (Lafferty and Shaw, 2013), alter activity levels (Hafer, 2016), adjust host appearance (Weinersmith and Faulkes, 2014), enhance sexual behavior (Heil, 2016), or reduce immunological responses to biological threats (Sánchez and Biron, 2019), which can magnify the probability of the risk of predator contact and affect the ability of a host to respond to predators (Lafferty and Shaw, 2013). The mechanisms of host manipulation express the same message by all parasitic species: survival at all costs.

Evolutionary pressures play a significant role in both the host and the parasite, but these may be variable (Øverli and Johansen, 2019). If a host is manipulated in a manner that is exclusively beneficial for the parasite, then that host may evolve greater biological responses in comparison to a host that is not placed at such a great disadvantage (Øverli and Johansen, 2019). Host systems have to strike a balance between controlling parasite loads without compromising their own health in the process by instigating potentially deadly immunological reactions (Graham et al., 2005; Øverli and Johansen, 2019). Manipulation of a host system through behavior modification or "body hijacking" has evolved through all lineages of parasites but can present quite differently (Heil, 2016; Poulin, 2013). One theory to explain the change in host behaviors is the Adaptive Host Manipulation hypothesis, which states that parasites evolve the ability to control host phenotypes and behavior to enhance the rate of contact among hosts, increasing their transmission (Heil, 2016; Hernandez-Caballero et al., 2022). What has largely been referred to as the "fatal attraction phenomenon" among species infected with the protozoan *Toxoplasma gondii* has strange, but fascinating results (Flegr et al., 2011; Heil, 2016). Rats infected with this parasite seemingly ignore all of their natural instincts to avoid the urine of their known predator, the cat, and instead become sexually attracted to feline odor (Dass et al., 2011; Flegr et al., 2011; Heil, 2016). This form of behavioral manipulation by this parasite has evolved because

Toxoplasma gondii reproduces in the gut of felines (Dass et al., 2011); since this bizarre change in behavior enhances attractiveness to felines, the "fatal" consequence often results from increased contact with their predators, which enhances *Toxoplasma gondii*'s transmission (Faulkner, 2007; Dass et al., 2011). But *Toxoplasma gondii* is not the only parasite that begets a deadly end for its host. On the surface, it appears that the parasitic hairworm *Spinochordodes tellinii* causes its cricket host to commit suicide (Biron et al., 2005; Øverli and Johansen, 2019) but there is more to this strange occurrence than meets the eye. This parasite, which develops inside the cricket, alters their response to aquatic environments, making the insect more likely to jump into water (Sanchez et al., 2008; Biron et al., 2005), which is costly for the cricket. This aberrant conduct caused by *Spinochordodes tellinii* allows the hairworm to return to water and complete its reproductive stage in search of a sexual partner (Heil, 2016; Thomas et al., 2002). These forms of behavioral manipulation yield two questions, (1) Who is *actually* in control?, and (2) How much of our behavior is really ours?

Although parasites and their hosts have a dynamic, ever-evolving give-and-take, the outcomes of these connections can often seem self-reinforcing (Hawley and Ezenwa, 2022). A bidirectional link emerges between parasite infections and host behavior (Hawley et al., 2021) where parasite traits affect host responses with consequences that select for the same type of parasite trait or vice versa (Hawley and Ezenwa, 2022); and when examining numerous species in nature, we see this become evident. For instance, aggressiveness in animals can result in an increased risk of contracting parasites, and those parasites may result in that same social behavior that is beneficial for transmission (Meyer et al., 2022; Hawley and Ezenwa, 2022; Zohdy et al., 2017). Aggressive lemurs have been found to exhibit increased parasitism and spread their parasites more frequently than docile individuals (Zohdy et al., 2017). A similar pattern was found in gray wolf populations infected with *Toxoplasma gondii*: seropositive wolves were more likely to increase risky behavior and aggression and become pack leaders (Meyer et al., 2022). This change in behavior causes the parasite-infected pack leader to lead uninfected wolves to cougars that are also infected with *Toxoplasma gondii*, thereby increasing transmission due to a parasitized host's behavior (Meyer et al., 2022). Parasites' ability to seemingly take over an organism's body to increase their own fitness by altering behavior and phenotype is unrivaled in the animal kingdom.

13. A Parasite-absent World

> It is derogatory that the Creator of countless systems of worlds should have created each of the myriads of creeping parasites and slimy worms which have swarmed each day of life on land and water (on) [this] one globe.
>
> **Charles Darwin, The Foundations of The Origin of Species (1842).**

At first glance, many of us would most likely agree with Darwin's wary view of parasites. Although a world without parasites may sound appealing at first, the dangers could be profound for the ecosystem (Zohdy et al., 2017). Studies have found that parasites serve a critical role in immunoregulating host populations, with a greater number of parasite species acting as a barrier against more virulent pathogens (Johnson et al., 2013; Carlson et al., 2020). In recent years, a parasite conservation

movement has emerged with active conservation strategies being offered, including the establishment of parasite refugia (Strona, 2015). The parasite preservation groups raise concerning questions: what would happen in a parasite-free world, and would this parasite-absent world have any host population heterogeneity?

Host heterogeneity is considered to be the commingling of different host types and the ability of these types to exchange and transmit parasites and pathogens (VanderWaal and Ezenwa, 2016). Heterogeneity among hosts can differ in susceptibility to infections and infectiousness, making up a large component of epidemics (Yates et al., 2006), and parasite diversity (Cattadori et al., 2008). Studies have found that parasitic infections that are mediated by host immune reactions are significant in continuing host heterogeneity (Cattadori et al., 2008). Not only do parasites affect heterogeneity but heterogeneity also impacts parasites. Across populations, heterogeneity in infection load, transmission, and resistance leads to variations among parasites for their speed, fitness, and number (Stephenson et al., 2017).

Despite the fact that a world without pesky parasites sounds like a reason to celebrate, the absence of these organisms would actually have repercussions on human health. On a smaller scale, let's examine the more current endangerment of human public lice due to "manscaping" and waxing. To pubic lice, the host hair is its entire world, and it spends its life on the hair shaft and reproducing (Sangaré et al., 2016; Anderson and Chaney, 2009). Because of the more modern emergence of "manscaping" and waxing (Prabhune et al., 2019), pubic lice may be becoming eradicated due to their natural habitat being taken away (Dholakia et al., 2014). Public lice may be an annoying problem, but they serve an unexpected purpose: conditioning the immune system (Jackson et al., 2009). Through feeding, lice actually initiate the immune system, and populations that are free of lice may be at a higher risk of immunopathology (Jackson et al., 2009). These immune responses are vital for preserving host heterogeneity (Cattadori et al., 2008). Even though a life free of parasites sounds ideal, a new plethora of problems impacting human health, environments, and diseases would emerge in its place.

14. Conclusions

Parasites are complex organisms with predators, mates, and unique habitats that have shaped the world around us, arguably more than hosts have. Gaining a further understanding of how parasites operate in the context of hosts, and hosts in the context of parasites is essential to gaining insight into biological systems. Hosts and parasites bandy back and forth with their adaptations, continually trying to gain the upper hand for survival. As evident by the evolutionary arms race, both host and parasite will continually coevolve with the other to increase survival. Hosts have evolved mechanisms to enhance their fitness, such as immunogenetic adaptations, mating strategies, drugs, and vaccines to counteract the injurious effects of parasitic infections. Although infinitesimal, parasites and pathogens have coevolved with their hosts through a variety of different adaptations. Some of these biological characteristics such as high mutation rates and short generation times. In response to host resistance and evolution, parasites have evolved the ability to manipulate host behavior and have introduced a new, concerning problem of drug resistance. Whether we like it or not, parasites are co-inhabitants of practically every organism on the planet, and although a universe without parasites sounds picture-perfect, we may find we rely on them just as much as they rely on us.

References

Agarwal, S., Baccam, P., Aggarwal, R., Veerapu, N.S., 2018. Novel synthesis and phenotypic analysis of mutant clouds for hepatitis E virus genotype 1. J. Virol. 92 (4). https://doi.org/10.1128/jvi.01932-17.

Anderson, A., Chaney, E., 2009. Pubic Lice (*Pthirus pubis*): history, biology and treatment vs. knowledge and beliefs of US college students. Int. J. Environ. Res. Publ. Health 6 (2), 592−600. https://doi.org/10.3390/ijerph6020592.

Anderson, R.M., May, R., 1981. The population dynamics of microparasites and their invertebrate hosts. Philos. Trans. R. Soc. Lond. B Biol. Sci. 291 (1054), 451−524. https://doi.org/10.1098/rstb.1981.0005.

Antonides, J., Mathur, S., Sundaram, M., Ricklefs, R., DeWoody, J.A., 2019. Immunogenetic response of the bananaquit in the face of malarial parasites. BMC Evol. Biol. 19 (1), 107. https://doi.org/10.1186/s12862-019-1435-y.

Anzia, E.L., Rabajante, J.F., 2018. Antibiotic-driven escape of host in a parasite- induced Red Queen dynamics. R. Soc. Open Sci. 5 (9), 180693. https://doi.org/10.1098/rsos.180693.

Ashby, B., Boots, M., 2015. Coevolution of parasite virulence and host mating strategies. Proc. Natl. Acad. Sci. USA 112 (43), 13290−13295. https://doi.org/10.1073/pnas.1508397112.

Balloux, F., van Dorp, L., 2017. Q&A: what are pathogens, and what have they done to and for us? BMC Biol. 15 (1), 91. https://doi.org/10.1186/s12915-017-0433-z.

Bartholomay, L.C., 2014. Infection barriers and responses in mosquito−filarial worm interactions. Curr. Opin. Insect Sci. 3, 37−42. https://doi.org/10.1016/j.cois.2014.08.006.

Birch, E.W., Ruggero, N.A., Covert, M.W., 2012. Determining host metabolic limitations on viral replication via integrated modeling and experimental perturbation. PLoS Comput. Biol. 8 (10), e1002746. https://doi.org/10.1371/journal.pcbi.1002746.

Biron, D.G., Marché, L., Ponton, F., Loxdale, H.D., Galéotti, N., Renault, L., Joly, C., Thomas, F., 2005. Behavioural manipulation in a grasshopper harbouring hairworm: a proteomics approach. Proc. Biol. Sci. 272 (1577), 2117−2126. https://doi.org/10.1098/rspb.2005.3213.

Bowman, D.D., 2012. Heartworms, macrocyclic lactones, and the specter of resistance to prevention in the United States. Parasites Vectors 5, 138. https://doi.org/10.1186/1756-3305-5-138.

Brajesh, R.G., Dutta, D., Saini, S., 2019. Distribution of fitness effects of mutations obtained from a simple genetic regulatory network model. Sci. Rep. 9, 9842. https://doi.org/10.1038/s41598-019-46401-7.

Buckingham, L.J., Ashby, B., 2022. Coevolutionary theory of hosts and parasites. J. Evol. Biol. 35 (2), 205−224. https://doi.org/10.1111/jeb.13981.

Carlson, C.J., Hopkins, S., Bell, K.C., Dona, J., Godfrey, S.S., Kwak, M.L., Lafferty, K.D., Moir, M.L., Speer, K.A., Strona, G., Wood, C.L., 2020. A global parasite conservation plan. Biol. Conserv. 250. https://doi.org/10.1016/j.biocon.2020.108596.

Carroll, L., 1865. Alice's Adventures in Wonderland. Macmillan & Co.

Carroll, L., 1871. Through the Looking-Glass. Macmillan & Co.

Castillo, J.C., Reynolds, S.E., Eleftherianos, I., 2011. Insect immune responses to nematode parasites. Trends Parasitol. 27 (12), 537−547. https://doi.org/10.1016/j.pt..2011.09.001.

Cattadori, I.M., Boag, B., Hudson, P.J., 2008. Parasite co-infection and interaction as drivers of host heterogeneity. Int. J. Parasitol. 38 (3−4), 371−380. https://doi.org/10.1016/j.ijpara.2007.08.004.

Chakrabarti, S., Panchenko, A.R., 2009. Coevolution in defining the functional specificity. Proteins: Struct., Funct., Bioinf. 75 (1), 231−240. https://doi.org/10.1002/prot.22239.

Clayton, D.H., Lee, P.L.M., Tompkins, D.M., Brodie III., E.D., 1999. Reciprocal natural selection on host-parasite phenotypes. Am. Nat. 154 (3), 261−270. https://doi.org/10.1086/303237.

Crilly, N., 2021. The evolution of "co-evolution" (Part II): the biological analogy, different kinds of co-evolution, and proposals for conceptual expansion. She Ji J. Design Econ. Innov. 7 (3), 333−355. https://doi.org/10.1016/j.sheji.2021.07.004.

Darwin, C., 1842. The Foundations of the Origin of Species. Essay.

Dass, S.A.H., Vasudevan, A., Dutta, D., Soh, L.J.T., Sapolsky, R.M., Vyas, A., 2011. Protozoan parasite *Toxoplasma gondii* manipulates mate choice in rats by enhancing attractiveness of males. PLoS One 6 (11), e27229. https://doi.org/10.1371/journal.pone.0027229.

de Moya, R.S., Allen, J.M., Sweet, A.D., Walden, K.K.O., Palma, R.L., Smith, V.S., Cameron, S.L., Valim, M.P., Galloway, T.D., Weckstein, J.D., Johnson, K.P., 2019. Extensive host-switching of avian feather lice following the Cretaceous-Paleogene mass extinction event. Commun. Biol. 2, 445. https://doi.org/10.1038/s42003-019-0689-7.

de Queiroz, K., 2013. Nodes, branches, and phylogenetic definitions. Syst. Biol. 62 (4), 625−632. https://doi.org/10.1093/sysbio/syt027.

Deroost, K., Pham, T.-T., Opdenakker, G., Van den Steen, P.E., 2016. The immunological balance between host and parasite in malaria. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Rev. 40 (2), 208−257. https://doi.org/10.1093/femsre/fuv046.

Dholakia, S., Shamik, Buckler, J., Jeans, J.P., Pillai, A., Eagles, N., Dholakia, S., 2014. Pubic lice: an endangered species? Sex. Transm. Dis. 41 (6), 388−391. https://doi.org/10.1097/QLQ0000000000000142.

Dolan, P.T., Whitfield, Z.J., Andino, R., 2018. Mapping the evolutionary potential of RNA viruses. Cell Host Microbe 23 (4), 435−446. https://doi.org/10.1016/j.chom.2018.03.012.

Domingo, E., Holland, J.J., 1997. RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. 51 (1), 151−178. https://doi.org/10.1146/annurev.micro.51.1.151.

Domingo, E., Perales, C., 2016. Species concepts: viral quasispecies. In: Encyclopedia of Evolutionary Biology. Elsevier.

Domingo, E., Sheldon, J., Perales, C., 2012. Viral quasispecies evolution. Microbiol. Mol. Biol. Rev. 76 (2), 159−216. https://doi.org/10.1128/MMBR.05023-11.

Drake, J., Wiseman, S., 2018. Increasing incidence of *Dirofilaria immitis* in dogs in USA with focus on the Southeast Region 2013−2016. Parasites Vectors 11, 39. https://doi.org/10.1186/s13071-018-2631-0.

Duffy, S., 2018. Why are RNA virus mutation rates so damn high? PLoS Biol. 16 (8), e3000003. https://doi.org/10.1371/journal.pbio.3000003.

Durzyńska, J., Goździcka-Józefiak, A., 2015. Viruses and cells intertwined since the dawn of evolution. Virol. J. 12, 169. https://doi.org/10.1186/s12985-015-0400-7.

Ebert, D., 2005. Host adaptations against the costs of parasitism. In: Ecology, Epidemiology, and Evolution of Parasitism in *Daphnia*. National Center for Biotechnology Information (US), Bethesda, Maryland.

Eizaguirre, C., Lenz, T.L., Kalbe, M., Milinski, M., 2012. Rapid and adaptive evolution of MHC genes under parasite selection in experimental vertebrate populations. Nat. Commun. 3. https://doi.org/10.1038/ncomms1632.

Faulkner, E., 2007. Fatal attraction. Nat. Rev. Microbiol. 5, 396. https://doi.org/10.1038/nrmicro1687.

Flegr, J., Lenochová, P., Hodný, Z., Vondrová, M., 2011. Fatal attraction phenomenon in humans − cat odour attractiveness increased for Toxoplasma-infected men while decreased for infected women. PLoS Neglected Trop. Dis. 5 (11), e1389. https://doi.org/10.1371/journal.pntd.0001389.

Foitzik, S., 2003. Arms races between social parasites and their hosts: geographic patterns of manipulation and resistance. Behav. Ecol. 14 (1), 80−88. https://doi.org/10.1093/beheco/14.1.80.

Graham, A.L., Allen, J.E., Read, A.F., 2005. Evolutionary causes and consequences of immunopathology. Annu. Rev. Ecol. Evol. Syst. 36, 373−397. https://doi.org/10.1146/annurev.ecolsys.36.102003.152622.

Gray, C.S., 1971. The arms race phenomenon. World Polit. 24 (1), 39−79. https://doi.org/10.2307/2009706.

Gregory, T.R., 2009. Understanding natural selection: essential concepts and common misconceptions. Evol. Educ. Outr. 2 (2), 156−175. https://doi.org/10.1007/s12052-009-0128-1.

Gutiérrez, S., Yvon, M., Thébaud, G., Monsion, B., Michalakis, Y., Blanc, S., 2010. Dynamics of the multiplicity of cellular infection in a plant virus. PLoS Pathog. 6 (9), e1001113. https://doi.org/10.1371/journal.ppat.1001113.

Gutiérrez, S., Pirolles, E., Yvon, M., Baecker, V., Michalakis, Y., Blanc, S., 2015. The multiplicity of cellular infection changes depending on the route of cell infection in a plant virus. J. Virol. 89 (18), 9665−9675. https://doi.org/10.1128/JVI.00537-15.

Guzik, H., McCoy-Simandle, K., 2017. Modeling microorganism transmission with Madagascar hissing cockroaches: an inquiry activity. Am. Biol. Teach. 79 (5), 365−374. https://doi.org/10.1525/abt.2017.79.5.365.

Hafer, N., 2016. Conflicts over host manipulation between different parasites and pathogens: investigating the ecological and medical consequences. Bioessays 38 (10), 1027−1037. https://doi.org/10.1002/bies.201600060.

Hafer-Hahmann, N., 2019. Experimental evolution of parasitic host manipulation. Proc. Biol. Sci. 286 (1895), 20182413. https://doi.org/10.1098/rspb.2018.2413.

Harbison, C.W., Clayton, D.H., 2011. Community interactions govern host-switching with implications for host-parasite coevolutionary history. Proc. Natl. Acad. Sci. USA 108 (23), 9525−9529. https://doi.org/10.1073/pnas.1102129108.

Harnos, A., Lang, Z., Petrás, D., Bush, S.E., Szabó, K., Rózsa, L., 2017. Size matters for lice on birds: coevolutionary allometry of host and parasite body size. Evolution 71 (2), 421−431. https://doi.org/10.1111/evo.13147.

Hart, B.L., Hart, L.A., 2018. How mammals stay healthy in nature: the evolution of behaviours to avoid parasites and pathogens. Phil. Trans. Biol. Sci. 373 (1751), 20170205. https://doi.org/10.1098/rstb.2017.0205.

Hawley, D.M., Ezenwa, V.O., 2022. Parasites, host behavior, and their feedbacks. In: Animal Behavior and Parasitism. Oxford University Press, Oxford, England, pp. 15−32.

Hawley, D.M., Gibson, A.K., Townsend, A.K., Craft, M.E., Stephenson, J.F., 2021. Bidirectional interactions between host social behaviour and parasites arise through ecological and evolutionary processes. Parasitology 148 (3), 274−288. https://doi.org/10.1017/S0031182020002048.

Heil, M., 2016. Host manipulation by parasites: cases, patterns, and remaining doubts. Front. Ecol. Evol. 4. https://doi.org/10.3389/fevo.2016.00080.

Hernandez-Caballero, I., Garcia-Longoria, L., Gomez-Mestre, I., Marzal, A., 2022. The adaptive host manipulation hypothesis: parasites modify the behaviour, morphology, and physiology of amphibians. Diversity 14 (9), 739. https://doi.org/10.3390/d14090739.

Imhof, M., Schlotterer, C., 2001. Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. Proc. Natl. Acad. Sci. USA 98 (3), 1113−1117. https://doi.org/10.1073/pnas.98.3.1113.

Irwin, K.K., Renzette, N., Kowalik, T.F., Jensen, J.D., 2016. Antiviral drug resistance as an adaptive process. Virus Evol. 2 (1). https://doi.org/10.1093/ve/vew014.

Jackson, J.A., Friberg, I.M., Bolch, L., Lowe, A., Ralli, C., Harris, P.D., Behnke, J.M., Bradley, J.E., 2009. Immunomodulatory parasites and toll-like receptor-mediated tumour necrosis factor alpha responsiveness in wild mammals. BMC Biol. 7, 16. https://doi.org/10.1186/1741-7007-7-16.

Johnson, P.T., Preston, D.L., Hoverman, J.T., LaFonte, B.E., 2013. Host and parasite diversity jointly control disease risk in complex communities. Proc. Natl. Acad. Sci. USA 110 (42), 16916−16921. https://doi.org/10.1073/pnas.1310557110.

Kochin, B.F., Bull, J.J., Antia, R., 2010. Parasite evolution and life history theory. PLoS Biol. 8 (10), e1000524. https://doi.org/10.1371/journal.pbio.1000524.

Konishi, T., 2022. Mutations in SARS-COV-2 are on the increase against the acquired immunity. PLoS One 17 (7), e0271305. https://doi.org/10.1371/journal.pone.0271305.

Kustin, T., Stern, A., 2021. Biased mutation and selection in RNA viruses. Mol. Biol. Evol. 38 (2), 575—588. https://doi.org/10.1093/molbev/msaa247.

Lafferty, K.D., Shaw, J.C., 2013. Comparing mechanisms of host manipulation across host and parasite taxa. J. Exp. Biol. 216 (1), 56—66. https://doi.org/10.1242/jeb.073668.

Lei, B.R., Olival, K.J., 2014. Contrasting patterns in mammal—bacteria coevolution: *Bartonella* and *Leptospira* in bats and rodents. PLoS Neglected Trop. Dis. 8 (3), e2738. https://doi.org/10.1371/journal.pntd.0002738.

Light, J.E., Hafner, M.S., 2007. Cophylogeny and disparate rates of evolution in sympatric lineages of chewing lice on pocket gophers. Mol. Phylogenet. Evol. 45 (3), 997—1013. https://doi.org/10.1016/j.ympev.2007.09.001.

Louten, J., 2016. Virus structure and classification. Essen. Human Virol. 19—29. https://doi.org/10.1016/B978-0-12-800947-5.00002-8.

Malik, J.A., Ahmed, S., Mir, A., Shinde, M., Bender, O., Alshammari, F., Ansari, M., Anwar, S., 2022. The SARS-CoV-2 mutations versus vaccine effectiveness: new opportunities to new challenges. J. Inf. Public Health 15 (2), 228—240. https://doi.org/10.1016/j.jiph.2021.12.014.

McGuire, M., 1977. A quantitative study of the strategic arms race in the missile age. Rev. Econ. Stat. 59 (3), 328—339. https://doi.org/10.2307/1925051.

Medel, R., Mendez, M.A., Ossa, C.G., Botto-Mahan, C., 2010. Arms race coevolution: the local and geographical structure of a host—parasite interaction. Evol. Educ. Outr. 3, 26—31. https://doi.org/10.1007/s12052-009-0191-7.

Meredith, T.A., Ulrich, N., 2013. Infectious endophthalmitis. In: Vol. 3Retina (Fifth Edition). Elsevier, pp. 2019—2039. (Accessed 15 December 2022).

Meyer, C.J., Cassidy, K.A., Stahler, E.E., Brandell, E.E., Anton, C.B., Stahler, D.R., Smith, D.W., 2022. Parasitic infection increases risk-taking in a social, intermediate host carnivore. Commun. Biol. 5 (1), 1180. https://doi.org/10.1038/s42003-022-04122-0.

Møller, A.P., Dufval, R., Allander, K., 1993. Parasites and the evolution of host social behavior. Adv. Stud. Behav. 22, 65—102.

Moya, A., Holmes, E.C., González-Candelas, F., 2004. The population genetics and evolutionary epidemiology of RNA viruses. Nat. Rev. Microbiol. 2, 279—288. https://doi.org/10.1038/nrmicro863.

Murray, G.G.R., Balmer, A.J., Herbert, J., Hadjirin, N.F., Kemp, C.L., Matuszewska, M., Bruchmann, S., Hossain, A.M.M., Gottschalk, M., Tucker, A.W., Miller, E., Weinert, L.A., 2021. Mutation rate dynamics reflect ecological change in an emerging zoonotic pathogen. PLoS Genet. 17 (11), e1009864. https://doi.org/10.1371/journal.pgen.1009864.

Naveed, A., Abdullah, S., 2021. Impact of parasitic infection on human gut ecology and immune regulations. Trans. Med. Commun. 6, 11. https://doi.org/10.1186/s41231-021-00091-4.

Nextstrain Team, 2022. "Genomic epidemiology of SARS-CoV-2 with subsampling focused globally over the past 6 months" is licensed by CC-BY 4.0. Auspice. https://nextstrain.org/ncov/gisaid/global/6m. (Accessed 16 December 2022).

Noack, S., Harrington, J., Carithers, D.S., Kaminsky, R., Selzer, P.M., 2021. Heartworm disease — overview, intervention, and industry perspective. Int. J. Parasitol.: Drugs Drug Resist. 16, 65—89. https://doi.org/10.1016/j.ijpddr.2021.03.004.

Nowak, M.A., 1992. What is a quasispecies? Trends Ecol. Evol. 7 (4), 118—121. https://doi.org/10.1016/0169-5347(92)90145-2.

Øverli, Ø., Johansen, I.B., 2019. Kindness to the final host and vice versa: a trend for parasites providing easy prey? Front. Ecol. Evol. 7. https://doi.org/10.3389/fevo.2019.00050.

Peck, K.M., Lauring, A.S., 2018. Complexities of viral mutation rates. J. Virol. 92 (14), e01031—17. https://doi.org/10.1128/JVI.01031-17.

Phillips, K.P., Cable, J., Mohammed, R.S., Herdegen-Radwan, M., Raubic, J., Przesmycka, K.J., Oosterhout, C.van, Radwan, J., 2018. Immunogenetic novelty confers a selective advantage in host−pathogen coevolution. Proc. Natl. Acad. Sci. USA 115 (7), 1552−1557. https://doi.org/10.1073/pnas.1708597115.

Papkou, A., Guzella, T., Yang, W., Koepper, S., Pees, B., Schalkowski, R., Barg, M.-C., Rosenstiel, P.C., Teotónio, H., Schulenburg, H., 2019. The genomic basis of Red Queen dynamics during rapid reciprocal host− pathogen coevolution. Proc. Natl. Acad. Sci. USA 116 (3), 923−928. https://doi.org/10.1073/pnas.1810402116.

Portillo, A., Maggi, R.G., Oteo, J.A., Bradley, J.M., García-Álvarez, L., Montserrat, S.-M., Roura, X., Breitschwerdt, E., 2020. *Bartonella* spp. prevalence (serology, culture, and PCR) in sanitary workers in La Rioja Spain. Pathogens 9 (3), 189. https://doi.org/10.3390/pathogens9030189.

Poulin, R., 1999. The functional importance of parasites in animal communities: many roles at many levels? Int. J. Parasitol. 29 (6), 903−914. https://doi.org/10.1016/s0020-7519(99)00045-4.

Poulin, R., 2013. Parasite manipulation of host personality and behavioural syndromes. J. Exp. Biol. 216 (Pt 1), 18−26. https://doi.org/10.1242/jeb.073353.

Prabhune, A., Nagrath, D., Vimal, P., 2019. Systematic review of prevalence, attitude, practices of pubic hair removal (PHR) activities. J. Behav. Health 1−7. https://doi.org/10.5455/jbh.20190828051852.

Quigley, B.J.Z., García López, D., Buckling, A., McKane, A.J., Brown, S.P., 2012. The mode of host−parasite interaction shapes coevolutionary dynamics and the fate of host cooperation. Proc. Biol. Sci. 279 (1743), 3742−3748. https://doi.org/10.1098/rspb.2012.0769.

Rabajante, J.F., Tubay, J.M., Uehara, T., Morita, S., Ebert, D., Yoshimura, J., 2015. Red Queen dynamics in multi-host and multi-parasite interaction system. Sci. Rep. 5 (1), 10004. https://doi.org/10.1038/srep10004.

Rachmadi, A.T., Kitajima, M., Watanabe, K., Okabe, S., Sano, D., 2018. Disinfection as a selection pressure on RNA virus evolution. Environ. Sci. Technol. 52 (5), 2434−2435. https://doi.org/10.1021/acs.est.8b00590.

Robinson, A., Busula, A.O., Voets, M.A., Beshir, K.B., Caulfield, J.C., Powers, S.J., Verhulst, N.O., Winskill, P., Muwanguzi, J., Birkett, M.A., Smallegange, R.C., Masiga, D.K., Mukabana, W.R., Sauerwein, R.W., Sutherland, C.J., Teun, B., Pickett, J.A., Takken, W., Logan, J.G., de Boer, Jetske, G., 2018. Plasmodium-associated changes in human odor attract mosquitoes. Proc. Natl. Acad. Sci. USA 115 (18), E4209−E4218. https://doi.org/10.1073/pnas.1721610115.

Rogalski, M.A., Gowler, C.D., Shaw, C.L., Hufbauer, R.A., Duffy, M.A., 2017. Human drivers of ecological and evolutionary dynamics in emerging and disappearing infectious disease systems. Phil. Trans. Biol. Sci. 372 (1712). https://doi.org/10.1098/rstb.2016.0043.

Rousseau, E., Moury, B., Mailleret, L., Senoussi, R., Palloix, A., Simon, V., Valière, S., Grognard, F., Fabre, F., 2017. Estimating virus effective population size and selection without neutral markers. PLoS Pathog. 13 (11), e1006702. https://doi.org/10.1371/journal.ppat.1006702.

Sánchez, M.I., Biron, D.G., 2019. Editorial: host manipulation by parasites. Front. Ecol. Evol. 7, 369. https://doi.org/10.3389/fevo.2019.00369.

Sanchez, M.I., Ponton, F., Schmidt-Rhaesa, A., Hughes, D.P., Misse, D., Thomas, F., 2008. Two steps to suicide in crickets harbouring hairworms. Anim. Behav. 76 (5), 1621−1624. https://doi.org/10.1016/j.anbehav.2008.07.018.

Sangaré, A.K., Doumbo, O.K., Raoult, D., 2016. Management and treatment of human lice. BioMed Res. Int. 2016, 1−12. https://doi.org/10.1155/2016/8962685.

Sanjuán, R., Domingo-Calap, P., 2016. Mechanisms of viral mutation. Cell. Mol. Life Sci. 73 (23), 4433−4448. https://doi.org/10.1007/s00018-016-2299-6.

Scheifler, M., Ruiz-Rodríguez, M., Sanchez-Brosseau, S., Magnanou, E., Suzuki, Marcelino, T., West, N., Duperron, S., Desdevises, Y., 2019. Characterization of ecto- and endoparasite communities of wild Mediterranean teleosts by a metabarcoding approach. PLoS One 14 (9), e0221475. https://doi.org/10.1371/journal.pone.0221475.

Schulte, R.D., Makus, C., Hasert, B., Michiels, N.K., Schulenburg, H., 2011. Host—parasite local adaptation after experimental coevolution of *Caenorhabditis elegans* and its microparasite *Bacillus thuringiensis*. Proc. Biol. Sci. 278 (1719), 2832—2839. https://doi.org/10.1098/rspb.2011.0019.

Seal, S., Dharmarajan, G., Khan, I., 2021. Evolution of pathogen tolerance and emerging infections: a missing experimental paradigm. eLife 10, e68874. https://doi.org/10.7554/elife.68874.

Siao, M.C., Borner, J., Perkins, S.L., Deitsch, K.W., Kirkman, L.A., 2020. Evolution of host specificity by malaria parasites through altered mechanisms controlling genome maintenance. mBio 11 (2), e03272—19. https://doi.org/10.1128/mbio.03272-19.

Sommer, S., 2005. The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front. Zool. 2 (16). https://doi.org/10.1186/1742-9994-2-16.

Stephenson, J.F., Young, K.A., Fox, J.F., Jokela, J., Cable, J., Perkins, S.E., 2017. Host heterogeneity affects both parasite transmission to and fitness on subsequent hosts. Philos. Trans. Royal Soc. B 372 (1719). https://doi.org/10.1098/rstb.2016.0093, 20160093—20160093.

Stern, A., Andino, R., 2016. Viral evolution. Viral Pathog. 233—240. https://doi.org/10.1016/B978-0-12-800964-2.00017-3.

Strona, G., 2015. Past, present and future of host—parasite co-extinctions. Int. J. Parasitol. Paras. Wildlife 4 (3), 431—441. https://doi.org/10.1016/j.ijppaw.2015.08.007.

Strotz, L.C., Simões, M., Girard, M.G., Breitkreuz, L., Kimmig, J., Lieberman, B.S., 2018. Getting somewhere with the Red Queen: chasing a biologically modern definition of the hypothesis. Biol. Lett. 14 (5), 20170734. https://doi.org/10.1098/rsbl.2017.0734.

Su, X.-Z., Zhang, C., Joy, D.A., 2020. Host-malaria parasite interactions and impacts on mutual evolution. Front. Cell. Infect. Microbiol. 10, 587933. https://doi.org/10.3389/fcimb.2020.587933.

Talapko, J., Škrlec, I., Alebić, T., Jukić, M., Včev, A., 2019. Malaria: the past and the present. Microorganisms 7 (6), 179. https://doi.org/10.3390/microorganisms7060179.

Taylor, M.W., 2014. Introduction: a short history of virology. In: Viruses and Man: A History of Interactions, pp. 1—22.

Thomas, F., Schmidt-Rhaesa, A., Martin, G., Manu, C., Durand, P., Renaud, F., 2002. Do hairworms (*Nematomorpha*) manipulate the water seeking behaviour of their terrestrial hosts? J. Evol. Biol. 15 (3), 356—361. https://doi.org/10.1046/j.1420-9101.2002.00410.x.

VanderWaal, K.L., Ezenwa, V.O., 2016. Heterogeneity in pathogen transmission: mechanisms and methodology. Funct. Ecol. 30 (10), 1606—1622. https://doi.org/10.1111/1365-2435.12645.

Villa, S.M., Evans, M.D., Subhani, Y.K., Altuna, J.C., Bush, S.E., Clayton, D.H., 2018. Body size and fecundity are correlated in feather lice (*Phthiraptera: ischnocera*): implications for Harrison's rule. Ecol. Entomol. 43 (3), 394—396. https://doi.org/10.1111/een.12511.

Villa, S.M., Altuna, J.C., Ruff, J.S., Beach, A.B., Mulvey, L.I., Poole, E.J., Campbell, H.E., Johnson, K.P., Shapiro, M.D., Bush, S.E., Clayton, D.H., 2019. Rapid experimental evolution of reproductive isolation from a single natural population. Proc. Natl. Acad. Sci. USA 116 (27), 13440—13445.

Vitiello, A., 2021. Sars-Cov-2 and risk of antiviral drug resistance. Ir. J. Med. Sci. 191 (5), 2367—2368. https://doi.org/10.10007/s11845-021-02820-y.

Weinersmith, K., Faulkes, Z., 2014. Parasitic manipulation of hosts' phenotype, or how to make a zombie–an introduction to the symposium. Integr. Comp. Biol. 54 (2), 93—100. https://doi.org/10.1093/icb/icu028.

Williams, E.P., Spruill-Harrell, B.M., Taylor, M.K., Lee, J., Nywening, A.V., Yang, Z., Nichols, J.H., Camp, J.V., Owen, R.D., Jonsson, C.B., 2021. Common themes in zoonotic spillover and disease emergence: lessons learned from bat- and rodent-borne RNA viruses. Viruses 13 (8), 1509. https://doi.org/10.3390/v13081509.

Wodarz, D., Levy, D.N., Komarova, N.L., 2018. Multiple infection of cells changes the dynamics of basic viral evolutionary processes. Evol. Lett. 3 (1), 104—115. https://doi.org/10.1002/evl3.95.

Yates, A., Antia, R., Regoes, R.R., 2006. How do pathogen evolution and host heterogeneity interact in disease emergence? Proc. Biol. Sci. 273 (1605), 3075–3083. https://doi.org/10.1098/rspb.2006.3681.

Young, A.D., Gillung, J.P., 2019. Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. Syst. Entomol. 45 (2), 225–247. https://doi.org/10.1111/syen.12406.

Zohdy, S., Bisanzio, D., Tecot, S., Wright, P.C., Jernvall, J., 2017. Aggression and hormones are associated with heterogeneity in parasitism and parasite dynamics in the brown mouse lemur. Anim. Behav. 132, 109–119. https://doi.org/10.1016/j.anbehav.2017.08.002.

Zohdy, S., Schwartz, T., Oaks, J.R., 2019. The coevolution effect as a driver of spillover. Trends Parasitol. 35 (6), 399–408.

This page intentionally left blank

CHAPTER

Human Genetic Variability and Transmissible Diseases

7

Michel Tibayrenc

Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle, MIVEGEC (IRD 224-CNRS 5290-UM1-UM2), IRD Center, Montpellier, France

*See abbreviations and definitions in the glossary of specialized terms.

1. Introduction

In spite of the major achievements of modern medicine, including vaccines and antibiotics, transmissible diseases still represent a major burden for mankind. In Southern countries, plagues, such as malaria, tuberculosis, dengue, yellow fever, and so on, still are the cause of a considerable number of annual casualties. Industrialized countries are not immune to these diseases. The recent epidemics of AIDS and COVID-19 are just a reminder of it. Nosocomial infections, contracted in the hospital, kill more than road accidents in a country like France.

Haldane (1949) has proposed that infectious diseases have been the major factor of mankind's natural selection* in the last 5000 years. This is conceivable. Microbes willingly kill young subjects before the age of reproduction. This is the case for malaria, which threatens many infants in highly endemic areas such as sub-Saharan Africa. Actually, it can be speculated that the increase in infectious risk dates back to the advent of agriculture and the Neolithic era, some 10,000 years ago (Cochran and Harpending, 2009). At the time, hunter-gatherer tribes turned to farmer societies. Populations grew a lot. Promiscuity worsened in villages, then towns and cities. Close contact with domestic animals favored the spread of zoonotic diseases, transmitted from animals to humans and vice versa. Lastly, new diets with less proteins and more carbohydrates may have affected immune defenses.

Environmental factors, such as nutrition, health education, general hygiene, access to health care, undoubtedly play a considerable role in the spread of infectious diseases. However, genetic make-up has also a considerable impact in it (Tibayrenc, 2017). Diseases are phenotypes*. Like any phenotype, they are the result of a close interaction between innate and acquired traits. This chapter will focus more specifically on the role of human genetic variation in transmissible diseases. This emphasis is all the more desirable, since in general, studies on infectious diseases focus more on the pathogen to the detriment of the host (Abdellaoui et al., 2023).

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00002-1
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

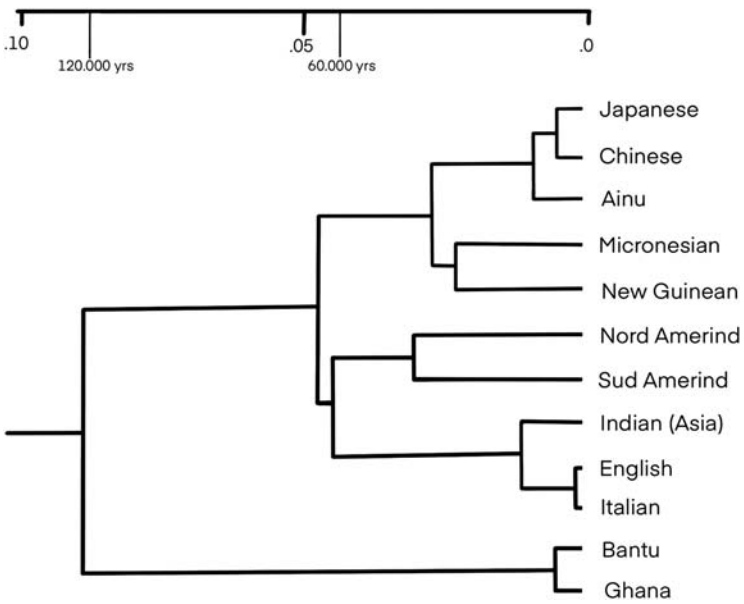


FIGURE 7.1

Phylogenetic* tree depicting the genetic distances* for 11 isoenzyme* loci and 11 blood groups among major geographical populations (inspired from Nei, 1978).

2. A Population Genetics* Framework for the Study of Transmissible Diseases

Homo sapiens (*H. sapien*)is a genetically and phenotypically diversified species. This diversify is uneven and is strongly linked to geographical variation (Fig. 7.1). This kind of species is called "polytypic." A monotypic species is a species with very little genetic and phenotypic variation, which obviously is not the case of *H. sapiens*.

It is recommendable, in a first time, to study how our species' genetic diversity is distributed before considering phenotypic traits, including infectious diseases. The goal is to design a reliable population genetics framework of human populations, upon which relevant phenotypic traits and the genes that command them are placed in a second time (Tibayrenc, 2017). This is the so-called "phylogenetic character mapping" approach (Avise, 2004). This framework should include the whole ecogeographical range of our species, and should obviously be based on the most recent technological advents, namely genomics* and megacomputing. As a matter of fact, these advances have upset our vision of human genetic variability. Abundant data has been harvested by three major international projects, namely the HapMap project (http://hapmap.ncbi.nlm.nih.gov/), the Human Genome Diversity project (HGDP; http://www.hagsc.org/hgdp/), and the 1000 genomes project (http://www.1000genomes.org/). These projects dealing with worldwide distributed populations are completed by geographically limited programs, such as the "UK10K consortium" (http://www.uk10k.org/). These projects, together with many outstanding pioneering studies, have made it possible to yield the main results summarized below.

2.1 Overall Genetic Variation Seems to be Greater Within Major Continental Groups Than Between Them

The seminal article by Lewontin (1972) has shown that about 85% of the total genetic variability is observed within the main continental populations, and only 15% occurs between them. However, although this result is widely used by antiracist activists, it is debatable. First, it is based on the so-called "historical variants," that is to say: these common genetic variants that are mere witnesses of past demographic events and have no direct role in phenotypic variability. These historical markers, such as isoenzymes, are termed "common variants* (CVs*)." They tend to be observed in all geographical populations, although with different frequencies. Statistical results are quite different when considering rarer variants (see: "Low-frequency variants* [LFVs*} and rare variants* {RVs*]").

Then, the statistics on which this result of 85% differences (within populations) versus 15% differences (between populations) is based, the so-called *F*st* statistics (Wright, 1951) is what is called a "black box" in statistics. It gives limited information on the way the genetic variability is actually distributed. If one enters the black box, a relevant trait of this variability is evidenced: it is correlated, that is to say: the variability of marker 1 goes the same direction than the one of markers 2, 3, 4, etc. This remarkable property makes that genetic markers are able to discriminate the main human geographical populations. This is already true with isoenzymes, which power of resolution is limited and are typical historical markers (see above). It is even better verified with modern genomic markers such as single-nucleotide polymorphisms* or SNPs*. Lewontin was an outstanding scientist. It is unexpected that he did not emphasize this remarkable fact in his 1972 paper. This has been called: "Lewontin's fallacy" (Edwards, 2003). With modern genomic markers, it is possible to accurately evidence the geographical origin of any individual (Rosenberg, 2020), even at tiny geographical scales (Novembre et al., 2008).

2.2 Clines* or Trees?

Some authors claim that human genetic variability is clinal*, with no gaps in a perfectly even continuum (Carlson et al., 2022; Lewis et al., 2022; Maglo et al., 2016). "Our evolutionary history is a continuous process of combining the new with the old, and the end result is a mosaic that is modified with each birth and death. This is why the process of using genetics to define "race" is like slicing soup: "You can cut wherever you want, but the soup stays mixed'" (Anonymous, 2002). In a seminal article, Livingstone (1962) claimed that "there are no races, there are only clines."

However, the use of historical markers such as isoenzymes suggests that this view is oversimplified (Fig. 7.1). This tree-like pattern has been fully confirmed by Rosenberg et al. (2002, 2005) and Rosenberg (2020), who used a wide range of microsatellite* markers to analyze 52 worldwide populations. Microsatellites, although still historical markers, have a much higher resolution power than isoenzymes. These studies fully confirmed that the 15% of variation observed among major continental groups (evaluated at 5% by Rosenberg et al. (2002, 2005), Rosenberg (2020) are highly correlated. The clusters evidenced by microsatellites clearly identify the major geographical groups classically distinguished by physical anthropology (Vallois, 1976): Africans, Europeans, Asians, and so on, as do the isoenzyme trees (Fig. 7.1). It is therefore wrong to claim that these major geographical groups can be distinguished from each other only by the use of a broad panel of the most modern parkers, as stated by Jordan (2008): they can be, by the old-fashioned isoenzyme technology.

It can be concluded that human genetic diversity does not behave like a true cline. Shiao et al. (2012) proposed the concept of "clinal classes": the distribution would be clinal within the continental groups, not between them. The gaps observed between the major continental groups would have been

caused by major geographical obstacles. This is the so-called "SHO hypothesis" (Sahara, Himalaya, Oceans) (Maglo et al., 2016), already proposed by Dobzhansky (1962). These major geographical obstacles would cause gaps equivalent to 3100 km of continental distances (Rosenberg et al., 2005).

The major geographical populations are not clear-cut separated: many intermediary populations are observed between them, as already observed by classic physical anthropology (Vallois, 1976).

2.3 Additional Variability and Stratifications

The major continental populations (Europeans, Africans, Asians) are far from homogeneous. For example, in Africa, ethnic groups such as the Fulani can be easily distinguished from neighboring populations by convenient genomic markers. This can be highly relevant when transmissible diseases are concerned (see further).

At an even finer scale, when high-resolution genomic markers are used, strong stratifications can be evidenced, for example, within the British population (Leslie et al., 2015). Such stratifications in Western European populations would be due to both geographical distance (Novembre and Ramachandran, 2011) and to linguistic obstacles (Cavalli-Sforza, 1997). Whatever their origin they mark obstacles to gene flow and, therefore, should be taken into account in the study of transmissible diseases.

2.4 Low-Frequency and Rare Variants

These stratifications at small geographical scales are mostly explained by the presence of low-frequency variants (LFVs) and rare variants (RVs) (Novembre and Ramachandran, 2011), which besides appear to have a strong impact on phenotypes and are usually population-specific, contrary to common variants* ("historical variants") such as isoenzymes and microsatellites. This impact seems to be verified also for the susceptibility to transmissible diseases (see further). Low-frequency and rare variants are defined by their frequencies (minor allele frequency 0.5%−2% and <0.5%, respectively; Nelson et al., 2012).

The work by Lewontin (1972) and its *F*st statistics (Wright, 1951) did not take into account these LFVs and RVs, which were unknown at the time. *F*st statistics designed after these kinds of variants would issue much higher values than the ones cited in Lewontin (1972). The frequency of these LFVs and RVs tends to be population-specific. According to Casals and Bertranpetit (2012), the proportion of RVs that are shared among all continents is only 10%−30%, while 70%−80% are continent-specific. When common variants are concerned, one recalls that these figures are the opposite (Lewontin, 1972).

2.5 "Russian Doll" Patterns

The population structure of human geographical populations appears therefore to exhibit a "Russian doll" pattern* (RDP*). This means that the population structure at limited geographical scales looks like a miniature of the whole species' population structure, with strong genetic stratifications. We have forged the term RDP (Tibayrenc and Ayala, 2013) to describe the population structure of various pathogens. However, in microbes, this kind of population structure is generated by in-built biological traits of these microorganisms, which "restrain their sexuality," since clonal propagation is many times favorable to them. In humans, this is not the case: all human populations are potentially interfertile, and restrictions to gene flow are due only to physical (time and/or space) and cultural obstacles.

2.6 Genetic Stratifications and Polygenic Scores*

Genetic stratifications are also visible at the level of "polygenic scores," which explore the impact of the cumulative effects, on the phenotype of a given individual, of multiple genes that have weak individual effects. The study of nine ancestral groups has shown that the predictive power of polygenic scores inferred from European populations decreases as a function of genetic distance, i.e., it is strongly reduced for Asians, and even more for Africans (Privé et al., 2022). The lower predictive value of polygenic scores inferred from one geographical population to other geographical populations emphasizes the urgent need to base all studies dealing with human genetics on diversified geographical samples (Oni Orisan et al., 2021). This impact of genetic stratifications is visible, not only at the level of these major geographical groups but also within European populations.

2.7 We are All Half-Breed: Mixtures are Ubiquitous

Modern genomic markers confirm that virtually all human populations at present are the result of deep past mixing. This had been already previewed by classic physical anthropology, with a terminology of the time, distinguishing between "primary races" (the exception) and "secondary races" (the rule, the result of deep mixing) (Hooton, 1926). The study of ancient DNA has fully confirmed this visionary view. For example, it shows that present European populations originate from three main ancestral groups, mainly the pastors of the Steppe (the so-called "Yamnaya" people, originating from the present Ukraine and Russia), the ancient middle-east farmers, and the paleo-European hunter-gatherers (Haak et al., 2015).

Recent historical events (wars, colonization, slave trade) have led to intense mixings. For example, the present inhabitants of Latin America feature three major components, namely European (mainly Spanish and Portuguese), Native American, and African. Almost all African Americans in the United States are of mixed ancestry. For them, the European component ranges from 7% to 20%, and varies according to the US states (González Burchard et al., 2003; Tishkoff et al., 2009). Such mixings should be taken into account when genetic susceptibility to infectious diseases is considered.

2.8 If We are All Mestizos, Don't Human Populations Tend Toward Homogenization?

Much less than you might expect: Extensive mixing did not lead to homogenization of populations. Even in multiracial societies such as the United States, cultural barriers, including racial discrimination, and perhaps biological factors (homogamy*) make that interracial marriages remain the exception rather than the rule (Ousley et al., 2009), although they are now increasing (Le Bars, 2020).

According to Risch et al. (2009), this phenomenon of assortative mating is also observed in Puerto Rico and Mexico, as evidenced by "ancestry informative markers*" (AIMS*: see below). In this study, the ethnic assortments of spouses could not be explained either by socioeconomic factors or by geographical proximity.

It is obvious that such lack of homogenization should be taken into account in all studies dealing with human genetic susceptibility to transmissible diseases.

2.9 Ancestry Informative Markers and Self-Reported Ancestry

The existence of alleles that are specific to given geographical populations, or that are very predominant in them (LFVs and RVs) makes it possible to design kits composed of a small number of markers that can reliably identify given populations. These markers are called: "ancestry informative markers" (AIMs). They have been widely used in forensic medicine and in police investigations (Sarich and Miele, 2004). AIMs are able to identify the geographical origin of individuals, and even their degree of interbreeding, down to the individual level (Pfaff et al., 2001). Such properties are precious for the study of infectious diseases. In Latin America, for example, the assessment of the degree of interbreeding (Galanter et al., 2012) is crucial for medical genome-wide association studies* (GWAS*) concerning pathological traits (see below).

AIMs are valuable, but costly. Many scientists limit themselves to self-reported ancestry (SRA). Unexpectedly, it works very well. In several localities in Taiwan and the United States, Tang et al. (2006) found only 0.14% misidentification by SRA versus AIMs, which is negligible.

2.10 General Conclusions

The overall picture that emerges about human population structure is that genetic diversity is tightly linked to geographical distribution. Main groups are associated with continents. They are termed now "continental populations" or "ancestry groups." The term "race," which is highly controversial for understandable historical reasons, tends to be abandoned in the scientific literature (Byeon et al., 2021). In a population genetics point of view, each continental population is highly heterogeneous, the most diverse one being the African group. Moreover, there are no sharp separations among continental populations, which have fuzzy limits. Still the fact remains that human geographical diversity should be cautiously taken into account in the epidemiology and control of infectious diseases, together with environmental parameters (socioeconomical status, education, nutrition, access to health structures). Recently, a team of five scientists called for the necessity to definitely consider human genetic variation in epidemiological studies with the goal to improve health management of the African American population (Oni Orisan et al., 2021). These five scientists are all of African ancestry and therefore cannot be suspected of "systemic racism." Their concern is that most genetic surveys are based on European samplings only. Since the genetic variability of people of African ancestry is much higher than the genetic diversity of European people, this bias leads to missing many genetic variants that are specific of African populations. The same pattern will be verified when Asian populations are concerned.

3. Communicable Diseases: A Phenomenon of Co-Evolution Involving Two or Three Actors

We have now a detailed population genetics framework that will allow us to better understand the way communicable diseases are distributed worldwide. One has to keep in mind two important things: (1) The distribution of genetic and genomic diversity in human populations is quite uneven, and is tightly linked to geography. (2) Geographical populations since the "Out-of-Africa" some 50,000 years ago have undergone different selective pressures caused by quite diverse environmental conditions. This is especially relevant for the case of infectious diseases, since pathogens and their

vectors have undergone also quite different selective pressures throughout the world. According to the so-called "Red Queen" hypothesis (Van Valen, 1973), our species is in a constant race against pathogens, and, of course, against the vectors that transmit pathogens. "Who does not move forward, moves back." It is therefore well ascertained that one will observe clear differences of genetic susceptibility to infectious diseases among geographical populations. The question comes of how to select the "relevant units of analysis" (Tibayrenc and Ayala, 2017) to analyze this phenomenon. Many authors recommend to base such studies on the major geographical groups, namely Africans, Europeans, Asians, etc., as units of analysis (Baer et al., 2013; Bednasz et al., 2019; Popejoy et al., 2018; Risch et al., 2002; Voigt et al., 2016). Recently, the American Medical Association's (AMA) (Press Release, 2020) and the American Academy of Pediatrics (AAP) (News at a glance, 2022) announced that they would eliminate race-based medicine in their recommendations and publications. It can be suspected that these decisions were taken in the name of morals and antiracism. However, from a medical point of view, they are debatable. Many genetic traits linked to transmissible diseases have very different frequencies in African versus European Americans. Sickle cell anemia is an obvious example.

Main geographical populations could be too broad proxies. We have seen in part 1 that these main populations are quite heterogeneous. For more refined studies, it can be relevant to focus on much lesser geographical/ethnic subdivisions. The west African group Fulani for the study of malaria is an example (Arama et al., 2015). The Fulani exhibit a higher resistance to malaria than neighboring ethnic groups.

3.1 The Considerable Role of Environment

This includes nutrition, access to health systems, education, political instability, and wars. Malaria, tuberculosis, and AIDS are called the "diseases of poverty" because these diseases strike hard in countries with weak economies. Similarly, Chagas disease, which occurs from southern United States to northern Argentina, is a striking example of the influence of socioeconomic factors on the impact of a disease. The vectors, triatomine bugs (Fig. 7.2), proliferate in traditional habitats made of "adobe" (Fig. 7.3), that I to say: bricks simply made of dried clay. Modern houses are never colonized by triatomine bugs. When studying the genetic susceptibility to infectious diseases, it is therefore crucial to take into account environmental factors.

Genetic diseases are classified into Mendelian diseases and complex diseases. The first category concerns diseases caused by the action of only one gene, or a few genes. The second category deals with diseases which cause is the action of various genes which individual effect is limited.

3.2 Genetic Inequality and Transmissible Diseases

The distribution and size of pathogen and vector populations varies considerably from one geographical area to another. This has been the cause of different selective pressures on the concerned human populations. Prugnolle et al. (2005) have postulated that the Human Leukocyte Antigens* (HLA)* polymorphism is proportional to the level of pathogen diversity in a given country.

We are not equal to infectious diseases, probably more than for any other disease. This is verified at the individual level (Abel and Dessein, 1997) and at the population level. For an equal infectious contact, all environmental conditions being equivalent, one individual will catch tuberculosis, another

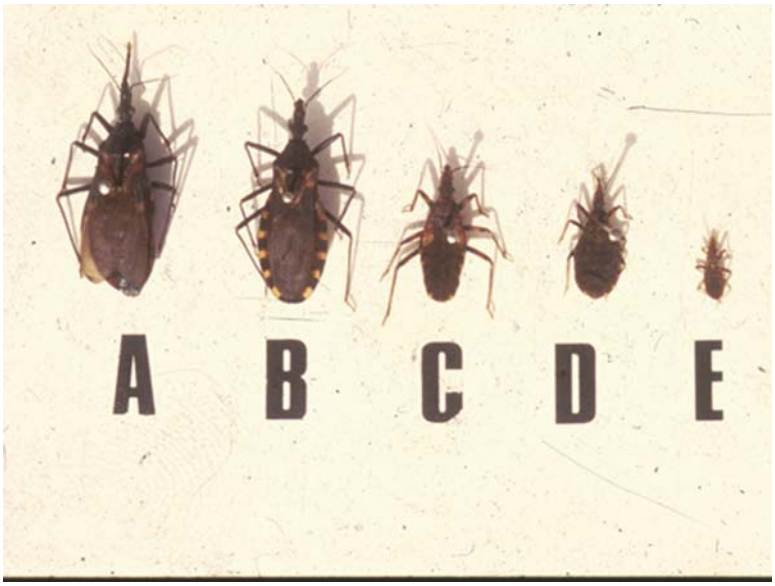


FIGURE 7.2

Triatomine bugs. The species *Triatoma infestans*, widespread in human habitats in South America. From *left* to *right*, male, female, and larvae. All are obligatorily blood-feeding. Note the impressive proboscis of the male. No mosquito net can resist it.



FIGURE 7.3

The author collecting triatomine bugs in Bolivia in the 1980s, with the help of young Aymara Indians. The house is built with adobe, a material made from dried clay. Triatomine bugs proliferate in the cracks and holes of adobe bricks.

one will not. And populations of different geographical origins, in the same environment, will not react in the same way to the same disease.

3.3 How do Humans Protect Themselves Against Pathogens?

Three major diseases illustrate the genetic adaptations of humans to their microscopic enemies. These diseases concern Mendelian traits, since in each case, only one gene is involved. In the three cases, association with ancestry groups is considerable.

- Malaria and sickle cell anemia or drepanocytosis: The cause of this disease is a unique mutation of the gene that codes for β-globin, located on chromosome 11. This results in red cells that have an anomalous shape. Subjects that are homozygous for the sickle cell trait present serious disorders (severe anemia), whereas heterozygous individuals are healthy and are protected against malaria. This is a classic case of balanced selection*. Such protection is of course valuable in tropical Africa. Now sickle cell anemia persists in African Americans (Cavalli-Sforza, 1974), although this population is no longer exposed to malaria.
- Human African trypanosomiasis (HAT, formerly sleeping sickness) is a calamitous disease in tropical Africa. Africans, unlike Europeans and Asians, have developed partially effective genetic tools to resist this disease. However, African Americans, who are no longer at risk for this disease, suffer from a higher rate of kidney disease than European Americans. This has been attributed to variants in the Apolipoprotein L1 (APOL1) gene, which plays a role in protection against sleeping sickness (Karlsson et al., 2014).
- AIDS, caused by human immunodeficiency viruses 1 and 2 (HIV 1 and 2), is a rather counterintuitive example of genetic protection, since Africans seem to be more susceptible to the disease than Europeans. The explanation would be a gene deletion called CCR5Δ32. This trait is much more common in Europeans than in Africans (Dean et al., 1996). Subjects who are heterozygous for this deletion have slower disease progression, while homozygous individuals are highly resistant to HIV infection (Chapman and Hill, 2012).

3.4 Other Diseases

The genes related to infection by *Mycobacterium leprae*, which causes leprosy, Mycobacterium *tuberculosis* (*M. tuberculosis)* (tuberculosis), *Salmonella typhimurium*, the agent of typhoid fever, Lassa and measles viruses, and *Yersinia pestis*, the agent of plague, have undergone recent evolution and are therefore bound to have an uneven geographical distribution (Grossman et al., 2013). The genes involved in leprosy resistance are different in Indian and Chinese populations (Karlsson et al., 2014). In Brazil, some genes are potential markers for susceptibility to leprosy, while others are potential markers for the severe clinical MB (multibacillary) form (Pinto et al., 2015). These markers are linked to ancestry. African admixture is protective against leprosy, while European admixture increases the risk of leprosy development (Pinto et al., 2015). Hepatitis C susceptibility differs among human geographical groups, due to HLA variation (Thio et al., 2001). A hepatitis C-protective C allele of the Interleukin 28B (*IL28B*) gene has its lowest prevalence in Africans, has an intermediary frequency in Europeans and is most prevalent in Asians. The higher frequency of this protective allele in non-African populations could be the result of positive selection (Manry and Quintana-Murci, 2013).

Smallpox resistance seems to have been positively selected in African populations. In African Americans, the cytokine* response to smallpox vaccination is stronger than in European Americans, which suggests a larger effect due to selection in Africa (Karlsson et al., 2014). Lastly, Heyn et al. (2013) evidenced different methylation* patterns (related to gene regulation*), among distinct geographical groups, in genes related to the different rates of infection by hepatitis B virus, measles, and enteropathogenic *Escherichia coli*.

The ACE2-receptor (angiotensin-converting enzyme) plays an important role in the infection by SARS-COV2. A relatively rare variant situated 60bp upstream of this gene may reduce the risk of infection by about 40% (Abdellaoui et al., 2023). Not all ancestry groups are equal before COVID-19 infection and severity. This is generally attributed to discrimination and structural racism, which would explain why African Americans suffered more than European American of COVID-19 infections (Graves and Goodman, 2022; Parens, 2020; Roberts, 2021). However, some results suggest that an ethnic component also plays a role at a biological, not environmental, level in the transmission and severity of COVID-19 (Niedzwiedz et al., 2020; Shelton et al., 2021). However, the precise biological background of these transancestry differences remains speculative.

3.5 Pathogen Strains as Markers of Human Populations

Hirsh et al. (2004) have evidenced that strains of *M. tuberculosis* (the bacterium responsible for tuberculosis) isolated from patients of different ancestry groups in San Francisco are a reflection of the genetic diversity of the strains of their country of origin. These authors evidenced two clusters having the dominant North American *M. tuberculosis* type, a Filipino cluster, and a Southeast Asian cluster. In admixed Colombian populations characterized by AIMs, the genotypes of *Mycobacterium leprae* (the bacterium responsible for leprosy) are linked with the degree of either European or African ancestry (Cardona-Castro et al., 2015). Lastly, since the advent of modern humans, the genetic evolution of the *Helicobacter pylori* bacterium (which causes gastric ulcers) is linked to the evolution of human populations. Different human populations harbor distinct genotypes of the bacterium (Moodley et al., 2012).

3.6 Influence of Human Behavior on Transmissible Disease Epidemiology: The Baldwin Effect and the Niche Construction Theory

The Baldwin effect evaluates the influence, on natural selection, of learned bthavior through cultural transmission. Richerson et al. (2010) have postulated that its impact on human biological evolution has been considerable. Obviously, newly acquired behavior may play an important role in transmissible disease epidemiology. For example, health education in intertropical Africa aims at changing the habit of keeping used tires in backyards. Tires full of rainwater indeed are enjoyed by mosquito larva, and therefore favor the transmission of malaria.

The niche construction theory (Laland et al., 2010) postulates that, by modifying their environments, organisms influence the natural selection acting on themselves. As an example, sanitation of the Pontine marshes in Italy before the Second World War stopped the transmission of malaria in the Rome region. In contrast, in tropical areas, sweet potato farming increased the density of water pools, which favored the proliferation of anopheline mosquito larvae and hence, the spread of malaria. The same probably obtained in Southeast Asia with rice farming and all diseases transmitted by mosquitoes.

Every time humans modify their environment, they influence the transmission and hence the selective pressure of infectious diseases. Both the Baldwin effect and the niche construction theory act differently in different geographical populations having different cultures and behaviors and will therefore have geographically differentiated influences on the natural selection caused by pathogens.

3.7 Transmissible Diseases and History

We have seen (see: "We are all half-breed: mixtures are ubiquitous") that historical facts such as slave trade have deeply modified the genetic make-up of various populations. Now infectious diseases may have influenced some parts of history. Cochran and Harpending (2009) have postulated that the conquest of Latin America by Spaniards and Portuguese has been made much easier by the fact that the *conquistadores* were resistant to certain germs from Europe, while the Amerindians were "naïve" before them. So, although the Incas and Aztecs had powerful armies, their defeat would be explainable by these microscopic weapons, which caused a true Amerindian genocide. Another example is the colonization of French Guiana: before the advent of modern medicine, all attempts to settle French populations in this territory were failures. French Guiana was called the "tomb of the white man." The causes were malaria and yellow fever.

3.8 Archaic Adaptive Introgression

This term designates the contribution of the DNA of archaic humans (Neanderthal, Denisovans) to the adaptation of modern humans. This contribution concerns almost exclusively non-African populations (Racimo et al., 2015). Dignat (2010) states that each European individual harbors 1%−4% Neanderthal DNA. Since these sequences are not the same from one individual to another, the present non-African human population harbors about a total of 30% Neanderthal DNA. According to Cochran and Harpending (2009) and Racimo et al. (2015), Archaic Adaptive Introgression (AAI) has offered modern humans genes useful to fight against pathogens. Quintana-Murci (2021) states that Neanderthal AAI plays a role in the appearance of serious COVID-19 cases. Gaudieri and John (2017) note that Neanderthal AAI had a strong impact on the HLA diversity of modern humans.

4. Conclusions on Human Genetic Variability and Susceptibility to Communicable Diseases

The presentation given in this chapter is purposely abbreviated and simplified. However, it is based on the most recent research results and benefits from the impressive technological progress that has enriched this field of research. Genome sequencing is now routine, and its price is deeply falling. All the most recent results show the same picture: The genetic composition of human populations is finely stratified, and this is verified down to very small geographical scales. Abundant migrations thoroughly shuffled the genes of mankind, but this did not result in homogenization. Genetic tools are able to identify the geographical origin of any individual with high precision, and also, its possible degree of admixture. All of these data have wide applications in forensic medicine. Moreover, we have seen that human genetic variability is of crucial importance in the monitoring and treatment of diseases, in particular transmissible ones. Generally, populations of European origin are overrepresented in genetic surveys (Oni Orisan et al., 2021; Quintana-Murci, 2021). This bias should be corrected.

The future of this medical field relies on more integrative approaches, taking into account jointly environmental and genetic factors. Rapid advances in modern technologies (high-throughput sequencing, proteomic* analysis of gene regulation and epigenetic* phenomena, mega-computing, cloud computing) are paving the way for personalized medicine, specifically targeted at given individuals or populations. To achieve this end, approaches that have been separated until now should be decompartmentalized and opened up to each other: evolutionary sciences and biomedical research, biology, and human sciences. Our knowledge on human diversity, both for basic science and for medicine, must not be limited to biology. This is all the truer since the environment in the broad sense (ecology, socioeconomic conditions, education) plays a major role in the transmission of infectious diseases, in addition to biology and genetics.

Glossary of Specialized Terms

(Gene) regulation Mechanisms that control the amount, location, and timing in which genes are expressed. It involves a variety of processes, including chemical modification of DNA (for example: methylation) and regulatory proteins. Gene regulation is key to the ability of an organism to respond to environmental changes

Ancestry informative marker (AIM) (Pfaff et al., 2001) AIMs are subsets of genetic markers that have very different allele frequencies across different geographical populations of the world. Most polymorphisms are shared among all populations, and for most loci, the most common allele is the same in each population. An ancestry informative marker is a unique combination of genetic markers that occurs mostly in particular founder population sets but may also be found at varying levels across all or some of the populations found in different parts of the world

Balanced polymorphism/balanced selection Genetic polymorphism that persists in a population because the heterozygous individuals for the concerned alleles have a higher fitness than either homozygote

Cline, clinal When a variation is continuous and evenly distributed, as the color spectrum is

Common variant (CV) Genetic variant whose frequency is $\geq 5\%$ in a given population. See "low-frequency variant" and "rare variant."

Cytokine An intercellular molecular messenger that permits communication between immune system cells and cells belonging to other tissue categories. Their role in the defense against infectious agents is crucial

Epigenetics Deals with the activation and deactivation of genes without any change in the corresponding DNA sequence of the organism. Epigenetic changes are sometimes heritable, which goes against the classic belief that only DNA changes are heritable. In other words, parents' life experiences (smoking or starving, for example), in the form of epigenetic tags, can be passed down to future generations

***F*st statistics** A statistics designed by Wright (1951), which purpose is to describe the genetic population structure of given populations by three *F*-statistics, *F*is, *F*it, and *F*st. To summarize, *F*st is a measure of the genetic differences between two populations, for example, between people of European versus African ancestry

Genetic distance Various statistical measures inferred from genetic data, estimating the genetic dissimilarities and hence, the evolutionary differences, among individuals or populations

Genome-wide Association study (GWAS) Compares the repartition of DNA markers (either microsatellites or SNPs) across the genome in people who have a disease or other phenotypic traits to people without the disease or traits. The goal is to identify the genomic region where the genes responsible for the traits under study are localized

Genomics Genetics deals with the analysis of individual genes, while genomics concerns the study of the whole genome

Homogamy The tendency to choose partners who are similar to yourself. Very tall men rarely marry very short women

Human leukocyte antigens (HLA) Cell-surface proteins identified by blood tests that exhibit a huge variability among humans and determine the individual's leukocyte type. The HLA set of genes commands the presentation of antigenic peptides to the immune system and plays a considerable role in tissue compatibility and in infectious processes

Isoenzymes Protein extracts of the biological samples under study are separated by electrophoresis (a method used to separate a biological sample on gels or plates by an electric field). These proteins migrate on the plate at a distance that is a direct result of the DNA sequence of the genes that determine their synthesis. Isoenzyme diversity is therefore a reliable, although gross, measure of the genetic variability of the organisms under study

Low-frequency variant (LFV) Genetic variant whose frequency is between 0.5% and 5% in a given population. See "common variant" and "rare variant."

Methylation A chemical process that adds methyl groups to the DNA molecule, leading to a modification of the expression of the concerned gene

Microsatellite A short DNA sequence, usually 1−4 bp long, which is repeated together in a row along the DNA sequence. In humans, there is a considerable microsatellite variation at both the individual and population levels

Natural selection Mechanism first hypothesized by Charles Darwin, that favors certain genotypes to the detriment of others over generations. It is caused by the interaction of an organism' genetic background with the environment

Phenotype All observable traits of a given individual or a given population, but the genome. The phenotype is not restrained to morphological features and includes, for example, physiological parameters (blood pressure, muscular strength, etc.) biochemical parameters (level of cholesterol, etc.), or cognitive properties. The phenotype is the result of the interaction between the genotype and the environment. The variable part that is specifically linked to the action of the genotype constitutes the heritability of the character considered

Phylogenetics A branch of genetics which goal is to reconstitute the evolutionary past and relationships of taxa or of separate evolutionary lines

Polygenic score In genetics, a polygenic score, also called a polygenic risk score, genetic risk score, or genome-wide score, is a value that summarizes the estimated effect of many genetic variants on a given phenotype of a given individual, for example the risk of occurrence of a given disease

Population genetics The analysis of genetic variation within and between populations. It involves the examination and modeling of changes in the frequencies of alleles and genes in populations over space and time. Many of the genes found within a population are polymorphic, that is, they will occur in a number of different molecular variants (alleles). Various mathematical and statistical models are designed to investigate and predict the occurrence of specific alleles or combinations of alleles within and between populations. The focus of population genetics is the population or the species rather than the individual

Proteomics This term is used to describe all the proteins expressed by the genome of a given organism. Proteomics, then, is the study of the expression, function, and interaction of proteins and is widely used to explore pathogenic processes (by comparing gene * expression in healthy and sick individuals). Proteomics explores gene expression at the scale of the entire genome and is therefore the logical complement of genomics

Rare variant (RV) Genetic variant whose frequency is less than 0.5% in a given population. See "low-frequency variant" and "common variant."

Russian doll pattern (RDP) (Tibayrenc and Ayala, 2013). When the population structure of the species under study appears at lower evolutionary scales as a miniature of the whole species' population structure

Single-nucleotide polymorphism (SNP) Polymorphisms of one-base variations in the DNA sequence. Most SNPs have no effect, others cause subtle differences in various traits, such as appearance, while some have an impact on the risk for certain diseases. SNPs are extensively used as high-resolution population markers

Acknowledgments

We thank Dr Jenny Telleria (IRD, Montpellier, France) for designing Figure 7.1.

References

Abdellaoui, A., et al., 2023. 15 years of GWAS discovery: realizing the promise. Am. J. Hum. Genet. 110 (2), 179–194. https://doi.org/10.1016/j.ajhg.2022.12.011.

Abel, L., Dessein, A.J., 1997. The impact of host genetics on susceptibility to human infectious diseases. Curr. Opin. Immunol. 9 (4), 509–516. https://doi.org/10.1016/S0952-7915(97)80103-3.

Anonymous, 2002. Slicing soup. Nature Biotechnol. 20, 637.

Arama, C., et al., 2015. Ethnic differences in susceptibility to malaria: what have we learned from immuno-epidemiological studies in West Africa? Acta Trop. 146, 152–156. https://doi.org/10.1016/j.actatropica.2015.03.023.

Avise, J.C., 2004. Molecular Markers, Natural History and Evolution. Chapman, and Hall, New York, London.

Baer, R.D., et al., 2013. Concepts of race and ethnicity among health researchers: patterns and implications. Ethn. Health 18 (2), 211–225. https://doi.org/10.1080/13557858.2012.713091.

Bednasz, C.J., et al., 2019. Race/ethnicity and protease inhibitor use influence plasma tenofovir exposure in adults living with HIV-1 in AIDS clinical trials group study A5202. Antimicrob. Agents Chemother. 63 (4). https://doi.org/10.1128/AAC.01638-18.

Byeon, Y.J.J., et al., 2021. Evolving use of ancestry, ethnicity, and race in genetics research—a survey spanning seven decades. Am. J. Hum. Genet. 108 (12), 2215–2223. https://doi.org/10.1016/j.ajhg.2021.10.008.

Cardona-Castro, N., et al., 2015. Human genetic ancestral composition correlates with the origin of *Mycobacterium leprae* strains in a leprosy endemic population. PLoS Neglected Trop. Dis. 9 (9), e0004045. https://doi.org/10.1371/journal.pntd.0004045.

Carlson, J., et al., 2022. Counter the weaponization of genetics research by extremists. Nature 610 (7932), 444–447. https://doi.org/10.1038/d41586-022-03252-z.

Casals, F., Bertranpetit, J., 2012. Human genetic variation, shared and private. Science 337 (6090), 39–40. https://doi.org/10.1126/science.1224528.

Cavalli-Sforza, L.L., 1974. The genetics of human populations. Sci. Am. 231 (3), 80–89. https://doi.org/10.1038/scientificamerican0974-80.

Cavalli-Sforza, L.L., 1997. Genes, peoples, and languages. Proc. Natl. Acad. Sci. U. S. A. 94 (15), 7719–7724. https://doi.org/10.1073/pnas.94.15.7719.

Chapman, S.J., Hill, A.V.S., 2012. Human genetic susceptibility to infectious disease. Nat. Rev. Genet. 13 (3), 175–188. https://doi.org/10.1038/nrg3114.

Cochran, G., Harpending, H., 2009. The 10,000 Year Explosion- How Civilization Accelerated Human Evolution. Basic books. A member of the Perseus Books Group, New York.

Dean, M., et al., 1996. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Science 273 (5283), 1856–1862. https://doi.org/10.1126/science.273.5283.1856.

Dignat, A., 2010. *Homo sapiens* Convole Avec Néandertal. Les Amis d'Herodote. Available at: https://www.herodote.net/histoire/evenement.php?jour=20100507&ID_dossier=250.

Dobzhansky, T., 1962. Mankind Evolving. Yale University Press.

Edwards, A.W.F., 2003. Human genetic diversity: lewontin's fallacy. Bioessays 25 (8), 798–801. https://doi.org/10.1002/bies.10315.

Galanter, J.M., et al., 2012. Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas. PLoS Genet. 8 (3). https://doi.org/10.1371/journal.pgen.1002554.

Gaudieri, S., John, M., 2017. Evolution of the human leukocyte antigen system. In: Tibayrenc, M., Ayala, F.J. (Eds.), On Human Nature: Biology, Psychology, Ethics, Politics, and Religion. Elsevier/Academic Press, San Diego, pp. 211–219. https://doi.org/10.1016/B978-0-12-420190-3.00013-2.

González Burchard, E., et al., 2003. The importance of race and ethnic background in biomedical research and clinical practice. N. Engl. J. Med. 348 (12), 1170–1175. https://doi.org/10.1056/NEJMsb025007.

Graves, J.L., Goodman, A.H., 2022. Racism, not Race; Answers to Frequently Asked Questions. Columbia University press.

Grossman, S.R., et al., 2013. Identifying recent adaptations in large-scale genomic data. Cell 152 (4), 703–713. https://doi.org/10.1016/j.cell.2013.01.035.

Haak, W., et al., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522 (7555), 207–211. https://doi.org/10.1038/nature14317.

Haldane, J.B.S., 1949. Disease and evolution. La Ricerca Scientifica 19, 68–76.

Heyn, H., et al., 2013. DNA methylation contributes to natural human variation. Genome Res. 23 (9), 1363–1372. https://doi.org/10.1101/gr.154187.112.

Hirsh, A.E., et al., 2004. Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. Proc. Natl. Acad. Sci. U. S. A. 101 (14), 4871–4876. https://doi.org/10.1073/pnas.0305627101.

Hooton, E.A., 1926. Methods of racial analysis. Science 63 (1621), 75–81. https://doi.org/10.1126/science.63.1621.75.

Jordan, B., 2008. L'humanité au pluriel. La génétique et la question des races. éditions du Seuil, Paris.

Karlsson, E.K., Kwiatkowski, D.P., Sabeti, P.C., 2014. Natural selection and infectious disease in human populations. Nat. Rev. Genet. 15 (6), 379–393. https://doi.org/10.1038/nrg3734.

Laland, K.N., Odling-Smee, J., Myles, S., 2010. How culture shaped the human genome: bringing genetics and the human sciences together. Nat. Rev. Genet. 11 (2), 137–148. https://doi.org/10.1038/nrg2734.

Le Bars, 2020. Ces Américains « multiraciaux » qui ne rentrent pas dans les cases (those « multiracial » Americans who do not fit the boxes). https://www.lemonde.fr/m-le-mag/article/2020/12/22/ces-americains-multiraciaux-qui-ne-rentrent-pas-dans-les-cases_6064182_4500055.html.

Leslie, S., et al., 2015. The fine-scale genetic structure of the British population. Nature 519 (7543), 309–314. https://doi.org/10.1038/nature14230.

Lewis, A.C.F., et al., 2022. Getting genetic ancestry right for science and society. Science 376 (6590), 250–252. https://doi.org/10.1126/science.abm7530.

Lewontin, R.C., 1972. The apportionment of human diversity. In: Evolutionary Biology, pp. 381–398. https://doi.org/10.1007/978-1-4684-9063-3_14.

Livingstone, F.B., 1962. On the non-existence of human races. Curr. Anthropol. 3 (3), 279–281. https://doi.org/10.1086/200290.

Maglo, K.N., Mersha, T.B., Martin, L.J., 2016. Population genomics and the statistical values of race: an interdisciplinary perspective on the biological classification of human populations and implications for clinical genetic epidemiological research. Front. Genet. 7. https://doi.org/10.3389/fgene.2016.00022.

Manry, J., Quintana-Murci, L., 2013. A genome-wide perspective of human diversity and its implications in infectious disease. Cold Spring Harbor Perspectives in Medicine 3 (1). https://doi.org/10.1101/cshperspect.a012450.

Moodley, Y., et al., 2012. Age of the association between *Helicobacter pylori* and man. PLoS Pathog. 8 (5). https://doi.org/10.1371/journal.ppat.1002693.

Nei, M., 1978. The theory of genetic distance and evolution of human races. Jpn. J. Hum. Genet. 23 (4), 341–369. https://doi.org/10.1007/bf01908190.

Nelson, M.R., et al., 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. Science 336 (6090), 100–104. https://doi.org/10.1126/science.1217876.

News at a Glance, 2022. Pediatricians take stand on race. Science 376, 677.
Niedzwiedz, C.L., et al., 2020. Ethnic and socioeconomic differences in SARS-CoV-2 infection: prospective cohort study using UK Biobank. BMC Med. 18 (1). https://doi.org/10.1186/s12916-020-01640-8.
Novembre, J., et al., 2008. Genes mirror geography within Europe. Nature 456 (7218), 98−101. https://doi.org/10.1038/nature07331.
Novembre, J., Ramachandran, S., 2011. Perspectives on human population structure at the cusp of the sequencing era. Annu. Rev. Genom. Hum. Genet. 12, 245−274. https://doi.org/10.1146/annurev-genom-090810-183123.
Oni Orisan, A., et al., 2021. Embracing genetic diversity to improve black health. N. Engl. J. Med. 384 (12), 1163−1167. https://doi.org/10.1056/NEJMms2031080.
Ousley, S., Jantz, R., Freid, D., 2009. Understanding race and human variation: why forensic anthropologists are good at identifying race. Am. J. Phys. Anthropol. 139 (1), 68−76. https://doi.org/10.1002/ajpa.21006.
Parens, E., 2020. The Inflated Promise of Genomic Medicine. Scientific American. Available at: https://blogs.scientificamerican.com/observations/the-inflated-promise-of-genomic-medicine/.
Pfaff, C.L., et al., 2001. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. Am. J. Hum. Genet. 68 (1), 198−207. https://doi.org/10.1086/316935.
Pinto, P., et al., 2015. Influence of genetic ancestry on INDEL markers of NFKβ1, CASP8, PAR1, IL4 and CYP19A1 genes in leprosy patients. PLoS Neglected Trop. Dis. 9 (9), e0004050. https://doi.org/10.1371/journal.pntd.0004050.
Popejoy, A.B., et al., Clinical Genome Resource (ClinGen) Ancestry and DiversityWorking Group (ADWG), 2018. The clinical imperative for inclusivity: race, ethnicity, and ancestry (REA) in genomics. Hum. Mutat. 39, 1713−1720.
Press Release, 2020. New AMA Policies Recognize Race as a Social, not Biological, Construct. AMA Media and Editorial.
Privé, F., et al., 2022. Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. Am. J. Hum. Genet. 109 (1), 12−23. https://doi.org/10.1016/j.ajhg.2021.11.008.
Prugnolle, F., et al., 2005. Pathogen-driven selection and worldwide HLA class I diversity. Curr. Biol. 15 (11), 1022−1027. https://doi.org/10.1016/j.cub.2005.04.050.
Quintana-Murci, L., 2021. Le peuple des humains. Editions Odile Jacob, Paris, France.
Racimo, F., et al., 2015. Evidence for archaic adaptive introgression in humans. Nat. Rev. Genet. 16 (6), 359−371. https://doi.org/10.1038/nrg3936.
Richerson, P.J., Boyd, R., Henrich, J., 2010. Gene-culture coevolution in the age of genomics. Proc. Natl. Acad. Sci. U. S. A. 107 (Suppl. ment_2), 8985−8992. https://doi.org/10.1073/pnas.0914631107.
Risch, N., et al., 2002. Categorization of humans in biomedical research: genes, race and disease. Genome Biol. 3 (7), 12.
Risch, N., et al., 2009. Ancestry related assortative mating in Latino populations. Genome Biol. 10, R132.
Roberts, D.E., 2021. Abolish race correction. Lancet 397 (10268), 17−18. https://doi.org/10.1016/S0140-6736(20)32716-1.
Rosenberg, N.A., 2020. A population-genetic perspective on the similarities and differences among worldwide human populations. Hum. Biol. 92 (3), 135−152. https://doi.org/10.13110/humanbiology.92.3.02.
Rosenberg, N.A., et al., 2005. Clines, clusters, and the effect of study design on the inference of human population structure. PLoS Genet. 1 (6), 0660−0671. https://doi.org/10.1371/journal.pgen.0010070.
Rosenberg, N.A., et al., 2002. Genetic structure of human populations. Science 298 (5602), 2381−2385. https://doi.org/10.1126/science.1078311.
Sarich, V., Miele, F., 2004. Race: The Reality of Human Differences. Westview Press, Boulder, CO.
Shelton, J.F., et al., 2021. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. Nat. Genet. 53 (6), 801−808. https://doi.org/10.1038/s41588-021-00854-7.

Shiao, J.L., et al., 2012. The genomic challenge to the social construction of race. Socio. Theor. 30 (2), 67–88. https://doi.org/10.1177/0735275112448053.

Tang, H., et al., 2006. Reconstructing genetic ancestry blocks in admixed individuals. Am. J. Hum. Genet. 79 (1), 1–12. https://doi.org/10.1086/504302.

Thio, C.L., et al., 2001. Racial differences in HLA class II associations with hepatitis C virus outcomes. JID J. Infect. Dis. 184 (1), 16–21. https://doi.org/10.1086/321005.

Tibayrenc, M., 2017. Human intergroup variation and disease genetics. In: On Human Nature: Biology, Psychology, Ethics, Politics, and Religion. Elsevier Inc., France, pp. 161–175. https://doi.org/10.1016/B978-0-12-420190-3.00010-7.

Tibayrenc, M., Ayala, F.J., 2013. How clonal are *trypanosoma* and *leishmania*? Trends Parasitol. 29 (6), 264–269. https://doi.org/10.1016/j.pt..2013.03.007.

Tibayrenc, M., Ayala, F.J., 2017. Relevant units of analysis for applied and basic research dealing with neglected transmissible diseases: the predominant clonal evolution model of pathogenic microorganisms. PLOS Neglected Trop. Dis. 11 (4), e0005293. https://doi.org/10.1371/journal.pntd.0005293.

Tishkoff, S.A., et al., 2009. The genetic structure and history of Africans and African Americans. Science 324 (5930), 1035–1044. https://doi.org/10.1126/science.1172257.

Vallois, H.V., 1976. Les Races Humaines. Presses Universitaires de France, Paris.

Van Valen, L., 1973. A new evolutionary law. Evol. Theor. 1, 1–30.

Voigt, E.A., et al., 2016. Genetically defined race, but not sex, is associated with higher humoral and cellular immune responses to measles vaccination. Vaccine 34 (41), 4913–4919. https://doi.org/10.1016/j.vaccine.2016.08.060.

Wright, S., 1951. The genetical structure of populations. Ann. Eug. 15 (4), 323–354.

This page intentionally left blank

CHAPTER

Pathogens as Tracers of Past Human Demography and Migrations

8

Charlotte J. Houldcroft[1,2]

[1]*Department of Genetics, University of Cambridge, Cambridge, United Kingdom;* [2]*McDonald Institute for Archaeological Research, Cambridge, United Kingdom*

1. Introduction

Human evolutionary biology addresses key questions surrounding human origins and our migratory history. The methods and polymorphic markers used in human population genetics to investigate these questions have evolved themselves over time: from proteins, uniparental markers such as the non-recombining portions of the Y chromosome (NRY) and mitochondrial DNA (mtDNA), microsatellites and single-nucleotide polymorphisms, to structural variants (insertions, deletions, and larger copy-number variants) and whole-genome sequences (Bergström et al., 2020; Cavalli-Sforza and Feldman, 2003; Li and Durbin, 2011; Schlebusch et al., 2017; Tishkoff et al., 2007). These genetic data have been combined with data from skeletal morphology, anthropological studies, and linguistics to arrive at the most accepted model of modern human evolution: the divergence of living populations of humans from one another 250−350 thousand years ago (kya) (Schlebusch et al., 2017); followed by a recent Out-of-Africa expansion of modern humans into Asia, Oceania, Europe, and finally the Americas from 70kya onwards (Bergström et al., 2021). This is a simplified model, and caveats and controversies will be discussed in more detail later.

By studying the genomes of thousands of living people, we have been able to show that human migrations have, over thousands of years, led to a reduction in genetic diversity. Each wave of migration caused a serial founder effect to occur, with genetic diversity decreasing with distance from East Africa. While human genetic differentiation does increase with geographical distance between populations (isolation-by-distance [IBD]), there is contention as to how high this Fst (a measure of population differentiation) between populations is. Most estimates are low (Fst <2%), whereas a more recent study suggested Fst values greater than 25% (Ochoa and Storey, 2021). Using DNA from contemporary populations, human genetic studies sometimes struggle to resolve patterns of human migration, particularly if populations are closely related and/or only recently diverged from one another.

Archeological evidence points to earlier migrations of *Homo sapiens* into Israel, the Levant, Saudi Arabia, and China 80kya (Bergström et al., 2021; Liu et al., 2015), with older fossil evidence of a more fragmentary nature from Greece and Israel (180−210kya (Harvati et al., 2019; Hershkovitz et al., 2018)).

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00032-X
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Whether these fossils are direct ancestors of any humans alive today is not clear, particularly for the oldest remains, and is a contentious topic (Higham and Douka, 2021; Sun et al., 2021). They may represent an earlier "Out of Africa" population expansion that failed (dying out without leaving surviving descendants), or was outcompeted by later expansions of modern humans (Bergström et al., 2021). Climate modeling suggests that there were multiple windows of possible expansion Out-of-Africa for *Homo sapiens* over the last 300kya, by northern or southern routes. Some of these periods were more hostile than others and might have led to "failed" waves of expansion, making no detectable contribution to the gene pool of humans alive today (Beyer et al., 2021).

There remain many aspects of human evolution and migration where, despite the extensive contributions of analyses of human genetic markers, recent patterns and changes cannot be established from human DNA studies alone. For example, the initial peopling of the Americas is still a topic of extensive research—to understand the number of migrations, their timings and sources (Moreno-Mayar et al., 2018; Reich et al., 2012; Skoglund et al., 2015; Waters, 2019). Human migrations to Oceania are also a subject of debate, including human interactions with other hominin groups (Choin et al., 2021; Jacobs et al., 2019) and the putative existence of an earlier *Homo sapiens* (Skoglund and Reich, 2016) or other hominin migration which left only a minimal contribution to the genomes of some populations alive today (Durvasula and Sankararaman, 2020; Rogers et al., 2020), as mentioned above. Recent migrations within western and northern Europe, such as the genetic contributions to recipient populations of Danish Viking, Saxon, and Norman expansions, are difficult to deconvolute because of the recent shared genetic histories of these populations and their low genetic divergence (Byrne et al., 2018; Jobling, 2012). Analysis of ancient DNA can identify ancestries or migrations that are "missing" from modern genetic diversity, and can highlight particular genetic signals which help to distinguish between related populations if sufficient diversity exists; where populations are too closely related, other sources of diversity need to be found to disentangle their relationships.

In order to overcome issues with relying on relatively slowly mutating human genomes alone, a number of pathogens have proven useful for inferring, tracing, and dating human migrations. For example, if contact between very early human expansions and the later, more successful Out of Africa migration from which the majority of human genetic diversity is derived *did* take place, it may be preserved in the genomes of human pathogens carried by these groups, which are still part of the modern human microbiome or virome.

Pathogen genomics can also allow us to distinguish between closely related human populations with a resolution not possible from currently available human genetic data alone, for example, due to patchy sampling or low resolution genotyping or marker studies. Here, I shall introduce the advantages and disadvantages of using pathogens as human migration markers; and discuss the lessons learned from using pathogens as tracers, with examples from bacterial and viral genomics.

2. Pathogen Genomics to Infer Host History

Many pathogens have been studied genetically and genomically to infer the evolutionary history of their hosts. Viruses are popular choices because, when carefully selected, they have a narrow relationship with their hosts (e.g., due to specializing in infecting a single species, owing to dependencies on species-specific factors such as cellular receptors). Pathogens may maintain higher genetic diversity than their hosts, with larger effective population sizes, higher mutation rates and shorter generation

times. This accumulated genetic diversity is particularly useful for phylogeographic analysis. Viral genomes have often been used to study deep evolutionary relationships across primate species where small but diverse fragments of viral genomes may convey more information than a similar-length fragment of the host genome. There has been considerable interest in using gammaherpesvirus genomes to study both host and virus phylogeny in primates; and while human gammaherpesvirus 4 (Epstein–Barr virus) shows strict host specificity, the 1:1 relationship between virus species and host species in primates breaks down as the phylogeny grows wider and deeper, revealing both viral lineage loss and host-switching events (Azab et al., 2018).

However, even a long and exclusive evolutionary history between pathogen and host may mask differences in their genealogies; and phylogenies may appear similar in pathogens and hosts through different mechanisms that do not reflect migrations or speciation events taking place on the same time scales, e.g., varicella-zoster virus or tuberculosis (Bos et al., 2014; Pontremoli et al., 2019; Weinert et al., 2015).

It is therefore essential to choose the relevant microbe for the research question in hand. Studies need to consider how strict the host–pathogen relationship is, the mode of transmission, the mutation rate, and the frequency of recombination (both within and between species, such as HSV1 donating genomic regions to the related virus HSV2 (Burrel et al., 2017; Koelle et al., 2017)). Choosing the appropriate microbe or virus means studying the pathogen through family transmission studies, ecological surveys, and in the laboratory and the clinic (for example, to obtain estimates of the mutation rate in vitro and in vivo). It may also mean studying closely related animal pathogens to act as outgroups. For example, studying the genome of *Yersinia pseudotuberculosis* has informed our understanding of the origins and evolution of the causative agent of plague, *Yersinia pestis* (Achtman et al., 1999b; Demeure et al., 2019).

When considering which pathogens will be most informative to understand human evolutionary history and migrations, those which are transmitted vertically (mother to child) and which cause long-term, persistent and/or latent infection are likely to be the most informative about long-term changes. Some horizontal transmission may be a tolerable aspect of pathogen biology if it tends to occur within host populations rather than between them, and if vertical transmission remains dominant, ensuring host and pathogen phylogenies remain consistent. Pathogens that are associated with horizontal sexual transmission may be important for understanding admixture between populations (e.g., human papillomavirus and human-Neanderthal sexual contact (Pimenoff et al., 2017)). Pathogens that are associated with epidemics, or that have become endemic after a previously epidemic cycle of infection (e.g., measles), do not well-reflect the history of their host. Epidemics lead to cycles of rapid increase in pathogen population size followed by sharp bottlenecks as the epidemic shrinks, sharply altering the demographic history of the pathogen compared with the host population.

To use pathogens as tracers of human migration most effectively, the rate of molecular evolution within the pathogen genome is particularly important. A low mutation rate means that there may be poor phylogenetic resolution of the pathogen family tree, especially if only subgenomic regions are studied; and therefore studying recent human migrations becomes impossible, although older events may still be reflected. A pathogen with a high mutation rate presents the opposite challenge: very recent migrations may be inferred phylogenetically (e.g., transmission of SARS-CoV-2 Delta around the world reflected the air-travel driven migration of humans in 2021). However, there is a risk of site saturation (multiple mutations have occurred or recurred at informative genomic sites (Saxenhofer et al., 2017)) and homoplasies (the same mutation occurring on different genomic backgrounds/

lineages despite not being present in the common ancestor). Choosing the right pathogen for the human timescale of interest therefore depends on having an accurate estimate of the pathogen mutation rate. If this rate is estimated incorrectly, then inferences drawn from the pathogen about human migrations will be wrong too.

The recombination rate is also an important factor in pathogen selection. Human genomes are, of course, recombinant; and this explains why uniparental markers much less prone to recombination (the mitochondrial genome and the NRY) have been preferred historically for understanding human migrations: they tell simpler stories. Some pathogens are more prone to recombination than others, and recombination presents challenges for phylogenetic reconstruction and dating analyses. Recombination is associated with genotypes that fall into "intermediate" phylogenetic positions, it generates longer terminal branches, and can lead to underestimation of the time to the most recent common ancestor (TMRCA), although analysis methods and software are increasingly able to take account of recombination (Spielman et al., 2019) and mask it or correct for it (e.g., (Li et al., 2019)). Because recombination typically takes place between populations that are related, genetic structure is typically maintained, albeit with a loss of signal-strength. It is therefore essential to test for recombination before using a pathogen phylogeny for host inferences, and to use appropriate downstream analysis methods for the recombination landscape of the virus or microbe of interest. It may also be necessary to restrict further analyses to nonrecombining regions of the genome or recombination cold spots.

Following considerations of mutation rate and recombination, selection is a third key factor to consider in choosing a tracer-pathogen. Selection is detected by comparing the ratios of nonsynonymous (dN) to synonymous (dS) site changes within protein-coding genes, with a range of computational tools. If the dN/dS ratio is 1, this indicates that the gene is evolving neutrally. Ratios less than one are indicative of purifying selection, in which amino acid variation is penalized; ratios greater than one are indicative of directional selection, in which certain amino acid changes are beneficial in the current landscape of selection pressures. Pathogen genomes can be screened before analysis to identify regions most likely to be evolving neutrally. It is important to account for selection: it has the potential to confound pathogen phylogenetics because it may expose populations with distinct evolutionary histories to the same evolutionary pressures. This may in turn cause populations to cluster together and appear more similar than their histories would suggest (Wirth et al., 2005). *Helicobacter pylori* provides a case study of this: the population structure of *H. pylori* has been studied using the *cag* pathogenicity island (*cag*PAI), which is under selection; and housekeeping genes. When Peruvian strains of *H. pylori* were analyzed using *cag*PAI alone, they were inferred to have a European origin because they carried the "western" *cag*PAI type. Analyzing the housekeeping genes showed that some Peruvian *H. pylori* strains in fact clustered with Ameriandian isolates (hpAmerind, likely of Asian origin), demonstrating that single markers such as *cag*PAI are not typically informative enough to understand the whole evolutionary history of a pathogen, and may in fact be misleading (Devi et al., 2006).

Taken together, it is desirable to use as much pathogen genomic information to reconstruct phylogenies as possible. By going beyond single markers, a more accurate phylogeny can be reconstructed, and researchers have the luxury of excluding highly recombinant regions and markers under strong directional selection, prioritizing portions of the genome with known mutation rates. Paired host and pathogen material (i.e., from the same sample) is the gold-standard for these kinds of analysis, although not often available in large numbers from a broad geographic area.

3. Bacteria

3.1 Helicobacter pylori

Helicobacter pylori is a Gram-negative bacterium closely associated with human diseases such as stomach ulcers, gastritis, and lymphomas of the lymphoid tissues of the stomach. *H. pylori* has been extensively studied as a marker of human history for a number of reasons. Despite its association with stomach cancer, most infections are asymptomatic; much (though not all) transmission is vertical, taking place during childhood; in many parts of the world, especially less industrialized areas, infection is common (>50%); and infection usually involves a dominant strain of *H. pylori*, although mixed infection also takes place (Osaki et al., 2015; Raymond et al., 2004; Schwarz et al., 2008). *H. pylori* also has a high degree of genetic diversity, partly driven by a high mutation rate, estimated to be between $0.5-6.5 \times 10^{-5}$/site/year. These factors are all favorable for a human demographic tracer; however, in many parts of the world people frequently acquire new strains and *H. pylori* is prone to frequent recombination in these mixed infection settings (Didelot et al., 2013), estimated to occur at a rate of 5.5×10^{-5} recombinations/initiation site/year (Kennemann et al., 2011). Recombination between closely related strains from the same geographical area of origin is likely to maintain phylogeography in *H. pylori* despite this tendency to recombination.

When analyzed on a whole-genome basis, the most divergent *H. pylori* lineages are found among the San hunter-gatherers (hpAfrica2); other strains diverge from hpAfrica2, leading to the hpSahul group of strains, while the more closely related central Asian (hpAsia2), European (hpEurope), and north-east African (hpNEAfrica) lineages branch off next. Amerindian and east Asian strains share a recent common ancestor (Montano et al., 2015). Further subtypes are known from multilocus sequence typing (MLST) or other subgenomic analyses. Increasing the number of *H. pylori* strains sequenced at a whole-genome level has given greater resolution to the population structure of many regions: for example, including genomes from the Iberian peninsula revealed a hspSWEurope subpopulation of strains that had not previously been recognized (Muñoz-Ramirez et al., 2021). Studying the prophages which are found within ~20% of *H. pylori* strains reveals further geographic structure, for example breaking the hpEurope multilocus sequence type into hpNEurope and hpSWEurope (Vale et al., 2015).

These bacterial population groups are hypothesized to arise from six ancestral populations: Ancestral Africa 1, Ancestral Africa 2, Ancestral Europe 1 (AE1), Ancestral Europe 2 (AE2), Ancestral East Asia, and Ancestral Sahul. These fit within a model of successive waves of human migration from Africa with an early branch between European and Asian ancestral populations, and strains of *H. pylori* arriving in the Americas originally from Asian populations through the Bering Strait, with later European and African lineages introduced in the Colonial Era (Falush et al., 2003; Moodley et al., 2021).

H. pylori population genetics may provide further insights into the peopling of the Americas, particularly the identity of source populations for the different ancestral components of modern Native American, indigenous American and First Nations peoples of different regions of North and South America (Flegontov et al., 2019; Posth et al., 2018; Skoglund et al., 2015). Siberia has a deep history of human habitation (for example, 45kyo modern human remains from Ust'-Ishim in western Siberia (Fu et al., 2014)) and is thought to be a source of some of the ancestry of the First Americans (Raghavan et al., 2014). The genetic diversity of living populations within Siberia has been sparsely sampled and often focused on uniparental markers. By comparing *H. pylori* diversity in different Siberian

populations alive today, Moodley and colleagues tested a number of models for the origins of the hspIndigenous Americas subpopulation and concluded that the most likely source population was the ancestors of peoples currently found in east and north-east Siberia. *H. pylori* from these peoples then traveled across Beringia into North America around 12kya (Moodley et al., 2021), a slightly younger date than many current estimates for the earliest peopling of the Americas (Ardelean et al., 2020; Becerra-Valdivia and Higham, 2020; Bennett et al., 2021). Studying *H. pylori* strains from a range of living American populations also identifies admixed isolates from across the continent, in common with the admixed history of local human populations (Muñoz-Ramirez et al., 2021).

Broadly, the population structure of *H. pylori* mirrors that of humans (Achtman et al., 1999a; Covacci et al., 1999) (Fig. 8.1). As a bacterium which has putatively been carried out of Africa by modern humans, there was evidence of a linear relationship between genetic diversity in *H. pylori* and geographic distance from east Africa (Linz et al., 2007), which is also true of human genetic diversity. East Africa was modeled as the most likely origin of circulating *H. pylori* strains. Working with

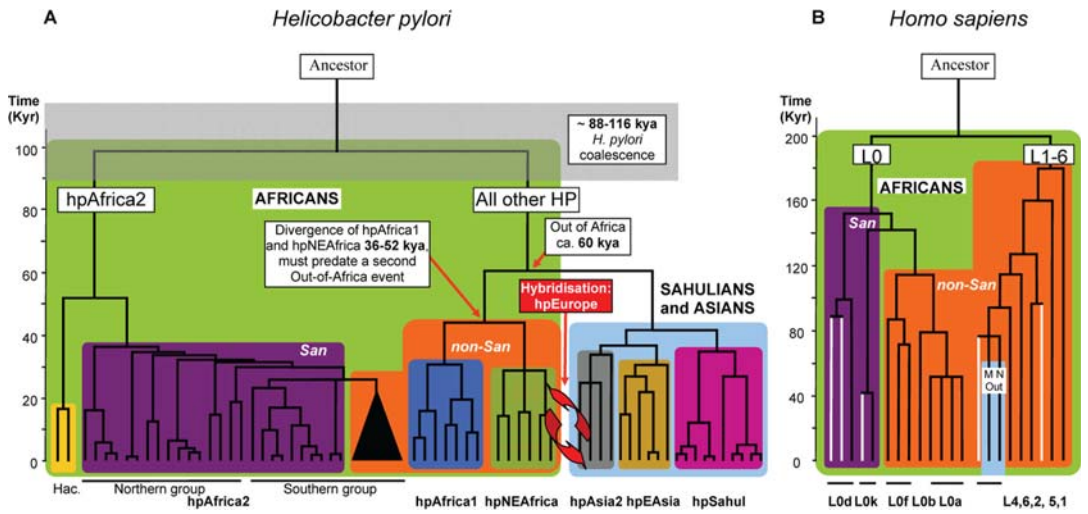


FIGURE 8.1

Global phylogenies of (A) *H. pylori* and related *Helicobacter* species; and (B) human mitochondrial DNA. (A) Phylogeny of *H. pylori*. Human population migrations and demographic events were used as calibration points to derive MRCA estimates of *H. pylori* population divergence times (Moodley et al., 2012). *Helicobacter acinonychis*, originally isolated from cheetahs with gastritis, is shown in yellow. *Helicobacter cetorum*, a pathogen of cetaceans, was used as the outgroup. *H. pylori* lineages associated with African populations are shown on a green background, with San clades further highlighted in purple and non-San Africa clades in orange. Non-African lineages are shown on a light blue background. (B) A human mtDNA phylogeny, adapted originally from (Behar et al., 2008). San-associated mtDNA types are shown in white and mtDNA types associated with human Out-of-Africa migrations have a blue background. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

Creative commons figure reused from Moodley, Y., et al., 2012. Age of the association between helicobacter pylori and man. PLoS Pathog. 8, e1002693. https://doi.org/10.1371/journal.ppat.1002693. https://journals.plos.org/plospathogens/article/figure?id=10.1371/journal.ppat.1002693.g006.

assumptions of 7.9×10^{-7} subs/site/year, and a 25 year human generation time, they modeled the divergence of *H. pylori* from East Africa to have occurred 58kya (Linz et al., 2007).

Human population calibration points were also used to date the divergence of different *H. pylori* types, with particular attention paid to migrations within the Pacific region (including Sahul—Papua New Guinea and Australia). This analysis suggested that the deepest divergences between *H. pylori* populations had occurred in the Pleistocene, with *H. pylori* lineages isolated from San individuals diverging between 88 and 116kya. The Sahul (PNG and Australia) lineages had diverged from Asian lineages 31−37kya (Moodley et al., 2009). The human Out of Africa migration was set at around 55kya in this study, and as more evidence accrues that this migration took place earlier (for example (O'Connell et al., 2018; Rito et al., 2019)), this would in turn push back the dates estimated by Moodley and colleagues (Moodley et al., 2009).

The oldest *H. pylori* genome sequenced to date (Maixner et al., 2016) comes from the mummified stomach of the Tyrolean Ice Man, found in glacier in the Eastern Italian Alps 5300 years after his death (Seidler et al., 1992). This genome is notable for being of the hpAsia2 type (putative descendants of Ancestral European AE1 strains), which is rarely found in Europe today; the Ice Man's *H. pylori* genome also contained <15% of the AE component more common of modern European isolates (Maixner et al., 2016). With only a single ancient genome available and very few SNP typed ancient samples (Maixner et al., 2019), this makes broad conclusions hard to draw. Previous hypotheses that AE1 and AE2 recombined to form hpEurope 52−10kya in the Fertile Crescent, before it was carried into Europe with Neolithic farmers, seem less likely given the Ice Man's genetic ancestry, age and the hpAsia2 infection he carried. This raises the possibility of a post-Neolithic introduction of AE2 from Africa to Europe (Maixner et al., 2016, 2019), with replacement of hpAsia2 over the course of several thousand years (Moodley et al., 2021).

Helicobacter pylori has many desirable characteristics of a marker of human demography, and seems a powerful tool to complement studies of population persistence in Siberia over the last 45ky and human migration in the Americas, alongside human genetic data. However, selection pressure from antibiotic treatment and evidence of recombination undoubtedly complicate any interpretations drawn from *H. pylori* genomic data alone. If further ancient *H. pylori* genomes can be sequenced, direct dating of *H. pylori* population divergences may become possible. For other human bacterial infections, such as *Mycobacterium tuberculosis* and *Mycobacterium leprae*, this has significantly challenged our interpretation of studies drawing on modern data alone (discussed below). It is currently unknown to what extent this will also apply to *H. pylori*.

The falling incidence of *H. pylori* in developed regions (Bureš et al., 2012; Inoue, 2017; van Blankenstein et al., 2013; Vyse et al., 2002), in part due to increasing global consumption of antimicrobial drugs (Klein et al., 2021) will contribute to changing patterns of *H. pylori* genetic diversity, creating a need to sample diversity before it is lost. Recent studies identifying other gut microbes that are human-specific, spread mother-to-child and geographically structured suggests a valuable research direction to complement studies of *H. pylori* (Suzuki et al., 2021).

3.2 Mycobacterium tuberculosis

Mycobacterium tuberculosis (MTB) caused 1.5 million deaths in 2018, more than malaria or HIV/AIDS (Global Tuberculosis Report 2019, n.d.; Harding, 2020). MTB is an acid-fast bacillus and is part of the *Mycobacterium tuberculosis* complex (MTBC), a growing group of related mycobacterial

species and subspecies capable of causing tuberculosis in humans and/or animals (Orgeur and Brosch, 2018). MTB is also an ancient human pathogen, although the minimum age of association is subject to debate. This has complicated its use as a tracer for human migrations, particularly in the last decade (Bos et al., 2014; Comas et al., 2013). Transmission of MTBC bacilli is airborne, as infected individuals breathe and cough out infectious aerosols; these are then inhaled by naïve hosts and the tubercle bacilli penetrate the lungs (Patterson and Wood, 2019; Yates et al., 2016). It was estimated that in 2019, 25% of the world's population was latently infected with TB, with the potential to progress to active infection later in life in 5%−10% of cases (Cohen et al., 2019). In many developing countries, the prevalence of TB is much higher, as is the burden of drug-resistant TB and mortality. Because TB is often acquired in childhood, transmitted within families and has the capacity for latency, MTBC meets many of the criteria set out for a suitable microbial tracer of human migrations.

One of the reasons that MTBC has been studied from an evolutionary and migrator perspective is because the population structure is clonal: recombination is rare (Chiner-Oms et al., 2019). While early studies of MTBC population diversity found little genetic diversity with which to resolve the phylogeny of these lineages, this was likely owing to insufficient sample sizes and low marker resolution (for example (Sreevatsan et al., 1997)). Studying MTBC at whole-genome resolution has identified sufficient genetic diversity to understand the evolution and history of MTBC lineages in much greater detail, and revealed extensive geographical structure within MTBC. Lineages 1−4 and 7 are human adapted and constitute *M. tuberculosis sensu stricto*; lineages 5 and 6 are highly prevalent in West Africa and are human adapted but referred to as *M. africanum*, with lineage 9 (identified as originating in East Africa) suggested as a third member of *M. africanum* (Coscolla et al., 2021); basal lineage 8 was recently identified in two humans cases from Rwanda and Uganda (Ngabonziza et al., 2020); and there are nine animal-adapted ecotypes. Africa is the only continent on which all MTBC lineages are present, suggesting an African origin of MTB as a human pathogen. *Mycobacterium canettii* (alias smooth tubercle bacilli or STB) is also found in Africa, and is usually considered an outgroup to MTBC. *M. canettii* is prone to recombination, unlike the rest of MTBC (Boritsch et al., 2014). MTBC genomes are 4.34−4.43 Mb in length and while strains within the complex are at most ∼2000 SNPs different from one another, *M. canettii* strains are at least 10 times more diverse and seem to represent an evolutionary step toward specialization as a human pathogen, with lower virulence and a possible environmental reservoir where *M. canettii* strains exchange genetic material with one another (Ngabonziza et al., 2020).

Many studies still use mycobacterial interspersed repetitive units (MIRUs) with variable numbers of tandem repeats (VNTRs) in order to characterize cost-effectively the variability of MTBC isolates within or between regions. The availability of data over time and space makes MIRU-VNTR studies a useful way to assess the global origin, timing, and spread of MTBC. These studies suggested that the majority of animal-infecting strains of MTBC were originally of human origin, because the East African-Indian MTBC strains were basal based on phylogenetic analysis. Among the exclusively human strains of MTBC, different regional clusters could be distinguished (African, Asian, African-European, and Latin American-Mediterranean), and a mutation rate was calculated for MIRU-VNTRs of 10^{-4} changes/locus/year. This was consistent with two clades within MTBC emerging around 40kya, suggesting MTB was a pathogen which traveled from Africa with human populations. SNP and indel analyses also identified six lineages of human-associated MTBC which were associated with particular geographic areas and thought to reflect human demographic history (Hershberg et al., 2008), before the genomic identification of lineages 7 and 8.

When using human demographic events as calibration points, the most recent common ancestor (MRCA) of MTBC was estimated to be between 50 and 96kya (Comas et al., 2013). However, using directly dated ancient DNA (aDNA) genomes and tip-dating to derive the MRCA of MTBC, Bos et al. (2014) suggest that the MRCA of circulating MTBC lineages 1−6 and *M. bovis* significantly postdates the Out of Africa migration and may in fact postdate the earliest origins of agriculture (Bos et al., 2014; Menardo et al., 2019), falling within the last 6000 years. Kay and colleagues used ancient genomes to reach similar conclusions, and estimated that lineage 4 strains share an MRCA around 400−500 Common Era (CE), i.e., the late Roman period (Kay et al., 2015), while Sabin and colleagues estimate a slightly earlier origin for lineage 4 and a younger age for the MRCA of all MTBC (Sabin et al., 2020). ADNA analysis of human remains dating to around 1000CE from Peru also identified that in South America, nonhuman-specific mycobacteria may have contributed to the burden of disease, such as *M. pinnipedii* (Bos et al., 2014). This further casts doubt on the idea that human populations carried MTBC out of Africa to the Americas during the Pleistocene. Most studies of MTBC since the inclusion of aDNA-derived genomes now estimate post-Neolithic divergence times for MTBC lineages (Fig. 8.2).

Understanding the evolutionary history of MTB, and accurately estimating the rate of molecular evolution for this pathogen, are important for more than just human demographic and migratory studies. They have clear implications for finding when drug-resistant lineages arrived in a region or began to expand, and dating the beginnings of outbreaks (Menardo et al., 2019). Using human historical or archeological calibration points for the nodes or root within the MTBC phylogeny give very different estimates of the mutation rate, compared with tip dating (sampling date of isolates), differing by as much as two orders of magnitude. This inconsistency between these two methods can be resolved by using historical or ancient samples to calibrate the MTB molecular clock. Sampling only within the last 15−20 years is not sufficient for molecular clock calibration. When using samples collected within the last 40 years or the last 1000 years, rates of 1×10^{-8} to 5×10^{-7} subs/site/year were obtained, although there were differences in rates between clades and lineages (Menardo et al., 2019). Mutation rate estimates which include ancient MTBC genome produce slightly lower rate estimates (Menardo et al., 2019), suggesting that time dependency of clock rates applies to MTBC (Ho et al., 2005, 2011). Until more ancient MTBC genomes are available, resolving the rate at which the molecular clock of MTBC ticks, there will continue to be questions asked about using MTBC genomes as a marker for human migrations. Accurately modeling the effect of latency postinfection, with lower replication rates within the host (Brites and Gagneux, 2017), may help to reconcile some differences between studies but cannot explain the largest differences in rate estimates. There are also difficulties in explaining the epidemiological and archeological evidence for an African origin of MTBC and dissemination to many species, with specialization, within the last 6000 years (Bos et al., 2014; Brites and Gagneux, 2017). A significant question mark currently hangs over the suitability of *Mycobacterium tuberculosis* as a marker of human demography and migration.

3.3 Mycobacterium leprae

Mycobacterium leprae is one of two species, with *Mycobacterium lepromatosis*, that can lead to leprosy (also referred to as Hansen's disease), a disease in which chronic infection damages the nerves, eyes, skin, and respiratory tract. *M. leprae* was once widespread around the world, but has been eliminated in many regions. In 2019, there were more than 200,000 new cases (Leprosy Hansen's disease, n.d.), and India, Brazil, and Indonesia now represent 80% of the cases reported each year (Blok et al., 2015).

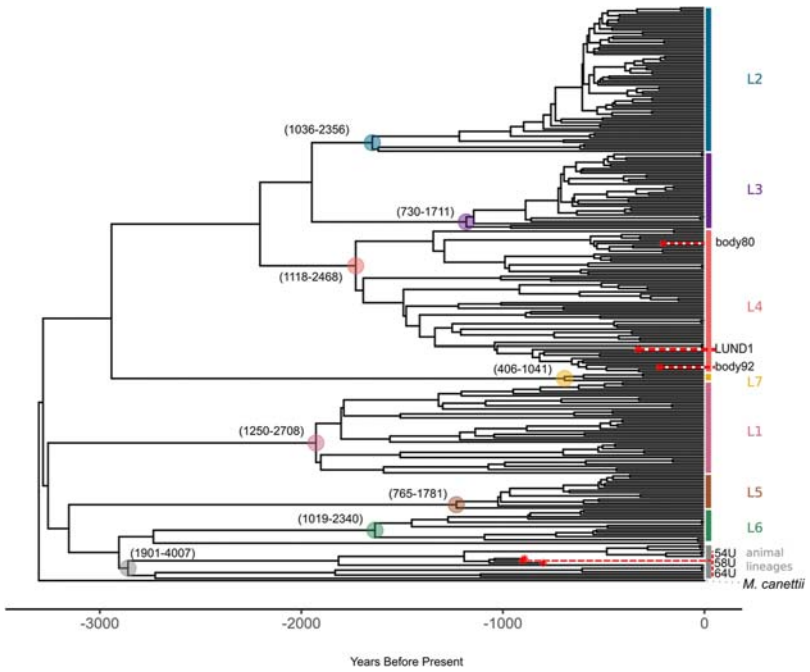


FIGURE 8.2

A maximum clade credibility tree of MTBC genomes, using a BDSKY + UCLD model and *M. canettii* as the outgroup. MTBC lineages are labeled on the right-hand side of the plot. Ancestral nodes for each lineage are shown as colored circles, with 95% highest posterior densities shown for divergence date estimates (*lower* to *upper bound* of years before present).

Creative commons figure reused from Sabin, S., et al., 2020. A seventeenth-century mycobacterium tuberculosis genome supports a neolithic emergence of the mycobacterium tuberculosis complex. Genome Biol. 21, 201. https://doi.org/10.1186/s13059-020-02112-1. https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02112-1/figures/3https://creativecommons.org/licenses/by/4.0/.

M. leprae is highly clonal with a low substitution rate. The *M. leprae* genome is 3.2 million base pairs in length, and has experienced reductive evolution: the genome has over 1000 pseudogenes and is the smallest genome of the mycobacteria, which is likely related to its life cycle as an obligate intracellular pathogen. *M. leprae* is not amenable to culture amplification, which has limited attempts to sequence its genome. This was because genome sequencing formerly required large amounts of genetic material as input. More recent *M. leprae* sequencing efforts have successfully used bait-based target enrichment strategies to sequence *M. leprae* directly from infected material without the need to culture the bacilli in a living organism such as a mouse or armadillo first (Honap et al., 2018; Schuenemann et al., 2018). PCR-based detection of infections has also broadened our knowledge of the epidemiology of *M. leprae* (Inskip et al., 2017).

M. leprae is thought to be most useful as a marker of population migrations for events within the last few thousand years, rather than ancient human migrations during, e.g., the Pleistocene. Hypotheses of *M. leprae* spreading from east Asia (possibly India (Robbins et al., 2009)) to Europe and the Mediterranean basin on the back of war and conquest have been suggested, with links made to the movements of the armies of Alexander the Great and the Roman empire, as well as the associated trade networks of these periods, and the dissemination of *M. leprae* (Mark, 2002). Putative European Bronze Age cases of leprosy, identified on the basis of skeletal changes, raise questions about an Asian origin of *M. leprae* and highlight the need for further study (Roberts, 2020).

Patterns of global *M. leprae* diversity were initially derived from PCR-based SNP typing, identifying 4 types, which were further resolved into 16 subgroups (A to P) (Monot et al., 2009). The using of genomic analysis and ancient DNA has led to the classification of *M. leprae* genomes into six major branches, 0−5. The use of whole genomes from a deeper time sample broke down previous assumptions about a simple correlation between branch or SNP type and geographical regions of the world (Monot et al., 2005). It was also once thought that humans were the sole reservoir species for *M. leprae*, but recent advances in genomics have changed this conclusion. There is evidence from Europe that red squirrels may have been an ancient host for *M. leprae* and are currently a host species for *M. lepromatosis* (Avanzi et al., 2016; Schilling et al., 2019), and these infections may have been a source of anthroponotic-zoonotic transmission cycles in medieval England, particularly for branch 3 (Inskip et al., 2017; Pfrengle et al., 2021). Within Africa, a number of nonhuman primates have also been shown to carry *M. leprae* and suffer human-like leprosy lesions in some cases (Honap et al., 2018).

M. leprae has long been noted to show geographic structure within its genetic diversity which suggested it may be suitable as a marker of human demography, using patterns of diversity tagged by three SNPs (Monot et al., 2009). These SNP-based studies and follow-ups suggested an African origin for *M. leprae*, which was speculated to reflect an Out of Africa initial expansion of *M. leprae* to Europe and Asia, followed by a second introduction in Europe more recently, following the route of the Silk Road trade route. *M. leprae* in the Americas was also speculated to be a post-Columbian introduction from Europe rather than a passenger with the First Peoples of the Americas (Cardona-Castro et al., 2015). Ancient DNA has challenged some of these models, while supporting others. Most ancient *M. leprae* genomes have been recovered from European individuals, and they reveal a considerable diversity of *M. leprae* in Europe in the Medieval period, including the basal branch 0 (Fig. 8.3). This suggests Europe, on the basis of current evidence, as the continent of origin for some branches or subbranches of *M. leprae*, although much wider geographic sampling is needed to assess this robustly. The most likely route by which is *M. leprae* arrived in the Americas is thought to be during the early modern Age of Discovery (Han and Silva, 2014; Monot et al., 2005; Schuenemann et al., 2013). This is supported by SNP based and whole genome analysis of *M. leprae* strains, which show American *M. leprae* lineages as having been common in Europe in the medieval period (Mark, 2017; Pfrengle et al., 2021).

Hypotheses of much earlier migrations of *M. leprae*, e.g., during the Pleistocene, have not been supported by aDNA evidence. Using directly dated ancient *M. leprae* genomes from ancient and contemporary infections suggested a relatively recent divergence *of M. leprae* genomes from a common ancestor 3600 years ago. The substitution rate of *M. leprae* was estimated to be 6.95×10^{-9} subs/site/year. This is at least an order of magnitude younger than estimates of the age of *M. leprae* which assumed a co-divergence with human populations. The presence of nonhuman hosts in Europe

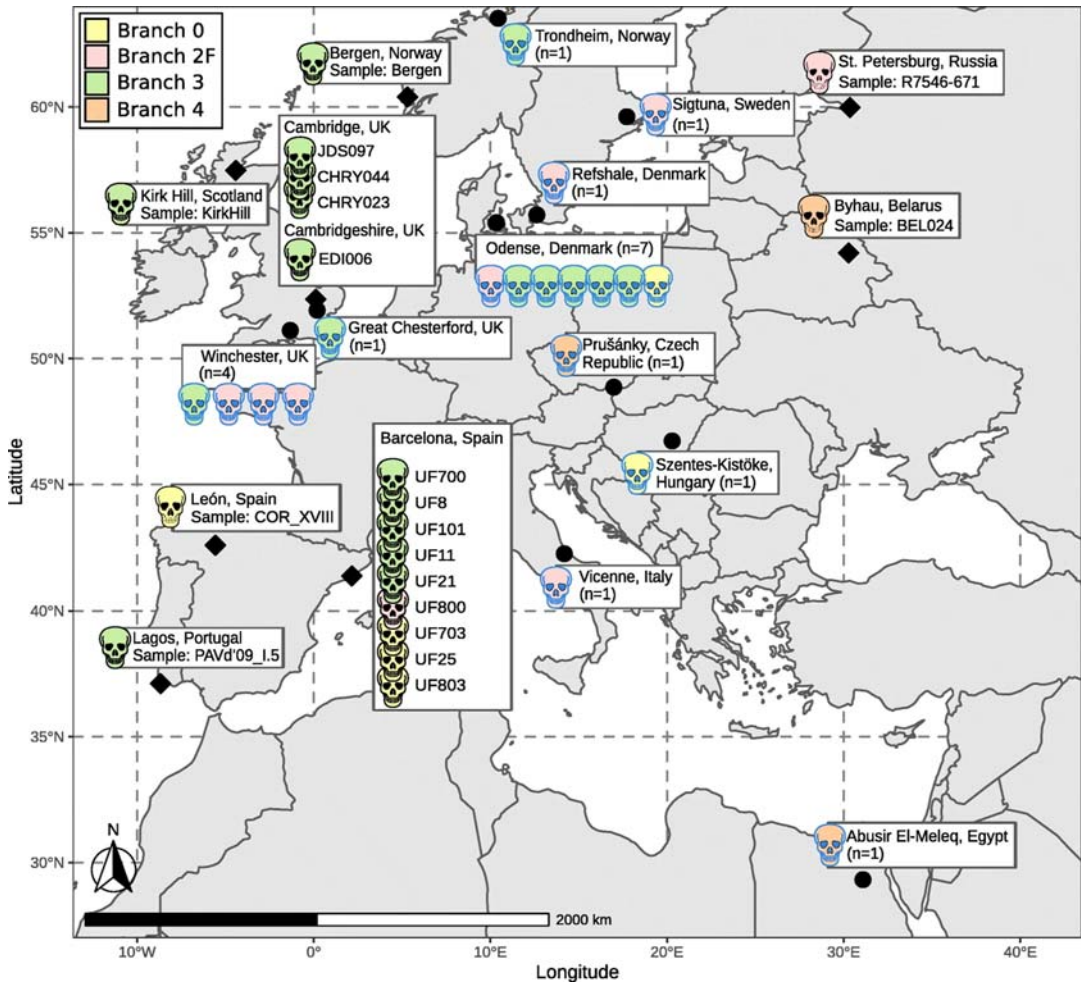


FIGURE 8.3

Map showing the locations and lineage assignments of ancient *M. leprae* genomes. Skull color is used to denote the specific phylogenetic branch assignment of each *M. leprae* genome recovered from ancient human remains at a given site.

Creative commons figure reused from Pfrengle, S., et al., 2021. Mycobacterium leprae diversity and population dynamics in medieval Europe from novel ancient genomes. BMC Biol. 19, 220. https://doi.org/10.1186/s12915-021-01120-2. https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-021-01120-2/figures/1http://creativecommons.org/licenses/by/4.0/.

(squirrels) and Africa and parts of Asia (nonhuman primates) for both leprosy-causing bacilli may muddy attempts to use *M. leprae* as a marker of past human demographic events. Further directly dated *M. leprae* genomes recovered from ancient human remains, particularly from outside Europe, will be a key tool to understand how ancient the association of *M. leprae* with humans is, and to what extent the genome of this bacillus is a faithful record of human demographic events.

4. Viruses

Human viruses are diverse and often highly prevalent, with genomes ranging in size from a few thousand to over 1 million base pairs in length. Humans are infected by viruses with a variety of genome types and structures: double and single stranded DNA or RNA genomes, and viruses that have a reverse-transcriptase capability, such as retroviruses. It is unsurprising, given this diversity, that many viruses are candidate markers for tracing human migrations. The focus has largely been on those viruses which infect for life or for a long period: DNA viruses such as herpesviruses (Aswad et al., 2021), papillomaviruses (Zehender et al., 2016), and polyomaviruses (Torres et al., 2018); partially double-stranded DNA viruses such as hepatitis B virus (Yuen et al., 2019); and retroviruses such as HTLV1 and 2 (Gessain et al., 1995). Many of these viruses are transmitted both horizontally and vertically, and transmission patterns are changing. Apparent co-evolutionary patterns may be the result of natural selection or even coincidence. There is often extensive evidence of recombination in DNA virus genomes. It is therefore important to be cautious in interpreting evidence from these pathogens, but modern and ancient viral DNA still has scope to reveal key insights.

4.1 Human Polyomaviruses

Polyomaviruses are nonenveloped, double-stranded DNA viruses, with small 4–6 kb genomes. Human polyomaviruses are highly prevalent, species-specific, often acquired in childhood and typically only cause disease in settings of immune suppression or immune senescence, i.e., old age. Human polyomaviruses have been extensively studied as potential markers of human migrations, and the scientific consensus over their usefulness in this regard has swung for and against them at different times. As with MTBC and *M. leprae*, the substitution rate of the human polyomaviruses has been estimated using different methods, and these methodologies affect the conclusions drawn from viral genomes. The polyomaviruses most likely to interest scientists from a human history perspective are BK and JC polyomaviruses (BKPyV and JCPyV), and Merkel cell polyomavirus (MCPyV).

JC polyomavirus (JCPyV) is linked to kidney disease in the immunosuppressed, such as after kidney transplant or in untreated HIV infections, where infection can progress to potentially fatal encephalopathy. It is also found, harmlessly, in the renal tissue of healthy people, as a persistent long-term infection with a prevalence of 44%–90% (Moens et al., 2017). JCV can be transmitted both vertically and horizontally, and in healthy people, mixed infection is uncommon (Kato et al., 1997; Kunitake et al., 1995; Suzuki et al., 2002). This in turn reduces the tendency for JCV to recombine, due to a reduction in opportunity. When JCV was sampled from the same individual over a period of up to 6 years, the strains recovered were identical (Kitamura et al., 1997).

Eight genotypes of JCPyV are currently recognized, with further subdivision into subtypes and lineages. Genotype 1 is associated geographically with Europe, East Asia, Siberia, and Arctic North America; genotype 2 with Europe, Asia, the Americas, and Oceania; genotype 3 with Africa, southeastern Europe, and Asia; genotype 4 with Europe and North Africa; genotype 6 with Africa; genotype 7 with Asia and Africa; and genotype 8 with Melanesia and Polynesia. Further types have been suggested on the basis of subgenomic regions only, suggesting further sampling of JCPyV diversity is required. The reference sequence for genotype 5 (AF015684.1) has had its record removed from GenBank "at the submitter's request because the sequence cannot be confirmed" and is likely a recombinant between genotypes 2 and 6.

Analysis of genomic diversity in JCPyV has revealed the geographic structure discussed above, and this has been interpreted as evidence of population-virus co-divergence. This led to the hypothesis that JCPyV migrated out of Africa with anatomically modern humans, and could be used as a marker of human demographic events. Calibrating JCPyV evolution using estimated dates for the major dispersal of modern humans from Africa gave estimated rates of 10^{-7} to 10^{-8} subs/site/year (Buck et al., 2016; Krumbholz et al., 2009; Madinda et al., 2016). The most recent study of the co-evolution of JCPyV and humans used the timings of two separate human population divergences to calibrate the JCPyV molecular clock: the divergence of Northern and Southern Native American populations at 16kya, and the earliest human migration to New Guinea and Australia at 51 kya (Forni et al., 2020a).

JCPyV mutation rates calculated from experimental data or longitudinal sequencing studies have been estimated to be significantly faster than those that assume co-divergence with humans. Using sample collection dates for calibration of the molecular clocks of human polyomaviruses yields very young divergence dates (300−350 years ago for the ancestor of JCPyV genotypes from Native American populations; a TMRCA for all circulating BKPyV of less than 1000 years ago) (Krumbholz et al., 2009; Sharp and Simmonds, 2011). These are not epidemiologically consistent with the high prevalence of the polyomaviruses, or their wide geographical spread. Using sample collection dates for calibration also gives fast substitution rates (approximately 10^{-5} substitutions/site/year (Firth et al., 2010; Kitchen et al., 2008; Shackelton et al., 2006)), more typical of RNA or single-stranded DNA viruses. This may be a reflection of the time-dependence of molecular rate estimates in viruses (Ghafari et al., 2021), which could be tested with aDNA polyomavirus genomes if they become available.

Patterns of geographic substructure seen in JCPyV are to some extent mirrored in Merkel cell polyomavirus. MCPyV is widespread in the healthy population, and prevalence increases with age (Šroller et al., 2014; Viscidi et al., 2011; Zhang et al., 2014). MCPyV is also associated with the majority of cases of Merkel cell carcinoma (an aggressive skin cancer), the tissue in which this virus was first identified (DeCaprio, 2017; Feng et al., 2008). As an oncogenic virus with high prevalence, there is interest in geographical or disease-specific patterns of genetic variation. Torres and colleagues recovered MCPyV from concentrated sewage collected in Argentina, Uruguay, and Spain (Torres et al., 2018). Analysis involved either the VP1 gene, the concatenated VP1-LT region or the complete genome sequence. When analyzed as whole genome sequences, monophyletic groups could be identified, where MCPyV sequences corresponded to the geographical area of sampling. Europe/North America, Africa, Asia, South America, and Oceania all formed distinct clusters (Martel-Jantin et al., 2014; Torres et al., 2018). Pairwise Fst showed a strong signal of subdivision between countries in different geographical regions, but little or no subdivision within geographic regions. Diversity in sequences from South America was thought to reflect genotypes introduced with the First Peoples of the Americas, genotypes originating with forcibly transported African individuals during the trans-Atlantic slave trade, and Europe/North American genotypes from European colonizers or later immigrants (Torres et al., 2018).

Because of the association between MCPyV phylogenetic groups and human populations in specific geographic regions, the substitution rate of MCPyV was calibrated using human demographic events: 50−100 kya for the major human migration out of Africa and the MCRA of MCPyV, 40−45kya for the divergence of the Europe/North America group of MCPyV sequences, and 13−15kya for the divergence of MCPyV genomes from South America. This yielded a genomic rate

for MCPyV of 5.1×10^{-8} subs/site/year (or approximately a ~5.1% divergence/mya, in line with other double-stranded DNA viruses) and an MRCA of circulating MCPyV strains ~72kya. Including samples of MCPyV from Merkel cell carcinoma was noted, in this analysis, to result in a younger estimate of the MRCA of circulating MCPyV (~57kya) and a higher substitution rate (5.8×10^{-8} subs/site/year). This suggests that using cancer-derived viral genome evolution rates to estimate rates of evolution in immune competent hosts may be problematic and lead to underestimates of the age of circulating viral diversity. More African MCPyV genome sequences will be required to establish whether African MCPyV genomes are basal to global diversity or a deeper split between European/North American and African MCPyV, suggested by phylogenetic analysis, reflects a different geographic origin for currently circulating MCPyV genotypes. If these issues of substitution rate and geographic origin can be independently derived or confirmed through ancient DNA analyses, MCPyV may be another suitable viral candidate for tracing human migrations (Torres et al., 2018).

A greater availability of nonhuman primate (especially chimpanzee and bonobo) polyomavirus genomes would also be beneficial to estimating the antiquity and co-divergence of human polyomaviruses, as these would provide alternative external calibration points for molecular clock estimates, as well as being important roots for resolving JCPyV, BKPyV, and MCPyV phylogeography. For example, both *Pan troglodytes verus* polyomavirus 8 (PtrovPyV8) and *Pan troglodytes troglodytes* polyomavirus 1 (PtrotPyV1) show >80% sequence identity to human BKPyV (Deuzing et al., 2010; van Persie et al., 2016). The pathology and genomic content of primate polyomaviruses are an important source of data in their own right to help us understand the diseases associated with human and great ape polyomavirus infections.

4.2 Human Herpesviruses

The nine human herpesviruses have been considered candidates for viral tracers of human evolution because they are globally distributed, many of them are highly prevalent, they cause lifelong infection, and some are strongly associated with vertical transmission or transmission within families. However, genomic analysis of hundreds of strains has shown that many are highly recombinant (HCMV (Cudini et al., 2019), HSV1 and 2 (Forni et al., 2020b; Renner and Szpara, 2018), EBV (Wegner et al., 2019)), making phylogenetic analysis challenging. Some families have evidence of widespread ancient host-switching between primate species (Azab et al., 2018).

Among the human herpesviruses, varicella-zoster virus (VZV) initially showed evidence of being a suitable candidate for tracing human migrations. VZV is the causative agent of both chicken pox (varicella) and shingles (zoster), the clinical manifestations of primary infection and reactivation, respectively. It is the only human herpesvirus currently known to be spread through aerosolization of infectious virus particles (Grose, 2012). Before the introduction of live attenuated vaccines, most children would have been infected at a young age, although the age at first infection can vary by geography and ethnicity (Goh et al., 2019; Pierik et al., 2012; Walker et al., 2017).

VZV genotyping studies detected a geographical structure within VZV isolates from different regions of the world (Breuer, 2018). Early studies used SNP typing to define five VZV clades, with further clades identified using whole-genome sequencing. Clades 1, 3, 6, and 9 are associated with transmission of VZV within Europe, Australia, and North America, although clade 9 is rarely detected (Jensen et al., 2017); clade 2 is associated with Asia; clades 4 and 5 are more cosmopolitan and found in many regions, although clade 5 is the only VZV clade found in most of Africa (Pontremoli et al., 2019).

Putative clades VII and VIII are labeled with Roman numerals until their status is confirmed by further identification; clade VIII may be a subclade of clade 6 (Breuer, 2018; Jensen et al., 2017). The worldwide distribution of VZV, the presence of primate varicella viruses (Mahalingam and Gilden, 2007), VZV's ability to latently infect a host for life suggested it might be a suitable DNA virus to trace human migrations (Wagenaar et al., 2003). Norberg et al. showed that VZV can recombine, and that recombination-aware phylogenetic analysis was needed when analyzing the geographic structure of VZV (Norberg et al., 2015).

The assumption of a co-divergence between human populations and circulating VZV diversity led to the dating of the divergence of the *Alphaherpesvirinae* using a range of calibration points. These included calibrations such as human populations diverging from one another 110kya to much further older events such as the divergence of Hominoidea and Old World monkeys 23 mya. This yielded a rate of 3.9×10^{-9} subs/site/year, suggesting that the modern human VZV clades were diverged by 12kya, the onset of the Neolithic (Zell et al., 2012). However, estimates of VZV substitution rates derived from other human alphaherpesviruses yielded a younger MRCA for the major VZV clades of between 3 and 19 kya (Muir et al., 2002). Muir and colleagues further estimated that the lowest the VZV substitution rate could be was 1.4×10^{-8} subs/site/year, but thought 1×10^{-7} was more plausible based on data from other human herpesviruses. ADNA analysis of four Eurasian ancient HSV1 genomes from the last 2000 years has yielded an estimated substitution rate of 2.5×10^{-6} to 2.4×10^{-7} subs/site/year (Guellil et al., 2022), suggesting some estimates of alphaherpesvirus substitution rates may be too low.

In vivo studies of VZV that take account of within (Weinert et al., 2015) and between-host (Depledge et al., 2014) evolution, including the effects of latency, have estimated VZV substitution rates which are too fast to reflect an ancient Out of Africa migration. They studied only regions of the viral genome with no evidence of recombination (Weinert et al., 2015). Estimates of the substitution rate of VZV are complicated by the finding that the viral genome experiences a mutation rate during initial infection of $\sim 3.2 \times 10^{-7}$ subs/site/day, but once the virus becomes latent, no further mutations arise until reactivation. Setting the time between infection and reactivation to 50 years, Weinert et al. calculated that the most recent common ancestor of currently circulating VZV could not be more than 8kya. Weinert and colleagues also demonstrated that "implausibly long latency periods are required to date the most recent common ancestor of extant VZV to an 'out-of-Africa' migration with humans, as has been previously suggested" (Weinert et al., 2015). Pontremoli et al. further showed that European clade 3 VZV genomes are less diverged from the putative MRCA of VZV than other clades found elsewhere, suggesting Europe (rather than Africa) as the home of the MRCA of circulating VZV. Thus, despite the degree of geographical clustering in VZV genomes which appears to match an Out of Africa dispersal, it seems that this structure is a red herring: the VZV genomes infecting children today share a much more recent common ancestor than human populations, and VZV genomes may also be of European rather than African origin (Pontremoli et al., 2019). If ancient VZV diversity has been replaced by much more recent VZV diversity originating in Europe, this suggests that VZV is not suitable for use as a marker of ancient human migrations. There are notable parallels with findings from analysis of ancient HSV1 genomes (Guellil et al., 2022). The conundrum of VZV evolution and worldwide dissemination will most likely only be resolved by global ancient DNA sampling strategies.

What of the other human herpesviruses? Among the betaherpesviruses, HCMV is an unsuitable candidate because of its tendencies toward recombination (Cudini et al., 2019). Human herpesviruses 6A (HHV6A) and 6B (HHV6B), however, have more scope as markers of human demography and

migration. HHV6A and HHV6B are Roseolaviruses, once thought to be a single species, now shown to have 90% interspecies genome sequence identify and formally recognized as separate species. Natural infection by circulating HHV6A and/or B occurs in high proportion of the population, and both viruses are frequently detected in the saliva of healthy carriers (Levy et al., 1990; Magalhães et al., 2010). Following primary infection, HHV6A and B have the unusual ability among human herpesviruses of being able to able to integrate into human telomeres as part of their infection and latency cycle, a phenomenon which has been also been reported for the less well-studied HHV7 (Houldcroft, 2019; Prusty et al., 2017). If HHV6A/B telomeric integration takes place in a germ cell, there is the scope for this integration to be passed on to offspring, inherited as a chromosomally integrated HHV6 copy (iciHHV6) (Arbuckle et al., 2010). HHV6A shows a stronger signal of geographic structure than HHV6B, which has more evidence of recombinant genomic regions than HHV6A (Telford et al., 2018). HHV6A also shows a clearer division in phylogenies between iciHHV6A and circulating HHV6A (Aswad et al., 2021).

In their ici forms, HHV6A and B may have scope for tracing human migrations, as more genomes become available (Houldcroft, 2019). Integration into the telomeres means that HHV6A and B are assumed to mutate at the same rate as the human genome (0.5×10^{-9} subs/site/year (Scally and Durbin, 2012)). iciHHV6A and B have fewer opportunities for recombination (post integration) than circulating HHV6A and B. Sporadic germline integration events followed by vertical transmission are rare (they are carried by less than 1% of people) but may reveal genetic bottlenecks and founder events in relatively deep time (e.g., over a 300−20kya timescale) (Aswad et al., 2021; Wood et al., 2021; Zhang et al., 2017). For example, carriers of group 3 iciHHV-6B share a common ancestor who lived before the Last Glacial Maximum ∼24.5kya but are found today across Europe, in Sardinia, England, Wales, and the Orkney Isles (Zhang et al., 2017). Another iciHHV6B type is found today in individuals who identify as Maasai Kenyan, Pakistani, and Native American and is dated to between 85 and 342kya. This suggests this specific integration event occurred in a common ancestor of these populations and predates (perhaps significantly) the human Out of Africa migration (Aswad et al., 2021). The same study found that the mean age of HHV6A integration events was 31−124kya. Further sequencing of iciHHV6A/B genomes will reveal whether these rare inherited copies can help us to understand human demography in greater resolution.

5. Conclusion

This chapter can only give a brief overview of the field of pathogens as tracers of human demography, both discoveries and controversies. This is a field that arguably grows more complex as more viral and bacterial genomes are sequenced. Ancient pathogen DNA has revolutionized the field by allowing the direct dating of microbial genomes and thus calibration of their molecular clocks without reliance on external (human) time points or laboratory estimates of mutation rates. Similarly, pathogen genomics (especially for neglected pathogens) is rapidly moving from isolated genomes to true, population-scale surveys, which will require both learning from human population genomics and the development of specialist analysis tools for genomes that evolve rather differently, be it with conspicuous levels of recombination (such as human cytomegalovirus (Cudini et al., 2019)) or remarkable clonality (for example, *M. leprae* (Schuenemann et al., 2013) and MTBC (Brites and Gagneux, 2017)).

Among bacteria, *Helicobacter pylori* is perhaps the best candidate for understanding human demography and migration at the present time. Its diversity, geographical structure, growing number of whole genome sequences and the possibility of further ancient genomes becoming available all speak in its favor. Understanding *H. pylori* phylogeography and history will complement the growing availability of human genetic data from regions with complex and contentious population histories, such as Oceania, Siberia, and the Americas. Its status as a cancer-associated pathogen is both its strength and its weakness. There is considerable interest in the genome of *H. pylori* and whether its genetic variation relates to its oncogenic potential, which means that the number of genomes available from around the world is likely to grow. At the same time, because *H. pylori* is potentially eradicable with antimicrobial therapy, worldwide diversity is being lost and the pace of this loss can be expected to increase as global antimicrobial consumption grows.

H. pylori also suffers, along with many of the pathogens discussed in this chapter, from a necessary reliance on using experimentally derived mutation rate estimates or else calibration of dating estimates using human demographic events (Comas et al., 2013; Krumbholz et al., 2008; Moodley et al., 2021; Torres et al., 2018). If ancient genomes are available, it is important to integrate them into both dating estimates and models of *H. pylori* evolution (Maixner et al., 2016). Our interpretations of the evolutionary histories of human mycobacteria have been fundamentally changed by ancient pathogen genomes and the same may be true of *H. pylori*: only time will tell.

M. leprae and *M. tuberculosis* seem, from ancient DNA estimates of their substitution rates, too young to be pathogens that humans carried Out of Africa and therefore less than ideal for tracking human migrations on ancient timescales. The antiquity of *M. leprae* and *M. tuberculosis* as human infections will remain a topic of debate even as more genomes are sequenced, revealing unexpected extinct (Bos et al., 2014) and extant (Ngabonziza et al., 2020) diversity. It will be challenging to reconcile ancient archaeological evidence of disease with recent divergence dates derived from molecular data without advances in aDNA recovery and the study of ancient biomolecules (Stone et al., 2020). However, with directly dated ancient *M. leprae* and MTBC genomes, we may able to track complex human migrations and interactions since the onset of the Neolithic, such as Europe's role as a putative hub for *M. leprae* dissemination to other regions (Pfrengle et al., 2021).

Turning to viruses, a strong candidate for a marker of human demography is not immediately available. Controversy surrounding the substitution rates of endemic human viruses currently limits their usefulness as demographic tracers (Krause-Kyora et al., 2018; Shackelton et al., 2006) in a similar manner to mycobacteria. Patterns of VZV genetic diversity appear too young to reflect the earliest expansions of humans from Africa (Pontremoli et al., 2019; Weinert et al., 2015), raising the question of when and by what routes current patterns of VZV diversity arose. The possibility that modern circulating diversity expanded from Europe may have parallels with evidence of ancient *M. leprae* diversity in Europe, which warrant further analysis. Dating the emergence of human polyomavirus clades is also challenged by uncertainty around estimates of the mutation rate for JCPyV, BKPyV, and MCPyV and their closest primate relatives (Madinda et al., 2016; Shackelton et al., 2006; Torres et al., 2018). It seems likely that only ancient DNA analysis and better sampling of modern diversity at a genomic level will reveal whether any of the herpesviruses or polyomaviruses can be used to understand human migrations. The first ancient human herpesvirus genomes from Pleistocene human remains testify to their antiquity as human infections, but are not yet of sufficiently high coverage to allowing accurate dating estimates (Nielsen et al., 2021).

With all these caveats and unanswered questions in mind, it is still the case that understanding the evolutionary history of human pathogens is a worthwhile endeavor in and of itself. Studying pathogen genomic evolution and diversity opens a window into the past, showing the mechanisms by which pathogens adapt and diversify within the human niche. Pathogen genome analysis also points to the ways in which emerging infectious diseases may evolve, and the classes of pathogens with the greatest potential for human disease in the future (Cousins, 2018; Gouglas et al., 2019). Following multiple Ebola outbreaks and the COVID pandemic, humans need to understand infectious disease ecology and evolution as much now as ever.

Acknowledgments

The author would like to thank Dr. Jordan Skittrall for providing critical reading; the author would also like to acknowledge J.-P. Rasigade, A. Gilabert, and T. Wirth, the authors of previous versions of this chapter (Gilabert and Wirth, 2011; Rasigade et al., 2017).

References

Achtman, M., Azuma, T., Berg, D.E., Ito, Y., Morelli, G., Pan, Z.J., Suerbaum, S., Thompson, S.A., Van Der Ende, A., Van Doorn, L.J., 1999a. Recombination and clonal groupings within *Helicobacter pylori* from different geographical regions. Mol. Microbiol. 32, 459−470. https://doi.org/10.1046/j.1365-2958.1999.01382.x.

Achtman, M., Zurth, K., Morelli, G., Torrea, G., Guiyoule, A., Carniel, E., 1999b. *Yersinia pestis*, the cause of plague, is a recently emerged clone of Yersinia pseudotuberculosis. Proc. Natl. Acad. Sci. U. S. A. 96, 14043−14048. https://doi.org/10.1073/pnas.96.24.14043.

Arbuckle, J.H., Medveczky, M.M., Luka, J., Hadley, S.H., Luegmayr, A., Ablashi, D., Lund, T.C., Tolar, J., De Meirleir, K., Montoya, J.G., Komaroff, A.L., Ambros, P.F., Medveczky, P.G., 2010. The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. Proc. Natl. Acad. Sci. 107, 5563−5568. https://doi.org/10.1073/pnas.0913586107.

Ardelean, C.F., Becerra-Valdivia, L., Pedersen, M.W., Schwenninger, J.-L., Oviatt, C.G., Macías-Quintero, J.I., Arroyo-Cabrales, J., Sikora, M., Ocampo-Díaz, Y.Z.E., Rubio-Cisneros, I.I., Watling, J.G., Medeiros, V.B. de, Oliveira, P.E. De, Barba-Pingarón, L., Ortiz-Butrón, A., Blancas-Vázquez, J., Rivera-González, I., Solís-Rosales, C., Rodríguez-Ceja, M., Gandy, D.A., Navarro-Gutierrez, Z., Rosa-Díaz, J.J.D. La, Huerta-Arellano, V., Marroquín-Fernández, M.B., Martínez-Riojas, L.M., López-Jiménez, A., Higham, T., Willerslev, E., 2020. Evidence of human occupation in Mexico around the last glacial maximum. Nat 2020, 1−6. https://doi.org/10.1038/s41586-020-2509-0.

Aswad, A., Aimola, G., Wight, D., Roychoudhury, P., Zimmermann, C., Hill, J., Lassner, D., Xie, H., Huang, M.-L., Parrish, N.F., Schultheiss, H.-P., Venturini, C., Lager, S., Smith, G.C.S.S., Charnock-Jones, D.S., Breuer, J., Greninger, A.L., Kaufer, B.B., 2021. Evolutionary history of endogenous human herpesvirus 6 reflects human migration out of Africa. Mol. Biol. Evol. 38, 96−107. https://doi.org/10.1093/molbev/msaa190.

Avanzi, C., Del-Pozo, J., Benjak, A., Stevenson, K., Simpson, V.R., Busso, P., McLuckie, J., Loiseau, C., Lawton, C., Schoening, J., Shaw, D.J., Piton, J., Vera-Cabrera, L., Velarde-Felix, J.S., McDermott, F., Gordon, S.V., Cole, S.T., Meredith, A.L., 2016. Red squirrels in the British Isles are infected with leprosy bacilli. Science 354, 744−747. https://doi.org/10.1126/science.aah3783, 80-.

Azab, W., Dayaram, A., Greenwood, A.D., Osterrieder, N., 2018. How host specific are herpesviruses? Lessons from herpesviruses infecting wild and endangered mammals. Annu. Rev. Virol. 5, 53−68. https://doi.org/10.1146/annurev-virology-092917-043227.

Becerra-Valdivia, L., Higham, T., 2020. The timing and effect of the earliest human arrivals in North America. Nature 584, 93−97. https://doi.org/10.1038/s41586-020-2491-6.

Behar, D.M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., Scozzari, R., Makkan, H., Tzur, S., Comas, D., Bertranpetit, J., Quintana-Murci, L., Tyler-Smith, C., Wells, R.S., Rosset, S., 2008. The dawn of human matrilineal diversity. Am. J. Hum. Genet. 82, 1130−1140. https://doi.org/10.1016/j.ajhg.2008.04.002.

Bennett, M.R., Bustos, D., Pigati, J.S., Springer, K.B., Urban, T.M., Holliday, V.T., Reynolds, S.C., Budka, M., Honke, J.S., Hudson, A.M., Fenerty, B., Connelly, C., Martinez, P.J., Santucci, V.L., Odess, D., 2021. Evidence of humans in north America during the last glacial maximum. Science 373, 1528−1531. https://doi.org/10.1126/science.abg7586, 80-.

Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., Blanché, H., Deleuze, J.F., Cann, H., Mallick, S., Reich, D., Sandhu, M.S., Skoglund, P., Scally, A., Xue, Y., Durbin, R., Tyler-Smith, C., 2020. Insights into human genetic variation and population history from 929 diverse genomes. Science 367. https://doi.org/10.1126/science.aay5012, 80-.

Bergström, A., Stringer, C., Hajdinjak, M., Scerri, E.M.L., Skoglund, P., 2021. Origins of modern human ancestry. Nature. https://doi.org/10.1038/s41586-021-03244-5.

Beyer, R.M., Krapp, M., Eriksson, A., Manica, A., 2021. Climatic windows for human migration out of Africa in the past 300,000 years. Nat. Commun. 12, 1−10. https://doi.org/10.1038/s41467-021-24779-1.

Blok, D.J., De Vlas, S.J., Richardus, J.H., 2015. Global elimination of leprosy by 2020: are we on track? Parasit. Vectors 8, 548. https://doi.org/10.1186/s13071-015-1143-4.

Boritsch, E.C., Supply, P., Honoré, N., Seeman, T., Stinear, T.P., Brosch, R., 2014. A glimpse into the past and predictions for the future: the molecular evolution of the tuberculosis agent. Mol. Microbiol. https://doi.org/10.1111/mmi.12720.

Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., Campbell, T.J., Majander, K., Wilbur, A.K., Guichon, R.A., Wolfe Steadman, D.L., Cook, D.C., Niemann, S., Behr, M.A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J.E., Gagneux, S., Stone, A.C., Krause, J., 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514, 494−497. https://doi.org/10.1038/nature13591.

Breuer, J., 2018. Molecular genetic insights into varicella zoster virus (VZV), the vOka vaccine strain, and the pathogenesis of latency and reactivation. J. Infect. Dis. 218, S75−S80. https://doi.org/10.1093/infdis/jiy279.

Brites, D., Gagneux, S., 2017. The nature and evolution of genomic diversity in the mycobacterium tuberculosis Complex. In: Advances in Experimental Medicine and Biology. Springer New York LLC, pp. 1−26. https://doi.org/10.1007/978-3-319-64371-7_1.

Buck, C.B., Van Doorslaer, K., Peretti, A., Geoghegan, E.M., Tisza, M.J., An, P., Katz, J.P., Pipas, J.M., McBride, A.A., Camus, A.C., McDermott, A.J., Dill, J.A., Delwart, E., Ng, T.F.F., Farkas, K., Austin, C., Kraberger, S., Davison, W., Pastrana, D.V., Varsani, A., 2016. The ancient evolutionary history of polyomaviruses. PLoS Pathog. 12. https://doi.org/10.1371/journal.ppat.1005574.

Bureš, J., Kopáčová, M., Koupil, I., Seifert, B., Fendrichová, M.Š., Špirková, J., Vořišek, V., Rejchrt, S., Douda, T., Král, N., Tachecí, I., 2012. Significant decrease in prevalence of *Helicobacter pylori* in the Czech Republic. World J. Gastroenterol. 18, 4412−4418. https://doi.org/10.3748/wjg.v18.i32.4412.

Burrel, S., Boutolleau, D., Ryu, D., Agut, H., Merkel, K., Leendertz, F.H., Calvignac-Spencer, S., 2017. Ancient recombination events between human herpes simplex viruses. Mol. Biol. Evol. 25, 1910−1920. https://doi.org/10.1093/molbev/msx113.

Byrne, R.P., Martiniano, R., Cassidy, L.M., Carrigan, M., Hellenthal, G., Hardiman, O., Bradley, D.G., McLaughlin, R.L., 2018. Insular Celtic population structure and genomic footprints of migration. PLoS Genet. 14, e1007152. https://doi.org/10.1371/journal.pgen.1007152.

Cardona-Castro, N., Cortés, E., Beltrán, C., Romero, M., Badel-Mogollón, J.E., Bedoya, G., 2015. Human genetic ancestral composition correlates with the origin of *Mycobacterium leprae* strains in a leprosy endemic population. PLoS Negl. Trop. Dis. 9, e0004045. https://doi.org/10.1371/journal.pntd.0004045.

Cavalli-Sforza, L.L., Feldman, M.W., 2003. The application of molecular genetic approaches to the study of human evolution. Nat. Genet. https://doi.org/10.1038/ng1113.

Chiner-Oms, Sánchez-Busó, L., Corander, J., Gagneux, S., Harris, S.R., Young, D., González-Candelas, F., Comas, I., 2019. Genomic determinants of speciation and spread of the *Mycobacterium tuberculosis* complex. Sci. Adv. 5. https://doi.org/10.1126/sciadv.aaw3307.

Choin, J., Mendoza-Revilla, J., Arauna, L.R., Cuadros-Espinoza, S., Cassar, O., Larena, M., Ko, A.M.S., Harmant, C., Laurent, R., Verdu, P., Laval, G., Boland, A., Olaso, R., Deleuze, J.F., Valentin, F., Ko, Y.C., Jakobsson, M., Gessain, A., Excoffier, L., Stoneking, M., Patin, E., Quintana-Murci, L., 2021. Genomic insights into population history and biological adaptation in Oceania. Nature 592, 583−589. https://doi.org/10.1038/s41586-021-03236-5.

Cohen, A., Mathiasen, V.D., Schön, T., Wejse, C., 2019. The global prevalence of latent tuberculosis: a systematic review and meta-analysis. Eur. Respir. J. https://doi.org/10.1183/13993003.00655-2019.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S., 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat. Genet. 45, 1176−1182. https://doi.org/10.1038/ng.2744.

Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., Otchere, I.D., Asante-Poku, A., Asare, P., Sánchez-Busó, L., Gehre, F., Sanoussi, C.N., Antonio, M., Affolabi, D., Fyfe, J., Beckert, P., Niemann, S., Alabi, A.S., Grobusch, M.P., Kobbe, R., Parkhill, J., Beisel, C., Fenner, L., Böttger, E.C., Meehan, C.J., Harris, S.R., de Jong, B.C., Yeboah-Manu, D., Brites, D., 2021. Phylogenomics of mycobacterium africanum reveals a new lineage and a complex evolutionary history. Microb. Genomics 7, 1−14. https://doi.org/10.1099/mgen.0.000477.

Cousins, S., 2018. WHO hedges its bets: the next global pandemic could be disease X. BMJ. https://doi.org/10.1136/bmj.k2015.

Covacci, A., Telford, J.L., Del Giudice, G., Parsonnet, J., Rappuoli, R., 1999. *Helicobacter pylori* virulence and genetic geography. Science. https://doi.org/10.1126/science.284.5418.1328, 80.

Cudini, J., Roy, S., Houldcroft, C.J., Bryant, J.M., Depledge, D.P., Tutill, H., Veys, P., Williams, R., Worth, A.J.J., Tamuri, A.U., Goldstein, R.A., Breuer, J., 2019. Human cytomegalovirus haplotype reconstruction reveals high diversity due to superinfection and evidence of within-host recombination. Proc. Natl. Acad. Sci. 116, 5693−5698. https://doi.org/10.1073/pnas.1818130116.

DeCaprio, J.A., 2017. Merkel cell polyomavirus and Merkel cell carcinoma. Philos. Trans. R. Soc. B Biol. Sci. https://doi.org/10.1098/rstb.2016.0276.

Demeure, C.E., Dussurget, O., Mas Fiol, G., Le Guern, A.-S., Savin, C., Pizarro-Cerdá, J., 2019. *Yersinia pestis* and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. Genes Immun. 20, 357−370. https://doi.org/10.1038/s41435-019-0065-0.

Depledge, D.P., Gray, E.R., Kundu, S., Cooray, S., Poulsen, A., Aaby, P., Breuer, J., 2014. Evolution of cocirculating varicella-zoster virus genotypes during a chickenpox outbreak in Guinea-Bissau. J. Virol. 88, 13936−13946. https://doi.org/10.1128/JVI.02337-14.

Deuzing, I., Fagrouch, Z., Groenewoud, M.J., Niphuis, H., Kondova, I., Bogers, W., Verschoor, E.J., 2010. Detection and characterization of two chimpanzee polyomavirus genotypes from different subspecies. Virol. J. 7, 347. https://doi.org/10.1186/1743-422X-7-347.

Devi, S.M., Ahmed, I., Khan, A.A., Rahman, S.A., Alvi, A., Sechi, L.A., Ahmed, N., 2006. Genomes of *Helicobacter pylori* from native Peruvians suggest admixture of ancestral and modern lineages and reveal a western type cag-pathogenicity island. BMC Genom. 7, 191. https://doi.org/10.1186/1471-2164-7-191.

Didelot, X., Nell, S., Yang, I., Woltemate, S., Van Der Merwe, S., Suerbaum, S., 2013. Genomic evolution and transmission of *Helicobacter pylori* in two South African families. Proc. Natl. Acad. Sci. U. S. A. 110, 13880−13885. https://doi.org/10.1073/pnas.1304681110.

Durvasula, A., Sankararaman, S., 2020. Recovering signals of ghost archaic introgression in African populations. Sci. Adv. 6.

Falush, D., Wirth, T., Linz, B., Pritchard, J.K., Stephens, M., Kidd, M., Blaser, M.J., Graham, D.Y., Vacher, S., Perez-Perez, G.I., Yamaoka, Y., Mégraud, F., Otto, K., Reichard, U., Katzowitsch, E., Wang, X., Achtman, M., Suerbaum, S., 2003. Traces of human migrations in *Helicobacter pylori* populations. Science 299, 1582−1585. https://doi.org/10.1126/science.1080857, 80.

Feng, H., Shuda, M., Chang, Y., Moore, P.S., 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 319, 1096−1100. https://doi.org/10.1126/science.1152586, 80.

Firth, C., Kitchen, A., Shapiro, B., Suchard, M.A., Holmes, E.C., Rambaut, A., 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol. Biol. Evol. 27, 2038−2051. https://doi.org/10.1093/molbev/msq088.

Flegontov, P., Altınışık, N.E., Changmai, P., Rohland, N., Mallick, S., Adamski, N., Bolnick, D.A., Broomandkhoshbacht, N., Candilio, F., Culleton, B.J., Flegontova, O., Friesen, T.M., Jeong, C., Harper, T.K., Keating, D., Kennett, D.J., Kim, A.M., Lamnidis, T.C., Lawson, A.M., Olalde, I., Oppenheimer, J., Potter, B.A., Raff, J., Sattler, R.A., Skoglund, P., Stewardson, K., Vajda, E.J., Vasilyev, S., Veselovskaya, E., Hayes, M.G., O'Rourke, D.H., Krause, J., Pinhasi, R., Reich, D., Schiffels, S., 2019. Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. Nature 570, 236−240. https://doi.org/10.1038/s41586-019-1251-y.

Forni, D., Cagliani, R., Clerici, M., Pozzoli, U., Sironi, M., 2020a. You will never walk alone: codispersal of JC polyomavirus with human populations. Mol. Biol. Evol. 37, 442−454. https://doi.org/10.1093/molbev/msz227.

Forni, D., Pontremoli, C., Clerici, M., Pozzoli, U., Cagliani, R., Sironi, M., 2020b. Recent out-of-Africa migration of human herpes simplex viruses. Mol. Biol. Evol. 37, 1259−1271. https://doi.org/10.1093/molbev/msaa001.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S.M., Bondarev, A.A., Johnson, P.L.F.F., Aximu-Petri, A., Prüfer, K., de Filippo, C., Meyer, M., Zwyns, N., Salazar-García, D.C., Kuzmin, Y.V., Keates, S.G., Kosintsev, P.A., Razhev, D.I., Richards, M.P., Peristov, N.V., Lachmann, M., Douka, K., Higham, T.F.G.G., Slatkin, M., Hublin, J.-J.J., Reich, D., Kelso, J., Viola, T.B., Pääbo, S., 2014. Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514, 445−449. https://doi.org/10.1038/nature13810.

Gessain, A., Mauclere, P., Froment, A., Biglione, M., Le Hesran, J.Y., Tekaia, F., Millan, J., De The, G., 1995. Isolation and molecular characterization of a human T-cell lymphotropic virus type II (HTLV-II), subtype B, from a healthy Pygmy living in a remote area of Cameroon: an ancient origin for HTLV-II in Africa. Proc. Natl. Acad. Sci. U. S. A. 92, 4041−4045. https://doi.org/10.1073/pnas.92.9.4041.

Ghafari, M., Simmonds, P., Pybus, O.G., Katzourakis, A., 2021. A mechanistic evolutionary model explains the time-dependent pattern of substitution rates in viruses. Curr. Biol. https://doi.org/10.1016/j.cub.2021.08.020.

Gilabert, A., Wirth, T., 2011. Elucidating human migrations by means of their pathogens. In: Genetics and Evolution of Infectious Diseases. Elsevier Inc., pp. 173−202. https://doi.org/10.1016/B978-0-12-384890-1.00007-8

Global Tuberculosis Report 2019. [WWW Document], n.d.. URL https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-report-2019 (accessed 29 August.2021).

Goh, A.E.N., Choi, E.H., Chokephaibulkit, K., Choudhury, J., Kuter, B., Lee, P.I., Marshall, H., Kim, J.O., Wolfson, L.J., 2019. Burden of varicella in the Asia-Pacific region: a systematic literature review. Expert Rev. Vaccines. https://doi.org/10.1080/14760584.2019.1594781.

Gouglas, D., Christodoulou, M., Plotkin, S.A., Hatchett, R., 2019. CEPI: Driving progress toward epidemic Preparedness and Response. Epidemiol. Rev. 41, 28−33. https://doi.org/10.1093/epirev/mxz012.

Grose, C., 2012. Pangaea and the out-of-Africa model of varicella-zoster virus evolution and phylogeography. J. Virol. 86, 9558−9565. https://doi.org/10.1128/jvi.00357-12.

Guellil, M., van Dorp, L., Inskip, S.A., Dittmar, J.M., Saag, L., Tambets, K., Hui, R., Rose, A., D'Atanasio, E., Kriiska, A., Varul, L., Koekkelkoren, A.M.H.C., Goldina, R.D., Cessford, C., Solnik, A., Metspalu, M., Krause, J., Herbig, A., Robb, J.E., Houldcroft, C.J., Scheib, C.L., 2022. Ancient herpes simplex 1 genomes reveal recent viral structure in Eurasia. Sci. Adv. 8. https://doi.org/10.1126/sciadv.abo4435.

Han, X.Y., Silva, F.J., 2014. On the age of leprosy. PLoS Negl. Trop. Dis. 8, e2544. https://doi.org/10.1371/journal.pntd.0002544.

Harding, E., 2020. WHO global progress report on tuberculosis elimination. Lancet Respir. Med. https://doi.org/10.1016/S2213-2600(19)30418-7.

Harvati, K., Röding, C., Bosman, A.M., Karakostis, F.A., Grün, R., Stringer, C., Karkanas, P., Thompson, N.C., Koutoulidis, V., Moulopoulos, L.A., Gorgoulis, V.G., Kouloukoussa, M., 2019. Apidima Cave fossils provide earliest evidence of *Homo sapiens* in Eurasia. Nature 571, 500−504. https://doi.org/10.1038/s41586-019-1376-z.

Hershberg, R., Lipatov, M., Small, P.M., Sheffer, H., Niemann, S., Homolka, S., Roach, J.C., Kremer, K., Petrov, D.A., Feldman, M.W., Gagneux, S., 2008. High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol. 6, e311. https://doi.org/10.1371/journal.pbio.0060311.

Hershkovitz, I., Weber, G.W., Quam, R., Duval, M., Grün, R., Kinsley, L., Ayalon, A., Bar-Matthews, M., Valladas, H., Mercier, N., Arsuaga, J.L., Martinón-Torres, M., De Castro, J.M.B., Fornai, C., Martín-Francés, L., Sarig, R., May, H., Krenn, V.A., Slon, V., Rodríguez, L., García, R., Lorenzo, C., Carretero, J.M., Frumkin, A., Shahack-Gross, R., Mayer, D.E.B.Y., Cui, Y., Wu, X., Peled, N., Groman-Yaroslavski, I., Weissbrod, L., Yeshurun, R., Tsatskin, A., Zaidner, Y., Weinstein-Evron, M., 2018. The earliest modern humans outside Africa. Science 359, 456−459. https://doi.org/10.1126/science.aap8369, 80.

Higham, T.F.G., Douka, K., 2021. The reliability of late radiocarbon dates from the Paleolithic of southern China. Proc. Natl. Acad. Sci. 118, e2103798118. https://doi.org/10.1073/pnas.2103798118.

Ho, S.Y.W., Lanfear, R., Bromham, L., Phillips, M.J., Soubrier, J., Rodrigo, A.G., Cooper, A., 2011. Time-dependent rates of molecular evolution. Mol. Ecol. 20, 3087−3101. https://doi.org/10.1111/j.1365-294X.2011.05178.x.

Ho, S.Y.W., Phillips, M.J., Cooper, A., Drummond, A.J., 2005. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. Mol. Biol. Evol. 22, 1561−1568. https://doi.org/10.1093/molbev/msi145.

Honap, T.P., Pfister, L.-A., Housman, G., Mills, S., Tarara, R.P., Suzuki, K., Cuozzo, F.P., Sauther, M.L., Rosenberg, M.S., Stone, A.C., 2018. Mycobacterium leprae genomes from naturally infected nonhuman primates. PLoS Negl. Trop. Dis. 12, e0006190. https://doi.org/10.1371/journal.pntd.0006190.

Houldcroft, C.J., 2019. Human herpesvirus sequencing in the genomic era: the growing ranks of the herpetic legion. Pathogens 8. https://doi.org/10.3390/pathogens8040186.

Inoue, M., 2017. Changing epidemiology of *Helicobacter pylori* in Japan. Gastric Cancer 20, 3−7. https://doi.org/10.1007/s10120-016-0658-5.

Inskip, S., Taylor, G.M., Anderson, S., Stewart, G., 2017. Leprosy in pre-norman Suffolk, UK: biomolecular and geochemical analysis of the woman from hoxne. J. Med. Microbiol. 66, 1640−1649. https://doi.org/10.1099/jmm.0.000606.

Jacobs, G.S., Hudjashov, G., Saag, L., Kusuma, P., Darusallam, C.C., Lawson, D.J., Mondal, M., Pagani, L., Ricaut, F.X., Stoneking, M., Metspalu, M., Sudoyo, H., Lansing, J.S., Cox, M.P., 2019. Multiple deeply divergent denisovan ancestries in papuans. Cell 177, 1010−1021.e32. https://doi.org/10.1016/j.cell.2019.02.035.

Jensen, N.J., Rivailler, P., Tseng, H.F., Quinlivan, M.L., Radford, K., Folster, J., Harpaz, R., LaRussa, P., Jacobsen, S., Scott Schmid, D., 2017. Revisiting the genotyping scheme for varicella-zoster viruses based on whole-genome comparisons. J. Gen. Virol. 98, 1434−1438. https://doi.org/10.1099/jgv.0.000772.

Jobling, M.A., 2012. The impact of recent events on human genetic diversity. Philos. Trans. R. Soc. B Biol. Sci. https://doi.org/10.1098/rstb.2011.0297.

Kato, A., Kitamura, T., Sugimoto, C., Ogawa, Y., Nakazato, K., Nagashima, K., Hall, W.W., Kawabe, K., Yogo, Y., 1997. Lack of evidence for the transmission of JC polyomavirus between human populations. Arch. Virol. 142, 875−882. https://doi.org/10.1007/s007050050125.

Kay, G.L., Sergeant, M.J., Zhou, Z., Chan, J.Z.M., Millard, A., Quick, J., Szikossy, I., Pap, I., Spigelman, M., Loman, N.J., Achtman, M., Donoghue, H.D., Pallen, M.J., 2015. Eighteenth-century genomes show that mixed infections were common at time of peak tuberculosis in Europe. Nat. Commun. 6, 1−9. https://doi.org/10.1038/ncomms7717.

Kennemann, L., Didelot, X., Aebischer, T., Kuhn, S., Drescher, B., Droege, M., Reinhardt, R., Correa, P., Meyer, T.F., Josenhans, C., Falush, D., Suerbaum, S., 2011. *Helicobacter pylori* genome evolution during human infection. Proc. Natl. Acad. Sci. U. S. A. 108, 5033−5038. https://doi.org/10.1073/pnas.1018444108.

Kitamura, T., Sugimoto, C., Kato, A., Ebihara, H., Suzuki, M., Taguchi, F., Kawabe, K., Yogo, Y., 1997. Persistent jc virus (jcv) infection is demonstrated by continuous shedding of the same jcv strains. J. Clin. Microbiol. 35, 1255−1257. https://doi.org/10.1128/jcm.35.5.1255-1257.1997.

Kitchen, A., Miyamoto, M.M., Mulligan, C.J., 2008. Utility of DNA viruses for studying human host history: case study of JC virus. Mol. Phylogenet. Evol. 46, 673−682. https://doi.org/10.1016/j.ympev.2007.09.005.

Klein, E.Y., Milkowska-Shibata, M., Tseng, K.K., Sharland, M., Gandra, S., Pulcini, C., Laxminarayan, R., 2021. Assessment of WHO antibiotic consumption and access targets in 76 countries, 2000−15: an analysis of pharmaceutical sales data. Lancet Infect. Dis. 21, 107−115. https://doi.org/10.1016/S1473-3099(20)30332-7.

Koelle, D.M., Norberg, P., Fitzgibbon, M.P., Russell, R.M., Greninger, A.L., Huang, M.-L., Stensland, L., Jing, L., Magaret, A.S., Diem, K., Selke, S., Xie, H., Celum, C., Lingappa, J.R., Jerome, K.R., Wald, A., Johnston, C., 2017. Worldwide circulation of HSV-2 × HSV-1 recombinant strains. Sci. Rep. 7, 44084. https://doi.org/10.1038/srep44084.

Krause-Kyora, B., Susat, J., Key, F.M., Kühnert, D., Bosse, E., Immel, A., Rinne, C., Kornell, S.-C., Yepes, D., Franzenburg, S., Heyne, H.O., Meier, T., Lösch, S., Meller, H., Friederich, S., Nicklisch, N., Alt, K.W., Schreiber, S., Tholey, A., Herbig, A., Nebel, A., Krause, J., 2018. Neolithic and Medieval virus genomes reveal complex evolution of Hepatitis B. Elife 7, e36666. https://doi.org/10.7554/eLife.36666.

Krumbholz, A., Bininda-Emonds, O.R.P., Wutzler, P., Zell, R., 2009. Phylogenetics, evolution, and medical importance of polyomaviruses. Infect. Genet. Evol. 9, 784−799. https://doi.org/10.1016/j.meegid.2009.04.008.

Krumbholz, A., Bininda-Emonds, O.R.P., Wutzler, P., Zell, R., 2008. Evolution of four BK virus subtypes. Infect. Genet. Evol. 8, 632−643. https://doi.org/10.1016/j.meegid.2008.05.006.

Kunitake, T., Kitamura, T., Guo, J., Taguchi, F., Kawabe, K., Yogo, Y., 1995. Parent-to-child transmission is relatively common in the spread of the human polyomavirus JC virus. J. Clin. Microbiol. 33, 1448−1451. https://doi.org/10.1128/jcm.33.6.1448-1451.1995.

Leprosy (Hansen's disease) [WWW Document], n.d.. URL https://www.who.int/news-room/fact-sheets/detail/leprosy (accessed 27.September.2021).

Levy, J.A., Ferro, F., Greenspan, D., Lennette, E.T., 1990. Frequent isolation of HHV-6 from saliva and high seroprevalence of the virus in the population. Lancet 335, 1047−1050. https://doi.org/10.1016/0140-6736(90)92628-U.

Li, G., Figueiró, H.V., Eizirik, E., Murphy, W.J., 2019. Recombination-aware phylogenomics reveals the structured genomic landscape of hybridizing cat species. Mol. Biol. Evol. 36, 2111−2126. https://doi.org/10.1093/molbev/msz139.

Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493−496. https://doi.org/10.1038/nature10231.

Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., Van Der Merwe, S.W., Yamaoka, Y., Graham, D.Y., Perez-Trallero, E., Wadstrom, T., Suerbaum, S., Achtman, M., 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature 445, 915−918. https://doi.org/10.1038/nature05562.

Liu, W., Martinón-Torres, M., Cai, Y.J., Xing, S., Tong, H.W., Pei, S.W., Sier, M.J., Wu, X.H., Edwards, R.L., Cheng, H., Li, Y.Y., Yang, X.X., De Castro, J.M.B., Wu, X.J., 2015. The earliest unequivocally modern humans in southern China. Nature 526, 696−699. https://doi.org/10.1038/nature15696.

Madinda, N.F., Ehlers, B., Wertheim, J.O., Akoua-Koffi, C., Bergl, R.A., Boesch, C., Akonkwa, D.B.M., Eckardt, W., Fruth, B., Gillespie, T.R., Gray, M., Hohmann, G., Karhemere, S., Kujirakwinja, D., Langergraber, K., Muyembe, J.-J., Nishuli, R., Pauly, M., Petrzelkova, K.J., Robbins, M.M., Todd, A., Schubert, G., Stoinski, T.S., Wittig, R.M., Zuberbühler, K., Peeters, M., Leendertz, F.H., Calvignac-Spencer, S., 2016. Assessing host-virus codivergence for close relatives of merkel cell polyomavirus infecting african great apes. J. Virol. 90, 8531−8541. https://doi.org/10.1128/jvi.00247-16.

Magalhães, I.M., Martins, R.V.N., Cossatis, J.J., Cavaliere, R.M., Afonso, L.A., Moysés, N., Oliveira, S.A., Cavalcanti, S.M.B., 2010. Detection of human herpesvirus 6 and 7 DNA in saliva from healthy adults from Rio de Janeiro, Brazil. Mem. Inst. Oswaldo Cruz 105, 925−927. https://doi.org/10.1590/S0074-02762010000700015.

Mahalingam, R., Gilden, D.H., 2007. Simian varicella virus. In: Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge University Press, pp. 1043−1050. https://doi.org/10.1017/CBO9780511545313.059.

Maixner, F., Krause-Kyora, B., Turaev, D., Herbig, A., Hoopmann, M.R., Hallows, J.L., Kusebauch, U., Vigl, E.E., Malfertheiner, P., Megraud, F., O'Sullivan, N., Cipollini, G., Coia, V., Samadelli, M., Engstrand, L., Linz, B., Moritz, R.L., Grimm, R., Krause, J., Nebel, A., Moodley, Y., Rattei, T., Zink, A., O'Sullivan, N., Cipollini, G., Coia, V., Samadelli, M., Engstrand, L., Linz, B., Moritz, R.L., Grimm, R., Krause, J., Nebel, A., Moodley, Y., Rattei, T., Zink, A., 2016. The 5300-year-old *Helicobacter pylori* genome of the Iceman. Science 351, 162−165. https://doi.org/10.1126/science.aad2545, 80.

Maixner, F., Thorell, K., Granehäll, L., Linz, B., Moodley, Y., Rattei, T., Engstrand, L., Zink, A., 2019. *Helicobacter pylori* in ancient human remains. World J. Gastroenterol. 25, 6289−6298. https://doi.org/10.3748/wjg.v25.i42.6289.

Mark, S., 2017. Early human migrations (ca. 13,000 Years ago) or postcontact Europeans for the earliest spread of *Mycobacterium leprae* and Mycobacterium lepromatosis to the Americas. Interdiscip. Perspect. Infect. Dis. 2017. https://doi.org/10.1155/2017/6491606.

Mark, S., 2002. Alexander the great, seafaring, and the spread of leprosy. J. Hist. Med. Allied Sci. 57, 285−311. https://doi.org/10.1093/jhmas/57.3.285.

Martel-Jantin, C., Filippone, C., Tortevoye, P., Afonso, P.V., Betsem, E., Descorps-Declere, S., Nicol, J.T.J., Touzé, A., Coursaget, P., Crouzat, M., Berthet, N., Cassar, O., Gessaina, A., 2014. Molecular epidemiology of

merkel cell polyomavirus: evidence for geographically related variant genotypes. J. Clin. Microbiol. 52, 1687−1690. https://doi.org/10.1128/JCM.02348-13.

Menardo, F., Duchêne, S., Brites, D., Gagneux, S., 2019. The molecular clock of *Mycobacterium tuberculosis*. PLoS Pathog. 15, e1008067. https://doi.org/10.1371/journal.ppat.1008067.

Moens, U., Krumbholz, A., Ehlers, B., Zell, R., Johne, R., Calvignac-Spencer, S., Lauber, C., 2017. Biology, evolution, and medical importance of polyomaviruses: an update. Infect. Genet. Evol. https://doi.org/10.1016/j.meegid.2017.06.011.

Monot, M., Honoré, N., Garnier, T., Araoz, R., Coppée, J.Y., Lacroix, C., Sow, S., Spencer, J.S., Truman, R.W., Williams, D.L., Gelber, R., Virmond, M., Flageul, B., Cho, S.N., Ji, B., Paniz-Mondolfi, A., Convit, J., Young, S., Fine, P.E., Rasolofo, V., Brennan, P.J., Cole, S.T., 2005. On the origin of leprosy. Science 308, 1040−1042. https://doi.org/10.1126/science/1109759, 80.

Monot, M., Honoré, N., Garnier, T., Zidane, N., Sherafi, D., Paniz-Mondolfi, A., Matsuoka, M., Taylor, G.M., Donoghue, H.D., Bouwman, A., Mays, S., Watson, C., Lockwood, D., Khamesipour, A., Dowlati, Y., Jianping, S., Rea, T.H., Vera-Cabrera, L., Stefani, M.M., Banu, S., MacDonald, M., Sapkota, B.R., Spencer, J.S., Thomas, J., Harshman, K., Singh, P., Busso, P., Gattiker, A., Rougemont, J., Brennan, P.J., Cole, S.T., 2009. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nat. Genet. 41, 1282−1289. https://doi.org/10.1038/ng.477.

Montano, V., Didelot, X., Foll, M., Linz, B., Reinhardt, R., Suerbaum, S., Moodley, Y., Jensen, J.D., 2015. Worldwide population structure, long-term demography, and local adaptation of *Helicobacter pylori*. Genetics 200, 947−963. https://doi.org/10.1534/genetics.115.176404.

Moodley, Y., Brunelli, A., Ghirotto, S., Klyubin, A., Maady, A.S., Tyne, W., Muñoz-Ramirez, Z.Y., Zhou, Z., Manica, A., Linz, B., Achtman, M., 2021. *Helicobacter pylori*'s historical journey through Siberia and the Americas. Proc. Natl. Acad. Sci. U. S. A. 118. https://doi.org/10.1073/pnas.2015523118.

Moodley, Y., Linz, B., Bond, R.P., Nieuwoudt, M., Soodyall, H., Schlebusch, C.M.C., Bernhöft, S., Hale, J., Suerbaum, S., Mugisha, L., van der Merwe, S.W., Achtman, M., Suerbaum, S., Michetti, P., Schwarz, S., Morelli, G., Kusecek, B., Manica, A., Balloux, F., Achtman, M., Azuma, T., Berg, D., Ito, Y., Morelli, G., Morelli, G., Didelot, X., Kusecek, B., Schwarz, S., Bahlawane, C., Falush, D., Stephens, M., Pritchard, J., Falush, D., Wirth, T., Linz, B., Pritchard, J., Stephens, M., Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H., Breurec, S., Wirth, T., Wang, X., Linz, B., Novick, R., Lum, J., Breurec, S., Guillard, B., Hem, S., Brisse, S., Dieye, F., Moodley, Y., Linz, B., Stauffer, R., Walker, A., Ryder, O., Lyons-Weiler, M., Hedges, S., Dewhirst, F., Shen, Z., Scimeca, M., Stokes, L., Boumenna, T., Eppinger, M., Baar, C., Linz, B., Raddatz, G., Lanz, C., Harper, C., Feng, Y., Xu, S., Taylor, N., Kinsel, M., Behar, D., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Gonder, M., Mortensen, H., Reed, F., De, S., Tishkoff, S., Tishkoff, S., Gonder, M., Henn, B., Mortensen, H., Knight, A., Schuster, S., Miller, W., Ratan, A., Tomsho, L., Giardine, B., Henn, B., Gignoux, C., Jobin, M., Granka, J., Macpherson, J., Guldemann, T., Stoneking, M., Diamond, J., Bellwood, P., Pereira, L., Macaulay, V., Torroni, A., Scozzari, R., Prata, M., Salas, A., Richards, M., Fe, T.D. la, Lareu, M., Sobrino, B., Rosa, A., Brehm, A., Delport, W., Cunningham, M., Olivier, B., Preisig, O., Merwe, S. Van Der, Didelot, X., Falush, D., Didelot, X., Maiden, M., Harris, S., Feil, E., Holden, M., Quail, M., Nickerson, E., Nubel, U., Dordel, J., Kurt, K., Strommenger, B., Westh, H., Liu, W., Li, Y., Learn, G., Rudicell, R., Robertson, J., Katzourakis, A., Tristem, M., Pybus, O., Gifford, R., Hey, J., Nielsen, R., Didelot, X., Barker, M., Falush, D., Priest, F., Hudson, R., Kaplan, N., Hey, J., Runemark, A., Hey, J., Hansson, B., Svensson, E., Foote, A., Morin, P., Durban, J., Willerslev, E., Orlando, L., Templeton, A., Cruciani, F., Trombetta, B., Massaia, A., Destro-Bisol, G., Sellitto, D., Schuster, S., Wittekindt, N., Linz, B., Liu, H., Prugnolle, F., Manica, A., Balloux, F., Gao, F., Bailes, E., Robertson, D., Chen, Y., Rodenburg, C., Bronsdon, M., Goodwin, C., Sly, L., Chilvers, T., Schoenknecht, F., Suerbaum, S., Kraft, C., Dewhirst, F., Fox, J., Handt, L., Fox, J., Yan, L., Shen, Z., Pouch, W., Doi, S., Kimbason, T., Reindel, J., Dubois, A., Solnick, J., Hansen, L., Salama, N., Boonjakuakul, J., Syvanen, M., Reindel, J., Fitzgerald, A., Breider, M., Gough, A., Yan, C., Shuto, R.,

Fujioka, T., Kubota, T., Nasu, M., O'Rourke, J., Solnick, J., Neilan, B., Seidel, K., Hayter, R., Fernandez, K., Hansen, L., Vandamme, P., Beaman, B., Solnick, J., Marini, R., Muthupalani, S., Shen, Z., Buckley, E., Alvarado, C., Foley, R., Lahr, M., Mellars, P., Oppenheimer, S., Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Momynaliev, K., Chelysheva, V., Akopian, T., Selezneva, O., Linz, B., Latifi-Navid, S., Ghorashi, S., Siavoshi, F., Linz, B., Massarat, S., Devi, S., Ahmed, I., Francalacci, P., Hussain, M., Akhter, Y., Higham, T., Compton, T., Stringer, C., Jacobi, R., Shapiro, B., Benazzi, S., Douka, K., Fornai, C., Bauer, C., Kullmer, O., Mellars, P., Clark, P., Dyke, A., Shakun, J., Carlson, A., Clark, J., Malyarchuk, B., Grzybowski, T., Derenko, M., Perkova, M., Vanecek, T., Torroni, A., Bandelt, H., D'Urbano, L., Lahermo, P., Moral, P., Torroni, A., Bandelt, H., Macaulay, V., Richards, M., Cruciani, F., Achilli, A., Rengo, C., Magri, C., Battaglia, V., Olivieri, A., Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Soares, P., Achilli, A., Semino, O., Davies, W., Macaulay, V., Bramanti, B., Thomas, M., Haak, W., Unterlaender, M., Jores, P., Haak, W., Forster, P., Bramanti, B., Matsumura, S., Brandt, G., Dominguez-Bello, M., Pérez, M., Bortolini, M., Salzano, F., Pericci, L., Ottoni, C., Primativo, G., Hooshiar, K., Achilli, A., Martinez-Labarga, C., Olivieri, A., Achilli, A., Pala, M., Battaglia, V., Fornarino, S., Henn, B., Botigue, L., Gravel, S., Wang, W., Brisbin, A., Schlebusch, C.M.C., Naidoo, T., Soodyall, H., Behar, D., Rosset, S., Blue-Smith, J., Balanovsky, O., Tzur, S., Rosenberg, N., Jobb, G., Haeseler, A. von, Strimmer, K., Rozas, J., Sanchez-DelBarrio, J., Messeguer, X., Rozas, R., Wiuf, C., Kumar, S., Nei, M., Dudley, J., Tamura, K., Peakall, R., Smouse, P., Tamura, K., Dudley, J., Nei, M., Kumar, S., Tishkoff, S., Dietzsch, E., Speed, W., Pakstis, A., Kidd, J., Pope, K., Terrell, J., Fagundes, N., Kanitz, R., Eckert, R., Valls, A., Bogo, M., Trejaut, J., Kivisild, T., Loo, J., Lee, C., He, C., Spriggs, M., Spriggs, M., Dillehay, T., Ramirez, C., Pino, M., Collins, M., Rossen, J., 2012. Age of the association between *Helicobacter pylori* and man. PLoS Pathog. 8, e1002693. https://doi.org/10.1371/journal.ppat.1002693.

Moodley, Y., Linz, B., Yamaoka, Y., Windsor, H.M., Breurec, S., Wu, J.Y., Maady, A., Bernhöft, S., Thiberge, J.M., Phuanukoonnon, S., Jobb, G., Siba, P., Graham, D.Y., Marshall, B.J., Achtman, M., 2009. The peopling of the pacific from a bacterial perspective. Science 323, 527−530. https://doi.org/10.1126/science.1166083, 80-.

Moreno-Mayar, J.V., Vinner, L., De Barros Damgaard, P., De La Fuente, C., Chan, J., Spence, J.P., Allentoft, M.E., Vimala, T., Racimo, F., Pinotti, T., Rasmussen, S., Margaryan, A., Orbegozo, M.I., Mylopotamitaki, D., Wooller, M., Bataille, C., Becerra-Valdivia, L., Chivall, D., Comeskey, D., Devièse, T., Grayson, D.K., George, L., Harry, H., Alexandersen, V., Primeau, C., Erlandson, J., Rodrigues-Carvalho, C., Reis, S., Bastos, M.Q.R.R., Cybulski, J., Vullo, C., Morello, F., Vilar, M., Wells, S., Gregersen, K., Hansen, K.L., Lynnerup, N., Lahr, M.M., Kjær, K., Strauss, A., Alfonso-Durruty, M., Salas, A., Schroeder, H., Higham, T., Malhi, R.S., Rasic, J.T., Souza, L., Santos, F.R., Malaspinas, A.-S.S., Sikora, M., Nielsen, R., Song, Y.S., Meltzer, D.J., Willerslev, E., Damgaard, P. de B., Fuente, C. de la, Chan, J., Spence, J.P., Allentoft, M.E., Vimala, T., Racimo, F., Pinotti, T., Rasmussen, S., Margaryan, A., Orbegozo, M.I., Mylopotamitaki, D., Wooller, M., Bataille, C., Becerra-Valdivia, L., Chivall, D., Comeskey, D., Devièse, T., Grayson, D.K., George, L., Harry, H., Alexandersen, V., Primeau, C., Erlandson, J., Rodrigues-Carvalho, C., Reis, S., Bastos, M.Q.R.R., Cybulski, J., Vullo, C., Morello, F., Vilar, M., Wells, S., Gregersen, K., Hansen, K.L., Lynnerup, N., Lahr, M.M., Kjær, K., Strauss, A., Alfonso-Durruty, M., Salas, A., Schroeder, H., Higham, T., Malhi, R.S., Rasic, J.T., Souza, L., Santos, F.R., Malaspinas, A.-S.S., Sikora, M., Nielsen, R., Song, Y.S., Meltzer, D.J., Willerslev, E., De Barros Damgaard, P., De La Fuente, C., Chan, J., Spence, J.P., Allentoft, M.E., Vimala, T., Racimo, F., Pinotti, T., Rasmussen, S., Margaryan, A., Orbegozo, M.I., Mylopotamitaki, D., Wooller, M., Bataille, C., Becerra-Valdivia, L., Chivall, D., Comeskey, D., Devièse, T., Grayson, D.K., George, L., Harry, H., Alexandersen, V., Primeau, C., Erlandson, J., Rodrigues-Carvalho, C., Reis, S., Bastos, M.Q.R.R., Cybulski, J., Vullo, C., Morello, F., Vilar, M., Wells, S., Gregersen, K., Hansen, K.L., Lynnerup, N., Lahr, M.M., Kjær, K., Strauss, A., Alfonso-Durruty, M., Salas, A., Schroeder, H., Higham, T., Malhi, R.S., Rasic, J.T., Souza, L., Santos, F.R., Malaspinas, A.-S.S., Sikora, M., Nielsen, R., Song, Y.S., Meltzer, D.J., Willerslev, E., 2018. Early human dispersals within the Americas. Science 362, eaav2621, 80.

Muir, W.B., Nichols, R., Breuer, J., 2002. Phylogenetic analysis of varicella-zoster virus: evidence of intercontinental spread of genotypes and recombination. J. Virol. 76, 1971−1979. https://doi.org/10.1128/JVI.76.4.1971-1979.2002.

Muñoz-Ramirez, Z.Y., Pascoe, B., Mendez-Tenorio, A., Mourkas, E., Sandoval-Motta, S., Perez-Perez, G., Morgan, D.R., Dominguez, R.L., Ortiz-Princz, D., Cavazza, M.E., Rocha, G., Queiroz, D.M.M., Catalano, M., Palma, G.Z. De, Goldman, C.G., Venegas, A., Alarcon, T., Oleastro, M., Vale, F.F., Goodman, K.J., Torres, R.C., Berthenet, E., Hitchings, M.D., Blaser, M.J., Sheppard, S.K., Thorell, K., Torres, J., 2021. A 500-year tale of co-evolution, adaptation, and virulence: *Helicobacter pylori* in the Americas. ISME J. 15, 78−92. https://doi.org/10.1038/s41396-020-00758-0.

Ngabonziza, J.C.S., Loiseau, C., Marceau, M., Jouet, A., Menardo, F., Tzfadia, O., Antoine, R., Niyigena, E.B., Mulders, W., Fissette, K., Diels, M., Gaudin, C., Duthoy, S., Ssengooba, W., André, E., Kaswa, M.K., Habimana, Y.M., Brites, D., Affolabi, D., Mazarati, J.B., de Jong, B.C., Rigouts, L., Gagneux, S., Meehan, C.J., Supply, P., 2020. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. Nat. Commun. 11, 1−11. https://doi.org/10.1038/s41467-020-16626-6.

Nielsen, S.H., Van Dorp, L., Houldcroft, C.J., Pedersen, A.G., Allentoft, M.E., Vinner, L., Margaryan, A., Pavlova, E., Chasnyk, V., Nikolskiy, P., Pitulko, V., Pimenoff, V.N., Balloux, F., Sikora, M., 2021. 31,600-year-old human virus genomes support a Pleistocene origin for common childhood infections. bioRxiv, 450199. https://doi.org/10.1101/2021.06.28.450199, 2021.06.28.

Norberg, P., Depledge, D.P., Kundu, S., Atkinson, C., Brown, J., Haque, T., Hussaini, Y., MacMahon, E., Molyneaux, P., Papaevangelou, V., Sengupta, N., Koay, E.S.C., Tang, J.W., Underhill, G.S., Grahn, A., Studahl, M., Breuer, J., Bergström, T., 2015. Recombination of globally circulating varicella-zoster virus. J. Virol. 89, 7133−7146. https://doi.org/10.1128/jvi.00437-15.

O'Connell, J.F., Allen, J., Williams, M.A.J., Williams, A.N., Turney, C.S.M., Spooner, N.A., Kamminga, J., Brown, G., Cooper, A., 2018. When did homo sapiens first reach Southeast Asia and Sahul? Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.1808385115.

Ochoa, A., Storey, J.D., 2021. Estimating FST and kinship for arbitrary population structures. PLoS Genet. 17, e1009241. https://doi.org/10.1371/journal.pgen.1009241.

Orgeur, M., Brosch, R., 2018. Evolution of virulence in the *Mycobacterium tuberculosis* complex. Curr. Opin. Microbiol. https://doi.org/10.1016/j.mib.2017.11.021.

Osaki, T., Konno, M., Yonezawa, H., Hojo, F., Zaman, C., Takahashi, M., Fujiwara, S., Kamiya, S., 2015. Analysis of intra-familial transmission of *Helicobacter pylori* in Japanese families. J. Med. Microbiol. 64, 67−73. https://doi.org/10.1099/jmm.0.080507-0.

Patterson, B., Wood, R., 2019. Is cough really necessary for TB transmission? Tuberculosis. https://doi.org/10.1016/j.tube.2019.05.003.

Pfrengle, S., Neukamm, J., Guellil, M., Keller, M., Molak, M., Avanzi, C., Kushniarevich, A., Montes, N., Neumann, G.U., Reiter, E., Tukhbatova, R.I., Berezina, N.Y., Buzhilova, A.P., Korobov, D.S., Suppersberger Hamre, S., Matos, V.M.J., Ferreira, M.T., González-Garrido, L., Wasterlain, S.N., Lopes, C., Santos, A.L., Antunes-Ferreira, N., Duarte, V., Silva, A.M., Melo, L., Sarkic, N., Saag, L., Tambets, K., Busso, P., Cole, S.T., Avlasovich, A., Roberts, C.A., Sheridan, A., Cessford, C., Robb, J., Krause, J., Scheib, C.L., Inskip, S.A., Schuenemann, V.J., 2021. *Mycobacterium leprae* diversity and population dynamics in medieval Europe from novel ancient genomes. BMC Biol. 19, 220. https://doi.org/10.1186/s12915-021-01120-2.

Pierik, J.G.J., Gumbs, P.D., Fortanier, S.A.C., Van Steenwijk, P.C.E., Postma, M.J., 2012. Epidemiological characteristics and societal burden of varicella zoster virus in The Netherlands. BMC Infect. Dis. 12, 110. https://doi.org/10.1186/1471-2334-12-110.

Pimenoff, V.N., De Oliveira, C.M., Bravo, I.G., 2017. Transmission between archaic and modern human ancestors during the evolution of the oncogenic human papillomavirus 16. Mol. Biol. Evol. 34, 4−19. https://doi.org/10.1093/molbev/msw214.

Pontremoli, C., Forni, D., Clerici, M., Cagliani, R., Sironi, M., 2019. Possible European origin of circulating varicella zoster virus strains. J. Infect. Dis. 221, 1286−1294. https://doi.org/10.1093/infdis/jiz227.

Posth, C., Nakatsuka, N., Lazaridis, I., Skoglund, P., Mallick, S., Lamnidis, T.C., Rohland, N., Nägele, K., Adamski, N., Bertolini, E., Broomandkhoshbacht, N., Cooper, A., Culleton, B.J., Ferraz, T., Ferry, M., Furtwängler, A., Haak, W., Harkins, K., Harper, T.K., Hünemeier, T., Lawson, A.M., Llamas, B., Michel, M., Nelson, E., Oppenheimer, J., Patterson, N., Schiffels, S., Sedig, J., Stewardson, K., Talamo, S., Wang, C.-C., Hublin, J.-J., Hubbe, M., Harvati, K., Nuevo Delaunay, A., Beier, J., Francken, M., Kaulicke, P., Reyes-Centeno, H., Rademaker, K., Trask, W.R., Robinson, M., Gutierrez, S.M., Prufer, K.M., Salazar-García, D.C., Chim, E.N., Müller Plumm Gomes, L., Alves, M.L., Liryo, A., Inglez, M., Oliveira, R.E., Bernardo, D.V., Barioni, A., Wesolowski, V., Scheifler, N.A., Rivera, M.A., Plens, C.R., Messineo, P.G., Figuti, L., Corach, D., Scabuzzo, C., Eggers, S., DeBlasis, P., Reindel, M., Méndez, C., Politis, G., Tomasto-Cagigao, E., Kennett, D.J., Strauss, A., Fehren-Schmitz, L., Krause, J., Reich, D., 2018. Reconstructing the deep population history of central and South America. Cell 175, 1185−1197.e22. https://doi.org/10.1016/j.cell.2018.10.027.

Prusty, B.K., Gulve, N., Rasa, S., Murovska, M., Hernandez, P.C., Ablashi, D.V., 2017. Possible chromosomal and germline integration of human herpesvirus 7. J. Gen. Virol. 98, 266−274. https://doi.org/10.1099/jgv.0.000692.

Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S., Stafford, T.W., Orlando, L., Metspalu, E., Karmin, M., Tambets, K., Rootsi, S., Mägi, R., Campos, P.F., Balanovska, E., Balanovsky, O., Khusnutdinova, E., Litvinov, S., Osipova, L.P., Fedorova, S.A., Voevoda, M.I., Degiorgio, M., Sicheritz-Ponten, T., Brunak, S., Demeshchenko, S., Kivisild, T., Villems, R., Nielsen, R., Jakobsson, M., Willerslev, E., 2014. Upper palaeolithic Siberian genome reveals dual ancestry of native Americans. Nature 505, 87−91. https://doi.org/10.1038/nature12736.

Rasigade, J.P., Gilabert, A., Wirth, T., 2017. Microbes as tracers of past human demography and migrations. In: Genetics and Evolution of Infectious Diseases, second ed. Elsevier Inc., pp. 141−165. https://doi.org/10.1016/B978-0-12-799942-5.00007-X

Raymond, J., Thiberge, J.M., Chevalier, C., Kalach, N., Bergeret, M., Labigne, A., Dauga, C., 2004. Genetic and transmission analysis of *Helicobacter pylori* Strains within a family. Emerg. Infect. Dis. 10, 1816−1821. https://doi.org/10.3201/eid1010.040042.

Reich, D., Patterson, N., Campbell, D., Tandon, A., Mazieres, S., Ray, N., Parra, M.V., Rojas, W., Duque, C., Mesa, N., García, L.F., Triana, O., Blair, S., Maestre, A., Dib, J.C., Bravi, C.M., Bailliet, G., Corach, D., Hünemeier, T., Bortolini, M.C., Salzano, F.M., Petzl-Erler, M.L., Acuña-Alonzo, V., Aguilar-Salinas, C., Canizales-Quinteros, S., Tusié-Luna, T., Riba, L., Rodríguez-Cruz, M., Lopez-Alarcón, M., Coral-Vazquez, R., Canto-Cetina, T., Silva-Zolezzi, I., Fernandez-Lopez, J.C., Contreras, A.V., Jimenez-Sanchez, G., Gómez-Vázquez, M.J., Molina, J., Carracedo, A., Salas, A., Gallo, C., Poletti, G., Witonsky, D.B., Alkorta-Aranburu, G., Sukernik, R.I., Osipova, L., Fedorova, S.A., Vasquez, R., Villena, M., Moreau, C., Barrantes, R., Pauls, D., Excoffier, L., Bedoya, G., Rothhammer, F., Dugoujon, J.-M., Larrouy, G., Klitz, W., Labuda, D., Kidd, J., Kidd, K., Di Rienzo, A., Freimer, N.B., Price, A.L., Ruiz-Linares, A., 2012. Reconstructing Native American population history. Nature 488, 370−374. https://doi.org/10.1038/nature11258.

Renner, D.W., Szpara, M.L., 2018. Impacts of genome-wide analyses on our understanding of human herpesvirus diversity and evolution. J. Virol. 92, 908−925. https://doi.org/10.1128/jvi.00908-17.

Rito, T., Vieira, D., Silva, M., Conde-Sousa, E., Pereira, L., Mellars, P., Richards, M.B., Soares, P., 2019. A dispersal of *Homo sapiens* from southern to eastern Africa immediately preceded the out-of-Africa migration. Sci. Rep. 9. https://doi.org/10.1038/s41598-019-41176-3.

Robbins, G., Tripathy, V.M., Misra, V.N., Mohanty, R.K., Shinde, V.S., Gray, K.M., Schug, M.D., 2009. Ancient skeletal evidence for leprosy in India (2000 B.C.). PLoS One 4, e5669. https://doi.org/10.1371/journal.pone.0005669.

Roberts, C., 2020. Leprosy: Past and Present.

Rogers, A.R., Harris, N.S., Achenbach, A.A., 2020. Neanderthal-Denisovan ancestors interbred with a distantly related hominin. Sci. Adv. 6. https://doi.org/10.1126/sciadv.aay5483.

Sabin, S., Herbig, A., Vågene, Å.J., Ahlström, T., Bozovic, G., Arcini, C., Kühnert, D., Bos, K.I., 2020. A seventeenth-century *Mycobacterium tuberculosis* genome supports a Neolithic emergence of the *Mycobacterium tuberculosis* complex. Genome Biol. 21, 201. https://doi.org/10.1186/s13059-020-02112-1.

Saxenhofer, M., de Melo, V.W., Ulrich, R.G., Heckel, G., 2017. Revised time scales of RNA virus evolution based on spatial information. Proc. R. Soc. B Biol. Sci. 284. https://doi.org/10.1098/rspb.2017.0857.

Scally, A., Durbin, R., 2012. Revising the human mutation rate: implications for understanding human evolution, 13, 745−753.

Schilling, A.-K., Avanzi, C., Ulrich, R.G., Busso, P., Pisanu, B., Ferrari, N., Romeo, C., Mazzamuto, M.V., McLuckie, J., Shuttleworth, C.M., Del-Pozo, J., Lurz, P.W.W., Escalante-Fuentes, W.G., Ocampo-Candiani, J., Vera-Cabrera, L., Stevenson, K., Chapuis, J.-L., Meredith, A.L., Cole, S.T., 2019. British red squirrels remain the only known wild rodent host for leprosy bacilli. Front. Vet. Sci. 6, 8. https://doi.org/10.3389/fvets.2019.00008.

Schlebusch, C.M., Malmström, H., Günther, T., Sjödin, P., Coutinho, A., Edlund, H., Munters, A.R., Vicente, M., Steyn, M., Soodyall, H., Lombard, M., Jakobsson, M., 2017. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science 358, 652−655. https://doi.org/10.1126/science.aao6266, 80.

Schuenemann, V.J., Avanzi, C., Krause-Kyora, B., Seitz, A., Herbig, A., Inskip, S., Bonazzi, M., Reiter, E., Urban, C., Dangvard Pedersen, D., Taylor, G.M., Singh, P., Stewart, G.R., Velemínský, P., Likovsky, J., Marcsik, A., Molnár, E., Pálfi, G., Mariotti, V., Riga, A., Belcastro, M.G., Boldsen, J.L., Nebel, A., Mays, S., Donoghue, H.D., Zakrzewski, S., Benjak, A., Nieselt, K., Cole, S.T., Krause, J., 2018. Ancient genomes reveal a high diversity of *Mycobacterium leprae* in medieval Europe. PLoS Pathog. 14, e1006997. https://doi.org/10.1371/journal.ppat.1006997.

Schuenemann, V.J., Singh, P., Mendum, T.A., Krause-Kyora, B., Jäger, G., Bos, K.I., Herbig, A., Economou, C., Benjak, A., Busso, P., Nebel, A., Boldsen, J.L., Kjellström, A., Wu, H.H.T.H., Stewart, G.R., Taylor, G.M., Bauer, P., Lee, O.Y.C., Wu, H.H.T.H., Minnikin, D.E., Besra, G.S., Tucker, K., Roffey, S., Sow, S.O., Cole, S.T., Nieselt, K., Krause, J., 2013. Genome-Wide Comparison of medieval and modern *Mycobacterium leprae*. Science 341, 179−184. https://doi.org/10.1126/science.1238286, 80.

Schwarz, S., Morelli, G., Kusecek, B., Manica, A., Balloux, F., Owen, R.J., Graham, D.Y., van der Merwe, S., Achtman, M., Suerbaum, S., 2008. Horizontal versus familial transmission of *Helicobacter pylori*. PLoS Pathog. 4, e1000180. https://doi.org/10.1371/journal.ppat.1000180.

Seidler, H., Bernhard, W., Teschler-Nicola, M., Platzer, W., Zur Nedden, D., Henn, R., Oberhauser, A., Sjøvold, T., 1992. Some anthropological aspects of the prehistoric Tyrolean ice man. Science 258, 455−457. https://doi.org/10.1126/science.1411539, 80.

Shackelton, L.A., Rambaut, A., Pybus, O.G., Holmes, E.C., 2006. JC virus evolution and its association with human populations. J. Virol. 80, 9928−9933. https://doi.org/10.1128/JVI.00441-06.

Sharp, P.M., Simmonds, P., 2011. Evaluating the evidence for virus/host co-evolution. Curr. Opin. Virol. 1 (5), 436−441. https://doi.org/10.1016/j.coviro.2011.10.018. ISSN 1879-6257. https://www.sciencedirect.com/science/article/pii/S1879625711001374.

Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L., Salzano, F.M., Patterson, N., Reich, D., 2015. Genetic evidence for two founding populations of the Americas. Nature 525, 104−108. https://doi.org/10.1038/nature14895.

Skoglund, P., Reich, D., 2016. A genomic view of the peopling of the Americas. Curr. Opin. Genet. Dev. https://doi.org/10.1016/j.gde.2016.06.016.

Spielman, S.J., Weaver, S., Shank, S.D., Magalis, B.R., Li, M., Kosakovsky Pond, S.L., 2019. Evolution of viral genomes: interplay between selection, recombination, and other forces. In: Methods in Molecular Biology. Humana Press Inc., pp. 427–468. https://doi.org/10.1007/978-1-4939-9074-0_14

Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., Musser, J.M., 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. U. S. A. 94, 9869–9874. https://doi.org/10.1073/pnas.94.18.9869.

Šroller, V., Hamšíková, E., Ludvíková, V., Vochozková, P., Kojzarová, M., Fraiberk, M., Saláková, M., Morávková, A., Forstová, J., Němečková, Š., 2014. Seroprevalence rates of BKV, JCV, and MCPyV polyomaviruses in the general Czech Republic population. J. Med. Virol. 86, 1560–1568. https://doi.org/10.1002/jmv.23841.

Stone, A.C., Lewis, C.M., Schuenemann, V.J., 2020. Insights into health and disease from ancient biomolecules. Philos. Trans. R. Soc. B Biol. Sci. 375, 20190568. https://doi.org/10.1098/rstb.2019.0568.

Sun, X.F., Wen, S.Q., Lu, C.Q., Zhou, B.Y., Curnoe, D., Lu, H.Y., Li, H.C., Wang, W., Cheng, H., Yi, S.W., Jia, X., Du, P.X., Xu, X.H., Lu, Y.M., Lu, Y., Zheng, H.X., Zhang, H., Sun, C., Wei, L.H., Han, F., Huang, J., Edwards, R.L., Jin, L., Li, H., 2021. Ancient DNA and multimethod dating confirm the late arrival of anatomically modern humans in southern China. Proc. Natl. Acad. Sci. U. S. A. 118. https://doi.org/10.1073/pnas.2019158118.

Suzuki, M., Zheng, H.-Y., Takasaka, T., Sugimoto, C., Kitamura, T., Beutler, E., Yogo, Y., 2002. Asian genotypes of JC virus in Japanese-Americans suggest familial transmission. J. Virol. 76, 10074–10078. https://doi.org/10.1128/jvi.76.19.10074-10078.2002.

Suzuki, T.A., Fitzstevens, L., Schmidt, V.T., Enav, H., Huus, K., Mbong, M., Adegbite, B.R., Zinsou, J.F., Esen, M., Velavan, T.P., Adegnika, A.A., Song, L.H., Spector, T.D., Muehlbauer, A.L., Marchi, N., Blekhman, R., Ségurel, L., Youngblut, N.D., Kremsner, P., Ley, R.E., 2021. Codiversification of gut microbiota with humans. bioRxiv 2021. https://doi.org/10.1101/2021.10.12.462973.

Telford, M., Navarro, A., Santpere, G., 2018. Whole genome diversity of inherited chromosomally integrated HHV-6 derived from healthy individuals of diverse geographic origin. Sci. Rep. 8, 3472. https://doi.org/10.1038/s41598-018-21645-x.

Tishkoff, S.A., Gonder, M.K., Henn, B.M., Mortensen, H., Knight, A., Gignoux, C., Fernandopulle, N., Lema, G., Nyambo, T.B., Ramakrishnan, U., Reed, F.A., Mountain, J.L., 2007. History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation. Mol. Biol. Evol. 24, 2180–2195. https://doi.org/10.1093/molbev/msm155.

Torres, C., Barrios, M.E., Cammarata, R.V., Victoria, M., Fernandez-Cassi, X., Bofill-Mas, S., Colina, R., Blanco Fernández, M.D., Mbayed, V.A., 2018. Phylodynamics of Merkel-cell polyomavirus and human polyomavirus 6: a long-term history with humans. Mol. Phylogenet. Evol. 126, 210–220. https://doi.org/10.1016/j.ympev.2018.04.025.

Vale, F.F., Vadivelu, J., Oleastro, M., Breurec, S., Engstrand, L., Perets, T.T., Mégraud, F., Lehours, P., 2015. Dormant phages of *Helicobacter pylori* reveal distinct populations in Europe. Sci. Rep. 5, 1–8. https://doi.org/10.1038/srep14333.

van Blankenstein, M., van Vuuren, A.J., Looman, C.W.N., Ouwendijk, M., Kuipers, E.J., 2013. The prevalence of *helicobacter pylori* infection in The Netherlands. Scand. J. Gastroenterol. 48, 794–800. https://doi.org/10.3109/00365521.2013.799221.

van Persie, J., Buitendijk, H., Fagrouch, Z., Bogers, W., Haaksma, T., Kondova, I., Verschoor, E.J., 2016. Complete genome sequence of a novel chimpanzee polyomavirus from a western common chimpanzee. Genome Announc. 4. https://doi.org/10.1128/genomeA.01406-15.

Viscidi, R.P., Rollison, D.E., Sondak, V.K., Silver, B., Messina, J.L., Giuliano, A.R., Fulp, W., Ajidahun, A., Rivanera, D., 2011. Age-specific seroprevalence of merkel cell polyomavirus, BK virus, and JC virus. Clin. Vaccine Immunol. 18, 1737−1743. https://doi.org/10.1128/CVI.05175-11.

Vyse, A.J., Gay, N.J., Hesketh, L.M., Andrews, N.J., Marshall, B., Thomas, H.I.J., Morgan-Capner, P., Miller, E., 2002. The burden of *Helicobacter pylori* infection in England and Wales. Epidemiol. Infect. 128, 411−417. https://doi.org/10.1017/S0950268802006970.

Wagenaar, T.R., Chow, V.T.K., Buranathai, C., Thawatsupha, P., Grose, C., 2003. The out of Africa model of varicella-zoster virus evolution: single nucleotide polymorphisms and private alleles distinguish Asian clades from European/North American clades. Vaccine 21, 1072−1081. https://doi.org/10.1016/S0264-410X(02)00559-5.

Walker, J.L., Andrews, N.J., Mathur, R., Smeeth, L., Thomas, S.L., 2017. Trends in the burden of varicella in UK general practice. Epidemiol. Infect. 145, 2678−2682. https://doi.org/10.1017/S0950268817001649.

Waters, M.R., 2019. Late Pleistocene exploration and settlement of the Americas by modern humans. Science. https://doi.org/10.1126/science.aat5447, 80.

Wegner, F., Lassalle, F., Depledge, D.P., Balloux, F., Breuer, J., 2019. Coevolution of sites under immune selection shapes epstein−barr virus population structure. Mol. Biol. Evol. 36, 2512−2521. https://doi.org/10.1093/molbev/msz152.

Weinert, L.A., Depledge, D.P., Kundu, S., Gershon, A.A., Nichols, R.A., Balloux, F., Welch, J.J., Breuer, J., 2015. Rates of vaccine evolution show strong effects of latency: implications for varicella zoster virus epidemiology. Mol. Biol. Evol. 32, 1020−1028. https://doi.org/10.1093/molbev/msu406.

Wirth, T., Meyer, A., Achtman, M., 2005. Deciphering host migrations and origins by means of their microbes. Mol. Ecol. https://doi.org/10.1111/j.1365-294X.2005.02687.x.

Wood, M.L., Veal, C.D., Neumann, R., Suárez, N.M., Nichols, J., Parker, A.J., Martin, D., Romaine, S.P., Codd, V., Samani, N.J., Voors, A.A., Tomaszewski, M., Flamand, L., Davison, A.J., Royle, N.J., 2021. Variation in human herpesvirus 6B telomeric integration, excision and transmission between tissues and individuals. Elife 10. https://doi.org/10.7554/elife.70452.

Yates, T.A., Khan, P.Y., Knight, G.M., Taylor, J.G., McHugh, T.D., Lipman, M., White, R.G., Cohen, T., Cobelens, F.G., Wood, R., Moore, D.A.J., Abubakar, I., 2016. The transmission of *Mycobacterium tuberculosis* in high burden settings. Lancet Infect. Dis. https://doi.org/10.1016/S1473-3099(15)00499-5.

Yuen, L.K.W., Littlejohn, M., Duchêne, S., Edwards, R., Bukulatjpi, S., Binks, P., Jackson, K., Davies, J., Davis, J.S., Tong, S.Y.C., Locarnini, S., 2019. Tracing ancient human migrations into Sahul using hepatitis B virus genomes. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msz021.

Zehender, G., Frati, E.R., Martinelli, M., Bianchi, S., Amendola, A., Ebranati, E., Ciccozzi, M., Galli, M., Lai, A., Tanzi, E., 2016. Dating the origin and dispersal of Human Papillomavirus type 16 on the basis of ancestral human migrations. Infect. Genet. Evol. 39, 258−264. https://doi.org/10.1016/j.meegid.2016.01.024.

Zell, R., Taudien, S., Pfaff, F., Wutzler, P., Platzer, M., Sauerbrei, A., 2012. Sequencing of 21 varicella-zoster virus genomes reveals two novel genotypes and evidence of recombination. J. Virol. 86, 1608−1622. https://doi.org/10.1128/jvi.06233-11.

Zhang, C., Liu, F., He, Z., Deng, Q., Pan, Y., Liu, Y., Zhang, C., Ning, T., Guo, C., Liang, Y., Xu, R., Zhang, L., Cai, H., Ke, Y., 2014. Seroprevalence of Merkel cell polyomavirus in the general rural population of Anyang, China. PLoS One 9. https://doi.org/10.1371/journal.pone.0106430.

Zhang, E., Bell, A.J., Wilkie, G.S., Suárez, N.M., Batini, C., Veal, C.D., Armendáriz-Castillo, I., Neumann, R., Cotton, V.E., Huang, Y., Porteous, D.J., Jarrett, R.F., Davison, A.J., Royle, N.J., 2017. Inherited chromosomally integrated human herpesvirus 6 genomes are ancient, intact, and potentially able to reactivate from telomeres. J. Virol. 91. https://doi.org/10.1128/JVI.01137-17. 011377-e1217.

CHAPTER

Evolutionary Responses to Infectious Disease

9

G. Cochran and H. Harpending
University of Utah, Salt Lake City, UT, United States

1. Introduction

Humans have a typical mammalian immune system, with three components: external barriers, the innate immune system, and the adaptive immune system. External barriers include physical barriers such as skin and mucosal surfaces as well as antibacterial secretions like lysozyme and defensins.

The second component, the innate immune system, responds quickly to attack but is not tailored to the specific attacking pathogen. It has built-in features that recognize and attack pathogens using pattern recognition receptors, which are triggered by characteristic molecular signatures associated with certain classes of disease organisms, such as the lipopolysaccharides found in the cell walls of Gram-negative bacteria. It also includes some very specific defenses against particular pathogens. The innate immune system knows that certain molecules are danger signs, but it does not learn. Its knowledge was generated by natural selection rather than individual experience, rather like the fear of snakes that is especially easily invoked in humans.

The third component, the adaptive immune system, can tailor responses to a specific pathogen and retains the ability to rapidly respond to future visitations by that pathogen. It acts as an individual immunological memory.

The human immune system has defenses which work against practically any conceivable pathogen (the adaptive system), defenses that are pretuned against traditional classes of pathogens such as bacteria, RNA and DNA viruses, protozoa, parasitic worms, and fungi, and a number of other specialized defenses that are aimed at specific pathogens. For example, there are genes that defend against herpes simplex (Sancho-Shimizu et al., 2007), Epstein−Barr virus (Rigaud et al., 2006), and certain dangerous strains of human papilloma virus (Ramoz et al., 2002). People with two broken copies of such a gene almost inevitably have serious or lethal infections of the associated pathogen.

Some human defenses protect against regional pathogens. Mainly that means malaria, which we discuss at length later, but we also know of a built-in defense that is effective against *Trypanosoma brucei*, the cause of a common African trypanosomal infection known as nagana in livestock. The molecule is apolipoprotein L1 (APOL-1) (Pays et al., 2006) 4 a lipoprotein: normally such molecules transport lipids, but this one also plays a role in killing trypanosomes, by chopping holes in their surface membranes. APOL-1 has, we suspect, played an important role in human evolution. First, it is a pretty clear signal that humans evolved in sub-Saharan Africa, since that is where nagana is found. It

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00015-X
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

probably says something about our preferred habitats in the early days of human evolution, since the tsetse flies carrying the disease are mainly found near rivers and lakes, in forests along watercourses, and in wooded savanna. Tsetse flies are not common in arboreal environments, and interestingly, chimpanzees do not have this defense (Puente et al., 2005) although gorillas do.

Innate human nagana defenses are not just another signpost pointing to Africa. They also suggest possible ecological explanations for some patterns in the fossil record, for example, for the fact that Neanderthal remains have never been found in Africa. Given a long time living in areas without tsetse flies, several hundred thousand years, Neanderthals probably had nonfunctional versions of this defense, due to relaxed selection and mutation accumulation. The loss of this defense (and most likely of other defenses against specific African pathogens) would have made it almost impossible to expand back into Africa. Since Africa was pathogen-rich compared to Europe and southwest Asia, the main Neanderthal homeland, the situation was not symmetrical. Hominids could leave Africa, but they could not go home again.

Interestingly, there are APOL-1 variants that protect against sleeping sickness and also cause human illness. There are two common mutations (G1 and G2). G2 gives strong protection against infection by the East African strain of sleeping sickness, but appears to increase vulnerability to the West African strain. G1 doesn't do much against the East African strain, but allow asymptomatic carriage of the West African strain—better survival, rather than resistance to infection, very much like hemoglobin variants and malaria. The problem is that people with two copies of these protective variants are much more likely to experience kidney failure in later life. African Americans are four to five times more likely to end up on dialysis than whites (Friedman and Pollak, 2020).

It is not just that Africa was pathogen-rich, compared to Eurasia, which might suggest that African immune systems were simply better. The optimal set of defenses depends on the environment: African defenses were better in Africa, while a different set would have been optimal for Eurasia. For one thing, Eurasia did have its own local pathogens, such as scrub typhus (Traub and Wisseman, 1974). Second, since strong immune defenses have costs and risks (autoimmune diseases such as lupus), lower total pathogen pressure in Eurasia would have favored a turned-down, not quite as aggressive immune response. Anatomically modern humans expanding out of Africa would certainly have developed the appropriate immune defenses, given time, but it may not have taken all that long. We now know that modern humans picked up useful variants through admixture with archaic humans such as Neanderthals (Green et al., 2010) and Denisovans (Reich et al., 2010) that had inhabited parts of Eurasia for several hundred thousand years, certainly enough time to adapt to local conditions. A Neanderthal variant of STAT2, an innate immune gene involved in interferon response to viral infection, is widely distributed in Eurasia and reaches a high frequency in New Guinea. Non-Africans also picked up new HLA alleles from both Neanderthals and Denisovans (Racimo et al., 2015). In retrospect, this should have been anticipated: HLA alleles are extremely varied, likely because of rare-allele advantage generated by host−pathogen coevolution. Since HLA alleles are so varied, humans are preadapted to new and different HLA variants. They are effectively interchangeable parts. This is in contrast to more typical archaic alleles apparently there had been enough evolutionary divergence so that the typical Neanderthal variant did not quite fit in modern humans, nor ours in theirs.

Many known or suspected genetic responses to infectious disease in humans are loss of function mutations, and damaged or broken genes. This is the case for nearly all the known malaria adaptations: we will discuss falciparum malaria in some detail in the following. Other protective variants that are known or suspected are also damaged versions of the wild type. Most tropical Africans have the

Duffy-negative chemokine receptor, which confers protection against vivax malaria. For many years, it has been thought that the deletion had no negative consequences, but (Reich et al., 2009) have shown that the Duffy-negative allele itself causes a significant reduction in neutrophil count in people of African descent. This reduced white cell count which almost certainly has some disadvantages.

2. Parasites as Our Friends

Of course parasites can also be helpful. Invading modern humans may have carried diseases that hit archaic humans harder during the original modern human diaspora out of Africa. This effect helped the Europeans expand into the New World. Venereal diseases would be good candidates, since they can propagate successfully at the low human densities typical of hunter-gatherers. Directly transmitted crowd diseases, on the other hand, would not have been particularly devastating to the low density archaic populations encountered in the Levant and Europe during that early expansion (as opposed to the high-density agricultural populations in the New World encountered by the Spanish).

Parasites may also have contributed to human success in hunting. We picked up three species of taenid tapeworms from African predators (another sign of our origin) several hundred thousand years ago (Hoberg, 2006) one from hyenas and two from lions. These tapeworms, like many other parasites, have a complex life cycle, forming cysts in herbivores (the intermediate host) and reaching maturity in the carnivorous definitive host. Obviously, their interests conflict with those of their intermediate hosts, since they benefit when their host is eaten.

In such a situation, there is an evolutionary incentive for parasites to manipulate the behavior of their intermediate host, for example, by making them easier to catch (Lagrue and Poulin, 2010) This may be the case for toxoplasma, a protozoan that uses many herbivores as intermediate hosts and cats as the definitive host. Toxoplasma has been shown to cause fearlessness in rats and mice (Berdoyet al., 2000) and to cause chimpanzees to develop a perverse attraction to leopards (Poirotte et al., 2016): who benefits? There is evidence that Echinococcus, another taeniid tape worm with canid definitive hosts, increases predation on its intermediate hosts (e.g., moose). Those human tapeworms may have played an important role in human hunting success, particularly in the olden days when human weapons and hunting skills were far less sophisticated than those used by contemporary hunter-gatherers. Before agriculture, those tapeworms used wild pigs and ungulates as intermediate hosts. Now they cycle through domesticated pigs and cows, suggesting another way in which those parasites could have aided humans by fostering domestication (Ivy Smith, personal communication).

Wild boars are quite formidable, but the aurochs (the wild ancestor of domesticated cattle) were simply terrifying, being 2 m high at the shoulder and weighing over a ton. Domestication sounds difficult and dangerous, but it might have been easier if a parasite was, for its own reasons, reducing the aurochs' fear of humans.

3. Demography and Parasites

Pathogen dynamics can have a major influence on long-term demographics—and the other way around of course. Pathogens typically require a minimum number of hosts in fairly close proximity (called the critical community size) in order to survive. Consider measles: it is infectious for no more than 10 days, and survivors have lifelong immunity. Clearly measles can only flourish in a situation where

there is a steady supply of fresh, never-infected hosts, that is to say children. Because of these facts, measles has a critical community size of roughly a quarter of a million people: it could not have existed in its present form back in hunter-gatherer times, since there were no such large population concentrations (Black, 1966).

At the opposite extreme, chickenpox, after infecting children, lingers in nerve ganglia. It often recurs much later in life as shingles, which causes excruciating pain. Children can catch chickenpox from their grandfather's case of shingles. Thus, due to its persistence and ability to wait, chickenpox has a critical community size around 1000 (Black et al., 1974).

These facts about infectious disease imply certain things about our ancestral demographics. For one thing, a population crash would have usually been followed by a boom, partly because resources become more abundant in such situations but also because infectious diseases become less important at low population densities. A mega-crash, one in which humans had a brush with extinction, could thus have had a silver lining: one or more human-specific parasites could have gone extinct. If those parasites had imposed a heavy fitness burden, humans would have flourished after the crash. Something similar (a bottleneck in space rather than time) happens sometimes when a species colonizes a new continent, the settler population is too small to carry along key parasites and thrives to a surprising degree in its new home.

Africa is rich in human pathogens. Since we originated there, African pathogens have had a long time to adapt to humans and other primates. We mentioned that populations such as Neanderthals that spent a long time outside of Africa probably lost defenses that would have been necessary in Africa, and thus could not go back. The other side of this coin is that those vigorous defenses against African pathogens had costs, costs that were no longer necessary in cooler climates. Leaving Africa may have had substantial payoffs, first for archaic humans in Eurasia and later for anatomically modern humans.

4. Agriculture

The biggest demographic change ever experienced by humans was the population explosion made possible by the development of agriculture. Our numbers increased by factors of 50−100, which had a fundamental (and highly unpleasant) impact on human infectious disease. Pathogens that already infected humans became more common and had greater impacts on fitness, while new pathogens arose that could only spread in high-density populations—crowd diseases. We acquired most of these crowd diseases from other animals. A few originated in the animals we domesticated, while a number of others came from African primates. Some probably evolved from older human pathogens moving into newly available ecological niches.

The human genome responded to the new disease pressures, and we have observed the resulting changes in many components of the immune system. The 35delG mutation of connexin-26 causes deafness in homozygotes but also changes characteristics of the skin (thicker) and sweat (saltier): it may protect against infections of the skin such as erysipelas (Meyer et al., 2002). It is also a common cause of deafness in homozygotes. There is evidence of selection on a number of genes in the innate immune system such as CR-1 (a malaria defense) and some of the Toll-like receptors (TLRs) that recognize characteristic pathogen molecules. Some changes, such as the mutations causing familiar Mediterranean fever (Aksentijevich et al., 1999) and alpha-1-antitrypsin deficiency (Lomas, 2006), loosen protective restrictions on some of the more aggressive components of the immune system—you

might compare these to unleashing the police, always a dangerous thing to do. There have been changes in the adaptive immune system as well, particularly in the major histocompatibility complex (MHC). We have recognized many of these changes because they cause serious Mendelian diseases that would hardly have reached such high frequencies unless there was some form of heterozygote advantage. Genomic scans have discovered other adaptive changes that do not have such high costs.

It is an odd fact that we seem to see fewer of these expensive disease defenses in East Asia, particularly considering only those that defend against something other than falciparum malaria. We know of no obvious reason why this should be so: conceivably it might a result of looking harder at European genetics (ascertainment bias) but right now it is something of a puzzle.

5. Some Lessons from Malaria

The case of malaria illustrates a number of general principles about the relationship between infectious disease, biological evolution, and social evolution in humans. We discuss aspects of malaria biology in some detail, but much the same story could be told for other infectious diseases of humans, for example, yellow fever. Falciparum malaria is the most serious human infectious disease and has been the strongest and best understood selective force acting on humans over the past few 1000 years. This selection pressure operated in the peoples of the Old World tropics and subtropics—but not elsewhere—and so caused those populations to diverge from the rest of humanity in some ways (Pennington et al., 2009). The most dramatic impact has been the rise to high frequency of many protective alleles. A number of those alleles (the best-studied ones) are overdominant and cause major health problems in homozygotes. The sickle-cell mutation is the most famous protective allele. Heterozygotes gain substantial protection against falciparum malaria while homozygotes suffer from a severe anemia that is usually lethal in childhood without modern medical treatment. Even so, it continues to cause substantial morbidity and mortality. It is the most common lethal mutation in humans, with a gene frequency of around 10% or more in many populations of tropical Africa.

There are a number of similar protective polymorphisms which are also disease alleles. Some change the hemoglobin molecule, either by amino acid changes (like hemoglobin C and hemoglobin E) or by changing the relative numbers of hemoglobin subunits, as in the thalassemias. Others change the red cell in different ways, Interfering with its metabolism (glucose-6-pohophodehydrogenase (G6PD) deficiency) or altering membrane proteins (Melanesian ovalocytosis). It seems likely that falciparum malaria has existed in its present form for 5000 years or less. The approximate age of some of the protective polymorphisms has been determined, and they all seem to be younger than that (Ohashi et al., 2004; Saunders et al., 2005). Increasingly, researchers are discovering alleles favored by malaria selection that apparently do not cause disease, not even in homozygotes. Some affect familiar targets, such as the red cell membrane (glycophorin C (Maier et al., 2003)) and type O blood (Rowe et al., 2009). We also see variants of immune-system molecules such as Cd36 (Pain et al., 2001) and CR-1 (Cockburn et al., 2004). This trend of discoveries is likely to continue, and we should eventually observe malaria-induced changes in the frequencies of many alleles, even those that have only weak effects on resistance. That is the typical pattern seen in artificial selection experiments. Strong selection for any trait other than fitness itself causes negative changes in other traits—so resistance to malaria has most likely had significant costs. Obviously, we know of the costs of many that take the form of Mendelian diseases, but there are likely others as well.

Falciparum malaria's unusual virulence can be explained in part by its means of transmission. Natural selection favors low virulence in many infectious diseases that are spread directly from person to person, since immobilizing the host interferes with transmission. Since malaria is a vector-borne disease (spread by mosquitoes), a severe infection can still spread, even if the host is bedridden (Ewald, 1994). If high parasite blood counts increase the probability of transmission, severe infection may be a favored strategy. The other major kind of human malaria, *Plasmodium vivax*, is also mosquito-borne. It is a fairly serious disease, although much less so than falciparum. It is often found in temperate climates, where it must survive winters without active mosquitoes. In order to do so, it has the ability to hide in liver cells for long periods, in some cases for decades. Of course, this strategy would not work if the host died, which explains why vivax malaria has relatively low virulence. *Plasmodium falciparum* mainly exists in warmer climates where mosquito transmission occurs through most or all of the year, so that it can keep moving to new hosts.

Malaria has another characteristic that increases its severity. Unlike most other pathogens, malaria repeatedly switches its surface proteins. A single parasite clone has about 60 antigenic variants and thus can stay ahead of the immune system for a year or more, while greater variety in the parasite population as a whole means that a single infection does not result in lasting immunity (Scherfet al., 2008). This defensive tactic of malaria has made the development of an effective vaccine very difficult: the first moderately effective vaccine became available in 2022.

Selection for malaria resistance in humans illustrates several key evolutionary principles: some of these are very well known, while others are not so obvious. First, it shows that adaptive evolution is a continuing process in humans, one that can cause significant changes over historical time and whose direction is not the same in every population.

This may have been especially the case over the Holocene, during which humans experienced substantial climate change, were exposed to the selective pressures associated with agriculture, and greatly increased in number.

Malaria selection is also a clear example of convergent evolution. The protective alleles in Southeast Asia are entirely different from those in Africa: some are different mutations of genes that have produced defensive alleles in Africa (e.g., G6PD deficiency), while others involve different genes. One sees the same thing in artificial selection experiments: the phenotypic changes are similar in different lines experiencing the same selective pressures (people in both Africa and South East Asia are resistant to malaria) but the genetic details are in general different. Another point is that strong selection evokes changes in many genes, changes that are concentrated in a few metabolic paths. In this case, we know of many polymorphisms that affect the red cell and hemoglobin, as well as a number that result in immunological changes. We have seen arguments that this pattern is somehow unparsimonious: one sweep might be caused by strong selective pressures, but surely not many! But in fact strong selection is likely to cause frequency changes in many genes—basically, every gene that significantly affects the trait under selection is a candidate for an adaptive mutation.

These convergent adaptations also show us something about the way in which advantageous alleles have spread through populations. Particular protective alleles have spread through much of sub-Saharan Africa, across New Guinea, or throughout the coastal regions of the Mediterranean but few have managed to cross the Sahara Desert or move between India and Southeast Asia. Strong geographical barriers have prevented high-fitness alleles from spreading to all the places they would have worked—that and limited time—and thus local protective variants took their place. Evolution was faster than gene flow.

Many of these protective alleles are overdominant, since homozygotes suffer from serious disease. Overdominance means that the heterozygote has higher fitness than the homozygote: such alleles never go to fixation. A recessive lethal like sickle cell is clearly overdominant, but some of the other defensive alleles that do not cause obvious disease may also have lower fitness in homozygotes. A number of domesticated animals also have overdominant alleles that are products of recent strong selection, such as myostatin mutations in whippets and cattle. This may be a general feature of strong selection: many of the sweeping alleles generated by such selection may therefore reach maximum frequencies well under 100%.

Another interesting point comes from a simple thought experiment: there must have been a time when falciparum malaria had not existed for long and protective alleles were as yet rare. In those days, sickle cell heterozygotes (for example) should have had a larger fitness advantage, relative to the population average, than they do today, since in those days the average person had no other protective alleles. Today, on the other hand, someone in Africa who does not carry the sickle cell allele is likely to have a number of other protective alleles—alpha thalassemia, G6PD deficiency, and so on. Africans who do not carry sickle cell are still far more resistant to falciparum malaria than northern Europeans or Amerindians (Mackinnon et al., 2005). Hence, the fitness advantage of being a sickle cell carrier (which was as high as 20% in recent centuries) must have been even larger thousands of years ago. This means that the rate of growth, and the equilibrium frequency, if overdominant, of every allele that protected against malaria slowed down as time passed, as the population acquired more and more resistance to malaria from other alleles. This effect can also stop a selective sweep short of fixation.

We think that falciparum malaria has had another interesting effect on human evolution, in that it often kept populations below the Malthusian limit—that is, kept population density below the level at which resource limitations would have stopped further growth. In a Malthusian situation, resources are short and individuals compete for them. Selection in that situation favors efficient use of available resources, which would involve improvements in metabolic and work efficiency—basically, farmers who can plow more acres per calorie. It also favors paternal investment. At the limit, you end up with hard-working peasant couples (both father and mother) who can just barely manage to raise enough food for themselves and the two children who replace them in the next generation.

In a sub-Malthusian ecology, where factors like disease and/or violence keep the population well below the subsistence limit, selection pushes in a different direction. Here the limiting factor might be health rather than wealth. Disease resistance in a mate could be more valuable than land, hence a father's genetic quality might be important than his provisioning ability. In much of Africa today, women do most of the farm work: this low level of paternal investment is only possible when resources are plentiful. Female self-sufficiency combined with a high value placed on genetic quality favors polygyny (multiple wives), since man's genes are more important than his wages. Polygyny is more common in West Africa than anywhere else.

6. The Columbian Exchange

Old World infectious diseases had a terrible impact upon the natives of the New World after contact, causing wave after wave of epidemics. The Amerindian population collapsed, decreasing by something like 90%. Invaders from the Old World were less vulnerable to these diseases, which aided them in their conquests.

Why, though, were these Amerindians more vulnerable? It wasn't just lack of personal immunity from previous infections—the course of the new diseases was different, more severe than in the Old World, and that stayed true for generations. Mortality from these diseases was often much higher among Amerindians than it was in the colonizing peoples (Blake and Garcia-Blanco, 2014; Vågene et al., 2018; Riley, 2010).

Immigrants from the Old World (mainly from Europe and Africa) often had some degree of immunity from previous personal exposure, but, more importantly, were more resistant because their ancestral populations had been exposed to these pathogens for long periods, and had developed genetic defenses. Such adaptation to be expected, as shown in the recent work that traces genetic changes in Europeans due to the Black Death (Klunk et al., 2022). And the Old World had *more* infectious diseases than the New World.

Again, why did the Old World have so many more infectious diseases? It has been suggested (Wolfe et al., 2007; McNeill, 1977) that most of these diseases originated from domesticated animals, and the disparity is explained by the fact that the Old world had many domesticated animals, while the New World had only a few. But this does not seem to have been the main reason: measles seems to have originated from rinderpest (Düx et al., 2020), a disease of cattle, but most other serious infectious diseases (such malaria, smallpox, bubonic plague, yellow fever) do not seem to have originated in this way.

It looks as if a number of different causes contribute to the larger number of infectious diseases in the Old World. One is that there were and are close relatives of humans there: gorillas and chimpanzees. It seem that falciparum malaria, the most virulent kind, was originally an infection of gorillas (Loy et al., 2017), while HIV originated in chimpanzees.

Second, the main ancestral group of Amerindians apparently spent quite a while in Beringia, during the Ice Age, before entering the Americas.: Most vector-borne diseases could not persist in those conditions, nor could any pathogens with a very large critical community size.

Third, many diseases could not spread in humans until population density became high enough for their critical community size. Such population concentrations only occurred after the development of agriculture, but agriculture developed earlier in the Old World. High human population density has existed longer there, and in multiple areas.

7. Disease and Standard of Living in Preindustrial Societies: A Simple Model

We can elaborate the role of disease in shaping human cultural diversity with a simple model. Disease in a population that would otherwise be Malthusian, that is to say resource limited, can have the effect of reducing the population size, leading to an increase in the standard of living of those who remain. A familiar example is the prosperity and high wages in Europe following the massive human die off following the great plague epidemics.

We start with a small group of 1000 colonizers in an empty environment. Initially the population is at such a low density that there is no competition for resources among people. Births and deaths occur at a constant per-person rate. There is no age structure—no youth, nor old people—so everyone is subject to the same rates; these assumptions make algebraic models easy and they reflect well what happens in more realistic (but more complicated) models. Plausible generic values for low-technology

human populations are 50 births per 1000 people per year and 30 deaths per 1000 people per year. The difference, 20 per 1000 per year, is the intrinsic growth rate, 20 per 1000 or 2 per 100, 2% per year. In the absence of any limitation the population grows according to this rate exactly like money at 8 compound interest. After a generation of 25 years the population size would be 1000 (1.02)25 or 1640 people. This population would double in about 35 years and would double slightly more than 14 times in 500 years to an implausibly large size of nearly 20 million people.

The customary models in population geneticists focus on gene frequency change, and mean fitness, population growth rate, is normalized away in the equations for genetic change. Here we need to acknowledge the demographic consequences of gene change and retain the mean-fitness parameter in order to study the interaction of demography and genetics. The mean fitness of 1.02 per year (or 1.64 per generation) occurs at low population densities, but gradually declines as the population grows to the carrying capacity. At this limit, the mean fitness is just 1.0, and the population remains stationary.

What are the long-term implications of this modest rate of growth? The rate of 2% per year is commonplace among human populations yet a growth of 2,000,000% over 500 years seems and is outlandish. Early in the process resources would become scarce and the rate of growth would slow. Assuming the initial colony occupied 100 square miles, the expanded population after five centuries would need to occupy nearly 2,000,000 (2 million) square miles, about the area of Argentina or Kazakhstan. This is explosive growth in historical time but it corresponds to everyday population growth today in many low-technology societies. We know that over the long period from the modern human diaspora out of Africa about 45,000 years ago to the industrial revolution about 200 years ago human numbers grew but at long term rates far below our modest 2% per year. On this long time scale, they hardly grew at all. It is likely that most of the time populations were growing at rates like our 2%, perhaps slower, but that there were frequent catastrophic events like wars, famines, and plagues that cut population sizes back.

8. Population Limitation

There is a convenient and standard way to make a model of population limited by resources, called the logistic model. This may not be very accurate but it is simple and, given our poor understanding of detailed dynamics, more than good enough. The idea is that there is some carrying capacity K of the environment. Populations below the carrying capacity in size can grow while populations above the carrying capacity decline until they reach K. If we write Pt for population in some generation t and Ptþ1 for population the following generation then simple population growth like compound interest, called geometric, follows this formula: Ptþ1 ¼ Pt ð1 þ Rþ where R is just the intrinsic growth rate and (1 þ R) is the mean fitness. We write R ¼ 0.641 in the expression for the intrinsic growth rate since a rate of 2% per year corresponds to growth of 64.1% per generation. The logistic model specifies that the growth rate R is damped by the current ratio of population to carrying capacity: Ptþ1 ¼ Pt 1 þ R 1 Pt K p0190. In an empty environment without intraspecific competition population P is much less than carrying capacity K and population growth is almost the same as the simple geometric case. But as population increases the ratio P/K becomes significant, growth slows down, and eventually stops when population reaches carrying capacity, that is, when P ¼ K. If the carrying capacity of the environment into which our population moved were 10,000 people, then the population would grow at a decreasing rate to reach 10,000.

What if the carrying capacity is not static but increases with the number of people? For example, we might imagine that more people bring more farmland under cultivation so that K itself changes. It turns out (Cohen, 1995) that nothing much changes if the increase in carrying capacity K is proportionally less than the increase in population P as would happen if the best land were cleared first while lower and lower quality land were subsequently brought under cultivation. The population still approaches a (new larger) carrying capacity so that as equilibrium is approached population P is equal to carrying capacity. The end result is that the standard of living, by which we mean the ratio of resources to people K/P, is still unity. There are more people but they are not living any better than they did before the new land was cleared. Such a population, limited by resources, is referred to as a Malthusian population.

An interesting variant of this model is to introduce a new source of mortality, perhaps disease or warfare (Armstrong and Gilpn ,1977). In areas of central Africa with high levels of falciparum malaria, malaria's fitness cost may be around 25%: that means that with malaria an average individual will leave 25% fewer living descendants one generation later. With a growth rate of 20 per 1000 per year an average individual has 1.64 daughters one generation later. If malaria now decreases fitness by 25% the average individual will only have 75% of 1.64 or 1.23 daughters one generation later. In terms of annual rates the malaria cuts population growth from 2% to 0.8% per year. (Notice that we count only daughters since our model is of a simple population that does not take into account sexual reproduction.) Now we can consider the fixed carrying capacity K and examine the consequences for the population and for individual well-being. The algebraic model now becomes (writing M for the extra density-independent death rate, from malaria in our example but also likely to be from violence and local warfare): Ptþ1 ¼ Pt 1 þ R RPt K.

We can find the equilibrium population, that is, the population that would remain unchanging in this environment with the extra mortality. We simply set Ptþ1 equal to Pt, rearrange some terms, and find that the new equilibrium is at P ¼ K ðR Mþ R. We substitute our assumed values, an intrinsic growth rate R of 0.64 and an extra mortality rate M of 0.25 to obtain P K ¼ 0:39 0:64w0:61: The population now equilibrates at 61% of the old carrying capacity. A more interesting way to summarize what we have found by manipulating the logistic model is in terms of the standard of living, where a value of 1 means the bare subsistence minimum compatible with life and the maintenance and population size and a value of, say, 5, means that there is five times the subsistence minimum amount of resources available to the average person. In our model population the standard of living is the reciprocal of 0.61 or 1.6. There are more than half again as many resources per person as there were before malaria appeared. What this means on the ground is that people do not have to work very hard to get enough to eat, that there is fruit on the trees for plucking, and that there are not great labor demands on anyone. Those who survive the malaria enjoy a much higher standard of living. Fig. 9.1 is a 0 2 4 6 8 10 12 14 16 18 20 0 5000 10000 Generation.

Clark (2007) points out that the medieval English had a higher standard than the medieval Japanese because there was much more sewage and filth in England, and so a heavier burden of disease. This extra disease translated, as in our malaria example, to a lower population density and higher standard of living.

What are the social consequences of this new disease for low-technology human populations? The most important immediate consequence is that there are plentiful resources for everyone and so, following the nature of the creature, males withdraw from subsistence work as they find that they can simply parasitize women for food. In much of central Africa, Oceania, and the Americas the result is or

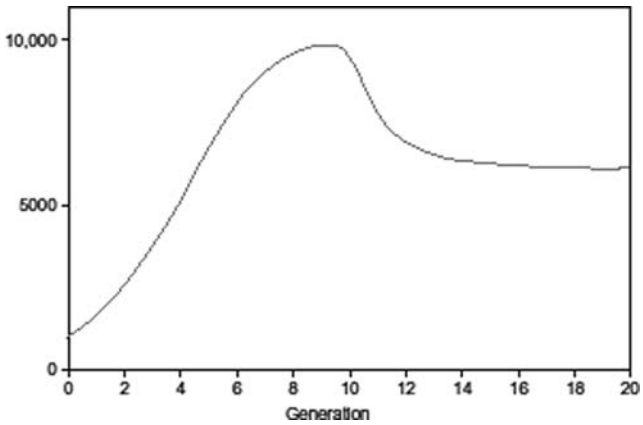


FIGURE 9.1

Model of population size over time of a population of 1000 introduced into an empty area with a carrying capacity of 10,000. After 250 years (10 generations) falciparum malaria appears, and population size quickly drops to about 6500. Generation is on the x-axis, population size on the y-axis. Simulation of this process using plausible numbers for a low-technology human population. The population grows tenfold from 1000 to 10,000 people in 10 generations, then quickly shrinks to the new equilibrium size of 6100 people after the introduction of falciparum malaria.

has been societies in which men do not do anything very useful and women provision themselves, their children, and the men. The euphemism in economics for this kind of society is "female farming system." Left free of the demands of subsistence the men start hanging out together, perhaps even all moving into a village men's house (not so uncommon in Africa). This leads to local and regional raiding and warfare, and an entrenched culture of local violence.

9. Disease, Mating, and Reproductive Strategy

Several decades ago (Hamilton and Zuk, 1982) showed a correlation in North American passerines between parasite burden and gaudiness. Their model was that a slowly changing parasite load leads to parent-offspring correlations in ducks in parasite resistance leading to mating preferences for bright colors as signals of that resistance. Subsequent literature suggests that a similar phenomenon occurs in human societies (Gangestad and Buss, 1993; Low, 1990).

The underlying logic is clear enough, much of it similar to that of the earlier discussion of endemic malaria. Human females, like all mammals, may obtain provisioning from a male for herself and her offspring. This is the pattern in many settled agricultural societies where male subsistence labor is necessary for successful reproduction: these are so-called dad societies. On the other hand, in societies that are far below the Malthusian limit, female mate preferences are more likely to favor males other than good providers. For example, in these societies, often characterized by chronic local raiding and warfare, males protect females from other males. In places with high endemic disease loads then, as in ducks, disease resistance can be heritable so that females prefer to mate with males with "good genes"

rather than with males who are "good providers." Of course if females are selecting males with good genes rather than males who are good providers then this is an open door for polygyny. A peasant farmer would have great difficulty provisioning several families but no difficulty at all simply mating with several females. Traditional societies of tropical Africa are indeed mostly polygynous.

There seem to be several important ecological routes to quasi-stable high standard of living non-Malthusian societies. One is warfare: in many well-documented cases deaths by violence approach a quarter to one-half of all deaths with the result that human densities remain below any purely subsistence population limit (Keeley, 1996). Much of highland New Guinea appears to be a classic example of this cultural ecosystem. Nothing much seems to have changed there in many millennia. The rough broken terrain contributes to the persistence of this system since the terrain makes it nearly impossible for an effective constabulary to suppress the chronic violence.

Another route to such a society is through a high burden of endemic disease as in the malaria example. Malaria is of course the classic case but yellow fever, hookworm, and many others should have similar effects. Much of sub-Saharan Africa is described by economists as "female farming systems," a euphemism for societies where men essentially parasitize women for subsistence while they commit more effort than males do elsewhere to subtle and not-so-subtle male-male competition. As females prefer males who appear healthy (for their "good genes") they may be selection in males to accommodate this preference, that is, there should be sexual selection for appearance. Several recently described myostatin mutants in Africa (Saunders et al., 2006) are probably recently evolved signals of male quality.

The worldwide fall of fertility rates following the industrial revolution in northern nations suggests that a stable non-Malthusian world is attainable without the unpleasantness and misery of violence and infectious disease. Meanwhile it is important to understand that a disease like falciparum malaria not only causes much human misery directly, it also leaves in its wake damaging genetic traces that may take hundreds of generations to dissipate. It also leaves in its wake a social order likely not so well suited for modern industrial society. While tough fierce physically attractive males may be favored in a social system where there are adequate resources for females to do all the provisioning, these same are not going to do so well in a subsistence ecology that demands hard agricultural labor and actively sanctions violent behavior.

10. Prosperity and the Postindustrial Era Mortality Decline

It has become apparent in the last decade that evolution in humans is an ongoing process that is even speeding up in the face of drastic cultural changes and the large number of humans on earth, each of whom is a potential target for mutations, including favorable mutations (Hawks et al., 2007). Clark (2007) has proposed that genetic change during the millennium or so before the industrial revolution led to essentially a new version of humans that made the revolution possible. The greatest change in human economic history since the origins of agriculture, the industrial revolution of around 1800, released our species from Malthusian constraints as income growth suddenly outstripped population growth.

This revolution in human society was accompanied by many changes, and no one has a very clear idea of how they are related to each other. Clark, whose focus is on Great Britain, emphasizes these changes as follows:

- A decline in propensity to violence, especially male violence. In the nations of Europe homicide rates fell by one to two orders of magnitude in the millennium before 1800. (Eisner, 2001, 2003). In many preindustrial societies violent males enjoyed a reproductive advantage through greater access to mates, but that advantage turned into a severe disadvantage in settled agricultural societies with effective constabularies.
- Declining interest rates reflecting declining time preference. People were more and more inclined to delay gratification.
- An increasing affinity for work.
- A strong correlation between wealth and reproductive success of males.

However, at the same time, according to Clark, there were other equally profound changes:

- A striking mortality decline, the cause of which is not well understood. Civil engineering and vaccination are often suggested as causes of this decline but the evidence is not very clear. The decline may also reflect in part genetic adaptation to new kinds of infectious disease.
- Birth rates fell drastically. The fall of birth rates lagged the fall in death rates by several decades. This is the so-called "demographic transition" from high mortality and high fertility to low mortality and low fertility. This led to the relationship between wealth and fertility to reverse, as it is today in industrial societies. Today wealthier people have fewer surviving offspring, and this reversal was the immediate precursor of the popular eugenics movement of the late 19th and early 20th centuries.

What is the role, if any, of the mortality decline in this seismic shift in the nature of society? A possibility is that the decline in pressure from infectious disease freed up much of the genome to evolve in new directions determined by the new social environment. Most of the genetic adaptations to infectious disease that we (think that we) understand involve major or minor damage to genes and to individuals. We discussed the sickle cell adaptation to malaria earlier with its high, purely genetic death toll on homozygotes, but there are many parallel adaptations to falciparum malaria that are known and almost certainly many more that we do not yet understand. In aggregate they must impose a large genetic burden on populations with a history of living with malaria.

As the prevalence of infectious disease declined in pre- and postindustrial societies there may have been widespread relaxation of the selection maintaining these damaging genetic polymorphisms, with the effect of releasing these constraints on the genome and facilitating selection to move phenotypes in different, more favorable directions in the phenotype space.

References

Aksentijevich, I., Torosyan, Y., Samuels, J., Centola, M., Pras, E., Chae, J.J., et al., 1999. Mutation and haplotype studies of familial Mediterranean fever reveal new ancestral relationships and evidence for a high carrier frequency with reduced penetrance in the Ashkenazi Jewish population. Am. J. Hum. Genet. 64 (4), 949–962.

Armstrong, R.A., Gilpin, M.E., 1977. Evolution in a time-varying environment. Science 195 (4278), 591–592.

Berdoy, M., Webster, J.P., Macdonald, D.W., August 7, 2000. Fatal attraction in rats infected with Toxoplasma gondii. Proc. Biol. Sci. 267 (1452), 1591–1594.

Black, F.L., July 1966. Measles endemicity in insular populations: critical community size and its evolutionary implication. J. Theor. Biol. 11 (2), 207–211.

Black, F.L., Hierholzer, W.J., Pinheiro, F.D.K., Evans, A.S., Woodall, J.P., Opton, E.M., et al., 1974. Evidence for persistence of infectious agents in isolated human populations. Am. J. Epidemiol. 100 (3), 230.

Blake, L.E., Garcia-Blanco, M.A., 2014. Human genetic variation and yellow fever mortality during 19th century US epidemics. mBio 5 (3), 10−1128.

Clark, G., 2007. A farewell to alms: a brief economic history of the world [Internet]. In: The Princeton Economic History of the Western World. Princeton University Press, Princeton. Available from: http://www.loc.gov/catdir/toc/ecip0715/2007015166.html.

Cockburn, I.A., Mackinnon, M.J., O'Donnell, A., Allen, S.J., Moulds, J.M., Baisor, M., et al., January 6, 2004. A human complement receptor 1 polymorphism that reduces Plasmodium falciparum rosetting confers protection against severe malaria. Proc. Natl. Acad. Sci. U S A 101 (1), 272−277.

Cohen, J.E., 1995. How many people can the earth support? Sciences 35 (6), 18−23.

Düx, A., Lequime, S., Patrono, L.V., Vrancken, B., Boral, S., Gogarten, J.F., Hilbig, A., et al., June 19, 2020. Measles virus and rinderpest virus divergence dated to the sixth century BCE. Science 368 (6497), 1367−1370.

Eisner, M., 2001. Modernization, self-control and lethal violence. The long-term dynamics of European homicide rates in theoretical perspective. Br. J. Criminol. 41 (4), 618−638.

Eisner, M., 2003. Long-term historical trends in violent crime. Crim. Justice 83−142.

Ewald, P.W., 1994. Evolution of Infectious Disease. Oxford University Press, USA.

Friedman, D.J., Pollak, M.R., 2020. APOL1 and kidney disease: from genetics to biology. Annu. Rev. Physiol. 82, 323−342.

Gangestad, S.W., Buss, D.M., 1993. Pathogen Prevalence and Human Mate Preferences. Elsevier.

Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., et al., May 7, 2010. A draft sequence of the Neandertal genome. Science 328 (5979), 710−722.

Hamilton, W.D., Zuk, M., 1982. Heritable true fitness and bright birds: a role for parasites? Science 218 (4570), 384e7.

Hawks, J., Wang, E.T., Cochran, G.M., Harpending, H.C., Moyzis, R.K., 2007. Recent acceleration of human adaptive evolution. Proc. Natl. Acad. Sci. U.S.A. 104 (52), 20753.

Hoberg, E.P., 2006. Phylogeny of Taenia: species definitions and origins of human parasites. Parasitol. Int. 55 (Suppl. 1), S23−S30.

Keeley, L.H., 1996. War before Civilization.

Klunk, J., Vilgalys, T.P., Demeure, C.E., Cheng, X., Shiratori, M., Madej, J., Beau, R., et al., November 2022. Evolution of immune genes is associated with the black death. Nature 611 (7935), 312−319.

Lagrue, C., Poulin, R., April 2010. Manipulative parasites in the world of veterinary science: implications for epidemiology and pathology. Vet. J. 184 (1), 9−13.

Lomas, D.A., 2006. The selective advantage of a1-antitrypsin deficiency. Am. J. Respir. Crit. Care Med. 173 (10), 1072−1077.

Low, B.S., 1990. Marriage systems and pathogen stress in human societies. Am. Zool. 30 (2), 325−340.

Loy, D.E., Liu, W., Li, Y., Learn, G.H., Plenderleith, L.J., Sundararaman, S.A., Sharp, P.M., Hahn, B.H., February 1, 2017. Out of Africa: origins and evolution of the human malaria parasites Plasmodium falciparum and Plasmodium vivax. Int. J. Parasitol. 47 (2), 87−97.

Mackinnon, M.J., Mwangi, T.W., Snow, R.W., Marsh, K., Williams, T.N., 2005. Heritability of malaria in Africa. PLoS Med. 2 (12), e340.

Maier, A.G., Duraisingh, M.T., Reeder, J.C., Patel, S.S., Kazura, J.W., Zimmerman, P.A., et al., January 2003. Plasmodium falciparum erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. Nat. Med. 9 (1), 87−92.

McNeill, W., 1977. Plagues and Peoples. Knopf Doubleday Publishing Group.

Meyer, C.G., Amedofu, G.K., Brandner, J.M., Pohland, D., Timmann, C., Horstmann, R.D., 2002. Selection for deafness? Nat. Med. 8, 1332−1333.

Ohashi, J., Naka, I., Patarapotikul, J., Hananantachai, H., Brittenham, G., Looareesuwan, S., et al., June 2004. Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. Am. J. Hum. Genet. 74 (6), 1198−1208.

Pain, A., Urban, B.C., Kai, O., Casals-Pascual, C., Shafi, J., Marsh, K., et al., 2001. A non-sense mutation in Cd36 gene is associated with protection from severe malaria. Lancet 357 (9267), 1502−1503.

Pays, E., Vanhollebeke, B., Vanhamme, L., Paturiaux-Hanocq, F., Nolan, D.P., Pérez-Morga, D., 2006. The trypanolytic factor of human serum. Nat. Rev. Microbiol. 4 (6), 477−486.

Pennington, R., Gatenbee, C., Kennedy, B., Harpending, H., Cochran, G., 2009. Group differences in proneness to inflammation. Infect. Genet. Evol. 9 (6), 1371−1380.

Poirotte, C., Kappeler, P.M., Ngoubangoye, B., Bourgeois, S., Moussodji, M., Charpentier, M.J., 2016. Morbid attraction to leopard urine in Toxoplasma-infected chimpanzees. Curr. Biol. 26 (3), R98−R99.

Puente, X.S., Gutiérrez-Fernandez, A., Ordõnez, G.R., Hillier, L.W., Lopez-Otín, C., December 2005. Comparative genomic analysis of human and chimpanzee proteases. Genomics 86 (6), 638−647.

Racimo, F., Sankararaman, S., Nielsen, R., Huerta-Sanchez, E., 2015. Evidence for archaic adaptive introgression in humans. Nat. Rev. Genet. 16 (6), 359−371.

Ramoz, N., Rueda, L.-A., Bouadjar, B., Montoya, L.-S., Orth, G., Favre, M., December 2002. Mutations in two adjacent novel genes are associated with epidermodysplasia verruciformis. Nat. Genet. 32 (4), 579−581.

Reich, D., Nalls, M.A., Kao, W.H.L., Akylbekova, E.L., Tandon, A., Patterson, N., et al., January 30, 2009. Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. PLoS Genet. 5 (1), e1000360.

Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., et al., December 23, 2010. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468 (7327), 1053−1060.

Rigaud, S., Fondaneche, M.C., Lambert, N., Pasquier, B., Mateo, V., Soulas, P., et al., 2006. XIAP deficiency in humans causes an X-linked lymphoproliferative syndrome. Nature 444 (7115), 110−114.

Riley, J.C., October 2010. Smallpox and American Indians revisited. J. Hist. Med. Allied Sci. 65 (4), 445−477.

Rowe, J.A., Opi, D.H., Williams, T.N., November 2009. Blood groups and malaria: fresh insights into pathogenesis and identification of targets for intervention. Curr. Opin. Hematol. 16 (6), 480−487.

Sancho-Shimizu, V., Zhang, S.-Y., Abel, L., Tardieu, M., Rozenberg, F., Jouanguy, E., et al., December 2007. Genetic susceptibility to herpes simplex virus 1 encephalitis in mice and humans. Curr. Opin. Allergy Clin. Immunol. 7 (6), 495−505.

Saunders, M.A., Slatkin, M., Garner, C., Hammer, M.F., Nachman, M.W., 2005. The extent of linkage disequilibrium caused by selection on G6PD in humans. Genetics 171 (3), 1219.

Saunders, M.A., Good, J.M., Lawrence, E.C., Ferrell, R.E., Li, W.H., Nachman, M.W., 2006. Human adaptive evolution at myostatin (GDF8), a regulator of muscle growth. Am. J. Hum. Genet. 79, 1089.

Scherf, A., Lopez-Rubio, J.J., Riviere, L., October 2008. Antigenic variation in Plasmodium falciparum. Annu. Rev. Microbiol. 62 (1), 445−470.

Traub, R., Wisseman, C.L., 1974. Review article: the ecology of chigger-borne rickettsiosis (scrub typhus). J. Med. Entomol. 11 (3), 237−303.

Vågene, A.J., et al., March 2018. Salmonella enterica genomes from victims of a major sixteenth-century epidemic in Mexico. Nat. Ecol. Evol. 2 (3), 3.

Wolfe, N.D., Dunavan, C.P., Diamond, J., May 2007. Origins of major human infectious diseases. Nature 447 (7142), 279−283.

This page intentionally left blank

CHAPTER

Infectious Disease Genomics 10

Yu-Tsueng Liu

Division of Infectious Diseases, Department of Medicine, University of California San Diego, La Jolla, CA, United States

1. Introduction

The history and development of infectious disease genomics are closely associated with the Human Genome Project (HGP) (Watson, 1990). A series of important discussions about the HGP were made from 1985 to 1986 (Dulbecco, 1986; Watson, 1990), which led to the appointment of a special National Research Council (NRC) committee by the National Academy of Sciences to address needs and concerns, such as impact, leadership, and funding sources. The committee recommended that the United States begin the HGP in 1988 (NRC, 1988). They emphasized the need for technological improvements in the efficiency of gene mapping, sequencing, and data analysis capabilities. In order to understand potential functions of human genes through comparative sequence analyses, they also advised that the HGP must not be restricted to the human genome and should include model organisms such as mouse, bacteria, yeast, fruit fly, and worm. In the meantime, the Office of Technology Assessment (OTA) of the US Congress also issued a similar report to support the HGP (OTA, 1988). In 1990, the Department of Energy (DOE) and the National Institutes of Health (NIH) jointly presented an initial 5-year plan for the HGP (DHHS and DOE, 1990). In October 1993, the Sanger Center/Institute (Hinxton, UK) was officially opened to join the HGP. The cost of DNA sequencing was about \$2 to \$5 per base in 1990, and the initial aim was to reduce the costs to less than \$0.50 per base before large-scale sequencing (DHHS and DOE, 1990). The sequencing cost gradually declined during the subsequent years. In 2004, the National Human Genome Research Institute (NHGRI) challenged scientists to achieve a \$100,000 human genome (3 Gb/haploid genome) by 2009 and a \$1000 genome by 2014 to meet the needs of genomic medicine. In early 2014, Illumina announced that the company would begin producing a new system to deliver full coverage human genomes for less than \$1000 (Hayden, 2014). Remarkably, sequencing costs have often dropped faster than Moore's law, and could be less than \$20 per human genome in 2030 (Denny and Collins, 2021).

The first complete genome to be sequenced was the phiX174 bacteriophage (5.4 kb) by Sanger's group in 1977 (Sanger et al., 1977). The complete genome sequence of the SV40 polyomavirus (5.2 kb) was published in 1978 (Fiers et al., 1978; Reddy et al., 1978). The human Epstein–Barr virus (170 kb) genome was determined in 1984 (Baer et al., 1984). The first completed free-living organism genome was *Haemophilus influenza* (1.8 Mb), sequenced through a whole-genome shotgun approach in 1995 (Fleischmann et al., 1995). The second sequenced bacterial genome, *Mycoplasma genitalium*

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00018-5

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

(600 kb), was completed in less than a month in the same year using the same approach (Smith, 2004). The DOE was the first to start a microbial genome program (MGP) as a companion to its HGP in 1994 (Stone, 1994). The initial focus was on nonpathogenic microbes. Along with the development of the HGP, there was exponential growth of the number of completely sequenced genomes. One of the major interests was to sequence organisms that are important to human health and commercial activities. With the explosion in the number of sequenced genomes, thanks to the development of next-generation sequencing methods, many genome-based studies have become popular (Gorbalenya et al., 2020; Liu, 2017).

Completed and ongoing genome projects (Rigden and Fernández, 2020) (https://www.oxfordjournals.org/nar/database/c/) provide enormous opportunities for the discovery of novel vaccines and drug targets against human pathogens, as well as the improvement of diagnosis and discovery of infectious agents and the development of new strategies for invertebrate vector control. Specific examples will be provided to illustrate how the information provided by various genome projects may help achieve the goal of promoting human health.

2. Vaccine Target

Meningococcal isolates produce 1 of 13 antigenically distinct capsular polysaccharides, but only five (A, B, C, W135, and Y) are commonly associated with disease (Lo et al., 2009). The polysaccharide capsule is important for meningococci to escape from complement-mediated killing. While conventional vaccines consisting of the conjugation of capsular polysaccharides to carrier proteins for meningococcus serogroups A, C, Y, and W-135 have been clinically successful, the same approach has failed to produce clinically useful vaccine for serogroup B (MenB). The capsule polysaccharide (α2-8 N-acetylneuraminic acid) of MenB is identical to human polysialic acid, and therefore is poorly immunogenic (Finne et al., 1987). Alternatively, vaccines consisting of outer-membrane vesicles (OMV) have been successfully developed to control MenB outbreaks in areas where epidemics are dominated by one particular strain (Bjune et al., 1991; Boslego et al., 1995; Jackson et al., 2009; Sierra et al., 1991). The most significant limitation of this type of vaccine is that the immune response is strain-specific, mostly directed against the porin protein, PorA, which varies substantially in both expression level and sequence across strains (Martin et al., 2000; Pizza et al., 2000).

With the completion of the genome sequence of a virulent MenB strain, a "reverse vaccinology" approach was applied for the development of a universal MenB vaccine by Novartis (Giuliani et al., 2006; Pizza et al., 2000; Tettelin et al., 2000). Through bioinformatic searching for surface exposed antigens, which may be the most suitable vaccine candidates due to their potential to be readily recognized by the immune system, 570 open reading frames (ORFs) were selected from a total of 2158 ORFs of the MC58 genome. Eventually, five antigens were chosen as vaccine components, based on a series of criteria including the ability of candidates to be expressed in *E. coli* as recombinant proteins (350 candidates), the confirmation of surface exposure by immunological analyses, the ability of induced protective antibodies in experimental animals (28 candidates) and the conservation of antigens within a panel of diverse meningococcal strains, primarily the disease-associated MenB strains (Giuliani et al., 2006; Pizza et al., 2000; Rinaudo et al., 2009). The vaccine formulation consists of an fHBP-GNA2091 fusion protein, a GNA2132-GNA1030 fusion protein, NadA, and OMV from the New Zealand MeNZB vaccine strain, which contains the immunogenic PorA. Initial phase II clinical

results in adults and infants showed that this vaccine could induce a protective immune response against three diverse MenB strains in 89%–96% of subjects following three vaccinations and in 93%–100% after four vaccinations (Rinaudo et al., 2009). This vaccine (Bexsero) has been recently approved in the United States and in more than 30 other countries (Bruno et al., 2015).

3. New Drug Discovery

Natural products, especially microbial secondary metabolites, are an important source of bioactive compounds. Actinomycetes have been a main source of natural-product discovery in bacteria. Consequently, the high rediscovery rate of known compounds and scaffolds was inevitable with activity-based screening. Genome mining of gene clusters that produce secondary metabolites has been a new approach to overcome this problem. For example, an antibiotic, clostrubin, was discovered through searching novel compounds from *Clostridium beijerinckii* due to the presence of several cryptic gene clusters for secondary metabolite biosynthesis (Pidot et al., 2014).

Genome mining starts with a genome-wide search for highly conserved members of the required biosynthesis gene cluster. Computational programs that support the prediction of operons help to assign boundaries of newly identified biosynthesis gene clusters. Recently, a large-scale, high-throughput genome mining for the genetic potential for producing phosphonic acids by screening more than 10,000 actinomycetes was achieved (Ju et al., 2015). It was believed that phosphonates would have greater potential to become pharmaceuticals, with a past commercialization rate of 15% (3/20) for products such as fosfomycin, compared to a 0.1% average for natural products as a whole (Berdy, 2012; Metcalf and van der Donk, 2009). In addition, bioinformatical discovery of phosphonate biosynthetic loci has been well established, as all but two previously characterized phosphonate biosynthetic pathways start with phosphoenolpyruvate (PEP) mutase that is encoded by *pepM*. Among 10,000 actinomycetes, only 278 strains were confirmed to have *pepM* by PCR screening and genome sequencing. A diverse collection of phosphonate biosynthetic gene clusters were identified within these strains. Remarkably, 55 out of the 64 distinct clusters direct the synthesis of unknown compounds. Characterization of strains within five of these groups resulted in the discovery of argolaphos, and other interesting compounds, including valinophos, and phosphonocystoximate. Argolaphos showed broad-spectrum antibacterial activity against *Salmonella typhimurium*, *Escherichia coli*, and *Staphylococcus aureus* (*S. aureus*).

4. Drug Target

Targeting an essential pathway is a necessary but not sufficient requirement for an effective antimicrobial agent (Brinster et al., 2009). Identification of essential genes in a completely sequenced genome has been actively pursued with various approaches (Hutchison et al., 1999; Ji et al., 2001). The indispensable fatty acid synthase (FAS) pathway in bacteria has been regarded as a promising target for the development of antimicrobial agents (Wright and Reynolds, 2007). The subcellular organization of the fatty acid biosynthesis components is different between mammals (type I FAS) and bacteria (dissociated type II FAS), which raises the likelihood of host specificity of the targeting drugs. Comparison of the available genome sequences of various species of prokaryotes reveals highly conserved FAS II systems, suggesting that the antimicrobial agent can be broad-spectrum (Zhang

et al., 2003). In addition, through computational analyses, new members of the FAS II system have been discovered in different bacterial species (Heath and Rock, 2000; Marrakchi et al., 2002). One of the protein components in this system, FabI, is the target of an antituberculosis drug, isonizid, and a general antibacterial and antifungal agent, triclosan (Banerjee et al., 1994; Levy et al., 1999; Zhang et al., 2006).

Through a systematic screening of 250,000 natural product extracts, a Merck team identified a potent and broad spectrum antibiotic, platensimycin, which is derived from *Streptomyces platensis* and a selective FabF/B inhibitor in FAS II system (Wang et al., 2006). Treatment with platensimycin eradicated *S. aureus* infection in mice. Platensimycin did not have cross-resistance to other antibiotic-resistant strains in vitro, including methicillin-resistant *S. aureus*, vancomycin-intermediate *S. aureus*, and vancomycin-resistant enterococci. No toxicity was observed using a cultured human cell line and the activity of platensimycin was not affected by the presence of human serum in this study. However, the FAS II system appears to be dispensable for another Gram-positive bacterium, *Streptococcus agalactiae*, when exogenous fatty acids are available, such as in human serum (Balemans et al., 2010; Brinster et al., 2009). The susceptibility to inhibitors targeting the FAS II system indicates heterogeneity in fatty acid synthesis or in acquiring exogenous fatty acids among Gram-positive pathogens (Balemans et al., 2010).

Afabicin (Debio 1450, AFN 1720) is a prodrug of Afabicin desphosphono (Debio 1452, AFN-1252) that specifically inhibits staphylococcal FabI. The initial lead was discovered by GlaxoSmithKline, which took a "genomics" approach to antibacterial discovery between 1995 and 2001 (Payne et al., 2007). Phase II clinical trials (NCT01519492 and NCT02426918) have been conducted to treat acute bacterial skin and skin structure infections (ABSSSI) due to staphylococci (Hafkin et al., 2015; Wittke et al., 2020).

5. Therapeutic Response and Drug Resistance

Emergence of drug-resistant malaria to chloroquine in the 1950s and to sulfadoxine-pyrimethamine in the 1960s occurred from western Cambodia to the Greater Mekong subregion (GMSR, including Cambodia, Laos, Myanmar, Thailand, and Vietnam) and to Africa. The recent findings of artemisinin-resistant malaria in Cambodia and GMSR raised concern regarding the global spread of these parasites. While a number of studies, including population genetics and laboratory-based investigations, were conducted, no reliable molecular marker was identified until a major breakthrough reported in early 2014 (Ariey et al., 2014). Clinical artemisinin resistance has been defined as a reduction of parasite clearance rate, which is expressed as an increase of parasite clearance half-life or a persistence of microscopically detectable parasites 3 days after artemisinin-based combination therapy (ACT). Although artemisinin was thought to have broad-stage specificity against malaria throughout the life cycle, it was shown that artemisinin-resistant parasites only had a decrease in artemisinin susceptibility at ring stages, which was demonstrated by the ring-stage survival assay ($RSA_{0-3\ h}$) (Witkowski et al., 2013).

An in vitro laboratory-based approach was conducted at a time when population-based genome-wide association studies (GWAS) did not clearly identify the genes responsible for artemisinin-resistance (Ariey et al., 2014). For 5 years, an artemisinin-resistant F32-ART5 parasite line was selected for by culturing an artemisinin-sensitive F32-Tanzania clone under a dose-escalating, 125-cycle regimen of artemisinin. Eight mutations in seven genes were eventually selected from the

result based on whole-genome sequence analysis of F32-ART5 and F32-TEM (its sibling clone cultured without artemisinin) at 460× and 500× average nucleotide coverage, respectively. To examine whether these in vitro selected mutations were associated with artemisinin resistance in Cambodia, sequence polymorphism in all 7 genes were analyzed from 49 culture-adapted clinical isolates related to their $RSA_{0-3\ h}$. Only polymorphisms of a single gene, K13-propeller, showed a significant association with $RSA_{0-3\ h}$ survival rates. In total, four mutant alleles, each harboring a single nonsynonymous SNP (Y493H, R539T, I543T, and C580Y) within a kelch repeat of the C-terminal K13-propeller domain, were identified. To confirm that K13-propeller polymorphism is a molecular marker of clinical artemisinin resistance, parasite clearance half-lives in patients were correlated with their K13 alleles. Of the 150 patients, 72 carried parasites with a wild-type allele and the others carried parasites with only one of the three single non-synonymous SNPs in the K13-propeller: C580Y (n = 51), R539T (n = 6), and Y493H (n = 21). The parasite clearance half-life in patients with wild-type parasites was significantly shorter (median 3.30 h) than those with these three mutant alleles (median 6.28−7.19 h). Subsequently, clinical studies have validated the association between K13 propeller mutations and artemisinin resistance (Ashley et al., 2014; Miotto et al., 2015; Tun et al., 2015; Wicht et al., 2020).

6. Vector Control

Early mathematical models for malaria control suggested that the most vulnerable element in the malaria cycle was survivorship of adult female mosquitos (Enayati and Hemingway, 2010; Macdonald, 1957). Therefore, insect control is an important part of reducing transmission. The use of DDT as an indoor residual spray in the global malaria eradication program from 1957 to 1969 reduced the population at risk of malaria to approximately 50% by 1975 compared with 77% in 1900 (Enayati and Hemingway, 2010; Hay et al., 2004). Engineering genetically modified mosquitoes refractory to malaria infection appeared to be an alternative approach (Curtis, 1968), given the environmental impact of DDT and the emergence of insecticide-resistant insects. The Vector Biology Network (VBN) was formed in 1989 and proposed a 20-year plan with the WHO in 2001 to achieve three major goals: 1. to develop basic tools for the stable transformation of anopheline mosquitoes by the year 2000; 2. to engineer a mosquito incapable of carrying the malaria parasite by 2005; and 3. to run controlled experiments to test how to drive the engineered genotype into wild mosquito populations by 2010 (Alphey et al., 2002; Beaty et al., 2009; Morel et al., 2002). While some proof-of-concept experiments had been achieved for the first two aims by 2002, when the *Anopheles gambiae* genome was completely sequenced (Catteruccia et al., 2000; Ito et al., 2002), further development has been relatively slow (Marshall and Taylor, 2009). Recently, the accomplishment of applying targeted gene-drive systems has shown encouraging progress.

Genetically modified mosquitoes (GMMs) to prevent pathogen transmission include those that decrease the number of mosquito vectors in a region (population suppression) and that make local mosquitoes less competent at transmitting a pathogen (population replacement). Through interbreeding with wild mosquitoes, transgenes from certain GMMs may spread and persist in targeted populations. Drive, or gene drive, is a phenomenon in sexually reproducing organisms whereby a particular gene can bias inheritance in its favor so that the gene becomes more prevalent in the population over successive generations (Alphey et al., 2020; World Health Organization, 2021). Gene

drive elements, such as transposons and homing endonucleases, are common in nature, and can be synthetically created. Among various synthetic gene-drive approaches, CRISPR editing is the most powerful tool. In this system, the transgenes carry an endonuclease, generally Cas9, and target-specific guide RNA (gRNA). The gRNA bound Cas9 is directed to the targeted genome to generate chromosome break. When the break is restored by homology-directed repair, using the targeted transgene as a template, all offspring will inherit the gene drive.

Only female mosquitos transmit diseases such as malaria, dengue and Zika. Anopheles males are heterogametic XY and females are XX. A population crash of the species *Drosophila neotestacea* in the field has been documented because of a natural X-chromosome distorter that generated a female-only population (Pinzone and Dyer, 2013). Therefore, sex-distorter drives have been proposed for pest or vector control. A successful male-biased distorter system that utilizes a CRISPR-gene drive has been reported recently (Kyrou et al., 2018). In *Anopheles gambiae*, the gene doublesex (Agdsx) encodes two alternatively spliced transcripts, dsx-female (AgdsxF) and dsx-male (AgdsxM), which control sex differentiation. The exon 5 is in the female, but not male, transcript. A CRISPR−Cas9-targeted disruption at the intron 4−exon 5 boundary for blocking the formation of functional AgdsxF did not affect male development or fertility. In contrast, females having homozygous targeted alleles showed an intersex phenotype with complete sterility. Remarkably, this gene drive spread rapidly in caged mosquitoes, reaching 100% prevalence within 7−11 generations, resulting in gradually reducing egg production and eventually total population collapse. Because of functional constraints of the target sequence, no selection of alleles resistant to the gene drive happened. While Cas9-resistant variants did arise in each generation at the target site, they did not prevent the spread of the drive.

Genomic loci of the *Anopheles gambiae* responsible for *Plasmodium falciparum* resistance have been identified through surveying a mosquito population in a West African malaria transmission zone (Riehle et al., 2006). A candidate gene, *Anopheles Plasmodium*-responsive leucine-rich repeat 1 (APL1), was discovered. Subsequently, other resistant genes have also been identified (Blandin et al., 2009; Povelones et al., 2009). Studying the genetic basis of resistance to malaria parasites and immunity of the mosquito vector will be important for controlling malaria transmission (Bartholomay and Michel, 2018).

While horizontal gene transfer is a well-known evolutionary mechanism for prokaryotes, gene transfer from plant to insect has also recently been discovered (Xia et al., 2021). The whitefly *Bemisia tabaci* is an agricultural pest. It serves as a vector for several plant pathogenic viruses. Strikingly, the whitefly can neutralize toxic plant phenolic glucosides, enabled by hijacking the plant-derived phenolic glucoside malonyltransferase gene BtPMaT1. This was experimentally verified by genetically transforming tomato plants to produce small interfering RNAs for silencing BtPMaT1, which eliminated the whiteflies' detoxification ability. This study provides new insights not only related to vector control but also to the evolutional strategy of horizontal gene transfer between remarkably different species.

7. Clinical Application

Perhaps the most immediate impact of a completely sequenced pathogen genome is for infectious disease diagnosis. The information may be of great importance to public health when a newly emerged or re-emerged pathogen is discovered. A few examples will be described.

A novel swine-origin influenza A virus (S-OIV) emerged in spring of 2009 in Mexico, and subsequently was discovered in specimens from two unrelated children in the San Diego area in mid-April 2009 (CDC, 2009; Dawood et al., 2009). The samples were positive for influenza A but negative for both human H1 and H3 subtypes. The complete genome sequence and a real-time PCR-based diagnostic assay were released to the public in late April. The outbreak evolved rapidly, and the World Health Organization (WHO) declared the highest Phase 6 worldwide pandemic alert on June 11, 2009. S-OIV has 3 genome segments (HA, NP, NS) from the classic North American swine (H1N1) lineage, 2 segments (PB2, PA) from the North American avian lineage, 1 segment (PB1) from seasonal H3N2 and, most notably, 2 segments (NA, M) from the Eurasian swine (H1N1) lineage (Dawood et al., 2009). With the available influenza genome database, diagnostic assays to distinguish from previous seasonal H1N1, H3N2, and S-OIV can be easily accomplished (Lu et al., 2009).

A comprehensive pathogen genome database is not only useful for infectious disease diagnosis but also for novel pathogen discovery (Liu, 2008). Homologous sequences within the same family or among different family members are important for new pathogen identification, even with the advent of third generation sequencing technology (Munroe and Harris, 2010). De novo pathogen discovery may also be complicated by coexisting microorganisms, such as commensal bacteria in the human body. Without prior knowledge of these microorganisms, one may be misled.

In 2003, a microarray-based assay, designated Virochip, was used to help discover the SARS conoronavirus (Wang et al., 2003). The Virochip contained the most highly conserved 70mer sequences from every fully sequenced reference viral genome in GenBank. The computational search for conservation was performed across all known viral families. A microarray hybridized with a reaction derived from a viral isolate cultivated from a SARS patient revealed that the strongest hybridizing array elements belonging to families astroviridae and coronaviridae. Alignment of the oligonucleotide probes with the highest signals showed that all four hybridizing oligonucleotides from the astroviridae and one oligonucleotide from avian infectious bronchitis virus, an avian coronavirus, shared a core consensus motif spanning 33 nucleotides. Interestingly, it had been known previously through bioinformatic analyses that this sequence was present in the 3′ UTR of all astroviruses, avian infectious bronchitis virus and an equine rhinovirus (Jonassen et al., 1998). Therefore, a new member of the coronavirus was identified through the unique hybridizing pattern and subsequent confirmations.

In 2019−20, a new SARS coronavirus, the etiological agent of COVID-19 (Coronavirus disease 2019), emerged. For the first time in the history of medicine, the molecular evolution of a pandemic virus could be observed in real time during the COVID-19 outbreak (Burioni and Topol, 2021). SARS-CoV-2 was first isolated in late December 2019 and completely sequenced in January 2020 (Zhou et al., 2020). While the evolutionary rate of SARS-CoV-2 is relatively low ($\sim 1 \times 10^{-3}$ substitutions/site/year) as a result of its proof-reading RNA polymerase, the massive number of infected populations presented abundant opportunities for SARS-CoV-2 mutations to occur (Boehm et al., 2021). As the early pandemic progressed, a variant with D614G substitution of the spike (S) protein receptor that led to 20% increased infectiousness developed into the global dominant strain (Korber et al., 2020; Zhou et al., 2021). The alpha (B.1.1.7) variant, with 50% increase in transmissibility beyond that of D614G, appeared in late 2020. In parallel, other variants also evolved in several regions around the world (Krause et al., 2021). Soon after, an even more transmissible delta (B.1.617.2) variant with enhanced ability for immuno-evasion became predominant in many countries, and gradually took over the alpha variant (Fischer et al., 2021; Tao et al., 2021). Pathogen surveillance by whole genome sequencing has become a reality (Armstrong et al., 2019; Becker et al., 2021). Remarkably, COVID-19 vaccines were

successfully developed at a record speed of less than a year, and their efficacy against emerging variants are continuously evaluated (Corey et al., 2020; Lurie et al., 2020; Slaoui and Hepburn, 2020; Tao et al., 2021).

Many human pathogens have animal origins and emerge through cross-species transmission events. Recently, a meta-transcriptomic sequencing study was conducted on more than 2400 wild bats, rodents, and shrews (Chen et al., 2023). The study expanded our understanding of mammalian viruses in local habitats by identifying 669 viruses, including 534 novel viruses, and elucidating their origins. It also revealed high levels of phylogenetic diversity and cross-species virus transmission events. Shrews were found to have a high richness of viruses, including many invertebrate-associated viruses with multiorgan distributions. On the other hand, rodents carried viruses that had a greater capacity for host jumping.

The finding of the seventh human oncogenic virus, Merkel cell polyomavirus (MCV) (Feng et al., 2008), in 2008 is another example of why conserved sequences are important for novel pathogen discovery. MCV is the etiological agent of Merkel cell carcinoma (MCC), which is a rare but aggressive skin cancer of neuroendocrine origin. Two cDNA libraries derived from MCC tumors were subjected to high throughput sequencing by a next-generation Roche/454 sequencer. Nearly 400,000 sequence reads were generated. The majority (99.4%) of the sequences derived from human origin were removed from further analyses. Only one of the remaining 2395 cDNA was homologous to the T antigen of two known polyomaviruses. One additional cDNA was subsequently identified to be part of the MCV sequence, once the complete viral sequence was known. Later analyses showed that 80% (8/10) of the MCC had integrated MCV in the human genome. Monoclonal viral integration was revealed by the patterns of Southern blot analysis. Only 8%−16% of control tissues had low copy number of MCV infection.

In 2015, an interesting and unexpected discovery of the malignant transformation of *Hymenolepis nana* (*H. nana*), a human tape worm, in a human host was reported by conventional and next-generation sequencing approaches (Muehlenbachs et al., 2015). Initially, examination of a 41-year-old HIV-infected man revealed extensive lymphadenopathy. *H. nana* eggs and *Blastocystis hominis* cysts were found in stool. The disease progressed to death, despite antiparasitic and antiretroviral treatment. Histological examination of biopsied lymph nodes revealed proliferative cells with overt malignant features. They were monomorphic with morphologic features characteristic of stem cells (a high nucleus-to-cytoplasm ratio). However, the small cell size (<10 μm) suggested infection with an unfamiliar, possibly unicellular, eukaryotic organism. Infection with a plasmodial slime mold rather than *H. nana* was considered because of the prominent syncytia formation and the primitive appearance of the atypical cells but lack of architecture identifiable as tapeworm tissue. PCR screening suggested that these cells were *H. nana*. Next-generation genome sequencing and comparative analysis revealed *H. nana* variants harboring mutations typically found in cancer.

Nowadays, next-generation sequencing technologies are gradually being applied for diagnosis and monitoring of infectious diseases, including genotypic resistance testing, direct detection of unknown disease-associated pathogens without culture, investigation of microbial population diversity in the host and strain typing (Lefterova et al., 2015). While promising, next-generation sequencing approaches for clinical diagnosis require further improvement for automation, standardization of technical and bioinformatic procedures and other practical issues, such as cost and turnaround time.

8. Conclusion

While we can expect that the efforts of a variety of genome projects may improve human health, the socioeconomic issues that are not discussed in this chapter may be substantial. In addition, the tremendous quantity of information derived from these projects will also be a challenge for scientists, as well as nonscientists, to follow and understand.

References

Alphey, L., Beard, C.B., Billingsley, P., Coetzee, M., Crisanti, A., Curtis, C., Eggleston, P., Godfray, C., Hemingway, J., Jacobs-Lorena, M., James, A.A., Kafatos, F.C., Mukwaya, L.G., Paton, M., Powell, J.R., Schneider, W., Scott, T.W., Sina, B., Sinden, R., Sinkins, S., Spielman, A., Toure, Y., Collins, F.H., 2002. Malaria control with genetically manipulated insect vectors. Science 298, 119−121.

Alphey, L.S., Crisanti, A., Randazzo, F.F., Akbari, O.S., 2020. Opinion: standardizing the definition of gene drive. Proc. Natl. Acad. Sci. U. S. A. 117, 30864−30867.

Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A.C., Khim, N., Kim, S., Duru, V., Bouchier, C., Ma, L., Lim, P., Leang, R., Duong, S., Sreng, S., Suon, S., Chuor, C.M., Bout, D.M., Menard, S., Rogers, W.O., Genton, B., Fandeur, T., Miotto, O., Ringwald, P., Le Bras, J., Berry, A., Barale, J.C., Fairhurst, R.M., Benoit-Vical, F., Mercereau-Puijalon, O., Menard, D., 2014. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature 505, 50−55.

Armstrong, G.L., MacCannell, D.R., Taylor, J., Carleton, H.A., Neuhaus, E.B., Bradbury, R.S., Posey, J.E., Gwinn, M., 2019. Pathogen genomics in public health. N. Engl. J. Med. 381, 2569−2580.

Ashley, E.A., Dhorda, M., Fairhurst, R.M., Amaratunga, C., Lim, P., Suon, S., Sreng, S., Anderson, J.M., Mao, S., Sam, B., Sopha, C., Chuor, C.M., Nguon, C., Sovannaroth, S., Pukrittayakamee, S., Jittamala, P., Chotivanich, K., Chutasmit, K., Suchatsoonthorn, C., Runcharoen, R., Hien, T.T., Thuy-Nhien, N.T., Thanh, N.V., Phu, N.H., Htut, Y., Han, K.T., Aye, K.H., Mokuolu, O.A., Olaosebikan, R.R., Folaranmi, O.O., Mayxay, M., Khanthavong, M., Hongvanthong, B., Newton, P.N., Onyamboko, M.A., Fanello, C.I., Tshefu, A.K., Mishra, N., Valecha, N., Phyo, A.P., Nosten, F., Yi, P., Tripura, R., Borrmann, S., Bashraheil, M., Peshu, J., Faiz, M.A., Ghose, A., Hossain, M.A., Samad, R., Rahman, M.R., Hasan, M.M., Islam, A., Miotto, O., Amato, R., MacInnis, B., Stalker, J., Kwiatkowski, D.P., Bozdech, Z., Jeeyapant, A., Cheah, P.Y., Sakulthaew, T., Chalk, J., Intharabut, B., Silamut, K., Lee, S.J., Vihokhern, B., Kunasol, C., Imwong, M., Tarning, J., Taylor, W.J., Yeung, S., Woodrow, C.J., Flegg, J.A., Das, D., Smith, J., Venkatesan, M., Plowe, C.V., Stepniewska, K., Guerin, P.J., Dondorp, A.M., Day, N.P., White, N.J., Tracking Resistance to Artemisinin, C, 2014. Spread of artemisinin resistance in Plasmodium falciparum malaria. N. Engl. J. Med. 371, 411−423.

Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Farrell, P.J., Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Seguin, C., et al., 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature 310, 207−211.

Balemans, W., Lounis, N., Gilissen, R., Guillemont, J., Simmen, K., Andries, K., Koul, A., 2010. Essentiality of FASII pathway for *Staphylococcus aureus*. Nature 463, E3 discussion E4.

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K.S., Wilson, T., Collins, D., de Lisle, G., Jacobs Jr., W.R., 1994. inhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. Science 263, 227−230.

Bartholomay, L.C., Michel, K., 2018. Mosquito immunobiology: the intersection of vector health and vector competence. Annu. Rev. Entomol. 63, 145−167.

Beaty, B.J., Prager, D.J., James, A.A., Jacobs-Lorena, M., Miller, L.H., Law, J.H., Collins, F.H., Kafatos, F.C., 2009. From Tucson to genomics and transgenics: the vector biology network and the emergence of modern vector biology. PLoS Neglected Trop. Dis. 3, e343.

Becker, S.J., Taylor, J., Sharfstein, J.M., 2021. Identifying and tracking SARS-CoV-2 variants - a challenge and an opportunity. N. Engl. J. Med. 385, 389−391.

Berdy, J., 2012. Thoughts and facts about antibiotics: where we are now and where we are heading. J. Antibiot. 65, 385−395.

Bjune, G., Hoiby, E.A., Gronnesby, J.K., Arnesen, O., Fredriksen, J.H., Halstensen, A., Holten, E., Lindbak, A.K., Nokleby, H., Rosenqvist, E., et al., 1991. Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway. Lancet 338, 1093−1096.

Blandin, S.A., Wang-Sattler, R., Lamacchia, M., Gagneur, J., Lycett, G., Ning, Y., Levashina, E.A., Steinmetz, L.M., 2009. Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. Science 326, 147−150.

Boehm, E., Kronig, I., Neher, R.A., Eckerle, I., Vetter, P., Kaiser, L., Geneva Centre for Emerging Viral, D, 2021. Novel SARS-CoV-2 variants: the pandemics within the pandemic. Clin. Microbiol. Infect. 27, 1109−1117.

Boslego, J., Garcia, J., Cruz, C., Zollinger, W., Brandt, B., Ruiz, S., Martinez, M., Arthur, J., Underwood, P., Silva, W., et al., 1995. Efficacy, safety, and immunogenicity of a meningococcal group B (15:P1.3) outer membrane protein vaccine in Iquique, Chile. Chilean National Committee for Meningococcal Disease. Vaccine 13, 821−829.

Brinster, S., Lamberet, G., Staels, B., Trieu-Cuot, P., Gruss, A., Poyart, C., 2009. Type II fatty acid synthesis is not a suitable antibiotic target for Gram-positive pathogens. Nature 458, 83−86.

Bruno, L., Cortese, M., Rappuoli, R., Merola, M., 2015. Lessons from reverse vaccinology for viral vaccine design. Curr. Opin. Virol. 11, 89−97.

Burioni, R., Topol, E.J., 2021. Has SARS-CoV-2 reached peak fitness? Nat. Med. 27, 1323−1324.

Catteruccia, F., Nolan, T., Loukeris, T.G., Blass, C., Savakis, C., Kafatos, F.C., Crisanti, A., 2000. Stable germline transformation of the malaria mosquito *Anopheles stephensi*. Nature 405, 959−962.

CDC, 2009. Swine influenza A (H1N1) infection in two children–Southern California, March-April 2009. MMWR Morb. Mortal. Wkly. Rep. 58, 400−402.

Chen, Y.M., Hu, S.J., Lin, X.D., Tian, J.H., Lv, J.X., Wang, M.R., Luo, X.Q., Pei, Y.Y., Hu, R.X., Song, Z.G., Holmes, E.C., Zhang, Y.Z., 2023. Host traits shape virome composition and virus transmission in wild small mammals. Cell 186, 4662−4675.

Corey, L., Mascola, J.R., Fauci, A.S., Collins, F.S., 2020. A strategic approach to COVID-19 vaccine R&D. Science 368, 948−950.

Curtis, C.F., 1968. Possible use of translocations to fix desirable genes in insect pest populations. Nature 218, 368−369.

Dawood, F.S., Jain, S., Finelli, L., Shaw, M.W., Lindstrom, S., Garten, R.J., Gubareva, L.V., Xu, X., Bridges, C.B., Uyeki, T.M., 2009. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N. Engl. J. Med. 360, 2605−2615.

Denny, J.C., Collins, F.S., 2021. Precision medicine in 2030-seven ways to transform healthcare. Cell 184, 1415−1419.

DHHS, DOE, 1990. Understanding Our Genetic Inheritance, the U.S. Human Genome Project: The First Five Years: Fiscal Years 1991−1995. http://www.ornl.gov/sci/techresources/Human_Genome/project/5yrplan/summary.shtml.

Dulbecco, R., 1986. A turning point in cancer research: sequencing the human genome. Science 231, 1055−1056.

Enayati, A., Hemingway, J., 2010. Malaria management: past, present, and future. Annu. Rev. Entomol. 55, 569−591.

Feng, H., Shuda, M., Chang, Y., Moore, P.S., 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 319, 1096−1100.

Fiers, W., Contreras, R., Haegemann, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G., Ysebaert, M., 1978. Complete nucleotide sequence of SV40 DNA. Nature 273, 113−120.

Finne, J., Bitter-Suermann, D., Goridis, C., Finne, U., 1987. An IgG monoclonal antibody to group B meningococci cross-reacts with developmentally regulated polysialic acid units of glycoproteins in neural and extraneural tissues. J. Immunol. 138, 4402−4407.

Fischer, W., Giorgi, E.E., Chakraborty, S., Nguyen, K., Bhattacharya, T., Theiler, J., Goloboff, P.A., Yoon, H., Abfalterer, W., Foley, B.T., Tegally, H., San, J.E., de Oliveira, T., , Network for Genomic Surveillance in South, A., Gnanakaran, S., Korber, B., 2021. HIV-1 and SARS-CoV-2: patterns in the evolution of two pandemic pathogens. Cell Host Microbe 29, 1093−1110.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al., 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269, 496−512.

Giuliani, M.M., Adu-Bobie, J., Comanducci, M., Arico, B., Savino, S., Santini, L., Brunelli, B., Bambini, S., Biolchi, A., Capecchi, B., Cartocci, E., Ciucchi, L., Di Marcello, F., Ferlicca, F., Galli, B., Luzzi, E., Masignani, V., Serruto, D., Veggi, D., Contorni, M., Morandi, M., Bartalesi, A., Cinotti, V., Mannucci, D., Titta, F., Ovidi, E., Welsch, J.A., Granoff, D., Rappuoli, R., Pizza, M., 2006. A universal vaccine for serogroup B meningococcus. Proc. Natl. Acad. Sci. U. S. A. 103, 10834−10839.

Gorbalenya, A.E., Krupovic, M., Mushegian, A., Kropinski, A.M., Siddell, S.G., Varsani, A., Adams, M.J., Davison, A.J., Dutilh, B.E., Harrach, B., Harrison, R.L., Junglen, S., King, A.M.Q., Knowles, N.J., Lefkowitz, E.J., Nibert, M.L., Rubino, L., Sabanadzovic, S., Sanfaçon, H., Simmonds, P., Walker, P.J., Zerbini, F.M., Kuhn, J.H., International Committee on Taxonomy of Viruses Executive, C, 2020. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. Nat. Microbiol. 5, 668−674.

Hafkin, B., Kaplan, N., Murphy, B., 2015. Efficacy and safety of AFN-1252, the first staphylococcus-specific antibacterial agent, in the treatment of acute bacterial skin and skin structure infections, including those in patients with significant comorbidities. Antimicrob. Agents Chemother. 60, 1695−1701.

Hay, S.I., Guerra, C.A., Tatem, A.J., Noor, A.M., Snow, R.W., 2004. The global distribution and population at risk of malaria: past, present, and future. Lancet Infect. Dis. 4, 327−336.

Hayden, E.C., 2014. The $1,000 genome. Nature 507, 294−295.

Heath, R.J., Rock, C.O., 2000. A triclosan-resistant bacterial enzyme. Nature 406, 145−146.

Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., Smith, H.O., Venter, J.C., 1999. Global transposon mutagenesis and a minimal Mycoplasma genome. Science 286, 2165−2169.

Ito, J., Ghosh, A., Moreira, L.A., Wimmer, E.A., Jacobs-Lorena, M., 2002. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. Nature 417, 452−455.

Jackson, C., Lennon, D.R., Sotutu, V.T., Yan, J., Stewart, J.M., Reid, S., Crengle, S., Oster, P., Ypma, E., Aaberge, I., Mulholland, K., Martin, D.R., 2009. Phase II meningococcal B vesicle vaccine trial in New Zealand infants. Arch. Dis. Child. 94, 745−751.

Ji, Y., Zhang, B., Van, S.F., Horn, Warren, P., Woodnutt, G., Burnham, M.K., Rosenberg, M., 2001. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. Science 293, 2266−2269.

Jonassen, C.M., Jonassen, T.O., Grinde, B., 1998. A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. J. Gen. Virol. 79 (Pt 4), 715−718.

Ju, K.S., Gao, J., Doroghazi, J.R., Wang, K.K., Thibodeaux, C.J., Li, S., Metzger, E., Fudala, J., Su, J., Zhang, J.K., Lee, J., Cioni, J.P., Evans, B.S., Hirota, R., Labeda, D.P., van der Donk, W.A., Metcalf, W.W., 2015. Discovery

of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. Proc. Natl. Acad. Sci. U. S. A. 112, 12175−12180.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Sheffield, C.-G.G., McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182, 812−827.

Krause, P.R., Fleming, T.R., Longini, I.M., Peto, R., Briand, S., Heymann, D.L., Beral, V., Snape, M.D., Rees, H., Ropero, A.M., Balicer, R.D., Cramer, J.P., Munoz-Fontela, C., Gruber, M., Gaspar, R., Singh, J.A., Subbarao, K., Van Kerkhove, M.D., Swaminathan, S., Ryan, M.J., Henao-Restrepo, A.M., 2021. SARS-CoV-2 variants and vaccines. N. Engl. J. Med. 385, 179−186.

Kyrou, K., Hammond, A.M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A.K., Nolan, T., Crisanti, A., 2018. A CRISPR-Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes. Nat. Biotechnol. 36, 1062−1066.

Lefterova, M.I., Suarez, C.J., Banaei, N., Pinsky, B.A., 2015. Next-generation sequencing for infectious disease diagnosis and management: a report of the association for molecular pathology. J. Mol. Diagn. 17, 623−634.

Levy, C.W., Roujeinikova, A., Sedelnikova, S., Baker, P.J., Stuitje, A.R., Slabas, A.R., Rice, D.W., Rafferty, J.B., 1999. Molecular basis of triclosan activity. Nature 398, 383−384.

Liu, Y.T., 2008. A technological update of molecular diagnostics for infectious diseases. Infect. Disord. Drug Targets 8, 183−188.

Liu, Y.T., 2017. 10 - infectious disease genomics. In: Tibayrenc, M. (Ed.), Genetics and Evolution of Infectious Diseases, second ed. Elsevier, London, pp. 211−225.

Lo, H., Tang, C.M., Exley, R.M., 2009. Mechanisms of avoidance of host immunity by Neisseria meningitidis and its effect on vaccine development. Lancet Infect. Dis. 9, 418−427.

Lu, Q., Zhang, X.Q., Pond, S.L., Reed, S., Schooley, R.T., Liu, Y.T., 2009. Detection in 2009 of the swine origin influenza A (H1N1) virus by a subtyping microarray. J. Clin. Microbiol. 47, 3060−3061.

Lurie, N., Saville, M., Hatchett, R., Halton, J., 2020. Developing Covid-19 vaccines at pandemic speed. N. Engl. J. Med. 382, 1969−1973.

Macdonald, G., 1957. The Epidemiology and Control of Malaria. Oxford University Press, Oxford.

Marrakchi, H., Choi, K.H., Rock, C.O., 2002. A new mechanism for anaerobic unsaturated fatty acid formation in Streptococcus pneumoniae. J. Biol. Chem. 277, 44809−44816.

Marshall, J.M., Taylor, C.E., 2009. Malaria control with transgenic mosquitoes. PLoS Med. 6, e20.

Martin, S.L., Borrow, R., van der Ley, P., Dawson, M., Fox, A.J., Cartwright, K.A., 2000. Effect of sequence variation in meningococcal PorA outer membrane protein on the effectiveness of a hexavalent PorA outer membrane vesicle vaccine. Vaccine 18, 2476−2481.

Metcalf, W.W., van der Donk, W.A., 2009. Biosynthesis of phosphonic and phosphinic acid natural products. Annu. Rev. Biochem. 78, 65−94.

Miotto, O., Amato, R., Ashley, E.A., MacInnis, B., Almagro-Garcia, J., Amaratunga, C., Lim, P., Mead, D., Oyola, S.O., Dhorda, M., Imwong, M., Woodrow, C., Manske, M., Stalker, J., Drury, E., Campino, S., Amenga-Etego, L., Thanh, T.N., Tran, H.T., Ringwald, P., Bethell, D., Nosten, F., Phyo, A.P., Pukrittayakamee, S., Chotivanich, K., Chuor, C.M., Nguon, C., Suon, S., Sreng, S., Newton, P.N., Mayxay, M., Khanthavong, M., Hongvanthong, B., Htut, Y., Han, K.T., Kyaw, M.P., Faiz, M.A., Fanello, C.I., Onyamboko, M., Mokuolu, O.A., Jacob, C.G., Takala-Harrison, S., Plowe, C.V., Day, N.P., Dondorp, A.M., Spencer, C.C., McVean, G., Fairhurst, R.M., White, N.J., Kwiatkowski, D.P., 2015. Genetic architecture of artemisinin-resistant Plasmodium falciparum. Nat. Genet. 47, 226−234.

Morel, C.M., Toure, Y.T., Dobrokhotov, B., Oduola, A.M., 2002. The mosquito genome—a breakthrough for public health. Science 298, 79.

Muehlenbachs, A., Bhatnagar, J., Agudelo, C.A., Hidron, A., Eberhard, M.L., Mathison, B.A., Frace, M.A., Ito, A., Metcalfe, M.G., Rollin, D.C., Visvesvara, G.S., Pham, C.D., Jones, T.L., Greer, P.W., Velez Hoyos, A., Olson, P.D., Diazgranados, L.R., Zaki, S.R., 2015. Malignant transformation of Hymenolepis nana in a human host. N. Engl. J. Med. 373, 1845−1852.

Munroe, D.J., Harris, T.J., 2010. Third-generation sequencing fireworks at Marco Island. Nat. Biotechnol. 28, 426−428.

NRC, 1988. Mapping and Sequencing the Human Genome. http://www.nap.edu/catalog.php?record_id=1097.

OTA, 1988. Mapping Our Genes—Genome Projects: How Big? How Fast? http://www.ornl.gov/sci/techresources/Human_Genome/publicat/OTAreport.pdf.

Payne, D.J., Gwynn, M.N., Holmes, D.J., Pompliano, D.L., 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat. Rev. Drug Discov. 6, 29−40.

Pidot, S., Ishida, K., Cyrulies, M., Hertweck, C., 2014. Discovery of clostrubin, an exceptional polyphenolic polyketide antibiotic from a strictly anaerobic bacterium. Angew Chem. Int. Ed. Engl. 53, 7856−7859.

Pinzone, C.A., Dyer, K.A., 2013. Association of polyandry and sex-ratio drive prevalence in natural populations of Drosophila neotestacea. Proc. Biol. Sci. 280, 20131397.

Pizza, M., Scarlato, V., Masignani, V., Giuliani, M.M., Arico, B., Comanducci, M., Jennings, G.T., Baldi, L., Bartolini, E., Capecchi, B., Galeotti, C.L., Luzzi, E., Manetti, R., Marchetti, E., Mora, M., Nuti, S., Ratti, G., Santini, L., Savino, S., Scarselli, M., Storni, E., Zuo, P., Broeker, M., Hundt, E., Knapp, B., Blair, E., Mason, T., Tettelin, H., Hood, D.W., Jeffries, A.C., Saunders, N.J., Granoff, D.M., Venter, J.C., Moxon, E.R., Grandi, G., Rappuoli, R., 2000. Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. Science 287, 1816−1820.

Povelones, M., Waterhouse, R.M., Kafatos, F.C., Christophides, G.K., 2009. Leucine-rich repeat protein complex activates mosquito complement in defense against Plasmodium parasites. Science 324, 258−261.

Reddy, V.B., Thimmappaya, B., Dhar, R., Subramanian, K.N., Zain, B.S., Pan, J., Ghosh, P.K., Celma, M.L., Weissman, S.M., 1978. The genome of simian virus 40. Science 200, 494−502.

Riehle, M.M., Markianos, K., Niare, O., Xu, J., Li, J., Toure, A.M., Podiougou, B., Oduol, F., Diawara, S., Diallo, M., Coulibaly, B., Ouatara, A., Kruglyak, L., Traore, S.F., Vernick, K.D., 2006. Natural malaria infection in *Anopheles gambiae* is regulated by a single genomic control region. Science 312, 577−579.

Rigden, D.J., Fernández, X.M., 2020. The 2021 Nucleic Acids Research database issue and the online molecular biology database collection. Nucleic Acids Res. 49, D1−D9.

Rinaudo, C.D., Telford, J.L., Rappuoli, R., Seib, K.L., 2009. Vaccinology in the genome era. J. Clin. Invest. 119, 2515−2525.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, C.A., Hutchison, C.A., Slocombe, P.M., Smith, M., 1977. Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265, 687−695.

Sierra, G.V., Campa, H.C., Varcacel, N.M., Garcia, I.L., Izquierdo, P.L., Sotolongo, P.F., Casanueva, G.V., Rico, C.O., Rodriguez, C.R., Terry, M.H., 1991. Vaccine against group B Neisseria meningitidis: protection trial and mass vaccination results in Cuba. NIPH (Natl. Inst. Public Health) Ann. (Oslo) 14, 195−207 discussion 208−110.

Slaoui, M., Hepburn, M., 2020. Developing safe and effective Covid vaccines—Operation Warp Speed's strategy and approach. N. Engl. J. Med. 383, 1701−1703.

Smith, H.O., 2004. History of microbial genomics. In: Fraser, C.M., Read, T.D., Nelson, K.E. (Eds.), Microbial Genomes. Humana, Totowa, N.J., pp. 3−16

Stone, R., 1994. DOE to sequence microbe genomes. Science 263, 311−311.

Tao, K., Tzou, P.L., Nouhin, J., Gupta, R.K., de Oliveira, T., Kosakovsky Pond, S.L., Fera, D., Shafer, R.W., 2021. The biological and clinical significance of emerging SARS-CoV-2 variants. Nat. Rev. Genet. 1−17.

Tettelin, H., Saunders, N.J., Heidelberg, J., Jeffries, A.C., Nelson, K.E., Eisen, J.A., Ketchum, K.A., Hood, D.W., Peden, J.F., Dodson, R.J., Nelson, W.C., Gwinn, M.L., DeBoy, R., Peterson, J.D., Hickey, E.K., Haft, D.H.,

Salzberg, S.L., White, O., Fleischmann, R.D., Dougherty, B.A., Mason, T., Ciecko, A., Parksey, D.S., Blair, E., Cittone, H., Clark, E.B., Cotton, M.D., Utterback, T.R., Khouri, H., Qin, H., Vamathevan, J., Gill, J., Scarlato, V., Masignani, V., Pizza, M., Grandi, G., Sun, L., Smith, H.O., Fraser, C.M., Moxon, E.R., Rappuoli, R., Venter, J.C., 2000. Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. Science 287, 1809−1815.

Tun, K.M., Imwong, M., Lwin, K.M., Win, A.A., Hlaing, T.M., Hlaing, T., Lin, K., Kyaw, M.P., Plewes, K., Faiz, M.A., Dhorda, M., Cheah, P.Y., Pukrittayakamee, S., Ashley, E.A., Anderson, T.J., Nair, S., McDew-White, M., Flegg, J.A., Grist, E.P., Guerin, P., Maude, R.J., Smithuis, F., Dondorp, A.M., Day, N.P., Nosten, F., White, N.J., Woodrow, C.J., 2015. Spread of artemisinin-resistant Plasmodium falciparum in Myanmar: a cross-sectional survey of the K13 molecular marker. Lancet Infect. Dis. 15, 415−421.

Wang, D., Urisman, A., Liu, Y.T., Springer, M., Ksiazek, T.G., Erdman, D.D., Mardis, E.R., Hickenbotham, M., Magrini, V., Eldred, J., Latreille, J.P., Wilson, R.K., Ganem, D., DeRisi, J.L., 2003. Viral discovery and sequence recovery using DNA microarrays. PLoS Biol. 1, E2.

Wang, J., Soisson, S.M., Young, K., Shoop, W., Kodali, S., Galgoci, A., Painter, R., Parthasarathy, G., Tang, Y.S., Cummings, R., Ha, S., Dorso, K., Motyl, M., Jayasuriya, H., Ondeyka, J., Herath, K., Zhang, C., Hernandez, L., Allocco, J., Basilio, A., Tormo, J.R., Genilloud, O., Vicente, F., Pelaez, F., Colwell, L., Lee, S.H., Michael, B., Felcetto, T., Gill, C., Silver, L.L., Hermes, J.D., Bartizal, K., Barrett, J., Schmatz, D., Becker, J.W., Cully, D., Singh, S.B., 2006. Platensimycin is a selective FabF inhibitor with potent antibiotic properties. Nature 441, 358−361.

Watson, J.D., 1990. The human genome project: past, present, and future. Science 248, 44−49.

Wicht, K.J., Mok, S., Fidock, D.A., 2020. Molecular mechanisms of drug resistance in Plasmodium falciparum malaria. Annu. Rev. Microbiol. 74, 431−454.

Witkowski, B., Khim, N., Chim, P., Kim, S., Ke, S., Kloeung, N., Chy, S., Duong, S., Leang, R., Ringwald, P., Dondorp, A.M., Tripura, R., Benoit-Vical, F., Berry, A., Gorgette, O., Ariey, F., Barale, J.C., Mercereau-Puijalon, O., Menard, D., 2013. Reduced artemisinin susceptibility of Plasmodium falciparum ring stages in western Cambodia. Antimicrob. Agents Chemother. 57, 914−923.

Wittke, F., Vincent, C., Chen, J., Heller, B., Kabler, H., Overcash, J.S., Leylavergne, F., Dieppois, G., 2020. Afabicin, a first-in-class antistaphylococcal antibiotic, in the treatment of acute bacterial skin and skin structure infections: clinical noninferiority to vancomycin/linezolid. Antimicrob. Agents Chemother. 64.

World Health Organization, 2021. Guidance Framework for Testing Genetically Modified Mosquitoes.

Wright, H.T., Reynolds, K.A., 2007. Antibacterial targets in fatty acid biosynthesis. Curr. Opin. Microbiol. 10, 447−453.

Xia, J., Guo, Z., Yang, Z., Han, H., Wang, S., Xu, H., Yang, X., Yang, F., Wu, Q., Xie, W., Zhou, X., Dermauw, W., Turlings, T.C.J., Zhang, Y., 2021. Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. Cell 184, 1693−1705.

Zhang, Y.M., Marrakchi, H., White, S.W., Rock, C.O., 2003. The application of computational methods to explore the diversity and structure of bacterial fatty acid synthase. J. Lipid Res. 44, 1−10.

Zhang, Y.M., White, S.W., Rock, C.O., 2006. Inhibiting bacterial fatty acid synthesis. J. Biol. Chem. 281, 17541−17544.

Zhou, B., Thao, T.T.N., Hoffmann, D., Taddeo, A., Ebert, N., Labroussaa, F., Pohlmann, A., King, J., Steiner, S., Kelly, J.N., Portmann, J., Halwe, N.J., Ulrich, L., Trueb, B.S., Fan, X., Hoffmann, B., Wang, L., Thomann, L., Lin, X., Stalder, H., Pozzi, B., de Brot, S., Jiang, N., Cui, D., Hossain, J., Wilson, M.M., Keller, M.W., Stark, T.J., Barnes, J.R., Dijkman, R., Jores, J., Benarafa, C., Wentworth, D.E., Thiel, V., Beer, M., 2021. SARS-CoV-2 spike D614G change enhances replication and transmission. Nature 592, 122−127.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F., Shi, Z.L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270−273.

CHAPTER

Proteomics and Host–Pathogen Interactions: Still a Bright Future?

11

David G. Biron[1,a], Dorothée Missé[2,a], Dobrin Nedelkov[3,a] and Philippe Holzmuller[4,a]

[1]*Laboratoire Microorganismes: Génome et Environnement, UMR 6023 CNRS/UCA, Campus Universitaire des Cézeaux, Aubière, France;* [2]*MIVEGEC (Maladies Infectieuses et Vecteurs: Écologie, Génétique, Évolution et Contrôle), UMR 5290 CNRS/IRD/UM, Montpellier, France;* [3]*Isoformix Inc., Phoenix, AZ, United States;* [4]*ASTRE (Animal, santé, territoires, risques et écosystèmes), UMR 117 CIRAD/INRAE/UM, TA A-117/E, Campus International de Baillarguet, Montpellier, France*

1. Introduction

Living organisms are constantly exposed to pathogens. In any environment, a molecular war begins when a host encounters a pathogen. In many host–pathogen associations, the molecular war is in progress a long time ago, and is even more complex for vector-borne pathogens. Nevertheless, a disease as an outcome of a pathogen attack remains an exception rather than a rule. Most host and vector species have acquired strategies by selective pressure to mislead the pathogen and to win fight during their crosstalk (i.e., molecular dialog), but that does not mean their life history traits are not affected. Actually, many pathogen species have acquired strategies by selective pressure to bypass the host defenses to win the molecular war and to ensure the completion of their life cycle, with collateral damage of very variable intensity. Pathogens remain a significant threat to any host species, and critical is the ability to rapidly detect, respond to, treat and contain the pathogen transmission. Since many centuries, some scientific fields (i.e., agroecology, evolutionary ecology, evolutionary medicine, biochemistry, microbiology, medicine, veterinary medicine, immunology, and molecular biology) have surveyed host–pathogen interactions to improve our understanding of infectious diseases and to prevent pathogen transmission in host populations.

Throughout human history, infectious diseases have severely and continue to affect most societies worldwide, and are a leading cause of premature death in the world. For instance, in Europe, one of the most dramatic disease events was the great plague pandemic of the mid-14th century (Achtman et al., 2004; Watts, 1997), before the recent COVID-19 pandemic reminded us of this cruel reality (Di Fazio et al., 2022; Alshubaith et al., 2022). Actually, pathogenic diseases result from an intimate relationship between a host and a pathogen, potentially a vector, which involves molecular "crosstalk." Clearly, elucidation of this complex molecular dialog between host, vector, and pathogen is still highly desirable in order to improve our understanding of pathogen virulence, to characterize pathogen-

[a]All authors contributed equally.

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00026-4

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

specific host biomarkers and to define novel diagnostic/prognostic, therapeutic and vaccine targets. Proteomics applications to decipher host–pathogen interactions are in their teenage in spite of important technological and scientific advances since the postgenomic era, and should led to new insights on host and vector specificity or plasticity, and on the evolution of pathogen virulence. In this chapter, we present the interest of proteomics to survey host–vector–pathogen interactions, a synthetic review of previous proteomics studies, the pitfalls of the current approach in surveys, new conceptual approaches to decipher host–vector–pathogen interactions (e.g., proteogenomics or metaproteomics), a new avenue to decipher the crosstalk diversity involved in trophic interactions in an habitat (i.e., the population proteomics) and 5 year view for future prospects on proteomics and host–pathogen interactions.

2. Interest of Proteomics to Study Host–(Vector)–Pathogen Interactions

Since the start of the genomic era in the early 1990s, many microbiologists and molecular biologists are confident that complete sequencing of the genome of the partners in host–pathogen associations for pathogens with simple life cycle (i.e., one host) and in host–vector–pathogen associations for pathogens with complex life cycle (i.e., at least two hosts) will enable to total understanding of the molecular mechanisms involved in most of the infectious diseases and will contribute to find new drugs for treating or new vaccines for preventing them (Hochstrasser, 1998; Degrave et al., 2001). Unfortunately, insufficient progress has been achieved in the control of such diseases as malaria, tuberculosis, or AIDS, despite decades of intensive genomic projects on host–pathogen interactions, vaccines, and chemotherapeutics. Pathogens continue to be a major cause of morbidity and mortality in humans and domestic livestock, more especially in developing countries (Ouma et al., 2001; Ryan, 2001; Guzman and Kouri, 2002; Gelfand and Callahan, 2003; WHO, 2016).

Until now, biologists have focused their studies on DNA analyses based on the central dogma of molecular biology—that is to say, the general pathway for the expression of genetic information stored in DNA. Although the basic blueprint of life is encoded in DNA, the execution of the genetic plan is carried out by the activities of proteins. The fabric of biological diversity is therefore protein-based and natural selection acts at the protein level (Karr, 2008). At the end of the 20th century, it had become clear that knowing genome sequences, while technically mandatory, was not in itself enough to fully understand complex biological events like the immune response of a host to a pathogen infection or the molecular strategies used by pathogens to thwart the host defenses during their interaction (Barrett et al., 2000; Ashton et al., 2001; Fell, 2001; Fields, 2001; Schmid-Hempel, 2008).

The evolution of any given species has tremendously increased complexity at the level of pre- (gene splicing, mRNA editing) and posttranslational (phosphorylation, glycosylation, acetylation, etc. …) gene–protein interaction. The genomics era has revealed that: (a) DNA sequences may be "fundamental," but can provide little information on the dynamic processes within and between host and a pathogen during their physical and molecular interaction (Barrett et al., 2000; Ashton et al., 2001); (b) the correlation between the expressed "transcriptome" (i.e., total mRNA transcription pattern) and the levels of translated proteins is poor (Anderson and Seilhaver, 1997; Gygi et al., 1999; Maniatis and Tasic, 2002); and (c) a single gene can produce different protein products (Fell, 2001; Fields, 2001; Maniatis and Tasic, 2002). Additionally, in terms of interaction, a given pathogen induce

a specific molecular signature by infected hosts these same hosts may also express a common response to different infections (Holzmuller et al., 2016). Moreover, the structure, function, abundance, and even the number of proteins in an organism cannot yet be predicted from the DNA sequence alone (Anderson and Anderson, 1996; Gygi et al., 1999; Barrett et al., 2000). Also, posttranslational modifications such as phosphorylation and glycosylation are often extremely important for the dynamic and function of many proteins, although most of these modifications cannot yet be predicted form genomic or mRNA sequences (Gygi et al., 1999). And to refine the whole process, epigenetic systems control and modify gene expression. Almost all the elements of epigenetic control systems are proteins (Anderson and Anderson, 1996). The cells of an organism are reactive systems in which information flows not only from genes to proteins but in the reverse direction as well (Hochstrasser, 1998). The proteome is the genome operating system by which the cells of an organism react to environmental signals (Anderson and Anderson, 1996). It comprises an afferent arm, the cytosensorium (i.e., many cellular proteins are sensors, receptors, and information transfer units from environmental signals) and an efferent arm, the cytoeffectorium (i.e., in cells, reaction of the genome via regulation of either individual proteins or a group of proteins in response to environmental changes). Thus, the biological phenotype of an organism is not directly related to its genotype (i.e., DNA sequences).

Proteomics is the study of the proteome (i.e., the genome operating system), which in a broad sense means all the proteins produced by a cell or tissue, at a given moment, in a given environment. Proteomics will contribute to bridge the gap between our understanding of genome sequence and cellular/organism behavior. Proteomics offers an excellent way to study the reaction of the host and pathogen proteomes (i.e., infectome) during their complex biochemical crosstalk (Biron et al., 2005a,b). Using the first-generation proteomics approach, two-dimensional electrophoresis (2-DE) and mass spectrometry (MS), posttranslational modifications of host and pathogen proteins (such as phosphorylation, glycosylation, acetylation, and methylation) in reaction to their interaction can be detected. Such modifications are vital for the correct activity of numerous proteins and are being increasingly recognized as a major mechanism in cellular regulation. Although 2-DE offers a high-quality approach for the study of host and/or pathogen proteomes, during the postgenomic era several proteomics approaches (e.g., bottom up; top down) and quantitative proteomics (e.g., subproteome enrichment) strategies have been developed that complement classical 2-DE (see Fig. 11.1) (Gygi et al., 1999; Fung et al., 2001; Lopez and Pluksal, 2003; Wu et al., 2003; Bischoff and Luider, 2004; Chetouhi et al., 2015). Table 11.1 presents a comparison of the most popular proteomics tools.

3. Retrospective of Pioneering Proteomics Studies

The host–pathogen crosstalks reflect the balance of host defenses and pathogen virulence mechanisms. Postgenomic technology promises to revolutionize many fields in biology by providing enormous amounts of genetic data from model and nonmodel organisms. Proteomics is a case point and promises to bridge the gap between our understanding of genome sequences and cellular behavior involved in host–pathogen interactions. Proteomics offers the possibility to characterize host–pathogen interactions from a global proteomic view. To date, most proteomics surveys on host–

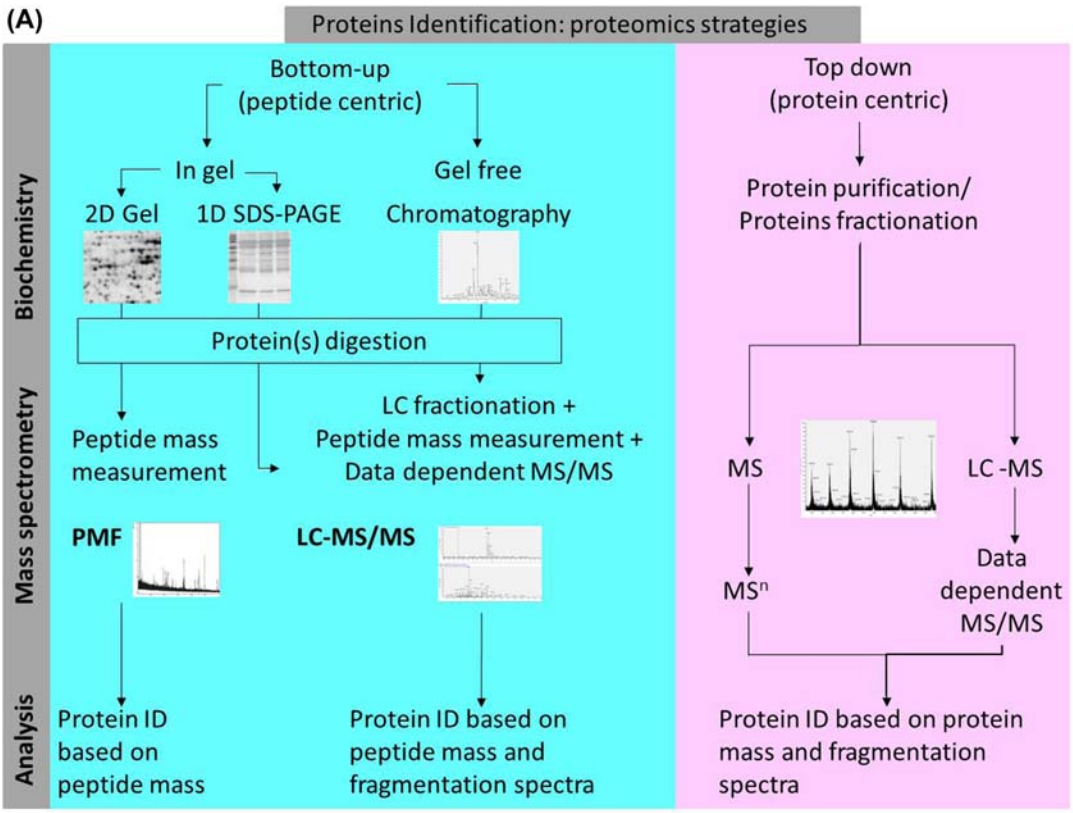


FIGURE 11.1

Deciphering of host parasite crosstalk with proteomics. *Bottom-up* and *Top-down* approaches (A); and quantitative proteomics strategies (B).

pathogen interactions have focused on cataloging protein content of pathogens and identifying virulence-associated proteins or proteomic alterations in host response to a pathogen. Also, many parasitologists, microbiologists, virologists, and molecular biologists have used proteomics to find pathogen-specific host biomarkers for rapid pathogen detection and characterization of host–pathogen crosstalks during the infection process. In this section, a synthetic retrospective of previous proteomics studies on host–pathogen interactions of these surveys is presented.

3.1 Deciphering of the Molecular Strategies Involved in Pathogen Immune Evasion

To elude the vigilance of the immune system of a host, particularly mammals, a causative microorganism must actually act as a double agent. Indeed, the broad immunity has a natural or innate and

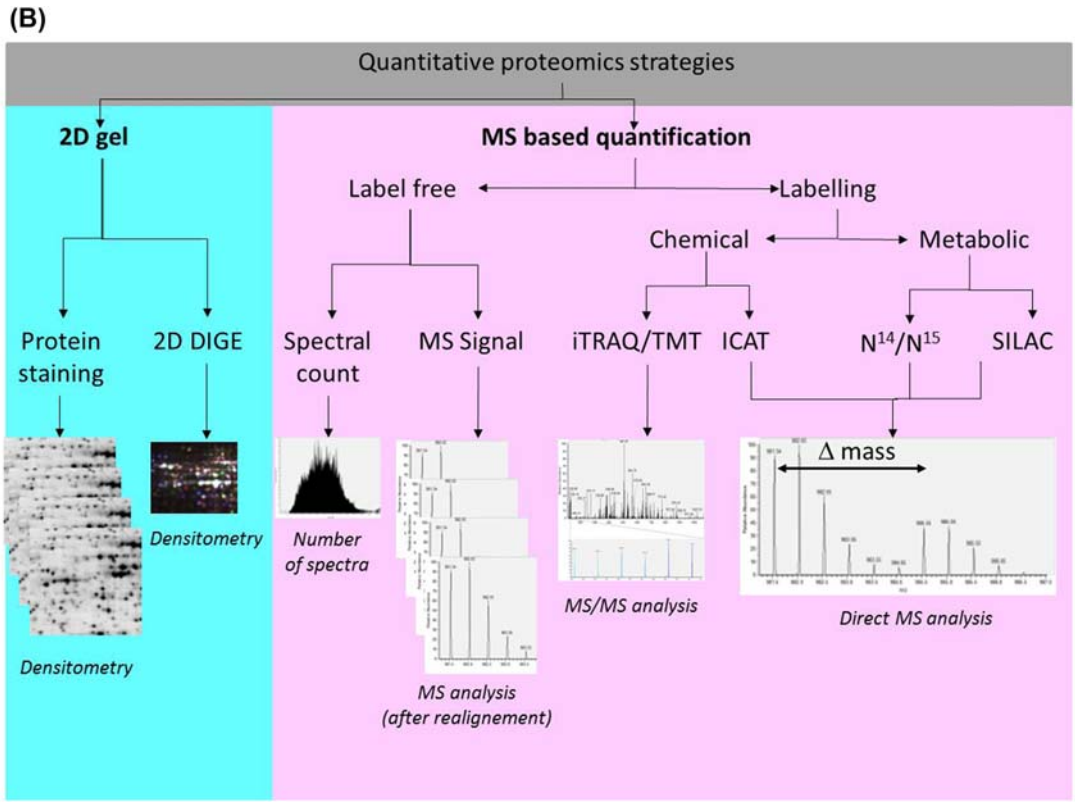


FIGURE 11.1 Cont'd

adaptive component. Innate immunity constitutes the first antimicrobial defense and rapidly induces soluble mediators such as complement, inflammatory cytokines and chemokines together with effector cells such as macrophages and natural killers, in order to control or delay the spreading of the infectious agent. Then a specific response of adaptive immunity will take place to eliminate pathogens that would have survived innate immune response (Roitt and Delves, 2001). These immune selective pressures have conducted pathogens to develop mechanisms to modulate and alter host responses or to evade phagocytosis. As a result of these host−pathogen interactions, protein expression profiles of the host immune system (susceptibility/tolerance factors) and of the pathogen (virulence/pathogenicity factors) mutually modified (Zhang et al., 2005; Coiras et al., 2008; Holzmuller et al., 2008; Marcelino et al., 2021).

Depending on the pathogen type (virus, bacteria, fungi, unicellular or multicellular parasites), strategies of interactions will be different and the subversion of the host immune responses will exhibit specificities at the protein level (for reviews see Walduck et al., 2004; Biron et al., 2005a; Viswanathan

Table 11.1 A Comparison of Proteomics Tools

	Assumptions		
	Mean	StdDev	
Hedge Price	100		
Hedge Volume	200	10	<— Hedge Volume is assumed to be the mean project output
Interest Rate	5%		
Basis Differential	-1.00		
Hub Price	100.00	10	
Node Price	99.00	10	
Correlation (Hub Price/Node Price)	80%		

Month	Output (MWh)	Hub Price ($/MWh)	Node Price ($/MWh)	Realized Revenues	Fixed Obligation	Floating Obligation	Beg Tr Acct Bal	Addition	Subtraction	Interest	End Tr Acct Bal	Revenues after TA Adjustment	Revenues w/o Tr Acct
											$0.00		
1	208.28	105.19	91.22	$19,000.17	$20,000.00	$21,037.89	$0.00	$2,037.72	$0.00	$0.00	$2,037.72	$20,000.00	$17,962.28
2	201.07	112.32	112.40	$22,600.51	$20,000.00	$22,463.90	$2,037.72	$0.00	$136.61	$8.49	$1,909.60	$20,000.00	$20,136.61
3	196.82	84.89	91.47	$18,002.15	$20,000.00	$16,978.73	$1,909.60	$0.00	$1,023.41	$7.96	$894.14	$20,000.00	$21,023.41
4	192.74	91.59	85.22	$16,424.55	$20,000.00	$18,318.16	$894.14	$1,893.62	$0.00	$3.73	$2,791.48	$20,000.00	$18,106.38
5	202.96	99.47	96.11	$19,506.74	$20,000.00	$19,893.55	$2,791.48	$386.81	$0.00	$11.63	$3,189.93	$20,000.00	$19,613.19
6	194.94	87.78	90.24	$17,591.75	$20,000.00	$17,555.15	$3,189.93	$0.00	$36.61	$13.29	$3,166.61	$20,000.00	$20,036.61
7	203.54	112.78	106.37	$21,650.01	$20,000.00	$22,555.51	$3,166.61	$905.50	$0.00	$13.19	$4,085.30	$20,000.00	$19,094.50
8	191.39	71.12	86.33	$16,522.78	$20,000.00	$14,223.40	$4,085.30	$0.00	$2,299.37	$17.02	$1,802.95	$20,000.00	$22,299.37
9	197.07	96.32	102.34	$20,168.35	$20,000.00	$19,263.13	$1,802.95	$0.00	$905.22	$7.51	$905.24	$20,000.00	$20,905.22
10	205.82	112.02	108.06	$22,240.90	$20,000.00	$22,403.38	$905.24	$162.48	$0.00	$3.77	$1,071.49	$20,000.00	$19,837.52
11	200.35	102.57	90.96	$18,223.78	$20,000.00	$20,513.21	$1,071.49	$2,289.43	$0.00	$4.46	$3,365.39	$20,000.00	$17,710.57
12	197.42	105.74	109.14	$21,547.11	$20,000.00	$21,148.90	$3,365.39	$0.00	$398.21	$14.02	$2,981.20	$20,000.00	$20,398.21
13	197.50	109.99	107.39	$21,209.61	$20,000.00	$21,997.23	$2,981.20	$787.61	$0.00	$12.42	$3,781.24	$20,000.00	$19,212.39
14	204.92	85.03	86.09	$17,642.01	$20,000.00	$17,006.63	$3,781.24	$0.00	$635.38	$15.76	$3,161.61	$20,000.00	$20,635.38
15	208.75	122.00	115.64	$24,140.32	$20,000.00	$24,400.77	$3,161.61	$260.45	$0.00	$13.17	$3,435.24	$20,000.00	$19,739.55
16	192.99	92.62	96.03	$18,532.00	$20,000.00	$18,524.60	$3,435.24	$0.00	$7.40	$14.31	$3,442.15	$20,000.00	$20,007.40
17	191.20	83.80	90.48	$17,299.51	$20,000.00	$16,760.72	$3,442.15	$0.00	$538.79	$14.34	$2,917.71	$20,000.00	$20,538.79
18	199.83	79.53	84.20	$16,825.76	$20,000.00	$15,906.05	$2,917.71	$0.00	$919.71	$12.16	$2,010.16	$20,000.00	$20,919.71
19	195.15	90.01	89.12	$17,391.46	$20,000.00	$18,002.59	$2,010.16	$611.13	$0.00	$8.38	$2,629.66	$20,000.00	$19,388.87
20	201.75	99.21	99.85	$20,144.98	$20,000.00	$19,842.78	$2,629.66	$0.00	$302.21	$10.96	$2,338.41	$20,000.00	$20,302.21
21	198.78	95.89	85.16	$16,927.82	$20,000.00	$19,178.34	$2,338.41	$2,250.52	$0.00	$9.74	$4,598.67	$20,000.00	$17,749.48
22	203.48	108.04	107.46	$21,865.28	$20,000.00	$21,608.69	$4,598.67	$0.00	$256.59	$19.16	$4,361.24	$20,000.00	$20,256.59
23	197.64	87.16	89.11	$17,611.58	$20,000.00	$17,432.17	$4,361.24	$0.00	$179.41	$18.17	$4,200.00	$20,000.00	$20,179.41
24	206.32	118.12	109.77	$22,648.46	$20,000.00	$23,623.60	$4,200.00	$975.14	$0.00	$17.50	$5,192.63	$20,000.00	$19,024.86
Standard Deviation				**$2,318.58**								**$0.00**	**$1,124.92**

Note: The example above illutsrates how a Tracking Account mechanism reduces the volatility of the revenues.

and Früh, 2007; Biron and Loxdale, 2013). In fact, these molecular dialogs and conflicts can be seen as a chess game between the host immune cell populations and the pathogen populations, in which the pathogen plays with the whites (i.e., it starts the game). Because of differences in host–pathogen organisms' size and ratio, leading to size differences of respective proteomes, the pathogen proteome could be considered as overwhelmed by the host proteome during the interactions. But in terms of immune evasion, this is not limiting because the immune system works on a qualitative basis, which constitutes a second advantage for the pathogen that can induce large-scale damages with low amounts of molecules. By contrast, this represents one major limitation to characterize host–pathogen interactions, and also a challenging perspective for proteomics technology. This is why, retrospectively, proteomics studies were mainly conducted to evidence pathogenic virulence and pathogenicity factors (Ouellette et al., 2003; Texier et al., 2005; Sims and Hyde, 2006; van Hellemond et al., 2007; Liu et al., 2008a; Bird and Opperman, 2009; Steuart, 2010; Weiss et al., 2009; Jagusztyn-Krynicka et al., 2009; Premsler et al., 2009; Bhavsar et al., 2010; Holzmuller et al., 2010).

Independently of the proteomics workflow used for analysis, immune evasion could be illustrated by at least three strategies that are commonly widespread among pathogens: (i) immune evasion based on antigenic variation, (ii) inhibition of adaptive immunity activation systems, and (iii) host mimicry. In African trypanosomes, the antigenic variation of the variant surface glycoprotrein (VSG) constituting the surface coat of the parasite is well described (Morrison et al., 2009). But as in proteomics study, the parasite population, which has switched the VSG, is so poorly represented it goes undetected, and therefore always keeps one step ahead of host immune responses. Also, in trypanosomatids, *Leishmania* amastigotes, which establish within macrophage (a major immune effector cell), developed the ability to degradating class II major histocompatibility molecules to prevent Th1-type immunity to be induced (Antoine et al., 2004). Another protozoan parasite, *Toxoplasma gondii*, generates its parasitophorous vacuole with elements of the plasma membrane from the targeted host cells, thus using the host "self" to evade immune recognition (Plattner and Soldati-Favre, 2008). These few examples actually perfectly illustrate how difficult it is to decipher, at the protein level during interactions, the pathogen molecular components involved in immune evasion. However, a new quantitative proteomics tools, the SILAC (stable isotope labeling by amino acids in cell culture) lately allowed the detection of 148 proteins of a microsporidian parasite during the kinetics of a host infection (Panek et al., 2014). Among these proteins many are involved in parasite proliferation, and an overrepresentation of putative secreted effectors proteins was observed. Finally, this SILAC survey also suggests that this microsporidia species could use a transposable element as a lure strategy to escape the host innate immune system. Recent advances in proteomics offer challenging perspectives to decipher the molecular war in host–pathogen interactions (Chetouhi et al., 2015).

3.2 Host Proteome Responses to Parasite Infection

Whereas it seems obvious to say that when a pathogen will infect a host, this last will react by expressing molecules that can be characterized by clinical proteomics, it is surprising how few studies are devoted to this research. Yet the discovery for biomarkers signing an infected state from a healthy state is the heart of the Infectious Disease Research (te Pas and Claes, 2004; Azad et al., 2006), and expression proteomics has quickly developed to characterize the differential expression of proteins

encoded by a particular gene and their posttranslational modifications in biological fluids and tissues (Fournier et al., 2008; Hood et al., 2009; Wilm, 2009; Mehta et al., 2015). In characterizing the host proteome responses to a pathogen infection, different levels of analysis have to be considered: soluble biomarkers expressed in biological fluids (e.g., serum, saliva, urine, and cerebrospinal fluid), tissue biomarkers indicative of an organ response and cellular biomarkers indicative of a cell-type response (e.g., immune cells).

Interestingly, the majority of the proteomics studies on host response to infection were performed on viral deregulation of host cells proteome ex vivo (Zheng et al., 2008; Liu et al., 2008b; Lee et al., 2009; Sun et al., 2008; Pastorino et al., 2009; Vester et al., 2009; Antrobus et al., 2009; Zhang et al., 2009; Zhang et al., 2010a). These works allowed to characterize at the molecular level the overall modifications in protein profiles of the target cells, and were of high interest to the better understanding of the pathogen influence on its host. In bacteria, studies have evaluated the mode of action of known toxins or bacterial components on host cells (Kuhn, et al., 2006; Shui et al., 2009). Concerning parasites, ex vivo experiments on host cells–parasite interactions have highlighted molecular details of manipulation strategies suffered by target cells during Toxoplasmosis Chagas' disease or malaria (Teixeira et al., 2006; Nelson et al., 2008; Wu et al., 2009). Curiously, few works directly focused on the subversion of the immune system, mainly through monocyte/macrophage deregulation (Oura et al., 2006; Fischer et al., 2007).

As a paradox, the most striking studies on host proteome response to parasite infection were performed on arthropod (infectious diseases vectors)–parasite interactions. Probably because the parasite induced a strong phenotype modification (Biron et al., 2005c; Rachinsky et al., 2007), particularly in the case of insect behavior manipulation (Lefevre et al., 2007a,b). Although few in number, taken together these pioneering analyses of the response of the proteome of the host to a pathogen pave the way for the dynamic analysis of host–pathogen interactions. These approaches deserve to be strengthened and extended to all infectious diseases to increase and improve knowledge of the molecular dialog and conflict that govern host–pathogen interactions.

On the other hand, the clinical aspect is important in infectious diseases, a number of studies have sought to characterize more comprehensively the proteome response of the host to infection in biological fluids, with a purpose diagnosis. One interesting pioneering study was performed in rabbits and allowed to detect intra-amniotic infection by proteomic-based amniotic fluid analysis (Klein et al., 2005). For human diseases or those of livestock, the biological fluid, which should enable the detection of infection, linked to host proteome response, in the host serum. Several studies performed on this biological sample have allowed discriminating host-commensal from host–pathogen interactions in *Candida albicans* (Pitarch et al., 2009) and determining the immunome of pathogens (Sakolvaree et al., 2007; Ju et al., 2009). Moreover, in African trypanosomiasis, proteomics analysis of the serum not only was indicative of the host response to infection but also was promising for characterizing disease progression toward neurological disorder (Papadopoulos et al., 2004; Agranoff et al., 2005). This illustrates how proteomics will help in considering at different analytical levels the host proteome response to a pathogen infection, with the prospect of benefits in improving diagnostic and therapeutic.

3.3 Biomarkers Linked to Infection Process by a Pathogen Using SELDI-TOF-MS Technology

High-throughput proteomic technology offers promise for the discovery of disease biomarkers and have extended our ability to unravel proteomes. In this section, we will focus on the surface-enhanced laser desorption time of flight mass spectrometry (SELDI-TOF-MS) technology. This mass spectrometric—based method requires a minimal amount of sample for analysis and allows the rapid high throughput analysis of complex protein samples (De Bock et al., 2010). SELDI-TOF-MS differs from conventional matrix-assisted laser desorption ionization (MALDI)-TOF-MS because the target surfaces, to which the proteins and matrices are applied to, are coated with various chemically active ProteinChip surfaces (ion exchange, immobilized metal affinity capture, and reverse phase arrays). Therefore, it is possible to fractionate proteins within a mixture, or particular classes of proteins, on the array surface prior to analysis. As with MALDI, different matrices can be used to facilitate the ionization and desorption of proteins from the SELDI array surface (Merchant and Weinberger, 2000).

This technology was initially applied to the discovery of early diagnostic or prognostic biomarkers of cancer (Petricoin and Liotta, 2004; Yang et al., 2005; Xiao et al., 2005). Recently, this technology has been used to discover fluid or tissue protein biomarkers for infectious diseases such as HIV-1 (Luo et al., 2003; Sun et al., 2004; Missé et al., 2007; Luciano-Montalvo et al., 2008; Toro-Nieves et al., 2009; Wiederin et al., 2008), hepatitis B and C viruses (Poon et al., 2005; Kanmura et al., 2007; Molina et al., 2008; Fujita et al., 2008; Wu et al., 2009), severe acute respiratory syndrome (Panek et al., 2006) and BK virus (Jahnukainen et al., 2006), African Trypanosomiasis (Stiles et al., 2004; Agranoff et al., 2005), infection of Artemia by cestodes (Sánchez et al., 2009), tuberculosis (Liu et al., 2010), bacterial endocarditis (Fenollar et al., 2006), and *Helicobacter pylori* infection (Das et al., 2005).

Certain individuals are resistant to HIV-1 infection, despite repeated exposure to the virus. The analysis of resistance to HIV infection is one of a research avenues which has the hope of resulting in the development of a more effective treatment or a successful preventive vaccine against HIV infection. However, the molecular mechanism underlying resistance in repeatedly HIV-1-exposed, uninfected individuals (EU) is unclear. A complementary transcriptome and SELDI-TOF-MS analyses on peripheral blood T cells, and plasma or serum from EU, their HIV-1-infected sexual partners, and healthy controls have been performed (Missé et al., 2007). This study detected a specific biomarker associated with innate host resistance to HIV infection, as an 8.6-kDa A-SAA cleavage product.

In the same vein, understanding the virus—host interactions that lead to patients with acute Hepatitis C Virus (HCV) infection to viral clearance is a key toward the development of more effective treatment and prevention strategies. SELDI-TOF-MS technology have been used to compare, at a proteomic level, plasma samples, respectively, from donors who had resolved their HCV infection after seroconversion, from donors with chronic HCV infection and from unexposed healthy donors (Molina et al., 2008). A candidate marker of about 9.4 kDa was found to be higher in donors with HCV clearance than in donors with chronic infection. This biomarker was identified by nanoLC-Q-TOF-MS/MS as Apolipoprotein C-III and validated by Western Blot analysis. Among the most strongly

up-regulated genes in Dengue virus–infected *Ae. aegypti* salivary glands, one study identified a gene belonging to the cecropin family. The overexpression of this antimicrobial peptide was confirm using the SELDI-TOF-MS technique (Luplertlop et al., 2011).

4. Evolution of Conceptual Approaches to Decipher the Host–Pathogen Interactions for Microorganisms With Simple or Complex Life Cycle

During the postgenomics era, the "parasito-proteomics" extended to "pathogeno-proteomics" have been suggested to study of the reaction of the host and pathogen (completed by an arthropod vector for vector-borne pathogens) genomes through the expression of the corresponding proteomes during their biochemical crosstalk.

4.1 From "Parasito-Proteomics" to "Pathogeno-Proteomics": An Holistic View to Disentangle the Hosts and Pathogens Genome Responses During Their Interactions

Studies in "parasito-proteomics" are performed either by following the expression of the parasite proteome during infection by a given parasite (Langley et al., 1987; Moura and Visvesvara, 2001; Cohen et al., 2002; Boonmee et al., 2003), by the reaction of the host proteome following an invasion by a parasite species (Moskalyk et al., 1996; Cohen et al., 2002; Thiel and Bruchhaus, 2001), or by the injection of immune elicitors (Han et al., 1999; Vierstraete et al., 2004).

Some elegant pioneering studies on the proteome responses of insect hosts during their molecular crosstalk with pathogens concluded that insects could rapidly react to infection by a given pathogen (i.e., bacteria or fungi) by producing immune-induced proteins including peptide or polypeptides (Hoffman, 1995). However, a key point is to define whether the host genomic responses elicited through activation of immune constitutive proteins, induction and/or suppression of proteins during the infection by a parasite represent a nonspecific response that might be induced by any pathogen. Many "parasite-proteomics" studies dealt with a limited framework by deciphering the host proteome response during the infection process by a specific pathogen (Huang et al., 2002).

Although 2DE and 2D-DIGE offer high-quality approaches for studying the host and/or pathogen proteomes and are still reference techniques (Meleady, 2018), several high-throughput proteomic tools have been developed that complement these traditional proteomic tool (Cui et al., 2022) (see Table 11.1). In the same way, mass spectrometry analysis could take advantages of combined techniques. For example, the widely used analysis of peptides via collisionally activated dissociation (CAD) is rapid and results in reproducible predictable fragmentation behavior for a given peptide sequence, a substantial proportion of peptide product ion mass spectra do not result in successful sequence identification (Steen and Mann, 2004). This is well illustrated by Hart et al. (2009), who showed that a substantial number of proteins from trypanosome flagellum were identified in their three independent flagellar proteome investigations and also that combining electron transfer dissociation (ETD) with CAD allowed the identification of 168 proteins that were not recognized in their first analysis. This strengthens the idea of integrating both approaches and technologies to reach exhaustive protein datasets from a given biological compartment.

One main goal of "parasite-proteomics" surveys is to find proteins for use as pathogen-specific host biomarkers and to decipher of host–pathogen crosstalks. Some recent papers emphasize that a significant number of surveys were done with a nonrigorous experimental design and without a conceptual approach to disentangle a general host proteome response from a specific host proteome response during the interaction with a pathogen (Tastet et al., 1999; Ashton et al., 2001; Huang et al., 2002; Biron et al., 2005a; Holzmuller et al., 2008; 2010). A new attitude is essential to improve the reliability of proteomics data on host–pathogen interactions. Lately, some conceptual approaches have been proposed to researchers working on host–pathogen interactions to improve the reliability of "parasite-proteomics" results and to stimulate the creation of proteomic database with a holistic view of host–pathogen interactions. Thus, in this section, three new avenues to decipher host–pathogen interactions for any pathogen species (i.e., with simple or complex life cycle) are presented.

Some proteomics studies have shown common features in the innate response of plants, insects, and mammals (Broekaert et al., 1995; Rock et al., 1998; Cao et al., 2001; Taylor et al., 2003). The plant defense response is mediated by disease resistance genes (R genes), which are abundant throughout the genome and confer resistance to many microorganisms, nematodes, and/or insects (Dixon et al., 2000). R genes of several families of plants studied to date show homology with the Drosophila receptor Toll and the mammalian interleukin-1 receptor (Rock et al., 1998). In addition, plants, invertebrates, and vertebrates produce a class of peptides call "defensins" which are pathogen-inducible (Broekaert et al., 1995; Hoffman, 1995). Some peptides and/or proteins used by phytophagous or animal parasites to modify the genome expression of their host, share many structural and functional homologies. Thus for example, phytoparasitic root-knot nematodes of the genus Meloidogyne secrete substances into their plant hosts in order to make a giant cell used as a feeding site (Abad et al., 2003; Doyle and Lambert, 2003). A similar system is observed for the zooparasite, *Trichinella spiralis* (Stichosomida: Trichinellidae) (Jasmer, 1993). Furthermore, the injection of a peptide isolated from nematode secretions to either plant protoplasts or human cells enhances cell division (Goverse et al., 2000). The mechanism is not yet well-known but protein induction is considered as a strong possibility.

These days, many data are obtained by genomic and proteomics projects concerned with host–parasite interactions. Nevertheless, as mentioned above, generally little effort is made to elaborate such projects with respect to a holistic view of the goal to increase knowledge concerning immune responses of a host along with the biochemical crosstalk between host and pathogen/parasite. Thus far, "parasito-proteomics" studies have already led to new insights concerning molecular pathogenesis and microorganism identification (Moura et al., 2003; Vierstraete et al., 2004; Levy et al., 2004; Biron et al., 2005d). However, many "parasito-proteomics" studies have been done with powerful tools but without a conceptual approach to disentangle the host and parasite genome responses during their interactions.

Lately, a new holistic approach was proposed to biologists based on evolutionary concepts of the immune response of a host to an invading parasite (for more details see Biron et al., 2005a). For instance, this new conceptual approach enables the classification of the host genomic response to infection by a parasite according to the immune mechanisms used (constitutive vs. induced) and the degree of specificity. From an evolutionary-ecological point of view, host immune responses to a particular parasite can be plotted on a chart according to the immune mechanisms used (constitutive

vs. induced) and degree of specificity. The first axis of the defense chart refers to the immune mechanisms employed by the host with the two extreme cases: (1) a constitutive immune mechanism used by the host to rapidly impair the invasion by a parasite; and (2), an induced immune mechanism which has the advantage of avoiding a costly defense system, yet meanwhile has the disadvantage that the parasite might escape host control (Schmid-Hempel and Ebert, 2003). The second axis of the defense chart refers to the degree of specificity of the host immune response.

Whatever the tactics used and the degree of specificity, the host genome ensures the adequate operation of the immune response via the proteome (genome operating system). For each immune tactic, many proteins are implicated. Consequently, any researcher in parasito-proteomics working with the immune defense chart will be able to categorize the host genome reaction for any given parasite at any given time. Also, for the pathogen, from an evolutionary-ecological point of view, parasite molecular strategies used to counteract host immune system can be plotted on a chart according to the infection mechanisms used (constitutive vs. induced) and degree of specificity. This type of approach should be as much hypothesis generating for parasito-proteomics as for evolutionary ecology itself.

Lately, other pioneer proteomics studies on parasite-induced alteration of host behavior (widespread transmission strategy among pathogens) have been carried out on six arthropod host–parasite associations: two orthoptera-hairworm associations, two insect vector–pathogen associations and two gammarid-parasite associations (Lefèvre et al., 2009). These "parasito-proteomics" studies were based on the conceptual approach suggested by Biron et al. (2005a,b). Thus, in each study, many biological treatments have been effected to control the potential confusion resulting from proteins that are nonspecific to the manipulative process and to find the protein potentially linked with host behavioral changes. Also, for each study, to limit the possible effects of multiple infection and/or host sex-specific factors on the host proteome response, only mono-infected host males were used for the protoemics analysis. These "parasito-proteomics" surveys on the parasitic manipulation hypothesis showed that proteomic tools and the conceptual approach suggested by Biron et al. (2005a,b) are sensitive enough to disentangle host proteome alterations and also the parasite proteome alterations linked to many factors such as the circadian cycle, the parasitic status, parasitic emergence, the quality of a habitat, and the manipulative process.

Relationships between pathogens and their hosts and vectors depend on a molecular dialog tightly regulated. The reciprocal influence of a pathogen with its host or vector will affect the level of their genomes and their expression, respectively (Holzmuller et al., 2008). Variability and cross-regulation increase from genomic DNA (mutations, rearrangement, methylations …) through RNA transcripts (initiation, splicing, maturation, editing, stability …) to functional proteins (initiation, folding, post-translational modifications, localization, function …). Pathogeno-proteomics is an extension of parasite-proteomics to decipher more generally host–vector–pathogen interactions as they have common characteristics, which integrates modifications at all analytical levels (genome, transcriptome, proteome: whole cell content, and secretome: naturally excreted-secreted molecules) through the analysis of their end-products' profile (Fig. 11.2). The concept is based on a management with drawers of the analytic workflow, from the determination of number of experimental treatments and design of the biological material preparation to the dedicated proteomics and bioinformatics tools

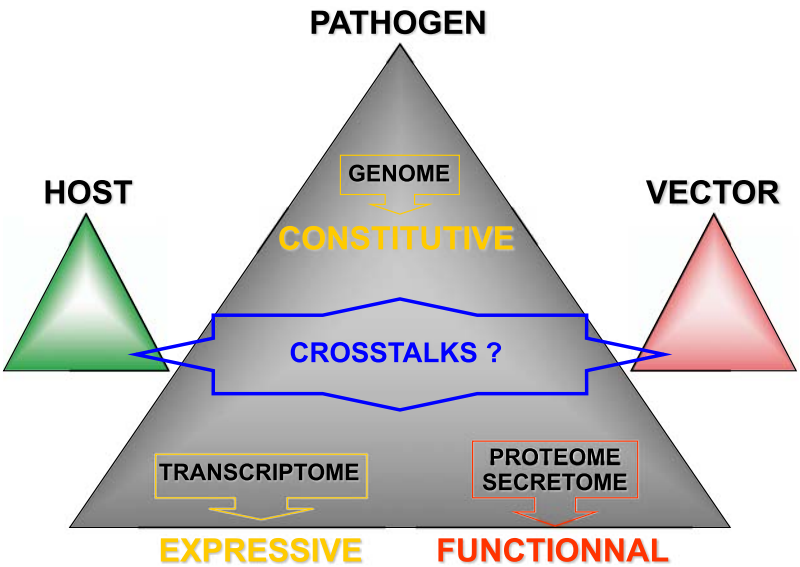


FIGURE 11.2

Pathogeno-proteomics: Integrating analytical levels in host–vector–pathogen interactions.

needed to answer a research question in cell immunobiology (directly involved in host–pathogen interactions) but also in ecology and evolution, population's biology and adaptive processes (Biron et al., 2006b; Holzmuller et al., 2008; Karr, 2008). Moreover, it has been proved that the results of this type of integrated approach has a concrete impact on the discovery of the causes of infectious diseases, as well as on improving the diagnosis, vaccine development, and rational drug design (Doytchinova et al., 2003; Bansal, 2005; Chautard et al., 2009). Despite a theoretical aspect (Kint et al., 2010), the pathogeno-proteomics concept brought new insights into important aspects of cell signaling Kleppe et al., 2006) and molecular medicine (Ahram and Petricoin, 2008; Ostrowski and Wyrwicz, 2009). As an example, proteomics and bioinformatics tools enable the formulation of relevant biological hypothesis on why part of the fungal population is killed while a significantly high percentage survives in *Candida albicans*–macrophage interactions (Diez-Orejas and Fernández-Arenas, 2008), leading to edition of a specific database for studying *Candida albicans*–host interactions (Vialás et al., 2009). Direct applications in terms of discovery of antifungal drug targets or design of new effective antibacterial vaccines become reality (Tournu et al., 2005; Jagusztyn-Krynicka et al., 2009). Other studies have also highlighted the pathogenic changes in the brain of SIV-infected monkeys (Pendyala et al., 2009), adaptive metabolic changes in *Trypanosoma cruzi* and *Trypanosoma congolense* (Grébaut et al., 2009; Roberts et al., 2009) or molecular biomarkers of intestinal disorder induced by *Helicobacter pylori* or *Tritrichomonas muris* (Wu et al., 2008; Kashiwagi et al., 2009). More recently, the use of model organisms interacting with infectious agent of medical importance emphasized the complexity

and pathogen-specificity of the worm's immune response (Bogaerts et al., 2010). Taken together these examples demonstrate the potential of the concept of pathogeno-proteomics and promote this new research avenue.

4.2 Proteomics of Arthropod Vectors: Understanding the Role of Sialomes in the Host–Vector Interactions

Many host–pathogen interactions are made possible through transmission by an hematophagous insect vector, and vector-borne diseases (VBDs) were responsible for more human deaths until the 20th century than all other causes combined, and are still a dramatic break to economic development in the tropics (Gubler, 1998). The current definition of VBDs could be as such: infectious diseases of humans and animals caused by pathogenic agents including viruses, bacteria, protists, protozoans, and helminths transmitted by the bite of blood-feeding arthropod vectors. The saliva and salivary glands are central to interactions between pathogens, vector, and mammalian host with an array of biologically active molecules, including antihemostatic, angiogenic, antiinflammatory, and immunomodulatory proteins, and therefore became in the last 10 years an attractive target to control VBDs, qualified as the ideal accomplice for corrupting host cell responses (Fontaine et al., 2011; Nuttall, 2019). As an illustration, climate change and human activities have modified the propagation of competent arthropods worldwide, making arboviruses become pathogens of major medical importance. Early transmission to vertebrates is initiated by skin puncture and deposition of virus together with arthropod saliva in the epidermis and dermis, where saliva components have the capacity to modulate skin cell responses by enhancing and/or counteracting initial replication and establishment of systemic viral infection (Briant et al., 2014; Guerrero et al., 2020). Similarly, it was demonstrated that *Ornithodoros* soft tick saliva modulated the host physiological and immunological responses during feeding on skin, thus affecting infection of pigs by African swine fever virus (Bernard et al., 2016). This can have disastrous financial consequences for farming since African swine fever is a hemorrhagic disease for which no vaccine is available, so animal slaughtering or area zoning to restrict risk-related movements are the only effective measures to prevent the spread of the disease.

Genome sequencing of some arthropod vectors combined with modern proteomic technologies expanded our knowledge of arthropod–pathogen interactions. Noteworthy, proteomic studies have identified proteins and biochemical pathways that may be involved in molecular crosstalk in vectors–pathogens associations (Aoki et al., 2022). Additionally, precision of the host–pathogen interactome will help to better characterize the molecular strategies and biochemical pathways used by arthropod vectors to cope with pathogens (Patramool et al., 2012). Sialome (or saliva transcriptome/proteome) studies also benefitted from recent advances in molecular biology, protein chemistry, and computational biology, and highlighted newly isolated genes that code for proteins with homologies to known proteins allowing identification or prediction of their function. However, most of these genes code for proteins with unknown functions therefore opening new ways to functional genomic approaches to identify their biological activities and roles both in blood feeding and pathogen transmission (Valenzuela, 2004). In a complementary way, proteomics provides information on the protein content of cells and tissues that may differ from results at the transcriptomics level and may be relevant

either for basic biological studies or for vaccine antigen discovery (Popara et al., 2015). For instance, study of the human antibody (Ab) response to *Aedes* salivary proteins can provide new biomarkers to evaluate human exposure to vector bites (Doucoure et al., 2013). Moreover, a recent proteomics study of *Amblyomma* spp. hard ticks evidenced that host and tick proteins involved in blood digestion, heme detoxification, development, and innate immunity were differentially represented between adults and nymphs, whereas proteins involved in tick attachment, feeding, heat shock response, protease inhibition, and heme detoxification were differentially represented between *Amblyomma* spp., suggesting adaptation processes to biotic and abiotic factors (Villar et al., 2014).

Finally, omics' input in improving the understanding of host–vector–pathogen interactions considering vectors as a primary pest for human and animals, and revisiting the concepts of virulence and pathogenesis, human and animal hosts' responses, arthropods' blood feeding and vector potency, will also positively influence development of global control measures against VBD.

4.3 Pitfalls of the Current Proteomics Approaches

The classical approaches in parasito-proteomics or pathogeno-proteomics make it possible to identify proteins of interest for a given host–parasite association. For example, Wattam and Christensen (Wattam and Christensen, 1992) associated some polypeptides with the genome response of the host mosquito, *Aedes aegypti* (Diptera: Culicidae) with the invasion of the filarial worm, *Brugia malayi*. This pioneering study provided important new information on the response of the host insect to invasion by a specific parasite species. Nevertheless, it was *not* possible to determine whether that the response detected in *Ae. aegypti* is specific to *B. malayi* (Spiruria, Filariidae), or whether it can be observed for any parasitic worm species invading a dipterous host. Other studies have revealed the limitations of the current approach in pathogeno-proteomics by showing that in the host–pathogen interaction, many immune mechanisms are involved (constitutive, induced, specific, or otherwise) (Haab et al., 2001; Levy et al., 2004; Vierstraete et al., 2004). By using two treatments, the injection of lipo-polysaccharides (LPS) and a sterile injury, Vierstraete et al. (2004) were able to disentangle proteome modifications induced by immunity from those induced by a physical stress. Levy et al. (2004) studied the immune response of the fruitfly, *Drosophila* to bacterial (*Micrococcus luteus* and *Escherichia coli*) and fungal (*Beauveria bassiana*) infections. The data revealed that 70 of the 160 protein spots detected were differentially expressed at least fivefold after a bacterial or fungal challenge. In addition, the majority of these spots were specifically regulated by one pathogen, whereas only a few spots corresponded to proteins altered in all cases of infection. In summary, the classical approach in parasito-proteomics has many benefits in terms of understanding fundamental aspects of gene–protein functional interactions. Unfortunately, it is not applicable to different pathogen species (and as such, does not encourage the growth of knowledge of general host proteome responses), nor does it necessarily help in the creation of a proteomic database with a holistic relevance to the understanding of host–parasite interactions.

Moreover, one classical pitfall in "parasite-proteomics" surveys is the use of a single proteomic technique. Recently, Bridges et al. (2008) have demonstrated by using a battery of proteomic techniques to characterize the plasma membrane subproteome of bloodstream form of *T. brucei* that these techniques are complementary since each one has identified a unique subset of proteins of the plasma

membrane. Finally, although 2-DE and MS have been very successfully employed to identify proteins involved in host–parasite crosstalks, many recent papers have emphasized the pitfalls of 2-DE experiments, especially in relation to experimental design, poor statistical treatment, and the high rate of "false-positive" results with regard to protein identification (Barrett et al., 2005; Biron et al., 2006; Holzmuller et al., 2008). It is necessary to be careful in the interpretation of results for the previous "parasito-proteomics" surveys on host–parasite interactions (see Biron et al., 2006a; Holzmuller et al., 2008).

Just as the host–vector–pathogen interaction is tripartite, the pitfalls that punctuate a proteomic analysis are of three orders; the data processing according to the conceptual approach as we have just seen above, also the biochemical processing of the samples as we will see later for metaproteomics or metabolomics, but above all the quality of the databases including the quantity and quality of annotation varies enormously between taxonomic entities. Actually, mass spectra generated by the biological samples analyzed are matched to theoretical mass spectra of the proteins of the targeted organism. But depending on the genome sequencing and corresponding annotation, the considered taxon will be poor or rich in information (Reeves et al., 2009; Heyer et al., 2017; Dupree et al., 2019). Moreover, to control the error rates in mass spectra identification, proteomics depends on the target-decoy search strategy, the de-facto method for false discovery rate (FDR) estimation, which can suffer from reduced sensitivity and specificity due to inflated database size (Aggarwal et al., 2022).

The pitfalls that have just been mentioned illustrate the importance of matching the proteomic strategy to the research question posed (from the experimental design of the biological samples to their analysis in mass spectrometry, taking into account taxonomic constraints).

4.4 Proteo-Genomics: Next-Generation Proteomics to Deepen the Characterization of Proteins, Their Biological Functions, and Their Interactions in Complex Environments

The last decade has seen the bridge between genomics and proteomics, which constitutes the last stage in the evolution of the means of dissecting interacting proteomes. This comes in response to the problems of annotation and the level of completeness of the databases, which are a major source of the pitfalls mentioned above for proteomic analyzes. Actually, high-throughput identification of proteins and their accurate partial sequencing by shotgun nanoLC-MS/MS are now feasible for any cellular model at a full genomic scale, and the integration of polypeptide sequence information together with the corresponding genome led to the definition of proteogenomics (Armengaud, 2009). It has recently been extended to proteo-transcriptomics to integrate the intermediate genome expression level (Levin and Butter, 2022). This brings new insights into mining microbial proteomes, with validation of predicted orphan genes and correction of genome annotation errors such as discovery of unannotated genes, reversal of reading frames and identification of translational start sites, stop codon read-throughs, or programmed frameshifts (Fig. 11.3). Increasing examples of the interest of proteogenomics on newly or already sequenced model genomes are available in the literature, adding value to understanding the mechanisms of pathogen virulence or host response to infection, as illustrated for microorganisms as different as *Escherichia coli* (Tomazella et al., 2012), *Mycoplasma*

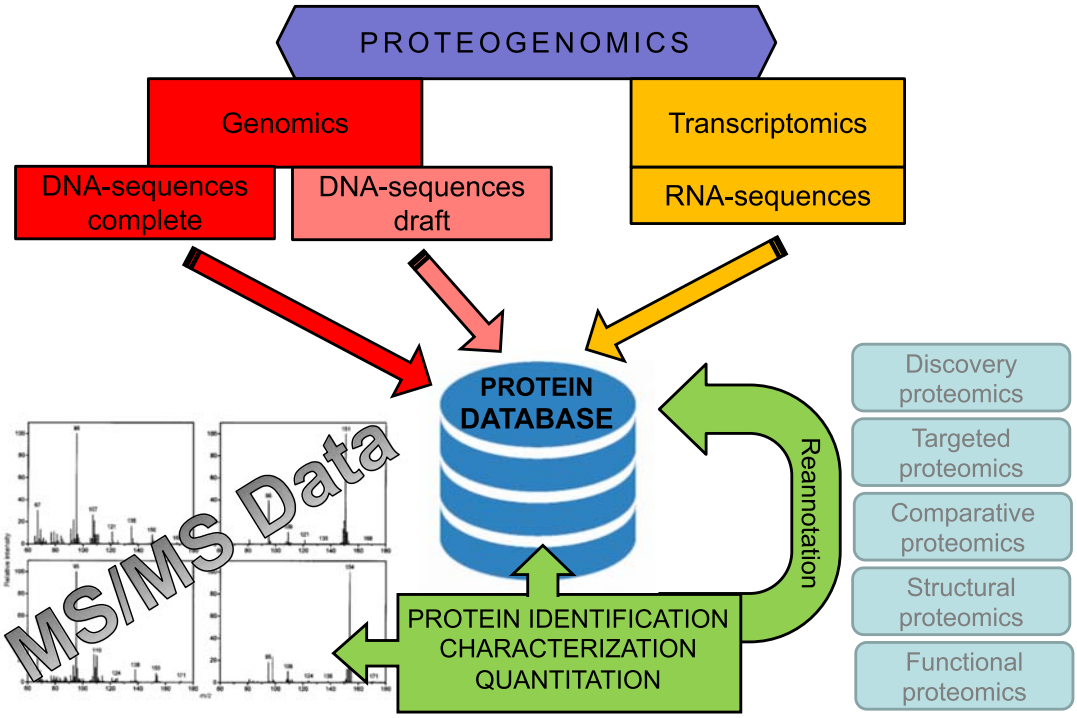


FIGURE 11.3

Proteogenomics: Assignment of MS/MS data to protein database designed from the nucleic acid information (either complete or draft DNA-sequence or RNA-sequences data). The resulting mass spectrometry–certified proteins participate in the database reannotation and improvement, and can be used for discovery proteomics, targeted proteomics, comparative proteomics, structural proteomics, or functional proteomics.

(Pendarvis et al., 2014), SARS-CoV-2 virus (Tushir et al., 2021), or *Toxoplasma* (Nie et al., 2022). Moreover, proteogenomics helps in establishing the most comprehensive and quantitative protein list of cellular processes, a mandatory prerequisite prior to analysis of the cell-wide dynamics of proteins, their posttranslational modifications, their complex network interactions, and interpretation of these data as a whole, constituting the basis for systems biology (Armengaud, 2010). An important element in host–vector–pathogen interactions, apart from the physical cellular interaction that occurs during the infectious cycle, is remote communication via soluble mediators, particularly in complex biological systems. These exoproteomes, also called secretomes, often contain key biomarkers of the host–pathogen interaction, which proteogenomics allows to finely characterize. It has been shown that these exoproteomes reflect the physiological state of the cells in a given condition and are indicators of

how living systems interact with their environments, such as microorganisms to modulate the host's response to its advantage (Armengaud et al., 2012; González-Fernández et al., 2015; Maffei et al., 2017; Vidal-Veuthey et al., 2022; Skelly and Da'dara, 2022).

A significant contribution of proteogenomics, linked to the current possibility to obtain draft genomes using next-generation sequencers in a rather straightforward way and at a reasonable cost for any organism, is the facilitation in studying nonmodel organisms (Armengaud et al., 2014). The use of experimental proteomics data to refine genome annotations, coupled to a higher integration of omics data, enabled these last years the molecular studies of nonmodel organisms at an unprecedented depth, bringing providing a mass of information on many pathogens, qualified as "exotic" or belonging to the family of "neglected diseases" allowing a better understanding of the interactions with their hosts and the development of new control tools (Holzmuller et al., 2013; Cross et al., 2016; Saviola et al., 2020; Nicholson et al., 2021).

4.5 Host–Pathogen Interactomes

The last few years have witnessed the birth of new biological entities named interactomes. They correspond in an "ideal world" to the complete set of protein–protein interactions existing between all the proteins of an organism. In reality, they are far from complete since an unknown number of interactions are yet to be discovered. Current interactomes are only a part of the whole set of possible interactions occurring within an organism or between organisms. They are generally assembled from (i) the results of large-scale two hybrid screens (LS-Y2H) (around 6000, 4000, 23,000, and 5500 interactions for yeast (Uetz et al., 2000; Ito et al., 2001), the nematode, *Caenorhabditis elegans* (Maupas) (Li et al., 2004), Drosophila spp. (Giot et al., 2003; Formstecher et al., 2005) and humans (Rual et al., 2005; Stelzl et al., 2005), respectively); and (ii) the interactions identified by low-scale experiments described in the literature that may be eventually compiled in specialized databases (e.g., INTACT (Hermjakob et al., 2004), MINT (Zanzoni et al., 2002), HPRD (Peri et al., 2003), BIND (Alfarano et al., 2005)). Consequently, they do not reflect temporal influences because interactions are gathered from different cell types, tissues, development stages and types of experiment.

Interactomes form large intricate networks leading to a renewed vision of cell biology as an integrated system. However, extracting and revealing the functional information they contain depends on our ability to analyze them in detail. For this, bioinformatics methods which partition the interaction network into functional modules have been proposed. These modules usually correspond to group of proteins involved in the same pathway, the same protein complex, or the same cellular process.

Since interaction networks are represented by complex graphs in which nodes correspond to proteins and edges to the interactions, a number of these network analysis methods have been grounded on principles that derive from graph theory.

Noticeably, a functional module or a class of protein that is functionally related and based on network analysis can be deduced from (i) a search for graph regions particularly densely populated by interactions (Bader and Hogue, 2003; Spirin and Mirny, 2003); (ii) the similarity between the shortest paths in the graph (Rives and Galitski, 2003); (iii) the progressive disconnection of the graph using a

calculation of edge "betweenness" (Girvan and Newman, 2002; Dunn et al., 2005); and (iv) the sharing of interactors (Brun et al., 2003; Samanta and Liang, 2003) or a combination thereof (Brun et al., 2004). Some of these methods use the functional annotations of the protein (such as Gene Ontology annotations) to annotate the functional modules they predict. Based on the characteristics of a protein of unknown function to some of these annotated modules or classes, a putative function for such proteins can be proposed (Brun et al., 2003). Currently, specialized methods are being developed to investigate the interactomes. Thus, some softwares, "plugins," or servers are available for a free use by the proteomics research community (for instance, MCODE in Cytoscape (Shannon et al., 2003); Prodistin (Baudot et al., 2006)). But it is clear that the field starts to move forward, using microarrays or protein–protein interaction bioanalyzers (Arnold et al., 2012; Uzoma and Zhu, 2013; Ghadie et al., 2018).

Although the deciphering of the interactomes of the main model organisms is not yet complete, studies of the interactomes of pathogens are increasing. The first pathogens to be investigated in terms of their interactomes were the hepatitis C virus (Flajolet et al., 2000) and the bacterium, *Helicobacter pylori* (Rain et al., 2001). More recently, the interactomes of the herpes viruses (Uetz et al., 2006) and the malaria parasite, *Plasmodium falciparum* (LaCount et al., 2005), have been determined. This starting point initiated by Uetz et al. (2006) launched the docking of the interactomes of pathogens onto those of their hosts, as illustrated by the interactomes of influenza virus (Chua et al., 2022) with an impact on antiviral drug development (Watanabe and Kawaoka, 2015), or Salmonella (Schleker et al., 2012) and recently SARS-CoV-2 (Jahanafrooz et al., 2022). The analysis of the host–pathogen interactome (i.e., "docked interactomes") during the host–pathogen crosstalk is certainly a very promising and exciting aspect of interactomics because of its obvious potential impacts on human and animal health (see Fig. 11.4). The host–pathogen interactome will permit to identify host and pathogen protein networks linked to specific functions during their interaction, as illustrated by the molecular elucidation of human hemostasis dysregulation during hematophagous triatomine meal (de Araujo et al., 2012).

4.6 Population Proteomics, a Growing Discipline, to Study Host–Pathogen Interactions

The host susceptibility to a pathogen and/or the pathogen virulence are often fluctuating within a host population even when infected hosts are collected in the same habitat and at the same time. Despite pathogen-specific signatures can be found, the global individual response of the host remains preponderant (Holzmuller et al., 2016). This host phenotypic variability can be caused by three factors: (i) host genotype and/or pathogen genotype, (ii) different environmental experiences (ex.: habitat fragmented in microclimates), and (iii) host coinfection by pathogens (i.e., competition or mutualism among co-infecting pathogens within hosts). It is now possible to detect and to decipher the host proteome variability within a habitat for the molecular mechanisms and for the protein networks involved in the host–pathogen interactions. In this section, a new growing discipline in proteomics, the population proteomics, and its prospects are presented with results of some pioneer studies on this topic, especially in human population proteomics.

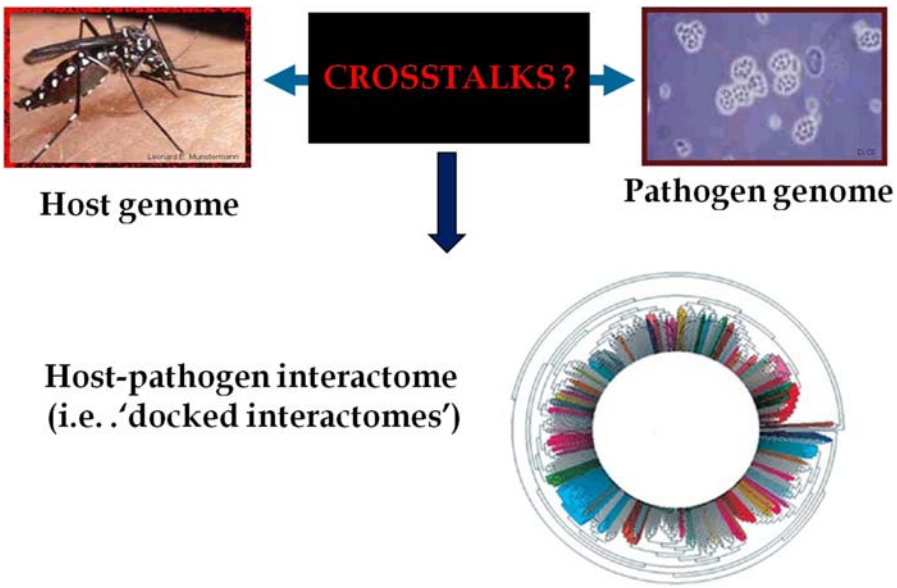


FIGURE 11.4

A new biological entities named host–pathogen interactome corresponding to complete set of protein–protein interactions existing between all the proteins of a host and a pathogen during their interaction.

4.6.1 Prospects With Population Proteomics for Any Living Organisms

One limiting factor for the first generation of proteomics tools (e.g., 2-DE) is the amount of proteins required to study the host and/or pathogen proteome expression(s) during their interactions. Most surveys were done by pooling many individuals for any treatment (e.g., infected and noninfected hosts) required to answer to a query. Thus, with this kind of experimental protocol, no data can be acquired on the interindividual variation in expression of host and pathogen proteomes during their crosstalk. New proteomics tools and methods have been developed as 2D-LC/MS that can permit to study the interindividual variation of molecular crosstalk in host–pathogen associations (Nedelkov et al., 2004; Predel et al., 2004; Brand et al., 2005).

Population proteomics was first defined by Dobrin Nedelkov as the study of protein diversity in human populations, or more specifically, targeted investigation of human proteins across and within populations to define and understand protein diversity with the main aim to discover disease-specific protein modulations (Nedelkov et al., 2004; 2008). Biron et al. (2006b) have proposed to broaden the "population proteomics" concept to all living organisms with the aims to complement the population genetics and to offer a new avenue to decipher the crosstalk diversity involved in trophic interactions in a habitat since the execution of the genetic plan is carried out by the activities of proteins and natural selection acts at the protein level (Karr, 2008; Cieslak and Ribera 2009).

The apparent separation between genomics and proteomics that leads to different perspective on the same ecological reality is a fundamental limitation that needs to be overcome if complex processes, like adaptation, pathogen virulence, and host susceptibility, are to be understood. Population proteomics coupled with population genetics has a great potential to resolve issues specific to the ecology, the evolution of natural populations, the dynamic of host susceptibility to pathogens, the evolution of pathogen virulence, and the range of host genotypes that can be infected with a given pathogen genotype in host–parasite interactions. Some perspectives for the population proteomics are resumed in Fig. 11.5. Even if we are yet far from this "promised land," a better understanding of the information contained in proteomics markers should permit an impressive amount of information to be gathered on the past as well as current environmental conditions experienced by a given population of a biological species, something that could be summarized as "show me your proteome and I will tell you who your are, where are from, and where you should go from here."

Pioneer surveys on population proteomics have been carried out with classical proteomic tools (i.e., 2-DE and mass spectrometry) (i) to determine the genetic variability between species and between

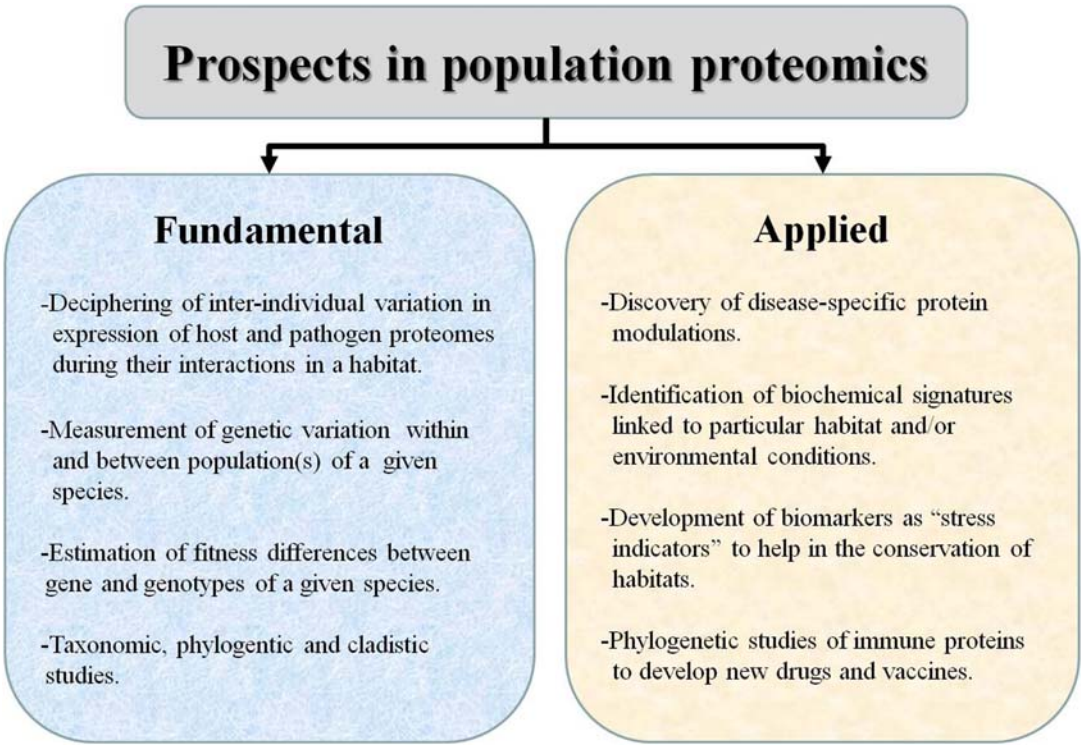


FIGURE 11.5

Potential of population proteomics as an emerging discipline in proteomics.

populations of a given species (Chevalier et al., 2004; Diz and Skibinski, 2007; Valcu et al., 2008), (ii) to identify biochemical signatures linked to particular habitat and/or environmental conditions (Thiellement et al., 1999; Pedersen et al., 2010), and (iii) to phylogenetic studies (Navas and Albar, 2004; Dorus et al., 2006). Nedelkov et al. (2005, 2007) have investigated the human plasma proteins diversity by using approaches similar to enzyme-linked immunosorbent assay but utilizing mass spectrometry as method of detection (Nedelkov, 2008). These pioneer results helped to discover disease-specific protein modulations but also to find pathogen-specific protein biomarkers. A growing number of studies have then been published on the characterization of biomarkers in the fields of host–pathogen interactions (Wu et al., 2008), the development of tools for controlling infectious diseases (Tanca et al., 2013), antimicrobial resistance (Park et al., 2016), or more recently the characterization of the pandemic coronavirus (Perrin-Cocon et al., 2020).

4.6.2 Human Population Proteomics

Human population proteomics deciphers protein diversity in human populations. In a broader term, human population proteomics can be compared to human population genomics, where individuals are interrogated with the aim of cataloging common genetic variants and determining how they are distributed among people within populations and among populations in different parts of the world (Nedelkov et al., 2004, 2006; Nedelkov, 2005).

Human population proteomics does not engage the study of entire proteomes because it is very likely that, for a specific cell or tissue proteome, there is no definitive set and number of proteins that is common to all within a group or a larger population. Instead, human population proteomics focuses on interrogation of a selected number of proteins but from a large number of individuals, to delineate the distribution of specific protein modifications within these subpopulations. Hence, targeted protein analysis approaches utilizing mass spectrometry as detection method are employed. Mass spectrometry measures a unique feature of each fully expressed protein—its molecular mass. Changes in the protein structure resulting from structural modifications are reflected in its molecular mass and can be detected via MS, without a priori knowledge of the modification. The MS methods utilized in human population proteomics must be capable of analyzing hundreds, if not thousands of samples per day, with high reproducibility and sensitivity. Hence, top-down MS approaches utilizing affinity ligands are the most likely methods of choice for population proteomics (Nedelkov, 2006). Surface-immobilized ligands can be utilized to affinity-retrieve a protein of interest from a biological sample, after which the protein (with or without the affinity ligand) is introduced in a mass spectrometer. One of the first affinity MS methods developed was Mass Spectrometric Immunoassay (MSIA) (Nelson et al., 1995). The approach combines targeted protein affinity extraction with rigorous characterization using MALDI-TOF mass spectrometry (Fig. 11.6). Protein(s) are extracted from a biological sample with the help of affinity pipettes derivatized with polyclonal antibodies. The proteins are eluted from the affinity pipettes with a MALDI matrix, and are MS-analyzed. Enzymatic digestion, if needed, is performed on the MALDI target itself. Specificity and sensitivity, as in traditional immunoassays, are dictated by the affinity-capture reagents—the antibodies.

However, a second measure of specificity is incorporated in the resulting mass spectra, wherein each protein registers at specific m/z value. During data analysis, the major signal in the mass spectrum that corresponds to the targeted protein is initially evaluated; it should be within a reasonable range (e.g., error of measurement of $<0.05\%$) from the value of the empirically calculated mass obtained from the sequence of the protein deposited in the Swiss-Prot databank. Once this mass value is

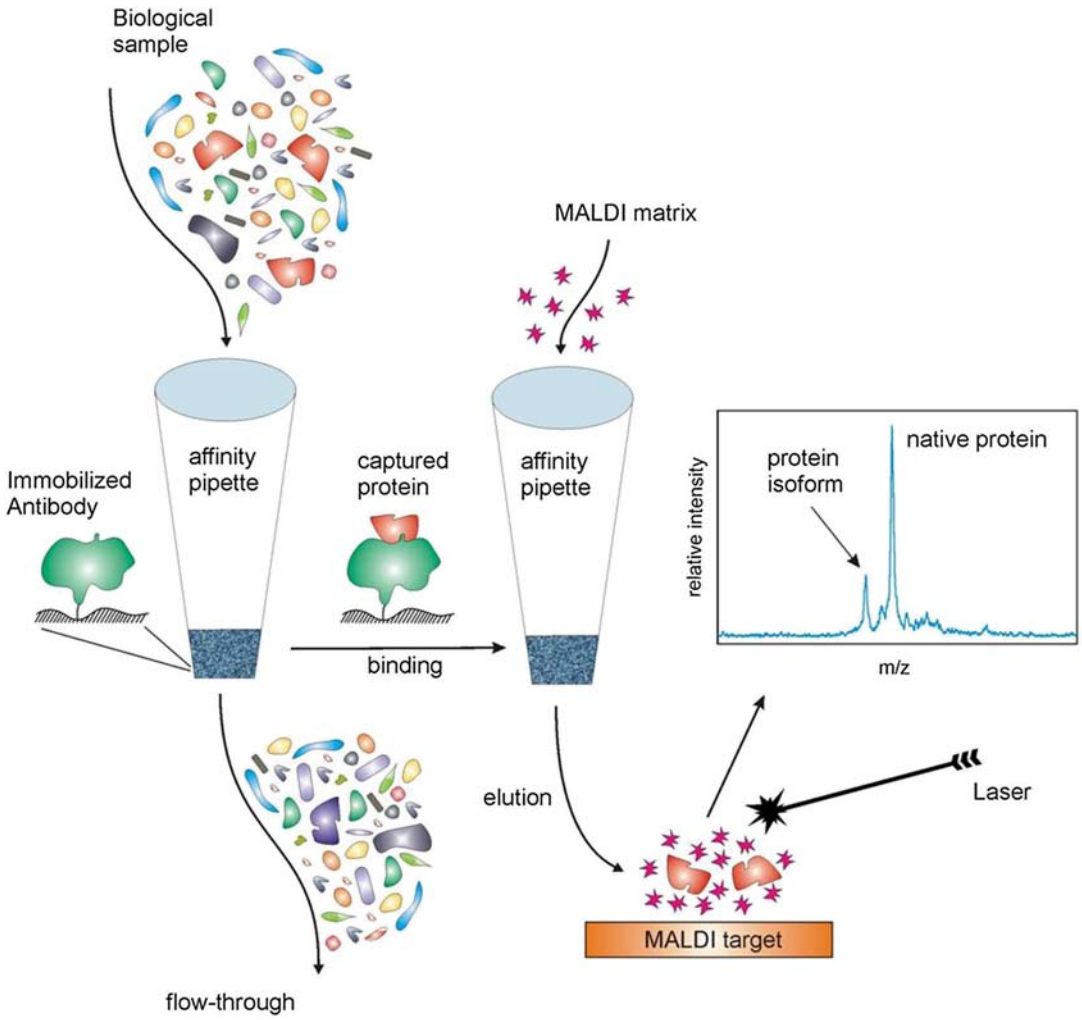


FIGURE 11.6

Schematics of the mass spectrometric immunoassay (MSIA) approach.

confirmed (or observed to be shifted), the presence of protein modifications is noted by the appearance of other signals in the mass spectra (usually in the vicinity of the native protein peaks), or by mass shifts of the major protein signal. Modifications can be tentatively assigned by accurate measurement of the observed mass shifts (from the wild-type protein signals and/or in-silico calculated mass) and knowledge of the protein sequence and possible modifications. The identity of the modifications is then verified using proteolytic digestion and mass mapping approaches in combination with high-performance mass spectrometry.

In an initial study of human protein diversity using mass spectrometric methods of detection, 25 plasma proteins from a cohort of 96 healthy individuals were investigated via mass spectrometric immunoassays (Nedelkov et al., 2005). The protocol and an example of the data generated for one of the proteins, transthyretin (TTR), are outlined in Fig. 11.7. The transthyretin MSIA assays were performed in parallel on the 96 human plasma samples using affinity pipettes derivatized with anti-transthyretin antibody. Following mass spectrometric analysis, data matrix containing all tentatively assigned modifications was assembled. Then, peptide-mapping experiments were performed on selected number of samples to identify the specific modifications and finalize the modifications database. The data for all 25 proteins is presented in Fig. 11.8, which lists the modifications observed for 18 of the 25 proteins studied (modifications were not observed for 7 proteins), and shows the frequency of each modification in the 96-samples cohort. A total of 53 protein variants were observed for these 18 proteins, stemming from posttranslational modifications and point mutations. The largest number of posttranslationally modified protein variants was found to be C- or N-terminal truncated protein isoforms. Deglycosylation, oxidation, and cysteinylation were also observed among several of the proteins. Among the point mutations detected for four of the proteins, notable was the high incidence of point mutations for apolipoprotein E and transthyretin, which is consistent with genomic studies that have found these proteins to be highly polymorphic. The overall frequency of the modifications in the 96-samples cohort was wide ranged. Fourteen modifications were observed in all 96 samples, suggesting that they must be regarded as wild-type protein forms. Others, such as most of the point mutations, were present in only few of the samples. Overall, 23 of the modifications were observed in more than 65% of the samples, and 20 in less than 15% of the 96 samples analyzed. Upon further data analysis, and taking into the consideration the gender, age, and ethnicity of the individuals who provided the samples, it was determined that the Gly6Ser mutation in transthyretin was detected only in individuals of Caucasian origin, which is consistent with existing knowledge about the occurrence of this common nonamyloidogenic population polymorphism in Caucasians (Connors et al., 2003). Another correlation was observed in regard to interprotein variations in specific individuals: all seven individuals for which carbohydrate deficient transferrin was detected were also characterized with deglycosylated antithrombin III.

Following this small-scale protein diversity study, a second study of human protein diversity was recently carried out wherein the number of samples was greatly expanded in order to get an accurate view of the distribution of some of the protein modifications in the general population (Nedelkov et al., 2007). One thousand individuals from four geographical regions in the United States (California, Florida, Tennessee, and Texas) were selected and the protein modifications for beta-2-microglobulin (b2m), cystatin C (cysC), retinol binding protein (RBP), transferrin (TRFE), and transthyretin (TTR) were delineated (in the 96-samples study, these 5 proteins accounted for 19 of the 53 protein variants observed). The results of the study are summarized in Fig. 11.9, which lists the protein modifications observed and the frequency of each in the 1000 samples cohort. A total of 27 protein modifications (20 posttranslational modifications and 7 point mutations) were detected, with various frequencies in the cohort of samples. Variants resulting from oxidation were observed most frequently, along with single amino acid truncations. Least frequent were variants arising from point mutations and extensive sequence truncations. In total, 6 modifications were observed with high frequency (present in $>80\%$ of the samples), 5 were of medium frequency (20%–50% of the samples), and 16 were low frequency

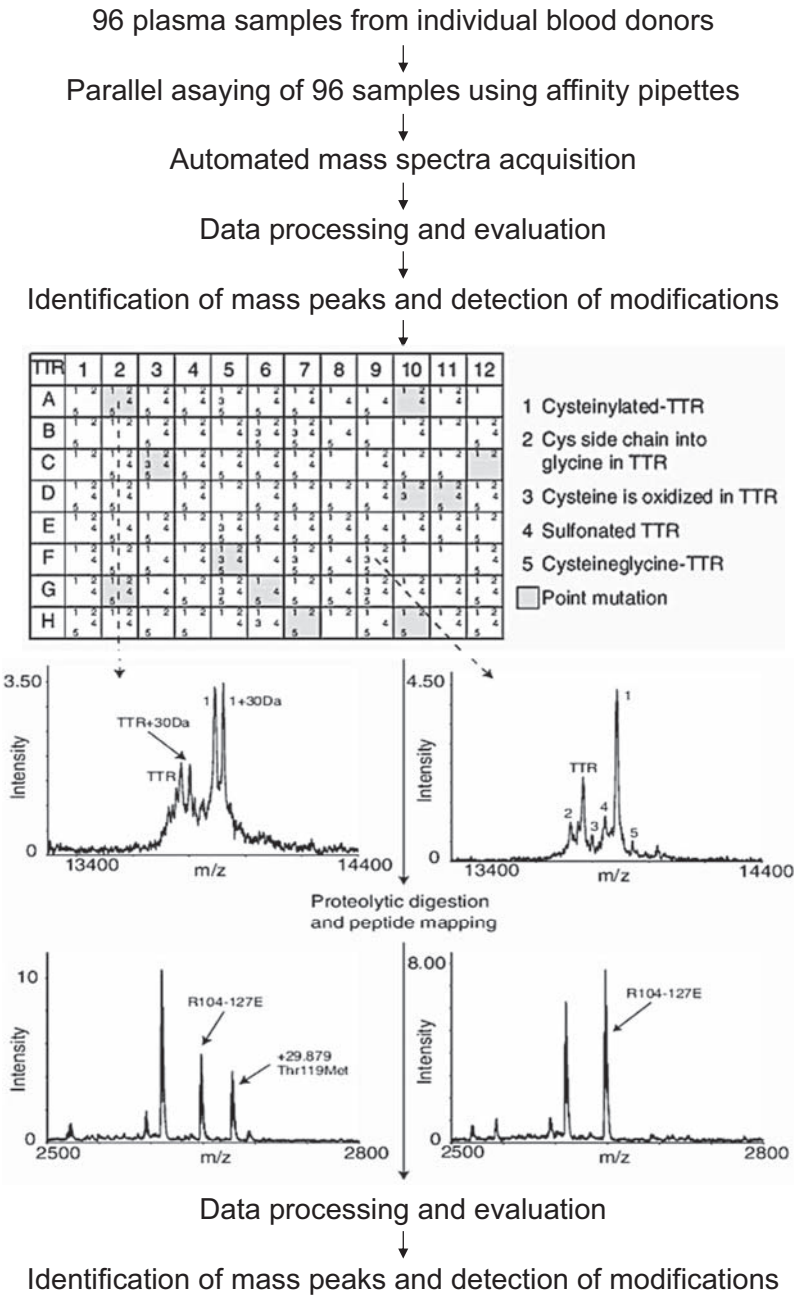


FIGURE 11.7

An outline of a population proteomics approach using TTR as an example.
m/z, Mass-to-charge ratio; *TTR*, Transthyretin.

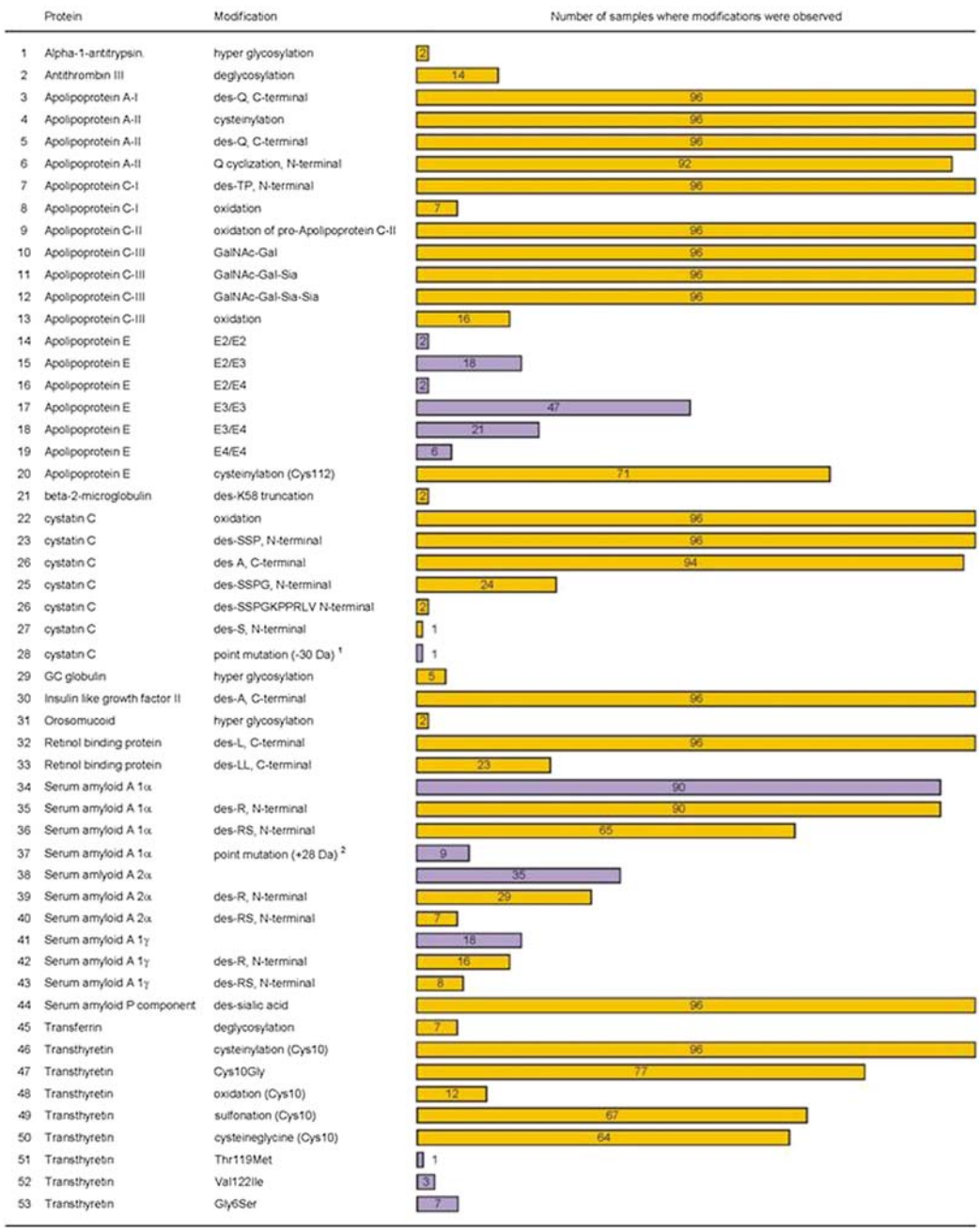


FIGURE 11.8

Modifications observed in 18 of the 25 proteins analyzed from 96 human plasma samples.

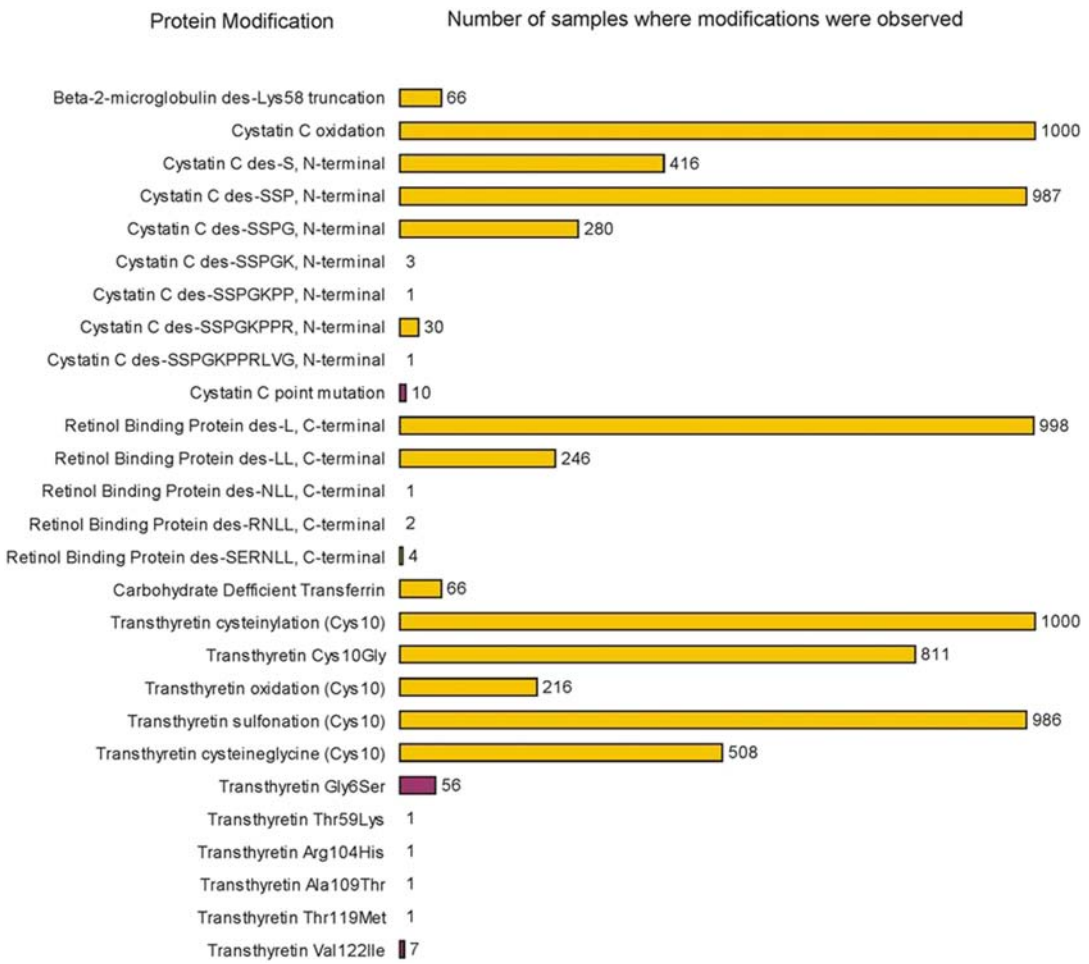


FIGURE 11.9

Modifications observed for five proteins studied from 1000 human plasma samples.

modifications observed in $<7\%$ of the samples. Nine of the low frequency modifications were not observed in the 96 individuals study. Thus, by increasing the size of the population it became possible to detect these low-occurrence protein modifications. When the frequencies of the modifications in the two studies were compared, an excellent correlation was obtained. For example, in both cohorts $\sim 7\%$ of the individuals were characterized with carbohydrate-deficient transferrin. Upon further data analysis based on the gender, age, and geographical origin of the individuals who provided the samples, it was determined that the samples obtained from California contained significantly less protein modifications than the samples obtained from Florida, Tennessee, and Texas, even though the samples from all four states were collected in the same way within a 3-month window in the spring of

2005, and stored under identical conditions until analysis. Correlations were also made in regard to the gender distribution of two protein modifications. Carbohydrate deficient transferrin was observed in ~1% of the females and ~10% of the males in the 1000 cohort. Carbohydrate deficient transferrin is an FDA-approved clinical biomarker for alcoholism, and this gender correlation can partially be explained by the higher prevalence of alcohol dependence in males than in females. The second gender correlation was related to cystatin C: all 10 of the cystatin C point mutations were found in males.

Two conclusions can be made from these two systematic studies of protein modifications and variants. First, mass spectrometry is capable of detecting structural protein modifications, and, when coupled to immunoaffinity separations, it can be employed in a high-throughput systematic study of human protein diversity. Second, the human protein diversity is far more complex than the variation observed at the genetic level. While it might be premature to declare the human proteins variation "the next big thing," it is reasonable to predict that assessing human proteome variations among and within populations will be a paramount effort that can facilitate biomarker discovery. Such endeavor would represent a paradigm shift in proteomics with significant clinical and diagnostic implications, as protein variations, quantitative and qualitative, begin to be associated with specific diseases.

While the healthy population protein diversity cataloging is a pretty straightforward proposition, the ultimate question is how human population proteomics can enable better disease diagnosis and management. Because of the one-protein-at-a-time approach, most studies involving cohorts of sick versus healthy control samples must first make an "educated guess" on what proteins to analyze via population proteomics. To start with, there is plenty of existing data on specific plasma proteins and their quantitative modulations with specific diseases. Hence, these proteins would be the first to benefit from qualitative reassessment via human population proteomics; next in line are proteins that are within the biomarker proteins' interaction network and pathways. To illustrate this point, structural isoforms have recently been discovered for some well-established biomarkers such as cardiac troponin I (Jaffe and Van Eyk, 2006; Peronnet et al., 2006) and B-type natriuretic peptide (Brandt et al., 2006; Lam et al., 2007). These structural variants might prove to be better sensitivity and specificity biomarkers than the native proteins themselves. To assess their modulation with disease, quantitative assessments of the isoforms abundance can be made by comparing the ratios of the isoforms and the native protein signals, or via standard curve approaches. The mass spectrometric immunoassays present straightforward means of looking into the protein microheterogeneity, using the well-established methods of immunoaffinity separation and mass spectrometric detection. As such, it is expected that MS-based immunoassays will readily be accepted into the clinical and diagnostic laboratories to study the effects of protein modifications in pathological processes and evaluate their potential as new biomarkers of disease (Trenchevska et al., 2016; Nedelkov, 2017).

5. Five-Year View

5.1 Metabolomics and Fluxomics, Key "Omics" Tools to Complete Deciphering of Host–Pathogen Crosstalks

Metabolomics (i.e., metabolic profiling) and fluxomics (i.e., rates of metabolic reactions within a biological entity) are concerned with the measurement of global sets of low-molecular-weight metabolites to detect changes in cell behavior and organ function. The term "metabolome" refers to the complete set of metabolites found in an organism (Peltonen and McKusick, 2001). Metabolomics

approaches use high-throughput analytical techniques such as chromatography, NMR spectrometry, and mass spectrometry to measure populations of low-molecular-weight metabolites in biological samples and are complementary to proteomics analyses. These large-scale efforts involve the identification and quantification of known and also yet unknown metabolites in host tissues and host fluids. Metabolite profiles can be important indicators of pathological states of a host and raise the possibility of identifying the host microbiota (endosymbionts, commensal microorganisms, pathogens) together with novel markers linked to the infection process of a specific pathogen (Pettersen et al., 2021; Wang et al., 2020). The current constraints in the development of metobolomics and metaproteomics lie in the sample preparation required for successful quantitative analyses of metabolic modifications (Kunath et al., 2019).

Moreover, pathogens, especially extracellular pathogens with a complex life cycle as the malaria and the sleeping sickness etiologic agents, must constantly monitor and respond to environmental changes in their intermediate (i.e., insect vector) and final (human and/or animal) hosts. How pathogens detect these changes is a black box, but they must have the ability to sample changes in nutrients and other small molecules. Afterward pathogens reprogram their gene expression profile in response to host environmental changes. The metabolomics is likely to bridge data from other "omics" tools. Correlation between the pathogen metabolites expressed during its life cycle in its hosts with global view of genome and proteome expression profiles may lead to new insight into how a pathogen interact with host cells during a host−vector−pathogen crosstalk.

5.2 Applied Outputs: New Diagnostic Tools and Identification of New Therapeutic Target

Despite the efforts of recent years, we still lack reliable biomarkers for diagnosis, prediction of clinical outcomes for many infectious diseases, and therapeutic follow-up of human diseases. Specific proteomic fingerprints might be present in biological fluids or tissues in response to the infection and could be useful for early detection of the disease, by noninvasive (saliva, urine, serum) or invasive (cerebrospinal fluid, tissue biopsies) sampling, or could constitute therapeutic targets. Before identifying a biomarker having great potential to become an important diagnostic tool, it is important to verify the clinical applicability of a technique. In fact, different criteria are necessary for a clinically applicable technique, the reproducibility, sensitivity, specificity, and rapidity.

Over the last few years, several studies have demonstrated that comparative protein profiling using a newly developed, high-throughput technique, without a priori knowledge of the proteins present, namely surface-enhanced laser desorption/ionization−time of flight mass spectrometry (SELDI-TOF-MS), breaks new ground in diagnostic (Fenollar et al., 2006; Buhimschi et al., 2007; Kanmura et al., 2007; Wu et al., 2009). This technique is potentially powerful investigative tool to improve the understanding as well as the diagnostic/prognostic capabilities in many human diseases. Nevertheless, if using such approach, it is important that potential new biomarkers, first of all for early diagnosis being validated in larger number of sample to avoid the risk of false-significant results.

Inspired from the genomes exploration, a metaproteomic strategy, namely proteomics shotgun, has also emerged in the last years to facilitate the understanding of host−pathogen interactions and accelerate the discovery of novel protein biomarkers with potential diagnostic and therapeutic potential (Swanson and Washburn, 2005; Spivey, 2009; Starr et al., 2018). Based on different MS workflows (e.g., capillary isotachophoresis-(CITP)-based multidimensional or multidimensional LC

separation systems coupled tandem MS), it promotes integration of proteomics and metabolomics datasets from direct analysis of a tissue or a biological fluid (Dowell et al., 2008; Pan et al., 2008; Fang et al., 2009; Kim and Moon, 2009; Mangé et al., 2009; Maccarrone et al., 2010). Although strategic, these technologies offer a "naïve" global approach of potential biomarkers but taking into accounts the level of expression, the posttranslational modifications and the maturation state (Ahrné et al., 2010; Park and Yates, 2010; Zhang et al., 2010b). These technologies are in their infancy in the analysis of host–pathogen interactions, but some pioneering studies have already highlighted some new biomarkers with diagnostic and therapeutic potential (Florens et al., 2004; Athanasiadou et al., 2008; Lal et al., 2009; Walters and Mobley, 2009; Vaezzadeh et al., 2010). Another promising aspect of proteogenomics is the ability to create gene–protein–disease maps, which provide frameworks to connect etiologically related diseases, to provide biological context for new or emerging disorders, and to integrate different biological domains to establish mechanisms for known disease-linked genetic regulation of proteins (Pietzner et al., 2021).

Also, understanding the three-dimensional structure of proteins, their posttranslational modifications, their biological functions, and the interaction between host and pathogen molecular constituents is necessary to validate a potential biomarker in its natural molecular environment. Advances in proteomics can consider a more functional understanding of molecular dialog and conflict that govern host–pathogen interactions, and thus develop ever more efficient both fishing tools to improve diagnosis and drug design to improve therapy of infectious diseases.

5.3 Bioterrorism and Proteomics

Following the events of "nine eleven" and the subsequent postal anthrax attacks in the United States, the possibility of further bioterrorism attacks became globally all too real. As a direct consequence of this, the US government expanded its "biodefense" program, with studies ranging from basic research to applications in detection, prevention, and treatment of diseases caused by such microbiological agents, the net result of which has been great progress in understanding their genomics (Fauci et al., 2005). Efforts were focused on the three major categories of Critical Biological Agents classified by the Centers for Disease Control & Prevention (http://www.bt.cdc.gov/agent/agentlist-category.asp#catdef) and proteomics emerge as a tool of interest (Drake et al., 2005). In the postgenomic era, the benefit of having the full sequence of the genomes of these agents is obvious. There are now genome sequences for a few isolates of each species, which have made studies of comparative genomics a reality and which are leading to important discoveries—such as the diversity of closed related isolates and the identification of new putative virulence genes (Fraser, 2004). In this way, breakthrough transcriptomics and proteomics studies promise further exciting results and surprises over the coming years as already illustrated by profiling bacteria at the strain level (Sandrin et al., 2013), which hopefully, will have highly beneficial applications in terms of combating the scourge of global bioterrorism.

5.4 Environment and HostMicroorganisms Interactions

For host–pathogen interactions, the main assumption is that, over ecological time-scales, host susceptibility, and pathogen virulence are fixed at the onset of the crosstalk (Bull, 1994; Dieckmann et al., 2002). Also, environmental factors are traditionally viewed as "setting the scene" for the crosstalk rather than having any explicit role one it is underway. As a result, the effect of extrinsic factors on host

susceptibility and pathogen virulence during a crosstalk has received little attention. However, it is common to find in populations of a pathogen species a substantial variation in the virulence, even when pathogens are collected in the same environment and at the same time. When a character as the virulence is variable for both genetic and environmental reasons, two individuals may differ because they differ in genotype, because they have had different environmental experiences or both (Elliot et al., 2002). Unfortunately, the extent to which different individual pathogens and pathogen ecotypes display different virulence abilities is poorly documented and deciphered.

Life-history traits of hosts and pathogens are shaped by coevolution processes (Wolinska & King, 2009). Infections measured under laboratory conditions have shown that the environment in which hosts and pathogens interact may affect the range of host genotypes that can be infected with a given pathogen genotype in host–pathogen associations. Despite this important fact, environmental fluctuations are often excluded in surveys on host–pathogen interactions. Since most host–pathogen interactions are in heterogeneous environments, there is a crucial need to consider environmental conditions in proteomics surveys. The population proteomics and metaproteomics will be a promising prospect to resolve interesting issues specific to host–pathogen crosstalks in a varying environment (see Fig. 11.10). For instance, the ability to quickly phenotype microbial agents will help in understanding their adaption to environmental changes (Gouveia et al., 2020). This kind of survey would bring pioneer molecular data to understand why pathogens sometimes evolve in a given environment toward high virulence and hosts toward high resistance. Also, these surveys would permit to assess the fixity or not of host–parasite interactomes involved in a host–pathogen association in a varying environment.

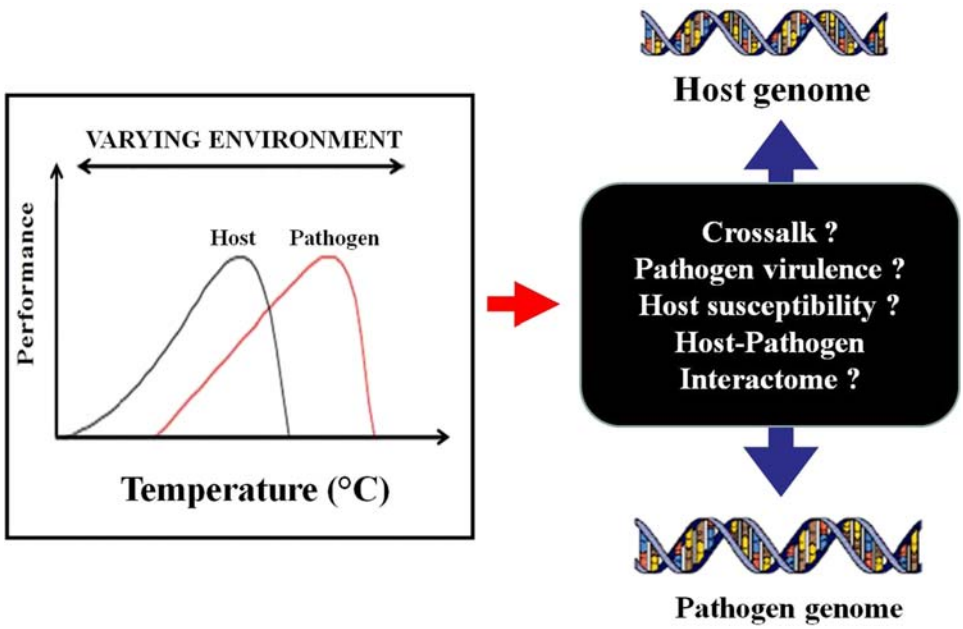


FIGURE 11.10

Host–pathogen interactions in a varying environment.

6. Conclusions

From the dawn of human evolution to the influenza, HIV/AIDS, and COVID-19 pandemics of the 20th and 21st centuries, infectious diseases have continued to emerge and re-emerge with great ferocity and by so doing, seriously affect populations as well as challenge our abilities to fight the responsible agents. Over the past decade, strains of many common pathogens have continued to develop resistance to the drugs that once were effective against them. In the battle against pathogens, humankind has created new mega-technologies such as massive sequencing, proteomics and bioinformatics, but without conceptual approaches based on the evolutionary concepts. Pathogen genome sequences do not themselves provide a full explanation of the biology of an organism and on the molecular war involved in host–pathogen associations. Since the 1990s, proteomic tools have been successfully employed in a large number of studies to find and identify proteins involved in biological phenomena, e.g., immunity, host–parasite interactions, etc. Even so, many studies have, as outlined above, revealed pitfalls in the approaches used. Thus, whatever the new technological advancements, it is apparent that biologists should attempt to improve their experimental design. This new attitude will surely improve the reliability of the data deriving from proteomics studies and will open the way for an enhanced comprehension of many biological mechanisms. In this chapter, new ways based on evolutionary concepts are suggested to enable further elucidation of the molecular complexities of host–pathogen genome interactions. These new ways could help to increase the knowledge about the molecular war involved in host–vector–pathogen interactions integrating the environmental factors.

Glossary

AIDS Acronym for acquired immune deficiency syndrome, the gravest of the sexually transmitted diseases, or STDs. It is caused by the human immunodeficiency virus (HIV), now known to be a retrovirus, an agent first identified in 1983. HIV is transmitted in body fluids, mainly blood and genital secretions.

Antigen Substances that are foreign to the body and cause the production of antibodies.

Bioinformatics The computational biology is the use of mathematical and informational techniques, including statistics, to solve biological problems, usually by creating or using computer programs, mathematical models, or both. One of the main areas of bioinformatics is the data mining and analysis of the data gathered by the various genome projects. Other areas are sequence alignment, protein structure prediction, systems biology, protein–protein interactions, and virtual evolution.

Defensin A substance with natural antibiotic effects found in human blood cells. There are three types of defensins. Other animal species have similar substances.

Elicitors Molecules produced by the host (or pathogen) that induce a response from the pathogen (or host).

Environment Elements of a habitat. In this text, "environment" refers to a broad range of biotic and abiotic conditions, including interactions with other species.

Genome The full complement of genes carried by a single (haploid) set of chromosomes. The term may be applied to the genetic information carried by an individual or to the range of genes found in a given species.

Genomics It is the study of an organism's genome and the use of the genes. It deals with the systematic use of genome information, associated with other data, to provide answers in biology, medicine, and industry.

Major histocompatibility complex Two classes of molecules on cell surfaces. MHC class I molecules exist on all cells and hold and present foreign antigens to CD8 cytotoxic T lymphocytes if the cell is infected by a virus or other microbe. MHC class II molecules are the billboards of the immune system. Peptides derived from foreign proteins are inserted into MHC's binding groove and displayed on the surface of antigen-presenting

cells. These peptides are then recognized by T lymphocytes so that the immune system is alerted to the presence of foreign material. See Histocompatibility Testing.

Mass spectrometry A technique for separating ions by their mass-to-charge (m/z) ratios.

Molecular crosstalk Molecular communications in a host–parasite system during their interaction.

Parasito(pathogeno)-proteomics The study of the reaction of the host and parasite genomes through the expression of the host and parasite (pathogen) proteomes (genome-operating systems) during their biochemical crosstalk.

Pathogenicity The capability of a pathogen to cause disease.

Population genetics Population genetics is the area of GENETICS that studies the distribution of genes (the units of genetic inheritance) and genotypes (the genetic complement at one or more loci), and the mechanisms determining genetic variability within a population.

Population proteomics The study of protein diversity in populations of any species; or more specifically, targeted investigation of proteins across and within populations of a species to define and understand protein diversity and to facilitate the discovery of disease-specific or pathogen-specific protein modulation.

Proteogenomics Field of biological research that combines genomics, transcriptomics, and proteomics data, to aid in the discovery and identification of set of peptides. Proteogenomics is used to identify new peptides assigning MS/MS spectra against a protein database that has been derived from genomic and transcriptomic information.

Proteome The term proteome was first used in 1995 and has been applied to several different types of biological systems. A cellular proteome is the collection of proteins found in a particular cell type under a particular set of environmental conditions such as exposure to hormone stimulation. It can also be useful to consider an organism's complete proteome. The complete proteome for an organism can be conceptualized as the complete set of proteins from all of the various cellular proteomes. This is very roughly the protein equivalent of the genome. The term "proteome" has also been used to refer to the collection of proteins in certain subcellular biological systems. For example, all of the proteins in a virus can be called a viral proteome.

Proteomics The large-scale study of proteins, particularly their structures and functions. This term was coined to make an analogy with genomics, and is often viewed as the "next step" but proteomics is much more complicated than genomics.

Reaction norm The set of phenotypes that can be produced by a genotype under various environmental settings.

Specificity An alternative concept to explain why hosts vary in their susceptibility to parasites is that host–parasite interactions have some degree of specificity.

Specificity of selection The range of host genotypes that can be infected with a given parasite genotype in host–parasite interactions.

Transcriptome Is the whole set of mRNA species in one or a population of cells.

Transcriptomics Techniques to identify mRNA from actively transcribed genes.

Two-dimensional gel electrophoresis Proteomics, the study of the proteome, has largely been practiced through the separation of proteins by two-dimensional gel electrophoresis. In the first dimension, proteins are separated by isoelectric focusing, resolved on the basis of charge. In the second dimension, they are separated by molecular weight using SDS-PAGE. To visualize the proteins, the gel is dyed with Coomassie Blue, silver, or other reagents. Spots on the gel are proteins that have migrated to specific locations.

Short Biographical Lines for Contributors

(1) David G. Biron (Ph. D.), LMGE, UMR 6023 CNRS/UCA, Aubière, France (david.biron1@univ-bpclermont.fr). D.G. Biron is a scientific leader with a well-established expertise in the field of host–parasite interactions, and especially on deciphering and understanding of host–parasite crosstalk by using "omics" tools.

(2) Dobrin Nedelkov (Ph. D.), Isoformix Inc., Phoenix, AZ, USA (dobrin.nedelkov@isoformix.com). D. Nedelkov is a pioneer and leader in human population proteomics.
(3) Dorothée Missé (Ph. D.), MIVEGEC, UMR 5290 CNRS/IRD/UM, Montpellier, France (dorothee.misse@ird.fr). D. Missé is a leader on the biochemical interactions in host–pathogen associations especially on the pathogen immune evasion to find diagnostic tools and therapeutic targets.
(4) Philippe Holzmuller (Ph. D.), ASTRE, UMR 117 CIRAD/INRAE/UM, Montpellier, France (philippe.holzmuller@cirad.fr). P. Holzmuller is a leader on the deciphering of crosstalk in host–pathogen associations especially on the pathogen immune evasion to find diagnostic tools and therapeutic targets/vaccine candidates.

References

Abad, P., Favery, B., Ross, M.N., Castagnone-Sereno, P., 2003. Root-knot nematode parasitism an host response: molecular basis of a sophisticated interaction. Mol. Plant Pathol. 4, 217–224.

Achtman, M., Morelli, G., Zhu, P., Wirth, T., Diehl, I., Kusecek, B., Vogler, A.J., Wagner, D.M., Allender, C.J., Easterday, W.R., Chenal-Francisque, V., Worsham, P., Thomson, N.R., Parkhill, J., Lindler, L.E., Carniel, E., Keim, P., 2004. Microevolution and history of the plague bacillus, Yersinia pestis. Proc. Natl. Acad. Sci. U. S. A. 101, 17837–17842.

Aggarwal, S., Raj, A., Kumar, D., Dash, D., Yadav, A.K., 2022. False discovery rate: the Achilles' heel of proteogenomics. Brief Bioinform. 23 (5), bbac163.

Agranoff, D., Stich, A., Abel, P., Krishna, S., 2005. Proteomic fingerprinting for the diagnosis of human African trypanosomiasis. Trends Parasitol. 21, 154–157.

Ahram, M., Petricoin, E.F., 2008. Proteomics discovery of disease biomarkers. Biomark. Insights 23, 325–333.

Ahrné, E., Müller, M., Lisacek, F., 2010. Unrestricted identification of modified proteins using MS/MS. Proteomics 10, 671–686.

Alfarano, C., Andrade, C.E., Anthony, K., Bahroos, N., Bajec, M., Bantoft, K., Betel, D., Bobechko, B., Boutilier, K., Burgess, E., Buzadzija, K., Cavero, R., D'Abreo, C., Donaldson, I., Dorairajoo, D., Dumontier, M.J., Dumontier, M.R., Earles, V., Farrall, R., Feldman, H., Garderman, E., Gong, Y., Gonzaga, R., Grytsan, V., Gryz, E., Gu, V., Haldorsen, E., Halupa, A., Haw, R., Hrvojic, A., Hurrell, L., Isserlin, R., Jack, F., Juma, F., Khan, A., Kon, T., Konopinsky, S., Le, V., Lee, E., Ling, S., Magidin, M., Moniakis, J., Montojo, J., Moore, S., Muskat, B., Ng, I., Paraiso, J.P., Parker, B., Pintilie, G., Pirone, R., Salama, J.J., Sgro, S., Shan, T., Shu, Y., Siew, J., Skinner, D., Snyder, K., Stasiuk, R., Strumpf, D., Tuekam, B., Tao, S., Wang, Z., White, M., Willis, R., Wolting, C., Wong, S., Wrong, A., Xin, C., Yao, R., Yates, B., Zhang, S., Zheng, K., Pawson, T., Ouellette, B.F., Hogue, C.W., 2005. The biomolecular interaction network database and related tools 2005 update. Nucleic Acids Res. 33, D418–D424.

Alshubaith, I.H., Alhajri, S., Alhajri, A., Alsultan, R.A., Azhar, E.I., Alhussaini, B.H., Al Solami, L.S., de Oliveira, M.C., Khafaga, A.F., Alqurashi, A.D., Attia, Y.A., September 6, 2022. The impact of COVID-19 on the sustainability of the environment, animal health and food security, and safety. Environ. Sci. Pollut. Res. Int. 1–10. https://doi.org/10.1007/s11356-022-22468-0.

Anderson, L., Seilhaver, J., 1997. A comparison of selected mRNA and protein abundance in human liver. Electrophoresis 18, 533–537.

Anderson, N.G., Anderson, N.L., 1996. Twenty years of two-dimensional electrophoresis: past, present and future. Electrophoresis 17, 443–453.

Antoine, J.C., Prina, E., Courret, N., Lang, T., 2004. *Leishmania* spp.: on the interactions they establish with antigen-presenting cells of their mammalian hosts. Adv. Parasitol. 58, 1–68.

Antrobus, R., Grant, K., Gangadharan, B., Chittenden, D., Everett, R.D., Zitzmann, N., Boutell, C., 2009. Proteomic analysis of cells in the early stages of herpes simplex virus type-1 infection reveals widespread changes in the host cell proteome. Proteomics 9, 3913−3927.

Aoki, V., Abdeladhim, M., Li, N., Cecilio, P., Prisayanh, P., Diaz, L.A., Valenzuela, J.G., 2022. Some good and some bad: sand fly salivary proteins in the control of leishmaniasis and in autoimmunity. Front. Cell. Infect. Microbiol. 12, 839932.

Armengaud, J., Christie-Oleza, J.A., Clair, G., Malard, V., Duport, C., 2012. Exoproteomics: exploring the world around biological systems. Expert Rev. Proteomics 9 (5), 561−575.

Armengaud, J., Trapp, J., Pible, O., Geffard, O., Chaumot, A., Hartmann, E.M., 2014. Non-model organisms, a species endangered by proteogenomics. J. Proteomics 105, 5−18.

Armengaud, J., 2009. A perfect genome annotation is within reach with the proteomics and genomics alliance. Curr. Opin. Microbiol. 12 (3), 292-300. doi: 10.1016/j.mib.2009.03.005. Epub 2009 May 4. PMID: 19410500.

Armengaud, J., 2010. Proteogenomics and systems biology: quest for the ultimate missing parts. Expert Rev. Proteomics 7 (1), 65−77.

Arnold, R., Boonen, K., Sun, M.G., Kim, P.M., 2012. Computational analysis of interactomes: current and future perspectives for bioinformatics approaches to model the host-pathogen interaction space. Methods 57 (4), 508−518.

Ashton, P.D., Curwen, R.S., Wilson, R.A., 2001. Linking proteome and genome: how to identify parasite proteins. Trends Parasitol. 17, 198−202.

Athanasiadou, S., Pemberton, A., Jackson, F., Inglis, N., Miller, H.R., Thévenod, F., Mackellar, A., Huntley, J.F., 2008. Proteomic approach to identify candidate effector molecules during the in vitro immune exclusion of infective *Teladorsagia circumcincta* in the abomasum of sheep. Vet. Res. 39, 58.

Azad, N.S., Rasool, N., Annunziata, C.M., Minasian, L., Whiteley, G., Kohn, E.C., 2006. Proteomics in clinical trials and practice: present uses and future promise. Mol. Cell. Proteomics 5, 1819−1829.

Bader, G.D., Hogue, C.W., 2003. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinf. 4, 2.

Bansal, A.K., 2005. Bioinformatics in microbial biotechnology-a mini review. Microb. Cell. Fact. 4, 19.

Barrett, J., Brophy, P.M., Hamilton, J.V., 2005. Analysing proteomic data. Int. J. Parasitol. 35, 543−553.

Barrett, J., Jefferies, J.R., Brophy, P.M., 2000. Parasite proteomics. Parasitol. Today 16, 400−403.

Baudot, A., Martin, D., Mouren, P., Chevenet, F., Guénoche, A., Jacq, B., Brun, C., 2006. PRODISTIN Web Site: a tool for the functional classification of proteins from interaction networks. Bioinformatics 22, 248−250.

Bernard, J., Hutet, E., Paboeuf, F., Randriamparany, T., Holzmuller, P., Lancelot, R., Rodrigues, V., Vial, L., Le Potier, M.F., 2016. Effect of *O. Porcinus* tick salivary gland extract on the african swine fever virus infection in domestic pig. PLoS One 11, e0147869.

Bhavsar, A.P., Auweter, S.D., Finlay, B.B., 2010. Proteomics as a probe of microbial pathogenesis and its molecular boundaries. Future Microbiol. 5, 253−265.

Bird, D.M., Opperman, C.H., 2009. The secret(ion) life of worms. Genome Biol. 10, 205.

Biron, D.G., Loxdale, H.D., 2013. Host-parasite molecular cross-talk during the manipulative process of a host by its parasite. J. Exp. Biol. 216 (Pt 1), 148−160.

Biron, D.G., Marché, L., Ponton, F., Loxdale, H.D., Galéotti, N., Renault, L., Joly, C., Thomas, F., 2005c. Behavioural manipulation in a grasshopper harbouring hairworm: a proteomics approach. Proc. R. Soc. Lond. B 272, 2117−2126.

Biron, D.G., Agnew, P., Marché, L., Renault, L., Sidobre, C., Michalakis, Y., 2005d. Proteome of *Aedes aegypti* larvae in response to infection by the intracellular parasite *Vavraia culicis*. Int. J. Parasitol. 35, 1385−1397.

Biron, D.G., Brun, C., Lefèvre, T., Lebarbenchon, C., Loxdale, H.D., Chevenet, F., Brizard, J.P., Thomas, F., 2006a. The pitfalls of proteomics experimentswithout the correct use of bioinformatics tools. Proteomics 6, 5577−5596.

Biron, D.G., Joly, C., Galéotti, N., Ponton, F., Marché, L., 2005b. The proteomics: a new prospect for studying parasitic manipulation. Behav. Process. 68, 249–253.

Biron, D.G., Loxdale, H.D., Ponton, F., Moura, H., Marché, L., Brugidou, C., Thomas, F., 2006b. Population proteomics: an emerging discipline to study metapopulation ecology. Proteomics 6, 1712–1715.

Biron, D.G., Moura, H., Marché, L., Hughes, A.L., Thomas, F., 2005a. Towards a new conceptual approach to Parasitoproteomics. Trends Parasitol. 21, 162–168.

Bischoff, R., Luider, T.M., 2004. Methodological advances in the discovery of protein and peptide disease markers. J. Chromatogr. B 803, 27–40.

Bogaerts, A., Beets, I., Temmerman, L., Schoofs, L., Verleyen, P., 2010. Proteome changes of *Caenorhabditis elegans* upon a *Staphylococcus aureus* infection. Biol. Direct 17 (5), 11.

Boonmee, S., Imtawil, K., Wongkham, C., Wongkham, S., 2003. Comparative proteomic analysis of juvenile and adult liver fluke, *Opisthorchis viverrini*. Act. Trop. 88, 233–238.

Brand, S., Hahner, D., Ketterlinus, R., 2005. Protein profiling and indentification in complex biological samples using LC-MALDI. Drug Plus Int. 6–8.

Brandt, I., Lambeir, A.M., Ketelslegers, J.M., Vanderheyden, M., Scharpe, S., De Meester, I., 2006. Dipeptidyl-peptidase IV converts intact B-type natriuretic peptide into its des-SerPro form. Clin. Chem. 52, 82–87.

Briant, L., Desprès, P., Choumet, V., Missé, D., 2014. Role of skin immune cells on the host susceptibility to mosquito-borne viruses. Virology 464–465, 26–32.

Bridges, D.J., Pitt, A.R., Hanrahan, O., Brennan, K., Voorheis, H.P., Herzyk, P., de Koning, H.P., Burchmore, R.J., 2008. Characterisation of the plasma membrane subproteome of bloodstream form *Trypanosoma brucei*. Proteomics 8, 83–99.

Broekaert, W.F., Terras, F.R.G., Cammue, B.P.A., Osborn, R.W., 1995. Plants defensins: novel antimicrobial peptides as components of the host defense system. Plant Physiol. 108, 1353–1358.

Brun, C., Herrmann, C., Guenoche, A., 2004. Clustering proteins from interaction networks for the prediction of cellular functions. BMC Bioinf. 5, 95.

Brun, C., Martin, D., Chevenet, F., Wojcik, J., Guénoche, A., Jacq, B., 2003. Functional classification of proteins for the prediction of cellular function from a protein-protein interaction network. Genome Biol. 5, R6.

Buhimschi, C.S., Bhandari, V., Hamar, B.D., Bahtiyar, M.O., Zhao, G., Sfakianaki, A.K., Pettker, C.M., Magloire, L., Funai, E., Norwitz, E.R., Paidas, M., Copel, J.A., Weiner, C.P., Lockwood, C.J., Buhimschi, I.A., 2007. Proteomic profiling of the amniotic fluid to detect inflammation, infection, and neonatal sepsis. PLoS Med 4, e18.

Bull, J.J., 1994. Perspective: virulence. Evolution 48, 1423–1437.

Cao, H., Baldini, R.L., Rahme, L.G., 2001. Common mechanism for pathogens of plant and animals. Annu. Rev. Phytopath. 39, 259–284.

Chautard, E., Thierry-Mieg, N., Ricard-Blum, S., 2009. Interaction networks: from protein functions to drug discovery. A review. Pathol. Biol. 57, 324–333.

Chetouhi, C., Panek, J., Bonhomme, L., ElAlaoui, H., Texier, C., Langin, T., de Bekker, C., Urbach, S., Demettre, E., Missé, D., Holzmuller, P., Hughes, D.P., Zanzoni, A., Brun, C., Biron, D.G., 2015. Cross-talk in host–parasite associations: what do past and recent proteomics approaches tell us? Infect. Genet. Evol. 33, 84–94.

Chevalier, F., Martin, O., Rofidal, V., Devauchelle, A.D., Barteau, S., Sommerer, N., Rossignol, M., 2004. Proteomic investigation of natural variation between *Arabidopsis* ecotypes. Proteomics 4, 1372–1381.

Chua, S.C.J.H., Cui, J., Engelberg, D., Lim, L.H.K., 2022. A review and meta-analysis of influenza interactome studies. Front. Microbiol. 13, 869406.

Cieslak, A., Ribera, I., 2009. Aplicaciones de proteómica en ecología y evolución. Ecosistemas 18, 34–43.

Cohen, A.M., Rumpel, K., Coombs, G.H., Wastling, J.M., 2002. Characterisationof global protein expression by two-dimensional electrophoresisa nd mass spectrometry: proteomics of *Toxoplasma gondii*. Int. J. Parasitol. 32, 39−51.

Coiras, M., Camafeita, E., López-Huertas, M.R., Calvo, E., López, J.A., Alcamí, J., 2008. Application of proteomics technology for analyzing the interactions between host cells and intracellular infectious agents. Proteomics 8, 852−873.

Connors, L.H., Lim, A., Prokaeva, T., Roskens, V.A., Costello, C.E., 2003. Tabulation of human transthyretin (TTR) variants, 2003. Amyloid 10, 160−184.

Cross, M., Klepzig, E., Dallaston, M., Young, N.D., Bailey, U.M., Mason, L., Jones, M.K., Gasser, R.B., Hofmann, A., 2016. Exploring the molecular mechanisms of parasite-host interactions with a view towards new therapeutics and vaccines. Postepy Biochem. 2 (3), 370−376.

Cui, M., Cheng, C., Zhang, L., August 3, 2022. High-throughput proteomics: a methodological mini-review. Lab. Invest. 1−12.

Das, S., Sierra, J.C., Soman, K.V., Suarez, G., Mohammad, A.A., Dang, T.A., Luxon, B.A., Reyes, V.E., 2005. Differential protein expression profiles of gastric epithelial cells following *Helicobacter pylori* infection using ProteinChips. J. Proteome Res. 4, 920−930.

de Araújo, C.N., Bussacos, A.C., Sousa, A.O., Hecht, M.M., Teixeira, A.R., 2012. Interactome: smart hematophagous triatomine salivary gland molecules counteract human hemostasis during meal acquisition. J Proteomics 75 (13), 3829−3841.

De Bock, M., de Seny, D., Meuwis, M.A., Chapelle, J.P., Louis, E., Malaise, M., Merville, M.P., Fillet, M., 2010. Challenges for biomarker discovery in body fluids using SELDI-TOF-MS. J. Biomed. Biotechnol. 2010, 906082.

Degrave, W.M., Melville, S., Ivens, A., Aslett, M., 2001. Parasite genome initiatives. Int. J. Parasitol. 31, 532−536.

Di Fazio, N., Morena, D., Delogu, G., Volonnino, G., Manetti, F., Padovano, M., Scopetti, M., Frati, P., Fineschi, V., July 30, 2022. Mental health consequences of COVID-19 pandemic period in the European population: an institutional challenge. Int. J. Environ. Res. Public. Health 19 (15), 9347. https://doi.org/10.3390/ijerph19159347.

Dieckmann, U., Metz, J.A.J., Sabelis, M.W., Sigmund, K. (Eds.), 2002. Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management. Cambridge University Press.

Diez-Orejas, R., Fernández-Arenas, E., 2008. *Candida albicans*-macrophage interactions: genomic and proteomic insights. Future Microbiol. 3, 661−681.

Dixon, M.S., Golstein, C., Thomas, C.M., Van Der Biezen, E.A., Jones, J.D., 2000. Genetic complexity of pathogen perception by plants: the example of Rcr3, a tomato gene required specifically by Cf-2. Proc. Nat. Acad. Sci. U. S. A. 97, 8807−8814.

Diz, A.P., Skibinski, D.O.F., 2007. Evolution of 2-DE protein patterns in a mussel hybrid zone. Proteomics 7, 2111−2120.

Dorus, S., Busby, S.A., Gerike, U., Shabanowitz, J., Hunt, D.F., Karr, T.L., 2006. Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. Nat. Genet. 38, 1440−1445.

Doucoure, S., Cornelie, S., Patramool, S., Mouchet, F., Demettre, E., Seveno, M., Dehecq, J.S., Rutee, H., Herve, J.P., Favier, F., Missé, D., Gasque, P., Remoue, F., 2013. First screening of *Aedes albopictus* immunogenic salivary proteins. Insect Mol. Biol. 22 (4), 411−423.

Dowell, J.A., Frost, D.C., Zhang, J., Li, L., 2008. Comparison of two-dimensional fractionation techniques for shotgun proteomics. Anal. Chem. 80, 6715−6723.

Doyle, E.A., Lambert, K.N., 2003. *Meloidogyne javanica* chorismate mutase 1 alters plant cell development. Mol. Plant Microbe Interact. 16, 123−131.

Doytchinova, I.A., Taylor, P., Flower, D.R., 2003. Proteomics in vaccinology and immunobiology: an informatics perspective of the immunone. J. Biomed. Biotechnol. 2003, 267–290.

Drake, R.R., Deng, Y., Schwegler, E.E., Gravenstein, S., 2005. Proteomics for biodefense applications: progress and opportunities. Expert Rev. Proteomics 2 (2), 203–213.

Dunn, R., Dudbridge, F., Sanderson, C.M., 2005. The use of edge-betweenness clustering to investigate biological function in protein interaction networks. BMC Bioinf. 6, 39.

Dupree, E.J., Crimmins, B.S., Holsen, T.M., Darie, C.C., 2019. Developing well-annotated species-specific protein databases using comparative proteogenomics. Adv. Exp. Med. Biol. 1140, 389–400.

Elliot, S.L., Blanford, S., Thomas, M.B., 2002. Host–pathogen interactions in a varying environment: temperature, behavioural fever and fitness. Proc. R. Soc. Lond. B 269, 1599–1607.

Fang, X., Balgley, B.M., Wang, W., Park, D.M., Lee, C.S., 2009. Comparison of multidimensional shotgun technologies targeting tissue proteomics. Electrophoresis 30, 4063–4070.

Fauci, A.S., Touchette, N.A., Folkers, G.K., 2005. Emerging infectious diseases: a 10- year perspective from the national institute of allergy and infectious diseases. Emerg. Infect. Dis. 11, 519–525.

Fell, D.A., 2001. Beyond genomics. Trends Genet. 17, 680–682.

Fenollar, F., Goncalves, A., Esterni, B., Azza, S., Habib, G., Borg, J.P., Raoult, D., 2006. A serum protein signature with high diagnostic value in bacterial endocarditis: results from a study based on surface-enhanced laser desorption/ionization time-of-flight mass spectrometry. J. Infect. Dis. 194, 1356–1366.

Fields, S., 2001. Proteomics in genomeland. Science 291, 1221–1224.

Fischer, J., West, J., Agochukwu, N., Suire, C., Hale-Donze, H., 2007. Induction of host chemotactic response by Encephalitozoon spp. Infect. Immun. 75, 1619–1625.

Flajolet, M., Rotondo, G., Daviet, L., Bergametti, F., Inchauspé, G., Tiollais, P., Transy, C., Legrain, P., 2000. A genomic approach of the hepatitis C virus generates a protein interaction map. Gene 242, 369–379.

Florens, L., Liu, X., Wang, Y., Yang, S., Schwartz, O., Peglar, M., Carucci, D.J., Yates 3rd, J.R., Wub, Y., 2004. Proteomics approach reveals novel proteins on the surface of malaria-infected erythrocytes. Mol. Biochem. Parasitol. 135, 1–11.

Fontaine, A., Diouf, I., Bakkali, N., Missé, D., Pagès, F., Fusai, T., Rogier, C., Almeras, L., 2011. Implication of haematophagous arthropod salivary proteins in host-vector interactions. Parasit. Vectors 4, 187.

Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C., Jacq, B., Arpin, M., Bellaiche, Y., Bellusci, S., Benaroch, P., Bornens, M., Chanet, R., Chavrier, P., Delattre, O., Doye, V., Fehon, R., Faye, G., Galli, T., Girault, J.A., Goud, B., de Gunzburg, J., Johannes, L., Junier, M.P., Mirouse, V., Mukherjee, A., Papadopoulo, D., Perez, F., Plessis, A., Rosse, C., Saule, S., Stoppa-Lyonnet, D.S., Vincent, A., White, M., Legrain, P., Wojcik, J., Camonis, J., Daviet, L., 2005. Protein interaction mapping: a Drosophila case study. Genome Res. 15, 376–384.

Fournier, I., Wisztorski, M., Salzet, M., 2008. Tissue imaging using MALDI-MS: a new frontier of histopathology proteomics. Expert Rev. Proteomics 5, 413–424.

Fraser, C.M., 2004. A genomics-based approach to biodefence preparedness. Nat. Rev. Genet. 5, 23–33.

Fujita, N., Sugimoto, R., Motonishi, S., Tomosugi, N., Tanaka, H., Takeo, M., Iwasa, M., Kobayashi, Y., Hayashi, H., Kaito, M., Takei, Y., 2008. Patients with chronic hepatitis C achieving a sustained virological response to peginterferon and ribavirin therapy recover from impaired hepcidin secretion. J. Hepatol. 49, 702–710.

Fung, E.T., Thulasiraman, V., Weinberger, S.R., Dalmasso, E.A., 2001. Protein biochips for differential profiling. Curr. Opin. Biotechnol. 12, 65–69.

Gelfand, J.A., Callahan, M.V., 2003. Babesiosis: an update on epidemiology and treatment. Curr. Infect. Dis. Rep. 5, 53–58.

Ghadie, M.A., Coulombe-Huntington, J., Xia, Y., 2018. Interactome evolution: insights from genome-wide analyses of protein-protein interactions. Curr. Opin. Struct. Biol. 50, 42–48.

Giot, L., Bader, J.S.,., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y.L., Ooi, C.E., Godwin, B., Vitols, E., Vijayadamodar, G., Pochart, P., Machineni, H., Welsh, M., Kong, Y., Zerhusen, B., Malcolm, R., Varrone, Z., Collis, A., Minto, M., Burgess, S., McDaniel, L., Stimpson, E., Spriggs, F., Williams, J., Neurath, K., Ioime, N., Agee, M., Voss, E., Furtak, K., Renzulli, R., Aanensen, N., Carrolla, S., Bickelhaupt, E., Lazovatsky, Y., DaSilva, A., Zhong, J., Stanyon, C.A., Finley, J.R.L., White, K.P., Braverman, M., Jarvie, T., Gold, S., Leach, M., Knight, J., Shimkets, R.A., McKenna, M.P., Chant, J., Rothberg, J.M., 2003. A protein interaction map of *Drosophila melanogaster*. Science 302, 1727−1736.

Girvan, M., Newman, M.E., 2002. Community structure in social and biological networks. Proc. Natl. Acad. Sci. U. S. A. 99, 7821−7826.

González-Fernández, R., Valero-Galván, J., Gómez-Gálvez, F.J., Jorrín-Novo, J.V., 2015. Unraveling the in vitro secretome of the phytopathogen Botrytis cinerea to understand the interaction with its hosts. Front. Plant Sci. 6, 839.

Gouveia, D., Grenga, L., Pible, O., Armengaud, J., 2020. Quick microbial molecular phenotyping by differential shotgun proteomics. Environ. Microbiol. 22 (8), 2996−3004.

Goverse, A., De Engler, J.A., Verhees, J., Van der Krol, S., Helder, J.H., Gheysen, G., 2000. Cell cycle activation by plant parasitic nematodes. Plant Mol. Biol. 43, 747−776.

Grébaut, P., Chuchana, P., Brizard, J.P., Demettre, E., Seveno, M., Bossard, G., Jouin, P., Vincendeau, P., Bengaly, Z., Boulangé, A., Cuny, G., Holzmuller, P., 2009. Identification of total and differentially expressed excreted-secreted proteins from *Trypanosoma congolense* strains exhibiting different virulence and pathogenicity. Int. J. Parasitol. 10, 1137−1150.

Gubler, D.J., 1998. Resurgent vector-borne diseases as a global health problem. Emerg. Infect. Dis. 4 (3), 442−450.

Guerrero, D., Cantaert, T., Missé, D., 2020. *Aedes* mosquito salivary components and their effect on the immune response to arboviruses. Front. Cell. Infect. Microbiol. 10, 407.

Guzman, M.G., Kouri, G., 2002. Dengue: an update. The Lancet Inf. Dis. 2, 33−42.

Gygi, S.P., Rochon, Y., Franza, B.R., Aebersold, R., 1999. Correlation between protein and mRNA abundance in yeast. Mol. Cell Biol. 19, 1720−1730.

Haab, B.B., Dunham, M.J., Brown, P.O., 2001. Protein microarrays for highly parallel detection and quantification of specific proteins and antibodies in complex solutions. Genome Biol. 2, Reserach004.

Han, Y.S., Chun, J., Schwartz, A., Nelson, S., Paskewitz, S.M., 1999. Induction of mosquito hemolymph proteins in response to immune challenge and wounding. Dev. Comp. Immunol. 23, 553−562.

Hart, S.R., Lau, K.W., Hao, Z., Broadhead, R., Portman, N., Hühmer, A., Gull, K., McKean, P.G., Hubbard, S.J., Gaskell, S.J., 2009. Analysis of the trypanosome flagellar proteome using a combined electron transfer/collisionally activated dissociation strategy. J. Am. Soc. Mass Spectrom. 20, 167−175.

Hermjakob, H., Montecchi-Palazzi, L., Bader, G., Wojcik, J., Salwinski, L., Ceol, A., Moore, S., Orchard, S., Sarkans, U., von Mering, C., Roechert, B., Poux, S., Jung, E., Mersch, H., Kersey, P., Lappe, M., Li, Y., Zeng, R., Rana, D., Nikolski, M., Husi, H., Brun, C., Shanker, K., Grant, S.G., Sander, C., Bork, P., Zhu, W., Pandey, A., Brazma, A., Jacq, B., Vidal, M., Sherman, D., Legrain, P., Cesareni, G., Xenarios, I., Eisenberg, D., Steipe, B., Hogue, C., Apweiler, R., 2004. The HUPO PSI's molecular interaction format–a community standard for the representation of protein interaction data. Nat. Biotechnol. 22, 177−183.

Heyer, R., Schallert, K., Zoun, R., Becher, B., Saake, G., Benndorf, D., 2017. Challenges and perspectives of metaproteomic data analysis. J. Biotechnol. 261, 24−36.

Hochstrasser, D.F., 1998. Proteome in perspective. Clin. Chem. Lab. Med. 36, 825−836.

Hoffman, J.A., 1995. Innate immunity of insects. Curr. Opin. Immunol. 7, 4−10.

Holzmuller, P., Grébaut, P., Semballa, S., Gonzatti, M.I., Geiger, A., 2013. Proteomics: a new way to improve human African trypanosomiasis diagnosis? Expert Rev. Proteomics 10 (3), 289−301.

Holzmuller, P., Nirdé, P., Vezilier, F., Chuchana, P., 2016. The transcriptomic analytical level determines the human monocyte-derived macrophage response toward either the infectious agent or the host. Infect. Genet. Evol. 45, 426–433.

Holzmuller, P., Grébaut, P., Brizard, J.P., Berthier, D., Chantal, I., Bossard, G., Bucheton, B., Vezilier, F., Chuchana, P., Bras-Gonçalves, R., Lemesre, J.L., Vincendeau, P., Cuny, G., Frutos, R., Biron, D.G., 2008. Pathogeno-proteomics: towards a new approach of host-vector-pathogen interactions. Ann. N.Y. Acad. Sci. 1149, 66–70.

Holzmuller, P., Grébaut, P., Cuny, G., Biron, D.G., 2010. Tsetse flies, trypanosomes, humans and animals: what is proteomics revealing about their crosstalks? Expert Rev. Proteomics 7, 113–126.

Hood, B.L., Malehorn, D.E., Conrads, T.P., Bigbee, W.L., 2009. Serum proteomics using mass spectrometry. Methods Mol. Biol. 520, 107–128.

Huang, S.H., Triche, T., Jong, A.Y., 2002. Infectomics: genomics and proteomics of microbial infections. Funct. Integr. Genomics 1, 331–344.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y., 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA 98, 4569–4574.

Jaffe, A.S., Van Eyk, J.E., 2006. Degradation of cardiac troponins. In: Morrow, D.A. (Ed.), Cardiovascular Biomarkers - Pathophysiology and Disease Management. Humana Press, Totowa, pp. 161–174.

Jagusztyn-Krynicka, E.K., Dadlez, M., Grabowska, A., Roszczenko, P., 2009. Proteomic technology in the design of new effective antibacterial vaccines. Expert Rev. Proteomics 6, 315–330.

Jahanafrooz, Z., Chen, Z., Bao, J., Li, H., Lipworth, L., Guo, X., 2022. An overview of human proteins and genes involved in SARS-CoV-2 infection. Gene 808, 145963.

Jahnukainen, T., Malehorn, D., Sun, M., Lyons-Weiler, J., Bigbee, W., Gupta, G., Shapiro, R., Randhawa, P.S., Pelikan, R., Hauskrecht, M., Vats, A., 2006. Proteomic analysis of urine in kidney transplant patients with BK virus nephropathy. J. Am. Soc. Nephrol. 17, 3248–3256.

Jasmer, D.P., 1993. *Trichinella spiralis* infected skeletal muscle cells arrest in G2/M and cease muscle gene expression. J. Cell Biol. 121, 785–793.

Ju, J.W., Joo, H.N., Lee, M.R., Cho, S.H., Cheun, H.I., Kim, J.Y., Lee, Y.H., Lee, K.J., Sohn, W.M., Kim, D.M., Kim, I.C., Park, B.C., Kim, T.S., 2009. Identification of a serodiagnostic antigen, legumain, by immunoproteomic analysis of excretory-secretory products of *Clonorchis sinensis* adult worms. Proteomics 9, 3066–3078.

Kanmura, S., Uto, H., Kusumoto, K., Ishida, Y., Hasuike, S., Nagata, K., Hayashi, K., Ido, A., Stuver, S.O., Tsubouchi, H., 2007. Early diagnostic potential for hepatocellular carcinoma using the SELDI ProteinChip system. Hepatology 45, 948–956.

Karr, T.L., 2008. Application of proteomics to ecology and population biology. Heredity 100, 200–206.

Kashiwagi, A.,K., urosaki, H., Luo, H., Yamamoto, H., Oshimura, M., Shibahara, T., 2009. Effects of *Tritrichomonas muris* on the mouse intestine: a proteomic analysis. Exp. Anim. 58, 537–542.

Kim, K.H., Moon, M.H., 2009. High speed two-dimensional protein separation without gel by isoelectric focusing-asymmetrical flow field flow fractionation: application to urinary proteome. J. Proteome Res. 8, 4272–4278.

Kint, G., Fierro, C., Marchal, K., Vanderleyden, J., De Keersmaecker, S.C., 2010. Integration of 'omics' data: does it lead to new insights into host-microbe interactions? Future Microbiol. 5, 313–328.

Klein, L.L., Freitag, B.C., Gibbs, R.S., Reddy, A.P., Nagalla, S.R., Gravett, M.G., 2005. Detection of intra-amniotic infection in a rabbit model by proteomics- based amniotic fluid analysis. Am. J. Obstet. Gynecol. 193, 1302–1306.

Kleppe, R., Kjarland, E., Selheim, F., 2006. Proteomic and computational methods in systems modeling of cellular signaling. Curr. Pharmaceut. Biotechnol. 7, 135–145.

Kuhn, J.F., Hoerth, P., Hoehn, S.T., Preckel, T., Tomer, K.B., 2006. Proteomics study of anthrax lethal toxin-treated murine macrophages. Electrophoresis 27, 1584−1597.

Kunath, B.J., Minniti, G., Skaugen, M., Hagen, L.H., Vaaje-Kolstad, G., Eijsink, V.G.H., Pope, P.B., Arntzen, M.Ø., 2019. Metaproteomics: sample preparation and methodological considerations. Adv. Exp. Med. Biol. 1073, 187−215.

LaCount, D.J., Vignali, M., Chettier, R., Phansalkar, A., Bell, R., Hesselberth, J.R., Schoenfeld, L.W., Ota, I., Sahasrabudhe, S., Kurschner, C., Fields, S., Hughes, R.E., 2005. A protein interaction network of the malaria parasite *Plasmodium falciparum*. Nature 438, 103−107.

Lal, K., Bromley, E., Oakes, R., Prieto, J.H., Sanderson, S.J., Kurian, D., Hunt, L., Yates, Wastling, J.M., Sinden, R.E., Tomley, F.M., 2009. Proteomic comparison of four *Eimeria tenella* life-cycle stages: unsporulated oocyst, sporulated oocyst, sporozoite and second-generation merozoite. Proteomics 9, 4566−4576.

Lam, C.S., Burnett Jr., J.C., Costello-Boerrigter, L., Rodeheffer, R.J., Redfield, M.M., 2007. Alternate circulating pro-B-type natriuretic peptide and B-type natriuretic peptide forms in the general population. J. Am. Coll. Cardiol. 49, 1193−1202.

Langley, R.C., Cali, A., Somberg, E.W., 1987. Two-dimensional electrophoretic analysis of spore proteins of the microsporida. J. Parasit. 73, 910−918.

Lee, S.R., Nanduri, B., Pharr, G.T., Stokes, J.V., Pinchuk, L.M., 2009. Bovine viral diarrhea virus infection affects the expression of proteins related to professional antigen presentation in bovine monocytes. Biochim. Biophys. Acta 1794, 14−22.

Lefèvre, T., Adamo, S.A., Biron, D.G., Missé, D., Hughes, D., Thomas, F., 2009. Invasion of the body snatchers: the diversity and evolution of manipulative strategies in host−parasite interactions. In: Webster, J.P., Rollinson, D., Hay, S.I. (Eds.), Advances in Parasitology, vol. 68. Academic Press, London, pp. 46−84.

Lefevre, T., Thomas, F., Ravel, S., Patrel, D., Renault, L., Le Bourligu, L., Cuny, G., Biron, D.G., 2007a. *Trypanosoma brucei brucei* induces alteration in the head proteome of the tsetse fly vector *Glossina palpalis gambiensis*. Insect Mol. Biol. 16, 651−660.

Lefevre, T., Thomas, F., Schwartz, A., Levashina, E., Blandin, S., Brizard, J.P., Le Bourligu, L., Demettre, E., Renaud, F., Biron, D.G., 2007b. Malaria *Plasmodium* agent induces alteration in the head proteome of their *Anopheles* mosquito host. Proteomics 7, 1908−1915.

Levin, M., Butter, F., 2022. Proteotranscriptomics - a facilitator in omics research. Comput. Struct. Biotechnol. J. 20, 3667−3675.

Levy, F., Bulet, P., Ehret-Sabatier, L., 2004. Proteomic analysis of the systemic immune response of Drosophila. Mol. Cell. Proteomics 3, 156−166.

Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., Goldberg, D.S., Li, N., Martinez, M., Rual, J.F., Lamesch, P., Xu, L., Tewari, M., Wong, S.L., Zhang, L.V., Berriz, G.F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H.W., Elewa, A., Baumgartner, B., Rose, D.J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S.E., Saxton, W.M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K.C., Harper, J.W., Cusick, M.E., Roth, F.P., Hill, D.E., Vidal, M., 2004. A map of the interactome network of the metazoan *C. elegans*. Science 303, 540−543.

Liu, H.C., Hicks, J., Yoo, D., 2008a. Proteomic dissection of viral pathogenesis. Dev. Biol. 132, 43−53.

Liu, N., Song, W., Wang, P., Lee, K., Chan, W., Chen, H., Cai, Z., 2008b. Proteomics analysis of differential expression of cellular proteins in response to avian H9N2 virus infection in human cells. Proteomics 8, 1851−1858.

Liu, Q., Chen, X., Hu, C., Zhang, R., Yue, J., Wu, G., Li, X., Wu, Y., Wen, F., 2010. Serum protein profiling of smear-positive and smear-negative pulmonary tuberculosis using SELDI-TOF mass spectrometry. Lung 188, 15−23.

Lopez, M.F., Pluskal, M.G., 2003. Protein micro- and macroarrays: digitizing the proteome. J. Chromatogr. B 787, 19–27.

Luciano-Montalvo, C., Ciborowski, P., Duan, F., Gendelman, H.E., Meléndez, L.M., 2008. Proteomic analyses associate cystatin B with restricted HIV-1 replication in placental macrophages. Placenta 29, 1016–1023.

Luo, X., Carlson, K.A., Wojna, V., Mayo, R., Biskup, T.M., Stoner, J., Anderson, J., Gendelman, H.E., Meléndez, L.M., 2003. Macrophage proteomic fingerprinting predicts HIV-1-associated cognitive impairment. Neurology 60, 1931–1937.

Luplertlop, N., Surasombatpattana, P., Patramool, S., Dumas, E., Wasinpiyamongkol, L., Saune, L., Hamel, R., Bernard, E., Sereno, D., Thomas, F., Piquemal, D., Ysse, H., Briant, L., Missé, D., 2011. Induction of a peptide with activity against a broad spectrum of pathogens in the *Aedes aegypti* salivary gland, following infection with Dengue virus. PLoS Pathog. 7 (1), e1001252.

Maccarrone, G., Turck, C.W., Martins-de-Souza, D., 2010. Shotgun mass spectrometry workflow combining IEF and LC-MALDI-TOF/TOF. Protein J. 29, 99–102.

Maffei, B., Francetic, O., Subtil, A., 2017. Tracking proteins secreted by bacteria: what's in the toolbox? Front. Cell. Infect. Microbiol. 7, 221.

Mangé, A., Chaurand, P., Perrochia, H., Roger, P., Caprioli, R.M., Solassol, J., 2009. Liquid chromatography-tandem and MALDI imaging mass spectrometry analyses of RCL2/CS100-fixed, paraffin-embedded tissues: proteomics evaluation of an alternate fixative for biomarker discovery. J. Proteome Res. 8, 5619–5628.

Maniatis, T., Tasic, B., 2002. Alternative pre-mRNA splicing and proteome expression in metazoans. Nature 418, 236–243.

Marcelino, I., Holzmuller, P., Coelho, A., Mazzucchelli, G., Fernandez, B., Vachiéry, N., 2021. Revisiting ehrlichia ruminantium replication cycle using proteomics: the host and the bacterium perspectives. Microorganisms 9 (6), 1144.

Mehta, S.M., Banerjee, S.M., Chowdhary, A.S., 2015. Postgenomics biomarkers for rabies—the next decade of proteomics. OMICS 19 (2), 67–79.

Meleady, P., 2018. Two-dimensional gel electrophoresis and 2D-DIGE. Methods Mol. Biol. 1664, 3–14.

Merchant, M., Weinberger, S.R., 2000. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis 21, 1164–1177.

Missé, D., Yssel, H., Trabattoni, D., Oblet, C., Lo Caputo, S., Mazzotta, F., Pène, J., Gonzalez, J.P., Clerici, M., Veas, F., 2007. IL-22 participates in an innate anti-HIV-1 host-resistance network through acute-phase protein induction. J. Immunol. 178, 407–415.

Molina, S., Missé, D., Roche, S., Badiou, S., Cristol, J.P., Bonfils, C., Dierick, J.F., Veas, F., Levayer, T., Bonnefont-Rousselot, D., Maurel, P., Coste, J., Fournier- Wirth, C., 2008. Identification of apolipoprotein C-III as a potential plasmatic biomarker associated with the resolution of hepatitis C virus infection. Proteonomics Clin. Appl. 2, 751–761.

Morrison, L.J., Marcello, L., McCulloch, R., 2009. Antigenic variation in the African trypanosome: molecular mechanisms and phenotypic complexity. Cell Microbiol. 11.

Moskalyk, L.A., Oo, M.M., Jacobs-Lorena, M., 1996. Peritrophic matrix proteins of *Anopheles gambiae* and *Aedes aegypti*. Ins. Mol. Biol. 5, 261–268.

Moura, H., Ospina, M., Woolfitt, A.R., Barr, J.R., Visvesvara, G.S., 2003. Analysis of four human microsporidian isolates by MALDI-TOF mass spectrometry. J. Eukaryot. Microbiol. 50, 156–163.

Moura, H., Visvesvara, G.S., 2001. A proteome approach to host-parasite interaction of the microsporidian *Encephalitozoom intestinalis*. J. Eukaryot. Microbiol. Suppl. 56S–59S.

Navas, A., Albar, J.P., 2004. Application of proteomics in phylogenetic and evolutionary studies. Proteomics 4, 299–302.

Nedelkov, D., 2005. Population proteomics: addressing protein diversity in humans. Expert Rev. Proteomics 2, 315–324.

Nedelkov, D., 2006. Mass spectrometry-based immunoassays for the next phase of clinical applications. Expert Rev. Proteomics 3, 631−640.

Nedelkov, D., 2008. Population proteomics: investigation of protein diversity in human populations. Proteomics 8, 779−786.

Nedelkov, D., 2017. Human proteoforms as new targets for clinical mass spectrometry protein tests. Expert Rev. Proteomics 14 (8), 691−699.

Nedelkov, D., Kiernan, U.A., Niederkofler, E.E., Tubbs, K.A., Nelson, R.W., 2005. Investigating human plasma proteins diversity. Proc. Natl. Acad. Sci. U. S. A. 102, 10852−10857.

Nedelkov, D., Kiernan, U.A., Niederkofler, E.E., Tubbs, K.A., Nelson, R.W., 2006. Population proteomics: the concept, attributes, and potential for cancer biomarker research. Mol. Cell Proteomics 5, 1811−1818.

Nedelkov, D., Phillips, D.A., Tubbs, K.A., Nelson, R.W., 2007. Investigation of human protein variants and their frequency in the general population. Mol. Cell. Proteomics 6, 1183−1187.

Nedelkov, D., Tubbs, K.A., Niederkofler, E.E., Kiernan, U.A., Nelson, R.W., 2004. High-throughput comprehensive analysis of human plasma proteins: a step toward population proteomics. Anal. Chem. 76, 1733−1737.

Nelson, M.M., Jones, A.R., Carmen, J.C., Sinai, A.P., Burchmore, R., Wastling, J.M., 2008. Modulation of the host cell proteome by the intracellular apicomplexan parasite *Toxoplasma gondii*. Infect. Immun. 76, 828−844.

Nelson, R.W., Krone, J.R., Bieber, A.L., Williams, P., 1995. Mass-spectrometric immunoassay. Anal. Chem. 67, 1153−1158.

Nicholson, K.R., Mousseau, C.B., Champion, M.M., Champion, P.A., 2021. The genetic proteome: using genetics to inform the proteome of mycobacterial pathogens. PLoS Pathog. 17 (1), e1009124.

Nie, L.B., Cong, W., He, J.J., Zheng, W.B., Zhu, X.Q., 2022. Global proteomic profiling of multiple organs of cats (*Felis catus*) and proteome-transcriptome correlation during acute *Toxoplasma gondii* infection. Infect. Dis. Poverty 11 (1), 96.

Nuttall, P.A., 2019. Wonders of tick saliva. Ticks Tick Borne Dis. 10 (2), 470−481.

Ostrowski, J., Wyrwicz, L.S., 2009. Integrating genomics, proteomics and bioinformatics in translational studies of molecular medicine. Expert Rev. Mol. Diagn 6, 623−630.

Ouellette, M., Olivier, M., Sato, S., Papadopoulou, B., 2003. Studies on the parasite *Leishmania* in the post-genomic era. Med. Sci. 19, 900−909.

Ouma, J.H., Vennervald, B.J., Butterworth, A.E., 2001. Morbidity in schistosomiasis: an update. Trends Parasitol. 17, 117−118.

Oura, C.A., McKellar, S., Swan, D.G., Okan, E., Shiels, B.R., 2006. Infection of bovine cells by the protozoan parasite *Theileria annulata* modulates expression of the ISGylation system. Cell Microbiol. 8, 276−288.

Pan, J., Chen, H.Q., Sun, Y.H., Zhang, J.H., Luo, X.Y., 2008. Comparative proteomic analysis of non-small-cell lung cancer and normal controls using serum label-free quantitative shotgun technology. Lung 186, 255−261.

Panek, J., El Alaoui, H., Mone, A., Urbach, S., Demettre, E., Texier, C., Brun, C., Pang, R.T., Poon, T.C., Chan, K.C., Lee, N.L., Chiu, R.W., Tong, Y.K., Wong, R.M., Chim, S.S., Ngai, S.M., Sung, J.J., Lo, Y.M., 2006. Serum proteomic fingerprints of adult patients with severe acute respiratory syndrome. Clin. Chem. 52, 421−429.

Panek, J., El Alaoui, H., Mone, A., Urbach, S., Demettre, E., Texier, C., Brun, C., Zanzoni, A., Peyretaillade, E., Parisot, N., Lerat, E., Peyret, P., Delbac, F., Biron, D.G., 2014. Hijacking of host cellular functions by an intracellular parasite, the microsporidian *Anncaliia algerae*. PLoS One 9, e100791.

Papadopoulos, M.C., Abel, P.M., Agranoff, D., Stich, A., Tarelli, E., Bell, B.A., Planche, T., Loosemore, A., Saadoun, S., Wilkins, P., Krishna, S., 2004. A novel and accurate diagnostic test for human African trypanosomiasis. Lancet 363, 1358−1363.

Park, A.J., Krieger, J.R., Khursigara, C.M., May 2016. Survival proteomes: the emerging proteotype of antimicrobial resistance. FEMS Microbiol. Rev. 40 (3), 323−342.

Park, S.K., Yates 3rd, J.R., 2010. Census for proteome quantification. Curr. Protoc. Bioinformatics 13.

Pastorino, B., Boucomont-Chapeaublanc, E., Peyrefitte, C.N., Belghazi, M., Fusaï, T., Rogier, C., Tolou, H.J., Almeras, L., 2009. Identification of cellular proteome modifications in response to West Nile virus infection. Mol. Cell. Proteomics 8, 1623–1637.

Patramool, S., Choumet, V., Surasombatpattana, P., Sabatier, L., Thomas, F., Thongrungkiat, S., Rabilloud, T., Boulanger, N., Biron, D.G., Missé, D., 2012. Update on the proteomics of major arthropod vectors of human and animal pathogens. Proteomics 12, 3510–3523.

Pedersen, K.S., Codrea, M.C., Vermeulen, C.J., Loeschcke, V., Bendixen, E., 2010. Proteomic characterization of a temperature-sensitive conditional lethal in *Drosophila melanogaster*. Heredity 104, 125–134.

Peltonen, L., McKusick, V.A., 2001. Genomics and medicine. Dissecting human disease in the postgenomic era. Science 291, 1224–1229.

Pendarvis, K., Padula, M.P., Tacchi, J.L., Petersen, A.C., Djordjevic, S.P., Burgess, S.C., Minion, F.C., 2014. Proteogenomic mapping of *Mycoplasma hyopneumoniae* virulent strain 232. BMC Genomics 15.

Pendyala, G., Trauger, S.A., Kalisiak, E., Ellis, R.J., Siuzdak, G., Fox, H.S., 2009. Cerebrospinal fluid proteomics reveals potential pathogenic changes in the brains of SIV-infected monkeys. J. Proteome Res. 5, 2253–2260.

Peri, S., Navarro, J.D., Amanchy, R., Kristiansen, T.Z., Jonnalagadda, C.K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T.K., Gronborg, M., Ibarrola, N., Deshpande, N., Shanker, K., Shivashankar, H.N., Rashmi, B.P., Ramya, M.A., Zhao, Z., Chandrika, K.N., Padma, N., Harsha, H.C., Yatish, A.J., Kavitha, M.P., Menezes, M., Choudhury, D.R., Suresh, S., Ghosh, N., Saravana, R., Chandran, S., Krishna, S., Joy, M., Anand, S.K., Madavan, V., Joseph, A., Wong, G.W., Schiemann, W.P., Constantinescu, S.N., Huang, L., Khosravi- Far, R., Steen, H., Tewari, M., Ghaffari, S., Blobe, G.C., Dang, C.V., Garcia, J.G., Pevsner, J., Jensen, O.N., Roepstorff, P., Deshpande, K.S., Chinnaiyan, A.M., Hamosh, A., Chakravarti, A., Pandey, A., 2003. Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res. 13, 2363–2371.

Peronnet, E., Becquart, L., Poirier, F., Cubizolles, M., Choquet-Kastylevsky, G., Jolivet-Reynaud, C., 2006. SELDI-TOF MS analysis of the Cardiac Troponin I forms present in plasma from patients with myocardial infarction. Proteomics 6, 6288–6299.

Perrin-Cocon, L., Diaz, O., Jacquemin, C., Barthel, V., Ogire, E., Ramière, C., André, P., Lotteau, V., Vidalain, P.O., 2020. The current landscape of coronavirus-host protein-protein interactions. J. Transl. Med. 18 (1), 319.

Pettersen, V.K., Antunes, L.C.M., Dufour, A., Arrieta, M.C., 2021. Inferring early-life host and microbiome functions by mass spectrometry-based metaproteomics and metabolomics. Comput. Struct. Biotechnol. J. 20, 274–286.

Petricoin, E.F., Liotta, L.A., 2004. SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. Curr. Opin. Biotechnol. 15, 24–30.

Pietzner, M., Wheeler, E., Carrasco-Zanini, J., Cortes, A., Koprulu, M., Wörheide, M.A., Oerton, E., Cook, J., Stewart, I.D., Kerrison, N.D., Luan, J., Raffler, J., Arnold, M., Arlt, W., O'Rahilly, S., Kastenmüller, G., Gamazon, E.R., Hingorani, A.D., Scott, R.A., Wareham, N.J., Langenberg, C., 2021. Mapping the proteo-genomic convergence of human diseases. Science 374 (6569), eabj1541.

Pitarch, A., Nombela, C., Gil, C., 2009. Proteomic profiling of serologic response to Candida albicans during host-commensal and host-pathogen interactions. Methods Mol. Biol. 470, 369–411.

Plattner, F., Soldati-Favre, D., 2008. Hijacking of host cellular functions by the Apicomplexa. Annu. Rev. Microbiol. 62, 471–487.

Poon, T.C., Hui, A.Y., Chan, H.L., Ang, I.L., Chow, S.M., Wong, N., Sung, J.J., 2005. Prediction of liver fibrosis and cirrhosis in chronic hepatitis B infection by serum proteomic fingerprinting: a pilot study. Clin. Chem. 51, 328–335.

Popara, M., Villar, M., de la Fuente, J., 2015. Proteomics characterization of tick- host-pathogen interactions. Methods Mol. Biol. 1247, 513−527.

Predel, R., Wegener, C., Russell, W.K., Tichy, S.E., Russell, D.H., Nachman, R.J., 2004. Peptidomics of CNS-associated neurohemal systems of adult *Drosophila melanogaster*: a mass spectrometric survey of peptides from individuals flies. J. Comp. Neurol. 474, 379−392.

Premsler, T., Zahedi, R.P., Lewandrowski, U., Sickmann, A., 2009. Recent advances in yeast organelle and membrane proteomics. Proteomics 9, 4731−4743.

Rachinsky, A., Guerrero, F.D., Scoles, G.A., 2007. Differential protein expression in ovaries of uninfected and Babesia-infected southern cattle ticks, *Rhipicephalus* (Boophilus) microplus. Insect Biochem. Mol. Biol. 37, 1291−1308.

Rain, J.C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schächter, V., Chemamam, Y., Labigne, A., Legrain, P., 2001. The protein-protein interaction map of *Helicobacter pylori*. Nature 409, 553.

Reeves, G.A., Talavera, D., Thornton, J.M., 2009. Genome and proteome annotation: organization, interpretation and integration. J. R. Soc. Interface 6 (31), 129−147.

Rives, A.W., Galitski, T., 2003. Modular organization of cellular networks. Proc. Natl. Acad. Sci. U. S. A. 100, 1128−1133.

Roberts, S.B., Robichaux, J.L., Chavali, A.K., Manque, P.A., Lee, V., Lara, A.M., Papin, J.A., Buck, G.A., 2009. Proteomic and network analysis characterize stage-specific metabolism in *Trypanosoma cruzi*. BMC Syst. Biol. 16 (3), 52.

Rock, F.L., Hardiman, G., Timans, J.C., Kastelein, R.A., Bazan, J.F., 1998. A family of human receptors structurally related to Drosophilla Toll. Proc. Natl. Acad. Sci. U. S. A. 95, 588−593.

Roitt, I.M., Delves, P.J., 2001. Essential Immunology. Blackwell Publishing, Inc., p. 496

Rual, J.F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G.F., Gibbons, F.D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D.S., Zhang, L.V., Wong, S.L., Franklin, G., Li, S., Albala, J.S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R.S., Vandenhaute, J., Zoghbi, H.Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M.E., Hill, D.E., Roth, F.P., Vidal, M., 2005. Towards a proteome-scale map of the human protein-protein interaction network. Nature 437, 1173−1178.

Ryan, E.T., 2001. Malaria: epidemiology, pathogenesis, diagnosis, prevention, and treatment-an update. Curr. Clin. Top. Infect. Dis. 21, 83−113.

Sakolvaree, Y., Maneewatch, S., Jiemsup, S., Klaysing, B., Tongtawe, P., Srimanote, P., Saengjaruk, P., Banyen, S., Tapchaisri, P., Chonsa-nguan, M., Chaicumpa, W., 2007. Proteome and immunome of pathogenic *Leptospira* spp. revealed by 2DE and 2DE-immunoblotting with immune serum. Asian Pac. J. Allergy Immunol. 25, 53−73.

Samanta, M.P., Liang, S., 2003. Predicting protein functions from redundancies in large-scale protein interaction networks. Proc. Natl. Acad. Sci. USA 100, 12579−12583.

Sánchez, M.I., Thomas, F., Perrot-Minnot, M.J., Biron, D.G., Bertrand-Michel, J., Missé, D., 2009. Neurological and physiological disorders in Artemia harboring manipulative cestodes. J. Parasitol. 95, 20−24.

Sandrin, T.R., Goldstein, J.E., Schumaker, S., 2013. MALDI TOF MS profiling of bacteria at the strain level: a review. Mass Spectrom. Rev. 32 (3), 188−217.

Saviola, A.J., Negrão, F., Yates 3rd, J.R., 2020. Proteomics of select neglected tropical diseases. Annu. Rev. Anal. Chem. 13 (1), 315−336.

Schleker, S., Sun, J., Raghavan, B., Srnec, M., Müller, N., Koepfinger, M., Murthy, L., Zhao, Z., Klein-Seetharaman, J., 2012. The current Salmonella-host interactome. Proteomics Clin. Appl. 6 (1−2), 117−133.

Schmid-Hempel, P., 2008. Parasite immune evasion: a momentous molecular war. Trends Ecol. Evol. 23, 318−326.

Schmid-Hempel, P., Ebert, D., 2003. On the evolutionary ecology of specific immune defence. Trends Ecol. Evol. 18, 27–32.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.

Shui, W., Gilmore, S.A., Sheu, L., Liu, J., Keasling, J.D., Bertozzi, C.R., 2009. Quantitative proteomic profiling of host-pathogen interactions: the macrophage response to *Mycobacterium tuberculosis* lipids. J. Proteome Res. 8, 282–289.

Sims, P.F., Hyde, J.E., 2006. Proteomics of the human malaria parasite *Plasmodium falciparum*. Expert Rev. Proteomics 3, 87–95.

Skelly, P.J., Da'dara, A.A., 2022. Schistosome secretomes. Acta. Trop. 236, 106676.

Spirin, V., Mirny, L.A., 2003. Protein complexes and functional modules in molecular networks. Proc. Natl. Acad. Sci.U S A. 100, 12123–12128.

Spivey, A., 2009. Amplify, amplify: shotgun proteomics boosts the signal for biomarker discovery. Environ. Health Perspect. 117, A206–A209.

Starr, A.E., Deeke, S.A., Li, L., Zhang, X., Daoud, R., Ryan, J., Ning, Z., Cheng, K., Nguyen, L.V.H., Abou-Samra, E., Lavallée-Adam, M., Figeys, D., 2018. Proteomic and metaproteomic approaches to understand host-microbe interactions. Anal. Chem. 90 (1), 86–109.

Steen, H., Mann, M., 2004. The ABC's (and XYZ's) of peptide sequencing. Nat. Rev. Mol. Cell Biol. 5, 699–711.

Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksöz, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., Wanker, E.E., 2005. A human protein-protein interaction network: a resource for annotating the proteome. Cell 122, 957–968.

Steuart, R.F., 2010. Proteomic analysis of Giardia: studies from the pre- and post- genomic era. Exp. Parasitol. 124, 26–30.

Stiles, J.K., Whittaker, J., Sarfo, B.Y., Thompson, W.E., Powell, M.D., Bond, V.C., 2004. Trypanosome apoptotic factor mediates apoptosis in human brain vascular endothelial cells. Mol. Biochem. Parasitol. 133, 229–240.

Sun, B., Rempel, H.C., Pulliam, L., 2004. Loss of macrophage-secreted lysozyme in HIV-1-associated dementia detected by SELDI-TOF mass spectrometry. AIDS 18, 1009–1012.

Sun, J., Jiang, Y., Shi, Z., Yan, Y., Guo, H., He, F., Tu, C., 2008. Proteomic alteration of PK-15 cells after infection by classical swine fever virus. J. Proteome. Res. 7, 5263–5269.

Swanson, S.K., Washburn, M.P., 2005. The continuing evolution of shotgun proteomics. Drug Discov. Today 10, 719–725.

Tanca, A., Deligios, M., Addis, M.F., Uzzau, S., 2013. High throughput genomic and proteomic technologies in the fight against infectious diseases. J. Infect. Dev. Ctries. 7 (3), 182–190.

Tastet, C., Bossis, M., Gauthier, J.P., Renault, L., Mugniéry, D., 1999. *Meloidogyne chitwoodi* and *M. fallax* protein variation assessed by two-dimensional electrophoregram computed analysis. Nematology 1, 301.

Taylor, J.E., Hatcher, P.E., Paul, N.D., 2003. Crosstalk between plant responses to pathogens and herbivores: a view from the outside in. J. Exp. Bot. 55, 159–168.

te Pas, M.F., Claes, F., 2004. Functional genomics and proteomics for infectious diseases in the post-genomics era. Lancet 363, 1337.

Teixeira, P.C., Iwai, L.K., Kuramoto, A.C., Honorato, R., Fiorelli, A., Stolf, N., Kalil, J., Cunha-Neto, E., 2006. Proteomic inventory of myocardial proteins from patients with chronic Chagas' cardiomyopathy. Braz. J. Med. Biol. Res. 39, 1549–1562.

Texier, C., Brosson, D., El Alaoui, H., Méténier, G., Vivarès, C.P., 2005. Post- genomics of microsporidia, with emphasis on a model of minimal eukaryotic proteome: a review. Folia. Parasitol. 52, 15–22.

Thiel, M., Bruchhaus, I., 2001. Comparative proteome analysis of *Leishmania donovani* at different stages of transformation from promastigotes to amastigotes. Med. Microbiol. Immunol. 190, 33−36.

Thiellement, H., Bahrman, N., Damerval, C., Plomion, C., Rossignol, M., Santoni, V., de Vienne, D., Zivy, M., 1999. Proteomics for genetic and physiological studies in plants. Electrophoresis 20, 2013−2026.

Tomazella, G.G., Risberg, K., Mylvaganam, H., Lindemann, P.C., Thiede, B., Souza, G.A., Wiker, H.G., 2012. Proteomic analysis of a multi-resistant clinical *Escherichia coli* isolate of unknown genomic background. J. Proteomics 75 (6), 1830−1837.

Toro-Nieves, D.M., Rodriguez, Y., Plaud, M., Ciborowski, P., Duan, F., Pérez Laspiur, J., Wojna, V., Meléndez, L.M., 2009. Proteomic analyses of monocyte-derived macrophages infected with human immunodeficiency virus type 1 primary isolates from Hispanic women with and without cognitive impairment. J. Neurovirol. 15, 36−50.

Tournu, H., Serneels, J., Van Dijck, P., 2005. Fungal pathogens research: novel and improved molecular approaches for the discovery of antifungal drug targets. Curr. Drug Targets 6, 909−922.

Trenchevska, O., Nelson, R.W., Nedelkov, D., 2016. Mass spectrometric immunoassays for discovery, screening and quantification of clinically relevant proteoforms. Bioanalysis 8 (15), 1623−1633.

Tushir, S., Kamanna, S., Nath, S.S., Bhat, A., Rose, S., Aithal, A.R., Tatu, U., 2021. Proteo-genomic analysis of SARS-CoV-2: a clinical landscape of single-nucleotide polymorphisms, COVID-19 proteome, and host responses. J. Proteome Res. 20 (3), 1591−1601.

Uetz, P., Dong, Y.A., Zeretzke, C., Atzler, C., Baiker, A., Berger, B., Rajagopala, S.V., Roupelieva, M., Rose, D., Fossum, E., Haas, J., 2006. Herpesviral protein networks and their interaction with the human proteome. Science 311, 239−242.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., Rothberg, J.M., 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 403, 623−627.

Uzoma, I., Zhu, H., 2013. Interactome mapping: using protein microarray technology to reconstruct diverse protein networks. Dev. Reprod. Biol. 11 (1), 18−28.

Vaezzadeh, A.R., Briscoe, A.C., Steen, H., Lee, R.S., 2010. One-step sample concentration, purification, and albumin depletion method for urinary proteomics. J. Proteome Res. 9, 6082−6089.

Valcu, C.M., Lalanne, C., Muller-Starck, G., Plomion, C., Schlink, K., 2008. Protein polymorphism between 2 *Picea abies* populations revealed by 2-dimensional gel electrophoresis and tandem mass spectrometry. Heredity 99, 364−375.

Valenzuela, J.G., 2004. Exploring tick saliva: from biochemistry to 'sialomes' and functional genomics. Parasitology 129 (Suppl. l), S83−S94.

Van Hellemond, J.J., van Balkom, B.W., Tielens, A.G., 2007. Schistosome biology and proteomics: progress and challenges. Exp. Parasitol. 117, 267−274.

Vester, D., Rapp, E., Gade, D., Genzel, Y., Reichl, U., 2009. Quantitative analysis of cellular proteome alterations in human influenza A virus-infected mammalian cell lines. Proteomics 9, 3316−3327.

Vialás, V., Nogales-Cadenas, R., Nombela, C., Pascual-Montano, A., Gil, C., 2009. Proteopathogen a protein database for studying *Candida albicans*–host interaction. Proteomics 9, 4664−4668.

Vidal-Veuthey, B., González, D., Cárdenas, J.P., 2022. Role of microbial secreted proteins in gut microbiota-host interactions. Front Cell Infect. Microbiol. 12, 964710.

Vierstraete, E., Verleyen, P., Baggerman, G., D'Hertog, W., Van den Bergh, G., Arckens, L., De Loof, A., Schoofs, L., 2004. A proteomic approach for the analysis of instantly released wound and immune proteins in *Drosophila melanogaster* hemolymph. Proc. Natl. Acad. Sci. U. S. A. 101, 470−475.

Villar, M., Popara, M., Mangold, A.J., de la Fuente, J., 2014. Comparative proteomics for the characterization of the most relevant Amblyomma tick species as vectors of zoonotic pathogens worldwide. J. Proteomics 105, 204−216.

Viswanathan, K., Früh, K., 2007. Viral proteomics: global evaluation of viruses and their interaction with the host. Expert Rev. Proteomics 4, 815–829.

Walduck, A., Rudel, T., Meyer, T.F., 2004. Proteomic and gene profiling approaches approaches to study host responses to bacterial infection. Curr. Opin. Microbiol. 7 (1), 33–38.

Walters, M.S., Mobley, H.L., 2009. Identification of uropathogenic *Escherichia coli* surface proteins by shotgun proteomics. J. Microbiol. Methods. 78, 131–135.

Wang, Y., Zhou, Y., Xiao, X., Zheng, J., Zhou, H., 2020. Metaproteomics: a strategy to study the taxonomy and functionality of the gut microbiota. J. Proteomics 219, 103737.

Watanabe, T., Kawaoka, Y., 2015. Influenza virus-host interactomes as a basis for antiviral drug development. Curr. Opin. Virol. 14, 71–78.

Wattam, A.R., Christensen, B.M., 1992. Induced polypetides associated with filarial worm refractoriness in *Aedes aegypti*. Proc. Natl. Acad. Sci. U. S. A. 89, 6502–6505.

Watts, S., 1997. Epidemics and History: Disease, Power and Imperialism. Yale University Press, New Haven, CT.

Wiederin, J., Rozek, W., Duan, F., Ciborowski, P., 2008. Biomarkers of HIV-1 associated dementia: proteomic investigation of sera. Proteome Sci. 17, 7–8.

Weiss, L.M., Fiser, A., Angeletti, R.H., Kim, K., 2009. *Toxoplasma gondii* proteomics. Expert Rev. Proteomics 6, 303–313.

WHO, 2016. Global Health Observatory (GHO) Data. https://www.who.int/data/gho.

Wilm, M., 2009. Quantitative proteomics in biological research. Proteomics 9, 4590–4605.

Wolinska, J., King, K.C., 2009. Environment can alter selection in host-parasite interactions. Trends Parasitol. 25, 236–244.

Wu, M.S., Chow, L.P., Lin, J.T., Chiou, S.H., 2008. Proteomic identification of biomarkers related to Helicobacter pylori-associated gastroduodenal disease: challenges and opportunities. J. Gastroenterol. Hepatol. 23 (11), 1657–1661.

Wu, C., Wang, Z., Liu, L., Zhao, P., Wang, W., Yao, D., Shi, B., Lu, J., Liao, P., Yang, Y., Zhu, L., 2009. Surface enhanced laser desorption/ionization profiling: new diagnostic method of HBV-related hepatocellular carcinoma. J. Gastroenterol. Hepatol. 24, 55–62.

Wu, C.C., MacCoss, M.J., Howell, K.E., Yates III, J.R., 2003. A method for the comprehensive proteomics analysis of membrane proteins. Nat. Biotechnol. 21, 532–538.

Wu, Y., Nelson, M.M., Quaile, A., Xia, D., Wastling, J.M., Craig, A., 2009. Identification of phosphorylated proteins in erythrocytes infected by the human malaria parasite *Plasmodium falciparum*. Malar. J. 8, 105.

Xiao, Z., Prieto, D., Conrads, T.P., Veenstra, T.D., Issaq, H.J., 2005. Proteomic patterns: their potential for disease diagnosis. Mol. Cell. Endocrinol. 230, 95–106.

Yang, S.Y., Xiao, X.Y., Zhang, W.G., Zhang, L.J., Zhang, W., Zhou, B., Chen, G., He, D.C., 2005. Application of serum SELDI proteomic patterns in diagnosis of lung cancer. BMC Cancer 5, 83.

Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M., Cesareni, G., 2002. MINT: a Molecular INTeraction database. FEBS Lett. 513, 135–140.

Zhang, C.G., Chromy, B.A., McCutchen-Maloney, S.L., 2005. Host-pathogen interactions: a proteomic view. Expert Rev. Proteomics 2, 187.

Zhang, L., Jia, X., Zhang, X., Sun, J., Peng, X., Qi, T., Ma, F., Yin, L., Yao, Y., Qiu, C., Lu, H., 2010a. Proteomic analysis of PBMCs: characterization of potential HIV-associated proteins. Proteome Sci. 8, 12.

Zhang, X., Zhou, J., Wu, Y., Zheng, X., Ma, G., Wang, Z., Jin, Y., He, J., Yan, Y., 2009b. Differential proteome analysis of host cells infected with porcine circovirus type 2. J. Proteome Res. 8, 5111–5119.

Zhang, Y., Wen, Z., Washburn, M.P., Florens, L., 2010b. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. Anal. Chem. 82, 2272–2281.

Zheng, X., Hong, L., Shi, L., Guo, J., Sun, Z., Zhou, J., 2008. Proteomics analysis of host cells infected with infectious bursal disease virus. Mol. Cell. Proteomics 7, 612–625.

CHAPTER

Genomic Epidemiology and Surveillance of Antimicrobial Resistance

12

Neris García-González[1], Irving Cancino-Muñoz[1], Leonor Sánchez-Busó[1,2] and Fernando González-Candelas[1,2]

[1]*Joint Research Unit "Infection and Public Health" FISABIO-University of Valencia, Institute for Integrative Systems Biology (I2SysBio, UV-CSIC), Valencia, Spain;* [2]*CIBER in Epidemiology and Public Health, Valencia, Spain*

1. Antimicrobial Resistance: A Current Major Threat

Antimicrobial resistance (AMR) is the ability of certain microorganisms, including bacteria, viruses, fungi, and parasites, to survive and grow in the presence of antimicrobial compounds that once killed them or stopped their growth.

AMR is one of the leading global public health threats of the 21st century (Antimicrobial Resistance Collaborators, 2022) and needs to be addressed urgently. The impact of AMR in the clinical setting and for public health is tremendous in terms of therapeutic failures, deaths, and associated healthcare costs. AMR is particularly important in bacteria (Prestinaci et al., 2015), but resistance is also found in viruses, fungi, and protozoa. When sensitive bacteria become resistant to an effective and established treatment, clinicians need to find new effective therapies. Treatment failures and the time it takes to find an effective one increase the morbidity and mortality of patients (Davies and Crombie, 1997). In the last decades, the emergence of infections by strains that have acquired resistance to a wide range of antibiotics, or multidrug-resistant (MDR) infections, has threatened treatment with last-line antibiotics, leading to a lack of effective treatments.

Bacteria possess a great genetic plasticity that allows them to adapt and compete for essential resources in different ecological niches. As a result of their adaptability, many microorganisms have developed resistance mechanisms to evade molecules that threaten their survival. Some bacterial species are intrinsically resistant to a broad range of antibiotics. The presence of bacteria highly resistant to three to four different classes of antibiotics has been demonstrated in ecosystems free from exposure to anthropogenic antimicrobials such as deep caves (Bhullar et al., 2012) and permafrost sediments of ancient DNA (D'Costa et al., 2011), supporting the idea that antimicrobial resistance is natural and ancient. Among the drugs tested, the list includes natural, broad-spectrum antibiotics used for a century, such as some β-lactam agents, but also novel, synthetic compounds, such as linezolid.

There are two major ways for AMR acquisition by bacteria. The first way is represented by de novo mutations in genes involved in any of the antibiotic resistance mechanisms (Munita and Arias, 2016; Darby et al., 2023). The second mechanism includes the acquisition of entirely new genes or fragments

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00030-6

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

of genes determining resistance by Horizontal Gene Transfer (HGT). HGT, mostly driven by plasmids, plays a crucial role in the development and dissemination of antimicrobial resistance among clinically relevant bacteria (Martinez and Baquero, 2000).

A global survey estimated that in 2019, resistant bacterial infections caused 1.27 million deaths and that 4.95 million deaths were associated with it (Antimicrobial Resistance Collaborators, 2022; Cassini et al., 2018, 2019). This survey found that the burden of bacterial AMR is comparable to the combined burden of influenza, tuberculosis, and AIDS.

In 2017, the WHO published a list of antibiotic-resistant "priority pathogens" in need of research and development of new antibiotics, thus posing the greatest AMR threat to human health (World Health Organization, 2017). This list is divided into three categories based on urgency and need for new antibiotics. The most critical group includes carbapenem-resistant bacteria such as *Acinetobacter*, *Pseudomonas*, and various Enterobacteriales (including *Klebsiella*, *Escherichia coli, Serratia*, and *Proteus*). The high and medium priority categories contain increasingly drug-resistant bacteria that cause more common diseases such as gonorrhea and salmonellosis. This list was modified and expanded by the CDC in 2019, when the list of most urgent antibiotic resistance threats in the United States was published (Centers for Disease Control and Prevention, 2019).

In the absence of an effective treatment to combat these priority pathogens, there is an urgent need for effective strategies to control the spread of multidrug-resistant organisms (MDROs). Many efforts have been made to develop surveillance systems to track changes in the prevalence of resistant infections, identify known and new resistance mechanisms, and monitor the spread of resistant pathogens (World Health Organization, 2020).

In this chapter, we describe in detail how the use of sequencing has improved resistance surveillance globally and locally. In addition, we will discuss how resistance is predicted from sequencing data. Finally, we will describe the resistance mechanisms developed by three pathogens according to their lifestyle and level of genomic complexity, from an adapted intracellular microorganism to an opportunistic commensal pathogen.

2. Genomic Epidemiology and Surveillance in the Fight Against AMR

AMR surveillance is essential to monitor and control the spread of resistance as it can help to establish with precision the occurrence and distribution of AMR at different geographic scales, pathogens, and sources. This information is essential for informing policies and interventions and evaluating the effectiveness of mitigation strategies, especially in infection prevention and control, public health actions, and treatment decisions.

Before the advent of generalized genome sequencing, epidemiological tracking of pathogens and AMR was performed by supplementing epidemiology data with a diverse suite of phenotypic and molecular typing tools. Many of these methods are still "gold standard" techniques because they work efficiently and require short-time and easy-to-follow protocols. However, while much has been learned about the epidemiology of MDROs using these methods, they have major limitations. For example, to identify an outbreak of carbapenemase-producing Enterobacteriaceae, several techniques need to be performed to identify the pathogen, the carbapenemase-determining gene, and the relationship between the isolates. With the traditional, marker-based techniques, this would be time-consuming and expensive, and still may provide insufficient resolution for epidemiological tracing and phenotype prediction.

The introduction of High-Throughput Sequencing (HTS) technologies made sequencing of complete genomes (Whole Genome Sequencing, WGS) at clinically relevant timescales and costs feasible (Croucher and Didelot, 2015; Loman et al., 2012; Reuter et al., 2013). In contrast to traditional "gold standard" techniques, for which studies are targeted at assessing specific, and usually small, genomic regions, WGS allows the sequencing of the entire genome. Therefore, WGS provides information on pathogen species, genetic determinants of AMR, as well as a greatly improved subtyping resolution for epidemiological research, and all these pieces of information can be retrieved simultaneously. For instance, in the epidemiological field, HTS-based methods have been demonstrated to be a highly discriminatory tool for identifying and studying human pathogens outbreaks. Some examples include *Legionella* (Sánchez-Busó et al., 2014), *Mycobacterium tuberculosis* (Cancino-Muñoz et al., 2022), and Methicillin-resistant *Staphylococcus aureus* (Harris et al., 2013).

However, one of the major advances of HTS and WGS technologies is that protocols are reproducible and easily applicable to most pathogens. In recent years, in line with its advantages and the increasingly reduced costs, WGS is being incorporated into routine public health practice as part of surveillance and control programs of AMR pathogens, especially in high-income countries. The potential uses of WGS in public health include: (a) the identification of resistant clones at different geographic scales, (b) the identification of genetic mechanisms of AMR, how they emerge and are transmitted between humans and animals through the environment ("One Health" approach), (c) outbreak identification and source tracing, (d) the identification of novel targets for vaccines and antimicrobials, as well as (e) the development of point-of-care tests to detect AMR (World Health Organization, 2020).

HTS technologies can also be applied to complex samples, not only to bacterial isolates, an approach known as metagenomics. However, further standardization of protocols, databases, and analysis tools are still needed to ensure reproducibility. Furthermore, the linkage of an AMR gene with a specific pathogen is not always straightforward, especially when the gene is located on a mobile element, and the identification of new AMR determinant genes is not possible.

3. Genotypic Prediction of Antimicrobial Resistance

Since the introduction of HTS technologies, the diversity of sequencing platforms and methodologies has exploded. HTS technologies vary greatly in the methodology, their accuracy, and the quantity and quality of data output. However, all current sequencing technologies have in common three main steps: sample preparation, which includes the generation of "libraries," the sequencing step, and the post-processing of the generated sequences, including the bioinformatic analyses. For each step, there are many different possibilities, and the choice between them depends on many factors, such as sequencing objective, input sample, cost, time, access to sequencing equipment, etc.

Although the current HTS platforms differ in their methodologies, all share some general steps. During the library preparation step, the genomic DNA of the isolate is fragmented and tagged so the sequencer can process it and return "reads," the inferred DNA sequences of each fragment. In general, sequencing technologies can be classified in two major categories regarding the length of the output reads: short-read and long-read sequencing. HTS is currently dominated by short-read-based technologies, with Illumina as the most widely used platform. Short-read sequencing process begins by generating millions of short sequences by fragmentation of the genomic sequence. Those fragments

will be later sequenced producing "reads" with lengths between 150 and 300 base pairs. Short-read platforms are very accurate and provide an effective tool for identifying AMR genes. However, as the genome is transformed into millions of short fragments, they are not suitable for chromosome or plasmid reconstruction, elucidating the synteny or order of genetic features within the genome or identifying if an AMR gene is present in a plasmid or a chromosome. On the other hand, long-read sequencing, currently dominated by Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) platforms, aims to sequence very long DNA fragments (average of 30 Kbp reads) (Amarasinghe et al., 2020). Long-read sequencing can overcome short-read assembly problems, as they sequence very large regions of the genome obtaining very good results in the structure and arrangement of chromosomes and plasmids. Here, we have to distinguish between the accurate results of both different platforms. ONT technologies have a high error rate—although this has reduced substantially lately, complicating the detection of variants/alleles, and usually have to be combined with other sequencing platforms to obtain good results, commonly Illumina. On the contrary, PacBio produces highly accurate long reads; however, it is expensive and more time-consuming.

The chosen platform is very relevant for AMR surveillance and will depend greatly on our objective. For example, if we want to know which are the most prevalent resistance genes in a certain region or environment, a short-read project would be sufficient. However, if the question is about the genetic context of these genes (plasmid, chromosome, linked to other resistance genes), we should opt for a long-read sequencing project. Cost is also a factor to be considered and it greatly influences decisions. In general, Illumina has the lowest costs per isolate, primarily because it has higher throughput, which allows sequencing hundreds of bacterial genomes in a single experiment thus reducing the sequencing cost per isolate. As of today, long-read technologies are restricted to sequence only a few tens of isolates in a batch.

Depending on the type of platform and, therefore, the methodology selected, we will have to perform different bioinformatics approaches because the tools used for short and long reads are not the same. Moreover, we have to consider that the bioinformatics workflow will also be affected by the genetic characteristics of the sequenced organism, the time, and the goals of the analysis. In AMR surveillance using HTS, one of the first questions we need to ask is whether we are doing surveillance or we want to decipher an outbreak. If we are looking for fast detection of AMR we could use the sequencing results directly from the platforms (raw reads); however, if the aim is AMR outbreak detection or plasmid reconstruction, we will need to perform more complex analyses.

3.1 Methods Based on Raw Reads

Raw data, that is the reads obtained directly from the sequencer, can be processed directly after a short quality evaluation and filtering step. Using raw data obviates the need for postprocessing steps such as assembly (further detailed below) and, therefore, it reduces the time needed to obtain the results. However, for some applications raw data cannot be as accurate as needed.

For short-read raw data, SRST2 (Inouye et al., 2014) is one of the most common bioinformatics tools. It performs a fast and accurate detection of genes and variants together with subtyping of microorganisms from HTS data. It is a mapping-based tool that predicts resistance determinants by aligning reads to a reference database. Similar software has been developed in the last years such as ResFinder (Zankari et al., 2012).

Mapping raw reads to reference databases has some limitations, being the most important the insertion of errors in the alignment by spurious mapping matches or sequencing errors. However, this is usually solved by the high throughput of short read methodologies. Other limitations derive from the completeness of the databases and the impossibility of identifying new AMR determinants not included therein yet.

There is no similar methodology for the long-read platforms. Remarkably and interestingly for diagnostics, ONT allows the processing of raw data in real time directly from the sequencer (Gardy and Loman, 2018). Real-time analysis allows the rapid detection of antimicrobial resistance genes, but not further subtyping of gene-variants (Taxt et al., 2020). Thus, ONT allows detecting rapidly an antimicrobial gene but not its variants. For PacBio, real-time surveillance is not a possibility. Nevertheless, accurate reads are obtained so information about the variants can be gained in the posterior phases.

3.2 Assembly-based Methods

Most AMR detection programs take assembled data as input, either from long or short reads. The assembly steps rely on overlapping short reads into longer contiguous sequences called contigs. Available methods often rely on interrogating those contigs using sequence search algorithms to detect genes of interest. As in the short-read AMR detection, sequence search tools use previously curated databases. The most popular software to detect AMR from assemblies are ResFinder (Zankari et al., 2012), ARIBA (Hunt et al., 2017), CARD RGI (Jia et al., 2017), ABRicate (Seemann, 2022), and NCBI AMRFinderPlus (Feldgarden et al., 2019), among others.

Although AMR detection from assemblies is fast, the process of de novo assembly and annotation is computationally expensive and time-consuming. The computational requirements for AMR detection from raw reads are low, whereas for assemblies the requirements are larger and more specialized resources are necessary.

However, for some methodologies and objectives, the assembly step is absolutely necessary. De novo assembly of nanopore reads can correct sequencing errors and correctly identify AMR genes and alleles. Therefore, using assembled data increases the precision of ONT sequencing substantially.

3.3 Databases

AMR databases gather the sequence of known AMR determinants. Generally speaking, there are generalized and specialized databases that can be coupled with or without phenotype information. Each type differs in the scope of species, resistance mechanisms, and the associated information included.

Generalized databases include information on a wide range of species and/or AMR mechanisms and, therefore, their use allows the identification of AMR determinants in species that can acquire new AMR by horizontal transfer. Two general databases are widely cited in the literature: the Comprehensive Antibiotic Resistance Database (CARD) (Mcarthur et al., 2013) and Resfinder 4.0 (Bortolaia et al., 2020). CARD collects antimicrobial resistance genes, proteins, and phenotypes for all types of antimicrobial classes for at least 377 pathogens (accessed December 2022). Resfinder was developed to include acquired resistance genes (Zankari et al., 2012). Nevertheless, it was updated and now also includes point mutations in chromosomal genes (Zankari et al., 2017), phenotypic data, and allows for

in silico antibiograms. However, it is a more restricted database, containing information for Enterobacteriales and a few other species such as *N. gonorrhoeae* and *M. tuberculosis*.

In contrast, specialized databases contain information on specific species, antimicrobial families, or groups of resistant mechanisms. Species-specific databases are a better choice if the species of interest is not affected by horizontal transfer and the AMR mechanisms are mainly chromosomal point mutations. This is the case, for instance, of *M. tuberculosis* for which specialized databases have been developed such as the Tuberculosis Drug Resistance Database (Sandgren et al., 2009) or MUBII-TB-DB (Flandrois et al., 2014).

3.4 New AMR Determinants

Several methods and tools have been published in recent years for detecting genetic determinants of antimicrobial resistance from whole genome sequencing. Knowing AMR determinants and their translation into phenotype opens the possibility for predicting antimicrobial resistance from WGS data. Recent studies have shown that in silico antimicrobial susceptibility testing (AST) of MDROs using WGS can be highly concordant to the observed phenotypic resistance (Monk, 2019). Nevertheless, different levels of accuracy have been found depending on the pathogen and antibiotic class tested (Guo et al., 2019; Nguyen et al., 2018, 2019). However, we are still unable to predict AMR exclusively from WGS data, because some mechanisms are complex and often involve epistatic interactions and the creation of mosaic genes through homologous recombination. As for today, phenotypic methods are still needed (World Health Organization, 2020).

For some bacterial species and/or antibiotics, AMR determinants and resistance mechanisms remain unknown. The currently available tools mentioned before may detect new AMR alleles, yet they are not capable of detecting new AMR genes or determinants. To pursue the identification of novel AMR determinants, kmer-based and machine-learning strategies have been employed. However, these approaches still require extensive, previous phenotypic and genotypic characterization of a high number of curated genomes to produce accurate results (Eyre et al., 2017; McDermott et al., 2016; Nguyen et al., 2018) .

4. Case Studies: Genomics and AMR in Clinically Relevant Bacteria

4.1 *Mycobacterium tuberculosis*

Tuberculosis (TB) has historically been one of the leading causes of death attributed to infectious diseases worldwide, even in the last decade. Unfortunately, the current COVID-19 pandemic has exacerbated this situation. It is considered that there has been a decline in reported cases worldwide, especially in low- and middle-income countries (LMIC), which are the most affected (World Health Organization, 2021). The increasing prevalence of cases not responding to standard antibiotic therapy is one of the main threats to global TB control efforts. During 2020, it was estimated that 150,903 new cases were caused by drug-resistant TB (World Health Organization, 2021). The treatment of these resistant cases is usually very expensive and time consuming (up to 24 months, depending on the antibiotic options available). Additionally, the side effects developed during drug management are often harmful to the patient (Fitzpatrick and Floyd, 2012; Pontali et al., 2017; Seddon et al., 2013).

Tuberculosis is caused by members of the *Mycobacterium tuberculosis* complex (MTBC). These intracellular well-adapted bacteria have a clonal population structure, with minimal or no genetic transfer between strains (Gagneux, 2018). MTBC has been shown to have low levels of recombination and horizontal transfer (HGT) (Boritsch et al., 2016; Supply et al., 2003). As a consequence of this genomic background and the physiological characteristics of the bacterium (the cell wall is formed by mycolic acids that make it more hydrophobic) (Gygli et al., 2017), the selection of resistance mutations at the intraindividual level is mainly due to the concentration of antibiotics that penetrate this cell wall (Andersson and Hughes, 2014). Hence, specific point mutations (such as SNPs, deletions, and insertions) have been identified in genes directly involved in antibiotic resistance. For instance, around 95% of all mutations conferring rifampicin resistance are located in an 81 bp region of the *rpoB* gene (Telenti et al., 1993). Nevertheless, less common mutations outside this region have been described elsewhere (Van Deun et al., 2015). Other factors related to resistance are noncompletion or low adherence to antibiotic treatment in some patients.

4.1.1 Drug Susceptibility Prediction

The standard therapy for susceptible tuberculosis consists in the administration of four different antibiotics (rifampicin, isoniazid, pyrazinamide, and ethambutol) during 6 months, and its clinical success is approximately 85% of all newly diagnosed TB cases (Tiberi et al., 2018). However, the increasing prevalence of cases that fail to respond to this first-line treatment is a main threat to global TB control efforts. Multidrug-resistant TB (MDR-TB) are those infections that are resistant to rifampicin and isoniazid simultaneously, the most powerful antibiotics against TB. The second-line drug regimen for these MDR cases includes bedaquiline, fluoroquinolones, ethionamide, clofazimine, in combination with effective first-line drugs for another 6–9 months (Programme, 2020).

The diagnosis and prediction of drug resistance (DR) in MTBC is based on the presence or absence of specific mutations in particular loci. Regarding this, complete well-curated mutation catalogs are needed. Over the last few years, high-confidence DR mutations have been described by using statistically significant genotype–phenotype associations derived by sequencing thousands of clinical isolates (Walker et al., 2022). Through initiatives and collaborations, international consortia have been formed with the main purpose of replacing standard phenotypic tests by a faster genomic diagnosis of DR, especially those related with first-line treatment (Genome-wide Association Studies of Global Mycobacterium Tuberculosis Resistance to 13 Antimicrobials in 10,228 Genomes Identify New Resistance Mechanisms, 2022). To this end, DR catalogs are enriched with clinical and genomic information from large-scale datasets. Recent meta-analyses have reported high sensitivity and specificity values for rifampicin (0.98 [95% CI 0.93–0.98], and 0.98 [95% CI 0.98–1.00] and isoniazid (0.97 [95% CI 0.94–0.99], and 0.93 [95% CI 0.91–0.96]) DR prediction (Coll et al., 2015; CRyPTIC Consortium and the 100, 2018). Nevertheless, these agreed values are lower depending on the antibiotic tested, particularly for second-line drugs or those recently introduced into the therapy (e.g., bedaquiline and linezolid). This discrepancy may be related to standardization issues in phenotypic-based assays or to the presence of unidentified mutations (Papaventsis et al., 2017).

Related to the latter, many studies have described novel mutations related to multidrug-resistant cases (Cancino-Muñoz et al., 2019; López et al., 2020; Torres et al., 2015). Interestingly, these clinical isolates present compensatory mutations in almost all the cases, suggesting a strong association between these resistance cases. Remarkably, such resistance mutations have to be corroborated and

confirmed by phenotypic assays, so it is necessary to standardize or improve routine phenotypic techniques.

4.1.2 The Role of Compensatory Mutations

Although drug resistance usually implies a cost for bacterial fitness through the modification or loss of some essential functions related to replication and survival in the absence of antibiotics, there are some mutations that are capable of compensating for these costs even without the selective pressure caused by antibiotics (Comas et al., 2011). These mutations may emerge in resistance-related genes or in other genomic regions that interact with these targets (e.g., promoters, plasmids, or even chromosomes) (Szamecz et al., 2014). Because MTBC strains are monoclonal (lacking horizontal gene transfer), with very low recombination rate, and that the first-line treatment is globally standard, these compensatory mutations are relatively simple to screen at the population level.

Experimental in vitro studies in TB have demonstrated that these compensatory variants do not confer resistance by themselves (Gagneux et al., 2006; Heym et al., 1997). However, some of them are involved in the evolution of drug resistance to first-line drugs in susceptible strains. Compensatory mutations related to rifampicin and isoniazid, the most effective anti-TB drugs, have been described (Merker et al., 2018; Sherman et al., 1996; Comas et al., 2011). The mechanism of action of rifampicin is to inhibit bacterial transcription by targeting RNA polymerase β subunit that is encoding by the *rpoB* gene (Goldstein, 2014). Thus, mutations detected in other polymerase subunits (encoded by *rpoA,* and *rpoC* genes) are responsible for compensating the fitness cost derived from *rpoB* variants (Comas et al., 2011). Between 83% and 96% of isoniazid-related mutations occur in the catalase-peroxidase *katG* gene (Seifert et al., 2015). However, other isoniazid-resistant strains harbor mutations in other promoter and coding genomic regions of *inhA, kasA* and *ahpC*, all of which are related to mycolic acid production and play a compensatory role for the loss of catalase-peroxidase activity (Sherman et al., 1996).

At the epidemiological level, compensatory mutations are often associated to the successful dissemination of large MDR outbreaks in high-burden TB settings such as South Africa (de Vos et al., 2013), Uzbekistan (Merker et al., 2018), and India (Dreyer et al., 2022), but not in low burden countries. This suggests that the interaction between compensatory mutations and the corresponding main resistance mutations may be influenced by the mutation rates, population sizes, and transmission of some MDR strains.

With this knowledge, it is essential to add these compensatory mutations to the catalogs of mutations associated with resistance to first-line drugs. Nevertheless, in the case of resistance to second-line treatment, it is not clear whether compensatory mutations are present in outbreaks or transmission chains. Recently, there have been studies based on functional genomics capable of predicting the evolution of strains resistant to these antibiotics that may be involved in MDR outbreaks (Furió et al., 2021; Papakonstantinou et al., 2021).

4.2 *Neisseria gonorrhoeae*

Over 350 million cases of sexually transmitted infections (STIs) are reported every year worldwide, and the most recent estimates indicate an incidence of 87 million cases of gonorrhea per year, the second highest among STIs after chlamydia infections(127 million cases) (Rowley et al., 2019; Rowley et al., 2019). The causative agent of gonorrhea is *Neisseria gonorrhoeae* and its global

incidence is higher in men (15−41 cases per 1000 population) but also considerably high in women (14−28 cases per 1000 population). Asymptomatic gonorrhea infections can become very problematic and they are common in women (in approximately 50% of cases) although they also appear in around 10% of men. Undetected or not properly treated gonorrhea infections can lead to severe complications especially in women, such as endometritis, pelvic inflammatory disease, ectopic pregnancy, and premature labor or miscarriage. It can also cause penile edema or epididymitis in men and ultimately septicemia in both genders (Unemo and Shafer, 2014). Conjunctivitis infections can also occur in adults but also in newborns (*ophtalmia neonatorum*), as they can get infected through the birth canal and result in blindness. Fortunately, cases of disseminated gonococcal infection are rare (Unemo and Shafer, 2014).

N. gonorrhoeae was categorized by the WHO in 2017 as a pathogen of high priority in the list of priority pathogens in need for research and development of new antibiotics (World Health Organization, 2017). This list was validated and expanded by the CDC in 2019, where *N. gonorrhoeae* was classified as one of the most urgent antibiotic resistance threats in the United States (Centers for Disease Control and Prevention, 2019). This is because it has successively acquired or evolved resistance to every antimicrobial used to treat gonorrhea, from the sulfonamides, which were introduced in 1930s as treatment, to penicillins, tetracyclines, spectinomycin, ciprofloxacin, and more recently to the extended-spectrum cephalosporins (ESCs) cefixime (oral) and ceftriaxone (injectable) and the macrolide azithromycin in the last years of the 20th century. Treatment with ESCs was one of the latest options. However, the rise in cefixime resistance led to a change in treatment recommendations in the EU/EEA in 2012 toward a dual therapy with ceftriaxone and azithromycin (Unemo et al., 2021; Wi et al., 2017). Since then, cases of decreased susceptibility to ceftriaxone or resistance to azithromycin have been reported worldwide (Unemo et al., 2021; Cole et al., 2017). These include recently described sporadic cases of ceftriaxone resistance without or with limited transmission, the international spread of the ceftriaxone-resistant strain FC428 (Lahra et al., 2018), the sustained transmission of high-level azithromycin-resistant gonorrhea infections, particularly in England (Fifer et al., 2018), and also the European expansion of an azithromycin-resistant lineage (Sánchez-Busó et al., 2022). The rise in azithromycin resistance has triggered a change in treatment recommendations in some countries such as the United Kingdom and the United States, which now advocate for a treatment with ceftriaxone monotherapy (Fifer et al., 2020; St Cyr et al., 2021). Fortunately, cases of resistance to the dual therapy are not common despite having been sporadically reported (Fifer et al., 2016; Pleininger et al., 2022; Whiley et al., 2018) and resistance to cephalosporins is decreasing significantly (European Centre for Disease Prevention and Control, 2022).

4.2.1 Genetic Mechanisms of Antimicrobial Resistance

Extensive investigations have been ongoing for years to unveil the genetic mechanisms that explain most of the observed patterns of antimicrobial resistance for the main classes of antimicrobials for *N. gonorrhoeae*: penicillins, tetracyclines, fluoroquinolones, macrolides, and ESCs. However, we still cannot explain all the observed cases of phenotypic resistance (Mortimer and Grad, 2019; Unemo et al., 2019). In fact, values of >87% sensitivity and >98% specificity in AMR genotype−phenotype congruence have been described for different pathogen−antimicrobial combinations, while predictive values of resistance of 100% are exceptional (Hendriksen et al., 2019). In the case of the gonococcus, for ciprofloxacin (a fluoroquinolone) nearly all of the resistance can be explained by the presence of one amino acid alteration, GyrA S91F (Lindbäck et al., 2002). However, this is not that straightforward

for other antibiotics. Known resistance mechanisms often involve additive or suppressive effects as well as epistatic interactions that jointly explain only part of this resistance (Sánchez-Busó et al., 2021). For example, there is good evidence that many mosaic structures of the penicillin-binding protein 2 (*penA*) gene (Ohnishi et al., 2011) are associated with a decreased susceptibility to ESCs. However, mosaics do not explain all the cases of ESC resistance, especially for ceftriaxone (Grad et al., 2016). On top of these, variants that overexpress the MtrCDE efflux pump, mutations in the porin-encoding *porB1b* gene that reduce drug influx, and non-mosaic mutations in penicillin-binding proteins also contribute to decreased susceptibility to ESCs (Lindberg et al., 2007). Mutations in the 23S rRNA gene (A2045G and C2597T in *N. gonorrhoeae* nomenclature, A2059G and C2611T in *Escherichia coli*) are frequently associated with azithromycin resistance, as do variants in the MtrCDE repressor-encoding *mtrR* gene or its promoter, which increase the expression of this pump (Zalucki et al., 2012). However, epistatic interactions between a mosaic *mtrR* promoter region and a mosaic *mtrD* gene have also been reported recently to increase the expression of this pump, contributing to macrolide resistance (Wadsworth et al., 2018).

Genetic mechanisms of resistance in the form of point mutations, insertions, deletions, or recombination events can be gathered into open access databases and genome analysis suites such as Pathogenwatch (Sánchez-Busó et al., 2021) or PubMLST (Jolley et al., 2018) and queried with high-throughput sequencing data obtained from a particular isolate, in order to infer its susceptibility or resistance to a panel of antibiotics. These genetic mechanisms can be used for AMR prediction but also to inform about the development of new treatment options and point-of-care tests of antimicrobial susceptibility. More sophisticated algorithms can also be used for phenotypic prediction of resistance from WGS data, such as regression-based or machine learning approaches (Demczuk et al., 2020; Hicks et al., 2019; Yasir et al., 2022). However, they can be only as good as the databases they rely on. This is why phenotypic antimicrobial susceptibility testing (AST) is still encouraged (World Health Organization, 2020), as it allows detecting cases of resistance caused by novel or still unknown mechanisms.

4.2.2 *Genomic Epidemiology of* N. gonorrhoeae

Several studies have used high-throughput sequencing approaches for the retrospective analysis of *N. gonorrhoeae* strain collections. These have provided a significant improvement in resolution over traditional molecular epidemiology and typing methods (Harris et al., 2018) and have provided important insight into the evolution of AMR in the gonococcus and also on its transmission, dispersion, and population structure and dynamics (Sánchez-Busó and Harris, 2019). Interestingly, a recent study on disseminated gonococcal infections revealed that the population structure of isolates causing these rare complications of the disease are similar to those causing cases of uncomplicated gonorrhea (Cartee et al., 2022). Without the new insights provided by genomics and with the lack of completeness in contact tracing, the epidemiological tracking of *N. gonorrhoeae* constituted a real challenge before WGS.

The use of WGS is giving further insight into how *N. gonorrhoeae* spreads in different sexual networks and has identified particular lineages associated to different risk populations and antimicrobial susceptibilities. A study on the analysis of a global gonococcal collection (Sánchez-Busó et al., 2019) revealed two major lineages likely representing two epidemiological strategies used by this pathogen. A major lineage accumulating significantly more genetic mechanisms of resistance spreads in high-risk populations such as men who have sex with men (MSM), which are often treated with

antimicrobials for gonorrhea infection and/or other STIs so the strains in this lineage are under the constant pressure of antimicrobials. These high-risk populations often perform particular sexual practices (e.g., oral sex during "sex and drug" parties (McCall et al., 2015) and are more prone to pharyngeal infections). In the pharynx, *N. gonorrhoeae* can recombine with other *Neisseria*, resulting in mosaic alleles that cause resistance or reduced susceptibility to key antibiotics such as ESCs or the macrolide azithromycin (Mortimer and Grad, 2019; Wadsworth et al., 2018). This study hypothesizes that the antimicrobial pressure plus its higher ability to acquire resistance mechanisms through homologous recombination and its higher opportunity to exchange DNA in coinfections, helps this lineage to evolve and acquire AMR. The second major lineage accumulates significantly less genetic mechanisms of resistance and spreads mostly in heterosexual groups and particular populations where rates of asymptomatic infections are high, e.g., women. These sublineages are less frequently under the pressure of antimicrobials and have lower recombination rates, maintaining their susceptibility while spreading unnoticed.

Interestingly, the genomic data from the European Gonococcal Antimicrobial Surveillance Program (Euro-GASP) 2013 survey (Harris et al., 2018) accordingly revealed a susceptible lineage strongly associated with heterosexual groups. The two major clades associated with different antimicrobial susceptibilities were also found in a gonococcal collection from Denmark from 1928−2013 (Golparian et al., 2020). In another dataset from the Gonococcal Resistance to Antimicrobials Surveillance Program (GRASP) from England from 2013−16, the results also revealed a lineage strongly associated with MSM and decreased susceptibility to antimicrobials, while heterosexual men and women were found more frequently infected by susceptible lineages (Town et al., 2020). A collection of isolates from 2011−15 collected in New York City (NYC) revealed the existence of the same two major clades, although they found extensive bridging between MSM and heterosexual populations (Mortimer et al., 2021). Very interestingly, they observed that the largest transmission clusters circulating in NYC were susceptible to the ESCs and azithromycin, proving the importance of studying both antimicrobial resistant and susceptible lineages, as they all have the mechanisms to become resistant.

Surveillance of these gonococcal lineages through genomics is imperative and already a reality. The European Gonococcal Antimicrobial Surveillance Program (Euro-GASP) monitors AMR in the EU/EEA by analyzing a combination of quality-assured AMR, epidemiological and clinical data of gonorrhea patients (Cole et al., 2019). Euro-GASP has produced three molecular surveys: one, in 2009−10, analyzed 1066 isolates from 21 EU/EEA countries using the *N. gonorrhoeae* multi-antigen sequence typing (NG-MAST) scheme (Chisholm et al., 2013), and two, in 2013 and 2018, that analyzed 1054 isolates from 20 EU/EEA countries (Harris et al., 2018) and 2375 isolates from 27 EU/EEA countries, respectively, using WGS (Sánchez-Busó et al., 2022). Comparative results among the three surveys have provided interesting insights into the dynamics of existing and new lineages. For example, the first molecular survey showed cefixime-resistant NG-MAST genogroup G1407 (mostly *N. gonorrhoeae* sequence typing for antimicrobial resistance (NG-STAR) clonal complex (CC) 90) as one major group and was found to be associated with MSM populations. In 2012, cefixime was replaced by ceftriaxone and azithromycin in dual therapy as main treatment recommendation for uncomplicated gonorrhea, and this was reflected in the 2013 genomic survey with a decrease in the prevalence of NG-MAST G1407, which was associated with heterosexual groups in this survey. Results from the 2018 survey confirmed the further decline of this genogroup until practically vanishing. However, it also revealed the emergence and European dispersion of a novel lineage, NG-

MAST G12302 (NG-STAR CC168/63), carrying an *N. lactamica*-like sequence of the *mtrR* promoter and *mtrD* gene (termed "*mtr* mosaic 2" (Sánchez-Busó et al., 2021; Wadsworth et al., 2018)). This mosaic *mtr* had already been identified as an emerging cause of azithromycin resistance during the genomic analysis of isolates from the United States Gonococcal Isolate Surveillance Project (GISP) (Grad et al., 2016; Schmerer et al., 2020) and a clonal expansion has been reported in Germany (Banhart et al., 2021), a reflection of the Europe-wide dispersion described in the Euro-GASP 2018 survey (Sánchez-Busó et al., 2022), as well as the United States (Gernert et al., 2020; Thomas et al., 2019), and Canada (Sawatzky et al., 2022).

Results from these genomic epidemiology studies demonstrate the benefits of applying WGS for pathogen surveillance as it allows the monitorization of the dynamics of particular risk lineages and the study of existing and emerging genetic mechanisms of resistance.

4.3 *Klebsiella pneumoniae*

Klebsiella pneumoniae is an opportunistic pathogen that mainly causes hospital-acquired (HA) infections in critically ill and immunocompromised individuals. The most common manifestations are urinary tract, pneumonia, surgical site, and soft tissue infections, any of which can develop into bacteremia (Jarvis et al., 1985; Podschun and Ullmann, 1998). *K. pneumoniae* accounts for one-third of all gram-negative nosocomial infections (Navon-Venezia et al., 2017) and is a major cause of neonatal sepsis (Zaidi et al., 2005). Although *K. pneumoniae* is not a major cause of life-threatening infections, it has gained notoriety as a pathogen due to its prevalence and its potential to rapidly acquire and accumulate antimicrobial resistance mechanisms to a variety of antimicrobial classes. Nowadays, antibiotic-resistant *K. pneumoniae* prevalence has increased globally. Consequently, we face an increasing scarcity of effective treatments where mild *K. pneumoniae* infections have become very challenging to treat and together with severe infections, have become life-threatening (Kuehn, 2013). In particular, the increasing rates of multidrug-resistant *K. pneumoniae* due to the acquisition of ESBLs and carbapenemase genes are a big concern.

The first-line therapy to treat *K. pneumoniae* infections are cephalosporins, aminoglycosides, or quinolones. However, due to the acquisition of extended-spectrum lactamases (ESBL) by *K. pneumoniae* strains over the last decade, there has been a significant increase in the non-susceptibility rate to first-line antibiotics. These strains have emerged worldwide, and currently, they can reach a prevalence of up to 75% in endemic areas (European Centre for Disease Control and Prevention, 2022; Cantón et al., 2008). To treat infections caused by ESBL strains, carbapenems are the most widely used antibiotics. But carbapenem-resistant *K. pneumoniae* (CRKp) strains have also been reported. Although other antimicrobial classes can be used to treat CRKp infections, carbapenem resistance is more frequently found in combination with resistance to those other antibiotic families than alone (European Centre for Disease Prevention and Control, 2022). Consequently, therapeutic options for treating infections caused by CRKp are very limited and often rely on less effective, more toxic, and/or not widely available agents such as polymyxins, fosfomycin, tigecycline, and select aminoglycosides (Kaewpoowat and Ostrosky-Zeichner, 2015). Indeed, it has been estimated that the mortality rate associated with hospital-acquired bacteremia caused by CRKp ranges from 48% to 65% (Ben-David et al., 2012; Borer et al., 2009; Fraenkel-Wandel et al., 2016; Marra et al., 2006; Chetcuti Zammit et al., 2014), which is significantly higher compared to carbapenem-susceptible infections (Xu et al., 2017). As a result of the growing burden of infections caused by CRKp strains, the WHO has

identified *K. pneumoniae* as one of the critical priorities for therapeutic development (World Health Organization, 2017). Alarmingly, the prevalence of CRKp has been increasing worldwide. In 2005, almost all European regions were free of CRKp; by 2020, almost a quarter of EU/EEA countries reported that more than 10% of *K. pneumoniae* isolates causing infections were resistant to carbapenems, and Greece reached nonsusceptibility rates above 50% (European Centre for Disease Control and Prevention, 2022). Indeed, in 2019, CRKp was reported as the fastest-growing cause of drug-resistant infections in European hospitals (Cassini et al., 2019).

Although the prevalence of carbapenem-resistant *K. pneumoniae* remains relatively low, its impact on health systems is large, mainly due to the high attributable mortality and mobility of these infections (Antimicrobial Resistance Collaborators, 2022). This situation not only affects immunodeficient individuals but also those who are healthy.

4.3.1 Genetic Mechanisms of Antimicrobial Resistance

Antimicrobial resistance in *K. pneumoniae* is influenced by a large number of antimicrobial determinants, mainly antimicrobial-resistant genes (ARGs). Over 100 acquired ARGs from different antimicrobial families have been reported in *K. pneumoniae* (Navon-Venezia et al., 2017).

β-lactam antibiotics are the most commonly used class to treat *K. pneumoniae* infections [42]. β-lactamases are the most important AMR mechanisms to β-lactams (Bonomo, 2017). These enzymes can inactivate β-lactam antibiotics by breaking the amide bond of the β-lactam ring. Since their discovery, hundreds of different β-lactamases have been reported (>2000) (Bush and Jacoby, 2010). The major health concerns among β-lactamases are ESBLs (Paterson and Bonomo, 2005) and carbapenemases (Arnold et al., 2011). ESBLs are β-lactamases that can hydrolyze all first-line β-lactams such as penicillins, first-, second-, and third-generation cephalosporins, and aztreonam but are inhibited by the β-lactamase inhibitors (e.g., clavulanic acid). Carbapenemases are capable of hydrolyzing nearly all available β-lactams, including carbapenems and even β-lactamase inhibitors. Carbapenemases are the primary mechanism of CRKp strains all over the world. A few carbapenemases have been identified in *K. pneumoni*ae: KPC (*Klebsiella pneumoniae* carbapenemase) (Munoz-Price et al., 2013), OXA-48 (oxacillinase-48) (Pitout et al., 2020), NDM (New Delhi metallo-β-lactamase) (Wu et al., 2019), IMP (active-on-imipenem) (Lowe et al., 2020), and VIM (Verona integron-mediated metallo-β-lactamase) (Walsh, 2005).

β-lactamase genes are frequently found in combination with other AMR determinants to first-line antibiotic classes such as aminoglycosides or quinolones. Aminoglycoside resistance in *K. pneumoniae* is mainly mediated by aminoglycoside-modifying enzymes and 16S rRNA methylases. Whereas aminoglycoside-modifying enzymes have a narrow spectrum of activity, 16S rRNA methylases confer high-level resistance to all widely used aminoglycosides, including those recently developed (Poulikakos and Falagas, 2013). In contrast, quinolone resistance is mostly mediated by point mutations in the DNA gyrase and the topoisomerase IV codifying genes (*gyrA*, *gyrB*, *parC*, *parE*) (Rodríguez-Martínez et al., 2016). However, *K. pneumoniae* can acquire plasmid-mediated resistance mechanisms to quinolones. The *qnr* genes can be acquired through plasmid transfer and encode peptides that physically protect gyrase and topoisomerase IV from quinolone activity. The acquired aminoglycoside acetyltransferase AAC(6′)-Ib-cr, which also reduces the susceptibility to aminoglycosides, can modify quinolones, avoiding their effects (Ruiz, 2019; Schultsz and Geerlings, 2012).

Antimicrobial resistance determinants to other antibiotic families used as last-line therapy, such as polymyxins (e.g., colistin), have also been reported in *K. pneumoniae*. Polymyxin resistance is mainly mediated by chromosomal mutations in the *pmrA*, *pmrB*, *phoP*, *phoQ*, *mrgB*, or *crrB* genes. Plasmid-mediated polymyxin resistance due to the *mcr* gene has also been reported.

4.3.2 The Role of Plasmids

Plasmids play a key role in the introduction of ARGs from environmental sources into clinically relevant bacterial species and later in the dissemination of these ARGs among other clinically relevant organisms (Carattoli, 2013; Wyres and Holt, 2018; Koonin et al., 2001). Indeed, the vast majority of AMR mechanisms in *K. pneumoniae* are associated with the acquisition and accumulation of ARGs as a result of plasmid transfer (Navon-Venezia et al., 2017). Plasmids are extrachromosomal circular DNA molecules with a backbone filled with core plasmid functions related to replication, survival, mobility, and plasmid stability, joined by recombination with accessory functions that allow the cell to adapt to specific environments (AMR, virulence, environmental, metabolisms, etc.). Accessory genes, especially ARGs, are usually harbored within other mobile and transposable elements such as gene cassettes, integrons, or transposons that can integrate and accumulate into the same plasmid.

The remarkable versatility of plasmids to rearrange accessory genes and the potential of *K. pneumoniae* to incorporate plasmids has allowed the emergence and spread of multidrug-resistant (MDR) and pan-drug-resistant (PDR) phenotypes (Galani et al., 2019; Mezzatesta et al., 2011; Naf-plioti et al., 2021; Ramos et al., 2014).

In *K. pneumoniae*, a high diversity of plasmids has been reported with different Inc and replicon types, where IncFII, IncL/M, IncN, IncR, and IncX3 are among the most common plasmid types carrying ARGs (Wyres et al., 2020). Transposons and plasmids increase the mobility of ARGs and the likelihood of their accumulation in a single plasmid. In addition, it has been shown that plasmids can affect the resistance level of a given ARG. When plasmid-encoded, β-lactamases have been shown to have higher resistance levels. This can be the result of the increased gene expression owing to the copy number (Millan et al., 2015, 2016). As plasmids are usually in more copies than the chromosome, plasmid genes tend to be more highly expressed (Chaves et al., 2001; San Millan et al., 2018).

4.3.3 Genomic Epidemiology of Klebsiella

In *K. pneumoniae*, the introduction of whole genome sequencing has provided enough resolution to elucidate the high genetic diversity within the species. Over 250 phylogenetically distinct lineages have been identified. Although hundreds of these lineages can cause human infections, only a few contribute to the majority of them (e.g., ST11, ST147, ST307). These lineages have been designated as "high-risk" lineages because they are highly prevalent in healthcare settings around the world (Woodford et al., 2011; Wyres et al., 2020). Most of these lineages emerged locally and were subsequently disseminated throughout the world, spreading AMR determinants (Mathers et al., 2015).

The spread of AMR in *K. pneumoniae*, especially carbapenemases, is driven by many different situations. For example, KPC enzymes have been found in more than 100 different lineages but the dissemination of KPC carbapenemases is associated with the expansion of the high-risk lineage ST258 (Marsh et al., 2019; Munoz-Price et al., 2013). The first KPC enzyme was identified in a non-ST258 *K. pneumoniae* sample collected in 1996 in the United States. However, by 2009, over 70% of the KPC-producing *K. pneumoniae* in the United States belonged to ST258 (Kitchel et al., 2009). By this year, reports from all over the world described the presence of the KPC linked to the ST258 lineage.

Nowadays, KPC and ST258 are considered endemic in many countries such as the United States, Israel, Greece, Italy, Poland, Colombia, Argentina, and Brazil (Munoz-Price et al., 2013). In contrast to what has been observed with the ST258 lineage and the *bla* $_{KPC}$ genes, the emergence and dissemination of other carbapenemases have been due to plasmid dispersion through different lineages.

The main reason for the global distribution of OXA-48 is the dissemination of the IncL/M-type plasmid pOXA-48a harboring the *bla*$_{OXA-48}$ gene (Poirel et al., 2012a). The pOXA-48a is a self-transmissible plasmid with a very high transmission rate (Potron et al., 2014), which enhances its ability to easily and widely disperse between lineages and even bacterial species (Pitout et al., 2020). pOXA-48a has been associated with successful high-risk lineages, such as ST101, ST395, ST405, ST11, ST14, and ST15 (Carattoli et al., 2015; Potron et al., 2013), that have contributed to the international spread of this plasmid by introducing it into local transmission networks. OXA-48-producing *K. pneumoniae* are endemic in Western European and North African countries (Poirel et al., 2012b).

The *bla* $_{NDM}$ gene presents a different case of plasmid-driven dispersion, in which the current global dissemination is linked to different host plasmids in different lineages. *bla* $_{NDM-1}$ was firstly identified in a *K. pneumoniae* strain isolated from a Swedish patient transferred from a New Delhi hospital in 2008 (Yong et al., 2009). Later, numerous reports of NDM-producing *K. pneumoniae* were reported in the UK; however, over half the cases had previously traveled to or were hospitalized in India or Pakistan. Later, a rapid spread of NDM genes was reported worldwide, although most carriers had epidemiological links to India (Kumarasamy et al., 2010). Today, NDM genes have been acquired by many different plasmids distributed across a large number of STs (Poirel et al., 2011). Moreover, no predominant lineages have been reported, remarking on the highly diverse genetic background responsible for the dissemination of these genes (Wu et al., 2019). A similar scenario has been linked to the spread of VIM and IMP carbapenemases (Lowe et al., 2020; Nordmann and Poirel, 2014).

5. Future Perspectives: A Gold Standard Becoming Routine?

Recent advances in clinical diagnostic methods and public health surveillance systems have been made possible by the application of WGS by HTS to the study of infectious pathogens. These advancements produce results that are highly replicable, faster, and more accurate than those obtained using traditional microbiological methods. Although sequencing prices are decreasing (the lowest reported is 10 USD (Perez-Sepulveda et al., 2021)), the implementation of WGS as a surveillance tool is not a reality yet. Only some high-income countries such as the United States (National Antimicrobial Resistance Monitoring System), the United Kingdom (Public Health England), and the Netherlands (National Institute for Public Health and the Environment, RIVM) have implemented WGS as part of their routine AMR diagnostic tools (Armstrong et al., 2019).

There are different considerations that are limiting the implementation of WGS as part of national surveillance systems all over the world:

- Infrastructure. WGS has complex technological and infrastructure requirements from laboratory and computational perspectives. Regarding the laboratory workflow, it requires enough space to set up a unidirectional workflow with the ability to regulate the temperature and humidity of the laboratory. Current platforms require vibration-free platforms. For computational equipment, there is a need for dust-free rooms to store the computer cabins and the ability to control adequate

temperatures (Meehan et al., 2019). In addition, there are some limitations specific to LMIC. WGS laboratories require a reliable internet connection, continuous electricity supply, cooling storage, or quality water for molecular biology. The lack of these resources is often a barrier to the use of WGS for AMR surveillance in LMIC.

- Expertise. Another requisite to setting up an AMR surveillance system is the need for specialized personnel with great expertise in microbiology, molecular biology, and bioinformatics. Microbiologists have the expertise to prepare the biological samples whereas molecular biologists are demanded to perform the WGS library preparation protocols. Bioinformaticians are needed to analyze the sequencing results and create, handle, and maintain the bioinformatic pipelines for AMR surveillance. This is a requirement as for today there are still no standardized or universally accepted "gold-standard" tools or pipelines for bioinformatics analysis of AMR. Furthermore, the interpretation of the results and the routine adjustments of protocols and parameters require the joint intervention of the three kinds of professionals.
- Cost. HTS equipment usually involves a high initial investment and, after, there are maintenance costs for continuous power, cooling and replacement of components. The cost of infrastructure preparation, purchase of computational resources, sequencing platforms, other required devices, and necessary reagents is huge. Moreover, if the surveillance system is not centralized, building suitable laboratories (or renovating the previous ones) in many places simultaneously can be challenging.
- Technical limitations. Current methodologies and bioinformatics analyses do not allow obtaining perfect AMR susceptibility profiles by using the genome content. First, databases do not contain every AMR determinant, as many remain unknown or difficult to identify. Second, the presence of an AMR determinant does not necessarily mean phenotypic resistance, as the bacteria may not be expressing it.
- Standardization. Although WGS laboratory protocols are mostly standardized for all bacterial pathogens and can be performed in different laboratories with reproducibility, bioinformatic analyses lack a gold-standard pipeline, with different possible implementations depending on the personnel responsible for the analysis.

In low- and middle-income countries, this implementation seems farther. Several low-middle income countries are still behind due to the multiple constraints in setting up suitable laboratories with WGS and bioinformatics capacities. However, initiatives such as that from the NIHR Global Health Research Unit on Genomic Surveillance of AMR(NIHR Global Health Research Unit on Genomic Surveillance of Amr, 2020; Kekre et al., 2021) (2020) are aimed at bridging these gaps, and the number of laboratories and countries using WGS as a routine technique for AMR evaluation and surveillance will no doubt increase in the coming years.

6. Conclusions

WGS using HTS for the detection of AMR determinants is essential to implement control and prevention strategies to combat the increasing threat of AMR. However, there is a need to standardize sequencing protocols, bioinformatic pipelines, and AMR databases to perform reproducible and accurate phenotypic predictions based on genomic data for clinically relevant pathogens. With a few exceptions, we are still far from their regular usage as diagnostic tools in real, clinical settings, for

which costs, and most notably, the time from sampling until accurate results are available, should be reduced considerably. Nevertheless, the use of HTS for surveillance is quickly replacing other alternative methods although more efforts are needed to implement AMR genomic surveillance worldwide, especially in the LMICs.

Acknowledgments

This work was supported by projects PID2021-127010OB-I00 from Spanish Ministry of Science and Innovation, CIPROM-2021-053 from Generalitat Valencia, and Conselleria de Sanitat Universal i Salut Pública, Generalitat Valenciana. LSB is funded by Conselleria de Sanitat Universal i Salut Pública, Generalitat Valenciana (CDEI-06/20-B), Conselleria d'Innovació, Universitats, Ciència i Societat Digital (CISEJI/2022/66) and Ministry of Science and Innovation (PID2020-120113RA-I00), Spain. ICM holds a Margarita Salas postdoctoral contract UNI/551/2021 from the Spanish Ministry of Universities.

References

Amarasinghe, S.L., et al., 2020. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 21 (1). https://doi.org/10.1186/s13059-020-1935-5.

Andersson, D.I., Hughes, D., 2014. Microbiological effects of sublethal levels of antibiotics. Nat. Rev. Microbiol. 12 (7), 465−478. https://doi.org/10.1038/nrmicro3270.

Antimicrobial Resistance Collaborators, 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399 (10325), 629−655. https://doi.org/10.1016/S0140-6736(21)02724-0.

Armstrong, G.L., et al., 2019. Pathogen genomics in public health. N. Engl. J. Med. 381 (26), 2569−2580. https://doi.org/10.1056/NEJMsr1813907.

Arnold, R.S., et al., 2011. Emergence of *Klebsiella pneumoniae* carbapenemase-producing bacteria. Southern Med. J. 104 (1), 40−45. https://doi.org/10.1097/SMJ.0b013e3181fd7d5a.

Banhart, S., et al., 2021. The mosaic mtr locus as major genetic determinant of azithromycin resistance of Neisseria gonorrhoeae —Germany, 2018. J. Infect. Dis. 224 (8), 1398−1404. https://doi.org/10.1093/infdis/jiab091.

Ben-David, D., et al., 2012. Outcome of carbapenem resistant *Klebsiella pneumoniae* bloodstream infections. Clin. Microbiol. Infection 18 (1), 54−60. https://doi.org/10.1111/j.1469-0691.2011.03478.x.

Bhullar, K., et al., 2012. Antibiotic resistance is prevalent in an isolated cave microbiome. PLoS One 7 (4), e0034953. https://doi.org/10.1371/journal.pone.0034953.

Bonomo, R.A., 2017. β-Lactamases: a focus on current challenges. Cold Spring Harbor Perspec. Med. 7 (1), a025239. https://doi.org/10.1101/cshperspect.a025239.

Borer, A., et al., 2009. Attributable mortality rate for carbapenem-resistant *Klebsiella pneumoniae* bacteremia. Infect. Control Hosp. Epidemiol. 30 (10), 972−976. https://doi.org/10.1086/605922.

Boritsch, E.C., et al., 2016. Key experimental evidence of chromosomal DNA transfer among selected tuberculosiscausing mycobacteria. Proc. Nat. Acad. Sci. United States Am. France Nat. Acad. Sci. 113 (35), 9876−9881. https://doi.org/10.1073/pnas.1604921113.

Bortolaia, V., et al., 2020. ResFinder 4.0 for predictions of phenotypes from genotypes. J. Antimicrob. Chemother. 75 (12), 3491−3500. https://doi.org/10.1093/jac/dkaa345.

Bush, K., Jacoby, G.A., 2010. Updated functional classification of β-lactamases. Antimicrob. Agents Chemother. 54 (3), 969−976. https://doi.org/10.1128/AAC.01009-09.

Cancino-Muñoz, I., et al., 2019. Cryptic resistance mutations associated with misdiagnoses of multidrug-resistant tuberculosis. J. Infect. Dis. 220 (2), 316−320. https://doi.org/10.1093/infdis/jiz104.

Cancino-Muñoz, et al., 2022. Population-based sequencing of reveals how current population dynamics are shaped by past epidemics. Elife 11, e76605. https://doi.org/10.7554/eLife.76605.

Cantón, R., et al., 2008. Prevalence and spread of extended-spectrum β-lactamase-producing enterobacteriaceae in Europe. Clin. Microbiol. Infection 14 (1), 144−153. https://doi.org/10.1111/j.1469-0691.2007.01850.x.

Carattoli, A., 2013. Plasmids and the spread of resistance. Int. J. Med. Microbiol. 303 (6−7), 298−304. https://doi.org/10.1016/j.ijmm.2013.02.001.

Carattoli, A., et al., 2015. "Differentiation of IncL and IncM plasmids associated with the spread of clinically relevant antimicrobial resistance. PLoS One 10 (5). https://doi.org/10.1371/journal.pone.0123063.

Cartee, J.C., et al., 2022. Phylogenomic comparison of causing disseminated gonococcal infections and uncomplicated gonorrhea in Georgia, United States. Open Forum Infect. Dis. 9 (7).

Cassini, A., et al., 2018. Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the Burden of Communicable Diseases in Europe study, European Union and European Economic Area countries, 2009 to 2013. Euro Surveill. 23 (16), 17. https://doi.org/10.2807/1560-7917.ES.2018.23.16.17-00454.

Cassini, A., et al., 2019. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. Lancet Infect. Dis. 19 (1), 56−66. https://doi.org/10.1016/S1473-3099(18)30605-4.

Centers for Disease Control and Prevention, 2019. Antibiotic Resistance Threats in the United States. U.S. Department of Health and Human Services. Available at: https://www.cdc.gov/drugresistance/pdf/threats-report/2019-ar-threats-report-508.pdf.

Chaves, J., et al., 2001. SHV-1 β-lactamase is mainly a chromosomally encoded species-specific enzyme in *Klebsiella pneumoniae*. Antimicrob. Agents Chemother. 45 (10), 2856−2861. https://doi.org/10.1128/AAC.45.10.2856-2861.2001.

Chetcuti Zammit, S., Azzopardi, N., Sant, J., 2014. Mortality risk score for *Klebsiella pneumoniae* bacteraemia. Eur. J. Intern. Med. 25 (6), 571−576. https://doi.org/10.1016/j.ejim.2014.04.008.

Chisholm, S.A., et al., 2013. Molecular epidemiological typing within the European Gonococcal Antimicrobial Resistance Surveillance Programme reveals predominance of a multidrugresistant clone. Euro Surveill. 18 (3). Available at: http://www.eurosurveillance.org/images/dynamic/EE/V18N03/art20358.pdf.

Cole, M.J., et al., 2017. Overall low extended-spectrum cephalosporin resistance but high azithromycin resistance in Neisseria gonorrhoeae in 24 European countries, 2015. BMC Infect. Dis. 17 (1), 617. https://doi.org/10.1186/s12879-017-2707-z.

Cole, M.J., et al., 2019. The European gonococcal antimicrobial surveillance programme (Euro-GASP) appropriately reflects the antimicrobial resistance situation for Neisseria gonorrhoeae in the European Union/European Economic Area. BMC Infect. Dis. 19 (1). https://doi.org/10.1186/s12879-019-4631-x.

Coll, F., et al., 2015. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med. 7 (1). https://doi.org/10.1186/s13073-015-0164-0.

Comas, I., et al., 2011. Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. Nat. Genet. 44 (1), 106−110. https://doi.org/10.1038/ng.1038.

Croucher, N.J., Didelot, X., 2015. The application of genomics to tracing bacterial pathogen transmission. Curr. Opin. Microbiol. 23, 62−67. https://doi.org/10.1016/j.mib.2014.11.004.

CRyPTIC Consortium and the 100, 000 Genomes Project, 2018. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. N. Engl. J. Med. 379 (15), 1403−1415. https://doi.org/10.1056/NEJMoa1800474.

Darby, E.M., et al., 2023. Molecular mechanisms of antibiotic resistance revisited. Nat. Rev. Microbiol. 21 (5), 280−295. https://doi.org/10.1038/s41579-022-00820-y.

Davies, H.T.O., Crombie, I.K., 1997. Interpreting health outcomes. J. Eval. Clin. Pract. 3 (3), 187−199. https://doi.org/10.1046/j.1365-2753.1997.00003.x.

de Vos, M., et al., 2013. Putative compensatory mutations in the rpoC gene of rifampin-resistant *Mycobacterium tuberculosis* are associated with ongoing transmission. Antimicrob. Agents Chemother. 57 (2), 827−832. https://doi.org/10.1128/aac.01541-12.

Demczuk, W., et al., 2020. Equations to predict antimicrobial MICs in Neisseria gonorrhoeae using molecular antimicrobial resistance determinants. Antimicrob. Agents Chemother. 64 (3). https://doi.org/10.1128/AAC.02005-19.

Dreyer, V., et al., 2022. High fluoroquinolone resistance proportions among multidrug-resistant tuberculosis driven by dominant L2 *Mycobacterium tuberculosis* clones in the Mumbai Metropolitan Region. Genom. Med. 14 (1). https://doi.org/10.1186/s13073-022-01076-0.

D'Costa, V.M., et al., 2011. Antibiotic resistance is ancient. Nature 477 (7365), 457−461. https://doi.org/10.1038/nature10388.

European Centre for Disease Control and Prevention, 2022. Antimicrobial Resistance in the EU/EEA (EARS-Net) - Annual Epidemiological Report 2020. European Centre for Disease Prevention and Control. Available at: https://www.ecdc.europa.eu/sites/default/files/documents/AER-EARS-Net-2020.pdf.

European Centre for Disease Prevention and Control, 2022. Antimicrobial Resistance Surveillance in Europe 2022 - 2020 Data. Available at: https://www.ecdc.europa.eu/en/publications-data/antimicrobial-resistance-surveillance-europe-2022-2020-data.

Eyre, D.W., et al., 2017. WGS to predict antibiotic MICs for Neisseria gonorrhoeae. J. Antimicrob. Chemother. 72 (7), 1937−1947. https://doi.org/10.1093/jac/dkx067.

Feldgarden, M., et al., 2019. Validating the AMRFINder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. Antimicrob. Agents Chemother. 63 (11). https://doi.org/10.1128/AAC.00483-19.

Fifer, H., et al., 2016. Failure of dual antimicrobial therapy in treatment of gonorrhea. N. Engl. J. Med. 374 (25), 2504−2506. https://doi.org/10.1056/NEJMc1512757.

Fifer, H., et al., 2018. Sustained transmission of high-level azithromycin-resistant Neisseria gonorrhoeae in England: an observational study. Lancet Infect. Dis. 18 (5), 573−581. https://doi.org/10.1016/s1473-3099(18)30122-1.

Fifer, H., et al., 2020. 2018 UK national guideline for the management of infection with Neisseria gonorrhoeae. Int. J. STD AIDS 31 (1), 4−15. https://doi.org/10.1177/0956462419886775.

Fitzpatrick, C., Floyd, K., 2012. A systematic review of the cost and cost effectiveness of treatment for multidrug-resistant tuberculosis. Pharmacoeconomics 30 (1), 63−80. https://doi.org/10.2165/11595340-000000000-00000.

Flandrois, J.-P., Lina, G., Dumitrescu, O., 2014. MUBII-TB-DB: a database of mutations associated with antibiotic resistance in *Mycobacterium tuberculosis*. BMC Bioinf. 15 (1), 107. https://doi.org/10.1186/1471-2105-15-107.

Fraenkel-Wandel, Y., et al., 2016. Mortality due to blaKPC *Klebsiella pneumoniae* bacteraemia. J. Antimicrob. Chemother. 71 (4), 1083−1087. https://doi.org/10.1093/jac/dkv414.

Furió, V., et al., 2021. An evolutionary functional genomics approach identifies novel candidate regions involved in isoniazid resistance in *Mycobacterium tuberculosis*. Commun. Biol. 4 (1). https://doi.org/10.1038/s42003-021-02846-z.

Gagneux, S., 2018. Ecology and evolution of *Mycobacterium tuberculosis*. Nat. Rev. Microbiol. 16 (4), 202−213. https://doi.org/10.1038/nrmicro.2018.8.

Gagneux, S., et al., 2006. The competitive cost of antibiotic resistance in *Mycobacterium tuberculosis*. Science 312 (5782), 1944−1946. https://doi.org/10.1126/science.1124410.

Galani, I., et al., 2019. Nationwide epidemiology of carbapenem resistant *Klebsiella pneumoniae* isolates from Greek hospitals, with regards to plazomicin and aminoglycoside resistance. BMC Infect. Dis. 19 (1). https://doi.org/10.1186/s12879-019-3801-1.

Gardy, J.L., Loman, N.J., 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat. Rev. Genet. 19 (1), 9−20. https://doi.org/10.1038/nrg.2017.88.

Genome-wide Association Studies of Global Mycobacterium Tuberculosis Resistance to 13 Antimicrobials in 10,228 Genomes Identify New Resistance Mechanisms, 2022. The CRyPTIC Consortium.

Gernert, K.M., et al., 2020. Azithromycin susceptibility of Neisseria gonorrhoeae in the USA in 2017: a genomic analysis of surveillance data. Lancet Microbe 1 (4), e154−e164. https://doi.org/10.1016/S2666-5247(20)30059-8.

Goldstein, B.P., 2014. Resistance to rifampicin: a review. J. Antibiot. 67 (9), 625−630. https://doi.org/10.1038/ja.2014.107.

Golparian, D., et al., 2020. Genomic evolution of Neisseria gonorrhoeae since the preantibiotic era (1928-2013): antimicrobial use/misuse selects for resistance and drives evolution. BMC Genom. 21 (1). https://doi.org/10.1186/s12864-020-6511-6.

Grad, Y.H., et al., 2016. Genomic epidemiology of gonococcal resistance to extended-spectrum cephalosporins, macrolides, and fluoroquinolones in the United States, 2000-2013. J. Infect. Dis. 214 (10), 1579−1587. https://doi.org/10.1093/infdis/jiw420.

Guo, S., et al., 2019. Phenotypic and genotypic characterization of antimicrobial resistant *Escherichia coli* isolated from ready-to-eat food in Singapore using disk diffusion, broth microdilution and whole genome sequencing methods. Food Contr. 99, 89−97. https://doi.org/10.1016/j.foodcont.2018.12.043.

Gygli, S.M., et al., 2017. Antimicrobial resistance in Mycobacterium tuberculosis: mechanistic and evolutionary perspectives. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Rev. 41 (3), 354−373. https://doi.org/10.1093/femsre/fux011.

Harris, S.R., et al., 2013. Whole-genome sequencing for analysis of an outbreak of meticillin-resistant *Staphylococcus aureus*: a descriptive study. Lancet Infect. Dis. 13 (2), 130−136. https://doi.org/10.1016/S1473-3099(12)70268-2.

Harris, S.R., et al., 2018. Public health surveillance of multidrug-resistant clones of Neisseria gonorrhoeae in Europe: a genomic survey. Lancet Infect. Dis. 18 (7), 758−768. https://doi.org/10.1016/S1473-3099(18)30225-1.

Hendriksen, R.S., et al., 2019. Using genomics to track global antimicrobial resistance. Front. Public Health 7. https://doi.org/10.3389/fpubh.2019.00242.

Heym, B., et al., 1997. Effects of overexpression of the alkyl hydroperoxide reductase AhpC on the virulence and isoniazid resistance of Mycobacterium tuberculosis. Infect. Immun. 65 (4), 1395−1401. https://doi.org/10.1128/iai.65.4.1395-1401.1997.

Hicks, A.L., et al., 2019. Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. PLoS Computat. Biol. 15 (9). https://doi.org/10.1371/journal.pcbi.1007349.

Hunt, M., et al., 2017. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. Microb. Genom. 3 (10). https://doi.org/10.1099/mgen.0.000131.

Inouye, M., et al., 2014. SRST2: rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 6 (11). https://doi.org/10.1186/s13073-014-0090-6.

Jarvis, W.R., et al., 1985. The epidemiology of nosocomial infections caused by *Klebsiella pneumoniae*. Infect. Control 6 (2), 68−74. https://doi.org/10.1017/S0195941700062639.

Jia, B., et al., 2017. Card 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 45 (1), D566−D573. https://doi.org/10.1093/nar/gkw1004.

Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2018. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications [version 1; referees: 2 approved]. Wellcome Open Res. 3. https://doi.org/10.12688/wellcomeopenres.14826.1.

Kaewpoowat, Q., Ostrosky-Zeichner, L., 2015. Tigecycline: a critical safety review. Expet Opin. Drug Saf. 14 (2), 335−342. https://doi.org/10.1517/14740338.2015.997206.

Kekre, M., et al., 2021. Integrating scalable genome sequencing into microbiology laboratories for routine antimicrobial resistance surveillance. Clin. Infect. Dis. 73, S258−S266. https://doi.org/10.1093/cid/ciab796.

Kitchel, B., et al., 2009. Molecular epidemiology of KPC-producing *Klebsiella pneumoniae* isolates in the United States: clonal expansion of multilocus sequence type 258. Antimicrob. Agents Chemother. 53 (8), 3365−3370. https://doi.org/10.1128/AAC.00126-09.

Koonin, E.V., Makarova, K.S., Aravind, L., 2001. Horizontal gene transfer in prokaryotes: quantification and classification. Annu. Rev. Microbiol. 55, 709−742. https://doi.org/10.1146/annurev.micro.55.1.709.

Kuehn, B.M., 2013. 'Nightmare' bacteria on the rise in US hospitals, long-term care facilities. JAMA 309 (15), 1573−1574. https://doi.org/10.1001/jama.2013.2922.

Kumarasamy, K.K., et al., 2010. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect. Dis. 10 (9), 597−602. https://doi.org/10.1016/S1473-3099(10)70143-2.

Lahra, M.M., et al., 2018. Cooperative recognition of internationally disseminated ceftriaxone-resistant Neisseria gonorrhoeae strain. Emerg. Infect. Dis. 24 (4), 735−740. https://doi.org/10.3201/eid2404.171873.

Lindbäck, E., et al., 2002. Mutations in gyrA, gyrB, parC, and parE in quinolone-resistant strains of Neisseria gonorrhoeae. APMIS 110 (9), 651−657. https://doi.org/10.1034/j.1600-0463.2002.1100909.x.

Lindberg, R., et al., 2007. Neisseria gonorrhoeae isolates with reduced susceptibility to cefixime and ceftriaxone: association with genetic polymorphisms in penA , mtrR , porB1b , and ponA. Antimicrob. Agents Chemother. 51 (6), 2117−2122. https://doi.org/10.1128/aac.01604-06.

Loman, N.J., et al., 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. Nat. Rev. Microbiol. 10 (9), 599−606. https://doi.org/10.1038/nrmicro2850.

López, M.G., et al., 2020. Tuberculosis in Liberia: high multidrug-resistance burden, transmission and diversity modelled by multiple importation events. Microb. Genom. 6 (1). https://doi.org/10.1099/mgen.0.000325.

Lowe, C.F., et al., 2020. The brief case: IMP, the uncommonly common carbapenemase. J. Clin. Microbiol. 58 (4). https://doi.org/10.1128/JCM.01094-19.

Marra, A.R., et al., 2006. Nosocomial bloodstream infections caused by *Klebsiella pneumoniae*: impact of extended-spectrum β-lactamase (ESBL) production on clinical outcome in a hospital with high ESBL prevalence. BMC Infect. Dis. 6. https://doi.org/10.1186/1471-2334-6-24.

Marsh, J.W., et al., 2019. Evolution of outbreak-causing carbapenem-resistant *Klebsiella pneumoniae* ST258 at a tertiary care hospital over 8 years. mBio 10 (5). https://doi.org/10.1128/mBio.01945-19.

Martinez, J.L., Baquero, F., 2000. Mutation frequencies and antibiotic resistance. Antimicrob. Agents Chemother. 44 (7), 1771−1777. https://doi.org/10.1128/aac.44.7.1771-1777.2000.

Mathers, A.J., Peirano, G., Pitout, J.D.D., 2015. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. Clin. Microbiol. Rev. 28 (3), 565−591. https://doi.org/10.1128/CMR.00116-14.

Mcarthur, A.G., et al., 2013. The comprehensive antibiotic resistance database. Antimicrob. Agents chemother. 57 (7), 3348−3357. https://doi.org/10.1128/AAC.00419-13.

McCall, H., et al., 2015. What is chemsex and why does it matter? BMJ 351. https://doi.org/10.1136/bmj.h5790.

McDermott, P.F., et al., 2016. Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal Salmonella. Antimicrob. Agents Chemother. 60 (9), 5515−5520. https://doi.org/10.1128/AAC.01030-16.

Meehan, C.J., et al., 2019. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. Nat. Rev. Microbiol. 17 (9), 533−545. https://doi.org/10.1038/s41579-019-0214-5.

Merker, M., et al., 2018. Compensatory evolution drives multidrug-resistant tuberculosis in central Asia. Elife 7. https://doi.org/10.7554/eLife.38200.

Mezzatesta, M.L., et al., 2011. Outbreak of KPC-3-producing, and colistin-resistant, *Klebsiella pneumoniae* infections in two Sicilian hospitals. Clin. Microbiol. Infection 17 (9), 1444−1447. https://doi.org/10.1111/j.1469-0691.2011.03572.x.

Millan, A.S., et al., 2015. Small-plasmid-mediated antibiotic resistance is enhanced by increases in plasmid copy number and bacterial fitness. Antimicrob. Agents Chemother. 59 (6), 3335−3341. https://doi.org/10.1128/AAC.00235-15.

Monk, J.M., 2019. Predicting antimicrobial resistance and associated genomic features from whole-genome sequencing. J. Clin. Microbiol. 57 (2). https://doi.org/10.1128/JCM.01610-18.

Mortimer, T.D., Grad, Y.H., 2019. Applications of genomics to slow the spread of multidrug-resistant neisseria gonorrhoeae. Ann. N. Y. Acad. Sci. 1435 (1), 93−109. https://doi.org/10.1111/nyas.13871.

Mortimer, T.D., et al., 2021. The distribution and spread of susceptible and resistant Neisseria gonorrhoeae across demographic groups in a major metropolitan center. Clin. Infect. Dis. 73 (9), E3146−E3155. https://doi.org/10.1093/cid/ciaa1229.

Munita, J.M., Arias, C.A., 2016. Mechanisms of antibiotic resistance. Microbiol. Spectr. 4 (2). https://doi.org/10.1128/microbiolspec.VMBF-0016-2015.

Munoz-Price, L.S., et al., 2013. Clinical epidemiology of the global expansion of *Klebsiella pneumoniae* carbapenemases. Lancet Infect. Dis. 13 (9), 785−796. https://doi.org/10.1016/S1473-3099(13)70190-7.

Nafplioti, K., et al., 2021. Characterization of 16S rRNA methylase genes in enterobacterales and *Pseudomonas aeruginosa* in athens metropolitan area, 2015−2016. Eur. J. Clin. Microbiol. Infect. Dis. 40 (1), 111−121. https://doi.org/10.1007/s10096-020-04006-3.

Navon-Venezia, S., Kondratyeva, Kira, Carattoli, A., 2017. *Klebsiella pneumoniae*: a major worldwide source and shuttle for antibiotic resistance. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Rev. 41 (3), 252−275. https://doi.org/10.1093/femsre/fux013.

Nguyen, M., et al., 2018. Developing an in silico minimum inhibitory concentration panel test for Klebsiella pneumonia. Sci. Rep. 8 (1). https://doi.org/10.1038/s41598-017-18972-w.

Nguyen, M., et al., 2019. Using machine learning to predict antimicrobial MICs and associated genomic features for nontyphoidal Salmonella. J. Clin. Microbiol. 57 (2). https://doi.org/10.1128/JCM.01260-18.

NIHR Global Health Research Unit on Genomic Surveillance of Amr, 2020. Whole-genome sequencing as part of national and international surveillance programmes for antimicrobial resistance: a roadmap. BMJ Glob. Health 5 (11), e002244. https://doi.org/10.1136/bmjgh-2019-002244.

Nordmann, P., Poirel, L., 2014. The difficult-to-control spread of carbapenemase producers among Enterobacteriaceae worldwide. Clin. Microbiol. Infection 20 (9), 821−830. https://doi.org/10.1111/1469-0691.12719.

Ohnishi, M., et al., 2011. Is Neisseria gonorrhoeae initiating a future era of untreatable gonorrhea?: detailed characterization of the first strain with high-level resistance to ceftriaxone. Antimicrob. Agents Chemother. 55 (7), 3538−3545. https://doi.org/10.1128/AAC.00325-11.

Papakonstantinou, D., et al., 2021. Mapping gene-by-gene single-nucleotide variation in 8,535 Mycobacterium tuberculosis genomes: a resource to support potential vaccine and drug development. mSphere 6 (2), 1−15. https://doi.org/10.1128/mSphere.01224-20.

Papaventsis, D., et al., 2017. Whole genome sequencing of Mycobacterium tuberculosis for detection of drug resistance: a systematic review. Clin. Microbiol. Infection 23 (2), 61−68. https://doi.org/10.1016/j.cmi.2016.09.008.

Paterson, D.L., Bonomo, R.A., 2005. Extended-spectrum β-lactamases: a clinical update. Clin. Microbiol. Rev. 18 (4), 657−686. https://doi.org/10.1128/CMR.18.4.657-686.2005.

Perez-Sepulveda, B.M., et al., 2021. An accessible, efficient and global approach for the large-scale sequencing of bacterial genomes. Genome Biol. 22 (1). https://doi.org/10.1186/s13059-021-02536-3.

Pitout, J.D.D., et al., 2020. The global ascendency of OXA-48-type carbapenemases. Clin. Microbiol. Rev. 33 (1). https://doi.org/10.1128/CMR.00102-19.

Pleininger, S., et al., 2022. Extensively drug-resistant (XDR) Neisseria gonorrhoeae causing possible gonorrhoea treatment failure with ceftriaxone plus azithromycin in Austria, April 2022. Euro Surveill. 27 (24). https://doi.org/10.2807/1560-7917.ES.2022.27.24.2200455.

Podschun, R., Ullmann, U., 1998. Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. Clin. Microbiol. Rev. 11 (4), 589−603. https://doi.org/10.1128/cmr.11.4.589.

Poirel, L., et al., 2011. Genetic features of bla NDM-1 -positive enterobacteriaceae. Antimicrob. Agents Chemother. 55 (11), 5403−5407. https://doi.org/10.1128/aac.00585-11.

Poirel, L., Potron, A., Nordmann, P., 2012a. OXA-48-like carbapenemases: the phantom menace. J. Antimicrob. Chemother. 67 (7), 1597−1606. https://doi.org/10.1093/jac/dks121.

Poirel, L., Bonnin, R.A., Nordmann, P., 2012b. Genetic features of the widespread plasmid coding for the carbapenemase OXA-48. Antimicrob. Agents Chemother. 56 (1), 559−562. https://doi.org/10.1128/AAC.05289-11.

Pontali, E., et al., 2017. Cardiac safety of bedaquiline: a systematic and critical analysis of the evidence. Eur. Respir. J. 50 (5). https://doi.org/10.1183/13993003.01462-2017.

Potron, A., et al., 2013. Intercontinental spread of OXA-48 beta-lactamase-producing Enterobacteriaceae over a 11-year period, 2001 to 2011. Euro Surveill. 18 (31). https://doi.org/10.2807/1560-7917.es2013.18.31.20549.

Potron, A., Poirel, L., Nordmann, P., 2014. Derepressed transfer properties leading to the efficient spread of the plasmid encoding carbapenemase OXA-48. Antimicrob. Agents Chemother. 58 (1), 467−471. https://doi.org/10.1128/aac.01344-13.

Poulikakos, P., Falagas, M.E., 2013. Aminoglycoside therapy in infectious diseases. Expet Opin. Pharmacother. 14 (12), 1585−1597. https://doi.org/10.1517/14656566.2013.806486.

Prestinaci, F., Pezzotti, P., Pantosti, A., 2015. Antimicrobial resistance: a global multifaceted phenomenon. Pathog. Glob. Health 109 (7), 309−318. https://doi.org/10.1179/2047773215Y.0000000030.

Programme, G.T., 2020. WHO Consolidated Guidelines on Tuberculosis: Module 4: Treatment: Drug-resistant Tuberculosis Treatment. Available at: https://www.who.int/publications/i/item/9789240007048.

Ramos, P.I.P., et al., 2014. Comparative analysis of the complete genome of KPC-2-producing *Klebsiella pneumoniae* Kp13 reveals remarkable genome plasticity and a wide repertoire of virulence and resistance mechanisms. BMC Genom. 15 (1). https://doi.org/10.1186/1471-2164-15-54.

Reuter, S., et al., 2013. Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. JAMA Intern. Med. 173 (15), 1397−1404. https://doi.org/10.1001/jamainternmed.2013.7734.

Rodríguez-Martínez, J.M., et al., 2016. Plasmid-mediated quinolone resistance: two decades on. Drug Resist. Updates 29, 13−29. https://doi.org/10.1016/j.drup.2016.09.001.

Rowley, J., et al., 2019. Chlamydia, gonorrhoea, trichomoniasis and syphilis: global prevalence and incidence estimates, 2016. Bull. World Health Organ. 97 (8), 548. https://doi.org/10.2471/BLT.18.228486.

Ruiz, J., 2019. Transferable mechanisms of quinolone resistance from 1998 onward. Clin. Microbiol. Rev. 32 (4). https://doi.org/10.1128/cmr.00007-19.

San Millan, A., et al., 2016. Multicopy plasmids potentiate the evolution of antibiotic resistance in bacteria. Nat. Ecol. Evol. 1 (1), 1−8. https://doi.org/10.1038/s41559-016-0010.

San Millan, A., et al., 2018. Integrative analysis of fitness and metabolic effects of plasmids in *Pseudomonas aeruginosa* PAO1. ISME J. 12 (12), 3014−3024. https://doi.org/10.1038/s41396-018-0224-8.

Sánchez-Busó, L., Harris, S.R., 2019. Using genomics to understand antimicrobial resistance and transmission in Neisseria gonorrhoeae. Microb. Genom. 5 (2). https://doi.org/10.1099/mgen.0.000239.

Sánchez-Busó, Leonor, et al., 2014. Recombination drives genome evolution in outbreak-related *Legionella pneumophila* isolates. Nat. Genet. 46 (11), 1205−1211. https://doi.org/10.1038/ng.3114.

Sánchez-Busó, L., et al., 2019. The impact of antimicrobials on gonococcal evolution. Nat. Microbiol. 4 (11), 1941−1950. https://doi.org/10.1038/s41564-019-0501-y.

Sánchez-Busó, L., et al., 2021. A community-driven resource for genomic epidemiology and antimicrobial resistance prediction of Neisseria gonorrhoeae at Pathogenwatch. Genome Med. 13 (1). https://doi.org/10.1186/s13073-021-00858-2.

Sánchez-Busó, L., et al., 2022. Europe-wide expansion and eradication of multidrug-resistant Neisseria gonorrhoeae lineages: a genomic surveillance study. Lancet Microbe 3 (6), e452−e463. https://doi.org/10.1016/S2666-5247(22)00044-1.

Sandgren, A., et al., 2009. Tuberculosis drug resistance mutation database. PLoS Med. 6 (2), e1000002. https://doi.org/10.1371/journal.pmed.1000002.

Sawatzky, P., et al., 2022. Increasing azithromycin resistance in Neisseria gonorrhoeae due to NG-MAST 12302 clonal spread in Canada, 2015 to 2018. Antimicrob. Agents Chemother. 66 (3), 016888-21. https://doi.org/10.1128/aac.01688-21.

Schmerer, M.W., et al., 2020. Genomic characterization of Neisseria gonorrhoeae strains from 2016 U.S. sentinel surveillance displaying reduced susceptibility to azithromycin. Antimicrob. Agents Chemother. 64 (5). https://doi.org/10.1128/AAC.02420-19.

Schultsz, C., Geerlings, S., 2012. Plasmid-mediated resistance in enterobacteriaceae: changing landscape and implications for therapy. Drugs 72 (1), 1−16. https://doi.org/10.2165/11597960-000000000-00000.

Seddon, J.A., et al., 2013. Hearing loss in children treated for multidrug-resistant tuberculosis. J. Infect. 66 (4), 320−329. https://doi.org/10.1016/j.jinf.2012.09.002.

Seemann, T., 2022. Abricate: Mass Screening of Contigs for Antimicrobial and Virulence Genes. Available at: https://github.com/tseemann/abricate.

Seifert, M., et al., 2015. Genetic mutations associated with isoniazid resistance in Mycobacterium tuberculosis: a systematic review. PLoS One 10 (3). https://doi.org/10.1371/journal.pone.0119628.

Sherman, D.R., et al., 1996. Compensatory ahpC gene expression in isoniazid-resistant Mycobacterium tuberculosis. Science 272 (5268), 1641−1643. https://doi.org/10.1126/science.272.5268.1641.

St Cyr, S., et al., 2021. Update to CDC's treatment guidelines for gonococcal infection, 2020. Morb. Mortal. Wkly. Rep. 69 (50), 1911−1916. https://doi.org/10.15585/MMWR.MM6950A6.

Supply, P., et al., 2003. Linkage disequilibrium between minisatellite loci supports clonal evolution of Mycobacterium tuberculosis in a high tuberculosis incidence area. Mol. Microbiol. 47 (2), 529−538. https://doi.org/10.1046/j.1365-2958.2003.03315.x.

Szamecz, B., et al., 2014. The genomic landscape of compensatory evolution. PLoS Biol. 12 (8). https://doi.org/10.1371/journal.pbio.1001935.

Taxt, A.M., et al., 2020. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. Sci. Rep. 10 (1). https://doi.org/10.1038/s41598-020-64616-x.

Telenti, A., et al., 1993. Detection of rifampicin-resistance mutations in Mycobacterium tuberculosis. Lancet 341 (8846), 647−651. https://doi.org/10.1016/0140-6736(93)90417-f.

Thomas, J.C., et al., 2019. Evidence of recent genomic evolution in gonococcal strains with decreased susceptibility to cephalosporins or azithromycin in the United States, 2014-2016. J. Infect. Dis. 220 (2), 294−305. https://doi.org/10.1093/infdis/jiz079.

Tiberi, S., et al., 2018. Tuberculosis: progress and advances in development of new drugs, treatment regimens, and host-directed therapies. Lancet Infect. Dis. 18 (7), e183−e198. https://doi.org/10.1016/S1473-3099(18)30110-5.

Torres, J.N., et al., 2015. Novel katG mutations causing isoniazid resistance in clinical M. Tuberculosis isolates. Emerg. Microb. Infect. 4 (7). https://doi.org/10.1038/emi.2015.42.

Town, K., et al., 2020. Genomic and phenotypic variability in neisseria gonorrhoeae antimicrobial susceptibility, England. Emerg. Infect. Dis. 26 (3), 505−515. https://doi.org/10.3201/eid2603.190732.

Unemo, M., Shafer, W.M., 2014. Antimicrobial resistance in Neisseria gonorrhoeae in the 21st Century: past, evolution, and future. Clin. Microbiol. Rev. 27 (3), 587−613. https://doi.org/10.1128/CMR.00010-14.

Unemo, M., et al., 2019. Antimicrobial resistance in Neisseria gonorrhoeae and treatment of gonorrhea. In: Neisseria Gonorrhoeae: Methods and Protocols. Springer, New York, NY, pp. 37−58. https://doi.org/10.1007/978-1-4939-9496-0_3. (Accessed 7 October 2023).

Unemo, M., et al., 2021. WHO global antimicrobial resistance surveillance for Neisseria gonorrhoeae 2017−18: a retrospective observational study. Lancet Microbe 2 (11), e627−e636. https://doi.org/10.1016/S2666-5247(21)00171-3.

Van Deun, A., et al., 2015. Disputed *rpo*B mutations can frequently cause important rifampicin resistance among new tuberculosis patients. Int. J. Tubercul. Lung Dis. 19 (2), 185−190. https://doi.org/10.5588/ijtld.14.0651.

Wadsworth, C.B., et al., 2018. Azithromycin resistance through interspecific acquisition of an epistasis-dependent efflux pump component and transcriptional regulator in Neisseria gonorrhoeae. mBio 9 (4). https://doi.org/10.1128/mBio.01419-18.

Walker, T.M., et al., 2022. The 2021 WHO catalogue of complex mutations associated with drug resistance: a genotypic analysis. Lancet 3 (4).

Walsh, T.R., 2005. The emergence and implications of metallo-β-lactamases in Gram-negative bacteria. Clin. Microbiol. Infect. 11 (s6), 2−9. https://doi.org/10.1111/j.1469-0691.2005.01264.x.

Whiley, D.M., et al., 2018. Genetic characterisation of Neisseria gonorrhoeae resistant to both ceftriaxone and azithromycin. Lancet Infect. Dis. 18 (7), 717−718. https://doi.org/10.1016/S1473-3099(18)30340-2.

Wi, T., et al., 2017. Antimicrobial resistance in Neisseria gonorrhoeae: global surveillance and a call for international collaborative action. PLoS Med. 14 (7). https://doi.org/10.1371/journal.pmed.1002344.

Woodford, N., Turton, J.F., Livermore, D.M., 2011. Multiresistant Gram-negative bacteria: the role of high-risk clones in the dissemination of antibiotic resistance. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Rev. 35 (5), 736−755. https://doi.org/10.1111/j.1574-6976.2011.00268.x. Available at:

World Health Organization, 2017. WHO Publishes List of Bacteria for Which New Antibiotics Are Urgently Needed. Available at: https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed.

World Health Organization, 2020. GLASS Whole-Genome Sequencing for Surveillance of Antimicrobial Resistance. Available at: https://apps.who.int/iris/bitstream/handle/10665/334354/9789240011007-eng.pdf.

World Health Organization, 2021. Global Tuberculosis Report 2021.

Wu, W., et al., 2019. NDM metallo-β-lactamases and their bacterial producers in health care settings. Clin. Microbiol. Rev. 32 (2). https://doi.org/10.1128/cmr.00115-18.

Wyres, K.L., Holt, K.E., 2018. *Klebsiella pneumoniae* as a key trafficker of drug resistance genes from environmental to clinically important bacteria. Curr. Opin. Microbiol. 45, 131−139. https://doi.org/10.1016/j.mib.2018.04.004.

Wyres, K.L., Lam, M.M.C., Holt, K.E., 2020. Population genomics of *Klebsiella pneumoniae*. Nat. Rev. Microbiol. 18 (6), 344−359. https://doi.org/10.1038/s41579-019-0315-1.

Xu, L., Sun, X., Ma, X., 2017. Systematic review and meta-analysis of mortality of patients infected with carbapenem-resistant *Klebsiella pneumoniae*. Ann. Clin. Microbiol. Antimicrob. 16 (1). https://doi.org/10.1186/s12941-017-0191-3.

Yasir, M., et al., 2022. Prediction of antimicrobial minimal inhibitory concentrations for Neisseria gonorrhoeae using machine learning models. Saudi J. Biol. Sci. 29 (5), 3687−3693. https://doi.org/10.1016/j.sjbs.2022.02.047.

Yong, D., et al., 2009. Characterization of a new metallo-β-lactamase gene, bla NDM-1, and a novel erythromycin esterase gene carried on a unique genetic structure in *Klebsiella pneumoniae* sequence type 14 from India. Antimicrob. Agents Chemother. 53 (12), 5046−5054. https://doi.org/10.1128/AAC.00774-09.

Zaidi, A.K.M., et al., 2005. Hospital-acquired neonatal infections in developing countries. Lancet 365 (9465), 1175−1188. https://doi.org/10.1016/S0140-6736(05)71881-X.

Zalucki, Y.M., Dhulipala, V., Shafer, W.M., 2012. Dueling regulatory properties of a transcriptional activator (MtrA) and repressor (MtrR) that control efflux pump gene expression in neisseria gonorrhoeae. mBio 3 (6). https://doi.org/10.1128/mBio.00446-12.

Zankari, E., et al., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67 (11), 2640−2644. https://doi.org/10.1093/jac/dks261.

Zankari, E., et al., 2017. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. J. Antimicrob. Chemother. 72 (10), 2764−2768. https://doi.org/10.1093/jac/dkx217.

CHAPTER

Evolution of Resistance to Insecticides in Disease Vectors

13

Philippe Nwane[1,2,3], Achille Binyang[2], Lili Ranaise Mbakop[2] and Josiane Etang[1,4,5]

[1]*Malaria Research Laboratory, Yaoundé Research Institute (IRY), Organization for the Coordination of Endemic Diseases' control in Central Africa (OCEAC), Yaoundé, Cameroon;* [2]*Laboratory of Animal Biology and Physiology, Faculty of Sciences, University of Yaoundé I, Yaoundé, Cameroon;* [3]*Centre de Recherche sur les Filarioses et autres Maladies Tropicales (CRFilMT), Yaoundé, Cameroon;* [4]*Department of Biological Sciences, Faculty of Medicine and Pharmaceutical Sciences, University of Douala, Douala, Cameroon;* [5]*Institute for Insect Biotechnology, Justus Liebig University Gießen, Gießen, Germany*

1. Introduction

Vector-borne diseases (VBDs) contribute substantially about 17% the global burden of infectious diseases and their prevalence tends to increase worldwide (WHO, 2006). Their burden directly impairs the public health and socioeconomic development, especially among the poorest populations living in low and middle income countries in tropical and subtropical regions (WHO, 2015). The per capita mortality rate from VBDs is almost 300 times greater in developing nations than in developed countries (WHO, 2008). Many of these VBDs are co-endemic, and it is estimated that more than half of the world's population live in areas where two or more VBDs are present (WHO, 2006).

Human activities have led to the proliferation of many arthropod vectors of pathogens, such as mosquitoes, ticks, and flies, owing to environmental modifications associated with extensive deforestation, irrigation, and urbanization. Mosquitoes are particularly prone to transfer viruses and parasites from human to animal hosts and vice versa. They are vectors of malaria, arboviruses (dengue, yellow fever, zika, Japanese encephalitis, St. Louis encephalitis, west Nile, and chikungunya), filariasis, and the bacterial disease tularemia. Other major vector-borne diseases (sleeping sickness, leishmaniasis, onchocerciasis, plague, bartonellosis, rickettsioses, Lyme disease, ehrlichiosis, babesiosis, anaplasmosis, chagas disease, and several viral diseases) are transmitted by nonmosquito arthropods such as tsetse flies (*Glossina* sp.), sand flies (Phlebotominae), black flies (Simulidae), houseflies, fleas, lice, cockroaches, and Triatomine bugs. While the climate (temperature, rainfall, and humidity) does influence disease transmission, the expansion of these diseases is mostly due to human factors, such as forest clearing, increased travel, transport, and economical activities (e.g., the geographic distribution of *Aedes albopictus* has considerably increased via worldwide commerce of used tires and because of its capacity of diapausing and the resistance of its eggs to desiccation (Chaturvedi and Shrivastava, 2004)). Overall, socioeconomic activities are among the major determinants of vector-borne diseases' prevalence (Sang et al., 2014). Therefore, some tropical vector-

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00006-9
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

borne diseases are increasingly reported in new geographical areas (e.g., Chikungunya or West Nile virus in Europe and United States, Zika in Brazil). While man-made ecological imbalance induces a reduction in biodiversity in most insect groups, mosquito species are thriving and transmitting diseases as well (Seibold et al., 2019; Johnson et al., 2015).

Yet, the control of vector-borne diseases represents one of the greatest global public health challenges of the 21st century. This ideally entails active case-detection and treatment of human infections, as well as prevention through vaccines, chemoprophylaxis, and vector control interventions. However, these interventions encounter number of challenges in the field. For instance, few vaccines are currently available (e.g., for yellow fever, Japanese encephalitis, malaria) and many pathogens, such as *Plasmodium*, are now resistant to antimicrobial drugs. Moreover, populations from endemic countries struggle to get access to drugs and vaccines, notably due to economic impediments. In addition to the poor conditions of populations in endemic areas, access to health services especially diagnosis and case management, is a serious problem. Thus, in many instances, vector control appears as the only affordable measure to prevent diseases transmitted by vectors. The first documented attempts to control malaria were based on environmental modifications aiming at reducing the number of breeding sites which resulted in a decrease of vector density. Since then, for about 7 decades, the common way of fighting the major disease vectors relies on the use of insecticides (e.g., DDT) (Roberts and Andre, 1994; Hemingway and Ranson, 2000; Beier et al., 2008). In terms of global health, insecticides are the first line weapon against many vector-borne diseases; they are also widely used to increase food and animal productivity (WHO, 2015). Many scientific investigations conducted during the two last decades have shown that the use of synthetic insecticides such as pyrethroids can dramatically reduce the risk of insect-borne diseases. Insecticides, combined with extensive use of drugs, have rapidly led to the eradication of many diseases (e.g., malaria) from most nontropical areas in the world, but despite initial successes, eradication has proven more elusive in the tropics (Dialynas et al., 2009).

Furthermore, vector resistance to insecticides and parasite resistance to drugs make increasingly difficult the elimination of many vector-borne diseases. Resistance to more than one class of synthetic insecticides is now widespread among pests of public health importance, and it is considered the most important impediment in the successful control of vector-borne diseases.

Insecticide resistance is defined as a heritable change in the sensitivity of a pest population that is reflected in the repeated failure of a product to achieve the expected level of control when used according to the label recommendation for that pest species (Nauen, 2007). This phenomenon is due to the existence of genes in a vector species that confer resistance characters, which subsequently increase in frequency as a response to insecticide selection pressure. The use of pesticides, especially insecticides, is very common in livestock, agriculture, and public health. Among all categories of pests, insects are known to exhibit resistance at alarming rates. Since the 1940s, the number of insect species with reported insecticide resistance has been rapidly increasing and passing 580 species in the 1908s (Wood, 1981). The first recorded attempt of insect pest control, the application of tobacco juice against sheep scabs, is found in the literature of the 18th century (Wood, 1981). The first case of resistance was reported in 1908, in a population of San Jose scale (*Aspidiotus perniciosus*) resistant to lime sulfur (Melander, 1914). A century later (2007), 553 arthropod species were reported resistant to at least one insecticide, among many disease vectors. More than 100 mosquito species are resistant to at least one insecticide (including 56 Anopheline species, 39 Culicine species); for instance, *Culex pipiens pipiens* and *An. albimanus* were found resistant to more than 30 different compounds (Whalon et al., 2008). Mechanisms allowing insect survival to insecticide exposure, also called insecticide

resistance mechanisms, have been selected in many species of insect vectors. In general, insects may use four main mechanisms to overcome insecticide effects: (1) naturally detoxify or destroy the toxin faster than susceptible insects, or quickly rid their bodies of the toxic molecules (metabolic resistance); (2) have a genetically modified target site that reduce the binding of the toxin and preventing its activity (altered target-site resistance), (3) absorb slowly the toxin (e.g., penetration resistance due to increased cuticle thickness), and (4) detect or recognize a hazard and avoid the toxin (behavioral resistance).

Few examples of behavioral (e.g., *Anopheles gambiae* on Bioko Island and Senegal (Reddy et al., 2011; Ndiath et al., 2014) or *An. funestus* in Benin and Tanzania (Moiroux et al., 2012; Russell et al., 2011; Sougoufara et al., 2014) and penetration resistances have been reported; whether they are heritable remains debated, and it is difficult to assess the level of resistance they provide. Metabolic and target site resistances typically result in relatively high level of insect survival and are genetically determined. Resistant individuals carry one or several genetic mutations that prevent insecticide disruption of the target functioning. As a result, the frequency of resistance gene(s)/allele (s) increases in the population over time. Insecticide resistance is confirmed by toxicological tests (bioassays) establishing resistance ratio (or RR corresponding to the number by which an insecticide dose must be multiplied in order to obtain the same mortality in resistant than in susceptible insects). It can be investigated at different levels, from the molecular characterization of genes/alleles conferring resistance and their biochemical products, to the effects of these genes on the fitness (i.e., mean reproductive success) of the individuals carrying resistance alleles, to the dynamics and evolution of these resistance alleles in natural vector populations and their effects on disease control.

2. Insecticide Classification and Modes of Action

Insecticides refer to a subgroup of compounds within the group called "pesticides" which are capable of controlling, attracting, repelling, destroying, or opposing the development of living organisms considered undesirable for agriculture, public health, or disease vectors. Indeed, the word "pesticides" refers to various types of compounds including insecticides, herbicides, bactericides, nematicides, acaricides, fungicides, molluscicides, and rodenticides; each is active against specific pests (e.g., insect, weed, bacteria, nematode, fungi, snail, and rat). For arthropods of medical and veterinary importance, control measures rely mainly on the use of insecticides which can be either natural or synthetic. However, application of natural insecticides either in agriculture or in public health sectors as alternative to synthetic insecticides has been increasing since the emergence of the resistance threat.

2.1 Natural Insecticides

Natural insecticides derived from compounds that are produced through natural processes; they include botanical insecticides (e.g., pyrethrins and azadirachtin), microbial insecticides (e.g., *Bacillus thuringiensis*, *Lecanicillium muscarium*, abamectin, and spinosad), mineral oils, and minerals. These compounds tend to break down quickly in the environment, reducing the risk of harm to nontarget organisms; however, they may require repeated applications to effectively control the target pests.

Among natural insecticides, one of the oldest and most widely used botanical insecticide is Pyrethrum obtained from the flowers of the Chrysanthemum family. It has been known for a long period, the very high toxicity of the active principles to insects and their harmlessness to man and warm-

blooded animals forming an almost unique combination of qualities. Pyrethrins derived from Pyrethrum and serve as efficient insecticides, particularly against flying insects (Casida, 1970). Pyrethrins are esters, with the acid moiety derived from the terpene pathway and the alcohol moiety derived from jasmonic acid (Casida, 1970; Rivera et al., 2001; Matsuda et al., 2005). They fall into two groups. The first, called type I pyrethrins, include jasmolin I, pyrethrin I, and cinerin I, and each contain the monoterpene acid trans-chrysanthemic acid, bound to one of three alcohols-respectively, jasmolone, pyrethrolone, and cinerolone (Casida, 1973). Type I pyrethrins generally comprise 50% of the total pyrethrins in the flower (Casida, 1973). A second class of pyrethrins, called type II, are esters that contain one of the same three alcohols, but linked to pyrethric acid (also known aschrysanthemum acid) instead of to chrysanthemic acid, and are called, respectively, jasmolin II, pyrethrin II, and cinerin II (Casida, 1973).

A well-known botanical insecticide is azadirachtin, derived from the seed produced by the neem tree; it is a feeding deterrent and an insect growth regulator that suppresses fecundity, moulting, pupation, and adult emergence (Mordue and Blackwell, 1993; Bezzar-Bendjazia et al., 2016). Compounds that selectively act on the insect nicotinic acetylcholine receptor (neonicotinoids), such as imidacloprid, acetamiprid and thiomethaxam are among the modern insecticides used in pest management (Matsuda et al., 2001; Tan et al., 2007). Avermectins, the insecticides of microbial origin, target GABA-gated chloride channels (Albrecht and Sherman, 1987; Lasota and Dybas, 1991), whereas the diamide insecticide acts on the ryanodine receptor (Cordova et al., 2006; Lahm et al., 2009). Spinosyns, a new class of insecticides derived from actinomycetes, show high selectivity and low mammalian toxicity with eco-friendly behavior (Thompson et al., 2000; Sparks et al., 2001).

2.2 Synthetic Insecticides

Synthetic pesticides are compounds produced through an industrial process; pyrethroids are synthetic derivatives of natural pyrethrins. Originally, only inorganic insecticides (such as lime sulfur) and natural products were available.

Today, four classes of organic (synthetic) insecticides are essentially used: the organochlorines (OCs), the organophosphates (OPs), the carbamates (CXs), and the pyrethroids (PYRs), with, respectively, 4429, 1375, 30, and 414 metric tonnes of active ingredient used annually for global vector control from 2000 to 2009 (WHO, 2011). The first synthetic insecticides, introduced during World War II for malaria control, belonged to the OC class. These include dichlorodiphenyltrichloroethane or DDT (introduced in 1943), which targets the voltage-gated sodium channels (Na-channels) and cyclodiene (CD) dieldrin, which targets the g-inobutyric acid (GABA) receptor; both targets being essential in the insect nervous system. In addition to their public health applications, enormous tonnages of DDT and dieldrin were used worldwide in agriculture. It was at first a great success with large WHO-led campaigns leading to reduction of morbidity and mortality from malaria in many endemic regions after World War II. Widely acclaimed, DDT and dieldrin rapidly selected resistance in insect vectors. In *An. gambiae*, resistance to DDT was first noted 11 years after its introduction (WHO, 1957), while a population from northern Nigeria was reported resistant to dieldrin soon after (Davidson, 1956). DDT resistance has now been reported in mosquitoes (*Aedes* sp., *Anopheles* sp., and *Culex* sp.), houseflies, sand flies, body lice, and head lice, while resistance to dieldrin (60% of reported cases of resistance before 1990) has been detected in more than 277 arthropods, including mosquitoes (*Aedes*

sp., *Anopheles* sp., and *Culex* sp.), fleas, ticks, biting flies, bedbugs, cockroaches, and human lice (Oakeshott et al., 2005; Feyereisen et al., 2015).

An important issue against these insecticides is their environmental impact. Rachel Carson's book "Silent Spring" (Carson, 1962) was a seminal work publicizing and politicizing the toxic effects of the accumulation of DDT and its metabolites in the food chain. In vertebrates, DDT can interfere with reproduction, and in humans it can have neurologic, carcinogenic, and reproductive effects, although the evidence remains debated. These insecticides are also extremely stable in the environment, contaminating groundwater and remaining in soil long after their use. In the 1970s, the Persistent Organic Pollution Treaty led to total banning of dieldrin and to the banning of DDT for all uses except malaria control when this disease is very frequent and there is no alternative. DDT use rapidly declined in the 1970s (it is no longer used in Latin America), but it gained new advocates due to the development of resistance to the alternative insecticides, and to its low cost (van den Berg, 2009; Coleman et al., 2008). Consequently, its use quadrupled between 2007 and 2009 (WHO, 2011).

From the late 1970s, OCs were replaced by the PYRs class for vector control, and these became widely used in agriculture and public health, and more particularly against malaria vector. They are today by far the most-used insecticides, with 81% of the world spray coverage (WHO, 2011). As DDT, these insecticides target the Na-channels (i.e., neurotoxic effect). Their rapid popularity comes from their very low toxicity to human, their rapid knock-down (KD) effect associated with an excito-repellent effect. PYR-based indoor residual spraying (IRS) and insecticide-treated nets and curtains (ITNs) are currently advocated as standard malaria vector control strategies (WHO, 2006). PYR resistance was reported in 1993, in *An. gambiae* populations from Cote d'Ivoire (Elissa et al., 1993) and later in *C. pipiens quinquefasciatus* also in West Africa (Chandre et al., 1998). Resistance is now widespread in mosquitoes (*Aedes* sp., *Anopheles* sp., and *Culex* sp., body and head lice, ticks (e.g., *Boophilus microplus*), and fleas (WHO, 2006; Roberts and Andre, 1994; Hemingway and Ranson, 2000). As PYR resistance developed, many control programs attempted to revert to DDT for disease control. However, these insecticides share a common target site, and there is cross-resistance to both insecticide classes in many locations (Coleman et al., 2008). Finally, two other classes of synthetic insecticides are used at a large scale worldwide: the OPs and the CXs, which were first used in the 1940s and the 1950s, respectively (WHO, 2006; Nauen, 2007). OPs and CXs target the synaptic acetylcholinesterase (AChE), an essential enzyme in the nervous system. They are usually used as larvicides (although some are now considered for ITN impregnation and IRS as an alternative to PYR (Oxborough et al., 2008)) and are particularly well suited for species with delimited breeding sites. However, they have a short half-life, and two to three rounds of IRS are needed per year. This, combined in some instances with their high price, can make these insecticides too costly for most malaria control programs (Coleman et al., 2008). Early resistance to these insecticides has been detected shortly after their first application: for example, first OP treatments in the Montpellier area (southern France) started in 1969, the first resistance being detected only 3 years later (Pasteur and Sinegre, 1975). Resistance has now been recorded in mosquitoes (*Aedes* sp., *Anopheles* sp., and *Culex* sp.), biting flies (e.g., *Simulium damnosum*, vector of onchocerciasis), sand flies, houseflies, and fleas.

Indeed, most often PYRs are available, essentially for economic cost reasons: the most recent PYR was introduced in mid-1980s and new synthetic insecticide are under investigations since mid-1990s. The shrinking availability of insecticides because of resistance is exacerbated by the removal from the market of insecticides that are no longer registered for public health use: some compounds are too costly, and insecticide use is restricted by regulatory agencies, due to environmental concerns.

Consequently, new environment-proof products (high selectivity, no effects on nontargets) are now required for sustainable vector control (Farenhorst et al., 2009).

2.3 Alternative Insecticides

Environmental pollution concerns and unresolved issues pertaining to the toxicity of synthetic insecticides to humans and nontarget species have led the public and researchers' interest to investigate alternative "biological" insecticides (Mittal, 2003). Three main types of these alternative insecticides are documented: (1) bacterial toxins, (2) essential oils, and (3) fungi. There are two main sources of bacterial toxins: *Bacillus sphaericus* (Bs) and *Bacillus thuringiensis* (Bt). They kill insect larvae by producing proteic toxins binding to various receptors on midgut epithelial cells. Bs toxicity is due to a binary toxin, whereas Bt toxicity is due to the interaction of four different toxins. These larvicides are presented as highly specific and effective at low doses and are thus expected to be safe for the environment. Toxins extracted from Bs and a variety of Bt (*B. thuringiensis* var *israelensis* or Bti) are used for mosquito control.

In these species, bacterial toxins show some differences in specificity: Bti is more effective against *Aedes* and *Culex* species than against *Anopheles*; whereas Bs is more effective against *Culex* than *Anopheles* species and has no effect on *Aedes* species that lack receptors. While the presence of several toxins was expected to delay resistance apparition, Bs and Bti resistances have been detected in various mosquitoes (Mittal, 2003; Paul et al., 2005), and resistance to Bt has been detected in several agricultural pests. Although less documented, essential oils are investigated as potential biological larvicides. They are advocated to be more specific than synthetic insecticides, and biodegradable, thus with reduced impact on the environment. Variable efficacies seem to represent a restraint for pest control; identifying the bioactive components instead of raw products could be the solution to this problem.

Finally, fungi can be used as biological insecticides: they target the adult stage of mosquitoes and are used essentially for malaria control. The fungus *Metarhizium anisopliae* has been shown to reduce *An. gambiae* adult life span in the laboratory and in the field in Tanzania (Scholte et al., 2005), while *Beauveria bassiana* decreases the survival of another malaria vector, *Anopheles stephensi* (Blanford et al., 2005). These agents have several advantages: they are cheap, easily stored for long term, and specific to insects. These fungal insecticides have a direct effect on *Plasmodium* transmission and are expected to decrease malaria prevalence. Finally, their acting late in life is considered by several authors to be an important advantage, as it will decrease selective pressure and reduce the risk of resistance development (potentially "evolution-proof" insecticides (Farenhorst et al., 2009; Michalakis and Renaud, 2009; Read et al., 2009)).

2.4 New Molecules Used in Vector Control

During 2006–2008s, few new insecticides were described: neonicotinoids, phthalic acid diamides, or anthranilic acid diamides; however, they were used mostly against agricultural pests and not against disease vectors (Nauen, 2007; Whalon et al., 2008; Weill et al., 2004).

Also, another type of synthetic insecticides is growth regulators (GR). It regroups synthetic products called juvenoids that mimic the juvenile hormone (JH) and chitine inhibitors. So far, only few cases of resistance have been reported in houseflies and mosquitoes (e.g., resistance to methoprene, a

JH analog in the mosquito *Ochlerotatus nigromaculi*) (Cornel et al., 2002). The new molecules currently prequalified for vector control are clothianidin, which belongs to the neonicotinoid family, pyriproxifen, which belongs to the growth regulator family, and chlorfenapyr, which belongs to the pyrroles. These insecticides are used either alone or in combination with other insecticides already registered.

2.4.1 Neonicotinoids

Neonicotinoids are synthetic nicotine analogs introduced to the market between 1991 and 2002 with the initial goal of combating insect pest resistance against previous classes of insecticides (organophosphates, carbamates, and pyrethroids) (Simon-Delso et al., 2015). Their mode of action selectively targets nicotinic cholinergic receptors (nAChR). Many insect pest species have not yet developed resistance against neonicotinoids, giving this class of insecticides an advantage in efficacy. In vector control for the prevention of malaria, clothianidin is used as an intradomiciliary spray sometimes a combination with deltamethrin (pyrethrinoid) (WHO, 2019).

2.4.2 Growth Regulators: Pyriproxyfen

Growth regulators are analogs of juvenile insect hormones. They are juvenoids (methoprene, pyriproxyfen) and ecdysoids (diflubenzuron). They prevent the normal development of larvae, either by inhibiting pupation or by blocking the synthesis of chitin. They also act on adult mosquitoes by inhibiting ovarian development. The action of pyriproxyfen causes a profound imbalance in the hormonal system of the insect that results in inhibition of the development of morphogenesis, embryogenesis, and significant decreases in fecundity and fertility of adults (Ishaaya and Horowitz, 2015). This compound has very low toxicity to humans and mammals (Najera and Zaim, 2002). Susceptibility tests performed on different strains of mosquitoes have so far not revealed any cross-resistance with pyrethroid, organophosphate and carbamate insecticides. It has been prequalified by WHO as a larvicide and adulticide (combination with conventional insecticides) for use in vector control (WHO, 2019).

2.4.3 3 Pyrroles: Chlorfenapyr

Pyrroles are a class of molecules found in various natural products as well as in the chemical composition of some drugs because of their biological properties. Chlorfenapyr belongs to this class of insecticides. It is toxic to insect pollinators such as honeybees and other beneficial arthropods (predatory mites and parasitoid wasps in particular), but it has low toxicity on mammals. It acts by preventing the synthesis of Adenosine Triphosphate (ATP) in the mitochondria (Anonymous, 1995) by uncoupling oxidative phosphorylation. It prevents the transformation of Adenosine Diphosphate (ADP) into ATP. Chlorfenapyr thus acts on the respiratory system by uncoupling the phosphorylation-oxidation function of the mitochondria in the cell. This leads to the disruption of ATP production and consequently to a decrease in energy, leading to a dysfunction of the cell, causing its death. It facilitates the loss of proton from inside the mitochondria to the rest of the body thus preventing the formation of ATP source of energy. This group of insecticides has been used in the past to control several species of insect and mite pests such as ants, termites, cereal crop protection, and garden flower insects (Guglielmone et al., 2000). Studies conducted on adult malaria vector species of the *An. gambiae* complex using chlorfenapyr-impregnated nets showed delayed (48−72 hours) but very high mortality (N'Guessan, 2009). Furthermore, there is no cross resistance with conventional insecticides such as

pyrethroids, carbamates and organophosphates (N'Guessan et al., 2007). Chlorfenapyr has been prequalified by the WHO for use in combination with pyrethroids in the impregnation of mosquito nets for vector control.

3. Mechanisms of Resistance

The targets of most insecticides are critical proteins of the insect nervous system. Insecticides bind to specific sites on their targets and disrupt their function. Any mechanism that decreases the insecticide effect will lead to resistance. This encompasses reduced penetration of the insecticide, increased excretion or sequestration of the insecticide, increased metabolism of the insecticide, and finally target modification limits the binding of the insecticide. However, a behavioral change resulting in a reduced exposure to the insecticide can also be viewed as a resistance mechanism, if it is heritable: for example, Anopheline mosquitoes have been reported to have changed their blood-feedings habits, by seeking hosts outdoor (exophily and exophagy) rather than indoor (endophily and endophagy); however, whether this behavior is heritable remains debated. The first three mechanisms are poorly documented and do not seem to play a prominent role in resistance (Nkya et al., 2013). Most studies aiming at understanding the mechanisms and the genetic bases of insecticide resistance focus on metabolic resistance and target site modifications. In this chapter, we present the various documented mechanisms of resistance. We specifically focus on disease vector species, although many mechanisms are common to agricultural pests. We insist on the evolutionary aspects of resistance, while the detailed mechanisms are treated more succinctly, and only for the major ones. More comprehensive reviews can be found (Feyereisen et al., 2015; Nkya et al., 2013; David et al., 2013). Moreover, the recent explosion of genomic studies on resistance frustrates any pretention to exhaustiveness.

3.1 Metabolic Resistance

Metabolic resistance regroups various mechanisms that lead to the degradation of the insecticides in less- or nontoxic products, thus decreasing the quantity of toxic molecules that reach the target. These so-called "detoxification enzymes" belong mainly to three large gene families, cytochrome P450 monooxygenases (P450s or CYPs for genes), glutathione S-transferases (GSTs), and carboxylesterases (COEs), and most studies focus on a small set of genes. Genomic studies can, however, access mechanisms that had previously proven intractable. They allow deeper description of known resistance gene families and help find new candidate genes. They have suggested that other enzyme families may be implicated, such as UDP-glycosyl-transferases (UGTs), sulfotransferases, aldehyde dehydrogenases, NADH-cytochrome b reductases, NADH dehydrogenases, NADH-ubiquinone oxidoreductases, nitrilase thioredoxin peroxidases, and cuticular genes (Nauen, 2007; Hemingway et al., 2004; Oakeshott et al., 2003; Li et al., 2007). However, in most cases the causal role of the candidates remains to be formally validated. Detoxification enzymes are frequently divided into phase I and phase II enzymes depending on their role in detoxification pathways with hydrolases and oxidases acting during phase I, and transferases acting during phase II (Faucon et al., 2015). These enzymes can act individually, synergically, or sequentially through complex insecticide degradation pathways. Such complexity is accentuated by the redundancy of insect detoxification systems. A given detoxification enzyme may indeed metabolize different insecticides (although with different kinetic parameters),

thus contributing to cross-resistance. On the other hand, different enzymes may degrade the same insecticide, and contribute in an additive manner to the resistance phenotype. In natural populations, several metabolic mechanisms can be present in the same species (Strode et al., 2008), and metabolic resistance often combines with target-site modifications leading to high-resistance levels and complex cross-resistance patterns.

3.1.1 Glutathione S-Transferases

Various xenobiotics contain the tripeptide glutathione; GSTs catalyze the reaction of the sulfhydryl group of this tripeptide. This sulfhydryl group reacts with electrophilic sites on xenobiotics, leading to formation of conjugates that are more readily excreted and typically less toxic than the parent insecticide. In addition to this direct detoxification, GSTs play a role in phase II detoxification (see later). GST enzymes are present in most insects. They represent a large family of generalist detoxifying enzymes (six classes of GSTs have been identified in the genome of *An. gambiae*) and have thus broad substrate specificities. The GST family expands either by alternative splicing or by local gene duplication, the last leading to clusters of GST genes. GSTs are primarily associated with resistance to OCs, particularly DDT, and OPs. GST-based resistance seems to be associated with an increased amount of enzyme resulting from gene duplication or, more often, upregulation. A constitutive GST overexpression was frequently reported in mosquito populations showing elevated resistance level to DDT (Nkya et al., 2013; Strode et al., 2008; Enayati et al., 2005). Quantitative genetic analyses identified a quantitative trait locus (QTL) for resistance to DDT in *An. gambiae*, within which there is a cluster of eight GSTs (Ranson et al., 2002). Among them, GSTE2 was then shown to metabolize DDT (Enayati et al., 2005), GSTE2 ortholog in *Ae. aegypti* and *An. funestus* was further shown to metabolize DDT (Lumjuan et al., 2005; Riveron et al., 2014), GSTs are also suspected to play a role in the resistance to PYRs in mosquitoes through sequestration. Lumjuan and colleagues (Lumjuan et al., 2011) showed that the partial KD of *Ae. Aegypti* GSTE2 and GSTE7 led to an increased susceptibility to the PYR deltamethrin. Similarly, Riveron and colleagues (Riveron et al., 2014) show that GSTE2 contributes to PYR resistance in *An. funestus* probably through sequestration.

3.1.2 Cytochrome P450 Monooxygenases

Cytochrome P450 monooxygenases (P450) are heme-thiolate enzymes found in all living organisms (Werck-Reichhart and Feyereisen, 2000). They are best known for their monooxygenase activity, but they can catalyze a wide range of reactions. In insects, P450s are associated with the metabolism of endogenous compounds, such as hormones, and are involved in the phase I detoxification of a variety of xenobiotics including plant toxins, pollutants, and chemical insecticides (David et al., 2013; Feyereisen, 2005; Despres et al., 2007). P450s are frequently represented by more than 100 genes in insect genomes, so that the identification of those involved in insecticide resistance is challenging. Some of them are inducible by xenobiotics and expressed at higher level in classical detoxification tissues (midgut, fat bodies, Malpighian tubules), although such properties do not ensure their actual contribution to insecticide resistance. Insecticide resistance is often linked to the overexpression of one or multiple P450s through upregulation or gene amplification, although mutations may also lead to resistance. P450s have been reported as responsible for resistance to most insecticide classes, particularly DDT, PYRs, and CXs. In addition, some P450s are also capable of activating particular OPs, such as malathion and diazinon (i.e., they become toxic when oxidized). The contribution of P450s in insecticide resistance can be estimated by combining the exposure of insects to the P450

inhibitor piperonyl butoxide (PBO) and subsequent bioassays with insecticides: if P450s are implicated, the resistance level is usually decreased in the presence of PBO. However, PBO does not equally inhibit all P450s, so that absence of PBO-induced resistance decrease does not mean that no P450 is implicated. The role of P450s in resistance may also be evidenced by biochemical assays measuring either the global heme content or more specific enzyme activities using known P450 substrates, such as ethoxy coumarin (ECOD method) or resorufin (EROD method). However, biochemical assays are not always capable of detecting P450-based resistance, because these assays have a low specificity, unlike some P450s. Following the sequencing of mosquito genomes and the development of microarrays (Strode et al., 2008), transcriptomics has been intensively used for detecting over transcribed P450s in resistant populations, leading to the identification of several CYP genes associated with resistance in mosquitoes and other insects (Feyereisen et al., 2015; Nkya et al., 2013; Ffrench-Constant et al., 2004; Vontas et al., 2012). In mosquitoes, some of them were validated as capable of contributing to insecticide metabolism by functional approaches, such as heterologous expression followed by in vitro insecticide metabolism, RNA interference, or transgenic expression. These include the Anopheles genes CYP6Z1, CYP6M2, CYP6P3, CYP6P9, CYP6P4, CYP6P7, CYP6AA3 81; Aedes genes CYP9J32, CYP9J24, CYP9J28, CYP6BB2 (Chiu et al., 2008; Ibrahim et al., 2016; Stevenson et al., 2012; Kasai et al., 2014); and the *Culex* gene CYP9M10. Interestingly, it was shown that Anopheles CYP6M2 and CYP6P3 can metabolize insecticides from different classes, supporting the role of P450s in cross-resistance, and raising concerns for insecticide-resistance management (Mitchell et al., 2012; Edi et al., 2014). Although gene expression studies have identified multiple P450s overexpressed in resistant insects, very few data are available on the genetic factors controlling their overexpression. Recently, the use of targeted deep sequencing allowed the identification of gene amplifications controlling the overproduction of P450s in multiple PYR-resistant population of *Ae*. *aegypti* worldwide (Faucon et al., 2015). High-throughput sequencing approaches also allowed identifying non-synonymous variations affecting P450s potentially linked to insecticide detoxification (Faucon et al., 2015).

3.1.3 Carboxylesterases

More than 30 genes coding COEs are found in insects (Oakeshott et al., 2005; Edi et al., 2014). Most COEs are serine esterases, that is, they have a serine residue within a catalytic triad necessary for hydrolysis. COEs bind to an ester group and then break the ester bound by a process of acylation diacylation. Multiple forms of COEs are found in insects, with broad and overlapping substrate specificities. Most insecticides, including almost all CXs and OPs, most PYRs, and some GRs bear ester groups. In most cases, hydrolysis of the ester group leads to a reduced toxicity of the insecticide. Consequently, COEs are often involved in metabolic resistance mechanisms, although the level of resistance conferred is relatively low (w10_) compared to target-site resistance. As for P450s, the role of COEs in resistance is usually diagnosed by the addition of a synergist, the S,S,S-tributyl phosphoro trithioate (DEF) to bioassays. DEF inhibits COEs (but also GSTs): if COEs contribute to resistance, insecticide toxicity is expected to increase in the presence of DEF, significantly more in resistant than in susceptible insects (Hollingworth and Dong, 2008). COE-based resistance has been detected in various species, mainly against OPs and to a lesser extent to PYRs (Nauen, 2007). OP resistance in Culex mosquitoes is generally caused by an elevated COE protein quantity, up to 80 times the level found in susceptible individuals (Pasteur et al., 1984). Two esterases, a-esterase (or esterase A) and b-esterase (or esterase B), have been recognized based on their higher affinity for, respectively, a- and b-

naphthylacetate (Mouches et al., 1987). Their overexpression is usually caused by an increased gene copy number (gene amplification) of one or both esterases, although upregulation may also contribute to overexpression (Georghiou and Pasteur, 1978; Mouches et al., 1986). The loci coding for the esterases A and B behave as a single locus named Ester (Vaughan et al., 1997). The number of gene copies within an amplification of the Ester locus can vary greatly, potentially in relation with the intensity of insecticide treatments (Hollingworth and Dong, 2008; Raymond et al., 1998; Guillemaud et al., 1999). Amplified esterases have also been described in the mosquitoes *An. gambiae* and *Ae. aegypti* in association with resistance to the OP temephos (Weill et al., 2000a,b). Orthologs of these genes were also found amplified in association with temephos resistance in the tiger mosquito *Ae. Albopictus* (Poupardin et al., 2014). Biochemical assays also pointed out the role of esterases in PYR hydrolysis in mosquitoes (Pocquet et al., 2014), although no particular esterase has yet been validated as able to hydrolyze PYRs. Moreover, it appears that the PYR metabolites produced by esterases could be further metabolized by overexpressed P450s of the subfamily CYP6Z in PYR-resistant populations, suggesting synergy between these two resistance mechanisms (Grigoraki et al., 2015). Because overexpressed COEs can represent a large percentage of the total protein of the insect (up to 12% of the soluble proteins in some resistant mosquitoes) (Chandor-Proust et al., 2013), it is difficult to disentangle their sequestration effect (i.e., binding to the insecticide without hydrolysis) from the direct hydrolysis of the insecticide. This appears to depend on the species and the esterase allele: hydrolysis appears predominant in the aphid E4 esterase, while in mosquitoes the EsterB1 and Ester2 alleles rather sequester the insecticide and show a lower hydrolysis activity (Somwang et al., 2011; Fournier et al., 1987). However, qualitative changes affecting COEs may also be responsible for resistance to insecticides. For example, resistance to the OP malathion in Anophelinae, *Musca domestica* and *Lucilia cuprina*, was associated with point mutations inducing a faster hydrolysis (Oakeshott et al., 2005; Hemingway et al., 2004; Feyereisen, 1995). In terms of population genetics, COE resistance to OPs in *C. pipiens* is probably one of the best-studied cases. In this species, resistance to OPs was monitored since late 1960s in the Montpellier area of Southern France (Claudianos et al., 1999; Pasteur et al., 1984). This long-term monitoring showed that several Ester-resistance alleles have been replacing each other across time: Ester1 was the first detected resistance allele in 1972, then Ester4 in 1986, and finally Ester2 arrived by migration in 1991. These alleles were selected in insecticide-treated areas, and also showed a fitness disadvantage or cost in absence of insecticide (e.g., lower mating success, lower survival) (Labbé et al., 2009; Chevillon et al., 1997). The quantification of their fitness cost showed that the various alleles correspond to different fitness trade-offs: Ester4 was first favored over Ester1 because of a lower cost (selection for a generalist allele). Then Ester2 appeared to be replacing the first two alleles because it confers a higher resistance level, despite its relatively high cost (selection for a specialist allele) (Pasteur et al., 1984). Overall, this example confirms that insecticide resistance is a dynamic process, as new haplotypes can be selected for adjusting the resistance phenotype and the fitness of resistant individuals to insecticide pressures and environmental factors.

3.2 Target-Site Modification

Resistance by target-site modification is due to point mutations in the insecticide target genes that result in reduced binding of insecticides, rather than to a change in expression level. Because most insecticide targets are vital molecules, there is generally only a limited number of mutations in the target able to decrease insecticide affinity without impeding its original function to an unsustainable

degree. A mutation conferring resistance while partly impairing the target's normal function leads to a fitness cost.

3.2.1 GABA Receptors

GABA is a major neurotransmitter in the insect's central and peripheral nervous system and in neuromuscular junctions. The GABA receptors are linked to chlorine-gated channels, causing hyperpolarization that blocks the nervous influx. GABA receptors are the target of cyclodiennes (CDs). CDs are noncompetitive inhibitors that bind to a site on the receptor close to the chlorine-gated channel, stabilizing it in an inactive closed state. This induces an overexcitation by removal of the inhibition and leads to convulsions and death of the insect. GABA receptors have also secondary-binding sites for some PYRs or insecticides of the avermectin family. Resistance to CDs is due to a decreased sensitivity to insecticide of the GABA receptor A, through a point mutation causing an amino acid change in the receptor coding gene. This gene, called Rdl (resistance to dieldrin, the most-used CD), has been first cloned in *Drosophila melanogaster*. In all *D. melanogaster* resistant individuals, the Rdl locus displays a similar mutation at position 302 in the channel-lining domain sequence, changing an alanine into a serine (A302S). The role of this mutation in CD resistance was confirmed by directed mutagenesis. The serine residue occupies the insecticide-binding site of the GABA receptor and destabilizes its conformation (Duron et al., 2006). The resistance allele (RdlR) is semidominant and can confer cross-resistance to other insecticides, such as fipronil (Hemingway et al., 2004; Duron et al., 2006). Due to an extensive use of CDs before their banning in the 1980s, resistance has been selected in several insect species, which all display a mutation at the same position (A302S or A302G) (Hemingway et al., 2004; Duron et al., 2006). Whether these mutations are costly depends on species: a fitness cost associated with resistance has been identified in *L. cuprina* (Ffrench-Constant et al., 2000) and has been suggested in *C. pipiens* and *An. albopictus* (McKenzie, 1996; Tantely et al., 2010), but no cost has been found in *D. melanogaster* (Ffrench-Constant et al., 2000), even if resistance affects temperature sensitivity. The Rdl locus has been found duplicated in the greenbug Myzus *persicae* (Pocquet et al., 2013) and in a strain of *D. melanogaster* (Anthony et al., 1998). In the latter, a tandem duplication of 113 kb associates a susceptible and a resistance copy of the locus. The phenotype associated to this duplication was shown to be close to that of a standard heterozygote, namely an intermediate resistance level and a reduced heat shock recovery time (Anthony et al., 1998).

3.2.2 Voltage-Gated Sodium Channels

Nerve action potentials are transmitted by a wave of depolarization along the neural axon. They are due to the movement of sodium ions (Naþ) crossing the axonal membrane through the opening of voltage-gated sodium channels (VGSCs) and stop when these channels are inactivated. VGSCs are glycoproteins with a pore for ion transport and can adopt three different states: resting, open, or inactivated; the Naþ ions pass only when the channels are opened. VGSC are the targets of DDT and PYRs. When these insecticides bind to the VGSC, they slow their closing speed, prolonging the depolarization (Remnant et al., 2013; Lund, 1984; Soderlund and Knipple, 2003; Vais et al., 2001). The intensity of the effect is dose-dependent, proportional to the number of Na-channels inactivated (Remnant et al., 2013). For PYRs, the magnitude of the effect depends on the type of insecticide molecules, type I (e.g., permethrin) or type II (e.g., lambda-cyhalothrin and deltamethrin), which, respectively, lack or not a cyano-group. During action potential, type II PYRs lengthen the sodium flux more than type I, and thus usually display a more intense effect (Soderlund and Knipple, 2003). At the

phenotypic level, inactivation of VSGC results in a rapid KD effect, the insect being incapacitated for some time, followed by recovery or death, depending on the species and development stages (in mosquitoes, the adults tend to recover, while larvae will drown). One major mechanism, named knockdown resistance (*kdr*), is responsible for PYR and DDT resistance, by reducing the receptors sensitivity (binding capacity) to these insecticides and modifying the action potential of the channel (Soderlund and Knipple, 2003; Vais et al., 2001). First discovered in *M. domestica*, this mechanism has been described in many agricultural pests and vectors. This resistance mechanism has several consequences: it decreases the irritant and the repellent effects, and either cancels or reduces the KD effect (Dong et al., 2014). Extension mutations affecting the VGSC gene are called *kdr* mutations. By sequencing the VGSC protein (>2000 amino acids), the first two *kdr* mutations were identified in *M. domestica*, both in the second protein domain. The first one (L1014F) is associated with moderate (10e30) PYR resistance; the second (M918T, also called *super-kdr*) is always associated with the L1014F and confers a higher resistance (up to 500) (Chandre et al., 2000). Substitution of the L1014 is found in a large variety of species (L1014F or L1014S, and also L1014H in *Heliothis virescens*) and corresponds to the kdrR alleles (Soderlund and Knipple, 2003; Vais et al., 2001; Chandre et al., 2000; Williamson et al., 1996). The phenotype conferred by *kdr*R is recessive or semirecessive (Dong et al., 2014), with higher resistance to type I than type II PYRs (Etang et al., 2009). However, the various mutations show some specificity, as L1014F confers a high resistance to both DDT and permethrin (PYR), while L1014S confers a lower resistance to permethrin than to DDT (Chandre et al., 1999; Martinez-Torres et al., 1999). Other *kdr* mutations (about 30 in total) have been described in various species, including *super-kdr* mutations (Soderlund and Knipple, 2003; Vais et al., 2001). Some of these mutations are conserved over a large array of organisms, while others are more specific and unique. In *Ae. aegypti*, the *kdr* phenotype has been observed, but it appears that a codon bias prevents the appearance of any L1014 mutation (Ranson et al., 2000). However, several other mutations have been observed associated with resistance in *Ae. aegypti* (e.g., V1023G/I, I1018 M/V, F1565C, D1794Y, or S996P) (Brengues et al., 2003). In *Ae. albopictus*, the F1565C mutation has been observed, while no mutation has been found at the 1018 site (Kasai et al., 2014). The importance of these various mutations in the different resistance phenotypes is thus still in debate. The role of the L1014 F/S mutations (kdr) as the sole cause of the kdr phenotype is still discussed (Brooke 2008), *kdr* is clearly associated with PYR and DDT resistance in *Blatella germanica*, *C. pipiens*, houseflies, horn flies, and some moths (Du et al., 2013). In *An. gambiae*, although metabolic resistance is often present, high resistance to PYR and DDT is most of the times associated with a high kdr frequency, and resistant insects carry at least one *kdr* copy (Martinez-Torres et al., 1999; Xu et al., 2006; Awolola et al., 2007). Moreover, *kdr* frequency usually increases when PYRs are used (Ramphul et al., 2009; Stump et al., 2004): two alleles are spreading in *An. gambiae* in Africa, L1014F and L1014S mutations and analyses of the noncoding regions of the kdr gene suggest that the two alleles occurred several times independently (at least three times for L1014F and two times for L1014S) (Williamson et al., 1996; Norris et al., 2015; Weill et al., 2000a,b). Similarly, in West African *C. p. quinquefasciatus*, resistance frequency follows a gradient of treatment intensity (Pinto et al., 2007). In the field, *An. gambiae* resistance to PYRs through *kdr* can lead to reduced repellent effect and decreased mortality. For example, *kdr* frequency is high in Benin and Cote d'Ivoire (Chandre et al., 1998). In contrast, other studies have found that despite the high correlation between *kdr* mutations and PYR resistance, PYR-treated bed nets remained somewhat efficient against resistant *An. gambiae* (Brooke, 2008; N'Guessan et al., 2007). This could be due to the ability of resistant mosquitoes to stay on a treated bed net longer than susceptible ones, and thus absorb

a high-enough quantity of insecticide to be killed (Etang et al., 2009). For example, in Kenya, the use of PYR-treated bed nets increased *kdr* frequency, but had no impact on malaria and mosquito population densities, as both decreased in treated and untreated villages (Ramphul et al., 2009). Similarly, two studies found that *kdr* alone (i.e., in the absence of metabolic resistance) did not reduce bed net efficiency against resistant *An. stephensi*, despite a reduced KD effect (Casimiro et al., 2007). The issue of the impact of *kdr* resistance on PYR-treated bed net efficiency to control malaria thus remains hotly debated.

3.2.3 Acetylcholinesterase

In the cholinergic synapses of invertebrate and vertebrate central nervous system, AChE terminates the synaptic transmission by rapidly hydrolyzing the neurotransmitter acetylcholine (ACh). AChE is the target of OPs and CXs insecticides, which are competitive inhibitors of ACh: when they bind to AChE, their very slow release prevents hydrolysis of the natural substrate. Consequently, ACh remains active in the synaptic cleft and the nervous influx is continued, leading to insect death by tetany. In most insects there are two genes, ace-1 and ace-2, coding for AChE1 and AChE2, respectively. In these species, AChE1 is the main synaptic enzyme while the physiological role of AChE2 is still uncertain. Diptera of the Cyclorrapha group or "true" flies (such as *D. melanogaster* and *M. domestica*) possess a single AChE, which is encoded by the ace-2 gene and is the synaptic enzyme in that case. Phylogenetic analyses have shown that the presence of two ace genes is probably the ancestral insect state (Enayati and Hemingway, 2006; Weill et al., 2002). The first molecular studies on an insensitive AChE conferring resistance to OPs and CXs were carried out on *D. melanogaster.* Several mutations were identified, each giving a low resistance when alone, and a higher resistance when in combination (Huchard et al., 2006). Similar results were later found with other Diptera that have only the ace-2 gene (e.g., *Musca domestica*) (Oakeshott et al., 2005). In mosquitoes where AChE1 is the synaptic enzyme, the most common resistance mutation (G119S) in the ace-1 gene is located just near the active site. In *C. pipiens*, G119S occurred at least 3 times independently, once in *C. p. pipiens* and twice in *C. p. quinquefasciatus* (Weill et al., 2004; Enayati and Hemingway, 2006; Fournier and Mutéro, 1994). However, two other mutations in ace-1 have been identified, both close to the active site: (1) F331W has been observed only in *Culex tritaeniorhynchus* (Labbé et al., 2007a,b; Nabeshima et al., 2004), (2) F290V has been observed only in *C. p. pipiens* (Alout et al., 2007a,b). The type of mutation appears highly constrained by the codon use: until recently the G119S mutation was never found in *Ae. aegypti*, *Ae. albopictus*, or *C. tritaeniorhynchus*, probably because it requires two mutational steps (Weill et al., 2004). It was, however, described in *Ae. aegypti* from India in 2015, apparently through a mutation from a different codon R119S (Alout et al., 2007a,b). The ace mutations are responsible for a decreased inhibition of the AChE by the insecticides (Muthusamy and Shivakumar, 2015). There are only few resistance mutations observed in various species, suggesting high constraints: those observed in the field are within the active gorge of the enzyme and cause steric problems with bulkier side-chains, while other substitutions (lab-engineered) often result in the inability of enzyme to degrade Ach (Oakeshott et al., 2005). The G119S ace-1 mutation has recently been shown to interact synergistically with an unknown sex-linked gene to allow a >40,000-fold resistance to chlorpyrifos (OP) (Alout et al., 2008). Similarly, the G119S mutation associated with the *kdr*R allele confers higher resistance levels in *An. gambiae* to both OPs and CXs insecticides (Alout et al., 2015). The evolution of insensitive AChE1 has been studied in depth in the mosquitoes *C. pipiens* and *An. gambiae*. In *C. pipiens*, it was first detected in Southern France in 1978, 9 years after the beginning of OP treatments

(Assogba et al., 2014). The gene coding for this G119S mutated AChE1 (*ace-1R*) rapidly spread in treated natural populations. However, its frequency remained low in adjacent untreated areas connected by migration, indicating a fitness cost associated with ace-1R (Raymond et al., 1986). The >60% reduction of AChE1 activity in G119S-resistant mosquitoes (Lenormand et al., 1999) may probably explain, at least partially, this cost, which is expressed phenotypically through various developmental and behavioral problems in individuals carrying ace-1R (Bourguet et al., 1997, 2004; Berticat et al., 2002). Similarly, the F290V mutation is probably associated with a fitness cost, although it does not appear to be due to activity reduction (Alout et al., 2008). Several independent heterogeneous duplications of the ace-1 gene, putting a susceptible and a resistant copy in tandem (ace-1D), have been identified in *C. p. pipiens* and *C. p. quinquefasciatus* (Fournier and Mutéro, 1994; Bourguet et al., 2004). These alleles are thought to be selected because they confer an alternative fitness trade-off, that is, reducing the cost of the ace-1R allele, but with a decreased resistance level as well (Lenormand et al., 1998). However, some ace-1D can be associated to extremely deleterious phenotypes when homozygotes (Lenormand et al., 1998; Labbé et al., 2014). Several other duplications have been observed recently in the Mediterranean area, with a F290V copy instead of a G119S copy (Lenormand et al., 1998; Labbé et al., 2014). In *An. gambiae*, the occurrence of ace-1R has been detected in several West African countries, and this allele is probably spreading from a single origin (Labbé et al., 2007a,b; Weill et al., 2003; Djogbénou et al., 2008). As in *C. pipiens*, this mutation is associated with a strong selective cost in *An. gambiae* (Dabiré et al., 2009). A duplication carrying a G119S copy has also been found, and appears to follow the same trajectory as in *C. pipiens* (Assogba et al., 2015) the *An. gambiae* ace-1D allele also provides an alternative phenotype, a reduced cost associated with a reduced resistance (Dabiré et al., 2009). In both species, it has been suggested that the relative fitness of the two alleles (ace-1R and ace-1D) may depend on the intensity of insecticide treatments. Finally, two studies reported in 2015 have suggested that resistance alleles with multiple ace-1R copies are segregating in Africa (Djogbénou et al., 2009; Weetman et al., 2015) the fitness consequences of such duplications remain, however, unknown.

3.3 Other Resistance Mechanisms

3.3.1 Hormonal Regulation mechanisms

Juvenoids mimic (juvenoid hormones) disrupt insect development. Few resistance cases have been described in various species. High resistance to methoprene has been described in the mosquito *Ochlerotatus nigromaculis* in California, potentially through target-site mutation (Cornel et al., 2002) while a 7.7-fold resistance to the same insecticide has been reported in *C. p. pipiens* from New York (Djogbénou et al., 2015).

3.3.2 Toxin Receptors

Bacillus thuringiensis (Bt) toxins have a complex mode of action not clearly understood. *Bt* resistance is increasing in the field in several pests (Paul et al., 2005). Presently, the only report of field resistance in mosquito is a 33-fold resistance to Bti (*B. thuringiensis* var. *israelensis*) detected in a natural population of *C. p. pipiens* from New York. However, the mechanism of this resistance was not investigated (Djogbénou et al., 2015). Genomic studies suggested several candidates for Bti resistance in *Ae. aegypti*, but they are not yet validated (Tabashnik et al., 2013). Finally, it appears that depending on the environmental conditions, some of the four Bti toxins may be inactivated (Després et al., 2014),

which could favor the emergence of full Bti resistance through intermediate bouts of selection to each toxin independently. For *Bacillus sphaerucus* (Bs) toxins, resistance has been described essentially in mosquitoes of the *C. pipiens* complex, due to mutation in the toxin receptor. It developed very rapidly within the first year of treatment in India (10e155_ resistance) (Mittal, 2003), and in Tunisia (Sp-T gene, >5000_ resistance) (Tetreau et al., 2012). Similarly, control using Bs toxins started in the early 1990s in Southern France and first failure was reported in 1994 in Port-Louis (near Marseille). This resistance (>10,000) was due to a recessive sex-linked gene, named sp-1. In 1996, Bs resistance was reported close to the Spain border (Perpignan, 200 km away from Port St Louis); it was due to a second gene, sp-2, which was recessive and sex-linked (Nielsen-Leroux et al., 2002). Now Bs resistance has been observed worldwide in the *C. pipiens* complex (Mittal, 2003). Two of the alleles identified (sp-2R and an allele selected in a laboratory strain from California (Tetreau et al., 2012) change the toxin receptor binding properties, and were found to be due to "stop" mutations or mobile element insertion in the toxin receptor (Chevillon et al., 2001; Darboux et al., 2007). The effect of the other alleles is unknown. *Bacillus sphaericus* resistance has also been selected in the laboratory in *An. stephensi* (Mittal, 2003).

4. Methods for Testing and Monitoring Insecticide Resistance in Mosquitoes

Insecticide resistance monitoring in field populations of mosquitoes is required to determine the frequencies, mechanisms, and geographical distribution of resistance to select appropriate insecticides for vector control. Together with information on adult mosquito density, larval and pupal indices, ecology and habitats, and efficacy of vector control interventions, insecticide resistance data help to develop appropriate responses, to prevent and control mosquito borne viruses. Currently, methods to test and monitor insecticide resistance include biological, biochemical, and molecular assays.

4.1 Bioassays

The first susceptibility bioassay was made available by WHO more than 60 years ago. Historically, several changes have occurred to the test procedures, which may seem minor but could impact bioassay results. Susceptibility test can be carried out either on larvae or on adult insects.

4.1.1 Larval Tests

Larvae used for bioassays should be collected from the productive breeding sites using dippers or ovitraps and sorted into different instars. Tests are then conducted at $27 \pm 2°C$ temperature, $75 \pm 10\%$ relative humidity and low illumination that are usually maintained in an insectary. Where possible, comparison test with susceptible laboratory colonies should be made under similar conditions of temperature and humidity. Late third/early fourth instar larvae are exposed to a wide range of insecticide concentrations to test the activity range of the insecticide in disposable test cups or vessels. After determining the mortality of larvae after 24 hours of exposure to this range of concentrations, a narrower range (of four to five concentrations, yielding between 10% and 95% mortality in 24 hours) is used to determine LC50 and LC90 values. For slow-acting insecticides, 48-hour exposure may be required. The same procedure is carried out for a susceptible strain and a field test

population. Four or more replicates are set up for each concentration and an equal number of controls are set up simultaneously with tap water, to which 1 mL alcohol (or the organic solvent used) is added. Each test should be run three times on different days. Resistance ratios (RR) are often calculated and useful to monitor the evolution of insecticide resistance in a field population. The way to calculate an RR is to divide the LC50 of the field population by the LC50 of a susceptible strain. When RR is < 5 the field population is considered susceptible, when RR is between 5 and 10, test mosquitoes are considered displaying moderate resistance, and when RR is > 10 they are highly resistant.

4.1.2 Adult

- **Tube assay**
  Procedures to monitor the resistance in adult mosquitoes to a range of insecticides in tube assays commonly used for mosquito control have been published and updated by WHO (Saavedra-Rodríguez et al., 2007). Resistance monitoring using this approach relies on the collection of wild mosquito larvae, which are then reared to adulthood in a test facility. They are then exposed for 1 hour to a discriminating concentration (DC) of insecticides on a treated filter paper so that their knockdown and mortality can be scored 24 hours (or 48 hours for slow acting compounds) post exposure. Diagnostic doses and diagnostic times have been determined for mosquitoes from many geographical regions.
  The WHO tube bioassay is a simple direct response-to-exposure test. The test kit was developed in 1958 (WHO, 2016), they can be purchased from University Sains in Malaysia. It was designed to expose a defined number of adult mosquitoes of known ages and physiological statuses to an insecticide impregnated on a filter paper for a standard exposure time (1 hour). While the WHO test procedures provide parameters for some key environmental conditions which should be kept constant while carrying out the bioassay, during insecticide resistance monitoring there are multiple potential sources of (nonresistance-associated) variability, which can influence the result of the bioassay.
  In 2018, WHO has updated the test procedures, introducing resistance intensity assays (WHO, 2018). Indeed, any resistance phenotypes detected using the discriminating insecticide concentrations can be assessed for their potential operational significance by exposing subsequent or additional mosquito samples to the applicable 5× and 10× higher concentrations of those insecticides. Exposures at the higher concentrations will yield information on the intensity of resistance, which can be defined as the "strength" of a resistance phenotype. It is suggested that resistance at 5× and especially at 10× the discriminating concentration may indicate or predict operational control failure and highlight a particularly urgent need to develop an appropriate resistance management strategy (WHO, 2012).
- **WHO Bottle assay**
  During a recently concluded formal WHO multicenter study to establish species-specific discriminating concentrations and procedures for new and existing insecticides, WHO bottle assays has been adapted the Center for Disease Control (CDC) bottle bioassay using a range of insecticides and multiple *Anopheles* and *Aedes* species (WHO, 2022). The 250-mL Wheaton bottles used for the bioassay need to be coated inside with the diagnostic dose of the insecticide then 10−25 female mosquitoes are tested in each bottle.

- **Other bioassays**
  New bioassays have been developed to monitor mosquito susceptibility to pyriproxyfen grow regulators (WHO, 2022). This susceptibility bioassay is an exposure response test, which measures the reduction in oviposition rate in adult female mosquitoes preexposed to a known standard concentration (e.g., discriminatory concentration) or to a series of concentrations of pyriproxyfen for a fixed duration of 1 hour.

4.2 Identification of Resistance Mechanisms

Resistance is assumed to be present if a portion of the test population survives the diagnostic dose at the diagnostic time (Saavedra-Rodríguez et al., 2007). Test and control mosquitoes used in bioassay can further be used for identification of mechanisms of resistance using enzymatic assays (control mosquitoes) or molecular methods.

Mosquitoes that will be further tested using enzymatic assays should be stored frozen. Mosquitoes to be used for molecular studies can be frozen, dried, or stored in 70% (or higher) ethanol. In addition, it may be necessary to use products like RNALater (Applied Biosystems [Ambion], Foster City, California) to preserve samples for measurement of RNA levels associated with up-regulated enzyme mechanisms.

4.2.1 Bioassays

Piperonyl butoxide (PBO) synergizes the effects of pyrethroid insecticides against mosquitoes presenting metabolic resistance mechanism particularly monooxygenases and is currently used in combination with pyrethroids to impregnate mosquito nets (known as pyrethroid-PBO nets), which are designed to mitigate the impact of this resistance mechanism on the effectiveness of insecticide treated mosquito nets.

Bioassays can be used for evaluating the ability of PBO or other synergists to restore the susceptibility of mosquitoes to pyrethroids in pyrethroid-resistant vector populations. This bioassay is a direct response-to-exposure test that compares mosquito mortality after exposure to a pyrethroids insecticide in one test group to mosquito mortality after preexposure to 4% PBO-treated papers followed by exposure to the same pyrethroid in the second test group. It provides an indirect assessment of the involvement of mono-oxygenases in conferring pyrethroid resistance, and results are meant to contribute toward decision-making on the deployment of pyrethroid-PBO nets (see file:///C:/Users/josye/Downloads/9,789,240,043,855-eng.pdf).

4.2.2 Molecular Assays

A multitude of molecular assays have been developed to detect resistance alleles (*kdr*, *Ace*, etc.) in mosquitoes, several of which were recently compared in a study by Bass et al. (Brengues et al., 2003). Unfortunately, despite the recent identification of the key enzymes responsible for metabolic resistance to pyrethroids, there are currently no simple DNA-based assays to detect these resistance mechanisms. Detection of these genes is presently dependent on RNA based approaches using relatively sophisticated equipment (e.g., RT-qPCR).

Assays to detect the genetic mutation(s) responsible for the resistance phenotype in individual insects can provide an early warning of the emergence of resistance which may not have been detectable by bioassays that can only record the population response. The presence of a single

individual with an allele known to confer resistance should be cause for concern as experience dictates that resistance can spread very rapidly in a population unless the selection pressure is eased and/or the genetic cost associated with the resistant allele is high. Conversely, a negative result from a molecular assay should not lead to complacency. As discussed above, molecular assays are presently only available for target site resistance and the failure to detect *kdr* clearly cannot be interpreted as an absence of resistance in a population. Hence, molecular assays should be seen as a complement rather than a substitute for bioassays.

4.2.3 Biochemical Assays

Biochemical tests to detect alterations in activities of enzyme families associated with insecticide resistance have been available for over 2 decades and are sometimes used in combination with insecticide bioassays (Dong et al., 2014). These assays employ model substrates to record the overall activity of glutathione transferases, carboxylesterases, or cytochrome P450s in individual insects. Biochemical assays are also available to detect target site resistance to organophosphate and carbamate insecticides caused by insensitive acetylcholinesterase (AChE). The enzymatic reaction produces a color change that is generally visible to the naked eye and hence these assays do not require access to expensive equipment (spectrophotometer is appropriate). However, there are sensitivity and specificity issues that limit the utility of some of these assays (Chandre et al.,1999, 2000).

5. Features of Insecticide Resistance

Some general patterns can be identified from the variety of mechanisms observed for insecticide resistance.

5.1 Evolvement of Resistance

The first characteristic of resistance is its rapid evolvement, with fast selective sweeps in field populations. Most of the times, resistance alleles are present in the field before insecticide treatments, at very low frequencies. They are selected locally but can spread very rapidly. A single resistance gene may have a large distribution (Ffrench-Constant et al., 2004; Duron et al., 2006; Williamson et al., 1996) for example, the worldwide migration of Ester2 in *C. pipiens* (Raymond et al., 2001). Alternatively, other resistance alleles have multiple origins: ace-1R mutations in *C. pipiens* (G119S (Labbé et al., 2007a,b) or F290V (Alout et al., 2007a,b) or *kdr* mutations in *Ae. Aegypti* (Darboux et al., 2007; Marcombe et al., 2009). It also seems that resistance evolution is quite constrained. For target-site resistance, most mutations are costly and compromise the performance of the native protein function, so that codon usage may prevent resistance apparition (Weill et al., 2004; Ranson et al., 2000).

5.2 Cross-Resistance

Cross-resistance occurs when resistance to one insecticide confers resistance to other insecticides, even where the insect has not been exposed to the latter product. It often refers to the instance where resistance is conferred to two or more classes of insecticides, it is commonly assumed within an insecticide class.

Cross-resistances between insecticide classes can be associated with the sharing of target sites. For example, *kdr*R causes cross-resistance between DDT and PYRs in *An. gambiae* (Etang et al., 2009) and ace-1R between OPs and CXs (Muthusamy and Shivakumar, 2015). Cross-resistance can even be a greater issue when considering metabolic resistance. First, different genes belonging to a same enzyme family can cause resistance to several insecticides ("gene family cross-resistance"), even from different classes: for example, different COE and P450 genes cause resistance to DDT, others to PYRs, OPs, and CXs in Anophelines (Brogdon and McAllister, 1998). However, a unique gene may also be involved in resistance to several insecticides, from different classes: this is the case, for example, of the CYP6M2 gene (P450), which can metabolize both deltamethrin (PYR) and DDT (OC57). The consequences of these cross resistances are a severe reduction of the availability of alternative insecticides, thereby gravely endangering vector control.

5.3 Multiple Resistance

Despite advances, a full analysis of resistance remains challenging due to the complexity of interactions, pleiotropy, and redundancy when several resistance mechanisms and/or resistance genes are present in the same insect (Edi et al., 2014). Interactions between resistance loci have been studied in houseflies or mosquitoes, and most of them appear to be synergistic. Such synergies have been observed, for example, in *C. pipiens* between COE and ace-1 for OP resistance, between ace-1 and an unknown gene, raising resistance to chlorpyrifos by more than 2000-fold compared to ace-1 alone (>40,000-fold compared to susceptible) (Alout et al., 2008) and between *kdr* and P450 for PYR resistance in *Ae. aegypti* between repellents (DEET) and CXs, in *An. gambiae* s.s. between ace-1 and kdr for OPs and CXs resistance (Alout et al., 2015) or in three Anopheline species between PYR resistance and susceptibility to fungus applications (Farenhorst et al., 2009). Moreover, these interactions may vary with environmental conditions (positive synergism for resistance in treated area but negative synergism for cost in nontreated areas) or with the genetic background of the insect. For example, the presence of *kdr* decreases the cost of ace-1R in *C. pipiens* (Berticat et al., 2008). In *An. gambiae* s.l. from North Cameroon, interactions between resistance mechanisms revealed the emergence of P450s on top of *Kdr* L995F in urban areas, while *Kdr* L995F overpassed the P450s already present in rural areas (Etang et al., 2022).

6. Factors Governing the Evolution of Insecticide Resistance

As a general statement, the use of insecticides does not create resistance by itself but select small proportion of individuals having a genetic mutation that allow them to resist and survive the effects of the insecticide. If this advantage is maintained by constant use of the same insecticide, the resistant insects will reproduce and the genetic changes that confer resistance will be transferred to offspring so that they become more prevalent within the population. This selection process will take longer time to occur if the gene conferring resistance is rare or present at a low prevalence. Resistance should not be confused with "induction" that can occur after sublethal (or low dose) exposure to any insecticide and/or xenobiotic and is not passed on to offspring (Poupardin et al., 2008). Main factors influencing resistance development. The evolution of insecticide resistance is complex and depends on several genetic, biological and operational factors (IRAC, 2010; Denholm and Rowland, 1992). The biological

factors relate by the life cycle of the insect (e.g., rate of reproduction, number of generation/offspring, rate of migration and isolation, etc.), while the genetic factors include the intrinsic characteristics of the resistant genes (e.g., mono vs. polygenic resistance, dominance, fitness cost, and gene interaction). Operational factors concern the treatment itself including the method and frequency of application, dosage, and residual activity of the insecticides as well as insecticide coverage.

6.1 Biological Factors

Rate of reproduction. Insect species that have a short life cycle and high rates of reproduction are likely to develop resistance more rapidly than species that have a lower rate of reproduction, as any resistance genes can rapidly spread throughout the population. Because mosquitoes can produce high number of offspring (i.e., females can lay several hundred eggs during their reproductive life) they are much likely to develop resistance to insecticides than other species.

Population migration/isolation. With mosquitoes, the goal is to eliminate all or the majority of the population; however, the greater the selection pressure that is put on a population, the faster susceptibility may be lost. Immigration of individuals possessing susceptible genes from untreated areas can beneficially dilute and compete with the resistance genes in the overall population. An early step in a malaria vector control program should therefore be to assess the susceptibility status of vector populations and detect potential immigration of untreated insects. This can be achieved by using genetic markers to estimate the gene flow (migrants) and genetic structure between populations. For example, an isolated area (e.g., island) where the entire area is treated would have a higher risk of developing resistance as few "susceptible" genotypes would join the treated population. The risk of insecticide resistance developing should be considered when planning a resistance management strategy. Awareness of and coordination with neighboring vector control programs and agricultural activities should be encouraged, so that the regional and potential "side effect" on the target population is considered.

6.2 Genetic Factors

Dominance. Resistance genes can range from dominant through semidominant to recessive. If dominant or semidominant, only one parent needs to possess the characteristic to be fully or partially expressed in the offspring. If recessive, both parents must possess the trait. Fortunately, most resistance mechanisms (e.g., *kdr*) are controlled by recessive or semirecessive genes, which slows their spread within the population at early stage of resistance development when most individuals are present at heterozygous state. In contrast, when the resistance is genetically dominant, e.g., the *Ace*.1R gene conferring cross-resistance to carbamates and OPs (Djogbenou et al., 2007), it can rapidly become established within the population and will be difficult to manage. Fortunately, strong genetic cost is often associated with dominant resistant gene that can compensate the effect of the dominance and slow down the increase of resistance gene frequency in natural populations (Berticat et al., 2008).

Gene interactions. Epistasis is the nonadditive interaction (synergistic vs. antagonistic) between different loci which contribute to a phenotype (Moore and Williams, 2005). Epistasis between independent loci conferring insecticide resistance is important to investigate as this phenomenon can shape the rate at which resistance evolves and can dictate the level of resistance in the field. Epistasis can be measured in laboratory studies on susceptible and resistant colonies, but without these data, it is

generally impossible to predict whether or not it will occur when two genes are being evaluated. Studies of the interactions between resistance loci have been most commonly conducted in house flies (Shono et al., 2004; Scott et al., 1984). Generally, a greater than additive interaction was observed between two loci that were both homozygous resistant, whereas additively (i.e., lack of epistasis) occurs between two loci that were both heterozygous. In mosquitoes, Harstone and colleagues (Hardstone et al., 2009) showed multiplicative interactions between *kdr* and P450 detoxification in *Culex pipiens quinquefasciatus* whether the resistance alleles were homozygous or heterozygous. For example, resistance ratio 50 (i.e., LC50 resistant strain/LC50 susceptible strain) to permethrin in the double homozygote mosquitoes (RR50 of 30,000) was much higher than that expected (RR50 of 1400) by simple additive effect of the two loci. Overall, interactions between independent resistant genes are complex. It is therefore important to better understand the interactions between resistant loci as well as to address how the fitness costs/benefits of the mechanisms can manipulate the observed interactions.

Fitness cost. Populations of insects that have never been exposed to insecticides are usually fully susceptible, and resistance genes within those populations are very rare. This usually occurs through a "fitness cost," which means that insects sharing the resistance allele lack some other attribute or "quality" such that it gives an advantage to the susceptible insects in an insecticide-free environment (Bourguet et al., 1997). For example, resistant insects may have lower mating success, be more susceptible to natural enemies (Agnew et al., 2004), or more prone to mortality during overwintering (Foster et al., 1996). Increased production of metabolic enzymes generally shows lower associated fitness cost than those associated with alterations in the structural genes most probably because the primary function of the enzyme is not disrupted (Shi et al., 2004). There is good laboratory and field evidence to suggest that the deficit of insecticide selection pressure, in most cases, selects for susceptible genotypes. For example, the absence of homozygote's resistant genotypes in *An. gambiae* populations in West Africa is most probably due to the strong genetic cost associated with the carbamate-resistance allele ace.1R (G119S) (Djogbenou et al., 2010). In addition, once resistance in the field has been selected it often rapidly reverts once the insecticide treatment regime is changed. A good example of this occurred in *An. arabiensis* in Sudan, where malathion specific insecticide resistance was selected in the early 1980s through antimalarial house spraying. The development of resistance prompted a switch of insecticide treatment to fenitrothion and the malathion resistance rapidly reverted in the following years. However, reversion rates are variable and may be very slow, particularly when an insecticide has been used for many years. For example, DDT was used extensively for malaria control over a 20-year period up to the 1960s in Sri Lanka to control *An. culicifacies* s.l. and *An. subpictus*. DDT was replaced by malathion in Sri Lanka in the early 1970s when a total and effective ban on DDT use was implemented. Subsequent regular monitoring has shown that DDT resistance has reverted very slowly toward susceptibility; around 80% of the adult mosquito population was resistant in the 1970s compared to about 50% in the 1990s. The same is true with the Rdl gene that was maintained in field mosquito populations despite the abandon of cyclodiene for mosquito control for more than 30 years (Mahande et al., 2012). Rate of reversion is an important parameter to consider before implementing any resistance management strategy in the field.

Presence of a lower fitness cost and high reproductive and short life cycles producing *several generations per season*: The length of time that an insecticide remains effective, also called its persistence, is dependent upon the physical chemistry of the insecticide, the type of formulation, the application rate and the substrate. Products which provide a persistent effect provide continual selection pressure in a similar manner to multiple treatments. For example, a space pray will persist for a

very short time and will select only against a single generation of mosquitoes. In contrast, a residual wall application (IRS) or insecticide treated nets treatment (especially Long-Lasting Nets) will persist for months or years providing a selection pressure against many generations of the same insect. For example, repeated application of DDT for indoor residual spraying has contributed to increase the number of DDT-resistant malaria vector species in various geographical settings. Several studies showed however that the use of insecticides in agriculture play a key role in the selection of resistance in mosquitoes (Yadouleton et al., 2011; Diabate et al., 2002). Indeed, most insecticides used in agriculture are of the same chemical classes and have the same targets and modes of action as those used in public health program. In practice, VC programs cannot influence the choice of the pesticide used for crop protection and the only thing that can be done is to appropriately select the most judicious insecticide for mosquito control. However, there is more published evidence that public health insecticides can contribute to select for pyrethroid resistance alleles. It is obvious that we can expect enhanced selection pressure on resistance genes through the scaling up of LLIN and/or IRS for malaria elimination.

6.3 Operational Factors

In practice, only operational factors such as the insecticide(s) used, the area of coverage (for example for IRS or LLIN), and the timing, rate, and method of application can be manipulated directly to reduce the selection pressure for resistance. Operational factors influence selection by determining the overall fraction of a population exposed (larvae/adults) to a selecting agent and the degree of contact and pick-up of toxicant by exposed pests at what has been termed the "interface between insects and insecticides" (Denholm and Rowland, 1992). At both stages, operational and intrinsic factors interact in complex ways to establish the net effect of a control treatment on both genetic composition and total population size. Management of resistance therefore entails resolving these interactions to anticipate with some confidence both the suppressive and selective effects of potential control strategies. Frequency of application, dosage, and persistence of effect. How often an insecticide is used is one of the most important factors that influence resistance development (IRAC, 2010). With each use, an advantage is given to the resistant insects within a population. The rate of increase of resistance on any population will generally be faster in the Anopheline mosquitoes.

7. Insecticide Resistance in Disease Vectors

Disease vectors include insects, rodents, birds, or other animals that can transmit infectious pathogens between humans, or from animals to humans. The most common vectors are insects that suck blood, such as mosquitoes and ticks. Mosquitoes which act as major vectors of severe human diseases belong to three genera: *Anopheles*, *Culex*, and *Aedes*.

Insecticide resistance has been observed in more than 500 insect species worldwide (Corbel and N'Guessan, 2013), including malaria, arboviruses, and lymphatic filariasis vectors. By 1980, resistance to at least one insecticide had been reported in over 150 species of arthropods of public-health importance, including 93 species of mosquitoes (Kakarunamoorthi and Sabesan, 2013). This led to failure of some control programs and a switch to alternative insecticides (Marcombe et al., 2011; Vontas et al., 2004). For example, DDT was first introduced for mosquito control in 1946 and the first

cases of DDT resistance occurred in *Aedes tritaeniorhynchus* and *Ae. solicitans* in 1947. DDT resistance led to its replacement by organophosphates (e.g., malathion and fenitrothion), carbamates (e.g., propoxur and bendiocarb), or synthetic pyrethroids (e.g., permethrin and deltamethrin). Resistance has tended to follow the switches Insecticide groups such as the benzylphenyl ureas and Bti.

7.1 Malaria Vectors

Insecticide resistance was first reported in malaria vectors in the 1950s (Elliot and Ramakrishna, 1956) to date, malaria vectors have developed resistance to the main chemical classes used in public health, i.e., pyrethroids (PYs), organochlorines (OCs), carbamates (CAs), and organophosphates (OPs). About 61 countries over 76 malaria endemic countries had detected resistance in at least one major malaria vector to at least one class of insecticide used for adult malaria vector control. A total of 18 countries from 4 WHO regions detected resistance to all four classes between 2010 and 2016 (WHO, 2018). Resistance to pyrethroids is widespread, while resistance to organochlorines, organophosphates, and carbamates is still increasing (Ranson and Lissenden, 2016; Hemingway et al., 2016).

Pyrethroid resistance in anopheline mosquitoes is due to two key resistance mechanisms, namely target site resistance and metabolic resistance. Both of these two mechanisms are regularly present in the same mosquito vector populations. Cross resistance between pyrethroids and the organochlorine dichloro-diphenyl-trichloroethane (DDT) appears to be common in malaria vectors across Africa with high resistance frequencies (Ranson et al., 2011; Ranson and Lissenden, 2016; Hancock et al., 2018). There is also evidence of low resistance frequencies to carbamates organophosphate (OP) in many countries.

7.2 Arbovirus Vectors

The conventional strategies for controlling *Aedes* mosquitoes are based on reduction of breeding sites and insecticide-based interventions. However, as in the case of malaria, many vector control programs are threatened by the development of insecticide resistance in *Ae. aegypti* and *Ae. albopictus* (Ranson et al., 2010; Paupy et al., 2010), the main epidemic arbovirus vectors worldwide (Paupy et al., 2010; Wong et al., 2016). *Aedes aegypti* resistance has emerged into all conventional classes of insecticides, particularly to pyrethroids and the organophosphate temephos. Resistance of *Ae. aegypti* and *Ae. albopictus* to DDT and pyrethroids is present across the world (Paupy et al., 2010; Ranson et al., 2010; Kamgang et al., 2013). Several *kdr* mutations have been identified in *Ae. aegypti*, and the association between the V1016G/I and the F1534C mutations, and pyrethroid resistance has been established (Darboux et al., 2007; Brengues et al., 2003). Also, metabolic resistance through upregulation of detoxification genes is a major resistance mechanism in both species.

7.3 Other Vectors

Other common vectors include black flies' fleas, filth flies, sandflies, midges, and lice.

Vector control programs are almost nonexistent and the resistance to insecticides observed in most of these vectors is the consequence of the indiscriminate use of pesticides. For example, resistance in blackflies (Andrade and Castello, 1990; Osei-Atweneboana et al., 2001) has apparently arisen from agricultural pesticide use.

Resistance to DDT and organophosphates is well documented in simuliids (Georghiou and Lagunes-Tejada; 1991). Permethrin resistance in larvae of *Simulium squamosum* was attributed to decreased sensitivity of the target site and increased detoxification mechanisms (Hougard et al., 1992). Elevated esterase activity (Andrade and Castello, 1990; Parker and Callaghan, 1997) hydrolytic activity and oxidative activity in blackflies (Siegfried and Young, 1993) are other potential resistance mechanisms. Chemical tools to control sand flies including indoor residual spraying (IRS), space spraying (predominantly ultralow volume (ULV) fogging), insecticide-treated nets (ITNs), topical and spatial repellents, and impregnated dog collars (IDCs) (Wilson et al., 2020; Alexander and Maroli, 2003) have triggered the development of insecticide resistance (IR) in key sand fly vector species' populations (Alexander and Maroli, 2003; Rocha et al., 2020). The underlying mechanisms of resistance in sand flies remain largely unknown. Most molecular studies focus on target-site mutations with fewer reports on metabolic resistance. Resistance to DDT and pyrethroids (type I and II) has been correlated with multiple point mutations in the voltage-gated sodium channel (*VGSC*) gene (*kdr* mutations), with the presence of both 1014F and 1014S mutations in the major VL vector *P. argentipes*(Gomes et al., 2017; Sardar et al., 2018), *P. papatasi* (Fotakis et al., 2018), and *P. argentipes* (Pathirage et al., 2020) Possible reduced AChE susceptibility in *P. argentipes* field populations has also been observed (Pathirage et al., 2020; Surendran et al., 2005). Cuticular or behavioral resistance mechanisms, which have been reported and described in other insect vectors (Balabanidou et al., 2019; Carrasco et al., 2019), have not been investigated in sand flies. Concerning metabolic resistance, Fawaz and colleagues (Fawaz et al., 2016) highlighted the role of metabolic detoxification pathways in a permethrin-resistant *P. papatasi* colony (established under laboratory selection). Biochemical and molecular analyses revealed that the observed resistance was mediated by elevated oxidase and esterase activity levels. Pyrethroids, including d-phenothrin and permethrin, have been registered as pediculicides since the 1970s, and have been widely available since the 1980s (Surendran et al., 2005; Mandal et al., 2019). The first reports of control failure with this insecticide were in the early 1990s Mandal et al., 2019; Feliciangeli et al., 2003).

The body louse *Pediculus humanus* has developed widespread resistance to organochlorines (Brown and Pal, 1971; Maunder, 1971) is malathion resistant in parts of Africa (WHO, 1992) and has low-level resistance to pyrethroids in several regions (Fine, 1963). Resistance to organochlorine insecticides, such as DDT and lindane, has been recorded in the human head louse *Pediculus capitis* in Israel, Canada, Denmark, and Malaysia (Brown and Pal, 1971; WHO, 1992). Permethrin resistance in head lice was mostly conferred by the recessive *kdr* trait (Gao et al., 2003; Lee et al., 2000). Three point mutations (M8151, T9171, and L920F) in the VGSC α-subunit gene associated with permethrin-resistant phenotypes were suggested to be responsible for *kdr*-type resistance (Lee et al., 2003).

8. Approaches to Insecticide Resistance Management

Historically, the practice of using an insecticide until resistance becomes a limiting factor has rapidly eroded the number of suitable/available insecticides for vector control. Rotations, mosaics, and mixtures have all been proposed as resistance management tools (Guillet et al., 2001; Mosha et al., 2008), but there are very few "success stories" in public health. Numerous mathematical models have been produced to estimate how these tools could be optimally used (Committee on Strategies, 1986; Chitnis et al., 2010), but these models have rarely been tested under field conditions due to the practical

difficulties in estimating changes in resistance gene frequencies (especially for metabolic resistance) in large samples of insects (Hemingway et al., 1997). With the advent of different molecular techniques for resistance-gene frequency estimation, field trials of resistance management strategies have now become more feasible.

Ideally insecticide resistance management should be undertaken using insecticide-based approaches in conjunction with other noninsecticidal vector control methods, i.e., as part of Integrated Vector Management (IVM) (WHO, 2004). In practice, most of IVM programs work well in experimental trials but become challenging when programs are scaling up into long-term (operational) control. The simplest form of resistance management is likely to be "insecticide" based, and this could take several forms.

Focus on rotations: Rotational strategies are based on the rotation over time of two or preferably more insecticide classes with different modes of action. This approach assumes that if resistance to each insecticide is rare, then multiple resistances will be extremely rare (Curtis, 1985). Rotation allows any resistance developed to the first insecticide to decline over time when the second insecticide class is introduced. As for other strategy, rotations are particularly effective if the resistance gene has an associated fitness cost. The timeframe for rotation needs to be sufficiently short to prevent significant levels of resistance to develop to any one rotation partner. Rotations have been successful in many applications in agriculture and are considered to be effective in slowing the evolution of resistance. The most striking example of "success story" using this strategy in public health was within the framework of the Onchocerciasis Control Program (OCP) carried in West Africa 40 years ago. Indeed, weekly application of unrelated larvicides in rivers was successful to kill the larvae of the blackfly vector and mitigate the spread of temephos resistance over the 17 years of its implementation (Hougard et al., 1993). However, the rotation was introduced at early stage of the OCP, as soon as the operators faced temephos resistance problems in pilot localities. As for all IRM strategies, the status of resistance of the insecticide used in the rotation must be known when implementing rotations and the chemicals used should not present any (known) cross-resistance. For LLIN, it is difficult to implement this method knowing that pyrethroids are the main class of insecticides recommended so far by WHO for the impregnation (WHO, 2006). For IRS, the pragmatic approach would be to rotate insecticides annually. Indeed, changing insecticides more than once a year (which could be the case in areas where two spray rounds are conducted each year) is not recommended, because of procurement and other financial and logistical challenges. Despite higher cost of implementing rotation than single spray (as available alternatives to pyrethroids the carbamates organophosphates, insect growth regulators, pyrroles are currently more expensive), this is probably the price to pay to preserve the arsenal of cost-effective insecticides for malaria vector control.

9. Conclusion

The natural history of mosquito-borne diseases is complex, and the interplay of climate, ecology, vector biology, and many other factors defies simplistic analyses. The recent resurgence of many of these diseases is of major concern. Its principal determinants are politics, economics, and human activities (rather than climate change). In order to control these diseases and ameliorate the socio-economic burden they cause in developing countries; vector control remains a powerful and accessible tool. However, any disease control strategy should take into account insecticide-resistance

management as it can greatly impact its success (vector control failures) and may have a direct effect on pathogen transmission (Vontas et al., 2004; Alout et al., 2014). This includes first establishing a regular survey of resistance at a local scale by implicating the local population, a difficult but essential task to set goals and evaluate success. Several survey sites in different conditions are required for sentinel purposes, together with some baseline information, to rapidly detect resistance, identify the mechanisms, and change the policies adequately (Kelly-Hope et al., 2008). In order to achieve this survey, basic tools, such as bioassays, remain most powerful, and should always be a preliminary step before more complex and more costly analyses. However, specific and validated molecular markers for the known resistance genes (e.g., some metabolic markers) are also required to rapidly identify the origin and follow the dynamics of resistance at a minimum cost. These local surveys should then be integrated at a more global scale for vector control coordination, allowing informed decisions for using alternative tools to insecticides and preserving the remaining insecticides by carefully planning their use to minimize resistance selection. Clearly, the greatest challenge for successful vector and disease control is the coordination of the different actors (chemical industries, researchers, politics, control agencies, and local populations), which do not have the same agendas, motivations, or economic interests. Besides its implications in public health and development, insecticide resistance remains a powerful evolutionary biology model to study the contemporary adaptation of organisms to a changing environment. It indeed allows a complete and integrative study, from the molecular mechanisms to the fitness consequences at the individual level and their impacts on insect population dynamics and interactions with pathogens. Moreover, it is for once pleasant to see that these rather fundamental approaches of evolutionary biology may have a direct impact in the society and help design new strategies for the successful control of some of the most threatening human diseases (Michalakis and Renaud, 2009).

References

Agnew, P., Berticat, C., Bedhomme, S., Sidobre, C., Michalakis, Y., 2004. Parasitism increases and decreases the costs of insecticide resistance in mosquitoes. Evol. Int. J. Org. 58, 579−586.

Albrecht, C.P., Sherman, M., 1987. Lethal and sub lethal effect of avermectin B1 on three fruit fly species (Diptera: Tephritidae). J. Econ. Entomol. 80, 344−347.

Alexander, B., Maroli, M., 2003. Control of phlebotomine sandflies. Med. Vet. Entomol. 17, 1−18. https://doi.org/10.1046/j.1365-2915.2003.00420.x.

Alout, H., Berthomieu, A., Berticat, C., 2007a. Different amino-acid substitutions confer insecticide resistance through acetylcholinesterase 1 insensitivity in *Culex vishnui* and Culex tritaeniorhynchus (Diptera: Culicidae) mosquitoes from China. J. Med. Entomol. 44, 463−469.

Alout, H., Berthomieu, A., Hadjivassilis, A., Weill, M., 2007b. A new amino-acid substitution in acetylcholinesterase 1 confers insecticide resistance to *Culex pipiens* mosquitoes from Cyprus. Insect Biochem. Mol. Biol. 37, 41−47.

Alout, H., Djogbénou, L., Berticat, C., Chandre, F., Weill, M., 2008. Comparison of *Anopheles gambiae* and *Culex pipiens* acetycholinesterase 1 biochemical properties. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 150, 271−277.

Alout, H., Dj_egb_e, I., Chandre, F., Djogbénou, L.S., Dabiré, R.K., Corbel, V., et al., 2014. Insecticide exposure impacts vector-parasite interactions in insecticide-resistant malaria vectors. Proc. R. Soc. A B 281, 20140389.

Alout, H., Labbé, P., Berthomieu, A., Makoundou, P., Fort, P., Pasteur, N., et al., 2015. High chlorpyrifos resistance in *Culex pipiens* mosquitoes: strong synergy between resistance genes. Heredity 116, 224−231.

Andrade, C.F.S., Castello Branco, A., 1990. Methods for field detection of resistance to temephos in simulidies. Larval esterase level and topical application of the insecticide to adults. Mem. Inst. Oswaldo Cruz. 85, 291−297.

Anonymous, 1995. Insecticide-Miticide' in American Cyanamid Company, pp. 1−20.

Anthony, N., Unruh, T., Ganser, D., Ffrench-Constant, R.H., 1998. Duplication of the Rdl GABA receptor subunit gene in an insecticide-resistant aphid, *Myzus persicae*. Mol. Gen. Genet. 260, 165−175.

Assogba, B.S., Djogbénou, L.S., Saizonou, J., Milesi, P., Djossou, L., Djegbe, I., et al., 2014. Phenotypic effects of concomitant insensitive acetylcholinesterase (ace-1R) and knockdown resistance (kdrR) in *Anopheles gambiae*: a hindrance for insecticide resistance management for malaria vector control. Parasit. Vectors 7, 548.

Assogba, B.S., Djogbénou, L.S., Milesi, P., Berthomieu, A., Perez, J., Ayala, D., et al., 2015. An ace-1 gene duplication resorbs the fitness cost associated with resistance in *Anopheles gambiae*, the main malaria mosquito. Sci. Rep. 5, 14529.

Awolola, T.S., Oduola, A.O., Oyewole, I.O., Obansa, J.B., Amajoh, C.N., Koekemoer, L.L., et al., 2007. Dynamics of knockdown pyrethroid insecticide resistance alleles in a field population of *Anopheles gambiae* s.s. in southwestern Nigeria. J. Vector Borne Dis. 44, 181−188.

Balabanidou, V., Kefi, M., Aivaliotis, M., Koidou, V., Girotti, J.R., Mijailovsky, S.J., et al., 2019. Mosquitoes cloak their legs to resist insecticides. Proc. Biol. Sci. 286, 20191091. https://doi.org/10.1098/rspb.2019.1091.

Beier, J., Keating, J., Githure, J.I., Macdonald, M., Impoinvil, D., Novak, R., 2008. Integrated vector management for malaria control. Malar. J. 7, S4.

Berticat, C., Boquien, G., Raymond, M., Chevillon, C., 2002. Insecticide resistance genes induced mating competition cost in *Culex pipiens* mosquitoes. Genet. Res. 79, 41−47.

Berticat, C., Bonnet, J., Duchon, S., Agnew, P., Weill, M., Corbel, V., 2008. Costs and benefits of multiple resistance to insecticides for *Culex quinquefasciatus* mosquitoes. BMC Evol. Biol. 8.

Bezzar-Bendjazia, R., Kilani-Morakchi, S., Aribi, N., 2016. Larval exposure to azadirachtin affects fitness and oviposition site preference of Drosophila melanogaster. Pestic. Biochem. Physiol. 133, 85−90. https://doi.org/10.1016/j.pestbp.02.009.

Blanford, S., Chan, B.H.K., Jenkins, N., Sim, D., Turner, R.J., Read, A.F., et al., 2005. Fungal pathogen reduces potential for malaria transmission. Science 308, 1638−1641.

Bourguet, D., Roig, A., Toutant, J.P., Arpagaus, M., 1997. Analysis of molecular forms and pharmacological properties of acetylcholinesterase in several mosquito species. Neurochem. Int. 31, 65−72.

Bourguet, D., Guillemaud, T., Chevillon, C., Raymond, M., 2004. Fitness costs of insecticide resistance in natural breeding sites of the mosquito *Culex pipiens*. Evolution 58, 128−135.

Brengues, C., Hawkes, N.J., Chandre, F., McCarroll, L., Duchon, S., Guillet, P., et al., 2003. Pyrethroid and DDT cross-resistance in *Aedes aegypti* is correlated with novel mutations in the voltage-gated sodium channel gene. Med. Vet. Entomol. 17, 87−94.

Brogdon, W.G., McAllister, J.C., 1998. Insecticide resistance and vector control. Emerg. Infect. Dis. 4, 605−613.

Brooke, B.D., 2008. kdr: can a single mutation produce an entire insecticide resistance phenotype? Trans. R. Soc. Trop. Med. Hyg. 102, 524−525.

Brown, A.W.A., Pal, R., 1971. Insecticide resistance in arthropods. WHO Monogr. Ser. 38.

Carrasco, D., Lefèvre, T., Moiroux, N., Pennetier, C., Chandre, F., Cohuet, A., 2019. Behavioural adaptations of mosquito vectors to insecticide control. Curr. Opin. Insect. Sci 34, 48−54. https://doi.org/10.1016/j.cois.2019.03.005.

Carson, R., 1962. Silent Spring. Houghton Mifflin, Boston.

Casida, J.E., 1970. Mixed-function oxydases involvement in the biochemistry. J. Agri. Food Chem. 18, 753−772.

Casida, J.E., 1973. Pyrethrum, the Natural Insecticide. Academic Press, New York, NY, USA.

Casimiro, S.L.R., Hemingway, J., Sharp, B.L., Coleman, M., 2007. Monitoring the operational impact of insecticide usage for malaria control on *Anopheles funestus* from Mozambique. Malar. J. 6, 142.

Chandor-Proust, A., Bibby, J., Régent-Kloeckner, M., Roux, J., Guittard-Crilat, E., Poupardin, R., et al., 2013. The central role of mosquito cytochrome P450 CYP6Zs in insecticide detoxification revealed by functional expression and structural modelling. Biochem. J. 455, 75−85.

Chandre, F., Darriet, F., Darder, M., Cuany, A., Doannio, J.M.C., Pasteur, N., et al., 1998. Pyrethroid resistance in *Culex quinquefasciatus* from West Africa. Med. Vet. Entomol. 12, 359−366.

Chandre, F., Darriet, F., Manguin, S., Brengues, C., Carnevale, P., Guillet, P., 1999. Pyrethroid cross resistance spectrum among populations of *Anopheles gambiae* s.s. from Cote d'Ivoire. J. Am. Mosq. Control Assoc. 15, 53−59.

Chandre, F., Darriet, F., Duchon, S., Finot, L., Richet, I.P., Manguin, S., et al., 2000. Modifications of pyrethroid effects associated with *kdr* mutation in *Anopheles gambiae*. Med. Vet. Entomol. 14, 81−88.

Chaturvedi, U.C., Shrivastava, R., 2004. Dengue haemorrhagic fever: a global challenge. Indian J. Med. Microbiol. 22, 5−6.

Chevillon, C., Bourguet, D., Rousset, F., Pasteur, N., Raymond, M., 1997. Pleiotropy of adaptive changes in populations: comparisons among insecticide resistance genes in *Culex pipiens*. Genet. Res. 70, 195−204.

Chevillon, C., Bernard, C., Marquine, M., Pasteur, N., 2001. Resistance to *Bacillus sphaericus* in *Culex pipiens* (Diptera: Culicidae): interaction between recessive mutants and evolution in southern France. J. Med. Entomol. 38, 657−664.

Chitnis, N., Schapira, A., Smith, T., Steketee, R., 2010. Comparing the effectiveness of malaria vector-control interventions through a mathematical model. Am. J. Trop. Med. Hyg. 83, 230−240.

Chiu, T.-L., Wen, Z., Rupasinghe, S.G., Schuler, M.A., 2008. Comparative molecular modeling of *Anopheles gambiae* CYP6Z1, a mosquito P450 capable of metabolizing DDT. Proc. Natl. Acad. Sci. 105, 8855−8860.

Claudianos, C., Russell, R.J., Oakeshott, J.G., 1999. The same amino acid substitution in orthologous esterases confers organophosphate resistance on the house fly and a blowfly. Insect Biochem. Mol. Biol. 29, 675−686.

Coleman, M., Casimiro, S.L.R., Hemingway, J., Sharp, B., 2008. Operational impact of DDT reintroduction for malaria control on *Anopheles arabiensis* in Mozambique. J. Med. Entomol. 45, 885−890.

Committee on Strategies for the Management of Pesticide Resistant Pest Populations National Research Council, 1986. Pesticide Resistance: Strategies and Tactics for Management. The National Academies Press.

Corbel, V., N'Guessan, R., 2013. Distribution, mechanisms, impact and management of insecticide resistance in malaria vectors: a pragmatic review. In: Manguin, S. (Ed.), *Anopheles* Mosquitoes—New Insights into Malaria Vectors. InTech Open Access, Rijeka.

Cordova, D., Benner, E.A., Sacher, M.D., Rauh, J.J., Sopa, J.S., Lahm, G.P., et al., 2006. Anthanilic diamides: a new class of insecticides with a novel mode of action, ryanodine receptor activation. Pestic. Biochem. Physiol. 84, 196−214.

Cornel, A.J., Stanich, M.A., McAbee, R.D., Mulligan, F.S., 2002. High level methoprene resistance in the mosquito Ochlerotatus nigromaculis (Ludlow) in Central California. Pest. Manag. Sci. 58, 791−798.

Curtis, C., 1985. Theoretical models of the use of insecticide mixtures for management of resistance. Bull. Entomol. Res. 75, 259−265.

Dabiré, K.R., Diabaté, A., Namontougou, M., Djogbenou, L., Kengne, P., Simard, F., et al., 2009. Distribution of insensitive acetylcholinesterase (ace-1R) in *Anopheles gambiae* s.l. populations from Burkina Faso (West Africa). Trop. Med. Int. Health 14, 396−403.

Darboux, I., Charles, J.F., Pauchet, Y., Pauron, D., 2007. Transposon-mediated resistance to *Bacillus sphaericus* in a field-evolved population of *Culex pipiens* (Diptera: Culicidae). Cell Microbiol. 9, 2022−2029.

David, J., Ismail, H.M., Chandor-proust, A., Paine, M.J.I., 2013. Role of cytochrome P450s in insecticide resistance: impact on the control of mosquito-borne diseases and use of insecticides on Earth. Philos. Trans. R. Soc. B Biol. Sci. 368, 20120429.

Davidson, G., 1956. Insecticide resistance in *Anopheles gambiae* Giles: a case of simple Mendelian inheritance. Nature 178, 863−864.

Despres, L., David, J.-P.P., Gallet, C., Després, L., David, J.-P.P., Gallet, C., 2007. The evolutionary ecology of insect resistance to plant chemicals. Trends Ecol. Evol. 22, 298−307.

Després, L., Stalinski, R., Tetreau, G., Paris, M., Bonin, A., Navratil, V., et al., 2014. Gene expression patterns and sequence polymorphisms associated with mosquito resistance to *Bacillus thuringiensis israelensis* toxins. BMC Genom. 15, 926.

Diabate, A., Baldet, T., Chandre, F., Akoobeto, M., Guiguemde, T.R., Darriet, F., Brengues, C., Guillet, P., Hemingway, J., Small, G.J., et al., 2002. The role of agricultural use of insecticides in resistance to pyrethroids in *Anopheles gambiae* s.l. in Burkina Faso. Am. J. Trop. Med. Hyg. 67, 617−622.

Dialynas, E., Topalis, P., Vontas, J., Louis, C., 2009. Miro and IRbase: IT tools for the epidemiological monitoring of insecticide resistance in mosquito disease vectors. PLoS Neglected Trop. Dis. 3, e465.

Djogbenou, L., Weill, M., Hougard, J.M., Raymond, M., Akogbeto, M., Chandre, F., 2007. Characterization of insensitive acetylcholinesterase (ace-1R) in *Anopheles gambiae* (Diptera:Culicidae): resistance levels and dominance. J. Med. Entomol. 44, 805−810.

Djogbénou, L., Chandre, F., Berthomieu, A., Dabiré, R.K., Koffi, A., Alout, H., et al., 2008. Evidence of introgression of the ace-1R mutation and of the ace-1 duplication in west African *Anopheles gambiae* s.s. PLoS One 3, e2172.

Djogbénou, L., Labbé, P., Chandre, F., Pasteur, N., Weill, M., 2009. Ace-I duplication in *Anopheles gambiae*: a challenge for malaria control. Malar. J. 8, 70.

Djogbenou, L., Noel, V., Agnew, P., 2010. Costs of insensitive acetylcholinesterase insecticide resistance for the malaria vector *Anopheles gambiae* homozygous for the G119S mutation. Malar. J. 9, 12.

Djogbénou, L.S., Assogba, B., Essandoh, J., Constant, E.A.V., Makoutodé, M., Akogbéto, M., et al., 2015. Estimation of allele-specific ace-1 duplication in insecticide-resistant *Anopheles* mosquitoes from West Africa. Malar. J. 14, 507.

Dong, K., Du, Y., Rinkevich, F., Nomura, Y., Xu, P., Wang, L., et al., 2014. Molecular biology of insect sodium channels and pyrethroid resistance. Insect Biochem. Mol. Biol. 50, 1−17.

Denholm, I., Rowland, M.W., 1992. Tactics for managing pesticide resistance in arthropods: theory and practice. Annu. Rev. Entomol. 37, 91−112.

Du, Y., Nomura, Y., Satar, G., Hu, Z., Nauen, R., He, S.Y., et al., 2013. Molecular evidence for dual pyrethroid-receptor sites on a mosquito sodium channel. Proc. Natl. Acad. Sci. U.S.A. 110, 11785−11790. https://doi.org/10.1073/pnas.1305118110.

Duron, O., Labbé, P., Berticat, C., Rousset, F., Guillot, S., Raymond, M., et al., 2006. High Wolbachia density correlates with cost of infection for insecticide resistant *Culex pipiens* mosquitoes. Evolution 60, 303−314.

Edi, C.V., Djogbénou, L., Jenkins, A.M., Regna, K., Muskavitch, M.A.T., Poupardin, R., et al., 2014. CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple insecticide resistance in the malaria mosquito *Anopheles gambiae*. PLoS Genet. 10, e1004236.

Elissa, N., Mouchet, J., Rivi_ere, F., Meunier, J.Y., Yao, K., 1993. Resistance of *Anopheles gambiae* s.s. to pyrethroids in Cote d'Ivoire. Ann. Soc. Belg. Med. Trop. 73, 291−294.

Elliot, R., Ramakrishna, V., 1956. Insecticide resistance in *Anophles gambiae* giles. Nature 177, 523−533.

Enayati, A.A., Ranson, H., Hemingway, J., 2005. Insect glutathione transferases and insecticide resistance. Insect Mol. Biol. 14, 3−8.

Enayati, A.A., Hemingway, J., 2006. Pyrethroid insecticide resistance and treated bednets efficacy in malaria control. Pestic. Biochem. Physiol. 84, 116−126.

Etang, J., Vicente, J.L., Nwane, P., Chouaibou, M., Morlais, I., Do Rosario, V.E., et al., 2009. Polymorphism of intron-1 in the voltage-gated sodium channel gene of *Anopheles gambiae* s.s. populations from Cameroon with emphasis on insecticide knockdown resistance mutations. Mol. Ecol. 18, 3076−3086.

Etang, J., Mandeng, S.E., Nwane, P., Awono-Ambene, H.P., Bigoga, J.D., Ekoko, W.E., Binyang, A.J., et al., 2022. Patterns of *kdr*-l995F allele emergence alongside detoxifying enzymes associated with deltamethrin resistance in *Anopheles gambiae* s.l. from North Cameroon. Pathogens 11, 253.

Farenhorst, M., Mouatcho, J.C., Kikankie, C.K., Brooke, B.D., Hunt, R.H., Thomas, M.B., et al., 2009. Fungal infection counters insecticide resistance in African malaria mosquitoes. Proc. Natl. Acad. Sci. 106, 17443−17447.

Faucon, F., Dusfour, I., Gaude, T., Navratil, V., Boyer, F., Chandre, F., et al., 2015. Identifying genomic changes associated with insecticide resistance in the dengue mosquito *Aedes aegypti* by deep targeted sequencing. Genome Res. 25, 1347−1359.

Fawaz, E.Y., Zayed, A.B., Fahmy, N.T., Villinski, J.T., Hoel, D.F., Diclaro 2nd, J.W., 2016. Pyrethroid insecticide resistance mechanisms in the adult Phlebotomus papatasi (Diptera: psychodidae). J. Med. Entomol. 53, 620−628. https://doi.org/10.1093/jme/tjv256.

Feliciangeli, M.D., Mazzarri, M.B., Campbell-Lendrum, D., Maroli, M., Maingon, R., 2003. Cutaneous leishmaniasis vector control perspectives using lambda-cyhalothrin residual house spraying in El Ingenio, Miranda State, Venezuela. Trans. R. Soc. Trop. Med. Hyg. 97, 641−646.

Feyereisen, R., 1995. Molecular biology of insecticide resistance. Toxicol. Lett. 82, 83−90, 98.

Feyereisen, R., 2005. In: Gilbert, L.I., Iatrou, K., Gill, S.S. (Eds.), Insect Cytochrome P450. Elsevier, Oxford, UK, pp. 1−77.

Feyereisen, R., Dermauw, W., Van Leeuwen, T., 2015. Genotype to phenotype, the molecular and physiological dimensions of resistance in arthropods. Pestic. Biochem. Physiol. 121, 61−77.

Ffrench-Constant, R.H., Anthony, N., Aronstein, K., Rocheleau, T., Stilwell, G., Richard, H., 2000. Cyclodiene insecticide resistance: from molecular to population genetics. Annu. Rev. Entomol. 45, 449−466.

Ffrench-Constant, R.H., Daborn, P.J., Le Goff, G., 2004. The genetics and genomics of insecticide resistance. Trends Genet. 20, 163−170.

Fine, B.C., 1963. The present status of resistance to pyrethroid insecticides. Pyrethrum Post. 7, 18−21.

Foster, S.P., Harrington, R., Devonshire, A.L., Denholm, I., Devine, G.J., Kenward, M.G., 1996. Comparative survival of insecticide-susceptible and resistant peach-potato aphids, *Myzus persicae* (Sulzer) (Hemiptera: aphididae), in low temperature field trials. Bull. Entomol. Res. 86, 17−27.

Fotakis, E.A., Giantsis, I.A., Demir, S., Vontas, J.G., Chaskopoulou, A., 2018. Detection of pyrethroid resistance mutations in the major leishmaniasis vector *Phlebotomus papatasi*. J. Med. Entomol. 55, 1225−1230. https://doi.org/10.1093/jme/tjy066.

Fournier, D., Bride, J.M., Mouches, C., Raymond, M., Magnin, M., Bergé, J.B., et al., 1987. Biochemical characterization of the esterases A1 and B1 associated with organophosphate resistance in the *Culex pipiens* complex. Pestic. Biochem. Physiol. 27, 211−217.

Fournier, D., Mutéro, A., 1994. Modification of acetylcholinesterase as a mechanism of resistance to insecticides. Comp. Biochem. Physiol. 108, 19−31.

Gao, J.R., Yoon, K.S., Lee, S.H., Takano-Lee, M., Edman, J.D., Meinking, T.L., et al., 2003. Increased frequency of the T929I and L932F mutations associated with knockdown resistance in permethrin-resistant populations of the human head louse, *Pediculus capitis*, from California, Florida, and Texas. Pestic. Biochem. Physiol. 77, 115−124.

Georghiou, G.P., Lagunes-Tejada, A., 1991. Cases of resistance in insecta. In: The Occurrence of Resistance to Pesticides in Arthropods. Food and Agriculture Organization of the United Nations, Rome, pp. 39−141.

Georghiou, G.P., Pasteur, N., 1978. Electrophoretic esterase patterns in insecticide-resistant and susceptible mosquitoes. J. Econ. Entomol. 71, 201−205.

Gomes, B., Purkait, B., Deb, R.M., Rama, A., Singh, R.P., Foster, G.M., et al., 2017. Knockdown resistance mutations predict DDT resistance and pyrethroid tolerance in the visceral leishmaniasis vector *Phlebotomus argentipes*. PLoS Neglected Trop. Dis. 11, e0005504. https://doi.org/10.1371/journal.pntd.0005504.

Grigoraki, L., Lagnel, J., Kioulos, I., Kampouraki, A., Morou, E., Labbé, P., et al., 2015. Transcriptome profiling and genetic study reveal amplified carboxylesterase genes implicated in temephos resistance, in the Asian tiger mosquito *Aedes albopictus*. PLoS Neglected Trop. Dis. 9, e0003771.

Guglielmone, A.A., Volpogni, M.M., Scherling, N., Cobeñas, M.M., Mangold, A.J., Anziani, O.S., Ioppolo, M., Doscher, M., 2000. Chlorfenapyr ear tags to control *Haematobia irritans* (L.) (Diptera: muscidae) on cattle. Vet. Parasitol. 93. https://doi.org/10.1016/S0304-4017(00)00335-6.

Guillemaud, T., Raymond, M., Tsagkarakou, A., Bernard, C., Rochard, P., Pasteur, N., 1999. Quantitative variation and selection of esterase gene amplification in *Culex pipiens*. Heredity 83, 87−99.

Guillet, P., N'Guessan, R., Darriet, F., et al., 2001. Combined pyrethroid and carbamate 'two-inone' treated mosquito nets: field efficacy against pyrethroid-resistant *Anopheles gambiae* and *Culex quinquefasciatus*. Med. Vet. Entomol. 15, 105−112.

Hancock, P.A., et al., 2018. Associated patterns of insecticide resistance in field populations of malaria vectors across Africa. Proc. Natl. Acad. Sci. 115, 5938−5943.

Hardstone, M.C., Leichter, C.A., Scott, J.G., 2009. Multiplicative interaction between the two major mechanisms of permethrin resistance, *kdr* and cytochrome P450-monooxyge-nase detoxification, in mosquitoes. J. Evol. Biol. 22, 416−423.

Hemingway, J., Ranson, H., 2000. Insecticide resistance in insect vectors of human disease. Annu. Rev. Entomol. 45, 371−391.

Hemingway, J., Penilla, R.P., Rodriguez, A.D., James, B.M., Edge, W., Rogers, H., Rodriguez, M.H., 1997. Resistance management strategies in malaria vector mosquito control. A largescale field trial in Southern Mexico. Pestic. Sci. 51, 375−382.

Hemingway, J., Hawkes, N.J., McCarroll, L., Ranson, H., 2004. The molecular basis of insecticide resistance in mosquitoes. Insect Biochem. Mol. Biol. 34, 653−665.

Hemingway, J., et al., 2016. Averting a malaria disaster: will insecticide resistance derail malaria control? Lancet 387, 1785−1788.

Hollingworth, R.M., Dong, K., 2008. In: Whalon, M.E., Mota-Sanchez, D., Hollingworth, R.M. (Eds.), The Biochemical and Molecular Genetic Basis of Resistance in Arthropods. CAB International, Cambridge, MA, p. 192.

Hougard, J.M., Escaffre, H., Darriet, F., Lochouarn, L., Riviere, F., Back, C., 1992. An episode of resistance to permethrin in larvae of *Simulium squamosum* (Diptera: Simuliidae) from Cameroon after 3 years of control. J. Am. Mosq. Control Assoc. 8, 184−186.

Hougard, J.M., Poudiougo, P., Guillet, P., Back, C., Akpoboua, L.K., Quillevere, D., 1993. Criteria for the selection of larvicides by the onchocerciasis control programme in West Africa. Ann. Trop. Med. Parasitol. 87, 435−442.

Huchard, E., Martinez, M., Alout, H., Douzery, E.J.P., Lutfalla, G., Berthomieu, A., et al., 2006. Acetylcholinesterase genes within the Diptera: takeover and loss in true flies. Proc. Biol. Sci. 273, 2595−2604.

Ibrahim, S.S., Riveron, J.M., Stott, R., Irving, H., Wondji, C.S., 2016. The cytochrome P450 CYP6P4 is responsible for the high pyrethroid resistance in knockdown resistance-free *Anopheles arabiensis*. Insect Biochem. Mol. Biol. 68, 23−32.

IRAC, 2010. Prevention and management of insecticide resistance in vectors of public health importance. In: Resistance Management for Sustainable Agriculture and Improved Public Health, second ed. Insecticide Resistance Action Commitee, p. 72.

Ishaaya, I., Horowitz, A.R., 2015. Novel phenoxy juvenile hormone analog (pyriproxyfen) suppresses embryogenesis and adult emergence of sweetpotato whitefly (Homoptera: Aleyrodidae). J. Econ. Entomol. 85. https://doi.org/10.1093/jee/85.6.2113.

Johnson, P.T.J., de Roode, J.C., Fenton, A., 2015. Why infectious disease research needs community ecology. Science 349, 1259504.

Kakarunamoorthi, K., Sabesan, S., 2013. Insecticide resistance in insect vectors of diseases with spatial reference to mosquitoes: a potential threat to global public health. Health Scope 2, 4−18. https://doi.org/10.17795/jhealthscope-9840.

Kamgang, B., Ngoagouni, C., Manirakiza, A., Nakoune, E., Paupy, C., Kazanji, M., 2013. Temporal patterns of abundance of *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae) and mitochondrial DNA analysis of *Ae. albopictus* in the Central African Republic. PLoS Neglected Trop. Dis. 7, e2590.

Kasai, S., Komagata, O., Itokawa, K., Shono, T., Ng, L.C., Kobayashi, M., et al., 2014. Mechanisms of pyrethroid resistance in the dengue mosquito vector, *Aedes aegypti*: target site insensitivity, penetration, and metabolism. PLoS Neglected Trop. Dis. 8, e2948.

Kelly-Hope, L.A., Ranson, H., Hemingway, J., 2008. Lessons from the past: managing insecticide resistancein malaria control and eradication programmes. Lancet Infect. Dis. 8, 387−389.

Labbé, P., Berthomieu, A., Berticat, C., Alout, H., Raymond, M., Lenormand, T., et al., 2007a. Independent duplications of the acetylcholinesterase gene conferring insecticide resistance in the mosquito *Culex pipiens*. Mol. Biol. Evol. 24, 1056−1067.

Labbé, P., Berticat, C., Berthomieu, A., Unal, S., Bernard, C., Weill, M., et al., 2007b. Forty years of erratic insecticide resistance evolution in the mosquito *Culex pipiens*. PLoS Genet. 3, e205.

Labbé, P., Sidos, N., Raymond, M., Lenormand, T., 2009. Resistance gene replacement in the mosquito *Culex pipiens*: fitness estimation from long term cline series. Genetics 182, 303−312.

Labbé, P., Milesi, P., Yébakima, A., Pasteur, N., Weill, M., Lenormand, T., 2014. Gene-dosage effects on fitness in recent adaptive duplications: ace-1 in the mosquito *Culex pipiens*. Evolution 68, 2092−2101.

Lahm, G.P., Cordova, D., Barry, J.D., 2009. New and selective ryanodine receptor activators for insect control. Bioorg. Med. Chem. 17, 4127.

Lasota, J.A., Dybas, R.A., 1991. Avermectins, a novel class of compounds: implications for use in arthropod pest control. Annu. Rev. Entomol. 36, 91−117.

Lee, S.H., Yoon, K.S., Williamson, M.S., et al., 2000. Molecular analysis of kdr-like resistance in permethrin-resistant strains of head lice, Pediculus capitis. Pestic. Biochem. Physiol. 66, 130−143.

Lee, S.H.G.J., Yoon, K.S., Mumcuoglu, K.Y., et al., 2003. Sodium channel mutations associated with knockdown resistance in the human head louse, Pediculus capitis (De Geer). Pestic. Biochem. Physiol. 75, 79−91.

Lenormand, T., Guillemaud, T., Bourguet, D., Raymond, M., 1998. Appearance and sweep of a gene duplication: adaptive response and potential for new functions in the mosquito *Culex pipiens*. Evolution 52, 1705−1712.

Lenormand, T., Bourguet, D., Guillemaud, T., Raymond, M., 1999. Tracking the evolution of insecticide resistance in the mosquito *Culex pipiens*. Nature 400, 861−864.

Li, X., Schuler, M.A., Berenbaum, M.R., 2007. Molecular mechanisms of metabolic resistance to synthetic and natural xenobiotics. Annu. Rev. Entomol. 52, 231−253.

Lumjuan, N., McCarroll, L., Prapanthadara, L., Hemingway, J., Ranson, H., 2005. Elevated activity of an Epsilon class glutathione transferase confers DDT resistance in the dengue vector, *Aedes aegypti*. Insect Biochem. Mol. Biol. 35, 861−871.

Lumjuan, N., Rajatileka, S., Changsom, D., Wicheer, J., Leelapat, P., Prapanthadara, L., et al., 2011. The role of the *Aedes aegypti* Epsilon glutathione transferases in conferring resistance to DDT and pyrethroid insecticides. Insect Biochem. Mol. Biol. 41, 203−209.

Lund, A.E., 1984. Pyrethroid modification of sodium channel: current concepts. Pestic. Biochem. Physiol. 22, 161−168.

Mahande, A.M., Dusfour, I., Matias, J.R., Kweka, E.J., 2012. Knockdown resistance, rdl alleles, and the annual entomological inoculation rate of wild mosquito populations from lower moshi, northern Tanzania. J. Global Infect. Dis. 4, 114−119.

Mandal, R., Kumar, V., Kesari, S., Das, P., 2019. Assessing the combined effects of household type and insecticide effectiveness for kala-azar vector control using indoor residual spraying: a case study from North Bihar. India Parasit. Vectors 12, 409. https://doi.org/10.1186/s13071-019-3670-x.

Marcombe, S., Poupardin, R., Darriet, F., Reynaud, S., Bonnet, J., Strode, C., et al., 2009. Exploring the molecular basis of insecticide resistance in the dengue vector *Aedes aegypti*: a case study in Martinique Island (French West Indies). BMC Genom. 10, 494.

Marcombe, S., Darriet, F., Agnew, P., Etienne, M., Yp-Tcha, M.M., Yébakima, A., Corbel, V., 2011. Field efficacy of new larvicide products for control of multi-resistant *Aedes aegypti* populations in Martinique (French West Indies). Am. J. Trop. Med. Hyg. 84, 118−126.

Martinez-Torres, D., Chevillon, C., Bergé, J.B., Pauron, D., Brun-Barale, A., Bergé, J.B., et al., 1999. Voltage-dependent Naþ channels in pyrethroid-resistant *Culex pipiens* L. mosquitoes. Pestic. Sci. 55, 1012−1020.

Matsuda, K., Buckingham, S.D., Kleier, D., Rauh, J.J., Grauso, M., Sattelle, D.B., 2001. Neonicotinoids: insecticides acting on insect nicotinic acetylcholine receptors. Trends Pharmacol. Sci. 22, 573−580.

Matsuda, K., Kikuta, Y., Haba, A., Nakayama, K., Katsuda, Y., Hatanaka, A., Komai, K., 2005. Biosynthesis of pyrethrin I in seedlings of *Chrysanthemum cinerariaefolium*. Phytochem 66, 1529−1535.

Maunder, J.W., 1971. Resistance to organochlorine insecticides in head lice and trials using alternative compounds. Med. Off. 125, 27−29.

McKenzie, J.A., 1996. Ecological and Evolutionary Aspects of Insecticide Resistance. Academic Press, Austin, Texas, USA.

Melander, A., 1914. Can insects become resistants to sprays? J. Econ. Entomol. 7, 167−172.

Michalakis, Y., Renaud, F., 2009. Malaria: evolution in vector control. Nature 462, 298−300.

Mitchell, S.N., Stevenson, B.J., Muller, P., Wilding, C.S., Egyir-Yawson, A., Field, S.G., et al., 2012. Identification and validation of a gene causing cross-resistance between insecticide classes in *Anopheles gambiae* from Ghana. Proc. Natl. Acad. Sci. U.S.A. 109, 6147−6152.

Mittal, P.K., 2003. Biolarvicides in vector control: challenges and prospects. J. Vector Borne Dis. 40, 20−32.

Moiroux, N., Gomez, M.B., Pennetier, C., Elanga, E., Dj_enontin, A., Chandre, F., et al., 2012. Changes in *Anopheles funestus* biting behavior following universal coverage of long-lasting insecticidal nets in Benin. J. Infect. Dis. 206, 1622−1629.

Moore, J.H., Williams, S.M., 2005. Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. Bioessays 27, 637−646.

Mordue, L.A.J., Blackwell, A., 1993. Azadirachtin: an update. J. Insect Physiol. 39, 903−924. https://doi.org/10.1016/0022-1910(93)90001-8.

Mosha, F.W., Lyimo, I.N., Oxborough, R.M., et al., 2008. Experimental hut evaluation of the pyrrole insecticide chlorfenapyr on bed nets for the control of *Anopheles arabiensis* and *Culex quinquefasciatus*. Trop. Med. Int. Health 13, 644−652.

Mouches, C., Pasteur, N., Bergé, J.B., Hyrien, O., Raymond, M., Robert de Saint Vincent, B., et al., 1986. Amplification of an esterase gene is responsible for insecticide resistance in a California Culex mosquito. Science 233, 778−780.

Mouches, C., Magnin, M., Bergé, J.B., de Silvestri, M., Beyssat, V., Pasteur, N., et al., 1987. Overproduction of detoxifying esterases in organophosphate-resistant Culex mosquitoes and their presence in other insects. Proc. Natl. Acad. Sci. U.S.A. 84, 2113−2116.

Muthusamy, R., Shivakumar, M.S., 2015. Susceptibility status of Aedes aegypti (L.) (Diptera: Culicidae) to temephos from three districts of Tamil Nadu, India. J. Vector Borne Dis. 52, 159−165.

Nabeshima, T., Mori, A., Kozaki, T., Iwata, Y., Hidoh, O., Harada, S., et al., 2004. An amino acid substitution attributable to insecticide-insensitivity of acetylcholinesterase in a Japanese encephalitis vector mosquito, *Culex tritaeniorhynchus*. Biochem. Biophys. Res. Commun. 313, 794−801.

Najera, J.A., Zaim, M., 2002. Malaria vector control decision making criteria and procedures for World Health Organization. In: Communicable Disease Control, Prevention and Eradication. WHO Pesticide Evaluation Scheme (WHOPES).

Nauen, R., 2007. Insecticide resistance in disease vectors of public health importance. Pest Manag. Sci. 63, 628−633.

Ndiath, M.O., Mazenot, C., Sokhna, C., Trape, J.-F., 2014. How the malaria vector *Anopheles gambiae* adapts to the use of insecticide-treated nets by African populations. PLoS One 9, e97700.

Nielsen-Leroux, C., Pasteur, N., Pretre, J., Charles, J., Sheikh, H.B., Chevillon, C., 2002. High resistance to *Bacillus sphaericus* binary toxin in *Culex pipiens* (Diptera: Culicidae): the complex situation of west Mediterranean countries. J. Med. Entomol. 39, 729−735.

Nkya, T.E., Akhouayri, I., Kisinza, W., David, J.-P., 2013. Impact of environment on mosquito response to pyrethroid insecticides: facts, evidences and prospects. Insect Biochem. Mol. Biol. 43, 407.

Norris, L.C., Main, B.J., Lee, Y., Collier, T.C., Fofana, A., Cornel, A.J., et al., 2015. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proc. Natl. Acad. Sci. 112, 201418892.

N'Guessan, R., 2009. Insecticide Resistance in the West African Malaria Vector *Anopheles gambiae* and Investigation of Alternative Tools for its Delay. Wageningen University.

N'Guessan, R., Corbel, V., Akogbéto, M., Rowland, M., 2007. Reduced efficacy of insecticide-treated nets and indoor residual spraying for malaria control in pyrethroid resistance area, Benin. Emerg. Infect. Dis. 13, 199−206. https://doi.org/10.3201/eid1302.060631.

Oakeshott, J.G., Home, I., Sutherland, T., Russell, R., 2003. The genomics of insecticide resistance. Genome Biol. 4, 202.

Oakeshott, J.G., Devonshire, A.L., Claudianos, C., Sutherland, T.D., Horne, I., Campbell, P.M., et al., 2005. Comparing the organophosphorus and carbamate insecticide resistance mutations in cholin- and carboxylesterases. Chem. Biol. Interact. 157e158, 269−275.

Osei-Atweneboana, M.Y., Wilson, M.D., Post, R.J., Boakye, D.A., 2001. Temephos-resistant larvae of *Simulium sanctipauli* associated with a distinctive new chromosome inversion in untreated drivers of south-western Ghana. Med. Vet. Entomol. 15, 113−116.

Oxborough, R.M., Mosha, F.W., Matowo, J., Mndeme, R., Feston, E., Hemingway, J., et al., 2008. Mosquitoes and bednets: testing the spatial positioning of insecticide on nets and the rationale behind combination insecticide treatments. Ann. Trop. Med. Parasitol. 102, 717−727.

Parker, P.J.A., Callaghan, A., 1997. Esterase activity and allele frequency in field populations of *Simulium equinum* (L.) (Diptera: Simuliidae) exposed to organophosphate pollution. Environ. Toxicol. Chem. 16, 2550−2555.

Pasteur, N., Sinegre, G., 1975. Esterase polymorphism and sensitivity to Dursban organophosphorous insecticide in *Culex pipiens pipiens* populations. Biochem. Genet. 13, 789−803.

Pasteur, N., Georghiou, G.P., Iseki, A., 1984. Variation in organophosphate resistance and esterase activity in *Culex quinquefasciatus* say from California. Genet. Sel. Evol. 16, 271−284.

Pathirage, D.R.K., Karunaratne, S.H.P.P., Senanayake, S.C., Karunaweera, N.D., 2020. Insecticide susceptibility of the sand fly leishmaniasis vector *Phlebotomus argentipes* in Sri Lanka. Parasit. Vectors 13, 246. https://doi.org/10.1186/s13071-020-04117-y.

Paul, A., Harrington, L.C., Zhang, L., Scott, J.G., 2005. Insecticide resistance in *Culex pipiens* from New York. J. Am. Mosq. Control Assoc. 21, 305−309.

Paupy, C., Ollomo, B., Kamgang, B., Moutailler, S., Rousset, D., Demanou, M., et al., 2010. Comparative role of *Aedes albopictus* and *Aedes aegypti* in the emergence of dengue and chikungunya in central Africa. Vector Borne Zoonotic Dis. 10, 259−266.

Pinto, J., Lynd, A., Vicente, J.L., Santolamazza, F., Randle, N.P., Gentile, G., et al., 2007. Multiple origins of knockdown resistance mutations in the Afrotropical mosquito vector *Anopheles gambiae*. PLoS One 2, e1243.

Pocquet, N., Milesi, P., Makoundou, P., Unal, S., Zumbo, B., Atyame, C., et al., 2013. Multiple insecticide resistances in the disease vector *Culex p. quinquefasciatus* from Western Indian Ocean. PLoS One 8, e77855.

Pocquet, N., Darriet, F., Zumbo, B., Milesi, P., Thiria, J., Bernard, V., et al., 2014. Insecticide resistance in disease vectors from Mayotte: an opportunity for integrated vector management. Parasit. Vect 7, 299.

Poupardin, R., Reynaud, S., Strode, C., Ranson, H., Vontas, J., David, J.P., 2008. Cross-induction ofdetoxification genes by environmental xenobiotics and insecticides in the mosquito Aedes aegypti: impact on larval tolerance to chemical insecticides. Insect Biochem. Mol. Biol. 38, 540−551.

Poupardin, R., Srisukontarat, W., Yunta, C., Ranson, H., 2014. Identification of carboxylesterase genes implicated in temephos resistance in the dengue vector Aedes aegypti. PLoS Neglected Trop. Dis. 8, e2743.

Ramphul, U., Boase, T., Bass, C., Okedi, L.M., Donnelly, M.J., Muller, P., 2009. Insecticide resistance and its association with target-site mutations in natural populations of *Anopheles gambiae* from eastern Uganda. Trans. R. Soc. Trop. Med. Hyg. 103, 1121−1126.

Ranson, H., Lissenden, N., 2016. Insecticide resistance in African *Anopheles* mosquitoes: a worsening situation that needs urgent action to maintain malaria control. Trends Parasitol. 32, 187−196.

Ranson, H., Jensen, B., Vulule, J.M., Wang, X., Hemingway, J., Collins, F.H., 2000. Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. Insect Mol. Biol. 9, 491−497.

Ranson, H., Claudianos, C., Ortelli, F., Abgrall, C., Hemingway, J., Sharakhova, M.V., et al., 2002. Evolution of supergene families associated with insecticide resistance. Science 298, 179−181.

Ranson, H., Burhani, J., Lumjuan, N., Black, W.C., 2010. Insecticide resistance in dengue vectors. Trop IKA. net 1, 277−284.

Ranson, H., N'guessan, R., Lines, J., Moiroux, N., Zinga Nkuni, J., Corbel, V., 2011. Pyrethroid resistance in African anopheline mosquitoes: what are the implications for malaria control? Trends Parasit. 27, 91−98.

Raymond, M., Fournier, D., Bride, J.M., Cuany, A., Bergé, J.B., Magnin, M., et al., 1986. Identification of resistance mechanisms in *Culex pipiens* (Diptera: Culicidae) from southern France: insensitive acetylcholinesterase and detoxifying oxidases. J. Econ. Entomol. 79, 1452−1458.

Raymond, M., Chevillon, C., Guillemaud, T., Lenormand, T., Pasteur, N.A., 1998. An overview of the evolution of overproduced esterases in the mosquito *Culex pipiens*. Phil. Trans. Roy. Soc. Lond. B 353, 1707−1711.

Raymond, M., Berticat, C., Weill, M., Pasteur, N., Chevillon, C., 2001. Insecticide resistance in the mosquito *Culex pipiens*: what have we learned about adaptation? Genetica 112−113, 1−10.

Read, A.F., Lynch, P.A., Thomas, M.B., 2009. How to make evolution-proof insecticides for malaria control. PLoS Biol. 7, e58.

Reddy, M.R., Overgaard, H.J., Abaga, S., Reddy, V.P., Caccone, A., Kiszewski, A.E., et al., 2011. Outdoor host seeking behaviour of *Anopheles gambiae* mosquitoes following initiation of malaria vector control on Bioko Island, Equatorial Guinea. Malar. J. 10, 184.

Remnant, E.J., Good, R.T., Schmidt, J.M., Lumb, C., Robin, C., Daborn, P.J., et al., 2013. Gene duplication in the major insecticide target site, Rdl, in Drosophila melanogaster. Proc. Natl. Acad. Sci. U.S.A. 110, 14705−14710.

Rivera, S.B., Swedlund, B.D., King, G.J., Bell, R.N., Hussey Jr., C.E., Shattuck-Eidens, D.M., Wrobel, W.M., Peiser, G.D., Poulter, C.D., 2001. Chrysanthemyl diphosphate synthase: isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from *Chrysanthemum cinerariaefolium*. Proc. Natl. Acad. Sci. U.S.A. 98, 4373−4378.

Riveron, J.M., Yunta, C., Ibrahim, S.S., Djouaka, R., Irving, H., Menze, B.D., et al., 2014. A single mutation in the GSTe2 gene allows tracking of metabolically based insecticide resistance in a major malaria vector. Genome Biol. 15, R27.

Roberts, D.R., Andre, R.G., 1994. Insecticide resistance issues in vector-borne disease control. Am. J. Trop. Med. Hyg. 50, 21–34.

Rocha, D.A., Costa, L.M.D., Pessoa, G.D.C., Obara, M.T., 2020. Methods for detecting insecticide resistance in sand flies: a systematic review. Acta Trop. 213, 105747. https://doi.org/10.1016/j.actatropica.2020.105747.

Russell, T.L., Govella, N.J., Azizi, S., Drakeley, C.J., Kachur, S.P., Killeen, G.F., 2011. Increased proportions of outdoor feeding among residual malaria vector populations following increased use of insecticide-treated nets in rural Tanzania. Malar. J. 10, 80.

Saavedra-Rodríguez, K., Urdaneta-Marquez, L., Rajatileka, S., Moulton, M., Flores, A.E., Fernandez-Salas, I., et al., 2007. A mutation in the voltage-gated sodium channel gene associated with pyrethroid resistance in Latin American *Aedes aegypti*. Insect Mol. Biol. 16, 785–798.

Sang, S., Yin, W., Bi, P., Zhang, H., Wang, C., Liu, X., et al., 2014. Predicting local dengue transmission in Guangzhou, China, through the influence of imported cases, mosquito density and climate variability. PLoS One 9, e102755.

Sardar, A.A., Saha, P., Chatterjee, M., Bera, D.K., Biswas, P., Maji, D., et al., 2018. Insecticide susceptibility status of *Phlebotomus argentipes* and polymorphisms in voltage-gated sodium channel (vgsc) gene in kalaazar endemic areas of West Bengal. India Acta Trop 185, 285–293. https://doi.org/10.1016/j.Acta Tropica 2018.06.005.

Scholte, E.-J., Ng'habi, K., Kihonda, J., Takken, W., Paaijmans, K., Abdulla, S., et al., 2005. An Entomopathogenic fungus for control of adult African malaria mosquitoes. Science 308, 1641–1642.

Scott, J.G., Shono, T., Georghiou, G.P., 1984. Genetic analysis of permethrin resistance in the house fly, *Musca domestica* L. Experientia 40, 1416–1418.

Seibold, S., et al., 2019. Arthropod decline in grasslands and forests is associated with landscape-level drivers. Nature 574, 671–674.

Shi, M.A., Lougarre, A., Alies, C., Fremaux, I., Tang, Z.H., Stojan, J., Fournier, D., 2004. Acetylcholinesterase alterations reveal the fitness cost of mutations conferring insecticide resistance. BMC Evol. Biol. 4, 5.

Shono, T., Zhang, L., Scott, J.G., 2004. Indoxacarb resistance in the house fly, Musca domestica. Pestic. Biochem. Physiol. 80, 106–112.

Siegfried, B.D., Young, L.J., 1993. Activity of detoxification enzymes in aquatic and terrestrial insects. Environ. Entomol. 22, 958–964.

Simon-Delso, N., Amaral-Rogers, V., Belzunces, L.P., Bonmatin, J.M., Chagnon, M., Downs, C., Furlan, L., et al., 2015. Systemic insecticides (Neonicotinoids and fipronil): Trends, uses, mode of action and metabolites. Envirl. Sci. Poll. Res. 22, 5–34.

Soderlund, D.M., Knipple, D.C., 2003. The molecular biology of knockdown resistance to pyrethroid insecticides. Insect Biochem. Mol. Biol. 33, 563–577.

Somwang, P., Yanola, J., Suwan, W., Walton, C., 2011. Enzymes-based resistant mechanism in pyrethroid resistant and susceptible *Aedes aegypti* strains from northern Thailand. Parasitol. Res. 109, 531–537.

Sougoufara, S., Diédhiou, S.M., Doucouré, S., Diagne, N., Semb_ene, P.M., Harry, M., et al., 2014. Biting by *Anopheles funestus* in broad daylight after use of long-lasting insecticidal nets: a new challenge to malaria elimination. Malar. J. 13, 125.

Sparks, T.C., Crouse, G.D., Durst, G., 2001. Natural products as insecticides: the biology, biochemistry and quantitative structure-activity relationships of spinosyns and spinosoids. Pest Manag. Sci. 57, 896–905.

Stevenson, B.J., Pignatelli, P., Nikou, D., Paine, M.J.I., 2012. Pinpointing P450s associated with pyrethroid metabolism in the dengue vector, *Aedes aegypti*: developing new tools to combat insecticide resistance. PLoS Neglected Trop. Dis. 6, e1595.

Strode, C., Wondji, C.S., David, J.-P., Hawkes, N.J., Lumjuan, N., Nelson, D.R., et al., 2008. Genomic analysis of detoxification genes in the mosquito *Aedes aegypti*. Insect Biochem. Mol. Biol. 38, 113–123.

Stump, A.D., Atieli, F.K., Vulule, J.M., Besansky, N.J., 2004. Dynamics of the pyrethroid knockdown resistance allele in western Kenyan populations of *Anopheles gambiae* in response to insecticide-treated bed net trials. Am. J. Trop. Med. Hyg. 70, 591–596.

Surendran, S.N., Karunaratne, S.H., Adams, Z., Hemingway, J., Hawkes, N.J., 2005. Molecular and biochemical characterization of a sand fly population from Sri Lanka: evidence for insecticide resistance due to altered esterases and insensitive acetylcholinesterase. Bull. Entomol. Res. 95, 371–380. https://doi.org/10.1079/ber2005368.

Tabashnik, B.E., Brévault, T., Carriere, Y., 2013. Insect resistance to Bt crops: lessons from the first billion acres. Nat. Biotechnol. 31, 510–521.

Tan, J., Galligan, J.J., Hollingworth, R.M., 2007. Agonist actions of neonicotinoids on nicotinic acetylcholine receptors expressed by cockroach neurons. Neurotoxicology 28, 829–842.

Tantely, M.L., Tortosa, P., Alout, H., Berticat, C., Berthomieu, A., Rutee, A., et al., 2010. Insecticide resistance in Culex pipiens quinquefasciatus and Aedes albopictus mosquitoes from La Reunion Island. Insect Biochem. Mol. Biol. 40, 317–324.

Tetreau, G., Alessi, M., Veyrenc, S., David, J., Reynaud, S., 2012. Fate of *Bacillus thuringiensis* subsp. *israelensis* in the field: evidence for spore recycling and differential persistence of toxins in leaf litters. Appl. Environ. Microbiol. 78, 8362.

Thompson, G.D., Dutton, R., Sparks, T.C., 2000. Spinosad—a case study: an example from a natural products discovery programme. Pest Manag. Sci. 56, 69.

Vais, H., Williamson, M.S., Devonshire, A.L., Usherwood, P.N.R., 2001. The molecular interactions of pyrethroid insecticides with insect and mammalian sodium channels. Pest Manag. Sci. 57, 877–888.

van den Berg, H., 2009. Global status of DDT and its alternatives for use in vector control to prevent disease. Environ. Health Perspect. 117, 1656–1663.

Vaughan, A., Hawkes, N.J., Hemingway, J., 1997. Co-amplification explains linkage disequilibrium of two mosquito esterase genes in insecticide-resistant. Culex quinquefasciatus Biochem. J. 325, 359–365.

Vontas, J.G., McCarroll, L., Karunaratne, S.H.P.P., Louis, C., Hurd, H., Hemingway, J., 2004. Does environmental stress affect insect-vectored parasite transmission? Physiol. Entomol. 29, 210–213.

Vontas, J., Kioulos, E., Pavlidi, N., Morou, E., della Torre, A., Ranson, H., 2012. Insecticide resistance in the major dengue vectors *Aedes albopictus* and *Aedes aegypti*. Pestic. Biochem. Physiol. 104, 126–131.

Weetman, D., Mitchell, S.N., Wilding, C.S., Birks, D.P., Yawson, A.E., Essandoh, J., et al., 2015. Contemporary evolution of resistance at the major insecticide target site gene ace-1 by mutation and copy number variation in the malaria mosquito *Anopheles gambiae*. Mol. Ecol. 24, 265–272.

Weill, M., Berticat, C., Raymond, M., Chevillon, C., 2000a. Quantitative polymerase chain reaction to estimate the number of amplified esterase genes in insecticide-resistant mosquitoes. Anal. Biochem. 285, 267–270.

Weill, M., Chandre, F., Brengues, C., Manguin, S., Akogbéto, M.C., Pasteur, N., et al., 2000b. The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae* s.s. through introgression. Insect Mol. Biol. 9, 451–455.

Weill, M., Fort, P., Berthomieu, A., Dubois, M.P., Pasteur, N., Raymond, M., 2002. A novel acetylcholinesterase gene in mosquitos' codes for the insecticide target and is non-homologous to the ace gene in Drosophila. Proc. Biol. Sci. 269, 2007–2016.

Weill, M., Lutfalla, G., Mogensen, K., Chandre, F., Berthomieu, A., Berticat, C., et al., 2003. Insecticide resistance in mosquito vectors. Nature 423, 423–426.

Weill, M., Berticat, C., Lutfalla, G., Pasteur, N., Philips, A., Fort, P., et al., 2004. Insecticide resistance: a silent base prediction. Curr. Biol. 14, R552–R553.

Werck-Reichhart, D., Feyereisen, R., 2000. Cytochromes P450: a success story. Genome Biol. 1.

Whalon, M.E., Mota-Sanchez, D., Hollingworth, R.M., 2008. Global Pesticide Resistance in Arthropods. CABI Publishing, Cambridge, MA.

Williamson, M.S., Martinez-Torres, D., Hick, C.A., Devonshire, A.L., 1996. Identification of mutations in the housefly para-type sodium channel gene associated with knockdown resistance (*kdr*) to pyrethroid insecticides. Mol. Gen. Genet. 252, 51−60.
Wilson, A.L., Courtenay, O., Kelly-Hope, L.A., Scott, T.W., Takken, W., Torr, S.J., et al., 2020. The importance of vector control for the control and elimination of vector-borne diseases. PLoS Neglected Trop. Dis. 14, e0007831. https://doi.org/10.1371/journal.pntd.0007831.
Wong, S.S., Poon, R.W., Wong, S.C., 2016. Zika virus infection-the next wave after dengue? J. Formos. Med. Assoc. 115, 226−242.
Wood, R.J., 1981. In: Bishop, J.A., Cook, L.M. (Eds.), Insecticide Resistance: Genes and Mechanisms. Academic Press, London, pp. 53−96.
World Health Organisation (WHO), 1957. Malaria section. Bull. World Health Organ. 16, 874.
World Health Organisation (WHO), 1992. Vector Resistance to Pesticides. Fifteenth Report of the Expert Committee on Vector Biology and Control Technical Series, vol. 818. WHO, Geneva.
World Health Organisation (WHO), 2006. Pesticides and Their Applications for the Control of Vectors and Pests of Public Health Importance, sixth ed. World Health Organization, Geneva, pp. 1−125.
World Health Organisation (WHO), 2008. The Global Burden of Disease: 2004 Update. World Health Organization, Geneva, Switzerland.
World Health Organisation (WHO), 2011. Global Insecticide Use for Vector-Borne Sisease Control, a 10-year Assessment (2000e2009). ISBN:9789241502153.
World Health Organisation (WHO), 2012a. The Global Plan for Insecticide Resistance Management in Malaria Vectors. WHO Global Malaria Programme.
World Health Organisation (WHO), 2012b. Insecticide resistance and vector control control: Tenth Report. In: Of the Expert Committee on Insecticides. Meeting World Health Organization.
World Health Organisation (WHO), 2015. World Health Statistics.
World Health Organisation (WHO), 2018. Global Report on Insecticide Resistance in Malaria Vectors. https://www.jstor.org/stable/resrep27882.10.
World Health Organisation (WHO), 2022. Determining Discriminating Concentrations of Insecticides for Monitoring Resistance in Mosquitoes: Report of a Multi-Centre Laboratory Study and WHO Expert Consultations.
World Health Organization (WHO), 2004. Global Strategic Framework for Integrated Vector Management. Geneva.
World Health Organization (WHO), 2019. Prequalified Lists: Vector Control Product*s*.
World Health Organisation (WHO), 2016. Test Procedures for Insecticide Resistance Monitoring in Malaria Vector Mosquitoes, second ed. World Health Organization, Geneva. updated in June 2018. file:///C:/Users/josye/Downloads/9789241511575-eng.pdf. (Accessed 27 September 2022).
Xu, Q., Wang, H., Zhang, L., Liu, N., 2006. *Kdr* allelic variation in pyrethroid resistant mosquitoes, *Culex quinquefasciatus* (S.). Biochem. Biophys. Res. Commun. 345, 774−780.
Yadouleton, A., Martin, T., Padonou, G., Chandre, F., Asidi, A., Djogbenou, L., Dabire, R., Aikpon, R., Boko, M., Glitho, I., et al., 2011. Cotton pest management practices and the selection of pyrethroid resistance in *Anopheles gambiae* population in northern Benin. Parasit. Vectors 4, 60.

Further reading

Alout, H., Labbé, P., Berthomieu, A., Pasteur, N., Weill, M., 2009. Multiple duplications of the rare ace-1 mutation F290V in *Culex pipiens* natural populations. Insect Biochem. Mol. Biol. 39, 884−891.
Hardstone, M.C., Scott, J.G., 2010. A review of the interactions between multiple insecticide resistance loci. Pestic. Biochem. Physiol. 97, 123−128.

Karunaratne, S.H.P.P., Hemingway, J., Jayawardena, K.G.I., Dassanayaka, V., Vaughan, A., 1995. Kinetic and molecular differences in the amplified and non-amplified esterases from insecticide-resistant and susceptible *Culex quinquefasciatus* mosquitoes. Biochem Biophys 270, 31124−31128.

Pasteur, N., Sin_egre, G., Gabinaud, A., 1981. Est-2 and Est-3 polymorphism in Culex pipiens L. from southern France in relation to organophosphate resistance. Biochem. Genet. 19, 499−508.

CHAPTER

14

Genetics of Tsetse Flies (Diptera: Glossinidae), Vectors of Human and Animal Trypanosomoses

Philippe Solano

Intertryp, University of Montpellier, IRD Cirad, Montpellier, France

The presence of the tsetse fly (*Diptera*: Glossinidae) in sub-Saharan Africa has greatly contributed to the demographic and economic shaping of the African continent throughout its history. Described since antiquity, this unique insect has never ceased, through the parasites it transmits, to kill the poorest, to repel invasions, to hinder the transports, to slow down settlement, to reduce to zero any possibility of breeding in certain regions. On the other hand, the tsetse poses as a "guardian of biodiversity," and its geographical restriction to sub-Saharan Africa has greatly limited the interest of the international community for this vector compared to others, for the diseases associated, and ultimately and most dramatically for the communities suffering from these, who are the "neglected" ones.

In this chapter, we will mostly focus on population genetics of some tsetse species because of their involvement as vectors of trypanosomes causing human and animal African Trypanosomoses, respectively, HAT and AAT. We will first remind the taxonomy of these insects as well as some of their important and sometimes unique life traits, before reviewing some recent results on tsetse genetics, and suggesting some possible directions for future work.

1. Introduction

HAT is a severe, deadly, meningoencephalitic disease that develops after an acute lymphatic infection. HAT comprises two distinct diseases: Gambian sleeping sickness (g-HAT, caused by *Trypanosoma b. gambiense*) is a chronic disease that occurs in West and Central Africa. Rhodesian sleeping sickness (r-HAT, caused by *T. b. rhodesiense*) is acute and rapidly fatal if not treated. It occurs in eastern and southern Africa. The annual incidence of HAT in the 1990s was estimated to be 300,000 but that had declined to 70,000 in the early years of the 21st century, and to less than 1000 from 2019 onwards (Franco et al., 2022). The 2030 objective of the WHO roadmap on Neglected Tropical Diseases (NTDs) is "elimination as interruption of transmission" verified for g-HAT, and "elimination as public health problem" for r-HAT (meaning less than 1 case out of 10,000 inhabitants per year in every health district) (WHO, 2020).

AAT affecting domestic cattle has been estimated to cost African agriculture US $4.5 billion per year via mortality, loss of food, dung, and drafting power (Swallow, 1998). AAT is endemic in 37 countries in sub-Saharan Africa and caused principally by four species of trypanosomes: *Trypanosoma*

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00033-1
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

(Trypanozoon) brucei, *T. (Nannomonas) congolense, T. (Duttonella) vivax*, and *T. (Nannomonas) simiae*. In addition, some other important trypanosome species are not tsetse-transmitted, including *T. (Trypanozoon) evansi* and *T. (Trypanozoon) equiperdum*. A progressive control pathway (PCP) has been recently proposed by FAO and Cirad, providing an interesting framework for countries who want to develop their strategy and tools for sustainable control-elimination of AAT (Diall et al., 2017).

2. Reminder on the Taxonomy of Tsetse Flies

Tsetse flies belong to the Glossinidae family that was created for these insects, comprising a single genus, *Glossina*. Glossinidae are close to Muscidae and belong to *Diptera*. Thirty-one species and subspecies of tsetse are recognized (Table 14.1 here). Their distribution is restricted to sub-Saharan Africa where they have been usually reported to extend over 10 million km^2, although field observations suggest they have nowadays lost a significant part of this distribution due to climate change and human pressure on their habitats (Courtin et al., 2010). Both sexes of tsetse are hematophagous. They have a characteristic mode of reproduction that brings them closer to pupiparous hematophagous Diptera (*Hippobosca*, *Melophagus*).

Tsetse are long, robust, brown-black to brown flies. Their length varies between 6 and 16 mm. Males are generally smaller than females. The wings, hyaline or slightly smoked, are, at rest, crossed above the abdomen, their posterior end exceeding its extremity. The classification of the genus *Glossina* is mainly based on the shape of the male and female genitalia and the geographic distribution, together with additional morphological characteristics such as color, shape of the antennae, presence

Table 14.1 Classification of tsetse Species and Subspecies

Genus	*Glossina*		
Subgenus	*Nemorhina* (palpalis group)	*Glossina* (morsitans group)	*Austenina* (fusca group)
Species/subspecies	*G. caliginea*	*G. austeni*	*G. brevipalpis*
	G. fuscipes fuscipes	*G. longipalpis*	*G. frezili*
	G. fuscipes martinii	*G. morsitans centralis*	*G. fusca congolensis*
	G. fuscipes quanzensis	*G. morsitans morsitans*	*G. fusca fusca*
	G. pallicera pallicera	*G. morsitans submorsitans*	*G. fuscipleuris*
	G. pallicera newsteadi	*G. pallidipes*	*G. haningtoni*
	G. palpalis gambiensis	*G. swynnertoni*	*G. longipennis*
	G. palpalis palpalis		*G. medicorum*
	G. tachinoides		*G. nashi*
			G. nigrofusca nigrofusca
			G. nigrofusca hopkinsi
			G. schwetzi
			G. severini
			G. tabaniformis
			G. vanhoofi

of bristles on thoracic pleura, etc. The genus was divided into three groups of species, or "subgenera" (reviewed in Solano et al., 2010a).

- Subgenus *Nemorhina* (Robineau-Desvoidy, 1830) ("Palpalis group")
  Almost all the species of this subgenus live in West and Central Africa, in vegetation close to water (forests, small islands forests, gallery forests, "sacred" woods, river thickets, banks of the lakes, mangroves, "niayes," etc.). Some species colonize coffee, cocoa plantations, mango, and banana plantations. Currently, only two tsetse species (and their subspecies) are vectors of g-HAT in this part of Africa, and they belong to this subgenus: *G. palpalis* and *G. fuscipes*. They are also becoming increasingly important for the transmission of animal trypanosomosis. They are the most resilient to humans and can adapt to peridomestic habitats and feed on humans and their domestic animals (pigs in particular).
- Subgenus *Glossina* sensu stricto (Zumpt, 1935) ("Morsitans" group)
  These species are less restricted to watercourses than the species of the precedent subgenus. They occur mainly in savannah woodland, dense thickets, or clear forests. Their distribution is linked to the presence of wildlife and cattle. Therefore, species relying on wildlife have lost most of their original habitat due to poaching and to human needs of land for agriculture. These species are dangerous vectors of AAT, and r-HAT in eastern and southern Africa. *G. longipalpis* has been suggested to be the "west African cousin" of *G. pallidipes*, although more work should be done on this (Krafsur, Maudlin, 2018).
- Subgenus *Austenina* (Townsend, 1921) ("Fusca" group)
  The species of this subgenus are the biggest in size (11−16 mm). They live in forest belts (rain forest, evergreen rain forest, forest−savannah mosaics, coastal mosaics, relic forests remainders, broad and dense gallery forests), except *G. longipennis* (arid savannahs of Kenya) and *G. brevipalpis* (thickets of East Africa). Increasing human activity and forest destruction tends to make them disappear. Some are excellent vectors of animal trypanosomoses, but the areas in which they occur have only a weak pastoral vocation. Very little work has been done tsetse of this group, although increasing interest is reported on *G. brevipalis* for possible regional control/eradication in some countries of southern Africa.

3. Some Important Life Traits of Tsetse regarding Population Genetics

Although life traits of tsetse have been described in several reviews or book chapters (e.g., Solano et al., 2010, E Krafsur in Tibayrenc 2017), we thought it would be useful for the reader to make a brief point on some of the most important ones that may be related to genetics.

3.1 Reproduction

Tsetse flies have a particular mode of reproduction named adenotrophic viviparity, which is only encountered in a small number of *Diptera* grouped in the biological group of pupiparous insects (Hippoboscidae*, Streblidae, Nycteribiidae*). In this group, females have a pocket similar to the uterus of the mammal, in which it preserves its larva until maturity. During its intrauterine life, the larva is fed by secretions of lactiferous glands annexed to the uterus, part of which relies on symbiotic bacteria (*Wigglesworthia* spp.). When the tsetse imago hatches, its genital organs are already formed and

mating can occur in the hours that follow. In natural conditions, almost all females are fertilized as soon as they exit from their puparium. One insemination only is generally sufficient to enable the female to produce larvae for several months. Some young females can accept several sexual unions to fill their spermatheca, but, in experimental conditions, beyond the 10th day, the proportion of females accepting more than one sexual union is only 0.7%.

The larvae are deposited by the mother in a favorable shaded place, generally corresponding to the diurnal resting places next to the hunting area. In riverine species (those of *Nemorhina* subgenus), laying places vary with seasons: they are generally located near water in the dry season and more distant in the rainy season to avoid flooding. In savannah species, they are much more diffuse and more difficult to detect. The anal product of the larvae of *G. m. centralis* and *G. m. morsitans* contains pheromones that attract pregnant females, leading to a concentration of larvae in particular sites. After a free life of a few minutes to 2 hours on the surface of the ground, the larvae bury itself a few centimeters, generally in a clay−sandy ground or at the bottom of palm trees, always in a shady place. In forest areas, pupae can be found at the base of the feather grasses of palm trees or in tree trunks containing humus. In savannah areas, they can be found under large roots, in holes, burrows, under fallen trunks, etc.

The entire process of metamorphosis occurs within the puparium, leading to the "perfect adult" insect. Duration is variable, according to species, sex and climatic conditions, in particular minimal and maximum ground temperatures. Pupae require sufficient ground moisture (over 60%), but flooding is fatal; deforestation, bush fires, compaction of the ground, are also disadvantages, together with parasitism. Within the critical temperature limits range allowing survival, the duration of pupation increases as the temperature decreases (between 20 days in the hot season and 80 days in the cold season). At 25°C, it is included between 25 and 30 days, females having shorter pupation durations than males, by approximately 2−3 days. During this period, the pupa lives only with the food stocks that the larva constituted during its intrauterine life. When metamorphoses are complete, the young fly leaves its puparium by breaking a circular slit in the anterior end, and the imago can fly in just a few hours after emergence.

3.2 Lifespan and Population Dynamics

Females live longer than males. Some 12-month survival records were obtained in the laboratory but, in natural conditions as a mean, females are believed to live for ∼3 months and males between one and two. The lifespan is very variable depending on the season: optimal in the rainy season, it decreases in the cold, and even more in the heat. Despite a sex-ratio close to one at hatching, females are generally more numerous in populations due to their longer lifespan. Fluctuations in the sex ratio that would be controlled by a factor related to the X chromosome were reported in *G. m. submorsitans* (Gooding and Krafsur, 2005). Apparent densities of tsetse populations depend on biotic and abiotic environmental factors, are very variable from one place to another and are mainly related to hatching rates of pupae and lifespan of adults, which themselves are governed by climatic factors and host availability.

One of the characteristics of tsetse flies is that every female will give birth to no more than four to five descendants in nature throughout their life. This very low number is compensated by their long life duration, a high probability of reaching the adult stage, and a very high resilience and ability to survive at low densities, especially for tsetse species of the palpalis group.

3.3 Natural Habitats

Tsetse flies need particular conditions to live, in particular temperature and moisture, that are encountered only in tropical areas. The normal ecoclimate corresponds to that of wooded areas receiving more than 600 mm of annual rainfall. The hygrometrical optima vary between 50% and 60% of relative hygrometry in savannah species and 65%−85% in riverine and forest species. The thermal optimum is about 25°C in all species. In general, pupae and adults suffer as soon as the temperature exceeds 36°C and die at temperatures that exceed 38−40°C. Adults are paralyzed by the cold and pupae cannot develop normally with temperatures lower than 16°C. Hot and dry climates (Saharan or sub-Saharan) and cold climates where the annual average temperature is below 20°C (highlands and southern Africa) are incompatible with imago needs and prevent preimaginal stages to reach the adult stage. Of course climate change is having significant impacts nowadays on tsetse distribution in both senses: in some areas tsetse disappear due to high temperatures and/or human pressure and its consequences on habitat, and in others (formerly too cold), they are able to colonize new areas (Lord et al., 2018).

Tsetse flies are thus closely related to vegetation, which is itself dependent on the presence of surface or underground water that increases the humidity of both atmosphere and soils. The type and density of canopy and underwood influences temperature, moisture, luminosity, easiness of flight, as well as the presence of animals that are tsetse fly hosts. Host abundance, availability and attractiveness will determine the maintenance, development, and density of tsetse populations.

The species of the subgenera *Nemorhina* and *Austenina*, which in majority need high relative hygrometry, are dependent on the ligneous vegetation of gallery forests or large forests. Those of the subgenus *Glossina s. str.*, which are more xerophilous, disperse largely in savannah woodlands during the rainy season but are restricted to vegetation neighboring water in the dry season. Within these habitats, flies choose places where the microclimate is the most favorable, and where there is enough host availability. In these vegetal associations, generally located at the border between two facies, of which one at least is woody, they find resting and reproduction places and hunting areas.

3.4 Dispersal, Diet, Activity Cycle

The activity of tsetse flies is mostly diurnal, and reduced to a few minutes per day, during moments of the day when environmental climatic conditions are the most favorable. The height of flight is low (0.5 m approximately) and the speed elevated (5−6 m per second). Displacements last a few consecutive seconds or minutes, in small successive jumps. Movements generally involve short distances because tsetse use an original energetic system based on proline-alanine that seems not very "performant." In 1 day, males fly for approximately 30−50 minutes, whereas females move for only 5 minutes. This explains the noticeable differences between sexes of the abrasion of the posterior edge of the wings, a method used to evaluate the average age of a population.

Riverine species (e.g., *G. palpalis gambiensis* in savannah habitat) mainly show linear dispersal. They follow the river by backward and forward motions, upstream or downstream, over short distances (200−300 m). In the dry season, displacements thus remain mainly longitudinal inside the galleries. In the rainy season, tsetse may exhibit radiate dispersal in the neighboring savannah and may be exceptionally able to cross from one hydrographic network to another if they are not too far. In the hot season, some individuals can fly long distances (17 km in 3 days for *G. tachinoides*; 22 km in 5 days

for *G. p. gambiensis*) along the same river, although this appears to constitute exceptions rather than a rule. The savannah and forest species move more randomly, generally over short distances (175 m per day for *G. pallidipes* in Kenya and less than 1 km for *G. m. submorsitans*). However, at the beginning of the rainy season, successive waves can occur, including a significant number of individuals. For example, recolonization of cleaned areas has been reported to occur at rates of approximately 5−15 km per annum for *G. m. submorsitans*, and of 3−10 km per annum for *G. m. morsitans* and *G. palpalis*.

In homogeneous landscapes of humid savannahs of West Africa, dispersal can be modeled as a random walk or a diffusion process, its continuous analogue, where the probability for a given fly to go in any direction is the same. In fragmented landscapes, this is no longer the case, flies being able to use environmental stimuli to find their way toward favorable vegetal formations.

Tsetse can follow herds or humans, or be passively transported over long distances on vehicles, boats, or trains, which can be source of reinvasions.

3.5 Tsetse Bacteriome

Three bacterial symbionts of tsetse have been described (reviewed in Krafsur and Maudlin, 2018). They are maternally inherited *Wolbachia pipientis (Proteobacteria: Alphaproteobacteria, Rickettsiales, Anaplasmataceae, Wolbachieae), Sodalis glossinidius (Proteobacteria: Gammaproteobacteria, Enterobacterales, Pectobacteriaceae),* and *Wigglesworthia glossinidia (Gammaproteobacteria: Enterobacterales, Erwiniaceae*). *W. glossinidia* is found in the guts of tsetse where it provides essential B-complex vitamins to its hosts and is thus an obligate symbiont and indicative of a long evolutionary relationship. *S. glossinidius* is a microaerophilic, secondary symbiont. Its prevalence and density vary greatly within and among species, and it seems to affect tsetse susceptibility to trypanosome infection. But many other bacteria species exist in tsetse, some of which just begin to be described (Geiger et al., 2015; Doudoumis et al., 2017). In addition, the exact role of bacteria in tsetse may go well beyond what has been classically described so far, since they may also influence mating choice through pheromone communication and ultimately play a role in speciation (Engl et al., 2018), while some of them may also be used in biological control such as paratransgenesis.

4. Populations Genetics of Tsetse Flies

Annotation of the tsetse genome in 2014 (IGGI, 2014) showed that it was a total size of 366 megabases (Mb), double the size of *Drosophila melanogaster*, although fewer genes are present: about 12,000, or 1/10th of the number of genes present in humans.

All tsetse flies examined cytologically have two pairs of metacentric autosomes and a sex bivalent. Many also have heterochromatic supernumerary chromosomes, and sex chromosome polymorphisms and aneuploidy have been recorded in wild *G. p. palpalis*.

E Krafsur (in Tibayrenc, 2017) reports a lower genetic variation in the tsetse flies compared to *Drosophila* and *Anopheles* mosquitoes, and attributes this to a smaller effective population size (*Ne*).

Low diversities in *G. morsitans centralis* and southern African *G. pallidipes* seem to reflect earlier demographic events including the 19th century rinderpest epizootic that virtually eliminated the morsitans group flies (Krafsur, 2009). However, this hardly applies to western Africa. In this area,

rinderpest did not have such impact, and riverine tsetse are known to be able to feed on any host, being opportunistic. They however show the same trend of low *Ne*. May be small effective population sizes are inherent to tsetse flies (at least to some species), and may also contribute to explain their resilience to control methods, and their ability to survive at low, undetectable, population densities during control campaigns.

4.1 Species Complexes, Subspecies and Ecotypes

Morsitans group taxa occur largely in wooded savannas of East and southern Africa. *G. morsitans* "subspecies" are strongly allopatric. *G. pallidipes* is widely but discontinuously distributed over East and southern Africa; many populations seem to occupy very small areas. Patchy distributions also characterize *G. austeni* and *G. brevipalpis*. Crosses among *G. morsitans* "subspecies" have been studied for years with unambiguous results: reduced fecundity, embryonic lethality and sterile, hybrid male progeny. For the *G. morsitans* subspecies, Krafsur proposed these should be elevated to species rank (Krafsur 2009) because of complete premating and postmating isolation between them. Using allozymes and microsatellites indicated that random mating (i.e., $F_{IS} \approx 0$) was the rule within populations, but that gene flow among populations within species of subgenus morsitans taxa was restricted (i.e., $F_{ST} >> 0$) and attributed to their inherently patchy distributions. As discussed, the effective population numbers of morsitans group taxa are small relative to the regions said to be environmentally favorable. Such populations demonstrate high levels of genetic drift, the direct consequence of generation-to-generation flux from chance alone in selectively neutral allele frequencies (reviewed by E Krafsur in Tibayrenc 2017).

In the palpalis group, *G. fuscipes* has also been described as an allopatric species complex, but prezygotic reproductive isolation among geographically and morphologically designated members of the *G. fuscipes* species complex has not been confirmed with genetic evidence (reviewed in Tibayrenc, 2017). Indeed, the genetic basis of such differentiation is unresolved and unclear. Five microsatellite loci suggested no significant reproductive isolation among *G. fuscipes* subspecies although $F_{ST} = 0.26$ was estimated between West DRC and western Kenya *G. f. fuscipes* (Dyer et al., 2011). Dyer et al. (2008) critically evaluated available data pertaining to speciation in *G. palpalis palpalis* and suggested the existence of cryptic species in these taxa.

Spatial variations in temperature and moisture conditions, and also host availability and density, govern the distribution of tsetse flies, and empirical evidence that spatially separated demes of *G. palpalis s.l.* have adapted to their different environments in western and central Africa may provide an explanation (de Meeus et al., 2015). This latter paper also highlights the importance and the link between landscape and population genetics for this species, which were then further studied in Bouyer et al. (2015) who pointed out the correlation between an analysis of "landscape friction" and genetic differentiation at the scale of western Africa for this species.

It may be useful for the reader to remind that when tsetse subspecies have been described (e.g., in *G. morsitans*, *G. fuscipes*, *G. palpalis*), crosses between subspecies always led to males hybrid sterility in the lab (Tibayrenc, 2017). In the fantastic fieldwork of Challier et al. (1983) on both species of *G. palpalis* throughout all western Africa, they found a small number of natural male hybrids of the two subspecies in Cote d'Ivoire. This hybrid male sterility, together with a sex-ratio distortion in this area, probably explain the incredibly thin hybridization zone described, and led Gouteux, Millet

(1984) to propose that these were the probable mechanisms of incipient speciation existing between *G. p. gambiensis* and *G. p. palpalis*.

4.2 Use of Genetics for Eradication/Control[1]

Genetic isolation is a prerequisite when eradication is considered as part of an area wide insect pest management framework. Some of the recent studies on vector population structure and gene flow have been conducted from the standpoint of specifying populations' degree of isolation from conspecific demes that may compromise tsetse eradication programs using the Sterile Insect Technique (SIT) in areas affected by AAT (e.g., Solano et al., 2010b; Opiro et al., 2017; de Beer et al., 2019). For this, F_{ST} and other parameters of genetic differentiation are usually measured between the targeted population and its neighbors, preceded by careful mapping of known tsetse distribution. However, for other tsetse control strategies based on sustainable decrease of tsetse densities, isolation is no longer a prerequisite. Estimation of effective population sizes N_e may also be helpful in predicting the likelihood of population recovery from an earlier control program. Bouyer et al. (2010), Solano et al. (2010c), Aksoy et al. (2013), Vreysen et al. (2013) offer reviews of the general subject. There has been a renewed interest in gathering this kind of genetic information to inform and help decision on control strategies of countries thanks to the recent development of a framework called "program control pathway" (PCP) for AAT (Diall et al., 2017), which might include and take into account these genetic information at country level.

Of interest and curiosity is to note that tsetse control programs have been more implemented for AAT rather than for g-HAT. This has begun to change recently, especially since tsetse control using "tiny targets" has been implemented in five countries in Africa since the 2010s, with a demonstrated impact on g-HAT transmission (Ndungu et al., 2020; Mahamat et al., 2017). Given the increasing importance of "one health" approaches, where applicable, trypanosomosis offers one of the most obvious examples to develop sustainable control programs via the tsetse vectors that may address both the human and animal forms of the disease. May be in the future, eradication programs at local scales compatible with biodiversity might constitute an interesting option for sustainable interruption of g-HAT transmission.

It may also be interesting to remind the following paradox. Despite the (sometimes huge) genetic differentiation observed between individuals, populations, demes, or ecotypes of the same species, mating and fecundation have always still been possible between males from colonies and females from the field, whatever the geographic origin of the colonies and the time elapsed since the colony began. This has been confirmed during the SIT campaigns that have been implemented for AAT, for which of course mating compatibilities are of utmost importance and are checked before field releases. This applies for instance to *G. tachinoides* (colony originating from Chad, successfully released in Burkina Faso in the 1980s), to *G. palpalis gambiensis* (colony originating from Burkina Faso, used on the field in Burkina Faso in the 1980s Politzar and Cuisance, 1984, and also in the Niayes in Senegal nowadays Bassene et al., 2017), and possibly others.

[1] Bouyer et al. (2010) define these terms as follows. ERADICATION: This strategy aims at totally and definitively removing the tsetse fly from a given area. Theoretically, eradication should save money when compared to control operations because it has to be done only once, whereas control operations have to be repeated. CONTROL: Control operations that aim to lower tsetse densities until there is no tsetse left, or until an acceptable level of transmission risk has been reached. Such interventions are usually applied either to the foci of g-HAT or to AAT on a small scale (livestock ranches, farms, etc.).

In addition although not used on the field so far to the best of our knowledge, hybrid sterility combined to satyrism (Mitchell et al., 2022) that has been described in tsetse colonies (Clutton-Brock and Langley, 1997) might open perspectives for the biological control of tsetse in some areas, since releasing one subspecies in an area where the other occurs might lead to population suppression in certain conditions (see also VanderPlank, 1944).

5. Future Work

The pronounced genetic variation and the low effective population sizes among natural tsetse populations argue for geographically more extensive sampling across the geographical range. This may be investigated using markers of better quality than those already being used or new genetic markers, to assess genetic variation and reciprocal crossing of different lines to assay fertilities and uncover additional sibling species, as well as unknown features in the field such as satyrism behaviors and the impact of the bacteriome, as examples. Vector−parasite coadaptations have important epidemiological and economic consequences, but how they vary spatially remains mostly unknown for both g-HAT and AAT. Metabarcoding applied to decipher eco-epidemiology in vector-borne diseases appears a very promising tool among others (Batovska et al., 2018; Dumonteil et al., 2020), since it allows to gather simultaneous information on the so intimately linked compartments composed by the vector population structure, the bloodmeals taken hence information on the hosts, the bacterial community of the vector, and the trypanosome infection.

Acknowledgments

I thank my colleagues Sophie Ravel, Thierry de Meeüs, Jeremy Bouyer, Antoine Barreaux, and many others, for fruitful discussions on this topic.

References

Aksoy, S., Caccone, A., Galvani, A.P., Okedi, L.M., 2013. *Glossina fuscipes* populations provide insights for human African trypanosomiasis transmission in Uganda. Trends Parasitol. 29, 394−406.

Bassène, M.D., Seck, M.T., Pagabeleguem, S., Fall, A.G., Sall, B., Vreysen, M.J.B., Gimonneau, G., Bouyer, J., 2017. Competitiveness and survival of two strains of *Glossina palpalis gambiensis* in an urban area of Senegal. PLoS Neglected Trop. Dis. 11, e0006172.

Batovska, J., Lynch, S.E., Cogan, N.O.I., Brown, K., Darbro, J.M., Kho, E.A., Blacket, M.J., 2018. Effective mosquito and arbovirus surveillance using metabarcoding. Mol. Ecol. Res. 18, 32−40.

Bouyer, J., Dicko, A., Cecchi, G., et al., 2015. Mapping landscape friction to locate isolated tsetse populations candidate for elimination. Proc. Natl. Acad. Sci. U.S.A. 112 (47), 14575−14580. https://doi.org/10.1073/pnas.1516778112, 1516778112v1-201516778.

Bouyer, J., Solano, P., Cuisance, D., Itard, J., Frézil, J.L., Authié, E., 2010. Control methods in trypanosomosis. In: Lefèvre, P.C., Blancou, R., Chermette, J., Uilenberg, G. (Eds.), Infectious and Parasitic Diseases of Livestock, Chap 131, vol.2. Lavoisier, France, pp. 1927−1959, 1967 pages.

Challier, A., Gouteux, J.P., Coosemans, M., 1983. La limite géographique entre les sous-espèces Glossina palpalis palpalis et *G. p. gambiensis* en Afrique Occidentale. Cah. O.R.S.T.O.M., sér. Ent. méd. et Parasitol. 21, 207−220.

Clutton-Brock, T., Langley, P., 1997. Persistent courtship reduces male and female longevity in captive tsetse flies *Glossina morsitans morsitans* Westwood (*Diptera: Glossinidae*). Behav. Ecol. 8, 392−395.

Courtin, F., Rayaisse, J.B., Serdebeogo, O., Tamboura, I., Koudougou, Z., Solano, P., Sidibé, I., 2010. Updating the northern tsetse limit distribution in Burkina Faso: impact of global change. Int. J. Env. Res. Publ. Health 7, 1708−1719.

De Beer, C.J., Venter, G.J., Vreysen, M.J.B., Mulandane, F.C., Neves, L., Mdluli, S., Koekemoer, O., 2019. Using genetic and phenetic markers to assess population isolation within the southernmost tsetse fly belt in Africa. Onderstepoort J. Vet. Res. 86, e1−e8.

De Meeus, T., Bouyer, J., Ravel, S., Solano, P., 2015. Ecotype evolution in *Glossina palpalis* subspecies, major vectors of sleeping sickness. PLoS Negl. Trop. Dis. 9, e0003497.

Diall, O., Cecchi, G., Wanda, G., Argiles-Herrero, R., Vreysen, M.J.B., Cattoli, G., Viljoen, G.J., Mattioli, R., Bouyer, J., 2017. Developing a progressive control pathway for African animal trypanosomosis. Trends Parasitol. 33, 499−509.

Doudoumis, V., Blow, F., Saridaki, A., Augustinos, A., Dyer, N.A., Goodhead, I., Solano, P., Rayaisse, J.B., Takac, P., Mekonnen, S., Parker, A.G., Abd-Alla, A.M.M., Darby, A., Bourtzis, K., Tsiamis, G., 2017. Challenging the *Wigglesworthia, Sodalis, Wolbachia* symbiosis dogma in tsetse flies: *Spiroplasma* is present in both laboratory and natural populations. Sci. Rep. 7, 4699.

Dumonteil, E., Pronovost, H., Bierman, E.F., Sanford, A., Majeau, A., Moore, R., Herrera, C., 2020. Interactions among *Triatoma sanguisuga* blood feeding sources, gut microbiota and *Trypanosoma cruzi* diversity in southern Louisiana. Mol. Ecol. 19, 3747−3761.

Dyer, N.A., Lawton, S.P., Ravel, S., Choi, K.S., Lehane, M.J., Robinson, A.S., Okedi, L.M., Hall, M.J., Solano, P., Donnelly, M.J., 2008. Molecular phylogenetics of tsetse flies (*Diptera: Glossinidae*) based on mitochondrial (COI, 16S, ND2) and nuclear ribosomal DNA sequences, with an emphasis on the palpalis group. Mol. Phylogenet. Evol. 49, 227−239.

Dyer, N.A., Ravel, S., Choi, K.S., Darby, A.C., Causse, S., Kapitano, B., et al., 2011. Cryptic diversity within the major trypanosomiasis vector *Glossina fuscipes* revealed by molecular markers. PLoS Negl. Trop. Dis. 5, e1266.

Engl, T., Michalkova, V., Weiss, B.L., Uzel, G.D., Takac, P., Miller, W.J., Abd-Alla, A.M.M., Aksoy, S., Kaltenpoth, M., 2018. Effect of antibiotic treatment and gamma-irradiation on cuticular hydrocarbon profiles and mate choice in tsetse flies (*Glossina m. morsitans*). BMC Microbiol. 18 (Suppl. 1), 145.

Franco, J.R., Cecchi, G., Paone, M., Diarra, A., Grout, L., Kadima Ebeja, A., Simarro, P.P., Zhao, W., Argaw, D., 2022. The elimination of human African trypanosomiasis: achievements in relation to WHO road map targets for 2020. PLoS Negl. Trop. Dis. 16, e0010047.

Geiger, A., Ponton, F., Simo, G., 2015. Adult blood-feeding tsetse flies, trypanosomes, microbiota and the fluctuating environment in sub-Saharan Africa. ISME J. 9, 1496−1507.

Gooding, R.H., Krafsur, E.S., 2005. Tsetse genetics: contributions to biology, systematics, and control of tsetse flies. Ann. Rev. Entomol. 50, 101e23.

Gouteux, J.P., Millet, P., 1984. Observations sur le contact entre *Glossina palpalis palpalis* (Rob.-Desv., 1830) et *G. palpalis gambiensis* Vanderplank, 1949 (Diptera), dans la région de Bouaké (Côte d'Ivoire). Tropenmed. Parasitol. 35, 157−159.

International Glossina Genome Initiative (IGGI), 2014. Genome sequence of the tsetse fly (*Glossina morsitans*): vector of African trypanosomiasis. Science 344, 380−386.

Krafsur, E.S., 2009. Tsetse flies: genetics, evolution, and role as vectors. Inf. Genet. Evol. 9, 124−141.

Krafsur, E., Maudlin, I., 2018. Tsetse fly evolution, genetics and the trypanosomiases − a Review. Infect. Genet. Evol. 64, 185−206.

Lord, J.S., Hargrove, J.W., Torr, S.J., Vale, G.A., 2018. Climate change and African trypanosomiasis vector populations in Zimbabwe's Zambezi Valley: a mathematical modelling study. PLoS Med. 15, e1002675.

Mahamat, M.H., Peka, M., Rayaisse, J.B., Rock, K.S., Toko, M.A., Darnas, J., Brahim, G.M., Alkatib, A.B., Yoni, W., Tirados, I., Courtin, F., Brand, S.P.C., Nersy, C., Alfaroukh, I.O., Torr, S.J., Lehane, M.J., Solano, P., 2017. Adding tsetse control to medical activities contributes to decreasing transmission of sleeping sickness in the Mandoul focus (Chad). PLoS Negl. Trop. Dis. 11, e0005792.

Mitchell, C., Leigh, S., Alphey, L., et al., 2022. Reproductive interference and Satyrisation: mechanisms, outcomes and potential use for insect control. J. Pest. Sci. 95, 1023−1036.

Ndung'u, J.M., Boulangé, A., Picado, A., Mugenyi, A., Mortensen, A., Hope, A., Mollo, B.G., Bucheton, B., Wamboga, C., Waiswa, C., Kaba, D., Matovu, E., Courtin, F., Garrod, G., Gimonneau, G., Bingham, G.V., Hassane, H.M., Tirados, I., Saldanha, I., Kabore, J., Rayaisse, J.B., Bart, J.M., Lingley, J., Esterhuizen, J., Longbottom, J., Pulford, J., Kouakou, L., Sanogo, L., Cunningham, L., Camara, M., Koffi, M., Stanton, M., Lehane, M.J., Kagbadouno, M.S., Camara, O., Bessell, P., Mallaye, P., Solano, P., Selby, R., Dunkley, S., Torr, S., Bieler, S., Lejon, V., Jamonneau, V., Yoni, W., Katz, Z., 2020. Trypa-NO! contributes to the elimination of gambiense human African trypanosomiasis by combining tsetse control with "screen, diagnose and treat" using innovative tools and strategies. PLoS Neglected Trop. Dis. 14, e0008738.

Opiro, R., Saarman, N.P., Echodu, R., Opiyo, E.A., Dion, K., Halyard, A., Dunn, A.W., Aksoy, S., Caccone, A., 2017. Genetic diversity and population structure of the tsetse fly *Glossina fuscipes fuscipes* (*Diptera: Glossinidae*) in Northern Uganda: implications for vector control. PLoS Negl. Trop. Dis. 11, e0005485.

Politzar, H., Cuisance, D., 1984. An integrated campaign against riverine tsetse flies *Glossina palpalis gambiensis* and *Glossina tachinoides* by trapping and the release of sterile males. Insect Sci. Applic. 5, 439−442.

Solano, P., Bouyer, J., Itard, J., Cuisance, D., 2010a. Cyclical vectors of trypanosomosis. In: Lefèvre, P.C., Blancou, R., Chermette, J., Uilenberg, G. (Eds.), Infectious and Parasitic Diseases of Livestock, Chap 013, 2 volumes. Lavoisier, France, pp. 155−180, 1967 pages.

Solano, P., Kaba, D., Ravel, S., Dyer, N., Sall, B., Vreysen, M.J., Seck, M.T., Darbyshir, H., Gardes, L., Donnelly, M.J., de Meeûs, T., Bouyer, J., 2010b. Population genetics as a tool to select tsetse control strategies: suppression or eradication of *Glossina palpalis gambiensis* in the Niayes of Senegal. PLoS Negl. Trop. Dis. 4, e692.

Solano, P., Ravel, S., de Meeûs, T., 2010c. How can tsetse population genetics contribute to African Trypanosomosis control? Trends Parasitol. 26, 255−263.

Swallow, B., 1998. Impact of trypanosomiasis on African agriculture. In: PAAT Tech. Sci. Series No. 2. FAO, Rome, 52 pp.

Tibayrenc, M., 2017. Genetics and Evolution of Infectious Diseases, second ed. Elsevier.

Vanderplank, F., 1944. Hybridization between *Glossina* species and suggested new method for control of certain species of tsetse. Nature 154, 607−608.

Vreysen, M.J.B., Balenghien, T., Saleh, K.M., Maiga, S., Koudougou, Z., Cecchi, G., Bouyer, J., 2013. Release recapture studies confirm dispersal of *Glossina palpalis gambiensis* between river basins in Mali. PLoS Negl. Trop. Dis. 7, e2022.

W.H.O., 2020. Ending the Neglect to Attain the Sustainable Development Goals. A Roadmap for Neglected Tropical Diseases 2021-2030. World Health Organisation, Geneva, 196 pp.

This page intentionally left blank

CHAPTER

Genetics of the Triatominae (Hemiptera: Reduviidae), Vectors of Chagas Disease

15

S.F. Brenière[1], E. Waleckx[1,2], S. Pita[3], F. Panzera[3], C. Barnabé[1] and P.L. Dorn[4]

[1]*Institut de Recherche pour le Développement, UMR INTERTRYP IRD, CIRAD, Université de Montpellier, Montpellier, France;* [2]*Laboratorio de Parasitología, Centro de Investigaciones Regionales "Dr Hideyo Noguchi", Universidad Autónoma de Yucatán, Mérida, Mexico;* [3]*Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay;* [4]*Loyola University New Orleans, New Orleans, LA, United States*

1. Introduction

Chagas disease, also known as American trypanosomiasis, is a zoonosis caused by the flagellated protozoan *Trypanosoma cruzi* (Kinetoplastea: Trypanosomatida) and is among the most serious Neglected Tropical Diseases in Latin America. Although rarely fatal in its early acute stage, in ~30% of those infected it progresses to a debilitating chronic disease that involves severe cardiac and intestinal lesions, that can be deadly. There is no vaccine available, and treatment is woefully inadequate, at least during the chronic stage.

T. cruzi is mainly transmitted to humans and other animals through the contaminated feces of hematophagous bugs belonging to the subfamily Triatominae (Hemiptera: Reduviidae) called triatomines. These triatomines, known in the United States as kissing or conenose bugs, have many names in Latin America such as vinchucas, barbeiros, pics, chinches, chinches besuconas, chinchorros, pitos, or chipos, among others. Vector transmission ranges from the southern United States to northern Argentina, mostly in poor rural areas where houses are infested with triatomine vectors (Fig. 15.1). Furthermore, with deforestation and human migrations, Chagas disease is increasingly found in urban areas and even in nonendemic countries, i.e., countries without triatomine occurrence. Beyond vectorial transmission, *T. cruzi* can be transmitted by blood transfusion, organ transplantation, and congenitally, bringing new health problems in nonendemic areas.

Regional Chagas control initiatives in Latin America have resulted in a dramatic reduction of disease prevalence due to decreased vector transmission and an increase in blood donation screening. Despite these gains, nearly six million people remain infected with this deadly parasite (WHO, 2022).

2. The Triatominae, Vectors of Chagas Disease

The Triatominae are the only proven vectors of American trypanosomiasis. They require a blood meal to molt and lay eggs and acquire the Chagas parasite when they feed on a *T. cruzi*–infected animal.

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00001-X

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

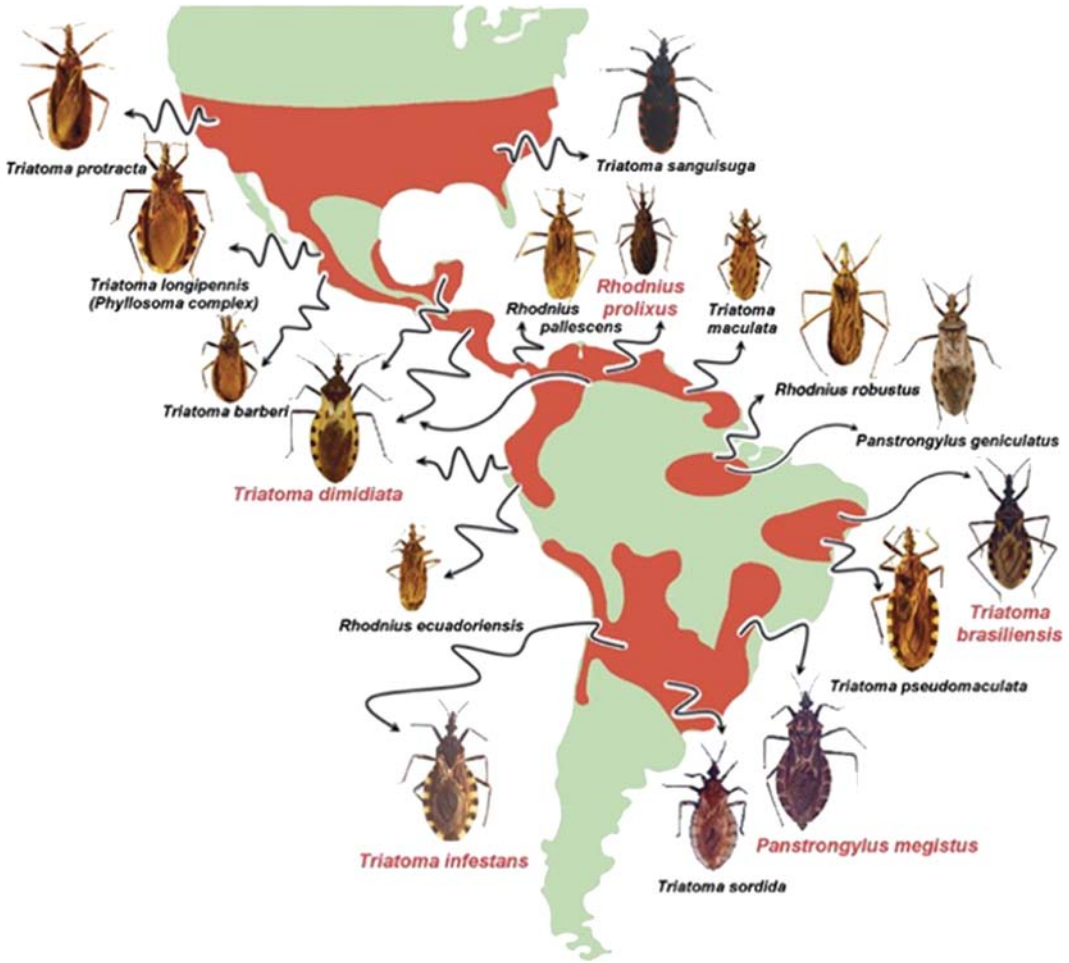


FIGURE 15.1

Approximate distribution of triatomine species of major epidemiological relevance. Only about 20 triatomine species are responsible for *T. cruzi* transmission to humans, because of their ability to infest and, for some species, to colonize human habitat. The red areas indicate the approximate species geographic distribution. Species highlighted in red are considered the most important vectors of the parasite. According to Gourbière (Gourbiere et al., 2012). (For interpretation of the references to color in this figure legend, please refer online version of this title.)

They transmit the parasite to a new host via parasite-contaminated feces deposited on the skin or mucous membrane of the new host during a blood meal. The level of domiciliation of triatomines, i.e., their level of adaptation to humans and their domestic environment, is a key determinant of vectorial capacity as it defines the level of human–vector contacts (Dujardin et al., 2002; Waleckx et al., 2015b). Less than a dozen of the over 150 Triatominae species are known to be well adapted to human

dwellings, so are the major vectors (Schofield and Galvao, 2009). Note, however, that anthropization and climate change influence the spatial distribution of triatomines and their relationship to humans. Both factors affect the natural habitats of wild populations of triatomines and their hosts. As a result, triatomine populations can be reduced or, alternatively, expand by colonization of new localities and even adaptation to human habitats (Coura, 2015; Escobar et al., 2016).

3. Evolution and Taxonomy of the Triatominae

The Triatominae are classified as a subfamily of Reduviidae (assassin bugs), which are mostly predators of other arthropods, and likely evolved from predators associated with nesting vertebrates, which allowed the shift from an arthropod prey to feeding on a vertebrate host. While the monophyletic/paraphyletic/polyphyletic origin of the group is still debated, most of the recent and most complete phylogenies favor the monophyletic origin (Justi et al., 2016; Kieran et al., 2021; Patterson and Gaunt, 2010; Weirauch and Munro, 2009).

Currently, the Triatominae subfamily includes 157 validated species, including 154 extant and 3 fossil species, classified into 5 tribes and 18 genera (Ceccarelli et al., 2022) (Table 15.1). Most of the species occur exclusively in the Americas. A few species are found exclusively in Asia (e.g., species of the genus *Linshcosteus*). *Triatoma rubrofasciata* is the only species globally distributed, found in America, Asia, and Africa. The most epidemiologically important species belong to the tribes Triatomini and Rhodniini. The Triatomini tribe is the most diverse, with over 100 described species. This tribe includes *Triatoma* and *Panstrongylus* genera, which include the main vector species. The Rhodniini tribe includes two genera, *Rhodnius* and *Psammolestes*. Overall, the epidemiologically most important species are *Triatoma infestans* and *Triatoma brasiliensis* in South America, *Triatoma dimidiata* and *Rhodnius prolixus* in Central America and northern South America (Fig. 15.1). However, several other species, which are less represented because of their more restricted geographical distribution, or because they are mainly found in sylvatic ecotopes, also play a major role. These include *Triatoma longipennis* in the State of Jalisco, Mexico (Walter et al., 2007), *Rhodnius ecuadoriensis* in the province of Loja, Ecuador (Villacis et al., 2017), and the mainly sylvatic species *Panstrongylus geniculatus*, which frequently invades human dwellings in South America (Vivas et al., 2021).

In a recent phylogeny including four of the five Triatominae tribes and constructed from sequences of ultraconserved elements (UCE), nine well-supported clades appear (Fig. 15.2). Interestingly, the main genera *Triatoma*, *Panstrongylus* and *Rhodnius* appear as nonmonophyletic clades (Kieran et al., 2021).

4. Cytogenetics of the Triatominae

Most of the 102 Triatominae species cytogenetically studied hitherto present a diploid chromosome number of 20 autosomes and XY sex chromosomes, which is considered as the ancestral karyotype (Panzera et al., 2021). Variations in the number of autosomes are due to the occurrence of a fusion between two autosomal pairs (originating species with 18 autosomes such as *Triatoma nitida* and *Panstrongylus megistus*) or one autosomal fission (*T. rubrofasciata* with 22 autosomes). The number of sex chromosomes varies due to the fragmentation of the ancestral X chromosome into two or three fragments (species with X1X2Y and X1X2X3Y sex systems). Considering that the variation observed in

Table 15.1 List of Species Included in the Subfamily Triatominae, According to Ceccarelli et al. (2022)

Tribe	Genera	Species
Alberprosenini	*Alberprosenia*	*goyovargasi*
		malheiroi
Bolboderini	*Belminus*	*corredori*
		costaricensis
		ferroae
		herreri
		laportei
		peruvianus
		pittieri
		rugulosus
		santosmalletae
	Bolbodera	*scabrosa*
	Microtriatoma	*borbai*
		trinidadensis
	Parabelminus	*carioca*
		yurupucu
Carvernicolini	*Cavernicola*	*lenti*
		pilosa
Rhodniini	*Psammolestes*	*arthuri*
		coreodes
		tertius
	Rhodnius	*amazonicus*
		barretti
		brethesi
		colombiensis
		dalessandroi
		domesticus
		ecuadoriensis
		marabaensis
		micki
		milesi
		montenegrensis
		nasutus
		neglectus
		neivai
		pallescens
		paraensis
		pictipes
		prolixus
		robustus
		stali
		zeledoni

Table 15.1 List of Species Included in the Subfamily Triatominae, According to Ceccarelli et al. (2022)—cont'd

Tribe	Genera	Species
Triatomini	*Dipetalogaster*	*maxima*
	Eratyrus	*cuspidatus*
		mucronatus
	Hermanlentia	*matsunoi*
	Linshcosteus	*carnifex*
		chota
		confumus
		costalis
		kali
		karupus
	Mepraia	*gajardoi*
		parapatrica
		spinolai
	Nesotriatoma	*confusa*
		flavida
		obscura
	Panstrongylus	*chinai*
		diasi
		geniculatus
		guentheri
		hispanoliae[a]
		howardi
		humeralis
		lenti
		lignarius
		lutzi
		martinezorum
		megistus
		mitarakaensis
		rufotuberculatus
		tupynambai
	Paleotriatoma	*metaxytaxa*[a]
	Paratriatoma	*hirsuta*
		lecticularia
	Triatoma	*amicitiae*
		arthurneivai
		bahiensis
		baratai
		barberi

Continued

Table 15.1 List of Species Included in the Subfamily Triatominae, According to Ceccarelli et al. (2022)—cont'd

Tribe	Genera	Species
		bassolsae
		bolivari
		boliviana
		bouvieri
		brailovskyi
		brasiliensis
		breyeri
		carcavalloi
		carrioni
		cavernicola
		circummaculata
		costalimai
		deaneorum
		delpontei
		dimidiata
		dispar
		dominicana[a]
		eratyrusiformis
		garciabesi
		gerstaeckeri
		gomeznunesi
		guasayana
		guazu
		hegneri
		huehuetenaguensis
		incrassata
		indictiva
		infestans
		jatai
		juazeirensis
		jurbergi
		klugi
		lenti
		leopoldi
		limai
		longipennis
		maculata
		matogrossensis
		mazzottii

Table 15.1 List of Species Included in the Subfamily Triatominae, According to Ceccarelli et al. (2022)—cont'd

Tribe	Genera	Species
		melanica
		melanocephala
		mexicana
		migrans
		mopan
		neotomae
		nigromaculata
		nitida
		oliveirai
		pallidipennis
		patagonica
		peninsularis
		petrochiae
		phyllosoma
		picturata
		pintodiasi
		platensis
		protracta
		pseudomaculata
		pugasi
		recurva
		rosai
		rubida
		rubrofasciata
		rubrovaria
		ryckmani
		sanguisuga
		sherlocki
		sinaloensis
		sinica
		sordida
		tibiamaculata
		vandae
		venosa
		vitticeps
		williami
		wygodzinskyi

[a]*Fossil species.*

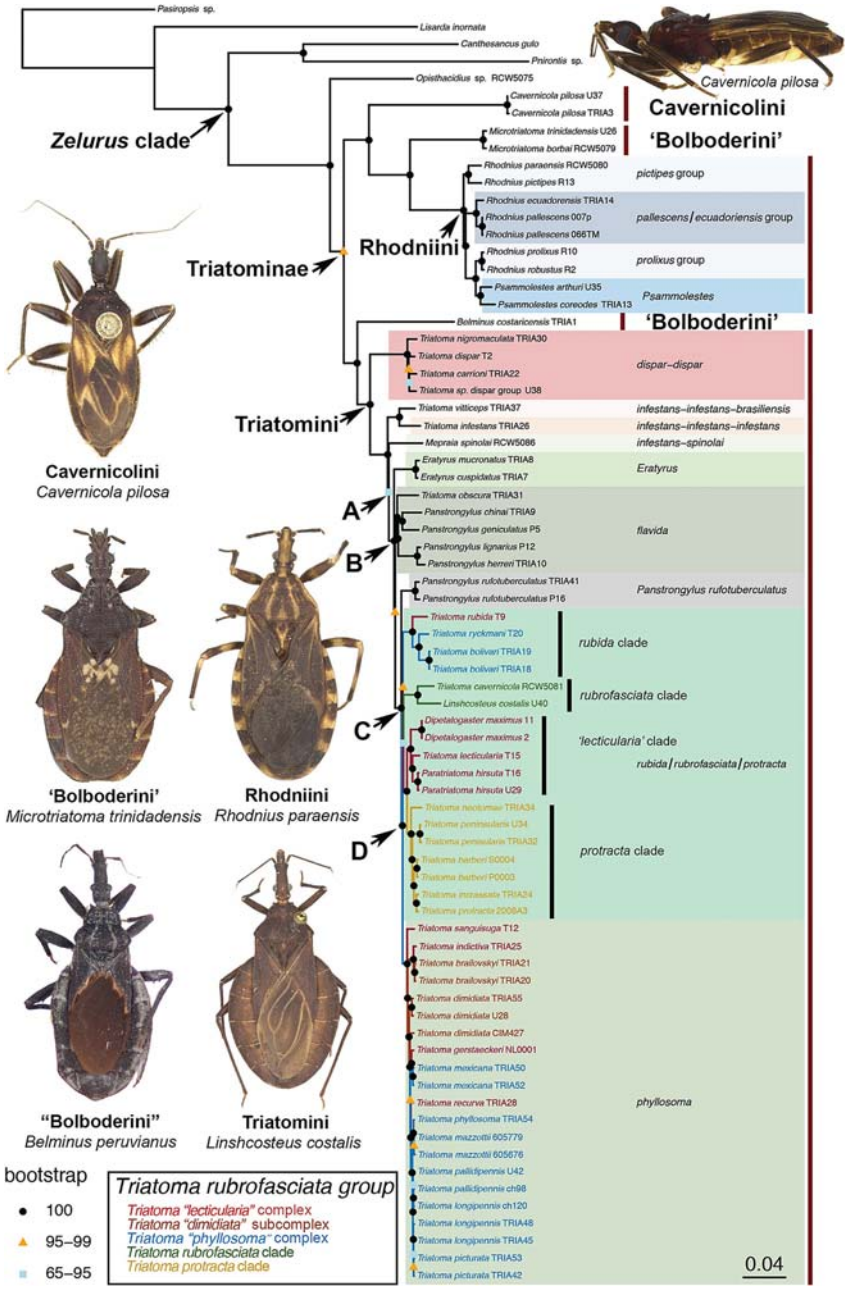


FIGURE 15.2

Phylogenetic hypothesis of Triatominae based on the best likelihood tree derived from the partitioned analysis of the 85% matrix of the UCE-only dataset according to Kieran et al. (2021). Capital letters next to nodes are clades: Clade A: *spinolai* clade + *Eratyrus* + *flavida* clade + *P. rufotuberculatus* + *phyllosoma* clade + *rubida* clade + *protracta* clade + *rubrofasciata* clade; Clade B: Clade A less *spinolai* clade; Clade C: *P. rufotuberculatus* + *phyllosoma* clade + *rubida* clade + *protracta* clade + *rubrofasciata* clade; Clade D: Clade C less *P. rufotuberculatus*. Clade E: *rubida* clade + *protracta* clade + *rubrofasciata* clade.

the number of autosomes and X chromosomes occurred in evolutionarily distant species, it is very likely that these chromosomal rearrangements (fissions and fusions) have occurred several times during the evolution of this insect group (Panzera et al., 2021). A distinctive feature of triatomines is that the amount and location of heterochromatin (tandemly repeated sequences) varies considerably among and even within some species. For example, Andean *T. infestans* specimens present 50% more C-banded heterochromatin and 30% more DNA content than non-Andean specimens. Several studies led to the suggestion that the Bolivian Andean valleys are the origin of *T. infestans* (Panzera et al., 2014). Overall, the extreme variation in the amount of C-heterochromatin in Triatominae, could be partially correlated with the marked variation in total genome DNA content measured by flow cytometry (between 700 and 2840 Mbp) (Panzera et al., 2021). Furthermore, other chromosomal traits that show unusual variability are the chromosomal location and the number of 45S ribosomal loci, reflecting that triatomine genomes have undergone a high rate of change with chromosomal rearrangements that do not involve variations in chromosome number (Pita et al., 2022). Chromosomal differences were also important in supporting cryptic speciation in *T. dimidiata* (Panzera et al., 2006), *Triatoma sordida* (Panzera et al., 2015), and *Panstrongylus rufotuberculatus* (Pita et al., 2021). The chromosome diversity of triatomines makes it a model system for understanding chromosome structure and its evolution, and the role of heterochromatin and other repetitive sequences during the speciation process.

5. Genetic Diversity and Population Structure of Triatominae Species

Numerous publications explore the genetic diversity of Triatominae species of known or presumed epidemiological importance. These generally use population genetics to evaluate the diversity of the species in time and space. Changes in genetic composition (genotype and phenotype frequencies) of populations due to natural selection processes, genetic drift, mutation, and gene flow provide information on phenomena such as adaptation, speciation, historical evolution, recent species history, and spatial population dynamics. Analyzing these phenomena is important in understanding the epidemiology and informing control interventions. Below, we present selected studies related to the genetic diversity and population structure of two species of interest, *T. infestans* and *T. dimidiata*, as well as the phenomena they elucidate.

Active and passive transport (by human activity and perhaps migratory birds) has been important in spreading and mixing triatomine populations, and consequently modulating their population structure and genetic diversity. Overall, studies in *T. dimidiata* and *T. infestans* show that at larger geographic scales (populations >50 km apart) there is generally a gradient of allele frequency differences among populations consistent with an "isolation by distance" model (Wright, 1943). Nevertheless, concerning *T. dimidiata*, it is largely accepted that the latter is in fact a complex including different genetic lineages (Bargues et al., 2008; Dorn et al., 2016; Justi et al., 2018) and two new cryptic species belonging to this complex have been recently described: *T. mopan* (Dorn et al., 2018) and *T. huehuetenanguensis* (Lima-Cordon et al., 2019). Different studies have also explored the geographical and ecological factors, which would explain the large-scale structure of *T. dimidiata* (sensu lato) (Gomez-Palacio and Triana, 2014; Landaverde-Gonzalez et al., 2020). For example, in Colombia, genetic differentiation did not strongly correlate with distance (isolation by distance model), instead subdivision may be more influenced by geological steps such as the formation of the Isthmus of Panama and the upwelling of the Andes (Gomez-Palacio and Triana, 2014). Concerning *T. infestans*, different population genetics and phylogeography studies have tried to identify the geographic origin of the species and arguments in favor of an

Andean or non-Andean origin are still under debate. In the most recent studies, a non-Andean origin has been favored, because of the presence of non-Andean haplotypes which appear as basal in the phylogenetic trees and networks, as well as the higher genetic diversity of *T. infestans* observed in this area (Ceballos et al., 2009; Fernández et al., 2019; Piccinali et al., 2011; Waleckx et al., 2011).

At smaller geographic scales, the picture of *T. infestans* and *T. dimidiata* population structure is sometimes more complicated, varies geographically, and may be affected by the insecticide application history. At a fine-scale genetic structure, *T. dimidiata* appears to be highly mobile, whether moving among domestic habitats (Stevens et al., 2015), or between domestic and sylvatic habitats (Dumonteil et al., 2007) with little subdivision among nearby populations. Rapid reinfestation post spraying, often observed in *T. dimidiata*, was attributed to dispersal from nearby houses as well as incomplete elimination of original populations in a recent study using genome-wide markers of *T. dimidiata* in Guatemala (Cahan et al., 2019). Interestingly, insecticide resistance has not yet been identified in *T. dimidiata*. Frequent migration among nearby villages was also suggested by a recent study of Colombian populations collected in the Boyacá department, which showed low genetic variability between localities so all were considered to be a single population (Velasquez-Ortiz et al., 2022).

T. infestans is the species subject to the most comprehensive insecticide spray campaign. The highly successful Southern Cone Initiative, and many studies have addressed the origin of reinfesting populations. In areas where sylvatic populations were absent, genetic analyses have generally concluded that reinfestations originated from survivor vectors (incomplete elimination of original populations or sometimes because of the presence of insecticide resistance) (Depickère et al., 2012; Dujardin et al., 1996; Garcia et al., 2003; Perez de Rosas et al., 2007; Piccinali et al., 2018; Pizarro et al., 2008; Quisberth et al., 2011; Rojas de Arias et al., 2022). In areas where sylvatic populations exist, at least part of the reinfestation can be attributed to these sympatric populations surrounding the villages (Breniere et al., 2013; Rojas de Arias et al., 2022).

In domestic Triatominae populations, a diminished genetic diversity is generally expected compared to sylvatic ones due to founder effects, because in a domestication process, only a portion of sylvatic genotypes may be successful in establishing enduring domestic populations (Schofield et al., 1999). Moreover, domestic populations are also the main target of insecticide spraying campaigns which are expected to result in a population bottleneck (i.e., a sharp reduction of the population size generally leading to a reduction of the genetic diversity) followed by new founder effects originating from surviving insects. Surprisingly, sylvatic *T. infestans* show lower, or comparable genetic diversity with that of domestic populations (Piccinali et al., 2009; Waleckx et al., 2011). Nevertheless, as proposed by Waleckx et al. (2011), this could be explained by numerous domestication events (i.e., colonization of anthropized spaces by the surrounding sylvatic populations in different regions, sometimes very distant, leading to a similar diversity of domestic and sylvatic populations because of their connectivity).

6. Conclusions and Future Directions

More than 111 years after its discovery, and despite notable successes of intergovernmental control initiatives, Chagas disease remains the most serious of the parasitic diseases affecting Latin America (Martins-Melo et al., 2021). Substantial challenges remain to reach the goals stated in the 2021−30 WHO Roadmap for Neglected Tropical Diseases concerning Chagas disease (WHO, 2020), because, among others, of the emergence of insecticide resistance, the re-infestation of dwellings after insecticide spraying, secondary vectors replacing eliminated primary vectors, the spread of Chagas as a

result of deforestation, climate change and global travel; unfortunately the challenges and the proposals for the control of Chagas disease from more than 10 years ago are still valid (Reithinger et al., 2009).

Completion of the first Triatominae genome, corresponding to *R. prolixus* (Mesquita et al., 2015), tissue-specific transcriptomes (Ons et al., 2016; Ribeiro et al., 2012, 2014; Schwarz et al., 2014) and complete mitochondrial genomes of several species, are studies which will provide many new tools to address these challenges. Comparative studies promise advances in understanding the genetic basis of vector competence/capacity, reproductive isolation among sympatric species, as well as genes and proteins involved in the switch to hematophagy, domestication, and insecticide resistance (Mougabure-Cueto and Ines Picollo, 2015). New hope for eventual elimination of Chagas disease comes from an integrative approach combining new tools in "-omics" with mathematical modeling to design evidence-based interventions, in conjunction with community-based development approaches as the ones developed in Southern Mexico and Central America (Castro-Arroyave et al., 2020; Lucero et al., 2013; Pennington et al., 2021; Peterson et al., 2019; Waleckx et al., 2015a, 2018) (Fig. 15.3).



FIGURE 15.3

Photos of the same house before and after the Ecohealth house improvements that make the houses refractory to the triatomine vectors in Guatemala.

Reproduced with permission from C. Monroy.

More broadly, a One Health/Ecohealth approach is essential since it addresses the complex relationships involved in vector-borne diseases between human health and (1) individuals and society, (2) animal health, (3) ecosystems, and (4) the new ecological factors such as anthropization, deforestation, and climate changes (Heukelbach et al., 2021; Rotureau et al., 2022). This integrated approach has shown tremendous promise in Central America (Dorn et al., 2022; Pereira et al., 2022) and is being implemented, modified for specific localities, to combat Chagas disease across the Americas.

Acknowledgments

The authors would like to acknowledge the contribution of Silvia Justi to a previous version of this chapter.

References

Bargues, M.D., Klisiowicz, D.R., Gonzalez-Candelas, F., Ramsey, J.M., Monroy, C., Ponce, C., Salazar-Schettino, P.M., Panzera, F., Abad-Franch, F., Sousa, O.E., Schofield, C.J., Dujardin, J.P., Guhl, F., Mas-Coma, S., 2008. Phylogeography and genetic variation of *Triatoma dimidiata,* the main Chagas disease vector in Central America, and its position within the genus *Triatoma*. PLoS Neglected Trop. Dis. 2, e233.

Breniere, S.F., Salas, R., Buitrago, R., Bremond, P., Sosa, V., Bosseno, M.F., Waleckx, E., Depickere, S., Barnabe, C., 2013. Wild populations of *Triatoma infestans* are highly connected to intra-peridomestic conspecific populations in the Bolivian Andes. PLoS One 8, e80786.

Cahan, S.H., Orantes, L.C., Wallin, K.F., Hanley, J.P., Rizzo, D.M., Stevens, L., Dorn, P.L., Rodas, A., Monroy, C., 2019. Residual survival and local dispersal drive reinfestation by *Triatoma dimidiata* following insecticide application in Guatemala. Infect. Genet. Evol. 74, 104000.

Castro-Arroyave, D., Monroy, M.C., Irurita, M.I., 2020. Integrated vector control of Chagas disease in Guatemala: a case of social innovation in health. Infect. Dis. Poverty 9, 25.

Ceballos, L.A., Piccinali, R.V., Berkunsky, I., Kitron, U., Gurtler, R.E., 2009. First finding of melanic sylvatic *Triatoma infestans* (Hemiptera: Reduviidae) colonies in the Argentine Chaco. J. Med. Entomol. 46, 1195−1202.

Ceccarelli, S., Balsalobre, A., Vicente, M.E., Curtis-Robles, R., Hamer, S.A., Ayala Landa, J.M., Rabinovich, J.E., Marti, G.A., 2022. American triatomine species occurrences: updates and novelties in the DataTri database. Gigabyte 2022, 1−8.

Coura, J.R., 2015. The main sceneries of Chagas disease transmission. The vectors, blood and oral transmissions–a comprehensive review. Mem. Inst. Oswaldo Cruz 110, 277−282.

Depickère, S., Buitrago, R., Siñani, E., Baune, M., Monje, M., Lopez, R., Waleckx, E., Chavez, T., Brenière, S.F., 2012. Susceptibility and resistance to deltamethrin of wild and domestic populations of *Triatoma infestans* (Reduviidae: Triatominae) in Bolivia: new discoveries. Mem. Inst. Oswaldo Cruz 107, 1042−1047.

Dorn, P.L., de la Rua, N.M., Axen, H., Smith, N., Richards, B.R., Charabati, J., Suarez, J., Woods, A., Pessoa, R., Monroy, C., Kilpatrick, C.W., Stevens, L., 2016. Hypothesis testing clarifies the systematics of the main Central American Chagas disease vector, *Triatoma dimidiata* (Latreille, 1811), across its geographic range. Infect. Genet. Evol. 44, 431−443.

Dorn, P.L., Justi, S.A., Dale, C., Stevens, L., Galvao, C., Lima-Cordon, R., Monroy, C., 2018. Description of *Triatoma mopan* sp. n. from a cave in Belize (Hemiptera, Reduviidae, Triatominae). ZooKeys 775, 69−95.

Dorn, P.L., Monroy, M.C., Stevens, L., 2022. Sustainable, integrated control of native vectors: the case of Chagas disease in Central America. Front. Trop. Dis. 3, 72.

Dujardin, J.-P., Schofield, J., Panzera, F., Matias, A., De La Riva, J., 2002. Los vectores de la enfermedad de Chagas. Académie Royale des Sciences d'Outre-Mer, Bruxelles.

Dujardin, J.P., Cardozo, L., Schofield, C., 1996. Genetic analysis of *Triatoma infestans* following insecticidal control interventions in central Bolivia. Acta Trop. 61, 263−266.

Dumonteil, E., Tripet, F., Ramirez-Sierra, M.J., Payet, V., Lanzaro, G., Menu, F., 2007. Assessment of *Triatoma dimidiata* dispersal in the Yucatan Peninsula of Mexico by morphometry and microsatellite markers. Am. J. Trop. Med. Hyg. 76, 930−937.

Escobar, L.E., Romero-Alvarez, D., Leon, R., Lepe-Lopez, M.A., Craft, M.E., Borbor-Cordova, M.J., Svenning, J.C., 2016. Declining prevalence of disease vectors under climate change. Sci. Rep. 6, 39150.

Fernández, C.J., González-Ittig, R.E., García, B.A., 2019. Genetic structure of the Chagas disease vector *Triatoma infestans* (Hemiptera: Reduviidae) based on nuclear and mitochondrial DNA sequences. Zool. J. Linn. Soc. 188, 1152−1171.

Garcia, B.A., Manfredi, C., Fichera, L., Segura, E.L., 2003. Short report: variation in mitochondrial 12S and 16S ribosomal DNA sequences in natural populations of *Triatoma infestans* (Hemiptera: Reduviidae). Am. J. Trop. Med. Hyg. 68, 692−694.

Gomez-Palacio, A., Triana, O., 2014. Molecular evidence of demographic expansion of the chagas disease vector *Triatoma dimidiata* (Hemiptera, Reduviidae, Triatominae) in Colombia. PLoS Neglected Trop. Dis. 8, e2734.

Gourbiere, S., Dorn, P., Tripet, F., Dumonteil, E., 2012. Genetics and evolution of triatomines: from phylogeny to vector control. Heredity 108, 190−202.

Heukelbach, J., de Sousa, A.S., Ramos Jr., A.N., 2021. New contributions to the elimination of Chagas disease as a public health problem: towards the Sustainable Development Goals by 2030. Trav. Med. Infect. Dis. 6.

Justi, S.A., Cahan, S., Stevens, L., Monroy, C., Lima-Cordon, R., Dorn, P.L., 2018. Vectors of diversity: genome wide diversity across the geographic range of the Chagas disease vector *Triatoma dimidiata* sensu lato (Hemiptera: Reduviidae). Mol. Phylogenet. Evol. 120, 144−150.

Justi, S.A., Galvao, C., Schrago, C.G., 2016. Geological changes of the Americas and their influence on the diversification of the neotropical kissing bugs (Hemiptera: Reduviidae: Triatominae). PLoS Neglected Trop. Dis. 10, e0004527.

Kieran, T.J., Gordon, E.R.L., Zaldívar-Riverón, A., Ibarra-Cerdeña, C.N., Glenn, T.C., Weirauch, C., 2021. Ultraconserved elements reconstruct the evolution of Chagas disease-vectoring kissing bugs (Reduviidae: Triatominae). Syst. Entomol. 46, 725−740.

Landaverde-Gonzalez, P., Menes, M., Melgar, S., Bustamante, D., Monroy, C., 2020. Common pattern of distribution for Mesoamerican *Triatoma dimidiata* suggest geological and ecological association. Acta Trop. 204, 105329.

Lima-Cordon, R.A., Monroy, M.C., Stevens, L., Rodas, A., Rodas, G.A., Dorn, P.L., Justi, S.A., 2019. Description of *Triatoma huehuetenanguensis* sp. n., a potential Chagas disease vector (Hemiptera, Reduviidae, Triatominae). ZooKeys 820, 51−70.

Lucero, D.E., Morrissey, L.A., Rizzo, D.M., Rodas, A., Garnica, R., Stevens, L., Bustamante, D.M., Monroy, M.C., 2013. Ecohealth interventions limit triatomine reinfestation following insecticide spraying in La Brea, Guatemala. Am. J. Trop. Med. Hyg. 88, 630−637.

Martins-Melo, F.R., Castro, M.C., Werneck, G.L., 2021. Levels and trends in Chagas disease-related mortality in Brazil, 2000-2019. Acta Trop. 220, 105948.

Mesquita, R.D., Vionette-Amaral, R.J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F.A., Minx, P., Spieth, J., Carvalho, A.B., Panzera, F., Lawson, D., Torres, A.Q., Ribeiro, J.M., Sorgine, M.H., Waterhouse, R.M., Montague, M.J., Abad-Franch, F., Alves-Bezerra, M., Amaral, L.R., Araujo, H.M., Araujo, R.N., Aravind, L., Atella, G.C., Azambuja, P., Berni, M., Bittencourt-Cunha, P.R., Braz, G.R., Calderon-Fernandez, G., Carareto, C.M., Christensen, M.B., Costa, I.R., Costa, S.G., Dansa, M., Daumas-Filho, C.R., De-Paula, I.F., Dias, F.A., Dimopoulos, G., Emrich, S.J., Esponda-Behrens, N., Fampa, P., Fernandez-Medina, R.D., da

Fonseca, R.N., Fontenele, M., Fronick, C., Fulton, L.A., Gandara, A.C., Garcia, E.S., Genta, F.A., Giraldo-Calderon, G.I., Gomes, B., Gondim, K.C., Granzotto, A., Guarneri, A.A., Guigo, R., Harry, M., Hughes, D.S., Jablonka, W., Jacquin-Joly, E., Juarez, M.P., Koerich, L.B., Lange, A.B., Latorre-Estivalis, J.M., Lavore, A., Lawrence, G.G., Lazoski, C., Lazzari, C.R., Lopes, R.R., Lorenzo, M.G., Lugon, M.D., Majerowicz, D., Marcet, P.L., Mariotti, M., Masuda, H., Megy, K., Melo, A.C., Missirlis, F., Mota, T., Noriega, F.G., Nouzova, M., Nunes, R.D., Oliveira, R.L., Oliveira-Silveira, G., Ons, S., Orchard, I., Pagola, L., Paiva-Silva, G.O., Pascual, A., Pavan, M.G., Pedrini, N., Peixoto, A.A., Pereira, M.H., Pike, A., Polycarpo, C., Prosdocimi, F., Ribeiro-Rodrigues, R., Robertson, H.M., Salerno, A.P., Salmon, D., Santesmasses, D., Schama, R., Seabra-Junior, E.S., Silva-Cardoso, L., Silva-Neto, M.A., Souza-Gomes, M., Sterkel, M., Taracena, M.L., Tojo, M., Tu, Z.J., Tubio, J.M., Ursic-Bedoya, R., Venancio, T.M., Walter-Nuno, A.B., Wilson, D., Warren, W.C., Wilson, R.K., Huebner, E., Dotson, E.M., Oliveira, P.L., 2015. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. Proc. Natl. Acad. Sci. U.S.A. 112, 14936−14941.

Mougabure-Cueto, G., Ines Picollo, M., 2015. Insecticide resistance in vector Chagas disease: evolution, mechanisms and management. Acta Trop. 149, 70−85.

Ons, S., Lavore, A., Sterkel, M., Pedro Wulff, J., Sierra, I., Martinez-Barnetche, J., Henry Rodriguez, M., Rivera-Pomar, R., 2016. Identification of G protein coupled receptors for opsines and neurohormones in *Rhodnius prolixus*. Genomic and transcriptomic analysis. Insect Biochem. Mol. Biol. 69, 34−50.

Panzera, F., Ferrandis, I., Ramsey, J., Ordonez, R., Salazar-Schettino, P.M., Cabrera, M., Monroy, M.C., Bargues, M.D., Mas-Coma, S., O'Connor, J.E., Angulo, V.M., Jaramillo, N., Cordon-Rosales, C., Gomez, D., Perez, R., 2006. Chromosomal variation and genome size support existence of cryptic species of *Triatoma dimidiata* with different epidemiological importance as Chagas disease vectors. Trop. Med. Int. Health 11, 1092−1103.

Panzera, F., Ferreiro, M.J., Pita, S., Calleros, L., Perez, R., Basmadjian, Y., Guevara, Y., Breniere, S.F., Panzera, Y., 2014. Evolutionary and dispersal history of *Triatoma infestans*, main vector of Chagas disease, by chromosomal markers. Infect. Genet. Evol. 27, 105−113.

Panzera, F., Pita, S., Lorite, P., 2021. Chromosome structure and evolution of Triatominae: a review. In: Guarneri, A.A., Lorenzo, M.G. (Eds.), Triatominae - the Biology of Chagas Disease Vectors, Entomology in Focus, vol. 5. Springer Nature, Cham, pp. 65−99.

Panzera, F., Pita, S., Nattero, J., Panzera, Y., Galvao, C., Chavez, T., Rojas De Arias, A., Cardozo Tellez, L., Noireau, F., 2015. Cryptic speciation in the *Triatoma sordida* subcomplex (Hemiptera, Reduviidae) revealed by chromosomal markers. Parasites Vectors 8, 495.

Patterson, J.S., Gaunt, M.W., 2010. Phylogenetic multi-locus codon models and molecular clocks reveal the monophyly of haematophagous reduviid bugs and their evolution at the formation of South America. Mol. Phylogenet. Evol. 56, 608−621.

Pennington, P.M., Rivera, E.P., De Urioste-Stone, S.M., Aguilar, T., Juarez, J.G., 2021. A successful community-based pilot programme to control insect vectors of Chagas disease in rural Guatemala. In: Area-Wide Integrated Pest Management. CRC Press, pp. 709−727.

Pereira, F.M., Penados, D., Dorn, P.L., Alcantara, B., Monroy, M.C., 2022. The long-term impact of an Ecohealth intervention: entomological data suggest the interruption of Chagas disease transmission in southeastern Guatemala. Acta Trop. 235, 106655.

Perez de Rosas, A.R., Segura, E.L., Garcia, B.A., 2007. Microsatellite analysis of genetic structure in natural *Triatoma infestans* (Hemiptera: Reduviidae) populations from Argentina: its implication in assessing the effectiveness of Chagas' disease vector control programmes. Mol. Ecol. 16, 1401−1412.

Peterson, J.K., Hashimoto, K., Yoshioka, K., Dorn, P.L., Gottdenker, N.L., Caranci, A., Stevens, L., Zuniga, C., Saldaña, A., Rodriguez, S., 2019. Chagas disease in Central America: recent findings and current challenges in vector ecology and control. Curr. Trop. Med. Rep. 6, 76−91.

Piccinali, R.V., Gaunt, M.W., Gurtler, R.E., 2018. A microsatellite-based analysis of house infestation with *Triatoma infestans* (Hemiptera: Reduviidae) after insecticide spraying in the Argentine Chaco. J. Med. Entomol. 55, 609−619.

Piccinali, R.V., Marcet, P.L., Ceballos, L.A., Kitron, U., Gurtler, R.E., Dotson, E.M., 2011. Genetic variability, phylogenetic relationships and gene flow in *Triatoma infestans* dark morphs from the Argentinean Chaco. Infect. Genet. Evol. 11, 895−903.

Piccinali, R.V., Marcet, P.L., Noireau, F., Kitron, U., Gurtler, R.E., Dotson, E.M., 2009. Molecular population genetics and phylogeography of the Chagas disease vector *Triatoma infestans* in South America. J. Med. Entomol. 46, 796−809.

Pita, S., Gomez-Palacio, A., Lorite, P., Dujardin, J.P., Chavez, T., Villacis, A.G., Galvao, C., Panzera, Y., Calleros, L., Pereyra-Mello, S., Burgueno-Rodriguez, G., Panzera, F., 2021. Multidisciplinary approach detects speciation within the kissing bug *Panstrongylus rufotuberculatus* populations (Hemiptera, Heteroptera, Reduviidae). Mem. Inst. Oswaldo Cruz 116, e210259.

Pita, S., Lorite, P., Cuadrado, A., Panzera, Y., De Oliveira, J., Alevi, K.C.C., Rosa, J.A., Freitas, S.P.C., Gomez-Palacio, A., Solari, A., Monroy, C., Dorn, P.L., Cabrera-Bravo, M., Panzera, F., 2022. High chromosomal mobility of rDNA clusters in holocentric chromosomes of Triatominae, vectors of Chagas disease (Hemiptera-Reduviidae). Med. Vet. Entomol. 36, 66−80.

Pizarro, J.C., Gilligan, L.M., Stevens, L., 2008. Microsatellites reveal a high population structure in *Triatoma infestans* from Chuquisaca, Bolivia. PLoS Neglected Trop. Dis. 2, e202.

Quisberth, S., Waleckx, E., Monje, M., Chang, B., Noireau, F., Breniere, S.F., 2011. "Andean" and "non-Andean" ITS-2 and mtCytB haplotypes of *Triatoma infestans* are observed in the Gran Chaco (Bolivia): population genetics and the origin of reinfestation. Infect. Genet. Evol. 11, 1006−1014.

Reithinger, R., Tarleton, R.L., Urbina, J.A., Kitron, U., Gurtler, R.E., 2009. Eliminating Chagas disease: challenges and a roadmap. BMJ 338, b1283.

Ribeiro, J.M.C., Assumpcao, T.C.F., Pham, V.M., Francischetti, I.M.B., Reisenman, C.E., 2012. An insight into the sialotranscriptome of *Triatoma rubida* (Hemiptera: heteroptera). J. Med. Entomol. 49, 563−572.

Ribeiro, J.M.C., Genta, F.A., Sorgine, M.H.F., Logullo, R., Mesquita, R.D., Paiva-Silva, G.O., Majerowicz, D., Medeiros, M., Koerich, L., Terra, W.R., Ferreira, C., Pimentel, A.C., Bisch, P.M., Leite, D.C., Diniz, M.M.P., Junior, J.L.d.S.G.V., Da Silva, M.L., Araujo, R.N., Gandara, A.C.P., Brosson, S., Salmon, D., Bousbata, S., Gonzalez-Caballero, N., Silber, A.M., Alves-Bezerra, M., Gondim, K.C., Silva-Neto, M.A.C., Atella, G.C., Araujo, H., Dias, F.A., Polycarpo, C., Vionette-Amaral, R.J., Fampa, P., Melo, A.C.A., Tanaka, A.S., Balczun, C., Oliveira, J.H.M., Goncalves, R.L.S., Lazoski, C., Rivera-Pomar, R., Diambra, L., Schaub, G.A., Garcia, E.S., Azambuja, P., Braz, G.R.C., Oliveira, P.L., 2014. An insight into the transcriptome of the digestive tract of the nloodsucking bug, *Rhodnius prolixus*. PLoS Neglected Trop. Dis. 8, e2594.

Rojas de Arias, A., Messenger, L.A., Rolon, M., Vega, M.C., Acosta, N., Villalba, C., Marcet, P.L., 2022. Dynamics of *Triatoma infestans* populations in the Paraguayan Chaco: population genetic analysis of household reinfestation following vector control. PLoS One 17, e0263465.

Rotureau, B., Waleckx, E., Jamonneau, V., Solano, P., Molia, S., Debre, P., Dellagi, K., Morand, S., 2022. Enhancing research integration to improve One Health actions: learning lessons from neglected tropical diseases experiences. BMJ Glob. Health 7, e008881.

Schofield, C.J., Diotaiuti, L., Dujardin, J.P., 1999. The process of domestication in Triatominae. Mem. Inst. Oswaldo Cruz 94 (Suppl. 1), 375−378.

Schofield, C.J., Galvao, C., 2009. Classification, evolution, and species groups within the Triatominae. Acta Trop. 110, 88−100.

Schwarz, A., Medrano-Mercado, N., Schaub, G.A., Struchiner, C.J., Bargues, D.M., Levy, M.Z., Ribeiro, J.M.C., 2014. An updated insight into the sialotranscriptome of *Triatoma infestans*: developmental stage and geographic variations. PLoS Neglected Trop. Dis. 8, e3372.

Stevens, L., Monroy, M.C., Rodas, A.G., Hicks, R.M., Lucero, D.E., Lyons, L.A., Dorn, P.L., 2015. Migration and gene flow among domestic populations of the Chagas insect vector *Triatoma dimidiata* (Hemiptera: Reduviidae) detected by microsatellite loci. J. Med. Entomol. 52, 419−428.

Velasquez-Ortiz, N., Hernandez, C., Cantillo-Barraza, O., Medina, M., Medina-Alfonso, M., Suescun-Carrero, S., Munoz, M., Vega, L., Castaneda, S., Cruz-Saavedra, L., Ballesteros, N., Ramirez, J.D., 2022. Estimating the genetic structure of *Triatoma dimidiata* (Hemiptera: Reduviidae) and the transmission dynamics of *Trypanosoma cruzi* in Boyaca, eastern Colombia. PLoS Neglected Trop. Dis. 16, e0010534.

Villacis, A.G., Marcet, P.L., Yumiseva, C.A., Dotson, E.M., Tibayrenc, M., Breniere, S.F., Grijalva, M.J., 2017. Pioneer study of population genetics of *Rhodnius ecuadoriensis* (Hemiptera: Reduviidae) from the central coastand southern Andean regions of Ecuador. Infect. Genet. Evol. 53, 116−127.

Vivas, R.J., Garcia, J.E., Guhl, F., Hernandez, C., Velasquez, N., Ramirez, J.D., Carranza, J.C., Vallejo, G.A., 2021. Systematic review on the biology, ecology, genetic diversity and parasite transmission potential of *Panstrongylus geniculatus* (Latreille 1811) in Latin America. Mem. Inst. Oswaldo Cruz 116, e200528.

Waleckx, E., Salas, R., Huaman, N., Buitrago, R., Bosseno, M.F., Aliaga, C., Barnabe, C., Rodriguez, R., Zoveda, F., Monje, M., Baune, M., Quisberth, S., Villena, E., Kengne, P., Noireau, F., Breniere, S.F., 2011. New insights on the Chagas disease main vector *Triatoma infestans* (Reduviidae, Triatominae) brought by the genetic analysis of Bolivian sylvatic populations. Infect. Genet. Evol. 11, 1045−1057.

Waleckx, E., Camara-Mejia, J., Jesus Ramirez-Sierra, M., Cruz-Chan, V., Rosado-Vallado, M., Vazquez-Narvaez, S., Najera-Vazquez, R., Gourbiere, S., Dumonteil, E., 2015a. An innovative Ecohealth intervention for Chagas disease vector control in Yucatan, Mexico. Trans. R. Soc. Trop. Med. Hyg. 109, 143−149.

Waleckx, E., Gourbiere, S., Dumonteil, E., 2015b. Intrusive versus domiciliated triatomines and the challenge of adapting vector control practices against Chagas disease. Mem. Inst. Oswaldo Cruz 110, 324−338.

Waleckx, E., Perez-Carrillo, S., Chavez-Lazo, S., Pasos-Alquicira, R., Camara-Heredia, M., Acuna-Lizama, J., Colli-Balam, F., Camara-Mejia, J., Ramirez-Sierra, M.J., Cruz-Chan, V., Rosado-Vallado, M., Vazquez-Narvaez, S., Najera-Vazquez, R., Gourbiere, S., Dumonteil, E., 2018. Non-randomized controlled trial of the long-term efficacy of an Ecohealth intervention against Chagas disease in Yucatan, Mexico. PLoS Neglected Trop. Dis. 12, e6605.

Walter, A., Lozano-Kasten, F., Bosseno, M.F., Ruvalcaba, E.G., Gutierrez, M.S., Luna, C.E., Baunaure, F., Phelinas, P., Magallon-Gastelum, E., Breniere, S.F., 2007. Peridomicilary habitat and risk factors for *Triatoma* infestation in a rural community of the Mexican occident. Am. J. Trop. Med. Hyg. 76, 508−515.

Weirauch, C., Munro, J.B., 2009. Molecular phylogeny of the assassin bugs (Hemiptera: Reduviidae), based on mitochondrial and nuclear ribosomal genes. Mol. Phylogenet. Evol. 53, 287−299.

WHO, 2020. Ending the Neglect to Attain the Sustainable Development Goals: A Road Map for Neglected Tropical Diseases 2021−2030. World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.

WHO, 2022. Chagas Disease (American Trypanosomiasis), Fact Sheet. 2022. http://www.who.int/mediacentre/factsheets/fs340/en/. (Accessed 23 May 2022).

Wright, S., 1943. Isolation by distance. Genetics 28, 114−138.

CHAPTER

16

Modern Morphometrics of Arthropods: A Phenotypic Approach to Species Recognition and Population Structure

Jean-Pierre Dujardin

Institut de Recherche pour le Développement (IRD), UMR INTERTRYP IRD-CIRAD, University of Montpellier, Montpellier, France

1. Introduction

The phenotype is the product of the interaction between genes and environment. Phenotypic variation is then an expected outcome of more than one factor. It can be scored by measurable changes in anatomy, morphology, physiology, life history, behavior, etc. (West-Eberhard, 1989; Gadagkar and Chandrashekara, 2005).

This chapter considers the phenotype as a set of metric properties and their variation. It presents briefly a few common statistical methods related to modern morphometrics, and it suggests some applications in the study of arthropods.

1.1 Modern and Traditional Morphometrics

Morphometric techniques applied to organismal forms aim at measuring their size, shape, and the relation between size and shape (allometry). Before the so-called "morphometric revolution" (Rohlf and Marcus, 1993), shape was a statistical abstraction computed from distances measured between anatomical points (landmarks). The replacement of linear measurements between landmarks by their coordinates in a common space, and the subsequent visualizing techniques, represented a giant step in the direct study of forms. The shift from traditional morphometrics to more complex geometric functions was facilitated by the development of image processing tools.

Two conceptually and statistically separate approaches are (i) the landmark-based morphometrics, based on the relative position of a few anatomical features ("true" or "traditional" landmarks, see [Rohlf, 1990; Bookstein, 1991a]) and (ii) the outline-based morphometrics, which captures the contour of forms as a sequence of close points collected along the contour itself ("pseudo-landmarks," see [Kuhl and Giardina, 1982; Rohlf, 1986; Lestrel, 2000]).

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00004-5
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

2. Landmark-Based Geometric Morphometry

The coordinates of anatomical landmarks contain not only size, such as distances between landmarks, but also shape, i.e., their relative position in a given system of orthogonal axes. In common practice, size, and shape are derived from a configuration of landmarks collected on a nonarticulated part, often a single organ (however see [Adams, 1999]). A few anatomical landmarks available on a wing (or any measurable part of the body) do not completely describe the wing, nor do they describe the complete body. However, provided there is anatomical correspondence among individual landmarks, only a partial capture of shape is needed to allow valid comparisons among populations and species.

2.1 The Landmark

For shape comparisons, great importance is given to the quality of landmarks in terms of comparability from one individual to another. In other words, the choice of suitable landmarks relies on their homology. In the morphometrics practice, homology is "correspondence of parts" with no specification about whether the parts correspond with respect to structure, development, or phylogeny (Smith, 1990). Anatomical landmarks are homologous in the sense that they are relocatable points, and according to this criterion various levels of quality have been recognized (see type I, II, and III landmarks, having decreasing levels of precision [Bookstein, 1991a]). If individuals belong to a single species, homologous landmarks are probably similar due to common descent because all members of the species come from a common ancestor. If they belong to different species, there is no guaranty that homologous landmarks are similar due to common descent, except maybe if the different species descend from a common ancestor (Lele and Richtsmeyer, 2001). This homology is one of the criteria making landmark-based morphometrics a suitable tool for systematics (see Section 7).

2.2 Landmark-Based Size

The global size of the body of an arthropod is hardly represented by only one of its many parts. In traditional morphometrics, to compare individuals, the measurement of size was preferably associated to the organ considered (head capsule, thorax, legs, wings, etc.). Alternatively, an estimator of global size could be constructed using various parts of the body: in that case, the first principal component combining these measurements was used. Nevertheless, entomologists often selected a single measurement as an estimator of body size. For instance, to estimate the relative size of different insects (or, at least, the relative size of their wings), the length of the wing along its largest axis has been traditionally used (Nasci, 1990; Siegel et al., 1992; Lehmann et al., 2006). For winged insects, the relationship between wing and body size was often assumed rather than demonstrated (Siegel et al., 1992; Morales Vargas et al., 2010).

Size variable: the centroid size (CS). The landmark-based morphometrics uses the coordinates of the landmarks themselves to produce an estimate of size. This estimate is the one delimited by the configuration of landmarks, and is called the centroid size (CS). It is the square root of the sum of the squared distances from the centroid point to each landmark (see Gower, 1971 in [Rohlf, 1990]). Depending on the relative position of all landmarks this measure is the most inclusive one. It has been shown that, in the case of small, circular variation at each landmark, this estimator of isometric change of size is not correlated to shape variation (Bookstein, 1991b). For these reasons, CS is the most

commonly used estimate of size in landmark-based studies. It is expressed in pixels, i.e., units relative to the resolution of the viewing device (most often a computer display). As a scalar it is less sensible to small digitization errors, and can be shared among systematists provided the pixels have been converted into absolute length units (inches, centimeters, millimeters, etc.). Thus, to allow for exchangeability of CS, an image on which landmarks may be collected should contain a scale for size (for instance the picture of a reticle) allowing the conversion of pixels to absolute units.

The same problem as for traditional morphometrics exists here also: the body size is not necessarily correlated to the size (here centroid size) of the anatomical part under study. However, it seems that the centroid size may be correlated with linear measurements taken on the same body part, to some extent. The relationship of CS values and the traditional wing length of the mosquito *Aedes aegypti* showed good correlation (Morales Vargas et al., 2010). Actually, the correlation of CS values with traditional interlandmark distances (ILD) is itself correlated to the relative dimensions of ILD: the largest the ILD, the highest its correlation with CS (Fig. 16.1).

2.3 Landmark-Based Shape

Not only in medical entomology but also in many fields where morphometrics is applied, shape has been traditionally described as the ratio of one dimension to another. Although intuitively the ratio may appear as capable of scaling for size, it often does not (Klingenberg, 1996; Albrecht et al., 1993; Burnaby, 1966; Dujardin and Slice, 2006). Moreover, the ratios may introduce some statistical

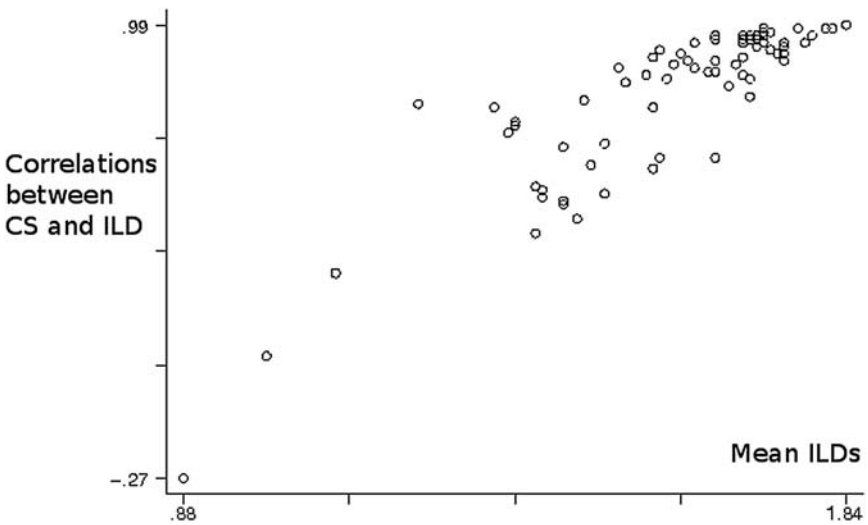


FIGURE 16.1

Each circle represents on the vertical axis the correlation found between an interlandmark distance (ILD) and the centroid size (CS) of the wing computed from the total set of landmarks. The value of each ILD on the horizontal axis is an average (Mean ILD) obtained from the total number of wings (78) examined in this sample. The wings belong to *Ae. aegypti*.

drawbacks (Albrecht et al., 1993). Angles do not improve the situation since they are another kind of ratio (Burnaby, 1966).

Residual coordinates, orthogonal projections, tangent space variables, Procrustes components, partial warps, relative warps. In geometric morphometrics, the shape of a configuration of landmarks is represented by their relative positions as contained in their coordinates. The raw landmarks coordinates contain the information about shape, but they also contain artifactual variation due to the image position (on the computer screen), its size, and its orientation. Therefore, shape must be described by new variables having removed these artifacts. This is obtained through the **Procrustes** superimposition on a consensus configuration. If using the least squares fit as an optimality criterion, the statistical procedure of superposition is called Generalized Procrustes Analysis (GPA). It is currently the most common procedure, but other techniques also exist (Zelditch et al., 2004).

The residual coordinates after a GPA provide a shape description relative to the consensus configuration of landmarks, they thus depend on the composition of the group under study. If other specimens (i.e., coordinates) are added to the analysis, the consensus configuration changes and shape variables must be recomputed accordingly (Rohlf and Marcus, 1993; Adams et al., 2004). This sample dependency represents a serious inconvenient to share landmark coordinates between biologists, especially when it is about systematics (Dujardin et al., 2010). Such limitation in the exchangeability of shape data is restricted to the landmark-based data, not to the pseudolandmarks-based data (see Section 3).

When working on a two-dimensional space (based on pictures), the dimensionality of the landmarks configuration is 2n (for n landmarks). The superimposition procedure induces the loss of 4 degrees of freedom (Rohlf, 1990) due to the constraints imposed by the scaling, translation, and rotation transformations (Procrustes superposition). The resulting residual coordinates lie in a curved, non-Euclidean space unsuitable for standard statistical tests (except resampling methods). They must be further modified by a rigid rotation so that they can be studied using classical statistical tools (Rohlf and Bookstein, 2003).

Thus, the aligned configurations (residual coordinates) are projected onto an Euclidean space that is tangent to the curved space (shape space) at the consensus configuration. The coordinates in the tangent space are then used as shape variables ("tangent space variables," "Orthogonal projections," or "Procrustes residuals") in all subsequent analyses.

An alternative, mathematically equivalent procedure, uses the eigenvectors of the bending-energy matrix (Bookstein, 1991b). The resulting shape variables are then called "partial warps" scores (PW). The PW, or their principal components, namely the "relative warp" (RW), may be used in classical statistical analyzes and in visualization of shape changes through the deformation grids, i.e., D'Arcy Thompson-like plots showing the geometry of shape changes between objects (Bookstein, 1991b).

In most cases a PCA on the "tangent space coordinates" generating "Procrustes components" will provide very similar results to a relative warps (RW) analysis. In summary, shape variables are principal components of either "tangent space variables" (Procrustes components) or of "partial warps" (relative warps).

The practical interest of using principal components (PC) is that the number of input variables can be reduced. The few first PC generally represent a significant fraction of shape. The number of first PC to use is rather subjective but it can follow some rules (Baylac and Frieβ, 2005).

2.4 Semilandmarks for Open Curves

A special development of type III landmarks, called "semilandmarks," allows the description of curved lines between two classical landmarks (Bookstein, 1990). The development of semilandmark-based analysis allows one to use the "Procrustes paradigm" as a unified framework for both interlandmarks curves and landmark analyses. It does not make obsolete the outline approach based on Fourier's analysis, since the latter applies to close curves and may apply to structures without any anatomical landmark (see Section 3).

The term "semilandmarks" refers to a series of points that are located relative to one another and collectively describing the shape of an open curve (Bookstein, 1997a,b) or a surface (Gunz et al., 2005). Semilandmarks and landmarks III both are defined relative to other landmarks, so that they are conceptually similar, even if there are minor differences between them. While landmarks III can focus on a single point, semilandmarks are exclusively about capturing curves (or surfaces) between landmarks. While the position of the semilandmarks along a curve is more or less arbitrary, depending on other landmarks and semilandmark, the position of landmarks III is defined according to some geometric construction rule, or template. In spite of these apparent differences, both kinds of landmarks can be treated the same way (Dujardin, 2019) because they suffer from the same problem, the loss of some level of independence.

Thus, semilandmarks are special points making it possible to capture curves or surfaces where landmarks are sparse. Superimposition of semilandmarks differ from what is currently applied to landmarks, ways, and tools for collecting them may also differ from collecting landmarks. Semilandmarks can be collected by simply digitizing points along a curve or a surface but can also be collected in a more systematic way by the use of a template.

Currently, the methods specifically developed for semilandmarks make use of a minimization criterion (Gunz and Mitteroecker, 2012), either minimum bending energy (Bookstein, 1991a) or minimum Procrustes distance (Rohlf and Slice, 1990). Both methods for semilandmarks processing do not produce exactly the same results (Perez et al., 2006; Gunz and Mitteroecker, 2012). As they imply the sliding of points along the tangent to the curve (or tangent plane to the surface), the term "sliding landmarks" is also used to refer to semilandmarks. Another statistical treatment of semilandmarks has been recently suggested which does not use any sliding procedure, based on the idea that landmarks III may be located along a template describing the shape of an open curve (Dujardin, 2019).

3. Pseudolandmark-Based Geometric Morphometry (Outlines)

Anatomical, "true" landmarks are opposed to "pseudo-landmarks" used in the outline-based approach. Pseudo-landmarks describe close contours or boundary outlines and do not depend on the presence of traditional landmarks. They can exist with no anatomical landmark at all, or can include one or more of them. They should not be confused with semilandmarks (see previous section): semilandmarks describe curves between landmarks (see above), pseudolandmarks describe closed curves.

Pseudolandmarks are of another nature than landmarks or semilandmarks because comparability is not expected from them separately, but from the structure they describe. Thus, if carefully chosen, like the contour of a tick's body (Dujardin et al., 2014), the genital leaflet of a mosquito (Boussès et al., 2013), or an internal cell of the wing (Francoy et al., 2012; Dujardin et al., 2015; Chonephetsarath

et al., 2021), a contour represents a homologous structure allowing interindividual and/or intergroup comparisons.

Because of these differences, it is expected that the statistical treatment applied to pseudolandmarks is different from the one applied to landmarks (and semilandmarks). The idea is however the same: from pseudolandmarks, one wants to extract separate information about size and shape, with the possibility of visualizing shape changes from one individual to another.

3.1 Pseudolandmark-Based Shape

Various techniques allow to derive shape variables from a digitized outline (Kuhl and Giardina, 1982; Lohmann, 1983). We will give a few words here about the most common one, the elliptic Fourier algorithm (EFA). To describe a close contour, the superiority of EFA over other methods (Rohlf, 1990) is now commonly accepted: the algorithm does not require the points (pseudolandmarks) to be equidistant (as in eigenshape analysis [Lohmann, 1983]), and the use of the normalized coefficients "saves the investigator from having to worry about aligning the images in a standard fashion" (Rohlf and Archie, 1984).

Briefly, the observed contour is decomposed in terms of sine and cosine curves of successive frequencies called harmonics, and each harmonic is described by four coefficients. With this method (Kuhl and Giardina, 1982), the first harmonic ellipse parameters are used to standardize the Fourier coefficients so that they are invariant to size, rotation, and the starting position of the outline trace. By doing this, the three first coefficients become constant (1, 0, and 0) and are not used in the remaining analyses. The fourth coefficient is related to the width-on-length ratio of the outline (Lestrel, 1989).

As for the conventional Fourier analysis, the maximum number of harmonics is half the number of sample points (pseudolandmarks). The optimal number of harmonics can be selected in various ways: by appraising the goodness-of fit using the sum of squared distances between the original data and reconstructed outline, by examining the spectrum of harmonic Fourier power, or simply by visualizing the reconstructed outlines (Lestrel, 1989, 1997; Claude, 2008).

Ideally, more harmonics should be able capture more shape parameters and produce a higher level of discrimination. However, more harmonics also inflate the digitization error and an optimal number should be selected (Crampton, 1995; Renaud et al., 1999; Firmat et al., 2010).

3.2 Pseudolandmark-Based Size

The Fourier's coefficients described above are normalized by the semi grand axis of the first harmonic ellipse. The latter may be used as an estimate of global size, as well as the root-squared area of this same starting ellipse (Rohlf, 1990), the root-squared area of the digitized contour or its perimeter.

4. Allometry

The study of allometry has a long history in biology (Huxley, 1932), which has mostly focused on the covariation among structures: for example, the relationship between fore- and hind limbs in theropod dinosaurs (e.g., *T. rex*), or between brain size and body height in humans (Rushton and Ankney, 2009).

Since organismal shape can be explained by changes in linear dimensions, it is expected that size and shape are not completely independent attributes. In morphometrics, allometry refers to the relationship between size and shape.

Various kinds of allometry have been described, which might result from the different developmental processes (Klingenberg, 1996). During development, organisms grow larger and in the meantime, because the rate of growth differs between their various parts, their shape also changes. Such covariation between size and shape throughout the ontogeny is called "ontogenetic" allometry. When considering different individuals of the same population and at the same developmental stage, there is also often some variation in size and shape. Such allometry is referred to as "static" allometry and is opposed to the "evolutionary" allometry that can be identified across species or geographic populations.

Thus, geometric shape variables as derived from either landmark-based or outline-based approaches are not allometry-free variables. There is an allometric residue contained in the shape variables, which is traduced by a correlation between size and shape components. Geometric morphometrics, by clearly separating shape and size variables, allows one to specifically investigate their relationship.

To tentatively remove the influence of size on shape, one might be tempted to remove the allometric residue of shape variables. In systematics studies, this could be justified when dealing with conspecific populations ("static" allometry) (Klingenberg, 1996; Caro-Riaño et al., 2009; Morales Vargas et al., 2010). Thus, assuming a common model of growth is not rejected, one can use the growth model to predict allometry-free shape variation among groups after fixing size to one value. The method is based on MANCOVA and described in the help files of the TPSregr software (J. Rohlf). The tentative removal of the allometric effect on shape seems less justified when comparing the metric properties of different species. For interspecific comparisons indeed, the allometric variation is likely to be part of the evolutionary process ("evolutionary allometry"), its inclusion in the variable is relevant to systematics.

5. Some Considerations About the Genetics of Metric Change

To what extent are shape differences among samples attributable to genetic differences? In other words, does shape have a strong genetic determinism or is it mostly plastic? And how many genes are involved in the shape variation? Are there few genes with major effects or many genes of small effects?

5.1 Shape as a Polygenic Character

Shape appears as a classical polygenic character (Klingenberg and Leamy, 2001). Evidence for strong genetic determinism of shape was suggested by significant association with chromosome polymorphism (Bitner-Mathé and Klaczko, 1999; Orengo and Prevosti, 2002; Hatadani and Klaczko, 2008), and confirmed by genetic studies (Breuker et al., 2006; Patterson and Klingenberg, 2007). When studies on quantitative trait loci (QTL) were applied to the shape and size of mouse mandible, many QTL were identified for shape (Klingenberg et al., 2004), many more than for size (Workman et al., 2002; Klingenberg et al., 2001). Few studies are found in insects, also fitting the idea of genetic

determinism (Iriarte et al., 2003) and polygenic inheritance (Shrimpton and Robertson, 1988; Long et al., 1995).

5.2 Genetic Drift

Since shape seems the output of a cascade of genes, it is expected that in natural conditions genetic drift be a common factor of shape variation (Falconer, 1981).

Laboratory experiment reproducing conditions favoring genetic drift between lines sharing the same environment were performed in *Ae. aegypti*. Using a set of three isofemale lines of *Ae. aegypti* monitored during 10 generations, a significant shift of shape appeared in one line, with nonsignificant changes in corresponding size (Jirakanjanakit et al., 2008). In this experiment, the change apparently produced by genetic drift did not affect the same landmarks as those affected by larval food or density variation (Jirakanjanakit et al., 2007).

Genetic drift has been the most parcimonious explanation to field observations reporting significant shape differences between geographic populations (De la Riva et al., 2001; Gumiel et al., 2003; Dujardin et al., 2003; Dujardin and Le Pont, 2004; Camara et al., 2006; Aytekin et al., 2007; Henry et al., 2010).

5.3 Heritability

Since heritability is an indirect measure of genetic variation, it is depending on the population under study. It is also depending on the trait under study. In insects, morphological traits commonly have the highest heritability values compared to other types of traits such as life history, probably because the former are less concerned with fitness (Roff and Mousseau, 1987). The consistent values of shape heritability measured so far (Bitner-Mathé and Klaczko, 1999; Roff and Mousseau, 1987; Gilchrist and Partridge, 2001; Hoffman and Shirriffs, 2002) suggest that a large fraction of morphometric divergence seen between natural populations of insects (Camara et al., 2006; Henry et al., 2010; Morales Vargas et al., 2010) may be due to additive effects of genes.

Size in insects may show consistent heritability values (Daly, 1992; Lehmann et al., 2006), so that they can be experimentally selected to constitute subpopulations genetically distinct for size (Anderson, 1973; Partridge et al., 1994). In quantitative genetics, the response to selection is embodied in the breeders' equation R = h2s (Falconer, 1981), where "s" is the selection differential, "R" is the response to selection, and "h2" is heritability. h2, the ratio of VG upon VP—the part of phenotypic variance (VP) that is due to genetic differences (VG)—plays a central role in this equation and conditions the response to selection (see [Houle, 1992] for a critic of the use of heritability).

It is therefore tempting to ask how can one measure shape heritability? How can a specific shape evolve under the action of natural selection?

The problem that arises here is that shape is inherently a multivariate trait. As such, its genetic variation cannot be assessed by a scalar heritability as for univariate quantitative traits. While it is theoretically possible to derive such a univariate measure of shape genetic variation using some kind of shape distance (e.g., Procrustes distance, see [Monteiro et al., 2002]), it has been shown to be misleading (Klingenberg, 2003; Klingenberg and Monteiro, 2005). For example, imagine a situation where such a global shape heritability would be high—say, 0.8. This would give the false impression

that shape could be selected in any direction. However, the multivariate version of the breeder's equation (Lande, 1979) shows that the multivariate response to selection is highly dependent on the genetic correlations among traits (the G matrix). For shape, G represents the genetic variances and covariances among landmarks coordinates, or shape dimensions. And these correlations are by no mean uniform in the shape space. In other words, some combinations of shape changes are generally genetically more variable than others, imposing selection a relative constraint. It is even conceivable that some directions are fully devoid of any variation at all, a situation referred to as an absolute constraint (Mezey and Houle, 2005). This means that even with a high value of global heritability, because of the anisotropic nature of genetic variation (i.e., genetic variation is not evenly distributed in the shape space), it might well be impossible to select at all in some specific direction (i.e., some shape changes are impossible to reach through selection).

Fortunately, the measurement of heritability is not mandatory to the epidemiological interpretation of natural metric variation. Without having attempted to measure the heritability, our experimental observations in the Triatominae (Dujardin et al., 2007) and in the mosquitoes (Jirakanjanakit et al., 2008) suggested that the shape of the wings (landmark approach) keeps the memory of the parents at least over a few generations. These laboratory observations supported the idea of using shape as an indicator of population structure (see Section 8.1.1).

5.4 Hidden Genetic Variability

Unexpectedly, some heritable changes seem to be triggered by environmental events. The mechanisms of such phenomenon, which is reminiscent of Lamarkian "inheritance of acquired characters," do not depart from orthodox genetics. They have been named "genetic assimilation" (Waddington, 1953) or "autonomization" (Schmalhausen, 1949 in [Levit et al., 2006]), and, later, "genetic accommodation" (West-Eberhard, 1989).

The mechanisms by which an environmentally induced phenotype may become heritable are entirely compatible with concepts of classical neo-Darwinian evolutionary biology. Indeed, the environmental trigger (since this is the disputable one) just uncovers previously cryptic genetic variation (Gibson and Dworkin, 2004).

Thus, there are genetic mutations that can remain masked until the environment (or another mutation) reveals them (Bergman and Siegal, 2003). Adaptationist ("capacitator," see for instance [Rutherford and Lindquist, 1998; Debat et al., 2006; Suzuki and Nijhout, 2006; Pennisi, 2006]) and other theories ("robustness" as a generic property of complex systems, see for instance [Kitano, 2004; Wagner, 2008; Masel, 2013]) provided plausible scenarios explaining the existence of hidden genetic variability.

5.5 Hybridism

Various observations have been made in the Triatominae (vectors of Chagas disease), where interspecific hybridism is relatively common (Usinger et al., 1966).

In the brasiliensis complex of Chagas disease vectors, historical hybridism has been suggested as a possible homoploid speciation mechanism for one of its members, *T. macromelasoma*, which would have arisen from the hybrid cross between *T. juazeirensis* and *T. brasiliensis* (Costa et al., 2008). This hypothesis has been raised after a geometric morphometric study showing the intermediate position of

the hybrid species, and is now consolidated by morphological, ecological and genetic data (Costa et al., 2016).

Centroid size was increased in hybrids obtained from two close species, initially considered as two subspecies (Costa and Felix, 2007): *T. brasiliensis* and *T. juazeirensis*. It was larger than the midparent size, and larger than the largest parent's size, suggesting heterosis pointing to a consistent genetic divergence of the parents. Extending this study to experimental hybrids among the four members of the Brasiliensis complex, it was possible to show a linear relationship between the genetic divergence of the parents (Costa et al., 2001; Costa and Felix, 2007) and the increase in size of their offspring (Fig. 16.2). Contrary to size, shape of the hybrids remained intermediate between parents (Costa et al., 2008).

Between cryptic species of *Diachasmimorpha longicaudata*, a hymenopteran parasitoid of fruit flies, the size of the hybrids was larger than that of midparents, although not significantly larger (Kitthawee and Dujardin, 2009), and hybrids showed intermediate shape on the morphological space obtained from the two first relative warps.

These studies indicate different genetic mechanisms affecting size and shape, suggesting size as a character prone to show heterosis in case of genetically differentiated parents.

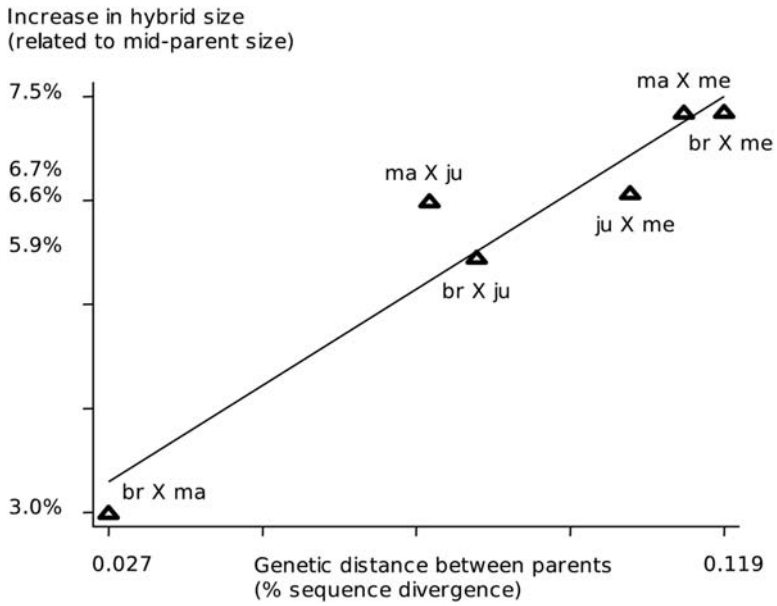


FIGURE 16.2

Relationship between size in hybrids and genetic distances between parents belonging to the Braziliensis complex. On vertical axis, the increasing of size relative to midparent size. On horizontal axis, the genetic distance between parents as inferred from mitochondrial DNA sequence. ma, *Triatoma brasiliensis macromelasoma*; me, *T. b. melanica*; br, *T. b. brasiliensis*; ju, *T. b. juazeirensis*.

From Dujardin, J.P., Costa, J., Bustamante, D., Jaramillo, N., Catalá, S., 2009. Deciphering morphology in Triatominae: the evolutionary signals. Acta Trop. Acta Tropica 110, 101–111.

No such heterosis was observed for the progeny between seven laboratory colonies of *T. protracta* (Dujardin et al., 2007), or between each of the five subspecies of *T. protracta*. Each hybrid had an intermediate centroid size between parents. Accordingly, geometric shape was also intermediate between parents (Dujardin, unpublished data). If we consider as a "rule" the linear relationship observed between hybrid increase of size and genetic distances between parents, these observations were in agreement with conspecific *T. protracta* specimens rather than with distinct taxa.

6. Some Considerations About Phenotypic Variation

It is common place that environment affects metric properties. Using geometric morphometrics, it is possible to specify which metric property is affected, size and/or shape, and how it is affected.

6.1 Phenotypic Plasticity

The genotype does not give rise to a single phenotype, but to a range of possible phenotypes. The "reaction norm" is the whole repertoire of possible phenotypes that may occur for a given genotype in all environments (Dobzhansky, 1971; Schlichting and Pigliucci, 1998). The reaction norm can easily be explored in laboratory experiments (Hillesheim and Stearns, 1991; David et al., 1994; Debat et al., 2003; Jirakanjanakit et al., 2007; Caro-Riaño et al., 2009). By definition, phenotypic plasticity (PP) is the occurrence of phenotypic variation of a single genotype interacting with different environments (Schlichting and Pigliucci, 1998).

A new phenotype expressed in a new environment may be adaptive. To this condition, PP can aid speciation by making available a different phenotype upon which natural selection can act. In such scenario, speciation would start with PP, not reproductive isolation (Görür, 2005). In Triatominae, the vectors of Chagas disease, such a scenario is apparent (Dujardin et al., 1999b) and many examples exist of morphologically and ecologically recognized species which can still interbreed (Dujardin et al., 2009). Understanding the causes and consequences of phenotypic variation is important for understanding the mechanisms of evolution. However, the genetic mechanisms underlying the evolutionary importance of PP (Waddington, 1953; Thompson, 1971; West-Eberhard, 1989; Schlichting and Pigliucci, 1998) received so far only few experimental confirmations (Rutherford and Lindquist, 1998; Suzuki and Nijhout, 2006).

Contrary to its evolutionary importance, the ecological importance of PP is easy to understand: populations or species having wider adaptive plastic responses can enlarge their ecological niches. For instance, among the more than 140 species of Triatominae, the vectors of Chagas disease, a few species have been able to colonize human structures. Within some of these species, the comparison of "domestic" and sylvatic subpopulations highlighted significant size differences, sylvatic insects being generally larger. Were these species more plastic than the others so that they could reduce their size as apparently required to survive in artificial ecotopes? Was the size a secondary event selected by the domestic environment? Using emerging adults of an inbred line of *R. pallescens* (Caro-Riaño et al., 2009), evaluated size and shape variation of the heads and the wings under simulated conditions of sylvatic (low population density, low feeding frequency) and domestic (higher density, higher feeding frequency) habitats. Results demonstrated that selection was not needed to account for observed changes between sylvatic and domestic ecotopes. The significant size reduction was obtained in a single generation as a direct response to combined population density and feeding frequency.

In addition to explain diversity and adaptation, PP also impacts our understanding of taxonomy, because it suggests that species characteristics are not immutable, but are influenced by the environment and can be highly variable (Ananthakrishnan, 2005).

6.2 The Regulation of Phenotype

To the many sources of phenotypic changes, the organism opposes homeostatic processes. Two components of this homeostasis are canalization and developmental stability. The two components are relatively easy to distinguish: canalization would be the between-individual stability of development, while developmental stability would refer to the within-individual stability. Canalization would then be defined as a buffering process against external and/or mutational perturbation from one (micro) environment to another (Wilkins, 1996), while developmental stability would refer to the organism withstanding random accidents during development in its own environment (Graham et al., 1993).

For a review of these concepts and their many definitions, please see Debat and David (2001). Canalization and developmental stability have been suggested to be independent processes (Debat et al., 2000; Réale and Roff D, 2003), but this position is not consensual, and others have suggested that canalization and developmental stability are the same thing (Breuker et al., 2006), or even, that there is no need for any specific buffering process (Félix and Barkoulas, 2015).

6.2.1 Canalization

The term "canalization" is due to Waddington, corresponding to the "stabilizing selection" of Schmalhausen (Levit et al., 2006). Here are the terms used by Schmalhausen himself: "*Every adaptive modification is an expression of a norm of reaction, which went the long way of historical development under changing conditions. It is connected with the establishment of 'canals' through which a certain modification develops (Waddington talks about the 'canalization' of development). An external factor operates only to switch the development into one of the existing canals*" (Schmalhausen, in [Levit et al., 2006]).

Thus, as for phenotypic plasticity, canalization is not a property of a species or of a population, but of a genotype (Dworkin, 2005). Different traits of a single organism can be examined for their relative canalization by studying their natural or experimental variation using different lines, populations, or species.

We showed that, contrary to size changes, shape divergence between two neotropical sandflies, *L. ayrozai* and *L. geniculata*, remained similar in spite of striking altitudinal variation as found between the Andes and the Amazon basin (Dujardin et al., 2003). A similar study comparing the wing shape of transcontinental populations of two close mosquito species, *Ae. aegypti* and *Ae. albopictus*, showed that species differentiation based on wing shape, but not on its size, was not altered by transcontinental migration during the last decades. Thus, in spite of the many possible situations supposed to affect shape, like environmental changes, possible environmental stress, likely founder effect, possible genetic drift, species competition, etc., two species were still distinguishable at the same landmark locations (Henry et al., 2010). This relative constancy of shape patterns within each species contrasted with the lability of size.

Experimentally, the size *Ae. aegypti* was significantly affected by a simple change in the food concentration or in the larval density; shape was not affected except as a passive consequence of size change (Jirakanjanakit et al., 2007). Another experiment comparing size and shape responses used

highly inbred lines of *R. pallescens* (Triatominae): the plastic response scored for the centroid size of the wing to the laboratory conditions of "domesticity" was not observed for the shape, except as an allometric change (Caro-Riaño et al., 2009).

Incidentally, the apparently higher canalization of shape makes this trait a suitable character for populations and species distinction (Dujardin and Le Pont, 2004).

6.2.2 Developmental Stability

Stress, either environmental or genetic, and other causes also such as relaxation of selection pressure (domestic animals), may have a disturbing effect on the organism development. The use of morphometrics as an indicator of stable development has been performed by estimating the frequency of abnormal phenotypes (phenodeviants). More frequently, it has been estimated by the amount of fluctuating asymmetry (Palmer and Strobeck, 1986).

Fluctuating asymmetry (FA) of size is a nondirectional asymmetry, as opposed to the classic directional asymmetry where one side is, on average, larger from the other side (which is common for internal organs of vertebrates). For a given sample or population, FA is computed as the variance of bilateral differences scored between individuals, either of size or shape. It requires to remove the measurement error and is obtained through an adapted ANOVA procedure (Palmer and Strobeck, 1986; Klingenberg and McIntyre, 1998). It has to be distinguished from antisymmetry, another kind of nondirectional asymmetry, less frequent however, where bilateral differences do not have a gaussian distribution (they are always more or less of the same amount).

Stress is the most commonly assigned cause to FA; it can be the infection by a virus or parasite, it can be a difficult conquest of a new habitat, or simply a different source of food (Nattero et al., 2015). Although FA has been widely used as a bioindicator of environmental and genetic stress, quite a strong controversy has occurred in the late 1990s, mostly around its genetic bases and in particular its heritability (Houle, 1998). It has been suggested that publications biases have occurred, casting doubts upon the reliability of the literature on FA and in particular its role as an indicator of fitness (Leamy and Klingenberg, 2005; Van Dongen, 2006). One should thus use FA very cautiously when attempting to relate it to any other population parameter.

In the few domestic populations of *T. sordida*, a potential vector of Chagas disease in Bolivia and Argentina, significant FA was found, whereas no FA at all could be disclosed in their sylvatic counterparts (Dujardin et al., 1999b). In *Glossina* flies (vectors of sleeping sickness and nagana), five species were compared between laboratory and natural conditions of life, showing a strikingly reduced FA in the laboratory (Kaba et al., unpublished data). In natural conditions, the adults caught during the rainy season showed higher FA than those collected in the dry season (Djohan et al., phD thesis).

Other measures of developmental stability could be used which are considered as developmental invariants, like fractal dimensions, although they were described for plants and vertebrates only (Graham et al., 1993). The use of shape variation itself in response to environmental stress has been advocated for insects, although these changes were considered as very subtle ones (Hoffmann et al., 2005). An advantage it could have on asymmetry analyzes would be a possible signature changes in landmarks characteristic of a specific environmental stress. A disadvantage is that, contrary to symmetry which is expected to be perfect, and contrary to the frequency of phenodeviants which is expected to be zero, there is no "expected shape," thus there is no way to use shape changes to measure the degree of stress.

7. Taxonomic Questions

According to the typological concept of species, different species have different morphologies. This concept has been progressively disputed because of the discovery of "sibling" or "isomorphic" species (Mayr, 2000). Sibling species are morphologically identical entities in spite of being recognized as different species according to other, modern concept(s) of species. The most frequently used modern species concept in entomology is the biological one (Mayr, 1969, 2000), which is based on the existence of reproductive isolation between species (as natural populations). Other frequently used concepts are the Hennigian concept (Hennig, 1965), the evolutionary (Wiley, 1978; Wiley and Mayden, 2000) and the various phylogenetic concepts (Mishler and Theriot, 2000).

Thus, the existence of sibling species represents one of the most important objection to the typological concept (i.e., to "morphospecies") (Mayr, 2000). However, this objection is weakened, for arthropods, in two ways: (1) many sibling species of arthropods, apparently not different in adults, actually show clear differences at the juvenile stages, and (2) electronic optical devices, and also—this is our topic—geometric morphometrics, have shown that the morphology of sibling species may differ at a hidden level, raising the concept of "cryptic species." The word "cryptic" comes from the greek "kryptos," meaning simply "hidden." Thus, two cryptic species are two species apparently identical under conventional optical devices, but apparently only because their differences lie elsewhere, or are too subtle to be perceived by the human eye. Geometric morphometry detects, measures and shows them (Baylac et al., 2003; Becerra and Valdecasas, 2004; Dujardin, 2008; Perrard et al., 2014).

According to this point of view, sibling species are cryptic species. Cryptic species have received other definitions, and even other names such as "semicryptic," "pseudo-cryptic," and "hypercryptic" (Struck et al., 2018b), raising a possible "terminological chaos" (Heethoff, 2018; Struck et al., 2018a; Zuñga Reinoso and Benítez, 2015). Moreover, some authors defended the idea of an evolutionary process hidden in cryptic species. Cryptic species would hide not only diagnostic characters, but also an evolutionary process on its own. This debate in turn hides a reality highlighted by geometric morphometry applied to arthropods: cryptic species obscure morphological divergences mainly because their species characters are not visible to the human eye. It is even possible that cryptic species hide nothing more. Past centuries of taxonomic practice have created a world of morphologically visible diversity. Current morphological characters separating species, such as those described in dichotomous keys, are all visible to the human eye. I guess this is actually no more than the result of a heritage, i.e., a taxonomy based on human eyes, limiting the ability to separate species (Skoracka et al., 2015; Karanovic et al., 2016). But to my knowledge, there is no biological law suggesting that the morphological divergence of species must consist of visible characters.

To explore morphological diversity beyond the limits of human eyes, modern morphometrics has an important role to play. With that objective, I do not focus on size differences, but on shape differences: they have proved useful in various arthropods, and often in situations of cryptic speciation, as for instance in kissing bugs (Villegas et al., 2002; Gumiel et al., 2003; Matias et al., 2001; Dujardin et al., 2009; Villacís et al., 2020), scythridids (Roggero and Passerin d'Entrèves, 2005), parasitoid hymenoptera (Baylac et al., 2003; Villemant et al., 2007; Kitthawee and Dujardin, 2009), syrphids (Francuski et al., 2009), fruit flies (Kitthawee and Dujardin, 2010), mosquitoes (Ruangsittichai et al., 2011; Vidal et al., 2011; Jaramillo-O et al., 2013; Demari-Silva et al., 2017; Chaiphongpachara et al., 2019; Chatpiyaphat et al., 2020), biting midges (Henni et al., 2014), screwworm flies (Lyra et al.,

2009), sandflies (Prudhomme et al., 2012; De la Riva et al., 2001; Prudhomme et al., 2016; Godoy et al., 2018; Mikery et al., 2019), tabanidae (Changbunjong et al., 2021), ants (Seifert et al., 2014; Samung et al., 2022), fireflies (Sumruayphol and Chaiphongpachara, 2019), and trombiculid mites (Sungvornyothin et al., 2019), among others. Thus, modern morphometry has become a fast and low-cost complementary tool, or sometimes an alternative mean, to identify cryptic species which often needed the molecular machinery to be distinguished.

In practice, although close to the traditional skills of entomologists, geometric morphometrics has still not entered the routine because of a few reasons. The main one is perhaps the lack of a relevant bank of reference images (see Section 7.2.3) (Dujardin et al., 2010). Another, more technical one, is that, unlike morphological qualitative traits, the diagnostic landmark-based shape features are difficult to share among biologists (Dujardin et al., 2010). Landmark-based shape is defined relative to the consensus of the specimens under study, as a consequence the resulting shape variables cannot be directly compared with another sample, except if the two samples are combined into a single one to extract shape data (Dujardin et al., 2010). Note that this restriction holds for the (semi)landmark data, not for the pseudo-landmark data (outlines). The computing of normalized elliptic Fourier coefficients (NEF) does not imply any knowledge about the sample from which the data are extracted.

For both the (semi)landmark-based and the pseudo-landmark-based techniques, the problem remains to freely exchange shape data (raw coordinates) because of the measurement error. The mathematical projection of one's own data into the discriminating reference space constructed by another user is problematic, and clearly discouraged for (semi)landmark-based data. Thus, the idea of a reference image bank involves the use of images, not coordinates (Dujardin et al., 2010). This way, any remaining interuser error would be due to the photography stage, which is easier to control (however see [Fox et al., 2020]).

7.1 The Taxonomic Signal of Shape

The many different techniques of exploratory data analyses (EDA) provide an interesting evaluation of the morphometric variables: they explore the internal structure of the data without using their labels, uncovering natural groupings or types (Dubes and Jain, 1980). The most recommended analysis in morphometrics is the covariance-based principal component analysis (PCA), an ordination method generating new variables (the principal components or PCs) that are uncorrelated and ordered by their decreasing contribution to the total variance. The PCs can be used as input to nonhierarchical (like the Kmeans algorithm [Pérez-Ortega et al., 2019]) and/or hierarchical agglomerative clustering (HAC) analyses. These clustering techniques would also detect structuring within the data, and provide different kinds of graphical illustrations (Seifert et al., 2013). These unsupervised classification techniques are not designed to identify "unknown" individuals, but they help to check the taxonomic signal naturally embedded in shape variables.

7.2 The Taxonomic Use of Morphometrics

The methodological approach to the taxonomic use of modern morphometrics is very different from the traditional dichotomous key as used in traditional taxonomy. It makes use of supervised classification techniques. It requires a reference space where relevant discriminatory coefficients can be computed, to which each single unknown individual can be projected. The "unknown" specimen has to

be assigned to its likely category or species, a method which requires an appropriate set of reference data. These data are grouped according to their labels and first projected in a discriminant space.

Using the discriminant model just created, an unknown individual will then be projected, and expected to lie close to its corresponding group. The most common method to create such discriminant space is by far the canonical variate analysis (CVA), also called linear discriminant analysis (Albrecht, 1980). Other methods exist, including the ones using machine learning methods (see Tables 16.2 and 16.3). Before entering the methods more closely, it is important to discuss the validity of the data involved in the identification process: it could be questioned indeed in case of measurement error.

7.2.1 Measurement Error

The measurement error exists at various steps of morphometric analysis: the mounting technique of specimens or of organs, the photographing conditions, and the user's skill to collect landmark coordinates (Arnqvist and Mårtensson, 1998; Dujardin et al., 2010; Robinson and Terhune, 2017). Generally, similar techniques are used to process similar organisms, and digital techniques of modern photography provide adequate resolution for correct recognition of landmarks under different conditions. However, it might be convenient to photograph a sight (chart) to make sure there is no distortion of the image. The image should be placed in the center of the visual field to avoid possible optical defects at the periphery of the lens.

Whatever the quality and reproducibility of landmark digitization, the recommended way to perform morphometric comparisons is to allow a single user to produce the data (see next paragraph). Even when performed by a single user, digitization should be repeated at least once, allowing one to measure his/her precision and to reduce the error by averaging the two digitizations. For size, the precision is estimated by the "repeatability" (R) index as described by (Arnqvist and Mårtensson, 1998), i.e., a Model II one way ANOVA on repeated measures, where "R" is provided by the ratio of the between individual variance and the total variance. For shape, represented by many variables, a Procrustes ANOVA is recommended (Klingenberg and McIntyre, 1998). At this step of the analyses, possible outliers can be easily detected by examining the factor map of the PCA of shape variables.

Table 16.1 Morphometric Identification of Culicidae Based on 13 Landmarks of the Wing

Genera	Ur, Ma	An	Mi	Cu	Ae, Ar, Co
Scores (%)	100	97	96	95	100
N	508(8)	446(6)	348(5)	317(4)	127(3)

The first column indicates that 100% of the genus Uranotaenia *(Ur) and 100% of the genus* Mansonia *(Ma) could be recognized when mixed with the 6 other genera:* Anopheles *(An),* Mimomyia *(Mi),* Culex *(Cu),* Aedes *(Ae),* Armigeres *(Ar), and* Coquillettidia *(Co). The second column indicates that 97% of the genus* Anopheles *could be recognized when mixed with the genera* Mimomyia, Culex, Aedes, *and* Armigeres. *The third column indicates that 96% of the genus* Mimomyia *could be recognized when mixed with the genera* Culex, Aedes, *and* Armigeres. *95% of the* Culex *could be distinguished from the genera* Aedes, Armigeres, *and* Coquillettidia. *The last column indicates that these three genera were perfectly discriminated by their wing geometry.* N, *total number of individuals in each analysis; between brackets, number of genera in the analysis. Mosquito collection by A. Henry and P. Thongsripong (University of Hawaii). Morphological identification of the genera by Dr. R. Rattanarithikul (AFRIMS, Thailand). Digitization of wings by J.-F. Lasnes (University of Montpellier).*

Table 16.2 Correct Species Attribution Scores Based on the Geometry of the Wings

Species	DA (%)	*n/N*	bgPCA (%)	*n/N*
Aedes				
(Stegomyia) aegypti	100	12/12	100	12/12
(Neomelaniconion) lineatopennis	93	14/15	93	14/15
(Aedimorphus) mediolineatus	100	12/12	100	12/12
(Aedimorphus) vexans	75	18/24	83	20/24
Anopheles				
(Anopheles) barbirostris	100	14/14	100	14/14
(Cellia) tesselatus	100	9/9	100	9/9
(Cellia) vagus	97	34/35	94	33/35
Culex				
(Culex) vishnui	50	26/52	27	14/52
(Culex) gelidus	100	15/15	93	14/15
(Culex) quinquefasciatus	83	10/12	92	11/12
(Oculeomyia) bitaeniorynchus	78	18/23	78	18/23
(Oculeomyia) sinensis	66	19/29	79	23/29
(Culiciomyia) nigropunctatus	100	11/11	92	11/11
Overall classification accuracy	81		78	
Adjusted classification accuracy (TAU)	79		75	

Scores based on adjusted validated reclassification (Kovarovic et al., 2011) are shown after two methods: the discriminant analysis (DA) and the between-groups principal components analysis (bgPCA). Thirteen species belonging to three genera, Anopheles *(An.),* Culex *(Cx.), and* Aedes *(Ae.), were analyzed for species identification, namely:* Ae. aegypti, Ae. lineatopennis, Ae. mediolineatus, Ae. vexans, An. barbirostris, An. tessellatus, An. vagus, Cx. bitaeniorhynchus, Cx. (C) gelidus, Cx. nigropunctatus, Cx. quinquefasciatus, Cx.(O) sinensis, *and* Cx. vishnui. n, *number of individuals correctly assigned to the species;* N, *total number of individuals in the species. Mosquito collection by A. Henry and P. Thongsripong (University of Hawaii). Species morphological identification by Dr. R. Rattanarithikul (AFRIMS, Thailand). Digitization of wings by J.-F. Lasnes (University of Montpellier).*

7.2.2 The "Inter-User Effect"

Among the sources of measurement error other than the ones arising from an incorrect photographing step, the user intervention is often the most important one. The error is generally due to small but systematic differences between different users in pointing to the exact localization of the same landmarks. These subtle discrepancies may be seriously amplified by the power of multivariate

Table 16.3 Correct Species Attribution Scores Based on the Geometry of the Wings

Species	MLP (%)	*n/N*	SVM (%)	*n/N*
Aedes				
(Stegomyia) aegypti	100	12/12	83	10/12
(Neomelaniconion) lineatopennis	93	14/15	93	13/15
(Aedimorphus) mediolineatus	100	12/12	100	12/12
(Aedimorphus) vexans	75	18/24	79	19/24
Anopheles				
(Anopheles) barbirostris	100	14/14	86	12/14
(Cellia) tesselatus	100	9/9	78	7/9
(Cellia) vagus	97	34/35	91	32/35
Culex				
(Culex) vishnui	38	20/52	83	43/52
(Culex) gelidus	100	15/15	80	12/15
(Culex) quinquefasciatus	83	10/12	83	10/12
(Oculeomyia) bitaeniorynchus	83	19/23	70	16/23
(Oculeomyia) sinensis	62	18/29	76	22/29
(Culiciomyia) nigropunctatus	100	11/11	100	11/11
Overall classification accuracy	78		83	
Adjusted classification accuracy (TAU)	75		81	

Scores based on adjusted validated reclassification (Kovarovic et al., 2011) are shown after two machine learning methods: the multilayer perceptron with backpropagation algorithm (MLP) and the support vector machine with kernel RGB (SVM). Thirteen species belonging to three genera, Anopheles *(An.),* Culex *(Cx.), and* Aedes *(Ae.), were analyzed for species identification, namely:* Ae. aegypti, Ae. lineatopennis, Ae. mediolineatus, Ae. vexans, An. barbirostris, An. tessellatus, An. vagus, Cx. bitaeniorhynchus, Cx. (C) gelidus, Cx. nigropunctatus, Cx. quinquefasciatus, Cx.(O) sinensis, *and* Cx. vishnui. n, *number of individuals correctly assigned to the species;* N, *total number of individuals in the species. Mosquito collection by A. Henry and P. Thongsripong (University of Hawaii). Species morphological identification by Dr. R. Rattanarithikul (AFRIMS, Thailand). Digitization of wings by J.-F. Lasnes (University of Montpellier).*

analysis like the discriminant analysis (Dujardin et al., 2010). Their impact can be reduced averaging repeated collections of the data (Arnqvist and Mårtensson, 1998). However, such correction might not be satisfactory when comparing very close specimens or groups, and measurement error may become a significant obstacle for different users (Jordaens et al., 2002; Rasmussen et al., 2001). As a consequence, user A should not enter her/his own measurements in a database of coordinates collected by user B, and vice versa.

7.2.3 The Need for a Bank of Reference Images

To circumvent the obstacle of poor interchangeability of morphometric variables, an alternative geometric descriptive system must be developed that separates data collection and analyses. This requires the creation of a reference image bank from which raw data can be extracted (by digitizing these images) and compared to specimens that one seeks to identify (Dujardin et al., 2010; Sonnenschein et al., 2015). The idea was born from a simple HTML page gathering images adapted to morphometric studies (CLIC bank, now on https://xyom-clic.eu/clic-bank/), but which is no longer maintained to this day. The initiative was then fully developed for South American mosquitoes (Virginio et al., 2021) and for bees (Baylac et al., 2008) (http://apiclass.mnhn.fr/). A successful initiative on a wider range of living organisms, Morphobank, should be mentioned here too https://morphobank.org/.

The chances of success of an identification using these reference images would then depend on their relevance, but also on the classification techniques used (Sonnenschein et al., 2015; Kitthawee and Dujardin, 2016). The relevance of the reference images must be underlined because many species are cryptic, their identification by dichotomous key can therefore be compromised and errors can contaminate the image bank: how to guarantee that the reference images have been correctly labeled? The construction of an adequate library is therefore not a simple task, it could require the help of molecular techniques in situations where the qualitative characters would be unreliable.

The need for such a morphometry-oriented image bank still arouses little enthusiasm because it requires the user to digitize the reference images himself (or herself), a tedious task that can be perceived as time-consuming. So far, however, this procedure is a reliable way to circumvent the (Dujardin et al., 2010) user-to-user digitizing error. One would expect, however, that automated digitizing software (Vandaele et al., 2018; Porto and Voje, 2020) could reduce this obstacle, at least in certain groups of arthropods (Palaniswamy et al., 2010; Macleod, 2017; Nawrocka et al., 2018; Geldenhuys et al., 2023).

7.3 Examples

In humans, some metric traits allow highly reliable individual identification (fingerprint, iris pattern, etc.). We can expect similar situation in animals. Using geometric shape comparisons, a single individual can generally be accurately classified using a database of images of the candidate species (Matias et al., 2001; Sonnenschein et al., 2015; Kitthawee and Dujardin, 2016). The efficiency of shape-based species recognition is so good that the question is not how good it is, but why is it so good (Perrard et al., 2014)? For the medically important group of mosquitoes, various publications suggest the use of morphometrics (Motoki et al., 2012; Laurito et al., 2015; Wilke et al., 2016; Lorenz et al., 2017; da Silva de Souza et al., 2020). As an indirect demonstration, I show here some data about morphometric classification of mosquitoes (Tables 16.1 and 16.2). These data were already presented in the very first edition of the book (Dujardin, 2011): they are resubmitted here to similar analyses under the online XYOM software (Dujardin and Dujardin, 2019) (https://xyom.io/).

7.3.1 Eight Genera of Mosquitoes

The taxa (8 mosquito genera, 508 specimens) have been primarily identified at the genus level by a morphologist (see the legend of Table 16.1). In this analysis (Table 16.1), each individual has been projected in the discriminant space of shape occupied by the remaining 507 individuals, grouped

according to their genus. Here, the 507 individuals could be seen as the reference data extracted from a reference bank of images for mosquitoes, except that the landmarks digitization were performed by the same user.

Different genera are expected to show important morphological divergence (as far, there are no cryptic genera in mosquitoes), and for a specialist some genera are recognizable at first sight. In other words, the recognition of mosquito genera on the basis of wing veins does not necessarily require the examination of subtle variations in shape. In such a case, the potential disruptive effect of digitization by different users is unlikely to be significant.

Our results close to perfection confirm that the use of morphometry by an entomologist who is not a specialist in mosquitoes can enable him to carry out a relevant classification. The possibility to perform satisfactory sorting of raw material without being an expert in taxonomy is very attractive, and encouraging results are increasingly published in various groups of arthropods (Dujardin et al., 2003; Kaba et al., 2017; Sumruayphol et al., 2017; Chatpiyaphat et al., 2020; Sauer et al., 2020; Champakaew et al., 2021).

7.3.2 Thirteen Species of Mosquitoes

This second example involves 263 specimens distributed among 13 species (3 genera). It had been presented in the first edition of this book, and is revisited here by allowing a direct comparison between various reclassification methods: the discriminant analysis (DA, or CVA for canonical variate analysis), the between-groups principal component analysis (bgPCA, see next paragraph), and two machine learning methods, the multilayer perceptron (MLP) and the support vector machine (SVM).

Linear, parametric classification techniques. We used two multivariate parametric analyses, the canonical variate analysis (CVA) also called (linear) discriminant analysis (DA) and the between-groups principal components analysis (bgPCA). Whatever the method, the input variables were the eight first Procrustes components (the principal components of residual coordinates after Procrustes superimposition to the consensus). The good practice indeed remains to first perform a principal component analysis (PCA) on the shape variables, then to select a set of principal components (PC) as input for the multivariate analysis (Schachter-Broide et al., 2004; McLeod, 2015).

To avoid possible overfitting, the number of first PCs should be less than the number of specimens in the smallest group. We used the eight first principal components of shape, which was lower than the number of specimens in the smallest group (9 *A. tesselatus* in a total sample where on average there were 20.6 individuals by group, see Table 16.2). This step (the PCA) also makes it possible to detect outliers: we removed 6 outliers from the 269 specimens analyzed in the previous editions. The resulting classification was tested using cross-validation (Manly, 1986) and another statistics taking into account a correct prediction by simple chance: the total adjusted accuracy (TAU), computed according to Klecka (1980):

$$\mathrm{TAU} = \left(N_c - \sum_{i=1}^{G} (N_i * P_i) \right) / \left(N - \sum_{i=1}^{G} (N_i * P_i) \right) \tag{16.1}$$

where N is total sample size; N_c the total number of cases correctly assigned by the classification method; P_i the prior probability of group membership in the i-th of the G groups; N_i the number of cases in the i-th group.

In our example, the total adjusted accuracy of the classification remained very close to that observed (Table 16.2). The numbers of individuals in each group, if they are too disparate, could also introduce a bias in the Mahalanobis distances (Pimentel, 1992), or any other Euclidean distance used to calculate the reclassification, or to construct a phenogram. The example presented here is not optimal because some sample sizes were too different (9,52).

Within genera the species determination of 263 individuals showed heterogeneous results: in some genera like *Aedes* or *Anopheles*, the species recognition was accurate; it was less convincing in the genus *Culex*, especially for the *Culex vishnui* species (Table 16.2). It is because of this local problem that the total accuracy could appear somewhat low: both methods could not separate a consistent proportion of *Cx. (C) vishnui* from *Cx. (C.) gelidus* and from *Cx. (O.) sinensis*.

This local problem can have several causes. It could be due to a bad statistical distribution of the data, not adapted to the parametric multivariate statistics, and it would then be necessary to seek other techniques of reclassification (cf. the paragraph 3). More worrying: the problem could highlight a possible lack of power of the wing venation, at least in some groups of mosquitoes. Finally, the problem could also be explained by a morphological classification error at the beginning of the study, and in this case the morphometric reclassification would have the merit of highlighting an unexpected taxonomic confusion.

Discriminant analysis and between-groups principal components analysis. The discriminant analysis (DA) has been characterized as unsuitable in the context of shape analysis by Mitteröcker and Bookstein (2011). One of the reasons for this position was that the discriminant analysis (CVA) does not preserve the Euclidean or Procrustes distances between specimens. The between-groups principal component analysis (bgPCA) preserves such distances and has been suggested as an alternative (Mitteröcker and Bookstein, 2011). Even if attractive, this proposal did not gain unanimity: as long as the objective was an optimal classification, it was perhaps not necessary to preserve the forms (McLeod, 2015). Later on, and more disturbing, the bgPCA has been shown to create spurious subdivisions: such anomaly could happen when there were fewer groups than variables (which is a relatively common situation, but not the one of the example Table 16.2), and especially when the number of variables was very high (Cardini et al., 2019; Bookstein, 2019). Even if some authors recently suggested special procedures to mitigate this issue (Cardini and Polly, 2020; Thioulouse et al., 2021), the discriminant analysis seems to remain the most suitable tool for an optimal group separation, provided it is validated by subsequent statistical testing of group separation (McLeod, 2015).

The machine learning approach. Perhaps as a valid alternative to the morphometric classification problems, there are the regularly proposed machine learning approaches (Marcondes and Borges, 2000; Dobigny et al., 2002; Platnick et al., 2005; Vañhara et al., 2007; MacLeod et al., 2007; Fedor et al., 2008; Lorenz et al., 2015; Soda et al., 2016; Macleod, 2017; Sumruayphol et al., 2020; Porto and Voje, 2020; Bellin et al., 2021). These supervised classification techniques do not suffer from the significant statistical constraints specific to parametric tests, but, and more than the other techniques, they require a reference database. It is in such a repository that the machine learning operates to then classify the external data as well as possible. Such reference data should be plentiful, far more than has been proposed here to distinguish 13 species. The scores presented in Table 16.3 were obtained according to the principle of validated classification: each of the 263 classified individuals was not part of the reference data. At each identification, therefore, the training data (the references) were 262 in number: the total number of individuals minus the one identified.

I show the results of two relatively common machine learning techniques (Table 16.3), the artificial neural network and the support vector machine (SVM). The neural network was composed of one hidden layer with three neurons, making use of a multi-layer perceptron with backpropagation algorithm (MLP). For each individual classification I used the "mlp" application, available at https://www.npmjs.com/package/mlp. The support vector machine algorithm was set with the kernel RGB. For each individual classification, I used the "node-svm" application available at https://www.npmjs.com/package/node-svm.

Despite the paucity of reference data, machine learning was able to come up with comparable global classification scores (see Table 16.3). The MLP algorithm provided similar performances as the ones obtained after linear transformations: it could probably perform better with more reference data. Interestingly, as in a recent comparison with neural network (Bellin et al., 2021), the SVM approach performed better, and could provide an acceptable recognition of the various *Culex* species, including *C. vishnui*. This could support one of the three hypotheses proposed above (see paragraph 7.3.2) to explain the unsatisfactory *C. vishnui* reclassification: an unsuitable statistical data distribution.

8. Population Structure Studies

Population structure accepts various definitions according to the field under study: ecology, parasitology, bacteriology, etc. The term is used here in relation to the epidemiological problem of vector or pest control. Each species has a known territory, but not necessarily a continuous one. Because of geographic barriers like mountains or rivers, the species territory can be subdivided into apparently separated areas. The epidemiological question is: do the various geographic areas of the species interchange individuals, or are they completely isolated one from another? In the first case, the control program must consider the total geographic distribution at the same time. In the second case, the control program can follow a sequential strategy. However, invisible frontiers do exist that could produce further subdivisions within geographic areas (see Section 8.1.2): does a morphometric approach detect them?

Although mark-release-recapture studies might be a valid option to evaluate the frequency of active migrants among subpopulations (Tapis and Hausermann, 1975; Harrington et al., 2005), it cannot account for passive migration of juvenile (nonflying) stages of the arthropod, so that this frequency is currently evaluated by indirect methods; the measurement of gene flow by genetic techniques is the technique of choice (Slatkin, 1981, 1985).

8.1 Gene Flow and the Flow of Migrants

Gene flow measurement after genetic studies provides indirect information on the level of migration among subpopulations. However, this information is of unequal value depending on its output, either "lack of" or "complete" gene flow. Lack of gene flow is an epidemiologically valid information since in that circumstance (genetic divergence) migrants are highly unlikely. Less reliable information is the case of genetic similarity suggesting "complete" gene flow. Indeed, it is rather reckless to affirm that such a (lack of) genetic structure is a reflection of the current situation. Gene flow could have taken place in the more or less distant past, and not at the time of genetic analysis. Yet it is the contemporary nature of the observations that constitutes the epidemiologically relevant information. How

contemporaneous or recent these supposed exchanges of genes have been is depending on the effective size of the populations under study, and on the evolutionary rate of the genetic marker (McKay and Latta, 2002). An additional inconvenient with genetic markers is that they are relatively costly and they need appropriate infrastructures. As an unfortunate consequence, genetic markers often remain inside research laboratories and have not yet found their way into routine entomology.

8.1.1 Neutral and Metric Characters

Modern morphometrics is tempting as a candidate population marker because it is a fast, low-cost, easily spread tool, it is informative about current or very recent population events (Dujardin et al., 2007). However, as long as morphometric traits have much higher environmental variance than genetic markers, they are not appropriate for gene flow estimation.

How then to interpret geographic variation of metric properties?

After what has been learned from natural and experimental studies on phenotypic plasticity, an entirely environmental origin of phenotypic variation is not likely among natural populations.

Theoretically, the importance of diversifying selection inflating size or shape variation among natural populations can be quantified by comparing on the same material (i) the *Fst* index as derived from neutral molecular markers and (ii) the *Qst* index as computed from metric characters. *Qst* separates quantitative genetic variation in a manner analogous to *Fst* for single gene markers (Spitze, 1993): if the quantitative characters and the molecular characters are neutral, *Qst* and *Fst* should converge to the same value (Hiernaux, 1977; Rogers and Harpending, 1983; Whitlock, 1999). Data comparing molecular *Fst* and quantitative *Qst* are few. They tend to show the following trends: (i) *Qst* is generally higher, or much higher, than *Fst*, and (ii) the value of *Qst* depends on character fitness (McKay and Latta, 2002). Within species, traits experiencing the strongest local selection pressures (diverging, or diversifying selection) are expected to be the most divergent from molecular *Fst* (McKay and Latta, 2002).

The small set of comparisons reported by Dujardin (2008) in medically important insects between *Qst* and *Fst* confirmed the importance of selection modifying the geometric variation among subpopulations. These comparisons allowed an important observation: the lower sensitivity of shape, relative to size, in response to diversifying selection.

8.1.2 Biogeographical Islands

Local elimination of an insect vector of disease is generally held to be feasible only for geographically constrained situations such as islands. The task of population genetics studies as applied to pest or vector control is, in a sense, to find and define on the mainland those biogeographical "islands."

Can modern morphometrics help defining these target areas, or biogeographical "islands" (de Paula et al., 2005)? A few studies on diverse organisms such as sandflies (Dujardin et al., 2003; Mikery et al., 2019; Sumruayphol et al., 2017), mosquitoes (Morales Vargas et al., 2012), kissing bugs (Schachter-Broide et al., 2004; Gómez-Palacio et al., 2012; Gaspe et al., 2013a), tsetse flies (Camara et al., 2006; Bouyer et al., 2007; Solano et al., 2010; Kaba et al., 2012), or soil mites (Baumann, 2021) brought encouraging arguments to use geometric morphometrics as a tool to examine population structure.

To discuss this application, it is important again to insist on which metric property is considered, either geometric shape or size. Here, modern morphometrics means "2D geometric shape": either a set of landmarks configuration, or a curve (open, or close). Moreover, since the populations compared are conspecific ones, and especially if size variation is important, allometry-free shape should be preferred

to just shape. As seen in the previous sections, geometric shape is made of homologous characters, it is a polygenic character and more canalized than size against environmental disturbances. The following two propositions are related to the geometric shape of the wings, but probably also would apply for the shape of other organs:

1. Because of genetic drift and the polygenic nature of shape, isolation in natural conditions will tend to quickly generate shape changes.
2. When frequent exchanges occur between populations, thus when genetic drift is not possible, shape is hardly different even if habitats are different.

Thus, the shape as captured by modern morphometrics, and as a population marker, would be more informative in case of no detectable or nonsignificant change. Indeed, shape similarity would be hardly explained by homogenizing selection (see Qst > Fst, Section 8.1.1). By contrast, shape difference may always be suspected to be depending more or less on some (known or unknown) environmental differences. In my point of view, shape similarity is maintained by genetic exchanges, otherwise, as a polygenic character, the similarity is likely to be quickly broken due to genetic drift (Falconer, 1981).

Between truly isolated populations, differences in shape should develop because of two main reasons: (i) genetic drift is likely to be a major force affecting shape, and (ii) homogenizing selection seems infrequent (or unable to counteract the effects of genetic drift).

Some experimental and natural observations agree with these propositions. For instance, two isolated isofemale laboratory lines of *Ae. aegypti* could diverge in shape (the geometry of wing venation) after less than 15 generations, in spite of a common laboratory environment (Jirakanjanakit et al., 2008).

Recent studies performed on natural populations of insects could show that shape similarity was suggestive of exchanges between compared populations, although they were collected in different habitats. Thus, between populations of houses and of palm trees, the geometry of the wing venation of *Rhodnius prolixus* did not show significant differences (Feliciangeli et al., 2007). Such similarity strongly suggested exchange of individuals, thus gene flow, which was confirmed later by genetic markers (Fitzpatrick et al., 2008).

Lack of isolation was also described by both genetic and metric markers for tsetse flies along the Mouhoun river in Burkina Faso (Bouyer et al., 2007). More recently, tsetse flies were collected in the city of Abidjan (Ivory Coast) from a primary forest relict, from the zoological garden and from the nearby university campus of Abobo Adjame. The three sites are a few miles away from each other. The question was: could the flies from forest reinvade the university campus or the zoological garden after vector control there? The wings of the flies, either males or females, in spite of size differences in one sex, could not show allometry-free shape differences. The likely connection among three sites was confirmed by microsatellite markers applied on the same specimens (Kaba et al., 2012).

On another insect, *Ae. albopictus*, the complete lack of landmark-based shape difference between Japan and USA populations induced Morales et al. (Morales Vargas et al., 2012) to suggest ongoing (passive) exchanges between these two continents.

8.2 The Need for a Heuristic

Considering the cost represented by the molecular machinery in developing countries, the above examples suggest that a faster and less expensive morphometric approach could be helpful, even as an

orientation method only. Thus, geometric shape variation could be our guide to quickly identify at low-cost areas where isolation is possible and where it is unlikely. Because the valid information here is the lack of significant allometry-free differences (and not the significant differences), it is possible to suggest two directives helping interpretation could be the following:

- If allometry-free shape does not show significant differences between conspecific allopatric populations, the most likely explanation is that populations are not isolated ones.
- If it shows significant differences, the isolation hypothesis cannot exclude the possible influence of the environment on shape variation.

The above guidelines are based on the hypothesis that genetic drift is the main force in nature producing fast differences in shape among conspecific populations. They refer to contemporaneous time, not to an undefined evolutionary past. They are easy to falsify, so that they invite for more natural observations related to population structure and shape variation.

8.3 Reinfestation Studies

Reinfestation studies refer to the situation where insects which have been controlled by any method are coming back to the treated area. Are they migrants coming from neighboring untreated foci, or are they descendants of a residual population which escaped the applied control methods.

Many studies could disclose significant shape differences between conspecific individuals according to geography (Henry et al., 2010; Morales Vargas et al., 2012, 2013, 2021; Hernandez and Acosta, 2021), suggesting it would be possible to identify the geographic origin of a few specimens.

The morphometric study of reinfesting populations of insects is a special case of population structure study. It is however a very special case since in this application both genetic drift and environmental influence are combining their force to sustain the main idea: an insect is more similar to its parents than to other insects, and more similar to insects having grown in the same microenvironment.

Various field experiments have been published, on tsetse flies, the vectors of sleeping sickness (Kaba et al., 2012; Mbewe et al., 2018), as well as on kissing bugs, the Chagas disease vectors. In the latter group, the shape of *Triatoma infestans* as extracted from traditional morphometrics (head linear measurements) provided information that could identify the source of reinfesting specimens (Dujardin et al., 1997), and such information has been shown to be in agreement with genetic markers (Dujardin et al., 1999a).

Later, an experimental study was conducted on the geometry of the wing of laboratory populations of the North American *T. protracta*. For the 313 wings tested, correct attribution to one of the seven candidate parental lines was four times higher than expected at random (Dujardin et al., 2007).

Provided that samples were available from the population before insecticide application (or any other kind of treatment), relative similarities could suggest the origin of reinfesting specimens: they would be either the descendants of previously killed bugs, i.e., a residual population, or immigrants from an external focus. A residual population is assumed to be the same generation as or the next generation to the individuals subjected to insecticide spraying. The reinfestation analysis is based on the assumption that an insect is more similar to its parents than to other insects (Dujardin et al., 2007). But the reasons for successful results are not only related to some heritability of shape. In the reinfestation studies, the objective is to distinguish local "inhabitants" from "immigrants." Thus, the

environmental effect on metric traits (the environmental covariance) is a welcome effect. Insects reared in the same microenvironmental conditions (a few houses, a village) would share a significantly larger amount of metric similarity, making a residual population easier to recognize.

Various studies related to this problematic have been informative (Dujardin et al., 1997, 1999a; Gaspe et al., 2013a; Hernández et al., 2013; de Arias et al., 2017). Of course, for morphometric characters to be applicable in reinfestation studies, the level of population structuring is an important condition (Gaspe et al., 2013b; Hernández et al., 2013). They would be less applicable to highly dispersing insects breaking the population structure at each new generation.

8.4 Species Richness

This topic could be a special topic of both taxonomy and epidemiology. The transmission of vector-borne diseases has obvious links with the environment. Studies exploring these links suggested that the reduction in global biodiversity is likely to contribute to vector-borne disease transmission through the "dilution effect." Briefly, it is a reduction of disease risk due to increased species diversity, whatever the underlying mechanism (Chivian and Bernstein, 2004; Keesing et al., 2006). It is therefore highly desirable to quantify and monitor the environment. In this kind of study, geometric morphometrics has two advantages to offer: one is its ability to help the identification of the taxa (see Tables 16.1 and 16.2), the other one is its ability to distinguish cryptic species.

Biodiversity is expressed as the combination of both species richness, the number of species in a specific environment, and species evenness, the proportion of each of them. Different indexes have been suggested to take into account both richness and evenness, from which the most commonly used are the Shannon−Wiener (Shannon and Weaver, 1949) and the Simpson (1949) indexes.

In addition to these estimates of biodiversity, complementary information has been looked for in the morphological disparity of organisms. The morphological disparity has been expressed in two ways, one considering the range of shape variation (the difference between extreme forms), the other one the amount of shape variation (the variance of shape). Modern morphometrics is giving these estimations a powerful quantitative tool for accurate measurements and comparisons (Roy et al., 2001; Neige, 2003).

Does metric disparity increase with the number of species? One could expect indeed a higher richness to be the cause of higher morphometric variation, but no such relationship could be confirmed. Trends in species richness (SR) generally did not match trends in metric disparity (MD). However, one could argue that if selection targets forms rather than species, some relationship is predictable. For a given clade's history, Foote (1993) predicts a high or low ratio MD/SR as depending mainly on the kind of selection during evolution: a selection precluding either intermediate or extreme forms, respectively.

Performing simulations from true data (43 mosquito species from 6 different environments), we observed that the correlation—always positive—between MD and species richness (SR) was not satisfactory when using the full set of landmarks available (Fig. 16.3), but could increase to very high values when a very limited set of LM was used (Dujardin et al., 2012). A two-factor ANOVA (species and ecotopes) suggested that with more landmarks, MD was also affected by the environmental heterogeneity of the sample. The metric disparity, when based on a large capture of shape, is under the

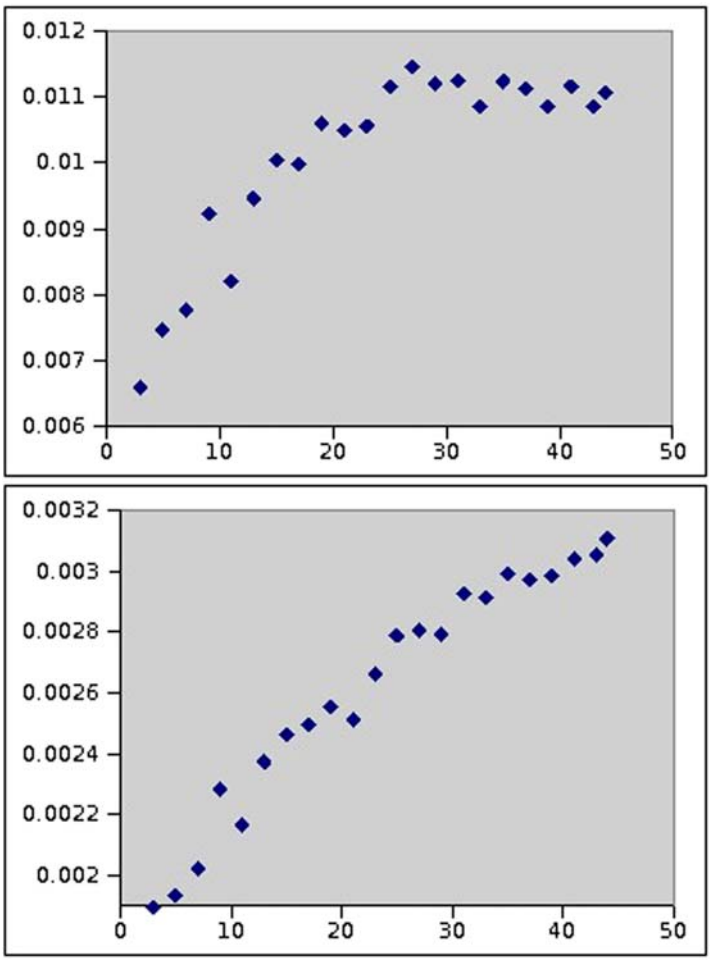


FIGURE 16.3

Two graphs representing the diversity structure (DS), i.e., the relationship between species richness (SR, on the horizontal axis) and metric diversity (MD, on vertical axis). The bottom graph shows the DS appearing when using a set of four landmarks. The top graph shows the DS using the total number (13) of LM. When using the totality of LM (top graph) there is a "first list" of increasing MD values up to an SR of 30 species. This first list is common to the two graphs. In the top one, after 30 species, there is a typical "plateau" breaking the correlation between SR and DM.

Data from Dujardin, J.P., Thongsripong, P., Henry, A.B., 2012. The mosquito fauna: from metric disparity to species diversity. In: Wahl, C. (Ed.), Morphometrics. ISBN: 978-953-51-0172-7. InTech. http://www.intechopen.com/books/morphometrics/the-mosquito-faunafrom-metric-disparity-to-species-diversity

influence of factors other than the mere species number, as for instance the likely influence of the habitat (Dujardin et al., 2012).

The relationship between richness of taxa and the geometric variation of the pronotum or the elytron, based on semilandmarks data, has been recently investigated in 1303 stag beetles across 4 subfamilies, showing no correspondence at either the species or the genus level (Zhang et al., 2019).

Studying shape variation among 1119 jewel beetles, the relationship with richness of taxa was not significant at the level of species, but positive and significant at the genus level (Tong et al., 2020).

The few studies exploring shape variation and species richness suggest that the possible link between shape and diversity is not simple, requiring more effort to understand the role of each factor affecting the relationship.

9. Conclusion

Through the use of coordinates of anatomical points instead of linear distances between them, modern morphometry has converted the abstract quantification of shape into a direct visual representation. In the same process, the geometric approach provided a global estimate of size, distinct from shape and independent of shape in the absence of allometry. Of these two metric properties, only the form has been mainly discussed here.

After a brief introduction to the technical aspects of modern morphometrics and some consideration of the genetic and phenotypic information it can provide, the following issues were discussed: the use of shape to answer taxonomic questions, the identification of species on a single specimen, some classification statistics, the relationship with biodiversity, the analyzes of reinfestation and the detection of "biogeographic islands."

Geometric shape, as opposed to size, appears to be the property of choice to answer these questions. Concrete proposals are made, one is a reference image bank intended for morphometry, the other is a heuristic for interpreting the structure of the population.

Glossary

- Procrustes, whose name means "he who stretches," was a thief in Greek mythology (the myth of Theseus). He preyed on travelers along the road to Athens. He offered his victims hospitality on a magical bed that would fit any guest. As soon as the guest lay down Procrustes went to work upon him, either stretching the guest or cutting off his limbs to make him fit perfectly onto the bed (Grose Educational Media, 1997–98)
- Partial warps, relative warps, etc. A complete glossary of the many technical terms related to GM can be found at http://life.bio.sunysb.edu/morph/

Acknowledgments

Thanks to V. Debat (MNHN, Paris) for his help on the subject of heritability of forms in the previous edition: this part has been left unchanged. S. Dujardin (co-founder of XYOM) helped me with the tricky parts of online programming. I also thank M. Tibayrenc (IRD, France), S. Sumruayphol and R. Morales Vargas (Mahidol University, Bangkok, Thailand) who encouraged this new edition. This study was supported by IRD grants number HC3165-3R165-GABI-ENT2 and HC3165-3R165-NV00-THA1.

References

Adams, D.C., 1999. Methods for shape analysis of landmark data from articulated structures. Evol. Ecol. Res. 1 (8), 959–970.

Adams, D.C., Rohlf, F.J., Slice, D.E., 2004. Geometric morphometrics: ten years of progress following the "revolution". Ital. J. Zool. 71, 5–16.

Albrecht, G.H., 1980. Multivariate analysis of the study of form with special reference to canonical variate analysis. Am. Zool. 20, 679–693.

Albrecht, G.H., Gelvin, B.R., Hartman, S.E., 1993. Ratios as a size adjustment in morphometrics. Am. J. Phys. Anthropol. 91 (4), 441−468.

Ananthakrishnan, T.N., 2005. Perspectives and dimensions of phenotypic plasticity in insects. In: Ananthakrishnan, T.N., Douglas, W. (Eds.), Insect Phenotypic Plasticity. Diversity of Responses. Science Publishers Inc, Enfield, NH, pp. 1−23 pp213.

Anderson, W., 1973. Genetic divergence in body size among experimental populations of *Drosophila pseudoobscura* kept at different temperatures. Evolution 2 (27), 278−284.

de Arias, A., Carbajal de la Fuente, A., Gómez, A., Cecere, M., Rolón, M., Gómez, M., Villalba, C., 2017. Morphometric wings similarity among sylvatic and domestic populations of *Triatoma infestans* (Hemiptera: Reduviidae) from the Gran Chaco Region of Paraguay. Am. J. Trop. Med. Hyg. 481−488. https://doi.org/10.4269/ajtmh.16-1013.

Arnqvist, G., Mårtensson, T., 1998. Measurement error in geometric morphometrics: empirical strategies to assess and reduce its impact on measure of shape. Acta Zool. Acad. Sci. Hung. 44 (1−2), 73−96.

Aytekin, A.M., Alten, B., Caglar, S., Ozbel, Y., Kaynas, S., Simsek, F.M., Kasap, O.E., Belen, A., 2007. Phenotypic variation among local populations of phlebotomine sand flies (Diptera: Psychodidae) in southern Turkey. J. Vector Ecol. 32 (2), 226−234.

Baumann, J., 2021. Patterns of intraspecific morphological variability in soil mites reflect their dispersal ability. Exp. Appl. Acarol. 83, 241−255.

Baylac, M., Frieß, M., 2005. Fourier descriptors, Procrustes superimposition, and data dimensionality: an example of cranial shape analysis in modern human populations. In: Slice (Ed.), Modern Morphometrics in Physical Anthropology, pp. 145−165. Chap. 6.

Baylac, M., Villemant, C., Simbolotti, G., 2003. Combining geometric morphometrics with pattern recognition for the investigation of species complexes. Biol. J. Linn. Soc. 80 (1), 89−98.

Baylac, M., Garnery, L., Tharavy, D., Pedraza-Acosta, J., Rortais, A., 2008. Apiclass, an Automatic Wing Morphometric Expert System for Honeybee Identification. http://apiclass.mnhn.fr.

Becerra, J.M., Valdecasas, A.G., 2004. Landmark superimposition for taxonomic identification. Biol. J. Linn. Soc. 81, 267−274.

Bellin, N., Calzolari, M., Callegari, E., Bonilauri, P., Grisendi, A., Dottori, M., Rossi, V., 2021. Geometric morphometrics and machine learning as tools for the identification of sibling mosquito species of the maculipennis complex (anopheles). Infect. Genet. Evol. 95, 105034. https://doi.org/10.1016/j.meegid.2021.105034.

Bergman, A., Siegal, M.L., 2003. Evolutionary capacitance as a general feature of complex gene networks. Nature 424 (6948), 501−504.

Bitner-Mathé, B.C., Klaczko, L.B., 1999. Size and shape heritability in natural populations of *Drosophila mediopunctata*: temporal and microgeographical variation. Genetica 105, 35−42.

Bookstein, F.L., 1990. Introduction to methods for landmark data. In: Rohlf, F.J., Bookstein, F.L. (Eds.), Proceedings, Michigan Morphometrics Workshop, 1988. The University of Michigan Museum of Zoology, Ann Arbor, MI, pp. 216−225. Special Publication No. 2.

Bookstein, F.L., 1991a. Morphometric Tools for Landmark Data. Geometry and Biology. Cambridge University Press, New York.

Bookstein, F.L., 1991b. Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press, Cambridge, 435pp.

Bookstein, F.L., 1997a. Landmark methods for forms without landmarks: Localizing group differences in outline shape. Med. Image Anal. 1, 225−243.

Bookstein, F.L., 1997b. Shape and the information in medical images: a decade of the morphometric synthesis. Comput. Vis. Image Understand. 66, 97−118.

Bookstein, F.L., 2019. Pathologies of between-groups principal components analysis in geometric morphometrics. Evol. Biol. 46, 271−302. https://doi.org/10.1007/s11692-019-09484-8.

Boussès, P., Dehecq, J.S., Brengues, C., Fontenille, D., 2013. Updated inventory of mosquitoes (Diptera: Culicidae) of the island of La Réunion, Indian Ocean. Bull. Soc. Pathol. Exot. 106, 113−125.

Bouyer, J., Ravel, S., Dujardin, J.P., de Meeus, T., Vial, L., Thevenon, S., Guerrini, L., Sidibe, I., Solano, P., 2007. Population structuring of *Glossina palpalis gambiensis* (Diptera: Glossinidae) according to landscape fragmentation in the Mouhoun river, Burkina Faso. J. Med. Entomol. 44 (5), 788−795.

Breuker, C.J., Patterson, J.S., Klingenberg, C.P., 2006. A single basis for developmental buffering of *Drosophila* wing shape. PLoS One 1 (1), e7.

Burnaby, T.P., 1966. Growth-invariant discriminant functions and generalized distances. Biometrics 22, 96−110.

Camara, M., Caro-Riaño, H., Ravel, S., Dujardin, J.P., Hervouet, J.P., de Meeus, T., Kagbadouno, M.S., Bouyer, J., Solano, P., 2006. Genetic and morphometric evidence for isolation of a tsetse (Diptera: Glossinidae) population (Loos islands, Guinea). J. Med. Entomol. 43 (5), 853−860.

Cardini, A., Polly, P., 2020. Cross-validated between group pca scatterplots: a solution to spurious group separation? Evol. Biol. 47. https://doi.org/10.1007/s11692-020-09494-x.

Cardini, A., O'Higgins, P., Rohlf, F.J., 2019. Seeing groups where there are none: spurious patterns from between-group PCA. Evol. Biol. https://doi.org/10.1101/706101.

Caro-Riaño, H., Jaramillo, N., Dujardin, J.P., 2009. Growth changes in *Rhodnius pallescens* under simulated domestic and sylvatic conditions. Infect. Genet. Evol. 9 (2), 162−168.

Chaiphongpachara, T., Sriwichai, P., Samung, Y., Ruangsittichai, J., Morales Vargas, R., Cui, L., Sattabongkot, J., Dujardin, J., Sumruayphol, S., 2019. Geometric morphometrics approach towards discrimination of three member species of Maculatus group in Thailand. Acta Trop. 192, 66−74. https://doi.org/10.1016/j.actatropica.2019.01.024.

Champakaew, D., Junkum, A., Sontigun, N., Sanit, S., Limsopatham, K., Saeung, A., Somboon, P., Pitasawat, B., 2021. Geometric morphometric wing analysis as a tool to discriminate female mosquitoes from different suburban areas of Chiang Mai province, Thailand. PLoS One 16 (11), e0260333. https://doi.org/10.1371/journal.pone.0260333.

Changbunjong, T., Prakaikowit, N., Maneephan, P., Kaewwiset, T., Weluwanarak, T., Chaiphongpachara, T., Dujardin, J., 2021. Landmark data to distinguish and identify morphologically close *Tabanus* spp. (Diptera: Tabanidae). Insects 12 (11), 974. https://doi.org/10.3390/insects12110974.

Chatpiyaphat, K., Sumruayphol, S., Dujardin, J., Samung, Y., Phayakkaphon, A., Cui, L., Ruangsittichai, J., Sungvornyothin, S., Sattabongkot, J., Sriwichai, P., 2020. Geometric morphometrics to distinguish the cryptic species Anopheles minimus and An. harrisoni in malaria hot spot villages, western Thailand. Med. Vet. Entomol. https://doi.org/10.1111/mve.12493.

Chivian, E., Bernstein, A.S., 2004. Embedded in nature: human health and biodiversity. Environ. Health Perspect. 112, A12−A13.

Chonephetsarath, S., Raksakoon, C., Sumruayphol, S., Dujardin, J.P., Potiwat, R., 2021. The unequal taxonomic signal of mosquito wing cells. Insects 12 (5), 376. https://doi.org/10.3390/insects12050376.

Claude, J., 2008. Morphometrics with R. Springer Science+Business Media, LLC, ISBN 978-0-387-77789-4, 330pp.

Costa, J., Felix, M., 2007. *Triatoma juazeirensis* sp. nov. from Bahia state, northeastern Brazil (Hemiptera: Reduviidae: Triatominae). Mem. Inst. Oswaldo Cruz 102, 87−90.

Costa, J., Monteiro, F., Beard, C.B., 2001. *Triatoma brasiliensis* Neiva, 1911 the most important Chagas disease vector in Brazil - phylogenetic and population analyzes correlated to epidemiologic importance. Am. J. Trop. Med. Hyg. 65, 280.

Costa, J., Peterson, A.T., Dujardin, J.P., 2008. Indirect evidences suggest homoploid hybridization as a possible mode of speciation in Triatominae (Hemiptera, Heteroptera, Reduviidae). Infect. Genet. Evol. 9 (2), 263−270.

Costa, J., Bargues, M.D., Neiva, V.L., Lawrence, G.G., Gumiel, M., Oliveira, G., Cabello, P., Lima, M.M., Dotson, E., Provance, D.W.J., Almeida, C.E., Mateo, L., Mas-Coma, S., Dujardin, J.P., 2016. Phenotypic variability confirmed by nuclear ribosomal DNA suggests a possible natural hybrid zone of *Triatoma brasiliensis* species complex. Infect. Genet. Evol. 37, 77−87.

Crampton, J.S., 1995. Elliptic Fourier shape analysis of fossil bivalves: some practical considerations. Lethaia 28, 179−186.

Daly, H.V., 1992. A statistical and empirical evaluation of some morphometric variables of honey bee classification. In: Foottit, R.G., Sorensen, J.T. (Eds.), Ordination in the Study of Morphology, Evolution and Systematics of Insects: Applications and Quantitative Genetic Rationales. Elsevier, New York, pp. 127−156, 418pp.

David, J.R., Moreteau, B., Gauthier, J.P., Petavy, G., Stockel, A., Imasheva, A.G., 1994. Reaction norms of size characters in relation to growth temperature in *Drosophila melanogaster* : an isofemale lines analysis. Genet. Sel. Evol. 26, 229−251.

Debat, V., David, P., 2001. Mapping phenotypes: canalization, plasticity and developmental stability. Trends Ecol. Evol. 16 (10), 555−561.

Debat, V., Alibert, P., David, P., Paradis, E., Auffray, J.C., 2000. Independence between developmental stability and canalization in the skull of the house mouse. Proc. R. Soc. Lond. B Biol. Sci. 267 (1442), 423−430.

Debat, V., Begin, M., Legout, H., David, J.R., 2003. Allometric and nonallometric components of *Drosophila* wing shape respond differently to developmental temperature. Evolution 57 (12), 2773−2784.

Debat, V., Milton, C.C., Rutherford, S., Klingenberg, C.P., Hoffmann, A.A., 2006. Hsp90 and the quantitative variation of wing shape in *Drosophila melanogaster*. Evolution 60, 2529−2538.

Demari-Silva, B., Multini, L.C., Suesdek, L., Oliveira, T., Sallum, M., Marrelli, M.T., 2017. Wing morphometry and genetic variability between *Culex coronator* and *Culex usquatus* (Diptera: Culicidae), two sibling species of the Coronator group. J. Med. Entomol. 54 (4), 901−908. https://doi.org/10.1093/jme/tjx033.

Dobigny, G., Baylac, M., Denys, C., 2002. Geometric morphometrics, neural networks and diagnosis of sibling *Taterillus* species (Rodentia, Gerbillinae). Biol. J. Linn. Soc. 77 (3), 319−327.

Dobzhansky, T., 1971. Genetics of Evolutionary Process. Colombia University Press, New York, 505pp.

Van Dongen, S.V., 2006. Fluctuating asymmetry and developmental instability in evolutionary biology: past, present and future. J. Evol. Biol. 19 (6), 1727−1743.

Dubes, R., Jain, A., 1980. Clustering methodologies in exploratory data analysis. In: Yovits, M.C. (Ed.), Advances in Computers, vol 19. Elsevier, pp. 113−228. https://doi.org/10.1016/S0065-2458(08)60034-0.

Dujardin, J.P., 2008. Morphometrics applied to medical entomology. Infect. Genet. Evol. 8, 875−890.

Dujardin, J.P., 2011. 16 - Modern morphometrics of medically important insects. In: Tibayrenc, M. (Ed.), Genetics and Evolution of Infectious Disease. Elsevier, London, pp. 473−501. https://doi.org/10.1016/B978-0-12-384890-1.00016-9.

Dujardin, J.P., 2019. A template-dependent semilandmarks treatment and its use in medical entomology. Infect. Genet. Evol. (70), 197−207. https://doi.org/10.1016/j.meegid.2019.03.002.

Dujardin, S., Dujardin, J.P., 2019. Geometric morphometrics in the cloud. Infect. Genet. Evol. (70), 189−196.

Dujardin, J.P., Le Pont, F., 2004. Geographic variation of metric properties within the Neotropical sandflies. Infect. Genet. Evol. 4 (4), 353−359.

Dujardin, J.P., Slice, D.E., 2006. Contributions of morphometrics to medical entomology. In: Tibayrenc, M. (Ed.), Encyclopedia of Infectious Diseases. John Wiley & Sons, Inc., pp. 435−447. ISBN 9780470114209.

Dujardin, J.P., Bermúdez, H., Schofield, C.J., 1997. The use of morphometrics in entomological surveillance of silvatic foci of *Triatoma infestans* in Bolivia. Acta Trop. 66, 145−153.

Dujardin, J.P., Bermúdez, H., Gianella, A., Cardozo, L., Ramos, E., Saravia, R., Quiroz, K., Forgues, G., Carazas, R., Hervas, D., Chavez, T., Machane, M., Martínez, E., Torrez, M., 1999a. Uso de marcadores genéticos en la vigilancia entomológica de la enfermedad de Chagas. In: Alfred Cassab, J., Noireau, F.,

Guillen, G. (Eds.), La Enfermedad de Chagas en Bolivia - Conocimientos científicos al inicio del Programa de Control (1998-2002). Ministerio de Salud y Prevision social, OMS/OPS, IRD and IBBA, La Paz, pp. 157−169, 259pp.

Dujardin, J.P., Panzera, P., Schofield, C.J., 1999b. Triatominae as a model of morphological plasticity under ecological pressure. Mem. Inst. Oswaldo Cruz 94, 223−228.

Dujardin, J.P., Le Pont, F., Baylac, M., 2003. Geographic versus interspecific differentiation of sandflies: a landmark data analysis. Bull. Entomol. Res. 93, 87−90.

Dujardin, J.P., Beard, C.B., Ryckman, R., 2007. The relevance of wing geometry in entomological surveillance of Triatominae, vectors of Chagas disease. Infect. Genet. Evol. 7, 161−167. https://doi.org/10.1016/j.meegid.2006.07.005.

Dujardin, J.P., Costa, J., Bustamante, D., Jaramillo, N., Catalá, S., 2009. Deciphering morphology in Triatominae: the evolutionary signals. Acta Trop. 110, 101−111.

Dujardin, J.P., Kaba, D., Henry, A.B., 2010. The exchangeability of shape. BMC Res. Notes 3, 266. https://doi.org/10.1186/1756−0500−3−266.

Dujardin, J.P., Thongsripong, P., Henry, A.B., 2012. The mosquito fauna: from metric disparity to species diversity. In: Wahl, C. (Ed.), Morphometrics. InTech. ISBN: 978-953-51-0172-7. http://www.intechopen.com/books/morphometrics/the-mosquito-faunafrom-metric-disparity-to-species-diversity.

Dujardin, J.P., Kaba, D., Solano, P., Dupraz, M., McCoy, K.D., Jaramillo-O, N., 2014. Outline-based morphometrics, an overlooked method in arthropod studies? Infect. Genet. Evol. 28.

Dujardin, J.P., Pham Thi, K., Truong Xuan, L., Panzera, F., Pita, S., Schofield, C.J., 2015. Epidemiological status of kissing-bugs in South East Asia: a preliminary assessment. Acta Trop. 151, 142−149.

Dworkin, I., 2005. Canalization, cryptic variation, and developmental buffering: a critical examination and analytical perspective. In: Hallgrimsson, B., Hall, B.K. (Eds.), Variation: A Central Concept in Biology. Academic Press, Oxford, UK, pp. 131−158.

Falconer, D.S., 1981. Introduction to Quantitative Genetics. Longman, London and New York, 300pp.

Fedor, P., Malenovský, I., Vañhara, J., Sierka, W., Havel, J., 2008. Thrips (Thysanoptera) identification using artificial neural network. Bull. Entomol. Res. https://doi.org/10.1017/S0007485308005750.

Feliciangeli, M.D., Sanchez-Martin, M., Marrero, R., Davies, C., Dujardin, J.P., 2007. Morphometric evidence for a possible role of *Rhodnius prolixus* from palm trees in house re-infestation in the State of Barinas (Venezuela). Acta Trop. 101, 169−177.

Félix, M.A., Barkoulas, M., 2015. Pervasive robustness in biological systems. Nat. Rev. Genet. 16 (8), 483−496.

Firmat, C., Gomes-Rodrigues, H., Renaud, S., Claude, J., Hutterer, R., Garcia-Talavera, F., Michaux, J., 2010. Mandible morphology, dental microwear, and diet of the extinct giant rats *Canariomys* (Rodentia: Murinae) of the Canary Islands (Spain). Biol. J. Linn. Soc. 101, 28−40.

Fitzpatrick, S., Feliciangeli, M., Sanchez-Martin, M.J., Monteiro, F.A., Miles, M.A., 2008. Molecular genetics reveal that silvatic *Rhodnius prolixus* do colonise rural houses. PLoS Neglected Trop. Dis. 2 (4), e210. https://doi.org/10.1371/journal.pntd.0000210.

Foote, M., 1993. Discordance and concordance between morphological and taxonomic diversity. Paleobiology 19, 185−204.

Fox, N.S., Veneracion, J.J., Blois, J.L., 2020. Are geometric morphometric analyses replicable? evaluating landmark measurement error and its impact on extant and fossil *Microtus* classification. Ecol. Evol. 10, 3260−3275. https://doi.org/10.1002/ece3.6063.

Francoy, T.M., Faria Franco, F., Roubik, D.W., 2012. Integrated landmark and outline-based morphometric methods efficiently distinguish species of *Euglossa* (Hymenoptera, Apidae, Euglossini). Apidologie 43, 609−617.

Francuski, L., Ludoski, J., Vujić, A., Milankov, V., 2009. Wing geometric morphometric inferences on species delimitation and intraspecific divergent units in the *Merodon ruficornis* group (Diptera, Syrphidae) from the Balkan peninsula. Zool. Sci. 26 (4), 301−308.

Gadagkar, R., Chandrashekara, K., 2005. Behavioral diversity and its apportionment in a primitively eusocial wasp. In: Ananthakrishnan, T.N., Douglas, W. (Eds.), Insect Phenotypic Plasticity. Diversity of Responses. Science Publishers, Inc, Enfield, NH, pp. 213 108−124.

Gaspe, M.S., Gurevitz, J.M., Gürtler, R.E., Dujardin, J.P., 2013a. Origins of house reinfestation with *Triatoma infestans* after insecticide spraying in the Argentine Chaco using wing geometric morphometry. Infect. Genet. Evol. 17, 93−100. https://doi.org/10.1016/j.meegid.2013.03.044.

Gaspe, M.S., Gurevitz, J.M., Gürtler, R.E., Dujardin, J.P., 2013b. Origins of house reinfestation with *Triatoma infestans* after insecticide spraying in the Argentine Chaco using wing geometric morphometry. Infect. Genet. Evol. 93−100.

Geldenhuys, D.S., Josias, S., Brink, W., Makhubele, M., Hui, C., Landi, P., Bingham, J., Hargrove, J., Hazelbag, M.C., 2023. Deep learning approaches to landmark detection in tsetse wing images. PLoS Comput. Biol. 19, 1−23. https://doi.org/10.1371/journal.pcbi.1011194.

Gibson, G., Dworkin, I., 2004. Uncovering cryptic genetic variation. Nat. Rev. Genet. 5, 681−690.

Gilchrist, A.S., Partridge, L., 2001. The contrasting genetic architecture of wing size and shape in *Drosophila melanogaster*. Heredity 86, 144−152.

Godoy, R., Shimabukuro, P., Dos Santos, T., Fac, P., Da Cunha, A., Santos, F., Vilela, M., Rangel, E., Galati, E., 2018. Geometric morphometry of the head in sand flies (Diptera: Psychodidae: Phlebotominae), an alternative approach to taxonomy studies. Zootaxa 4504 (4), 566−576. https://doi.org/10.11646/zootaxa.4504.4.7.

Gómez-Palacio, A., Jaramillo-O, N., Caro-Riaño, H., Diaz, S., Monteiro, F., Pérez, R., Panzera, F., Triana, O., 2012. Morphometric and molecular evidence of intraspecific biogeographical differentiation of *Rhodnius pallescens* (Hemiptera: Reduviidae: Rhodniini) from Colombia and Panama. Infect. Genet. Evol. 12 (8), 1975−1983.

Görür, G., 2005. The importance of phenotypic plasticity in herbivorous insect speciation. In: Anantakrishnan, T.N., Whitman, D. (Eds.), Insect Phenotypic Plasticity. Diversity of Responses. Science Publishers, Inc, Enfield, NH, pp. 145−172.

Graham, J., Freeman, D., Emlen, J., 1993. Developmental stability: a sensitive indicator of populations under 'stress'. In: Landis, W.G., Hugues, J.S., Lewis, M.A. (Eds.), Environmental Toxicology and Risk Assessment, ASTM STP 1179. American Society for Testing and Materials, Philadelphia, pp. 136−158.

Gumiel, M., Catalá, S., Noireau, F., de Arias, A.R., Garcia, A., Dujardin, J.P., 2003. Wing geometry in *Triatoma infestans* (Klug) and *T. melanosoma* Martinez, Olmedo and Carcavallo (Hemiptera: Reduviidae). Syst. Entomol. 28 (2), 173−179.

Gunz, P., Mitteroecker, P., 2012. Semilandmarks: a method for quantifying curves and surfaces. Hystrix. https://doi.org/10.4404/hystrix-24.1-6292.

Gunz, P., Mitteroeker, P., Bookstein, F.L., 2005. Semilandmarks in three dimensions. In: Slice, D.E. (Ed.), Modern Morphometrics in Physical Anthropology. Kluwer Academic/Plenum Publishers, New York.

Harrington, L., Scott, T.W., Lerdthusnee, K., Coleman, R.C., Costero, A., Clark, G.G., Jones, J.J., Kitthawee, S., Kittayapong, P., Sithiprasasna, R., Edman, J.D., 2005. Dispersal of the dengue vector *Aedes aegypti* within and between rural communities. Am. J. Trop. Med. Hyg. 72 (2), 209−220.

Hatadani, L.M., Klaczko, L.B., 2008. Shape and size variation on the wing of *Drosophila mediopunctata*: influence of chromosome inversions and genotype-environment interaction. Genetica 133, 335−342.

Heethoff, M., 2018. Cryptic species - conceptual or terminological chaos? a response to struck et al. Trends Ecol. Evol. 33 (5), 310. https://doi.org/10.1016/j.tree.2018.02.006.

Henni, L.H., Sauvage, F., Ninio, C., Depaquit, J., Augot, D., 2014. Wing geometry as a tool for discrimination of *Obsoletus* group (Diptera: Ceratopogonidae: Culicoides) in France. Infect. Genet. Evol. 21, 110−117. https://doi.org/10.1016/j.meegid.2013.10.008.

Hennig, W., 1965. Phylogenetic systematics. Annu. Rev. Entomol. 10, 97−116.

Henry, A., Thongsripong, P., Fonseca-Gonzalez, I., Jaramillo-Ocampo, N., Dujardin, J.P., 2010. Wing shape of dengue vectors from around the world. Infect. Genet. Evol. https://doi.org/10.1016/j.meegid.2009.12.001.

Hernandez, M., Acosta, L., 2021. Caracterización del dimorfismo sexual y reconocimiento de machos dimórficos en el complejo Discocyrtus prospicuus (Arachnida: Opiliones: Gonyleptidae): una aproximación desde la morfometría geométrica. Revista mexicana de biodiversidad versión On-line ISSN 2007-8706versión impresa ISSN 1870-3453 Rev. Mex. Biodiv 92. https://doi.org/10.22201/ib.20078706e.2021.92.3545.

Hernández, M.L., Dujardin, J.P., Gorla, D.E., Catalá, S.S., 2013. Potential sources of Triatoma infestans reinfesting peridomiciles identified by morphological characterization in los Llanos, La Rioja, Argentina. Mem. Inst. Oswaldo Cruz 91−97.

Hiernaux, J., 1977. Long-term biological effects of human migration from the African savanna to the equatorial forest: a case study of human adaptation to a hot and wet climate. In: Harrison, G.A. (Ed.), Population Structure and Human Variation. Cambridge University Press, Cambridge, pp. 187−218.

Hillesheim, E., Stearns, S.C., 1991. The responses of *Drosophila melanogaster* to artificial selection on body weight and its phenotypic plasticity in two larval food environments. Evolution 45 (8), 1909−1923.

Hoffman, A.A., Shirriffs, J., 2002. Geographic variation for wing shape in *Drosophila serrata*. Evolution 56, 1068−1073.

Hoffmann, A.A., Woods, R.E., Collins, E., Wallin, K., White, A., McKenzie, J.A., 2005. Wing shape versus asymmetry as an indicator of changing environmental conditions in insects. Aust. J. Entomol. 44, 233−243.

Houle, D., 1992. Comparing evolvability and variability of quantitative traits. Genetics 130 (1), 195−204.

Houle, D., 1998. High enthusiasm and low r-squared. Evolution 52 (6), 1872−1876.

Huxley, J.S., 1932. Problems of Relative Growth. Methuen, London.

Iriarte, P.F., Norry, F.M., Hasson, E.R., 2003. Chromosomal inversions effect body size and shape in different breeding resources in *Drosophila buzzatii*. Heredity 91, 51−59.

Jaramillo-O, N., Dujardin, J.P., Calle-Londoño, D., Fonseca-González, I., 2013. Geometric morphometrics for the taxonomy of 11 species of *Anopheles (Nyssorhynchus)* mosquitoes. Med. Vet. Entomol. 29 (1), 26−36.

Jirakanjanakit, N., Leemingsawat, S., Thongrungkiat, S., Apiwathnasorn, C., Singhaniyom, S., Bellec, C., Dujardin, J.P., 2007. Influence of larval density or food variation on the geometry of the wing of *Aedes (Stegomyia) aegypti*. Trop. Med. Int. Health 12 (11), 1354−1360.

Jirakanjanakit, N., Leemingsawat, S., Dujardin, J.P., 2008. The geometry of the wing of *Aedes (Stegomyia) aegypti* in isofemale lines through successive generations. Infect. Genet. Evol. 8, 414−421.

Jordaens, K., Van Dongen, S., Van Riel, P., Geenen, S., Verhagen, R., Backeljau, T., 2002. Multivariate morphometrics of soft body parts in terrestrial slugs: comparison between two datasets, error assessment and taxonomic implications. Biol. J. Linn. Soc. 75 (4), 533−542.

Kaba, D., Ravel, S., Acapovi-Yao, G., Solano, P., Allou, K., Bosson-Vanga, H., Gardes, L., N'Goran, E.K., Schofield, C.J., Koné, M., Dujardin, J., 2012. Phenetic and genetic structure of tsetse fly populations (*Glossina palpalis palpalis*) in southern Ivory Coast. Parasites Vectors 5, 153.

Kaba, D., Berté, D., Ta, B.T.D., Telleria, J., Solano, P., Dujardin, J.P., 2017. The wing venation patterns to identify single tsetse flies. Infect. Genet. Evol. 47, 132−139. https://doi.org/10.1016/j.meegid.2016.10.008.

Karanovic, T., Djurakic, M., Eberhard, S., 2016. Cryptic species or inadequate taxonomy? Implementation of 2D geometric morphometrics based on integumental organs as landmarks for delimitation and description of Copepod taxa. Syst. Biol. 65 (2), 304−327. https://doi.org/10.1093/sysbio/syv088.

Keesing, F., Holt, R.D., Ostfeld, R.S., 2006. Effects of species diversity on disease risk. Ecol. Lett. 9, 485−498.

Kitano, H., 2004. Biological robustness. Nat. Rev. Genet. 5, 826−837.

Kitthawee, S., Dujardin, J.P., 2009. *Diachasmimorpha longicaudata*: reproductive isolation and geometric morphometrics of the wings. Biol. Control 51 (1), 191−197.

Kitthawee, S., Dujardin, J.P., 2010. The geometric approach to explore the *Bactrocera tau* complex (Diptera: Tephritidae) in Thailand. Zoology 113 (4), 243−249.

Kitthawee, S., Dujardin, J., 2016. The *Diachasmimorpha longicaudata* complex in Thailand discriminated by its wing venation. Zoomorphology 135. https://doi.org/10.1007/s00435-016-0307-x.

Klecka, W.R., 1980. Discriminant Analysis. Sage University Paper Series on Quantitative Applications in the Social Sciences, No 07-019. Sage, Beverly Hills, CA.

Klingenberg, C., 1996. Multivariate allometry. NATO ASI Ser. Life Sci. 284. https://doi.org/10.1007/978-1-4757-9083-2_3.

Klingenberg, C., 2003. Quantitative genetics of geometric shape: heritability and the pitfalls of the univariate approach. Evolution 57 (1), 191−195.

Klingenberg, C.P., Leamy, L.J., 2001. Quantitative genetics of geometric shape in the mouse mandible. Evolution 55, 2342−2352.

Klingenberg, C.P., McIntyre, G., 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. Evolution 52 (5), 1363−1375.

Klingenberg, C., Monteiro, L., 2005. Distances and directions in multidimensional shape spaces: implications for morphometric applications. Syst. Biol. 54 (4), 678−688.

Klingenberg, C.P., Leamy, L.J., Routman, E.J., Cheverud, J.M., 2001. Genetic architecture of mandible shape in mice: effects of quantitative trait loci analyzed by geometric morphometrics. Genetics 157 (2), 785−802.

Klingenberg, C.P., Leamy, L.J., Cheverud, J.M., 2004. Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible. Genetics 166, 1909−1921.

Kovarovic, K., Aiello, L.C., Cardini, A., Lockwood, C.A., 2011. Discriminant function analyses in archaeology: are classification rates too good to be true? J. Archaeol. Sci. 38 (11), 3006−3018.

Kuhl, F.P., Giardina, C.R., 1982. Elliptic fourier features of a closed contour. Comput. Graph. Image Process. 18, 236−258.

Lande, R., 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution 402−416.

Laurito, M., Almirón, W.A., Ludueña Almeida, F.F., 2015. Discrimination of four *Culex (Culex*) species from the Neotropics based on geometric morphometrics. Zoomorphology 134, 447−455.

Leamy, L.J., Klingenberg, C.P., 2005. The genetics and evolution of fluctuating asymmetry. Annu. Rev. Ecol. Evol. Syst. 1−21.

Lehmann, T., Dalton, R., Kim, E., Dahl, E., Diabate, A., Dabire, R., Dujardin, J., 2006. Genetic contribution to variation in larval development time, adult size, and longevity of starved adults of *Anopheles gambiae*. Infect. Genet. Evol. 6, 410−416.

Lele, S.R., Richtsmeyer, J., 2001. An Invariant Approach to Statistical Analysis of Shape. Chapman and Hall/CRC, Boca Raton, pp. VIII−308.

Lestrel, P.E., 1989. Methods for analyzing complex two-dimensional forms: elliptic Fourier functions. Am. J. Hum. Biol. 1, 149−164.

Lestrel, P.E., 1997. Introduction and overview of Fourier descriptors. In: Chapter 2: In: Fourier Descriptors and Their Applications in Biology. Cambridge University Press, Cambridge, pp. 22−44.

Lestrel, P.E., 2000. Morphometrics for Life Sciences. World Scientific Publishing.

Levit, G.S., Hossfeld, U., Olssonzã, L., 2006. From the "modern synthesis" to cybernetics: Ivan Ivanovich Schmalhausen (1884-1963) and his research program for a synthesis of evolutionary and developmental biology. J. Exp. Zool. 306B, 89−106.

Lohmann, G.P., 1983. Eigenshape analysis of microfossils: a general morphometric procedure for describing changes in shape. Math. Geol. 15, 659−672.

Long, A., Mullaney, S., Reid, L., Fry, J., Langley, C., Mackay, T.F.C., 1995. High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. Genetics 139.

Lorenz, C., Ferraudo, A.S., Suesdek, L., 2015. Artificial neural network applied as a methodology of mosquito species identification. Acta Trop. 152, 165−169.

Lorenz, C., Almeida, F., Almeida-Lopes, F., Louise, C., Pereira, S.N., Petersen, V., Vidal, P.O., Virginio, F., Suesdek, L., 2017. Geometric morphometrics in mosquitoes: what has been measured? Infect. Genet. Evol. 54, 205–215.

Lyra, M.L., Hatadani, L.M., de Azeredo-Espin, A.M., Klaczko, L.B., 2009. Wing morphometry as a tool for correct identification of primary and secondary new world screwworm fly. Bull. Entomol. Res. 23, 1–8.

Macleod, N., 2017. On the use of machine learning in morphometric analysis. In: Biological Shape Analysis: Proceedings of the 4th International Symposium. World Scientific, pp. 134–171.

MacLeod, N., O'Neill, M., Walsh, S.A., 2007. A comparison between morphometric and artificial neural network approaches to the automated species recognition problems in systematics. Chapter V. In: Curry, G.B., Humphries, C.J. (Eds.), Biodiversity Databases: Techniques, Politics, and Applications (Systematics Association Special Volume) -US, pp. 37–62. ISBN:9780415332903.

Manly, B.F.J., 1986. Multivariate Statistical Methods. A Primer. Chapman and Hall, London, p. 154.

Marcondes, C.B., Borges, P.S.S., 2000. Distinction of males of the *Lutzomyia intermedia* (Lutz & Neiva, 1912) species complex by ratios between dimensions and by an artificial neural network (Diptera: Psychodidae, Phlebotominae). Mem. Inst. Oswaldo Cruz 95 (5), 685–688.

Masel, J., 2013. Q&A: evolutionary capacitance. BMC Biol. 11, 103.

Matias, A., De la Riva, J.X., Torrez, M., Dujardin, J.P., 2001. *Rhodnius robustus* in Bolivia identified by its wings. Mem. Inst. Oswaldo Cruz 96 (7), 947–950.

Mayr, E., 1969. Principles of Systematic Zoology. McGraw-Hill, New York.

Mayr, E., 2000. The biological species concept. In: Wheeler, Q.D., Meier, R. (Eds.), Species Concepts and Phylogenetic Theory. A Debate. Columbia University Press, New York, pp. 17–29.

Mbewe, N.J., Saini, R.K., Torto, B., Irungu, J., Yusuf, A.A., Pirk, C., 2018. Effects of vector control on the population structure of tsetse (*Glossina fuscipes fuscipes*) in western Kenya. Acta Trop. 179, 1–9. https://doi.org/10.1016/j.actatropica.2017.12.015.

McKay, J.K., Latta, R.G., 2002. Adaptive population divergence: markers, QTL and traits. Trends Ecol. Evol. 17.

McLeod, N., 2015. The direct analysis of digital images (eigenimage) with a comment on the use of discriminant analysis in morphometrics. In: Lestrel, P.E. (Ed.), Proceedings of the Third International Symposium on Biological Shape Analysis. World Scientific, Singapore, pp. 156–182.

Mezey, J., Houle, D., 2005. The dimensionality of genetic variation for wing shape in *Drosophila melanogaster.* Evolution 1027–1038.

Mikery, O., Rebollar-Téllez, E., Cruz-López, L., Marina, C., Castillo, A., 2019. Traditional and geometric morphometry analyses of *Lutzomyia cruciata (Diptera: Psychodidae: Phlebotominae)* populations of Chiapas, Mexico. J. Med. Entomol. 56, 697–707. https://doi.org/10.1093/jme/tjy227.

Mishler, B.D., Theriot, E.C., 2000. The phylogenetic species concept (*sensu* mishler and theriot): monophyly, apomorphy, and phylogenetic species concept. In: Wheeler, Q.D., Meier, R. (Eds.), Species Concepts and Phylogenetic Theory. A Debate. Columbia University Press, New York, pp. 44–54.

Mitteröcker, P., Bookstein, F., 2011. Linear discrimination, ordination, and the visualization of selection gradients in modern morphometrics. Evol. Biol. 38 (1), 100–114.

Monteiro, L., Diniz-Filho, J.A.F., Dos Reis, S.F., Araújo, E.D., 2002. Geometric estimates of heritability in biological shape. Evolution 56 (3), 563–572.

Morales Vargas, E.R., Ya-umphan, P., Phumala-Morales, N., Komalamisra, N., Dujardin, J.P., 2010. Climate associated size and shape changes in *Aedes aegypti* (Diptera: Culicidae) populations from Thailand. Infect. Genet. Evol. 10 (4), 580–585.

Morales Vargas, R.E., Phumala-Morales, N., Tsunoda, T., Apiwathnasorn, C., Dujardin, J.P., 2012. The phenetic structure of *Aedes albopictus*. Infect. Genet. Evol. 13, 242–251. https://doi.org/10.1016/j.meegid.2012.08.008.

Morales Vargas, R.E., Phumala-Morales, N., Tsunoda, T., Apiwathnasorn, C., Jean-Pierre, D., 2013. The phenetic structure of *Aedes albopictus*. Infect. Genet. Evol. 13, 242−251.

Morales Vargas, R., Tsunoda, T., Noda, J., Bousses, P., Nguyen, T., Hasebe, F., Dujardin, J., 2021. Shape relatedness between geographic populations of *Culex tritaeniorhynchus*, the primary vector of Japanese encephalitis virus: a landmark study. Infect. Genet. Evol. https://doi.org/10.1016/j.meegid.2021.104764.

Motoki, M.T., Suesdek, L., Bergo, E.S., Sallum, M.A., 2012. Wing geometry of *Anopheles darlingi* Root (Diptera: Culicidae) in five major Brazilian ecoregions. Infect. Genet. Evol. 12 (6), 1246−1252.

Nasci, R.S., 1990. Relationship of wing length to adult dry weight in several mosquito species (Diptera: Culicidae). J. Med. Entomol. 27, 716−719.

Nattero, J., Dujardin, J.P., del Pilar, F.M., Gürtler, R.E., 2015. Host-feeding sources and habitats jointly affect wing developmental stability depending on sex in the major Chagas disease vector *Triatoma infestans*. Infect. Genet. Evol. 36, 539−546.

Nawrocka, A., Kandemir, I., Fuchs, S., Tofilski, A., 2018. Computer software for identification of honey bee subspecies and evolutionary lineages. Apidologie 49, 172−184. https://doi.org/10.1007/s13592-017-0538-y.

Neige, P., 2003. Spatial patterns of disparity and diversity of the recent cuttlefishes (cephalopoda) across the Old World. J. Biogeogr. 30, 1125−1137.

Orengo, D.J., Prevosti, A., 2002. Relationship between chromosomal polymorphism and wing size in a natural population of *Drosophila subobscura*. Genetica 115, 311−318.

Palaniswamy, S., Thacker, N.A., Klingenberg, C.P., 2010. Automatic identification of morphometric landmarks in digital images. IET Comput. Vis. 4 (4), 247−260.

Palmer, A.R., Strobeck, C., 1986. Fluctuating asymmetry: measurement, analysis, patterns. Annu. Rev. Ecol. Systemat. 17, 391−421.

Partridge, L., Barrie, B., Fowler, K., French, V., 1994. Evolution and development of body size and cell size in *Drosophila melanogaster* in response to temperature. Evolution 48, 1269−1276.

Patterson, J., Klingenberg, C., 2007. Developmental buffering: how many genes? Evol. Dev. 9 (6), 525−526.

de Paula, A.S., Diotaiuti, L., Schofield, C.J., 2005. Testing the sister-group relationship of the Rhodniini and Triatomini (Insecta: Hemiptera: Reduviidae: Triatominae). Mol. Phylogenet. Evol. 35, 712−718.

Pennisi, E., 2006. Evolution: hidden genetic variation yields caterpillar of a different color. Science 591a.

Perez, S.Y., Bernal, V., Gonzalez, P.N., 2006. Differences between sliding semi-landmark methods in geometric morphometrics, with an application to human craniofacial and dental variation. J. Anat. 208, 769−784.

Pérez-Ortega, J., Almanza-Ortega, N.N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C., Martínez-Rebollar, A., 2019. The k-means algorithm evolution. In: Sud, K., Erdogmus, P., Kadry, S. (Eds.), Introduction to Data Science and Machine Learning. IntechOpen, Rijeka. https://doi.org/10.5772/intechopen.85447 (chapter 5).

Perrard, A., Baylac, M., Carpenter, J.M., Villemant, C., 2014. Evolution of wing shape in hornets: why is the wing venation efficient for species identification? J. Evol. Biol. 27, 2665−2675.

Pimentel, R.A., 1992. An introduction to ordination, principal components analysis and discriminant analysis. p. 11-28. In: Foottit, R.G., Sorensen, J.T. (Eds.), Ordination in the Study of Morphology, Evolution and Systematics of Insects: Applications and Quantitative Genetic Rationales. Elsevier, New York, p. 418.

Platnick, N., Russel, K.N., Do, M.T., 2005. Spida-Web. Species Identified Automatically. A Neural Network Based Automated Identification System for Biological Species. https://research.amnh.org/invertzoo/spida/common/index.htm.

Porto, A., Voje, K.L., 2020. Ml-morph: a fast, accurate and general approach for automated detection and landmarking of biological structures in images. Methods Ecol. Evol. 11, 500−512.

Prudhomme, J., Gunay, F., Rahola, N., Ouanaimi, F., Guernaoui, S., Boumezzough, A., Bañuls, A.L., Sereno, D., Bulent, A., 2012. Wing size and shape variation of *Phlebotomus papatasi* (Diptera: Psychodidae) populations from the south and north slopes of the Atlas Mountains in Morocco. J. Vector Ecol. 37 (1), 137−147.

Prudhomme, J., Cassan, C., Hide, M., Toty, C., Rahola, N., Vergnes, B., Dujardin, J., Alten, B., Sereno, D., Bañuls, A.L., 2016. Ecology and morphological variations in wings of *Phlebotomus ariasi (Diptera: Psychodidae)* in the region of Roquedur (Gard, France): a geometric morphometrics approach. Parasit. Vectors 9, 578. https://doi.org/10.1186/s13071016-1872-z.

Rasmussen, P., Wheeler, W., Moser, T., Vine, L., Sullivan, B., Rusch, D., 2001. Measurements of Canada goose morphology - sources of error and effects on classification of subspecies. J. Wildl. Manag. 65 (4), 716−725.

Réale, D., Roff D, A., 2003. Inbreeding, developmental stability, and canalization in the sand cricket *Gryllus firmus*. Evolution 57 (3), 597−605.

Renaud, S., Michaux, J., Mein, P., Aguilar, J.P., Aufray, C., 1999. Patterns of size and shape differentiation during the evolutionary radiation of the european miocene murine rodents. Lethaia 32, 61−71.

De la Riva, J., Le Pont, F., Ali, V., Matias, R., Mollinedo, S., Dujardin, J.P., 2001. Wing geometry as a tool for studying the *Lutzomyia longipalpis* (Diptera: Psychodidae) complex. Mem. Inst. Oswaldo Cruz. 96 (8), 1089−1094.

Robinson, C., Terhune, C.E., 2017. Error in geometric morphometric data collection: combining data from multiple sources. Am. J. Phys. Anthropol. 164, 62−75. https://doi.org/10.1002/ajpa.23257.

Roff, D.A., Mousseau, T.A., 1987. Quantitative genetics and fitness: lessons from *Drosophila*. Heredity 58, 103−118.

Rogers, A., Harpending, H.C., 1983. Population structure and quantitative characters. Genetics 105, 985−1002.

Roggero, A., Passerin d'Entrèves, P., 2005. Geometric morphometric analysis of wings variation between two populations of the *Scythris obscurella* species-group: geographic or interspecific differences ? (Lepidotera: Scythrididae). SHILAP Rev. Lepidopterol. 33 (130), 101−112.

Rohlf, F.J., 1986. Relationships among eigenshape analysis, Fourier analysis and analysis of coordinates. Math. Geol. 18 (8), 845−854.

Rohlf, F.J., 1990. Rotational fit (Procrustes) methods. In: Rohlf, F., Bookstein, F. (Eds.), Proceedings of the Michigan Morphometrics Workshop. Special Publication Number 2. The University of Michigan Museum of Zoology, Ann Arbor, MI, pp. 227−236 pp380.

Rohlf, F.J., Archie, J.W., 1984. A comparison of Fourier methods for the description of wing shape in mosquitoes (Diptera:Culicidae). Syst. Zool. 33, 302−317.

Rohlf, F.J., Bookstein, F.L., 2003. Computing the uniform component of shape variation. Syst. Biol. 52 (1), 66−69.

Rohlf, F.J., Marcus, L.F., 1993. A revolution in morphometrics. Trends Ecol. Evol. 8 (4), 129−132.

Rohlf, F.J., Slice, D.E., 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39, 40−59.

Roy, K., Balch, D.P., Hellberg, M.E., 2001. Spatial patterns of morphological diversity across the Indo-Pacific: analyses using strombid gastropods. Proc. R. Soc. Lond. 268, 2503−2508.

Ruangsittichai, J., Apiwathnasorn, C., Dujardin, J.P., 2011. Interspecific and sexual shape variation in the filariasis vectors *Mansonia dives* and *Ma. bonneae*. Infect. Genet. Evol. 11 (8), 2089−2094.

Rushton, J., Ankney, C., 2009. Whole brain size and general mental ability: a review. Int. J. Neurosci. 119, 691−731. https://doi.org/10.1080/00207450802325843.

Rutherford, S.L., Lindquist, S., 1998. Hsp90 as a capacitor for morphological evolution. Nature 396, 336−342.

Samung, Y., Chaiphongpachara, T., Ruangsittichai, J., Sriwichai, P., Phayakkaphon, A., Jaitrong, W., Dujardin, J.P., Sumruayphol, S., 2022. Species discrimination of three *Odontomachus* (Formicidae: Ponerinae) species in Thailand using outline morphometrics. Insects 13. https://doi.org/10.3390/insects13030287.

Sauer, F., Jaworski, L., Erdbeer, L., Heitmann, A., Schmidt-Chanasit, J., Kiel, E., Lühken, R., 2020. Geometric morphometric wing analysis represents a robust tool to identify female mosquitoes (Diptera: Culicidae) in Germany. Sci. Rep. 10, 17613. https://doi.org/10.1038/s41598-020-72873-z.

Schachter-Broide, J., Dujardin, J.P., Kitron, U., Gürtler, R.E., 2004. Spatial structuring of *Triatoma infestans* (Hemiptera, Reduviidae) populations from northwestern Argentina using wing geometric morphometry. J. Med. Entomol. 41 (4), 643−649.

Schlichting, C.D., Pigliucci, M., 1998. Phenotypic Evolution: A Reaction Norm Perpective. Sinauer Associates, Inc., Sunderland, MA.

Seifert, B., Ritz, M., Csõsz, S., 2013. Application of exploratory data analyses opens a new perspective in morphology-based alpha-taxonomy of eusocial organisms. Myrmecol. News 19, 1−15.

Seifert, B., Yazdi, A.B., Schultz, R., 2014. *Myrmica martini* sp. n.−a cryptic species of the *Myrmica scabrinodis* species complex (Hymenoptera: Formicidae) revealed by geometric morphometrics and nest-centroid clustering. Myrmecol. News 19, 171−183.

Shannon, C.E., Weaver, W., 1949. The Mathematical Theory of Communication. University of Illinois Press, Urbana.

Shrimpton, A., Robertson, A., 1988. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*: I. Allocation of third chromosome bristle effects to chromosome sections. Genetics 118, 437−443.

Siegel, J.P., Novak, R.J., Lampman, R.L., Steinly, B.A., 1992. Statistical appraisal of the weight-wing length relationship of mosquitoes. J. Med. Entomol. 29 (4), 711−714.

da Silva de Souza, A., Multini, L., Marrelli, M., Wilke, A., 2020. Wing geometric morphometrics for identification of mosquito species (diptera: Culicidae) of neglected epidemiological importance. Acta Trop. 211, 105593. https://doi.org/10.1016/j.actatropica.2020.105593.

Simpson, E.H., 1949. Measurement of diversity. Nature 163, 688.

Skoracka, A., Magalhães, S., Rector, B., Kuczyński, L., 2015. Cryptic speciation in the Acari: a function of species lifestyles or our ability to separate species? Exp. Appl. Acarol. 67, 165−182. https://doi.org/10.1007/s10493-015-9954-8.

Slatkin, M., 1981. Estimating levels of gene flow in natural populations. Genetics 99, 323−335.

Slatkin, M., 1985. Rare alleles as indicators of gene flow. Evolution 39 (1), 53−65.

Smith, G.R., 1990. Homology in morphometrics and phylogenetics. In: Rohlf, F.J., Bookstein, F.L. (Eds.), Proceedings of the Michigan Morphometrics Workshop, Special Publiation Number 2. The University of Michigan Museum of Zoology, Ann Arbor MI, pp. 325−338 pp380.

Soda, K., Slice, D.E., Naylor, G.J.P., 2016. Artificial neural networks and geometric morphometric methods as a means for classification: a case-study using teeth from *Carcharhinus* sp. (carcharhinidae). J. Morphol. 278 (1). https://doi.org/10.1002/jmor.20626.

Solano, P., Kaba, D., Ravel, S., Dyer, N.A., Sall, B., Vreysen, M.J.B., Seck, M.T., Darbyshir, H., Gardes, L., Donnelly, M.J., De Meeus, T., Bouyer, J., 2010. Population genetics as a tool to select tsetse control strategies: Suppression or eradication of *Glossina palpalis gambiensis* in the niayes of Senegal. PLOS Negl. Trop. Dis. 4, e692.

Sonnenschein, A., VanderZee, D., Pitchers, W.R., Chari, S., Dworkin, I., 2015. An image database of *Drosophila melanogaster* wings for phenomic and biometric analysis. GigaScience 4, 25. https://doi.org/10.1186/s13742-015-0065-6.

Spitze, K., 1993. Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. Genetics 135, 367−374.

Struck, T., Feder, J., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V., Kistenich, S., Larsson, K., Liow, L., Nowak, M., Stedje, B., Bachmann, L., Dimitrov, D., 2018a. Cryptic species - more than terminological chaos: a reply to heethoff. Trends Ecol. Evol. 33 (5), 310−312. https://doi.org/10.1016/j.tree.2018.02.008.

Struck, T., Feder, J., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V., Kistenich, S., Larsson, K., Liow, L.H., Nowak, M., Stedje, B., Bachmann, L., Dimitrov, D., 2018b. Finding evolutionary processes hidden in cryptic species. Trends Ecol. Evol. 33 (3), 153−163. https://doi.org/10.1016/j.tree.2017.11.007.

Sumruayphol, S., Chaiphongpachara, T., 2019. Geometric morphometrics as a tool for three species identification of the firefly (Coleoptera: Lampyridae) in Thailand. Biodiversitas 20 (8), 2388−2395. https://doi.org/10.13057/biodiv/d200837.

Sumruayphol, S., Chittsamart, B., Polseela, R., Sriwichai, P., Samung, Y., Apiwathnasorn, C., Dujardin, J., 2017. Wing geometry of *Phlebotomus stantoni* and *Sergentomyia hodgsoni* from different geographical locations in Thailand. C. R. Biol. 340 (1), 37−46. https://doi.org/10.1016/j.crvi.2016.10.002. Epub 2016 Nov 24. Erratum in: C. R. Biol. March 2017;340(3):195.

Sumruayphol, S., Siribat, P., Dujardin, J., Dujardin, S., Komalamisra, C., Thaenkham, U., 2020. *Fasciola gigantica, F. hepatica* and *Fasciola* intermediate forms: geometric morphometrics and an artificial neural network to help morphological identification. PeerJ 8, e8597. https://doi.org/10.7717/peerj.8597.

Sungvornyothin, S., Kumlert, R., Paris, D., Prasartvit, A., Sonthayanon, P., Apiwathnasorn, C., Morand, S., Stekolnikov, A.A., Sumruayphol, S., 2019. Geometric morphometrics of the scutum for differentiation of trombiculid mites within the genus *Walchia* (Acariformes: Prostigmata: Trombiculidae), a probable vector of scrub typhus. Ticks Tick Borne Dis. 10 (2), 495−503. https://doi.org/10.1016/j.ttbdis.2018.11.013.

Suzuki, Y., Nijhout, H.F., 2006. Evolution of a polyphenism by genetic accommodation. Science 311 (5761), 650−652.

Tapis, M., Hausermann, W., 1975. Demonstration of differential domesticity of *Aedes aegypti* (L.) (Diptera: Culicidae) in Africa by mark-releaserecapture. Bull. Entomol. Res. 65, 199−208.

Thioulouse, J., Renaud, S., Dufour, A., Dray, S., 2021. Overcoming the spurious groups problem in between-group pca. Evol. Biol. 48, 458−471. https://doi.org/10.1007/s11692-021-09550-0.

Thompson, J.D., 1971. Phenotypic plasticity as a component of evolutionary change. Trends Ecol. Evol. 6, 246−249.

Tong, Y., Yang, H., Yang, X., Bai, M., 2020. The relationship between category richness and morphological diversity in jewel beetles. Preprints 2020, 2020090378. https://doi.org/10.20944/preprints202009.0378.v1.

Usinger, R.L., Wygodzinsky, P., Ryckman, R.E., 1966. The biosystematics of Triatominae. Annu. Rev. Entomol. 11, 309−330.

Vandaele, R., Aceto, J., Muller, M., Péronnet, F., Debat, V., Wang, C.W., Huang, C.T., Jodogne, S., Martinive, P., Geurts, P., Marée, R., 2018. Landmark detection in 2d bioimages for geometric morphometrics: a multi-resolution tree-based approach. Sci. Rep. 8, 538. https://doi.org/10.1038/s41598-017-18993-5.

Vañhara, J., Murárikoválá, N., Malenovský, I., Havel, J., 2007. Artificial neural networks for fly identification: a case study from the genera *Tachina* and *Ectophasia* (Diptera, Tachinidae). Biologia 62, 462−469.

Vidal, P.O., Peruzin, M.C., Suesdek, L., 2011. Wing diagnostic characters for *Culex quinquefasciatus* and *Culex nigripalpus* (Diptera, Culicidae). Rev. Bras. Med.Trop. 55 (1), 134−137.

Villacís, A., Dujardin, J., Panzera, F., Yumiseva, C., Pita, S., SantillánGuayasamín, S., Orozco, M., Mosquera, K., Grijalva, M., 2020. Chagas vectors *Panstrongylus chinai* (Del Ponte, 1929) and *Panstrongylus howardi* (Neiva, 1911): chromatic forms or true species? Parasites Vectors 13 (1), 226. https://doi.org/10.1186/s13071-020-04097-z.

Villegas, J., Feliciangeli, M.D., Dujardin, J.P., 2002. Wing shape divergence between *Rhodnius prolixus* from Cojedes (Venezuela) and *R. robustus* from Mérida (Venezuela). Infect. Genet. Evol. 2, 121−128.

Villemant, C., Simbolotti, G., Kenis, M., 2007. Discrimination of *Eubazus* (Hymenoptera, Braconidae) sibling species using geometric morphometrics analysis of wing venation. Syst. Entomol. 32 (4), 625−634.

Virginio, F., Domingues, V., da Silva, L.C.G., Andrade, L., Braghetto, K.R., Suesdek, L., 2021. Wingbank: a wing image database of mosquitoes. Front. Ecol. Evol. 9. https://doi.org/10.3389/fevo.2021.660941.

Waddington, C.H., 1953. Genetic assimilation for an acquired character. Evolution 7, 118−126.

Wagner, A., 2008. Neutralism and selectionism: a network-based reconciliation. Nat. Rev. Genet. 9 (12), 965−974.

West-Eberhard, M.J., 1989. Phenotypic plasticity and the origins of diversity. Annu. Rev. Ecol. Systemat. 20, 249–278.
Whitlock, M.C., 1999. Neutral additive variance in a metapopulation. Genet. Res. 74, 215–221.
Wiley, E.O., 1978. The evolutionary species concept reconsidered. Syst. Zool. 27, 17–26.
Wiley, E.O., Mayden, R.L., 2000. The evolutionary species concepts. In: Wheeler, Q.D., Meier, R. (Eds.), Species Concepts and Phylogenetic Theory. A Debate. Columbia University Press, New York, pp. 70–92.
Wilke, A., Christe, R., Multini, L., Vidal, P.O., Wilk-da Silva, R., de Carvalho, G.C., Marrelli, M., 2016. Morphometric wing characters as a tool for mosquito identification. PLoS One 11. https://doi.org/10.1371/journal.pone.0161643.
Wilkins, A.S., 1996. Canalization: a molecular genetic perspective. Bioessays 19, 257–262.
Workman, M.S., Leamy, L.J., Routman, E.J., Cheverud, J.M., 2002. Analysis of quantitative trait locus effects on the size and shape of mandibular molars in mice. Genetics 160 (4), 1573–1586.
Zelditch, M.L., Swiderski, D.L., Sheets, H.D., Fink, W.L., 2004. Geometric Morphometrics for Biologists: A Primer. Elsevier, Academic Press, New York.
Zhang, M., Ruan, Y., Wan, X., Tong, Y., Yang, X., Bai, M., 2019. Geometric morphometric analysis of the pronotum and elytron in stag beetles: insight into its diversity and evolution. ZooKeys 833, 21–40. https://doi.org/10.3897/zookeys.833.26164.
Zuñga Reinoso, A., Benítez, H.A., 2015. The overrated use of the morphological cryptic species concept: an example with *Nyctelia* darkbeetles (Coleoptera: Tenebrionidae) using geometric morphometrics. Zool. Anz. 47–53.

This page intentionally left blank

CHAPTER

Microbial Sequence Typing Strategies for Genetic and Genomic Data

17

Marcos Pérez-Losada[1], Eduardo Castro-Nallar[2,3] and Miguel Arenas[4,5]

[1]*Computational Biology Institute, Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, George Washington University, Washington DC, United States;* [2]*Departamento de Microbiologíá, Facultad de Ciencias de la Salud, Universidad de Talca, Talca, Chile;* [3]*Centro de Ecologíá Integrativa, Universidad de Talca, Talca, Chile;* [4]*CINBIO, Universidade de Vigo, Vigo, Spain;* [5]*Department of Biochemistry, Genetics and Immunology, Universidade de Vigo, Vigo, Spain*

1. Introduction

Microbial typing techniques are greatly enhancing our insights into microbial population epidemiology and microbial diversity assessment and are widely used in diagnostics, genomics, and pathogenesis related to microbiology research (Boers et al., 2012; Van Belkum, 2002). In fact, our ability to accurately distinguish among strains of infectious pathogens is crucial for efficient epidemiological and surveillance analysis, studying microbial population structure and dynamics and, ultimately, developing improved public health strategies and treatments (Cooper and Feil, 2004). To achieve these goals, several molecular typing methods have been proposed that can identify isolates worldwide (global epidemiology) or in localized disease outbreaks (local epidemiology)—see Foley et al. (2009) for a review.

Since 1998, the established standard for molecular typing is Multilocus Sequence Typing (MLST) (Maiden et al., 1998), which has proven to be an effective method for characterizing bacterial isolates. However, next-generation sequencing (NGS) technologies are drastically changing the field of microbial genomics research (Forde and O'Toole, 2013). These new sequencing methods supply a range of applications (Glenn, 2011), allowing for a more comprehensive and in-depth analysis of the diversity, structure, and content of microbial genomes at a substantially lower cost compared to traditional automated Sanger capillary sequencers (Metzker, 2010; Mardis, 2008). The reader is referred to the following reviews for further information about NGS technologies (Metzker, 2010; Didelot and Falush, 2007; Mardis, 2013; Goodwin et al., 2016). As a consequence, in June 2022, a total of 196,962, 12,895, and 173,035 sequencing projects in bacteria, viruses, and metagenomes (i.e., the study of microbial communities directly in their natural environments), respectively, have been deposited in the Genomes OnLine Database (GOLD) (Mukherjee et al., 2017). NGS platforms have proven to be effective tools for the resequencing and de novo sequencing reference microbial species and strains (pathogens and underrepresented taxa), but also assembling genomes of entire microbial communities

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00023-9
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

of unculturable microbes (microbiotas) (Kyrpides et al., 2014; Sharon and Banfield, 2013; Sangwan et al., 2016; Wu et al., 2022).

Noncultured organisms represent the vast majority of the total microbial diversity which exists in the world (Pace, 2009). Hence, to characterize them we must rely on their genomic material. Microbial genomic studies aim to identify microbial taxa and describe their diversity and structure at the species (or strain) and community levels through targeted sequencing of gene amplicons (e.g., housekeeping genes, 16S/18S rRNA, ITS) or shotgun sequencing of (nearly) full genomes (Petrosino et al., 2009; Caporaso et al., 2012; Vincent et al., 2017; Chun and Rainey, 2014; Kwong et al., 2015; MacCannell, 2013).

Concerning microbial typing, NGS strategies have also expanded the versatility of widely used typing approaches, such as MLST, to accommodate high-throughput data (e.g., NGMLST and HiMLST). The drawback is that this massive amount of data comes in the form of short reads with relatively high sequencing errors (Stoler and Nekrutenko, 2021)—although emerging long-read sequencing technologies have the potential to fix these problems (Delahaye et al., 2021; Karst et al., 2021); hence, one needs to invest heavily in computational analysis. Additionally, novel analytic and statistical techniques that can handle these data are continually being developed. Over the last 10−15 years we have witnessed the blooming of new typing approaches that take advantage of parallel amplicon and whole genome sequencing.

In this review, we describe and compare classical (e.g., MLST) and newer (e.g., HiMLST) DNA typing approaches (Section 2) of microbial sequence typing. Then we present statistical methods and computational tools for the analysis of nucleotide, locus, and genome data generated via Sanger and NGS platforms (Section 3). In the last section (Section 4), we present several applications of these DNA-based approaches to the fields of genotyping, phylogenetics and molecular epidemiology. We also refer the reader to other reviews on MLST for complementary information (Boers et al., 2012; Cooper and Feil, 2004; Pérez-Losada et al., 2006, 2013, 2017; Jolley and Maiden, 2014; Maiden, 2006; Urwin and Maiden, 2003; Sullivan et al., 2005; Jolley et al., 2012; Larsen et al., 2012).

2. DNA-Based Typing Approaches

2.1 Standard Typing Approaches: MLST and MLVA

Multilocus Sequence Typing (MLST) examines nucleotide variation in sequences of internal fragments of usually seven housekeeping genes, i.e., those encoding fundamental metabolic functions—although the number of genes may vary across strains, species, and other particularities of the studied sample. The different sequences of each of those genes present in a particular species are assigned unique allele numbers and, for each isolate, all the alleles of the combined seven loci define its allelic profile or sequence type (ST). Each isolate is, therefore, unambiguously characterized by a series of seven integers, which correspond to the alleles at the seven housekeeping loci.

MLST is widely used for molecular typing (Pérez-Losada et al., 2013; Jolley and Maiden, 2014; Maiden, 2006; Maiden et al., 2013). This was made possible by three advances in molecular microbiology (Maiden, 2006) involving (1) bacterial evolution and population biology insights; (2) high-throughput nucleotide sequencing; and (3) genetic sequence databases. The bacterial population studies undertaken from the 1980s onwards were central to the development of MLST. Those studies showed that genetic exchange among bacteria was more common than previously thought, leading to a

reassessment of the role of sexual processes in the structuring of bacterial populations. Using sequence data, it has been shown that recombination (mosaic genes) was not only frequent in genes under diversifying selection (e.g., antigen-encoding and antibiotic resistant genes) but also in genes under purifying selection (housekeeping genes) (Maiden, 2006). This suggested that the central dogma of the clonal model (variation can only arise by mutation) was not universal and led to the proposal of new nonclonal or panmictic (variation is mainly generated by recombination) and partially clonal models of bacterial population structure (Spratt and Maiden, 1999; Feil and Enright, 2004). Consequently, typing methods needed to accommodate and distinguish among a broader spectrum of population structures, hence providing not only discriminatory power but also information about the clonal structure of the organism under study. Therefore, only molecular techniques that can contrast results across independent markers (such as MLST) would be adequate for bacterial typing and population genetic analyses.

The length of the nucleotide sequence amplified for each housekeeping locus is generally in the range of 400−600 bp and is determined largely by the parameters of automated sequencing instruments available at the time. Most MLST nucleotide sequence data were generated by Sanger sequencing, however, this technology is being replaced by high-throughput technologies such as pyrosequencing (Boers et al., 2012; Margulies et al., 2005)—already discontinued, sequencing-by-synthesis (Illumina/Ion Torrent) and single-molecule sequencing (PacBio/Nanopore) (Pérez-Losada et al., 2013; Chen et al., 2015) for targeted-amplicon and whole-genome sequencing. These technologies can generate accurate read lengths of ∼50−300 bp to 10−25 kb (Illumina and PacBio, respectively) and up to 25−50 million paired-end reads (Illumina MiniSeq/MiSeq platforms) per run (Foox et al., 2021a,b;). Moreover, the design of barcoded primers allows simultaneous and efficient sequencing of homologous products from hundreds of samples in the same run (Kozich et al., 2013; Rao et al., 2016; Taylor et al., 2016).

One of the goals of the MLST approach was the development of online platforms containing MLST databases to which public health officials and researchers could both have access and contribute and from which clinical, epidemiological, and population studies could benefit (Maiden et al., 1998; Maiden, 2006; Urwin and Maiden, 2003). The first MLST online platforms were based on single databases implemented in the MLSTdB software (Chan et al., 2001), but as MLST schemes began to expand several limitations became apparent: redundant information (each record contained the ST designation and the allelic profile), isolate bias (single databases were dominated by specific studies), and access (all databases were stored at a single location). To overcome these limitations, a new network-based database software, MLSTdBNet (Jolley et al., 2004) was developed and implemented on the PubMLST site (http://pubmlst.org/)—as of June 2022 hosting alleles and isolates from 674,450 genomes. This site includes two databases: (i) a profiles database with the sequences of each MLST allele for each locus linked to an allele number, and (ii) an allelic profiles database with the corresponding ST designations. The profile database can then interact with other isolate databases. For each scheme on the PubMLST site there is a PubMLST isolate database that aims to include at least one isolate for each ST. MLST databases are hence different to other repository databases such as GenBank, not only in organization but also in active curation for accuracy. It is important to highlight that MLST databases do not embody the global diversity of an organism but the extent of its diversity at the time of access. Moreover, stored data are unstructured and do not necessarily represent natural populations either. As high-throughput sequencing becomes more affordable, PubMLST is increasingly including whole genome sequences, e.g., BIGSdb (Larsen et al., 2012; Jolley and Maiden, 2010).

As the number of schemes available has increased, MLST has become the most used method of pathogen typing. In comparison to older methods (serotyping; multilocus enzyme electrophoresis analysis), the use of genetic variation gives MLST the advantage of producing variable data (more resolution) that are universally comparable (within schemes), easily validated, and readily shared across laboratories. The use of sequencing makes MLST a broadly applicable methodology that can be fully automated and scalable from single isolates to thousands of samples. Importantly, the material needed for MLST analysis (DNA or dead cells) is easily transported among labs, without the problems associated with infective materials. Furthermore, the use of online databases to store and curate MLST schemes makes them a globally and highly accessible resource.

The number of loci that should be evaluated to confidently assign an ST has been minimized to reduce the expense and time required for characterization, with most studies using 6−10 loci. But if performed manually, evaluating even these many loci can be time consuming. However, fully automated systems, e.g., robotics (Jefferies et al., 2003; Sullivan et al., 2006) provide a high-throughput pipeline for data collection that can run large volumes of samples with increased reliability. Likewise, commercial solutions such as Ion Torrent AmpliSeq panels targeting MLST schemes (www.ampliseq.com) can reduce costs down to cents per marker. As sequencing technology progresses, we expect the cost of automation to decrease, thus data interpretation rather than data generation will be the likely limiting factor in our understanding of pathogen population dynamics.

By focusing on sequence variation, MLST provides a highly replicable and reproducible typing method. Additionally, a set of housekeeping-genes can provide significant amounts of genetic data to calculate pathogen population genetic parameters at both local and global scales. Those parameters can be then used to construct more sophisticated models of pathogen evolution and epidemiology that can improve our understanding about the spread of disease. However, there is no single set of universal housekeeping genes that can be used for all pathogens as the recombination rates, substitution rates, and levels of selection vary across loci and species (Pérez-Losada et al., 2006). Therefore, a unique set of loci must be identified for each novel, untyped pathogen under study. The rapid increase of available microbial genomes will make data mining for housekeeping genes more feasible, reducing the time and cost required for constructing new MLST schemes.

Currently, the main drawback of the MLST method is that the selection of housekeeping loci requires reference genomes (Parkhill et al., 2003). Moreover, not all pathogens are suitable for MLST methods. Some pathogens (e.g., *M. tuberculosis*, *Y. pestis*) exhibit very little variation throughout their entire genome, most likely representing "evolutionarily young" pathogens that have not yet accumulated sufficient genetic variation to differentiate strains. For typing these pathogens, more rapidly evolving loci (e.g., insertion sequences or antibiotic-resistance determinants) or more markers (e.g., genome-wide single nucleotide polymorphisms or SNPs) are needed. Conversely, some bacterial genomes have accumulated so much variation that MLST housekeeping genes do not provide adequate information for typing. As we advance MLST schemes in the genomic era, we should be able to combine information-rich and widely adopted MLST schemes with cost-effective whole-genome sequencing.

Given the above limitations of MLST, over the last few years other typing approaches have been developed based on similar principles. *Multilocus Variable number of tandem repeats Analysis (MLVA)* uses polymorphic repeated sequences (VNTR) instead of housekeeping genes. Comparative studies between MLVA and MLST have yielded similar results (van Cuyck et al., 2012) and in recently evolved species, the MLVA approach can provide higher discriminatory power (Marsh et al., 2010).

Finally, to achieve even greater resolution, other approaches have been developed based on core/accessory genes or distributed genes among bacterial species with the same MLST profile (Hall et al., 2010; Leekitcharoenphon et al., 2012). This new approach could skip the laborious and time-consuming steps needed to develop bacteria-specific MLST schemes; this is, however, replaced by in silico work.

2.2 NGS-Based Approaches: WGS, SNP, rMLST, cgMLST, HiMLST, NGMLST, and MetaMLST

Whole-genome sequencing (WGS) has emerged as a powerful technology for the comparison of isolates in outbreak analysis. The applications of WGS in clinical and public health microbiology have already been demonstrated in proof-of-concept studies conducted retrospectively or in response to an emerging outbreak (Walker et al., 2013; Didelot et al., 2012). Since then WGS is becoming an effective method for investigating bacterial genomes and is being implemented worldwide and across sectors (human, veterinary, food, and environment) in prospective surveillance, risk characterization models and bacterial typing (Dallman et al., 2015; den Bakker et al., 2014; Uelze et al., 2020). Salipante et al. (2015) explored the utility of WGS as a strain typing approach for the clinical laboratory using a single, universal protocol (encompassing library preparation, sequencing, and data analysis) for three distinct bacterial species: methicillin-resistant *Staphylococcus aureus*, vancomycin-resistant *Enterococcus*, and multidrug-resistant *Acinetobacter baumannii*. They showed that WGS was highly reproducible, which enabled a functional, quantitative definition for determining clonality. Then, others (Kwong et al., 2016) compared routine prospective WGS to other conventional typing methods, including MLST and MLVA, for the national epidemiologic surveillance of *Listeria monocytogenes*. MLST inferred in silico from the WGS data was highly concordant (>99%) with laboratory typing performed in parallel. However, WGS could identify distinct nested clusters within groups of isolates that were otherwise indistinguishable using traditional typing methods. As in previous studies, WGS provided a greater level of discrimination, than that from conventional typing, for surveillance or inferring linkage to point source outbreaks.

Single nucleotide polymorphism (SNP) is a variation in a single nucleotide that occurs at a specific position in the genome, where each variation is present to some appreciable degree within a population (usually at ≥1%). SNP-typing is widely used for bacteria and has also been improved by using genome sequences. When coupled with NGS of DNA, genome-wide screening of SNPs is a powerful discriminatory technique that enables the identification of strain-specific genetic markers. Genome-wide SNP typing has been, for example, applied to the genotyping of strains and species of *Bacillus anthracis* and *B. cereus* (Kuroda et al., 2010), *Streptococcus suis* (Chen et al., 2013), or *Coxiella burnetii* (Huijsmans et al., 2011) among others, outbreak attribution (Hendriksen et al., 2011), phylogeography of recently emerged diseases (Keim and Wagner, 2009), or genome-wide association studies of SNPs associated with acquisition of bacteremia in healthcare settings (Nelson et al., 2014). Pipelines for phylogenetic typing of SNP genotypes have also been developed to identify bacterial strains including metagenomic samples (Sahl et al., 2015).

Although MLST genotyping is a superb approach to characterize microbial species and strains, their methodological implementation can be costly, time-consuming, and laborious. To accelerate automation and expand the versatility of the current MLST method, other approaches that take advantage of NGS technology have been developed to produce millions of high-quality bases at low

cost within a single sequence run. The *Ribosomal Multilocus Sequence Typing method (rMLST)* has been proposed to index the molecular variation of 53 genes encoding bacterial ribosome protein subunits (Jolley et al., 2012). This method pursues the integration of a taxonomic and typing method in a similar curated MLST scheme. Although more expensive, rMLST is likely to provide better resolution than standard MLST methodologies. Likewise, *core-genome (cg) MLST* has been developed to overcome a lack of resolution of MLST schemes for certain taxa. By collecting a sample of genome sequences representing extant diversity, the cgMLST scheme applies more than 1000 genes to create sequence types with increased resolution for clonal populations of bacteria (de Been et al., 2015).

High-throughput MLST (HiMLST) is another approach that uses the Genome Sequencer Junior (Roche) and generates up to 70,000 amplicon reads with average read lengths of around 400−500 bases (Boers et al., 2012). This long-read sequencing performance, in combination with a sample pooling strategy that uses "bar-coded" amplicons for parallel analysis of pooled samples, generate MLST profiles from multiple bacterial isolates in a single NGS-run. They then demonstrate the successful parallel sequencing of MLST alleles that were amplified by PCR from 96 bacterial isolates from four species in single NGS-runs. Roche 454 sequencing (pyrosequencing) technology has been discontinued and replaced by the Illumina MiSeq platform (sequencing-by-synthesis), which provides greater sequencing depths and a substantial reduction of labor and costs.

More recently, another high-throughput *Next-Generation sequencing MLST (NGMLST)* approach and concomitant automated software (MLSTEZ) have been developed for microbial typing (Chen et al., 2015). Essentially, they coupled an efficient multiplex PCR approach with PacBio circular consensus sequencing technology, which can generate relatively inexpensive single-molecule consensus reads of 1−2 kbp with lower error rates (after multiple sequencing cycles). The software MLSTEZ can then automatically identify the barcodes and primers used in the PCR, correct sequencing errors, generate the MLST profile for each isolate, and predict potentially heterozygous loci. Next, they compared NGMLST to conventional MLST. The major advantages of the NGMLST approach at the time of analyses were (i) the application of multiplex PCR greatly reduces the amount of labor; (ii) PacBio greatly extends the maximum read length of target loci or genes from 500 bp to 2 kb without requiring fragmentation into shorter sequences; (iii) the NGMLST workflow is optimized to reduce unnecessary steps; (iv) MLSTEZ can be easily implemented and does not require technical expertise or a background in bioinformatics; and (v) for analysis of hybrid isolates, unlike most programs, MLSTEZ can detect heterozygous loci and sequence their alleles.

Finally, another software tool for microbial typing of haploid organisms called *MetaMLST* that combines the effectiveness of the MLST approach with the high throughput of metagenomics has entered the field (Zolfo et al., 2017). MetaMLST overcomes the computational limitations and lowers the limit of detection (i.e., strain level) compared to metagenomic assembly. The authors then tested the pipeline on synthetic and spiked-in real metagenome datasets and showed that MetaMLST reconstructed the MLST sequences with high accuracy at low coverage (as low as 1×). When applied to real biological samples, MetaMLST also showed higher sensitivity than assembly-based approaches allowing the identification of pathogenic strains in epidemic outbreaks.

3. DNA-Based Typing Analyses

MLST data is usually analyzed by two main strategies depending on type of applied molecular marker (Fig. 17.1). One of them relies on the consideration of allele and ST designations (the number of

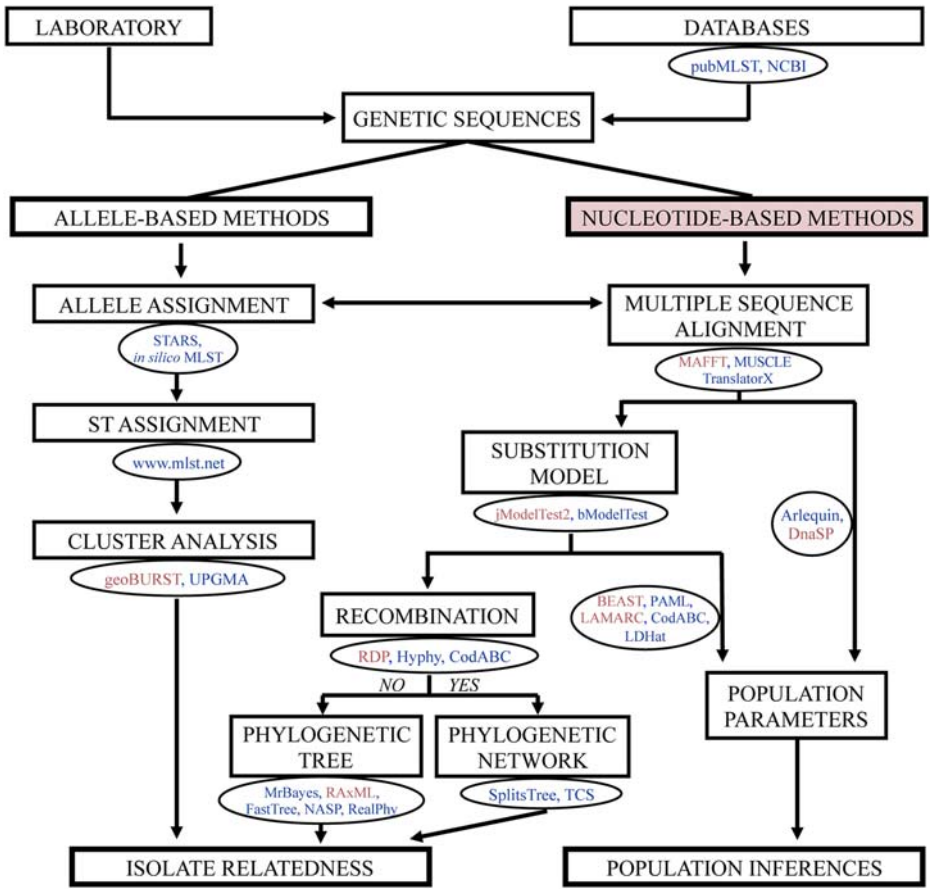


FIGURE 17.1 Standard workflow for typing analysis of microbial genetic sequences using allele- and nucleotide-based approaches.

The workflow shows data and tasks in *boxes*, and databases and computer frameworks in *circles*. We highlighted in *light red* our recommended approach and computer frameworks. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

nucleotide differences between alleles is ignored) to estimate relatedness among isolates (*allele-based methods*), where nucleotide differences among alleles are considered as an ST unit. The other one directly applies nucleotide sequences to estimate genetic relatedness and population genetics parameters (*nucleotide-based methods*). The allele-based approach was derived from Multilocus Enzyme Electrophoresis data and thus it was the first method applied to the analysis of MLST data (Maiden et al., 1998; Enright and Spratt, 1999) while nucleotide-based methods emerged later and provides accurate estimates by considering the number and type of nucleotide differences between sequences (Gurjar et al., 2021; Bezdicek et al., 2021). In general, the allele-based approach is thought to work well in nonclonal organisms (e.g., *Helicobacter pylori*), while nucleotide-based approaches are

preferable for clonal organisms (e.g., *Escherichia coli*) (Maiden, 2006). In practice, most microbes show some degree of clonality (clonal complex) in their populations and, therefore, we suggest that both types of analysis can be considered in population and epidemiological studies (Bezdicek et al., 2021; Tazi et al., 2010; Coipan et al., 2020; Miro et al., 2020). In this section, we present a brief description of the most used approaches for analyzing MLST data, including allele, nucleotide, SNP, and WGS data.

3.1 Allele-Based Analytical Methods

Allele-based methods consider the allele as the unit of analysis and, consequently, they first require assigning an allele number to each DNA sequence from each locus (Fig. 17.1). This can be performed by matching the study sequences against sequences stored in public MLST databases. If no match is found a number can be assigned following the order of discovery. Several computational programs were developed to perform this task, although *Sequence Typing Analysis and Retrieval System* (STARS) seems to be the most functional and popular (Sullivan et al., 2005). The STARS interface was specifically designed for typing and implements the assembly of many sequences. Once the alleles are assigned, the data can be incorporated to MLST curated databases (Jolley et al., 2004; Aanensen and Spratt, 2005; Jolley and Maiden, 2006) to acquire an ST profile. At this point, exploratory analyses (e.g., allele and profile frequencies, polymorphism estimates and codon usage) of the data can be performed. The software package *Sequence Type Analysis and Recombinational Tests* (START2) implements all these tasks (Jolley et al., 2001). Next, relatedness among STs can be obtained with heuristic approaches of cluster reconstruction such as *Based Upon Related Sequences Types* (eBURST) (Feil et al., 2004), the simple *Unweighted Pair Group Method with Arithmetic Mean* (UPGMA), *minimum-spanning trees* (MST), or network-based methods such as *NeighborNet* (Bryant and Moulton, 2004) and split decomposition (Bandelt and Dress, 1992), among others (Huson and Bryant, 2006).

The method eBURST implements a simple model of clonal expansion and diversification (Feil et al., 2004). First, it identifies mutually exclusive groups of related STs and next, it tries to identify the founding ST of each group. Bootstrap estimates can also be calculated to assess statistical confidence in the grouping. The algorithm predicts the descent from the previously identified founding ST considering the other STs in the group and finally provides a radial diagram centered on the predicted founding ST. A globally optimized version (goeBURST) identifies alternative patterns of descent with a graphic matroid approach (Francisco et al., 2009). The framework PHYLOViZ integrates allelic profiles from MLST or MLVA methods (although SNP data can also be included) and associated epidemiological data (Francisco et al., 2012). PHYLOViZ uses goeBURST for representing the estimated evolutionary relationships among strains. The recently developed clustering framework HierCC (Zhou et al., 2021) is based on core genome multilocus typing and implements incremental, static, multi-level cluster assignments of genomes. This framework was implemented in EnteroBase database (Zhou et al., 2020) and genotyped more than 530,000 genomes from diverse microorganisms (Zhou et al., 2021).

The traditional phylogenetic methods based on distances (i.e., UPGMA) infer isolate relatedness from the distances calculated for each pair of STs considering the number of allelic differences and next, groups can be sequentially clustered considering that STs similarity (i.e., allelic matches) to generate a phylogenetic tree. Other distance-based methods, such as NeighborNet and split

decomposition, provide clustering through phylogenetic networks, considering signatures of past reticulate evolutionary events or alternative evolutionary histories (Huson and Bryant, 2006; Fitch, 1997). These phylogenetic networks are widely used to analyze MLST data (Jolley and Maiden, 2014; Maiden et al., 2013; Patiño et al., 2018). Additional distance and parsimony-based methods have been proposed to estimate relatedness from allele frequencies but distance methods generally outperform parsimony methods (Wiens, 2000).

Allele-based methods provide the advantage of simplicity and speed (especially relevant when dealing with large datasets like those derived from NGS), which are crucial for efficient epidemiological surveillance and public health management, but ignore much of the evolutionary information present at the nucleotide level. A larger and more accurate plethora of nucleotide-based methods is available to study isolate genetic relationships and population genetics processes.

3.2 Nucleotide-Based Analytical Methods

The analysis of nucleotide data usually starts with the generation of a multiple sequence alignment (MSA), this is the identification and phasing of the homologous nucleotide sites. Since the loci traditionally used for MLST evolve slowly and codify for proteins, this step becomes straightforward especially at the amino acid level (Fig. 17.1). If needed, a variety of fast, accurate, and user-friendly aligning methods are implemented in the frameworks MAFFT (Katoh and Standley, 2013; Katoh et al., 2019), MUSCLE (Edgar, 2004), and TranslatorX for translated-alignment of coding sequences (Abascal et al., 2010), among others (Deorowicz et al., 2016; Sievers and Higgins, 2018).

After obtaining an MSA, a substitution model of molecular evolution that best fits with the data can be identified. During the last decades, substitution models have increased in complexity by incorporating more informative parameters to better explain the real observations (Arenas, 2015b). The consideration of the best fitting substitution model is critical because it can affect diverse subsequent phylogenetic and population analyses (Lemmon and Moriarty, 2004; Buckley, 2002), despite there is a recent debate concerning the influence of substitution model selection on the accuracy of phylogenetic tree reconstruction (Buckley, 2002). Substitution model choice is currently assessed with evolutionary frameworks such as ModelTest-NG (Darriba et al., 2020), which implements confidence sets of models (model averaging) (Posada and Buckley, 2004) and several criteria for model selection such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Decision Theory (DT), and hierarchical Likelihood Ratio Test (hLRT). Although AIC is the most broadly used criterion for evaluating model fit, BIC and DT are recommended (Luo et al., 2010). Conveniently for the analyses of large amounts of data, ModelTest-NG can run in parallel on multiple processors and other substitution model selection frameworks implement heuristics to avoid extensive calculations (Lefort et al., 2017). It is known that substitution models can vary among genomic regions (Arbiza et al., 2011) and thus frameworks for inferring nucleotide partitions with different substitution models like PartitionFinder2 (Lanfear et al., 2016), as well as for inferring substitution models under a Bayesian framework (Bouckaert and Drummond, 2017) were developed. Several phylogenetic tree reconstruction frameworks, such as IQ-TREE (Nguyen et al., 2015) and MrBayes (Ronquist et al., 2012), internally implement substitution model selection, simplifying the phylogenetic tree reconstruction pipeline. However, substitution model selection is required for many other evolutionary analyses such as the reconstruction of ancestral sequences and the detection of molecular adaptation (Kosakovsky Pond et al., 2005; Yang, 2007).

3.2.1 Phylogenetic Relatedness

Phylogenetic reconstruction methods are traditionally classified into two categories considering the genetic data and mathematical algorithm, methods based on distances (e.g., UPGMA and Neighbor-Joining, NJ; described in the previous section) and methods based on optimality criteria (e.g., maximum likelihood and Bayesian analysis). In this section, we focus on methods that implement maximum likelihood and Bayesian optimality criteria and that allow for considering multiple data partitions each under a best-fit model, which is a feature particularly important for analyzing MLST data and that can improve the accuracy of phylogenetic reconstructions (Zoller et al., 2015).

Phylogenetic maximum likelihood (ML) inference identifies the topology that explains the evolution of the aligned sequences, under a given substitution model of evolution, with the highest likelihood (Felsenstein, 1981). A variety evolutionary frameworks implement phylogenetic tree reconstruction (Yang and Rannala, 2012; De Bruyn et al., 2014), but only a few consider genetic partitions with evolution modeled under different substitution models. RAxML (Stamatakis, 2006) implements the ML criterion efficiently and can handle large datasets (more than 1000 taxa with more than 20,000,000 bp) (Stamatakis et al., 2012). Confidence in the estimated relationships (i.e., clade support) is usually assessed with a nonparametric bootstrap procedure (Felsenstein, 1985), which should be repeated more than 1000 times to achieve reasonable precision. Thus, RAxML is widely used to analyze MLST data (Zolfo et al., 2017). Next, RAxML has been extended to a new version RAxML-NG (Kozlov et al., 2019), which among diverse improvements (including accuracy, flexibility, speed, scalability, and advanced bootstrap support metrics), allows phylogenetic tree reconstruction under a user-specified site-specific substitution model of DNA or protein evolution. Another ML framework oriented to analyze large datasets and heterogeneous evolution among partitions is FastTree2 (Price et al., 2010). This program can internally optimize phylogenetic tree inferences by the joint application of several approaches (NJ, minimum-evolution and ML) and it is also frequently considered to analyze MLST data (Kwong et al., 2016; Skarp-de Haan et al., 2014). Interestingly, FastTree can infer phylogenetic trees with similar accuracy to RAxML but several orders of magnitude faster (Liu et al., 2011), but it was not yet compared with RAxML-NG. Another phylogenetic tree reconstruction framework that is increasing in popularity is IQ-TREE (Nguyen et al., 2015). This tool is slower than RAxML-NG but produces likelihoods with better stability (lower variance among higher-scoring trees) (Kozlov et al., 2019).

Bayesian inference (BI) combines the prior probability and likelihood of phylogenies to produce a posterior probability distribution of trees, which can be interpreted as the probability of those trees being correct (Huelsenbeck et al., 2001). Clade statistical support is estimated by summarizing the posterior distribution of trees through consensus analysis. Bayesian phylogenies are usually reconstructed with the Metropolis-coupled Markov chain Monte Carlo (MCMC) method and are implemented in diverse programs such as MrBayes (Fredrik Ronquist et al., 2012), RevBayes (Höhna et al., 2016), BEAST (Drummond et al., 2012; Bouckaert et al., 2014), and PhyloBayes (Lartillot et al., 2013). The output of the BI analysis must be evaluated to ensure that the MCMC chains are well mixed and converged. This task can be performed by the own BI framework (i.e., MrBayes) or can be performed by an external tool like TRACER (Rambaut and Drummond, 2009), for example, for inferences with BEAST. Importantly, as indicated previously, the best fitting substitution model can vary across partitions. In this concern, PhyloBayes implements priors for the internal assignment of across-partition heterogeneity (Lartillot and Philippe, 2004). Importantly for many datasets, Bayesian

approaches can perform phylogenetic inferences accounting for longitudinal sampling (Navascués et al., 2010; Rieux and Balloux, 2016) and relaxed molecular clocks (Zhang and Drummond, 2020), leading to phylogenetic tree reconstructions under much more realistic evolutionary scenarios (Zhang and Drummond, 2020).

It is common to observe that gene trees differ even when sampled from the same population. This can be the result of molecular processes (e.g., recombination and horizontal gene transfer) or stochastic variation (e.g., lineage sorting). Whatever the case, it may be necessary to check if individual gene topologies are significantly different because ignoring these processes may lead to biased phylogenetic trees and analyses derived (Schierup and Hein, 2000; Arenas and Posada, 2010b; Mallo et al., 2015). A number of ML-based phylogenetic incongruence tests were developed (Leigh et al., 2011), such as those implemented in CONSEL (Shimodaira and Hasegawa, 2001). In the section *GWS-based analyses*, we describe the phylogenetic tree reconstruction from genome-scale data accounting for genomic evolutionary events.

3.2.2 Population Dynamics

The evolution of genetic sequences is affected by processes occurring at the population level such as recombination, population size dynamics and selection, among others. Consequently, the accurate estimation of population genetics parameters is key for understanding the dynamics of evolutionary processes occurring in populations, including the generation of diversity at the population level and its consequences and applications such as epidemiology and the design of efficient public health control strategies (Omenn, 2010). Population genetics parameters are efficiently estimated with explicit statistical models such as the coalescent (Kingman, 1982) and birth-death (Novozhilov et al., 2006) and, therefore, most of the well-established evolutionary frameworks are based on these models. Next, we describe the estimation of relevant population genetics parameters from nucleotide data.

The rate of evolution quantifies the frequency of genetic change over time. Traditionally, this parameter was assumed constant over time (strict molecular clock). However, observations from diverse real data that the rate of evolution can vary over time (Bello et al., 2007; dos Reis et al., 2016). The violations of the molecular clock can be caused by multiple factors such as changes in the generation time, population size or metabolic rate. The rate of evolution can be accurately estimated with Bayesian approaches (i.e., with BEAST) under models of a relaxed molecular clock (Drummond et al., 2006; Rannala et al., 2007). Moreover, the framework BEAST allows to identify if DNA sequences evolved with or without following a strict molecular clock as well as the strength of molecular clock violation. Not only Bayesian methods can be used for testing the molecular clock (Ho and Duchêne, 2014) but the advantage of current Bayesian implementations is to accommodate multiple sources of uncertainty (prior distributions) with a dating analysis (Heled and Drummond, 2012).

Genetic recombination constitutes a relevant evolutionary force and produces diverse biological consequences (i.e., by increasing genetic diversity (Spencer et al., 2006)) and altering the estimation of other evolutionary parameters and processes (i.e., selection (Arenas and Posada, 2010a; Anisimova et al., 2003)) and ancestral sequence reconstruction (Arenas and Posada, 2010b). Statistical methods for analyzing recombination (Martin et al., 2011; Posada et al., 2002) have recommended the use of multiple recombination analyses methods to obtain well-supported conclusions. Next, we describe some of the most relevant and recent frameworks to detect recombination, predict recombination breakpoints and estimate recombination rates. The detection of presence of recombination can be performed with fast statistical tests such as Phi, Max χ^2 and NSS (Bruen et al., 2006). The framework

RDP (Martin et al., 2021) implements a large variety of tests and methods to study recombination at different levels. It includes multiple methods to estimate recombination breakpoints with statistical confidence, and identifies the specific sequences involved in the recombination events. Another ML method to detect recombination is GARD (Kosakovsky Pond et al., 2006), which is implemented in the HYPHY package (Pond et al., 2020) and in the Datamonkey webserver (Weaver et al., 2018), and outperformed previously developed methods. However, in HYPHY, GARD requires a multiprocessor machine and in Datamonkey, it can only analyze small datasets. Other programs can be used to estimate recombination rates and, therefore, quantify the amount of observed recombination:

- LAMARC (Kuhner, 2006), which is based on ML and Bayesian estimation although it can show problems for reaching convergence.
- LDhat (McVean et al., 2004), which is based on a composite likelihood method and it is widely used to study recombination in diploids. However, it is very slow in the analysis of large datasets such as more than 300 sequences and thousands of sites.
- CodABC (Arenas et al., 2015), which allows the analysis of codon sequences based on the approximate Bayesian computation approach (ABC) (Beaumont et al., 2002). It implements analysis of selection considering recombination (dN/dS) but can be very slow when dealing with many and long sequences due to the application of substitution models of codon evolution.
- OmegaMap (Wilson and McVean, 2006), which also allows the analysis of codon sequences. It is based on Bayesian analysis and, as CodABC, it implements analysis of selection considering recombination (dN/dS) but also can be very slow when dealing with many and long sequences due to the application of substitution models of codon evolution and Bayesian computation.
- ProteinEvolverABC (Arenas, 2022), which allows the analysis of protein sequences based on ABC. To our knowledge this is the only method to estimate the recombination rate from protein data. It is faster than CodABC but still very slow when dealing large datasets due to the simulations for the ABC method.

Conveniently, all these recombination rate estimation methods can also estimate the substitution rate because of the relationships between the population recombination rate ($\rho = 2nNrl$, where $n = 1$ or 2 for haploid or diploid populations, N is the effective population size, r is the recombination rate per site per generation and l is the alignment length) and the population substitution rate ($\theta = 2nN\mu l$, where μ is substitution rate per site per generation) through the effective population size. It is very important to mention that the observed recombination is highly influenced by the substitution rate. Under low substitution levels, recombination can be underestimated due to lack of genetic diversity (Posada and Crandall, 2001). A recent example is the analysis of recombination in the SARS-CoV-2, where only few recombination events were detected probably due to low genetic diversity (Dorp et al., 2020; VanInsberghe et al., 2021).

Another relevant population genetics parameter for characterizing microbial population dynamics is the population growth rate, which influences the variation of genetic diversity over time. Population growth rate can be estimated under a certain demographic model (e.g., exponential variation of the population size over time). Actually, it is possible to estimate the evolution of the population size over time with BEAST (Minin et al., 2008). The exponential growth rate can also be estimated with LAMARC (under both ML and Bayesian approaches) and with also with ABC-based methods.

In pathogen populations, selection is usually estimated from protein-coding sequences with the nonsynonymous (d_N) to synonymous (d_S) substitution rates ratio d_N/d_S (also named as ω). In

particular, $\omega > 1$ is considered as positive or diversifying selection, $\omega < 1$ suggests negative or purifying selection and, $\omega \approx 1$ indicates lack of selection (neutral evolution). Accurate estimation of ω can be obtained with ML (Kosakovsky Pond and Frost, 2005), Bayesian (Wilson and McVean, 2006), and ABC (Lopes et al., 2014) approaches under the assumption of an explicit substitution model of codon substitution. These substitution models can be very complex, allowing, for example, ω to vary across codon sites and/or tree branches under diverse probability distributions (Pond and Frost, 2005; Yang and Nielsen, 2002) although these estimates require many taxa to provide enough information about codon substitutions (Kosakovsky Pond and Frost, 2005). The estimation of ω is implemented in a wide variety of computational frameworks such as PAML (Yang, 2007), HYPHY, Datamonkey, OmegaMap and CodABC. Importantly, if recombination is detected in the data, it should be considered when estimating ω to avoid identifying false positively selected sites (Arenas and Posada, 2010a; Anisimova et al., 2003; Arenas and Posada, 2014a). In particular, if recombination is detected, one could co-estimate recombination and selection rates simultaneously with frameworks such as OmegaMap or CodABC, or account for the former while estimating the latter (e.g., HYPHY and Datamonkey).

Other relevant processes involved in microbial population genetics and dynamics include time since the onset of the expansion (e.g., pathogen outbreaks) and geographical distribution of pathogens. Probabilistic models based on the Bayesian approach (i.e., implemented in BEAST, BPEC, and SPREAD) (Bielejec et al., 2011) are currently applied for the inference and hypothesis testing of divergence times, ancestral locations and historical patterns of migration (i.e., phylogeographic history) (Lemey et al., 2020; Oaks, 2019).

3.3 SNPs-Based Analyses

Typing data can also be presented as strings of single nucleotide polymorphisms (SNPs) that vary among sequences (Maiden et al., 2013). SNPs can be detected by mapping reads against a related reference sequence and identifying the nucleotide differences (den Bakker, 2011). These strings of SNPs can be used to infer evolutionary relatedness and to study the evolution and dynamics of populations (Fig. 17.2). In comparison to nucleotide sequences (that include four character states), evolutionary analyses based on SNPs (based on only two states) are much faster (particularly convenient for dealing with whole genome data) and could be less affected by homoplasy (Pearson et al., 2009) but, they could also be less accurate due to considering less information. For example, analyses based on four states data can consider substitution models of DNA evolution leading to inferences more accurate than analyses ignoring substitution models (Yang, 1994). Indeed, SNPs can present low mutation rates and therefore accurate analyses may require large sequences (Pearson et al., 2009).

The arrival of NGS transformed microbiology, allowing genome sequences for a huge number of species and pathogenic strains. However, converting millions of sequencing reads per sample into meaningful biological data is not trivial, and methodological choices made to perform genome assembly, sequence alignment, and SNP calling can affect the resulting whole-genome sequences and results derived (Sutton et al., 2019). SNP calling from microbial genomes includes some major challenges, such as the selection of a reference genome (if available), presence of rare polymorphisms within a culture, and use of de novo genome assemblies (Olson et al., 2015; Yoshimura et al., 2019). Considering the variety of methods available to call SNPs from microbial genomes (Yoshimura et al.,

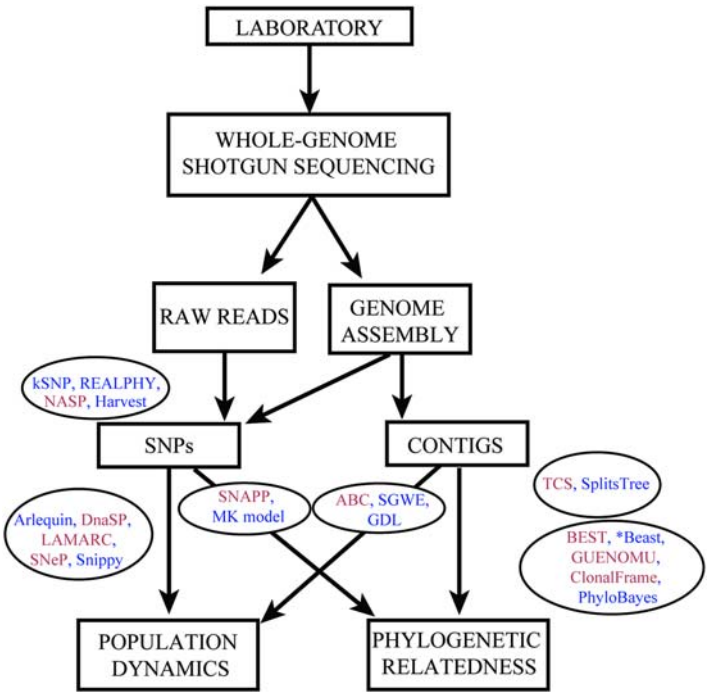


FIGURE 17.2 Standard workflow for typing analysis of microbial genomic sequences using whole-genome shotgun sequencing (WGS).

Current methods usually study SNPs or gene regions (contigs) for assessing phylogenomics relatedness and/or population dynamics patterns. The workflow shows data and tasks in *boxes*, and databases and, computer frameworks in *circles*. We highlighted in *light red* our recommended computer frameworks. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

2019; Nielsen et al., 2011), the optimization of computational pipelines for specific organisms and/or experiments is frequently required (Olson et al., 2015). First, a reference sequence should be identified for the sequences alignment (i.e., by mapping read sequences to a reference genome with frameworks such as Snippy (Kwong et al., 2016) followed by a refinement of the alignment (Wegrzyn et al., 2009; Chang et al., 2009)). However, some pipelines can infer SNPs directly from K-mers without the use of genomic references (Gardner et al., 2015). These steps of the pipeline are crucial and errors can lead to biased estimates (Pettengill et al., 2014; Castro-Nallar et al., 2015).

3.3.1 Phylogenetic Relatedness

SNP sequences can be used to construct a distance matrix that generates (i.e., by hierarchical clustering) a phylogenetic tree (Hendriksen et al., 2011; Gardy et al., 2011) or a phylogenetic network (Huson and Bryant, 2006; Huson, 1998). Commonly used frameworks to perform phylogenetic inference from a distance matrix are GARLI (Bazinet et al., 2014), FastTree, and SplitsTree (Huson, 1998). Another interesting program is SNAPP (Bryant et al., 2012), which performs phylogenetic

inferences from SNP data under a Bayesian approach similar to that implemented in BEAST for DNA sequences. Alternatively, researchers can use substitution models originally designed for analysis of morphological data such as the MK model (Lewis, 2001). Interestingly, the program BioNumerics (http://www.applied-maths.com) implements the steps to analyze SNP data including the alignment with a reference sequence, filter out SNPs and the phylogenetic reconstruction. Additional frameworks to perform phylogenetic inference from SNP data are (see also Fig. 17.2)

- NASP, which is particularly useful for phylogenomics data and supports a variety of input formats (Sahl et al., 2016).
- REALPHY, which is particularly useful for raw sequencing reads (Bertels et al., 2014).
- kSNP, which can analyze genomes without requiring a genome alignment or a reference genome (Gardner et al., 2015).
- Harvest, which analyses genomic data accounting for core-genome alignment, variant calls and recombination (Treangen et al., 2014).
- SNPhylo, which can reconstruct an ML phylogenetic tree from millions of SNP loci (Lee et al., 2014).

3.3.2 Population Dynamics

Genetic diversity in populations can be easily estimated from SNP data with traditional frameworks such as Arlequin (Excoffier and Lischer, 2010) and DnaSP (Rozas et al., 2017). Note that these programs implement the calculation of a variety of genetic summary statistics to measure genetic diversity in populations (i.e., number of alleles, heterozygosity and pairwise differences) and genetic differentiation among populations (i.e., F_{ST} and F_{CT}).

Indeed, population genetics parameters such as effective population size, population growth rate and migration rate can also be estimated from SNP data. The traditional procedure to estimate the effective population size was based on estimates of linkage disequilibrium (LD; the nonrandom association between alleles at different loci in a population) (Do et al., 2014; Hill, 1981). However, some evolutionary processes, such as admixture and genetic drift (Wang, 2005), hitchhiking during selective sweeps and background selection (Charlesworth et al., 1997), can affect the estimation of LD. Consequently, computer programs that consider these processes are recommended (e.g., SNeP Barbato et al., 2015). Additionally, ML and Bayesian approaches can be used to estimate population genetics parameters such as population growth rate, recombination rate, and migration rate from SNP data (Kuhner, 2006) (e.g., with the frameworks LAMARC and LDHat). Finally, ABC can also be used to estimate population genetics parameters from SNP data (Theunert et al., 2012; Cornuet et al., 2014). In this concern, a variety of computer simulators (i.e., fastsimcoal2 (Excoffier, 2021) and SPLATCHE3 (Currat et al., 2019)) implement the simulation of SNPs under diverse evolutionary scenarios (i.e., complex demographics, population structure and migration). Indeed, as noted above, genetic summary statistics required for ABC can be obtained from SNP data with programs such as Arlequin and then applied to the study of pathogens, for example (Labbé et al., 2017).

3.4 GWS-Based Analyses

Genome-wide sequences can decipher complex evolutionary histories (Pérez-Losada et al., 2020), among many other applications including in the fields of comparative and functional genomics (Gray

et al., 2015; Kobras et al., 2021). However, large datasets can lead to systematic errors by statistically well-supported incorrect estimates (Phillips et al., 2004; Rodréguez-Ezpeleta et al., 2007; Kumar et al., 2011). Importantly, the impact of several evolutionary processes specific to GWS evolution (i.e., horizontal gene transfer (HGT), gene duplication and loss (GDL), incomplete lineage sorting (ILS) or heterogeneous evolution among genomic regions) should be examined to avoid biased analyses (Mallo et al., 2015; Jeffroy et al., 2006; Galtier and Daubin, 2008). We next describe methodologies to consider these processes in evolutionary and population genetics analyses (Fig. 17.2).

3.4.1 Phylogenetic Relatedness

A variety of coalescent-based methods were developed to deal with the stochastic variation among gene trees from multilocus molecular data and to infer species trees (Fig. 17.2). Among these, BEST (Liu, 2008) and *BEAST (Heled and Drummond, 2010) incorporate the effect of ILS by implementing multispecies coalescent into a Bayesian hierarchical model and can be robust to the presence of paralogs (Yan et al., 2022). Unfortunately, these programs require extensive computation (Ogilvie et al., 2016). Concatenation methods assuming that all recovered gene trees share a common evolutionary history (Degiorgio and Degnan, 2010; Larget et al., 2010) can be an alternative. However, subsampled datasets analyzed with *BEAST produced more reliable results than full datasets analyzed with concatenation methods (Ogilvie et al., 2016). A few algorithms explored additional processes occurring in phylogenomic datasets (ILS, GDL, or HGT) (Rasmussen and Kellis, 2012; Yu et al., 2013; Szöllosi and Daubin, 2012) although, to our knowledge, none analytical program implements all of them excepting the framework GUENOMU (de Oliveira Martins and Posada, 2017). This framework implements a hierarchical Bayesian model that jointly considers ILS, GDL, and HGT and has the advantages of not requiring identification of orthologs and considers multiple individuals from the same species.

The evolutionary analysis of microbes based on DNA sequences often includes recombination (Arenas et al., 2018; Pérez-Losada et al., 2015). In presence of recombination, the evolutionary history of those sequences is no longer captured by a bifurcating model and, therefore, a phylogenetic tree representation may fail (in terms of both topology and branch lengths) to accurately represent the evolutionary history (Schierup and Hein, 2000). Under this scenario, two strategies can be used to consider recombination in phylogenetic reconstruction and the choice depends on the aims of the study:

1. Inference of a phylogenetic network (Huson and Bryant, 2006) such as the ancestral recombination graph (ARG) (Arenas, 2013). The computer programs TCS (Templeton et al., 1992) and SplitsTree can infer phylogenetic networks under different mathematical algorithms. Didelot and Falush (2007) developed a Bayesian coalescent approach (ClonalFrame) that takes homologous recombination into account while inferring clonal relationships between the members of a sample. The inference of the traditional recombination network (the ARG) is not straightforward and computationally demanding, only some studies explored algorithms for this inference from which only a few implemented them into analytical frameworks (Rasmussen et al., 2014; Hubisz and Siepel, 2020). These phylogenetic networks are especially interesting for identifying the presence of clusters and their genetic relationships (Huson and Bryant, 2006). Illustrative examples of reconstructed phylogenetic networks based on recombination can be found here (Rivailler et al., 2019; Kolb et al., 2017; Castelhano et al., 2017).

2. Independent phylogenetic tree reconstruction for each recombinant fragment (Arenas and Posada, 2010b). This methodology is based on two steps. First, recombination breakpoints are detected with programs such as RDP and Hyphy (through GARD, see above). Next, a phylogenetic tree is inferred for each recombinant fragment. This strategy is particularly useful for performing subsequent evolutionary analyses accounting for recombination, such as ancestral sequence reconstruction (Arenas and Posada, 2010b) and molecular adaptation (Pérez-Losada et al., 2009, 2011).

The pattern of the substitution process along the genome can be highly heterogeneous and thus different genomic regions can fit with different substitution models of evolution (Arbiza et al., 2011). Conveniently, the best-fitting substitution models for different genomic regions can be selected with frameworks like ModelTest-NG and PartitionFinder2. Since the selection of the best fitting substitution model is crucial for accurate phylogenetic inferences (Lemmon and Moriarty, 2004), phylogenetic programs that consider heterogeneous substitution across sites or regions (e.g., RAxML-NG and MrBayes) are recommended.

3.4.2 Population Dynamics

The previously described methods to analyze population dynamics from short nucleotide sequences can be extended to analyze GWS. Next, genomes can experience particular evolutionary events such as duplications, insertions, deletions, inversions, and translocations of genetic regions, among others. These events are difficult to model, leading to inefficient or intractable ML functions (Marjoram et al., 2003; Wegmann et al., 2009). To deal with complex evolutionary scenarios, ABC can be used as an alternative (Arenas, 2015a; Boitard et al., 2016). In this concern, some frameworks (Fig. 17.2) were developed to simulate GWS data accounting for complex evolutionary genomic events. For example, the program ALF can simulate genome evolution accounting for GDL, gene fusion and fission, lateral gene transfer (LGT), genome rearrangement and speciation under a birth-death process (Dalquen et al., 2012). Indeed, the program SGWE can simulate genome evolution accounting for heterogeneous recombination and substitution rates along sequences and can combine different gene trees into a species tree (Arenas and Posada, 2014b). Next, prior distributions and summary statistics for the ABC analysis could be based on the whole genome but also could be specific for genomic regions (Arenas, 2015a). To our knowledge, there is not yet an ABC framework available to analyze GWS data accounting for all the cited genomic events, but recent studies already analyzed this data with ABC under simplified models (Boitard et al., 2016; Smith and Flaxman, 2020) indicating that progresses in this direction are being made.

4. Microbial Sequence Typing Applications

Most contemporary studies use more than one approach to typing, epidemiology, and phylogenetic inference, with the aim of maximizing compatibility with current and past data and genetic resolution down to the strain level (e.g., MLST, SNPs, and WGS) (Castro-Nallar et al., 2015). Moreover, considering the current antibiotic crises highlighted by recent reports by the World Health Organization, researchers are also turning to in silico MLST schemes from whole-genome sequences to assign new sequence types to clinically important isolates while appreciating the value of genome sequences to typing. Additionally, with constantly decreasing sequencing costs, genome-scale

microbial typing studies are becoming more affordable. The analysis of WGS data tends to lead to high statistical confidence (*P* value). However, as indicated above, increasing reports are showing highly significant *P* values for contrasting phylogenetic hypotheses depending on the evolutionary model and inference method used. Additionally, genomes can experience unique evolutionary events (e.g., duplications, translocations, etc.) that can also bias epidemiological and phylogenetic inferences. Therefore, when applying WGS-base typing approaches, emphasizing effect size and biological relevance, rather than the *P* value, may help to alleviate systematic error (Kumar et al., 2011). Similarly, using ABC methods instead of ML functions to estimate population parameters may also help to accommodate complex evolutionary scenarios. The use of genome wide sequence data is a trend that is becoming the new standard in microbial typing; it not only includes generating genomic sequence data with ease but also involves mining hundreds of metagenomes to obtain metagenome-assembled genomes (MAGs). In the following sections, we show current examples of modern use of molecular typing for both epidemiology and phylogenetic inference. Other examples can be found in previous studies cited in the Introduction of this review.

4.1 Molecular Epidemiology

There has been increasing attention paid to tracking and identifying sources of (opportunistic) pathogens in hospital-based settings in the last few years. One such opportunistic pathogen is *Klebsiella pneumoniae*, a human commensal whose hyper-virulent and multidrug resistance members have emerged worldwide. Recently, it was studied the short-term evolution and transmission of *K. pneumoniae* over a 40-week period in a hospital from the United States, finding that isolates showed low intra-host diversity (up to 15 single nucleotide variants) (Yang et al., 2017). This level of resolution proved sufficient for following the transmission chain back to its source (endoscopic device); this was only possible due to the use of WGS, with the added benefit of detecting the *blaCTX* $_{\text{-M-15}}$ gene (involved in Carbapenem resistance) in 27 out of 32 isolates. Other studies compared the power of electrophoresis-based methods to WGS making evident that previously clonal isolates are distinguishable through WGS (Salipante et al., 2015). Similarly, another study presented a 5-year single-institution outbreak investigation revealing the molecular epidemiology of *K. pneumoniae* and where they could follow isolates both by sequence type (in silico determined) and SNP dataset derived from the core genome (Mathers et al., 2015). The authors highlighted the practicality of linking MLST types with antimicrobial resistance determinants and the power of a whole-genome SNP dataset for increased phylogenetic resolution. The combined use of WGS and MLST can provide valuable information regarding origin, clinical phenotype, and potential treatment of nosocomial infectious disease.

K. pneumoniae is one of six pathogens which are leading cause of nosocomial infections throughout the world (Pendleton et al., 2013). Although *K. pneumoniae* is traditionally acquired through nosocomial infection, it has elicited the interest of researchers to unravel potential zoonotic transmission pathways. Previous work has compared *K. pneumoniae* isolated from retail meat from grocery stores and from human urine and blood specimens (both sources from Flagstaff, AZ) (Davis et al., 2015). By combining traditional MLST and WGS, they observed that meat source isolates were more likely multidrug resistant than clinical isolates, even though isolates from both sources shared MLST profiles and were phylogenetically intermingled. Their results suggested potential food-borne transmission routes that carry the risk of spreading multidrug resistance into the general population.

For a review on *K. pneumoniae* population genomics, see Wyres and Holt (2016) and for a review on food animal production and antibiotic resistance, see Silbergeld et al. (2008).

The spread of multidrug resistance (MDR) has also been studied for old foes including the causative agent of Typhoid Fever (Wong et al., 2015). They explored the intra- and intercontinental spread of an H58 MDR *Salmonella* Typhi clade using WGS. Based on a dataset with more than 20 thousand SNPs, the authors showed that multiple transfers from Asia to Africa have occurred and are still occurring and that MDR isolates are replacing drug sensitive isolates. Interestingly, another study by the same group did not find this clade in Nigeria, where multiple introductions could better explain the *Salmonella* genotypes present (Wong et al., 2016). Overall, these and other similar studies highlight the need for unbiased sampling in molecular epidemiology studies, as often studies select isolates for sequencing and typing based on pathogenicity and convenience, which tend to overlook much of the variation needed for informed public health policy decisions (Holt et al., 2008).

Another Enterobacteriaceae that has been studied by modern typing methods is *Escherichia coli*, like in the German outbreak of May–July 2011 (Rohde et al., 2011). Within 24 h, the DNA sequences of *E. coli* infecting patient zero (TY2482) were assembled and MLST genes identified, which allowed researchers to rapidly genotype the causative strain and find the close phylogenetic relationship between this new strain and a previous one reported in 2001 in Germany. Later whole-genome comparisons showed that TY2482 was nearly identical to an African strain that may (or may not) harbor the Shiga toxin gene (Mossoro et al., 2002). While traditional typing approaches pointed to the outbreak strain as being enterohemorrhagic *E. coli*, it was only after more detailed whole-genome inspection that researchers discovered it corresponded to an enteroaggregative *E. coli* harboring two conjugative plasmids, a small plasmid, and a *stx2* prophage (Rohde et al., 2011). Along these same lines, linking outbreaks from different localities has been possible due to the increased resolution that WGS allows. For example, a small *E. coli* outbreak was reported in southwest France in June 2011, which was indistinguishable from the German one by traditional methods. Researchers could only separate the two variants by WGS, revealing that the German outbreak isolates were limited in genetic diversity (2 SNPs from four individuals) compared to the French isolates (19 SNPs from seven individuals) (Grad et al., 2012). Therefore, slow-evolving pathogens or pathogen outbreaks over short periods of time are difficult to type and present challenges for traditional MLST approaches.

Similarly, *Neisseria gonorrhea* represents an extremely slow-evolving pathogen. Scientists studied patients infected with *N. gonorrhea* from diverse UK locations and found that in 26% of the infections, *N. gonorrhea* isolates differed by zero nucleotide substitutions at the genome level; however, in 76% of the infections, contact-tracing demonstrated local, national, and international transmissions (De Silva et al., 2016). Epidemiological and population dynamic patterns have also been inferred from MLST data, where a combination of several housekeeping and hypervariable genes were used for increased resolution at the local level (Pérez-Losada et al., 2005). Other studies have contrasted traditional typing with WGS for populations from different localities but also different epidemiological properties. For instance, one study (Didelot et al., 2016) analyzed 237 isolates from predominantly heterosexual men from populations in Sheffield and London, respectively. Interestingly, all isolates resolved into a single sequence type per population (ST12 and ST225, respectively) by the most widely used tool for *N. gonorrhea* detection, multi-antigen sequence typing (for in silico version see Kwong, Gonçalves da Silva et al. (2016)). In contrast, WGS could resolve relationships among isolates at the intra- and interspecific levels (less than 200 substitutions genome-wide) (Didelot et al., 2016).

Microbial typing is enjoying the wealth of data provided by the affordable acquisition of genomic and metagenomic data. Accordingly, many studies now use metagenomic sequencing coupled with the recovery of single-taxon genome assemblies of reads obtained from a metagenome, a Metagenome-assembled Genome or MAG, to conduct molecular epidemiology studies. For detail discussion on how to obtain MAGs, the reader is referred to Setubal (2021), Meziti et al. (2021).

In a pioneer study, authors investigated the role of *Staphylococcus* and other bacteria of pathogenic potential in psoriasis (Tett et al., 2017). Interestingly, changes in microbial composition between psoriatic sites and unaffected sites showed only an increase in *Staphylococcus*. However, by reconstructing microbial genomes with strain-level resolution, the authors hypothesized that within-species genetic heterogeneity might explain part of the disease phenotype. This was confirmed after a more careful examination of the single-nucleotide differences among members of *S. epidemidis*. For instance, 50 out of 128 single nucleotide substitutions were more prevalent in strains inhabiting psoriatic plaques. In turn, 78 substitutions were more prevalent in *S. epidermidis* strains from unaffected skin. Likewise, *S. epidermidis* strains from psoriatic plaques exhibited virulence genes that were absent from strains from unaffected sites, such as the genes *ssaA* and *esxA*.

Similarly, gene dosage can be a marker for molecular epidemiology. Previous work (Ward et al., 2016) studied shotgun metagenomes of 144 preterm and 22 term infants and interrogated accessory genes and sequence types potentially associated with necrotizing enterocolitis (NEC). Using both a pangenome analysis of *Escherichia coli* and a metaMLST, the authors found that NEC patients were preferentially colonized by uropathogenic *E. coli* (UPEC), usually related to sequence types associated with urinary tract infections.

Frequently, the human microbiome has been studied under a community ecology framework, with little regard for within- and among-host evolutionary processes (Garud et al., 2019). Garud et al. asked whether evolutionary processes, i.e., mutation, recombination, adaptation/selection, are relevant forces shaping the microbiome. Using a panel of 40 prevalent species in the microbiome, they found that over short periods of time (6 months), species accumulate changes rarely attributable to replacement by distantly related strains. In turn, resident strains tend to acquire nucleotide substitutions, gene gains and losses that rapidly increase to high frequency. However, using data from twin studies, the authors inferred that microbial strain replacement overwhelms evolution over decadal time periods. Epidemiological studies of the microbiome, thus, are useful over periods of time that are also relevant for human health.

In another study (Zhao et al., 2019) authors investigated the population dynamics of the very abundant gut microbe *Bacteroides fragilis*. Using a combination of WGS and metagenomics, the authors reveal that *B. fragilis* populations harbor and maintain nucleotide diversity, representing years of within-person history. Additionally, signatures of adaptation were apparent in 16 genes; however, different alleles co-existed in the gene pool over time. Interestingly, *B. fragilis* seemed to colonize individuals only once, as each individual's population was dominated by only one bacterial lineage.

4.2 Genotyping and Phylogenetic Inference

Molecular markers can be used for both genotyping and inferring evolutionary relationships. More comprehensive genotyping frameworks link genetic variation with phylogenetic placement to obtain more information regarding origin and pathogenicity. For pathogens with limited genetic diversity such as *Salmonella* Typhi, new genotyping frameworks have been developed where researchers have

identified genome-wide SNPs that link isolates to geographic source populations (Roumagnac et al., 2006; Wong et al., 2016). Wong et al. used nearly 2000 isolates from over 60 countries to identify 68 phylogenetically informative SNPs. Using this framework, the authors predicted geographic origin at the country level for a subset of novel isolates, paving the way for future developments aimed at increasing accuracy and empowering clinicians and public health officials.

Another group of enteric pathogens, *Shigella* spp., have been studied regarding their evolutionary history and adaptation to human hosts. Four species exist that cause dysentery: *S. sonnei*, *S. flexneri*, *S. boydii*, and *S. dysenteriae*, all of which are phylogenetically nested within the *E. coli* clade. Previous work (Yang et al., 2007) has tested the monophyly and phylogenetic relationships of *Shigella* spp. by using up to 23 housekeeping chromosomal genes from *Shigella* and *E. coli*. Their results supported the hypothesis of multiple independent origins (probably four) of *Shigella* members from diverse *E. coli* strains. This would explain why *Shigella* spp. harbor diverse genomes but a similar phenotype.

In particular, *S. sonnei* is a human pathogen that diverged less than 500 years ago (Holt et al., 2012). The researchers collected samples from four continents and sequenced 132 genomes from *S. sonnei* isolates from 1943 to 2008, which allowed them to detect more than 10,000 SNPs for increased phylogenetic resolution. Interestingly, the authors defined four lineages from the SNP phylogeny that correlated with more traditional typing methods such as Biotypes and CRISPR types (Nastasi et al., 1993; Touchon et al., 2011), suggesting that for *S. sonnei*, traditional typing methods provide sufficient resolution to distinguish lineages even over short periods of time.

Similar approaches have been applied to *S. dysenteriae* type 1 (dysenteriae bacillus), a pathogen responsible for major epidemics during the 20th century (Njamkepo et al., 2016). Njamkepo et al. reconstructed the historical spread and geographic distribution of *S. dysenteriae* by sequencing 331 isolates collected from 1915 to 2011 (66 countries) and found that the global spread of the bacterium predates the First World War and the global expansion of *S. sonnei* (Holt et al., 2012; Njamkepo et al., 2016). While these studies cannot establish causal relationships, the major expansions from Europe to the rest of the world coincide with periods of intense European migration due to colonialism. It is important to note that while traditional typing techniques and WGS were congruent in *S. sonnei*, the opposite was true for *S. dysenteriae*, where single lineages were separated by tens or hundreds of SNPs. See The et al. (2016) for a review of *Shigella* spp. evolution, adaptation, and historical geographic spread.

Phylogenetic relationships of other globally important pathogens have also been elucidated using molecular typing methods. *Vibrio cholerae* epidemics have been characterized by several waves of global transmission with the latest Haiti outbreak belonging to the seventh cholera pandemic (caused by *V. cholerae* El Tor biotype of serogroup O1). In the Fall of 2010, the Haiti outbreak was reported and initial Pulse-Field Gel Electrophoresis (PFGE; a technique used for the separation of large DNA molecules by applying to a gel matrix an electric field that periodically changes direction) typing indicated that different Haiti isolates were indistinguishable. Using WGS, the authors concluded that Haiti isolates were related and were not identical to isolates from India and Cameroon (Reimer et al., 2011). Then, other study (Chin et al., 2011) established an association between Haiti and Bangladesh isolates by means of WGS. More importantly, Chin et al. refuted the hypothesis that the Haiti outbreak was related to indigenous *V. cholerae*, as it was presented by others (Hasan et al., 2012), who reported both *V. cholerae* O1 and non-O1/O139 early in the Haiti cholera epidemic with samples collected from victims of 18 towns of Haiti. Further work (Katz et al., 2013) sequenced longitudinal samples from the Haiti outbreak and concluded that it corresponded to a single source introduction; while others

(Hendriksen et al., 2011) demonstrated that the Nepalese isolates formed a monophyletic group with other isolates from Haiti and Bangladesh, findings that were consistent with PFGE patterns and antibiotic susceptibility tests. We refer the reader to other reviews on the Haiti outbreak (Frerichs et al., 2012; Orata et al., 2014).

As *V. cholerae* remains a highly relevant human pathogen, the development of higher resolution tools and more comprehensive sampling will aid researchers in establishing global routes of transmission and large-scale patterns of gene flow. For instance, some researchers (Mutreja et al., 2011) collected global genomic data on *V. cholerae* and developed high-resolution markers (SNPs) to genotype *V. cholerae* lineages. Using phylogenetic inference and SNP markers, they showed that the seventh *V. cholerae* pandemic has spread from the Bay of Bengal on multiple overlapping waves, some of which reached such faraway places as Haiti.

Investigating the origin of pathogenic populations is also important in health care settings. Phylogenetic approaches can pinpoint to reservoir and stable populations of hospital-acquired (nosocomial) pathogens and monitor them over time. A recent study (Gorrie et al., 2022) conducted a year-long study of *Klebsiella pneumoniae* in hospital patients and found that most infections corresponded to opportunistic infections rather than nosocomial (only 10% of patients). Interestingly, nosocomial onset was associated with extended spectrum betalactamases, which in turn, were indicative of other antimicrobial genes. A phylogenetic analysis showed that strains of *K. pneumoniae* came from different lineages and that those lineages shared a common ancestor over many years of evolution. Sequence type assignment also correlated with those lineages, giving support to the hypothesis of a mainly opportunistic origin of *K. pneumoniae* infections. Conversely, others (Chng et al., 2020) found that microbial communities, when followed by up to 1.5 years in a tertiary hospital, show remarkable stable patterns of colonization and persistence. Using phylogenetic inference, the authors found that multidrug-resistant strains were stable and widely distributed. Moreover, when comparing metagenome-derived microbial genomes with archival isolate genomes, the authors concluded that many microbes could persist in the hospital environment for extended periods of time (>8 years). Microbial genotyping and phylogenetic inference are key to assess whether microbial species have similar origins in the environment, built or natural, and to assess the potential impacts of those origins on human health.

5. Conclusions and Prospects

MLST is a flexible approach for characterizing bacteria and some eukaryotes. It has become a staple mainly due to the existence of comprehensive databases and its broad implementation in clinical practice and molecular diagnostics. MLST is widely used in basic research labs (PCR + Sanger) and core sequencing facilities performing genome sequencing. MLST has broadened its basic scheme from using housekeeping genes to incorporate more and new molecular markers, such as ribosomal proteins (rMLST) and polymorphic repeated sequences (MLVA); most recently, it has begun integrating draft and full genomes (cgMLST).

Over the last years, whole-genome sequencing (WGS) has emerged as a powerful technology for microbial sequence typing and is increasingly applied in clinical and public health microbiology. New MLST-genome strategies that take advantage of NGS technology to accelerate and automate WGS will expand the power and versatility of DNA typing. Such strategies will also allow for calculations of

more accurate and robust estimates of phylogenies and population genetic parameters under more complex (realistic) statistical models using, for example, Bayesian phylogenomics and approximate Bayesian computation. Those statistical frameworks will also integrate epidemiologic and geographic information, allowing the estimation of the spatial and temporal dynamics of pathogens. Several examples of WGS typing have already been published assessing the time of emergence, origin, and dissemination of pathogen outbreaks and the spread of antibiotic resistance. Moreover, with the rapid development of both genome-resolved and strain-level metagenomic analysis of microbial communities, researchers now can access microbial diversity from the uncultured world, including "microbial dark matter" (Jiao et al., 2021). Studies will increasingly rely on unbiased metagenomic sequencing for answering population genomics, phylogenetic, and epidemiological questions.

While the exploitation of WGS for typing microbial diversity presents great opportunities, it also brings major challenges; some of these challenges are computational, such as genome data analysis, sharing and storage; but others are conceptual, such as relating WGS data to typing and microbial taxonomy. The genomic era opens the door to new types of holistic microbiology research, i.e., a systems biology (a.k.a. ecology) framework, in which taxonomic, epidemiological, and evolutionary information are integrated with other *Omic* information from both the microbial communities and the host. In coming years, one can only assume that classical or expanded forms of MLST will remain a key component of the microbial genomicist's toolkit used for understanding the diversity and dynamics of infectious diseases.

Acknowledgments

MP-L was funded by a DC D-CFAR Research Award from the District of Columbia Developmental Center for AIDS Research (P30AI087714), by a University Facilitating Fund award from George Washington University, the Margaret Q. Landenberger Research Foundation, a K12 Career Development Program 5 K12 HL119994 award, and by the FCT under the Programa Operacional Potencial Humano—Quadro de Referência Estratégico Nacional funds from the European Social Fund and Portuguese Ministério da Educação e Ciência IF/00764/2013. MA was supported by the "Ramón y Cajal" grant RYC-2015-18241 from the Spanish Government. ECN was funded by "ANID-FONDECYT Regular 1240871" and by "ANID-PIA-Anillo INACH ACT192057."

References

Aanensen, D.M., Spratt, B.G., 2005. The multilocus sequence typing network: mlst.net. Nucleic Acids Res. 33 (2), W728−W733. https://doi.org/10.1093/nar/gki415.

Abascal, F., Zardoya, R., Telford, M.J., 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38 (2), W7−W13. https://doi.org/10.1093/nar/gkq291.

Anisimova, M., Nielsen, R., Yang, Z., 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164 (3), 1229−1236. https://doi.org/10.1093/genetics/164.3.1229.

Arenas, M., 2022. ProteinEvolverABC: coestimation of recombination and substitution rates in protein sequences by approximate Bayesian computation. Bioinformatics 38 (1), 58−64. https://doi.org/10.1093/bioinformatics/btab617.

Arenas, M., Posada, D., 2010a. Coalescent simulation of intracodon recombination. Genetics 184 (2), 429−437. https://doi.org/10.1534/genetics.109.109736.

Arenas, M., Posada, D., 2010b. The effect of recombination on the reconstruction of ancestral sequences. Genetics 184 (4), 1133−1139. https://doi.org/10.1534/genetics.109.113423.

Arenas, M., Posada, D., 2014a. The influence of recombination on the estimation of selection from coding sequence alignments.. In: Natural Selection: Methods and Applications. CRC Press, pp. 112−125.

Arenas, M., Posada, D., 2014b. Simulation of genome-wide evolution under heterogeneous substitution models and complex multispecies coalescent histories. Mol. Biol. Evol. 31 (5), 1295−1301. https://doi.org/10.1093/molbev/msu078.

Arenas, M., et al., 2015. CodABC: a computational framework to coestimate recombination, substitution, and molecular adaptation rates by approximate Bayesian computation. Mol. Biol. Evol. 32 (4), 1109−1112. https://doi.org/10.1093/molbev/msu411.

Arenas, M., et al., 2018. Mutation and recombination in pathogen evolution: relevance, methods and controversies. Infect. Genet. Evol. 63, 295−306. https://doi.org/10.1016/j.meegid.2017.09.029.

Arbiza, L., et al., 2011. Genome-wide heterogeneity of nucleotide substitution model fit. Genome Biol. Evol. 3 (1), 896−908. https://doi.org/10.1093/gbe/evr080.

Arenas, M., 2013. The importance and application of the ancestral recombination graph. Front. Genet. 4. https://doi.org/10.3389/fgene.2013.00206.

Arenas, M., 2015a. Advances in computer simulation of genome evolution: toward more realistic evolutionary genomics analysis by approximate bayesian computation. J. Mol. Evol. 80 (3−4), 189−192. https://doi.org/10.1007/s00239-015-9673-0.

Arenas, M., 2015b. Trends in substitution models of molecular evolution. Front. Genet. 6. https://doi.org/10.3389/fgene.2015.00319.

Bandelt, H.J., Dress, A.W.M., 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Mol. Phylogenet. Evol. 1 (3), 242−252. https://doi.org/10.1016/1055-7903(92)90021-8.

Barbato, M., et al., 2015. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Front. Genet. 6. https://doi.org/10.3389/fgene.2015.00109.

Bazinet, A.L., Zwickl, D.J., Cummings, M.P., 2014. A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. Syst. Biol. 63 (5), 812−818. https://doi.org/10.1093/sysbio/syu031.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162 (4), 2025−2035.

Bello, G., et al., 2007. Lack of temporal structure in the short term HIV-1 evolution within asymptomatic naïve patients. Virology 362 (2), 294−303. https://doi.org/10.1016/j.virol.2006.11.039.

Bertels, F., et al., 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. Mol. Biol. Evol. 31 (5), 1077−1088. https://doi.org/10.1093/molbev/msu088.

Bezdicek, M., et al., 2021. Rapid high-resolution melting genotyping scheme for *Escherichia coli* based on MLST derived single nucleotide polymorphisms. Sci. Rep. 11 (1). https://doi.org/10.1038/s41598-021-96148-3.

Bielejec, F., et al., 2011. SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. Bioinformatics 27 (20), 2910−2912. https://doi.org/10.1093/bioinformatics/btr481.

Boers, S.A., et al., 2012. High-throughput multilocus sequence typing: bringing molecular typing to the next level. PLoS One 7 (7). https://doi.org/10.1371/journal.pone.0039630.

Boitard, S., et al., 2016. Inferring population size history from large samples of genome-wide molecular data - an approximate bayesian computation approach. PLoS Genet. 12 (3). https://doi.org/10.1371/journal.pgen.1005877.

Bouckaert, R.R., Drummond, A.J., 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. 17 (1), 1−11. https://doi.org/10.1186/s12862-017-0890-6.

Bouckaert, R., et al., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10 (4). https://doi.org/10.1371/journal.pcbi.1003537.

Bruen, T.C., Philippe, H., Bryant, D., 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172 (4), 2665−2681. https://doi.org/10.1534/genetics.105.048975.

Bryant, D., Moulton, V., 2004. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. 21 (2), 255−265. https://doi.org/10.1093/molbev/msh018.

Bryant, D., et al., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29 (8), 1917−1932. https://doi.org/10.1093/molbev/mss086.

Buckley, T.R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51 (3), 509−523. https://doi.org/10.1080/10635150290069922.

Caporaso, J.G., et al., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6 (8), 1621−1624. https://doi.org/10.1038/ismej.2012.8.

Castelhano, N., Araujo, N.M., Arenas, M., 2017. Heterogeneous recombination among hepatitis B virus genotypes. Infect. Genet. Evol. 54, 486−490. https://doi.org/10.1016/j.meegid.2017.08.015.

Castro-Nallar, E., et al., 2015. Concordance and discordance of sequence survey methods for molecular epidemiology. PeerJ 2015 (2). https://doi.org/10.7717/peerj.761.

Chan, M.S., Maiden, M.C.J., Spratt, B.G., 2001. Database-driven multi locus sequence typing (MLST) of bacterial pathogens. Bioinformatics 17 (11), 1077−1083. https://doi.org/10.1093/bioinformatics/17.11.1077.

Chang, H.W., et al., 2009. Seq-SNPing: multiple-alignment tool for SNP discovery, SNP ID identification, and RFLP genotyping. OMICS A J. Integr. Biol. 13 (3), 253−260. https://doi.org/10.1089/omi.2008.0058.

Charlesworth, B., Nordborg, M., Charlesworth, D., 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. 70 (2), 155−174. https://doi.org/10.1017/s0016672397002954.

Chen, C., et al., 2013. Minimum core genome sequence typing of bacterial pathogens: a unified approach for clinical and public health microbiology. J. Clin. Microbiol. 51 (8), 2582−2591. https://doi.org/10.1128/JCM.00535-13.

Chen, Y., et al., 2015. Next generation multilocus sequence typing (NGMLST) and the analytical software program MLSTEZ enable efficient, cost-effective, high-throughput, multilocus sequencing typing. Fungal Genet. Biol. 75, 64−71. https://doi.org/10.1016/j.fgb.2015.01.005.

Chin, C.S., et al., 2011. The origin of the Haitian cholera outbreak strain. N. Engl. J. Med. 364 (1), 33−42. https://doi.org/10.1056/NEJMoa1012928.

Chng, K.R., et al., 2020. Cartography of opportunistic pathogens and antibiotic resistance genes in a tertiary hospital environment. Nat. Med. 26 (6), 941−951. https://doi.org/10.1038/s41591-020-0894-4.

Chun, J., Rainey, F.A., 2014. Integrating genomics into the taxonomy and systematics of the Bacteria and Archaea. Int. J. Syst. Evol. Microbiol. 64 (2), 316−324. https://doi.org/10.1099/ijs.0.054171-0.

Coipan, C.E., et al., 2020. Concordance of SNP- and allele-based typing workflows in the context of a large-scale international *Salmonella* Enteritidis outbreak investigation. Microb. Genom. https://doi.org/10.1099/mgen.0.000318.

Cooper, J.E., Feil, E.J., 2004. Multilocus sequence typing - what is resolved? Trends Microbiol. 12 (8), 373−377. https://doi.org/10.1016/j.tim.2004.06.003.

Cornuet, J.M., et al., 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. Bioinformatics 30 (8), 1187−1189. https://doi.org/10.1093/bioinformatics/btt763.

Currat, M., et al., 2019. SPLATCHE3: simulation of serial genetic data under spatially explicit evolutionary scenarios including long-distance dispersal. Bioinformatics 35 (21), 4480−4483. https://doi.org/10.1093/bioinformatics/btz311.

Dallman, T.J., et al., 2015. Whole-genome sequencing for national surveillance of Shiga toxin-producing *Escherichia coli* O157. Clin. Infect. Dis. 61 (3), 305−312. https://doi.org/10.1093/cid/civ318.

Dalquen, D.A., et al., 2012. ALF-A simulation framework for genome evolution. Mol. Biol. Evol. 29 (4), 1115–1123. https://doi.org/10.1093/molbev/msr268.

Darriba, D., et al., 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. 37 (1), 291–294. https://doi.org/10.1093/molbev/msz189.

Davis, G.S., et al., 2015. Intermingled *Klebsiella* pneumoniae populations between retail meats and human urinary tract infections. Clin. Infect. Dis. 61 (6), 892–899. https://doi.org/10.1093/cid/civ428.

de Been, M., et al., 2015. Core genome multilocus sequence typing scheme for high-resolution typing of Enterococcus faecium. J. Clin. Microbiol. 53 (12), 3788–3797. https://doi.org/10.1128/jcm.01946-15.

De Bruyn, A., Martin, D.P., Lefeuvre, P., 2014. Phylogenetic reconstruction methods: an overview. Methods Mol. Biol. 1115, 257–277. https://doi.org/10.1007/978-1-62703-767-9_13.

de Oliveira Martins, L., Posada, D., 2017. Species tree estimation from genome-wide data with guenomu. In: Methods in Molecular Biology. Humana Press Inc., pp. 461–478

De Silva, D., et al., 2016. Whole-genome sequencing to determine transmission of *Neisseria gonorrhoeae*: an observational study. Lancet Infect. Dis. 16 (11), 1295–1303. https://doi.org/10.1016/S1473-3099(16)30157-8.

Degiorgio, M., Degnan, J.H., 2010. Fast and consistent estimation of species trees using supermatrix rooted triples. Mol. Biol. Evol. 27 (3), 552–569. https://doi.org/10.1093/molbev/msp250.

Delahaye, C., Nicolas, J., Andrés-León, E., 2021. Sequencing DNA with nanopores: troubles and biases. PLoS One 16 (10). https://doi.org/10.1371/journal.pone.0257521.

den Bakker, H.C., et al., 2011. A whole-genome single nucleotide polymorphism-based approach to trace and identify outbreaks linked to a common *Salmonella enterica* subsp. enterica serovar Montevideo pulsed-field gel electrophoresis type. Appl. Environ. Microbiol. 77 (24). https://doi.org/10.1128/AEM.06538-11.

den Bakker, H.C., et al., 2014. Rapid whole-genome sequencing for surveillance of *Salmonella enterica* serovar Enteritidis. Emerg. Infect. Dis. 20 (8), 1306–1314. https://doi.org/10.3201/eid2008.131399.

Deorowicz, S., Debudaj-Grabysz, A., Gudys, A., 2016. FAMSA: fast and accurate multiple sequence alignment of huge protein families. Sci. Rep. 6. https://doi.org/10.1038/srep33964.

Didelot, X., Falush, D., 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics 175 (3), 1251–1266. https://doi.org/10.1534/genetics.106.063305.

Didelot, X., et al., 2012. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. Genome Biol. 13 (12). https://doi.org/10.1186/gb-2012-13-12-r118.

Didelot, X., et al., 2016. Genomic analysis and comparison of two gonorrhea outbreaks. mBio 7 (3). https://doi.org/10.1128/mBio.00525-16.

Do, K.-T., et al., 2014. Estimation of effective population size using single-nucleotide polymorphism (SNP) data in Jeju horse. J. Anim. Sci. Technol. 56 (1). https://doi.org/10.1186/2055-0391-56-28.

Dorp, et al., 2020. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. Nat. Commun. 11 (1). https://doi.org/10.1038/s41467-020-19818-2. PubMed PMID: 33239633.

dos Reis, Mario, D., Philip, C.J., Yang, Z., 2016. Bayesian molecular clock dating of species divergences in the genomics era. Nat. Rev. Genet. 17 (2), 71–80. https://doi.org/10.1038/nrg.2015.8.

Drummond, A.J., et al., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4 (5). https://doi.org/10.1371/journal.pbio.0040088.

Drummond, A.J., et al., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29 (8), 1969–1973. https://doi.org/10.1093/molbev/mss075.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797. https://doi.org/10.1093/nar/gkh340.

Enright, M.C., Spratt, B.G., 1999. Multilocus sequence typing. Trends Microbiol. 7 (12), 482–487. https://doi.org/10.1016/S0966-842X(99)01609-1.

Excoffier, L., et al., 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics 37 (24), 4882–4885. https://doi.org/10.1093/bioinformatics/btab468.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Res. 10 (3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x.

Feil, E.J., Enright, M.C., 2004. Analyses of clonality and the evolution of bacterial pathogens. Curr. Opin. Microbiol. 7 (3), 308–313. https://doi.org/10.1016/j.mib.2004.04.002.

Feil, E.J., et al., 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 186 (5), 1518–1530. https://doi.org/10.1128/JB.186.5.1518-1530.2004.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17 (6), 368–376. https://doi.org/10.1007/bf01734359.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39 (4), 783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x.

Fitch, W.M., 1997. Networks and viral evolution. J. Mol. Evol. 44 (1), S65–S75. https://doi.org/10.1007/pl00000059.

Foley, S.L., Lynne, A.M., Nayak, R., 2009. Molecular typing methodologies for microbial source tracking and epidemiological investigations of Gram-negative bacterial foodborne pathogens. Infect. Genet. Evol. 9 (4), 430–440. https://doi.org/10.1016/j.meegid.2009.03.004.

Foox, J., Tighe, S.W., et al., 2021a. Author correction: performance assessment of DNA sequencing platforms in the ABRF next-generation sequencing study. Nat. Biotechnol. 39 (11), 1466. https://doi.org/10.1038/s41587-021-01122-z.

Foox, J., Tighe, S.W., et al., 2021b. Performance assessment of DNA sequencing platforms in the ABRF next-generation sequencing study. Nat. Biotechnol. 39 (9), 1129–1140. https://doi.org/10.1038/s41587-021-01049-5.

Forde, B.M., O'Toole, P.W., 2013. Next-generation sequencing technologies and their impact on microbial genomics. Brief. Funct. Genomics 12 (5), 440–453. https://doi.org/10.1093/bfgp/els062.

Francisco, A.P., et al., 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinf. 10. https://doi.org/10.1186/1471-2105-10-152.

Francisco, A.P., et al., 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinf. 13 (1). https://doi.org/10.1186/1471-2105-13-87.

Frerichs, R.R., et al., 2012. Nepalese origin of cholera epidemic in Haiti. Clin. Microbiol. Infection 18 (6), E158–E163. https://doi.org/10.1111/j.1469-0691.2012.03841.x.

Galtier, N., Daubin, V., 2008. Dealing with incongruence in phylogenomic analyses. Phil. Trans. Biol. Sci. 363 (1512), 4023–4029. https://doi.org/10.1098/rstb.2008.0144.

Gardner, S.N., et al., 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics 31 (17), 2877.

Gardy, J.L., et al., 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N. Engl. J. Med. 364 (8), 730–739. https://doi.org/10.1056/NEJMoa1003176.

Garud, N.R., et al., 2019. Evolutionary dynamics of bacteria in the gut microbiome within and across hosts. PLoS Biol. 17 (1). https://doi.org/10.1371/journal.pbio.3000102.

Glenn, T.C., 2011. Field guide to next-generation DNA sequencers. Mol. Ecol. Res. 11 (5), 759–769. https://doi.org/10.1111/j.1755-0998.2011.03024.x.

Goodwin, S., McPherson, J.D., McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. 17 (6), 333–351. https://doi.org/10.1038/nrg.2016.49.

Gorrie, C.L., et al., 2022. Genomic dissection of *Klebsiella* pneumoniae infections in hospital patients reveals insights into an opportunistic pathogen. Nat. Commun. 13 (1). https://doi.org/10.1038/s41467-022-30717-6.

Grad, Y.H., et al., 2012. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. Proc. Natl. Acad. Sci. U.S.A. 109 (8), 3065−3070. https://doi.org/10.1073/pnas.1121491109.

Gray, A.N., et al., 2015. High-throughput bacterial functional genomics in the sequencing era. Curr. Opin. Microbiol. 27, 86−95. https://doi.org/10.1016/j.mib.2015.07.012.

Gurjar, M.S., et al., 2021. Multilocus sequence typing and single nucleotide polymorphism analysis in *Tilletia indica* isolates inciting karnal bunt of wheat. J. Fungi 7 (2), 1−10. https://doi.org/10.3390/jof7020103.

Hall, B.G., Ehrlich, G.D., Hu, F.Z., 2010. Pan-genome analysis provides much higher strain typing resolution than multi-locus sequence typing. Microbiology 156 (4), 1060−1068. https://doi.org/10.1099/mic.0.035188-0.

Hasan, N.A., et al., 2012. Genomic diversity of 2010 Haitian cholera outbreak strains. Proc. Natl. Acad. Sci. U.S.A. 109 (29), E2010−E2017. https://doi.org/10.1073/pnas.1207359109.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27 (3), 570−580. https://doi.org/10.1093/molbev/msp274.

Heled, J., Drummond, A.J., 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Syst. Biol. 61 (1), 138−149. https://doi.org/10.1093/sysbio/syr087.

Hendriksen, R.S., et al., 2011. Population genetics of vibrio cholerae from Nepal in 2010: evidence on the origin of the Haitian outbreak. mBio 2 (4). https://doi.org/10.1128/mBio.00157-11.

Hill, W.G., 1981. Estimation of effective population size from data on linkage disequilibrium. Genet. Res. 38 (3), 209−216. https://doi.org/10.1017/S0016672300020553.

Ho, S.Y.W., Duchêne, S., 2014. Molecular-clock methods for estimating evolutionary rates and timescales. Mol. Ecol. 23 (24), 5947−5965. https://doi.org/10.1111/mec.12953.

Höhna, S., et al., 2016. RevBayes: bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst. Biol. 65 (4), 726−736. https://doi.org/10.1093/sysbio/syw021.

Holt, K.E., et al., 2008. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. Nat. Genet. 40 (8), 987−993. https://doi.org/10.1038/ng.195.

Holt, K.E., et al., 2012. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat. Genet. 44 (9), 1056−1059. https://doi.org/10.1038/ng.2369.

Hubisz, M., Siepel, A., 2020. Inference of ancestral recombination graphs using ARGweaver. In: Methods in Molecular Biology. Humana Press Inc., pp. 231−266

Huelsenbeck, J.P., et al., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294 (5550), 2310−2314. https://doi.org/10.1126/science.1065889.

Huijsmans, C.J.J., et al., 2011. Single-nucleotide-polymorphism genotyping of *Coxiella burnetii* during a Q fever outbreak in The Netherlands. Appl. Environ. Microbiol. 77 (6), 2051−2057. https://doi.org/10.1128/AEM.02293-10.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23 (2), 254−267. https://doi.org/10.1093/molbev/msj030.

Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14 (1), 68−73. https://doi.org/10.1093/bioinformatics/14.1.68.

Jefferies, J., et al., 2003. Automated pneumococcal MLST using liquid-handling robotics and a capillary DNA sequencer. Appl. Biochem. Biotechnol. Part B Mol. Biotechnol. 24 (3), 303−307. https://doi.org/10.1385/MB:24:3:303.

Jeffroy, O., et al., 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22 (4), 225−231. https://doi.org/10.1016/j.tig.2006.02.003.

Jiao, J.Y., et al., 2021. Microbial dark matter coming to light: challenges and opportunities. Natl. Sci. Rev. 8 (3). https://doi.org/10.1093/nsr/nwaa280.

Jolley, K.A., Maiden, M.C., 2006. AgdbNet−antigen sequence database software for bacterial typing. BMC Bioinf. 7 (1).

Jolley, K.A., Maiden, M.C.J., 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinf. 11. https://doi.org/10.1186/1471-2105-11-595.

Jolley, K.A., Maiden, M.C.J., 2014. Using MLST to study bacterial variation: prospects in the genomic era. Future Microbiol. 9 (5), 623–630. https://doi.org/10.2217/fmb.14.24.

Jolley, K.A., et al., 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17 (12), 1230–1231. https://doi.org/10.1093/bioinformatics/17.12.1230.

Jolley, K.A., Chan, M.S., Maiden, M.C.J., 2004. mlstdbNet - distributed multi-locus sequence typing (MLST) databases. BMC Bioinf. 5. https://doi.org/10.1186/1471-2105-5-86.

Jolley, K.A., et al., 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology 158 (4), 1005–1015. https://doi.org/10.1099/mic.0.055459-0.

Karst, S.M., et al., 2021. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nat. Methods 18 (2), 165–169. https://doi.org/10.1038/s41592-020-01041-y.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. 20 (4), 1160–1166. https://doi.org/10.1093/bib/bbx108.

Katz, L.S., et al., 2013. Evolutionary dynamics of *Vibrio cholerae* O1 following a single-source introduction to Haiti. mBio 4 (4). https://doi.org/10.1128/mBio.00398-13.

Keim, P.S., Wagner, D.M., 2009. Humans, evolutionary and ecologic forces shaped the phylogeography of recently emerged diseases. Nat. Rev. Microbiol. 7 (11).

Kingman, J.F.C., 1982. The coalescent. Stoch. Process. Appl. 13 (3), 235–248. https://doi.org/10.1016/0304-4149(82)90011-4.

Kobras, C.M., Fenton, A.K., Sheppard, S.K., 2021. Next-generation microbiology: from comparative genomics to gene function. Genome Biol. 22 (1). https://doi.org/10.1186/s13059-021-02344-9.

Kolb, A.W., et al., 2017. Phylogenetic and recombination analysis of the *herpesvirus* genus *varicellovirus*. BMC Genom. 18 (1). https://doi.org/10.1186/s12864-017-4283-4.

Kosakovsky Pond, S.L., Frost, S.D.W., 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22 (5), 1208–1222. https://doi.org/10.1093/molbev/msi105.

Kosakovsky Pond, S.L., Frost, S.D.W., Muse, S.V., 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21 (5), 676–679. https://doi.org/10.1093/bioinformatics/bti079.

Kosakovsky Pond, S.L., et al., 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22 (24), 3096–3098. https://doi.org/10.1093/bioinformatics/btl474.

Kozich, J.J., et al., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79 (17), 5112–5120. https://doi.org/10.1128/AEM.01043-13.

Kozlov, A.M., et al., 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35 (21), 4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

Kuhner, M.K., 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22 (6), 768–770. https://doi.org/10.1093/bioinformatics/btk051.

Kumar, S., et al., 2011. Statistics and truth in phylogenomics. Mol. Biol. Evol. 29 (2), 457–472.

Kuroda, M., et al., 2010. Genome-wide single nucleotide polymorphism typing method for identification of Bacillus anthracis species and strains among B. Cereus group species. J. Clin. Microbiol. 48 (8), 2821–2829. https://doi.org/10.1128/jcm.00137-10.

Kwong, J.C., et al., 2015. Whole genome sequencing in clinical and public health microbiology. Pathology 47 (3), 199–210. https://doi.org/10.1097/pat.0000000000000235.

Kwong, J.C., Gonçalves da Silva, A., et al., 2016a. NGMASTER: in silico multi-antigen sequence typing for *Neisseria* gonorrhoeae. Microb. Genom. 2 (8). https://doi.org/10.1099/mgen.0.000076.

Kwong, J.C., Mercoulia, K., et al., 2016b. Prospective whole-genome sequencing enhances national surveillance of listeria monocytogenes. J. Clin. Microbiol. 54 (2), 333−342. https://doi.org/10.1128/JCM.02344-15.

Kyrpides, N.C., et al., 2014. Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. PLoS Biol. 12 (8). https://doi.org/10.1371/journal.pbio.1001920.

Labbé, F., et al., 2017. Genetic signatures of variation in population size in a native fungal pathogen after the recent massive plantation of its host tree. Heredity 119 (6), 402−410. https://doi.org/10.1038/hdy.2017.58.

Lanfear, R., et al., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34 (3). https://doi.org/10.1093/molbev/msw260.

Larget, B.R., et al., 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26 (22), 2910−2911. https://doi.org/10.1093/bioinformatics/btq539.

Larsen, M.V., et al., 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J. Clin. Microbiol. 50 (4), 1355−1361. https://doi.org/10.1128/JCM.06094-11.

Lartillot, N., Philippe, H., 2004. A bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21 (6), 1095−1109. https://doi.org/10.1093/molbev/msh112.

Lartillot, N., et al., 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. Syst. Biol. 62 (4), 611−615. https://doi.org/10.1093/sysbio/syt022.

Lee, T.H., et al., 2014. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genom. 15 (1). https://doi.org/10.1186/1471-2164-15-162.

Leekitcharoenphon, P., et al., 2012. Genomic variation in *Salmonella enterica* core genes for epidemiological typing. BMC Genom. 13 (1). https://doi.org/10.1186/1471-2164-13-88.

Lefort, V., et al., 2017. SMS: smart model selection in PhyML. Mol. Biol. Evol. 34 (9), 2422−2424. https://doi.org/10.1093/molbev/msx149.

Leigh, J.W., et al., 2011. Evaluating phylogenetic congruence in the post-genomic era. Genome Biol. Evol. 3 (1), 571−587. https://doi.org/10.1093/gbe/evr050.

Lemey, P., et al., 2020. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. Nat. Commun. 11 (1). https://doi.org/10.1038/s41467-020-18877-9.

Lemmon, A.R., Moriarty, E.C., 2004. The importance of proper model assumption in bayesian phylogenetics. Syst. Biol. 53 (2), 265−277. https://doi.org/10.1080/10635150490423520.

Lewis, P.O., 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50 (6), 913−925. https://doi.org/10.1080/106351501753462876.

Liu, K., et al., 2011. RAxML and FastTree: comparing two methods for large-scale maximum likelihood phylogeny estimation. PLoS One 6 (11). https://doi.org/10.1371/journal.pone.0027731.

Liu, L., 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24 (21), 2542−2543. https://doi.org/10.1093/bioinformatics/btn484.

Lopes, J.S., et al., 2014. Coestimation of recombination, substitution and molecular adaptation rates by approximate Bayesian computation. Heredity 112 (3), 255−264. https://doi.org/10.1038/hdy.2013.101.

Luo, A., et al., 2010. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. BMC Evol. Biol. 10 (1). https://doi.org/10.1186/1471-2148-10-242.

MacCannell, D., 2013. Bacterial strain typing. Clin. Lab. Med. 33 (3), 629−650. https://doi.org/10.1016/j.cll.2013.03.005.

Maiden, M.C.J., et al., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U.S.A. 95 (6), 3140−3145. https://doi.org/10.1073/pnas.95.6.3140.

Maiden, M.C.J., et al., 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat. Rev. Microbiol. 11 (10), 728−736. https://doi.org/10.1038/nrmicro3093.

Maiden, M.C.J., 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. 60 (1), 561−588. https://doi.org/10.1146/annurev.micro.59.030804.121325.

Mallo, D., Sánchez-Cobos, A., Arenas, M., 2015. Diverse considerations for successful phylogenetic tree reconstruction: impacts from model misspecification, recombination, homoplasy, and pattern recognition. In: Pattern Recognition in Computational Molecular Biology: Techniques and Approaches. wiley, pp. 439−456.

Mardis, E.R., 2008. Next-generation DNA sequencing methods. Annu. Rev. Genom. Hum. Genet. 9, 387−402. https://doi.org/10.1146/annurev.genom.9.081307.164359.

Mardis, E.R., 2013. Next-generation sequencing platforms. Annu. Rev. Anal. Chem. 6, 287−303. https://doi.org/10.1146/annurev-anchem-062012-092628.

Margulies, M., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437 (7057), 376−380. https://doi.org/10.1038/nature03959.

Marjoram, P., et al., 2003. Markov chain Monte Carlo without likelihoods. Proc. Natl. Acad. Sci. U.S.A. 100 (26), 15324−15328. https://doi.org/10.1073/pnas.0306899100.

Marsh, J.W., et al., 2010. Multilocus variable-number tandem-repeat analysis and multilocus sequence typing reveal genetic relationships among *Clostridium difficile* isolates genotyped by restriction endonuclease analysis. J. Clin. Microbiol. 48 (2), 412−418. https://doi.org/10.1128/JCM.01315-09.

Martin, D.P., Lemey, P., Posada, D., 2011. Analysing recombination in nucleotide sequences. Mol. Ecol. Res. 11 (6), 943−955. https://doi.org/10.1111/j.1755-0998.2011.03026.x.

Martin, D.P., et al., 2021. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. Virus Evol. 7 (1). https://doi.org/10.1093/ve/veaa087.

Mathers, A.J., et al., 2015. *Klebsiella pneumoniae* carbapenemase (KPC)-producing *K. pneumoniae* at a single institution: insights into endemicity from Whole-genome sequencing. Antimicrob. Agents Chemother. 59 (3), 1656−1663. https://doi.org/10.1128/AAC.04292-14.

McVean, G.A.T., et al., 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304 (5670), 581−584. https://doi.org/10.1126/science.1092500.

Metzker, M.L., 2010. Sequencing technologies the next generation. Nat. Rev. Genet. 11 (1), 31−46. https://doi.org/10.1038/nrg2626.

Meziti, A., et al., 2021. The reliability of metagenome-assembled genomes (MAGs) in representing natural populations: insights from comparing MAGs against isolate genomes derived from the same fecal sample. Appl. Environ. Microbiol. 87 (6), 1−15. https://doi.org/10.1128/AEM.02593-20.

Minin, V.N., Bloomquist, E.W., Suchard, M.A., 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25 (7), 1459−1471. https://doi.org/10.1093/molbev/msn090.

Miro, E., et al., 2020. Core/whole genome multilocus sequence typing and core genome SNP-based typing of OXA-48-producing *Klebsiella pneumoniae* clinical isolates from Spain. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02961.

Mossoro, C., et al., 2002. Chronic diarrhea, hemorrhagic colitis, and hemolytic-uremic syndrome associated with HEp-2 adherent *Escherichia coli* in adults infected with human immunodeficiency virus in Bangui, Central African Republic. J. Clin. Microbiol. 40 (8), 3086−3088. https://doi.org/10.1128/JCM.40.8.3086-3088.2002.

Mukherjee, S., et al., 2017. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. Nucleic Acids Res. 45 (D1), D446−D456. https://doi.org/10.1093/nar/gkw992.

Mutreja, A., et al., 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. Nature 477 (7365), 462−465. https://doi.org/10.1038/nature10392.

Nastasi, A., et al., 1993. rRNA gene restriction patterns and biotypes of *Shigella sonnei*. Epidemiol. Infect. 110 (1), 23−30. https://doi.org/10.1017/s0950268800050640.

Navascués, M., Depaulis, F., Emerson, B.C., 2010. Combining contemporary and ancient DNA in population genetic and phylogeographical studies. Mol. Ecol. Res. 10 (5), 760−772. https://doi.org/10.1111/j.1755-0998.2010.02895.x.

Nelson, C.L., et al., 2014. A genome-wide association study of variants associated with acquisition of *Staphylococcus aureus* bacteremia in a healthcare setting. BMC Infect. Dis. 14 (1). https://doi.org/10.1186/1471-2334-14-83.

Nguyen, L.T., et al., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32 (1), 268−274. https://doi.org/10.1093/molbev/msu300.

Nielsen, R., et al., 2011. Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12 (6), 443−451. https://doi.org/10.1038/nrg2986.

Njamkepo, E., et al., 2016. Global phylogeography and evolutionary history of *Shigella dysenteriae* type 1. Nat. Microbiol. 1 (4). https://doi.org/10.1038/nmicrobiol.2016.27.

Novozhilov, A.S., Karev, G.P., Koonin, E.V., 2006. Biological applications of the theory of birth-and-death processes. Briefings Bioinf. 7 (1), 70−85. https://doi.org/10.1093/bib/bbk006.

Oaks, J.R., 2019. Full Bayesian comparative phylogeography from genomic data. Syst. Biol. 68 (3), 371−395. https://doi.org/10.1093/sysbio/syy063.

Ogilvie, H.A., et al., 2016. Computational performance and statistical accuracy of BEAST and comparisons with other methods. Syst. Biol. 65 (3), 381−396. https://doi.org/10.1093/sysbio/syv118.

Olson, N.D., et al., 2015. Best practices for evaluating single nucleotide variant calling methods for microbial genomics. Front. Genet. 6. https://doi.org/10.3389/fgene.2015.00235.

Omenn, G.S., 2010. Evolution and public health. Proc. Natl. Acad. Sci. U.S.A. 107 (1), 1702−1709. https://doi.org/10.1073/pnas.0906198106.

Orata, F.D., et al., 2014. The 2010 cholera outbreak in Haiti: how science solved a controversy. PLoS Pathog. 10 (4). https://doi.org/10.1371/journal.ppat.1003967.

Pace, N.R., 2009. Mapping the tree of life: progress and prospects. Microbiol. Mol. Biol. Rev. 73 (4), 565−576. https://doi.org/10.1128/MMBR.00033-09.

Parkhill, J., et al., 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat. Genet. 35 (1), 32−40. https://doi.org/10.1038/ng1227.

Patiño, L.H., et al., 2018. Unveiling the multilocus sequence typing (MLST) schemes and core genome phylogenies for genotyping chlamydia trachomatis. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.01854.

Pearson, T., et al., 2009. Phylogenetic understanding of clonal populations in an era of whole genome sequencing. Infect. Genet. Evol. 9 (5), 1010−1019. https://doi.org/10.1016/j.meegid.2009.05.014.

Pendleton, J.N., Gorman, S.P., Gilmore, B.F., 2013. Clinical relevance of the ESKAPE pathogens. Expert Rev. Anti-Infect. Ther. 11 (3), 297−308. https://doi.org/10.1586/eri.13.12.

Pérez-Losada, M., et al., 2005. Population genetics of Neisseria gonorrhoeae in a high-prevalence community using a hypervariable outer membrane porB and 13 slowly evolving housekeeping genes. Mol. Biol. Evol. 22 (9), 1887−1902. https://doi.org/10.1093/molbev/msi184.

Pérez-Losada, M., et al., 2006. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect. Genet. Evol. 6 (2), 97−112. https://doi.org/10.1016/j.meegid.2005.02.003.

Pérez-Losada, M., et al., 2009. Ethnic differences in the adaptation rate of HIV gp120 from a vaccine trial. Retrovirology 6 (67). https://doi.org/10.1186/1742-4690-6-67.

Pérez-Losada, M., et al., 2011. Phylodynamics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand. PLoS One 6 (3), e16902. https://doi.org/10.1371/journal.pone.0016902.

Pérez-Losada, M., et al., 2013. Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. Infect. Genet. Evol. 16, 38−53. https://doi.org/10.1016/j.meegid.2013.01.009.

Pérez-Losada, M., et al., 2015. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. Infect. Genet. Evol. 30, 296−307. https://doi.org/10.1016/j.meegid.2014.12.022.

Pérez-Losada, M., Arenas, M., Castro-Nallar, E., 2017. Multilocus sequence typing of pathogens: methods, analyses, and applications. Genetics and Evolution of Infectious Diseases, Second. Elsevier, pp. 383−404.

Pérez-Losada, M., et al., 2020. High-throughput sequencing (HTS) for the analysis of viral populations. Infect. Genet. Evol. 80. https://doi.org/10.1016/j.meegid.2020.104208.

Petrosino, J.F., et al., 2009. Metagenomic pyrosequencing and microbial identification. Clin. Chem. 55 (5), 856−866. https://doi.org/10.1373/clinchem.2008.107565.

Pettengill, J.B., et al., 2014. An evaluation of alternative methods for constructing phylogenies from whole genome sequence data: a case study with *Salmonella*. PeerJ 2 (1). https://doi.org/10.7717/peerj.620.

Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. Mol. Biol. Evol. 21 (7), 1455−1458. https://doi.org/10.1093/molbev/msh137.

Pond, S.L.K., Frost, S.D.W., 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol. Biol. Evol. 22 (3), 478−485. https://doi.org/10.1093/molbev/msi031.

Pond, K., et al., 2020. HyPhy 2.5-A customizable platform for evolutionary hypothesis testing using phylogenies. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msz197.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53 (5), 793−808. https://doi.org/10.1080/10635150490522304.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U.S.A. 98 (24), 13757−13762. https://doi.org/10.1073/pnas.241370698.

Posada, D., Crandall, K.A., Holmes, E.C., 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. 36, 75−97. https://doi.org/10.1146/annurev.genet.36.040202.111115.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2−approximately maximum-likelihood trees for large alignments. PLoS One 5 (3).

Rambaut, Drummond, A.J., 2009. Tracer: MCMC Trace Analysis Tool, 1.5 ed. Institute of Evolutionary Biology, Edinburgh.

Rannala, B., Yang, Z., Anderson, F., 2007. Inferring speciation times under an episodic molecular clock. Syst. Biol. 56 (3), 453−466. https://doi.org/10.1080/10635150701420643.

Rao, P.N., et al., 2016. A method for amplicon deep sequencing of drug resistance genes in plasmodium falciparum clinical isolates from India. J. Clin. Microbiol. 54 (6), 1500−1511. https://doi.org/10.1128/JCM.00235-16.

Rasmussen, M.D., Kellis, M., 2012. Unified modeling of gene duplication, loss, and coalescence using a locus tree. Genome Res. 22 (4), 755−765. https://doi.org/10.1101/gr.123901.111.

Rasmussen, M.D., et al., 2014. Genome-wide inference of ancestral recombination graphs. PLoS Genet. 10 (5). https://doi.org/10.1371/journal.pgen.1004342.

Reimer, A.R., et al., 2011. Comparative genomics of *Vibrio cholerae* from Haiti, Asia, and Africa. Emerg. Infect. Dis. 17 (11), 2113−2121. https://doi.org/10.3201/eid1711.110794.

Rieux, A., Balloux, F., 2016. Inferences from tip-calibrated phylogenies: a review and a practical guide. Mol. Ecol. 25 (9), 1911−1924. https://doi.org/10.1111/mec.13586.

Rivailler, P., et al., 2019. Recombination analysis of Human *Mastadenovirus* C whole genomes. Sci. Rep. 9 (1). https://doi.org/10.1038/s41598-019-38719-z.

Rodréguez-Ezpeleta, N., et al., 2007. Detecting and overcoming systematic Errors in genome-scale phylogenies. Syst. Biol. 56 (3), 389−399. https://doi.org/10.1080/10635150701397643.

Rohde, H., et al., 2011. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. N. Engl. J. Med. 365 (8), 718−724. https://doi.org/10.1056/NEJMoa1107643.

Ronquist, F., et al., 2012. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (3), 539–542. https://doi.org/10.1093/sysbio/sys029.

Roumagnac, P., et al., 2006. Evolutionary history of *Salmonella* Typhi. Science 314 (5803), 1301–1304. https://doi.org/10.1126/science.1134933.

Rozas, J., et al., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34 (12), 3299–3302. https://doi.org/10.1093/molbev/msx248.

Sahl, J.W., et al., 2015. Phylogenetically typing bacterial strains from partial SNP genotypes observed from direct sequencing of clinical specimen metagenomic data. Genome Med. 7 (1). https://doi.org/10.1186/s13073-015-0176-9.

Sahl, J.W., et al., 2016. NASP: an accurate, rapid method for the identification of SNPs in WGS datasets that supports flexible input and output formats. Microb. Genom. 2 (8). https://doi.org/10.1099/mgen.0.000074.

Salipante, S.J., et al., 2015. Application of whole-genome sequencing for bacterial strain typing in molecular epidemiology. J. Clin. Microbiol. 53 (4), 1072–1079. https://doi.org/10.1128/JCM.03385-14.

Sangwan, N., Xia, F., Gilbert, J.A., 2016. Recovering complete and draft population genomes from metagenome datasets. Microbiome 4 (1). https://doi.org/10.1186/s40168-016-0154-5.

Schierup, M.H., Hein, J., 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics 156 (2), 879–891.

Setubal, J.C., 2021. Metagenome-assembled genomes: concepts, analogies, and challenges. Biophys. Rev. 13 (6), 905–909. https://doi.org/10.1007/s12551-021-00865-y.

Sharon, I., Banfield, J.F., 2013. Genomes from metagenomics. Science 342 (6162), 1057–1058. https://doi.org/10.1126/science.1247023.

Shimodaira, H., Hasegawa, M., 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17 (12), 1246–1247. https://doi.org/10.1093/bioinformatics/17.12.1246.

Sievers, F., Higgins, D.G., 2018. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 27 (1), 135–145. https://doi.org/10.1002/pro.3290.

Silbergeld, E.K., Graham, J., Price, L.B., 2008. Industrial food animal production, antimicrobial resistance, and human health. Annu. Rev. Publ. Health. https://doi.org/10.1146/annurev.publhealth.29.020907.090904.

Skarp-de Haan, C.P.A., et al., 2014. Comparative genomics of unintrogressed *Campylobacter coli* clades 2 and 3. BMC Genom. 15 (1). https://doi.org/10.1186/1471-2164-15-129.

Smith, C.C.R., Flaxman, S.M., 2020. Leveraging whole genome sequencing data for demographic inference with approximate Bayesian computation. Mol. Ecol. Res. 20 (1), 125–139. https://doi.org/10.1111/1755-0998.13092.

Spencer, C.C.A., et al., 2006. The influence of recombination on human genetic diversity. PLoS Genet. 2 (9). https://doi.org/10.1371/journal.pgen.0020148.

Spratt, B.G., Maiden, M.C.J., 1999. Bacterial population genetics, evolution and epidemiology. Phil. Trans. Biol. Sci. 354 (1384), 701–710. https://doi.org/10.1098/rstb.1999.0423.

Stamatakis, A., et al., 2012. RAxML-Light: a tool for computing terabyte phylogenies. Bioinformatics 28 (15), 2064–2066. https://doi.org/10.1093/bioinformatics/bts309.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22 (21), 2688–2690. https://doi.org/10.1093/bioinformatics/btl446.

Stoler, N., Nekrutenko, A., 2021. Sequencing error profiles of Illumina sequencing instruments. NAR Genom. Bioinform. 3 (1). https://doi.org/10.1093/nargab/lqab019.

Sullivan, C.B., Diggle, M.A., Clarke, S.C., 2005. Multilocus sequence typing: data analysis in clinical microbiology and public health. Mol. Biotechnol. 29 (3), 245–254. https://doi.org/10.1385/MB:29:3:245.

Sullivan, C.B., et al., 2006. Automation of MLST using third-generation liquid-handling technology. Mol. Biotechnol. 32 (3), 219–225. https://doi.org/10.1385/MB:32:3:219.

Sutton, T.D.S., et al., 2019. Choice of assembly software has a critical impact on virome characterisation. Microbiome 7 (1). https://doi.org/10.1186/s40168-019-0626-5.

Szöllosi, G.J., Daubin, V., 2012. Modeling gene family evolution and reconciling phylogenetic discord. Methods Mol. Biol. 856, 29–51. https://doi.org/10.1007/978-1-61779-585-5_2.

Taylor, D.L., et al., 2016. Accurate estimation of fungal diversity and abundance through improved lineage-specific primers optimized for Illumina amplicon sequencing. Appl. Environ. Microbiol. 82 (24), 7217–7226. https://doi.org/10.1128/AEM.02576-16.

Tazi, L., et al., 2010. Population dynamics of *Neisseria gonorrhoeae* in Shanghai, China: a comparative study. BMC Infect. Dis. 10. https://doi.org/10.1186/1471-2334-10-13.

Templeton, A.R., Crandall, K.A., Sing, C.F., 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132 (2), 619–633. https://doi.org/10.1093/genetics/132.2.619.

Tett, A., et al., 2017. Unexplored diversity and strain-level structure of the skin microbiome associated with psoriasis. Npj Biofilms Microbiomes 3 (1). https://doi.org/10.1038/s41522-017-0022-5.

The, H.C., et al., 2016. The genomic signatures of Shigella evolution, adaptation and geographical spread. Nat. Rev. Microbiol. 14 (4), 235–250. https://doi.org/10.1038/nrmicro.2016.10.

Theunert, C., et al., 2012. Inferring the history of population size change from genome-wide SNP data. Mol. Biol. Evol. 29 (12), 3653–3667. https://doi.org/10.1093/molbev/mss175.

Touchon, M., et al., 2011. CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. J. Bacteriol. 193 (10), 2460–2467. https://doi.org/10.1128/jb.01307-10.

Treangen, T.J., et al., 2014. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol. 15 (11). https://doi.org/10.1186/s13059-014-0524-x.

Uelze, L., et al., 2020. Typing methods based on whole genome sequencing data. One Health Outlook 2 (1). https://doi.org/10.1186/s42522-020-0010-1.

Urwin, R., Maiden, M.C.J., 2003. Multi-locus sequence typing: a tool for global epidemiology. Trends Microbiol. 11 (10), 479–487. https://doi.org/10.1016/j.tim.2003.08.006.

Van Belkum, A., 2002. Molecular typing of micro-organisms: at the centre of diagnostics, genomics and pathogenesis of infectious diseases? J. Med. Microbiol. 51 (1), 7–10. https://doi.org/10.1099/0022-1317-51-1-7.

van Cuyck, H., et al., 2012. Multiple-locus variable-number tandem-repeat analysis of *Streptococcus* pneumoniae and comparison with multiple loci sequence typing. BMC Microbiol. 12 (1). https://doi.org/10.1186/1471-2180-12-241.

VanInsberghe, et al., 2021. Recombinant SARS-CoV-2 genomes are currently circulating at low levels. bioRxiv. https://doi.org/10.1101/2020.08.05.238386.

Vincent, A.T., et al., 2017. Next-generation sequencing (NGS) in the microbiological world: how to make the most of your money. J. Microbiol. Methods 138, 60–71. https://doi.org/10.1016/j.mimet.2016.02.016.

Walker, T.M., et al., 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect. Dis. 13 (2), 137–146. https://doi.org/10.1016/S1473-3099(12)70277-3.

Wang, J., 2005. Estimation of effective population sizes from data on genetic markers. Phil. Trans. Biol. Sci. 360 (1459), 1395–1409. https://doi.org/10.1098/rstb.2005.1682.

Ward, D.V., et al., 2016. Metagenomic sequencing with strain-level resolution implicates uropathogenic *E. coli* in necrotizing enterocolitis and mortality in preterm infants. Cell Rep. 14 (12), 2912–2924. https://doi.org/10.1016/j.celrep.2016.03.015.

Weaver, S., et al., 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol. Biol. Evol. 35 (3), 773–777. https://doi.org/10.1093/molbev/msx335.

Wegmann, D., Leuenberger, C., Excoffier, L., 2009. Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. Genetics 182 (4), 1207–1218. https://doi.org/10.1534/genetics.109.102509.

Wegrzyn, J.L., et al., 2009. PineSAP - sequence alignment and SNP identification pipeline. Bioinformatics 25 (19), 2609–2610. https://doi.org/10.1093/bioinformatics/btp477.

Wiens, J.J., 2000. Reconstructing phylogenies from allozyme data: comparing method performance with congruence. Biol. J. Linn. Soc. 70 (4), 613–632. https://doi.org/10.1006/bijl.1999.0420.

Wilson, D.J., McVean, G., 2006. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics 172 (3), 1411–1425. https://doi.org/10.1534/genetics.105.044917.

Wong, V.K., et al., 2015. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter-and intracontinental transmission events. Nat. Genet. 47 (6), 632–639. https://doi.org/10.1038/ng.3281.

Wong, V.K., Baker, S., et al., 2016a. An extended genotyping framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid. Nat. Commun. 7. https://doi.org/10.1038/ncomms12827.

Wong, V.K., Holt, K.E., et al., 2016b. Molecular surveillance identifies multiple transmissions of typhoid in West Africa. Trop. Dis. 10 (9). https://doi.org/10.1371/journal.pntd.0004781.

Wu, J., et al., 2022. Annotating unknown species of urban microorganisms on a global scale unveils novel functional diversity and local environment association. Environ. Res. 207. https://doi.org/10.1016/j.envres.2021.112183.

Wyres, K.L., Holt, K.E., 2016. *Klebsiella pneumoniae* population genomics and antimicrobial-resistant clones. Trends Microbiol. 24 (12), 944–956. https://doi.org/10.1016/j.tim.2016.09.007.

Yan, Z., et al., 2022. Species tree inference methods intended to deal with incomplete lineage sorting are robust to the presence of paralogs. Syst. Biol. 71 (2), 367–381. https://doi.org/10.1093/sysbio/syab056.

Yang, Z., Nielsen, R., 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19 (6), 908–917. https://doi.org/10.1093/oxfordjournals.molbev.a004148.

Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nat. Rev. Genet. 13 (5), 303–314. https://doi.org/10.1038/nrg3186.

Yang, J., et al., 2007. Revisiting the molecular evolutionary history of *Shigella* spp. J. Mol. Evol. 64 (1), 71–79. https://doi.org/10.1007/s00239-006-0052-8.

Yang, S., et al., 2017. Evolution and transmission of carbapenem-resistant *Klebsiella pneumoniae* expressing the blaOXA-232 gene during an institutional outbreak associated with endoscopic retrograde cholangiopancreatography. Clin. Infect. Dis. 64 (7), 894–901. https://doi.org/10.1093/cid/ciw876.

Yang, Z., 1994. Statistical properties of the maximum likelihood method of phylogenetic estimation and comparison with distance matrix methods. Syst. Biol. 43 (3), 329–342. https://doi.org/10.1093/sysbio/43.3.329.

Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24 (8), 1586–1591. https://doi.org/10.1093/molbev/msm088.

Yoshimura, D., et al., 2019. Evaluation of SNP calling methods for closely related bacterial isolates and a novel high-accuracy pipeline: BactSNP. Microb. Genom. 5 (5). https://doi.org/10.1099/mgen.0.000261.

Yu, Y., Ristic, N., Nakhleh, L., 2013. Fast algorithms and heuristics for phylogenomics under ILS and hybridization. BMC Bioinf. 14. https://doi.org/10.1186/1471-2105-14-S15-S6.

Zhang, R., Drummond, A., 2020. Improving the performance of Bayesian phylogenetic inference under relaxed clock models. BMC Evol. Biol. 20 (1). https://doi.org/10.1186/s12862-020-01609-4.

Zhao, S., et al., 2019. Adaptive evolution within gut microbiomes of healthy people. Cell Host Microbe 25 (5), 656–667.e8. https://doi.org/10.1016/j.chom.2019.03.007.

Zhou, Z., et al., 2020. The EnteroBase user's guide, with case studies on *Salmonella* transmissions, *Yersinia pestis* phylogeny, and *Escherichia* core genomic diversity. Genome Res. 30 (1), 138−152. https://doi.org/10.1101/gr.251678.119.

Zhou, Z., et al., 2021. HierCC: a multi-level clustering scheme for population assignments based on core genome MLST. Bioinformatics 37 (20), 3645−3646. https://doi.org/10.1093/bioinformatics/btab234.

Zolfo, M., et al., 2017. MetaMLST: multi-locus strain-level bacterial typing from metagenomic samples. Nucleic Acids Res. 45 (2). https://doi.org/10.1093/nar/gkw837.

Zoller, S., Boskova, V., Anisimova, M., 2015. Maximum-likelihood tree estimation using codon substitution models with multiple partitions. Mol. Biol. Evol. 32 (8), 2208−2216. https://doi.org/10.1093/molbev/msv097.

This page intentionally left blank

CHAPTER

Proteomics, Bioinformatics, and Infectious Diseases

18

Jennifer Geddes-McAlister[1,4], Florence Roux-Dalvai[2,3,4] and Arnaud Droit[2,3,4]

[1]*Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON, Canada;* [2]*Proteomics Platform, CHU de Québec - Université Laval Research Center, Québec City, QC, Canada;* [3]*Computational Biology Laboratory, CHU de Québec - Université Laval Research Center, Québec City, QC, Canada;* [4]*Canadian Proteomics and Artificial Intelligence Consortium, Guelph, ON, Canada*

1. Introduction

The field of mass spectrometry–based proteomics has rapidly advanced over the past few decades to discover new technological approaches, develop complex bioinformatics pipelines, and push the boundaries of application. For many infectious diseases, pathogen identification is an initial critical step for diagnosing a patient and devising a plan for treatment. For decades, tools, including matrix-assisted laser desorption/ionization–time of flight mass spectrometry (MALDI-TOF MS) have been used for accurate, cost-effective, and relatively rapid identification of cultured bacteria and fungi in clinical microbiology (Patel, 2015). The technology is automated, high throughput, and applicable to a broad range of pathogens but presents limitations when distinguishing between species and strains and assessing nonculturable pathogens (e.g., viruses) (Sloan et al., 2017).

Current state-of-the-art diagnostic methods for infections, known as proteotyping, can identify the causative microbial species within approximately 24 hours (Karlsson et al., 2015). However, depending on the microbe, proteotyping may take several days, and determination of the correct antimicrobial therapy during the first 6–12 hours of infection is crucial for the treatment of life-threatening infections (Burnham et al., 2017; Kollef, 2008). With such temporal delays, patients often receive broad-spectrum antibiotics, which may be ineffective if the infection is not caused by a bacterial pathogen (e.g., fungi or virus), the antibiotic may be inefficient if a species-specific drug is required (e.g., Gram-positive vs. Gran-negative bacterial pathogens are susceptible to different drugs), or irrelevant if the condition is not due to an infection at all. In these instances, time delays and inaccurate antimicrobial administration contribute to the evolution of resistance within microbial populations. Alternative strategies, such as rapid, and robust liquid chromatography (LC) separation combined with high-sensitivity and high-resolution tandem mass spectrometry provides unmatched identification specificity and removes the need for microbial culturing, reducing diagnostic wait times (Boulund et al., 2017; Wang et al., 2016). Moreover, combining this technology with artificial intelligence, such as machine learning and deep learning algorithms, allows for the prediction of pathogenic species in shorter periods of time (Roux-Dalvai et al., 2019).

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00024-0
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Beyond pathogen identification, mass spectrometry–based proteomics approaches are also used for assay development, for instance, with biomarkers to detect the presence of a pathogen from a biological sample, as well as identify the causative agent of disease (Hawkridge and Muddiman, 2009; Nakayasu et al., 2021). Discovery-based approaches can detect proteins within various sample matrices and quantify changes in protein production to identify biomarkers of disease and to improve our understanding of pathogenesis and the relationship between the host and pathogen during disease (Muselius et al., 2023). Specifically, mass spectrometry–based proteomics platforms combined with advanced bioinformatics tools can profile how a pathogen adapts to the hostile environment of the host and how the host remodels its proteome to defend from invasion. This dual or multi species approach provides new biological insights into relationships driving disease from differing perspectives within a single experiment (Sukumaran et al., 2022).

Another critical component of infectious disease research is the ever-increasing pressure of antimicrobial resistance, which threatens the effective prevention and treatment of infections caused by diverse microorganisms (Murray et al., 2022). Importantly, by 2050, antimicrobial resistance is predicted to be the leading cause of death around the world costing between $300 billion and $1.2 trillion, annually, for health care costs (Council of Canadian Academies, 2019; World Health Organization, 2021). These predictions are based on over prescription and misuse of antimicrobials within the clinic, use of antibiotics for animal growth and production, lack of surveillance of resistance development, and the application of fungicides and pesticides within the field for agricultural purposes (Chokshi et al., 2019; Davies et al., 2013; Manyi-Loh et al., 2018; World Health Organization, 2021; Woods et al., 2023). Additionally, our knowledge of the mechanisms that drive antimicrobial resistance is constantly shifting due to the emergence of new pathogens with intrinsic resistance (e.g., *Candida auris*) and the continual arms race against the evolution of multidrug resistance (e.g., ESKAPE pathogens) (Chaabane et al., 2019; Mulani et al., 2019; Bermas and Geddes-McAlister, 2020). Defining signatures of resistance is not trivial and traditionally, resistance levels of a pathogen are determined with microbial culturing and plating on singular or combinatory antimicrobial plates. However, mass spectrometry–based proteomics provides a new avenue for exploration of protein drivers of antimicrobial resistance (Goodyear et al., 2021).

In this chapter, we provide an overview of mass spectrometry–based proteomics technologies and applications in the context of infectious disease, which has steadily increased over the past two decades (Fig. 18.1). We focus on bacterial and fungal pathogens and their relationship with the host during disease, while acknowledging and providing resources for additional exploration into diseases attributed to viral infections, including the globally devastating, HIV and SARS-CoV-2. We also explore a plethora of bioinformatics tools used to interrogate proteomics data for rapid analysis, creative visualization, and accurate interpretation, as well as leading-edge applications of artificial intelligence, such as machine learning models. Finally, we highlight the use of proteomics and advanced bioinformatics to investigate infectious diseases, applications within the clinic, and to increase our depth of understanding, as well as harness the power of proteomics to define and treat diseases.

1.1 Mass Spectrometry–Based Proteomics

The field of mass spectrometry–based proteomics is diverse in its instrumentation and applications (Aebersold and Mann, 2003, 2016). When combined with annotated genomes or de novo sequencing,

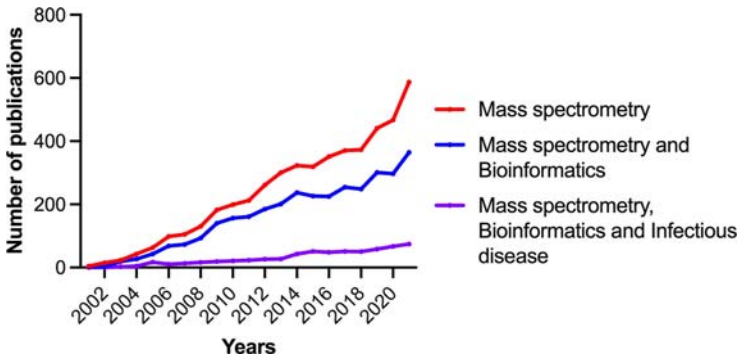


FIGURE 18.1 Mass spectrometry–based proteomics and infectious disease research publications from 2000–2021.

Search parameters set to: "Mass spectrometry" (*red*), "mass spectrometry" and "bioinformatics" (*blue*), and "mass spectrometry" and "bioinformatics" and "infectious disease" (*purple*) in the Title and/or Abstract. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

Based on data collected from the National center for biotechnology information, database: PubMed. Figure generated in GraphPad Prism version 9.0.

proteomics can identify and distinguish disease-causing species or investigate complex biological systems to define changes in protein abundance and production under observed conditions. Additionally, proteomics provides insight into modifications (reversible and irreversible) and the downstream implications, as well as interactions between proteins and other molecules (e.g., DNA, RNA, proteins). Processable sample types are plentiful, including tissues, organs, fluids, cells, and supernatants, which promotes diverse application of the technology to address global challenges, such as climate change, health, drug discovery, and food security, among others (Ball et al., 2020, 2019; Geddes-McAlister et al., 2022; Meissner et al., 2022; Sukumaran et al., 2020, 2019).

Mass spectrometry–based proteomics is commonly defined as top-down (i.e., analyzes intact proteins for the identification of protein complexes and proteoforms) or bottom-up (i.e., classical discovery-driven approach that uses sequence-specific enzymes to digest proteins into peptides followed by separation via liquid chromatography and measurement by mass spectrometry) (Fig. 18.2). Targeted proteomics also analyzes digested proteins and is often applied for detection and accurate quantification of predefined peptides in complex mixtures, such as for biomarker discovery and detection. Traditionally, proteomics experiments were, by default, performed in data-dependent acquisition (DDA) mode, which selects the top *n* (e.g., top 10) most abundant peptides from a survey scan for sequential fragmentation (Gillet et al., 2012). However, as advances in instrumentation and computation are made, data-independent acquisition (DIA), which fragments all peptides within a cycling *mass*-to-*charge* (*m/z)* window over the entire *m/z* range, has increased in popularity. DIA presents advantageous peptide identification rates compared to DDA but unambiguous peptide identification from the complex fragmentation spectra is an ongoing challenge (Hu et al., 2016).

To measure changes in protein abundance, different quantification strategies exist. For instance, a popular option due to limited sample handling and lack of needed a priori sample knowledge, label-free quantification (LFQ), uses signal intensities directly from measured peptide mixtures in both DDA and DIA modes (Nahnsen et al., 2013). Although LFQ provides flexibility and post-spectra data

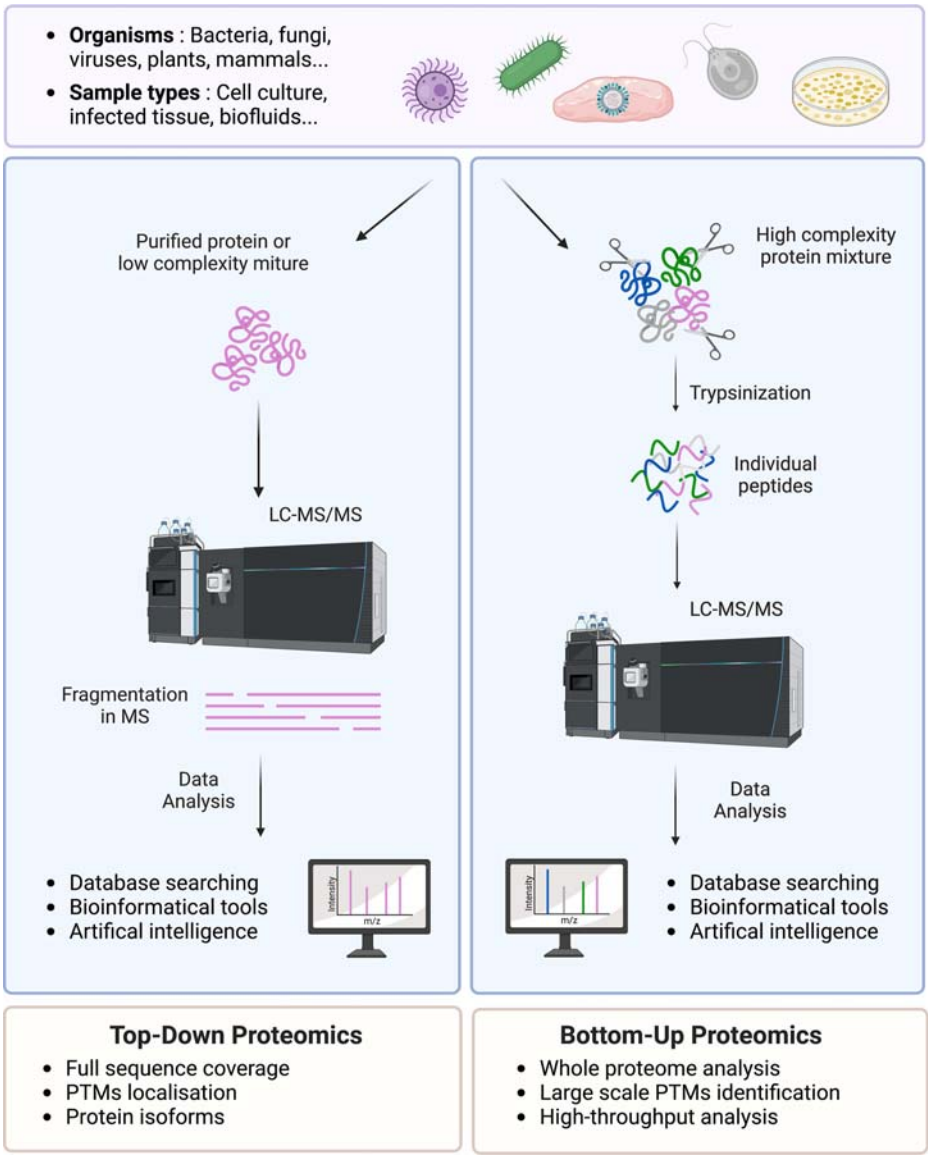


FIGURE 18.2

Summary of mass spectrometry–based platforms and techniques.
Definitions: *LC-MC/MS*, liquid chromatography tandem mass spectrometry; *PTMs*, posttranslational modifications.

Figure generated with Biorender.com.

processing, a limitation of missing values and the use of imputation are drawbacks. Alternative labeling methods include metabolic and chemical approaches that differ in sample handling and applications. For example, metabolic labeling with isotopically stable and nonradioactive forms of amino acids, such as SILAC (stable isotope labeling with amino acids in cell culture), incorporate at the cellular or organismal level during growth to assess protein synthesis and turnover (Savitski et al., 2018). Conversely, chemical labeling strategies employ derivatized peptides for quantification and multiplexing capabilities, such as tandem mass tags (Thompson et al., 2003), iTRAQ (isobaric tags for relative and absolute quantification) (Ross et al., 2004), and EASI-tag (easily abstractable sulfoxide-based isobaric tag) (Virreira Winter et al., 2018). Each of these approaches requires additional sample handling to incorporate the selected tags, which has potential for introducing technical variability; however, a reduction in instrument time due to multiplexing and increased sensitivity for quantification coverage despite common losses in proteome identification rates, provide benefits to labeling strategies that warrant consideration.

Beyond identification and quantification, mass spectrometry–based proteomics also enables detection of posttranslational modification (PTM) events based on the associated mass changes (Larance and Lamond, 2015). PTMs play diverse roles toward biological processes by changing the structure and function of proteins and subsequent regulation of signaling cascades and other downstream events (Retanal et al., 2021). PTMs encompass both covalent and noncovalent functional group additions to a protein, the cleavage of proteins or peptides by proteases or peptidases, respectively, or protein degradation. Commonly explored PTMs include phosphorylation (phosphoryl group addition), acetylation (acetyl functional group addition), ubiquitination (attachment of a ubiquitin molecule), glycosylation (introduction of glycosyl donor groups of varying sizes), methylation (methyl group addition), among others. Such modifications are interrogated from proteomics sample preparations via physical enrichment techniques (e.g., TiO_2, antibodies) or computational (e.g., mass changes) approaches to detect stoichiometrically limiting modifications among the highly abundant nonmodified peptides from complex samples.

Proteomics also plays critical roles in affinity purification and localization experiments to define interactions and explore spatial–temporal relationships among proteins and complexes. For instance, subcellular compartment profiling may involve enrichment of desired regions or depletion of contaminating organelles via lysis, gradient centrifugation, and washing (Andersen et al., 2003). Additionally, protein correlation profiling informs protein localization based on co-purification assumed from shared subcellular compartmentalization (Geladaki et al., 2019; Itzhak et al., 2016; Kristensen et al., 2012). Alternative methods include proximity-based labeling, which rely on the fusion of bait proteins with enzymes that generate reactive affinity handles to modify proteins covalently in close spatial radius. Examples include ascorbate peroxidase (APEX) (Rhee et al., 2013) and biotin ligase (BioID) (Kim et al., 2016; Roux et al., 2012, 2018) for identification of direct and indirect protein interactors. Spatial organization of proteins is also applicable for imaging mass spectrometry to map the distribution of biomolecules for quantitative cell-type–specific images (Piehowski et al., 2020). Together, the dynamic and extensive applications of mass spectrometry–based proteomics to explore global and cell-type–specific changes in protein production, along with regulation of PTMs, and interactions among proteins and complexes provides a powerful platform to interrogate complex biological systems.

1.2 Bioinformatics Tools in Proteomics

The value of proteomics is only fully recognized in combination with advanced bioinformatics tools that enable the rapid and reliable interpretation of raw mass spectrometry data. Bioinformatics is an interdisciplinary field of science integrating computer science, statistics, mathematics, and engineering to analyze and interpret complex datasets (Schmidt et al., 2014). Foundational bioinformatics tools that support the development of proteomics experiments include databases of annotated genomes (e.g., NCBI; https://www.ncbi.nlm.nih.gov) and proteomes (e.g., Uniprot; https://www.uniprot.org), as well as homology tools (e.g., Blast; https://blast.ncbi.nlm.nih.gov), structure prediction databases (e.g., AlphaFold) (Jumper et al., 2021), and functional analyses based on biological processes, molecular functions, and cellular compartmentalization (e.g., Gene Ontology) (Ashburner et al., 2000). Combining such biological insights with proteomics datasets enhances the complexity and depth of the experimental findings. Further, publicly available repositories promote consistency, transparency, and sharing of proteomics data, such as ProteomeXchange PRIDE (Perez-Riverol et al., 2019) and MASSive (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) for extracting maximal information.

For analyzing proteomics datasets, many bioinformatics tools have specific and overlapping features that depend upon user expertise and preference, experimental parameters, and the desired outcomes. For example, Skyline (MacLean et al., 2010), a freely available and open-source platform, was designed for building and searching spectral libraries for selected reaction monitoring (SRM)/multiple reaction monitoring (MRM), parallel reaction monitoring (PRM), DIA, and DDA mass spectrometry data. OpenMS (Röst et al., 2016) and OpenSWATH (sequential window acquisition of all theoretical mass spectra) (Röst et al., 2014) are open-source libraries that enable the analysis of DIA mass spectrometry data for management, analysis, and visualization, as well as integration across proteomics and metabolomics experiments. Other examples include MaxQuant (Cox et al., 2011) and Perseus (Tyanova et al., 2016), quantitative proteomics software packages that are publicly available and designed for analyzing and visualizing large, high-resolution mass spectrometry datasets. MaxQuant integrates the Andromeda search engine (Cox et al., 2011) for peptide identification from raw mass spectrometry data, whereas Perseus supports interpretation of protein quantification, interaction, and PTM data sets using a comprehensive suite of statistical tools for high-dimensional analysis (e.g., normalization, imputation, cross-omics comparisons, multiple hypothesis testing, clustering, principal component plotting). Fragpipe (https://www.nesvilab.org/software.html), a complete proteomics pipeline built upon the ultrafast MSFragger (Kong et al., 2017) search engine, is a java graphical user interface with capabilities for downstream postprocessing of search results for multiple-hypothesis testing, filtering, label-based quantification, and multi-experiment report generation while supporting plug-ins for additional processing options. Further, ProteinProspector (Chalkley et al., 2005), Peptide Atlas (Desiere, 2006), the Trans Proteomics Pipeline (Deutsch et al., 2015) encompass suites of proteomics tools for mining sequences databases and analyzing mass spectrometry datasets.

For capitalizing on protein interaction data acquired from affinity purification proteomics experiments or similar, multiple tools are available to identify direct and indirect interacting proteins and protein complexes, and to visualize interaction networks across species. For instance, SAINT (Significance Analysis of INTeractome) (Choi et al., 2011) and DAVID (Database for Annotation, Visualization, and Integrated Discovery) (Huang et al., 2009) assigns confidence scores to protein-protein interactions based on quantitative proteomics data in affinity purification experiments. Other

interactome software tools include BioPlex (Huttlin et al., 2021), STRING (Szklarczyk et al., 2019), Cytoscape (Shannon et al., 2003) for assessment and visualization of protein−protein interaction networks, and PIGNON (Nadeau et al., 2021) for functional enrichment analysis of protein−protein interaction maps. Moreover, as our understanding of biological systems grows, as does the potential to explore protein level changes in multiple species in a single experiment, as designed for microbiome and metaproteome experiments. For instance, MetaLab (Cheng et al., 2017) and iMetaLab (Liao et al., 2018) are user-friendly web-based platforms for the comprehensive and automated analysis of meta-proteomics datasets from compiled sample specific databases to define abundances of microbial peptides and proteins, taxonomic compositions, and functional profiling. In addition, mPies (Werner et al., 2019) creates relevant protein databases and automated protein annotation for meta-proteomics experiments, whereas MetaProClust-MS1 (Simopoulos et al., 2022) is a computational framework for screening features of the microbiome for sample prioritization and follow up by tandem mass spectrometry.

As the field of bioinformatics rapidly advances, as does the opportunity to integrate new strategies, such as artificial intelligence, with examples of machine and deep learning to further foster the discoveries. For example, deep learning applications to proteomics datasets now accurately predict physiochemical properties of peptides from sequence information, including tandem mass spectra and retention time, informing experimental design and biomarker development (Meyer, 2021). For such advances, large volumes of training data are required (Ma et al., 2018), along with Deep Learning models, such as pDeep (Zhou et al., 2017), and DIA-NN (Demichev et al., 2020), which is dedicated to DIA analysis and exploits deep neural network to generate an in silico library from a protein sequence database that extracts specific signal from complex MS/MS spectra. Recently, the development of AlphaPept (Strauss et al., 2021) and specifically, AlphaPeptDeep (Zeng et al., 2022) an open-source modular deep learning framework to predict peptide properties for proteomics, has furthered these applications. The AlphaPeptDeep platform contains pretrained models for predicting MS2 intensities, retention time, and collisional cross sections of peptide sequences or entire proteomes, along with peptides supporting PTMs, and uses transfer learning strategies to reduce the amount of training data required. Overall, a plethora of freely available platforms exist, along with commercial options, to process, analyze, interpret, and visualize proteomics datasets with a transition toward machine learning and deep learning strategies to better equip researchers with leading-edge tools to extract the most valuable biological information from their datasets. Further, machine learning and deep learning can be used to extract biomarker signatures from large-scale proteomics data (e.g., BioDiscML) (Leclercq et al., 2019) and to make predictions for disease progression, outcome, and response to treatment.

2. Proteomics of Infectious Diseases

2.1 Disease Applications

2.1.1 Host Perspective

Mass spectrometry−based proteomics provides an advantage when exploring infectious diseases: profiling of protein level changes during infection is universal and can investigate the host's defense response upon invasion, the pathogen's ability to invade, as well as the interaction between the host and pathogen. These multi-perspective approaches lend themselves to better explore the relationship

between a host and pathogen during infection and provide new biological insight into disease (Greco and Cristea, 2017; Jean Beltran et al., 2017). For example, from the host's perspective, mass spectrometry–based proteomics has defined communication patterns among activated immune cells by characterizing 28 primary human hematopoietic cell populations and detecting over 10,000 proteins (Rieckmann et al., 2017). This study uncovered immune cell specializations associated with ligand–receptor expression to connect distinct immune cell functions and revealed communication signals via protein secretion between previously unconnected immune cell types. To enhance the value of this dataset, a publicly available proteomic resource was developed (http://www.immprot.org) for researchers to explore these new connections and better understand host response and signaling associated with immune cell activation.

Profiling of the host upon microbial infection can also highlight differences between effective versus ineffective immune responses to infection. For example, proteomic profiling of neutrophils derived from mouse models that are susceptible (C57BL6/N) or resistant (Swiss Webster) to bacterial infection of the cornea by *Pseudomonas aeruginosa,* revealed strain-specific and infection-induced alterations between the neutrophil proteomics (Kugadas et al., 2019). Specifically, within the resistant mice, levels of α-mannosidases, enzymes associated with digestion of polysaccharides comprising bacterial biofilms (Hill et al., 2017), were distinctly elevated compared to neutrophils derived from susceptible mice. Moreover, treatment of ocular infection with exogenous mannosidase promoted clearance of the bacteria. Another example explored global phosphoproteome reprogramming of host signaling cascades upon fungal infection with *Cryptococcus neoformans* (Pandey et al., 2017). This study revealed differential phosphorylation following fungal ingestion by macrophages, including changes to the host autophagy initiative complex to promote intracellular fungal trafficking and replication. These data support the role of autophagy in regulating susceptibility of host cells to fungal infection and support new strategies to restore host signaling networks for an efficient defense response to pathogen invasion.

2.1.2 Pathogen Perspective

For a pathogen to survive the hostile and competitive environment of the host, adaptive mechanisms are often employed, including new strategies for nutrient acquisition. For instance, mass spectrometry–based proteomics was used to define global proteome changes of the bacterial pathogen, *Klebsiella pneumoniae*, under environments with limited essential ions, including biological important metal ions (e.g., iron and zinc) (Muselius et al., 2020; Sukumaran et al., 2021). Investigation into iron limitation on the cellular proteome and secretome reported an impact on transcriptional regulation and modulation of production toward proteins associated with iron acquisition, transport, and binding (Muselius et al., 2020). Importantly, a novel connection between the well-characterized lon protease, which regulates cellular homeostasis through degradation of abnormal proteins, and degrades physiological targets involved in cell division and capsule transcription (Mizusawa and Gottesman, 1983; Nicoloff and Andersson, 2013), and iron homeostasis was reported with predicted function toward degradation of bacterial and host proteins for improved iron uptake by the pathogen. Similarly, under zinc limitation, *K. pneumoniae* demonstrates altered transcriptional regulation that is influenced by protein abundance along with characterization of a novel cation transport regulator, ChaB (Sukumaran et al., 2021). Further phenotypic profiling confirmed a role for ChaB in metal ion homeostasis, as well as a connection between zinc availability, capsule regulation, and virulence of the pathogen.

Notably, fungal pathogens are also influenced by fluctuations in metal ion availability that drive pathogen adaptation and survival within the host. Changes to the proteome were demonstrated for the fungal pathogen, *C. neoformans*, under altered zinc availability to reveal a core proteome combined with distinct zinc-regulated protein level signatures (Ball et al., 2021). These findings highlight a shift away from metal ion transport and binding under replete conditions and toward transcription and metal acquisition under limited conditions. Further, the study characterized the role of a novel co-chaperone, Wos2, under a limiting environment influencing fungal thermotolerance, as well as capsule and melanin production. Together, probing the ability of pathogens to adapt and survive under host-mimicking conditions highlight new avenues of antivirulence (i.e., disarming virulence factors of the pathogen but not killing the pathogen to reduce selective pressure toward the evolution of resistance) strategies by exposing points of weakness for each pathogen.

2.1.3 Dual Perspective

Applying mass spectrometry−based proteomics to explore host−pathogen interactions can highlight novel interactions between the biological systems with novel function (Sukumaran et al., 2020). For instance, in *Salmonella enterica* serovar Typhimurium infections, effector proteins are injected by the pathogen into host cells to hijack host cellular processes and promote bacterial survival (Jennings et al., 2017; Watson and Holden, 2010). The interactions between pathogen effector proteins and the host targets may represent novel opportunities to perturb infection using antivirulence strategies by disarming the bacteria and promoting host cell clearance (Mühlen and Dersch, 2016; Rasko and Sperandio, 2010). To define such interactions, affinity purification combined with mass spectrometry−based proteomics reported over 440 interactions, including previously defined controls, and novel interactions between *S.* Typhimurium and host macrophages and epithelial cells (Walch et al., 2021). These data demonstrate redundancy in effector function by multiple effectors targeting the same host cellular processes, as well as new and unexpected interactions, such as host cholesterol transport to promote protection of the bacterium through vacuole formation. Together, this study describes a method for probing host-pathogen interactions during infection applied to *S*. Typhimurium but with universal applicability, along with novel mechanisms used by pathogens to overcome host defenses, representing potential new therapeutic targets.

Other examples include defining global changes in protein abundance upon infection, as demonstrated during ocular keratitis caused by the bacterial pathogen, *P. aeruginosa* (Yeung et al., 2020). In this study, proteomic profiling of the ocular surface using eye wash fluid detected neutrophil marker proteins, supporting recruitment of the immune cell to the site of infection as well as bacteria-specific responses to the host during bacterial biofilm formation of the eye. The study also revealed multiple potential diagnostic markers of *P. aeruginosa* infection associated with biofilm and bacterial vesicles, with potential use in the clinic using noninvasive collection methods. Proteomic profiling of coinfection models (e.g., host infected with multiple pathogens) can aid in the treatment of infections that primarily occur within hospital settings and greatly impact immunocompromised individuals to effectively treat the host for multiple infections at once. This concept was recently explored during infection of macrophages with the fungal pathogen, *C. neoformans*, and the bacterial pathogen, *K. pneumoniae*, both of which, target the lungs of infected patients and employ similar pathogenic protective mechanisms (e.g., formation of polysaccharide capsules) (Sukumaran et al., 2022). This study revealed global reprogramming of host cells upon immediate (e.g., acute) and established (e.g., chronic) singular and co-infections, including initial defense responses to protect macrophages from

invasion followed by adaptation to a new homeostasis prior to further disruption. Further, changes in well-characterized virulence factors (e.g., extracellular enzymes, melanin production) were influenced under differential infection states to modulate host response. Together, these studies suggest opportunities to effectively treat infections at multiple stages by leveraging changes in either the host or pathogen proteomes that represent points of empowerment or weaknesses, respectively, to overcome infections.

2.2 Clinical Applications

Combatting infectious diseases and effectively treating drug-resistant pathogens is critical for global health and food security. To identify proteins and pathways important for pathogenesis and modulation of host immune responses, and to reveal new candidate targets for therapy, mass spectrometry–based proteomics provides an optimal platform. Application of proteomics using in vivo models of disease and with clinical samples provide an opportunity to develop personalized medicine to diagnose, track, and treat infections (Chen et al., 2017; Greco and Cristea, 2017; Muselius et al., 2021). For example, beyond microbe identification by proteotyping, mass spectrometry can detect species-specific candidate biomarker proteins using MALDI-Tandem MS (Wynne et al., 2009) and combinatory approaches, such as PCR-electrospray ionization mass spectrometry (PCR-ESI/MS), has potential for improved diagnostics complementary to cell culture and microscopy (Wolk et al., 2012). Movement toward effective detection of pathogens directly from human biological fluids (e.g., blood, urine) without the need for microbial culturing shows promise in the clinic (Íñigo et al., 2016; la Scola and Raoult, 2009), as well as identification of biomarkers of infection. For instance, cryptococcal infection produces potential clinical biomarkers to detect and diagnose infection from collected blood (Geddes et al., 2016), or tissue samples (Muselius et al., 2023) or host biomarkers of sepsis from profiling of urine (Su et al., 2013), which can also be used to monitor treatment efficacy (Nahid et al., 2014). Further, proteome mapping of pathogen-specific host response patterns in meningitis (i.e., inflammation of the leptomeningeal membranes in the brain caused by pathogenic invasion) from cerebrospinal fluid (CSF) demonstrated signatures of tissue-, cell-, and plasma proteins in the CSF associated with distinct pathogens, leading to the design of a potential diagnostic panel (Bakochi et al., 2021).

Recently, plasma proteomics, which was comprehensively developed to assess human health and disease (Geyer et al., 2016), was applied to COVID-19 patients to predict infections with persistent symptoms (Captur et al., 2022), outline biomarkers of early infection (Staessen et al., 2022), and identify patterns of dysregulation (Ciccosanti et al., 2022). The latter study revealed significant dysregulation of the host inflammatory response and suppression of general metabolic pathways but was unable to provide a clear indication of patient stratification correlated with disease severity. Interestingly, the study also explored plasma depletion prior to proteome analysis and discovered association between neutrophil and platelet activity and acute phase response with disease severity and induction of the complement cascade. Plasma proteomics applied to *Mycobacterium tuberculosis* infection also defined novel candidate biomarkers of infection across independent patient cohorts for the design of a five-protein biosignature for tuberculosis (Garay-Baquero et al., 2020). Another example used discovery mass spectrometry–based proteomics with TMT labeling for quantification combined with a targeted assay to profile serum of active tuberculosis patients and close contacts to establish a panel of protein biomarkers to differentiate active versus latent

infections (Mateos et al., 2020). The signature of active tuberculosis infections included heightened production of proteins related to complement activation and inflammation and modulation of immune responses, along with reduced production of apolipoprotein A and serotransferrin, supporting the importance of lipid transport and iron acquisition in disease progression. Together, these examples support the value of proteomics in the clinic for disease diagnosis, monitoring, and treatment but highlight a disconnect between laboratory discoveries, regulatory approval, and implementation of such tools in the clinic.

2.3 Technological Advances

Over the past two decades, advances in instrumentation and technology associated with mass spectrometry platforms have made substantial strides and directly impact the value and diversity of proteomics experiments to address complex biological problems (Yates, 2019). For instance, adaptation of ion mobility for differential mobility spectrometry or field asymmetric applications are used to create separation of ions to decrease the sample complexity entering the mass spectrometer (Shvartsburg et al., 2014). Additional advances include integration of multiomics techniques for a systems level view of biological processes (Ball et al., 2020), high-throughput methods that lend themselves to clinical applications (Cui et al., 2022; Wang et al., 2022), and single cell proteomics methods that aim to push technological boundaries and enhance the in-depth analysis of samples only available in miniscule quantities (Chappell et al., 2018; Derks et al., 2022; Kelly, 2020; Ma et al., 2020; Mund et al., 2022a,b; Pham et al., 2021; Specht and Slavov, 2018; Vistain and Tay, 2021).

2.3.1 High-Throughput Analyses

The application of mass spectrometry–based proteomics techniques toward infectious disease research is still developing but technological advances are improving our depth of understanding of host-pathogen interactions that drive disease (Jean Beltran et al., 2017). For example, an ultra-high-throughput platform for analysis of serum and plasma proteomes was designed for diagnostic markers of diseases, such as COVID-19, while assuring standardization with international guidelines and facilitation of sample collection and preparation in clinical proteomics laboratories (Messner et al., 2020). The platform is designed to be low-cost, handle close to 200 samples/day, maintain consistency across samples and time, and provide quantitative findings to aid in clinical decision making with potential widespread applications across diseases. For HIV research, the combination of high-throughput approaches with quantitative proteomics has furthered research into diagnostic and prognostic measures. For instance, proteomics supports detection of HIV-specific proteins and general inflammation responses (Rodríguez-Gallego et al., 2019), differentiation between active and latent infections (Kok et al., 2017), and characterizing the role of enzymes (e.g., matrix metalloproteinase) (Ciborowski et al., 2007) and signal transduction cascades (e.g., signal transducer and activator of transcription1) (Rivera et al., 2016) in mediating disease.

2.3.2 Multiomics Integration

For multiomics approaches, profiling of different biological molecules (e.g., gene, transcript, protein, metabolite) can provide insight into mechanistic regulation of both the pathogen and host during infection (Hasin et al., 2017; Prudhomme et al., 2024; Ward et al., 2021). However, many investigations are commonly performed on a single molecular level and retroactively compared

(Freiberg et al., 2016; Oshota et al., 2017), which limits the ability to comprehensively track the progression of a cellular signal from transcript expression to protein abundance to metabolite production. Moreover, this approach may bias the data based on the selected molecular level and mask communication patterns across molecules, therefore supporting a proactive, integrated omics strategy to profile infectious diseases. This concept has been extensively explored over the past few years toward SARS-CoV-2 infections for advances in virology, immunology, and pathogenic mechanisms (Byeon et al., 2022; Lu et al., 2022; Singh et al., 2021; Wang et al., 2021). Other examples include fingerprinting systems-level alterations associated with Crimean-Congo hemorrhagic fever orthonairovirus using a combination of untargeted quantitative proteomics, RNA-sequencing, and network-based analysis (Neogi et al., 2022). This study discovered that blocking two key central carbon and energy metabolism pathways (i.e., glycolysis and glutaminolysis), viral replication was inhibited in vitro and informed a role for type I and II interferon−mediated antiviral mechanisms during progressive viral replication.

2.3.3 Single-Cell Proteomics

For single-cell proteomics, applications toward COVID-19 have been plentiful and informative, profiling changes in patient immune response from minimal sample volumes to observe coordinated immune responses, predict disease severity, and identify potential new therapeutic targets (Stephenson et al., 2021). Other examples integrated multi-omics approaches (e.g., RNAseq, proteomics) with single-cell technology to profile shifts in disease states (Su et al., 2020) and patterns of differentiation between mild versus severe cases (Schulte-Schrepping et al., 2020). For instance, in the latter study, single-cell RNAseq combined with single-cell proteomics of whole- and peripheral-blood mononuclear cells reported changes in immune cell composition and activation across patient samples of COVID-19 comparing mild versus severe disease profiles over time. The study observed drastic changes in neutrophil presentation through the presence of neutrophil precursors associated with myelopoiesis, dysfunction, and altered monocytes affiliated with severe disease compared to interferon-stimulated genetic signature elevation in mild cases. Overall, by exploring and combining each of these technological advances, we gain valuable new insight into pathogenesis, immune response, disease progression, and opportunities for new therapeutic intervention.

3. Combining Proteomics and Bioinformatics Towards Fighting Infectious Diseases

In Section 2, we explored the capabilities and value of mass spectrometry−based proteomics in advancing our understanding and ability to fight infectious diseases. In this section, we highlight the growing applications of machine learning in leveraging and extrapolating proteomics datasets to dig deeper into pathogenic mechanisms of disease, improve diagnostic and prognostic methods, and uncover new strategies to eradicate infections. Importantly, machine learning algorithms and software packages can be used for processing of proteomics datasets, as well as providing new biological insight, biomarkers for discovery, and opportunities to track treatment efficacy (Rahmatbakhsh et al., 2021). Moreover, in silico techniques, including virtual screening and machine learning, have great potential to discover small molecules that can inhibit pathogens, as well as optimize compounds for improved target interactions and therapeutic value (An et al., 2021).

3.1 Diagnostic and Prognostic Markers

A critical area for intervention against infectious diseases and antimicrobial resistance includes the correct and rapid diagnosis of microbes causing disease in the clinic. For urinary tract infections (UTIs), an important confounding factor is the ability of both bacterial and fungal pathogens to cause disease. These infections, which affect over 50% of all women at least once in their lifetime (Foxman, 2010), accounting for 14% of all antibiotic prescriptions in developed countries (Hickling, 2021; Saatchi et al., 2021), cause death in approximately 6.5% of cases if antimicrobial resistance prevents treatment and sepsis develops (Public Health Agency of Canada, 2022). By combining proteomics technology with machine learning and deep learning algorithms, prediction of bacterial species (with 95% accuracy) can be performed in less than 4 hours using MICROB-AI technology (Roux-Dalvai et al., 2019). This approach defines unique signatures for each pathogen based on measurement of large MS-DIA libraries that train machine learning models with BioDiscML (Leclercq et al., 2019) for rapid prediction of the pathogen to enable a high-throughput analysis of 200−300 samples/day.

Another study focused on prediction of antimicrobial resistance profiles for common disease-causing bacteria (Weis et al., 2022). As current antimicrobial resistance testing using culture-based techniques can take up to 72 hours to generate results and discerning resistance profiles of a pathogen is paramount for optimal antibiotic therapy, a novel machine learning approach predicts antimicrobial resistance directly from MALDI-TOF mass spectra profiles of clinical isolates. Here, over 300,000 mass spectra with more than 750,000 antimicrobial resistance phenotypes were used to train calibrated classifiers followed by validation in *Staphylococcus aureus*, *Escherichia coli*, and *K. pneumoniae* followed by retroactive testing on clinical case studies. Another study exemplified the importance of antimicrobial resistance prediction by focusing on methicillin-resistant *S. aureus* and combining MALDI-TOF MS data with a machine learning model for rapid prediction (Yu et al., 2022). This model was evaluated on clinical data with high classification accuracy, supporting its future integration into clinical workflows for proper antibiotic therapy selection. Such technological and computational advances are critical for improving our ability to detect and predict the pathogens and potential antimicrobial resistance profiles responsible for causing disease.

For COVID-19, machine learning models present a valuable opportunity to leverage multi-omics datasets for patient outcome prediction based on disease severity. For a complicated disease that induces highly variable immune responses across individuals, such tools are critical for effective and timely patient care. For instance, profiling of the proteome and metabolome from sera of COVID-19 infected versus healthy individuals integrated with a random forest machine learning model, accurately classified disease severity in 70% of patients (Shen et al., 2020). The model was further validated with targeted proteomics and metabolomics assays in a second test cohort, which increased classification accuracy to 84%. Specifically, dysregulation of macrophages, platelet degranulation, completement system pathway activation, and metabolic suppression were distinctions across patient cohorts. Another study used supervised machine learning approaches with a linear support vector machine model combined with plasma proteomics to classify patients for disease severity associated with COVID-19 (Suvarna et al., 2021). Further, combining plasma proteomics with the development of a new machine learning pipeline (Prioritization of Optimal biomarker Combinations for COVID-19, POC-19) identified 11 individual protein biomarkers, as well as combinatory biomarkers to predict COVID-19 associated patient outcomes derived from surviving and nonsurviving patients with mild to severe symptoms (Shu et al., 2020). Interestingly, each of these studies showed consistent patterns in

host response dysregulation across immune and inflammatory responses, platelet degranulation and coagulation, and metabolism.

More recently, untargeted plasma proteomics analysis was combined with blood cell counts and enzyme activities for time-dependent progression profiling of COVID-19 across 139 patients (Demichev et al., 2021). The study identified prognostic marker signatures for devising risk-adapted treatment strategies using machine learning models based on gradient boosted trees to classify therapeutic needs and validated these findings with an independent cohort. Conversely, a multiplexed targeted mass spectrometry-based assay on a robust triple quadrupole mass spectrometer determined accurate concentrations for hundreds of proteins and metabolites in plasma from hospitalized COVID-19 patients to inform disease severity and mortality (Richard et al., 2022). Here, development of a machine learning multi-OMICs model based on concentrations of 10 proteins and five metabolites was used to predict patient survival made on the day of hospitalization with 92% accuracy. Given the uncertainty of patient outcomes during COVID-19, such standardization assays present valuable information for the clinic to assess optimal treatment options and early intervention strategies to decrease patient mortality rates.

3.2 Target and Drug Discovery

For the discovery of new drugs, target identification and defining the mode of action are two critical components driving success. A recent study used LFQ proteomics of *E. coli* challenged with 19 individual antibiotics to build a comprehensive reference map of proteomic signatures (Yu et al., 2020). These findings were combined with machine learning algorithms, including random forest, LibLinear, LibSVM, IBK, and SMO (Hall et al., 2009) to select the best fit method with the SMO-based method achieving the greatest performance. Optimization of the SMO classifiers led to derivatization of a panel of 14 proteins (regulating diverse bacterial cellular and metabolic processes) for classification of antibiotics into distinct modes of action with close to 100% accuracy to aid in the identification of mechanism of action for unknown compounds in the future.

For HIV research, unbiased approaches, including spatial proteomics, can provide valuable insight into protein relocalization profiles upon infection. Such context was explored using differential centrifugation for fractionation of cellular proteomes combined with testing of machine learning pipelines for classification of the space proteome and identification of protein translocations (Oom et al., 2022). This study determined that the success of machine learning classification varied by organelle within the tested models (i.e., Bayesian t-augmented Gaussian mixture modeling, vector machine learning, QSep analysis). Moreover, the findings showed improved efficacy of the Bayesian model for protein translocation identification and support vector machine classification demonstrated higher consistency and less sensitivity to HIV-dependent noise. Importantly, this study highlights the value of testing and evaluating machine learning algorithms against specific datasets for accurate and insightful predictions.

To identify new druggable targets, mapping protein−protein interactions become important for prioritization of pathogen-specific targets that are not present within the host. Recently, the development of new tools that predict proteome-wide interaction networks by combining structural and sequence similarity, functional association data, and machine learning−based classification, such as PEPPI (Bell et al., 2022), are gaining popularity for computational screening of protein−protein interactions. For bacterial−human protein−protein interactions computational

prediction methods using conventional sequence-based encoding schemes and host network property—related encoding schemes provide improved models in *Yersinia pestis*, supporting new knowledge about pathogenesis and potential drug targets (Lian et al., 2019). For malarial parasites, the discovery of new drug targets is limited by a lack of genome annotation for pathways critical to *Plasmodium falciparum* pathogenesis (Aurrecoechea et al., 2017). By combining proximity biotinylation-based proteomics with an improved neural network prediction algorithm, Apico-PLAST Neural Network (PlastNN) (Boucher et al., 2018), the authors generated a high-confidence proteome of the apicoplast, an unusual organelle for the parasite that is enriched with novel cellular pathways and parasite-specific drug targets (van Dooren and Striepen, 2013). This study revealed proteins with roles in normal blood-stage growth, including an ATP-binding cassette protein, ABCF1, previously unreported in apicoplast biogenesis, with potential for antimalarial drug target prioritization.

4. Concluding Remarks and Future Directions

Immense advances in technology and bioinformatics have transitioned the applications of mass spectrometry—based proteomics from the lab to the clinic. In addition, these advances have expanded the usefulness of proteomics to address several layers of infectious disease research, including how a pathogen infects the host, how the host defends itself from infection, critical points of interaction between the host and pathogen driving disease, and weaknesses in the pathogen that can be targeted for antivirulence strategies to empower the host and clear the infection. Moreover, new druggable targets can be defined, disease progression can be monitored in real-time, and improvements to diagnosis, prognosis, and treatment are being explored. Enormous growth in our understanding of biological processes in microbiology and immunology underpinning the outcome of disease has resulted from such discoveries.

Importantly, the diversity of applications in the field of mass spectrometry—based proteomics now stands at a critical junction: to continue increasing accessibility of the technology through user-friendly and freely available bioinformatics platforms for accurate, reliable, and robust data analysis, while moving into new territory, such as the integration of multi-omics strategies, high-throughput applications, and single-cell technologies, as well as combining proteomics datasets with artificial intelligence methods (e.g., machine learning). Machine learning applied to proteomics data opens a new avenue for discovery in infectious disease research, including diagnostic and prognostic markers to reduce invasiveness of current methodologies, patient stratification for optimal treatment assignment early in disease, as well as the discovery of new targets and drugs to expand the current repertoire of therapeutic options. Critically, appropriate applications and boundaries for machine learning need to be integrated with its use to ensure ethical guidelines are met and interpretation of proteomics datasets is accurate, for optimal and safe clinical care.

Acknowledgments

We thank members of the Geddes-McAlister and Droit labs for constructive feedback on the chapter preparation and extend appreciation to Dr. Jan Rieckmann for writing the R script used to generate Fig. 18.1 and Mr. Ben Muselius (Geddes-McAlister lab) for processing the data.

Author Contributions

J.G.-M. and A.D. conceptualized the chapter. J.G.-M. and F.R.-D. prepared figures. J.G.-M. wrote the first chapter draft. J.G.-M., F.R.-D., and A.D. edited and wrote the final chapter version. All authors have read and approved the submitted manuscript.

Funding

This project is supported, in part, by the University of Guelph, Canadian Institutes of Health Research (Project Grant), Natural Sciences and Engineering Research Council of Canada (Discovery Grant), and the Canada Research Chairs program for J.G.-M., and by Université Laval (Québec, QC), CHU de Québec—Université Laval research center, Génome Québec (Genomic Integration Program) and Fonds de Recherche du Québec—Nature et Technologies (Team Grant) for A.D.

References

Aebersold, R., Mann, M., 2016. Mass-spectrometric exploration of proteome structure and function. Nature 537, 347−355. https://doi.org/10.1038/nature19949.

Aebersold, R., Mann, M., 2003. Mass spectrometry-based proteomics. Nature 422.

An, A.Y., Choi, K.-Y.G., Baghela, A.S., Hancock, R.E.W., 2021. An overview of biological and computational methods for designing mechanism-informed anti-biofilm agents. Front. Microbiol. 12. https://doi.org/10.3389/fmicb.2021.640787.

Andersen, J.S., Wilkinson, C.J., Mayor, T., Mortensen, P., Nigg, E.A., Mann, M., 2003. Proteomic characterization of the human centrosome by protein correlation profiling. Nature 426, 570−574. https://doi.org/10.1038/nature02166.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the unification of biology. Nat. Genet. https://doi.org/10.1038/75556.

Aurrecoechea, C., Barreto, A., Basenko, E.Y., Brestelli, J., Brunk, B.P., Cade, S., Crouch, K., Doherty, R., Falke, D., Fischer, S., Gajria, B., Harb, O.S., Heiges, M., Hertz-Fowler, C., Hu, S., Iodice, J., Kissinger, J.C., Lawrence, C., Li, W., Pinney, D.F., Pulman, J.A., Roos, D.S., Shanmugasundram, A., Silva-Franco, F., Steinbiss, S., Stoeckert, C.J., Spruill, D., Wang, H., Warrenfeltz, S., Zheng, J., 2017. EuPathDB: the eukaryotic pathogen genomics database resource. Nucleic Acids Res. 45, D581−D591. https://doi.org/10.1093/nar/gkw1105.

Bakochi, A., Mohanty, T., Pyl, P.T., Gueto-Tettay, C.A., Malmström, L., Linder, A., Malmström, J., 2021. Cerebrospinal fluid proteome maps detect pathogen-specific host response patterns in meningitis. Elife 10. https://doi.org/10.7554/eLife.64159.

Ball, B., Bermas, A., Carruthers-Lay, D., Geddes-McAlister, J., 2019. Mass spectrometry-based proteomics of fungal pathogenesis, host−fungal interactions, and antifungal development. J. Fungi 5, 52. https://doi.org/10.3390/jof5020052.

Ball, B., Langille, M., Geddes-McAlister, J., 2020. Fun(gi)OMICS: advanced and diverse technologies to explore emerging fungal pathogens. mBio 11, e01020−20.

Ball, B., Woroszchuk, E., Sukumaran, A., West, H., Afaq, A., Carruthers-Lay, D., Muselius, B., Gee, L., Langille, M., Pladwig, S., Kazi, S., Hendriks, A., Geddes-McAlister, J., 2021. Proteome and secretome

profiling of zinc availability in *Cryptococcus neoformans* identifies Wos2 as a subtle influencer of fungal virulence determinants. BMC Microbiol. 21, 341. https://doi.org/10.1186/s12866-021-02410-z.

Bell, E.W., Schwartz, J.H., Freddolino, P.L., Zhang, Y., 2022. PEPPI: whole-proteome protein-protein interaction prediction through structure and sequence similarity, functional association, and machine learning. J. Mol. Biol. 434, 167530. https://doi.org/10.1016/j.jmb.2022.167530.

Bermas, A., Geddes-McAlister, J., 2020. Combatting the evolution of antifungal resistance in *Cryptococcus neoformans*. Mol. Microbiol. 114 (5), 721−734. https://doi.org/10.1111/mmi.14565.

Boucher, M.J., Ghosh, S., Zhang, L., Lal, A., Jang, S.W., Ju, A., Zhang, S., Wang, X., Ralph, S.A., Zou, J., Elias, J.E., Yeh, E., 2018. Integrative proteomics and bioinformatic prediction enable a high-confidence apicoplast proteome in malaria parasites. PLoS Biol. 16, e2005895. https://doi.org/10.1371/journal.pbio.2005895.

Boulund, F., Karlsson, R., Gonzales-Siles, L., Johnning, A., Karami, N., AL-Bayati, O., Åhrén, C., Moore, E.R.B., Kristiansson, E., 2017. Typing and characterization of bacteria using bottom-up tandem mass spectrometry proteomics. Mol. Cell. Proteomics 16, 1052−1063. https://doi.org/10.1074/mcp.M116.061721.

Burnham, C.-A.D., Leeds, J., Nordmann, P., O'Grady, J., Patel, J., 2017. Diagnosing antimicrobial resistance. Nat. Rev. Microbiol. 15, 697−703. https://doi.org/10.1038/nrmicro.2017.103.

Byeon, S.K., Madugundu, A.K., Garapati, K., Ramarajan, M.G., Saraswat, M., Kumar-M, P., Hughes, T., Shah, R., Patnaik, M.M., Chia, N., Ashrafzadeh-Kian, S., Yao, J.D., Pritt, B.S., Cattaneo, R., Salama, M.E., Zenka, R.M., Kipp, B.R., Grebe, S.K.G., Singh, R.J., Sadighi Akha, A.A., Algeciras-Schimnich, A., Dasari, S., Olson, J.E., Walsh, J.R., Venkatakrishnan, A.J., Jenkinson, G., O'Horo, J.C., Badley, A.D., Pandey, A., 2022. Development of a multiomics model for identification of predictive biomarkers for COVID-19 severity: a retrospective cohort study. Lancet Digit Health 4, e632−e645. https://doi.org/10.1016/S2589-7500(22)00112-1.

Captur, G., Moon, J.C., Topriceanu, C.-C., Joy, G., Swadling, L., Hallqvist, J., Doykov, I., Patel, N., Spiewak, J., Baldwin, T., Hamblin, M., Menacho, K., Fontana, M., Treibel, T.A., Manisty, C., O'Brien, B., Gibbons, J.M., Pade, C., Brooks, T., Altmann, D.M., Boyton, R.J., McKnight, Á., Maini, M.K., Noursadeghi, M., Mills, K., Heywood, W.E., 2022. Plasma proteomic signature predicts who will get persistent symptoms following SARS-CoV-2 infection. EBioMedicine 85, 104293. https://doi.org/10.1016/j.ebiom.2022.104293.

Chaabane, F., Graf, A., Jequier, L., Coste, A.T., 2019. Review on antifungal resistance mechanisms in the emerging pathogen *Candida auris*. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02788.

Chalkley, R.J., Baker, P.R., Hansen, K.C., Medzihradszky, K.F., Allen, N.P., Rexach, M., Burlingame, A.L., 2005. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer. Mol. Cell. Proteomics 4, 1189−1193. https://doi.org/10.1074/mcp.D500001-MCP200.

Chappell, L., Russell, A.J.C., Voet, T., 2018. Single-cell (Multi)omics technologies. Annu. Rev. Genom. Hum. Genet. 19, 15−41. https://doi.org/10.1146/annurev-genom-091416-035324.

Chen, B., Zhang, D., Wang, X., Ma, W., Deng, S., Zhang, P., Zhu, H., Xu, N., Liang, S., 2017. Proteomics progresses in microbial physiology and clinical antimicrobial therapy. Eur. J. Clin. Microbiol. Infect. Dis. 36, 403−413. https://doi.org/10.1007/s10096-016-2816-4.

Cheng, K., Ning, Z., Zhang, X., Li, L., Liao, B., Mayne, J., Stintzi, A., Figeys, D., 2017. MetaLab: an automated pipeline for metaproteomic data analysis. Microbiome 5, 157. https://doi.org/10.1186/s40168-017-0375-2.

Choi, H., Larsen, B., Lin, Z.-Y., Breitkreutz, A., Mellacheruvu, D., Fermin, D., Qin, Z.S., Tyers, M., Gingras, A.-C., Nesvizhskii, A.I., 2011. SAINT: probabilistic scoring of affinity purification−mass spectrometry data. Nat. Methods 8, 70−73. https://doi.org/10.1038/nmeth.1541.

Chokshi, A., Sifri, Z., Cennimo, D., Horng, H., 2019. Global contributors to antibiotic resistance. J. Global Infect. Dis. 11, 36. https://doi.org/10.4103/jgid.jgid_110_18.

Ciborowski, P., Kadiu, I., Rozek, W., Smith, L., Bernhardt, K., Fladseth, M., Ricardo-Dukelow, M., Gendelman, H.E., 2007. Investigating the human immunodeficiency virus type 1-infected monocyte-derived macrophage secretome. Virology 363, 198−209. https://doi.org/10.1016/j.virol.2007.01.013.

Ciccosanti, F., Antonioli, M., Sacchi, A., Notari, S., Farina, A., Beccacece, A., Fusto, M., Vergori, A., D'Offizi, G., Taglietti, F., Antinori, A., Nicastri, E., Marchioni, L., Palmieri, F., Ippolito, G., Piacentini, M., Agrati, C., Fimia, G.M., 2022. Proteomic analysis identifies a signature of disease severity in the plasma of COVID-19 pneumonia patients associated to neutrophil, platelet and complement activation. Clin. Proteonomics 19, 38. https://doi.org/10.1186/s12014-022-09377-7.

Council of Canadian Academies, 2019. When Antibiotics Fail: The Expert Panel on the Potential Socio-Economic Impacts of Antimicrobial Resistance in Canada, 2019.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., Mann, M., 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. 10, 1794−1805. https://doi.org/10.1021/pr101065j.

Cui, M., Cheng, C., Zhang, L., 2022. High-throughput proteomics: a methodological mini-review. Lab. Invest. 102, 1170−1181. https://doi.org/10.1038/s41374-022-00830-7.

Davies, S.C., Fowler, T., Watson, J., Livermore, D.M., Walker, D., 2013. Annual report of the Chief Medical Officer: infection and the rise of antimicrobial resistance. Lancet 381, 1606−1609. https://doi.org/10.1016/S0140-6736(13)60604-2.

Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S., Ralser, M., 2020. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat. Methods 17, 41−44. https://doi.org/10.1038/s41592-019-0638-x.

Demichev, V., Tober-Lau, P., Lemke, O., Nazarenko, T., Thibeault, C., Whitwell, H., Röhl, A., Freiwald, A., Szyrwiel, L., Ludwig, D., Correia-Melo, C., Aulakh, S.K., Helbig, E.T., Stubbemann, P., Lippert, L.J., Grüning, N.-M., Blyuss, O., Vernardis, S., White, M., Messner, C.B., Joannidis, M., Sonnweber, T., Klein, S.J., Pizzini, A., Wohlfarter, Y., Sahanic, S., Hilbe, R., Schaefer, B., Wagner, S., Mittermaier, M., Machleidt, F., Garcia, C., Ruwwe-Glösenkamp, C., Lingscheid, T., Bosquillon de Jarcy, L., Stegemann, M.S., Pfeiffer, M., Jürgens, L., Denker, S., Zickler, D., Enghard, P., Zelezniak, A., Campbell, A., Hayward, C., Porteous, D.J., Marioni, R.E., Uhrig, A., Müller-Redetzky, H., Zoller, H., Löffler-Ragg, J., Keller, M.A., Tancevski, I., Timms, J.F., Zaikin, A., Hippenstiel, S., Ramharter, M., Witzenrath, M., Suttorp, N., Lilley, K., Mülleder, M., Sander, L.E., Ralser, M., Kurth, F., Kleinschmidt, M., Heim, K.M., Millet, B., Meyer-Arndt, L., Hübner, R.H., Andermann, T., Doehn, J.M., Opitz, B., Sawitzki, B., Grund, D., Radünzel, P., Schürmann, M., Zoller, T., Alius, F., Knape, P., Breitbart, A., Li, Y., Bremer, F., Pergantis, P., Schürmann, D., Temmesfeld-Wollbrück, B., Wendisch, D., Brumhard, S., Haenel, S.S., Conrad, C., Georg, P., Eckardt, K.-U., Lehner, L., Kruse, J.M., Ferse, C., Körner, R., Spies, C., Edel, A., Weber-Carstens, S., Krannich, A., Zvorc, S., Li, L., Behrens, U., Schmidt, S., Rönnefarth, M., Dang-Heine, C., Röhle, R., Lieker, E., Kretzler, L., Wirsching, I., Wollboldt, C., Wu, Y., Schwanitz, G., Hillus, D., Kasper, S., Olk, N., Horn, A., Briesemeister, D., Treue, D., Hummel, M., Corman, V.M., Drosten, C., von Kalle, C., 2021. A time-resolved proteomic and prognostic map of COVID-19. Cell Syst. 12, 780−794.e7. https://doi.org/10.1016/j.cels.2021.05.005.

Derks, J., Leduc, A., Wallmann, G., Huffman, R.G., Willetts, M., Khan, S., Specht, H., Ralser, M., Demichev, V., Slavov, N., 2022. Increasing the throughput of sensitive proteomics by plexDIA. Nat. Biotechnol. https://doi.org/10.1038/s41587-022-01389-w.

Desiere, F., 2006. The PeptideAtlas project. Nucleic Acids Res. 34, D655−D658. https://doi.org/10.1093/nar/gkj040.

Deutsch, E.W., Mendoza, L., Shteynberg, D., Slagel, J., Sun, Z., Moritz, R.L., 2015. Trans-proteomic pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. Proteonomics Clin. Appl. 9, 745−754. https://doi.org/10.1002/prca.201400164.

Foxman, B., 2010. The epidemiology of urinary tract infection. Nat. Rev. Urol. 7, 653−660. https://doi.org/10.1038/nrurol.2010.190.

Freiberg, J.A., le Breton, Y., Tran, B.Q., Scott, A.J., Harro, J.M., Ernst, R.K., Goo, Y.A., Mongodin, E.F., Goodlett, D.R., McIver, K.S., Shirtliff, M.E., 2016. Global analysis and comparison of the transcriptomes and proteomes of group a *Streptococcus* biofilms. mSystems 1. https://doi.org/10.1128/mSystems.00149-16.

Garay-Baquero, D.J., White, C.H., Walker, N.F., Tebruegge, M., Schiff, H.F., Ugarte-Gil, C., Morris-Jones, S., Marshall, B.G., Manousopoulou, A., Adamson, J., Vallejo, A.F., Bielecka, M.K., Wilkinson, R.J., Tezera, L.B., Woelk, C.H., Garbis, S.D., Elkington, P., 2020. Comprehensive plasma proteomic profiling reveals biomarkers for active tuberculosis. JCI Insight 5. https://doi.org/10.1172/jci.insight.137427.

Geddes, J.M.H., Caza, M., Croll, D., Stoynov, N., Foster, L.J., Kronstad, J.W., 2016. Analysis of the protein kinase a-regulated proteome of *Cryptococcus neoformans* identifies a role for the ubiquitin-proteasome pathway in capsule formation. mBio 7, 1−15. https://doi.org/10.1128/mBio.01862-15.

Geddes-McAlister, J., Prudhomme, N., Dutierrez Gongora, D., Cossar, D., McLean, M.D., 2022. The emerging role of mass spectrometry-based proteomics in molecular pharming. Curr. Opin. Chem. Biol. 68, 102133.

Geladaki, A., Kočevar Britovšek, N., Breckels, L.M., Smith, T.S., Vennard, O.L., Mulvey, C.M., Crook, O.M., Gatto, L., Lilley, K.S., 2019. Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. Nat. Commun. https://doi.org/10.1038/s41467-018-08191-w.

Geyer, P.E., Kulak, N.A., Pichler, G., Holdt, L.M., Teupser, D., Mann, M., 2016. Plasma proteome profiling to assess human health and disease. Cell Syst. 2, 185−195. https://doi.org/10.1016/j.cels.2016.02.015.

Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., Aebersold, R., 2012. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics 11, O111.016717. https://doi.org/10.1074/mcp.O111.016717.

Goodyear, M.C., Garnier, N., Krieger, J.R., Geddes-McAlister, J., Khursigara, C.M., 2021. Label-free quantitative proteomics identifies unique proteomes of clinical isolates of the Liverpool Epidemic Strain of *Pseudomonas aeruginosa* and laboratory strain PAO1. Proteomics Clin. Appl. 15, 2100062. https://doi.org/10.1002/prca.202100062.

Greco, T.M., Cristea, I.M., 2017. Proteomics tracing the footsteps of infectious disease. Mol. Cell. Proteomics 16, S5−S14. https://doi.org/10.1074/mcp.O116.066001.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H., 2009. The WEKA data mining software. ACM SIGKDD Explor. Newsl. 11, 10−18. https://doi.org/10.1145/1656274.1656278.

Hasin, Y., Seldin, M., Lusis, A., 2017. Multi-omics approaches to disease. Genome Biol. 18, 83. https://doi.org/10.1186/s13059-017-1215-1.

Hawkridge, A.M., Muddiman, D.C., 2009. Mass spectrometry−based biomarker discovery: toward a global proteome index of individuality. Annu. Rev. Anal. Chem. 2, 265−277. https://doi.org/10.1146/annurev.anchem.1.031207.112942.

Hickling, D.R., 2021. Recurrent urinary tract infection: rates and prescription patterns in the antibiotic stewardship era. Can. Urol. Assoc. J. 15. https://doi.org/10.5489/cuaj.7697.

Hill, P.J., Scordo, J.M., Arcos, J., Kirkby, S.E., Wewers, M.D., Wozniak, D.J., Torrelles, J.B., 2017. Modifications of *Pseudomonas aeruginosa* cell envelope in the cystic fibrosis airway alters interactions with immune cells. Sci. Rep. 7, 4761. https://doi.org/10.1038/s41598-017-05253-9.

Hu, A., Noble, W.S., Wolf-Yadlin, A., 2016. Technical advances in proteomics: new developments in data-independent acquisition. F1000Res 5, 419. https://doi.org/10.12688/f1000research.7042.1.

Huang, D.W., Sherman, B.T., Lempicki, R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44−57. https://doi.org/10.1038/nprot.2008.211.

Huttlin, E.L., Bruckner, R.J., Navarrete-Perea, J., Cannon, J.R., Baltier, K., Gebreab, F., Gygi, M.P., Thornock, A., Zarraga, G., Tam, S., Szpyt, J., Gassaway, B.M., Panov, A., Parzen, H., Fu, S., Golbazi, A., Maenpaa, E.,

Stricker, K., Guha Thakurta, S., Zhang, T., Rad, R., Pan, J., Nusinow, D.P., Paulo, J.A., Schweppe, D.K., Vaites, L.P., Harper, J.W., Gygi, S.P., 2021. Dual proteome-scale networks reveal cell-specific remodeling of the human interactome. Cell 184, 3022−3040.e28. https://doi.org/10.1016/j.cell.2021.04.011.

Íñigo, M., Coello, A., Fernández-Rivas, G., Rivaya, B., Hidalgo, J., Quesada, M.D., Ausina, V., 2016. Direct identification of urinary tract pathogens from urine samples, combining urine screening methods and matrix-assisted laser desorption ionization−time of flight mass spectrometry. J. Clin. Microbiol. 54, 988−993. https://doi.org/10.1128/JCM.02832-15.

Itzhak, D.N., Tyanova, S., Cox, J., Borner, G.H.H., 2016. Global, quantitative and dynamic mapping of protein subcellular localization. Elife 5. https://doi.org/10.7554/eLife.16950.

Jean Beltran, P.M., Federspiel, J.D., Sheng, X., Cristea, I.M., 2017. Proteomics and integrative omic approaches for understanding host−pathogen interactions and infectious diseases. Mol. Syst. Biol. 13, 922. https://doi.org/10.15252/msb.20167062.

Jennings, E., Thurston, T.L.M., Holden, D.W., 2017. Salmonella SPI-2 type III secretion system effectors: molecular mechanisms and physiological consequences. Cell Host Microbe. https://doi.org/10.1016/j.chom.2017.07.009.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583−589. https://doi.org/10.1038/s41586-021-03819-2.

Karlsson, R., Gonzales-Siles, L., Boulund, F., Svensson-Stadler, L., Skovbjerg, S., Karlsson, A., Davidson, M., Hulth, S., Kristiansson, E., Moore, E.R.B., 2015. Proteotyping: proteomic characterization, classification and identification of microorganisms − a prospectus. Syst. Appl. Microbiol. 38, 246−257. https://doi.org/10.1016/j.syapm.2015.03.006.

Kelly, R.T., 2020. Single-cell proteomics: progress and prospects. Mol. Cell. Proteomics 19, 1739−1748. https://doi.org/10.1074/mcp.R120.002234.

Kim, D.I., Jensen, S.C., Noble, K.A., Kc, B., Roux, K.H., Motamedchaboki, K., Roux, K.J., 2016. An improved smaller biotin ligase for BioID proximity labeling. Mol. Biol. Cell. https://doi.org/10.1091/mbc.E15-12-0844.

Kok, Y.L., Ciuffi, A., Metzner, K.J., 2017. Unravelling HIV-1 latency, one cell at a time. Trends Microbiol. 25, 932−941. https://doi.org/10.1016/j.tim.2017.06.002.

Kollef, M.H., 2008. Broad-spectrum antimicrobials and the treatment of serious bacterial infections: getting it right up front. Clin. Infect. Dis. 47, S3−S13. https://doi.org/10.1086/590061.

Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D., Nesvizhskii, A.I., 2017. MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry−based proteomics. Nat. Methods 14, 513−520. https://doi.org/10.1038/nmeth.4256.

Kristensen, A.R., Gsponer, J., Foster, L.J., 2012. A high-throughput approach for measuring temporal changes in the interactome. Nat. Methods 9, 907−909. https://doi.org/10.1038/nmeth.2131.

Kugadas, A., Geddes-McAlister, J., Guy, E., DiGiandomenico, A., Sykes, D.B., Mansour, M.K., Mirchev, R., Gadjeva, M., 2019. Frontline science: employing enzymatic treatment options for management of ocular biofilm-based infections. J. Leukoc. Biol. https://doi.org/10.1002/JLB.4HI0918-364RR.

la Scola, B., Raoult, D., 2009. Direct identification of bacteria in positive blood culture bottles by matrix-assisted laser desorption ionisation time-of-flight mass spectrometry. PLoS One 4, e8041. https://doi.org/10.1371/journal.pone.0008041.

Larance, M., Lamond, A.I., 2015. Multidimensional proteomics for cell biology. Nat. Rev. Mol. Cell Biol. 16, 269−280. https://doi.org/10.1038/nrm3970.

Leclercq, M., Vittrant, B., Martin-Magniette, M.L., Scott Boyer, M.P., Perin, O., Bergeron, A., Fradet, Y., Droit, A., 2019. Large-scale automatic feature selection for biomarker discovery in high-dimensional OMICs data. Front. Genet. 10. https://doi.org/10.3389/fgene.2019.00452.

Lian, X., Yang, S., Li, H., Fu, C., Zhang, Z., 2019. Machine-learning-based predictor of human−bacteria protein−protein interactions by incorporating comprehensive host-network properties. J. Proteome Res. 18, 2195−2205. https://doi.org/10.1021/acs.jproteome.9b00074.

Liao, B., Ning, Z., Cheng, K., Zhang, X., Li, L., Mayne, J., Figeys, D., 2018. iMetaLab 1.0: a web platform for metaproteomics data analysis. Bioinformatics 34, 3954−3956. https://doi.org/10.1093/bioinformatics/bty466.

Lu, T., Wang, Y., Guo, T., 2022. Multi-omics in COVID-19: seeing the unseen but overlooked in the clinic. Cell Rep. Med. 3, 100580. https://doi.org/10.1016/j.xcrm.2022.100580.

Ma, A., McDermaid, A., Xu, J., Chang, Y., Ma, Q., 2020. Integrative methods and practical challenges for single-cell multi-omics. Trends Biotechnol. 38, 1007−1022. https://doi.org/10.1016/j.tibtech.2020.02.013.

Ma, C., Ren, Y., Yang, J., Ren, Z., Yang, H., Liu, S., 2018. Improved peptide retention time prediction in liquid chromatography through deep learning. Anal. Chem. 90, 10881−10888. https://doi.org/10.1021/acs.analchem.8b02386.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., MacCoss, M.J., 2010. Skyline: an open-source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26, 966−968. https://doi.org/10.1093/bioinformatics/btq054.

Manyi-Loh, C., Mamphweli, S., Meyer, E., Okoh, A., 2018. Antibiotic use in agriculture and its consequential resistance in environmental sources: potential public health implications. Molecules 23, 795. https://doi.org/10.3390/molecules23040795.

Mateos, J., Estévez, O., González-Fernández, Á., Anibarro, L., Pallarés, Á., Reljic, R., Mussá, T., Gomes-Maueia, C., Nguilichane, A., Gallardo, J.M., Medina, I., Carrera, M., 2020. Serum proteomics of active tuberculosis patients and contacts reveals unique processes activated during *Mycobacterium tuberculosis* infection. Sci. Rep. 10, 3844. https://doi.org/10.1038/s41598-020-60753-5.

Meissner, F., Geddes-McAlister, J., Mann, M., Bantscheff, M., 2022. The emerging role of mass spectrometry-based proteomics in drug discovery. Nat. Rev. Drug Discov. 21, 637−654. https://doi.org/10.1038/s41573-022-00409-3.

Messner, C.B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., Textoris-Taube, K., Vernardis, S.I., Egger, A.-S., Kreidl, M., Ludwig, D., Kilian, C., Agostini, F., Zelezniak, A., Thibeault, C., Pfeiffer, M., Hippenstiel, S., Hocke, A., von Kalle, C., Campbell, A., Hayward, C., Porteous, D.J., Marioni, R.E., Langenberg, C., Lilley, K.S., Kuebler, W.M., Mülleder, M., Drosten, C., Suttorp, N., Witzenrath, M., Kurth, F., Sander, L.E., Ralser, M., 2020. Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. Cell Syst. 11, 11−24.e4. https://doi.org/10.1016/j.cels.2020.05.012.

Meyer, J.G., 2021. Deep learning neural network tools for proteomics. Cell Rep. Methods 1, 100003. https://doi.org/10.1016/j.crmeth.2021.100003.

Mizusawa, S., Gottesman, S., 1983. Protein degradation in *Escherichia coli*: the lon gene controls the stability of sulA protein. Proc. Natl. Acad. Sci. U. S. A. https://doi.org/10.1073/pnas.80.2.358.

Mühlen, S., Dersch, P., 2016. Anti-virulence strategies to target bacterial infections. Curr. Top. Microbiol. Immunol. https://doi.org/10.1007/82_2015_490.

Mulani, M.S., Kamble, E.E., Kumkar, S.N., Tawre, M.S., Pardesi, K.R., 2019. Emerging strategies to combat ESKAPE pathogens in the era of antimicrobial resistance: a review. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.00539.

Mund, A., Brunner, A.-D., Mann, M., 2022a. Unbiased spatial proteomics with single-cell resolution in tissues. Mol. Cell 82, 2335−2349. https://doi.org/10.1016/j.molcel.2022.05.022.

Mund, A., Coscia, F., Kriston, A., Hollandi, R., Kovács, F., Brunner, A.-D., Migh, E., Schweizer, L., Santos, A., Bzorek, M., Naimy, S., Rahbek-Gjerdrum, L.M., Dyring-Andersen, B., Bulkescher, J., Lukas, C., Eckert, M.A., Lengyel, E., Gnann, C., Lundberg, E., Horvath, P., Mann, M., 2022b. Deep visual proteomics defines single-cell identity and heterogeneity. Nat. Biotechnol. 40, 1231−1240. https://doi.org/10.1038/s41587-022-01302-5.

Murray, C.J., Ikuta, K.S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S.C., Browne, A.J., Chipeta, M.G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B.H., Kumaran, E.A.P., McManigal, B., Agarwal, R., Akech, S., Albertson, S., Amuasi, J., Andrews, J., Aravkin, A., Ashley, E., Bailey, F., Baker, S., Basnyat, B., Bekker, A., Bender, R., Bethou, A., Bielicki, J., Boonkasidecha, S., Bukosia, J., Carvalheiro, C., Castañeda-Orjuela, C., Chansamouth, V., Chaurasia, S., Chiurchiù, S., Chowdhury, F., Cook, A.J., Cooper, B., Cressey, T.R., Criollo-Mora, E., Cunningham, M., Darboe, S., Day, N.P.J., de Luca, M., Dokova, K., Dramowski, A., Dunachie, S.J., Eckmanns, T., Eibach, D., Emami, A., Feasey, N., Fisher-Pearson, N., Forrest, K., Garrett, D., Gastmeier, P., Giref, A.Z., Greer, R.C., Gupta, V., Haller, S., Haselbeck, A., Hay, S.I., Holm, M., Hopkins, S., Iregbu, K.C., Jacobs, J., Jarovsky, D., Javanmardi, F., Khorana, M., Kissoon, N., Kobeissi, E., Kostyanev, T., Krapp, F., Krumkamp, R., Kumar, A., Kyu, H.H., Lim, C., Limmathurotsakul, D., Loftus, M.J., Lunn, M., Ma, J., Mturi, N., Munera-Huertas, T., Musicha, P., Mussi-Pinhata, M.M., Nakamura, T., Nanavati, R., Nangia, S., Newton, P., Ngoun, C., Novotney, A., Nwakanma, D., Obiero, C.W., Olivas-Martinez, A., Olliaro, P., Ooko, E., Ortiz-Brizuela, E., Peleg, A.Y., Perrone, C., Plakkal, N., Ponce-de-Leon, A., Raad, M., Ramdin, T., Riddell, A., Roberts, T., Robotham, J.V., Roca, A., Rudd, K.E., Russell, N., Schnall, J., Scott, J.A.G., Shivamallappa, M., Sifuentes-Osornio, J., Steenkeste, N., Stewardson, A.J., Stoeva, T., Tasak, N., Thaiprakong, A., Thwaites, G., Turner, C., Turner, P., van Doorn, H.R., Velaphi, S., Vongpradith, A., Vu, H., Walsh, T., Waner, S., Wangrangsimakul, T., Wozniak, T., Zheng, P., Sartorius, B., Lopez, A.D., Stergachis, A., Moore, C., Dolecek, C., Naghavi, M., 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399, 629−655. https://doi.org/10.1016/S0140-6736(21)02724-0.

Muselius, B., Durand, S., Geddes-McAlister, J., 2021. Proteomics of *Cryptococcus neoformans*: from the lab to the clinic. Int. J. Mol. Sci. 22, 12390. https://doi.org/10.3390/ijms222212390.

Muselius, B., Roux-Dalvai, F., Droit, A., Geddes-McAlister, J., 2023. Resolving the temporal splenic proteome during fungal infection for discovery of putative dual perspective biomarker signatures. JASMS 34 (9), 1928−1940. https://doi.org/10.1021/jasms.3c00114.

Muselius, B., Sukumaran, A., Yeung, J., Geddes-McAlister, J., 2020. Iron limitation in *Klebsiella pneumoniae* defines new roles for lon protease in homeostasis and degradation by quantitative proteomics. Front. Microbiol. 11, 546. https://doi.org/10.3389/fmicb.2020.00546.

Nadeau, R., Byvsheva, A., Lavallée-Adam, M., 2021. PIGNON: a protein−protein interaction-guided functional enrichment analysis for quantitative proteomics. BMC Bioinf. 22, 302. https://doi.org/10.1186/s12859-021-04042-6.

Nahid, P., Bliven-Sizemore, E., Jarlsberg, L.G., de Groote, M.A., Johnson, J.L., Muzanyi, G., Engle, M., Weiner, M., Janjic, N., Sterling, D.G., Ochsner, U.A., 2014. Aptamer-based proteomic signature of intensive phase treatment response in pulmonary tuberculosis. Tuberculosis 94, 187−196. https://doi.org/10.1016/j.tube.2014.01.006.

Nahnsen, S., Bielow, C., Reinert, K., Kohlbacher, O., 2013. Tools for label-free peptide quantification. Mol. Cell. Proteomics 12, 549−556. https://doi.org/10.1074/mcp.R112.025163.

Nakayasu, E.S., Gritsenko, M., Piehowski, P.D., Gao, Y., Orton, D.J., Schepmoes, A.A., Fillmore, T.L., Frohnert, B.I., Rewers, M., Krischer, J.P., Ansong, C., Suchy-Dicey, A.M., Evans-Molina, C., Qian, W.-J., Webb-Robertson, B.-J.M., Metz, T.O., 2021. Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. Nat. Protoc. 16, 3737−3760. https://doi.org/10.1038/s41596-021-00566-6.

Neogi, U., Elaldi, N., Appelberg, S., Ambikan, A., Kennedy, E., Dowall, S., Bagci, B.K., Gupta, S., Rodriguez, J.E., Svensson-Akusjärvi, S., Monteil, V., Vegvari, A., Benfeitas, R., Banerjea, A., Weber, F., Hewson, R., Mirazimi, A., 2022. Multi-omics insights into host-viral response and pathogenesis in Crimean-Congo hemorrhagic fever viruses for novel therapeutic target. Elife 11. https://doi.org/10.7554/eLife.76071.

Nicoloff, H., Andersson, D.I., 2013. Lon protease inactivation, or translocation of the lon gene, potentiate bacterial evolution to antibiotic resistance. Mol. Microbiol. https://doi.org/10.1111/mmi.12429.

Oom, A.L., Stoneham, C.A., Lewinski, M.K., Richards, A., Wozniak, J.M., Shams-Ud-Doha, K., Gonzalez, D.J., Krogan, N.J., Guatelli, J., 2022. Comparative analysis of T-cell spatial proteomics and the influence of HIV expression. Mol. Cell. Proteomics 21, 100194. https://doi.org/10.1016/j.mcpro.2022.100194.

Oshota, O., Conway, M., Fookes, M., Schreiber, F., Chaudhuri, R.R., Yu, L., Morgan, F.J.E., Clare, S., Choudhary, J., Thomson, N.R., Lio, P., Maskell, D.J., Mastroeni, P., Grant, A.J., 2017. Transcriptome and proteome analysis of *Salmonella enterica* serovar Typhimurium systemic infection of wild type and immune-deficient mice. PLoS One 12, e0181365. https://doi.org/10.1371/journal.pone.0181365.

Pandey, A., Ding, S.L., Qin, Q.M., Gupta, R., Gomez, G., Lin, F., Feng, X., Fachini da Costa, L., Chaki, S.P., Katepalli, M., Case, E.D., van Schaik, E.J., Sidiq, T., Khalaf, O., Arenas, A., Kobayashi, K.S., Samuel, J.E., Rivera, G.M., Alaniz, R.C., Sze, S.H., Qian, X., Brown, W.J., Rice-Ficht, A., Russell, W.K., Ficht, T.A., de Figueiredo, P., 2017. Global reprogramming of host kinase signaling in response to fungal infection. Cell Host Microbe 21, 637−649.e6. https://doi.org/10.1016/j.chom.2017.04.008.

Patel, R., 2015. MALDI-TOF MS for the diagnosis of infectious diseases. Clin. Chem. 61, 100−111. https://doi.org/10.1373/clinchem.2014.221770.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yilmaz, Ş., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A.F., Ternent, T., Brazma, A., Vizcaíno, J.A., 2019. The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. https://doi.org/10.1093/nar/gky1106.

Pham, T., Tyagi, A., Wang, Y., Guo, J., 2021. Single-cell proteomic analysis. WIREs Mech. Dis. 13. https://doi.org/10.1002/wsbm.1503.

Piehowski, P.D., Zhu, Y., Bramer, L.M., Stratton, K.G., Zhao, R., Orton, D.J., Moore, R.J., Yuan, J., Mitchell, H.D., Gao, Y., Webb-Robertson, B.-J.M., Dey, S.K., Kelly, R.T., Burnum-Johnson, K.E., 2020. Automated mass spectrometry imaging of over 2000 proteins from tissue sections at 100-μm spatial resolution. Nat. Commun. 11, 8. https://doi.org/10.1038/s41467-019-13858-z.

Prudhomme, N., Pastora, R., Thomson, S., Zheng, E., Sproule, A., Krieger, J.R., Murphy, J.P., Overy, D.P., Cossar, D., McLean, M.D., Geddes-McAlister, J., 2024. Bacterial growth-mediated systems remodelling of *Nicotiana benthamiana* defines unique signatures of target protein production in molecular pharming. Plant Biotechnol. J 1−9. https://doi.org/10.1111/pbi.14342.

Public Health Agency of Canada, 2022. Canadian Antimicrobial Resistance Surveillance System Report 2021.

Rahmatbakhsh, M., Gagarinova, A., Babu, M., 2021. Bioinformatic analysis of temporal and spatial proteome alternations during infections. Front. Genet. 12. https://doi.org/10.3389/fgene.2021.667936.

Rasko, D.A., Sperandio, V., 2010. Anti-virulence strategies to combat bacteria-mediated disease. Nat. Rev. Drug Discov. 9, 117−128. https://doi.org/10.1038/nrd3013.

Retanal, C., Ball, B., Geddes-McAlister, J., 2021. Post-translational modifications drive success and failure of fungal−host interactions. J. Fungi. https://doi.org/10.3390/jof7020124.

Rhee, H.W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., Ting, A.Y., 2013. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. Science (1979). https://doi.org/10.1126/science.1230593.

Richard, V.R., Gaither, C., Popp, R., Chaplygina, D., Brzhozovskiy, A., Kononikhin, A., Mohammed, Y., Zahedi, R.P., Nikolaev, E.N., Borchers, C.H., 2022. Early prediction of COVID-19 patient survival by targeted

plasma multi-omics and machine learning. Mol. Cell. Proteomics 21, 100277. https://doi.org/10.1016/j.mcpro.2022.100277.

Rieckmann, J.C., Geiger, R., Hornburg, D., Wolf, T., Kveler, K., Jarrossay, D., Sallusto, F., Shen-Orr, S.S., Lanzavecchia, A., Mann, M., Meissner, F., 2017. Social network architecture of human immune cells unveiled by quantitative proteomics. Nat. Immunol. 18, 583−593. https://doi.org/10.1038/ni.3693.

Rivera, L.E., Kraiselburd, E., Meléndez, L.M., 2016. Cystatin B and HIV regulate the STAT-1 signaling circuit in HIV-infected and INF-β-treated human macrophages. J. Neurovirol. 22, 666−673. https://doi.org/10.1007/s13365-016-0443-6.

Rodríguez-Gallego, E., Tarancón-Diez, L., García, F., del Romero, J., Benito, J.M., Alba, V., Herrero, P., Rull, A., Dominguez-Molina, B., Martinez-Madrid, O., Martin-Pena, L., Pulido, F., León, A., Rodríguez, C., Rallón, N., Peraire, J., Viladés, C., Leal, M., Vidal, F., Ruiz-Mateos, E., 2019. Proteomic profile associated with loss of spontaneous human immunodeficiency virus type 1 elite control. J. Infect. Dis. 219, 867−876. https://doi.org/10.1093/infdis/jiy599.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., Pappin, D.J., 2004. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol. Cell. Proteomics 3, 1154−1169. https://doi.org/10.1074/mcp.M400129-MCP200.

Röst, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinoviä, S.M., Schubert, O.T., Wolski, W., Collins, B.C., Malmström, J., Malmström, L., Aebersold, R., 2014. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat. Biotechnol. https://doi.org/10.1038/nbt.2841.

Röst, H.L., Sachsenberg, T., Aiche, S., Bielow, C., Weisser, H., Aicheler, F., Andreotti, S., Ehrlich, H.C., Gutenbrunner, P., Kenar, E., Liang, X., Nahnsen, S., Nilse, L., Pfeuffer, J., Rosenberger, G., Rurik, M., Schmitt, U., Veit, J., Walzer, M., Wojnar, D., Wolski, W.E., Schilling, O., Choudhary, J.S., Malmström, L., Aebersold, R., Reinert, K., Kohlbacher, O., 2016. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat. Methods. https://doi.org/10.1038/nmeth.3959.

Roux, K.J., Kim, D.I., Burke, B., May, D.G., 2018. BioID: a screen for protein-protein interactions. Curr. Protoc. Protein Sci. 91, 19−23. https://doi.org/10.1002/cpps.51.

Roux, K.J., Kim, D.I., Raida, M., Burke, B., 2012. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. J. Cell Biol. 196, 801−810. https://doi.org/10.1083/jcb.201112098.

Roux-Dalvai, F., Gotti, C., Leclercq, M., Hélie, M.-C., Boissinot, M., Arrey, T.N., Dauly, C., Fournier, F., Kelly, I., Marcoux, J., Bestman-Smith, J., Bergeron, M.G., Droit, A., 2019. Fast and accurate bacterial species identification in urine specimens using LC-MS/MS mass spectrometry and machine learning. Mol. Cell. Proteomics 18, 2492−2505. https://doi.org/10.1074/mcp.TIR119.001559.

Saatchi, A., Yoo, J.W., Marra, F., 2021. Outpatient prescribing and prophylactic antibiotic use for recurrent urinary tract infections in British Columbia, Canada. Can. Urol. Assoc. J. 15. https://doi.org/10.5489/cuaj.7162.

Savitski, M.M., Zinn, N., Faelth-Savitski, M., Poeckel, D., Gade, S., Becher, I., Muelbaier, M., Wagner, A.J., Strohmer, K., Werner, T., Melchert, S., Petretich, M., Rutkowska, A., Vappiani, J., Franken, H., Steidel, M., Sweetman, G.M., Gilan, O., Lam, E.Y.N., Dawson, M.A., Prinjha, R.K., Grandi, P., Bergamini, G., Bantscheff, M., 2018. Multiplexed proteome dynamics profiling reveals mechanisms controlling protein homeostasis. Cell 173, 260−274.e25. https://doi.org/10.1016/j.cell.2018.02.030.

Schmidt, A., Forne, I., Imhof, A., 2014. Bioinformatic analysis of proteomics data. BMC Syst. Biol. 8, S3. https://doi.org/10.1186/1752-0509-8-S2-S3.

Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., de Domenico, E., Wendisch, D., Grasshoff, M., Kapellos, T.S., Beckstette, M., Pecht, T., Saglam, A., Dietrich, O., Mei, H.E., Schulz, A.R., Conrad, C., Kunkel, D., Vafadarnejad, E., Xu, C.-J., Horne, A., Herbert, M., Drews, A., Thibeault, C., Pfeiffer, M., Hippenstiel, S., Hocke, A., Müller-

Redetzky, H., Heim, K.-M., Machleidt, F., Uhrig, A., Bosquillon de Jarcy, L., Jürgens, L., Stegemann, M., Glösenkamp, C.R., Volk, H.-D., Goffinet, C., Landthaler, M., Wyler, E., Georg, P., Schneider, M., Dang-Heine, C., Neuwinger, N., Kappert, K., Tauber, R., Corman, V., Raabe, J., Kaiser, K.M., Vinh, M.T., Rieke, G., Meisel, C., Ulas, T., Becker, M., Geffers, R., Witzenrath, M., Drosten, C., Suttorp, N., von Kalle, C., Kurth, F., Händler, K., Schultze, J.L., Aschenbrenner, A.C., Li, Y., Nattermann, J., Sawitzki, B., Saliba, A.-E., Sander, L.E., Angelov, A., Bals, R., Bartholomäus, A., Becker, A., Bezdan, D., Bonifacio, E., Bork, P., Clavel, T., Colome-Tatche, M., Diefenbach, A., Dilthey, A., Fischer, N., Förstner, K., Frick, J.-S., Gagneur, J., Goesmann, A., Hain, T., Hummel, M., Janssen, S., Kalinowski, J., Kallies, R., Kehr, B., Keller, A., Kim-Hellmuth, S., Klein, C., Kohlbacher, O., Korbel, J.O., Kurth, I., Landthaler, M., Li, Y., Ludwig, K., Makarewicz, O., Marz, M., McHardy, A., Mertes, C., Nöthen, M., Nürnberg, P., Ohler, U., Ossowski, S., Overmann, J., Peter, S., Pfeffer, K., Poetsch, A.R., Pühler, A., Rajewsky, N., Ralser, M., Rieß, O., Ripke, S., Nunes da Rocha, U., Rosenstiel, P., Saliba, A.-E., Sander, L.E., Sawitzki, B., Schiffer, P., Schulte, E.-C., Schultze, J.L., Sczyrba, A., Stegle, O., Stoye, J., Theis, F., Vehreschild, J., Vogel, J., von Kleist, M., Walker, A., Walter, J., Wieczorek, D., Ziebuhr, J., 2020. Severe COVID-19 is marked by a dysregulated myeloid cell compartment. Cell 182, 1419−1440.e23. https://doi.org/10.1016/j.cell.2020.08.001.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498−2504. https://doi.org/10.1101/gr.1239303.

Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., Qian, L., Ge, W., Liu, W., Liang, S., Chen, H., Zhang, Y., Li, J., Xu, J., He, Z., Chen, B., Wang, J., Yan, H., Zheng, Y., Wang, D., Zhu, J., Kong, Z., Kang, Z., Liang, X., Ding, X., Ruan, G., Xiang, N., Cai, X., Gao, H., Li, L., Li, S., Xiao, Q., Lu, T., Zhu, Y., Liu, H., Chen, H., Guo, T., 2020. Proteomic and metabolomic characterization of COVID-19 patient sera. Cell 182, 59−72. https://doi.org/10.1016/j.cell.2020.05.032.

Shu, T., Ning, W., Wu, D., Xu, J., Han, Q., Huang, M., Zou, X., Yang, Q., Yuan, Y., Bie, Y., Pan, S., Mu, J., Han, Y., Yang, X., Zhou, H., Li, R., Ren, Y., Chen, X., Yao, S., Qiu, Y., Zhang, D.-Y., Xue, Y., Shang, Y., Zhou, X., 2020. Plasma proteomics identify biomarkers and pathogenesis of COVID-19. Immunity 53, 1108−1122.e5. https://doi.org/10.1016/j.immuni.2020.10.008.

Shvartsburg, A.A., Ibrahim, Y.M., Smith, R.D., 2014. Differential ion mobility separations in up to 100% helium using microchips. J. Am. Soc. Mass Spectrom. 25, 480−489. https://doi.org/10.1007/s13361-013-0797-4.

Simopoulos, C.M.A., Ning, Z., Li, L., Khamis, M.M., Zhang, X., Lavallée-Adam, M., Figeys, D., 2022. MetaProClust-MS1: an MS1 profiling approach for large-scale microbiome screening. mSystems 7, e0038122. https://doi.org/10.1128/msystems.00381-22.

Singh, R., Singh, P.K., Kumar, R., Kabir, MdT., Kamal, M.A., Rauf, A., Albadrani, G.M., Sayed, A.A., Mousa, S.A., Abdel-Daim, M.M., Uddin, MdS., 2021. Multi-omics approach in the identification of potential therapeutic biomolecule for COVID-19. Front. Pharmacol. 12. https://doi.org/10.3389/fphar.2021.652335.

Sloan, A., Wang, G., Cheng, K., 2017. Traditional approaches versus mass spectrometry in bacterial identification and typing. Clin. Chim. Acta 473, 180−185. https://doi.org/10.1016/j.cca.2017.08.035.

Specht, H., Slavov, N., 2018. Transformative opportunities for single-cell proteomics. J. Proteome Res. 17, 2565−2571. https://doi.org/10.1021/acs.jproteome.8b00257.

Staessen, J.A., Wendt, R., Yu, Y.-L., Kalbitz, S., Thijs, L., Siwy, J., Raad, J., Metzger, J., Neuhaus, B., Papkalla, A., von der Leyen, H., Mebazaa, A., Dudoignon, E., Spasovski, G., Milenkova, M., Canevska-Taneska, A., Salgueira Lazo, M., Psichogiou, M., Rajzer, M.W., Fuławka, Ł., Dzitkowska-Zabielska, M., Weiss, G., Feldt, T., Stegemann, M., Normark, J., Zoufaly, A., Schmiedel, S., Seilmaier, M., Rumpf, B., Banasik, M., Krajewska, M., Catanese, L., Rupprecht, H.D., Czerwieńska, B., Peters, B., Nilsson, Å., Rothfuss, K., Lübbert, C., Mischak, H., Beige, J., Staessen, J.A., Wendt, R., Yu, Y.-L., Kalbitz, S., Thijs, L., Siwy, J., Raad, J., Metzger, J., Neuhaus, B., Papkalla, A., von der Leyen, H., Mebazaa, A., Dudoignon, E., Spasovski, G., Milenkova, M., Canevska-Taneska, A., Lazo, M.S., Psichogiou, M., Rajzer, M.W., Fulawka, L., Dzitkowska-Zabielska, M., Weiss, G., Feldt, T., Stegemann, M., Normark, J., Zoufaly, A., Schmiedel, S.,

Seilmaier, M., Rumpf, B., Banasik, M., Krajewska, M., Catanese, L., Rupprecht, H., Czerwienska, B., Peters, B., Nilsson, Å., Rothfuss, K., Lübbert, C., Mischak, H., Beige, J., Ermisch, J., Kellner, N., Peruth-Stutzmann, L., Schroth, S., Schmidt, J., Schmidt, U., Breuer, D., Abeud, F., Fournier, M.-C., Louadah, B., Molas, R., Rojas, F.L., García, F.A., Sánchez, I.G., Hrom, I.C., Wieczek, A., Schwab, M., K Asayama, K., Hansen, T.W., Maestre, G.E., Basoulis, D., Karamanakos, G., Lis, P., Olszanecka, A., Bellmann-Weiler, R., Lanser, L., Edin, A., Forsell, M.N., Stegmayr, B., Jensen, B.-E.O., Orth, H.-M., Borstel, S., Mikolajewska, A., Hecking, M., Schmölz, L., Hoffmann, M., Narkiewicz, K., Matera-Witkiewicz, A., Zachciał, J., Litwin, M., Marciniak, P., 2022. Predictive performance and clinical application of COV50, a urinary proteomic biomarker in early COVID-19 infection: a prospective multicentre cohort study. Lancet Digit Health 4, e727−e737. https://doi.org/10.1016/S2589-7500(22)00150-9.

Stephenson, E., Reynolds, G., Botting, R.A., Calero-Nieto, F.J., Morgan, M.D., Tuong, Z.K., Bach, K., Sungnak, W., Worlock, K.B., Yoshida, M., Kumasaka, N., Kania, K., Engelbert, J., Olabi, B., Spegarova, J.S., Wilson, N.K., Mende, N., Jardine, L., Gardner, L.C.S., Goh, I., Horsfall, D., McGrath, J., Webb, S., Mather, M.W., Lindeboom, R.G.H., Dann, E., Huang, N., Polanski, K., Prigmore, E., Gothe, F., Scott, J., Payne, R.P., Baker, K.F., Hanrath, A.T., Schim van der Loeff, I.C.D., Barr, A.S., Sanchez-Gonzalez, A., Bergamaschi, L., Mescia, F., Barnes, J.L., Kilich, E., de Wilton, A., Saigal, A., Saleh, A., Janes, S.M., Smith, C.M., Gopee, N., Wilson, C., Coupland, P., Coxhead, J.M., Kiselev, V.Y., van Dongen, S., Bacardit, J., King, H.W., Rostron, A.J., Simpson, A.J., Hambleton, S., Laurenti, E., Lyons, P.A., Meyer, K.B., Nikolić, M.Z., Duncan, C.J.A., Smith, K.G.C., Teichmann, S.A., Clatworthy, M.R., Marioni, J.C., Göttgens, B., Haniffa, M., 2021. Single-cell multi-omics analysis of the immune response in COVID-19. Nat. Med. 27, 904−916. https://doi.org/10.1038/s41591-021-01329-2.

Strauss, M.T., Bludau, I., Zeng, W.-F., Voytik, E., Ammar, C., Schessner, J., Ilango, R., Gill, M., Meier, F., Willems, S., Mann, M., 2021. AlphaPept, a modern and open framework for MS-based proteomics. bioRxiv, 453379. https://doi.org/10.1101/2021.07.23.453379.

Su, L., Zhou, R., Liu, C., Wen, B., Xiao, K., Kong, W., Tan, F., Huang, Y., Cao, L., Xie, L., 2013. Urinary proteomics analysis for sepsis biomarkers with iTRAQ labeling and two-dimensional liquid chromatography−tandem mass spectrometry. J. Trauma Acute Care Surg. 74, 940−945. https://doi.org/10.1097/TA.0b013e31828272c5.

Su, Y., Chen, D., Yuan, D., Lausted, C., Choi, J., Dai, C.L., Voillet, V., Duvvuri, V.R., Scherler, K., Troisch, P., Baloni, P., Qin, G., Smith, B., Kornilov, S.A., Rostomily, C., Xu, A., Li, J., Dong, S., Rothchild, A., Zhou, J., Murray, K., Edmark, R., Hong, S., Heath, J.E., Earls, J., Zhang, R., Xie, J., Li, S., Roper, R., Jones, L., Zhou, Y., Rowen, L., Liu, R., Mackay, S., O'Mahony, D.S., Dale, C.R., Wallick, J.A., Algren, H.A., Zager, M.A., Wei, W., Price, N.D., Huang, S., Subramanian, N., Wang, K., Magis, A.T., Hadlock, J.J., Hood, L., Aderem, A., Bluestone, J.A., Lanier, L.L., Greenberg, P.D., Gottardo, R., Davis, M.M., Goldman, J.D., Heath, J.R., 2020. Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. Cell 183, 1479−1495.e20. https://doi.org/10.1016/j.cell.2020.10.037.

Sukumaran, A., Ball, B., Krieger, J.R., Geddes-McAlister, J., 2022. Cross-kingdom infection of macrophages reveals pathogen- and immune-specific global reprogramming and adaptation. mBio 13. https://doi.org/10.1128/mbio.01687-22.

Sukumaran, A., Coish, J.M., Yeung, J., Muselius, B., Gadjeva, M., MacNeil, A.J., Geddes-McAlister, J., 2019. Decoding communication patterns of the innate immune system by quantitative proteomics. J. Leukoc. Biol. https://doi.org/10.1002/JLB.2RI0919-302R.

Sukumaran, A., Pladwig, S., Geddes-McAlister, J., 2021. Zinc limitation in *Klebsiella pneumoniae* profiled by quantitative proteomics influences transcriptional regulation and cation transporter-associated capsule production. BMC Microbiol. 21, 1−15. https://doi.org/10.1186/s12866-021-02091-8.

Sukumaran, A., Woroszchuk, E., Ross, T., Geddes-McAlister, J., 2020. Proteomics of host-bacterial interactions: new insights from dual perspectives. Can. J. Microbiol. 1−43.

Suvarna, K., Biswas, D., Pai, M.G.J., Acharjee, A., Bankar, R., Palanivel, V., Salkar, A., Verma, A., Mukherjee, A., Choudhury, M., Ghantasala, S., Ghosh, S., Singh, A., Banerjee, A., Badaya, A., Bihani, S., Loya, G., Mantri, K., Burli, A., Roy, J., Srivastava, A., Agrawal, S., Shrivastav, O., Shastri, J., Srivastava, S., 2021. Proteomics and machine learning approaches reveal a set of prognostic markers for COVID-19 severity with drug repurposing potential. Front. Physiol. 12. https://doi.org/10.3389/fphys.2021.652799.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J., von Mering, C., 2019. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47, D607−D613. https://doi.org/10.1093/nar/gky1131.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Hamon, C., 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1895−1904. https://doi.org/10.1021/ac0262560.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M., Cox, J., 2016. The perseus computational platform for comprehensive analysis of (prote) omics data. Nat. Methods 13, 731−740.

van Dooren, G.G., Striepen, B., 2013. The algal past and parasite present of the apicoplast. Annu. Rev. Microbiol. 67, 271−289. https://doi.org/10.1146/annurev-micro-092412-155741.

Virreira Winter, S., Meier, F., Wichmann, C., Cox, J., Mann, M., Meissner, F., 2018. EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. Nat. Methods 15, 527−530. https://doi.org/10.1038/s41592-018-0037-8.

Vistain, L.F., Tay, S., 2021. Single-cell proteomics. Trends Biochem. Sci. 46, 661−672. https://doi.org/10.1016/j.tibs.2021.01.013.

Walch, P., Selkrig, J., Knodler, L.A., Rettel, M., Stein, F., Fernandez, K., Viéitez, C., Potel, C.M., Scholzen, K., Geyer, M., Rottner, K., Steele-Mortimer, O., Savitski, M.M., Holden, D.W., Typas, A., 2021. Global mapping of *Salmonella enterica*-host protein-protein interactions during infection. Cell Host Microbe 29, 1316−1332.e12. https://doi.org/10.1016/j.chom.2021.06.004.

Wang, H., Drake, S.K., Yong, C., Gucek, M., Tropea, M., Rosenberg, A.Z., Dekker, J.P., Suffredini, A.F., 2016. A novel peptidomic approach to strain typing of clinical *Acinetobacter baumannii* isolates using mass spectrometry. Clin. Chem. 62, 866−875. https://doi.org/10.1373/clinchem.2015.253468.

Wang, X., Xu, G., Liu, X., Liu, Y., Zhang, S., Zhang, Z., 2021. Multiomics: unraveling the panoramic landscapes of SARS-CoV-2 infection. Cell. Mol. Immunol. 18, 2313−2324. https://doi.org/10.1038/s41423-021-00754-0.

Wang, Z., Mülleder, M., Batruch, I., Chelur, A., Textoris-Taube, K., Schwecke, T., Hartl, J., Causon, J., Castro-Perez, J., Demichev, V., Tate, S., Ralser, M., 2022. High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. Elife 11. https://doi.org/10.7554/eLife.83947.

Ward, R.A., Aghaeepour, N., Bhattacharyya, R.P., Clish, C.B., Gaudillière, B., Hacohen, N., Mansour, M.K., Mudd, P.A., Pasupneti, S., Presti, R.M., Rhee, E.P., Sen, P., Spec, A., Tam, J.M., Villani, A.-C., Woolley, A.E., Hsu, J.L., Vyas, J.M., 2021. Harnessing the potential of multiomics studies for precision medicine in infectious disease. Open Forum Infect. Dis. 8. https://doi.org/10.1093/ofid/ofab483.

Watson, K.G., Holden, D.W., 2010. Dynamics of growth and dissemination of *Salmonella* in vivo. Cell Microbiol. https://doi.org/10.1111/j.1462-5822.2010.01511.x.

Weis, C., Cuénod, A., Rieck, B., Dubuis, O., Graf, S., Lang, C., Oberle, M., Brackmann, M., Søgaard, K.K., Osthoff, M., Borgwardt, K., Egli, A., 2022. Direct antimicrobial resistance prediction from clinical MALDI-TOF mass spectra using machine learning. Nat. Med. 28, 164−174. https://doi.org/10.1038/s41591-021-01619-9.

Werner, J., Géron, A., Kerssemakers, J., Matallana-Surget, S., 2019. mPies: a novel metaproteomics tool for the creation of relevant protein databases and automatized protein annotation. Biol. Direct 14, 21. https://doi.org/10.1186/s13062-019-0253-x.

Wolk, D.M., Kaleta, E.J., Wysocki, V.H., 2012. PCR−electrospray ionization mass spectrometry. J. Mol. Diagn. 14, 295−304. https://doi.org/10.1016/j.jmoldx.2012.02.005.

Woods, M., McAlister, J.A., Geddes-McAlister, J., 2023. A one health approach to overcoming fungal disease and antifungal resistance. WIREs Mech. Dis. 15 (4), e1610. https://doi.org/10.1002/wsbm.1610.

World Health Organization, 2021. Antimicrobial Resistance.

Wynne, C., Fenselau, C., Demirev, P.A., Edwards, N., 2009. Top-down identification of protein biomarkers in bacteria with unsequenced genomes. Anal. Chem. 81, 9633−9642. https://doi.org/10.1021/ac9016677.

Yates, I.J.R., 2019. Recent technical advances in proteomics. F1000Res 8, 351. https://doi.org/10.12688/f1000research.16987.1.

Yeung, J., Gadjeva, M., Geddes-McAlister, J., 2020. Label-free quantitative proteomics distinguishes general and site-specific host responses to *Pseudomonas aeruginosa* infection at the ocular surface. Proteomics. https://doi.org/10.1002/pmic.201900290.

Yu, J., Tien, N., Liu, Y.-C., Cho, D.-Y., Chen, J.-W., Tsai, Y.-T., Huang, Y.-C., Chao, H.-J., Chen, C.-J., 2022. Rapid identification of methicillin-resistant *Staphylococcus aureus* using MALDI-TOF MS and machine learning from over 20,000 clinical isolates. Microbiol. Spectr. 10. https://doi.org/10.1128/spectrum.00483-22.

Yu, Y., O'Rourke, A., Lin, Y.-H., Singh, H., Eguez, R.V., Beyhan, S., Nelson, K.E., 2020. Predictive signatures of 19 antibiotic-induced *Escherichia coli* proteomes. ACS Infect. Dis. 6, 2120−2129. https://doi.org/10.1021/acsinfecdis.0c00196.

Zeng, W.-F., Zhou, X.-X., Willems, S., Ammar, C., Wahle, M., Bludau, I., Voytik, E., Strauss, M.T., Mann, M., 2022. AlphaPeptDeep: a modular deep learning framework to predict peptide properties for proteomics. Nat. Commun. 13, 7238. https://doi.org/10.1038/s41467-022-34904-3.

Zhou, X.-X., Zeng, W.-F., Chi, H., Luo, C., Liu, C., Zhan, J., He, S.-M., Zhang, Z., 2017. pDeep: predicting MS/MS spectra of peptides with deep learning. Anal. Chem. 89, 12690−12697. https://doi.org/10.1021/acs.analchem.7b02566.

CHAPTER

Genomics of Infectious Diseases and Private Industry

19

Alex van Belkum[1] **and Guy Vernet**[2]

[1]*BaseClear, Leiden, the Netherlands;* [2]*Institut Pasteur de Bangui, Bangui, Central African Republic*

1. Introduction

Genomics comprises the science focusing on structure, expression, evolution, and mapping of genomes of essentially all living organisms. The main current driver for companies that specialize in eukaryotic genomics is human genome sequencing and annotation. Still, genomics also concerns full genome sequencing and bio-informatic analysis of the sequencing data for microorganisms. This includes the integral sequencing of microbiomes as well as targeted or whole-exome sequencing, single-cell DNA profiling, transcriptome sequencing, targeted gene expression profiling, miRNA and small RNA analysis, and DNA–protein interaction analysis. Genomic science feeds basic and applied research, pharmaceutical R&D, investigations into personalized medicine, genealogy and diagnosis, surveillance, and the continuous monitoring of globally disseminating infectious diseases.

Most of the nucleic acid sequencing technology breakthroughs have led to improvements in the speed and quality of human Whole Genome Sequencing (WGS) and to reductions in its costs and experimental turnaround time. Consequently, noneukaryotic genomics, and especially pathogen genomics, has made considerable progress in the recent years. The first complete genome ever sequenced was that from an RNA-virus, the bacteriophage MS2 (Fiers et al., 1976). This was just 1 year before the invention of dideoxy chain termination sequencing by Frederic Sanger, which, in principle, was the first major technological breakthrough in pathogen genomics (Sanger et al., 1977). The first bacterial WGS was obtained in 1995 for *Haemophilus influenzae* (Fleischmann et al., 1995). Many have followed during the past decades and sequencing plus annotation of a microbial genome has become a routine laboratory activity.

Detecting genomic variation: The availability, through extensive genome sequencing, of reference genomic data for prokaryotic and eukaryotic organisms has allowed the identification by resequencing of genomic mutations including Single Nucleotide Polymorphisms (SNPs) and the detection of more extensive chromosomal rearrangements, insertions or deletions, which all can have a profound impact on microbial phenotypes. Detection of such genomic variation has major diagnostic impact. For instance, the *Mycobacterium tuberculosis* Drug Resistance tests (MTBDR; Hain Lifescience, Bruker, Nehren, Germany) which are available since 2001 and which are based on the detection of chromosomal variability, are still in use in many clinical laboratories (Kardan-Yamchi et al., 2021). It is a simple line probe assay detecting signature SNPs for *Mycobacterium tuberculosis* resistance to the

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00022-7

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

major antibiotics. Microarrays have also been considered a powerful tool for such diagnosis and genotyping (resistance mutations, molecular epidemiology). However, they are now mainly used for targeted sequencing and transcriptome analysis, mostly of animals and plants. Multiplex PCR technologies (covering 15−50 targets) which partially or fully integrate automatic sample preparation, amplification, detection, and analysis such as the xMAP platforms (Luminex, DiaSorin, Austin, TX, USA), the BIOFIRE FILMARRAY platform (bioMérieux, Salt-Lake City, UT, USA) or the QIAstat-Dx platform (QIAGEN, Hilden, Germany) have the potential to detect multiple pathogens in a syndrome-based approach including functionally relevant SNPs.

Next-Generation Sequencing: A further breakthrough in genomics has been the development of Next-Generation Sequencing (NGS) in the late 1990s of the former century. The short length of reads generated by the first NGS technologies, either sequencing-by-synthesis (SBS) or semiconductor sequencing, made them resequencing rather than de novo sequencing technologies. However, with the availability of more reference genomes to compare with, NGS rapidly became a very powerful technology for high-throughput whole-genome or targeted sequencing as well as for all applications listed above. This further enhanced both the direct and indirect influence of sequencing on development in the in vitro diagnostic (IVD) industry.

Interestingly, NGS technologies and instruments were first used to sequence the complete genomes of two opportunistic human pathogens, *Mycoplasma genitalia* (Margulies et al., 2005) and *Escherichia coli* (Shendure et al., 2005). Hundreds of thousands of genomes from viruses, bacteria, and protists have now been fully or partially sequenced. The International Nucleotide Sequence Database Collaboration (INSDC; https://www.insdc.org//) has been created by the DNA Databank of Japan, the European EMBL's European Bioinformatics Institute and the US National Center for Biotechnology Information (NCBI) to deposit and access these sequences. Other databases, such as the Global Initiative on Sharing Avian Influenza Data (GISAID) for influenza viruses and more recently several SARS-CoV-2 genome data depositories, target specific pathogens (https://www.gisaid.org/; http://www.cogconsortium.uk/).

Commercial sequencing facilities: The cost and time-to-result of WGS both decreased dramatically during recent years. Sanger sequencing has been used for the 13-year Human Genome Project, which resulted in the first human genome in 2003 for a budget of about $2.7 billion, about a dollar per base pair. According to NIH, in 2021, the cost per human "draft" genome, i.e., covering ∼90% of the genome at ∼99.9% accuracy, approaches a figure less than 1000$ (Wetterstrand, 2021). In practice, is all conditions are favorable, this can be done within a working day even. Genomics service companies provide custom bacterial whole-genome sequencing at a cost of ∼60$/genome and do-it-yourself sequencing costs probably less than 20$/genome currently. Theoretically, WGS now takes hours per genome although in practice bacterial WGS usually takes a couple of days. Some genomics service companies guaranty a 2- or even 1-week turn-around time for bacterial WGS after sample reception.

Genomics companies have considerably changed since the publication of the previous edition of this book in 2016. Manufacturers often propose a range of instruments covering different specific application needs and tunable throughput. They now mostly target benchtop or even smaller format instruments. The currently available portable sequencers allow sequencing in clinical laboratories for diagnosis as well as at the point of care during epidemics for implementation of detailed genomic surveillance. Companies are usually qualified as Certified Service Providers, endorsed by the companies that provide the sequencing equipment. NGS has been especially helpful during the COVID-19 pandemic to monitor the emergence and spread of SARS-CoV-2 variants. Manufacturers propose

reagents and software covering the entire process from nucleic acids extraction to gene annotation. Some companies specialize on preanalytical instruments or on reagents only, others have a more generic offer. Some also provide services under the form of certification of Genomics Service Companies (GSC). Numerous GSC have been created in the last decade. A few of them specializes on microbial genomics services. Finally, Direct-to-Consumer companies are addressing the needs of individual consumers for personalized medicine and genealogy.

The size of the genomics market (including PCR and nucleic acids extraction) was estimated between 20 and 23 billion $ in 2020 and the annual growth rates until 2027 was estimated to be somewhere between 9.9% and 19% (https://www.fortunebusinessinsights.com/industry-reports/genomics-market-100941; https://www.grandviewresearch.com/industry-analysis/genomics-market; https://www.marketdataforecast.com/market-reports/global-genomics-market). Products and services related to nucleic acid sequencing will likely represent the highest value in this market.

The genomics market now provides numerous products and services for customers interested in infectious diseases. During the COVID-19 pandemic, genomics has shown that it is an indispensable tool, not only for research but also for public health monitoring and even individual patient management (Saravanan et al., 2022). The expected strong growth of the genomics market suggests that there will be continuous changes in the coming 5 years. Fig. 19.1 shows how private and public actors involved in genomics interact.

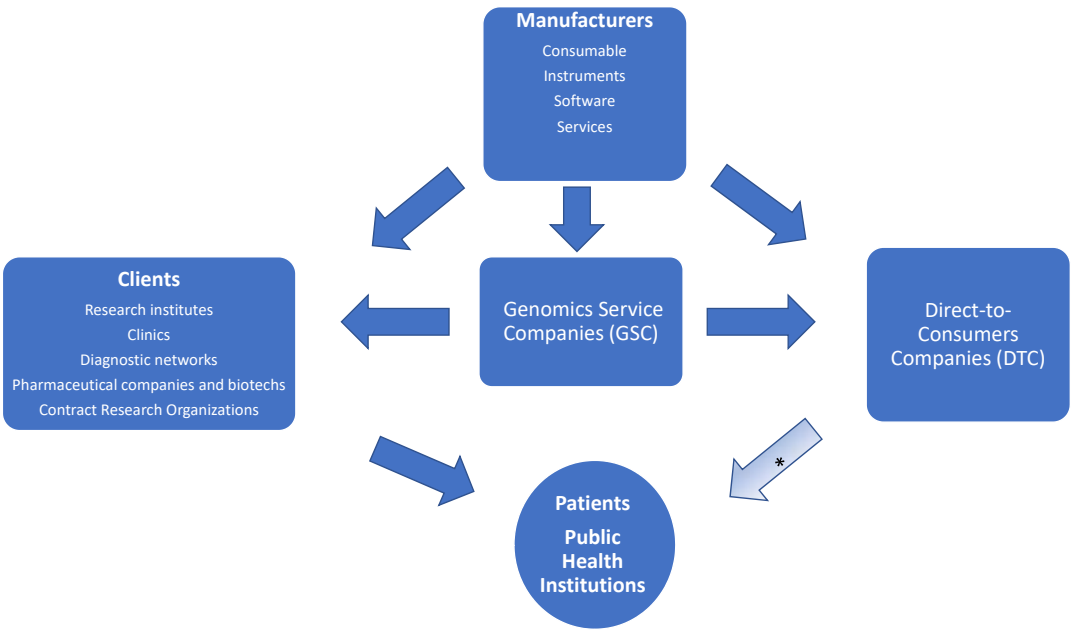


FIGURE 19.1

Interactions of NGS actors.

2. Customers and Applications

Pathogen genomics: Customers of commercial companies involved in pathogen genomics can both be public or private organizations. Final users may be either research laboratories, clinical laboratories, forensic laboratories, public health laboratories, hospital laboratories involved in infection control and pharmaceutical but also food-oriented companies. However, because pathogen genomics integrates different competencies for preanalytical, analytical, and ultimate data analysis, and because costly investments are needed in equipment and software, major customers of companies that commercialize genomics solutions are very often organizations that provide services to the above cited laboratories (Genomics Service Companies-GSC), or even to individuals (Direct-to-Consumers Companies-DTC) (see below). Still, because of the evolution toward the use of benchtop formats, even the smaller laboratories may also equip themselves and turn toward less outsourcing of sequencing services.

Research laboratories have been the first to use genomic tools to study human, animal and plant pathogens. The major applications are WGS of viruses, bacteria, parasites, and fungi, and exome and large gene panel sequencing following enrichment of pathogens and targeted (often amplicon-based) gene sequencing. An important current need of microbiology scientists is to sequence microbial genomes and build databases consisting of genome sequences. WGS has been made possible by using Sanger sequencing first and now next-generation, short-reads or single-molecule sequencing have been made available. Some of these polymorphisms (co-)define phenotypic characteristics of a strain or an isolate (Herring and Palsson, 2007). Scientists may use these data for fundamental research: to annotate all genes and better understand their function, to comprehend genome organization, to taxonomically classify species, and to study their evolution and overall genome dynamics. More fundamental approaches such as single cell profiling, transcriptome sequencing, targeted gene expression profiling, miRNA and small RNA analysis, DNA−protein interaction analysis or 16S metagenomic sequencing allow microbiology scientists to study the response of hosts to the infection by microbes and host−pathogens interactions. The knowledge of genome organization, combined with functional genomics is the basis to understand the detailed pathology of infectious agents. Among downstream applications are the development of new antimicrobial drugs, vaccines, and improvement of in vitro diagnostics including the establishment of resistance and escape or virulence profiles.

Microbiome research: A new field of investigation is the study of microbiomes, often referred to as "metagenomics" (Holt and Jones, 2008), which consists in the systematic sequencing of all nucleic acids originating from microorganisms in a given ecological "niche"—gut, upper respiratory tracts, and skin but also environmental samples—to identify and further characterize all microbes present. This allows to define the "normal" resident flora, for example at different ages in life, and interactions between this flora in pathogenic agents before, during and after infection, in particular the exchange of genetic material conferring resistance and virulence. Innovative applications by pharmaceutical, diagnostic, and agro-food companies are very important. The first studies showing correlations between the composition of the microbiome and certain clinical conditions have been published to date. This type of information points toward possible use of certain microorganisms as life biotherapeutic agents and this is opening up an entirely new field of microbiological intervention. For instance, recurrent infection with *Clostridioides difficile* in patients with life-threatening gut conditions can be very adequately treated with fecal transplantation. Exposing patients to what is considered to be healthy flora has strong prophylactic and therapeutic effect (Rakotonirina et al., 2022). Many additional therapies are being designed and several of these will reach clinical practice anytime soon.

It is beyond the scope of this chapter to provide a full review of this huge field of investigation. Next to niche specific research increasing effort is invested in the assessment and functional description using all genomics tools (holo-omics) of the holomicrobiome, the concerted presence and action of the various niche specific microbiota that have been described thus far (Wilson et al., 2019). Research on the correlation between the nature of the holomicrobiome and diverse (infectious) pathologies will be deciphered in increasing detail over the years to come. This will generate new hypothesis for the relevance of microorganisms on overall human, animal, and plant health, also including diseases that are currently being considered as being noninfectious or nonmicrobiological in origin. This will range from the inflammatory bowel diseases to Alzheimer and Parkinson diseases and many others. New paradigms for the relevance of our microbiota in all pathologies will emerge with regularity.

Genomics and diagnostics: Healthcare facilities are using genomic tools with increasing frequency. Genome sequencing is generally not the first-in-line technology to identify a bacterial infection (except maybe for mycobacterial infections in some specific cases), whereas viral diagnosis often makes use of molecular nucleic acid amplification-based diagnostics and then sequencing may come in if higher resolution is needed. On the contrary, metagenomics NGS (mNGS) may prove useful in analyzing all sequences from a given sample and identifying pathogens that could be at the origin of the disease without any prior suspicion and the need for a culture step (for a review, see Nyholm et al., 2020). This approach is excellently suited for the detection of new and unknown or otherwise hard to detect pathogens in clinical specimens. However, there are still several practical obstacles. First, the excessive number of short sequences generated requires complex computational approaches for sequence assembly which are costly and time-consuming. In several cases, it is also difficult to separate microbial DNA from human DNA and the amount of DNA from a given pathogen may be very low, requiring complex extraction methods to allow the high depth coverage of its genome required for assembly. This indicates that improvements in the preanalytical phase of NGS-based diagnostic are urgently required. Finally, the detection of a genome is not sufficient to identify it as the true or unique cause of a disease. As a matter of fact, many human pathogens also know a more innocent colonization status. Again, bio-informatic approaches for distinguishing the infectious from the colonization state are urgently needed and this may be based on sequencing of host more than pathogen-derived nucleic acids (Govender et al., 2021).

Genomic antimicrobial susceptibility testing: The use of WGS has been clearly shown to be of value for antimicrobial susceptibility testing (Rogers et al., 2021). It has been shown that WGS of *Klebsiella pneumoniae* and *Escherichia coli* isolates was able to predict resistance to different antibiotics with a correlation of 97% with phenotypic testing (Stoesser et al., 2013). Of note, Public Health England (PHE) has now turned to WGS for the detection of antibiotic resistance to first-line antimicrobial agents in tuberculosis and bas embedded WGS in its strategy to control outbreaks, and provide improved information for the diagnosis and treatment of infectious diseases (https://www.gov.uk/government/publications/phe-infectious-diseases-strategy). All screening cultures that turn positive are translated into a mycobacterial genome sequence and this sequence is then screened for resistance associated mutations. It has been shown that sensitivity and specificity of this method is at least as good as classical phenotypic methods and equally important, this new method is also cost-effective. Of course, this use of WGS requires prior culture and updated databases of resistance genes (Walker et al., 2022). In addition, new solutions for detecting resistance markers with increasing efficacy and specificity are in constant need. Recently, the use of k-mer mediated approaches has proven useful in this respect (Jaillard et al., 2018).

WGS-based surveillance: Genomics already finds applications in the public health sector, for example in the surveillance and control of emerging infectious diseases, including those caused by resistant pathogens. An international study conducted by scientists from the Wellcome Sanger Institute, Emory University (Atlanta, USA), and the US Centers for Disease Control and Prevention on more than 3000 samples has shown that WGS sequencing of *Streptococcus pneumoniae* strains can adequately predict the impact of pneumococcal vaccination (Lo et al., 2019). Genomic epidemiology has already been proven useful, and probably easier to use than current typing methods, for outbreak investigations in the hospital ecosystem or the community (Luheshi et al., 2015). By providing hospitals an improved understanding of the genetic markers of virulence and resistance, this service can help them understand how bacteria are transmitted, while enabling better containment of ongoing outbreaks, limiting the spread of infectious agents, and improving surveillance approaches.

Genomics can be used by pharmaceutical and food companies for both R&D and quality assessment and control. Of course, results generated by research laboratories using genomic technologies can help find industrial applications such as new and better tailored drugs and vaccines as well as improved environmental hygiene monitoring and maintenance. Reverse vaccinology is a good example (Masignani et al., 2019). Pharmaceutical companies which need to verify sterility of injectable drugs and vaccines, for example, vaccines produced by viral culture, may take advantage of metagenomics analysis of their bulk solutions. Diagnostic companies can base the identification of new biomarkers for diagnosis or treatment or epidemiological monitoring of diseases on genomics. For the food industry a highly parallelized workflow has been developed where WGS is the final step in the analysis of microbial contaminants. Once organisms have been cultured from foodstuffs or consumers (with a food-related infection) WGS is used for molecular tracing. Linking contamination in food with clinical diseases is extremely important for improvement of the food production process and for prevention of future food-associated disease outbreaks. Also, for tracing already contaminated food and for stopping dissemination of ongoing outbreaks WGS is an essential tool. Finally, WGS data can be used as a forensic tool in court cases of litigation.

3. Commercial Sequencing Instruments, Reagents, and Software

Since the first description of nucleic acid sequencing technologies by Maxam and Gilbert in 1977 (Maxam and Gilbert, 1977) and a little later by Sanger et al. (1977), continuous progress has been made and the offer of chemistries, instruments and software by major manufacturers now use a variety of different approaches. For a very detailed description of these technologies there are excellent technical reviews available (Slatko et al., 2018), so we will here only give a relatively brief review of key aspects of the currently most popular laboratory methodologies.

Sanger sequencing: Sanger sequencing uses the sequencing-by-synthesis (SBS) approach in which a DNA polymerase generates new DNA reads from a template which is the targeted DNA molecule to analyze. The nature of the nucleotide at a given position is usually targeted by sequence specific oligonucleotides that serve as sequencing starting points and determined using specific dyes that are coupled to chain-terminating nucleotide triphosphates. Frederic Sanger was awarded a Nobel prize, and the huge impact of this technology is still reflected by the fact that currently automated Sanger sequencing remains a preferred option for many sequencing applications.

Sanger sequencing, although too laborious and expensive for WGS, remains routinely used when sequencing of specific genes or fragment of genes is needed, for example, for viral or bacterial

genotyping or for resistance testing when specific SNPs are associated with specific phenotypes. The length of such targets is usually restricted to a length of 100 basepairs maximum. For bacterial WGS, biological amplification by culture and single colony picking is needed whereas PCR amplification of specific genes can be done for both viruses and bacteria before culture and more complete sequencing. Since 1987 and during the last 4 decades, Sanger sequencing has been mostly performed using Applied Biosystems (ABI) sequencers (now a company of ThermoFisher Scientific, Waltham, MA, USA), a brand which proposes a series of capillary electrophoresis sequencers with different numbers of capillaries to cover the needs of different laboratories in terms of throughput: the SeqStudio (4 capillaries), the SeqStudio Flex Series (4 or 24 capillaries), the Applied Biosystems 3500 Series (8 or 24 capillaries), and the Applied Biosystems 3730 Series (46 or 96 capillaries). All of those use fluorescent dyes with two different chemistries: dye-primer chemistry or dye-terminator chemistry. Accurate sequences of 600−1000 bases are generated on average after optimalization of assays (https://www.thermofisher.com/be/en/home/brands/applied-biosystems.html). Promega (Madison, WI, USA) also commercializes a 4-capillary benchtop sequencer, the Spectrum Compact CE System with associated reagents (https://be.promega.com/products/sequencing/sanger-sequencing/spectrum-cce-instrument/?catNum=CE1304).

Next generation sequencing: From 2005 onward, instruments and chemistries allowing sequencing with higher throughput than Sanger sequencing were successfully commercialized. They are referred to as Next Generation Sequencing and are based on three different principles: short reads Sequencing-By-Synthesis, Single-Molecule Real-Time sequencing (SMRT) and nanopore-based sequencing. A detailed description of the different NGS principles is available from multiple sources (see for example Slatko et al., 2018). A limited number of companies has been involved in commercialization of these technologies, facilitating and rendering genomics accessible from in the "fields" (point of care or point of need testing) to laboratories in large hospitals, research institutes, and private or public service companies (Table 19.1). The main experimental characteristics of sequencing

Table 19.1 Major NGS Manufacturers

Technology	Company	Products
Sanger capillary electrophoresis sequencing	ThermoFisher Scientific	Applied Biosystems 3500 Series, Applied Biosystems 3730 Series, SeqStudio, SeqStudio Flex Series
	Promega	Spectrum Compact CE System
NGS: short fragments sequencing by synthesis	ThermoFisher Scientific	Ion Torrent NGS: GeneStudio S5, Genexus Integrated Sequencer, Ion PGM Dx
	Illumina	iSeq 100, MiniSeq, MiSeq, NextSeq 550, 1000&2000, NovaSeq 6000, Ion Proton, NextSeq (NGS and microarray)
	MGI Tech	DNBSEQ T7, G400, G50, MGIFLP-L200
	Qiagen	GeneReader
NGS: large fragment single molecule sequencing	PacBio	Sequel system, II and IIe

Continued

Table 19.1 Major NGS Manufacturers—cont'd

Technology	Company	Products
NGS: nanopore sequencing	Oxford Nanopore	MinION, GridION, PromethION, nVIDIA
Microarrays	ThermoFisher Scientific	Applied Biosystems Axiom, GeneChip System 3000Dx v.2
	Illumina	iScan, NextSeq (NGS and microarray)
	Agilent	SureScan Microarray Scanner
Integrated pre-NGS platforms	Roche	AVENIO Edge System
	Qiagen	GeneRead QIAcube
	ThermoFisher	Ion Chef
Reagents for NGS	Qiagen	QIAseq Multimodal kits (DNA/RNA libraries, Illumina compatible)
	Agilent	SureSelect NGS panels (integrated sequencing workflow)
	IDT (Danaher)	xGen NGS (integrated sequencing workflow)
	Roche	KAPA EvoPlus kits (DNA/RNA libraries)

platforms are detailed in Table 19.2. Numerous publications or case studies describe the way these technologies are used in microbiology for research, to characterize various microbiomes or for infectious disease surveillance. In addition, several companies provide kits for the preparation of libraries for NGS application to infectious diseases such as COVID-19.

Some of the main companies (Illumina, Oxford Nanopore Technologies, PacBio) partner with certified genomics service companies and propose a list of these companies that is accessible on their web site. Of note, NGS is a very active domain, and companies emerge, prosper, and are acquired. New technologies are proposed and integrated into existing or new NGS workflows. This implies that the data provided below on a diversity of companies and technologies do evolve at a relatively high speed. This also implies that there may be inaccuracies due to (very) recent developments. For those the authors apologize in advance.

Short-Reads Sequencing-by-Synthesis: Next-Generation Sequencing based on SBS differs from Sanger sequencing in that it generates shorter reads. Although NGS can be used for de novo sequencing and produce "reference genomes" much alike Sanger sequencing, it is mostly used for resequencing and comparison with preexisting reference genomes. Massively parallel sequencing of millions to billions of short sequences of 50−300 (up to 600) nucleotides, which are then precisely aligned to reconstitute genome information, compensates for the higher error rates of NGS. To achieve massively parallel sequencing, these technologies rely on the distribution of millions of individual DNA molecules into wells, chambers or at specific locations on solid substrates. The difference between Sanger sequencing and NGS is shown in Fig. 19.2.

Pyrosequencing AB (Sweden) has been the first company to develop an NGS platform in 2005 (454 pyrosequencing) and it was ultimately acquired by Roche Sequencing (Pleasanton, CA, USA) which discontinued Pyrosequencing in 2013. Roche is now commercializing a range of products for library preparation upstream of Illumina sequencers (see below).

Table 19.2 Characteristics of the Main NGS Platforms

	ThermoFisher			Illumina						Qiagen	MGI Tech				PacBio		Oxford Nanopore		
	GeneStudio S5	**Genexus**	**Ion PGM Dx**	**iSeq 100**	**MiniSeq**	**MiSeq**	**NextSeq 550**	**NextSeq 1000& 2000**	**NovaSeq 6000**	**GeneReader**	**DNBSEQ T7**	**DNBSEQ G400**	**DNBSEQ G400 Fast**	**DNBSEQ G50**	**Sequel**	**Sequel II and Iie**	**MinION**	**GridION**	**PromethION 2/24/48**
Library preparation automation	Proprietary: Ion Chef	Proprietary: Genexus Purification	Proprietary: Ion Chef	Bravo (Agilent); Biomek (Beckman Coulter); epMotion (Eppendorf); NG STAR (Hamilton); SciClone (PerkinElmer); Fleunt or Freesom EVO (Tecan)						Yes	Proprietary: MGISP-100 or -960	Proprietary: MGISP-100 or -960	Proprietary: MGISP-100 or -960	Proprietary: MGISP-100 or -960	no	no	N/A	N/A	N/A
Minimum starting material (ng)	1	1	10	1	1	1	1	1	1		1	1	1	1	5	5	0.01−10,000	0.01−10000	0.01−10000
Maximum number of reads/lane (million)	2−130 (depending on the chip)	12−15 per lane, 1−4 lanes/run	4.5−5	4	25	25	400	1200	20,000		5000	550	1500−1800	100−500	0.5	8	0,1/hour	0,1/hour	0,1/hour
Maximum read length (bp)	Up to 200 or 400 (depending on applications)	Up to 200 or 400 (depending on applications)	200	2 × 150	2 × 150	2 × 300	2 × 150	2 × 150	2 × 250		150	200−400	100	150	100k	20	10k (up to >4M)	10k (up to >4M)	10k (up to >4M)
Maximum output (Gb/run)	0.3−50	ND	0.6−1	1.2	7.5	15	120	360	6000		6000	1440	330	150	50	50	30−50	30−250	170−14,000
Analysis time including analysis (hours)	3 (prime system) to 12	12	4−7	9.5−19	4−24	4−55	12−30	11−48	13−44		24−30	13−109	15.5−50.5	9−40	10−20	20−30	1 mn−72 hours	1 mn−72 hours	1 mn−72 hours
Maximum samples per run[a]	1 to 60−64 (depending on the chip and application)	1−32	1	1−48	5−384	1−96	1−96	2−108	4−500	42	1	2 or 4	2	1	48	384	1	5	2-24-48

[a] *depends on needs for number of reads and depth of coverage.*

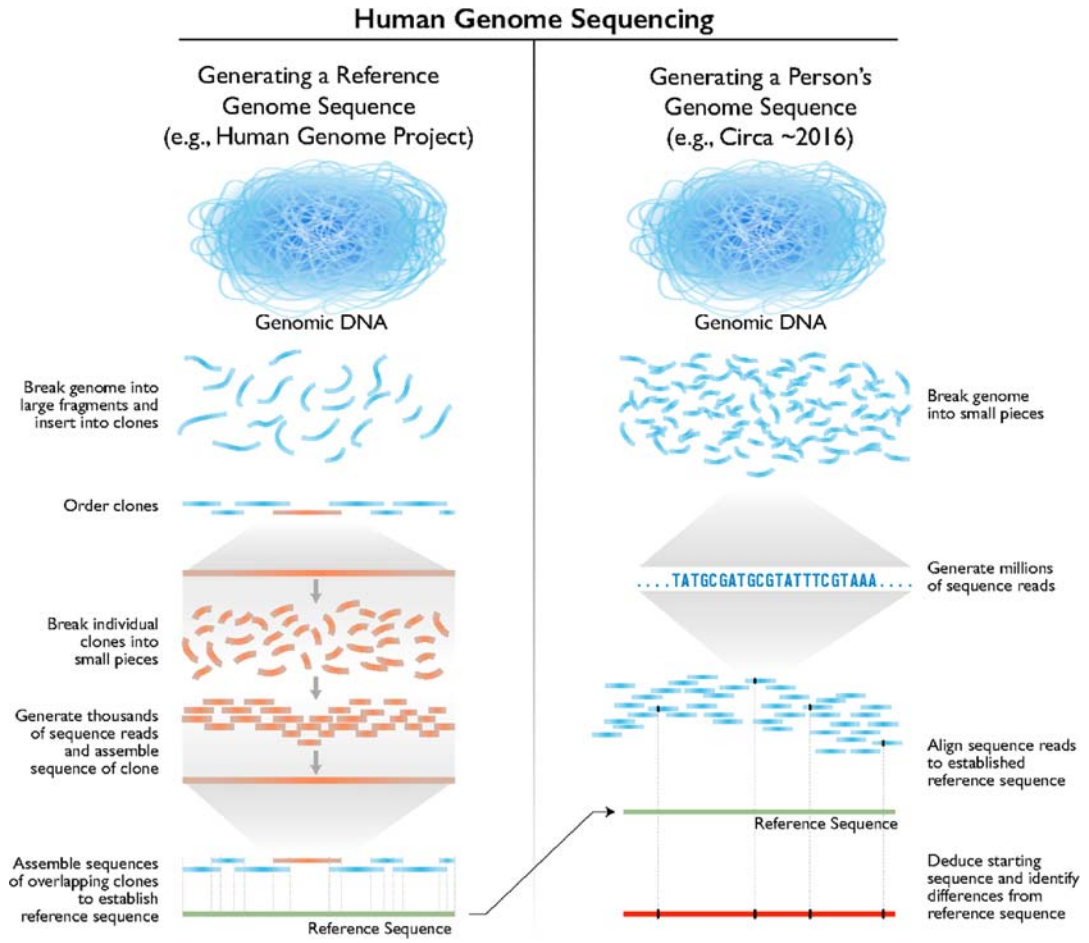


FIGURE 19.2

Comparison of Sanger sequencing and NGS.

Illumina (San Diego, CA, USA) was created in 1998 to exploit rights on the BeadArray technology developed at Tufts University. It's first NGS instrument, the Genetic Analyzer II platform, has been released in 2007. The technology is based on "bridge amplification": DNA molecules of about 500 bp are used as in situ substrates for repeated amplification directly on glass slides. Millions of clusters of about 1000 copies of each short DNA fragment are detected, thanks to the incorporation of fluorescent labels at each synthesis step (Slatko et al., 2018). Currently, Illumina, which is still the nondisputed leader in genomics, proposes a complete range of systems including small benchtop formats: iSeq 100 and MiniSeq and MiSeq as well as instruments providing higher throughputs: NextSeq 550, 1000, and 2000 and NovaSeq 6000. The benchtop systems are especially designed for targeted and small genome sequencing whereas the NextSeq and NovaSeq instruments are well suited for WGS of more complex

organisms. Two platforms, MiSeqDX and NextSeq 550Dx are approved for in vitro Diagnostic (IVD) applications in clinical laboratories (so far, with no approved applications for infectious diseases) in addition to Research Use Only (RUO) applications. Library preparation can be automated using instruments from several partner companies (see Table 19.2). Illumina actively commercializes two library preparation kits to detect mutations in SARS-CoV-2 and identify the emergence of new strains of the virus. It also commercializes a kit for the detection and characterization of common respiratory viruses and a panel which targets more than 280 respiratory pathogens and more than 1200 antimicrobial resistance mutations.

Ion Torrent, created in 2007, has commercialized its first NGS platform in 2010. The company has been rapidly absorbed by Life Technologies, now part of ThermoFisher Scientific (Waltham, MA, USA). Also based on SBS, the Ion Torrent technology uses semiconductor detection of pH changes in sequencing chambers when a hydrogen ion is released if the newly incorporated base matches its complementary base. DNA fragments, amplified by PCR are attached to a bead which enters one unique well among millions of others on the chip (http://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course/what-next-generation-dna-sequencing/ion-torrent).
ThermoFisher is currently commercializing several sequencing platforms as well as instruments to automate library preparation. The Ion Personal Genome Machine (ion PGM) is a benchtop instrument which is well suited for de novo microbial sequencing, bacterial typing research, multilocus sequence typing (MLST), and viral typing. It has been approved for IVD applications in addition to assay development for clinical research applications. The IVD applications so far are targeted at sequencing human genomic DNA derived from peripheral whole-blood and DNA and RNA extracted from formalin-fixed, paraffin-embedded (FFPE) samples.

The Ion GeneStudio S5 system has higher throughputs and outputs. The different chips cover different needs: targeted and RNA sequencing, small RNA sequencing, de novo microbial sequencing, bacterial typing, viral typing, metagenomics, sequencing by genotyping, exome sequencing, transcriptome sequencing, and copy number analysis. It allows sequencing of whole bacterial genomes from isolates or the direct sequencing of specific viral, bacterial, or fungal genes (e.g., 16S genes or antibiotic resistance genes) from biological samples without the need for culturing. All of this can be done in as little as 24 hours.

The Genexus system is fully integrated system allowing anything from nucleic acid extraction to the generation of reports through library preparation and sequencing. Its CE-IVD approved version, Genexus Dx, is intended for clinical laboratories and has a very short hands-on time.

The Ion Chef system and the Genexus Purification system (intended for use with the Genexus sequencer) are commercialized to automate library preparation and template preparation.

Ion AmpliSeq panels contain pools of oligonucleotide primer pairs designed to amplify a specified genomic region and interrogate all bases of specific mutations spots. The Ion AmpliSeq SARS-CoV-2 Research Panel covers more than 99% of the SARS-CoV-2 genome for complete genome sequencing and variant detection for epidemiological studies. The Ion AmpliSeq Microbiome Health Research Kit can be used for characterizing the human gut microbiome. It provides an extensive species-level resolution for the comprehensive profiling of microbial diversity in the human gut microbiome.

In 2013, the Beijing Genome Institute (BGI) acquired the US company Complete Genomics and the DNA nanoball sequencing technology. MGI Tech (Shenzhen, China), a subsidiary of BGI is now commercializing instruments based on DNBSEQ using DNA balls. Briefly, DNA fragments are circularized using adapters and amplified by rolling circle replication resulting in nanoparticles (the

DNA ball). DNA balls are attached to a flow cell in a highly ordered way. Each ball is sequenced by SBS, and fluorescence is used to detect the newly incorporated base. A variant of this chemistry involves labeled antibodies instead of labeled nucleotides in the SBS process. For a more detailed description of MGI sequencing technology see https://en.mgi-tech.com/products/. Several instruments are proposed: DNBSEQ-G50 (150GB/run), DNBSEQ-G400 (720GB/run), and DNBSEQ-T7 (6TB/run). Multiplexing samples are possible during library preparation using up to 96 barcode adapters. DNBSEQ-G50 platforms currently contribute to the surveillance of SARS-CoV-2 variants in China.

Qiagen (Hilden, Germany) commercializes the Sample-to-Insight GeneReader NGS System which includes an NGS platform based on highly parallel fluorescence-based sequencing chemistry (terminator-dNTP sequencing-by-synthesis) using templates that have been clonally amplified using the GeneRead QIAcube. For more details on the GeneReader and associated technologies see https://www.qiagen.com/us/products/instruments-and-automation/genereader-system/qiagen-genereader-platform/.

Single molecule sequencing: real time long read sequencing-by-synthesis: In contrast to NGS based on massive parallel sequencing of short fragments, the Single Molecule Real-Time (SMRT) sequencing technology developed by PacBio (Menlo Park, CA, USA) generates long DNA or RNA sequences. Fragments up to 100 kbp can be generated by binding a DNA polymerase and the DNA to be sequenced in the presence of fluorescently labeled nucleotides in a well in which fluorescence can be detected only at the bottom of the well. Each newly incorporated base is successively detected, and the label is then released outside of the detection area and the next nucleotide can be incorporated. During template preparation, a circular double-stranded element is generated from the fragment to be sequenced including an adapter which can initiate DNA synthesis in the well. Synthesis occurs many times through this circular template and generates DNA from both strands. SMRT can also detect the presence of modified bases which affects the rate of nucleotide addition and provides important epigenetic information (see http://www.pacb.com/smrt-science/smrt-sequencing/). The first commercially available system, the PacBio RSII system, was released in 2011. It has been used by Public Health England in partnership with the Wellcome Trust Sanger Institute to sequence its complete collection of bacterial strains (http://www.sanger.ac.uk/resources/downloads/bacteria/nctc/). PacBio now commercializes the so-called Sequel platforms: Sequel, Sequel II, and Sequel IIe with different read lengths, throughput, and samples per run (Table 19.2).

This technology has been used for several applications in microbiology, including infectious diseases. De novo sequencing of the *Plasmodium falciparum* whole genome has been successfully performed despite it AT-bias which makes more difficult sequencing based on assembly of short fragments (Vembar et al., 2016). The emergence of variants of SARS-CoV-2 has been identified during acute COVID-19 using SMRT (Ko et al., 2021). Full-length 16S sequencing can be used to identify bacteria at the species and strain level (Callahan et al., 2019). The HiFiViral SARS-CoV-2 kit allows enrichment of SARS-CoV-2 viral RNA for subsequent NGS for subsequent and direct NGS.

Nanopore-based sequencing: This technology, which also generates single-molecule long-reads, does not rely on SBS but on the use of nanopores through which a single DNA molecule passes. Each base going through the nanopore deviates an electrical current to an extent which is specific and measurable so that the technique can determine the sequence. Oxford Nanopore Technologies (Oxford, UK) uses this technology for real-time sequencing. Oxford Nanopore devices are made of a nanopore set in an arrayed sensor chip. The nanopore is made of proteins and is embedded in a polymer membrane. Each nanopore is connected with an integrated circuit which measures the experiments.

The same instrument can be used for protein sequencing and that is why ONT is promoting its instruments as sensors rather than nucleic acid sequencers. Oxford Nanopore commercializes several sequencing instruments. MinION is a portable, low-cost device, which has been widely used in decentralized laboratories during the recent Ebola and SARS-CoV-2 epidemics. GridION and PromethION are modular and are suited for high-throughput sequencing for different applications, especially for microbiology. Each platform can generate reads of up to 4 Mbp although DNA molecules of about 10 kbp are generally analyzed through the nanopores. Long nanopore sequencing reads facilitate genome assembly of microorganisms. Data are provided in real-time allowing rapid species identification and detection of antimicrobial resistance. The real-time aspect allows for selective sequencing: sequences that are of noninterest can be suppressed by inverting the current for an individual pore. This enlarges the efficiency of the nanopore sequencing. Targeted, long-reads sequencing of 16S rRNA is a powerful tool for species and strain identification. Publications on Oxford Nanopore instruments are available at https://nanoporetech.com/resource-centre?tags[value][0]=publications&investigations-[value][0]=Identification. Recently, it has been used to sequence monkeypox virus in Central African Republic where the virus is endemic although it has recently spread to many other countries outside Africa (Vandenbogaert et al., 2022). Oxford Nanopore commercializes three kits for sequencing 576, 2304 or 9216 SARS-CoV-2 samples and monitor the emergence of virus variants. The SARS-CoV-2 genome is amplified in 29 overlapping amplicons (~1200 bp long). Samples can be easily multiplexed. The 16S Barcoding Kit amplifies the entire ~1500 bp 16S rRNA gene to allow genus-level bacterial identification.

4. Library Preparation

Several companies commercialize instruments and/or reagents which can be used upstream of genome sequencing. Companies such as Qiagen or Thermofisher Scientific propose solutions for nucleic acid extraction and purification. Again, this is a highly dynamic field of research and development, and the authors apologize for any inaccuracies in the section below.

Qiagen has developed a range of panels for the preparation of libraries upstream of Illumina sequencers. The QIAseq multimodal kits contain reagents for total nucleic acid extraction from FFPE tissue sections, cells and fresh biopsies, and other sample types. Nucleic acid fragmentation, end repair, differential tagging for genomic DNA and cDNA molecules and target enrichment are performed. Separate DNA and RNA libraries are generated in the final amplification step using specific PCR primers, which are then ready for quantification and sequencing. Solutions specific for the detection of SARS-CoV-2 based on sequences from the ARTIC databases are also commercialized by the company. The QIAseq xHYB Respiratory Panel targets 89 viruses including SARS-CoV-2, influenza viruses, rhinovirus, parainfluenza viruses or enteroviruses. The QIAseq xHYB Viral STI Panel targets HBV, HIV-1, and 19 types of high-risk HPV, including types 16 and 18. The QIAseq xHYB AMR Panel offers detection of more than 2700 separate AMR gene targets. The QIAseq xHYB Adventitious Agent Panel allows NGS detection and characterization of human adenovirus, norovirus, rotavirus, influenza viruses, SARS-CoV-2, HPV, Epstein−Barr virus, HIV, Zika virus, hepatitis viruses, and more.

The SureSelect range of Agilent (Santa Clara, CA, USA) is intended to simplify and automate the workflow for different NGS technologies and applications. employs probes to select sequences of

interest in an NGS library. Agilent SureSelect sequencing panels, which are hybridization capture-based target enrichment reagents, are intended for library preparation and target enrichment. Agilent NGS workflow solutions include automation platforms, sample QC, and analysis software. SureSelect Ion Proton is optimized for the platform distributed by Thermofisher. SureSelectQXT Kits offer uniform coverage of the genome, especially in challenging AT-rich regions compatible with Illumina platforms.

Roche Diagnostics, which has commercialized the first 454 NGS platform in 2011, now focuses on reagents and automates for all steps upstream of sequencing, from sample collection to library quantification. These reagents are intended to work well in combination with Illumina platforms.

Coralville (Iowa, USA; a Danaher company) also proposes a range of reagents for library preparation, xGen NGS, including reagents for SARS-CoV-2. Variants detection and 16 seconds rRNA sequencing for bacterial typing is offered as well.

Companies with a strong experience in automated pipetting have developed solutions to automate library preparation. Especially, Agilent (Bravo), Beckman Coulter (Biomek), Eppendorf (epMotion), Hamilton (NG STAR), PerkinElmer (SciClone); Tecan (Flnt or Freesom EVO) have partnered with Illumina for that purpose.

5. Genomics Service Companies and Infectious Diseases

The number of platforms or companies that provide genomics services has exploded in the recent years. Many of them address the needs of scientists who handle projects on microbes or industries for improving their quality control. It is beyond the scope of this chapter to exhaustively list these companies. Most major universities and life sciences research institutes in industrialized and emerging countries have developed genomics services which are opened to research teams in public institutes, pharmaceutical or agro-food industry. These services are leveraging on technological platforms and competencies in their host department at the university. A map of genomics companies in France can be found on the website of France Genomique (https://www.france-genomique.org/plateformes-et-equipements/). Spin-off companies are sometimes created to cover specific R&D needs, marketing, sales and administrative aspects such as ViroScan 3D, associated to ProfileXpert, the genomic platform of Université Claude Bernard in Lyon France (https://www.viroscan3d.com/).

A nonexhaustive list of the major companies providing genomics services for infectious diseases is presented in Table 19.3. Services may be the only activity of some of them but companies like Qiagen or Illumina propose services in addition to equipment and reagents for NGS. These two companies also certify genomics services companies or platform using their technologies, listed on their websites.

6. Conclusion

Genomics has become indispensable in all research domains covering infectious agents. Hundreds of thousands of microbial genome sequences have been defined and this has hugely impacted on our fundamental knowledge of microbial genetics and biology in general. This will continue to evolve over the years to come and scientists will profit from increasingly simple and efficient NGS technologies to help progress their research. The next revision of this chapter will be documenting this rapid evolution in even more detail.

Table 19.3 Genomics Service Compagnies Offering Solutions for Infectious Diseases

Name	Genomics Services
THERAGENE BIO	Genome sequencing Transcriptome sequencing Epigenome sequencing Metagenome sequencing Clinical sequencing
Novogene	
CD Genomics MicrobioSeq	Microbial sequencing services
MiGS	Microbial genome sequencing
Microbes NG	Microbial whole-genome sequencing
Creative Biogene	Microbial sequencing services
Illumina	Partnering with certified sequencing service providers https://www.illumina.com/services/service-partners-and-programs/propel-certified-labs.html Clinical sequencing: CLIA-certified, CAP-accredited laboratory
Qiagen	Partnering with certified sequencing service providers https://www.qiagen.com/us/applications/next-generation-sequencing/qiagen-reference-core-program

References

Callahan, B.J., Wong, J., Heiner, C., Oh, S., Theriot, C.M., Gulati, A.S., McGill, S.K., Dougherty, M.K., 2019. High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Res. 47 (18), e103. https://doi.org/10.1093/nar/gkz569.

Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G., Ysebaert, M., 1976. Complete nucleotide-sequence of bacteriophage MS2-RNA - primary and secondary structure of replicase gene. Nature 260, 500−507.

Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al., 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. Science 269 (5223), 496−512.

Govender, K.N., Street, T.L., Sanderson, N.D., Eyre, D.W., 2021. Metagenomic sequencing as a pathogen-agnostic clinical diagnostic tool for infectious diseases: a systematic review and meta-analysis of diagnostic test accuracy studies. J. Clin. Microbiol. 59 (9), e0291620. https://doi.org/10.1128/JCM.02916-20.

Herring, C.D., Palsson, B.Ø., 2007. An evaluation of Comparative Genome Sequencing (CGS) by comparing two previously-sequenced bacterial genomes. BMC Genom. 14, 274.

Holt, R.A., Jones, S.J., 2008. The new paradigm of flow cell sequencing. Genome Res. 18, 839846.

Jaillard, M., Lima, L., Tournoud, M., Mahé, P., van Belkum, A., Lacroix, V., Jacob, L., 2018. A fast and agnostic method for bacterial genome-wide association studies: bridging the gap between k-mers and genetic events. PLoS Genet. 14 (11), e1007758.

Kardan-Yamchi, J., Amini, S., Hamzelou, G., Rahimi Foroushani, A., Ghodousi, A., Cirillo, D.M., Feizabadi, M.M., 2021. Assessment of the genotype mtbdrsl ver 2.0 compared to the phenotypic drug susceptibility testing and whole genome sequencing for the rapid detection of resistance to fluoroquinolone and

second-line injectable drugs among rifampicin-resistant *Mycobacterium tuberculosis* isolates. Arch. Microbiol. 203 (7), 3989−3996.

Ko, S.H., Bayat Mokhtari, E., Mudvari, P., Stein, S., Stringham, C.D., Wagner, D., Ramelli, S., Ramos-Benitez, M.J., Strich, J.R., Davey Jr., R.T., Zhou, T., Misasi, J., Kwong, P.D., Chertow, D.S., Sullivan, N.J., Boritz, E.A., 2021. High-throughput, single-copy sequencing reveals SARS-CoV-2 spike variants coincident with mounting humoral immunity during acute COVID-19. PLoS Pathog. 17 (4), e1009431. https://doi.org/10.1371/journal.ppat.1009431.

Lo, S.W., Gladstone, R.A., van Tonder, A.J., Lees, J.A., du Plessis, M., Benisty, R., Givon-Lavi, N., Hawkins, P.A., Cornick, J.E., Kwambana-Adams, B., Law, P.Y., Ho, P.L., Antonio, M., Everett, D.B., Dagan, R., von Gottberg, A., Klugman, K.P., McGee, L., Breiman, R.F., Bentley, S.D., Global Pneumococcal Sequencing Consortium, July 2019. Pneumococcal lineages associated with serotype replacement and antibiotic resistance in childhood invasive pneumococcal disease in the post-PCV13 era: an international whole-genome sequencing study. Lancet Infect. Dis. 19 (7), 759−769.

Luheshi, L., Raza, S., Moorthie, S., Hall, A., Blackburn, L., Rands, C., Sagoo, G., Chowdhury, S., Kroese, M., Burton, H., 2015. Pathogen Genomics in Practice. PGH Foundation (2015). ISBN 978-1-907198-18-2.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376380.

Masignani, V., Pizza, M., Moxon, E.R., 2019. The development of a vaccine against meningococcus B using reverse vaccinology. Front. Immunol. 10, 751.

Maxam, A.M., Gilbert, W., February 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U. S. A. 74 (2), 560−564.

Nyholm, L., Koziol, A., Marcos, S., Botnen, A.B., Aizpurua, O., Gopalakrishnan, S., Limborg, M.T., Gilbert, M.T.P., Alberdi, A., 2020. Holo-omics: integrated host-microbiota multi-omics for basic and applied biological research. iScience 23 (8), 101414.

Rakotonirina, A., Galperine, T., Allémann, E., 2022. Fecal microbiota transplantation: a review on current formulations in *Clostridioides difficile* infection and future outlooks. Expet Opin. Biol. Ther. 22 (7), 929−944.

Rogers, L.A., Strong, K., Cork, S.C., McAllister, T.A., Liljebjelke, K., Zaheer, R., Checkley, S.L., 2021. The role of whole genome sequencing in the surveillance of antimicrobial resistant Enterococcus spp.: a scoping review. Front. Public Health 9, 599285. https://doi.org/10.3389/fpubh.2021.599285.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U. S. A. 74 (12), 5463−5467.

Saravanan, K.A., Panigrahi, M., Kumar, H., Rajawat, D., Nayak, S.S., Bhushan, B., Dutt, T., 2022. Role of genomics in combating COVID-19 pandemic. Gene 823, 146387.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., et al., 2005. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309, 1728−1732.

Slatko, B.E., Gardner, A.F., Ausubel, F.M., 2018. Overview of next-generation sequencing technologies. Curr. Protoc. Mol. Biol. 122 (1), e59. https://doi.org/10.1002/cpmb.59.

Stoesser, N., Batty, E.M., Eyre, D.W., Morgan, M., Wyllie, D.H., Del Ojo Elias, C., Johnson, J.R., Walker, A.S., Peto, T.E., Crook, D.W., 2013. Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates using whole genomic sequence data. J. Antimicrob Chemother. Oct 68 (10), 2234−2244.

Vandenbogaert, M., Kwasiborski, A., Gonofio, E., Descorps-Declère, S., Selekon, B., Nkili Meyong, A.A., Ouilibona, R.S., Gessain, A., Manuguerra, J.C., Caro, V., Nakoune, E., Berthet, N., Gonofio, E., et al., 2022. Nanopore sequencing of a monkeypox virus strain isolated from a pustular lesion in the Central African Republic. Sci. Rep. 12, 10768. https://doi.org/10.1038/s41598-022-15073-1.

Vembar, S.S., Seetin, M., Lambert, C., Nattestad, M., Schatz, M.C., Baybayan, P., Scherf, A., Smith, M.L., 2016. Complete telomere-to-telomere de novo assembly of the *Plasmodium falciparum* genome through long-read (>11 kb), single molecule, real-time sequencing. DNA Res. 23 (4), 339−351.

Walker, T.M., Miotto, P., Köser, C.U., Fowler, P.W., Knaggs, J., Iqbal, Z., Hunt, M., Chindelevitch, L., Farhat, M., Cirillo, D.M., Comas, I., Posey, J., Omar, S.V., Peto, T.E., Suresh, A., Uplekar, S., Laurent, S., Colman, R.E., Nathanson, C.M., Zignol, M., Walker, A.S., , CRyPTIC Consortium; Seq and Treat Consortium, Crook, D.W., Ismail, N., Rodwell, T.C., 2022. The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. Lancet Microbe e265−e273. https://doi.org/10.1016/S2666-5247(21)00301-3.

Wetterstrand, K.A., 2021. The Cost of Sequencing a Human Genome. https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost.

Wilson, M.R., Sample, H.A., Zorn, K.C., Arevalo, S., Yu, G., Neuhaus, J., Federman, S., Stryke, D., Briggs, B., Langelier, C., Berger, A., Douglas, V., Josephson, S.A., Chow, F.C., Fulton, B.D., DeRisi, J.L., Gelfand, J.M., Naccache, S.N., Bender, J., Dien Bard, J., Murkey, J., Carlson, M., Vespa, P.M., Vijayan, T., Allyn, P.R., Campeau, S., Humphries, R.M., Klausner, J.D., Ganzon, C.D., Memar, F., Ocampo, N.A., Zimmermann, L.L., Cohen, S.H., Polage, C.R., DeBiasi, R.L., Haller, B., Dallas, R., Maron, G., Hayden, R., Messacar, K., Dominguez, S.R., Miller, S., Chiu, C.Y., June 13, 2019. Clinical metagenomic sequencing for diagnosis of meningitis and encephalitis. N. Engl. J. Med. 380 (24), 2327−2340.

This page intentionally left blank

CHAPTER

Emerging Zoonoses, Biodiversity, and Ecosystems

20

Serge Morand[1,2,3]

[1]*IRL HealthDEEP (Health, Disease Ecology, Environment and Policy) CNRS – Kasetsart University – Mahidol University, Bangkok, Thailand;* [2]*Faculty of Veterinary Technology, University of Kasetsart, Bangkok, Thailand;* [3]*Faculty of Tropical Medicine, University of Mahidol, Bangkok, Thailand*

1. Introduction

Zoonoses account for more than 70% of human infectious diseases (Jones et al., 2008). The microbial and parasitic agents at the origin of these diseases have been acquired throughout the evolution of the human species, but particularly during the recent history of humanity, since the dawn of the Neolithic revolution until the present time (Morand, 2015). The environmental impact of the development of agriculture allowed the first major infectious and zoonotic diseases to emerge. Malaria, yellow fever, trypanosomiasis, or filariasis have benefited from deforestation and the extension of agricultural land and the increase in human densities.

Domestic animals constitute privileged epidemiological bridges favoring the overflow of infectious agents harbored in wildlife toward humans and ensure their amplification (Morand et al., 2014). The spectacular increase in farm animals in recent decades is first of all a major cause of the decline in biodiversity and the alteration of natural habitats, but any increase in the densities of farm animals in countries hosting high biodiversity also promotes zoonotic risks (Morand et al., 2020; Morand, 2021). Recent comparative studies show that domestic animals, and nonhuman primates, are central in the networks for sharing infectious agents between wildlife and humans (Wells et al., 2020). Thus, even if wild animals can constitute intermediate hosts favoring the passage of bat viruses to humans, as in the case of the civet and SARS-Cov1, for the vast majority, these adaptations have occurred via domestic reservoirs: Nipah virus with pigs, Hendra virus with horses, MERS-CoV with camels (Latinne and Morand, 2022).

The increasing number of disease outbreaks (Smith et al., 2014) and disease emergence (Jones et al., 2008) that occurred during the last decades is linked to several drivers. Some drivers have to do with local conditions while others are more regional or global, such as the increasing number of drug-resistant microbes, higher contact rates with wildlife, increased livestock and global food production, loss and fragmentation of natural habitats caused by urbanization and agricultural intensification, extreme climate events, and others (Plowright et al., 2008; Keesing et al., 2010; McMahon et al., 2018; Rohr et al., 2019). Outbreaks of infectious disease epidemics over the past decades appear to be associated with the increase in livestock farming and loss of biodiversity, deforestation and commercial plantation expansion (Morand and Lajaunie, 2021a).

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00027-6
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Humans constantly interact with their environment, which role in human health has long been recognized. Globally, nearly 25% of all deaths and the total burden of disease can be attributed to environmental factors (WHO, 2022). Yet, it is only recently that maintaining a healthy environment is seen as essential to improve quality of life and healthy life expectancy. Studies are increasingly demonstrating the impact of global change (climate change, loss of biodiversity, habitat degradation, urbanization) on human health (Morand and Lajaunie, 2017).

This chapter deals with the evolution of approaches linking health, environment, and ecosystems. Environmental history teaches us how the fixist view of the environment has evolved into a dynamic vision of our interaction with it. The current global changes and their impacts on human health require the implementation of renewed approaches based on ecosystems capable of responding to new health challenges. These approaches and new strategies will be all the more efficient if they are based on shared knowledge bases.

2. A Long History

2.1 Environment as a Health Factor

The Western recognition that health is linked to the environment dates back to the advice given by Hippocrates to the physicians of Ancient Greece in the fifth century BC: "*First he ought to consider what effects each season of the year can produce; for the seasons are not at all alike, but differ widely both in themselves and at their changes … He must also consider the properties of the waters; for as these differ in taste and in weight, so the property of each is far different from that of any other … He must consider with the greatest care both these things and how the natives are off for water, whether they use marshy, soft waters, or such as are hard and come from rocky heights, or brackish and harsh. The soil too, whether bare and dry or wooded and watered, hollow and hot or high and cold. The mode of life also of the inhabitants that is pleasing to them, whether they are heavy drinkers, taking lunch,1 and inactive, or athletic, industrious, eating much and drinking little. …*" (Hippocrates in Jones, 1868).

The Hippocratic idea with the four humors (blood, yellow bile, black bile, and phlegm), the four qualities (hot, cold, dry, and wet) and the four seasons constitutes one of the first doctrines formalizing the links between environment and health. However, as Berridge and Gorsky (2012) pointed out, this doctrine is deeply fatalistic. Hippocrates recognized the connections between environment and health but these cannot be modified, human beings are subject to it.

In the middle of the 19th century, the significance of the links between the environment and health was still very present in Thomas Logan, an American doctor from the Californian "New Frontier" who wrote in a text from 1859: "*A knowledge of the etiology of diseases can best be attained by studying the affections of different localities in connections with every condition and circumstance calculated to operate prejudicially or otherwise upon the health of the inhabitants. Such philosophical investigation is particularly useful in tracing the modifications diseases may undergo from the agency of causes of a local or special character; and being also calculated to elucidate the relationships of diseases to climate, to the prevailing geological formations, the fauna, the vegetables, the minerals, the water, which vary with the earth's crust, …* " (quoted by Nash, 2006). Like Hippocrates, colonial physicians also saw the environment as fundamental to understanding risks to public health. But, it will be necessary to wait for the development of hygiene for policies of intervention on the environment to be put in place, such as water purification.

2.2 The Rise of Hygienism

Hygienism really began in 1893, with the chemist and physician Theodor Weyl of the Technical University of Berlin-Charlottenburg. Weyl inaugurated an important publishing enterprise with the "*Handbuch der Hygiene*" the first great manual of public hygiene. The undertaking would last almost 10 years with the publication of 10 volumes summarizing the scientific and medical knowledge of the time (Schott, 2012). Weyl identified the lack of academic training in the field of health and the need to combine the activities of doctors with those of engineers. Weyl also proposed that health and public hygiene be dealt with in a dedicated ministry with a minister of health who should of course be a doctor (Schott, 2012).

The "Handbuch der Hygiene" is sometimes referred to as a textbook of preenvironmental science. Some environmental historians point out that movements for the protection of nature, like the movements for the protection of the landscape and natural habitats, could have converged with the hygienist movement in an effective policy of environmental protection. However, the weight of the chemical industry and the First World War delayed this advent. Thus, for Schott (2012), hygiene plays in the history of the environment a role similar to that which the concepts of ecology played at the end of the 1960s. The ecological vision of health will be carried by the Unesco (United Nations Educational, Scientific and Cultural Organization) and IUCN (International Union for Conservation of Nature). But this also finds its origin in the colonial sciences.

2.3 Environment and Health in Colonial Sciences

In the 1930s, the African Research Survey brought together a network of academics from the schools of tropical medicine in London and Liverpool, and managers from the Imperial Forestry Institute in Oxford and the Imperial Agricultural Bureaux. Under the direction of Malcolm Hailey and scientific adviser Edgar Barton Worthington, a book "Science in Africa" summarizes the work carried out by the African Research Survey (Worthington, 1938). A diagram included in this book illustrates the scientific project of development in colonial Africa (Fig. 20.1). The change in vision of the links between health and the environment is in total rupture with past doctrines. As Worthington points out "*the picture really presented by Africa is one of movement, all branches of physical, biological and human activity reacting upon each other, to produce what biologists would call an ecological complex*" (cited by Tilley, 2011). The report stressed among others that "I*t should be urged that knowledge of mental and physical development, though of great academic interest, is not sufficient as a guide in formulating policy. It should be combined with a survey of diseases, of nutrition, of agricultural and animal husbandry methods, and even botany, zoology, soils, etc. In fact, the ecological outlook would be all-important in such an inquiry in order to understand the several factors working on native development, and to appreciate which of them are susceptible to control*" (Worthington, 1938).

Human development, that of colonial Africa, is conceived in an ecological and ecosystemic approach. Oxford University ecologist Julien Huxley was one of the mentors of this approach. Treasurer of the Zoological Society London, President of the British Eugenics Society, Julian Huxley co-founded the World Wide Fund for Nature (WWF), initiated the creation of the International Union for Conservation of Nature (IUCN), and was appointed the first Director of Unesco (United Nations Educational, Scientific and Cultural Organization) in 1946.

In 1920, Julian Huxley wrote for the Committee on Native Education in Tropical Africa: "*At the present moment, it is clear that many if not most problems of applied biology can only be satisfactorily*

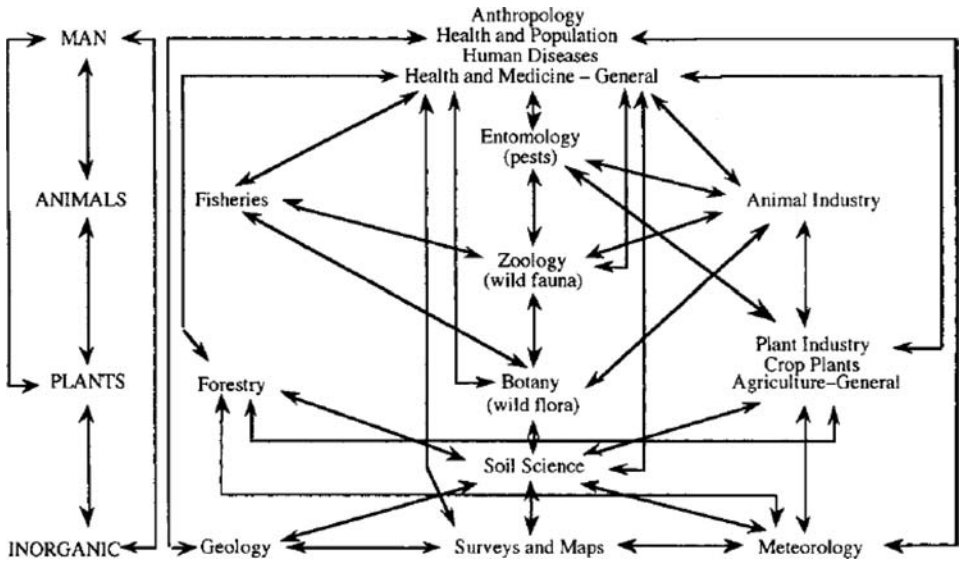


FIGURE 20.1

The colonial environmental and health management scientific network of "science for Africa" (Worthington, 1938). For contemporary health policy makers this type of figure resembles many of today's (see also Rajan, 2006; Tilley, 2011; Lainé and Morand, 2020).

solved by reference to a background of ecological ideas, by whose aid the interrelations of different branches of biological science can be studied …. it is often possible for the ecologist to point out to this or that specialist new lines of approach to his particular problem—disease of man or of domestic animals may prove to be correlated with a cycle of abundance and scarcity in some wild animal … game migrations or … climatic cycles or variations in mineral content of food plants."

Environmental influences on health were also investigated by two French geographers, Maximilien Sorre, who developed the concept of pathogenic complex, and Jacques May, the founder of modern medical geography. Sorre (1943) argued that the emergence of diseases depends on physical, biological, and social factors and more specifically on the climate, the natural environment, and the anthropological and geographical environment (Oppong and Harold, 2010). May started his career as a medical doctor in French Indochina, where he investigated the role of the environment in human diseases and the importance of geography in mapping epidemiology. May (1950) provided a theoretical framework for studying the interplay of environment and geographical factors (in his words, "geogenetics") of infectious disease emergence. He continued his career in the United States working for the United States Agency for International Development (USAID) and the WHO.

2.4 The Unesco Conference on the Biosphere

For the first time, the links between the development of societies, health, and biological conservation were addressed in 1968 during the Unesco Conference on the Biosphere. The scientific bases for the

rational use and conservation of the resources of the biosphere are elaborated there, in particular in a section devoted to human ecology (Unesco, 1970).

In terms of health, the loss of biological diversity is for the first time directly implicated in the deterioration of physical and mental health: "*Whether the challenges come from physical or social forces, the diversity of environments is of crucial importance for the evolution of man and his societies because the ultimate results of a stereotyped and equalized environment can be and often are an impoverishment of life, a progressive loss of the qualities that we identify with humanness and a weakening of physical and mental health. Our policy should be to preserve or to create as many diversified environments as possible.*"

Recommendation three on "*Research on human ecology*" of the final report of the conference after considering that "*man is an integral part of most ecosystems, not only influencing but being influenced by his environment; that his physical and mental health, now and in the future, are intimately linked with the dynamic systems of natural objects, forces and processes that interact with the biosphere and including also of man's culture*," recommended to the Member States, to Unesco, WHO and the international organizations concerned "*that continued research and intensified be undertaken on the ecology of human disease, with special reference to those associated with environmental change and to the zoonotic diseases arising from interaction between man and animals*" and "*that this research be directed at solving increasingly important problems of the establishment of the necessary balance between man and his environment in relation to the maintenance of his health and well-being in their broad connotations.*"

This recommendation calls for the establishment of an ecology of zoonotic diseases integrating the issue of environmental changes. A recommendation that will materialize years later with the Ecosystem Approach and then with the One Health concept.

2.5 Ecosystem Services and Health

The Millennium Development Goals and more particularly the Millennium Ecosystem Assessment in 2005 (MA, 2005) establish a functional link between biodiversity, health, and the well-being of human populations. This functional link is expressed by the ecosystem approach, also promoted by Unesco and by the Convention on Biological Diversity (CBD), and by the notion of ecosystem services.

The ecosystem approach is a strategy for the integrated management of land, water, and living resources, which aims to promote the conservation and sustainable use of these resources. The ecosystem approach is based on the application of scientific methods appropriate to the various levels of biological organization (from populations to communities), which should include the essential processes, functions and interactions between organisms and their environment.

The Millennium Ecosystem Assessment identifies scientific work demonstrating how the degradation of ecosystem functioning and the associated loss of biodiversity have negative impacts on the quality of services provided by ecosystems on the health and well-being of populations. Ecohealth recognizes the interdependence of human health, animal health and ecosystem health. Charron (2012) identifies the fundamental principles of the Ecohealth approach, which are transdisciplinarity, systems thinking, participatory research, sustainability, social and gender equity, and the passage from knowledge to action.

The literature on ecosystem services is considerable, but studies directly analyzing the links between ecosystem services and health are still few (Morand and Lajaunie, 2017). We can also note the

absence of developed and validated indicators for ecosystem services regulating diseases, whether infectious or noncommunicable.

Studies concerning ecosystem services and human well-being are more numerous and show positive effects of the environment, often in interaction with socioeconomic factors. These effects on well-being have been particularly studied in urban ecosystems The work shows that the ecosystem service of well-being depends on the presence of elements of biodiversity, such as the sight of a beautiful landscape or the presence of a garden in a city. Urban dwellers in contact with elements of nature, such as urban parks and gardens, are significantly in better physical and psychological health than those who do not benefit from these contacts.

3. One Health

3.1 The Origin of One Health

In 2004, the World Conservation Society (WCS), one of the major international nongovernmental organizations on biodiversity, promoted the One World One Health initiative and the "*Twelve Manhattan Principles"* After recognizing "*the essential link between human, domestic and wildlife health and the threat disease poses to people, their food supplies and economies, and the biodiversity essential to maintaining the healthy environments and functioning ecosystems we all require*" and "*that decisions regarding land and water use have real implications for health. Alterations in resilience of ecosystems and shifts in patterns of disease emergence and spread manifest themselves when we fail to recognize this relationship,*" the fifth principle recommends the development of proactive and integrative approaches to understand and prevent the development of emerging and reemerging infectious diseases, taking better account of all the complex interconnections that exist between species in ecosystems (Lajaunie and Morand, 2021).

The One Health approach was officially launched in Vietnam in 2008, as an agreement in principle aimed at better tripartite collaboration between the World Organization for Animal Health (WOAH, former OIE), the Food and Agriculture Organization of the United Nations (FAO), and WHO to improve surveillance of zoonotic diseases at the interface of human health, animal health, and environmental health. A global early warning system for major animal and zoonotic diseases was created in 2013 in GLEWS + to address health risks at the human−animal−ecosystem interface. A tool for intersectoral and multidisciplinary collaboration, the goal of GLEWS+ is to link the fields of public, animal, and wildlife health (Morand and Lajaunie, 2021b). The "Berlin One Health Principles" are released in 2019 at the "One Planet, One Health, One Future" conference as an "update" of the Manhattan Principles (Morand and Lajaunie, 2021b). The ecosystem approach to health is promoted as a tool in the service of shared health between humans, animals, and their environment in complement to other approaches (see Box 20.1).

3.2 A New One Health Definition

Since 2008, the "One Health" approach supported by the tripartite agreement WHO-WOAH (former OIE)-FAO attempts to respond to common challenges and issues for animal health and human health. It is clear that an essential pillar supporting the environmental dimension was missing (Garine-Wichatitsky et al., 2020; Destoumieux-Garzón et al., 2022). The science and policy dialog on health, environment and ecosystems can only be strengthened by greater commitment by UNEP to the One Health approach.

BOX 20.1 THE DIFFERENT APPROACHES TO HEALTH (CHARRON, 2012; ZINSSTAG, 2012; MORAND AND LAJAUNIE, 2017; MORAND ET AL., 2020; OHHLEP, 2022; DESTOUMIEUX-GARZÓN ET AL., 2018).

Environmental Health

Environmental health is defined by the World Health Organization as those aspects of human health and disease that are determined by environmental factors. Environmental health also refers to the assessment and controls of environmental factors that can potentially affect health. Environmental health is concerned with the direct pathological effects of chemicals, radiation, and certain biological agents present in housing, the urban, agricultural, or natural environment, and also with their indirect effects on well-being. It takes into account the social and cultural environment, as well as the genetic components.

Ecological Health

Ecological health is used in human health to describe the state of the environment. In human health, it appears very close to environmental health by focusing on exposure to synthetic chemicals, such as pesticides or residues from food packaging, released into the environment. Ecological health is concerned with addressing the causes rather than paying attention to the consequences of the impacts of environmental factors on health (Charron, 2012). Urban planning has appropriated ecological health in the context of green cities, and agriculture in the context of organic production.

Ecosystem Health

The health of ecosystems is a metaphor used to describe the state of an ecosystem which sees its origin in the work of Aldo Leopold (1949) who spoke of the health of the Earth, sick of the violence of practices, and who presented the preservation of biodiversity like a medicine of the Earth. The medical metaphor is claimed as a communication tool that is accompanied by indicators for evaluating desirable environmental objectives: conservation of species, maintenance of evolutionary and ecological processes, or even functioning of ecosystems (Costanza, 1992).

Global Health

Advocated by the WHO, global health reflects the awareness of the interrelationships between globalization and human health, emphasizing the social dimensions of health. Global health aims to reduce health, economic and gender injustices in order to improve individual or population health in the face of global risks, such as those caused by climate change (Nichter, 2008).

Planetary Health

Planetary health has been promoted by the Lancet and the Rockefeller Foundation (Whitmee et al., 2015). It starts from the notion of the Anthropocene and the planetary boundaries, the crossing of which could upset the global planetary functioning. Planetary health offers an integrative approach linking human health to the major cycles of global functioning altered by climate change, biodiversity loss, air pollution or the acidification of the oceans (Myers and Frumkin, 2020).

One Health

One Health is an important initiative that brings together international organizations with a certain capacity to intervene. The new collaboration between the WHO-WOAH-FAO-UNEP or One Health quadripartite offers new opportunities to integrate environment and ecosystem thinking in a more integrated One Health approach (OHHLEP, 2022).

In June 2021, the United Nations Environment Program (UNEP), the world's leading environmental authority, joins forces with the WHO-WOAH-FAO Tripartite to jointly launch a High Level Expert Group on One Health (OHHLEP). This group of experts is responsible for providing relevant scientific assessment in order to improve policies for preventing the emergence and epidemics of zoonotic diseases resulting from changes in the interfaces between humans, wild and domestic animals. OHHLEP (2022) provided a new definition: "*One Health is an integrated, unifying approach that aims to sustainably balance and optimize the health of people, animals, and ecosystems. It recognizes the health of humans, domestic and wild animals, plants, and the wider environment (including ecosystems) are closely linked and interdependent. The approach mobilizes multiple sectors,*

disciplines, and communities at varying levels of society to work together to foster well-being and tackle threats to health and ecosystems, while addressing the collective need for healthy food, water, energy, and air, taking action on climate change and contributing to sustainable development."

OHHLEP emphasize several underlying principles including: "*(1) equity between sectors and disciplines; (2) sociopolitical and multicultural parity and inclusion and engagement of communities and marginalized voices; (3) socioecological equilibrium that seeks a harmonious balance between human− animal−environment interaction and acknowledging the importance of biodiversity, access to sufficient natural space and resources, and the intrinsic value of all living things within the ecosystem; (4) stewardship and the responsibility of humans to change behavior and adopt sustainable solutions that recognize the importance of animal welfare and the integrity of the whole ecosystem, thus securing the well-being of current and future generations; (5) transdisciplinarity and multisectoral collaboration, which includes all relevant disciplines, both modern and traditional forms of knowledge and a broad representative array of perspectives.*" A definition and principles summarize in a diagram (Fig. 20.2).

The four organizations formally joined in 2022 as the One Health quadripartite.

3.3 Preventing Emerging Zoonoses

Emergence is a local problem of transmission of an infectious agent. In the case of zoonoses, emergence occurs at the interface between wild animals and/or domestic animals and humans in rapidly changing environments under the combined effects of climate change, habitat modification, agricultural intensification, and biodiversity losses. However, we are still far from having a complete understanding of the ecological processes underlying the dynamics of emerging infectious diseases. An

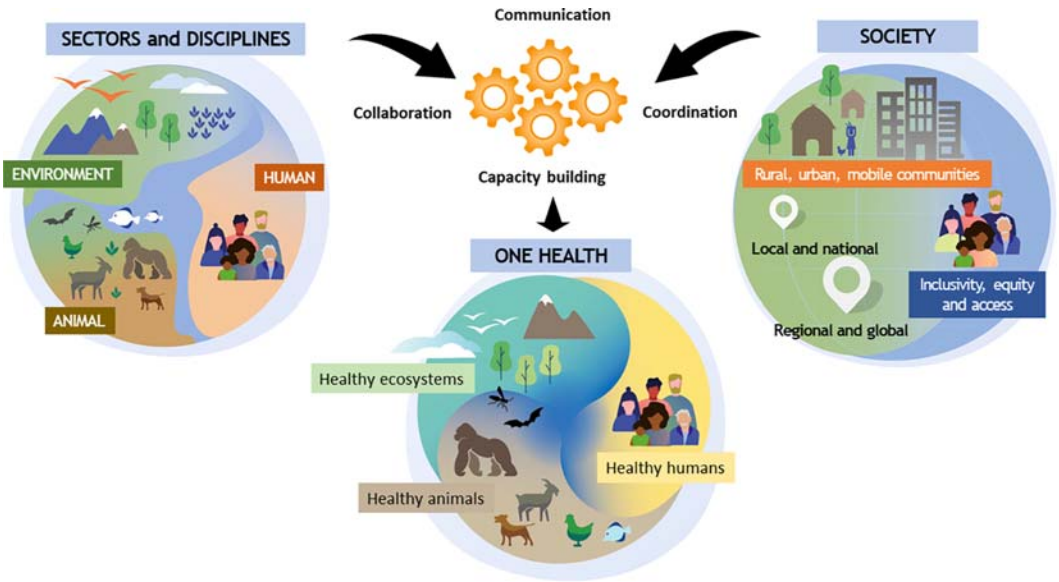


FIGURE 20.2

One Health as developed by the OHHLEP (2022).

increased research effort must examine zoonoses in the dynamics of changes at the interface between environment, biodiversity, livestock, and societies.

A new international initiative for the prevention of the emergence of zoonoses (PREZODE) is officially launched during the "*One Planet Summit for Biodiversity*" organized on January 11, 2021. The PREZODE program (https://prezode.org/) aims to federate international research teams to prevent the risk of emergence of zoonoses at the interface between the environment, wildlife, and livestock. PREZODE can become a research and development model to prevent health risks related to the emergence of infectious agents from animals.

4. Conclusion

This chapter has shown the significance of the environment in the history of health and medicine. It is really with the colonial sciences, and in particular ecology and geography, that the links between biodiversity and health have been theorized and put into practice. Later, the ecosystem approach has underlined the importance of ecological dynamics for the health of humans but also of animals and plants. Finally, the new definition of One Health is a final deepening.

As emphasized by Morand and Lajaunie (2021b), in order to prevent the emergence and spread of zoonotic diseases, it is crucial to develop boundary spanning in relation to human health, animal health, environment health, and the socioeconomic model. Boundary spanning presupposes system thinking, to be more knowledgeable, adaptable, flexible, and to foster interconnections between organizations, institutions, or disciplines in order to tackle complex issues and wicked problems. Boundary spanning should facilitate new integrative research in support of the One Health approach at the interface of disciplines and political sectors (Fig. 20.3). The new One Health and other approaches

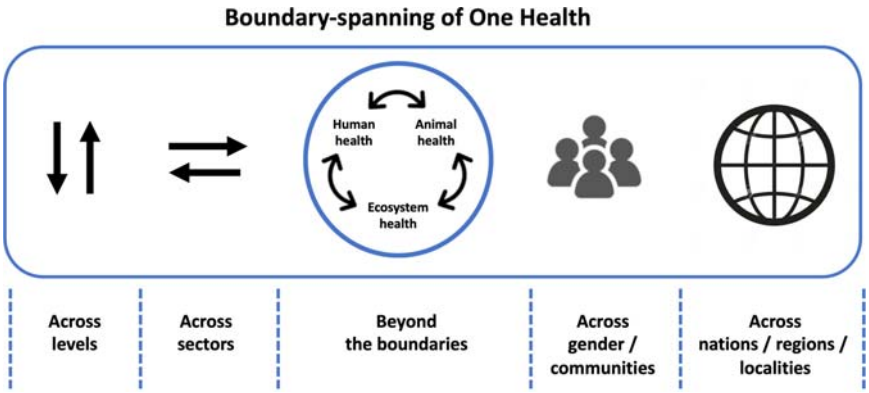


FIGURE 20.3

Boundary spanning helps us to think about a new way for the One Health approach to tackle the risk of emerging zoonotic diseases at the interface between humans, animals, and the environment. The use of boundary spanning in One Health fosters interconnection across levels of organizations, sectors, academic disciplines, communities, gender, and nations in order to tackle complex issues of emerging zoonotic diseases at the boundary of human health, animal health, and ecosystem health (from Morand and Lajaunie 2021b).

by promoting interconnections between organizations, institutions, and disciplines, offer the opportunity to consider a new way of conducting health research at the interface between humans and ecosystems to better inform public health policies. However, as emphasized by Giraudoux et al. (2022), the One Health approach requires a real expansion of collaborations between disciplines and sectors to achieve its goals.

Acknowledgments

SM is supported by the French ANR project FutureHealthSEA and by the Thailand International Cooperation Agency (TICA) project "Innovative Animal Health."

References

Berridge, V., Gorsky, M.n, 2012. Environment, Health and History. Palgrave Macmillan, New York.

Charron, D.F., 2012. Ecohealth Research in Practice. Insight and Innovation in International Development. Springer, New York.

Costanza, R., 1992. Towards an operational definition of health. In: Costanza, R., Norton, B., Haskell, B.D. (Eds.), Ecosystem Health. New Goals for Ecosystem Management. Island Press, Washington DC, pp. 239–256.

Destoumieux-Garzón, D., Mavingui, P., Boetsch, G., Boissier, J., Darriet, F., Duboz, P., Fritsch, C., Giraudoux, P., Le Roux, F., Morand, S., et al., 2018. The One Health concept: 10 Years Oold and a long road ahead. Front. Vet. Sci. 5, 14.

Destoumieux-Garzón, D., Matthies-Wiesler, F., Bierne, N., Binot, A., Boissier, J., Devouge, A., Garric, J., Gruetzmacher, K., Grunau, C., Guégan, J.-F., et al., 2022. Getting out of crises: environmental, social-ecological and evolutionary research needed to avoid future risks of pandemics. Environ. Int. 158, 106915.

Garine-Wichatitsky, M., Binot, A., Morand, S., Kock, R., Roger, F., Wilcox, B.A., Caron, A., 2020. Will Covid-19 crisis trigger one health coming-of-age? Lancet Planet. Health 4, e377–e378.

Giraudoux, P., Besombes, C., Bompangue, D., Guégan, J.-F., Mauny, F., Morand, S., 2022. One Health or "One Health Washing"? an Alternative More than Ever to Overcome. CABI One Health (in press).

Hippocrates, Jones, W.H.S., 1868. Hippocrates Collected Works I. Harvard University Press, Cambridge. http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0251%3Atext%3DAer. (Accessed September 2022).

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P., 2008. Global trends in emerging infectious diseases. Nature 451, 990–993.

Keesing, F., Belden, L.K., Daszak, P., Dobson, A., Harvell, C.D., Holt, R.D., Hudson, P., Jolles, A., Jones, K.E., Mitchell, C.E., et al., 2010. Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature 468, 647–652.

Lainé, N., Morand, S., 2020. Linking humans, their animals, and the environment again: a decolonized and more-than-human approach to 'One Health. Parasite 27, 55.

Lajaunie, C., Morand, S., 2021. Biodiversity targets, SDGs and health: a new turn after the coronavirus pandemic? Sustainability 13, 4353.

Latinne, A., Morand, S., 2022. Climate anomalies and spillover of bat-borne viral diseases in the Asia-Pacific region and Arabian Peninsula. Viruses 14, 1100.

Leopold, A., 1949. A Land County Almanac and Sketches Here and There. Oxford University Press, Oxford.

MA (Millennium Ecosystem Assessment), 2005. Millennium Ecosystem Assessment. Ecosystems and Human Well-Being: Biodiversity Synthesis. Island Press, Washigton DC.

May, J.M., 1950. Medical geography: its methods and objectives. Geogr. Rev. 40, 9–41.
McMahon, B.J., Morand, S., Gray, J.S., 2018. Ecosystem change and zoonoses in the Anthropocene. Zoonoses Public Health 65, 755–765.
Morand, S., 2015. Diversity and origins of human infectious diseases. In: Muehlenbein, M.P. (Ed.), Basics in Human Evolution. Academic Press, London, pp. 405–414.
Morand, S., 2021. Emerging diseases, livestock expansion and biodiversity loss are positively related at global scale. Biol. Conserv. 248, 108707.
Morand, S., Lajaunie, C., 2017. Biodiversity and Health. Linking Life, Ecosystems and Societies. Elsevier, London.
Morand, S., Lajaunie, C., 2021a. Outbreaks of vector-borne and zoonotic diseases are associated with changes in forest cover and oil palm expansion at global scale. Front. Vet. Parasitol. 24, 661063.
Morand, S., Lajaunie, C., 2021b. Biodiversity and Covid-19: a report and a long road ahead to avoid a next pandemic. One Earth 4, 920–923.
Morand, S., McIntyre, K.M., Baylis, M., 2014. Domesticated animals and human infectious diseases of zoonotic origins: domestication time matters. Infect. Genet. Evol. 24, 76–87.
Morand, S., Guégan, J.-F., Laurans, Y., 2020. From One Health to Ecohealth, mapping the incomplete integration of human, animal and environmental health. IDDRI Décryptage 4, 1–4.
Myers, S., Frumkin, H., 2020. Planetary Health: Protecting Nature to Protect Oourselves. Island Press, Washington DC.
Nash, L., 2006. Inescapable Ecologies. A History of Environment, Diseases, and Knowledge. University of California Press, Berkeley.
Nichter, M., 2008. Global Health: Why Cultural Perceptions, Social Representations, and Biopolitics Matter. University of Arizona Press, Tucson.
One Health High-Level Expert Panel (OHHLEP), Adisasmito, W.B., Almuhairi, S., Behravesh, C.B., Bilivogui, P., Bukachi, S.A., et al., 2022. One Health: a new definition for a sustainable and healthy future. PLoS Pathog. 18, e1010537.
Oppong, J.R., Harold, A., 2010. Disease, ecology, and environment. In: Brown, T., Mclafferty, S., Moon, G. (Eds.), A Companion to Health and Medical Geography. Willey, London, pp. 181–195.
Plowright, R.K., Sokolow, S.H., Gorman, M.E., Daszak, P., Foley, J.E., 2008. Causal inference in disease ecology: investigating ecological drivers of disease emergence. Front. Ecol. Environ. 6, 420–429.
Rajan, R.S., 2006. Modernizing Nature: Forestry and Imperial Ecodevelopment 1800–1950. Oxford University Press, Oxford.
Rohr, J.R., Barrett, C.B., Civitello, D.J., Craft, M.E., Delius, B., DeLeo, G.A., Hudson, P.J., Jouanard, N., Nguyen, K.H., Ostfeld, R.S., et al., 2019. Emerging human infectious diseases and the links to global food production. Nat. Sustain. 2, 445–456.
Schott, D., 2012. The handbuch der Hygiene: a manual of proto-environmental science in Germany of 1900? In: Berridge, V., Gorsky, V.E. (Eds.), Health and History. Palgrave Macmillan, New York, pp. 69–93.
Smith, K.F., Goldberg, M., Rosenthal, S., Carlson, L., Chen, J., Chen, C., Ramachandran, S., 2014. Global rise in human infectious disease outbreaks. J. Roy. Soc. Interface 11, 20140950.
Sorre, M., 1943. Les fondements de la Géographie humaine. In: Les Fondements Biologiques: Essai d'une Écologie del'Homme, vol. 1. Armand Colin, Paris.
Tilley, H., 2011. Africa as a Living Laboratory Empire, Development, and the Problem of Scientific Knowledge, 1870–1950. The University of Chicago Press, Chigaco.
UNESCO (United Nations Educational, Scientific and Cultural Organization), 1970. Intergovernmental Conference of Experts on the Scientific Basis for Rational Use and Conservation of the Resources of the Biosphere, Paris, 4-13 September 1968. Unesco, Paris.

Wells, K., Morand, S., Wardeh, M., Baylis, M., 2020. Distinct spread of DNA and RNA viruses among mammals amid prominent role of domestic species. Glob. Ecol. Biogeogr. 29, 470−481.

Whitmee, S., Haines, A., Beyrer, C., Boltz, F., Capon, A.G., Ferreira de Souza Dias, B., Ezeh, A., Frumkin, H., Gong, P., Head, P., et al., 2015. Safeguarding human health in the Anthropocene epoch: report of the Rockefeller Foundation−Lancet Commission on planetary health. Lancet 386, 1973−2028.

WHO (World Health Organization), 2022. Public Health and Environment. https://www.who.int/data/gho/data/themes/public-health-and-environment.

Worthington, E.B., 1938. Science in Africa. A Review of Scientific Research Relating to Tropical and Southern Africa. Oxford University Press, London. https://archive.org/details/scienceinafricar00wort.

Zinsstag, J., 2012. Convergence of Ecohealth and one health. EcoHealth 9, 371−373.

CHAPTER

Epigenetics and Transmissible Diseases

21

Mohamed Kamel[1,2,3], Josiah Levi Davidson[1,3], Rachel A. Munds[1,4], Jiangshan Wang[1,3], Simerdeep Kaur[1,3], Virendra Kumar[1,3] and Mohit S. Verma[1,3,4,5]

[1]*Department of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN, United States;* [2]*Department of Medicine and Infectious Diseases, Faculty of Veterinary Medicine, Cairo University, Giza, Egypt;* [3]*Birck Nanotechnology Center, Purdue University, West Lafayette, IN, United States;* [4]*Krishi Inc., West Lafayette, IN, United States;* [5]*Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN, United States*

1. Introduction and Definition of Epigenetics

1.1 Definition of Epigenetics

Epigenetics is an intricate field of study and potentially revolutionary new science (Bird, 2007). This emerging field explores heritable alterations in gene expression that cannot be attributed to changes in DNA sequence (Bird, 2007). The processes underlying epigenetic phenomena encompass both transient and enduring modifications, which induce structural adaptations in chromosomal regions, enabling them to record, transmit signals, or sustain altered activity states (Bird, 2007). The environment can influence epigenetic changes, and epigenetic modifications can buffer genetic variation, allowing the same combination of genes to produce different developmental outcomes (Bird, 2007; Sollars et al., 2003). Epigenetic systems have the capability to mark changes in chromosomal status and help preserve changes in genetic programming (Bird, 2007). Examples of epigenetic systems that can be inherited across generations include DNA methylation and the Polycomb/Trithorax systems (Bird, 2007; Schuettengruber et al., 2007; Ringrose and Paro, 2004). Epigenetic marks are generally responsive, not proactive, and this is a key unifying feature of such systems (Bird, 2007; Esterhuyse and Kaufmann, 2013). Epigenetics can be defined in consensus terms by the scientific community as a field concerned with unraveling how genomic function is modulated through control, which entails a multitude of mechanistic events leading to the inheritance of an individual's characteristics (Berger et al., 2009; Abhimanyu et al., 2021). Disruption to the epigenetic reprogramming at any stage of germ cell development or embryogenesis can alter the epigenetic phenotype (Dupont et al., 2009; Shi and Wu, 2009; Nestorov et al., 2013). In contrast to genetics, there is no obvious "epigene," and epigenetics provides an antidote to the idea that we are hard-wired by our genes (Bird, 2007; Kim, 2014; Ng et al., 2019).

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00013-6
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1.2 The Impact of Epigenetic Modifications on Gene Expression

The significance of epigenetic processes in gene regulation is an expanding field of understanding, with prevailing thought suggesting these effects are key players in natural species variation (Sideri et al., 2017; Eichten et al., 2014). Epigenetics offer a more fluid and transient approach to gene regulation when compared to genetic variation (Sideri et al., 2017; Eichten et al., 2014). The belief holds that epigenetic mechanisms may induce alterations in gene expression related to development or environmental responses (Eichten et al., 2014). This occurs because these mechanisms yield varying, interchangeable structural states that adapt gene activity without altering the DNA sequence (Venditti et al., 2020). For instance, the changing patterns of acetylation and deacetylation on histones can modulate chromatin, forming either "open" or "closed" domains governing gene activity (Venditti et al., 2020; Shvedunova and Akhtar, 2022). Similarly, DNA methylation can compress chromatin causing DNA unavailability for transcriptional machinery leading to gene silencing (Venditti et al., 2020; Moore et al., 2013; Irizarry et al., 2009). Furthermore, microRNAs manage the stability and accessibility of mRNAs for protein production translation machinery (Venditti et al., 2020; Tbeishat, 2022). Taken together, factors such as histone modifications, DNA methylation, and small noncoding RNAs all serve regulatory roles in gene expression via their epigenetic marks thereby influencing overall gene output (Gibney and Nolan, 2010; Venditti et al., 2020). Recent research in the field of epigenetics has suggested a significant role for RNA-based mechanisms in more widespread epigenetic regulation, despite their understanding being less comprehensive than that of DNA methylation and histone-based mechanisms (Gibney and Nolan, 2010; Wei et al., 2017). RNA has been identified as a key regulator in monoallelic expression of imprinted genes and X-chromosome inactivation (Gibney and Nolan, 2010). Furthermore, studies have noted a priming effect in response to various abiotic and biotic environmental triggers (Eichten et al., 2014). Essentially, this priming or training response leads to changes in gene expression or phenotype upon recurring exposure to stressful environmental conditions (Eichten et al., 2014). The initial stress exposure sets the stage for future adaptations, leading to more marked or quicker responses during subsequent encounters with similar stressors (Eichten et al., 2014). Several scenarios have linked chromatin modifications or DNA methylation changes to the priming response phenomenon (Eichten et al., 2014). The role of epigenetics in how organisms respond to their environment cannot be overlooked, as it often leads to changes in gene expression patterns (Eichten et al., 2014). An illustration of this is seen when rodents exposed to an enriched environment showed reduced levels of DNA methylation at their glucocorticoid receptor gene promoter, which coincided with improvements in their social interactions and ability to cope with stressful conditions (Venditti et al., 2020; Turecki and Meaney, 2016).

The microRNA-mediated upregulation in brain-derived neurotrophic factor (BDNF) was particularly evident under conditions of environmental enrichment. At the same time, various mechanisms, including education-based interventions, regular exercise routines, appropriate dietary plans, and adequate sleep, led to changes in DNA methylation profiles (Venditti et al., 2020; Bai et al., 2015; Gutierrez et al., 2020). Furthermore, the phenomenon of environmental enrichment resulted in increased histone acetylation in the hippocampus and neocortex (Venditti et al., 2020), underscoring the fundamental role of chromatin modifications in numerous epigenetic phenomena (Eichten et al., 2014).

Instances where chromatin modifications at a single gene differ between two plant species, tissue types, or under different treatments can be identified as an instance of epigenetic regulation for that

gene (Eichten et al., 2014; Hunter et al., 2012). Environmental triggers are also known to induce epigenetic modifications that ultimately affect the expression of genes (Venditti et al., 2020).

1.3 The Role of Epigenetics in Disease Development

Epigenetic modifications, including changes in histone modification patterns, are prone to potential loss within one generation and demonstrate tissue-specific variability (Gómez-Díaz et al., 2012). Thus, delineating an unequivocal relationship between a particular disease phenotype and the complex interplay of the epigenome presents significant obstacles (Gómez-Díaz et al., 2012). Moreover, epigenetic changes may be expressed in a sex-specific, age-dependent, or parent-of-origin-specific manner, causing further complications (Gómez-Díaz et al., 2012; Mahan et al., 2017). Thus, longitudinal and transgenerational studies accompanied by comprehensive data are needed to overcome these difficulties (Gómez-Díaz et al., 2012). Consequently, new epidemiological approaches are being developed to control cause−effect relationships (Gómez-Díaz et al., 2012). Epigenetic inheritance is believed to play a crucial role in the dynamics of host−pathogen interactions (Al Akeel, 2013; Reisenauer et al., 1999). This implies that the reprogramming of host genes at an epigenetic level is implicated in bacterial pathogenesis; however, the full extent and significance of this mode of inheritance are still not fully comprehended (Gómez-Díaz et al., 2012). Moreover, there exists an urgent demand for comprehensive investigations into epigenetics to establish robust theoretical evolutionary models (Hunter et al., 2012; Gómez-Díaz et al., 2012). Furthermore, it is likely that transgenerational immune priming in plants and invertebrates involves an epigenetic component, but the precise mechanism by which this inheritance occurs remains unknown (Gómez-Díaz et al., 2012). Finally, through epigenetic inheritance, favorable combinations of traits can be transmitted across generations directly to the offspring, including those acquired during an individual's lifetime, thereby providing an additional pool of phenotypic variation on which natural selection can act (Klosin et al., 2017; Hernández López et al., 2014; Gómez-Díaz et al., 2012). Thus, it can be concluded that exploring the field of epigenetic epidemiology holds great promise for future research on infectious diseases, as both epigenetics and infectious diseases have reciprocal influences on each other.

2. Epigenetics and Its Role in Infectious Diseases

2.1 The Impact of Epigenetic Modifications on Infectious Disease Transmission and Host Susceptibility

The role of epigenetics in some diseases, such as cancer and other diseases, is well understood to some extent in regards to the impact of the environment on organisms and their offspring. For example, studies on humans have shown how parental generations that faced food scarcity passed on epigenetic signatures for increased insulin production to their progeny. This is beneficial to those following generations that face famine but can be detrimental to offspring that live in resource-rich environments. That is because when food is abundant, an overproduction of insulin can lead to obesity, diabetes, and other health conditions (Liotti et al., 2022; Jablonka and Raz, 2009; Dominguez-Andres and Netea, 2019). Other studies have demonstrated how mental stressors, such as anxiety or depression, can alter the epigenome and result in progeny being more prone to express anxiety or depression (Bohacek and Mansuy, 2013). These previous studies were more focused on the transmission of

heritable phenotypes through epigenetics, but newer studies are showing even infectious pathogens (bacteria, viruses, and parasites) along with the environment are able to modify an organism's epigenome, and this epigenomic modification can be passed on to the offspring through epigenetic inheritance. Commonly noted epigenetic modifications resulting from infectious diseases are modifications found in the promoter regions of genes (CpG sites), changes in DNA methylation patterns, histone modification, and changes in noncoding RNA molecules. Modifications in these regions affect the susceptibility of a host, and sometimes their offspring, to be more or less susceptible to the pathogen, as well as even the severity of the infection (Liotti et al., 2022; Bohacek and Mansuy, 2013; Syn et al., 2016).

2.2 Epigenetics Role in Resistance to Infectious Pathogens

A healthy immune system and immunity rely on the ability of an organism to adapt to a constantly changing environment, which is a hallmark of epigenetic modifications (Liotti et al., 2022). As stated in the introduction, there are several types of epigenetic modifications that can affect the phenotypic expression of an organism in its lifetime, and such modifications play a crucial role in the transmission of and susceptibility to infectious diseases. This is because epigenetic modifications can alter the expression of genes that are involved in immune responses, leading to an organism's increase or decrease of susceptibility to a pathogen after exposure, known as trained immunity. After exposure to a pathogen or attenuated vaccine, an epigenetic (and metabolic) reprogramming of the immune system occurs, which can provide enhanced resistance to the pathogen in the future (Liotti et al., 2022; Netea et al., 2011; Kondilis-Mangum and Wade, 2013). Additionally, recent studies have found pathogenic susceptibility is not limited to just the host, as epigenetic modifications can be passed down to offspring. Epigenetic inheritance is the passing down of epigenetic modifications to offspring, with such modifications affecting the gene activity or expression due to environmental, developmental, and functionality of cells, or, in the case of infectious pathogens, in response to an infection (Bohacek and Mansuy, 2013; Jablonka, 2004). Thus, it is not always the case that the generation that is exposed to the pathogen is the only generation with epigenetic modifications (Skinner, 2011). Epigenetic inheritance can be advantageous over classical inheritance as it provides a rapid adaptation opportunity for an organism and its offspring in the face of new environmental challenges or pathogens (Gayen and Betal, 2022; Atlante et al., 2020; Mongelli et al., 2021).

2.2.1 Epigenetic Host Response to Infections

Upon confronting a foreign pathogen invasion, an organism's immune system is stimulated. Primarily, the innate immune cells that have a limited life span take up the responsibility of sparking off the adaptive immune cells. The adaptive immune system is composed of B and T cells, which display a distinct class of antigen surface receptors and bring forth naïve B and T cells through various complex processes that keep circulating within the organism. In reaction to an infection, the naïve B and T cells specific to that particular pathogen undergo multiplication and respond until they successfully eliminate the infection. Consequently, all mobilized innate immune cells along with some adaptive B and T cells perish, while a small population of pathogen-experienced B and T cells are retained by the immune system, giving rise to memory B and T cells for better-equipped responses against future same-pathogen invasions. DNA methylation can control gene expression involved in T cell differentiation and function while histone modifications can alter gene expression associated with immunity

response. There is also chromatin remodeling essential for effective T cell activation and antigen recognition—critical determinants for improved responses to subsequent infections.

Numerous studies have noted that vaccine-dependent trained immunity through epigenetic modifications protects individuals from reinfection, which has been further supported by numerous SARS-CoV-2 studies which highlight how vaccinated individuals were less susceptible to secondary infections (Youngblood et al., 2013; Giamarellos-Bourboulis et al., 2020; Doesschate, 2020), but there are fewer studies that have examined the role of epigenetics in disease resistance after exposure. Stein et al. (2021) conducted an exhaustive investigation on the methylome landscape in individuals infected with Human Immunodeficiency Virus (HIV) and exposed to tuberculosis, aiming to uncover the influence of epigenetic modulation on amplified immune responses against TB. The study elucidated three distinct regions exhibiting differential methylation in individuals infected with HIV and exposed to TB, as opposed to those unexposed. The specific SNP regions discovered exhibited associations with methylation levels, TB, and SNPs' interplay with the methylation level linked to TB. Additionally, these loci had a regulatory role in immune cells responding to TB infections. The concluding inference was that these epigenetically altered, methylated sites could potentially be a biological determinant for resistance against TB among immunocompromised individuals (Stein et al., 2021).

2.2.2 Epigenetic Inheritance in Resistance to Infectious Pathogens

Although the current state of knowledge regarding the role of epigenetics in improving resistance through trained immunity is still limited, the field of epigenetic inheritance in the context of immunity to infectious diseases is only growing. This expanding knowledge base relies heavily on studies conducted on rodents, *C. elegans*, and other model organisms (de Candia and Matarese, 2021; Katzmarski et al., 2021).

Studies on *C. elegans* found that these worms can pass on improved resistance to bacterial infection phenotypes to their offspring through histone modifications (Belicard et al., 2018). Insect research has found that insects that are exposed to a low dose of a virus are more resistant to it, a similar finding in human vaccine studies, except insects can pass on this improved resistance to their offspring via epigenetic modification (Tidbury et al., 2011). Although a few studies have noted insects and mice are able to pass on vaccine-related adaptive immune responses to their offspring, this has not been discovered in humans yet (Liotti et al., 2022; Tidbury et al., 2011), but epigenetic inheritance does play a role in improved resistance in other ways.

Often, epigenetic inheritance occurs through the germ line, usually during crucial moments of development (Skinner, 2011). The majority of epigenetic inheritance work has focused on the impact of environmental toxins, often during pregnancy, on subsequent generations. These pioneering studies provide a foundational understanding of epigenetic inheritance and a base for more recent studies that now seek to understand the influence of pathogens on epigenetic inheritance (Manikkam et al., 2012; Thompson et al., 2020). Much of our understanding of epigenetic inheritance and its role in pathogen resistance is based on plants (Jablonka and Lamb, 1998; Tirnaz and Batley, 2019). In these studies, on maize, *Arabidopsis*, rice, and other crops, what was found was that DNA methylation and DNA demethylation altered gene expression and transposon silencing. Many studies noted that miRNAs were responsible for the improved resistance response in plants after exposure. Many of these pathogen-induced DNA methylation changes resulted in phenotypic and genotypic changes through transcriptome reprogramming and a change in genome stability. These epigenetic modifications permit the plant to adapt fast to stressors and improve its defense against pathogens through hypo- and

hypermethylation processes. Furthermore, studies have noted these epigenetic modifications are passed on to subsequent generations, improving generational resistance to pathogens (Jablonka and Lamb, 1998; Tirnaz and Batley, 2019). More studies are needed to determine the role of epigenetic inheritance in mammals and humans in improved infectious disease resistance (Jablonka and Raz, 2009).

2.3 Epigenetics Role in Greater Susceptibility to Infectious Pathogens

Although studies have noted how epigenetic modifications are improving host resistance to pathogens, research has also found how pathogens, in particular, as well as environmental variables, are modifying a host's epigenome to increase susceptibility to infection. Furthermore, as noted, epigenetic inheritance and epigenetic changes can be positive, but they can also be maladaptive if the stressors are no longer an issue, but the adaptation remains (Jablonka and Raz, 2009; Bohacek and Mansuy, 2013), thus these initial modifications in response to infections may be beneficial, but in the long run, they can prove detrimental (Kondilis-Mangum and Wade, 2013). Epigenetic modification of or by pathogens or by a host in response to the environment or pathogen not only impacts the susceptibility of the host or its progeny to infection but it can also impact the progress of the infection and its severity in the host or offspring (Jablonka and Raz, 2009; Minarovits and Niller, 2013).

2.3.1 Pathogen Modification of Host Epigenome to Increase Host Susceptibility

Recent studies are now examining how infectious pathogens incorporate epigenetic mechanisms in their life cycles to modify host processes to create optimal environments for survival within the host. In the previous section, it was noted how exposure to TB can improve a host's resistance to TB, but other studies have highlighted how *Mycobacterium tuberculosis* (*Mtb*) can alter the host's epigenome to modulate the transcriptional processes of the host's immune response using several epigenetic mechanisms. Histone modification by *Mtb* alters both innate and adaptive immune cells and acts as an agonist to *TLR2*, resulting in the downregulation of antigens in T cells (Yadav, 2015; Kathirvel and Mahadevan, 2016). The *Mtb* also uses DNA methylation processes through hypermethylation of CpG islands around a gene known to increase the pathogenesis of TB (the *VDR* gene), ultimately resulting in an increase in host susceptibility to TB (Andraos et al., 2011). Lastly, *Mtb* disrupts the miRNA expression profiles of the host to persist and survive at the intracellular level inside macrophages (Yadav, 2015; Kathirvel and Mahadevan, 2016). Yet, *Mtb* is not the only pathogen to evade the host's immune system through epigenetics; there are numerous other viruses, bacteria, and parasites that take similar measures to *Mtb* to persist and survive (Bohacek and Mansuy, 2013; Niller and Minarovits, 2018).

2.3.2 Increased Severity of Infection to Host and Offspring

Of growing concern are newer findings demonstrating that the epigenetic modification of the host by pathogens does not only result in increased susceptibility but also contribute to the proliferation and severity of the infection in the host too, as well as the later life development of tumors or cancers. Much attention has been given to the herpesvirus Epstein–Barr (EBV), commonly known as mononucleosis, one of the most infective viruses, as it is estimated that 90% of adults worldwide have been exposed to it (Scott, 2017). This is concerning, as research is finding that EBV is excellent at manipulating the host epigenome for its survival, and such epigenetic modifications can be severe and

detrimental to the host. Furthermore, EBV-induced epigenetic modifications can be passed on from one generation to the next, providing a negative example of epigenetic inheritance (Scott, 2017).

EBV is able to persist and enter latent stages in the host by invading memory B cells or even creating new memory B cells through naïve B cells. The latency of the virus is based on guided promotor silencing or switching, resulting in restricted viral gene expression. Many of the epigenetic modifications EBV makes to the host genome promote EBV's ability to evade the immune system. But EBV is a tumor virus, and newer studies are finding that the epigenetic changes by EBV through DNA methylation and chromatin structure is altering the expression of growth-promoting or suppressing genes to potential oncogenic consequences to the host. This is because EBV is able to immortalize B cells into lymphoblastoid cells (LCLs), resulting in immortal growth. In fact, EBV is responsible for several forms of benign and malignant lymphoma, such as Burkitt's lymphoma. In Burkitt's lymphoma, the research found the majority of CpGs of EBV-positive Burkitt's lymphoma were methylated, and many Burkitt's lymphoma tumors show DNA hypermethylation. Furthermore, EBV-positive Burkitt's lymphoma tumors carried fewer mutations in comparison to non-EBV Burkitt's lymphoma tumors, suggesting EBV epigenetics are replacing mutations (Scott, 2017; Kang et al., 2002; Giulino-Roth et al., 2012). Similar findings have been discovered in other EBV-related tumors, as well as in the most highly DNA-methylated tumors (Scott, 2017). EBV is not the only pathogen that influences epigenetic modification of the host, which can result in oncogenic concerns as other viruses are known to silence tumor suppressor genes or impact the expression of lymphomas (Al Akeel, 2013; Niller and Minarovits, 2018). Such findings have even inspired a new field: Patho-epigenetics, the study of pathologic changes elicited by epigenetic dysregulation (Minarovits and Niller, 2013).

2.3.3 The Tradeoff of Epigenetic Inheritance

For the most part, studies on epigenetic inheritance focus on the positive outcome of passing on improved resistance to a pathogen through generations. As noted, this has been strongly supported in crop studies, but murine models and other model organisms have been discovered to transmit epigenetic modifications to offspring for improved resistance. Yet, epigenetic inheritance is not optimal, this is because the trained immunity being passed down comes at a cost: increased risk of developing atherosclerosis and cardiovascular diseases. More studies are necessary to determine the risk of increased trained immunity through epigenetic modifications (de Candia and Matarese, 2021; Katzmarski et al., 2021; Mukherjee et al., 2015).

3. The Role of Infectious Diseases in Epigenetic Alterations

Infectious diseases have recently been identified as a major driver of profound epigenetic changes. Recent research highlights that pathogens can modify the chromatin structure and transcriptional program of host cells by affecting a wide range of epigenetic factors (Bierne et al., 2012; Goyal et al., 2021).

A growing body of research strongly implies that the pathological process in acute or chronic infectious diseases is related to infection-induced cellular epigenomic changes. For instance, many viruses induce epigenetic alterations like DNA methylation in the host genome to avoid host immune reactions like inflammation (Nehme et al., 2019). This concept is particularly important given that some infections can be prolonged and that pathogens modify their host cells to ensure their own

persistence. Furthermore, due to the heritable nature of certain epigenetic modifications, infection-induced epigenetic changes can modify the function of the host cell for a long time after the primary infections have been eliminated.

Although it is apparent that infections can elicit epigenetic modifications in the immune cells of the host, discerning whether these alterations are instigated by the pathogen or by the host cell's response to infection remains uncertain. Moreover, while the presence of epigenetic changes has been observed, there is limited understanding regarding the specific mechanisms or molecules responsible for mediating these modifications and their origins. Advancements in technology such as next-generation sequencing, microarray, chromatin immunoprecipitation sequencing, CRISPR-Cas9 gene-editing, and single-cell sequencing technologies have facilitated significant advancements in comprehending the role of epigenetic pathways in the pathogenesis of infectious diseases. Given that epigenetic modifications cannot be detected the DNA sequence level, investigating genome-wide chromatin modification patterns through epigenome mapping may prove instrumental in identifying disease-associated genes and facilitating innovative development of therapeutic and diagnostic protocols.

3.1 Mechanisms by Which Infectious Diseases Induce Epigenetic Change

Infectious diseases can have a wide array of effects on gene expression. When a pathogen infects a host, it can trigger a cascade of signaling events of the host's immune response and interfere with the host's epigenetic regulation to favor the pathogen's survival or replication.

When a pathogen invades a host organism, it initiates a cascade of signaling processes, leading to alterations in the epigenetic profile of the host's immune cells (Ramendra et al., 2021). One plausible mechanism for this phenomenon involves the activation of pathogen-associated molecular patterns (PAMPs). PAMPs are distinct molecular motifs that are specifically linked to pathogens and are detected by pattern recognition receptors (PRRs) present on the host's cells. The binding of PAMPs to PRRs initiates a cascade of signaling events that lead to the induction of expression of proinflammatory cytokines, which in turn can regulate the immune response. Cytokines are small polypeptides or glycoproteins that act as signaling molecules that bind to receptors on immune and nonimmune cells, activating intracellular signaling pathways and regulating gene expression. Different cytokines can have different effects on the immune response to infectious pathogens. For example, interferons (IFNs) possess the ability to activate antiviral responses, thereby exerting control over viral infections. Their activity involves the induction of gene expression responsible for encoding antiviral proteins, as well as the enhancement of immune cell function involved in eliminating viral infections. Additionally, other cytokines such as tumor necrosis factor alpha (TNF−α) and interleukins (ILs) contribute to the activation of the inflammatory response during infection. These cytokines induce the expression of genes associated with inflammation, including other cytokines and adhesion molecules that recruit immune cells to the infection site and stimulate their clearance of pathogens. The recognition of pathogens by PRRs is crucial for initiating innate immunity, which serves as the primary defense mechanism until adaptive immunity develops. While substantial knowledge exists regarding the immune response following pathogen recognition, our understanding of how epigenetics shapes transcriptional responses downstream remains limited. Notably, recent research on toll-like receptors (TLRs), a extensively studied class of PRRs, has firmly established that TLR-induced epigenetic modifications represent a pivotal regulatory landscape governing both positive and negative regulation of TLR-induced genes.

Another way that infectious diseases can affect epigenetic changes is by altering the host's epigenetic machinery and enzymes. For example, some viruses can inhibit or degrade the host's DNA methyltransferases, which function to add methyl groups to the DNA and silencing gene expression (Kuss-Duerkop et al., 2018). This can result in global hypomethylation and activation of viral genes or oncogenes. Similarly, infectious diseases can affect epigenetic changes through expression of proteins, toxins, metabolites, or antigens (Abhimanyu et al., 2021; Hardy and Mann, 2016). For example, the hepatitis B virus (HBV) HBx protein functions as an epigenetic deregulator that can alter host microRNA (miRNA) profile, which in turn modulates the viral load and increases persistence (Sivasudhan et al., 2022). The E6 and E7 oncoproteins of the human papillomavirus (HPV) are viral proteins that are essential in the development of cervical carcinogenesis. E6 and E7 can alter the expression of genes involved in cell cycle regulation, DNA repair, and immune system function, ultimately promoting the development and progression of HPV-associated cancers (Yim and Park, 2005).

Infection by pathogens can induce various cellular stresses, leading to alterations in cellular homeostasis and triggering specific stress responses (Rodrigues et al., 2018; Trinh, 2022). For example, infections can trigger oxidative stress by generating an excess of reactive oxygen species (ROS). ROS can activate various signaling pathways that regulate inflammation and immune response genes (Zheng and Sawalha, 2022). Oxidative stress can also induce epigenetic changes, such as DNA methylation, as seen in HPV infection-induced DNA methylation of the promoter region of p16INK4a, a tumor suppressor gene (Carestiato et al., 2013). Inflammation is another common response to infection that aims to eliminate pathogens and repair tissue damage. Inflammation can impact the structure of chromatin as well as the epigenetic alterations of DNA and histones (Tan et al., 2022).

3.2 Epigenetic Alterations Associated With Different Infectious Diseases

Epigenetic modifications induced by infectious diseases have been documented in different categories of microbial pathogens encompassing viruses, bacteria, fungi, and parasites (Zune et al., 2021). These infectious agents exert an influence on the host's epigenome, thereby instigating changes in gene expression patterns that contribute to an augmented predisposition toward diverse chronic disease such as cancer.

For example, epigenetic changes induced by coronaviruses such as Middle East respiratory syndrome (MERS) and severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1) can occur through interference with host antigen presentation or activation of interferon-response genes (Menachery et al., 2018). In the case of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), older individuals aged 65 and above face a significantly higher risk of severe illness and death, accounting for approximately 80% of hospitalizations and exhibiting a 23-fold increased mortality risk compared to those under 65 (Mueller et al., 2020). This vulnerability among the elderly to SARS-CoV-2 is thought to be influenced by the role played by the epigenome in viral entry (Mueller et al., 2020). Investigating alterations in DNA methylation age within immune cells and various blood cell types prior to, during, and subsequent to infection can yield pivotal revelations pertaining to the role of age-associated epigenetic modifications in disease severity, as well as the impact of viral infection on aging-related epigenetic alterations. Moreover, viruses such as hepatitis C virus (HCV), Epstein−Barr virus (EBV), HBV, and Human Immunodeficiency Virus type-1 (HIV-1) have well-established

associations with cancer (White et al., 2014). EBV, for instance, has been found to manipulate host epigenetic machinery, leading to enduring reprogramming of the host's epigenome (Birdwell et al., 2014). Similarly, HCV, HBV, and HIV-1 have been demonstrated to utilize epigenetic pathways to ensure gene regulation during chronic or latent infection (Fischer, 2020). These modifications can result in the development of particular cancers in patients.

Not only do viruses exploit host cells for their survival but parasites and specific bacteria also depend on cellular machinery to replicate and accomplish their life cycles. *Listeria monocytogenes* is a well-studied example of bacterial interference with cellular epigenetic machinery. Foodborne infections can have severe consequences, especially in individuals who are immunocompromised. *Listeria monocytogenes* can alter the "histone code" of the host cell through the bacterial toxin listeriolysin O, leading to epigenetic effects (Hamon et al., 2007). Moreover, the genome of the bacterium harbors additional virulence factors that can potentially affect the epigenetic machinery of the host cell. For instance, the surface-expressed protein internalin B (InlB) of *Listeria monocytogenes* can aid in the infiltration of nonphagocytic cells by binding to c-Met, a tyrosine kinase receptor present on the plasma membrane of diverse human cell types. Another protein, namely Listeria nuclear-targeted protein A (LntA), which is secreted by the bacterium, can also interact with c-Met and trigger a signaling cascade that indirectly leads to histone deacetylation, ultimately resulting in gene silencing (Eskandarian et al., 2013). These epigenetic changes induced by *Listeria monocytogenes* can have significant consequences for the host cell's gene expression and ultimately lead to disease pathogenesis.

3.3 Long-term Effects of Epigenetic Changes Induced by Infectious Diseases

There has been substantial discussion over whether infectious diseases can cause long-lasting epigenetic modifications that last even after the infection has been resolved (Fischer, 2020). While direct molecular evidence is still lacking, there are several indications that such mechanisms exist in bacterial, viral, and parasitic infection systems. One example of persistent epigenetic changes is seen in parasitic infections from *Leishmania*, *Toxoplasma*, and *Theileria*. Studies have shown that these parasites cause persistent epigenetic changes in the host's cells even after the parasite has been cleared (Cheeseman and Weitzman, 2015; Robert McMaster et al., 2016). Researchers hypothesize that these protozoan parasites cause epigenetic changes at the DNA methylation level of the host cell genome, which permanently downregulate and inhibit the host's defense mechanisms and promote the pathogen's intracellular survival and replication (Robert McMaster et al., 2016). Similar outcomes were observed with *Helicobacter pylori* infections, the type of bacteria implicated as a risk driver for developing gastric cancer (Cheng et al., 2013). *Helicobacter pylori* can alter gene expression in the human gastric mucosa through DNA methylation. However, many of the host genes remain methylated long after the bacterium has been eradicated via antibiotic treatment (Capparelli et al., 2022).

The existence of long-term epigenetic effects of infectious diseases has important implications for public health. Individuals who have been infected with certain viruses or bacteria may be at increased risk of developing certain types of cancer or other chronic diseases, even after the infection has been cleared. Understanding the molecular mechanisms underlying these effects could lead to the development of new interventions, such as epigenetic therapies or lifestyle modifications, to mitigate the long-term health effects of infectious diseases.

4. Potential Role of Laboratory Tests in Detecting Infectious Diseases Based on Epigenetic Signatures

4.1 Epigenetic Biomarkers and Laboratory Tests in Detecting Infectious Diseases

The use of epigenetic biomarkers to detect infectious diseases is a fast-paced field in diagnostic medicine. Numerous studies have revealed a high degree of correlation between infectious diseases and epigenetic alterations, which may be specific to certain infectious agents and can be employed as biomarkers for their detection, for understanding the severity of the disease, or for monitoring the effectiveness of treatment. The epigenetic alterations are associated with a wide range of infections caused by bacterial, viral, fungal, and protozoan pathogens (Barturen et al., 2022; Tolg et al., 2011; Marr et al., 2014).

Epigenetic biomarkers have arisen as a promising strategy for diagnosing bacterial infections and identifying the multidrug-resistant nature of the pathogen (Crimi et al., 2020). For example, increased levels of certain miRNAs in *M. tuberculosis* are associated with drug resistance (Alipoor et al., 2019; Muhammad et al., 2022). Similarly, Yan et al. identified differentially expressed long noncoding RNAs in blood mononuclear cells in response to drug-resistant TB (tuberculosis) infection and drug-sensitive TB infection (Yan et al., 2018). Epigenetic biomarkers can not only provide diagnosis but they can also pave the way for developing new therapeutic targets that can overcome the challenge of multidrug resistance in pathogens (Crimi et al., 2020). For example, increased acetylation of H4 and phosphorylation/acetylation of H3 in *L. monocytogenes* provides both diagnostic and therapeutic targets (Schmeck et al., 2005). This is critical in this era where drug resistance to pathogens is on the rise and is a serious concern. Research on epigenetic alterations in individuals infected with human papillomavirus (HPV) variants suggests that they can also be used as a biomarker to distinguish the virus genotype causing the infections (Yin et al., 2016; Silva Da, 2021). Furthermore, alterations in the epigenetic markers can serve as an additional source of information to identify the progression of various infectious diseases (Santos et al., 2019). For example, diverse patterns of CpG sites with altered methylation were observed in both severe and mild cases of SARS-CoV-2, predominantly linked to interferon pathway activation and B- and T-lymphocyte hyper-activation (Barturen et al., 2022).

One significant advantage of utilizing epigenetic biomarkers for diagnostic purposes is their inherent stability. Giménez et al. has shed light on the remarkable stability displayed by these markers in a multitude of biological contexts. Specifically, microRNAs and posttranslational modifications in histones are shown to endure fluctuations with exceptional resilience across various body fluids such as plasma, serum, urine, saliva, semen, and vaginal secretions (García-Giménez et al., 2017). Furthermore, this stability extends to primary tissue preparations encompassing fresh and frozen tissues as well as dried blood spots (Guthrie cards) and formalin-fixed paraffin-embedded (FFPE) tissues (García-Giménez et al., 2017). This stability of epigenetic biomarkers in a range of sample types and tissue preparations makes them ideal for use in diagnostic assays, as they can be reliably detected and quantified even after extended storage periods or suboptimal storage conditions. Moreover, the stability of these markers enables the use of noninvasive sample collection methods, such as saliva or urine, which are less invasive and more easily obtained than other types of clinical specimens. This could improve screening rates in patients by making it convenient for them to be tested. Therefore, this could lead to earlier detection of diseases, ultimately leading to better health outcomes.

4.2 Epigenetic Lab Tests for Detecting Gene Expressions Related to Infectious Diseases

There has been tremendous effort and development in new laboratory technologies for epigenetic analysis. Next-generation sequencing (NGS) and Polymerase Chain Reaction (PCR) have advanced epigenetic studies to the next level. It has become a common choice among all types of epigenetic biomarkers for their detection and discovery. For instance, several laboratory methods have been developed for the detection of DNA methylation (Pajares et al., 2021). The commonly used method for DNA methylation analysis is sodium bisulfite conversion. Bisulfite-converted genomic DNA techniques have revolutionized the analysis of DNA methylation in both high-throughput technologies and gene-specific studies. This process involves converting unmethylated cytosine to uracil by using sodium bisulfite, which can be detected using sequencing-based technologies and methyl-specific PCR (MS-qPCR). Two types of bisulfite sequencing are performed: whole-genome bisulfite sequencing (WGBS) and targeted bisulfite sequencing (TBS). For the diagnosis of infectious disease, detection of a specific epigenetic alteration needs to be detected. TBS is used to get methylation information for a specific region in the genome with single-nucleotide information. MS-qPCR is also used to detect methylation at a region of interest by designing primers specific to the methylated or unmethylated DNA sequence, MSP can selectively amplify and detect only the target sequence of interest, depending on its methylation status (Chiara, 2020).

Similarly, for the detection of histone modification, the technique commonly used is chromatin immunoprecipitation (ChIP), in which specific antibodies are used to enrich chromatin fragments containing a particular histone modification, and then PCR, microarrays, or NGS can then be used to identify genomic regions associated with that modification. miRNA biomarker studies usually employ qPCR and NGS, using reverse transcription in both (Tribolet et al., 2020). The use of this method allows the quantification of all microRNAs in a given sample, even those with novel nucleotide sequences. However, its main application lies as the method of choice for the initial identification of microRNA biomarkers and using a subsequent validation process, that often includes the use of qPCR assays. Recently, the sensitivity of the PCR methods has been improved by using digital PCR. The technologies mentioned above can be easily adapted to more advanced versions of molecular techniques like digital PCR for enhancing signals or analysis (Jelinek et al., 2018; Hashimoto et al., 2007; Leemkolk, 2022). Furthermore, to enhance the results obtained from laboratory technologies, a plethora of commercial kits are available that allow protocols for a specific type of sample and epigenetic alteration, maximizing the yield and stability of the biomarker (García-Giménez et al., 2017).

4.3 Diagnostic Use of Gene Expression Profiling in Epigenetics for Infectious Diseases

Gene expression profiling has emerged as a valuable tool for diagnosing infectious diseases by identifying the specific genes that are activated or suppressed in response to an infection. For instance, recent studies by Lee et al. and Momeni et al. have demonstrated the potential of gene expression profiling in identifying active tuberculosis patients and specific biomarkers for COVID-19, respectively (Lee et al., 2016; Momeni et al., 2023). Gene expression profiling involves the analysis of messenger RNAs (mRNAs) or translated proteins in a tissue, such as blood. It allows the quantification

of gene expression levels by detection and measurement of the abundance of the final mRNA product or its precursor (Gliddon et al., 2018). Several methods are available for quantifying cellular RNA, including microarrays, RT-qPCR, RNA sequencing (RNA-seq), and third-generation sequencing. Among these methods, RT-qPCR is the preferred choice for studying a relatively small, previously identified set of genes and is broadly recognized as the gold standard for RNA quantification (Gliddon et al., 2018). Infectious diseases can be diagnosed using gene expression profiling by identifying the genes that are activated or suppressed during an infection. This information can help differentiate between different pathogens and predict antibiotic sensitivity in bacterial infections (Suzuki et al., 2014). In conclusion, gene expression profiling stands poised to bring about a paradigm shift in both the diagnosis and management of infectious diseases. By offering comprehensive and real-time information on the underlying molecular processes involved in infections, this cutting-edge technique holds great promise for improving disease detection and treatment outcomes. The continual development of novel RNA quantification technologies will undoubtedly continue to be a critical driving force behind the advancement and application of gene expression profiling.

5. Strategies to Limit the Impact of Infectious Diseases on Epigenetics

Various pathogens alter the epigenetic landscape of host cells during their infection lifecycle to enable and promote the pathogens' propagation (Menachery et al., 2018). Additionally, epigenetic modifications allow for viruses to enter latent states where they are "integrated" into the host genome and lie in wait for the appropriate stimuli before reactivating. Therapies targeting these epigenetic modifications provide novel avenues to control and potentially cure infections that have previously been untreatable. This section will begin by reviewing the epigenetic modifications of certain pathogens and how these modifications are countered and mitigated. Strategies to prevent future epigenetic modifications by pathogens are then examined, and finally, the role of research in the development of novel therapies is discussed.

5.1 Strategies to Reduce the Impact of Infectious Diseases on Epigenetics

The role of epigenetics in regulating and, at times, enhancing the infectious life cycle of human pathogens is becoming increasingly important as infectious mechanisms are elucidated (Liu et al., 2019; Pei et al., 2020; Kee et al., 2022). One such example is the HIV, which during its latent stage of infection in $CD4^+$ memory T cells and myeloid cells, is silenced by deacetylases and methyltransferases, preventing its purging (Bouchat et al., 2012; Margolis, 2011; Moron-Lopez et al., 2021).

More generally, DNA viruses are known to form chromatin structures during their infection cycle. Consequently, inhibitors of demethylation have been found to block viral genome transcription of DNA viruses. Interestingly, these inhibitors have been shown to enhance the infections and severity of some RNA viruses, such as influenza A virus (IAV) (Shan et al., 2017). This shows the capability of targeted antiviral treatments for certain classes of viruses. Furthermore, methyltransferase inhibitors of the methyltransferases EZH2 and EZH1 have been found to prevent a broad range of viral infections, ranging from herpes simplex virus (HSV) to adenovirus to Zika virus both in vitro and in vivo (Arbuckle et al., 2017). This indicates that, at least for viral infection, a therapeutic target exists that can prevent infection from multiple classes of viruses (RNA, DNA, etc.).

Epigenetic alternations are also being found to play a major role in other pathogens' infectious cycles (such as fungi), not just viruses (Habig et al., 2021). One such example is the emergence of drug resistance in fungi which utilizes both RNA-based and chromatin-based epigenetic modifications to gain and propagate drug resistance in fungi. RNA-based drug resistance can be achieved via a process known as epimutation (Villalobos-Escobedo et al., 2016; Chang et al., 2019). Epimutation uses silencing RNA (sRNA) that binds to the target DNA and induces degradation of the mRNA coding for the target protein. In this case, the fkbA gene, the gene encoding for FKBP12 which, is the therapeutic target of various antifungal agents, FK506 and rapamycin (Calo et al., 2014). The result is a phenotypic resistance to these antifungal agents, which can be reversed when the fungus stops producing the sRNA caused by environmental triggers (Chang et al., 2019).

5.2 Preventative Measures for Minimizing Epigenetic Changes Caused by Infectious Diseases

Preventative measures for minimizing epigenetic changes caused by infectious diseases involve a comprehensive approach that includes vaccination, early diagnosis and treatment, epigenetic-targeted therapy, a healthy lifestyle, environmental regulation, personal hygiene, probiotics usage, and research. Other important strategies include immune system support, avoidance of environmental toxins, antiviral medications usage, genetic counseling and screening, public health measures implementation such as social distancing, mental health support provision, alternative therapies exploration, personal protective equipment usage. Additionally, collaboration and communication enhancement along with education and awareness campaigns play a significant role.

By implementing these strategies, it is possible to reduce the impact of infectious diseases on epigenetics while promoting overall health and well-being. Moreover, hygiene practices in healthcare settings, vector control measures like quarantine and isolation protocols implementation can be effective. Health surveillance and monitoring are crucial aspects alongside international cooperation efforts. Contact tracing methods should also be employed to track the spread of emerging infectious diseases. Antibiotic stewardship programs need to be encouraged while providing healthcare worker education and training. Community engagement initiatives must be prioritized along with social distancing practices to ensure mental health support availability. Exploring alternative therapies can also contribute positively. Personal protective equipment usage should be emphasized in addition to collaboration and communication improvements.

Furthermore, global vaccination efforts should be strengthened alongside surveillance of vaccine-preventable diseases. Research on vaccine development and animal health monitoring are essential components as well. Water and sanitation infrastructure improvement is vital for overall disease prevention while emergency preparedness plans need to be in place.

Whereas therapeutics do not yet exist to prevent epigenetic modifications in the host, a variety of measures can be taken to mitigate the impact of epigenetic diseases postinfection. Two broad strategies are outlined below in the context of HIV.

Clinical therapeutics targeting the epigenetic modifications induced by pathogens are beginning to emerge and offer promising alternatives to conventional small molecule therapeutics. Two strategies have been created to combat HIV infection: the "shock and kill" strategy and the "block and lock" strategy (Nehme et al., 2019). Both of these strategies target the latent phase of HIV, which can cause

rebound when treated with conventional HIV therapies, such as combination antiretroviral therapy (cART) (Monforte et al., 2000).

The "shock and kill" approach seeks to reactivate HIV expression to allow the host immune system to target and eradicate HIV and HIV-infected cells (Archin and Margolis, 2014). This is done by using combination inhibitors (such as suberanilohydroxamic acid and vorinostat) targeting histone deacetylase, histone methyl transferase, and DNA methyltransferase inhibitors. In general, these compounds inhibit epigenetic modifications that favor the repression of HIV genes and thus enable latently infected cells to evade the host immune system (Nehme et al., 2019).

The "block and lock" strategy diametrically opposes the "shock and kill" strategy in that it seeks to lock HIV promoter genes and prevent them from re-activation, even under appropriate stimuli. This is accomplished by using a compound, didehydro-coristatin A (dCA), which binds and inactivates an HIV transactivator and prevents transcription of viral mRNA. Additionally, the Tat promoter with dCA binds to does not have a homolog found in humans and thus does not interfere with other immune responses during non-HIV infections (Das et al., 2011).

5.3 Reducing the Impact of Infectious Diseases on Epigenetics

Promising directions for diminishing the influence of infectious diseases on epigenetics involve: I. Continuous research and progress in relation to epigenetic-based therapies; II. Precision medicine approaches; III. Incorporation of epigenetics into public health strategies; IV. Creation of epigenetic vaccines; V. Implementation of environmental interventions; VI. Integration into global health initiatives; VII. Technological advancements; VIII. Management of viral infections via targeted therapy focused on epigenetics; IX. Interplay between host−pathogen regarding their respective epigenomes; X. Preservation of epigenetic memories concerning the evolution of infectious diseases; XI. Ongoing research as well as progress in vaccination protocols and techniques for diagnosis purposes. Furthermore, addressing climate change while concurrently developing novel antimicrobial agents are also integral components. The primary objective is to deepen our understanding pertaining to the alterations occurring at an epigenetic level due to infectious diseases while concurrently devising effective strategies with a view to minimizing their impact on human health.

Modulation of host epigenetics by infectious agents has been well documented over the past decade (Cole et al., 2016). Through modulation of histones and genetic content accessibility by methylation and acetylation of DNA, along with RNA silencing to affect posttranslational protein-producing capabilities described in the previous sections, infectious agents have demonstrated a profound interaction with host epigenetics. Consequently, these interactions provide novel opportunities for therapeutic targets (Cole et al., 2016). Indeed, as an example, clinical trials are underway to provide hope for the elimination of HIV and a cure for AIDS (Oriol-Tordera et al., 2022). Continued research and clinical trials in the area of epigenetic therapeutics provide a glimmer of hope for cures that have thus far been unattainable for a variety of infectious diseases.

6. Potential Applications of Epigenetic Factors in Preventive Medicine

Understanding the multilayer regulation of the human genome by epigenetic factors and their interactions with environmental stimuli, infectious agents, nutrition, and human behavior can provide a

bigger picture to fight human diseases. Intelligent and careful interpretation of these interactions can provide functional insight into the prevention and treatment policies for human disease. In this chapter, we will be studying epigenetics to prevent and treat infectious diseases, potential therapies associated with epigenetic manipulations, and health policies guided by epigenetic interactions.

6.1 Use of Epigenetic Modifications for Prevention and Treatment of Infectious Diseases

The epigenetic processes get influenced by environmental factors and infectious agents. The epigenetic processes, immune system, and infectious agents can affect each other, and the perturbation of one can influence the other in a tightly knit web of interactions. Infectious agents, like viruses, bacteria, and parasites, interact with the immune system and impact epigenetic processes. Pathogens can directly affect the epigenetic regulation of host genes, manipulating the immune response for their own benefit. Moreover, the immune response to infections can trigger epigenetic changes in immune cells, leading to enduring effects on the functionality of the immune system.

Exposure to environmental pollutants may alter epigenetic modifications, leading to changes in immune function and potentially increasing susceptibility to infections. Similarly, dysregulation of the immune system due to genetic or epigenetic factors can result in a wide range of diseases, including cancer, autoimmune disorders, and chronic inflammation.

Understanding the interplay between the immune system, infectious agents, and epigenetic processes is crucial for unraveling the mechanisms underlying health and disease. These interactions are being explored in preventing and treating infectious diseases. While the field of using epigenetic modifications for preventing or treating infectious diseases is still evolving, there are several promising avenues of research that have been explored. Here are some examples:

Modulating immune response: Epigenetic modifications can be used to boost the host's immune response against pathogens. For instance, histone deacetylase inhibitors (HDACi) have been shown to enhance the expression of interferon-stimulated genes (ISGs), which play a crucial role in antiviral immunity (Sakamoto et al., 2004). Epigenetic modifications can be targeted to enhance the immune response against infectious agents. TET (ten-eleven translocation) enzyme can reverse stable DNA methylation in cells. The TET enzyme has regulatory functions in gene expression, differentiation, and suppressing tumor formation. Agents like vitamin C and 2-hydroxyglutarate can modulate TET enzyme activities. This modulation can be used to enhance the immune system to fight infectious diseases (Lio and Rao, 2019). Antigen processing and presentation are impaired in tumor cells due to epigenetic silencing. The expression of immunosuppressive cytokines in the tumor microenvironment also hinders the immune system's ability to recognize and act on the infected cells. Epigenetic modulation to restore the natural recognition of infected cells can be a great approach to fighting infected cells (HÃ©ninger et al., 2015). Hypomethylating agents like Azacitidine and 5-aza-2′-deoxycitidine have been successfully used to restore gene expression of hypermethylated genes in tumors (Jones and Taylor, 1980; Momparler et al., 1990).

Epigenetic modifications can influence the interactions between the host and pathogens. The parasite Toxoplasma gondii alters host DNA methylation patterns, leading to immune dysregulation and increased susceptibility to infection (Hakimi et al., 2017). By understanding these modifications, targeted interventions can potentially be developed to disrupt pathogen-mediated epigenetic changes and restore normal immune responses.

Vaccination strategies: Epigenetic modifications can influence the efficacy of vaccines. Specific DNA methylation patterns can serve as epigenetic markers to predict vaccine responsiveness. DNA methylation patterns at certain gene loci in immune cells can predict the immune response to the influenza vaccine (Liu et al., 2019). This knowledge could be utilized to develop personalized vaccination strategies or interventions to enhance vaccine efficacy.

Epigenetic regulation of antimicrobial peptides: Antimicrobial peptides (AMPs) are pivotal in the innate immune response against pathogenic invaders. The expression of AMPs can be influenced by epigenetic modifications, which in turn can have a significant impact on the host's defense mechanisms. Research has revealed that the regulation of expression of cathelicidin, a vital AMP, is controlled by DNA methylation and histone modifications. Manipulating these epigenetic markers could enhance the production of cathelicidin and bolster the host's ability to combat microbial infections (Wei et al., 2019).

Inhibiting pathogen replication: Epigenetic drugs can also target the pathogens directly by inhibiting their replication. For example, DNA methyltransferase inhibitors (DNMTi) have been shown to suppress the replication of the HIV by inducing hypermethylation of the viral promoter (Kauder et al., 2009).

Modulating host−pathogen interactions: Epigenetic modifications can also be used to modulate the interactions between the host and the pathogen. For instance, small noncoding RNA molecules, such as microRNAs (miRNAs), have been shown to regulate the expression of both host and viral genes, thereby influencing the outcome of viral infections (Trobaugh and Klimstra, 2017).

Epigenetic regulation of autophagy: Autophagy, a fundamental cellular process involved in degrading and repurposing cellular components, emerges as a critical factor in mounting an effective immune response against intracellular pathogens. In relation to bacterial infections, recent studies have elucidated the role of epigenetic modifications, particularly histone acetylation, in governing the regulation of autophagy (Mandhair et al., 2021).

Epigenetic control of inflammation: Inflammation serves as an integral part of the host's immune reaction to infections. Nonetheless, excessive inflammation can lead to tissue injury and disease. Recent investigations indicate that epigenetic modifications, particularly histone methylation, are involved in governing the expression of proinflammatory genes during infection scenarios (Chen et al., 2009).

6.2 Potential Therapies Based on Epigenetic Factors Against Infectious Diseases

Investigating the intricate interplay among gene expression, epigenetic modifications, and environmental and infectious agents is pivotal in formulating efficacious approaches to combat both infectious and noninfectious diseases. Recent findings have substantiated the critical involvement of epigenetic modifications in governing gene expression and immune response, thereby rendering them potential targets for combating infectious diseases.

Here are some potential therapies for the treatment and prevention of infectious diseases based on the epigenetic factors:

Histone deacetylase (HDAC) inhibitors: HDAC inhibitors have demonstrated their ability to bolster the immune response against viral infections through the upregulation of antiviral genes. Notably, vorinostat, an HDAC inhibitor, has exhibited its capacity to elevate the expression of ISGs and enhance the eradication of HBV in vitro (Lucera et al., 2014). Similarly, valproic acid, another

HDAC inhibitor, has been shown to augment the immune response to influenza virus infection in mice (Andreu et al., 2020).

DNA methyltransferase (DNMT) inhibitors: DNMT inhibitors can reverse the hypermethylation of immune-related genes that occur during chronic viral infections, thereby restoring their expression and enhancing the immune response. For example, the DNMT inhibitor 5-azacytidine has been shown to increase the expression of ISGs and improve the clearance of HBV in vitro (Hong et al., 2017).

MicroRNA (miRNA) modulators: MicroRNAs (miRNAs) represent a class of noncoding small RNA molecules that govern gene expression by selectively binding to target messenger RNAs (mRNAs). Numerous miRNAs have been implicated in the immune response to viral infections, and modulation of their expression has been proposed as a promising therapeutic approach (Trobaugh and Klimstra, 2017; Bartel, 2009). Specifically, inhibition of miRNA-122 by miravirsen has demonstrated remarkable efficacy in reducing viral load in patients with HCV infection (Janssen et al., 2013).

CRISPR/Cas9-based epigenome editing: CRISPR/Cas9 technology can be used to target specific epigenetic modifications and modulate gene expression. For example, a recent study showed that CRISPR/Cas9-mediated demethylation of the interleukin-2 (IL-2) gene promoter increased IL-2 production and improved clearance of chronic viral infections in mice (Jung et al., 2018; Dimitri et al., 2022).

6.3 Informing Public Health Policy Through Epigenetic Research

Epigenetic research exhibits significant potential for shaping public health policies by identifying environmental determinants capable of modulating gene expression and the epigenome, thereby heightening the risk of developing diseases (Rozek et al., 2014). As an illustration, air pollution exposures during pregnancy has been implicated in DNA methylation pattern changes in infants. Such changes may contribute to increased susceptibility to childhood asthma (Aguilera et al., 2022; Isaevska et al., 2022). These investigations underline the importance of implementing public health interventions aimed at reducing air pollution as a preventive measure against childhood asthma. Some research found that maternal stress during pregnancy was associated with changes in DNA methylation patterns in the placenta, which could affect fetal development and increase the risk of adverse birth outcomes (Monk et al., 2016; Provenzi et al., 2022). This research suggests that reducing maternal stress during pregnancy could be an important public health intervention to improve birth outcomes.

Epigenetic research can also inform public health policy by identifying potential targets for intervention. For example, a study found that site-specific methylation of the placental promotor region is linked with maternal smoking during pregnancy, which deregulates infant cortisol regulation and results in infant neurobehavioral development (Stroud et al., 2016). This research suggests intervention in the health policies associated with maternal care, awareness, and guidance.

In summary, epigenetic research can inform public health policy by identifying environmental factors that can influence gene expression and contribute to disease risk, identifying potential targets for intervention, and suggesting interventions to improve health outcomes.

7. Conclusion and Future Perspectives

In conclusion, epigenetic changes play a crucial role in regulating gene expression during infections. These modifications not only influence host–pathogen interactions but also contribute to disease

progression and treatment responses. The insights gained from studying the epigenetic landscape of infectious diseases have immense implications for developing innovative therapeutic strategies.

Advancements and advantages: I. Progress has been made in unraveling how DNA methylation patterns are altered by infections, providing valuable insights into disease pathology at the molecular level. II. The discovery of histone modifications as essential players in host−pathogen interactions has expanded our understanding of infectious diseases. III. NGS technologies have enabled genome-wide profiling of epigenetic changes associated with infections, facilitating a comprehensive view of the disease process. IV. MicroRNA regulation during infection has unveiled potential diagnostic markers for early detection of certain infectious diseases. V. Epigenetic drugs, such as HDAC inhibitors, exhibit promising efficacy against various types of infections in preclinical models.

Limitations of current knowledge: I. Despite advancements, there are still gaps in our knowledge regarding the precise mechanisms through which epigenetic changes drive disease progression. II. Epigenetic regulation is a complex process involving multiple layers of modifications and interactions, making it challenging to fully understand and manipulate. III. Epigenetic modifications induced by infectious diseases may have transgenerational consequences, but the mechanisms and implications of these changes are still not well understood. IV. Translating laboratory findings into clinically applicable therapies remains challenging due to potential off-target effects and toxicity concerns associated with some epigenetic drugs. V. Global mapping technologies often require large sample sizes and sophisticated analytical tools, limiting accessibility for researchers working on less common or neglected tropical diseases. VI. Interindividual variations in epigenetic modifications pose challenges when developing universally effective therapeutics.

Future research to address these limitations should focus on: I. Further exploring how specific pathogens manipulate host epigenetic machinery to better understand disease pathogenesis. II. Developing more precise technologies capable of detecting subtle yet clinically significant epigenetic changes. III. Expanding studies to encompass a broader range of patients, accounting for interindividual variations in epigenetic modifications. IV. Investigating combination therapies that integrate conventional antimicrobial agents with emerging epigenetics-based drugs. V. Establishing robust computational models that predict disease outcomes based on observed epigenetic alterations. There is also a need for greater collaboration between researchers in different fields, which would allow for a more comprehensive approach to the study of infectious diseases using epigenetic techniques.

In summary, the study of epigenetics in infectious diseases has provided valuable insights into the intricate dynamics between pathogens and their hosts. While challenges persist, continued research efforts and advancements in technology hold tremendous potential for revolutionizing our understanding and management of infectious diseases.

References

Abhimanyu, et al., 2021. Reversing post-infectious epigenetic-mediated immune suppression. Front. Immunol. 12. https://doi.org/10.3389/fimmu.2021.688132.

Aguilera, J., et al., 2022. Increases in ambient air pollutants during pregnancy are linked to increases in methylation of IL4, IL10, and IFNγ. Clin. Epigenet. 14 (1). https://doi.org/10.1186/s13148-022-01254-2.

Al Akeel, R., 2013. Role of epigenetic reprogramming of host genes in bacterial pathogenesis. Saudi J. Biol. Sci. 20 (4), 305−309. https://doi.org/10.1016/j.sjbs.2013.05.003.

Alipoor, S.D., et al., 2019. Serum exosomal miRNAs are associated with active pulmonary tuberculosis. Dis. Markers 2019. https://doi.org/10.1155/2019/1907426.

Andraos, C., et al., 2011. Vitamin D receptor gene methylation is associated with ethnicity, tuberculosis, and TaqI polymorphism. Hum. Immunol. 72 (3), 262−268. https://doi.org/10.1016/j.humimm.2010.12.010.

Andreu, S., et al., 2020. Valproic acid and its amidic derivatives as new antivirals against alphaherpesviruses. Viruses 12 (12). https://doi.org/10.3390/v12121356.

Arbuckle, J.H., et al., 2017. Inhibitors of the histone methyltransferases EZH2/1 induce a potent antiviral state and suppress infection by diverse viral pathogens. mBio 8 (4). https://doi.org/10.1128/mBio.01141-17.

Archin, N.M., Margolis, D.M., 2014. Emerging strategies to deplete the HIV reservoir. Curr. Opin. Infect. Dis. 27 (1), 29−35. https://doi.org/10.1097/QCO.0000000000000026.

Atlante, S., et al., 2020. The epigenetic implication in coronavirus infection and therapy. Clin. Epigenet. 12 (1). https://doi.org/10.1186/s13148-020-00946-x.

Bai, G., et al., 2015. Epigenetic dysregulation of hairy and enhancer of split 4 (HES4) is associated with striatal degeneration in postmortem Huntington brains. Hum. Mol. Genet. 24 (5), 1441−1456. https://doi.org/10.1093/hmg/ddu561.

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136 (2), 215−233. https://doi.org/10.1016/j.cell.2009.01.002.

Barturen, G., et al., 2022. Whole blood DNA methylation analysis reveals respiratory environmental traits involved in COVID-19 severity following SARS-CoV-2 infection. Nat. Commun. 13 (1). https://doi.org/10.1038/s41467-022-32357-2.

Belicard, T., Jareosettasin, P., Sarkies, P., 2018. The piRNA pathway responds to environmental signals to establish intergenerational adaptation to stress. BMC Biol. 16 (1). https://doi.org/10.1186/s12915-018-0571-y.

Berger, S.L., et al., 2009. An operational definition of epigenetics: figure 1. Gene Dev. 23 (7), 781−783. https://doi.org/10.1101/gad.1787609.

Bierne, H., Hamon, M., Cossart, P., 2012. Epigenetics and bacterial infections. Cold Spring Harb Perspect Med 2.

Bird, A., 2007. Perceptions of epigenetics. Nature 447 (7143), 396−398. https://doi.org/10.1038/nature05913.

Birdwell, C.E., et al., 2014. Genome-wide DNA methylation as an epigenetic consequence of Epstein-Barr virus infection of immortalized keratinocytes. J. Virol. 88 (19), 11442−11458. https://doi.org/10.1128/JVI.00972-14.

Bohacek, J., Mansuy, I.M., 2013. Epigenetic inheritance of disease and disease risk. Neuropsychopharmacology 38 (1), 220−236. https://doi.org/10.1038/npp.2012.110.

Bouchat, S., et al., 2012. Histone methyltransferase inhibitors induce HIV-1 recovery in resting CD4+ T cells from HIV-1-infected HAART-treated patients. AIDS 26 (12), 1473−1482. https://doi.org/10.1097/qad.0b013e32835535f5.

Calo, S., et al., 2014. Antifungal drug resistance evoked via RNAi-dependent epimutations. Nature 513 (7519), 555−558. https://doi.org/10.1038/nature13575.

de Candia, P., Matarese, G., 2021. Reimagining an immunological dogma. Nat. Immunol. 22 (11), 1355−1358. https://doi.org/10.1038/s41590-021-01046-5.

Capparelli, R., et al., 2022. Int. J. Mol. Sci. 23.

Carestiato, F.N., et al., 2013. AN upward trend in DNA P16INK4A methylation pattern and high risk HPV infection according to the severity of the cervical lesion. Rev. do Inst. Med. Trop. São Paulo 55 (5), 329−334. https://doi.org/10.1590/s0036-46652013000500006.

Chang, Z., et al., 2019. Epigenetic mechanisms of drug resistance in fungi. Fungal Genet. Biol. 132. https://doi.org/10.1016/j.fgb.2019.103253.

Cheeseman, K., Weitzman, J.B., 2015. Host-parasite interactions: an intimate epigenetic relationship. Cell Microbiol. 17 (8), 1121−1132. https://doi.org/10.1111/cmi.12471.

Chen, X., et al., 2009. The NF-κB factor RelB and histone H3 lysine methyltransferase G9a directly interact to generate epigenetic silencing in endotoxin tolerance. J. Biol. Chem. 284 (41), 27857−27865. https://doi.org/10.1074/jbc.M109.000950.

Cheng, A.S.L., et al., 2013. *Helicobacter pylori* causes epigenetic dysregulation of FOXD3 to promote gastric carcinogenesis. Gastroenterology 144 (1), 122−133.e9. https://doi.org/10.1053/j.gastro.2012.10.002.

Chiara, 2020. Comparison of bisulfite pyrosequencing and methylation-specific qPCR for methylation assessment. Int. J. Mol. Sci. 21.

Cole, J., et al., 2016. The therapeutic potential of epigenetic manipulation during infectious diseases. Pharmacol. Therapeut. 167, 85−99. https://doi.org/10.1016/j.pharmthera.2016.07.013.

Crimi, E., et al., 2020. Clinical epigenetics and multidrug-resistant bacterial infections: host remodelling in critical illness. Epigenetics 15 (10), 1021−1034. https://doi.org/10.1080/15592294.2020.1748918.

Das, A.T., Harwig, A., Berkhout, B., 2011. The HIV-1 Tat protein has a versatile role in activating viral transcription. J. Virol. 85 (18), 9506−9516. https://doi.org/10.1128/JVI.00650-11.

Dimitri, A., Herbst, F., Fraietta, J.A., 2022. Engineering the next-generation of CAR T-cells with CRISPR-Cas9 gene editing. Mol. Cancer 21 (1). https://doi.org/10.1186/s12943-022-01559-z.

Doesschate, 2020. Correction to: two Randomized Controlled Trials of Bacillus Calmette-Guérin Vaccination to reduce absenteeism among health care workers and hospital admission by elderly persons during the COVID-19 pandemic: a structured summary of the study protocols for two randomised controlled trials. Trials 21.

Dominguez-Andres, J., Netea, M.G., 2019. Long-term reprogramming of the innate immune system. J. Leukoc. Biol. 105 (2), 329−338. https://doi.org/10.1002/JLB.MR0318-104R.

Dupont, C., Armant, D.R., Brenner, C.A., 2009. Epigenetics: definition, mechanisms and clinical perspective. Semin. Reprod. Med. 27 (5), 351−357. https://doi.org/10.1055/s-0029-1237423.

Eichten, S.R., Schmitz, R.J., Springer, N.M., 2014. Epigenetics: beyond chromatin modifications and complex genetic regulation. Plant Physiol. 165 (3), 933−947. https://doi.org/10.1104/pp.113.234211.

Eskandarian, H.A., et al., 2013. A role for SIRT2-dependent histone H3K18 deacetylation in bacterial infection. Science 341 (6145). https://doi.org/10.1126/science.1238858.

Esterhuyse, M.M., Kaufmann, S.H.E., 2013. Diagnostic biomarkers are hidden in the infected host's epigenome. Expert Rev. Mol. Diagn. 13 (6), 625−637. https://doi.org/10.1586/14737159.2013.811897.

Fischer, N., 2020. Infection-induced epigenetic changes and their impact on the pathogenesis of diseases. Semin. Immunopathol. 42 (2), 127−130. https://doi.org/10.1007/s00281-020-00793-1.

García-Giménez, J.L., et al., 2017. Epigenetic biomarkers: current strategies and future challenges for their use in the clinical laboratory. Crit. Rev. Clin. Lab Sci. 54 (7−8), 529−550. https://doi.org/10.1080/10408363.2017.1410520.

Gayen, Betal, S., 2022. COVID-19 infection during pregnancy induces differential gene expression in human cord blood cells from term neonates. Front. Pediatr. 10.

Giamarellos-Bourboulis, E.J., et al., 2020. Complex immune dysregulation in COVID-19 patients with severe respiratory failure. Cell Host Microbe 27 (6), 992−1000.e3. https://doi.org/10.1016/j.chom.2020.04.009.

Gibney, E.R., Nolan, C.M., 2010. Epigenetics and gene expression. Heredity 105 (1), 4−13. https://doi.org/10.1038/hdy.2010.54.

Giulino-Roth, L., et al., 2012. Targeted genomic sequencing of pediatric Burkitt lymphoma identifies recurrent alterations in antiapoptotic and chromatin-remodeling genes. Blood 120 (26), 5181−5184. https://doi.org/10.1182/blood-2012-06-437624.

Gliddon, H.D., et al., 2018. Genome-wide host RNA signatures of infectious diseases: discovery and clinical translation. Immunology 153 (2), 171−178. https://doi.org/10.1111/imm.12841.

Gómez-Díaz, E., et al., 2012. Epigenetics of host-pathogen interactions: the road ahead and the road behind. PLoS Pathog. 8 (11). https://doi.org/10.1371/journal.ppat.1003007.

Goyal, N., Prathap, V.M., Prakash, A., 2021. Epigenetics and viral infectious diseases. In: Epigenetics and Metabolomics. Elsevier, India, pp. 353–369. https://doi.org/10.1016/B978-0-323-85652-2.00003-8.

Gutierrez, A., et al., 2020. Evaluation of biochemical and epigenetic measures of peripheral brain-derived neurotrophic factor (BDNF) as a biomarker in huntington's disease patients. Front. Mol. Neurosci. 12. https://doi.org/10.3389/fnmol.2019.00335.

Hãninger, E., Krueger, T.E.G., Lang, J.M., 2015. Augmenting antitumor immune responses with epigenetic modifying agents. Front. Immunol. 6. https://doi.org/10.3389/fimmu.2015.00029.

Habig, M., et al., 2021. Epigenetic modifications affect the rate of spontaneous mutations in a pathogenic fungus. Nat. Commun. 12 (1). https://doi.org/10.1038/s41467-021-26108-y.

Hakimi, M.A., Olias, P., Sibley, L.D., 2017. Toxoplasma effectors targeting host signaling and transcription. Clin. Microbiol. Rev. 30 (3), 615–644. https://doi.org/10.1128/CMR.00005-17.

Hamon, M.A., et al., 2007. Histone modifications induced by a family of bacterial toxins. Proc. Natl. Acad. Sci. USA 104 (33), 13467–13472. https://doi.org/10.1073/pnas.0702729104.

Hardy, T., Mann, D.A., 2016. Epigenetics in liver disease: from biology to therapeutics. Gut 65 (11), 1895–1905. https://doi.org/10.1136/gutjnl-2015-311292.

Hashimoto, K., et al., 2007. Improved quantification of DNA methylation using methylation-sensitive restriction enzymes and real-time PCR. Epigenetics 2 (2), 86–91. https://doi.org/10.4161/epi.2.2.4203.

Hernández López, J., et al., 2014. Trans-generational immune priming in honeybees. Proc. Biol. Sci. 281 (1785). https://doi.org/10.1098/rspb.2014.0454.

Hong, X., Kim, E.S., Guo, H., 2017. Epigenetic regulation of hepatitis B virus covalently closed circular DNA: implications for epigenetic therapy against chronic hepatitis B. Hepatology 66 (6), 2066–2077. https://doi.org/10.1002/hep.29479.

Hunter, B., Hollister, J.D., Bomblies, K., 2012. Epigenetic inheritance: what news for evolution? Curr. Biol. 22 (2), R54–R56. https://doi.org/10.1016/j.cub.2011.11.054.

Irizarry, R.A., et al., 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. 41 (2), 178–186. https://doi.org/10.1038/ng.298.

Isaevska, E., et al., 2022. Prenatal exposure to PM10 and changes in DNA methylation and telomere length in cord blood. Environ. Res. 209. https://doi.org/10.1016/j.envres.2022.112717.

Jablonka, E., 2004. Epigenetic epidemiology. Int. J. Epidemiol. 33 (5), 929–935. https://doi.org/10.1093/ije/dyh231.

Jablonka, E., Lamb, M.J., 1998. Epigenetic inheritance in evolution. J. Evol. Biol. 11 (2), 159–183. https://doi.org/10.1046/j.1420-9101.1998.11020159.x.

Jablonka, E., Raz, G., 2009. Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. Q. Rev. Biol. 84 (2), 131–176. https://doi.org/10.1086/598822.

Janssen, H.L.A., et al., 2013. Treatment of HCV infection by targeting microRNA. N. Engl. J. Med. 368 (18), 1685–1694. https://doi.org/10.1056/NEJMoa1209026.

Jelinek, J., et al., 2018. Digital restriction enzyme analysis of methylation (DREAM). In: Methods in Molecular Biology. Humana Press Inc., United States, pp. 247–265. https://doi.org/10.1007/978-1-4939-7481-8_13.

Jones, P.A., Taylor, S.M., 1980. Cellular differentiation, cytidine analogs and DNA methylation. Cell 20 (1), 85–93. https://doi.org/10.1016/0092-8674(80)90237-8.

Jung, I.Y., et al., 2018. CRISPR/Cas9-Mediated knockout of DGK improves antitumor activities of human T cells. Cancer Res. 78 (16), 4692–4703. https://doi.org/10.1158/0008-5472.CAN-18-0030.

Kang, G.H., et al., 2002. Epstein-Barr virus-positive gastric carcinoma demonstrates frequent aberrant methylation of multiple genes and constitutes CpG island methylator phenotype-positive gastric carcinoma. Am. J. Pathol. 160 (3), 787–794. https://doi.org/10.1016/S0002-9440(10)64901-2.

Kathirvel, M., Mahadevan, S., 2016. The role of epigenetics in tuberculosis infection. Epigenomics 8 (4), 537–549. https://doi.org/10.2217/epi.16.1.

Katzmarski, N., et al., 2021. Transmission of trained immunity and heterologous resistance to infections across generations. Nat. Immunol. 22 (11), 1382−1390. https://doi.org/10.1038/s41590-021-01052-7.

Kauder, S.E., et al., 2009. Epigenetic regulation of HIV-1 latency by cytosine methylation. PLoS Pathog. 5 (6). https://doi.org/10.1371/journal.ppat.1000495.

Kee, J., et al., 2022. SARS-CoV-2 disrupts host epigenetic regulation via histone mimicry. Nature 610 (7931), 381−388. https://doi.org/10.1038/s41586-022-05282-z.

Kim, Y.Z., 2014. Altered histone modifications in gliomas. Brain Tumor Res. Treat. 2 (1). https://doi.org/10.14791/btrt.2014.2.1.7.

Klosin, A., et al., 2017. Transgenerational transmission of environmental information in *C. elegans*. Science 356 (6335), 320−323. https://doi.org/10.1126/science.aah6412.

Kondilis-Mangum, H.D., Wade, P.A., 2013. Epigenetics and the adaptive immune response. Mol. Aspect. Med. 34 (4), 813−825. https://doi.org/10.1016/j.mam.2012.06.008.

Kuss-Duerkop, S.K., Pyeon, D., Westrich, J.A., 2018. DNA tumor virus regulation of host dna methylation and its implications for immune evasion and oncogenesis. Viruses 10 (2). https://doi.org/10.3390/v10020082.

Lee, S.W., et al., 2016. Gene expression profiling identifies candidate biomarkers for active and latent tuberculosis. BMC Bioinf. 17 (1). https://doi.org/10.1186/s12859-015-0848-x.

Leemkolk, 2022. Quantification of unmethylated insulin DNA using methylation sensitive restriction enzyme digital polymerase Chain reaction. Transpl. Int. 35.

Lio, C.W.J., Rao, A., 2019. TET enzymes and 5hMC in adaptive and innate immune systems. Front. Immunol. 10. https://doi.org/10.3389/fimmu.2019.00210.

Liotti, A., et al., 2022. Epigenetics: an opportunity to shape innate and adaptive immune responses. Immunology 167 (4), 451−470. https://doi.org/10.1111/imm.13571.

Liu, S., et al., 2019. Epigenetic modification is regulated by the interaction of influenza A virus nonstructural protein 1 with the de novo DNA methyltransferase DNMT3B and subsequent transport to the cytoplasm for K48-linked polyubiquitination. J. Virol. 93 (7). https://doi.org/10.1128/jvi.01587-18.

Lucera, M.B., et al., 2014. The histone deacetylase inhibitor vorinostat (SAHA) increases the susceptibility of uninfected CD4+ T cells to HIV by increasing the kinetics and efficiency of postentry viral events. J. Virol. 88 (18), 10803−10812. https://doi.org/10.1128/JVI.00320-14.

Mahan, M.J., et al., 2017. Epigenetic programming by microbial pathogens and impacts on acute and chronic disease. In: Epigenetics and Human Health. Springer Verlag, United States, pp. 89−112. https://doi.org/10.1007/978-3-319-55021-3_5.

Mandhair, H.K., Novak, U., Radpour, R., 2021. Epigenetic regulation of autophagy: a key modification in cancer cells and cancer stem cells. World J. Stem Cell. 13 (6), 542−567. https://doi.org/10.4252/wjsc.v13.i6.542.

Manikkam, M., et al., 2012. Dioxin (TCDD) induces epigenetic transgenerational inheritance of adult onset disease and sperm epimutations. PLoS One 7 (9). https://doi.org/10.1371/journal.pone.0046249.

Margolis, D.M., 2011. Histone deacetylase inhibitors and HIV latency. Curr. Opin. HIV AIDS 6 (1), 25−29. https://doi.org/10.1097/COH.0b013e328341242d.

Marr, A.K., et al., 2014. Leishmania donovani infection causes distinct epigenetic DNA methylation changes in host macrophages. PLoS Pathog. 10 (10). https://doi.org/10.1371/journal.ppat.1004419.

Menachery, V.D., et al., 2018. MERS-CoV and H5N1 influenza virus antagonize antigen presentation by altering the epigenetic landscape. Proc. Natl. Acad. Sci. USA 115 (5). https://doi.org/10.1073/pnas.1706928115.

Minarovits, J., Niller, H.H., 2013. Patho-epigenetics. Med. Epigenet. 1 (1), 37−45. https://doi.org/10.1159/000354284.

Momeni, M., et al., 2023. A comprehensive analysis of gene expression profiling data in COVID-19 patients for discovery of specific and differential blood biomarker signatures. Sci. Rep. 13 (1). https://doi.org/10.1038/s41598-023-32268-2.

Momparler, R.L., Onetto-Pothier, N., Momparler, L.F., 1990. Comparison of antineoplastic activity of cytosine arabinoside and 5-aza-2′-deoxycytidine against human leukemic cells of different phenotype. Leuk. Res. 14 (9), 755−760. https://doi.org/10.1016/0145-2126(90)90068-K.

Monforte, A.d'Arminio, et al., 2000. Insights into the reasons for discontinuation of the first highly active antiretroviral therapy (HAART) regimen in a cohort of antiretroviral naïve patients. AIDS 14 (5), 499−507. https://doi.org/10.1097/00002030-200003310-00005.

Mongelli, A., et al., 2021. Evidence for biological age acceleration and telomere shortening in COVID-19 survivors. Int. J. Mol. Sci. 22 (11). https://doi.org/10.3390/ijms22116151.

Monk, C., et al., 2016. Distress during pregnancy: epigenetic regulation of placenta glucocorticoid-related genes and fetal neurobehavior. Am. J. Psychiatr. 173 (7), 705−713. https://doi.org/10.1176/appi.ajp.2015.15091171.

Moore, L.D., Le, T., Fan, G., 2013. DNA methylation and its basic function. Neuropsychopharmacology 38 (1), 23−38. https://doi.org/10.1038/npp.2012.112.

Moron-Lopez, S., et al., 2021. The genome-wide methylation profile of CD4+ T cells from individuals with human immunodeficiency virus (HIV) identifies distinct patterns associated with disease progression. Clin. Infect. Dis. 72 (9), e256−e264. https://doi.org/10.1093/cid/ciaa1047.

Mueller, A.L., Mcnamara, M.S., Sinclair, D.A., 2020. Why does COVID-19 disproportionately affect older people? Aging 12 (10), 9959−9981. https://doi.org/10.18632/aging.103344.

Muhammad, J.S., et al., 2022. Epigenetic-mediated antimicrobial resistance: host versus pathogen epigenetic alterations. Antibiotics 11 (6). https://doi.org/10.3390/antibiotics11060809.

Mukherjee, K., Twyman, R.M., Vilcinskas, A., 2015. Insects as models to study the epigenetic basis of disease. Prog. Biophys. Mol. Biol. 118 (1−2), 69−78. https://doi.org/10.1016/j.pbiomolbio.2015.02.009.

Nehme, Z., Pasquereau, S., Herbein, G., 2019. Control of viral infections by epigenetic-targeted therapy. Clin. Epigenet. 11 (1). https://doi.org/10.1186/s13148-019-0654-9.

Nestorov, P., Tardat, M., Peters, A.H.F.M., 2013. H3K9/HP1 and Polycomb. Two key epigenetic silencing pathways for gene regulation and embryo development. In: Current Topics in Developmental Biology. Academic Press Inc., Switzerland, pp. 243−291. https://doi.org/10.1016/B978-0-12-416027-9.00008-5.

Netea, M.G., Quintin, J., Van Der Meer, J.W.M., 2011. Trained immunity: a memory for innate host defense. Cell Host Microbe 9 (5), 355−361. https://doi.org/10.1016/j.chom.2011.04.006.

Ng, Y.-Q., et al., 2019. Dietary restriction and epigenetics: Part I. Cond Med. 2, 284−299.

Niller, H.H., Minarovits, J., 2018. Epigenetics and human infectious diseases. In: Epigenetics in Human Disease. Elsevier, Germany, pp. 643−687. https://doi.org/10.1016/B978-0-12-812215-0.00021-2.

Oriol-Tordera, B., et al., 2022. Epigenetic landscape in the kick-and-kill therapeutic vaccine BCN02 clinical trial is associated with antiretroviral treatment interruption (ATI) outcome. EBioMedicine 78. https://doi.org/10.1016/j.ebiom.2022.103956.

Pajares, M.J., et al., 2021. Methods for analysis of specific DNA methylation status. Methods 187, 3−12. https://doi.org/10.1016/j.ymeth.2020.06.021.

Pei, Y., Wong, J.H.Y., Robertson, E.S., 2020. Herpesvirus epigenetic reprogramming and oncogenesis. Ann. Rev. Virol. 7, 309−331. https://doi.org/10.1146/annurev-virology-020420-014025.

Provenzi, L., et al., 2022. Maternal and infant NR3C1 and SLC6A4 epigenetic signatures of the COVID-19 pandemic lockdown: when timing matters. Res. Square. https://doi.org/10.21203/rs.3.rs-1792374/v1.

Ramendra, R., et al., 2021. Glutathione metabolism is a regulator of the acute inflammatory response of monocytes to (1→3)-β-D-glucan. Front. Immunol. 12. https://doi.org/10.3389/fimmu.2021.694152.

Reisenauer, A., et al., 1999. Bacterial DNA methylation: a cell cycle regulator? J. Bacteriol. 181 (17), 5135−5139. https://doi.org/10.1128/jb.181.17.5135-5139.1999.

Ringrose, L., Paro, R., 2004. Epigenetic regulation of cellular memory by the Polycomb and trithorax group proteins. Annu. Rev. Genet. 38 (1), 413−443. https://doi.org/10.1146/annurev.genet.38.072902.091907.

Robert McMaster, W., Morrison, C.J., Kobor, M.S., 2016. Epigenetics: a new model for intracellular parasite-host cell regulation. Trends Parasitol. 32 (7), 515−521. https://doi.org/10.1016/j.pt..2016.04.002.

Rodrigues, L.O.C.P., Graça, R.S.F., Carneiro, L.A.M., 2018. Integrated stress responses to bacterial pathogenesis patterns. Front. Immunol. 9. https://doi.org/10.3389/fimmu.2018.01306.

Rozek, L.S., et al., 2014. Epigenetics: relevance and implications for public health. Annu. Rev. Publ. Health. https://doi.org/10.1146/annurev-publhealth-032013-182513.

Sakamoto, S., Potla, R., Larner, A.C., 2004. Histone deacetylase activity is required to recruit RNA polymerase II to the promoters of selected interferon-stimulated early response genes. J. Biol. Chem. 279 (39), 40362−40367. https://doi.org/10.1074/jbc.M406400200.

Santos, E.J.M.d., et al., 2019. Editorial: immuno-epigenetic markers for infectious diseases. Front. Immunol. 10. https://doi.org/10.3389/fimmu.2019.02719.

Schmeck, B., et al., 2005. Intracellular bacteria differentially regulated endothelial cytokine release by MAPK-dependent histone modification. J. Immunol. 175 (5), 2843−2850. https://doi.org/10.4049/jimmunol.175.5.2843.

Schuettengruber, B., et al., 2007. Genome regulation by Polycomb and trithorax proteins. Cell 128 (4), 735−745. https://doi.org/10.1016/j.cell.2007.02.009.

Scott, R.S., 2017. Epstein−Barr virus: a master epigenetic manipulator. Curr. Opin. Virol. 26, 74−80. https://doi.org/10.1016/j.coviro.2017.07.017.

Shan, J., et al., 2017. Histone demethylase LSD1 restricts influenza A virus infection by erasing IFITM3-K88 monomethylation. PLoS Pathog. 13 (12). https://doi.org/10.1371/journal.ppat.1006773.

Shi, L., Wu, J., 2009. Epigenetic regulation in mammalian preimplantation embryo development. Reprod. Biol. Endocrinol. 7 (1). https://doi.org/10.1186/1477-7827-7-59.

Shvedunova, M., Akhtar, A., 2022. Modulation of cellular processes by histone and non-histone protein acetylation. Nat. Rev. Mol. Cell Biol. 23 (5), 329−349. https://doi.org/10.1038/s41580-021-00441-y.

Sideri, T., et al., 2017. The copper transport-associated protein Ctr4 can form prion-like epigenetic determinants in *Schizosaccharomyces pombe*. Microb. Cell 4 (1), 16−28. https://doi.org/10.15698/mic2017.01.552.

Silva Da, R.,M.L., 2021. The role of HPV-induced epigenetic changes in cervical carcinogenesis (Review). Biomed. Rep. 15.

Sivasudhan, E., et al., 2022. Hepatitis B viral protein HBx and the molecular mechanisms modulating the hallmarks of hepatocellular carcinoma: a comprehensive review. Cells 11 (4). https://doi.org/10.3390/cells11040741.

Skinner, M.K., 2011. Role of epigenetics in developmental biology and transgenerational inheritance. Birth Defects Res. Part C Embryo Today - Rev. 93 (1), 51−55. https://doi.org/10.1002/bdrc.20199.

Sollars, V., et al., 2003. Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. Nat. Genet. 33 (1), 70−74. https://doi.org/10.1038/ng1067.

Stein, C.M., et al., 2021. Methylome-wide analysis reveals epigenetic marks associated with resistance to tuberculosis in human immunodeficiency virus-infected individuals from East africa. JID (J. Infect. Dis.) 224 (4), 695−704. https://doi.org/10.1093/infdis/jiaa785.

Stroud, L.R., et al., 2016. Epigenetic regulation of placental NR3C1: mechanism underlying prenatal programming of infant neurobehavior by maternal smoking? Child Dev. 87 (1), 49−60. https://doi.org/10.1111/cdev.12482.

Suzuki, S., Horinouchi, T., Furusawa, C., 2014. Prediction of antibiotic resistance by gene expression profiles. Nat. Commun. 5 (1). https://doi.org/10.1038/ncomms6792.

Syn, G., Blackwell, J.M., Jamieson, S.E., 2016. Epigenetics in Infectious Diseases. Elsevier BV, pp. 377−400. https://doi.org/10.1016/b978-0-12-801899-6.00019-x.

Tan, S.Y.X., Zhang, J., Tee, W.W., 2022. Epigenetic regulation of inflammatory signaling and inflammation-induced cancer. Front. Cell Dev. Biol. 10. https://doi.org/10.3389/fcell.2022.931493.

Tbeishat, H.A., 2022. Novel in Silico mRNA vaccine design exploiting proteins of *M. tuberculosis* that modulates host immune responses by inducing epigenetic modifications. Sci. Rep. 12.

Thompson, R.P., Nilsson, E., Skinner, M.K., 2020. Environmental epigenetics and epigenetic inheritance in domestic farm animals. Anim. Reprod. Sci. 220. https://doi.org/10.1016/j.anireprosci.2020.106316.

Tidbury, H.J., Pedersen, A.B., Boots, M., 2011. Within and transgenerational immune priming in an insect to a DNA virus. Proc. Biol. Sci. https://doi.org/10.1098/rspb.2010.1517.

Tirnaz, S., Batley, J., 2019. DNA methylation: toward crop disease resistance improvement. Trends Plant Sci. 24 (12), 1137−1150. https://doi.org/10.1016/j.tplants.2019.08.007.

Tolg, C., et al., 2011. Uropathogenic *E. coli* infection provokes epigenetic downregulation of CDKN2A (p16INK4A) in uroepithelial cells. Lab. Invest. 91 (6), 825−836. https://doi.org/10.1038/labinvest.2010.197.

Tribolet, L., et al., 2020. MicroRNA biomarkers for infectious diseases: from basic research to biosensing. Front. Microbiol. 11. https://doi.org/10.3389/fmicb.2020.01197.

Trinh, Q.D., 2022. Recent research in cell stress and microbial infection. Microorganisms 10 (3). https://doi.org/10.3390/microorganisms10030622.

Trobaugh, D.W., Klimstra, W.B., 2017. MicroRNA regulation of RNA virus replication and pathogenesis. Trends Mol. Med. 23 (1), 80−93. https://doi.org/10.1016/j.molmed.2016.11.003.

Turecki, G., Meaney, M.J., 2016. Effects of the social environment and stress on glucocorticoid receptor gene methylation: a systematic review. Biol. Psychiatr. 79 (2), 87−96. https://doi.org/10.1016/j.biopsych.2014.11.022.

Venditti, S., et al., 2020. Molecules of silence: effects of meditation on gene expression and epigenetics. Front. Psychol. 11. https://doi.org/10.3389/fpsyg.2020.01767.

Villalobos-Escobedo, J.M., Herrera-Estrella, A., Carreras-Villaseñor, N., 2016. The interaction of fungi with the environment orchestrated by RNAi. Mycologia 108 (3), 556−571. https://doi.org/10.3852/15-246.

Wei, J.W., et al., 2017. Non-coding RNAs as regulators in epigenetics (Review). Oncol. Rep. 37 (1), 3−9. https://doi.org/10.3892/or.2016.5236.

Wei, R., et al., 2019. PU.1 and epigenetic signals modulate 1,25-dihydroxyvitamin D 3 and C/EBPα regulation of the human cathelicidin antimicrobial peptide gene in lung epithelial cells. J. Cell. Physiol. 234 (7), 10345−10359. https://doi.org/10.1002/jcp.27702.

White, M.K., Pagano, J.S., Khalili, K., 2014. Viruses and human cancers: a long road of discovery of molecular paradigms. Clin. Microbiol. Rev. 27 (3), 463−481. https://doi.org/10.1128/CMR.00124-13.

Yadav, V., 2015. Understanding the host epigenetics in *Mycobacterium tuberculosis* infection. J. Genet. Genome Res. 2 (1). https://doi.org/10.23937/2378-3648/1410016.

Yan, H., et al., 2018. Identifying differentially expressed long non-coding RNAs in PBMCs in response to the infection of multidrug-resistant tuberculosis. Infect. Drug Resist. 11, 945−959. https://doi.org/10.2147/IDR.S154255.

Yim, E.-K., Park, J.-S., 2005. The role of HPV E6 and E7 oncoproteins in HPV-associated cervical carcinogenesis. Cancer Res. Treat. 37 (6). https://doi.org/10.4143/crt.2005.37.6.319.

Yin, F.F., et al., 2016. Serine/threonine kinases 31(STK31) may be a novel cellular target gene for the HPV16 oncogene E7 with potential as a DNA hypomethylation biomarker in cervical cancer. Virol. J 13 (1). https://doi.org/10.1186/s12985-016-0515-5.

Youngblood, B., Hale, J.S., Ahmed, R., 2013. T-cell memory differentiation: insights from transcriptional signatures and epigenetics. Immunology 139 (3), 277−284. https://doi.org/10.1111/imm.12074.

Zheng, X., Sawalha, A.H., 2022. The role of oxidative stress in epigenetic changes underlying autoimmunity. Antioxidants Redox Signal. 36 (7−9), 423−440. https://doi.org/10.1089/ars.2021.0066.

Zune, A.L.D., et al., 2021. Pathogen-induced epigenetic modifications in cancers: implications for prevention, detection and treatment of cancers in africa. Cancers 13 (23). https://doi.org/10.3390/cancers13236051.

CHAPTER

The Evolution of SARS-CoV-2 and Influenza A Virus at the Human–Animal Interface

22

Tavis K. Anderson[1], Rafael A. Medina[2,3,4] and Martha I. Nelson[5]

[1]*Virus and Prion Research Unit, National Animal Disease Center, USDA-ARS, Ames, IA, United States;* [2]*Department of Pathology and Experimental Medicine, School of Medicine, Emory University, Atlanta, GA, United States;* [3]*Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, United States;* [4]*Departamento de Enfermedades Infecciosas e Inmunología Pediátrica, Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile;* [5]*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, United States*

1. Introduction: New Perspectives on the Bidirectional Human–Animal Interface

As the 1918 "Spanish flu" pandemic swept the world, killing more than 20 million people, American veterinarian J.S. Koen noticed pigs with suspiciously similar flu-like respiratory symptoms (Koen, 1919). Koen wondered if the pigs and people were infected with the same organism and, if so, whether pigs had transmitted the disease to people (*zoonosis*) or people had infected the pigs (*reverse zoonosis*). Over the next century, technological advances in genetic sequencing confirmed that the pigs and people in 1918 were infected with similar influenza A viruses of the H1N1 subtype (Shope, 1931), prompting questions about the human–animal interface. *How frequently do humans and animals exchange pathogens? Is the transfer bidirectional? Or does animal-to-human transmission occur more frequently than human-to-animal?* Rapidly evolving RNA viruses were found to be particularly adept at host-jumping and represent the bulk of emerging infectious diseases in humans, raising another suite of questions (Taylor et al., 2001). *What molecular changes are required for viruses to invade new host cells that differ in acidity, surface receptors, and even routes of transmission? What evolutionary processes, such as recombination, help viruses make evolutionary leaps to infect new host species?*

Influenza is considered a model system for understanding pathogen spillover (Plowright et al., 2017) because of the wealth of historical influenza data available from humans, wild birds, pigs, horses, and poultry—including genomes resurrected from the 1918 pandemic virus (Taubenberger et al., 2005)—that reveal the virus's continuous shapeshifting and host-switching (Kawaoka et al., 1989; Smith et al., 2009b; Vijaykrishna et al., 2011; Webster et al., 1992; Wilson et al., 1965; Worobey et al., 2014). While public health research understandably focuses on the risk of animal pathogens infecting people, the next section describes how our perspective on the human–animal interface has changed in light of new genomic data. It turns out humans actually transmit more flu to pigs than the

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00016-1

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

reverse, turning swine into breeding grounds for new chimeric viruses with genetic components from humans, birds, and pigs (Anderson et al., 2021; Nelson and Vincent, 2015). Eventually, the new swine viruses spill back from pigs to humans, with notable examples including the H1N1 "swine flu" pandemic in 2009 (Garten et al., 2009) and zoonotic outbreak of H3N2v in 2012 (Jhung et al., 2013). Understanding the molecular basis for why humans and pigs share so many flu viruses, including similarities in their flu-binding sialic acid receptors (Nicholls et al., 2008; Wasik et al., 2017), is discussed in the second section of this chapter. Which molecular changes help avian viruses adapt to mammals (Herfst et al., 2012; Imai et al., 2012; Medina and García-Sastre, 2011; Moncorgé et al., 2013) is the focus of the third section.

During the COVID-19 pandemic, lessons from influenza helped to contextualize the large number of observed SARS-CoV-2 spillovers (and spillback) (Sooksawasdi Na Ayudhya and Kuiken, 2021). Frequent reverse zoonosis of SARS-CoV-2 viruses from humans to animals (Garigliany et al., 2020; Grome et al., 2022; McAloose et al., 2020; Sit et al., 2020) echoed the events of the 2009 H1N1 flu pandemic, when humans also infected their pets, zoo animals, poultry, and livestock (Boedeker et al., 2017; Crossley et al., 2012; Lin et al., 2012). Early in 2020, SARS-CoV-2-related deaths in zoo animals alerted veterinarians that infected humans present a risk to diverse wildlife species. Reverse zoonosis became an even bigger concern when European mink farms were infected with SARS-CoV-2. Mink-to-mink transmission led to new variants with mutations in the spike protein that quickly spilled back to mink farmers (Lu et al., 2021). Alarmed public health officials responded with a decisive, but controversial nationwide mink culling.

The real shockwave arrived during the summer of 2021 when high SARS-CoV-2 infection rates were observed in wild North American white-tailed deer across the eastern and central United States (Chandler et al., 2021; Hale et al., 2022). Stunned public health officials could not answer basic questions about the routes by which the virus transmitted from humans to deer or how the virus spread deer-to-deer. Up to this point, the COVID-19 pandemic had been a proving ground for "genomic epidemiology" and the power of high-throughput genetic sequencing and data analysis to trace pathogen evolution and transmission with unprecedented precision and inform public health responses in real time (Bedford et al., 2020; du Plessis et al., 2021; Lemieux et al., 2021). But of the 15 million SARS-CoV-2 genomes published on GISAID by 2023, less than 0.1% were from non-human hosts. Many wildlife infections were potentially going undetected. It was unknown whether deer sustain SARS-CoV-2 transmission on their own or are merely a sentinel that is convenient to sample via hunters or wildlife managers conducting population control or surveillance for chronic wasting disease. It is possible deer are just one node within a broader transmission network spanning multiple wildlife species. The black box of SARS-CoV-2 transmission and interspecies spillover is just beginning to be uncovered, as described in the fourth section of this chapter.

In the wake of the COVID-19 pandemic, there has never been stronger universal recognition of the need for more studies of viral spillover at the human−animal interface to combat the growing threat of emerging zoonotic diseases. The clock is ticking. The next pandemic could arrive in 10 years, 20 years, or 6 months. A decade of advances in high-throughput genetic sequencing, bioinformatics, and experimental risk assessment provide new tools for preventing virus spillover (COVID-19 Genomics UK (COG-UK) 2020; Gardy and Loman, 2018). While influenza and SARS-CoV-2 are particularly well-characterized disease systems for spillover, and therefore the focus of this chapter, new virus discovery surveys are uncovering previously unknown pathogens with host-jumping capabilities at an

unprecedented pace (Lednicky et al., 2021). New human–animal disease systems will likely be featured in the future editions of this textbook.

2. A Tangled History of Influenza A Virus Evolution in Humans and Swine

For more than a century, pigs and people have passed influenza viruses back and forth. In most cases, humans were infecting the pigs, but in 2009 a new H1N1 influenza virus (H1N1pdm) jumped from pigs to people, causing the first pandemic of the 21st century. The virus was the descendant of the 1918 pandemic virus after it continued to circulate in pigs, before leaping back to people in 2009. Next-generation genetic sequencing and new "phylodynamic" methods of data analysis reconstructed a vivid picture of how the H1N1pdm virus had evolved in pigs (Mena et al., 2016). Over many decades, the virus hopped back and forth between birds, pigs, and people, shuffling its genome through "reassortment" to create the three-part chimera that caused the global H1N1 pandemic in humans in 2009 (Garten et al., 2009; Smith et al., 2009b). Continued monitoring of the new virus revealed frequent reverse zoonosis, with hundreds of spillover events from humans to pigs and other animals (Nelson et al., 2012; Markin et al., 2023). As genetic sequencing improves, it is uncovering a "black box" of hidden viral spillover that occurs routinely between animals and humans in both directions (Anderson et al., 2021). With it, scientists are gaining further appreciation for the evolutionary plasticity of RNA viruses like influenza A and the molecular changes that allow viruses to invade new host cells that differ in acidity, surface receptors, and even routes of transmission.

Influenza viruses in humans often find their way into pigs and other species. During the 2009 influenza pandemic, humans transmitted H1N1pdm viruses to a variety of pets, zoo animals, and wildlife. Zookeepers infected sloth bears at a US zoo, and somehow human-like H1N1pdm viruses wound up in free-ranging California sea lions (Boyce et al., 2013). Most H1N1pdm transmissions from humans to animals resulted in dead-end infections that did not continue to transmit animal-to-animal. Pigs, however, were the exception. The continued transmission of H1N1pdm in farmed pigs provided opportunities for the virus to evolve in new ways all over the world. One rapid form of virus evolution via gene shuffling is termed genomic reassortment, which occurred in pigs when the new H1N1pdm virus and an endemic swine influenza virus coinfected a single animal and swapped genes during virus replication to create new hybrid offspring with vastly different genomic arrangements (Watson et al., 2015; Gao et al., 2017; Henritzi et al., 2020). New genomic arrangements that emerged in pigs created problems not just for pigs (Neveau et al., 2022), but for people, as the variants eventually transmitted to humans. In the United States, pigs bred for competition at agricultural fairs harbor the same cocktail of genetic variants found in the larger commercial swine population but amplified in a unique setting (Nelson et al., 2016). This becomes a problem because agricultural fair pigs are handled daily by susceptible children, providing ample opportunity for zoonotic transmission. In 2012, a major outbreak broke out in youth swine exhibitors of "variant" influenza A viruses, predominantly in Ohio and Indiana (Wong et al., 2012; Jhung et al., 2013; Bowman et al., 2014). Hundreds of kids were infected, many were hospitalized, and one adult fair attendee died. Fortunately, no human-to-human transmission chains were sustained and a global pandemic was averted thanks to coordinated efforts by public health, local physicians, and the swine agricultural community. But the outbreak provided a case study for how virus evolution within the interconnected populations of commercial swine and show pigs has created ideal conditions for spillover (McBride et al., 2021).

The influenza viruses in swine reflect genes of human seasonal and avian influenza strains followed by decades of evolution in pigs. As most genes are derived from human strains, they have maintained mammalian adaptation properties. Yet, the evolution and divergence in the swine host from currently or recently circulating influenza in humans has resulted in an assemblage of influenza to which humans lack immunity, presenting a continual pandemic threat (Markin et al., 2023). And the problem is perpetuated with near constant recharging of the virus diversity "battery" with new spillovers from humans (e.g., Sharma et al., 2022; Zeller et al., 2022). With each of the spillovers, nearly 500 detected between 2010 and 2023, a new opportunity is provided for the virus to antigenically drift and/or reassort with endemic swine viruses. Given that human infection by swine influenza is possible and has been documented more than 700 times (Chen et al., 2022), if these viruses transmit in humans, the next limiting factor to preventing efficient human transmission and progression of an outbreak to a pandemic is existing population immunity. And, there is moderate to little immunity in human populations to many of the viruses circulating in swine (Souza et al., 2022; Venkatesh et al., 2022). Consequently, reducing the diversity of viruses circulating in swine is important as with every spillover, there is an increased likelihood that one of those viruses has the capacity to be the single propagule that initiates the next pandemic.

The goal to reduce influenza diversity in swine is challenged by heterogeneity in vaccination programs, vaccine efficacy, and the rapid movement of pigs across the landscape. Generally, the vaccines that are used do not completely block infection under typical field situations (Vincent et al., 2016, 2020). Passive or previously acquired immunity against similar viruses tends to reduce clinical signs but infection and subsequent transmission within densely housed animals remains possible (Allerson et al., 2013; Li et al., 2022). Additionally, prior immunity from infection, vaccination, or maternal antibodies shapes the immunological landscape: the active immune response to a vaccine may be muted or skewed whereby enhanced disease is possible (e.g., Rajão et al., 2016), and protection from circulating strains can be minimal (e.g., Everett et al., 2021). These fundamental immunological issues are further confounded by determining what viruses or proteins to include as components in a vaccine. Currently, no single viral antigen can induce sufficiently broad immunity to all diversity (Markin et al., 2023), and incorporating all known naturally circulating variants into a single multivalent vaccine is not feasible (Anderson et al., 2012). Epidemiological analyses have demonstrated at least 30 clades across the H1N1, H1N2, and H3N2 subtypes (Anderson et al., 2021; Arendsee et al., 2021), and constant reassortment between endemic viruses appears to have driven rapid evolution of the surface proteins (Henritzi et al., 2020; Zeller et al., 2021; Hufnagel et al., 2023). Currently, there are multiple research avenues being explored to improve animal health: vaccine platforms have been developed to incorporate custom antigens (e.g., Vander Veen et al., 2012); the impact of maternal antibodies may be obviated through including additional vaccine antigens targets (Wymore Brand et al., 2022; Kitikoon et al., 2023); and veterinarians and farmers are pivoting to make custom vaccines that reflect viruses detected in on-farm surveillance. Though this is an improvement and many of the new vaccine approaches work adequately in controlled experimental systems, modern swine agriculture has unique conditions that change transmission dynamics, and intensified swine production frequently moves pigs and their viruses across regions (Nelson et al., 2011, 2015). A consequence is that unique viruses are introduced into naive host populations, reassort, and some of the viruses that emerge from the "battle royale" have genetic components that enhance transmission and are antigenically novel, further reducing vaccine efficacy (Campbell et al., 2014; Neveau et al., 2022; Thomas et al., 2023). At this point, preemptively controlling influenza in pigs to minimize the chances of

interspecies transmission appears unlikely; there are constant influenza detections across the year with predictable seasonal epidemics distinct from the human flu season and the detection frequency of flu in pigs has not changed for more than 10 years (Arendsee et al., 2021). Consequently, in an effort to prepare for zoonotic infections, public and animal health agencies are now focused on developing strategies that would minimize the impact of a swine influenza spillover (Cox et al., 2014; TIPRA, n.d.).

Biannually at the WHO Information Meeting on the Composition of Influenza Virus Vaccines, animal influenza activity data are presented concurrently with human seasonal influenza activity data. If an animal influenza clade has been associated with recent human cases, is detected at high levels, and it is significantly drifted from previously recommended prepandemic candidate vaccine viruses (CVV) and current human seasonal vaccine strains, a new CVV may be generated (O'Neill and Donis, 2009; Anderson et al., 2021). Though these CVVs are effective against the viruses and clades that they are derived from, the degree to which they provide protection against other groups of viruses is uncertain. Further, as swine are surveilled for influenza, additional unique groups of human-origin viruses are identified that are evolving in swine and drifting from CVVs and human population immunity (Cheung et al., 2023; Deng et al., 2020; Tochetto et al., 2023; Zeller et al., 2023). All of the newly characterized viruses retain genetic components that are related to human-seasonal influenza, and each new clade has a different evolutionary trajectory away from the seeding human influenza lineage driven by agricultural practices and reassortment (Encinas et al., 2022). Though more genomic surveillance is likely to further complicate pandemic preparedness efforts, they must continue as the data can elucidate spatial and temporal patterns of virus diversity (Anderson et al., 2013; Vincent et al., 2014), determine whether circulating viruses retain the capacity for human transmission (Pearce et al., 2012; Kimble et al., 2022; Souza et al., 2023; X. Sun et al., 2023), and isolates can be used to assess whether human population immunity will prevent or limit the impact of a swine-origin influenza. As resources are always limited, surveillance should be prioritized to geographic areas with high levels of diversity, production practices that support virus movement and transmission, as well as at interfaces that promote greater contact between pigs and people.

The human−swine interface and the problem it represents is not going away. US youth swine exhibitors and those with close swine contact continue to be infected each year with new genetic flu variants belonging to the H1N1, H1N2, and H3N2 subtypes that circulate in pigs. In certain locations, such as in the agricultural shows in the US Midwest (e.g., Rambo-Martin et al., 2020; VanInsberghe et al., 2023), swine are closely monitored. Veterinarians surveil the pigs and local physicians look for abnormal clusters of flu cases in children that occur during summer fairs and outside the typical winter flu season. In contrast, agricultural fairs in the southern United States are held during winter months to avoid the heat, where variant flu cases that arise in swine exhibitors are difficult to distinguish clinically from the standard human seasonal flu viruses that infect kids in much larger numbers. It is exceedingly difficult to capture rare zoonotic spillovers of acute, transient infections causing normal flu-like illness. For every zoonotic infection among swine exhibitors and those with close swine contacts that is detected, hundreds—possibly thousands—are missed. The lower number of variant flu cases in swine exhibitors in southern states compared to northern states probably does not reflect real geographical differences in zoonotic risk, but rather the difficulty in southern states of finding the signal in the noise during prime flu season. This problem is likely worse in areas that are currently growing their public health infrastructure to reflect the emergence of new tools in modern genomic surveillance (Ladner and Sahl, 2023). Indeed, there is a clear linear

relationship between the burden of flu in people and the number of spillovers into pigs (Markin et al., 2023); if the reciprocal were to be true, then our understanding of influenza evolution in both hosts will be dramatically impacted. The bidirectional exchange of viruses between people and pigs has established a dynamic evolutionary landscape, each spillover spikes the genetic pool with new genes that initiate a surge of evolution: resolving which of these viruses have zoonotic potential is a critical and unresolved puzzle.

3. Viral Adaptations Required for Influenza A Virus Host Switches

Host switching events arise from spillover infections of a new host. Influenza A viruses (IAV) have demonstrated great plasticity in their ability to infect and adapt to diverse host species (Medina and García-Sastre, 2011; Long et al., 2019), including a wide variety of avian and mammalian hosts. Spillover events typically occur in nature among sympatric animal species, e.g., populations that share the same habitat, and these host jumps can occasionally lead to zoonotic events, resulting in the infection of humans at the animal−human interface. However, infections of humans in direct contact with infected animal species are rare (e.g., H5N1 virus−infected birds), highlighting the role of host factors in limiting influenza virus infection across species boundaries (Kuiken et al., 2006). The occurrence of the 2009 H1N1 pandemic and prior pandemics remind us of the ability of certain viruses to overcome these host-restriction barriers. As such, many host adaptive mutations have been identified for IAVs; however, some of them are required to overcome infection and transmission bottlenecks, including species-specific tropism, replication, and host restriction factors that can modulate viral infection in different species.

A major function of the IAV hemagglutinin (HA) glycoprotein is to recognize sialic acid receptors which mediate viral entry into the host cells. Mammalian adapted HAs, such as those from seasonal influenza H1N1 and H3N2, bind primarily to receptors containing a-2,6 sialic acids (SA) located in epithelial cells of the human upper respiratory tract (Stevens et al., 2006a; Chandrasekaran et al., 2008), whereas avian adapted viruses bind predominantly to receptors containing a-2,3 SAs found on epithelial cells in the intestine of birds (Matrosovich et al., 1999). Importantly, a-2,3 SA receptors can be found in cells of the lower respiratory tract of humans (Shinya et al., 2006; van Riel et al., 2006), allowing the potential for avian viruses to infect the human respiratory tract by aerosol transmission (e.g., aerosolized virus particles that reach deep into the lung alveoli). In fact, the sporadic human H5Nx and H7Nx cases observed are likely to arise from events in which high viral loads are inhaled by exposed individuals. Additionally, a-2,3 SA and a-2,6 SA receptors are also present in cells of the respiratory tract of pigs, providing the possibility for mammalian and avian viruses to reassort and generate novel IAV genotypes, some of which can result in potential epidemic or pandemic strains (Smith et al., 2009a).

Data arising from numerous surveillance and experimental studies with pandemic and zoonotic viruses have allowed to define residue changes in the HA that modulate the specificity and affinity for its receptor (Table 22.1). Changing in amino acid residues at position 190 and 225 of HA of the 1918 H1N1 influenza pandemic, from the human adapted virus to the consensus avian virus sequences, mutations D190E and D225G (H3 numbering), were found to modulate its receptor binding specificity from a-2,6 SA to a-2,3 SA (Stevens et al., 2006b; Srinivasan et al., 2008), and to impede transmission of the virus in ferrets (Tumpey et al., 2007). Conversely, residue

Table 22.1 Mammalian Adaptations of Influenza A Virus

Protein		HA			PB2			
Function		Receptor Binding, Fusion to the Host Cell Membrane and Antigenic Properties			Viral Replication			
Virus		Sialic Acid Linkage Specificity	Residues Involved in Binding Specificity Switch	Multibasic Cleavage Site	Mammalian Adaptations			
					D701N	E627K	Q591R	Q591K
Pandemic Strains	1918 H1N1	α2,3 and α2,6	E190D and G225D	No		X		
	1957 H2N2	α2,6	Q226L and N186K or Q226L, G228S	No		X		
	1968 H3N2	α2,6	Q226L, G228S	No		X		
	2009 H1N1	α2,6 and α2,3	E190D and G225D	No			X	
HPAIV	H5Nx	α2,3	N158D/K, N186K, N224K, Q226L, G228S	Yes	X	X		X
	H7Nx	α2,3	Q226L	Yes	X	X		

changes in HPAI H5N1 viruses from G225D and E190D modulate its avidity to a-2,3 SAs, but do not confer it specificity to the a-2,6 SA receptor (Stevens et al., 2006b). However, residue changes Q226L together with N224K resulted in a complete switch from a-2,3 SA to a-2,6 SA receptor binding of an H5 HA, indicating that various residues in or near the receptor binding domain play a role in host receptor adaptation (Herfst et al., 2012; Imai et al., 2012). Similarly, the combination of residue changes Q226L and G228S changed the receptor preference of a recently characterized seal H10N7 virus from a-2,3 SA to a-2,6 SA and render it transmissible by aerosols (Herfst et al., 2020). Of interest, a recent human-isolated avian H3N8 virus was also found to contain the G228S HA mutation (H. Sun et al., 2023), underscoring the importance of receptor binding on host transmission. Of interest, while the 2009 H1N1 pandemic virus binds preferentially to a-2,6 SA, some virus isolates that contained the D225G mutation had increased binding to a-2,3 SA receptors (Childs et al., 2009; Soundararajan et al., 2009; Chen et al., 2011); and therefore had dual receptor activity and the capability to infect cells of the upper and lower respiratory tracts (Maines et al., 2009; Munster et al., 2009).

Apart from HA, the virus polymerase complex, particularly the PB2 subunit, is implicated in host range determination and viral fitness during replication. Specific amino acid residues in PB2 have been associated with increased transmission in mammals. The virus's ability to replicate in the human host is influenced by interactions between viral proteins and host factors, as well as the host's immune responses (Long et al., 2019). It has been shown that the viral polymerase requires the acquisition of one or more host adaptations that provide enhanced replication efficiency (Table 22.1). The most notable, and possibly important mutation for viral mammalian adaptation has been the identification of the PB2 protein 627K amino acid change (Subbarao et al., 1993). This mutation allows the virus to replicate efficiently at 33°C in the upper respiratory tract (Hatta et al., 2007). This mutation has been found in human HPAI H5N1 isolates, in an HPAI H7N7 virus isolate from a fatal case (de Wit et al., 2010) and a recent H3N8 avian influenza human case (H. Sun et al., 2023). Two additional residues in PB2, amino acids 701 (Gabriel et al., 2005; de Jong et al., 2006) and 591 have also been implicated in the mammalian adaptation. The D701N mutation together with a mutation on NP (N319K) have been shown to enhance nuclear import of the viral ribonucleoprotein (vRNP) by binding to human importin-α1 (Sediri et al., 2015; Tarendeau et al., 2007). Residue change Q591R/K of PB2 has been identified as a compensatory mutation for the lack of lysine at position 627, enabling efficient viral replication in mammals. During the 2009 H1N1 pandemic it was established that the 591R mutation could confer efficient replication in mammals and that it compensated for the lack of the 627K and 701N adaptations (Mehle and Doudna, 2009; Yamada et al., 2010). The positive charge associated with PB2-627K or 591R/K is thought to disrupt an interaction with an inhibitory host protein (Yamada et al., 2010). Recently, the 591K was isolated from an HPAI H5N1 human case in Chile (CDC, 2023) and from South American sea lions in Peru and Chile (Leguia et al., 2023; Pardo-Roa et al., 2023), in an ongoing epizootic of these viruses in the Americas, thus highlighting the continuous evolution and host adaptation of IAV in nature.

The complex interplay between viral factors and host restriction factors determines the host range, transmissibility, and pathogenicity of influenza A viruses. Modifications in the polymerase proteins and HA enable the effective binding, entry, and replication of the virus in pertinent human respiratory tract cells. Understanding viral adaptations leading to host switch events is essential for developing effective surveillance strategies to monitor and mitigate the emergence of zoonotic influenza viruses.

4. SARS-CoV-2 Evolution at the Human–Animal Interface

In February 2020, two pet dogs—a Pomeranian and a German shepherd—were infected by their owners with the SARS-CoV-2 virus (Sit et al., 2020), alerting the scientific community to the possibility that the virus could transmit from humans to their pets. Later that year, SARS-CoV-2 outbreaks in farmed mink in Denmark and the Netherlands revealed that the virus could transmit from humans to mustelids, transmit animal-to-animal, and accrue new mutations that could transmit from mink back to their human handlers (Hammer et al., 2021; Larsen et al., 2021; Lu et al., 2021). Large-scale culls of mink were conducted in both countries to mitigate the threat of spillback of mutated viruses to humans. Hong Kong authorities also took an aggressive approach when hamsters imported into the island from the Netherlands arrived infected with human-like strains of the SARS-CoV-2 delta variant (Yen et al., 2022). At the time, Hong Kong was pursuing a zero-COVID-19 strategy, so an unexpected cluster of delta variant cases in humans triggered intensive genetic sleuthing and contact tracing, which led back to a pet store. There, Hong Kong officials traced the source of a new delta outbreak back to hamsters, which were infected with genetically similar viruses. The viruses had unusual mutations in the spike protein, which is recognized by antibodies, which prompted an aggressive public health response. Thousands of hamsters sold from the infected pet store were rounded up across the city and euthanized. For farmed mink and pet hamsters and dogs, there is a clear human–animal interface characterized by high human–animal contact rates and it is easy to reconstruct how a SARS-CoV-2 virus could transmit from humans to these animals. The detection of human-origin SARS-CoV-2 viruses in wild North American white-tailed deer is a different story. The absence of direct human–animal contact exposes significant gaps in our understanding of how acute respiratory viruses transmit at high levels between humans and wildlife, either directly or through the environment.

In 2022, Canadian researchers published the first report of SARS-CoV-2 spillback from a white-tailed deer to a human (Pickering et al., 2022). Leading up to that event, researchers at the US Department of Agriculture and several academic institutions had been tracking high SARS-CoV-2 infection rates in North American white-tailed deer (Chandler et al., 2021; Hale et al., 2022; Kuchipudi et al., 2022; Marques et al., 2022). Several studies found active SARS-CoV-2 infection by PCR test in more than one-third of the sampled wild deer. Moreover, the virus appeared to be spreading deer-to-deer, across ranges spanning ~50–100 square miles. Plus, the SARS-CoV-2 virus appeared to be evolving in deer along a trajectory that differed from humans—and three times faster (McBride et al., 2023). For that reason, the deer-like virus found in the Ontario man was substantially genetically drifted compared to the SARS-CoV-2 viruses circulating in humans at that time, forming a cluster of outliers on the phylogenetic tree along with other viruses collected from deer in the area (Pickering et al., 2022). Although deer-to-human transmission appears to be rare, there is concern that SARS-CoV-2 will be maintained in North American white-tailed deer, evolve to become phenotypically distinct, and spill back to humans, livestock, or other wildlife. White-tailed deer could potentially serve as a reservoir for SARS-CoV-2 lineages that go extinct in humans (Caserta et al., 2023; McBride et al., 2023), just as swine have done for human-origin influenza A viruses (see Section 2). Older SARS-CoV-2 lineages have also been detected in European mink farms, raising questions about whether mink could also serve as reservoirs for SARS-CoV-2 lineages that go extinct in humans (Domańska-Blicharz et al., 2023). In July 2023, the first report of SARS-CoV-2 detection by antibodies in Eurasian wild fallow deer was published, raising additional questions about the potential for

SARS-CoV-2 to establish in deer in Europe (Purves et al., 2023). Going forward, there is a need to resolve basic questions about the human−animal interface for SARS-CoV-2, including the mystery of how human SARS-CoV-2 viruses transmit so frequently from humans to free-ranging white-tailed deer in North America, despite the infrequency of direct contact.

Transmission of SARS-CoV-2 in mink and deer provides opportunities to examine the adaptive changes viruses undergo following host-jumps. Convergent evolution has been observed, as certain mutations occurred repeatedly in deer following the host-switch from humans, such as L18F in the spike protein, which also appeared in the Hong Kong hamster outbreak (McBride et al., 2023; Yen et al., 2022). But no mutation was found universally among all deer viruses, indicating that no single molecular change is required for a human SARS-CoV virus to transmit in deer. Experimental studies of deer viruses confirm that the SARS-CoV-2 viruses found in deer differ little in phenotype (e.g., replication efficiency or infectivity) compared to human viruses (McBride et al., 2023). The conclusion is that human-like viruses are fit enough to replicate and transmit in deer without undergoing major evolutionary changes. Still, the evolutionary rate of SARS-CoV-2 (substitutions per site per year) was three-times higher in deer compared to humans, suggesting that the virus evolves differently in deer, for reasons that are still being explored. Most substitutions in deer were synonymous and did not change the amino acid encoded. As a result, the ratio of nonsynonymous to synonymous substitutions (dN/dS) was lower in deer than in humans, where building global population immunity after years of human circulation has increased selective pressure for mutations in epitopes located in the spike protein. What drives the higher rate of SARS-CoV-2 evolution in deer, if not adaptive changes needed after a host shift or immune selection? Mutational biases were observed among the synonymous substitutions, with elevated rates of C-to-T substitutions. High C-to-T substitution rates observed in human pathogens are thought to result from APOBEC activity, which is a host defense mechanism found in many diverse host species, including some ungulates (Harris and Liddament, 2004). Dramatic shifts in C-to-T substitution rates across branches of the phylogenetic tree can provide an indication of when host-shifts occur between species with differing levels of APOBEC activity (O'Toole et al., 2023). As the SARS-CoV-2 virus continues to evolve in humans, it is possible that the virus is becoming more human adapted and less capable of infecting other species. Evading high immune selection pressure in humans could entail evolutionary trade-offs that reduce fitness in other hosts. Early evidence suggests that the omicron variant does not infect North American white-tailed deer as readily as prior variants such as delta and alpha (Vandegrift et al., 2022).

Going forward, large-scale opportunistic genomic sequencing will be helpful in exposing the high frequency of virus spillover between humans and animals and refine our understanding of its bidirectional nature. However, answering deeper questions about the human−animal interface requires more costly experiments and targeted surveillance. Experimental SARS-CoV-2 infections in live deer have been performed to confirm airborne transmission, measure pathogenesis, and explore why deer are such good hosts for the virus and how viruses transmit so frequently from humans to deer. However, experiments with large cervids are costly and logistically complicated, and therefore limited (Cool et al., 2022; Martins et al., 2022). While deer are relatively easy to sample through hunter-harvested donations, large-scale sampling of other wildlife is more costly and can be difficult to justify when many wildlife surveys yield no SARS-CoV-2-positive samples. SARS-CoV-2 has been described as a "generalist" virus that appears to be capable of infecting a wide range of host species (MacLean et al., 2021), in part due to the ubiquity and conservation of the ACE2 receptors the virus binds for cellular entry (Damas et al., 2020). However, in contrast to influenza, SARS-CoV-2 does not

appear to be a problem for livestock, as of November 2023. Moreover, most serosurveys of wildlife (except white-tailed deer) find low evidence of SARS-CoV-2 across many species in North America, Europe, or Asia (Apaa et al., 2023; Bourret et al., 2022; Despres et al., 2023; Ehrlich et al., 2023; Greenhorn et al., 2022; Miot et al., 2022). Understanding the determinants of SARS-CoV-2's host range remains a central biological question going forward. An important lesson of the SARS-CoV-2 pandemic is that we have barely scratched the surface of virus diversity circulating in animals; therefore, studies at the human−animal interface are likely to yield more surprises.

5. Future Perspectives

As of this writing, a new clade (2.3.4.4.b) of highly pathogenic avian H5N1 influenza A viruses is sweeping the globe at an unprecedented pace, causing mass mortality events in sea birds, poultry, farmed mink, and marine mammals, from Finland to Peru and Chile (Kandeil et al., 2023; Leguia et al., 2023; Lindh et al., 2023). Human activities continue to drive pathogen evolution at an unprecedented pace. Loss of habitat crowds wildlife into smaller patches and into urban areas and farmlands that creates opportunities for spillover between wildlife, humans, and livestock. Over the years, influenza researchers have gained critical insights into the mutations that help an avian influenza virus successfully infect, replicate, and transmit in a mammalian host. Some of these concerning mutations, including D701N in the polymerase complex, have already emerged during the current H5N1 outbreak in seals, sea lions, and humans (Pardo-Roa et al., 2023). Intermediary mammalian hosts are thought to play a key role in the evolutionary jump of avian influenza viruses to humans (Webster et al., 1992), just as swine did during the 2009 H1N1 pandemic. Fortunately, the 2.3.4.4.b virus has not been detected in commercial or feral pigs as of November 2023. However, a different mammalian species could serve as the intermediary host, such as a marine mammals or mustelid. An outbreak of 2.3.4.4.b on a mink farm in Spain raised particular concern because mink have human-like distributions of sialic acid receptors in their respiratory tracts and could serve as an intermediary host between wildlife and humans (Agüero et al., 2023). It has also been suggested that flu pandemics of avian origin in 1957 and 1968 could have transmitted directly from birds to humans, without the need for an intermediary host (Nelson and Worobey, 2018). Will H5N1 viruses cause a global pandemic? Some indicators suggest not. There have been few human H5N1 cases during the current 2021−23 outbreak. Moreover, H5N1 viruses have caused major poultry outbreaks for decades without ever establishing transmission in humans. Still, there are alarming trends to the current H5N1 outbreak, including frequent spillover to diverse mammalian species and the emergence of mutations that help avian viruses replicate in mammalian cells that require careful monitoring.

How can lessons from the COVID-19 pandemic prepare us for future pandemics, or possibly prevent them? One legacy of the COVID-19 pandemic is that it fostered unprecedented levels of teamwork and cooperation among disparate research groups that previously were in competition. There are beautiful examples of how disease forecasters created centralized hubs that combined models generated independently by 10−20 academic and government research teams to improve predictions of future COVID-19 activity (Reich et al., 2022). Evolutionary biologists created similar research hubs to combine different computational approaches developed by multiple teams to improve predictions of future SARS-CoV-2 variants from emerging pools of diversity in real time (DeGrace et al., 2022). Interdisciplinary pipelines were developed to feed the variants selected by computational biologists into experimental animal models for phenotypic characterization and potentially vaccine

strain updates. Herding scientists is like herding cats, but major public health emergencies can galvanize cooperation. However, cooperation between human and animal health faces additional challenges, since the priorities of animal health and public health agencies do not always align. Pathogens of greatest concern for human health are not necessarily the ones that sicken animals. During the 2009 influenza and COVID-19 pandemics, public health officials sometimes enacted indiscriminate animal culls out of an abundance of caution that took a major economic toll on farmers. Countries that discover pathogens transmitting silently in their livestock risk or decide to vaccinate poultry farms for avian influenza losing their export markets. As a result, farmers can be incentivized to keep pathogen data private or not test at all. Retrieving viral samples from free-ranging wildlife does not encounter the same economic liabilities, but the work is laborious and lower disease prevalence means that surveillance in wildlife sometimes comes up empty, discouraging future efforts. Researchers studying SARS-CoV-2 or influenza at the human-animal interface must navigate conflicting priorities between public health and farmers' economic viability. The term "One Health" was coined decades ago to synergize the efforts of human, animal, and environmental health, given that pathogens do not respect these boundaries (Cook et al., 2004). The better scientists document the free flow of pathogens between humans and animals in well characterized disease systems like SARS-CoV-2 and influenza, the more One Health is validated as a concept. But efforts to put One Health into practice still must contend with siloed government agencies that espouse different, sometimes conflicting, priorities.

Going forward, it is crucial to recognize that much of what is currently known about the human−animal interface for pathogens like influenza and SARS-CoV-2 is shaped by opportunistic sampling. Wildlife surveillance is even lower outside of North America and Europe, and the recent detection of SARS-CoV-2 in wild coatis in Brazil illustrates how wide our surveillance net ought to be (Stoffella-Dutra et al., 2023). The hosts where we have data are not necessarily the ones to be most concerned about. The blind spots are what keep scientists up at night.

Acknowledgments

This work was supported by the Centers of Excellence for Influenza Research and Response, National Institute of Allergy and Infectious Diseases, National Institutes of Health (NIH), Department of Health and Human Services, under contracts 75N93021C00015, 75N93021C00017, and 75N93021C00014. This work was supported by USDA-ARS and the Intramural Research Program of the US National Library of Medicine at the NIH. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA or NIH. USDA is an equal opportunity provider and employer.

References

Agüero, M., Monne, I., Sánchez, A., Zecchin, B., Fusaro, A., Ruano, M.J., Del Valle Arrojo, M., Fernández-Antonio, R., Souto, A.M., Tordable, P., Cañás, J., Bonfante, F., Giussani, E., Terregino, C., Orejas, J.J., 2023. Highly pathogenic avian influenza A(H5N1) virus infection in farmed minks, Spain, October 2022. Euro Surveill. 28. https://doi.org/10.2807/1560-7917.ES.2023.28.3.2300001.

Allerson, M., Deen, J., Detmer, S.E., Gramer, M.R., Joo, H.S., Romagosa, A., Torremorell, M., 2013. The impact of maternally derived immunity on influenza A virus transmission in neonatal pig populations. Vaccine 31, 500−505.

Anderson, T.K., Chang, J., Arendsee, Z.W., Venkatesh, D., Souza, C.K., Kimble, J.B., Lewis, N.S., Davis, C.T., Vincent, A.L., 2021. Swine influenza a viruses and the tangled relationship with humans. Cold Spring Harb. Perspect. Med. 11, 1−25.

Anderson, T.K., Laegreid, W.W., Cerutti, F., Osorio, F.A., Nelson, E.A., Christopher-Hennings, J., Goldberg, T.L., 2012. Ranking viruses: measures of positional importance within networks define core viruses for rational polyvalent vaccine development. Bioinformatics 28, 1624−1632.

Anderson, T.K., Nelson, M.I., Kitikoon, P., Swenson, S.L., Korslund, J.A., Vincent, A.L., 2013. Population dynamics of cocirculating swine influenza A viruses in the United States from 2009 to 2012. Influenza Other Respi. Viruses 7 (Suppl. 4), 42−51.

Apaa, T., Withers, A.J., Mackenzie, L., Staley, C., Dessi, N., Blanchard, A., Bennett, M., Bremner-Harrison, S., Chadwick, E.A., Hailer, F., Harrison, S.W.R., Lambin, X., Loose, M., Mathews, F., Tarlinton, R., 2023. Lack of detection of SARS-CoV-2 in British wildlife 2020-21 and first description of a stoat (Mustela erminea) Minacovirus. J. Gen. Virol. 104. https://doi.org/10.1099/jgv.0.001917.

Arendsee, Z.W., Chang, J., Hufnagel, D.E., Markin, A., Janas-Martindale, A., Vincent, A.L., Anderson, T.K., 2021. octoFLUshow: an interactive tool describing spatial and temporal trends in the genetic diversity of influenza A virus in U.S. swine. Microbiol. Resour. Announc. 10, e0108121.

Bedford, T., Greninger, A.L., Roychoudhury, P., Starita, L.M., Famulare, M., Huang, M.L., Nalla, A., Pepper, G., Reinhardt, A., Xie, H., Shrestha, L., Nguyen, T.N., Adler, A., Brandstetter, E., Cho, S., Giroux, D., Han, P.D., Fay, K., Frazar, C.D., Ilcisin, M., Lacombe, K., Lee, J., Kiavand, A., Richardson, M., Sibley, T.R., Truong, M., Wolf, C.R., Nickerson, D.A., Rieder, M.J., Englund, J.A., Hadfield, J., Hodcroft, E.B., Huddleston, J., Moncla, L.H., Müller, N.F., Neher, R.A., Deng, X., Gu, W., Federman, S., Chiu, C., Duchin, J.S., Gautom, R., Melly, G., Hiatt, B., Dykema, P., Lindquist, S., Queen, K., Tao, Y., Uehara, A., Tong, S., MacCannell, D., Armstrong, G.L., Baird, G.S., Chu, H.Y., Shendure, J., Jerome, K.R., 2020. Cryptic transmission of SARS-CoV-2 in Washington state. Science 370, 571−575.

Boedeker, N.C., Nelson, M.I., Killian, M.L., Torchetti, M.K., Barthel, T., Murray, S., 2017. Pandemic (H1N1) 2009 influenza A virus infection associated with respiratory signs in sloth bears (Melursus ursinus). Zoonoses Public Health 64. https://doi.org/10.1111/zph.12370.

Bourret, V., Dutra, L., Alburkat, H., Mäki, S., Lintunen, E., Wasniewski, M., Kant, R., Grzybek, M., Venkat, V., Asad, H., Pradel, J., Bouilloud, M., Leirs, H., Colombo, V.C., Sluydts, V., Stuart, P., McManus, A., Eccard, J.A., Firozpoor, J., Imholt, C., Nowicka, J., Goll, A., Ranc, N., Castel, G., Charbonnel, N., Sironen, T., 2022. Serological surveillance for wild rodent infection with SARS-CoV-2 in Europe. bioRxiv. https://doi.org/10.1101/2022.08.02.502439.

Bowman, A.S., Nelson, S.W., Page, S.L., Nolting, J.M., Killian, M.L., Sreevatsan, S., Slemons, R.D., 2014. Swine-to-human transmission of influenza A(H3N2) virus at agricultural fairs, Ohio, USA, 2012. Emerg. Infect. Dis. 20, 1472−1480.

Boyce, W.M., Mena, I., Yochem, P.K., Gulland, F.M., García-Sastre, A., Moreno, N., Perez, D.R., Gonzalez-Reiche, A.S., Stewart, B.S., 2013. Influenza A(H1N1)pdm09 virus infection in marine mammals in California. Emerg. Microb. Infect. 2, e40.

Campbell, P.J., Danzy, S., Kyriakis, C.S., Deymier, M.J., Lowen, A.C., Steel, J., 2014. The M segment of the 2009 pandemic influenza virus confers increased neuraminidase activity, filamentous morphology, and efficient contact transmissibility to A/Puerto Rico/8/1934-based reassortant viruses. J. Virol. 88, 3802−3814.

Caserta, L.C., Martins, M., Butt, S.L., Hollingshead, N.A., Covaleda, L.M., Ahmed, S., Everts, M.R.R., Schuler, K.L., Diel, D.G., 2023. White-tailed deer (Odocoileus virginianus) may serve as a wildlife reservoir for nearly extinct SARS-CoV-2 variants of concern. Proc. Natl. Acad. Sci. U. S. A. 120, e2215067120.

CDC, 2023. Human Infection with Highly Pathogenic Avian Influenza A(H5N1) Virus in Chile. Centers for Disease Control and Prevention. https://www.cdc.gov/flu/avianflu/spotlights/2022-2023/chile-first-case-h5n1-addendum.htm. (Accessed 12 December 2023).

Chandler, J.C., Bevins, S.N., Ellis, J.W., Linder, T.J., Tell, R.M., Jenkins-Moore, M., Root, J.J., Lenoch, J.B., Robbe-Austerman, S., DeLiberto, T.J., Gidlewski, T., Kim Torchetti, M., Shriner, S.A., 2021. SARS-CoV-2 exposure in wild white-tailed deer (Odocoileus virginianus). Proc. Natl. Acad. Sci. U. S. A. 118. https://doi.org/10.1073/pnas.2114828118.

Chandrasekaran, A., Srinivasan, A., Raman, R., Viswanathan, K., Raguram, S., Tumpey, T.M., Sasisekharan, V., Sasisekharan, R., 2008. Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. Nat. Biotechnol. 26, 107−113.

Chen, L.-M., Rivailler, P., Hossain, J., Carney, P., Balish, A., Perry, I., Davis, C.T., Garten, R., Shu, B., Xu, X., Klimov, A., Paulson, J.C., Cox, N.J., Swenson, S., Stevens, J., Vincent, A., Gramer, M., Donis, R.O., 2011. Receptor specificity of subtype H1 influenza A viruses isolated from swine and humans in the United States. Virology 412, 401−410.

Chen, X., Wang, W., Qin, Y., Zou, J., Yu, H., 2022. Global epidemiology of human infections with variant influenza viruses, 1959-2021: a descriptive study. Clin. Infect. Dis. 75, 1315−1323.

Cheung, J., Bui, A.N., Younas, S., Edwards, K.M., Nguyen, H.Q., Pham, N.T., Bui, V.N., Peiris, M., Dhanasekaran, V., 2023. Long-term epidemiology and evolution of swine influenza viruses. Vietnam. Emerg. Infect. Dis. 29, 1397−1406.

Childs, R.A., Palma, A.S., Wharton, S., Matrosovich, T., Liu, Y., Chai, W., Campanero-Rhodes, M.A., Zhang, Y., Eickmann, M., Kiso, M., Hay, A., Matrosovich, M., Feizi, T., 2009. Receptor-binding specificity of pandemic influenza A (H1N1) 2009 virus determined by carbohydrate microarray. Nat. Biotechnol. 27, 797−799.

Cook, R., Karesh, W., Osofsky, S., 2004. One World, One Health: Building Interdisciplinary Bridges to Health in a Globalized World. Wildlife Conservation Society, Bronx, NY, USA.

Cool, K., Gaudreault, N.N., Morozov, I., Trujillo, J.D., Meekins, D.A., McDowell, C., Carossino, M., Bold, D., Mitzel, D., Kwon, T., Balaraman, V., Madden, D.W., Artiaga, B.L., Pogranichniy, R.M., Roman-Sosa, G., Henningson, J., Wilson, W.C., Balasuriya, U.B.R., García-Sastre, A., Richt, J.A., 2022. Infection and transmission of ancestral SARS-CoV-2 and its alpha variant in pregnant white-tailed deer. Emerg. Microb. Infect. 11, 95−112.

COVID-19 Genomics UK (COG-UK) consortiumcontact@cogconsortium.uk, 2020. An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe 1, e99−e100.

Cox, N.J., Trock, S.C., Burke, S.A., 2014. Pandemic preparedness and the influenza risk assessment tool (IRAT). Curr. Top. Microbiol. Immunol. 385, 119−136.

Crossley, B., Hietala, S., Hunt, T., Benjamin, G., Martinez, M., Darnell, D., Rubrum, A., Webby, R., 2012. Pandemic (H1N1) 2009 in captive cheetah. Emerg. Infect. Dis. 18, 315−317.

Damas, J., Hughes, G.M., Keough, K.C., Painter, C.A., Persky, N.S., Corbo, M., Hiller, M., Koepfli, K.-P., Pfenning, A.R., Zhao, H., Genereux, D.P., Swofford, R., Pollard, K.S., Ryder, O.A., Nweeia, M.T., Lindblad-Toh, K., Teeling, E.C., Karlsson, E.K., Lewin, H.A., 2020. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. Proc. Natl. Acad. Sci. U. S. A. 117, 22311−22322.

DeGrace, M.M., Ghedin, E., Frieman, M.B., Krammer, F., Grifoni, A., Alisoltani, A., Alter, G., Amara, R.R., Baric, R.S., Barouch, D.H., Bloom, J.D., Bloyet, L.-M., Bonenfant, G., Boon, A.C.M., Boritz, E.A., Bratt, D.L., Bricker, T.L., Brown, L., Buchser, W.J., Carreño, J.M., Cohen-Lavi, L., Darling, T.L., Davis-Gardner, M.E., Dearlove, B.L., Di, H., Dittmann, M., Doria-Rose, N.A., Douek, D.C., Drosten, C., Edara, V.-V., Ellebedy, A., Fabrizio, T.P., Ferrari, G., Fischer, W.M., Florence, W.C., Fouchier, R.A.M., Franks, J., García-Sastre, A., Godzik, A., Gonzalez-Reiche, A.S., Gordon, A., Haagmans, B.L., Halfmann, P.J., Ho, D.D., Holbrook, M.R., Huang, Y., James, S.L., Jaroszewski, L., Jeevan, T., Johnson, R.M., Jones, T.C., Joshi, A., Kawaoka, Y., Kercher, L., Koopmans, M.P.G., Korber, B., Koren, E., Koup, R.A., LeGresley, E.B., Lemieux, J.E., Liebeskind, M.J., Liu, Z., Livingston, B., Logue, J.P., Luo, Y., McDermott, A.B., McElrath, M.J., Meliopoulos, V.A., Menachery, V.D., Montefiori, D.C., Mühlemann, B., Munster, V.J., Munt, J.E., Nair, M.S., Netzl, A., Niewiadomska, A.M., O'Dell, S., Pekosz, A., Perlman, S.,

Pontelli, M.C., Rockx, B., Rolland, M., Rothlauf, P.W., Sacharen, S., Scheuermann, R.H., Schmidt, S.D., Schotsaert, M., Schultz-Cherry, S., Seder, R.A., Sedova, M., Sette, A., Shabman, R.S., Shen, X., Shi, P.-Y., Shukla, M., Simon, V., Stumpf, S., Sullivan, N.J., Thackray, L.B., Theiler, J., Thomas, P.G., Trifkovic, S., Türeli, S., Turner, S.A., Vakaki, M.A., van Bakel, H., VanBlargan, L.A., Vincent, L.R., Wallace, Z.S., Wang, L., Wang, M., Wang, P., Wang, W., Weaver, S.C., Webby, R.J., Weiss, C.D., Wentworth, D.E., Weston, S.M., Whelan, S.P.J., Whitener, B.M., Wilks, S.H., Xie, X., Ying, B., Yoon, H., Zhou, B., Hertz, T., Smith, D.J., Diamond, M.S., Post, D.J., Suthar, M.S., 2022. Defining the risk of SARS-CoV-2 variants on immune protection. Nature 605, 640−652.

de Jong, M.D., Simmons, C.P., Thanh, T.T., Hien, V.M., Smith, G.J.D., Chau, T.N.B., Hoang, D.M., Chau, N.V.V., Khanh, T.H., Dong, V.C., Qui, P.T., Van Cam, B., Ha, D.Q., Guan, Y., Peiris, J.S.M., Chinh, N.T., Hien, T.T., Farrar, J., 2006. Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. Nat. Med. 12, 1203−1207.

Deng, Y.-M., Wong, F.Y.K., Spirason, N., Kaye, M., Beazley, R., Grau, M.L.L., Shan, S., Stevens, V., Subbarao, K., Sullivan, S., Barr, I.G., Dhanasekaran, V., 2020. Locally acquired human infection with swine-origin influenza A(H3N2) variant virus, Australia, 2018. Emerg. Infect. Dis. 26, 143−147.

Despres, H.W., Mills, M.G., Schmidt, M.M., Gov, J., Perez, Y., Jindrich, M., Crawford, A.M.L., Kohl, W.T., Rosenblatt, E., Kubinski, H.C., Simmons, B.C., Nippes, M.C., Goldenberg, A.J., Murtha, K.E., Nicoloro, S., Harris, M.J., Feeley, A.C., Gelinas, T.K., Cronin, M.K., Frederick, R.S., Thomas, M., Johnson, M.E., Murphy, J., Lenzini, E.B., Carr Jr., P.A., Berger, D.H., Mehta, S.P., Floreani, C.J., Koval, A.C., Young, A.L., Fish, J.H., Wallace, J., Chaney, E., Ushay, G., Ross, R.S., Vostal, E.M., Thisner, M.C., Gonet, K.E., Deane, O.C., Pelletiere, K.R., Rockafeller, V.C., Waterman, M., Barry, T.W., Goering, C.C., Shipman, S.D., Shiers, A.C., Reilly, C.E., Duff, A.M., Shirley, D.J., Jerome, K.R., Pérez-Osorio, A.C., Greninger, A.L., Fortin, N., Mosher, B.A., Bruce, E.A., 2023. Surveillance of Vermont wildlife in 2021-2022 reveals no detected SARS-CoV-2 viral RNA. bioRxiv. https://doi.org/10.1101/2023.04.25.538264.

de Wit, E., Munster, V.J., van Riel, D., Beyer, W.E.P., Rimmelzwaan, G.F., Kuiken, T., Osterhaus, A.D.M.E., Fouchier, R.A.M., 2010. Molecular determinants of adaptation of highly pathogenic avian influenza H7N7 viruses to efficient replication in the human host. J. Virol. 84, 1597−1606.

Domańska-Blicharz, K., Oude Munnink, B.B., Orłowska, A., Smreczak, M., Opolska, J., Lisowska, A., Trebas, P., Socha, W., Giza, A., Bomba, A., Iwan, E., Rola, J., Koopmans, M., 2023. Cryptic SARS-CoV-2 lineage identified on two mink farms as a possible result of long-term undetected circulation in an unknown animal reservoir, Poland, November 2022 to January 2023. Euro Surveill. 28. https://doi.org/10.2807/1560-7917.ES.2023.28.16.2300188.

du Plessis, L., McCrone, J.T., Zarebski, A.E., Hill, V., Ruis, C., Gutierrez, B., Raghwani, J., Ashworth, J., Colquhoun, R., Connor, T.R., Faria, N.R., Jackson, B., Loman, N.J., O'Toole, Á., Nicholls, S.M., Parag, K.V., Scher, E., Vasylyeva, T.I., Volz, E.M., Watts, A., Bogoch, I.I., Khan, K., , COVID-19 Genomics UK (COG-UK) Consortium, Aanensen, D.M., Kraemer, M.U.G., Rambaut, A., Pybus, O.G., 2021. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science 371, 708−712.

Ehrlich, M., Madden, C., McBride, D.S., Nolting, J.M., Huey, D., Kenney, S., Wang, Q., Saif, L.J., Vlasova, A., Dennis, P., Lombardi, D., Gibson, S., McLaine, A., Lauterbach, S., Yaxley, P., Winston, J.A., Diaz-Campos, D., Pesapane, R., Flint, M., Flint, J., Junge, R., Faith, S.A., Bowman, A.S., Hale, V.L., 2023. Lack of SARS-CoV-2 viral RNA detection among a convenience sampling of Ohio wildlife, companion, and agricultural animals, 2020-2021. Animals (Basel) 13. https://doi.org/10.3390/ani13162554.

Encinas, P., Del Real, G., Dutta, J., Khan, Z., van Bakel, H., Del Burgo, M.Á.M., García-Sastre, A., Nelson, M.I., 2022. Evolution of influenza A virus in intensive and free-range swine farms in Spain. Virus Evol. 7, veab099.

Everett, H.E., van Diemen, P.M., Aramouni, M., Ramsay, A., Coward, V.J., Pavot, V., Canini, L., Holzer, B., Morgan, S., , Dynamics sLoLa Consortium, Woolhouse, M.E.J., Tchilian, E., Brookes, S.M., Brown, I.H., Charleston, B., Gilbert, S., 2021. Vaccines that reduce viral shedding do not prevent transmission of H1N1

pandemic 2009 swine influenza A virus infection to unvaccinated pigs. J. Virol. 95. https://doi.org/10.1128/JVI.01787-20.

Gabriel, G., Dauber, B., Wolff, T., Planz, O., Klenk, H.-D., Stech, J., 2005. The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. Proc. Natl. Acad. Sci. U. S. A. 102, 18590−18595.

Gao, S., Anderson, T.K., Walia, R.R., Dorman, K.S., Janas-Martindale, A., Vincent, A.L., 2017. The genomic evolution of H1 influenza A viruses from swine detected in the United States between 2009 and 2016. J. Gen. Virol. 98, 2001−2010.

Gardy, J.L., Loman, N.J., 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat. Rev. Genet. https://doi.org/10.1038/nrg.2017.88.

Garigliany, M., Van Laere, A.-S., Clercx, C., Giet, D., Escriou, N., Huon, C., van der Werf, S., Eloit, M., Desmecht, D., 2020. SARS-CoV-2 natural transmission from human to cat, Belgium, March 2020. Emerg. Infect. Dis. 26, 3069−3071.

Garten, R.J., Davis, C.T., Russell, C.a., Shu, B., Lindstrom, S., Balish, A., Sessions, W.M., Xu, X., Skepner, E., Deyde, V., Okomo-Adhiambo, M., Gubareva, L., Barnes, J., Smith, C.B., Emery, S.L., Hillman, M.J., Rivailler, P., Smagala, J., de Graaf, M., Burke, D.F., Fouchier, R.a.M., Pappas, C., Alpuche-Aranda, C.M., López-Gatell, H., Olivera, H., López, I., Myers, C.a., Faix, D., Blair, P.J., Yu, C., Keene, K.M., Dotson, P.D., Boxrud, D., Sambol, A.R., Abid, S.H., St George, K., Bannerman, T., Moore, A.L., Stringer, D.J., Blevins, P., Demmler-Harrison, G.J., Ginsberg, M., Kriner, P., Waterman, S., Smole, S., Guevara, H.F., Belongia, E.a., Clark, P.a., Beatrice, S.T., Donis, R., Katz, J., Finelli, L., Bridges, C.B., Shaw, M., Jernigan, D.B., Uyeki, T.M., Smith, D.J., Klimov, A.I., Cox, N.J., 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science 325, 197−201.

Greenhorn, J.E., Kotwa, J.D., Bowman, J., Bruce, L., Buchanan, T., Buck, P.A., Davy, C.M., Dibernardo, A., Flockhart, L., Gagnier, M., Hou, A., Jardine, C.M., Lair, S., Lindsay, L.R., Massé, A., Muchaal, P.K., Nituch, L.A., Sotto, A., Stevens, B., Yip, L., Mubareka, S., 2022. SARS-CoV-2 wildlife surveillance in Ontario and Québec. Can. Comm. Dis. Rep. 48, 243−251.

Grome, H.N., Meyer, B., Read, E., Buchanan, M., Cushing, A., Sawatzki, K., Levinson, K.J., Thomas, L.S., Perry, Z., Uehara, A., Tao, Y., Queen, K., Tong, S., Ghai, R., Fill, M.-M., Jones, T.F., Schaffner, W., Dunn, J., 2022. SARS-CoV-2 outbreak among Malayan tigers and humans, Tennessee, USA, 2020. Emerg. Infect. Dis. 28, 833−836.

Hale, V.L., Dennis, P.M., McBride, D.S., Nolting, J.M., Madden, C., Huey, D., Ehrlich, M., Grieser, J., Winston, J., Lombardi, D., Gibson, S., Saif, L., Killian, M.L., Lantz, K., Tell, R.M., Torchetti, M., Robbe-Austerman, S., Nelson, M.I., Faith, S.A., Bowman, A.S., 2022. SARS-CoV-2 infection in free-ranging white-tailed deer. Nature 602, 481−486.

Hammer, A.S., Quaade, M.L., Rasmussen, T.B., Fonager, J., Rasmussen, M., Mundbjerg, K., Lohse, L., Strandbygaard, B., Jørgensen, C.S., Alfaro-Núñez, A., Rosenstierne, M.W., Boklund, A., Halasa, T., Fomsgaard, A., Belsham, G.J., Bøtner, A., 2021. SARS-CoV-2 transmission between mink (Neovison vison) and humans, Denmark. Emerg. Infect. Dis. 27, 547−551.

Harris, R.S., Liddament, M.T., 2004. Retroviral restriction by APOBEC proteins. Nat. Rev. Immunol. 4, 868−877.

Hatta, M., Hatta, Y., Kim, J.H., Watanabe, S., Shinya, K., Nguyen, T., Lien, P.S., Le, Q.M., Kawaoka, Y., 2007. Growth of H5N1 influenza A viruses in the upper respiratory tracts of mice. PLoS Pathog. 3, 1374−1379.

Henritzi, D., Petric, P.P., Lewis, N.S., Graaf, A., Pessia, A., Starick, E., Breithaupt, A., Strebelow, G., Luttermann, C., Parker, L.M.K., Schröder, C., Hammerschmidt, B., Herrler, G., Beilage, E.G., Stadlbauer, D., Simon, V., Krammer, F., Wacheck, S., Pesch, S., Schwemmle, M., Beer, M., Harder, T.C., 2020. Surveillance of European domestic pig populations identifies an emerging reservoir of potentially zoonotic swine influenza A viruses. Cell Host Microbe 28, 614−627.e6.

Herfst, S., Schrauwen, E.J.A., Linster, M., Chutinimitkul, S., de Wit, E., Munster, V.J., Sorrell, E.M., Bestebroer, T.M., Burke, D.F., Smith, D.J., Rimmelzwaan, G.F., Osterhaus, A.D.M.E., Fouchier, R.A.M., 2012. Airborne transmission of influenza A/H5N1 virus between ferrets. Science 336, 1534−1541.

Herfst, S., Zhang, J., Richard, M., McBride, R., Lexmond, P., Bestebroer, T.M., Spronken, M.I.J., de Meulder, D., van den Brand, J.M., Rosu, M.E., Martin, S.R., Gamblin, S.J., Xiong, X., Peng, W., Bodewes, R., van der Vries, E., Osterhaus, A.D.M.E., Paulson, J.C., Skehel, J.J., Fouchier, R.A.M., 2020. Hemagglutinin traits determine transmission of avian A/H10N7 influenza virus between mammals. Cell Host Microbe 28, 602−613.e7.

Hufnagel, D.E., Young, K.M., Arendsee, Z.W., Gay, L.C., Caceres, C.J., Rajão, D.S., Perez, D.R., Vincent Baker, A.L., Anderson, T.K., 2023. Characterizing a century of genetic diversity and contemporary antigenic diversity of N1 neuraminidase in influenza A virus from North American swine. Virus Evol. 9, vead015.

Imai, M., Watanabe, T., Hatta, M., Das, S.C., Ozawa, M., Shinya, K., Zhong, G., Hanson, A., Katsura, H., Watanabe, S., Li, C., Kawakami, E., Yamada, S., Kiso, M., Suzuki, Y., Maher, E.A., Neumann, G., Kawaoka, Y., 2012. Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. Nature 486, 420−428.

Jhung, M.A., Epperson, S., Biggerstaff, M., Allen, D., Balish, A., Barnes, N., Beaudoin, A., Berman, L., Bidol, S., Blanton, L., Blythe, D., Brammer, L., D'Mello, T., Danila, R., Davis, W., de Fijter, S., Diorio, M., Durand, L.O., Emery, S., Fowler, B., Garten, R., Grant, Y., Greenbaum, A., Gubareva, L., Havers, F., Haupt, T., House, J., Ibrahim, S., Jiang, V., Jain, S., Jernigan, D., Kazmierczak, J., Klimov, A., Lindstrom, S., Longenberger, A., Lucas, P., Lynfield, R., McMorrow, M., Moll, M., Morin, C., Ostroff, S., Page, S.L., Park, S.Y., Peters, S., Quinn, C., Reed, C., Richards, S., Scheftel, J., Simwale, O., Shu, B., Soyemi, K., Stauffer, J., Steffens, C., Su, S., Torso, L., Uyeki, T.M., Vetter, S., Villanueva, J., Wong, K.K., Shaw, M., Bresee, J.S., Cox, N., Finelli, L., 2013. Outbreak of variant influenza A(H3N2) virus in the United States. Clin. Infect. Dis. 57, 1703−1712.

Kandeil, A., Patton, C., Jones, J.C., Jeevan, T., Harrington, W.N., Trifkovic, S., Seiler, J.P., Fabrizio, T., Woodard, K., Turner, J.C., Crumpton, J.-C., Miller, L., Rubrum, A., DeBeauchamp, J., Russell, C.J., Govorkova, E.A., Vogel, P., Kim-Torchetti, M., Berhane, Y., Stallknecht, D., Poulson, R., Kercher, L., Webby, R.J., 2023. Rapid evolution of A(H5N1) influenza viruses after intercontinental spread to North America. Nat. Commun. 14, 3082.

Kawaoka, Y., Krauss, S., Webster, R.G., 1989. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. J. Virol. 63, 4603−4608.

Kimble, J.B., Souza, C.K., Anderson, T.K., Arendsee, Z.W., Hufnagel, D.E., Young, K.M., Lewis, N.S., Davis, C.T., Thor, S., Vincent Baker, A.L., 2022. Interspecies transmission from pigs to ferrets of antigenically distinct swine H1 influenza A viruses with reduced reactivity to candidate vaccine virus antisera as measures of relative zoonotic risk. Viruses 14. https://doi.org/10.3390/v14112398.

Kitikoon, P., Knetter, S.M., Mogler, M.A., Morgan, C.L., Hoehn, A., Puttamreddy, S., Strait, E.L., Segers, R.P.A.M., 2023. Quadrivalent neuraminidase RNA particle vaccine protects pigs against homologous and heterologous strains of swine influenza virus infection. Vaccine 41, 6941−6951.

Koen, J.S., 1919. A practical method for field diagnosis of swine diseases. Am. J. Vet. Med. 14, 468−470.

Kuchipudi, S.V., Surendran-Nair, M., Ruden, R.M., Yon, M., Nissly, R.H., Vandegrift, K.J., Nelli, R.K., Li, L., Jayarao, B.M., Maranas, C.D., Levine, N., Willgert, K., Conlan, A.J.K., Olsen, R.J., Davis, J.J., Musser, J.M., Hudson, P.J., Kapur, V., 2022. Multiple spillovers from humans and onward transmission of SARS-CoV-2 in white-tailed deer. Proc. Natl. Acad. Sci. U. S. A. 119. https://doi.org/10.1073/pnas.2121644119.

Kuiken, T., Holmes, E.C., McCauley, J., Rimmelzwaan, G.F., Williams, C.S., Grenfell, B.T., 2006. Host species barriers to influenza virus infections. Science 312, 394−397.

Ladner, J.T., Sahl, J.W., 2023. Towards a post-pandemic future for global pathogen genome sequencing. PLoS Biol. 21, e3002225.

Larsen, H.D., Fonager, J., Lomholt, F.K., Dalby, T., Benedetti, G., Kristensen, B., Urth, T.R., Rasmussen, M., Lassaunière, R., Rasmussen, T.B., Strandbygaard, B., Lohse, L., Chaine, M., Møller, K.L., Berthelsen, A.-S.N., Nørgaard, S.K., Sönksen, U.W., Boklund, A.E., Hammer, A.S., Belsham, G.J., Krause, T.G.,

Mortensen, S., Bøtner, A., Fomsgaard, A., Mølbak, K., 2021. Preliminary report of an outbreak of SARS-CoV-2 in mink and mink farmers associated with community spread, Denmark, June to November 2020. Euro Surveill. 26. https://doi.org/10.2807/1560-7917.ES.2021.26.5.210009.

Lednicky, J.A., Tagliamonte, M.S., White, S.K., Elbadry, M.A., Alam, M.M., Stephenson, C.J., Bonny, T.S., Loeb, J.C., Telisma, T., Chavannes, S., Ostrov, D.A., Mavian, C., Beau De Rochars, V.M., Salemi, M., Morris Jr., J.G., 2021. Independent infections of porcine deltacoronavirus among Haitian children. Nature 600, 133−137.

Leguia, M., Garcia-Glaessner, A., Muñoz-Saavedra, B., Juarez, D., Barrera, P., Calvo-Mac, C., Jara, J., Silva, W., Ploog, K., Amaro, L., Colchao-Claux, P., Johnson, C.K., Uhart, M.M., Nelson, M.I., Lescano, J., 2023. Highly pathogenic avian influenza A (H5N1) in marine mammals and seabirds in Peru. Nat. Commun. 14, 5489.

Lemieux, J.E., Siddle, K.J., Shaw, B.M., Loreth, C., Schaffner, S.F., Gladden-Young, A., Adams, G., Fink, T., Tomkins-Tinch, C.H., Krasilnikova, L.A., DeRuff, K.C., Rudy, M., Bauer, M.R., Lagerborg, K.A., Normandin, E., Chapman, S.B., Reilly, S.K., Anahtar, M.N., Lin, A.E., Carter, A., Myhrvold, C., Kemball, M.E., Chaluvadi, S., Cusick, C., Flowers, K., Neumann, A., Cerrato, F., Farhat, M., Slater, D., Harris, J.B., Branda, J.A., Hooper, D., Gaeta, J.M., Baggett, T.P., O'Connell, J., Gnirke, A., Lieberman, T.D., Philippakis, A., Burns, M., Brown, C.M., Luban, J., Ryan, E.T., Turbett, S.E., LaRocque, R.C., Hanage, W.P., Gallagher, G.R., Madoff, L.C., Smole, S., Pierce, V.M., Rosenberg, E., Sabeti, P.C., Park, D.J., MacInnis, B.L., 2021. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. Science 371. https://doi.org/10.1126/science.abe3261.

Li, C., Culhane, M.R., Schroeder, D.C., Cheeran, M.C.-J., Galina Pantoja, L., Jansen, M.L., Torremorell, M., 2022. Vaccination decreases the risk of influenza A virus reassortment but not genetic variation in pigs. Elife 11. https://doi.org/10.7554/eLife.78618.

Lindh, E., Lounela, H., Ikonen, N., Kantala, T., Savolainen-Kopra, C., Kauppinen, A., Österlund, P., Kareinen, L., Katz, A., Nokireki, T., Jalava, J., London, L., Pitkäpaasi, M., Vuolle, J., Punto-Luoma, A.-L., Kaarto, R., Voutilainen, L., Holopainen, R., Kalin-Mänttäri, L., Laaksonen, T., Kiviranta, H., Pennanen, A., Helve, O., Laamanen, I., Melin, M., Tammiranta, N., Rimhanen-Finne, R., Gadd, T., Salminen, M., 2023. Highly pathogenic avian influenza A(H5N1) virus infection on multiple fur farms in the South and Central Ostrobothnia regions of Finland, July 2023. Euro Surveill. 28. https://doi.org/10.2807/1560-7917.ES.2023.28.31.2300400.

Lin, D., Sun, S., Du, L., Ma, J., Fan, L., Pu, J., Sun, Y., Zhao, J., Sun, H., Liu, J., 2012. Natural and experimental infection of dogs with pandemic H1N1/2009 influenza virus. J. Gen. Virol. 93, 119−123.

Long, J.S., Mistry, B., Haslam, S.M., Barclay, W.S., 2019. Host and viral determinants of influenza A virus species specificity. Nat. Rev. Microbiol. 17, 67−81.

Lu, L., Sikkema, R.S., Velkers, F.C., Nieuwenhuijse, D.F., Fischer, E.A.J., Meijer, P.A., Bouwmeester-Vincken, N., Rietveld, A., Wegdam-Blans, M.C.A., Tolsma, P., Koppelman, M., Smit, L.A.M., Hakze-van der Honing, R.W., van der Poel, W.H.M., van der Spek, A.N., Spierenburg, M.A.H., Molenaar, R.J., Rond, J. de, Augustijn, M., Woolhouse, M., Stegeman, J.A., Lycett, S., Oude Munnink, B.B., Koopmans, M.P.G., 2021. Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in The Netherlands. Nat. Commun. 12, 6802.

MacLean, O.A., Lytras, S., Weaver, S., Singer, J.B., Boni, M.F., Lemey, P., Kosakovsky Pond, S.L., Robertson, D.L., 2021. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. PLoS Biol. 19, e3001115.

Maines, T.R., Jayaraman, A., Belser, J.A., Wadford, D.A., Pappas, C., Zeng, H., Gustin, K.M., Pearce, M.B., Viswanathan, K., Shriver, Z.H., Raman, R., Cox, N.J., Sasisekharan, R., Katz, J.M., Tumpey, T.M., 2009. Transmission and pathogenesis of swine-origin 2009 A(H1N1) influenza viruses in ferrets and mice. Science 325, 484−487.

Markin, A., Ciacci Zanella, G., Arendsee, Z.W., Zhang, J., Krueger, K.M., Gauger, P.C., Vincent Baker, A.L., Anderson, T.K., 2023. Reverse-zoonoses of 2009 H1N1 pandemic influenza A viruses and evolution in United States swine results in viruses with zoonotic potential. PLoS Pathog. 19, e1011476.

Marques, A.D., Sherrill-Mix, S., Everett, J.K., Adhikari, H., Reddy, S., Ellis, J.C., Zeliff, H., Greening, S.S., Cannuscio, C.C., Strelau, K.M., Collman, R.G., Kelly, B.J., Rodino, K.G., Bushman, F.D., Gagne, R.B., Anis, E., 2022. Multiple introductions of SARS-CoV-2 alpha and delta variants into white-tailed deer in Pennsylvania. mBio 13, e0210122.

Martins, M., Boggiatto, P.M., Buckley, A., Cassmann, E.D., Falkenberg, S., Caserta, L.C., Fernandes, M.H.V., Kanipe, C., Lager, K., Palmer, M.V., Diel, D.G., 2022. From Deer-to-Deer: SARS-CoV-2 is efficiently transmitted and presents broad tissue tropism and replication sites in white-tailed deer. PLoS Pathog. 18, e1010197.

Matrosovich, M., Zhou, N., Kawaoka, Y., Webster, R., 1999. The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. J. Virol. 73, 1146−1155.

McAloose, D., Laverack, M., Wang, L., Killian, M.L., Caserta, L.C., Yuan, F., Mitchell, P.K., Queen, K., Mauldin, M.R., Cronk, B.D., Bartlett, S.L., Sykes, J.M., Zec, S., Stokol, T., Ingerman, K., Delaney, M.A., Fredrickson, R., Ivančić, M., Jenkins-Moore, M., Mozingo, K., Franzen, K., Bergeson, N.H., Goodman, L., Wang, H., Fang, Y., Olmstead, C., McCann, C., Thomas, P., Goodrich, E., Elvinger, F., Smith, D.C., Tong, S., Slavinski, S., Calle, P.P., Terio, K., Torchetti, M.K., Diel, D.G., 2020. From people to Panthera: natural SARS-CoV-2 infection in tigers and lions at the Bronx Zoo. mBio 11. https://doi.org/10.1128/mBio.02220-20.

McBride, D.S., Perosfky, A.C., Nolting, J.M., Nelson, M.I., Bowman, A.S., 2021. Tracing the source of influenza A virus zoonoses in interconnected circuits of swine exhibits. J. Infect. Dis. 3, 458−468.

McBride, D.S., Garushyants, S.K., Franks, J., Magee, A.F., Overend, S.H., Huey, D., Williams, A.M., Faith, S.A., Kandeil, A., Trifkovic, S., Miller, L., Jeevan, T., Patel, A., Nolting, J.M., Tonkovich, M.J., Genders, J.T., Montoney, A.J., Kasnyik, K., Linder, T.J., Bevins, S.N., Lenoch, J.B., Chandler, J.C., DeLiberto, T.J., Koonin, E.V., Suchard, M.A., Lemey, P., Webby, R.J., Nelson, M.I., Bowman, A.S., 2023. Accelerated evolution of SARS-CoV-2 in free-ranging white-tailed deer. Nat. Commun. 14, 5105.

Medina, R.A., García-Sastre, A., 2011. Influenza A viruses: new research developments. Nat. Rev. Microbiol. 9, 590−603.

Mehle, A., Doudna, J.A., 2009. Adaptive strategies of the influenza virus polymerase for replication in humans. Proc. Natl. Acad. Sci. U. S. A. 106, 21312−21316.

Mena, I., Nelson, M.I., Quezada-Monroy, F., Dutta, J., Cortes-Fernández, R., Lara-Puente, J.H., Castro-Peralta, F., Cunha, L.F., Trovão, N.S., Lozano-Dubernard, B., Rambaut, A., Van Bakel, H., García-Sastre, A., 2016. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. Elife 5. https://doi.org/10.7554/eLife.16777.

Miot, E.F., Worthington, B.M., Ng, K.H., de Lataillade, L. de G., Pierce, M.P., Liao, Y., Ko, R., Shum, M.H., Cheung, W.Y., Holmes, E.C., Leung, K.S., Zhu, H., Poon, L.L., Peiris, M.J., Guan, Y., Leung, G.M., Wu, J.T., Lam, T.T., 2022. Surveillance of rodent pests for SARS-CoV-2 and other coronaviruses, Hong Kong. Emerg. Infect. Dis. 28, 467−470.

Moncorgé, O., Long, J.S., Cauldwell, A.V., Zhou, H., Lycett, S.J., Barclay, W.S., 2013. Investigation of influenza virus polymerase activity in pig cells. J. Virol. 87, 384−394.

Munster, V.J., de Wit, E., van den Brand, J.M.A., Herfst, S., Schrauwen, E.J.A., Bestebroer, T.M., van de Vijver, D., Boucher, C.A., Koopmans, M., Rimmelzwaan, G.F., Kuiken, T., Osterhaus, A.D.M.E., Fouchier, R.A.M., 2009. Pathogenesis and transmission of swine-origin 2009 A(H1N1) influenza virus in ferrets. Science 325, 481−483.

Nelson, M.I., Gramer, M.R., Vincent, A.L., Holmes, E.C., 2012. Global transmission of influenza viruses from humans to swine. J. Gen. Virol. 93, 2195−2203.

Nelson, M.I., Lemey, P., Tan, Y., Vincent, A., Lam, T.T.-Y., Detmer, S., Viboud, C., Suchard, M.a., Rambaut, A., Holmes, E.C., Gramer, M., 2011. Spatial dynamics of human-origin H1 influenza A virus in North American swine. PLoS Pathog. 7, e1002077.

Nelson, M.I., Viboud, C., Vincent, A.L., Culhane, M.R., Detmer, S.E., Wentworth, D.E., Rambaut, A., Suchard, M.A., Holmes, E.C., Lemey, P., 2015. Global migration of influenza A viruses in swine. Nat. Commun. 6. https://doi.org/10.1038/ncomms7696.

Nelson, M.I., Vincent, A.L., 2015. Reverse zoonosis of influenza to swine: new perspectives on the human-animal interface. Trends Microbiol. 23, 142−153.

Nelson, M.I., Wentworth, D.E., Das, S.R., Sreevatsan, S., Killian, M.L., Nolting, J.M., Slemons, R.D., Bowman, A.S., 2016. Evolutionary dynamics of influenza a viruses in US exhibition swine. J. Infect. Dis. 213. https://doi.org/10.1093/infdis/jiv399.

Nelson, M.I., Worobey, M., 2018. Origins of the 1918 pandemic: revisiting the swine "mixing vessel" hypothesis. Am. J. Epidemiol. 187. https://doi.org/10.1093/aje/kwy150.

Neveau, M.N., Zeller, M.A., Kaplan, B.S., Souza, C.K., Gauger, P.C., Vincent, A.L., Anderson, T.K., 2022. Genetic and antigenic characterization of an expanding H3 influenza A virus clade in U.S. swine visualized by Nextstrain. mSphere 7, e0099421.

Nicholls, J.M., Chan, R.W.Y., Russell, R.J., Air, G.M., Peiris, J.S.M., 2008. Evolving complexities of influenza virus and its receptors. Trends Microbiol. 16, 149−157.

O'Neill, E., Donis, R.O., 2009. Generation and characterization of candidate vaccine viruses for prepandemic influenza vaccines. Curr. Top. Microbiol. Immunol. 333, 83−108.

O'Toole, Á., Neher, R.A., Ndodo, N., Borges, V., Gannon, B., Gomes, J.P., Groves, N., King, D.J., Maloney, D., Lemey, P., Lewandowski, K., Loman, N., Myers, R., Omah, I.F., Suchard, M.A., Worobey, M., Chand, M., Ihekweazu, C., Ulaeto, D., Adetifa, I., Rambaut, A., 2023. APOBEC3 deaminase editing in mpox virus as evidence for sustained human transmission since at least 2016. Science 382, 595−600.

Pardo-Roa, C., Nelson, M.I., Ariyama, N., Aguayo, C., Almonacid, L.I., Munoz, G., Navarro, C., Avila, C., Ulloa, M., Reyes, R., Luppichini, E.F., Mathieu, C., Vergara, R., González, Á., González, C.G., Araya, H., Fernández, J., Fasce, R., Johow, M., Medina, R.A., Neira, V., 2023. Cross-species transmission and PB2 mammalian adaptations of highly pathogenic avian influenza A/H5N1 viruses in Chile. bioRxiv. https://doi.org/10.1101/2023.06.30.547205.

Pearce, M.B., Jayaraman, A., Pappas, C., Belser, J.A., Zeng, H., Gustin, K.M., Maines, T.R., Sun, X., Raman, R., Cox, N.J., Sasisekharan, R., Katz, J.M., Tumpey, T.M., 2012. Pathogenesis and transmission of swine origin A(H3N2)v influenza viruses in ferrets. Proc. Natl. Acad. Sci. U. S. A. 109, 3944−3949.

Pickering, B., Lung, O., Maguire, F., Kruczkiewicz, P., Kotwa, J.D., Buchanan, T., Gagnier, M., Guthrie, J.L., Jardine, C.M., Marchand-Austin, A., Massé, A., McClinchey, H., Nirmalarajah, K., Aftanas, P., Blais-Savoie, J., Chee, H.-Y., Chien, E., Yim, W., Banete, A., Griffin, B.D., Yip, L., Goolia, M., Suderman, M., Pinette, M., Smith, G., Sullivan, D., Rudar, J., Vernygora, O., Adey, E., Nebroski, M., Goyette, G., Finzi, A., Laroche, G., Ariana, A., Vahkal, B., Côté, M., McGeer, A.J., Nituch, L., Mubareka, S., Bowman, J., 2022. Divergent SARS-CoV-2 variant emerges in white-tailed deer with deer-to-human transmission. Nat. Microbiol. 7, 2011−2024.

Plowright, R.K., Parrish, C.R., McCallum, H., Hudson, P.J., Ko, A.I., Graham, A.L., Lloyd-Smith, J.O., 2017. Pathways to zoonotic spillover. Nat. Rev. Microbiol. https://doi.org/10.1038/nrmicro.2017.45.

Purves, K., Brown, H., Haverty, R., Ryan, A., Griffin, L.L., McCormack, J., O'Reilly, S., Mallon, P.W., Gautier, V., Cassidy, J.P., Fabre, A., Carr, M.J., Gonzalez, G., Ciuti, S., Fletcher, N.F., 2023. First Eurasian cases of SARS-CoV-2 seropositivity in a free-ranging urban population of wild fallow deer. bioRxiv. https://doi.org/10.1101/2023.07.07.547941.

Rajão, D.S., Chen, H., Perez, D.R., Sandbulte, M.R., Gauger, P.C., Loving, C.L., Shanks, G.D., Vincent, A., 2016. Vaccine-associated enhanced respiratory disease is influenced by haemagglutinin and neuraminidase in whole inactivated influenza virus vaccines. J. Gen. Virol. 97, 1489−1499.

Rambo-Martin, B.L., Keller, M.W., Wilson, M.M., Nolting, J.M., Anderson, T.K., Vincent, A.L., Bagal, U.R., Jang, Y., Neuhaus, E.B., Davis, C.T., Bowman, A.S., Wentworth, D.E., Barnes, J.R., 2020. Influenza A virus field surveillance at a swine-human interface. mSphere 5. https://doi.org/10.1128/msphere.00822-19.

Reich, N.G., Lessler, J., Funk, S., Viboud, C., Vespignani, A., Tibshirani, R.J., Shea, K., Schienle, M., Runge, M.C., Rosenfeld, R., Ray, E.L., Niehus, R., Johnson, H.C., Johansson, M.A., Hochheiser, H., Gardner, L., Bracher, J., Borchering, R.K., Biggerstaff, M., 2022. Collaborative hubs: making the most of predictive epidemic modeling. Am. J. Publ. Health 112, 839−842.

Sediri, H., Schwalm, F., Gabriel, G., Klenk, H.-D., 2015. Adaptive mutation PB2 D701N promotes nuclear import of influenza vRNPs in mammalian cells. Eur. J. Cell Biol. 94, 368−374.

Sharma, A., Zeller, M.A., Souza, C.K., Anderson, T.K., Vincent, A.L., Harmon, K., Li, G., Zhang, J., Gauger, P.C., 2022. Characterization of a 2016-2017 human seasonal H3 influenza A virus spillover now endemic to U.S. swine. mSphere 7, e0080921.

Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N., Kawaoka, Y., 2006. Avian flu: influenza virus receptors in the human airway. Nature 440, 435−436.

Shope, R.E., 1931. The etiology of swine influenza. Science 73, 214−215.

Sit, T.H.C., Brackman, C.J., Ip, S.M., Tam, K.W.S., Law, P.Y.T., To, E.M.W., Yu, V.Y.T., Sims, L.D., Tsang, D.N.C., Chu, D.K.W., Perera, R.A.P.M., Poon, L.L.M., Peiris, M., 2020. Infection of dogs with SARS-CoV-2. Nature 586, 776−778.

Smith, G.J.D., Bahl, J., Vijaykrishna, D., Zhang, J., Poon, L.L.M., Chen, H., Webster, R.G., Peiris, J.S.M., Guan, Y., 2009a. Dating the emergence of pandemic influenza viruses. Proc. Natl. Acad. Sci. U. S. A. 106, 11709−11712.

Smith, G.J.D., Vijaykrishna, D., Bahl, J., Lycett, S.J., Worobey, M., Pybus, O.G., Ma, S.K., Cheung, C.L., Raghwani, J., Bhatt, S., Peiris, J.S.M., Guan, Y., Rambaut, A., 2009b. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 459, 1122−1125.

Sooksawasdi Na Ayudhya, S., Kuiken, T., 2021. Reverse zoonosis of COVID-19: lessons from the 2009 influenza pandemic. Vet. Pathol. https://doi.org/10.1177/0300985820979843.

Soundararajan, V., Tharakaraman, K., Raman, R., Raguram, S., Shriver, Z., Sasisekharan, V., Sasisekharan, R., 2009. Extrapolating from sequence–the 2009 H1N1 "swine" influenza virus. Nat. Biotechnol. 27, 510−513.

Souza, C.K., Anderson, T.K., Chang, J., Venkatesh, D., Lewis, N.S., Pekosz, A., Shaw-Saliba, K., Rothman, R.E., Chen, K.-F., Vincent, A.L., 2022. Antigenic distance between North American swine and human seasonal H3N2 influenza A viruses as an indication of zoonotic risk to humans. J. Virol. 96, e0137421.

Souza, C.K., Kimble, J.B., Anderson, T.K., Arendsee, Z.W., Hufnagel, D.E., Young, K.M., Gauger, P.C., Lewis, N.S., Davis, C.T., Thor, S., Vincent Baker, A.L., 2023. Swine-to-ferret transmission of antigenically drifted contemporary swine H3N2 influenza A virus is an indicator of zoonotic risk to humans. Viruses 15. https://doi.org/10.3390/v15020331.

Srinivasan, A., Viswanathan, K., Raman, R., Chandrasekaran, A., Raguram, S., Tumpey, T.M., Sasisekharan, V., Sasisekharan, R., 2008. Quantitative biochemical rationale for differences in transmissibility of 1918 pandemic influenza A viruses. Proc. Natl. Acad. Sci. U. S. A. 105, 2800−2805.

Stevens, J., Blixt, O., Glaser, L., Taubenberger, J.K., Palese, P., Paulson, J.C., Wilson, I.A., 2006a. Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. J. Mol. Biol. 355, 1143−1155.

Stevens, J., Blixt, O., Tumpey, T.M., Taubenberger, J.K., Paulson, J.C., Wilson, I.A., 2006b. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science 312, 404−410.

Stoffella-Dutra, A.G., de Campos, B.H., Bastos E Silva, P.H., Dias, K.L., da Silva Domingos, I.J., Hemetrio, N.S., Xavier, J., Iani, F., Fonseca, V., Giovanetti, M., de Oliveira, L.C., Teixeira, M.M., Lobato, Z.I.P., Ferreira, H.L., Arns, C.W., Durigon, E., Drumond, B.P., Alcantara, L.C.J., de Carvalho, M.P.N., de Souza Trindade, G., 2023. SARS-CoV-2 spillback to wild coatis in sylvatic-urban hotspot, Brazil. Emerg. Infect. Dis. 29, 664−667.

Subbarao, E.K., London, W., Murphy, B.R., 1993. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. J. Virol. 67, 1761−1764.

Sun, H., Li, H., Tong, Q., Han, Q., Liu, J., Yu, H., Song, H., Qi, J., Li, J., Yang, J., Lan, R., Deng, G., Chang, H., Qu, Y., Pu, J., Sun, Y., Lan, Y., Wang, D., Shi, Y., Liu, W.J., Chang, K.-C., Gao, G.F., Liu, J., 2023. Airborne transmission of human-isolated avian H3N8 influenza virus between ferrets. Cell 186, 4074−4084.e11.

Sun, X., Belser, J.A., Pulit-Penaloza, J.A., Brock, N., Pappas, C., Zanders, N., Jang, Y., Jones, J., Tumpey, T.M., Davis, C.T., Maines, T.R., 2023. Pathogenesis and transmission assessment of three swine-origin influenza A(H3N2) viruses with zoonotic risk to humans isolated in the U.S from 2017-2020. J. Infect. Dis. https://doi.org/10.1093/infdis/jiad359.

Tarendeau, F., Boudet, J., Guilligay, D., Mas, P.J., Bougault, C.M., Boulo, S., Baudin, F., Ruigrok, R.W.H., Daigle, N., Ellenberg, J., Cusack, S., Simorre, J.-P., Hart, D.J., 2007. Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. Nat. Struct. Mol. Biol. 14, 229−233.

Taubenberger, J.K., Reid, A.H., Lourens, R.M., Wang, R., Jin, G., Fanning, T.G., 2005. Characterization of the 1918 influenza virus polymerase genes. Nature 437, 889−893.

Taylor, L.H., Latham, S.M., Woolhouse, M.E., 2001. Risk factors for human disease emergence. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 983−989.

Thomas, M.N., Zanella, G.C., Cowan, B., Joaquin Caceres, C., Rajao, D.S., Perez, D.R., Gauger, P.C., Vincent Baker, A.L., Anderson, T.K., 2023. Nucleoprotein reassortment enhanced transmissibility of H3 1990.4.a clade influenza A virus in swine. bioRxiv. https://doi.org/10.1101/2023.10.30.564250.

Tochetto, C., Junqueira, D.M., Anderson, T.K., Gava, D., Haach, V., Cantão, M.E., Vincent Baker, A.L., Schaefer, R., 2023. Introductions of human-origin seasonal H3N2, H1N2 and pre-2009 H1N1 influenza viruses to swine in Brazil. Viruses 15. https://doi.org/10.3390/v15020576.

Tool for Influenza Pandemic Risk Assessment (TIPRA), n.d. https://www.who.int/teams/global-influenza-programme/avian-influenza/tool-for-influenza-pandemic-risk-assessment-(tipra). (Accessed 12 December 2023).

Tumpey, T.M., Maines, T.R., Van Hoeven, N., Glaser, L., Solórzano, A., Pappas, C., Cox, N.J., Swayne, D.E., Palese, P., Katz, J.M., García-Sastre, A., 2007. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. Science 315, 655−659.

Vandegrift, K.J., Yon, M., Surendran Nair, M., Gontu, A., Ramasamy, S., Amirthalingam, S., Neerukonda, S., Nissly, R.H., Chothe, S.K., Jakka, P., LaBella, L., Levine, N., Rodriguez, S., Chen, C., Sheersh Boorla, V., Stuber, T., Boulanger, J.R., Kotschwar, N., Aucoin, S.G., Simon, R., Toal, K.L., Olsen, R.J., Davis, J.J., Bold, D., Gaudreault, N.N., Dinali Perera, K., Kim, Y., Chang, K.-O., Maranas, C.D., Richt, J.A., Musser, J.M., Hudson, P.J., Kapur, V., Kuchipudi, S.V., 2022. SARS-CoV-2 omicron (B.1.1.529) infection of wild white-tailed deer in New York City. Viruses 14, 2770.

Vander Veen, R.L., Harris, D.L.H., Kamrud, K.I., 2012. Alphavirus replicon vaccines. Anim. Health Res. Rev. 13, 1−9.

VanInsberghe, D., McBride, D.S., DaSilva, J., Stark, T.J., Lau, M.S.Y., Shepard, S.S., Barnes, J.R., Bowman, A.S., Lowen, A.C., Koelle, K., 2023. Genetic drift and purifying selection shape within-host influenza A virus populations during natural swine infections. bioRxiv. https://doi.org/10.1101/2023.10.23.563581.

van Riel, D., Munster, V.J., de Wit, E., Rimmelzwaan, G.F., Fouchier, R.A.M., Osterhaus, A.D.M.E., Kuiken, T., 2006. H5N1 virus attachment to lower respiratory tract. Science 312, 399.

Venkatesh, D., Anderson, T.K., Kimble, J.B., Chang, J., Lopes, S., Souza, C.K., Pekosz, A., Shaw-Saliba, K., Rothman, R.E., Chen, K.-F., Lewis, N.S., Vincent Baker, A.L., 2022. Antigenic characterization and pandemic risk assessment of North American H1 influenza A viruses circulating in swine. Microbiol. Spectr. 10, e0178122.

Vijaykrishna, D., Smith, G.J.D., Pybus, O.G., Zhu, H., Bhatt, S., Poon, L.L.M., Riley, S., Bahl, J., Ma, S.K., Cheung, C.L., Perera, R.A.P.M., Chen, H., Shortridge, K.F., Webby, R.J., Webster, R.G., Guan, Y., Peiris, J.S.M., 2011. Long-term evolution and transmission dynamics of swine influenza A virus. Nature 473, 519–522.

Vincent, A., Awada, L., Brown, I., Chen, H., Claes, F., Dauphin, G., Donis, R., Culhane, M., Hamilton, K., Lewis, N., Mumford, E., Nguyen, T., Parchariyanon, S., Pasick, J., Pavade, G., Pereda, A., Peiris, M., Saito, T., Swenson, S., Van Reeth, K., Webby, R., Wong, F., Ciacci-Zanella, J., 2014. Review of influenza A virus in swine worldwide: a call for increased surveillance and research. Zoonoses Public Health 61, 4–17.

Vincent, A.L., Anderson, T.K., Lager, K.M., 2020. A brief introduction to influenza A virus in swine. Methods Mol. Biol. 2123, 249–271.

Vincent, A.L., Perez, D.R., Rajao, D., Anderson, T.K., Abente, E.J., Walia, R.R., Lewis, N.S., 2016. Influenza A virus vaccines for swine. Vet. Microbiol. https://doi.org/10.1016/j.vetmic.2016.11.026.

Wasik, B.R., Barnard, K.N., Ossiboff, R.J., Khedri, Z., Feng, K.H., Yu, H., Chen, X., Perez, D.R., Varki, A., Parrish, C.R., 2017. Distribution of O-acetylated sialic acids among target host tissues for influenza virus. mSphere 2, e00379–16.

Watson, S.J., Langat, P., Reid, S.M., Lam, T.T.-Y., Cotten, M., Kelly, M., Van Reeth, K., Qiu, Y., Simon, G., Bonin, E., Foni, E., Chiapponi, C., Larsen, L., Hjulsager, C., Markowska-Daniel, I., Urbaniak, K., Dürrwald, R., Schlegel, M., Huovilainen, A., Davidson, I., Dán, Á., Loeffen, W., Edwards, S., Bublot, M., Vila, T., Maldonado, J., Valls, L., Brown, I.H., Pybus, O.G., Kellam, P., 2015. Molecular epidemiology and evolution of influenza viruses circulating within European swine between 2009 and 2013. J. Virol. 89, 9920–9931.

Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., Kawaoka, Y., 1992. Evolution and ecology of influenza A viruses. Microbiol. Rev. 56, 152–179.

Wilson, J.C., Bryans, J.T., Doll, E.R., 1965. Recovery of influenza virus from horses in the equine influenza epizootic of 1963. Am. J. Vet. Res. 26, 1466–1468.

Wong, K.K., Greenbaum, A., Moll, M.E., Lando, J., Moore, E.L., Ganatra, R., Biggerstaff, M., Lam, E., Smith, E.E., Storms, A.D., Miller, J.R., Dato, V., Nalluswami, K., Nambiar, A., Silvestri, S.A., Lute, J.R., Ostroff, S., Hancock, K., Branch, A., Trock, S.C., Klimov, A., Shu, B., Brammer, L., Epperson, S., Finelli, L., Jhung, M.A., 2012. Outbreak of influenza A (H3N2) variant virus infection among attendees of an agricultural fair, Pennsylvania, USA, 2011. Emerg. Infect. Dis. 18, 1937–1944.

Worobey, M., Han, G.-Z., Rambaut, A., 2014. A synchronized global sweep of the internal genes of modern avian influenza virus. Nature. https://doi.org/10.1038/nature13016.

Wymore Brand, M., Anderson, T.K., Kitikoon, P., Brian Kimble, J., Otis, N., Gauger, P.C., Souza, C.K., Kaplan, B., Mogler, M., Strait, E., Vincent Baker, A.L., 2022. Bivalent hemagglutinin and neuraminidase influenza replicon particle vaccines protect pigs against influenza a virus without causing vaccine associated enhanced respiratory disease. Vaccine 40, 5569–5578.

Yamada, S., Hatta, M., Staker, B.L., Watanabe, S., Imai, M., Shinya, K., Sakai-Tagawa, Y., Ito, M., Ozawa, M., Watanabe, T., Sakabe, S., Li, C., Kim, J.H., Myler, P.J., Phan, I., Raymond, A., Smith, E., Stacy, R., Nidom, C.A., Lank, S.M., Wiseman, R.W., Bimber, B.N., O'Connor, D.H., Neumann, G., Stewart, L.J., Kawaoka, Y., 2010. Biological and structural characterization of a host-adapting amino acid in influenza virus. PLoS Pathog. 6, e1001034.

Yen, H.-L., Sit, T.H.C., Brackman, C.J., Chuk, S.S.Y., Gu, H., Tam, K.W.S., Law, P.Y.T., Leung, G.M., Peiris, M., Poon, L.L.M., HKU-SPH study team, 2022. Transmission of SARS-CoV-2 delta variant (AY.127) from pet hamsters to humans, leading to onward human-to-human transmission: a case study. Lancet 399, 1070–1078.

Zeller, M.A., Chang, J., Vincent, A.L., Gauger, P.C., Anderson, T.K., 2021. Spatial and temporal coevolution of N2 neuraminidase and H1 and H3 hemagglutinin genes of influenza A virus in US swine. Virus Evol. 7, veab090.

Zeller, M.A., Ma, J., Wong, F.Y., Tum, S., Hidano, A., Holt, H., Chhay, T., Sorn, S., Koeut, D., Seng, B., Chao, S., Ng, G.G.K., Yan, Z., Chou, M., Rudge, J.W., Smith, G.J.D., Su, Y.C.F., 2023. The genomic landscape of swine influenza A viruses in Southeast Asia. Proc. Natl. Acad. Sci. U. S. A. 120, e2301926120.

Zeller, M.A., Saxena, A., Anderson, T.K., Vincent, A.L., Gauger, P.C., 2022. Use of the ISU FLUture multi-sequence identity tool for rapid interpretation of swine influenza A virus sequences in the United States. J. Vet. Diagn. Invest. 34, 874−878.

CHAPTER

Emerging Viruses

23

Rafael Sanjuán

Institute for Integrative Systems Biology (I^2SysBio), Universitat de València − CSIC, Paterna, Spain

1. Introduction: Emergent Viruses and the Virosphere

Emerging viruses have a major impact on public health, as evidenced by the recent COVID-19 pandemic and other diseases that emerged previously such as AIDS and the "Spanish" influenza, among others. Over the past decades, outbreaks have revealed a wide diversity of novel viruses, some of which pose important threats. These emerging viruses have a zoonotic origin, meaning that they have been transmitted from an animal reservoir to humans. Indeed, only a small fraction of all human pathogenic viruses have been inherited vertically over geological timescales (Holmes, 2022). Hence, instead of classifying viruses as zoonotic or nonzoonotic, the most relevant question is when these viruses were transmitted from animals to humans (Forni et al., 2022). The timescales of these events vary amply. For instance, it is believed that the measles virus originated from cattle rinderpest virus approximately 2500 years ago (Düx et al., 2020), whereas herpes simplex virus 2 was probably transmitted between bonobos and human ancestors around a million years ago (Wertheim et al., 2021).

Viral metagenomics is helping us achieve a broader and less human-centric view of the virosphere. Remarkably, metagenomics has revealed the zoonotic nature of viruses that were believed to circulate exclusively in humans. For instance, hepatitis E virus genotypes 3 and 4 are now known to infect a wide range of mammals, including nonhuman primates, pigs, rabbits, and rodents, and cause sporadic zoonotic infections in humans, especially from pork consumption (Meng, 2016). Viral metagenomics has revealed approximately 400 new virus-mammal associations per year since the beginning of this century, according to data deposited in the VIRION database (www.viralemergence.org/virion). Most emerging human viruses originate from other mammals (ca. 85%), followed by birds and, very rarely, other vertebrates. Still, the total number of viral species known to infect humans is a minute fraction of the estimated 40,000 mammalian viruses present in nature and of the estimated 10,000 viruses with the potential to become zoonotic (Carlson et al., 2019). Hence, the vast majority of animal viruses may never infect people, either because we are not exposed to these viruses or because they are not human-infective. Even if infection occurs, human−human transmissibility may be inefficient, leading to dead-end or "spillover" infections. Only those viruses capable of causing significant outbreaks might end up becoming new human viruses (Fig. 23.1).

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00031-8
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

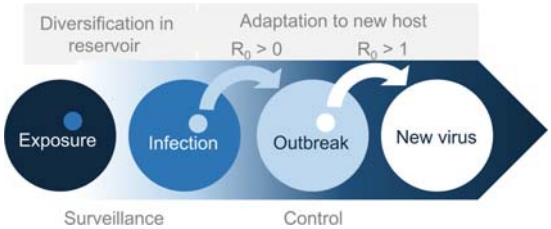


FIGURE 23.1 Bottlenecks in viral emergence.
Humans or other potential hosts are exposed to a small fraction of the tens of thousands of mammalian viruses estimated to circulate in nature. Of these, most are not infectious in the new host, and among those which are infectious, only a small fraction will cause an outbreak (basic reproductive number, R_0, larger than zero). Most outbreaks will be self-limited contained (only viruses for which $R_0 > 1$ can become established in the host population). Spillover infections can be defined as those showing $R_0 < 1$ in the new host. Before infections occur in the new host, the virus might diversify in the reservoir host, potentially leading to the emergence of variants with increased cross-species transmissibility. Once the virus infects and transmits among members of the new host, specific adaptation should take place under natural selection. Surveillance of reservoir hosts and spillover infections can help prevent viral emergence, but early epidemiological control of outbreaks is also highly important.

Viral cross-species transmission and emergence in a new host is a universal process that has been observed not only in humans but also in animals, plants, other eukaryotes, and prokaryotes. Since the topic is too vast, this chapter will focus on mammalian and human viruses. It should be noted, though, that zoonosis is just a special case of many possible viral cross-species transmissions, and that humans are not necessarily the end stage of such processes, as shown by several examples of reverse zoonosis (i.e., transmission of human viruses to domestic or wild animals). Indeed, viruses are ubiquitous components of all ecosystems and often circulate among different host species, on many occasions leading to commensal interactions (Holmes, 2022). Ecological disturbances may increase the risk of cross-species transmission and disease emergence in new hosts, which in turn can have important effects on ecosystem functioning.

Therefore, the diversity, ecology, and evolvability of animal viruses should be viewed as cornerstones in viral emergence, including zoonosis. For instance, it has been suggested the diversification of the SARS-CoV-2 spike protein in animal reservoirs and the subsequent selection of human-infective variants probably played a decisive role in COVID-19 emergence (Andersen et al., 2020). A snapshot of viral composition at a given place and moment may thus lead to a poor understanding of the mechanisms leading to viral emergence. Viral genetic diversification in natural ecosystems can produce potentially human-infective variants, which may find their way into humans after critical habitat perturbations, subsequently allowing natural selection to promote the evolution of virus variants that are better adapted to us. However, our understanding of the complex processes leading to viral emergence is still limited, and information about how new viruses function is often lacking.

This chapter addresses different aspects of viral emergence. The importance of the ecological and societal contexts is discussed first. The role of the host from which viruses emerge is discussed next, including wildlife reservoirs and domestic animals that function as intermediate hosts before transmission to humans occurs. Then, cross-species transmission is tackled from the point of view of the virus and the molecular determinants of virus–host interactions. Finally, the steps leading from initial cross-species transmission events to viral emergence are discussed.

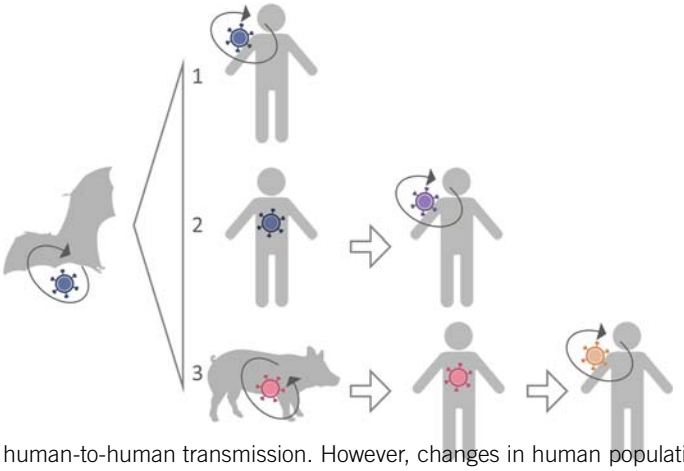


FIGURE 23.2 Possible alternative scenarios leading to viral emergence in humans.

1. Contact between humans and a wildlife species that serves as the viral reservoir directly leads to viral cross-species transmission and establishment in humans. Increased contact between the reservoir and humans can be driven by ecological factors such as changes in land use, etc. 2. Emergence following spillover infections. The virus is transmitted to humans once or multiple times but is initially unable to achieve human-to-human transmission. However, changes in human populations (increased traveling, crowding, famines, etc.) can promote human transmissibility. Alternatively, the virus can evolve to become more transmissible in humans. 3. Domesticated animals can serve as intermediate hosts that facilitate viral zoonosis. The virus is transmitted from wild to domestic species, where it starts to circulate and potentially evolves. Frequent exposure to domesticated animals or increased similarity with humans facilitates zoonosis and viral establishment. Viral circulation in a given host is indicated by *thin curved arrows*, cross-species transmission by broad *white-filled arrows*, and viral evolution by changes in the color of the virion symbol.

2. Ecological and Societal Drivers of Viral Emergence

2.1 Viral Emergence as an Ecological Process

Viral emergence is a trigger for viral evolution since cross-species transmission sets into motion new factors to which viruses adapt through natural selection. However, the reverse is not always true, meaning that changes in the genetic composition of viral populations are not strictly required for viral emergence to occur (Fig. 23.2). Hence, viral emergence can in principle be solely a result of changes in the ecology of the reservoir host, vectors, or recipient host (Plowright et al., 2017).

Ecologists aim to establish a general theory that links ecosystem properties with disease emergence risks. One such attempt relates host species diversity with disease prevalence. A simple first approach to this question is the dilution hypothesis, which states that virus transmission should be reduced in more diverse host communities. The reason is that most viruses infect one or a few host species, and the population density of each species should be lower in richer communities as a result of interspecies competition (Keesing et al., 2010). Supporting this idea, it was shown that a higher diversity of birds correlates with a lower incidence of human infection with West Nile virus in the United States (Swaddle and Calos, 2008). Conversely, in a field study carried out in Panama, where zoonotic hantaviruses are endemic, small-mammal biodiversity was experimentally reduced by removing non-reservoir species in field plots. This perturbation led to increases in host population density and infection prevalence in wild rodent reservoir species (Suzán et al., 2009). However, the relationship between species richness and cross-species transmission rates might be more complex. As the number

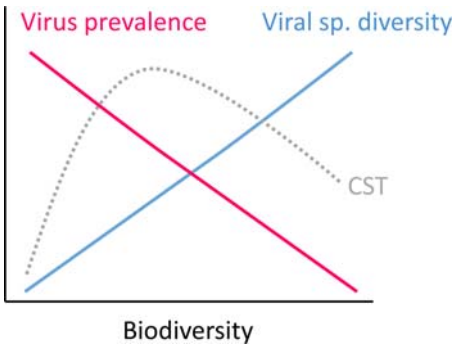


FIGURE 23.3 Interpretation of the dilution effect hypothesis.

This hypothesis states that parasite transmission and, hence, prevalence should correlate negatively with biodiversity because competition with nonsusceptible species limits the population density of the host. This is shown qualitatively with the red line. However, it is also expected that viral species diversity should increase with host species diversity (*blue line*). As a result, maximal cross-species transmission (CST) might occur at intermediate levels of biodiversity (*gray dotted line*). The asymmetrical shape of this line as a function of biodiversity aims to reflect the idea that species-species contacts should increase with biodiversity, making cross-species transmission increase faster than linearly with biodiversity. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

of host species decreases, so does the diversity of their associated viruses, therefore reducing the number of possible virus–host species encounters. Therefore, one can hypothesize maximal cross-species transmission might take place at intermediate biodiversity levels (Fig. 23.3).

Several studies have examined ecological scenarios that correlate with zoonotic risk. For instance, machine learning was applied to explain the location and time of viral discovery for 223 ICTV-approved human-infective RNA viruses as a function of climate (rainfall, temperature, latitude), demography (human population size and growth rate), ecology (land use, mammal biodiversity, livestock abundance), as well as research effort and disease management capabilities (country gross domestic product, number of universities). The explanatory variables with the highest weights were gross domestic product and urbanized land, probably reflecting research effort bias, as illustrated by the fact that most viruses have been historically first reported in the United States and Europe. However, climate and ecological factors such as temperature, mammal species richness, and changes in precipitation also played a role (Zhang et al., 2020a). Correcting for research effort, the study identified Southeast Asia as the highest-risk area in the world, followed by tropical Africa and America (Allen et al., 2017).

In general, tropical forests are usually identified as hotspots for zoonosis and viral emergence at large because these are regions with not only high wildlife diversity but also frequent land-use alterations and bush meat consumption. It should be noted, though, that some important emergent viral diseases, such as Middle-East respiratory syndrome (MERS) and the 2009 swine flu, did not originate in tropical areas (Gibbs et al., 2009; Han et al., 2016). In Europe, wildlife biodiversity is lower and the number of potentially zoonotic viruses is reduced accordingly, but deforestation caused by changes in land use and forest fires has also promoted contacts between humans and wild mammals. Examples of

animal-to-human viral transmission events in Europe include Louping ill, tick-borne encephalitis, hantaviruses such as Puumala and Dobrava viruses, and Granada virus (Kallio-Kokko et al., 2005). Ecological interactions are complex and, in some specific cases, relatively low population densities and dry climate can unexpectedly promote viral transmission. For instance, red-necked avocets from the arid interior in Australia had a greater viral diversity and abundance than those from temperate Australia, probably because water sources in arid areas are visited by a higher diversity and density of species (Wille et al., 2018). In some cases, such as Hendra virus transmitted from bats to horses, cold weather may favor cross-species transmission by increasing the stability of viral particles deposited in the ground, grass, leaves, or fruits, which increases environmental exposure rates (Fogarty et al., 2008).

Global warming is an important driver of disease emergence for vector-borne viruses. A clear example is the expansion of Chikungunya virus to Europe, and of West Nile virus to northern latitudes (Paz, 2015; Tjaden et al., 2021). The global dissemination of arboviruses in recent years has been particularly dramatic for bluetongue and epizootic hemorrhagic disease viruses infecting ruminants and transmitted by hematophagous midges (Maclachlan et al., 2019). Hence, in the decades to come, diseases caused by arboviruses that were circumscribed to the tropics may become endemic in areas such as the Mediterranean basin. Moreover, deforestation is also forcing some wild species to invade urban areas, such as foxes and wild dogs which are major sources of rabies transmission to humans in certain countries.

2.2 Role of Societal Changes in Viral Emergence and Evolution

In general, humans are exposed to a small fraction of the thousands of viruses that circulate in wild animals, but exposure rates can vary strongly. For instance, wildlife markets are an important interface for zoonoses (He et al., 2022). Other known risk activities include hunting and wildlife management practices. Changes in human societies, particularly traveling, have played a major role in the upsurge of viral zoonoses. This has been extensively studied in the case of HIV-1 (Faria et al., 2014; Tebit and Arts, 2011). The virus emerged multiple times in a relatively short period around the beginning of the 20th century. Throughout history, spillover events in which simian immunodeficiency virus was transmitted from chimpanzees to humans might have been relatively frequent in Central Africa, mainly as a result of bush meat activities. However, when Europeans established colonial cities, a largely masculine labor force was recruited, increasing migration, disrupting tribal societies, and favoring the spread of sexually transmitted diseases across many regions. Injected medical treatments and vaccinations without proper sterilization also became frequent at the time, which might have further promoted viral transmission. Finally, frequent long-distance traveling allowed the spread of HIV-1 outside Africa to Haiti and major US cities in the late 1970s (Worobey et al., 2016). These societal changes offered a favorable scenario for HIV-1 transmission and provided new opportunities for the virus to evolve in humans.

Another example of how societal changes may determine viral emergence is provided by the 2022 global outbreak of monkeypox. Occasional monkeypox virus transmission from rodents or other mammals to humans occurred intermittently in central Africa for decades without causing major outbreaks (Tiee et al., 2018). The smallpox vaccine, universally administered in the 1970s, confers cross-protection against monkeypox. However, cessation of vaccination after smallpox eradication resulted in a gradual decline of population immunity, offering the opportunity for monkeypox emergence (Taube et al., 2023).

3. The Sources of Viral Emergence

3.1 Zoonotic Viruses Originate Preferentially From Some Host Taxa

Certain mammalian taxa cause zoonoses more often than others, but in many cases, the reasons for these differences remain unclear. One study questioned the idea that certain mammals tend to be at higher risk of causing viral zoonoses (Mollentze and Streicker, 2020). Systematic analysis of virus–reservoir relationships for hundreds of viruses supported a largely neutral model, in which the distribution of human-infective viruses across taxonomical orders simply obeyed viral richness and host abundance.

In contrast to this neutral model, similar host species should tend to share more viruses than distantly related species, which is expected based on their resemblance in terms of cellular receptors, physiology, immune responses, and so on (Albery et al., 2020). Therefore, nonhuman primates should be particularly high-risk mammals. Their contribution to new human viruses, though, is limited by the fact that primates are globally less abundant than other taxa such as bats or rodents (Olival et al., 2017; Wang and Anderson, 2019; Woolhouse et al., 2012). Indeed, it has been found that host species that occupy wider geographical ranges and are more densely populated tend to share more viruses with other host species (Albery et al., 2021).

Another general host feature that predicts zoonotic propensity is host lifespan. Species with short lives, fast reproduction, small body sizes, and high population sizes harbor on average more parasites and are more likely to transfer zoonotic viruses (Albery and Becker, 2021). Moreover, fast-lived mammals tend to invest less in some aspects of adaptive immunity, which could make them more susceptible to viral infections. These species are also more often found in disturbed ecosystems and have adapted better to live close to human settlements (Keesing et al., 2010). Lastly, urban-adapted mammals have more known pathogens, but this may be attributed to research effort bias, since these species are more easily captured and, therefore, their viromes are better characterized (Albery et al., 2022).

Several rodent species live in close contact with humans and domestic animals. For instance, the house mouse has been identified as an important source of virus cross-species transmission (Luis et al., 2015). The main source of human contact with rodent pathogens is exposure to the urine or feces of infected animals (Temmam et al., 2014). Dwellings and agricultural settings are among the highest-risk interfaces, particularly for rodents (Kreuder Johnson et al., 2015). For bats, it has been suggested that the metabolic consequences of flight evolution generated pleiotropic costs related to immunity that increased tolerance to viruses, making them more prone to being asymptomatic carriers of viruses (Brook and Dobson, 2015). Moreover, many bat species form large and densely populated colonies with frequent species cohabitation, offering a favorable scenario for viral transmission. Furthermore, some bats frequently travel large distances, fostering the establishment of viruses in other populations. Altogether, these features might contribute to the higher propensity of bats to be at the origin of cross-species viral transmission events, beyond the obvious fact that bats are abundant and diverse.

Based on the perception that they are frequent sources of viral zoonoses, the role of certain mammals as carriers of specific human-infective viruses is relatively well characterized. Zoonotic viruses from rodents include arenaviruses and hantaviruses and, to a lesser extent, coronaviruses, and flaviviruses (Wu et al., 2018). Rodents are the source of human infections with lymphocytic

choriomeningitis, Junin, and Lassa arenaviruses, Sin Nombre, Puumala, and Hantaan hantaviruses, as well as some coronaviruses. Bats carry a wider variety of SARS-related coronaviruses, the families Rhinolophidae and Hipposideridae being two well-studied groups of insectivorous bats with a very broad geographical distribution. For instance, RaTG13 and RmYN02 found in *Rhinolophus affinis* (horseshoe bats) animals captured in the Chinese province of Yunan are closely related to SARS-CoV-2 (Zhou et al., 2020a,b). It has been suggested that coronavirus positivity in bat populations can be as high as 10% (Geng and Zhou, 2021). New bat coronaviruses are continuously discovered, such as 19 alphacoronaviruses in a single recent study, which also reported over 300 new viruses from other families after sequencing libraries from over 2000 individual animals including bats, rodents, pikas, insectivores, pangolins, and zoo animals (Cui et al., 2023). Well-known zoonoses with a suspected or confirmed bat origin include not only coronaviruses but also Ebola and Marburg filoviruses, Nipah and Hendra paramyxoviruses, lyssaviruses causing rabies-like diseases, bunyaviruses and hantaviruses (Kohl and Kurth, 2014; Shipley et al., 2019; Simons et al., 2014; Wang and Anderson, 2019). Some older zoonoses such as measles, mumps, and hepatitis C are also believed to have originated in bats (Drexler et al., 2012; Quan et al., 2013). Bats have also transmitted important pathogens to domestic animals, such as swine acute diarrhea syndrome (SADS) coronavirus, which in 2016−17 caused a major outbreak in piglets from Chinese farms (Zhou et al., 2018).

3.2 The Importance of Viral Diversification in Reservoir Hosts

Most zoonotic viruses probably have a limited ability to spread in human populations, yet certain variants with increased transmissibility might cause significant outbreaks. How these variants emergence is unknown in most cases, albeit two theoretical scenarios can be devised in principle (Woolhouse et al., 2016). First, diversification in the initially infected human subjects might allow the natural selection of viruses with increased transmissibility. Second, such diversification might take place in nonhuman reservoir hosts, producing variants with increased human transmissibility which are not necessarily advantageous in the reservoir but are particularly likely to originate zoonotic outbreaks (Fig. 23.4). The first scenario seems unlikely since there are no well-documented cases of viruses typically causing dead-end infection in humans (such as rabies or avian influenza virus H5N1) that evolve the ability to undergo sustained human-to-human transmission. In contrast, there are multiple examples of zoonotic viruses that are human-transmissible in the first place. This suggests that, in principle, studying viral diversity in reservoir hosts should provide important clues about viral zoonosis.

However, in terms of viral genetics, in many cases little is known beyond mere detection of a particular virus with potential zoonotic risk in a given reservoir, mainly because the number of viruses to be monitored is extremely large and reservoir species are often undersampled. In the case of viruses that are already known to spillover frequently from animals to humans, important efforts have been made to understand their genetic diversity in the reservoir. Arguably, the best-studied case is influenza A virus circulating in aquatic birds (Mostafa et al., 2018). Genetically and antigenically diverse influenza A viruses are found in wild avian species around the world. Infections occur mainly in waterfowl of the orders Anseriformes, in which they usually produce asymptomatic infections transmitted via the fecal−oral route (Taubenberger and Kash, 2010). Clinical signs in poultry are similar and include slight losses in body weight and egg production. However, H5, H7, and H9 subtypes can be highly pathogenic, with mortality rates >50% (Alexander, 2000). Another group of

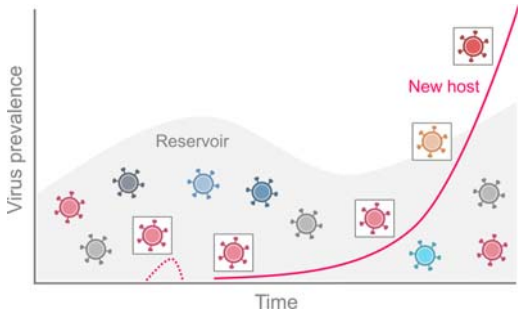


FIGURE 23.4 Role of genetic diversification in reservoir hosts.

A potentially emerging virus can circulate and diversify for long periods in a host that serves as the viral reservoir. This diversity is represented by virions of different colors. By chance, this might lead to the appearance of mutations that increase cross-species transmissibility (pink virion). Contact with the new host can lead to infections (virion in box), but on many occasions, this may not cause major outbreaks (spillover infections, *dotted pink line*). Under the right conditions (e.g., increase in host density or behavioral change), though, the virus might become established in the new host (*solid pink line*). Selection for increased transmissibility in the new host may rapidly promote genetic changes in the emerging virus (yellow and red virions in boxes). (For interpretation of the references to color in this figure legend, please refer online version of this title.)

viruses that is becoming very well studied in terms of their genetics in reservoir hosts are SARS-related viruses, which have been detected in many bat samples. Their genetic variation in the receptor binding domain is characterized by insertions and deletions, which determine the ability of these viruses to use Angiotensin I Converting Enzyme 2 (ACE2) as a receptor for entry, and human ACE2 in particular (Starr et al., 2022; Wells et al., 2021).

In general, though, the genetic and ecological factors determining viral circulation in reservoirs and their cross-species transmission to humans remain poorly studied, including well-known pathogens such as mosquito-borne flaviviruses (Marklewitz and Junglen, 2019). For example, Chikungunya virus circulates between humans and *Aedes* mosquitoes in urban areas but also between *Aedes* and nonhuman primates together with other mammals in sylvatic transmission cycles (Hakim et al., 2022). Spillover from the sylvatic cycles to humans is mediated by mosquito species biting both humans and wild mammals, such as *Aedes furcifer*. However, little is known about the genetic diversity of Chikungunya virus in the African forests, and sylvatic transmission cycles in other parts of the world may also exist but are uncharacterized. The Asian genotype of Chikungunya virus was isolated in long-tailed Malaysian macaques, and phylogenetic analyses showed that these were the result of a reverse zoonosis from humans, adding further complexity to the ecology and evolution of this virus (Suhana et al., 2019).

3.3 The Role of Intermediate Hosts in Viral Zoonosis

A virus is sometimes first transmitted from a wild reservoir species to domestic animals and, then, from domestic animals to humans (Glennon et al., 2018) (Fig. 23.2). Domestic animals are in closer contact

with wild species than humans since they can share drink or food sources. If they are susceptible to infection, their usually high and dense populations promote the spread of the new virus to high numbers. This, together with close contact with humans, favors zoonosis. This two-stage transmission can take place with or without selection for viral genetic variants that increase infectivity in the new hosts.

A viral emergence event in which an intermediate host was identified was the 2002−03 SARS outbreak. A virus with 99.8% sequence identity with SARS-CoV was found in palm civets (*Paguma larvata*) from wet markets (Guan et al., 2003). Civets probably acquired the virus from bats, and in turn, civets transmitted it to humans. Intermediate hosts for SARS-CoV-2, in contrast, remain undiscovered. Although a virus nearly identical to SARS-CoV-2 was isolated from pangolins, it is unclear whether these animals transmitted the virus to humans or experienced an independent dead-end spillover infection from bats or other species (Banerjee et al., 2021).

Several mammals have played a role in the cross-species transmission of influenza virus from birds to humans. Avian influenza strains have been isolated from pigs, horses, dogs, mink, seals, and whales, among others. Sustained circulation of the virus in domestic poultry is also considered an important zoonosis risk factor. Viral adaptation to gallinaceous birds involves the selection of mutations in the hemagglutinin (HA), neuraminidase (NA), and ribonucleoprotein genes. For instance, a deletion of approximately 20 residues in the stalk region of the NA protein has been associated with successful infection of poultry, probably by modulating the interactions with sialic acid, which serves as the main attachment factor required for viral entry (Taubenberger and Kash, 2010). Pigs have also hosted viral strains transmitted from birds, as well as from humans and, therefore, they have been suggested to function as "mixing vessels" where avian and human strains recombine by segment reassortment.

Pigs are important amplifier hosts for additional viral diseases, including Japanese encephalitis, Nipah, and hepatitis E (Glud et al., 2021). Indeed, it has been estimated that humans exchange over 30 different viruses with pigs (Johnson et al., 2020). Nipah virus is a paramyxovirus (genus henipavirus) first introduced in 1999 into pig herds from fruit bats in Malaysia and Singapore. Isolates from pigs, bats and humans showed >99% sequence identity, indicating recent transmission events and probably little species-specific viral adaptation. However, mortality rates were far higher in humans (40%) than in pigs (<5%). Increased severity was linked to higher expression of the receptor Ephrin B2 (EFNB2) in the human respiratory tract compared to pigs (Sauerhering et al., 2016). While there have been no other known outbreaks of Nipah in Malaysia and Singapore since 1999, outbreaks have been recorded almost annually in other parts of Asia since, especially in Bangladesh and India. The virus has been shown to spread from person-to-person, raising concerns about pandemic risks. The virus may be transmitted from bats to pigs through discarded fruit pulp but also through contaminated date palm sap (Islam et al., 2016; Luby et al., 2006). In other cases, viruses have been transmitted presumably from bats to pigs and caused major outbreaks in these domestic animals, but without pig-to-human transmission. This is the case of SADS coronavirus, a virus considered to be at high risk of future zoonosis because it can infect various human cell lines (Yang et al., 2020).

Horses have been shown to function as intermediate hosts for the zoonotic transmission of Hendra virus, another henipavirus (Plowright et al., 2015). Hendra virus is endemic to Australian *Pteropus* bats (fruit bats). Occasional human infections have been reported after viral amplification in horses. Transmission from bats to horses is believed to take place via viral shedding from bats onto surfaces, or through aerosolized urine or feces. Consequently, horses may get exposed when eating grass or fruit, drinking water, or even sniffling contaminated surfaces.

4. Genetic and Molecular Determinants of Viral Emergence

4.1 Virological Features

Viral cross-species transmission may promote or be driven by genetic changes that increase viral fitness in the new host. Since RNA viruses show extremely high rates of spontaneous mutation (Sanjuan and Domingo-Calap, 2016) and since these rates should afford them high adaptability, it is expected that most cross-species transmission events involve RNA viruses. Supporting this view, a metaanalysis of 586 mammalian viruses, 754 mammal species, and 2805 virus−host associations reported that RNA viruses represent 85% of all viral zoonoses and that the proportion of zoonotic viruses is threefold higher among RNA viruses than among DNA viruses (Olival et al., 2017).

However, more careful statistical analysis of these data revealed instead two other features associated with increased zoonotic propensity, namely replication in the cytoplasm and vector-borne transmission (Babayan et al., 2018; Olival et al., 2017). Replication in the cytoplasm was also suggested in a previous study as a relevant trait associated with zoonotic risk (Pulliam and Dushoff, 2009). As it turns out, most vector-borne viruses are RNA viruses, and most RNA viruses replicate in the cytoplasm, whereas most DNA viruses replicate in the nucleus. These correlations between viral features make it particularly difficult to reach unambiguous conclusions regarding which are the actual determinants of zoonotic risk. In a more recent study, a larger dataset of 12,000 mammalian virus−host associations was used to examine the role of several viral features (Valero-Rello and Sanjuán, 2022). It was concluded that possessing an envelope was the most significant factor associated with cross-species transmissibility and zoonotic propensity, whereas genomic features such as genome size, structure (singe/double-stranded), composition (RNA/DNA), as well replication in the cytoplasm played a less obvious role. The mechanism by which viral envelopes influence viral host range is unclear, though. Viral envelope proteins might be more flexible in terms of binding different cellular receptors, or in their ability to accommodate host-shift mutations without altering their stability and function. Previously analyzed datasets indicate that, within RNA viruses, 85% of the viruses exhibit a broad host range and 92% of vector-borne viruses are enveloped (Woolhouse and Brierley, 2018).

It has also been suggested that viruses that recombine more frequently have a greater ability to cross species boundaries because recombination allows exploring new combinations of mutations more rapidly (Patiño-Galindo et al., 2021; Strobel et al., 2022). Genetic reassortment of segmented viruses such as orthomyxoviruses, arenaviruses, and rotaviruses, is an efficient mechanism for achieving this. Influenza A virus is a paradigmatic example, since the various subtypes of the HA and NA segments can, in principle, form around 200 different combinations, many of which have been observed in birds (AbuBakar et al., 2023; Taubenberger and Kash, 2010). Acquisition of new segments from avian reservoirs or pigs has been at the origin of the major flu pandemics in humans. New combinations of mutations can also be produced by polymerase template switching, a process that occurs frequently during plus-strand RNA virus replication. This recombination mechanism has been implicated in the cross-species transmission of several coronaviruses, including the 2003 SARS, 2012 MERS, and COVID-19 outbreaks (Graham and Baric, 2010; Li et al., 2020a), as well as the emergence of new alphacoronavirus 1 strains in cats, dogs, and pigs (Jaimes et al., 2020). Despite these notorious examples, no overall statistical association between recombination rate and zoonotic virus emergence has been demonstrated, and it has been pointed out that, despite their high tendency to recombine,

coronaviruses do not have higher rates of cross-species transmission than rhabdoviruses, flaviviruses, or paramyxoviruses, which show low recombination rates (Holmes, 2022).

A systematic approach to quantifying long-term cross-species transmissibility was undertaken by examining the congruence between host and viral phylogenies (Geoghegan et al., 2017). Viral families that rarely undergo cross-species transmission should tend to coevolve with their hosts, thus producing phylogenetic trees that mirror those of their hosts. In contrast, frequent host jumps should lead to dissimilar phylogenies for the viruses and their current hosts. It was found that double-stranded DNA viruses including polyomaviruses, poxviruses, papillomaviruses, and adenoviruses showed higher levels of phylogenetic congruence with their hosts than RNA viruses, rhabdoviruses and picornaviruses being the viruses with the lowest levels of coevolution. This analysis also suggested that segmented viruses undergo host jumps more frequently than nonsegmented viruses, and a similar trend was found for negative-stranded RNA viruses when compared to positive-stranded RNA viruses.

Viral emergence risk does not depend only on cross-species transmissibility but also on the ability of the virus to produce significant outbreaks and become established in the new host. A systematic study involving 203 RNA and DNA viruses found that human transmissibility is generally associated with chronic infection and low virulence (Geoghegan et al., 2016). This was expected because these traits increase the number of secondary contacts per infected host. The use of efficient transmission routes such as respiratory droplets or insect vectors should also in principle increase outbreak risk. However, the above study found that the fraction of human-transmissible viruses was lower among arboviruses, probably because these viruses function as ecological generalists and can produce human spillover infections even if their natural transmission cycle does not involve humans. Probably for a similar reason, it was found that the faction of viruses showing sustained human transmission was lower among enveloped viruses than among nonenveloped viruses. Alternatively, this could be due to the higher stability of nonenveloped virions, which increases the chances of retaining infectiousness in the air, in water, and on surfaces.

4.2 Receptor Specificity as a Determinant of Cross-Species Transmission

A key event in viral emergence takes place when a new virus starts replicating in human cells since this first enables host-specific adaptation (Warren and Sawyer, 2019). For this initial stage to occur, the virus must find a compatible cell surface receptor, a process dependent on the viral receptor-binding protein located in the viral envelope or capsid. The genetic divergence of cellular receptors between potential host species can therefore be an important obstacle for viral cross-species transmission, providing a molecular basis for the observation that viral transmission is more likely between phylogenetically related hosts. There are many well-documented cases in which receptor specificity plays a key role in viral cross-species transmission. For MERS coronavirus, it has been shown that variations among mammalian species in several residues of the cellular receptor Dipeptidyl Peptidase 4 (DPP4) determine viral host tropism (Lu et al., 2015). Permissive species share most of these residues with humans, whereas nonpermissive species such as mouse, hamster, and ferret show more divergent sequences. Yet, changing these key residues is sufficient to make cells susceptible to the virus (van Doremalen et al., 2014, p. 4). Another example is provided by Tacaribe mammarenavirus, which uses the Transferrin Receptor (TFRC) of rodents and several other mammalian species, but not the human ortholog. It was shown though that substitution of a single residue in the human transferrin receptor renders it a functional receptor for this virus (Abraham et al., 2009).

As a result of the selective pressure exerted by viruses, many host genes encoding viral receptors have evolved rapidly. Examples include the filovirus receptor Niemann-Pick C1 Cholesterol Transporter 1 (NPC1) in bats (Ng et al., 2015), the arenavirus receptor TFRC in rodents (Kerr et al., 2015), and the avian Xenotropic and Polytropic Retrovirus Receptor 1 (XPR1) (Martin et al., 2013). This concept was examined systematically by testing for positive selection in nearly a 100 viral receptor genes from primates (Wang et al., 2020). Evidence for episodic positive selection was found in half of these genes. Moreover, a codon-wise analysis showed that receptor residues that directly interact with the viral surface are more often under positive selection, including specific residues of the HIV-1 receptor CD4, the HBV receptor Solute Carrier Family 10 Member 1 (SLC10A1), the adenovirus receptor CD46, and the enterovirus receptor CD55, among others. Virus receptors showed higher average nonsynonymous to synonymous substitution ratios and a higher fraction of codon under positive selection than other host genes performing similar functions but that are not known virus receptors. These analyses were later extended to other orders of mammals beyond primates (Wang and Han, 2021).

Conversely, point mutations in the viral receptor-binding protein can play a key role in the host-specificity of receptor usage. For instance, measles virus uses human Signaling Lymphocytic Activation Molecule Family Member 1 (SLAMF1) but not the orthologous protein of other mammals, whereas the related canine distemper virus does not use human SLAMF1. However, changing a single residue of the viral hemagglutinin protein allows canine distemper virus to infect human cells (Fukuhara et al., 2019). The genetics of receptor usage has been studied in many other viruses, two particularly well-studied cases being SARS-related viruses and influenza A virus.

Several SARS-related viruses use ACE2 as a cellular receptor, and the receptor binding domain (RBD) of the viral spike protein has been characterized in detail. SARS-related viruses often show insertions or deletions in the RBD, and this variability is an important determinant of the ability to interact with ACE2 (Lu et al., 2015). Moreover, the use of ACE2 from different hosts depends on specific RBD residues. For instance, it was found that the RBD sequences of SARS-CoVs from humans and civets (the probable intermediate host) differed at six residues, of which three interact directly with ACE2 (Graham and Baric, 2010). Two of these mutations, K479N and S487T mutations, were found to drive adaptation to the human receptor (Li et al., 2005). For SARS-CoV-2, several mutations in the spike are known to increase infectivity in humans, for instance, D614G (Zhang et al., 2020b). However, these mutations need not be human-specific adaptations, as they can also increase infectivity in other animals (Bashor et al., 2021). It has been suggested that one of the first amino acid substitutions in the spike protein of SARS-CoV-2 favoring viral emergence in humans was T372A (Kang et al., 2021). This change is present in all human SARS-CoV-2 sequences but not in SARS-related viruses isolated from bats or pangolins, and was predicted to remove a glycosylation site that may influence affinity for human ACE2.

For influenza virus, it has been extensively shown that viral entry is strongly dependent on the interaction between the viral HA and sialic acids that decorate proteins at the cell surface, although entry can also be driven by the viral NA (Karakus et al., 2020; Sempere Borau and Stertz, 2021). Sialic acids function as key attachment factors that allow the virus to further interact with certain proteins, such as the Epidermal Growth Factor Receptor (EGFR) (Eierhoff et al., 2010), the Calcium Voltage-Gated Channel Subunit Alpha1 C (CACNA1C) (Fujioka et al., 2018), or the Carcinoembryonic Antigen-Related Cell Adhesion Molecule 6 (CEACAM6) (Rahman et al., 2021) to trigger entry. Avian influenza viruses also use sialic acids but have a preference toward N-glycolylneuraminic acid (also

known as α2,3 sialic acid), whereas viral strains adapted to humans exhibit a preference toward N-acetylneuraminic acid (α2,6 sialic acid). This is a major barrier to zoonosis since the human upper respiratory tract contains mainly α2,6 sialic acid (Thompson and Paulson, 2021). Hence, the avian influenza virus HA needs to switch preference from α2,3 to α2,6 sialic acids to drive human infection. This can be achieved by several mutations in the HA gene. Specific such changes have indeed been associated with cross-species transmissions from birds to humans that gave rise to influenza pandemics, such as E190D and G225D for H1N1 (1918 and 2009) and Q226L and G228S for H2N2 (1957) and H3N2 (1968) (AbuBakar et al., 2023) (Table 23.1). Adaptation to human-to-human transmission has not been achieved by some avian influenza viruses, such as H5N1 or H7N9, which are of particular concern due to their high pathogenicity. Nevertheless, specific mutations that could confer the virus this capacity have already been identified (Song et al., 2017; Thompson and Paulson, 2021) (Table 23.1). An exception to sialic acid dependency is bat influenza virus H7N10, which uses the Major Histocompatibility Complex Class II DR Alpha Chain (HLA-DR) as a cellular receptor for entry in a sialic acid−independent manner (Giotis et al., 2019).

In principle, the species barrier imposed by receptor specificity should be weaker for viruses using receptors that are less variable across hosts. Indeed, several broad-range viruses use conserved receptors. For instance, EFNB2, used by Nipah virus, is involved in key processes such as embryogenesis, vasculogenesis, and neurogenesis, and is relatively conserved among mammals (Escudero-Pérez et al., 2023; Negrete et al., 2005). This idea was tested using data from 64 human viruses with at least one known receptor (Woolhouse et al., 2012). Viruses were classified as having a narrow host range if they only infected primates and as broad-range viruses otherwise. It was found that viruses using less conserved receptors (<90% human-mouse homology) were less likely to be broad-range viruses. However, the scenario can be more complex, since viral entry is often mediated by alternative receptors, as well as coreceptors, and accessory receptors. Many viruses are known to exhibit receptor usage plasticity. For instance, several betacoronaviruses can use CD209L (a C-type lectin) in addition to ACE2 for cell entry (Jeffers et al., 2004, 2006), and different cell surface receptors can participate in Zika virus entry (Lee et al., 2018). Viral host range may correlate not only with receptor conservation across species but also with the ability of the virus to use alternative receptors.

4.3 Additional Molecular Determinants of Cross-Species Transmission: Examples from Influenza Virus

Although receptor specificity is believed to be a major factor limiting cross-species viral transmission, additional barriers exist in other stages of the viral infection cycle. These barriers have been extensively investigated in influenza A virus, which provides a highly illustrative example.

One such barrier involves nuclear import. Following membrane fusion and release into the cytosol, the influenza virus ribonucleoprotein needs to be transported into the nucleus, a process that is dependent on cellular α-importins (Staller and Barclay, 2021). Avian influenza viruses show a preference for importin α3, but the human respiratory tract is richer in importin α7. Several point mutations, such as D701N and E627K in the PB2 subunit of the influenza A H7N9 RNA polymerase, or N319K in the nucleoprotein of influenza A virus H7N7 (Table 23.1), have been shown to shift the preference from importin α3 to importin α7 (Gabriel et al., 2011). The N319K substitution has also been reported in human cases of H5N1, supporting the role of these mutations in viral zoonosis

Table 23.1 Mutations Involved in Influenza Virus Cross-Species Transmission From Birds to Humans

Gene	Mutations	Strain	Effect
HA	E190D, G225D	H1N1	Shift in the preference from α2,3 SA (birds) to α2,6 SA (human upper respiratory tract) to promote viral attachment
	Q226L, G228S	H2N2, H3N2	
	N154S, N182K, Q192R, Q222L, S223N, G224S, Q226L, S227N	H5N1	
	K153D, S223N, G272S	H5N1	
	V186G/K, K193T, G228S	H7N9	
	V186N, N224K, G228S	H7N9	
	Q226L	H10N7	
HA	Polybasic sequence at the HA cleavage site, loss of N11	H7N7, H7N9	HA cleavage and activation by ubiquitously expressed proteases (e.g., furin), leading to expanded cellular tropism and increased pathogenicity
NA	N308S, A346V, T442A, P458S	H7N7	Shift in the preference from α2,3 SA (birds) to α2,6 SA (human upper respiratory tract) to assist viral release from cells
	I275V	H3N2, H2N2	
PB2	D701N, E627K	H7N9	Shift in the preference from importin α3 (birds) to importin α1 or α7 (human upper respiratory tract) for nuclear import of ribonucleoproteins
NP	N319K	H7N7, H5N1	
NP	R100V, F313Y/V, E53D	2009 H1N1	Evasion from human MX1
PA	S388R, A448E	H5N1	Increased transcription efficiency and host range expansion
PA	N409S	H7N9	Increased polymerase activity and efficient replication in mammals
PB2	E627K	H7N9	Interaction with human ANP32A protein, required for recruitment of viral PA and viral replication
	G590S, Q591R	2009 H1N1	

Table 23.1 Mutations Involved in Influenza Virus Cross-Species Transmission From Birds to Humans—cont'd

Gene	Mutations	Strain	Effect
M	Loss of intronic splicing silencer GAAARGARR	H5N1	Prevents binding of human TRA2A and allows M mRNA splicing in humans, increasing viral release
PB2	E627K	H7N9	Increased function at 33C, improving replication in the human upper respiratory tract relative to the avian gut

(Gabriel et al., 2005). Another cellular factor involved in nuclear import is the Myxoma Resistance Protein 1 (MX1), which is a well-characterized antiviral effector. MX1 inhibits viral ribonucleoprotein import, and it has been shown that this inhibitory effect acts more potently on avian H5N1 viruses than on human-adapted strains, because species-specific viral adaptation has led to the fixation of certain sequence changes in the viral nucleoprotein gene that allow partial evasion from this antiviral effector (Mänz et al., 2013) (Table 23.1).

Acidic nuclear phosphoprotein 32 family member A (ANP32A) is another cellular protein required for efficient replication of influenza A virus that differs between human and avian hosts (Long et al., 2016). Interaction of the viral RNA polymerase subunit PB2 with ANP32A is involved in the formation of a functional replication complex. Human ANP32A has a 33-residue deletion compared to the avian ortholog, which makes the interaction of avian influenza PB2 inefficient. This barrier is overcome by mutation E627K in PB2, which allows avian influenza polymerase activity to be supported by the shorter human ANP32. The E627K substitution was a hallmark of all influenza virus pandemics in the 20th century, including the 1918 H1N1, 1957 H2N2, and 1968 H3N2 viruses. In contrast, the 2009 H1N1 pandemic virus showed other mutations (G590S and Q591R) that probably resulted in a similar phenotype (AbuBakar et al., 2023; Long et al., 2019). Host range mutations have also been described in the viral polymerase PA subunit (Arai et al., 2020) (Table 23.1).

Additional barriers to the cross-species transmission of influenza A virus exist at the levels of gene expression, viral budding, and HA stability, among others. Several splicing regulators interact with viral mRNAs, one such regulator being human transformer-2 protein homolog alpha (huTRA2A). This protein regulates the splicing of viral gene M by binding to a specific sequence motif. When avian influenza virus infects human cells, this interaction leads to an underexpression of the viral product M2, resulting in reduced viral yield (Zhu et al., 2020). By mutating this binding motif, avian influenza viruses can increase their fitness in human cells. At the level of viral budding, a species-specific interaction may occur between the viral protein M1 and host G Protein Subunit Beta 1 (GNB1). The M1 protein migrates to the plasma membrane, where it oligomerizes and induces a curvature that promotes budding. Recruitment of M1 is dependent on GNB1, but human GNB1 does not interact with the M1 of certain avian influenza strains (Li et al., 2021). At the level of viral egress, avian influenza virions tend to aggregate at the cell surface, probably due to inefficient NA activity.

Similar to HA, the NA from bird flu shows a preference for α2,3 sialic acids but, to efficiently release egressed viral particles from the surface of human cells, the viral NA needs to process α2,6 sialic acids. As for many other viral proteins, specific mutations allow breaching this barrier (Zanin et al., 2017) (Table 23.1).

Yet another barrier to influenza A virus zoonosis has to do with the stability of the viral HA under acidic conditions (Russell et al., 2018). Influenza virus enters the cell following the endosome–lysosome pathway, where acidification activates the HA at a pH between 5.0 and 6.0 to induce fusion of the viral and endosomal membranes and release of the virion content to the cytosol. The human respiratory mucosa can however reach pH values as low as 5.5, which could trigger premature HA activation in the extracellular milieu and loss of viral infectivity. For this reason, the HA of human influenza virus has evolved to be stable at this pH and is typically activated at pH values slightly below 5.5, whereas this is not the case for avian influenza viruses, since the avian respiratory mucosa is less acidic.

5. Viral Establishment Following Cross-Species Transmission

5.1 Unique Versus Multiple and Overt Versus Cryptic Zoonoses

Because the number of potentially zoonotic viruses is much larger than the actual number of zoonotic events, the per-virus average risk of zoonosis is small. Therefore, in principle, the chance that the same virus is introduced multiple times in humans from an animal reservoir should be very small. However, some viruses have indeed caused multiple independent zoonoses, highlighting that risk varies amply among viruses, depending on factors such as ecological opportunity and host species compatibility (Woolhouse et al., 2016). For instance, Lassa virus is frequently transmitted from the *Mastomys natalensis* mouse to humans in sub-Saharan Africa, causing over 100,000 infections yearly, the majority of cases deriving from independent zoonotic events (Olayemi and Fichet-Calvet, 2020). Similarly, different lineages of MERS coronavirus have spilled over from camels probably hundreds of times (Dudas et al., 2018). In both cases, despite multiple introductions, the virus has experienced limited or no human-to-human transmission. This suggests that these viruses are unlikely to cause significant outbreaks without the appearance of specific mutations that increase human transmissibility.

In some cases, there is evidence for intermittent zoonotic events, of which most ended in rapid lineage extinction but some led to major epidemics, with or without the advent of adaptive mutations. For instance, in 2003, it was estimated that 13% of Chinese animal traders and >50% of traders specialized in civets had high levels of immunoglobulin G to SARS coronavirus, suggesting very frequent exposure to SARS-related viruses (Centers for Disease Control and Prevention (CDC), 2003). Indeed, using human-bat contact estimates and human seroprevalence data, it has been estimated that over 60,000 people are infected annually with SARS-related viruses in Southeast Asia (Sánchez et al., 2022).

HIV underwent multiple zoonotic events originating in wild primates, and the resulting four major HIV-1 lineages show widely different prevalence in humans (Sharp and Hahn, 2011). Moreover, the virus remained undetected for decades. The HIV-1 main group M presumably jumped from chimpanzees in Southeast Cameroon after which the first human-to-human transmission events occurred mainly in Kinshasa around the 1920s (Faria et al., 2014; Keele et al., 2006), although the first AIDS

cases were not reported until 1981 in the United States after the virus had diversified extensively in Central Africa and entered the Americas through Haiti (Worobey et al., 2016).

Discrepancies between the dates of viral discovery and the actual cross-transmission events have also been noticed for coronaviruses. Molecular clock analyses suggest that some human and closely related animal coronaviruses have diverged from their last known common ancestor decades or centuries ago. For instance, the most recent common ancestor of SARS coronavirus and the closest SARS-related bat coronavirus was traced back to the 1990s, whereas the SARS epidemic was declared in 2002 (Lau et al., 2015). Human coronavirus HKU1 was discovered in a pneumonia patient in 2005, but the virus might have been diversifying since the 1950s (Al-Khannaq et al., 2016). These discrepancies might be due to viral evolution in an intermediate species from which viral sequences are missing, such as palm civets for SARS coronavirus. Alternatively, the virus might have been circulating in humans inadvertently.

5.2 Selection for Increased Transmissibility in Emerging Viruses

Arguably, the most important selective pressure acting on a virus that has recently undergone a host jump is the ability to achieve efficient transmission in the new host. Many zoonotic viruses cause spillover human infections but have limited transmissibility among humans. Examples include many rodent arenaviruses (Junin virus, Lassa virus, Lujo virus), hantaviruses (Hantaan virus, Sin Nombre virus, Andes virus), tick-borne nairoviruses (Crimean-Congo hemorrhagic fever virus), flaviriruses (West Nile virus, Usutu virus, and Japanese encephalitis virus), coronaviruses (MERS coronavirus), togaviruses (Semliki virus, Ross River virus), orthomyxoviruses (influenza A virus H7N9 and H5N1, Oz virus), filoviruses (Marburg virus, Ebola virus), and rhabdoviruses (Bas-Congo virus, rabies virus), among others (Woolhouse et al., 2016). A genetic variant that improves human-to-human transmissibility might emerge in wildlife animal reservoirs, in intermediate hosts, or, less likely, during viral replication in human spillover infections. Alternatively, successful transmission in human populations might result solely from ecological or societal changes, without the need for a particular genetic variant of the virus to emerge (Figs. 23.2 and 23.4).

Regardless of the trigger leading to an outbreak, increased human transmissibility might be one of the first traits to evolve. There have been tens of Ebola outbreaks since 1976 with large differences in size. During the largest outbreak, in 2013−16, the virus acquired a mutation in the receptor-binding domain of the glycoprotein (A82V) that increased transmission in humans, probably by increasing affinity for the human homolog of the viral receptor NPC1 (Diehl et al., 2016). In SARS-CoV-2, the spike mutation D614G emerged recurrently and became dominant as early as February 2020, only 3 months after the first reported COVID-19 cases. This mutation did not increase disease severity, but it enhanced viral replication in the upper human respiratory tract and the stability of virions, therefore increasing viral shedding and transmission opportunities (Korber et al., 2020; Plante et al., 2021). These stages of the COVID-19 pandemic were characterized by gradual viral diversification and selection of several mutations in the viral spike but also in other regions of the genome such as, for instance, the nuclear localization signals of the viral nucleoprotein (Rochman et al., 2021). By optimizing receptor binding, entry, or nuclear import, among other processes, these mutations resulted in increased viral yields, shedding, and transmission.

The successive SARS-CoV-2 variants that became dominant showed increasing transmissibility, as shown by estimates of the basic reproductive number (Du et al., 2022a), the Omicron variant being one

of the most transmissible human viruses today. Omicron provides another illustrative example in which the first evolutionary steps of an emergent virus are largely missed. When the Omicron variant was first declared in South Africa in November 2021, an unusually high number of mutations were found, particularly in the Spike protein, suggesting that this virus had been diverging from other SARS-CoV-2 variants for at least 1 year. This is particularly surprising given the unprecedented sequencing effort carried out during the COVID-19 pandemic. Several hypotheses were put forward to explain this cryptic behavior (Du et al., 2022b). First, Omicron might have circulated exclusively in low-income countries in which sequencing was infrequent, which seems unlikely given the rapid global spread of the virus. Second, the variant could have emerged from a long-COVID case, in which chronic viral replication could have allowed for the accumulation of mutations, particularly Spike mutations conferring antibody evasion (Chaguza et al., 2023). Third, the virus might have jumped to another host species and then spilled back to humans, the candidate alternate species being a rodent (Wei et al., 2021).

Selection for increased transmissibility has also been reported in emerging arboviruses. A well-studied example is Chikungunya virus. The first cases were reported in Tanzania in the 1950s, followed by sporadic outbreaks in Africa and Southeast Asia in the following decades. Throughout this period, transmission was driven essentially by the *Aedes aegypti* mosquito. However, the geographic range of the virus has recently expanded considerably, including India, Europe, and the Americas. A key factor responsible for this expansion is that the virus acquired the ability to infect *Aedes albopictus*, a geographically widespread mosquito with a biting preference for humans. This shift has been attributed to a single amino acid substitution (A226 V) in the viral envelope glycoprotein E1 (Higgs and Vanlandingham, 2015). It is noteworthy that this specific mutation has appeared at least three times independently in different regions, demonstrating a strongly positive effect on viral fitness (Hakim et al., 2022).

Changes in the envelope glycoprotein of other arboviruses have also resulted in increased viral transmission and geographic range. This was the case of Venezuelan equine encephalitis virus, which has evolved toward higher viremia in equids and increased infection of the *Aedes taeniorhynchus* vector (Marklewitz and Junglen, 2019). In West Nile virus, substitution V159A in the envelope protein has been associated with reduced incubation times in *Culex* mosquitoes and accelerated viral spread in North America (Moudy et al., 2007). Aside from viral envelope proteins, mechanisms leading to increased vector-mediated transmission have been identified in the Zika virus NS1 protein, a secreted protein that helps the virus overcome the mosquito midgut immune barrier (Liu et al., 2016). The A188V mutation increases NS1 secretion and is believed to have contributed to the infection of *Aedes aeypti* mosquitoes and the spread of Zika virus in South America (Liu et al., 2017).

Another process that can promote viral transmissibility and increase the chances of global viral dissemination is a change in the transmission route. In general, the transmission route is a relatively conserved feature in viruses, but there are exceptions. This has been suggested for monkeypox virus, which is one of several zoonotic poxviruses (Li et al., 2022). Historical outbreaks of monkeypox were largely restricted to central Africa and originated from cross-species transmission events from wild mammals, possibly rodents, with self-limited chains of transmission between humans. However, the clinical symptoms of the 2022 outbreak differed from previous ones, since patients primarily presented pustular rashes in the genital area often followed by more disseminated systemic rashes. Together with the strong epidemiological bias toward men who have sex with men, these observations suggested sexual transmission of the virus, in contrast to previous monkeypox outbreaks. Independent of the

possible shift in transmission route, cessation of immunization campaigns against the related smallpox virus probably contributed to the global spread of monkeypox (Rimoin et al., 2010).

5.3 Longer-Term Evolution of Emerging Viruses: Virulence and Immune Escape

Hosts that recover from infection are protected from reinfection for a period that varies from months to lifelong immunity. This can exert a major selective pressure on highly prevalent pathogens, favoring immune escape variants, as amply shown for influenza A virus, in which antigenic evolution is a key factor to be considered in vaccine updating (Kim et al., 2018). Because antigenic evolution becomes a relevant factor only when a virus becomes sufficiently prevalent in host populations, immune escape is not expected to be highly significant during the earliest stages of viral emergence. During the COVID-19 pandemic, antibody escape mutations began to be selected in July 2020, that is, months after the selection of mutations leading to increased viral transmissibility (Rochman et al., 2021). The scenario may be different for viruses causing chronic infections, since in this case, the virus has to face the onset of adaptive immunity in each individual, as extensively shown for HIV-1, hepatitis C virus, and other viruses.

Viral transmissibility is connected to multiple other viral traits, including the environmental stability of viral particles, the ability to evade host barriers, and the viral replication compartment, among others. A major related trait is the ability of the virus to replicate to high levels in the host which, in turn, is expected to increase virulence. On the other hand, high virulence can reduce transmission opportunities by reducing the social interactions or life expectancy of the host. Hence, as emerging viruses evolve, virulence should be adjusted in response to these fitness trade-offs (Geoghegan and Holmes, 2018; Visher et al., 2021). Optimal virulence depends not only on virus features but also on host factors such as population density and immune status, as well as on environmental factors.

During cross-species transmission, a virus might become more virulent or less virulent in the new host compared to the previous host. Although initial virulence varies strongly and is largely unpredictable, some patterns have been suggested. Spillover infections with zoonotic viruses originating from animal species more distantly to humans (e.g., rodents) tend to be more virulent (Guth et al., 2019). When a virus is introduced in a new host that is distantly related to the reservoir host, severe symptoms can triggered by inappropriate immune responses such as cytokine storms, as shown in patients infected with avian influenza viruses (Long et al., 2019). Due to transmission costs, outbreaks of highly virulent zoonotic viruses are expected to rapidly die out or evolve toward lower virulence.

The evolution of virulence might be anticipated if its relationship to transmissibility is understood (Holmes, 2022). For instance, preferential replication of SARS-CoV-2 in the upper respiratory tract not only increases viral shedding and thus transmissibility but also reduces virulence compared to virus variants replicating preferentially in the lungs (Hui et al., 2022). Interestingly, the highly transmissible Omicron variant is highly sensitive to high temperatures, possibly contributing to preferential replication in the upper respiratory tract (Muramoto et al., 2023). Such association between viral tropism at the intrahost level, virulence, and transmissibility may apply to other respiratory viruses. For instance, preferential replication in the lower respiratory tract might explain both the poor transmissibility and pathogenicity of avian influenza viruses in humans. In contrast, high levels of virulence might evolve with little or no transmission costs in viruses with indirect transmission modes, such as arboviruses and environmentally transmitted viruses. This appears to be the case of rabbit hemorrhagic disease virus, which was released as a second biocontrol agent against European rabbits in Australia after myxoma

virus (Di Giallonardo and Holmes, 2015). In this case, virulence has increased through time probably because the virus is carried by blow flies that feed on animal carcasses, making host death advantageous for viral transmission.

6. Closing Remarks and Future Perspectives

Several important topics in viral emergence remain unanswered despite having been addressed in different studies. For instance, the influence of viral mutation and recombination rates on zoonotic risk has been studied using different statistical and phylogenetic methods but, as of today, no solid conclusions have been reached. Available data are limited and biased toward certain highly studied pathogens, and many confounders exist. Although RNA viruses are more likely to cause emerging diseases than DNA viruses, this observation might also be explained by the contribution of correlated traits such as replication in the cytoplasm, vector-borne transmission, and the presence of a viral envelope. Disentangling the roles of ecological factors and reservoir hosts is also challenging because many biases are also present at these levels. Viral discovery is highly dependent on research effort, which varies strongly across regions. Also, the common perception that certain animals such as bats are particularly likely to cause zoonoses might have been distorted by a disproportionate research focus on this group, which indeed constitutes a very abundant and diverse mammalian order.

These uncertainties can be reduced by combining descriptive work with experimental virology studies to characterize key steps of viral cross-species transmission, such as receptor use, viral entry, and other molecular details of virus−host interactions. Moreover, viral metagenomics campaigns have provided a new perspective on the virosphere, in which the bias toward pathogens and human-infective viruses has been mitigated. However, reporting viral sequences per se provides limited information about viral biology and emergence risks. Virus−host associations need to be better established, a type of information that is sometimes not available or even misleading in viral metagenomics such as, for instance, when viruses present in feces are associated with a given host but could merely be part of the diet or the microbiota. A very illustrative example is the Picobirnaviridae family. These viruses were initially defined as gastroenteritis-causing mammalian viruses, but may indeed be bacteriophages infecting bacteria present in the gut (Krishnamurthy and Wang, 2018).

Despite being a very challenging topic, there have been major recent advances in our ability to predict viral emergence. Several studies have shown that viral hosts can be predicted to a certain level based on viral sequences alone. An obvious but important remark to achieve this goal is that similar viruses behave similarly, meaning that they tend to use related hosts. Moreover, viruses infecting widely different hosts (e.g., primates vs. rodents) show subtle differences in genome composition, including dinucleotide, codon, and amino acid biases. Taking into account this information, it was possible to predict with >90% accuracy whether a virus is vector-borne based solely on sequence data (Babayan et al., 2018). The specific identity of the vector and the reservoir was less certain, but some broad taxonomical groups could be predicted with >80% accuracy. Predicting which viruses can infect humans is far more challenging because it requires a finer assessment of viral tropism, but encouraging results have been obtained for flu, a virus for which it was possible to distinguish avian influenza viruses capable of infecting humans (Li et al., 2020b). A study that considered many different viral families showed that it was possible to identify human-infective viruses with 70% accuracy (Mollentze et al., 2021).

Parallel evolution is yet another relevant process to be considered for viral emergence predictions. If certain mutations appear repeatedly or the same regions of the viral genome tend to change during independent host shifts, these sites could be used for surveilling viral populations in reservoirs or assessing the risk that human infections give rise to large outbreaks. For example, the rabies virus lineage circulating in dogs has undergone multiple host jumps, and it was found that the same two amino acid replacements in the nucleoprotein and viral polymerase occurred independently in two viral lineages transmitted from dogs to ferret-badgers, strongly suggesting host-specific adaptation (Troupin et al., 2016). In West Nile virus, the same mutation in the helicase protein has been found repeatedly in different outbreaks (Brault et al., 2007). Such deterministic evolution has been shown for other viruses as well, including major human pathogens such as influenza A virus, HIV-1 and SARS-CoV-2 (Gutierrez et al., 2019; Jones et al., 2021; Valesano et al., 2021), as well as during independent zoonotic events caused by viruses at risk of emergence, such as porcine deltacoronavirus (Lednicky et al., 2021).

In sum, the field of viral emergence might strongly benefit from a cross-talk between viral metagenomics, ecology, applied statistics, and experimental virology. This could help us achieve a less biased, more comprehensive, and more mechanistic understanding of the complex processes leading to viral emergence and, ultimately, to increase our ability to predict and manage emergent viruses.

Acknowledgments

The author is funded by the European Research Council (ERC) through the Advanced Grant 101019724 (EVADER), by the Spanish Ministerio Ciencia e Innovación (MICINN) through grant PID2020-118602RB-I00 (ZooVir), and by the Generalitat Valenciana through grant CIAICO/2022/110.

References

Abraham, J., Kwong, J.A., Albariño, C.G., Lu, J.G., Radoshitzky, S.R., Salazar-Bravo, J., Farzan, M., Spiropoulou, C.F., Choe, H., 2009. Host-species transferrin receptor 1 orthologs are cellular receptors for nonpathogenic new world clade B arenaviruses. PLoS Pathog. 5, e1000358. https://doi.org/10.1371/journal.ppat.1000358.

AbuBakar, U., Amrani, L., Kamarulzaman, F.A., Karsani, S.A., Hassandarvish, P., Khairat, J.E., 2023. Avian influenza virus tropism in humans. Viruses 15, 833. https://doi.org/10.3390/v15040833.

Albery, G.F., Becker, D.J., 2021. Fast-lived hosts and zoonotic risk. Trends Parasitol. 37, 117−129. https://doi.org/10.1016/j.pt.2020.10.012.

Albery, G.F., Becker, D.J., Brierley, L., Brook, C.E., Christofferson, R.C., Cohen, L.E., Dallas, T.A., Eskew, E.A., Fagre, A., Farrell, M.J., Glennon, E., Guth, S., Joseph, M.B., Mollentze, N., Neely, B.A., Poisot, T., Rasmussen, A.L., Ryan, S.J., Seifert, S., Sjodin, A.R., Sorrell, E.M., Carlson, C.J., 2021. The science of the host-virus network. Nat. Microbiol. 6, 1483−1492. https://doi.org/10.1038/s41564-021-00999-5.

Albery, G.F., Carlson, C.J., Cohen, L.E., Eskew, E.A., Gibb, R., Ryan, S.J., Sweeny, A.R., Becker, D.J., 2022. Urban-adapted mammal species have more known pathogens. Nat. Ecol. Evol. 6, 794−801. https://doi.org/10.1038/s41559-022-01723-0.

Albery, G.F., Eskew, E.A., Ross, N., Olival, K.J., 2020. Predicting the global mammalian viral sharing network using phylogeography. Nat. Commun. 11, 2260. https://doi.org/10.1038/s41467-020-16153-4.

Alexander, D.J., 2000. A review of avian influenza in different bird species. Vet. Microbiol. 74, 3–13. https://doi.org/10.1016/s0378-1135(00)00160-7.

Al-Khannaq, M.N., Ng, K.T., Oong, X.Y., Pang, Y.K., Takebe, Y., Chook, J.B., Hanafi, N.S., Kamarulzaman, A., Tee, K.K., 2016. Molecular epidemiology and evolutionary histories of human coronavirus OC43 and HKU1 among patients with upper respiratory tract infections in Kuala Lumpur, Malaysia. Virol. J. 13, 33. https://doi.org/10.1186/s12985-016-0488-4.

Allen, T., Murray, K.A., Zambrana-Torrelio, C., Morse, S.S., Rondinini, C., Di Marco, M., Breit, N., Olival, K.J., Daszak, P., 2017. Global hotspots and correlates of emerging zoonotic diseases. Nat. Commun. 8, 1124. https://doi.org/10.1038/s41467-017-00923-8.

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., 2020. The proximal origin of SARS-CoV-2. Nat. Med. 26, 450–452. https://doi.org/10.1038/s41591-020-0820-9.

Arai, Y., Kawashita, N., Elgendy, E.M., Ibrahim, M.S., Daidoji, T., Ono, T., Takagi, T., Nakaya, T., Matsumoto, K., Watanabe, Y., 2020. PA mutations inherited during viral evolution act cooperatively to increase replication of contemporary H5N1 influenza virus with an expanded host range. J. Virol. 95, e01582–20. https://doi.org/10.1128/JVI.01582-20.

Babayan, S.A., Orton, R.J., Streicker, D.G., 2018. Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. Science 362, 577–580. https://doi.org/10.1126/science.aap9072.

Banerjee, A., Doxey, A.C., Mossman, K., Irving, A.T., 2021. Unraveling the zoonotic origin and transmission of SARS-CoV-2. Trends Ecol. Evol. 36, 180–184. https://doi.org/10.1016/j.tree.2020.12.002.

Bashor, L., Gagne, R.B., Bosco-Lauth, A.M., Bowen, R.A., Stenglein, M., VandeWoude, S., 2021. SARS-CoV-2 evolution in animals suggests mechanisms for rapid variant selection. Proc. Natl. Acad. Sci. U.S.A. 118, e2105253118. https://doi.org/10.1073/pnas.2105253118.

Brault, A.C., Huang, C.Y.-H., Langevin, S.A., Kinney, R.M., Bowen, R.A., Ramey, W.N., Panella, N.A., Holmes, E.C., Powers, A.M., Miller, B.R., 2007. A single positively selected West Nile viral mutation confers increased virogenesis in American crows. Nat. Genet. 39, 1162–1166. https://doi.org/10.1038/ng2097.

Brook, C.E., Dobson, A.P., 2015. Bats as "special" reservoirs for emerging zoonotic pathogens. Trends Microbiol. 23, 172–180. https://doi.org/10.1016/j.tim.2014.12.004.

Carlson, C.J., Zipfel, C.M., Garnier, R., Bansal, S., 2019. Global estimates of mammalian viral diversity accounting for host sharing. Nat. Ecol. Evol. 3, 1070–1075. https://doi.org/10.1038/s41559-019-0910-6.

Centers for Disease Control and Prevention (CDC), 2003. Prevalence of IgG antibody to SARS-associated coronavirus in animal traders–Guangdong Province, China, 2003. MMWR Morb. Mortal. Wkly. Rep. 52, 986–987.

Chaguza, C., Hahn, A.M., Petrone, M.E., Zhou, S., Ferguson, D., Breban, M.I., Pham, K., Peña-Hernández, M.A., Castaldi, C., Hill, V., , Yale SARS-CoV-2 Genomic Surveillance Initiative, Schulz, W., Swanstrom, R.I., Roberts, S.C., Grubaugh, N.D., 2023. Accelerated SARS-CoV-2 intrahost evolution leading to distinct genotypes during chronic infection. Cell Rep. Med. 4, 100943. https://doi.org/10.1016/j.xcrm.2023.100943.

Cui, X., Fan, K., Liang, X., Gong, W., Chen, W., He, B., Chen, X., Wang, H., Wang, X., Zhang, P., Lu, X., Chen, R., Lin, K., Liu, J., Zhai, J., Liu, D.X., Shan, F., Li, Y., Chen, R.A., Meng, H., Li, X., Mi, S., Jiang, J., Zhou, N., Chen, Z., Zou, J.-J., Ge, D., Yang, Q., He, K., Chen, T., Wu, Y.-J., Lu, H., Irwin, D.M., Shen, X., Hu, Y., Lu, X., Ding, C., Guan, Y., Tu, C., Shen, Y., 2023. Virus diversity, wildlife-domestic animal circulation and potential zoonotic viruses of small mammals, pangolins and zoo animals. Nat. Commun. 14, 2488. https://doi.org/10.1038/s41467-023-38202-4.

Di Giallonardo, F., Holmes, E.C., 2015. Viral biocontrol: grand experiments in disease emergence and evolution. Trends Microbiol. 23, 83–90. https://doi.org/10.1016/j.tim.2014.10.004.

Diehl, W.E., Lin, A.E., Grubaugh, N.D., Carvalho, L.M., Kim, K., Kyawe, P.P., McCauley, S.M., Donnard, E., Kucukural, A., McDonel, P., Schaffner, S.F., Garber, M., Rambaut, A., Andersen, K.G., Sabeti, P.C., Luban, J., 2016. Ebola virus glycoprotein with increased infectivity dominated the 2013-2016 epidemic. Cell 167, 1088−1098. https://doi.org/10.1016/j.cell.2016.10.014.

Drexler, J.F., Corman, V.M., Müller, M.A., Maganga, G.D., Vallo, P., Binger, T., Gloza-Rausch, F., Cottontail, V.M., Rasche, A., Yordanov, S., Seebens, A., Knörnschild, M., Oppong, S., Adu Sarkodie, Y., Pongombo, C., Lukashev, A.N., Schmidt-Chanasit, J., Stöcker, A., Carneiro, A.J.B., Erbar, S., Maisner, A., Fronhoffs, F., Buettner, R., Kalko, E.K.V., Kruppa, T., Franke, C.R., Kallies, R., Yandoko, E.R.N., Herrler, G., Reusken, C., Hassanin, A., Krüger, D.H., Matthee, S., Ulrich, R.G., Leroy, E.M., Drosten, C., 2012. Bats host major mammalian paramyxoviruses. Nat. Commun. 3, 796. https://doi.org/10.1038/ncomms1796.

Du, P., Gao, G.F., Wang, Q., 2022a. The mysterious origins of the Omicron variant of SARS-CoV-2. Innov. Camb. Mass 3, 100206. https://doi.org/10.1016/j.xinn.2022.100206.

Du, Z., Hong, H., Wang, S., Ma, L., Liu, C., Bai, Y., Adam, D.C., Tian, L., Wang, L., Lau, E.H.Y., Cowling, B.J., 2022b. Reproduction number of the omicron variant triples that of the delta variant. Viruses 14, 821. https://doi.org/10.3390/v14040821.

Dudas, G., Carvalho, L.M., Rambaut, A., Bedford, T., 2018. MERS-CoV spillover at the camel-human interface. Elife 7, e31257. https://doi.org/10.7554/eLife.31257.

Düx, A., Lequime, S., Patrono, L.V., Vrancken, B., Boral, S., Gogarten, J.F., Hilbig, A., Horst, D., Merkel, K., Prepoint, B., Santibanez, S., Schlotterbeck, J., Suchard, M.A., Ulrich, M., Widulin, N., Mankertz, A., Leendertz, F.H., Harper, K., Schnalke, T., Lemey, P., Calvignac-Spencer, S., 2020. Measles virus and rinderpest virus divergence dated to the sixth century BCE. Science 368, 1367−1370. https://doi.org/10.1126/science.aba9411.

Eierhoff, T., Hrincius, E.R., Rescher, U., Ludwig, S., Ehrhardt, C., 2010. The epidermal growth factor receptor (EGFR) promotes uptake of influenza A viruses (IAV) into host cells. PLoS Pathog. 6, e1001099. https://doi.org/10.1371/journal.ppat.1001099.

Escudero-Pérez, B., Lalande, A., Mathieu, C., Lawrence, P., 2023. Host-pathogen interactions influencing zoonotic spillover potential and transmission in humans. Viruses 15, 599. https://doi.org/10.3390/v15030599.

Faria, N.R., Rambaut, A., Suchard, M.A., Baele, G., Bedford, T., Ward, M.J., Tatem, A.J., Sousa, J.D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O.G., Lemey, P., 2014. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science 346, 56−61. https://doi.org/10.1126/science.1256739.

Fogarty, R., Halpin, K., Hyatt, A.D., Daszak, P., Mungall, B.A., 2008. Henipavirus susceptibility to environmental variables. Virus Res. 132, 140−144. https://doi.org/10.1016/j.virusres.2007.11.010.

Forni, D., Cagliani, R., Clerici, M., Sironi, M., 2022. Disease-causing human viruses: novelty and legacy. Trends Microbiol. 30, 1232−1242. https://doi.org/10.1016/j.tim.2022.07.002.

Fujioka, Y., Nishide, S., Ose, T., Suzuki, T., Kato, I., Fukuhara, H., Fujioka, M., Horiuchi, K., Satoh, A.O., Nepal, P., Kashiwagi, S., Wang, J., Horiguchi, M., Sato, Y., Paudel, S., Nanbo, A., Miyazaki, T., Hasegawa, H., Maenaka, K., Ohba, Y., 2018. A sialylated voltage-dependent Ca2+ channel binds hemagglutinin and mediates influenza A virus entry into mammalian cells. Cell Host Microbe 23, 809−818.e5. https://doi.org/10.1016/j.chom.2018.04.015.

Fukuhara, H., Ito, Y., Sako, M., Kajikawa, M., Yoshida, K., Seki, F., Mwaba, M.H., Hashiguchi, T., Higashibata, M.-A., Ose, T., Kuroki, K., Takeda, M., Maenaka, K., 2019. Specificity of *Morbillivirus* hemagglutinins to recognize SLAM of different species. Viruses 11, 761. https://doi.org/10.3390/v11080761.

Gabriel, G., Dauber, B., Wolff, T., Planz, O., Klenk, H.-D., Stech, J., 2005. The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. Proc. Natl. Acad. Sci. U.S.A. 102, 18590−18595. https://doi.org/10.1073/pnas.0507415102.

Gabriel, G., Klingel, K., Otte, A., Thiele, S., Hudjetz, B., Arman-Kalcek, G., Sauter, M., Shmidt, T., Rother, F., Baumgarte, S., Keiner, B., Hartmann, E., Bader, M., Brownlee, G.G., Fodor, E., Klenk, H.-D., 2011. Differential use of importin-α isoforms governs cell tropism and host adaptation of influenza virus. Nat. Commun. 2, 156. https://doi.org/10.1038/ncomms1158.

Geng, R., Zhou, P., 2021. Severe acute respiratory syndrome (SARS) related coronavirus in bats. Anim. Dis. 1, 4. https://doi.org/10.1186/s44149-021-00004-w.

Geoghegan, J.L., Duchêne, S., Holmes, E.C., 2017. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. PLoS Pathog. 13, e1006215. https://doi.org/10.1371/journal.ppat.1006215.

Geoghegan, J.L., Holmes, E.C., 2018. The phylogenomics of evolving virus virulence. Nat. Rev. Genet. 19, 756–769. https://doi.org/10.1038/s41576-018-0055-5.

Geoghegan, J.L., Senior, A.M., Di Giallonardo, F., Holmes, E.C., 2016. Virological factors that increase the transmissibility of emerging human viruses. Proc. Natl. Acad. Sci. U.S.A. 113, 4170–4175. https://doi.org/10.1073/pnas.1521582113.

Gibbs, A.J., Armstrong, J.S., Downie, J.C., 2009. From where did the 2009 "swine-origin" influenza A virus (H1N1) emerge? Virol. J. 6, 207. https://doi.org/10.1186/1743-422X-6-207.

Giotis, E.S., Carnell, G., Young, E.F., Ghanny, S., Soteropoulos, P., Wang, L.-F., Barclay, W.S., Skinner, M.A., Temperton, N., 2019. Entry of the bat influenza H17N10 virus into mammalian cells is enabled by the MHC class II HLA-DR receptor. Nat. Microbiol. 4, 2035–2038. https://doi.org/10.1038/s41564-019-0517-3.

Glennon, E.E., Restif, O., Sbarbaro, S.R., Garnier, R., Cunningham, A.A., Suu-Ire, R.D., Osei-Amponsah, R., Wood, J.L.N., Peel, A.J., 2018. Domesticated animals as hosts of henipaviruses and filoviruses: a systematic review. Vet. J. Lond. Engl. 233, 25–34. https://doi.org/10.1016/j.tvjl.2017.12.024.

Glud, H.A., George, S., Skovgaard, K., Larsen, L.E., 2021. Zoonotic and reverse zoonotic transmission of viruses between humans and pigs. APMIS Acta Pathol. Microbiol. Immunol. Scand. 129, 675–693. https://doi.org/10.1111/apm.13178.

Graham, R.L., Baric, R.S., 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. 84, 3134–3146. https://doi.org/10.1128/JVI.01394-09.

Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., Butt, K.M., Wong, K.L., Chan, K.W., Lim, W., Shortridge, K.F., Yuen, K.Y., Peiris, J.S.M., Poon, L.L.M., 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302, 276–278. https://doi.org/10.1126/science.1087139.

Guth, S., Visher, E., Boots, M., Brook, C.E., 2019. Host phylogenetic distance drives trends in virus virulence and transmissibility across the animal-human interface. Philos. Trans. R. Soc. Lond. B Biol. Sci. 374, 20190296. https://doi.org/10.1098/rstb.2019.0296.

Gutierrez, B., Escalera-Zamudio, M., Pybus, O.G., 2019. Parallel molecular evolution and adaptation in viruses. Curr. Opin. Virol. 34, 90–96. https://doi.org/10.1016/j.coviro.2018.12.006.

Hakim, M.S., Annisa, L., Gazali, F.M., Aman, A.T., 2022. The origin and continuing adaptive evolution of chikungunya virus. Arch. Virol. 167, 2443–2455. https://doi.org/10.1007/s00705-022-05570-z.

Han, H.-J., Yu, H., Yu, X.-J., 2016. Evidence for zoonotic origins of Middle East respiratory syndrome coronavirus. J. Gen. Virol. 97, 274–280. https://doi.org/10.1099/jgv.0.000342.

He, W.-T., Hou, X., Zhao, J., Sun, J., He, H., Si, W., Wang, J., Jiang, Z., Yan, Z., Xing, G., Lu, M., Suchard, M.A., Ji, X., Gong, W., He, B., Li, J., Lemey, P., Guo, D., Tu, C., Holmes, E.C., Shi, M., Su, S., 2022. Virome characterization of game animals in China reveals a spectrum of emerging pathogens. Cell 185, 1117–1129.e8. https://doi.org/10.1016/j.cell.2022.02.014.

Higgs, S., Vanlandingham, D., 2015. Chikungunya virus and its mosquito vectors. Vector Borne Zoonotic Dis. 15, 231–240. https://doi.org/10.1089/vbz.2014.1745.

Holmes, E.C., 2022. The ecology of viral emergence. Annu. Rev. Virol. 9, 173−192. https://doi.org/10.1146/annurev-virology-100120-015057.

Hui, K.P.Y., Ho, J.C.W., Cheung, M.-C., Ng, K.-C., Ching, R.H.H., Lai, K.-L., Kam, T.T., Gu, H., Sit, K.-Y., Hsin, M.K.Y., Au, T.W.K., Poon, L.L.M., Peiris, M., Nicholls, J.M., Chan, M.C.W., 2022. SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. Nature 603, 715−720. https://doi.org/10.1038/s41586-022-04479-6.

Islam, M.S., Sazzad, H.M.S., Satter, S.M., Sultana, S., Hossain, M.J., Hasan, M., Rahman, M., Campbell, S., Cannon, D.L., Ströher, U., Daszak, P., Luby, S.P., Gurley, E.S., 2016. Nipah virus transmission from bats to humans associated with drinking traditional liquor made from date palm sap, Bangladesh, 2011-2014. Emerg. Infect. Dis. 22, 664−670. https://doi.org/10.3201/eid2204.151747.

Jaimes, J.A., Millet, J.K., Stout, A.E., André, N.M., Whittaker, G.R., 2020. A tale of two viruses: the distinct spike glycoproteins of feline coronaviruses. Viruses 12, 83. https://doi.org/10.3390/v12010083.

Jeffers, S.A., Hemmila, E.M., Holmes, K.V., 2006. Human coronavirus 229E can use CD209L (L-SIGN) to enter cells. Adv. Exp. Med. Biol. 581, 265−269. https://doi.org/10.1007/978-0-387-33012-9_44.

Jeffers, S.A., Tusell, S.M., Gillim-Ross, L., Hemmila, E.M., Achenbach, J.E., Babcock, G.J., Thomas, W.D., Thackray, L.B., Young, M.D., Mason, R.J., Ambrosino, D.M., Wentworth, D.E., Demartini, J.C., Holmes, K.V., 2004. CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. Proc. Natl. Acad. Sci. U.S.A. 101, 15748−15753. https://doi.org/10.1073/pnas.0403812101.

Johnson, C.K., Hitchens, P.L., Pandit, P.S., Rushmore, J., Evans, T.S., Young, C.C.W., Doyle, M.M., 2020. Global shifts in mammalian population trends reveal key predictors of virus spillover risk. Proc. Biol. Sci. 287, 20192736. https://doi.org/10.1098/rspb.2019.2736.

Jones, J.E., Le Sage, V., Padovani, G.H., Calderon, M., Wright, E.S., Lakdawala, S.S., 2021. Parallel evolution between genomic segments of seasonal human influenza viruses reveals RNA-RNA relationships. Elife 10, e66525. https://doi.org/10.7554/eLife.66525.

Kallio-Kokko, H., Uzcategui, N., Vapalahti, O., Vaheri, A., 2005. Viral zoonoses in Europe. FEMS Microbiol. Rev. 29, 1051−1077. https://doi.org/10.1016/j.femsre.2005.04.012.

Kang, L., He, G., Sharp, A.K., Wang, X., Brown, A.M., Michalak, P., Weger-Lucarelli, J., 2021. A selective sweep in the Spike gene has driven SARS-CoV-2 human adaptation. Cell 184, 4392−4400.e4. https://doi.org/10.1016/j.cell.2021.07.007.

Karakus, U., Pohl, M.O., Stertz, S., 2020. Breaking the convention: sialoglycan variants, coreceptors, and alternative receptors for influenza A virus entry. J. Virol. 94, e01357−19. https://doi.org/10.1128/JVI.01357-19.

Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., Loul, S., Ngole, E.M., Bienvenue, Y., Delaporte, E., Brookfield, J.F.Y., Sharp, P.M., Shaw, G.M., Peeters, M., Hahn, B.H., 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313, 523−526. https://doi.org/10.1126/science.1126531.

Keesing, F., Belden, L.K., Daszak, P., Dobson, A., Harvell, C.D., Holt, R.D., Hudson, P., Jolles, A., Jones, K.E., Mitchell, C.E., Myers, S.S., Bogich, T., Ostfeld, R.S., 2010. Impacts of biodiversity on the emergence and transmission of infectious diseases. Nature 468, 647−652. https://doi.org/10.1038/nature09575.

Kerr, S.A., Jackson, E.L., Lungu, O.I., Meyer, A.G., Demogines, A., Ellington, A.D., Georgiou, G., Wilke, C.O., Sawyer, S.L., 2015. Computational and functional analysis of the virus-receptor interface reveals host range trade-offs in New World arenaviruses. J. Virol. 89, 11643−11653. https://doi.org/10.1128/JVI.01408-15.

Kim, H., Webster, R.G., Webby, R.J., 2018. Influenza virus: dealing with a drifting and shifting pathogen. Viral Immunol. 31, 174−183. https://doi.org/10.1089/vim.2017.0141.

Kohl, C., Kurth, A., 2014. European bats as carriers of viruses with zoonotic potential. Viruses 6, 3110−3128. https://doi.org/10.3390/v6083110.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de

Silva, T.I., Sheffield COVID-19 Genomics Group, McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell 182, 812−827.e19. https://doi.org/10.1016/j.cell.2020.06.043.

Kreuder Johnson, C., Hitchens, P.L., Smiley Evans, T., Goldstein, T., Thomas, K., Clements, A., Joly, D.O., Wolfe, N.D., Daszak, P., Karesh, W.B., Mazet, J.K., 2015. Spillover and pandemic properties of zoonotic viruses with high host plasticity. Sci. Rep. 5, 14830. https://doi.org/10.1038/srep14830.

Krishnamurthy, S.R., Wang, D., 2018. Extensive conservation of prokaryotic ribosomal binding sites in known and novel *Picobirnaviruses*. Virology 516, 108−114. https://doi.org/10.1016/j.virol.2018.01.006.

Lau, S.K.P., Feng, Y., Chen, H., Luk, H.K.H., Yang, W.-H., Li, K.S.M., Zhang, Y.-Z., Huang, Y., Song, Z.-Z., Chow, W.-N., Fan, R.Y.Y., Ahmed, S.S., Yeung, H.C., Lam, C.S.F., Cai, J.-P., Wong, S.S.Y., Chan, J.F.W., Yuen, K.-Y., Zhang, H.-L., Woo, P.C.Y., 2015. Severe acute respiratory syndrome (SARS) coronavirus ORF8 protein is acquired from SARS-related coronavirus from greater horseshoe bats through recombination. J. Virol. 89, 10532−10547. https://doi.org/10.1128/JVI.01048-15.

Lednicky, J.A., Tagliamonte, M.S., White, S.K., Elbadry, M.A., Alam, M.M., Stephenson, C.J., Bonny, T.S., Loeb, J.C., Telisma, T., Chavannes, S., Ostrov, D.A., Mavian, C., Beau De Rochars, V.M., Salemi, M., Morris, J.G., 2021. Independent infections of porcine deltacoronavirus among Haitian children. Nature 600, 133−137. https://doi.org/10.1038/s41586-021-04111-z.

Lee, I., Bos, S., Li, G., Wang, S., Gadea, G., Desprès, P., Zhao, R.Y., 2018. Probing molecular insights into Zika Virus−Host interactions. Viruses 10. https://doi.org/10.3390/v10050233.

Li, F., Liu, J., Yang, J., Sun, H., Jiang, Z., Wang, C., Zhang, X., Yu, Y., Zhao, C., Pu, J., Sun, Y., Chang, K.-C., Liu, J., Sun, H., 2021. H9N2 virus-derived M1 protein promotes H5N6 virus release in mammalian cells: mechanism of avian influenza virus inter-species infection in humans. PLoS Pathog. 17, e1010098. https://doi.org/10.1371/journal.ppat.1010098.

Li, H., Zhang, H., Ding, K., Wang, X.-H., Sun, G.-Y., Liu, Z.-X., Luo, Y., 2022. The evolving epidemiology of monkeypox virus. Cytokine Growth Factor Rev. 68, 1−12. https://doi.org/10.1016/j.cytogfr.2022.10.002.

Li, J., Zhang, S., Li, B., Hu, Y., Kang, X.-P., Wu, X.-Y., Huang, M.-T., Li, Y.-C., Zhao, Z.-P., Qin, C.-F., Jiang, T., 2020a. Machine learning methods for predicting human-adaptive influenza A viruses based on viral nucleotide compositions. Mol. Biol. Evol. 37, 1224−1236. https://doi.org/10.1093/molbev/msz276.

Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.-K., Huang, I.-C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W.A., Guan, Y., Choe, H., Farzan, M., 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. 24, 1634−1643. https://doi.org/10.1038/sj.emboj.7600640.

Li, X., Giorgi, E.E., Marichannegowda, M.H., Foley, B., Xiao, C., Kong, X.-P., Chen, Y., Gnanakaran, S., Korber, B., Gao, F., 2020b. Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci. Adv. 6, eabb9153. https://doi.org/10.1126/sciadv.abb9153.

Liu, J., Liu, Y., Nie, K., Du, S., Qiu, J., Pang, X., Wang, P., Cheng, G., 2016. Flavivirus NS1 protein in infected host sera enhances viral acquisition by mosquitoes. Nat. Microbiol. 1, 16087. https://doi.org/10.1038/nmicrobiol.2016.87.

Liu, Y., Liu, J., Du, S., Shan, C., Nie, K., Zhang, R., Li, X.-F., Zhang, R., Wang, T., Qin, C.-F., Wang, P., Shi, P.-Y., Cheng, G., 2017. Evolutionary enhancement of Zika virus infectivity in Aedes aegypti mosquitoes. Nature 545, 482−486. https://doi.org/10.1038/nature22365.

Long, J.S., Giotis, E.S., Moncorgé, O., Frise, R., Mistry, B., James, J., Morisson, M., Iqbal, M., Vignal, A., Skinner, M.A., Barclay, W.S., 2016. Species difference in ANP32A underlies influenza A virus polymerase host restriction. Nature 529, 101−104. https://doi.org/10.1038/nature16474.

Long, J.S., Mistry, B., Haslam, S.M., Barclay, W.S., 2019. Host and viral determinants of influenza A virus species specificity. Nat. Rev. Microbiol. 17, 67−81. https://doi.org/10.1038/s41579-018-0115-z.

Lu, G., Wang, Q., Gao, G.F., 2015. Bat-to-human: spike features determining "host jump" of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends Microbiol. 23, 468−478. https://doi.org/10.1016/j.tim.2015.06.003.

Luby, S.P., Rahman, M., Hossain, M.J., Blum, L.S., Husain, M.M., Gurley, E., Khan, R., Ahmed, B.-N., Rahman, S., Nahar, N., Kenah, E., Comer, J.A., Ksiazek, T.G., 2006. Foodborne transmission of Nipah virus, Bangladesh. Emerg. Infect. Dis. 12, 1888−1894. https://doi.org/10.3201/eid1212.060732.

Luis, A.D., O'Shea, T.J., Hayman, D.T.S., Wood, J.L.N., Cunningham, A.A., Gilbert, A.T., Mills, J.N., Webb, C.T., 2015. Network analysis of host-virus communities in bats and rodents reveals determinants of cross-species transmission. Ecol. Lett. 18, 1153−1162. https://doi.org/10.1111/ele.12491.

Maclachlan, N.J., Zientara, S., Wilson, W.C., Richt, J.A., Savini, G., 2019. Bluetongue and epizootic hemorrhagic disease viruses: recent developments with these globally re-emerging arboviral infections of ruminants. Curr. Opin. Virol. 34, 56−62. https://doi.org/10.1016/j.coviro.2018.12.005.

Mänz, B., Dornfeld, D., Götz, V., Zell, R., Zimmermann, P., Haller, O., Kochs, G., Schwemmle, M., 2013. Pandemic influenza A viruses escape from restriction by human MxA through adaptive mutations in the nucleoprotein. PLoS Pathog. 9, e1003279. https://doi.org/10.1371/journal.ppat.1003279.

Marklewitz, M., Junglen, S., 2019. Evolutionary and ecological insights into the emergence of arthropod-borne viruses. Acta Trop. 190, 52−58. https://doi.org/10.1016/j.actatropica.2018.10.006.

Martin, C., Buckler-White, A., Wollenberg, K., Kozak, C.A., 2013. The avian XPR1 gammaretrovirus receptor is under positive selection and is disabled in bird species in contact with virus-infected wild mice. J. Virol. 87, 10094−10104. https://doi.org/10.1128/JVI.01327-13.

Meng, X.-J., 2016. Expanding host range and cross-species infection of hepatitis E virus. PLoS Pathog. 12, e1005695. https://doi.org/10.1371/journal.ppat.1005695.

Mollentze, N., Babayan, S.A., Streicker, D.G., 2021. Identifying and prioritizing potential human-infecting viruses from their genome sequences. PLoS Biol. 19, e3001390. https://doi.org/10.1371/journal.pbio.3001390.

Mollentze, N., Streicker, D.G., 2020. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. Proc. Natl. Acad. Sci. U.S.A. 117, 9423−9430. https://doi.org/10.1073/pnas.1919176117.

Mostafa, A., Abdelwhab, E.M., Mettenleiter, T.C., Pleschka, S., 2018. Zoonotic potential of influenza A viruses: a comprehensive overview. Viruses 10, 497. https://doi.org/10.3390/v10090497.

Moudy, R.M., Meola, M.A., Morin, L.-L.L., Ebel, G.D., Kramer, L.D., 2007. A newly emergent genotype of West Nile virus is transmitted earlier and more efficiently by Culex mosquitoes. Am. J. Trop. Med. Hyg. 77, 365−370.

Muramoto, Y., Takahashi, S., Halfmann, P.J., Gotoh, S., Noda, T., Kawaoka, Y., 2023. Replicative capacity of SARS-CoV-2 omicron variants BA.5 and BQ.1.1 at elevated temperatures. Lancet Microbe 4, e486. https://doi.org/10.1016/S2666-5247(23)00100-3.

Negrete, O.A., Levroney, E.L., Aguilar, H.C., Bertolotti-Ciarlet, A., Nazarian, R., Tajyar, S., Lee, B., 2005. EphrinB2 is the entry receptor for Nipah virus, an emergent deadly paramyxovirus. Nature 436, 401−405. https://doi.org/10.1038/nature03838.

Ng, M., Ndungo, E., Kaczmarek, M.E., Herbert, A.S., Binger, T., Kuehne, A.I., Jangra, R.K., Hawkins, J.A., Gifford, R.J., Biswas, R., Demogines, A., James, R.M., Yu, M., Brummelkamp, T.R., Drosten, C., Wang, L.-F., Kuhn, J.H., Müller, M.A., Dye, J.M., Sawyer, S.L., Chandran, K., 2015. Filovirus receptor NPC1 contributes to species-specific patterns of ebolavirus susceptibility in bats. Elife 4, e11785. https://doi.org/10.7554/eLife.11785.

Olayemi, A., Fichet-Calvet, E., 2020. Systematics, ecology, and host switching: attributes affecting emergence of the Lassa virus in rodents across Western Africa. Viruses 12, 312. https://doi.org/10.3390/v12030312.

Olival, K.J., Hosseini, P.R., Zambrana-Torrelio, C., Ross, N., Bogich, T.L., Daszak, P., 2017. Host and viral traits predict zoonotic spillover from mammals. Nature 546, 646−650. https://doi.org/10.1038/nature22975.

Patiño-Galindo, J.Á., Filip, I., Rabadan, R., 2021. Global patterns of recombination across human viruses. Mol. Biol. Evol. 38, 2520−2531. https://doi.org/10.1093/molbev/msab046.

Paz, S., 2015. Climate change impacts on West Nile virus transmission in a global context. Philos. Trans. R. Soc. Lond. B Biol. Sci. 370, 20130561. https://doi.org/10.1098/rstb.2013.0561.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., Mirchandani, D., Scharton, D., Bilello, J.P., Ku, Z., An, Z., Kalveram, B., Freiberg, A.N., Menachery, V.D., Xie, X., Plante, K.S., Weaver, S.C., Shi, P.-Y., 2021. Spike mutation D614G alters SARS-CoV-2 fitness. Nature 592, 116−121. https://doi.org/10.1038/s41586-020-2895-3.

Plowright, R.K., Eby, P., Hudson, P.J., Smith, I.L., Westcott, D., Bryden, W.L., Middleton, D., Reid, P.A., McFarlane, R.A., Martin, G., Tabor, G.M., Skerratt, L.F., Anderson, D.L., Crameri, G., Quammen, D., Jordan, D., Freeman, P., Wang, L.-F., Epstein, J.H., Marsh, G.A., Kung, N.Y., McCallum, H., 2015. Ecological dynamics of emerging bat virus spillover. Proc. Biol. Sci. 282, 20142124. https://doi.org/10.1098/rspb.2014.2124.

Plowright, R.K., Parrish, C.R., McCallum, H., Hudson, P.J., Ko, A.I., Graham, A.L., Lloyd-Smith, J.O., 2017. Pathways to zoonotic spillover. Nat. Rev. Microbiol. 15, 502−510. https://doi.org/10.1038/nrmicro.2017.45.

Pulliam, J.R.C., Dushoff, J., 2009. Ability to replicate in the cytoplasm predicts zoonotic transmission of livestock viruses. J. Infect. Dis. 199, 565−568. https://doi.org/10.1086/596510.

Quan, P.-L., Firth, C., Conte, J.M., Williams, S.H., Zambrana-Torrelio, C.M., Anthony, S.J., Ellison, J.A., Gilbert, A.T., Kuzmin, I.V., Niezgoda, M., Osinubi, M.O.V., Recuenco, S., Markotter, W., Breiman, R.F., Kalemba, L., Malekani, J., Lindblade, K.A., Rostal, M.K., Ojeda-Flores, R., Suzan, G., Davis, L.B., Blau, D.M., Ogunkoya, A.B., Alvarez Castillo, D.A., Moran, D., Ngam, S., Akaibe, D., Agwanda, B., Briese, T., Epstein, J.H., Daszak, P., Rupprecht, C.E., Holmes, E.C., Lipkin, W.I., 2013. Bats are a major natural reservoir for hepaciviruses and pegiviruses. Proc. Natl. Acad. Sci. U.S.A. 110, 8194−8199. https://doi.org/10.1073/pnas.1303037110.

Rahman, S.K., Ansari, M.A., Gaur, P., Ahmad, I., Chakravarty, C., Verma, D.K., Sharma, A., Chhibber, S., Nehal, N., Wirth, D., Lal, S.K., 2021. The immunomodulatory CEA cell adhesion molecule 6 (CEACAM6/CD66c) is a protein receptor for the influenza a virus. Viruses 13, 726. https://doi.org/10.3390/v13050726.

Rimoin, A.W., Mulembakani, P.M., Johnston, S.C., Lloyd Smith, J.O., Kisalu, N.K., Kinkela, T.L., Blumberg, S., Thomassen, H.A., Pike, B.L., Fair, J.N., Wolfe, N.D., Shongo, R.L., Graham, B.S., Formenty, P., Okitolonda, E., Hensley, L.E., Meyer, H., Wright, L.L., Muyembe, J.-J., 2010. Major increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo. Proc. Natl. Acad. Sci. U.S.A. 107, 16262−16267. https://doi.org/10.1073/pnas.1005769107.

Rochman, N.D., Wolf, Y.I., Faure, G., Mutz, P., Zhang, F., Koonin, E.V., 2021. Ongoing global and regional adaptive evolution of SARS-CoV-2. Proc. Natl. Acad. Sci. U.S.A. 118, e2104241118. https://doi.org/10.1073/pnas.2104241118.

Russell, C.J., Hu, M., Okda, F.A., 2018. Influenza hemagglutinin protein stability, activation, and pandemic risk. Trends Microbiol. 26, 841−853. https://doi.org/10.1016/j.tim.2018.03.005.

Sánchez, C.A., Li, H., Phelps, K.L., Zambrana-Torrelio, C., Wang, L.-F., Zhou, P., Shi, Z.-L., Olival, K.J., Daszak, P., 2022. A strategy to assess spillover risk of bat SARS-related coronaviruses in Southeast Asia. Nat. Commun. 13, 4380. https://doi.org/10.1038/s41467-022-31860-w.

Sanjuan, R., Domingo-Calap, P., 2016. Mechanisms of viral mutation. Cell. Mol. Life Sci. 73, 4433−4448.

Sauerhering, L., Zickler, M., Elvert, M., Behner, L., Matrosovich, T., Erbar, S., Matrosovich, M., Maisner, A., 2016. Species-specific and individual differences in Nipah virus replication in porcine and human airway epithelial cells. J. Gen. Virol. 97, 1511−1519. https://doi.org/10.1099/jgv.0.000483.

Sempere Borau, M., Stertz, S., 2021. Entry of influenza A virus into host cells - recent progress and remaining challenges. Curr. Opin. Virol. 48, 23−29. https://doi.org/10.1016/j.coviro.2021.03.001.

Sharp, P.M., Hahn, B.H., 2011. Origins of HIV and the AIDS pandemic. Cold Spring Harb. Perspect. Med. 1, a006841. https://doi.org/10.1101/cshperspect.a006841.

Shipley, R., Wright, E., Selden, D., Wu, G., Aegerter, J., Fooks, A.R., Banyard, A.C., 2019. Bats and viruses: emergence of novel *Lyssaviruses* and association of bats with viral zoonoses in the EU. Trop. Med. Infect. Dis. 4, 31. https://doi.org/10.3390/tropicalmed4010031.

Simons, R.R.L., Gale, P., Horigan, V., Snary, E.L., Breed, A.C., 2014. Potential for introduction of bat-borne zoonotic viruses into the EU: a review. Viruses 6, 2084−2121. https://doi.org/10.3390/v6052084.

Song, H., Qi, J., Xiao, H., Bi, Y., Zhang, W., Xu, Y., Wang, F., Shi, Y., Gao, G.F., 2017. Avian-to-Human receptor-binding adaptation by influenza A virus hemagglutinin H4. Cell Rep. 20, 1201−1214. https://doi.org/10.1016/j.celrep.2017.07.028.

Staller, E., Barclay, W.S., 2021. Host cell factors that interact with influenza virus ribonucleoproteins. Cold Spring Harb. Perspect. Med. 11, a038307. https://doi.org/10.1101/cshperspect.a038307.

Starr, T.N., Zepeda, S.K., Walls, A.C., Greaney, A.J., Alkhovsky, S., Veesler, D., Bloom, J.D., 2022. ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. Nature 603, 913−918. https://doi.org/10.1038/s41586-022-04464-z.

Strobel, H.M., Stuart, E.C., Meyer, J.R., 2022. A trait-based approach to predicting viral host-range evolvability. Annu. Rev. Virol. 9, 139−156. https://doi.org/10.1146/annurev-virology-091919-092003.

Suhana, O., Nazni, W.A., Apandi, Y., Farah, H., Lee, H.L., Sofian-Azirun, M., 2019. Insight into the origin of chikungunya virus in Malaysian non-human primates via sequence analysis. Heliyon 5, e02682. https://doi.org/10.1016/j.heliyon.2019.e02682.

Suzán, G., Marcé, E., Giermakowski, J.T., Mills, J.N., Ceballos, G., Ostfeld, R.S., Armién, B., Pascale, J.M., Yates, T.L., 2009. Experimental evidence for reduced rodent diversity causing increased hantavirus prevalence. PLoS One 4, e5461. https://doi.org/10.1371/journal.pone.0005461.

Swaddle, J.P., Calos, S.E., 2008. Increased avian diversity is associated with lower incidence of human West Nile infection: observation of the dilution effect. PLoS One 3, e2488. https://doi.org/10.1371/journal.pone.0002488.

Taube, J.C., Rest, E.C., Lloyd-Smith, J.O., Bansal, S., 2023. The global landscape of smallpox vaccination history and implications for current and future orthopoxvirus susceptibility: a modelling study. Lancet Infect. Dis. 23, 454−462. https://doi.org/10.1016/S1473-3099(22)00664-8.

Taubenberger, J.K., Kash, J.C., 2010. Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe 7, 440−451. https://doi.org/10.1016/j.chom.2010.05.009.

Tebit, D.M., Arts, E.J., 2011. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. Lancet Infect. Dis. 11, 45−56. https://doi.org/10.1016/S1473-3099(10)70186-9.

Temmam, S., Davoust, B., Berenger, J.M., Raoult, D., Desnues, C., 2014. Viral metagenomics on animals as a tool for the detection of zoonoses prior to human infection? Int. J. Mol. Sci. 15, 10377−10397. https://doi.org/10.3390/ijms150610377.

Thompson, A.J., Paulson, J.C., 2021. Adaptation of influenza viruses to human airway receptors. J. Biol. Chem. 296, 100017. https://doi.org/10.1074/jbc.REV120.013309.

Tiee, M.S., Harrigan, R.J., Thomassen, H.A., Smith, T.B., 2018. Ghosts of infections past: using archival samples to understand a century of monkeypox virus prevalence among host communities across space and time. R. Soc. Open Sci. 5, 171089. https://doi.org/10.1098/rsos.171089.

Tjaden, N.B., Cheng, Y., Beierkuhnlein, C., Thomas, S.M., 2021. Chikungunya beyond the tropics: where and when do we expect disease transmission in Europe? Viruses 13, 1024. https://doi.org/10.3390/v13061024.

Troupin, C., Dacheux, L., Tanguy, M., Sabeta, C., Blanc, H., Bouchier, C., Vignuzzi, M., Duchene, S., Holmes, E.C., Bourhy, H., 2016. Large-scale phylogenomic analysis reveals the complex evolutionary history of rabies virus in multiple Carnivore hosts. PLoS Pathog. 12, e1006041. https://doi.org/10.1371/journal.ppat.1006041.

Valero-Rello, A., Sanjuán, R., 2022. Enveloped viruses show increased propensity to cross-species transmission and zoonosis. Proc. Natl. Acad. Sci. U.S.A. 119, e2215600119. https://doi.org/10.1073/pnas.2215600119.

Valesano, A.L., Rumfelt, K.E., Dimcheff, D.E., Blair, C.N., Fitzsimmons, W.J., Petrie, J.G., Martin, E.T., Lauring, A.S., 2021. Temporal dynamics of SARS-CoV-2 mutation accumulation within and across infected hosts. PLoS Pathog. 17, e1009499. https://doi.org/10.1371/journal.ppat.1009499.

van Doremalen, N., Miazgowicz, K.L., Milne-Price, S., Bushmaker, T., Robertson, S., Scott, D., Kinne, J., McLellan, J.S., Zhu, J., Munster, V.J., 2014. Host species restriction of Middle East respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4. J. Virol. 88, 9220−9232. https://doi.org/10.1128/JVI.00676-14.

Visher, E., Evensen, C., Guth, S., Lai, E., Norfolk, M., Rozins, C., Sokolov, N.A., Sui, M., Boots, M., 2021. The three Ts of virulence evolution during zoonotic emergence. Proc. Biol. Sci. 288, 20210900. https://doi.org/10.1098/rspb.2021.0900.

Wang, L.-F., Anderson, D.E., 2019. Viruses in bats and potential spillover to animals and humans. Curr. Opin. Virol. 34, 79−89. https://doi.org/10.1016/j.coviro.2018.12.007.

Wang, W., Han, G.-Z., 2021. Pervasive positive selection on virus receptors driven by host-virus conflicts in mammals. J. Virol. 95, e0102921. https://doi.org/10.1128/JVI.01029-21.

Wang, W., Zhao, H., Han, G.-Z., 2020. Host-virus arms races drive elevated adaptive evolution in viral receptors. J. Virol. 94, e00684−20. https://doi.org/10.1128/JVI.00684-20.

Warren, C.J., Sawyer, S.L., 2019. How host genetics dictates successful viral zoonosis. PLoS Biol. 17, e3000217. https://doi.org/10.1371/journal.pbio.3000217.

Wei, C., Shan, K.-J., Wang, W., Zhang, S., Huan, Q., Qian, W., 2021. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. J. Genet. Genomics Yi Chuan Xue Bao 48, 1111−1121. https://doi.org/10.1016/j.jgg.2021.12.003.

Wells, H.L., Letko, M., Lasso, G., Ssebide, B., Nziza, J., Byarugaba, D.K., Navarrete-Macias, I., Liang, E., Cranfield, M., Han, B.A., Tingley, M.W., Diuk-Wasser, M., Goldstein, T., Johnson, C.K., Mazet, J.a.K., Chandran, K., Munster, V.J., Gilardi, K., Anthony, S.J., 2021. The evolutionary history of ACE2 usage within the coronavirus subgenus Sarbecovirus. Virus Evol. 7, veab007. https://doi.org/10.1093/ve/veab007.

Wertheim, J.O., Hostager, R., Ryu, D., Merkel, K., Angedakin, S., Arandjelovic, M., Ayimisin, E.A., Babweteera, F., Bessone, M., Brun-Jeffery, K.J., Dieguez, P., Eckardt, W., Fruth, B., Herbinger, I., Jones, S., Kuehl, H., Langergraber, K.E., Lee, K., Madinda, N.F., Metzger, S., Ormsby, L.J., Robbins, M.M., Sommer, V., Stoinski, T., Wessling, E.G., Wittig, R.M., Yuh, Y.G., Leendertz, F.H., Calvignac-Spencer, S., 2021. Discovery of novel herpes simplexviruses in wild Gorillas, bonobos, and chimpanzees supports zoonotic origin of HSV-2. Mol. Biol. Evol. 38, 2818−2830. https://doi.org/10.1093/molbev/msab072.

Wille, M., Eden, J.-S., Shi, M., Klaassen, M., Hurt, A.C., Holmes, E.C., 2018. Virus-virus interactions and host ecology are associated with RNA virome structure in wild birds. Mol. Ecol. 27, 5263−5278. https://doi.org/10.1111/mec.14918.

Woolhouse, M., Scott, F., Hudson, Z., Howey, R., Chase-Topping, M., 2012. Human viruses: discovery and emergence. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 2864−2871. https://doi.org/10.1098/rstb.2011.0354.

Woolhouse, M.E.J., Brierley, L., 2018. Epidemiological characteristics of human-infective RNA viruses. Sci. Data 5, 180017. https://doi.org/10.1038/sdata.2018.17.

Woolhouse, M.E.J., Brierley, L., McCaffery, C., Lycett, S., 2016. Assessing the epidemic potential of RNA and DNA viruses. Emerg. Infect. Dis. 22, 2037−2044. https://doi.org/10.3201/eid2212.160123.

Worobey, M., Watts, T.D., McKay, R.A., Suchard, M.A., Granade, T., Teuwen, D.E., Koblin, B.A., Heneine, W., Lemey, P., Jaffe, H.W., 2016. 1970s and "Patient 0" HIV-1 genomes illuminate early HIV/AIDS history in North America. Nature 539, 98−101. https://doi.org/10.1038/nature19827.

Wu, Z., Lu, L., Du, J., Yang, L., Ren, X., Liu, B., Jiang, J., Yang, J., Dong, J., Sun, L., Zhu, Y., Li, Y., Zheng, D., Zhang, C., Su, H., Zheng, Y., Zhou, H., Zhu, G., Li, H., Chmura, A., Yang, F., Daszak, P., Wang, J., Liu, Q.,

Jin, Q., 2018. Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. Microbiome 6, 178. https://doi.org/10.1186/s40168-018-0554-9.

Yang, Y.-L., Yu, J.-Q., Huang, Y.-W., 2020. Swine enteric alphacoronavirus (swine acute diarrhea syndrome coronavirus): an update three years after its discovery. Virus Res. 285, 198024. https://doi.org/10.1016/j.virusres.2020.198024.

Zanin, M., Duan, S., Wong, S.-S., Kumar, G., Baviskar, P., Collin, E., Russell, C., Barman, S., Hause, B., Webby, R., 2017. An amino acid in the stalk domain of N1 neuraminidase is critical for enzymatic activity. J. Virol. 91, e00868−16. https://doi.org/10.1128/JVI.00868-16.

Zhang, F., Chase-Topping, M., Guo, C.-G., van Bunnik, B.A.D., Brierley, L., Woolhouse, M.E.J., 2020a. Global discovery of human-infective RNA viruses: a modelling analysis. PLoS Pathog. 16, e1009079. https://doi.org/10.1371/journal.ppat.1009079.

Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Peng, H., Quinlan, B.D., Rangarajan, E.S., Pan, A., Vanderheiden, A., Suthar, M.S., Li, W., Izard, T., Rader, C., Farzan, M., Choe, H., 2020b. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nat. Commun. 11, 6013. https://doi.org/10.1038/s41467-020-19808-4.

Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E.C., Hughes, A.C., Bi, Y., Shi, W., 2020a. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. Curr. Biol. CB 30, 2196−2203.e3. https://doi.org/10.1016/j.cub.2020.05.023.

Zhou, P., Fan, H., Lan, T., Yang, X.-L., Shi, W.-F., Zhang, W., Zhu, Y., Zhang, Y.-W., Xie, Q.-M., Mani, S., Zheng, X.-S., Li, B., Li, J.-M., Guo, H., Pei, G.-Q., An, X.-P., Chen, J.-W., Zhou, L., Mai, K.-J., Wu, Z.-X., Li, D., Anderson, D.E., Zhang, L.-B., Li, S.-Y., Mi, Z.-Q., He, T.-T., Cong, F., Guo, P.-J., Huang, R., Luo, Y., Liu, X.-L., Chen, J., Huang, Y., Sun, Q., Zhang, X.-L.-L., Wang, Y.-Y., Xing, S.-Z., Chen, Y.-S., Sun, Y., Li, J., Daszak, P., Wang, L.-F., Shi, Z.-L., Tong, Y.-G., Ma, J.-Y., 2018. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. Nature 556, 255−258. https://doi.org/10.1038/s41586-018-0010-9.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., Shi, Z.-L., 2020b. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270−273. https://doi.org/10.1038/s41586-020-2012-7.

Zhu, Y., Wang, R., Yu, L., Sun, H., Tian, S., Li, P., Jin, M., Chen, H., Ma, W., Zhou, H., 2020. Human TRA2A determines influenza A virus host adaptation by regulating viral mRNA splicing. Sci. Adv. 6, eaaz5764. https://doi.org/10.1126/sciadv.aaz5764.

This page intentionally left blank

PART

Specialized Chapters 2

This page intentionally left blank

CHAPTER

Genetic Exchange in Trypanosomatids and its Relevance to Epidemiology

24

Wendy Gibson[1], Michael D. Lewis[2], Matthew Yeo[2] and Michael A. Miles[2]

[1]*University of Bristol, School of Biological Sciences, Bristol, United Kingdom;* [2]*Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine (LSHTM), London, United Kingdom*

1. Introduction

Genetic exchange has now been demonstrated experimentally in the so-called TriTryps, three of the trypanosomatids for which genome sequences have been published (Berriman et al., 2005; El-Sayed et al., 2005; Ivens et al., 2005): *Trypanosoma brucei*, *T. cruzi*, *Leishmania major*, and also within and between other *Leishmania* species (Romano et al., 2014). As well as in these human pathogens, the related bumble bee parasite, *Crithidia bombi*, has also been shown to undergo genetic exchange (Schmid-Hempel et al., 2011). For *T. brucei* and *Leishmania* spp. genetic exchange occurs in the insect vector and appears to follow Mendelian rules of inheritance (Akopyants et al., 2009; Jenni et al., 1986; MacLeod et al., 2005). In contrast, for *T. cruzi* genetic exchange has been demonstrated experimentally in *in vitro* cultures and hybrid formation appears to result from fusion of diploid cells followed by genome erosion (Gaunt et al., 2003; Matos et al., 2022).

Before the experimental confirmation of genetic exchange in these parasites, evidence for the natural occurrence of hybrids had accumulated from molecular epidemiological analysis of isolates collected from the field. Tait (1980) showed that isoenzyme data from *T. brucei* conformed to Hardy−Weinberg equilibrium and concluded that the population was undergoing random mating. Later analyses threw doubt on the extent of panmictic mating in *T. brucei* with the widespread acceptance of the concept of clonality in parasitic protozoa (Tibayrenc et al., 1990). However, genetic exchange was considered to have led to demonstrably hybrid *T. cruzi* recovered from the field (Carrasco et al., 1996; Machado and Ayala, 2001; Sturm et al., 2003; Westenberger et al., 2005), including TcV and TcVI interlineage natural hybrids, subsequently indicated to be of relatively recent origin (Lewis et al., 2011). Natural hybrids were also demonstrated between several *Leishmania* species and subspecies (Belli et al., 1994; Kelly et al., 1991; Ravel et al., 2006).

Here we review results from these two complementary avenues of study—analysis of naturally occurring hybrids among field isolates and experimental genetic exchange in the laboratory—for the pathogenic trypanosomatids *T. brucei*, *T. cruzi*, and several *Leishmania* species.

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00007-0
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

2. Trypanosoma brucei

2.1 Genetic Crosses

Compelling evidence for mating in *Trypanosoma brucei sensu lato* came from population genetics studies based on isoenzyme data. Gibson et al. (1980) described isoenzyme patterns consistent with those expected from homo- and heterozygotes in an extensive analysis of *T. brucei* isolates from East and West Africa, and Tait (1980) showed that similar data from 17 Ugandan isolates of *T. brucei* conformed to Hardy−Weinberg equilibrium, indicating that the population was undergoing random mating. The first successful laboratory cross was reported in 1986 when Jenni and colleagues co-transmitted two genetically distinct clones of *T. brucei s.l.* through tsetse flies and demonstrated hybrid progeny that had inherited a mixture of genetic markers from both parents (Jenni et al., 1986). The hybrids were found among metacyclics recovered from the salivary glands showing that genetic exchange had occurred sometime during the trypanosome's developmental cycle in the tsetse fly. However, the metacyclic population contained a mixture of parental and hybrid genotypes indicating that mating is not an obligatory event in the life cycle. This contrasts with the situation for the malaria parasite, *Plasmodium* spp., where gamete formation and production of a zygote is a normal part of the transmission cycle.

Subsequent trypanosome crosses have followed the same general plan: two genetically distinct parental trypanosome clones are fed to groups of newly emerged tsetse flies in their first blood meal (Fig. 24.1). Tsetse are typically refractory to trypanosome infection, but are at their most susceptible as very young flies before they have fed (Maudlin, 1991). Not all flies become infected after the infected feed and only some infected flies produce hybrids. Thus, large numbers of flies and trypanosome populations need to be screened to identify those containing hybrids. To avoid the laborious job of

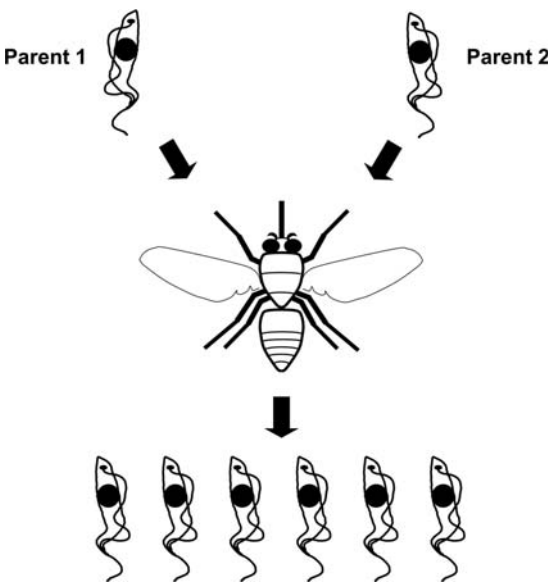


FIGURE 24.1 Design of experimental cross.

Design of an experimental cross for *Trypanosoma brucei*. The two parental trypanosomes are co-transmitted via tsetse flies and hybrid trypanosomes are found among the infective metacyclics from the salivary glands.

identifying hybrid-producing flies by screening every fly, selectable markers were incorporated into the experimental design. This became feasible following the development of methods for the stable transformation of trypanosomes with exogenous DNA in the early 1990s (Eid and Sollner-Webb, 1991; Lee and Van Der Ploeg, 1990; Ten Absroek et al., 1990). In the cross described by Gibson and Whittington (1993), each of the parental clones was transformed with a different construct designed to integrate a gene for drug resistance into the tubulin locus by homologous recombination. In this way, parental clones resistant to the antibiotics hygromycin or G418 were created. After co-transmission through the fly, hybrid progeny was selected by resistance to both drugs. This strategy has obvious advantages over the previous "finding a needle in a haystack" approach and was that used for the discovery of hybrids in *Leishmania major* (Akopyants et al., 2009).

The discovery of green fluorescent protein (GFP) and the development of methods to image the protein in living cells opened up exciting new approaches for studying genetic exchange in trypanosomes. Incorporation of genes for GFP in one parental line and red fluorescent protein (RFP) in the other led to the production of hybrids with both genes, which appear yellow making them immediately distinct from the parental cells by simple fluorescence microscopy of live cell preparations. This approach was a boon for the analysis of *T. brucei* hybrids within the tsetse fly and allowed the location and timing of genetic exchange to be determined precisely (Gibson et al., 2008).

It has proved possible to cross all subspecies of *T. brucei* except the human pathogen *T. b. gambiense* Group 1. The difficulty in setting up crosses of *T. b. gambiense* Group 1 is that it transmits poorly or not at all through *Glossina morsitans* ssp., the standard laboratory tsetse fly (Dukes et al., 1989; Gibson, 1986). The more virulent Group 2 *T. b. gambiense* is readily transmissible via *G. morsitans* ssp. and has featured in several crosses, including the original cross of (Jenni et al., 1986), where the parents were *T. b. brucei* STIB 247 and *T. b. gambiense* Group 2 STIB 386 (TH114). *T. b. gambiense* Group 2 (TH2) was also mated with both *T. b. brucei* and *T. b. rhodesiense* (Bingle et al., 2001; Gibson et al., 1997a). Several crosses have involved parental lines of *T. b. brucei* (Degen et al., 1995; Gibson and Garside, 1991; Gibson et al., 2008; Schweizer et al., 1994; Turner et al., 1990), or *T. b. brucei* and *T. b. rhodesiense* (Gibson, 1989; Gibson and Whittington, 1993; Gibson et al., 2015). Inheritance of the trait for human infectivity has been analyzed in crosses of *T. b. brucei* and *T. b. rhodesiense* (Gibson and Mizen, 1997), and *T. b. brucei* and Group 2 *T. b. gambiense* (Turner et al., 2004). A single gene, *SRA* (serum resistance associated), confers human infectivity on *T. b. rhodesiense* (De Greef et al., 1989; Van Xong et al., 1998), and it was possible to follow the inheritance of this gene, along with the trait of human serum resistance, in crosses of *T. b. brucei* and *T. b. rhodesiense* (Gibson et al., 2015). The *SRA* gene was transferred to new recombinant genetic backgrounds, potentially creating new strains of the human pathogen with larger and more diverse antigenic repertoires.

2.2 Location of Genetic Exchange

Jenni and colleagues originally showed that genetic exchange took place in the fly, since cloned metacyclics from the saliva of infected flies were hybrid (Jenni et al., 1986); however, the exact location and lifecycle stage remained undefined. While there were reports of hybrid formation in mixed midgut procyclic populations both in vitro and in vivo (Evans and Ellis, 1983; Schweizer and Jenni, 1991; Schweizer et al., 1991), the bulk of evidence pointed to the salivary glands as the site of

genetic exchange. Firstly the timing of hybrid appearance: hybrids were most likely to be found in flies infected for at least 28 days, i.e., sufficient time for salivary gland invasion and colonization, despite there being a large population of procyclic trypanosomes continuously present in the midgut throughout this time (Gibson and Whittington, 1993; Gibson et al., 1997a; Schweizer et al., 1988). Selection by double drug resistance revealed that hybrids were present in populations derived from the salivary glands but not from the midgut (Gibson and Bailey, 1994; Gibson and Whittington, 1993; Gibson et al., 1997a). The direct visualization of trypanosome hybrids using fluorescent reporter proteins unequivocally established that hybrids are formed in the salivary glands and are not present in the midgut or among the migratory stages in the proventriculus and foregut (Bingle et al., 2001; Gibson et al., 2008).

The use of red and green fluorescent reporters also explained why hybrids are found infrequently, with less than a quarter of infected flies producing hybrids (Gibson and Stevens, 1999). In order to mate, both parental trypanosomes need to reach and colonize the same salivary gland, and it became evident that this is not always the case. While most flies co-infected with red and green fluorescent trypanosomes developed a mixed midgut infection, only about a third of these flies also had a mixed infection in the salivary glands (Gibson et al., 2008; Peacock et al., 2007). In a number of cases, the composition of the trypanosome population in the two salivary glands of the pair differed, perhaps with only one salivary gland infected or a mixed infection in one gland but not the other. Interestingly, all progeny from the red/green cross were hybrid and no trypanosomes with parental genotypes were recovered as in previous crosses (Gibson et al., 2008). The design of this cross may have increased the probability of finding hybrids, as analysis could be focused on salivary glands containing both parental genotypes, which was not possible in previous crosses.

These observations highlight the fact that few trypanosomes complete the journey from the midgut to the salivary duct, and only some of these then succeed in establishing an infection in either of the salivary glands. When together in the same salivary gland, the trypanosomes readily mate, as demonstrated by the fact that most salivary glands with a mixed infection of red and green trypanosomes also contained yellow fluorescent hybrids (Gibson et al., 2008). Compatibility of different trypanosome strains may then depend on them reaching the salivary glands simultaneously. In some early crosses, the two parental clones differed substantially in their speed of colonizing the salivary glands (Gibson and Whittington, 1993; Gibson et al., 1997a; Schweizer et al., 1988). The fact that mating in *T. brucei* occurs in the salivary glands makes the prospect of producing hybrids in vitro more remote, as reliable culture systems do not exist for the life cycle stages that occur in the salivary glands.

2.3 Mendelian Inheritance and Meiosis

It is generally accepted that *T. brucei* is diploid with respect to the 11 pairs of large chromosomes that contain the housekeeping genes, although this arrangement probably does not apply to the intermediate and mini-chromosomes (Berriman et al., 2005; Gottesdiener et al., 1990; Melville et al., 1998). Analysis of the inheritance of genetic markers in hybrid progeny from crosses of *T. brucei* is consistent with Mendelian genetics for the most part (Gibson, 1989; Gibson and Bailey, 1994; MacLeod et al., 2005; Schweizer et al., 1994; Sternberg et al., 1989; Turner et al., 1990), leading to the assumption that meiosis occurs. Indeed, genetic linkage maps for *T. b. brucei* and Group 2 *T. b. gambiense* have been constructed from detailed analysis of microsatellite inheritance and frequency of crossing-over (Cooper et al., 2008; MacLeod et al., 2005).

It has proved more difficult to directly visualize trypanosomes undergoing meiosis. While hybrids were easily detected in a cross of red and green fluorescent trypanosomes from day 13 onwards (Gibson et al., 2008), putative intermediate stages were neither abundant nor obvious, necessitating an alternative approach to detect trypanosomes undergoing meiosis. Phylogenomic studies had identified the presence in *T. brucei* of homologs of several genes crucial for meiosis in yeast and other eukaryotes such as Spo11, Hop1, Dmc1, and Mnd1 (Ramesh et al., 2005). By constructing fusions of the *T. brucei* homologs with the gene for yellow fluorescent protein (YFP), expression of three putative meiosis genes (*MND1*, *DMC1*, and *HOP1*) was monitored through the developmental cycle of *T. brucei* in the fly (Peacock et al., 2011). Expression was observed only among trypanosomes during the early stage of colonization of the salivary glands, consistent with the first appearance of hybrids, and was localized to the nucleus; morphologically, the meiotic cells were epimastigotes and had a large nucleus occupying the posterior end of the cell, with two anterior kinetoplasts, each associated with a flagellum. These dividing epimastigotes were presumed to be in Meiosis I, as all three meiosis-specific proteins function during chromosomal recombination. Surprisingly, the meiotic trypanosomes occurred with similar frequency in both single and mixed strain transmissions and it seems probable that meiosis is a normal part of the trypanosome developmental cycle in the fly. Expression of meiosis-specific genes has now been observed in all subspecies of *T. brucei*, including *T. b. gambiense* group 1 (Peacock et al., 2014b).

The identification of trypanosomes undergoing meiosis was the first step in elucidating the mechanism of genetic exchange in *T. brucei* at the cell biology level, and eventually led to the identification of haploid gametes (Peacock et al., 2014b), a stage that had initially proved elusive (Kooy et al., 1989; Shapiro et al., 1984; Zampetti-Bosseler et al., 1986). Observation of trypanosomes from the early phase of salivary gland colonization when meiotic stages are present, revealed the presence of distinctive pear-shaped cells with a long flagellum; in mixtures of red and green fluorescent trypanosomes, these cells were seen to interact by intertwining their flagella and transfer of cytoplasmic material was also evident as some cells had yellow fluorescence (Peacock et al., 2014b). Measurement of DNA contents of individual nuclei showed that the pear-shaped cells had half the nuclear DNA content of metacyclics, which are nondividing trypanosomes arrested in G_0. Curiously, about half the haploid cells observed had one nucleus and two kinetoplasts (1N2K), but only a single flagellum; the rest were 1N1K cells with one flagellum. Investigation of the intermediate stages involved in meiosis revealed that gametes are produced sequentially during asynchronous Meiosis II divisions (Peacock et al., 2021). This diverges from textbook accounts of eukaryote meiosis that describe symmetrical and synchronous nuclear divisions, resulting in four haploid gametes. Evolution of this unusual form of meiosis may have been constrained by the asymmetrical nature of the trypanosome cell.

Gametes and preceding meiotic stages are mixed together with other cell types including epimastigotes and metacyclics in parasitized tsetse salivary glands, making it difficult to isolate and characterize the sexual stages. Numbers also fluctuate over time with sexual stages most abundant early in infection, while attached epimastigotes and infective metacyclics dominate later (Peacock et al., 2014b, 2021). A prerequisite for the further investigation of gametes and other sexual stages was therefore the identification of markers for these cell types. Peacock et al. (2021) were able to demonstrate that gametes, as well as presumed meiotic intermediates, expressed the gamete fusogenic protein HAP2 (hapless two or generative cell specific1, GCS1) (Fédry et al., 2017; Fedry et al., 2018) by tagging the endogenous HAP2 gene with YFP. Expression of this gene then served as a tool to identify these cell populations in single cell RNA sequence analysis, enabling further characterization

of their transcriptomes (Howick et al., 2022; Hutchinson et al., 2021). Of particular interest are highly expressed genes encoding proteins of currently unknown function, as such proteins may be involved in sexual processes in trypanosomatids.

Previously, indirect evidence for haploid nuclei came from the observation that many *T. brucei* sspp. crosses produced triploid hybrid progeny, which most probably arose from fusion of a haploid nucleus with one that is diploid (Gibson and Bailey, 1994; Gibson and Whittington, 1993; Gibson et al., 1992, 1995, 1997a; Kay et al., 2022). Hybrids with high DNA contents relative to the parents were found even in the first experimental cross (Jenni et al., 1986), which created some initial confusion about the mechanism of genetic exchange (Paindavoine et al., 1986; Wells et al., 1987), as it pointed to fusion of diploids with subsequent genome erosion, rather than meiosis and fusion of haploid gametes. Analysis of progeny clones with high DNA contents from several further crosses demonstrated that these hybrids were probably triploid with DNA contents that clustered at the 3N value; in addition, trisomy was confirmed for several chromosomes (Gibson and Bailey, 1994; Gibson and Whittington, 1993; Gibson et al., 1992, 1995, 1997a; Kay et al., 2022). The availability of genome sequences from parental and progeny clones from an experimental cross allowed (Kay et al., 2022) to infer from analysis of single nucleotide polymorphisms (SNPs) that a 3N hybrid clone had inherited a haploid set of megabase chromosomes from one parent and two sets from the other. As most chromosomes showed evidence of crossing over, it was clear that nuclear fusion had occurred postmeiotically, perhaps involving the combination of three haploid gametes, or alternatively, one haploid gamete with a meiotic intermediate, prior to its division into two gametes. Such meiotic intermediates, in addition to gametes, were shown to express HAP2 (Peacock et al., 2021) making this scenario quite plausible.

As well as triploid hybrids, several putative tetraploid hybrids with a 4N DNA content were recovered from a cross of red and green fluorescent trypanosomes (Gibson et al., 2008). Whereas the presence of triploid hybrids was obvious from the demonstration of three alleles at some loci, only two microsatellite alleles were present in each of the tetraploid hybrids and subsequent SNP analysis of one 4N clone also failed to detect the presence of more than one homologue from each parent trypanosome (Kay et al., 2022). However, as there was evidence of crossing over on several chromosomes, these 4N hybrids must have been formed postmeiotically and not by fusion of the two diploid parental genomes, but their precise origin remains uncertain. While triploids appear to be stable during growth and fly transmission (Gibson et al., 1992; Wells et al., 1987), the tetraploids may be unstable, as flow cytometry analysis of the DNA contents of tetraploid clones frequently revealed an extra G1 peak at the 2N position (Gibson et al., 2008).

Intraclonal mating was initially thought not to occur in *T. brucei* except in the presence of outcrossing trypanosomes, leading to the hypothesis that some kind of diffusible factor produced by recognition of nonself-induced mating (Gibson et al., 1997a; Tait et al., 1996). Intraclonal mating explained the occasional anomalies where hybrid progeny were homozygous instead of heterozygous as expected if the parents were different homozygotes (Gibson and Bailey, 1994; Sternberg et al., 1988). Experiments using red and green fluorescent clones of a single *T. brucei* strain have shown that intraclonal mating occurs with some frequency in the absence of a second trypanosome strain (Peacock et al., 2009). The assumption that a *T. brucei* clone can be tsetse-transmitted without change is therefore doubtful. Indeed, genome sequencing unexpectedly picked up evidence of selfing in a hybrid progeny clone after fly transmission (Kay et al., 2022).

In other protists such as ciliates, strain compatibility is determined by a system of mating types, but the question of trypanosome mating types remains open. Analysis of a series of F1 and back crosses revealed no systematic pattern of compatible matings (Peacock et al., 2014a). When red and green haploid gametes from intraclonal crosses were observed, they behaved exactly like the gametes from interclonal crosses, coming into close proximity and intertwining their flagella, except that no mixing of cytoplasm (yellow fluorescence) was evident among gametes of the same trypanosome strain (Peacock et al., 2014a).

2.4 Inheritance of Kinetoplast DNA

Kinetoplast DNA (kDNA) is the mitochondrial DNA of trypanosomatids and consists of an interlocked network of about 50 (20−25 kb) maxicircles and 5000 (1 kb) minicircles (reviewed by (Shapiro and Englund, 1995). The kDNA is contained within an organelle, the kinetoplast, which is inside the mitochondrion. Initial results supported the hypothesis that inheritance of kDNA was uniparental, with the kDNA of either parent being passed on to the hybrid progeny (Gibson, 1989; Sternberg et al., 1988, 1989). However, detailed analysis of both maxi- and minicircles showed that although maxicircles were of a single parental type, the minicircles had been inherited from both parents (Gibson and Garside, 1990; Gibson et al., 1997b). Recovery of maxi- and minicircle sequences from whole genome sequences of parents and progeny from an experimental cross has confirmed this surprising mode of inheritance (Kay et al., 2022). To explain this result, it was assumed that the hybrid kDNA network initially consists of both maxi- and minicircles from both parents in an equal proportion; in this scenario, the small number of maxicircles resolves to a single type by random segregation during subsequent mitotic divisions, while the much greater number of minicircles endures as a hybrid network. This idea is supported by the observation that hybrids at an early stage of growth have mixed maxicircle networks (Gibson et al., 2008; Turner et al., 1995).

The initial formation of the hybrid kDNA network remains an intriguing problem. The first requirement is fusion of the mitochondria from both parental trypanosomes to allow the kDNA to mix. Then mini- and maxicircles from both parental kDNA networks would need to combine into a single hybrid kDNA network, which would involve detachment and reattachment of all the individual DNA circles. An alternative hypothesis is that only some minicircles are swapped while the kDNA networks of the two parents are adjacent, leaving the core maxicircle network intact; the partially hybrid kDNA networks would then be inherited by individual progeny trypanosomes (Shapiro and Englund, 1995). However, this hypothesis does not fit with the observation of mixed maxicircle networks above.

2.5 Implications for Epidemiology

Although genetic exchange in *T. brucei* has been amply demonstrated in the laboratory, its importance in natural trypanosome populations remains controversial. One problem has been sampling bias, with collection of trypanosome isolates usually focused on epidemics of human disease rather than collecting all the *T. brucei* strains circulating in an area (Cibulskis, 1992; Hide et al., 1994; Stevens and Tibayrenc, 1996; Stevens and Welburn, 1993). In epidemics of human trypanosomiasis caused by *T. b. rhodesiense*, transmission may well be direct from human to human rather than from an animal reservoir, or involve only local cattle reservoir hosts, allowing the clonal expansion of particular trypanosome genotypes. This is very different from the endemic scenario where humans and their

livestock are occasional hosts in the natural circulation of *T. brucei* ssp. strains in wild mammals and tsetse.

Through the analysis of trypanosome mating in the laboratory, we can now define the biological circumstances in which genetic exchange will be found: at least two trypanosome strains must be present in the salivary glands of a tsetse fly for mating to occur; since flies are most readily infected on their first bloodmeal (Maudlin, 1991; Wijers, 1958), a mixed infection is likely to be acquired from one infected mammal carrying multiple trypanosome strains. Few mixed infections have been reported from humans (Balmer and Tanner, 2011; Truc et al., 1998), but mixed infections of more than one trypanosome strain or species are frequently encountered in livestock (Godfrey and Killick-Kendrick, 1961; Nyeko et al., 1990), tsetse (Adams et al., 2006; Lehane et al., 2000; MacLeod et al., 1999; Masiga et al., 1996), and presumably also occur frequently in the large wild mammals that sustain many tsetse populations. In-depth analysis of trypanosome samples from these transmission cycles would be informative.

Even though the frequency of genetic exchange in natural populations may be low, there is potential for significant epidemiological consequences. For example, the trait of human infectivity in *T. b. rhodesiense* is conferred by a single gene, *SRA* (De Greef et al., 1989; Van Xong et al., 1998). Any cross between *T. b. rhodesiense* and *T. b. brucei* would place this key virulence gene in new genetic backgrounds, thus creating new genotypes of *T. b. rhodesiense*. There is abundant evidence of strain heterogeneity in *T. b. rhodesiense* from several foci of human trypanosomiasis in East Africa (Enyaru et al., 1993; Gibson and Gashumba, 1983; Gibson and Wellde, 1985; Hide et al., 1994; Komba et al., 1997), and quite different *T. b. rhodesiense* genotypes have been found in neighboring foci in Uganda and Kenya (Echodu et al., 2015; Gibson and Gashumba, 1983; Gibson and Wellde, 1985). We know little about the genetic basis of most phenotypic characteristics of *T. brucei s.l.*, but undoubtedly genetic exchange provides the opportunity for more virulent, pathogenic, or fly transmissible strains of *T. b. rhodesiense* to arise. Population genetics analysis of *T. brucei* sspp. strains from across Africa provided strong evidence of genetic admixture between *SRA*-positive *T. b. rhodesiense* and *SRA*-negative *T. b. brucei* in East Africa (Balmer et al., 2011), and transfer of the *SRA* gene was demonstrated in several *T. b. rhodesiense* x *T. b. brucei* lab crosses, thus creating new genetic strains of the human pathogen (Gibson et al., 2015). Genetic exchange among *T. brucei* strains will also serve to increase the diversity of the variant surface glycoprotein (VSG) repertoire. Considering the prodigious energy and resources allocated to antigenic variation by the trypanosome, including an estimated 1000 *VSG* genes, the wholesale import of new *VSG* genes brought about by genetic exchange would constitute a major boost for survival in the mammalian host.

3. Trypanosoma cruzi

3.1 *T. cruzi* Diversity

Trypanosoma cruzi is considered to be a single species encompassing six distinct genetic lineages (also referred to as discrete typing units, DTUs). The current international consensus nomenclature for the six *T. cruzi* lineages is TcI to TcVI (Zingales et al., 1998). The definition of these six *T. cruzi* subgroups was originally on the basis of phenotyping using multilocus enzyme electrophoresis (MLEE) (Messenger et al., 2015b; Miles et al., 2009a). Subsequently the same six genetic lineages were supported by comparative analyses of a wide range of nuclear DNA targets from across the parasite's genome (Messenger et al., 2015b,c; Schwabl et al., 2020).

TcI to TcVI show broadly distinctive but not entirely exclusive geographical, ecological, transmission cycle and disease associations; overlaps occur, and mixed infections are reported from humans, reservoir mammal hosts and arthropod vectors (Bosseno et al., 1996; Burgos et al., 2008; Cardinal et al., 2008; Messenger et al., 2015c; Yeo et al., 2005, 2007). TcI is the principal agent of Chagas disease in Latin America north of the Amazon basin, whereas TcII, TcV, and TcVI are the main causes of Chagas disease in the southern cone countries of South America (Barnabé et al., 2000; Fernandes et al., 1998; Messenger et al., 2015c; Miles et al., 1978, 1981; Zingales et al., 1998). The sylvatic TcI transmission cycles are widespread throughout Latin America and are largely arboreal; the common opossums (*Didelphis* species) are obvious and abundant reservoir hosts but many other mammal species may be infected, with some transmission among rodent species and triatomines with terrestrial habitats (Yeo et al., 2005). The natural transmission cycles of TcII, TcV, and TcVI appear to be rare or difficult to discover and characterize; however, presence of TcII in primates of the Atlantic forest region of Brazil has recently been confirmed by a combination genotyping and *T. cruzi* lineage-specific serology (Bhattacharyya et al., 2015; Lisboa et al., 2004). TcIII is predominantly sylvatic and seldom infects humans. TcIV is a secondary cause of Chagas disease in Venezuela, after TcI (Carrasco et al., 2012). TcIII appears to have the most clearly defined and exclusive natural ecological and host association, with the burrowing armadillo, *Dasypus novemcinctus*, and rarely infects humans (Messenger et al., 2015c; Yeo et al., 2005).

The various *T. cruzi* isolates that were characterized to establish the TcI−VI subgroups were collected from disparate geographical locations and usually with small numbers of isolates gathered from each site. Although this striking genetic heterogeneity of *T. cruzi* was a landmark discovery, which has transformed subsequent research on the epidemiology of Chagas disease, this sampling strategy was not ideally suited for determining the structure of *T. cruzi* populations. Nevertheless, repeated analyses of the molecular diversity of *T. cruzi* based on such samples have supported the discrete nature of the six genetic lineages and demonstrated strong linkage disequilibrium between them (Tibayrenc et al., 1990). These observations fostered the tenet that *T. cruzi* was propagated clonally, both between and within the genetic lineages, and that genetic exchange, if it occurred, was rare and of little epidemiological consequence (Tibayrenc and Ayala, 2002). On the other hand, despite this clonal theory, MLEE, and later MLST, analyses indicated that TcV and TcVI resembled natural interlineage hybrids between TcII and TcIII (Machado and Ayala, 2001; Miles, 1985). The hybrid nature of TcVI was conclusively shown when the CL Brener genome was sequenced and proven to comprise widespread TcII−TcIII heterozygosity (El-Sayed et al., 2005). This finding was broadly confirmed in wider panels of TcV and TcVI strains (Flores-López and Machado, 2011; Lewis et al., 2011), but these studies also showed that after the founding recombination events, which were dated to within the last 60,000 years, these F_1-like hybrid lineages have been maintained clonally. Further evidence of interlineage recombination has come in the form of natural TcI, TcIII and TcIV strain collections for which there is clear discordance between nuclear and mitochondrial DNA phylogenies (Machado and Ayala, 2001; Messenger and Miles, 2015; Messenger et al., 2015c).

The question of intralineage genetic exchange has proved more difficult to address definitively, largely because of the greater scarcity of informative markers among more closely related strains. Over the last decade, genome scale analysis combined with more intensive sampling, often at the microscale of individual, undisturbed transmission cycles, revealed evidence of sexually recombining TcI and TcII populations (Barnabe et al., 2013; Berry et al., 2019; Ocaña-Mayorga et al., 2011; Rusman et al.,

2020; Schwabl et al., 2019). In some circumstances recombination appears so widespread that marker frequencies are compatible with panmixia, while in others intra-TcI hybrid genotypes have formed and then spread clonally, as seen for TcV and TcVI. These distinct population structures may even occur in geographically overlapping areas (Schwabl et al., 2019). Lastly, intralineage mitochondrial introgression indicates that some parasite strains are the products of repeated backcrossing of hybrids into one parental genetic background (Rusman et al., 2020).

In summary, there is a broad base of evidence that genetic exchange is an active force shaping *T. cruzi* diversity, with significant epidemiological implications (Messenger et al., 2015a; Ramírez and Llewellyn, 2014, 2015). In some natural populations sex is frequent among closely related *T. cruzi* strains, in others it can give rise to intralineage and, apparently very rarely, interlineage, hybrids that can be maintained clonally as new types. Introgression of kDNA within and across the major lineages, likely a result of backcrossing and/or selfing among hybrids, adds to the complexity. The factors that determine the balance of clonal and sexual reproduction within and between *T. cruzi* lineages remain unknown. Progress may depend on studies of the mechanism of genetic exchange in the laboratory.

3.2 Genetic Crosses

In contrast with *T. brucei*, only two successful genetic crossing experiments have so far been reported for *T. cruzi*. The earliest, and unsuccessful, attempts at genetic crosses involved passaging TcI and TcII together in vitro, through triatomine bugs, or through mice, followed by phenotypic analysis of the resultant populations (Miles, 1985). As with *T. brucei*, such experiments were revolutionized by the ability to genetically transform *T. cruzi* strains to carry different drug resistance markers, permitting selection of any double drug resistant populations emerging from genetic crossing experiments. Perhaps the best experimental strategy for application of this new technology would be an attempt to cross *T. cruzi* strains from within the same DTU, an approach encouraged by the discovery of both putative hybrid and parental TcI phosphoglucomutase isoenzyme phenotypes among clones of *T. cruzi* isolates from a single undisturbed locality in Amazonian forest (Carrasco et al., 1996). Accordingly, a pair of these putative TcI parental isolates was selected, biological clones were prepared, and they were transfected to carry different episomal drug resistance markers for hygromycin and neomycin (G418). The transgenic parental isolates were passaged together through triatomine bugs, mice, and mammalian cell cultures and the recovered populations were cultured in media containing both drugs to select possible hybrids. The only double-drug-resistant clones were derived from the mammalian cell cultures (Stothard et al., 1999). MLEE, karyotyping, gene sequencing, microsatellite profiling and later, whole genome sequencing, demonstrated that the six double-drug-resistant clones were hybrids of the two parental strains (Gaunt et al., 2003; Matos et al., 2022). Thus, it was demonstrated experimentally for the first time that *T. cruzi* has an extant capacity to undergo genetic exchange, at least within TcI.

The next major advance did not come until 2018 when Alves et al. (2018) showed it was possible to generate fused cell hybrids in axenic cell cultures containing only insect stage *T. cruzi* parasites. In these experiments epimastigote cultures were split into two separate flasks and supplemented with a halogenated thymidine analogue, either 5′-chloro-2′-deoxyuridine (CldU) or 5′-iodo-2′-deoxyuridine (IdU); these modified nucleotides become integrated into newly synthesized DNA as the parasites multiply and they can be detected downstream using specific monoclonal antibodies. After 12 hours the parasites were washed to remove unincorporated CldU/IdU and then mixed together again in fresh

medium without CldU/IdU. After 24 hours of co-culture the parasites were fixed and a dual immunofluorescence approach was used to label CldU and IdU with secondary antibodies conjugated with green and red fluorophores, respectively. While the majority of parasites were positive for only one of the analogs, hybrid parasites with both analogs present were also identified. For TcVI-CL Brener the frequency of hybrids was ~5% and for TcII-Y strain it was ~1% of all parasites. Interestingly, when CL Brener parasites were genetically engineered to overexpress the gene *Rad51*, which has a key role in homologous recombination and DNA repair, the frequency of hybrid parasites increased to ~12%. In separate experiments using transgenic parasites carrying G418 or Hygromycin B drug resistance markers, the axenic epimastigote co-culture approach again yielded hybrid parasites with inheritance of both reporter genes.

3.3 Location of Genetic Exchange

In the original *T. cruzi* cross described by Gaunt et al. (2003), nonconfluent mammalian cell cultures were infected with *T. cruzi* cultures containing both infective metacyclic trypomastigotes and noninfective epimastigotes. Once the metacyclic trypomastigotes had invaded the mammalian cells, the epimastigotes were removed by washing the cell monolayers. The infected mammalian cell cultures were maintained for more than 20 days, allowing completion of several rounds of intracellular replication, each of which may take as little as 5 days. The double-drug-resistant hybrids were recovered from the culture supernatant, which contained trypomastigotes released from lysed mammalian cells. Thus, the most obvious interpretation was that hybridization took place intracellularly in a co-infected host cell, and that it therefore involved either amastigotes or trypomastigotes. However, it was also possible that genetic exchange could have taken place between extracellular forms prior to host cell invasion (i.e., epimastigotes or metacyclic trypomastigotes), or after infected cells had ruptured (i.e., new trypomastigotes, including short or slender forms). The production of hybrids in fully axenic epimastigote cultures by Alves et al. (2018) lends weight to this interpretation, although it cannot be ruled out that hybridization can occur at more than one stage of the life cycle.

In terms of opportunity in endemic areas, where natural recombinants do exist, mixed interlineage *T. cruzi* infections and multiclonal intralineage infections are by no means rare in humans or other mammals (Bosseno et al., 1996; Brenière et al., 1998; Fernandes et al., 1999; Gaunt et al., 2003; Messenger et al., 2015c; Vago et al., 2000). Furthermore, sylvatic mammals are likely to be subject to multiple challenge infections, some of which may be orally acquired by consuming triatomine bugs. *Didelphis* can have anal gland infections that include morphological stages typically restricted to the insect vector. Both Gaunt et al. (2003), Alves et al. (2018) reported unsuccessful attempts to cross *T. cruzi* using co-passage through triatomine bugs. Nevertheless, by analogy with *T. brucei* and *Leishmania* where genetic exchange takes place in the tsetse or sand fly vectors respectively, along with the successful in vitro hybridization experiments using epimastigotes (Alves et al., 2018), it would be surprising if genetic exchange in *T. cruzi* did not occur in its insect vector. There would be abundant opportunity for *T. cruzi* to undergo genetic exchange in triatomine bugs. Whereas tsetse flies have limited susceptibility to *T. brucei* and most readily acquire infection during their first feed, triatomine bugs may acquire *T. cruzi* infection at any of the five nymphal stages or as adults. Each bug takes many feeds, any one of which may be infective; infection rates rise with instar examined, overall infection prevalence rates are often 50% and may be much higher; as for mammals, mixed infections in bugs are common (Bosseno et al., 1996; Cardinal et al., 2008; Yeo et al., 2007). Although infection of the

triatomine salivary glands is not described for *T. cruzi*, there is analogous intense infection of the rectum with abundant epimastigotes attached to the epithelium and free in the lumen, as well as metacyclic trypomastigotes. Salivary gland infections cannot be considered a prerequisite for genetic exchange to occur in triatomines—they are not a feature of the *Leishmania* life cycle in sand flies and many examples exist of genetic exchange in other protozoan flagellates in the gut lumen of insect species (Corliss and Sleigh, 1974). There are also prominent candidate physiological triggers in triatomines, for example hormonal changes that govern moulting, known to precipitate genetic exchange in flagellates of other insects, or starvation/nutritional depletion, which appears to encourage dispersive flight in male triatomine bugs.

Clearly, the occurrence of genetic exchange in *T. cruzi* during the vector stage of the life cycle deserves further investigation. With approximately 127 of the 140 or more triatomine species native to the Americas and at least six *T. cruzi* genetic lineages, there are multiple scenarios to be explored and indubitably new discoveries to be made, aided by the latest technological advances.

3.4 Behavior of Experimental Hybrids

The phenomenon of hybrid vigor (heterosis), in which hybrids tolerate unusually severe conditions or thrive and out-compete nonhybrids, is well-known for a variety of organisms (Mallet, 2007). On the other hand, hybrids may have significantly reduced viability and be difficult to recover from experimental or natural populations. Hybrid vigor was explored theoretically by Widmer et al. (1987) for a naturally occurring hybrid *T. cruzi* by comparing the catalytic efficiency of the three individual glucose phosphate isomerase isoenzymes in hybrids, and against the individual isoenzymes in nonhybrids. Although the isoenzymes clearly differed in temperature stabilities their catalytic efficiencies appeared to be similar. Comprehensive phenotypic comparisons of natural TcV and TcVI hybrid strains with TcII and TcIII representative parental strains will be required to establish whether heterosis contributed to the success of TcV and TcVI in becoming established in domestic transmission cycles. In this context it is clearly of interest to know whether experimentally derived hybrid *T. cruzi* clones have decreased or increased vigor. The experimental hybrid clones described by Gaunt et al. (2003) grew vigorously in vitro, and in preliminary in vivo comparisons in immunocompromised SCID mice, the hybrids readily established infections and produced abundant pseudocysts in heart and skeletal muscle; pseudocysts were also seen in smooth muscle of the alimentary tract (Lewis et al., 2009). The hybrid clones therefore appear to be at least as virulent as their parents. Genotype of the infecting strain or mixture of strains may have an important impact upon disease pathology in humans, for instance, as a result of differences in tissue tropism (Vago et al., 1996, 2000). Results from experimental investigations of mixed infection dynamics also suggest that dual-clone mixtures of *T. cruzi* behave differently than would be expected simply based on the behavior of each clone individually, both in bugs (Da Silveira Pinto et al., 2000; Pinto et al., 1998), and in animal models (De Lana et al., 2000; Martins et al., 2006, 2007). Imaging of transgenic bioluminescent biological clones of *T. cruzi* parents and derived hybrid progeny now provides a means of following their comparative behavior and virulence in vivo, throughout the acute and prolonged chronic phases of mouse model infections (Lewis et al., 2014).

3.5 Mechanism of Genetic Exchange in Experimental and Natural Hybrids

The mechanism of genetic exchange involved in experimentally generated hybrid *T. cruzi* clones is at least partially well understood. MLEE, karyotype analysis, random amplification of polymorphic DNA (RAPD) and multilocus microsatellite typing (MLMT) of the hybrids generated by Gaunt et al. (2003) suggested they had not inherited alleles in typical Mendelian ratios, i.e., one allele per locus from each parent. In fact, for most loci tested the hybrids had inherited at least two alleles from both parents, although a minority of the parental alleles were absent. *T. cruzi* hybrid clones each displayed one of the parental mitochondrial maxicircle genotypes but not both, as is reported (above) for *T. brucei*. The dynamics for inheritance of minicircles remains unexplored in *T. cruzi* hybrids. Flow cytometric analysis demonstrated that the DNA content of hybrid clones was approximately intermediate between 3n and 4n, though the method could not differentiate between nuclear and kinetoplast DNA contributions to the total (Lewis et al., 2009). Whole genome sequencing revealed that almost all chromosomes in the hybrids were tetrasomic, with equal contribution from both parents, consistent with the model of diploid cell fusion (Matos et al., 2022). The DNA content of these hybrids was relatively stable after passage through mice (Lewis et al., 2009), though following 800 generations of in vitro culture, progressive declines in DNA content were observed and more aneuploid karyotypes emerged, with a mixed pattern of trisomies and tetrasomies (Matos et al., 2022). Thus, tetraploid hybrids undergo genome erosion through apparently stochastic loss of individual chromosome copies, without any evidence of a meiotic reductional division, which would be expected to result in a halving of ploidy in a single step.

This mode of hybridization is in contrast with the general consensus for the typical program of genetic exchange in *T. brucei*, which evidently does involve meiosis and Mendelian inheritance. Furthermore, analysis of natural *T. cruzi* hybrids (TcV and TcVI) showed that their genotypes were consistent with them being typical meiotic F_1 progeny from a TcII x TcIII cross (Lewis et al., 2011). Meiosis was strongly implicated in the natural panmictic population of TcI discovered in Ecuador, for example by highly mosaic karyotypes reflecting frequent chromosomal crossovers (Schwabl et al., 2019). Nevertheless, more than one mechanism may occur. Schwabl et al. (2019) also described other TcI genomes that are more consistent with the parasexual diploid fusion−erosion mechanism of hybridization, and this has a biological precedent in the pathogenic yeast *Candida albicans* (Bennett and Johnson, 2003). In *C. albicans* the genome erosion following fusion of diploids is by random, concerted loss of whole chromosomes over the course of repeated mitotic replication and results in diploid or near-diploid recombinants.

The interesting results of the DNA content analysis of the experimental cross, in particular the elevated DNA content of the hybrids, led us to undertake a much wider study of DNA content among field isolates representing all six of the known *T. cruzi* genetic lineages. As described by Lewis et al. (2009) and in agreement with the pioneering observations by Dvorak et al. (1982), extraordinarily wide diversity in DNA content was observed (Fig. 24.2). Some trends were observed relating to the DNA content of cloned strains from the different genetic lineages, for example, TcI on average clearly had the lowest DNA content, with a few outlying isolates, while TcII and TcIV presented a wide range of DNA contents. The key question, however, was whether naturally occurring hybrid TcV and TcVI strains had elevated DNA contents, compatible with the fusion of diploids, as deduced from the experimental cross. This proved not to be the case. TcV and TcVI had DNA contents equivalent to those of their parental TcII and TcIII lineages, implying diploidy. This was confirmed by genome-scale

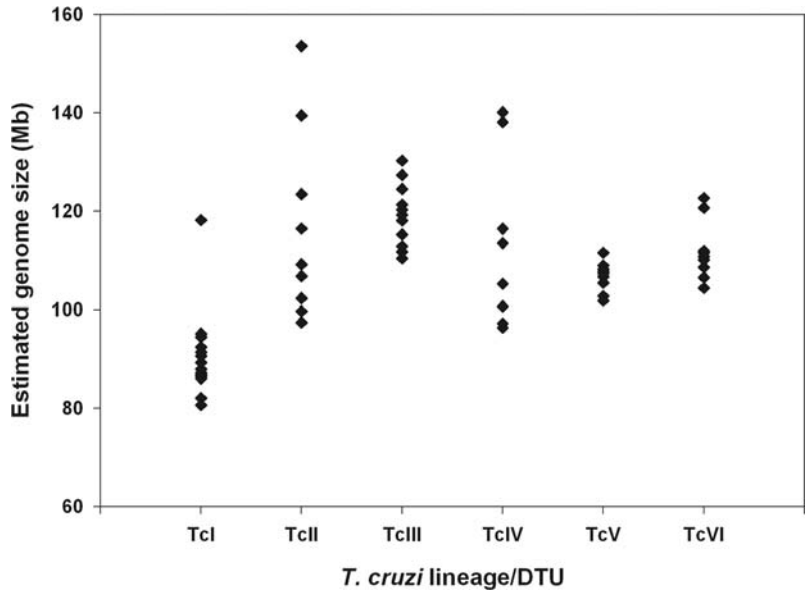


FIGURE 24.2 Graph of DNA contents.

Variation in DNA content within and between *Trypanosoma cruzi* genetic lineages or DTUs. Each diamond indicates a different *T. cruzi* clone. Flow cytometry was used to measure the relative DNA content of parasites labeled with the fluorescent dye propidium iodide. Genome sizes were estimated based on the predicted size of the CL Brener genome (El-Sayed et al., 2005).

Adapted from Lewis, M.D. et al., 2009. Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in Trypanosoma cruzi populations and expose contrasts between natural and experimental hybrids. International Journal for Parasitology 39 (12), pp. 1305–1317. https://doi.org/10.1016/j.ijpara.2009.04.001

MLMT and DNA sequencing of nuclear and mitochondrial (maxicircle) genes in TcII, TcIII, TcV, and TcVI isolates, which indicated TcV and TcVI carried one TcII allele and one TcIII allele per locus, with only rare instances of allelic aneuploidy (Lewis et al., 2009). Only TcIII kDNA maxicircles were inherited, whereas a mixture of TcII and TcIII kDNA minicircles have been retained (Rusman et al., 2019). TcV and TcVI present very little intralineage diversity but the extensive genotyping clearly shows they are distinct from each other, confirming the original MLEE-based distinction. TcV and TcVI thus resemble normal F_1 meiotic progeny of hybridization events between TcII and TcIII that have undergone clonal expansion within the domestic niche—they are essentially diploid with fixed heterozygous genotypes comprising equal proportions of TcII and TcIII alleles. There are therefore contrasts with the aneuploid experimental hybrids. These could stem from mechanistic differences in interlineage, as opposed to intralineage, recombination. Alternatively, the physiological cues that might be required for meiotic reductive division may well have been absent under the conditions used for experimental crosses so far. It is not known how the diploid state was reached by TcV/VI and so operation of the fusion-erosion mechanism or other genetic exchange mechanisms in natural recombination events should not be ruled out at this stage.

3.6 Implications for Epidemiology

The successful experimental *T. cruzi* genetic crosses have proved that this organism has an extant capacity for genetic exchange, with consequent epidemiological implications, and the unusual genetic mechanism continues to be of considerable fundamental interest. However, the main implications of genetic exchange in *T. cruzi* come not from examination of the experimental hybrids but from genetic analysis of natural populations. The hybrid DTUs, TcV and TcVI, have thrived and spread dramatically among human populations in Bolivia, Argentina, Chile, Paraguay, and the extreme south of Brazil, where severe clinical manifestations of Chagas disease are common, including chagasic cardiomyopathy, megaesophagus, megacolon, and congenital transmission (Messenger et al., 2015c; Miles et al., 2009a,b; WHO, 2002). TcV and TcVI have probably been propagated in conjunction with the dispersion of *Triatoma infestans*, the principal domestic triatomine vector in the Southern Cone region, which was itself spread in association with the activities and migration of humans (Bargues et al., 2006; Cortez et al., 2010). There is therefore potentially a great epidemiological risk attached to the emergence of new recombinant lineages, which would have unpredictable phenotypes, for example, with respect to pathogenicity or transmission potential. Furthermore, new high impact hybridization events cannot be excluded as the dynamics of sylvatic and domestic transmission cycles alter with development, environmental change, and migration human of populations.

4. Leishmania

4.1 Evidence of Genetic Exchange in Natural Populations

Numerous species of *Leishmania* have been named as agents of leishmaniasis. The multiple species of *Leishmania* and their complex taxonomy contrasts with *T. cruzi* as the sole designated species causing Chagas disease. This taxonomic contrast is despite the remarkable diversity of the *T. cruzi* genetic lineages, exceeding the level of diversity that is used to support definition of several distinct *Leishmania* species (Miles et al., 2009). In fact, there has been a tendency to describe new *Leishmania* species based on minor genotypic differences; some species have been shown to be synonyms, e.g., *Leishmania infantum* (synonym *Leishmania chagasi)*; *Leishmania donovani* (synonym *Leishmania archibaldi*), and several other species names are of questionable validity (Fraga et al., 2010; Lukeš et al., 2007; Mauricio et al., 2006). Nevertheless, molecular methods have also revealed hitherto unexpected genetic diversity within species, for example within *L. donovani*.

Two of four subgenera of *Leishmania*, subgenus *Leishmania* and subgenus *Viannia*, are agents of human disease, which were initially distinguished based on differences in clinical presentation, and subsequently on the distribution of infection within the alimentary tract of the sand fly vectors. The *Leishmania donovani* complex (subgenus *Leishmania*; *L. donovani* and *L. infantum*) causes human visceral leishmaniasis (VL, kala-azar), which is the most devastating form of human leishmaniasis. *Leishmania infantum* is also responsible for widespread severe canine visceral leishmaniasis (CVL). Symptomatic human VL is nearly always fatal in the absence of treatment, with a current estimate of 50,000 to 90,000 cases annually (Alvar et al., 2012; Valdivia et al., 2022; WHO, 2010). Several other species of the subgenus *Leishmania* are responsible for diverse clinical presentations of cutaneous disease in Africa-Eurasia (*Leishmania major*; *Leishmania tropica*; *Leishmania aethiopica*) and the Americas (*Leishmania amazonensis*; *Leishmania mexicana*), with some species (*L. aethiopica*, *L. amazonensis*) occasionally presenting with disseminated cutaneous lesions.

The subgenus *Viannia* is confined to the Americas, where *Leishmania braziliensis* causes the most severe form of the cutaneous disease, mucocutaneous leishmaniasis (MCL), with potentially catastrophic metastatic invasion and destruction of the nasopharynx. *Leishmania peruviana* is very closely related to *L. braziliensis* but is more prevalent in higher altitude regions of Peru, where in contrast it is associated with simple, self-healing cutaneous lesions. Infections of *Leishmania guyanensis* and *Leishmania panamensis,* both within the subgenus *Viannia*, may spread beyond the initial site of infection but are usually less aggressive than *L. braziliensis*.

In parallel with research on *T. cruzi* diversity, a similar range of molecular methods has been applied to the genus *Leishmania*. Historically, as with *T. cruzi*, population structure of all *Leishmania* was considered fundamentally clonal. However, evidence for genetic exchange in natural populations of *Leishmania* has repeatedly emerged in the form of both interspecific and intraspecific hybrids bearing recombinant genotypes. For example, natural hybrids have been described between *L. braziliensis* and *L. panamensis* in Nicaragua (Belli et al., 1994), between *L. braziliensis* and *L. guyanensis* in Venezuela (Delgado et al., 1997), and between *L. braziliensis* and *L. peruviana* in Peru (Dujardin et al., 1995). In the latter case, the *L. braziliensis/peruviana* hybrids were highly prevalent among patients and occurred in some patients with mucosal disease. Furthermore, nine hybrid genotypes were discovered in a single Peruvian endemic region (Nolder et al., 2007). Inbreeding was detected in natural populations of both *Leishmania braziliensis* and *Leishmania guyanensis* based on linkage disequilibrium (LD), with a deficit of heterozygosity (Rougeron et al., 2011). In a more recent wide-ranging study of the *Leishmania braziliensis/peruviana* complex, in Andean and Amazonian Peru, genomic hybrids were detected and were compatible with meiotic recombination (Van den Broeck et al., 2020).

In Africa-Eurasia *L. major/arabica* hybrids have been described (Evans et al., 1987; Kelly et al., 1991), and *L. major/donovani* hybrids (originally designated *L. major/infantum* hybrids) from patients with HIV infection (Ravel et al., 2006). A widespread lineage of *L. tropica* appears to be disseminated from a recombination event (Schwenkenbecher et al., 2006). Comparative analyses of *L. tropica* genomic diversity in the Middle East (Glans et al., 2021) indicated presence of frequent meiotic recombination, and also implied parasexual genetic exchange.

Putative parental and hybrid phenotypes of the *L. donovani* complex (*L. donovani*; "*L.* archibaldi") occur sympatrically in East Africa and sequencing of housekeeping genes encoding enzymes shows mosaic characters across such strains (Lukeš et al., 2007). Comparative genome sequencing to detect recombination showed evidence of hybridization and subsequent selfing among vector-isolated *L. infantum* from south-eastern Turkey (Rogers et al., 2014). Multilocus sequence typing and microsatellites confirmed the presence of sympatric putative parents and hybrid progeny in an endemic focus of VL in Ethiopia (Gelanew et al., 2014). Follow-up genome sequencing analyses of isolates of *L. donovani* from Ethiopia showed recombinant hybrids between two natural lineages (Cotton et al., 2020). In comparative genomics of 204 strains of *L. donovani* from the Indian subcontinent, Imamura et al. (2016) found evidence of genetic exchange, associated with the spread of the LdAQP1 mutation marker of resistance to antimonials. Thus, there is an accumulation of evidence of intra- and interspecific genetic exchange in natural populations of *Leishmania*. Also in India, *L. donovani* hybrids were associated with cutaneous clinical presentation (Lypaczewski et al., 2022). Comparative genome analyses of *L. donovani* and *L. infantum*, based on 151 global isolates, showed extraordinary diversity and multiple hybrid lineages, although lesser diversity within *L. infantum* (Franssen et al., 2020).

4.2 Genetic Crosses in Sand Flies

Akopyants et al. (2009) finally demonstrated recombination in *Leishmania* experimentally. They co-transmitted pairs of transgenic *L. major* strains resistant to different selective drugs through the natural sand fly vector, *Phlebotomus duboscqi*, and recovered parasites that were resistant to both drugs. Nine double drug resistant populations were recovered directly from co-infected sand flies and two from mice bitten by such flies; these yielded a total of 18 progeny clones. Genetic analysis of homozygous parental markers, assigned to several chromosomes, showed that all clones carried both parental alleles, proving they were hybrid. Maxicircle kDNA genotypes were uniparentally inherited. Phenotypic analysis also demonstrated clear segregation of dominant traits in different hybrid progeny. The genotypes were consistent with canonical meiosis generating heterozygous F1 progeny. However, 7 of the 18 clones were triploid. Experimental crosses using parental clones carrying red and green fluorescence markers in conjunction with separation of single cells by fluorescent activated cell sorting revealed the presence of "yellow" hybrid cells with expression of both fluorescent markers (Sadlova et al., 2011).

Subsequently, it was shown that genetic crosses, in both the natural vector *P. duboscqi* and a nonnatural permissive vector *L. longipalpis*, could be performed with facility and between geographically diverse *L. major* strains, with similar efficiencies in all pairwise crosses (Inbar et al., 2015). Of 96 derived clonal lines the majority were diploid, but some were triploid or tetraploid. A low frequency of uniparental allelic inheritance, more so in the nonnatural vector species, was attributed to loss of heterozygosity. Again, only uniparental kinetoplastid maxicircle inheritance occurred. No distinct male or female gametes have been described but of the developmental stages in the sand fly (procyclic, nectomonad, haptomonad, and metacyclic) the nectomonad stage is considered the most likely stage at which genetic exchange takes place (Inbar et al., 2015). In further intraspecific and interspecific crosses of *L. major*, *L. infantum*, and *L. tropica* chromosomal inheritance patterns were 97%–99% compatible with meiosis (Inbar et al., 2019).

Consistent with the observation of *L. major*/*L. donovani* hybrids in natural populations (Ravel et al., 2006; Romano et al., 2014) achieved interspecies crosses of *L. major* and *L. infantum* in the permissive vector *L. longipalpis*, yielding hybrid progeny that displayed similar patterns of inheritance and ploidy among hybrids from the *L. major* crosses. Most importantly, trait segregation of cutaneous or visceral dissemination and pathology was seen among the progeny. Calvo-Álvarez et al. (2014) showed that intraclonal crosses of *L. infantum* in a natural vector *Phlebotomus perniciosus* could produce a viable hybrid, somewhat less virulent in a mouse model than the parental clone. In permissive or appropriately specific sand fly species other clonal hybrid *Leishmania* species can be generated (Dostálová and Volf, 2012).

4.3 Genetic Crosses In Vitro

In vitro production of genetic hybrids of *Leishmania* has recently been achieved by the novel approaches of two research groups. Previously hybrids were only generated in vitro by conducting crosses between selected robust strains of *L. tropica*. Louradour et al. (2022) have applied irradiation-stress to *Leishmania* DNA to encourage hybridization, and have successfully obtained intraspecific and interspecific hybrids of *L. donovani*, *L. infantum*, *L. braziliensis*, and *L. tropica*. Although productive, the hybrids were predominantly tetraploid, without meiosis, which limits the

overall comparative applications of the genomics. Single cell RNA sequencing indicated expression of the HAP2 gamete fusogen in the hybrids. The methodology is summarized by Ferreira and Sacks (2022), Monte-Neto et al. (2022), Negreira et al. (2022), also applying single cell sequencing, found rapid changes in karyotype aneuploidy, possibly associated with hybridization.

Gutiérrez-Corbo et al. (2022) used *L. donovani*, *L. tropica*, and *L. major*, carrying fluorescence and antibiotic resistance markers, for intraspecies and interspecific crosses in vitro. Some crosses were set up in the presence of an embryonic *Lutzomyia longipalpis* cell line. Intra-*L. tropica* crosses and inter-*L. donovani*/*L. major* and *L. donovani*/*L. tropica* were obtained. Compared to parents, the DNA contents of hybrids were increased, and phenotypes diverged from those of the parents.

4.4 Implications for Epidemiology

Research progress on genetic exchange in *Leishmania* and in trypanosomes has now been dramatic and far reaching, albeit much remains to be done to understand precise alternative mechanisms. Among others, Volf et al. (2007), Volf and Sadlova (2009) have shown that *L. major/donovani* hybrids have increased transmission potential, since they were transmitted efficiently by the otherwise *L. major*-specific vector, *Phlebotomus papatasi*. Such results highlight the potential epidemiological impact of genetic exchange in *Leishmania*: hybrids may have the disturbing potential to expand transmission of visceral disease and invade new geographical regions.

Thus, the fact that *Leishmania* spp. can undergo genetic exchange is of profound epidemiological importance (Miles et al., 2009). Genetic exchange has implications for heterosis (hybrid vigor), the emergence and spread of virulent strains, resistance to chemotherapeutics, exploitation of different vectors and hosts, and adaptation to new ecological niches that may provide a selective advantage. As with *T. brucei* and *T. cruzi*, virulent clones may emerge and predominate in some epidemiologically important populations. The perception that genetic exchange is rare and of little evolutionary or extant epidemiological consequence is no longer tenable. Analysis of genomic diversity in experimental genetic crosses and in natural populations can provide insight that may facilitate control of leishmaniasis.

Abbreviations

MLEE Multilocus enzyme electrophoresis
MLMT Multilocus microsatellite typing
MLST Multilocus sequence typing

Acknowledgments

We thank the Wellcome Trust, the European Union Seventh Framework Program, contract number 223034 (ChagasEpiNet) for financial support. We especially thank many current and previous research collaborators working on *Trypanosoma brucei*, *Trypanosoma cruz*i, and *Leishmania*.

References

Adams, E.R., et al., 2006. Trypanosome identification in wild tsetse populations in Tanzania using generic primers to amplify the ribosomal RNA ITS-1 region. Acta Tropica 100 (1–2), 103–109. https://doi.org/10.1016/j.actatropica.2006.10.002.

Akopyants, N.S., et al., 2009. Demonstration of genetic exchange during cyclical development of Leishmania in the sand fly vector. Science 324 (5924), 265–268. https://doi.org/10.1126/science.1169464.

Alvar, J., et al., 2012. Leishmaniasis worldwide and global estimates of its incidence. PLoS One 7 (5), e35671. https://doi.org/10.1371/journal.pone.0035671.

Alves, C.L., et al., 2018. The recombinase Rad51 plays a key role in events of genetic exchange in Trypanosoma cruzi. Sci. Rep. 8 (1). https://doi.org/10.1038/s41598-018-31541-z.

Balmer, O., et al., 2011. Phylogeography and taxonomy of Trypanosoma brucei. PLoS Neglected Trop. Dis. 5 (2). https://doi.org/10.1371/journal.pntd.0000961.

Balmer, O., Tanner, M., 2011. Prevalence and implications of multiple-strain infections. Lancet Infect. Dis. 11 (11), 868–878. https://doi.org/10.1016/S1473-3099(11)70241-9.

Bargues, M.D., et al., 2006. Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size. Infect. Genet. Evol. 6 (1), 46–62. https://doi.org/10.1016/j.meegid.2005.01.006.

Barnabe, C., et al., 2013. Putative panmixia in restricted populations of Trypanosoma cruzi isolated from wild Triatoma infestans in Bolivia. PLoS One 8 (11), e82269. https://doi.org/10.1371/journal.pone.0082269.

Barnabé, C., Brisse, S., Tibayrenc, M., 2000. Population structure and genetic typing of Trypanosoma cruzi, the agent of Chagas disease: a multilocus enzyme electrophoresis approach. Parasitology 120 (5), 513–526. https://doi.org/10.1017/S0031182099005661.

Belli, A.A., Miles, M.A., Kelly, J.M., 1994. A putative Leishmania panamensis/Leishmania braziliensis hybrid is a causative agent of human cutaneous leishmaniasis in Nicaragua. Parasitology 109 (4), 435–442. https://doi.org/10.1017/S0031182000080689.

Bennett, R.J., Johnson, A.D., 2003. Completion of a parasexual cycle in Candida albicans by induced chromosome loss in tetraploid strains. EMBO J. 22 (10), 2505–2515. https://doi.org/10.1093/emboj/cdg235.

Berriman, M., et al., 2005. The genome of the African trypanosome Trypanosoma brucei. Science 309 (5733), 416–422. https://doi.org/10.1126/science.1112642.

Berry, A.S.F., et al., 2019. Sexual reproduction in a natural Trypanosoma cruzi population. PLoS Neglected Trop. Dis. 13 (5), e0007392. https://doi.org/10.1371/journal.pntd.0007392.

Bhattacharyya, T., et al., 2015. Prospects for T. cruzi lineage-specific serological surveillance of wild mammals. Acta Trop. 151 (1), 182–186. https://doi.org/10.1016/j.actatropica.2015.06.017.

Bingle, L.E.H., et al., 2001. A novel GFP approach for the analysis of genetic exchange in trypanosomes allowing the in situ detection of mating events. Microbiology 147 (12), 3231–3240. https://doi.org/10.1099/00221287-147-12-3231.

Bosseno, M.F., et al., 1996. Trypanosoma cruzi: study of the distribution of two widespread clonal genotypes in Bolivian Triatoma infestans vectors shows a high frequency of mixed infections. Exp. Parasitol. 83 (3), 275–282. https://doi.org/10.1006/expr.1996.0075.

Brenière, S.F., et al., 1998. Trypanosoma cruzi genotypes associated with domestic Triatoma sordida in Bolivia. Acta Trop. 71 (3), 269–283. https://doi.org/10.1016/S0001-706X(98)00061-8.

Burgos, J.M., et al., 2008. Case report: molecular identification of Trypanosoma cruzi i tropism for central nervous system in chagas reactivation due to AIDS. Am. J. Trop. Med. Hyg. 78 (2), 294–297. https://doi.org/10.4269/ajtmh.2008.78.294.

Calvo-Álvarez, E., et al., 2014. First evidence of intraclonal genetic exchange in trypanosomatids using two leishmania infantum fluorescent transgenic clones. PLoS Neglected Trop. Dis. 8 (9), e3075. https://doi.org/10.1371/journal.pntd.0003075.

Cardinal, M.V., et al., 2008. Molecular epidemiology of domestic and sylvatic Trypanosoma cruzi infection in rural northwestern Argentina. Int. J. Parasitol. 38 (13), 1533–1543. https://doi.org/10.1016/j.ijpara.2008.04.010.

Carrasco, H.J., et al., 1996. Genetic exchange as a possible source of genomic diversity in sylvatic populations of Trypanosoma cruzi. Am. J. Trop. Med. Hyg. 54 (4), 418–424. https://doi.org/10.4269/ajtmh.1996.54.418.

Carrasco, H.J., et al., 2012. Geographical distribution of trypanosoma cruzi genotypes in Venezuela. PLoS Neglected Trop. Dis. 6 (6). https://doi.org/10.1371/journal.pntd.0001707.

Cibulskis, R.E., 1992. Genetic variation in trypanosoma brucei and the epidemiology of sleeping sickness in the lambwe valley, Kenya. Parasitology 104 (1), 99–109. https://doi.org/10.1017/S0031182000060844.

Cooper, A., et al., 2008. Genetic analysis of the human infective trypanosome Trypanosoma brucei gambiense: chromosomal segregation, crossing over, and the construction of a genetic map. Genome Biol. 9 (6). https://doi.org/10.1186/gb-2008-9-6-r103.

Corliss, J.O., Sleigh, M.A., 1974. The biology of Protozoa. Trans. Am. Microsc. Soc. 93 (2), 291. https://doi.org/10.2307/3225321.

Cortez, M.R., Monteiro, F.A., Noireau, F., 2010. New insights on the spread of Triatoma infestans from Bolivia-implications for Chagas disease emergence in the Southern Cone. Infect. Genet. Evol. 10 (2), 350–353. https://doi.org/10.1016/j.meegid.2009.12.006.

Cotton, J.A., et al., 2020. Genomic analysis of natural intra-specific hybrids among Ethiopian isolates of Leishmania donovani. PLoS Neglected Trop. Dis. 14 (4), 1–26. https://doi.org/10.1371/journal.pntd.0007143.

Da Silveira Pinto, A., et al., 2000. Experimental Trypanosoma cruzi biclonal infection in Triatoma infestans: detection of distinct clonal genotypes using kinetoplast DNA probes. Int. J. Parasitol. 30 (7), 843–848. https://doi.org/10.1016/S0020-7519(00)00058-8.

De Greef, C., et al., 1989. A gene expressed only in serum-resistant variants of Trypanosoma brucei rhodesiense. Mol. Biochem. Parasitol. 36 (2), 169–176. https://doi.org/10.1016/0166-6851(89)90189-8.

De Lana, M., et al., 2000. Trypanosoma cruzi: infectivity of clonal genotype infections in acute and chronic phases in mice. Exp. Parasitol. 96 (2), 61–66. https://doi.org/10.1006/expr.2000.4552.

Degen, R., et al., 1995. Sexual compatibility among Trypanosoma brucei isolates from an epidemic area in southeastern Uganda. Parasitol. Res. 81 (3), 253–257. https://doi.org/10.1007/BF00937118.

Delgado, O., et al., 1997. Cutaneous leishmaniasis in Venezuela caused by infection with a new hybrid between Leishmania (Viannia) braziliensis and L. (V.) guyanensis. Memórias do Instituto Oswaldo Cruz 92 (5), 581–582. https://doi.org/10.1590/S0074-02761997000500002.

Dostálová, A., Volf, P., 2012. Leishmania development in sand flies: parasite-vector interactions overview. Parasit. Vectors 5 (1). https://doi.org/10.1186/1756-3305-5-276.

Dujardin, J.C., et al., 1995. Putative Leishmania hybrids in the Eastern Andean valley of Huanuco, Peru. Acta Trop. 59 (4), 293–307. https://doi.org/10.1016/0001-706X(95)00094-U.

Dukes, P., et al., 1989. A new method for isolating trypanosoma brucei gambiense from sleeping sickness patients. Trans. R. Soc. Trop. Med. Hyg. 83 (5), 636–639. https://doi.org/10.1016/0035-9203(89)90379-9.

Dvorak, J.A., et al., 1982. Trypanosoma cruzi : flow cytometric analysis. I. Analysis of total DNA/organism by means of Mithramycin-induced fluorescence 1, 2. J. Protozool. 9 (3), 430–437. https://doi.org/10.1111/j.1550-7408.1982.tb05427.x.

Echodu, R., et al., 2015. Genetic diversity and population structure of Trypanosoma brucei in Uganda: implications for the epidemiology of sleeping sickness and Nagana. PLoS Neglected Trop. Dis. 9 (2), e0003353. https://doi.org/10.1371/journal.pntd.0003353.

Eid, J., Sollner-Webb, B., 1991. Stable integrative transformation of Trypanosoma brucei that occurs exclusively by homologous recombination. Proc. Natl. Acad. Sci. U. S. A. 88 (6), 2118–2121. https://doi.org/10.1073/pnas.88.6.2118.

El-Sayed, N.M., et al., 2005. The genome sequence of Trypanosoma cruzi, etiologic agent of chagas disease. Science 309 (5733), 409−435. https://doi.org/10.1126/science.1112631.

Enyaru, J.C.K., et al., 1993. Isoenzyme comparison of Trypanozoon isolates from two sleeping sickness areas of south-eastern Uganda. Acta Trop. 55 (3), 97−115. https://doi.org/10.1016/0001-706X(93)90072-J.

Evans, D.A., et al., 1987. Hybrid formation within the genus Leishmania? Parassitologia 29 (2−3), 165−173.

Evans, D.A., Ellis, D.S., 1983. Recent observations on the behaviour of certain trypanosomes within their insect hosts. Adv. Parasitol. 22, 1−42. https://doi.org/10.1016/S0065-308X(08)60460-1.

Fedry, J., et al., 2018. Evolutionary diversification of the HAP2 membrane insertion motifs to drive gamete fusion across eukaryotes. PLoS Biol. 16 (8). https://doi.org/10.1371/journal.pbio.2006357.

Fédry, J., et al., 2017. The ancient gamete fusogen HAP2 is a eukaryotic class II fusion protein. Cell. France 168 (5), 904−915. https://doi.org/10.1016/j.cell.2017.01.024.

Fernandes, O., et al., 1998. Brazilian isolates of Trypanosoma cruzi from humans and triatomines classified into two lineages using mini-exon and ribosomal RNA sequences. Am. J. Trop. Med. Hyg. 58 (6), 807−811. https://doi.org/10.4269/ajtmh.1998.58.807.

Fernandes, O., et al., 1999. The complexity of the sylvatic cycle of Trypanosoma cruzi in Rio de Janeiro state (Brazil) revealed by the non-transcribed spacer of the mini-exon gene. Parasitology 118 (2), 161−166. https://doi.org/10.1017/S0031182098003709.

Ferreira, T.R., Sacks, D.L., 2022. Experimental hybridization in Leishmania: tools for the study of genetic exchange. Pathogens 11 (5). https://doi.org/10.3390/pathogens11050580.

Flores-López, C.A., Machado, C.A., 2011. Analyses of 32 loci clarify phylogenetic relationships among Trypanosoma cruzi lineages and support a single hybridization prior to human contact. PLoS Neglected Trop. Dis. 5 (8). https://doi.org/10.1371/journal.pntd.0001272.

Fraga, J., et al., 2010. Phylogeny of Leishmania species based on the heat-shock protein 70 gene. Infect. Genet. Evol. 10 (2), 238−245. https://doi.org/10.1016/j.meegid.2009.11.007.

Franssen, S.U., et al., 2020. Global genome diversity of the Leishmania donovani complex. Elife 9. https://doi.org/10.7554/eLife.51243.

Gaunt, M.W., et al., 2003. Mechanism of genetic exchange in American trypanosomes. Nature 421 (6926), 936−939. https://doi.org/10.1038/nature01438.

Gelanew, T., et al., 2014. Multilocus sequence and microsatellite identification of intra-specific hybrids and ancestor-like donors among natural Ethiopian isolates of Leishmania donovani. Int. J. Parasitol. 44 (10), 751−757. https://doi.org/10.1016/j.ijpara.2014.05.008.

Gibson, W., et al., 2015. Genetic recombination between human and animal parasites creates novel strains of human pathogen. PLoS Neglected Trop. Dis. 9 (3), e0003665. https://doi.org/10.1371/journal.pntd.0003665.

Gibson, W., et al., 1997a. Intraclonal mating in Trypanosoma brucei is associated with out-crossing. Microbiology 143 (3), 909−920. https://doi.org/10.1099/00221287-143-3-909.

Gibson, W., et al., 2008. The use of yellow fluorescent hybrids to indicate mating in Trypanosoma brucei. Parasit. Vectors 1 (1). https://doi.org/10.1186/1756-3305-1-4.

Gibson, W., Bailey, M., 1994. Genetic exchange in Trypanosoma brucei: evidence for meiosis from analysis of a cross between drug-resistant transformants. Mol. Biochem. Parasitol. 64 (2), 241−252. https://doi.org/10.1016/0166-6851(94)00017-4.

Gibson, W.C., 1989. Analysis of a genetic cross between Trypanosoma brucei rhodesiense and T. b. brucei. Parasitology 99 (3), 391−402. https://doi.org/10.1017/S0031182000059114.

Gibson, W.C., 1986. Will the real Trypanosoma b. gambiense please stand up. Parasitol. Today 2 (9), 255−257. https://doi.org/10.1016/0169-4758(86)90011-6.

Gibson, W.C., Gashumba, J.K., 1983. Isoenzyme characterization of some Trypanozoon stocks from a recent trypanosomiasis epidemic in Uganda. Trans. R. Soc. Trop. Med. Hyg. 77 (1), 114−118. https://doi.org/10.1016/0035-9203(83)90033-0.

Gibson, W.C., Marshall, T.F. de C., Godfrey, D.G., 1980. Numerical analysis of enzyme polymorphism: a new approach to the epidemiology and taxonomy of trypanosomes of the subgenus trypanozoon. Adv. Parasitol. 18 (C), 175−246. https://doi.org/10.1016/S0065-308X(08)60400-5.

Gibson, W.C., Mizen, V.H., 1997. Heritability of the trait for human infectivity in genetic crosses of Trypanosoma brucei ssp. Trans. R. Soc. Trop. Med. Hyg. 91 (2), 236−237. https://doi.org/10.1016/S0035-9203(97)90236-4.

Gibson, W.C., Wellde, B.T., 1985. Characterization of trypanozoon stocks from the south nyanza sleeping sickness focus in western Kenya. Trans. R. Soc. Trop. Med. Hyg. 79 (5), 671−676. https://doi.org/10.1016/0035-9203(85)90187-7.

Gibson, W., Crow, M., Kearns, J., 1997b. Kinetoplast DNA minicircles are inherited from both parents in genetic crosses of Trypanosoma brucei. Parasitol. Res. 83 (5), 483−488. https://doi.org/10.1007/s004360050284.

Gibson, W., Garside, L., 1991. Genetic exchange in Trypanosoma brucei brucei: variable chromosomal location of housekeeping genes in different trypanosome stocks. Mol. Biochem. Parasitol. 45 (1), 77−89. https://doi.org/10.1016/0166-6851(91)90029-6.

Gibson, W., Garside, L., 1990. Kinetoplast DNA mini-circles are inherited from both parents in genetic hybrids of Trypanosoma brucei. Mol. Biochem. 42, 45−54.

Gibson, W., Garside, L., Bailey, M., 1992. Trisomy and chromosome size changes in hybrid trypanosomes from a genetic cross between Trypanosoma brucei rhodesiense and T. b. brucei. Mol. Biochem. Parasitol. 51 (2), 189−199. https://doi.org/10.1016/0166-6851(92)90069-V.

Gibson, W., Kanmogne, G., Bailey, M., 1995. A successful backcross in Trypanosoma brucei. Mol. Biochem. Parasitol. 69 (1), 101−110. https://doi.org/10.1016/0166-6851(94)00196-T.

Gibson, W., Stevens, J., 1999. Genetic exchange in the Trypanosomatidae. Adv. Parasitol. 43, 1−46. https://doi.org/10.1016/s0065-308x(08)60240-7.

Gibson, W., Whittington, H., 1993. Genetic exchange in Trypanosoma brucei: selection of hybrid trypanosomes by introduction of genes conferring drug resistance. Mol. Biochem. Parasitol. 60 (1), 19−26. https://doi.org/10.1016/0166-6851(93)90024-R.

Glans, H., et al., 2021. High genome plasticity and frequent genetic exchange in Leishmania tropica isolates from Afghanistan, Iran and Syria. PLoS Neglected Trop. Dis. 15 (12), e0010110. https://doi.org/10.1371/journal.pntd.0010110.

Godfrey, D.G., Killick-Kendrick, R., 1961. Bovine trypanosomiasis in Nigeria. I. The inoculation of blood into rats as a method of survey in the Donga Valley, Benue province. Ann. Trop. Med. Parasitol. 55, 287−292.

Gottesdiener, K., et al., 1990. Chromosome organization of the protozoan Trypanosoma brucei. Mol. Cell Biol. 10 (11), 6079−6083. https://doi.org/10.1128/MCB.10.11.6079.

Gutiérrez-Corbo, C., et al., 2022. Axenic interspecies and intraclonal hybrid formation in Leishmania: successful crossings between visceral and cutaneous strains. PLoS Neglected Trop. Dis. 16 (2). https://doi.org/10.1371/journal.pntd.0010170.

Hide, G., et al., 1994. Epidemiological relationships of trypanosoma brucei stocks from south east Uganda evidence for different population structures in human infective and non-human infective isolates. Parasitology 109 (1), 95−111. https://doi.org/10.1017/S0031182000077805.

Howick, V.M., et al., 2022. Single-cell transcriptomics reveals expression profiles of Trypanosoma brucei sexual stages. PLoS Pathog. 18 (3). https://doi.org/10.1371/journal.ppat.1010346.

Hutchinson, S., et al., 2021. The establishment of variant surface glycoprotein monoallelic expression revealed by single-cell RNA-seq of Trypanosoma brucei in the tsetse fly salivary glands. PLoS Pathog. 17 (9), e1009904. https://doi.org/10.1371/journal.ppat.1009904.

Imamura, H., et al., 2016. Evolutionary genomics of epidemic visceral leishmaniasis in the Indian subcontinent. Elife 5. https://doi.org/10.7554/eLife.12613.

Inbar, E., et al., 2015. The mating competence of geographically diverse Leishmania major strains in their natural and unnatural sand fly vectors. PLoS Genet. 9.

Inbar, E., et al., 2019. Whole genome sequencing of experimental hybrids supports meiosis-like sexual recombination in leishmania. PLoS Genet. 15 (5). https://doi.org/10.1371/journal.pgen.1008042.

Ivens, A.C., et al., 2005. The genome of the kinetoplastid parasite, Leishmania major. Science 309 (5733), 436−442. https://doi.org/10.1126/science.1112680.

Jenni, L., et al., 1986. Hybrid formation between African trypanosomes during cyclical transmission. Nature 322 (6075), 173−175. https://doi.org/10.1038/322173a0.

Kay, C., et al., 2022. Signatures of hybridization in Trypanosoma brucei. PLoS Pathog. 18 (2). https://doi.org/10.1371/journal.ppat.1010300.

Kelly, J.M., et al., 1991. Evidence of genetic recombination in Leishmania. Mol. Biochem. Parasitol. 46 (2), 253−263. https://doi.org/10.1016/0166-6851(91)90049-C.

Komba, E.K., et al., 1997. Genetic diversity among Trypanosoma brucei rhodesiense isolates from Tanzania. Parasitology 115 (6), 571−579. https://doi.org/10.1017/S0031182097001856.

Kooy, R.F., et al., 1989. Evidence for diploidy in metacyclic forms of African trypanosomes. Proc. Natl. Acad. Sci. U. S. A. 86 (14), 5469−5472. https://doi.org/10.1073/pnas.86.14.5469.

Lee, M.G.S., Van Der Ploeg, L.H.T., 1990. Homologous recombination and stable transfection in the parasitic protozoan Trypanosoma brucei. Science 250 (4987), 1583−1587. https://doi.org/10.1126/science.2177225.

Lehane, M.J., et al., 2000. Grouping of trypanosome species in mixed infections in Glossina pallidipes. Parasitology 120 (6), 583−592. https://doi.org/10.1017/S0031182099005983.

Lewis, M.D., et al., 2014. Bioluminescence imaging of chronic Trypanosoma cruzi infections reveals tissue-specific parasite dynamics and heart disease in the absence of locally persistent infection. Cell Microbiol. 16 (9), 1285−1300. https://doi.org/10.1111/cmi.12297.

Lewis, M.D., et al., 2009. Flow cytometric analysis and microsatellite genotyping reveal extensive DNA content variation in Trypanosoma cruzi populations and expose contrasts between natural and experimental hybrids. Int. J. Parasitol. 39 (12), 1305−1317. https://doi.org/10.1016/j.ijpara.2009.04.001.

Lewis, M.D., et al., 2011. Recent, independent and anthropogenic origins of Trypanosoma cruzi hybrids. PLoS Neglected Trop. Dis. 5, e1363. https://doi.org/10.1371/journal.pntd.0001363.

Lisboa, C.V., et al., 2004. Distinct patterns of Trypanosoma cruzi infection in Leontopithecus rosalia in distinct Atlantic coastal rainforest fragments in Rio de Janeiro - Brazil. Parasitology 129 (6), 703−711. https://doi.org/10.1017/S0031182004005918.

Louradour, I., et al., 2022. Stress conditions promote Leishmania hybridization in vitro marked by expression of the ancestral gamete fusogen HAP2 as revealed by singlecell RNA-seq. Elife 11. https://doi.org/10.7554/eLife.73488.

Lukeš, J., et al., 2007. Evolutionary and geographical history of the Leishmania donovani complex with a revision of current taxonomy. Proc. Natl. Acad. Sci. U. S. A. 104 (22), 9375−9380. https://doi.org/10.1073/pnas.0703678104.

Lypaczewski, P., et al., 2022. An intraspecies Leishmania donovani hybrid from the Indian subcontinent is associated with an atypical phenotype of cutaneous disease. iScience 25 (2). https://doi.org/10.1016/j.isci.2022.103802.

Machado, C.A., Ayala, F.J., 2001. Nucleotide sequences provide evidence of genetic exchange among distantly related lineages of Trypanosoma cruzi. Proc. Natl. Acad. Sci. U. S. A. 98 (13), 7396−7401. https://doi.org/10.1073/pnas.121187198.

MacLeod, A., Tweedie, A., McLellan, S., Taylor, S., Cooper, A., et al., 2005a. Allelic segregation and independent assortment in T. brucei crosses: proof that the genetic system is Mendelian and involves meiosis. Mol. Biochem. Parasitol. 143 (1), 12−19. https://doi.org/10.1016/j.molbiopara.2005.04.009.

MacLeod, A., Tweedie, A., McLellan, S., Taylor, S., Hall, N., et al., 2005b. The genetic map and comparative analysis with the physical map of Trypanosoma brucei. Nucleic Acids Res. 3 (21), 6688−6693. https://doi.org/10.1093/nar/gki980.

MacLeod, A., Turner, C.M.R., Tait, A., 1999. A high level of mixed Trypanosoma brucei infections in tsetse flies detected by three hypervariable minisatellites. Mol. Biochem. Parasitol. 102 (2), 237−248. https://doi.org/10.1016/S0166-6851(99)00101-2.

Mallet, J., 2007. Hybrid speciation. Nature 446 (7133), 279−283. https://doi.org/10.1038/nature05706.

Martins, H.R., et al., 2007. Impact of dual infections on chemotherapeutic efficacy in BALB/c mice infected with major genotypes of trypanosoma cruzi. Antimicrob. Agents Chemother. 51 (9), 3282−3289. https://doi.org/10.1128/aac.01590-06.

Martins, H.R., et al., 2006. Trypanosoma cruzi: impact of dual-clone infections on parasite biological properties in BALB/c mice. Exp. Parasitol. 112 (4), 237−246. https://doi.org/10.1016/j.exppara.2005.11.006.

Masiga, D.K., et al., 1996. A high prevalence of mixed trypanosome infections in tsetse flies in Sinfra, Côte d'Ivoire, detected by DNA amplification. Parasitology 112 (1), 75−80. https://doi.org/10.1017/s0031182000065094.

Matos, G.M., et al., 2022. Microevolution of Trypanosoma cruzi reveals hybridization and clonal mechanisms driving rapid genome diversification. Elife 11. https://doi.org/10.7554/eLife.75237.

Maudlin, I., 1991. Transmission of African Trypanosomiasis: Interactions Among Tsetse Immune System, Symbionts, and Parasites. Springer Science and Business Media LLC, pp. 117−148. https://doi.org/10.1007/978-1-4613-9044-2_6.

Mauricio, I.L., et al., 2006. Towards multilocus sequence typing of the Leishmania donovani complex: resolving genotypes and haplotypes for five polymorphic metabolic enzymes (ASAT, GPI, NH1, NH2, PGD). Int. J. Parasitol. 36 (7), 757−769. https://doi.org/10.1016/j.ijpara.2006.03.006.

Melville, S.E., et al., 1998. The molecular karyotype of the megabase chromosomes of Trypanosoma brucei and the assignment of chromosome markers. Mol. Biochem. Parasitol. 94 (2), 155−173. https://doi.org/10.1016/S0166-6851(98)00054-1.

Messenger, L.A., Garcia, L., et al., 2015a. Ecological host fitting of Trypanosoma cruzi TcI in Bolivia: mosaic population structure, hybridization and a role for humans in Andean parasite dispersal. Mol. Ecol. 24 (10), 2406−2422. https://doi.org/10.1111/mec.13186.

Messenger, L.A., Yeo, M., et al., 2015b. Molecular genotyping of Trypanosoma cruzi for lineage assignment and population genetics. Methods Mol. Biol. 1201, 297−337. https://doi.org/10.1007/978-1-4939-1438-8_19.

Messenger, L.A., Miles, M.A., 2015. Evidence and importance of genetic exchange among field populations of Trypanosoma cruzi. Acta Trop. 151 (1), 150−155. https://doi.org/10.1016/j.actatropica.2015.05.007.

Messenger, L.A., Miles, M.A., Bern, C., 2015c. Between a bug and a hard place: Trypanosoma cruzi genetic diversity and the clinical outcomes of Chagas disease. Exp. Rev. Anti-Infect. Ther. 13 (8), 995−1029. https://doi.org/10.1586/14787210.2015.1056158.

Miles, M.A., et al., 1981. DO radically dissimilar Trypanosoma cruzi strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' disease? Lancet 317 (8234), 1338−1340. https://doi.org/10.1016/s0140-6736(81)92518-6.

Miles, M.A., et al., 1978. Isozymic heterogeneity of Trypanosoma cruzi in the first autochthonous patients with Chagas' disease in Amazonian Brazil. Nature 272 (5656), 819−821. https://doi.org/10.1038/272819a0.

Miles, M.A., 1985. Ploidy, 'heterozygosity' and antigenic expression of South American trypanosomes. Parassitologia 27 (1−2), 87−104.

Miles, M.A., et al., 2009a. The molecular epidemiology and phylogeography of Trypanosoma cruzi and parallel research on Leishmania: looking back and to the future. Parasitology 136 (12), 1509−1528. https://doi.org/10.1017/S0031182009990977.

Miles, M.A., Yeo, M., Mauricio, I.L., 2009b. Leishmania exploit sex. Science 324 (5924), 187−189. https://doi.org/10.1126/science.1172789.

Monte-Neto, R.L., Fernandez-Prada, C., Moretti, N.S., 2022. Sex under pressure: stress facilitates Leishmania in vitro hybridization. Trends Parasitol. 38 (4), 274−276. https://doi.org/10.1016/j.pt..2022.02.001.

Negreira, G.H., et al., 2022. High throughput single-cell genome sequencing gives insights into the generation and evolution of mosaic aneuploidy in Leishmania donovani. Nucleic Acids Res. 50 (1), 293−305. https://doi.org/10.1093/nar/gkab1203.

Nolder, D., et al., 2007. Multiple hybrid genotypes of Leishmania (Viannia) in a focus of mucocutaneous leishmaniasis. Am. J. Trop. Med. Hyg. 76 (3), 573−578. https://doi.org/10.4269/ajtmh.2007.76.573.

Nyeko, J.H.P., et al., 1990. Characterization of trypanosome isolates from cattle in Uganda using species-specific DNA probes reveals predominance of mixed infections. Int. J. Trop. Insect Sci. 11 (03), 271−280. https://doi.org/10.1017/s1742758400012674.

Ocaña-Mayorga, C., et al., 2011. Sex, subdivision, and domestic dispersal of Trypanosoma cruzi lineage I in Southern Ecuador. PLoS Negl. Trop. 4.

Paindavoine, P., et al., 1986. Trypanosome hybrids generated in tsetse flies by nuclear fusion. EMBO J. 5 (13), 3631−3636. https://doi.org/10.1002/j.1460-2075.1986.tb04692.x.

Peacock, L., et al., 2007. Dynamics of infection and competition between two strains of Trypanosoma brucei brucei in the tsetse fly observed using fluorescent markers. Kinetoplastid Biol. Dis. 6. https://doi.org/10.1186/1475-9292-6-4.

Peacock, L., et al., 2011. Identification of the meiotic life cycle stage of Trypanosoma brucei in the tsetse fly. Proc. Natl. Acad. Sci. USA 108 (9), 3671−3676. https://doi.org/10.1073/pnas.1019423108.

Peacock, L., et al., 2009. Intraclonal mating occurs during tsetse transmission of Trypanosoma brucei. Parasit. Vectors 2 (1). https://doi.org/10.1186/1756-3305-2-43.

Peacock, L., Ferris, V., et al., 2014a. Mating compatibility in the parasitic protist Trypanosoma brucei. Parasit. Vectors 7 (1). https://doi.org/10.1186/1756-3305-7-78.

Peacock, L., Bailey, M., et al., 2014b. Meiosis and haploid gametes in the pathogen Trypanosoma brucei. Curr. Biol. 24 (2), 181−186. https://doi.org/10.1016/j.cub.2013.11.044.

Peacock, L., et al., 2021. Sequential production of gametes during meiosis in trypanosomes. Commun. Biol. 4 (1). https://doi.org/10.1038/s42003-021-02058-5.

Pinto, A.S., et al., 1998. Compared vectorial transmissibility of pure and mixed clonal genotypes of Trypanosoma cruzi in Triatoma infestans. Parasitol. Res. 84 (5), 348−353. https://doi.org/10.1007/s004360050409.

Ramesh, M.A., Malik, S.B., Logsdon, J.M., 2005. A phylogenomic inventory of meiotic genes: evidence for sex in Giardia and an early eukaryotic origin of meiosis. Curr. Biol. 15 (2), 185−191. https://doi.org/10.1016/j.cub.2005.01.003.

Ramírez, J.D., Llewellyn, M.S., 2014. Reproductive clonality in protozoan pathogens - truth or artefact? Mol. Ecol. 23 (17), 4195−4202. https://doi.org/10.1111/mec.12872.

Ramírez, J.D., Llewellyn, M.S., 2015. Response to Tibayrenc and Ayala: reproductive clonality in protozoan pathogens - truth or artefact? Mol. Ecol. 24 (23), 5782−5784. https://doi.org/10.1111/mec.13442.

Ravel, C., et al., 2006. First report of genetic hybrids between two very divergent Leishmania species: Leishmania infantum and Leishmania major. Int. J. Parasitol. 36 (13), 1383−1388. https://doi.org/10.1016/j.ijpara.2006.06.019.

Rogers, M.B., et al., 2014. Genomic confirmation of hybridisation and recent inbreeding in a vector-isolated leishmania population. PLoS Genet. 10 (1). https://doi.org/10.1371/journal.pgen.1004092.

Romano, A., et al., 2014. Cross-species genetic exchange between visceral and cutaneous strains of Leishmania in the sand fly vector. Proc. Natl. Acad. Sci. USA 111 (47), 16808−16813. https://doi.org/10.1073/pnas.1415109111.

Rougeron, V., et al., 2011. Reproductive strategies and population structure in Leishmania: substantial amount of sex in Leishmania Viannia guyanensis. Mol. Ecol. 20 (15), 3116−3127. https://doi.org/10.1111/j.1365-294X.2011.05162.x.

Rusman, F., et al., 2019. Elucidating diversity in the class composition of the minicircle hypervariable region of Trypanosoma cruzi: new perspectives on typing and kDNA inheritance. PLoS Neglected Trop. Dis. 13 (6). https://doi.org/10.1371/journal.pntd.0007536.

Rusman, F., et al., 2020. Evidence of hybridization, mitochondrial introgression and biparental inheritance of the kDNA minicircles in Trypanosoma cruzi I. PLoS Neglected Trop. Dis. 14 (1), e0007770. https://doi.org/10.1371/journal.pntd.0007770.

Sadlova, J., et al., 2011. Visualisation of leishmania donovani fluorescent hybrids during early stage development in the sand fly vector. PLoS One 6 (5). https://doi.org/10.1371/journal.pone.0019851.

Schmid-Hempel, R., et al., 2011. Genetic exchange and emergence of novel strains in directly transmitted trypanosomatids. Infect. Genet. Evol. 11 (3), 564−571. https://doi.org/10.1016/j.meegid.2011.01.002.

Schwabl, P., et al., 2020. Culture-free genome-wide locus sequence typing (GLST) provides new perspectives on Trypanosoma cruzi dispersal and infection complexity. PLoS Genet. 16 (12), e1009170. https://doi.org/10.1371/journal.pgen.1009170.

Schwabl, P., et al., 2019. Meiotic sex in Chagas disease parasite Trypanosoma cruzi. Nat. Commun. 10 (1). https://doi.org/10.1038/s41467-019-11771-z.

Schweizer, J., et al., 1994. Analysis of a new genetic cross between two East African Trypanosoma brucei clones. Parasitology 109 (1), 83−93. https://doi.org/10.1017/S0031182000077799.

Schweizer, J., Jenni, L., 1991. Hybrid formation in the lifecycle of Trypanosoma (T.) brucei: detection of hybrid trypanosomes in a midgut-derived isolate. Acta Trop. 48, 319−321.

Schweizer, J., Pospichal, H., Jenni, L., 1991. Hybrid formation between African trypanosomes in vitro. Acta Trop. 49 (3), 237−240. https://doi.org/10.1016/0001-706X(91)90043-J.

Schweizer, J., Tait, A., Jenni, L., 1988. The timing and frequency of hybrid formation in African trypanosomes during cyclical transmission. Parasitol. Res. 75 (2), 98−101. https://doi.org/10.1007/BF00932707.

Schwenkenbecher, J.M., et al., 2006. Microsatellite analysis reveals genetic structure of Leishmania tropica. Int. J. Parasitol. 36 (2), 237−246. https://doi.org/10.1016/j.ijpara.2005.09.010.

Shapiro, S.Z., et al., 1984. Analysis by flow cytometry of DNA synthesis during the life cycle of African trypanosomes. Acta Trop. 41 (4), 313−323.

Shapiro, T.A., Englund, P.T., 1995. The structure and replication of kinetoplast DNA. Annu. Rev. Microbiol. 49, 117−143. https://doi.org/10.1146/annurev.mi.49.100195.001001.

Sternberg, J., et al., 1988. Gene exchange in African trypanosomes: characterisation of a new hybrid genotype. Mol. Biochem. Parasitol. 27 (2−3), 191−200. https://doi.org/10.1016/0166-6851(88)90038-2.

Sternberg, J., et al., 1989. Gene exchange in African trypanosomes: frequency and allelic segregation. Mol. Biochem. Parasitol. 34 (3), 269−279. https://doi.org/10.1016/0166-6851(89)90056-X.

Stevens, J.R., Tibayrenc, M., 1996. Trypanosoma brucei s.l.: evolution, linkage and the clonality debate. Parasitology 112 (5), 481−488. https://doi.org/10.1017/s0031182000076940.

Stevens, J.R., Welburn, S.C., 1993. Genetic processes within an epidemic of sleeping sickness in Uganda. Parasitol. Res. 79 (5), 421−427. https://doi.org/10.1007/BF00931833.

Stothard, J.R., Frame, I.A., Miles, M.A., 1999. Genetic diversity and genetic exchange in Trypanosoma cruzi: dual drug-resistant \progeny\ from episomal transformants. Mem. Inst. Oswaldo Cruz 94 (1), 189−193. https://doi.org/10.1590/S0074-02761999000700027.

Sturm, N.R., et al., 2003. Evidence for multiple hybrid groups in Trypanosoma cruzi. Int. J. Parasitol. 33 (3), 269−279. https://doi.org/10.1016/S0020-7519(02)00264-3.

Tait, A., 1980. Evidence for diploidy and mating in trypanosomes. Nature 287 (5782), 536−538. https://doi.org/10.1038/287536a0.

Tait, A., et al., 1996. Self-fertilisation in Trypanosoma brucei. Mol. Biochem. Parasitol. 76 (1−2), 31−42. https://doi.org/10.1016/0166-6851(95)02528-6.

Ten Asbroek, A.L.M.A., Ouellette, M., Borst, P., 1990. Targeted insertion of the neomycin phosphotransferase gene into the tubulin gene cluster of Trypanosoma brucei. Nature 348 (6297), 174–175. https://doi.org/10.1038/348174a0.

Tibayrenc, M., Ayala, F.J., 2002. The clonal theory of parasitic protozoa: 12 years on. Trends Parasitol. 18 (9), 405–410. https://doi.org/10.1016/S1471-4922(02)02357-7.

Tibayrenc, M., Kjellberg, F., Ayala, F.J., 1990. A clonal theory of parasitic protozoa: the population structures of Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas, and Trypanosoma and their medical and taxonomical consequences. Proc. Natl. Acad. Sci. U. S. A. 87 (7), 2414–2418. https://doi.org/10.1073/pnas.87.7.2414.

Truc, P., et al., 1998. Trypanosoma brucei ssp. and T. congolense: mixed human infection in Cote d'Ivoire. Trans. R. Soc. Trop. Med. Hyg. 92 (5), 537–538. https://doi.org/10.1016/S0035-9203(98)90904-X.

Turner, C.M.R., et al., 1990. Evidence that the mechanism of gene exchange in Trypanosoma brucei involves meiosis and syngamy. Parasitology 101 (3), 377–386. https://doi.org/10.1017/S0031182000060571.

Turner, C.M.R., et al., 2004. Human infectivity trait in Trypanosoma brucei: stability, heritability and relationship to sra expression. Parasitology 129 (4), 445–454. https://doi.org/10.1017/S0031182004005906.

Turner, C.M.R., et al., 1995. Trypanosoma brucei: inheritance of kinetoplast DNA maxicircles in a genetic cross and their segregation during vegetative growth. Exp. Parasitol. 80 (2), 234–241. https://doi.org/10.1006/expr.1995.1029.

Vago, A.R., et al., 2000. Genetic characterization of Trypanosoma cruzi directly from tissues of patients with chronic Chagas disease: differential distribution of genetic types into diverse organs. Am. J. Pathol. 156 (5), 1805–1809. https://doi.org/10.1016/S0002-9440(10)65052-3.

Vago, A.R., et al., 1996. Kinetoplast DNA signatures of Trypanosoma cruzi strains obtained directly from infected tissues. Am. J. Pathol. 149 (6), 2153–2159.

Valdivia, H.O., et al., 2022. Replacement of Leishmania (Leishmania) infantum populations in an endemic focus of visceral leishmaniasis in Brazil. Front. Cell. Infect. Microbiol. 12. https://doi.org/10.3389/fcimb.2022.900084.

Van den Broeck, F., et al., 2020. Ecological divergence and hybridization of neotropical Leishmania parasites. Proc. Natl. Acad. Sci. U. S. A. 117 (40), 25159–25168. https://doi.org/10.1073/pnas.1920136117.

Van Xong, H., et al., 1998. A VSG expression site-associated gene confers resistance to human serum in Trypanosoma rhodesiense. Cell 95 (6), 839–846. https://doi.org/10.1016/S0092-8674(00)81706-7.

Volf, P., et al., 2007. Increased transmission potential of Leishmania major/Leishmania infantum hybrids. Int. J. Parasitol. 37 (6), 589–593. https://doi.org/10.1016/j.ijpara.2007.02.002.

Volf, P., Sadlova, J., 2009. Sex inLeishmania. Science 324 (5935), 1644. https://doi.org/10.1126/science.324_1644b.

Wells, J.M., et al., 1987. DNA contents and molecular karyotypes of hybrid Trypanosoma brucei. Mol. Biochem. Parasitol. 24 (1), 103–116. https://doi.org/10.1016/0166-6851(87)90121-6.

Westenberger, S.J., et al., 2005. Two hybridization events define the population structure of Trypanosoma cruzi. Genetics 171 (2), 527–543. https://doi.org/10.1534/genetics.104.038745.

WHO, 2002. Control of Chagas Disease. WHO, Geneva.

WHO, 2010. Control of the Leishmaniases, Technical Report Series. WHO, Geneva.

Widmer, G., Dvorak, J.A., Miles, M.A., 1987. Temperature modulation of growth rates and glucosephosphate isomerase isozyme activity in Trypanosoma cruzi. Mol. Biochem. Parasitol. 23 (1), 55–62. https://doi.org/10.1016/0166-6851(87)90187-3.

Wijers, D.J.B., 1958. Factors that may influence the infection rate of glossina palpalis with Trypanosoma gambiense. Ann. Trop. Med. Parasitol. 52 (4), 385–390. https://doi.org/10.1080/00034983.1958.11685878.

Yeo, M., et al., 2005. Origins of Chagas disease: didelphis species are natural hosts of Trypanosoma cruzi I and armadillos hosts of Trypanosoma cruzi II, including hybrids. Int. J. Parasitol. 35 (2), 225−233. https://doi.org/10.1016/j.ijpara.2004.10.024.

Yeo, M., et al., 2007. Resolution of multiclonal infections of Trypanosoma cruzi from naturally infected triatomine bugs and from experimentally infected mice by direct plating on a sensitive solid medium. Int. J. Parasitol. 37 (1), 111−120. https://doi.org/10.1016/j.ijpara.2006.08.002.

Zampetti-Bosseler, F., et al., 1986. Evidence for haploidy in metacyclic forms of Trypanosoma brucei. Proc. Natl. Acad. Sci. USA 83 (16), 6063−6064. https://doi.org/10.1073/pnas.83.16.6063.

Zingales, B., et al., 1998. Molecular epidemiology of American trypanosomiasis in Brazil based on dimorphisms of rRNA and mini-exon gene sequences. Int. J. Parasitol. 28. https://doi.org/10.1016/S0020-7519(97)00178-1.

CHAPTER

Genomic Insights Into the Past, Current, and Future Evolution of Human Parasites of the Genus *Plasmodium*

25

Colin J. Sutherland[1], **Khalid B. Beshir**[1] **and Spencer D. Polley**[2]

[1]*Department of Infection Biology, London School of Hygiene & Tropical Medicine, London, United Kingdom;*
[2]*Department of Clinical Parasitology, Hospital for Tropical Diseases, London, United Kingdom*

1. Introduction

1.1 Overview of Plasmodium Phylogeny and Description of Species Infecting *Homo sapiens*

The protozoan genus *Plasmodium* comprises chromalveolate protists of the Phylum Apicomplexa, Order Haemosporida, Family Plasmodidae. Members of the genus are obligate parasites of vertebrate hosts including lizards, snakes, birds, rodents, and primates. Amphibians, marsupials, carnivores, and ungulates are major vertebrate groups not known to host *Plasmodium* sp. parasites. Natural infections of *Homo sapiens* are caused by six species: *Plasmodium falciparum*, *P. knowlesi*, *P. malariae*, *P. vivax* and the two closely related species *P. ovale curtisi* and *P. ovale wallikeri*, which are genetically distinct (Sutherland et al., 2010; Sutherland, 2016). In naïve human hosts each of these six parasites cause an acute febrile illness of varying severity and duration, known as malaria.

The evolution of the genus *Plasmodium* was punctuated by a series of host transitions, as the radiation into more than 270 current species occurred through a variety of vertebrate hosts. The primate malarias are probably the best studied group of species, and have been well described both in natural infections of simian, ape, and human hosts (Garnham, 1966; Duval et al., 2009; Prugnolle et al., 2010; Liu et al., 2014; Plenderleith et al., 2022; Scully et al., 2022) and in experimental infections in chimpanzees, baboons, rhesus, and *Aotus* monkeys (Coatney et al., 2003). Investigations of the biochemistry and cell biology of *P. falciparum* have been possible in vitro since a system for continuous culture was devised by Trager and Jensen (1976), and this is therefore by far the best characterized of the primate malaria parasites. However, as *P. knowlesi* has now been successfully adapted to continuous culture in human erythrocytes (Moon et al., 2013), analysis of parasite biology in vitro is now routinely derived from both of these contrasting members of the genus, expanding our understanding of the comparative evolutionary history of members of the genus, six of which will be considered in this chapter.

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00028-8
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Due to the significant number of severe and fatal cases of malaria caused by *P. falciparum*, and its ability to be propagated as blood stage parasites in vitro, Laveran's parasite remains the most closely studied of the human-infecting plasmodia. Geographically widespread, the parasite occurs well north of the Tropic of Cancer (e.g., Afghanistan) and also south of the Tropic of Capricorn (e.g., South Africa, Namibia). Human blood stage infections display a 48 h cycle in experimental infections in volunteers, such as those studied in detail in the Georgia State Penitentiary in the mid-20th century (Coatney et al., 2003), hence the term "malignant tertian" malaria, referring to the periodicity of fevers in *P. falciparum*. Two particular biological features of this parasite distinguish it from the other human malaria infections. The first is that parasite-encoded adhesins, capable of binding to host endothelium, are expressed on the surface of the infected erythrocyte from the midtrophozoite stage right through to schizogony and release of merozoites. This enables the mature forms of the parasite to sequester in small blood vessels in a variety of host tissues, for a period of 30−36 hours, and thus only the young trophozoites ("ring forms") are observed in smears of peripheral blood whereas, for other human malaria species, intraerythrocytic parasites at all stages of maturity circulate in the periphery. The second distinguishing feature of *P. falciparum* is the production of crescent-shaped gametocytes (transmissible stages) which do not develop in synchrony with the asexual parasite stages, but appear late in infection after an extended period of development (typically 6−10 days) sequestered primarily in bone marrow by virtue of the rigid undeformability of these stages. This is reversed on gaining maturity, allowing release into the peripheral circulation in order to have access to biting mosquitoes (Tiburcio et al., 2013; Neveu et al., 2020). *P. vivax* infection occurs across the broadest geographical range of all the human malaria parasites, aided partly by an important survival strategy which this parasite shares with *P. ovale* spp.: the ability of liver schizonts to arrest development and remain dormant for weeks, months, or years as hypnozoites. Reactivation of hypnozoites sometime after the primary infection following an infective mosquito bite can thus initiate a fresh blood-stage malaria infection in a subsequent season favorable for transmission to mosquitoes. This mechanism is thought particularly important for the continued transmission of *P. vivax* in temperate areas with a long winter in which anophelines are scarce or absent or in areas with extreme or extended dry, hot seasons, such as Mauritania (Garnham, 1966; Coatney et al., 2003; Lekweiry et al., 2009). Malaria caused by *P. vivax* shares the 48 hours periodicity of falciparum malaria, but is responsible for many fewer severe or fatal cases, hence the label "benign tertian" malaria. There is new evidence of a sequestered reservoir of parasitized erythrocytes in the spleen (Kho et al., 2021), although all asexual and sexual forms of *P. vivax* are seen in peripheral blood smears. The lack of a reliable continuous in vitro culture system has hampered research on this parasite, but transient 48 hours schizont maturation cultures have been useful in studying antimalarial drug response phenotypes in *P. vivax* (Wirjanata et al., 2015).

Plasmodium malariae occurs in humans throughout malaria endemic regions of the world, and has earned the name "quartan malaria" for its 72-hour fever periodicity (reviewed in Collins and Jeffery, 2007; Mueller et al., 2007). This species occurs at low parasite density in the peripheral blood, frequently occurs with *P. vivax* or *P. falciparum* as mixed infections, causes generally a relatively mild malaise compared to other species, and is likely to be substantially underreported due to misdiagnosis as *P. vivax* or *P. falciparum* (Garnham, 1966; Lindo et al., 2007). *P. malariae* also occurs as a well-described zoonosis in American monkeys, under the species designation *P. brasilianum*. Although generally considered a minor species, *P. malariae* is very common in some locations in PNG, Indonesia, and Africa, contributing substantially to overall malaria morbidity (Dinko et al., 2013; Bruce et al., 2008). One of the most puzzling aspects of the biology of *P. malariae* is the ability to recur

years, and even decades, after the last possible exposure of the individual to an infected mosquito bite (Tsuchida et al., 1982; Vinetz et al., 1998). Analysis of imported *P. malariae* infections in three nonendemic countries has led to the suggestion that this species may also, under certain circumstances, form latent liver-stages analogous to the hypnozoites of *P. vivax, P. ovale* spp., and *P. cynomologi* (Teo et al., 2015).

Ovale malaria is caused by two closely related but distinct parasite species, *P. ovale curtisi* and *P. ovale wallikeri* (Sutherland et al., 2010). First described by Stephens (1922), *P. ovale* spp. cause acute febrile malaria with a tertian periodicity, but only rarely are associated with severe or life-threatening complications (Collins and Jeffery, 2005; Rojo-Marcos et al., 2018). Ovale malaria does not occur outside of the tropics, and is not known in the Americas, but in seasonal and arid settings the two causative species benefit from the ability to form hypnozoites, allowing emergence during periods of mosquito abundance (Lekweiry et al., 2009; Nabarro et al., 2018). The absence of *P. ovale* spp. transmission in temperate zones despite its ability to form hypnozoites suggests that the mosquito stages of this parasite are not able to complete development at lower temperatures. Relapse episodes of ovale malaria can occur months or years apart (Coldren et al., 2007; Nolder et al., 2014), and therefore this species may pose a particular challenge for the eradication of malaria from sub-Saharan Africa.

The sixth human malaria, *Plasmodium knowlesi*, whose natural host is certain species of Asian macaques, is a relatively common zoonotic agent of both clinical malaria and chronic asymptomatic parasitemia in humans in South East Asia (Singh et al., 2004; Cox-Singh et al., 2008; Fornace et al., 2016; Lubis et al., 2017). Human infections have been described in almost all countries in southeast Asia where macaques occur (Cuenca et al., 2021).

1.2 Population Genetics and Design of Public Health Interventions

Whereas much attention has been paid to population-level studies of polymorphic protein-coding genes as a means to empirical identification of potential candidate vaccine antigens in both *P. falciparum* and *P. vivax*, these studies have also provided invaluable insights into the nature and complexity of historical selective forces in shaping populations of *Plasmodium* parasites (Polley and Conway, 2001; Ord et al., 2008). These insights from population genetics inform not only the design of vaccines but also the processes used to monitor the efficacy over time of vaccines put into widespread use.

Population genetic tools have also been powerfully deployed to assist understanding of the spread over recent decades of gene variants encoding parasite resistance to antimalarial drugs, particularly chloroquine (CQ) and sulphadoxine-pyramethamine (SP) (Roper et al., 2004; Anderson and Roper, 2005). Further, with the availability of whole genome sequences of and the accumulating genome databases with multiple isolates of each species (Robinson et al., 2011; Ansari et al., 2016; Diez Benavente et al., 2017, 2021; Ibrahim et al., 2020), genome-wide variation can now be studied among all *Plasmodium* species infecting humans (Section 22.1.4).

1.3 Genomic Signals of Selection due to Host Immunity

Co-evolution of *Plasmodium* species and their vertebrate hosts has left significant signals of selection on the host genome, well-known examples being the hemoglobin structural variants of *Homo sapiens* (such as sickle-cell anemia, the thalassemias, and G6PD deficiency) which provide a measure of

protection against malaria. These genotypes have therefore become established in populations with ancient or recent history of *Plasmodium* infection risk (reviewed by Weatherall et al., 2002). Conversely, both mammalian and insect hosts have imposed selection upon the *Plasmodium* parasite genome as their respective immune systems adapt to minimize the harmful effects of parasitization. Perhaps one of the most spectacular examples of this is the genus-wide expansion of genes encoding large families of proteins involved in immune evasion. The SICAvar antigens of *P. knowlesi* were the first such proteins to be discovered, followed by a number of families in other members of the genus, notably including the PfEMP1 proteins of *P. falciparum*, encoded by the ~60-strong *var* gene family, the *surf* family ($N = 10$) and the *Pfmc-2tm* family ($N = 11$) (Winter et al., 2005). Genome analysis of a number of *Plasmodium* species genomes has identified members of the *pir* (*Plasmodium* interspersed repeat) family of small variant proteins, and among the parasites of primates these reach their greatest expansion in *P. ovale curtisi* and *P. ovale wallikeri* (Sutherland, 2016; Ansari et al., 2016). Thus, subtelomeric gene families have expanded to encode hundreds of protein variants per genome, suggesting that the benefit of avoiding immune responses from the vertebrate host has outweighed the cost of amplifying up such large gene families, and developing the complex gene regulatory pathways required to express them effectively.

Single-copy genes throughout *Plasmodium* genomes also bear signals of immune selection, and these are exemplified by diversification of the well-characterized nucleotide sequences in both *P. falciparum* and *P. vivax* that encode the immunogenic merozoite adhesion/invasion proteins such as AMA1, MSP2, MSP3, EBA-175, and the duffy-binding proteins (Polley and Conway, 2001; Ord et al., 2008; Weatherall et al., 2002). Other single-copy genes displaying evidence of selection for diversification include those coding for the sporozoite protein thrombospondin-related protein (TRAP), and $SURFIN_{4.2}$, a protein expressed in late stage intraerythrocytic asexual parasites (Winter et al., 2005; Ochola et al., 2010). In contrast, evidence for balancing selection is not seen in population genetic analyses of the locus encoding circumsporozoite protein (CSP), despite its high degree of polymorphism (Weedall et al., 2007).

1.4 The Impact of New Technologies on Genomic Studies of the Genus

Since the first whole genome sequence of *P. falciparum* was published (Gardner et al., 2002), genome data from all human-infecting species in the genus have become partly or wholly available. New sequencing technologies have facilitated direct studies of chromosomal rearrangements and speciation events, and chromosomal level evolution of drug and diagnostic escape variants within species, and even within specific populations.

1.4.1 Deep and Long-read Sequencing of Parasite Genomes

The advent of long-read technologies, such as PacBio and Oxford Nanopore, have enabled the sequencing and improved, rapid assembly of the entire genome of many *Plasmodium* species and allowed illumination of structural variation at genomic level. The technology provides greater depth of genomic information which has allowed researchers to understand the genetic mechanisms of drug resistance, virulence and disease transmission as well as temporal and spatial origins of new Plasmodium variants (Vembar et al., 2016). NGS using long read sequencing technology provides additional information on genome structural variation including deletions, duplications, inversions, and translocations.

1.4.2 Chromosomal Rearrangements and Speciation Events

In the past, investigation of chromosomal rearrangement was conducted through a painstakingly long process (Pologe and Ravetch, 1986). Short-read sequencing are not ideal for chromosomal level analysis of genome sequences due to the inability to assemble regions of high homology and polymorphic regions such as repetitive subtelemeric sequences, and are often excluded from analysis (Gardner et al., 2002). High homology regions enable recombination processes, a crucial strategy for rapid genome evolution. Using long read sequencing technology it is now possible to fully assemble chromosomes including the subtelomeric regions with uniform coverage in a single run allowing genome comparison among different plasmodium species (Otto et al., 2018). The coding potential of related genomes and the evolutionary factors that have shaped genome organization can be inferred via comparative genomics (Ramaprasad et al., 2021). The successful genome sequencing of *P. malariae* and the two *P. ovale* species using PacBio sequencing allowed elucidation of the interrelationships and evolutionary history of human-infecting *Plasmodium* species (Rutledge et al., 2017).

1.4.3 Chromosomal Level Evolution of Drug and Diagnostic Escape Variants Within Species

Gene duplications and deletions are known mechanisms of drug and diagnostic resistance in *P. falciparum* and these can lead to the emergence of new parasite populations (Beshir et al., 2017; Grignard et al., 2020). Whole genome sequencing of the variants has revealed that the duplication and deletion events do not only affect the gene of interest but also the flanking regions (Nolder et al., 2021). Research is currently underway to understand the break point of common deletions and the mechanism of the duplication and deletion events at the chromosomal level. Long-read sequencing technologies are vital to give insight into the events and whether the genomic structural variations impact virulence, transmission, and evolutionary relationships among *Plasmodium* species, a perhaps contribute to speciation events themselves (Section 22.2.3).

2. Evolution of *Plasmodium*: The Last 10 Million Years

2.1 Role of Host Transitions in Speciation Events Within the Genus

A species is a group of organisms which can mate to produce fertile offspring. For speciation to occur, two populations of a species must become reproductively isolated, thereby preventing gene flow between them. The result of reproductive isolation is that over time the two populations will accumulate private mutations which result in the inability to form viable offspring should interpopulation breeding occur again at some time after the original separation. Allopatric speciation occurs when reproductive isolation is a direct result of the geographical separation of two (or more) populations through a physical barrier such as an ocean. In contrast, sympatric speciation has a genetic origin, preventing admixture within two populations found in the same location. Putative mechanisms for sympatric speciation may include temporal isolation through a shift in the timing of gamete release, behavior isolation through different courtship routines, physical isolation through noncompatible genetalia and gametic imcompatibility mediated by mating incompatibility loci in plant, fungi, and marine organisms such as the sea urchin *Echinometra* and marine diatoms (Palumbi and Metz, 1991). For an obligate parasite, an obvious mechanism for sympatric speciation is host switching. In the case of *Plasmodium,* such genetic isolation would be exacerbated if the two vertebrate hosts are targeted by

different species of vector, given that genetic recombination between parasite genotypes occurs exclusively in the mosquito midgut.

Experimental infections, the comparison of *Plasmodium* and host phylogenies and most recently transgenic manipulation of parasites, have all shed light on the role of host switching in the evolution of the genus. Phylogenetic studies have allowed the evolution of mammalian and bird malarias to be elucidated and the common origin of these species to be calculated at between 120 and 160 million years ago (mya) (Perkins and Schall, 2002). Such a time point is much later than the divergence of sauropsid (ancestral linage of mammals) and synapsid (ancestral lineage of birds, lizards, and snakes) vertebrate lineages some 315 million years ago, suggesting that host switching has occurred during this time period. Human malaria species themselves appear not to be monophyletic but instead have multiple independent evolutionary origins. At least three separate simian to human host transitions are apparent from the phylogenetic reconstruction of mammalian malarial mitochondrial sequences, when only *P. vivax*, *P. knowlesi*, and *P. malariae* are considered (Mu et al., 2005). Recent studies on both bird and bat malaria species also show evidence for both host plasticity and host switching (Ricklefs and Fallon, 2002; Iezhova et al., 2005; Duval et al., 2007).

2.2 *P. falciparum* and *P. reichenowi*: Divergent Host Specificities?

The closest relative of the human malaria parasite *P. falciparum* is the chimpanzee parasite *P. reichenowi.* The ancestors of *Homo sapiens* (modern day humans) and *Pan troglodytes* (chimpanzees) are thought to have diverged around four to eight mya according to phylogenetic and fossil evidence (Brunet et al., 2002; Hacia, 2001). This time point has often been used as a date for the divergence of *P. falciparum* and *P. reichenowi*, a time point which has in turn been used to date other nodes within phylogenetic studies. When *P. reichenowi* was discovered in 1917 in the blood of wild chimpanzees and gorillas, it was morphologically indistinguishable from *P. falciparum*. However, in vitro experiments showed that humans could not be infected with *P. reichenowi*, and nor could chimpanzees be infected with *P. falciparum* parasitized patient blood (Martin et al., 2005). The two infections were therefore deemed to be separate species, an interpretation which was subsequently supported by genetic analysis of the *csp* (circumsporozoite protein) locus (Lal and Goldman, 1991).

As early as 1977, differences in the erythrocyte receptors of humans and chimpanzees were postulated as the driving force behind the apparent species specificity of *P. reichenowi* and *P. falciparum* (Miller, 1977), and this has since been related to differences in the sialic acid structure of Glycophorin A between humans and chimpanzees (Martin et al., 2005). However, recent insights gained from studies of the genomes of the Laveranian parasites *P. reichenowi* and *P. gaboni* from DNA fragments isolated from wild ape feces show that *P. falciparum* has only recently transferred into *Homo sapiens*, and shows relatively little genetic diversity compared to the former two species, or even compared to *P. falciparum*-like parasites from wild gorillas (Liu et al., 2010; Sundararaman et al., 2016). This strongly suggests a more profound basis to speciation than Miller's concept based on specialized erythrocyte invasion, which is more likely to be a recent adaptation after the original host transition.

While early experiments showed that chimpanzees were refractory to *P. falciparum*, experimental infections can be established in splenectomized chimpanzees. Indeed several isolates of *P. falciparum* including Vietnam Oak Knoll and Uganda Palo Alto have been adapted by serial passage through new world monkey hosts for use with animal models (Geiman and Meagher, 1967; Siddiqui et al., 1972).

The last few years have also seen the discovery of *P. knowlesi* infections in humans (Singh et al., 2004), confirmation of *P. falciparum* infection in both great apes and monkeys (Duval et al., 2009; Liu et al., 2010), and both *P. malariae* and *P. vivax* in chimpanzees (Hayakawa et al., 2006; Liu et al., 2014), suggesting that the host specificity of certain *Plasmodium* species may be more fluid than was once thought: Such infections, at the very least, suggest that mechanisms such as specific adaptation to erythrocytic receptors specific to a particular primate host may be insufficient to fully explain host switching and potential sympatric speciation events.

What alternatives are there to host switching that would facilitate the genetic isolation of sympatric populations? Mutations within mating incompatibility loci are known to produce genetic isolation within sympatric populations of animals and plants. The surface gamete proteins Pfs48/45 are known to be critical for the fertility of male microgametes in *P. falciparum* (Van Dijk et al., 2001), and gene distributions at this locus show significant skewing at the population level (Conway et al., 2001). Following fertilization, a diploid ookinete is formed which buries into the mosquito midgut and undergoes meiotic reduction to form a sporozoite filled oocyte. Analysis of Pfs48/45 allele frequencies in oocysts gathered from Tanzanian mosquitos shows a significantly higher inbreeding coefficients (F(IS)) compared with 11 unlinked microsatellite loci locus. Thus, it would appear that assortative mating is occurring in these populations (Anthony et al., 2007), raising the possibility that genetic isolation may occur through mutations in gametic surface receptors at the time of host switching by the novel species.

2.3 Speciation Between *P. ovale curtisi* and *P. ovale wallikeri*: Separate Host Transitions?

The parasites previously defined as *Plasmodium ovale,* on the basis of their morphology when viewed on Giemsa-stained blood films, are now known to comprise two genetically distinct, nonrecombining species which diverged around 2 mya. These two newly recognized species were shown to be sympatric at country-level in their distribution, thus ruling out classic allopatric barriers as an explanation for the current pattern of endemicity, and suggesting that reproductive isolation between the two parasites is due to other mechanisms such as host partitioning (Sutherland et al., 2010). Nevertheless, as only country-level distribution data was available for this first study of the dimorphic parasite species, it could not be ruled out that continuing physical separation between the two species is maintained due to more subtle mechanisms such as microgeographic discontinuities or specific habitat or seasonal requirements preventing admixture of the two types in the present. Subsequent examination of parasites from clinics in Congo-Brazzaville, and community parasitological surveys in Equatorial Guinea and two sites in Uganda demonstrated conclusively that both *P. ovale curtisi* and *P. ovale wallikeri* occur in different people in the same village at the same time, providing stronger support for the existence of a robust biological barrier (or barriers) between the two parasite species (Oguike et al., 2011). A similar picture has been seen in Ethiopia (Alemu et al., 2013).

It is prudent to recognize that whatever the current barriers to mating and recombination might be, it is not necessarily true that these were also the *cause* of the original speciation event between the two forms of ovale malaria parasites. In our view, the most parsimonious explanation for the existence of these two related but distinct parasite lineages in the same primate host, *Homo*, is that their common ancestor in the most recent nonhuman primate host (*Plasmodium ovale sensu stricto*) underwent successful transit to *Homo* on two separate occasions. The two lineages antecedent to the recently

identified *P. ovale curtisi* and *P. ovale wallikeri* species were thus prevented from recombination because each occupied a different primate host, and so the two lineages never occurred simultaneously in a single mosquito blood-meal, the absolute requirement for hybridization. During this time apart, differences arose which meant that recombination was no longer possible, once both lineages had become parasites of humans. It is also valid to argue an alternative to our favored "two transition hypothesis," namely that separation was due to a geographic separation of two populations *after* host transition to *Homo* (or the ancestors of *Homo*). This could be dubbed the "transient allopatry hypothesis," in that it requires sympatric, early *Homo*-parasitizing (hypothetical) *P. ovale s.s.* populations to have developed into two geographically isolated lineages that came back into sympatry due to migration of the hominid hosts or expansion of the separated parasite populations. During the period of separation, the pair of lineages underwent sufficient divergent evolution for differences to arise, and these now keep the two lineages apart, despite their current close co-habitation across a global range (Sutherland et al., 2010; Sutherland, 2016; Oguike et al., 2011; Oguike and Sutherland, 2015). Other authors have suggested an alternative form of this hypothesis in which, given that wild apes harbor ovale-like parasites similar to both *P. ovale curtisi and P. ovale wallikeri*, it seems more likely that the two forms had already evolved as separate species in different ape hosts prior to their introduction into humans (Sharp et al., 2020).

What could be the nature of these differences that maintain the species-barrier between *P. ovale curtisi* and *P. ovale wallikeri* now they exist in sympatry? This is a matter of great interest, and a number of explanations can be hypothesized; five have been considered by Sutherland (2016; Box 1). The explanation we favor is through genetic drift, leading to divergence of specific molecules required during the mating process, or to gross changes at chromosomal level that prevent viable chromosome pairing at meiosis.

This hypothesis, that divergence in genome primary and secondary structure during a period of separation prevents effective meiotic pairing between the two current lineages, now that they are circulating in the same host species (*Homo sapiens*), is based on the assumption that recombination does not occur between *P. ovale curtisi* and *P. ovale wallikeri.* Species-specific restriction to different subsets of the human population had appeared to be a plausible mechanism for the current maintenance of the species barrier between *P. ovale curtisi* and *P. ovale wallikeri*. A precedent exists in the form of the well-described reliance of *P. vivax* on the Duffy blood group antigen for invasion of host reticulocytes. However, evidence from field studies shows that both species can be identified simultaneously in a single individual (Fuehrer et al., 2012; Dinko et al., 2013), findings which are not compatible with any restriction occurring between different host subgroups (Sutherland, 2016).

Species-specific restriction of *P. ovale curtisi* and *P. ovale wallikeri* to different species of *Anopheles* mosquito is also unable to explain the lack of recombination between the two parasites. There are a large variety of competent malaria vector species within the genus *Anopheles*, and the malaria vectors of medical importance differ within and between nations, and within and between continents. Thus the distribution of both *P. ovale curtisi* and *P. ovale wallikeri* in Asia, Africa, and the Pacific demonstrates unequivocally that both parasites are transmitted by a variety of mosquito species across this broad range (Sutherland et al., 2010; Fuehrer et al., 2012). Mosquito host restriction, although possibly important in some endemic localities, therefore cannot explain the observed lack of recombination between the two ovale species, although vector incompatibility may be one of the factors that have prevented *P. ovale* spp. From becoming endemic in the Americas (Fuehrer et al.,

2022), this despite the frequent trans-Atlantic introductions of these species that must have occurred from the 1500s on.

Allopatric speciation events can generate two related taxa, physically separated, which can become secondarily sympatric due to migration or changes in the extent of suitable habitat (Section 2.1). If during the period of separation substantial genetic drift occurs in the sequences of genes determining mating compatibility, or chromosomal rearrangements occur, interrupting synteny such that hybrid offspring are extremely unlikely to be viable, then the species barrier will remain intact, and recombination between the two forms will not occur. As argued above, we consider that the most likely reason for the evolution of two forms of ovale malaria parasite is that at least two independent host transitions into ancestors of modern humans occurred, separated by a lengthy period of time. This scenario is of course analogous to allopatry, in that the two lineages would have been "physically" separated by occupying different hosts during the period between the first and second transitions. Thus, the lack of recombination between 21st century human-dwelling populations of *P. ovale curtisi* and *P. ovale wallikeri* is most likely due to genomic changes accrued over this period of separation, thus rendering meiotic chromosome pairing impossible, and hybrid zygotes unviable. The recent genomic data of Ansari et al. (2016) does identify divergence among the greatly expanded subtelomeric *oir* genes, encoding members of the PIR super family of variant antigens found in all members of the genus, and areas of syntenic discontinuity, that together could hinder effective homologous chromosome pairing in heterologous matings between the two ovale parasite species. Long-read genome data from both species, such as the output of Oxford nanopore technology (Section 22.1.4), should help to identify such regions of the two genomes.

Genomic studies are therefore poised to answer important questions about genome-wide interspecies polymorphisms—which loci have diverged the most between *P. ovale curtisi* and *P. ovale wallikeri*? Are erythrocyte invasion molecules prominent among them? Are the greatly expanded *oir* loci, located in the subtelomeres which accrue diversity and structural rearrangements at a faster rate than other chromosomal regions, helping to prevent efficient meiotic pairing between the two species (Sutherland, 2016)? Close examination of intraspecies polymorphism will directly address the hypothesis of clonal propagation in ovale parasites. It is certainly true that, to date, there is little evidence that either ovale parasite species is highly diverse genetically, and our first study of interspecies genetic variation found only a very low level of diversity at the seven loci examined (Sutherland et al., 2010). However, the genes studied were mostly identified due to their high degree of sequence conservation with homologous loci in other better characterized *Plasmodium* species; such genes are likely to be under purifying selection and thus poor candidates with which to estimate levels of linkage disequilibrium, and coefficients of inbreeding. Some intraspecific polymorphism has since been described in the *potra* locus (Oguike et al., 2011; Tanomsing et al., 2013), *cox1* (Fuehrer et al., 2012), and *msp1* (Putaporntip et al., 2013). Thus, although it remains possible that frequent clonal propagation has exacerbated the isolation of these two genomes, this is very unlikely to be an explanation for the existence of a species barrier per se.

2.4 Importance of Host Specificity

Understanding the mechanisms of host specificity, and host switching, will allow a greater understanding of the pathology of *Plasmodium* spp. infections, identifying key genes required for both asexual and sexual reproduction, and possible targets for disruption through the production of novel

vaccines and chemotherapeutics. The establishment of SARS, HIV, Ebola, Zika, and SARS-CoV-2 viral infections in human populations shows how host switching can lead to the emergence of a highly virulent pathogen, and a relatively recent switch to human hosts might explain the increased pathogenicity of *P. falciparum* compared to *P. ovale* spp., *P. vivax*, and *P. malariae*. Ultimately, detailed understanding of the parasite genes contributing to the vertebrate host specificity of each *Plasmodium* parasite species may come from careful comparisons among the genomes now being sequenced. Given current understanding of the importance of erythrocyte-binding proteins and their specificity for certain host receptors (Section 2.2), elucidation of the reticulocyte-binding proteins, Duffy-binding proteins and erythrocyte binding protein families in each species is a good place to start. The zoonotic species *Plasmodium simium* appears to have evolved from *P. vivax* via a recent host switch to nonhuman primates. The parasite maintains the ability to infect humans. A 38-amino acid deletion of the *P. simium* Duffy Binding Protein 1 not found in *P. vivax* has been found to be associated with human infection in some (e.g., Mourier et al., 2021) but not all (e.g., de Oliveira et al., 2021) studies. All *P. simium* isolates (both human and simian) show a 1004 amino acid deletion in the RPP2a gene not found in *P. vivax*. In vitro it has been shown that cultured lines of *P. knowlesi* that have been adapted to propagate in human erythrocytes express particular variants of the reticulocyte-binding protein, but that gene-editing of these loci can restore the preference for macaque erythrocytes (Moon et al., 2013; Ahmed et al., 2014; Mohring et al., 2019). However, we should bear in mind that an uncomfortably large proportion of each *Plasmodium* genome thus far sequenced encodes for "hypothetical proteins" of unknown function; it may be that comparative empirical approaches are required to identify many of the key molecules involved in parasite host specificity (Hall et al., 2005).

2.5 Contrasting Drug Susceptibilities: Divergent Biochemical Adaptation Within the Genus

Speciation in the genus *Plasmodium* permits divergence between emerging species of cellular adaptation to the host environment. An interesting new research area bringing such divergent biochemistry to the attention of biologists and chemists alike, is that of comparative drug susceptibility. The availability of *P. knowlesi* as a second human-infecting malaria parasite species that can be propagated in long-term in vitro culture has led to new studies directly comparing susceptibility of this parasite with *P. falciparum*, using carefully harmonized assay conditions that allow comparability, despite the contrasting 27h and 48h intra-erythrocytic cycles of the two species, respectively (van Schalkwyk et al., 2017, 2019, 2020). These findings were consistent with a limited number of ex vivo susceptibility assays performed on other species, in transient cultures derived directly from infected blood from malaria patients (van Schalkwyk et al., 2021). A markedly smaller effect against *P. knowlesi* was noted for a group of investigational compounds, including cipargamin, that are known to targeting the ATP4 cation transporter in *P. falciparum* with high potency. Using a new approach of ortholog exchange, in which the *atp4* locus from each species replaced the endogenous locus in *P. knowlesi* by CRISPR-Cas9 gene editing, it was established that these contrasting susceptibilities were due to differences in the primary amino acid sequence of the transporter (Mohring et al., 2022). Although most drug classes are equipotent for all *Plasmodium* species tested, these studies have now established that, for a small number of compounds, potency should be examined across multiple species.

3. Evolution of *Plasmodium*: The 20th Century

3.1 Habitat Destruction and Ecological Transitions

Host transitions in malaria parasites require contact between the novel vertebrate host species and insect vectors that have bitten infected individuals of the primary vertebrate host species. Thus, human (and prehuman) migration, settlement, and the resulting encroachment of human activity into the habitats of different nonhuman primates has been the probable driving force behind the prehistorical transitions discussed in Section 2. In the 21st century, human migrations still occur, and encroachment into the habitats of wild simian and ape populations continue. Coupled with a recent improvement in our ability to discern the presence of unusual *Plasmodium* species in both humans and nonhuman primates, these continued close encounters between humans, and the parasites of beasts, will be more commonly recognized. Further, the destruction of habitat and the resulting decline in numbers of the great apes means that the parasite species dependent on them as hosts are under selective pressure to expand their host range. This may increase the likelihood of new transitions into human-centered parasitism. An added factor to consider here is that the impacts of global warming on weather and climate patterns worldwide in the decades ahead are certain to lead to large human migrations, and to both expansions and contractions of malaria vector distributions. Both these dynamic forces, although difficult to predict in magnitude with any accuracy from our current knowledge, are certain to exert additional pressure on malaria parasite populations to adapt to new circumstances as they arise.

On the other hand, the human parasite *Plasmodium falciparum* is known to be present in wild African apes, with identification of this "human" parasite in bonobos (*Pan paniscus*) from the Democratic Republic of Congo (Krief et al., 2010), in *Gorilla gorilla dielhi* from Cameroon and *G. g. gorilla* from Gabon (Prugnolle et al., 2010) and other central African locations (Liu et al., 2010). Genetic analyses of these strongly suggest continued cycling of the Laverania grouping of malaria parasites among different hominids, including *Homo*, in Africa. Conversely, as these studies also identified additional examples of *P. reichenowi* in three subspecies of *Pan* in Ivory Coast and Cameroon, and a closely related parasite in a bonobo from DRC, as well as two new species, *P. billbrayi* and *P. billcollinsi*, the possibility now occurs that a variety of hitherto overlooked parasite species, with *Pan* and *Gorilla* as primary hosts, may be zoonotic in humans. This poses two interesting questions. Firstly, have we seriously underestimated the number of malaria parasite species that naturally infect humans in central Africa, by ignoring the possibility of frequent zoonotic infections? Secondly, will the great apes provide an eradication-proof reservoir of human malaria parasites, particularly *P. falciparum*, *P. ovale curtisi*, and *P. ovale wallikeri* (Section 4; Duval et al., 2009, 2010)?

The recent descriptions of contemporary wild isolates of *P. reichenowi* (Prugnolle et al., 2010; Sundararaman et al., 2016) raise the possibility that human infections with this parasite may occasionally occur in some parts of Africa, but have failed to be recognized. This is not least because remnant populations of the nonhuman hominids exist in relatively remote locations, where people suffering from malaria who do come in contact with health services are likely to be treated (if at all) without any diagnosis, and certainly without a species-specific one (Sutherland et al., 2010). The prospect of such previously unrecognized zoonoses occurring is an interesting one from a scientific point of view, but may also have important evolutionary and public health consequences. Although at least one recent study found no evidence of ape parasites among human communities near populations of wild apes in Gabon (Délicat-Loembet et al., 2015), studies of likely vectors in the forest canopy do

suggest there is no entomological barrier to zoonotic transmission of these ape parasite species to humans (Makanga et al., 2016). As long as such infections are at least possible, the evolutionary change to permit efficient human-to-human transmission of a previously ape-confined species may occur. From a public health point of view, this could lead to rapid expansion of a "new" malaria pathogen in human populations with no species-specific immunity; this threat is even more relevant were malaria elimination/eradication to occur in these regions in the near future. Removal of human parasite species through vector control, immunization, and effective drug deployment will lead to a human populations with no naturally acquired immunity, and perhaps with greatly reduced access to effective malaria treatment as the number of cases dwindle. Thus, understanding of these zoonotic pathogens should be vigorously pursued not only because of its great scientific interest, which would be enough for most of us, but through human self-interest, which may be needed to persuade science funders. Indeed, numerous isolates of *P. inui* and some *P. fieldi* were recently found among symptomatic patients screened by molecular methodologies in Thailand, suggesting that either species can establish human infections under natural circumstances and that *P. inui* infections in Thailand are far more common than previously thought (Putaporntip et al., 2022).

3.2 Emergence of *P. knowlesi* as an Important Zoonosis

As mentioned above (Section 1.1), the story of *P. knowlesi* in Southeast Asia is one particularly important example, becoming well understood only in the past decade, of a zoonotic malaria parasite being recognized as a significant public health problem for humans. The parasite was thought to only naturally infect long-tailed and pig-tailed macaques, which are widely distributed across Southeast Asia and in which infection is generally considered benign (Coatney et al., 2003), despite one or two sporadic reports of human cases in Singapore and peninsular Malaysia in the 1960s (Garnham, 1966). It had been demonstrated during the use of induced malaria infections for syphilis therapy in the early 20th century that *P. knowlesi* would infect humans with effective blood-stage multiplication, and cause full-blown clinical malaria, but the true extent of naturally occurring human cases had not been understood, and was assumed to be negligible. Molecular investigations during a malaria epidemic in Malaysian Borneo in the late 1990s confirmed *P. knowlesi* as the causative agent of a significant number of human cases (Singh et al., 2004). Molecular studies have since identified human cases of *P. knowlesi* in several Asian countries in addition to Malaysia, including Burma, The Philippines, Singapore, Vietnam, Indonesia, and Thailand (Zheng et al., 2006; Luchavez et al., 2008; Ng et al., 2008; Putaporntip et al., 2009; Van den Eede et al., 2009; Setiadi et al., 2016). *P. knowlesi* infections in humans can cause severe and fatal disease (Cox-Singh et al., 2008), but it is now becoming clear from studies in Malaysian Borneo (Fornace et al., 2016), Myanmar (Ghinai et al., 2017), and Sumatra (Lubis et al., 2017) that *P. knowlesi* infections can occur widely as asymptomatic chronic infections in communities in proximity to macaque populations.

It is probably the case that infections of humans in areas where macaques abound have always occurred, and that human *P. knowlesi* infections may have been either ignored, or extensively misdiagnosed in the past (Lee et al., 2009). However, as there is as yet no evidence of human-to-mosquito transmission, the disease is best described as a well-established zoonosis with a measurable public health impact in some areas. Should efficient human-to-vector transmission arise de novo, then *P. knowlesi* could easily spread via human movements beyond the current range, which for now is absolutely restricted by the distribution of suitable host macaque species. Such a "break-out" of this

species may carry a threat of substantial mortality, as this parasite has a very rapid replication cycle and can reach life-threatening parasite densities in a small proportion of infected human individuals (Cox-Singh et al., 2008).

3.3 Chemotherapy and the Pharmaceutical Selection of New Parasite Gene Variants

A well-studied example of recent evolution in the genus *Plasmodium* is the worldwide development of resistance to the antimalarial drug chloroquine during the latter half of the 20th century. The evolutionary aspects of this drug selection has been well examined (Escalante et al., 2009), establishes principles relevant to understanding the currently emerging impact of new-generation antimalarial drugs that have been in use during the present century.

Firstly, although there is evidence that CQ-resistant alleles of the transporter gene *pfcrt* evolved in different lineages of *P. falciparum* on multiple occasions, one or two of these alleles were particularly successful, spreading through contiguous parasite populations, and moving intercontinental distances, presumably in human travelers, to dominate parasite populations world-wide (Mehlotra et al., 2008). Interestingly, in the case of at least one of these alleles, that encoding the amino acid haplotype SVMNT at codons 72–76 of *pfcrt*, it now seems likely that heavy use of the related amino-quinoline drug amodiaquine (AQ) in some countries in the mid-20th century provides a plausible alternative explanation for the spread of this parasite type (Sá et al., 2009; Beshir et al., 2010). Thus, CQ resistance per se may be a secondary characteristic of the CRT protein variant encoded by this allele.

Secondly, antimalarial selection is far from uniform in endemic areas, and so there are both temporal and geographic variations in the intensity of drug pressure, broadly meaning the proportion of infected people likely to use a particular drug type. This modulates the selective advantage enjoyed by parasites carrying drug resistant alleles of key genes, particularly if there is a fitness cost to the development of resistance. Thus, poor access to treatment, high levels of antiparasite immunity leading to asymptomatic parasite carriage (particularly in African adults), and extended periods in the dry season with no transmission of new infections have contributed to the successful survival and continual transmission of wild-type, CQ sensitive *P. falciparum* in many parts of Africa throughout the period that CQ was the main treatment option for malaria (Ord et al., 2007). In the extreme case of complete removal of CQ from an entire health system, as was achieved in Malawi, the fitness advantage of wild-type CQ-sensitive parasites ensures their relatively rapid resurgence to replace CQ-resistant genotypes that no longer have a survival advantage (Laufer et al., 2006).

Thirdly, different antimalarial regimens generate different, sometimes opposite, selective forces on the parasite genome (Duraisingh et al., 2000; Humphreys et al., 2007). Thus, as drug use diversifies in the current post-chloroquine era, the parasite genome appears to be showing evidence of "balancing selection" at drug resistance loci, a concept in direct opposition to the theoretical notion that drug pressure always drives advantageous alleles to "fixation" (Section 3.2.2).

Finally, although genetically determined resistance to CQ has been described in several species of *Plasmodium*, only in *P. falciparum* are mutations in the *crt* locus directly linked to altered response to the drug. Stable resistance to CQ in both *P. vivax* and in *P. chabaudi* has been described, but the *pvcrt* and *pccrt* loci, respectively, remain unchanged in resistant parasites (Suwanarusk et al., 2007; Hunt et al., 2007). This is in contrast to resistance to the antifolate drugs pyrimethamine and sulfadoxine, which involve nonsynonymous mutations in the *dhfr* and *dhps* loci of each *Plasmodium* species so far

investigated. Thus, the use of model host-parasite systems, and even in vitro studies of drug susceptibility, may not provide adequate understanding of drug resistance of *Plasmodium* parasites in vivo, nor are findings in one species translatable to all species.

The last 15−20 years has seen the introduction of artemisinin-based combination therapy (ACT) as the main replacement for CQ and SP in the formal health service by the governments of most endemic countries. The most frequently used ACT in Africa are artemether-lumefantrine (AL) and artesunate-amodiaquine (ASAQ), whereas in Asian countries dihydroartemisinin-piperaquine (DP) and mefloquine-artesunate (MA) are widely deployed. The emergence of *P. falciparum* parasites in Cambodia with markedly increased time to parasite clearance in vivo following artesunate monotherapy (Noedl et al., 2008; Dondorp et al., 2013) has caused great alarm, and there have been important efforts to systematically monitor parasite clearance across the global range of *P. falciparum*, to both artesuante alone and to ACT. Despite the coverage afforded by partner drugs, the almost universal adoption of ACT in endemic countries does effectively guarantee that selection on parasite genomes by artemisinins is essentially worldwide. Interestingly, these slow clearance phenotypes, associated with specific mutations in the propeller domain of the *P. falciparum* K13 kelch protein, appear to have arisen more than once in SE Asia under artemisinin selection, rather than spreading from a single focus (Ye et al., 2016), and now a similar phenomenon has been recognized in Africa, where K13 propeller domain variants have arisen de novo in both Rwanda and Uganda (Uwimana et al., 2020; Balikagala et al., 2021; Tumwebaze et al., 2022). However, in African settings treatment failures following ACT treatment have been reported without associated *pfk13* mutations (Muwanguzi et al., 2016; Sonden et al., 2017; Sutherland et al., 2017), and there is growing evidence of reduced susceptibility to the widely used partner drug lumefantrine (Tumwebaze et al., 2022; van Schalkwyk et al., 2023). Multigenic haplotypes including variants at the loci *pfcrt*, *pfmdr1, pfap2mu*, and *pfubp1* have been implicated in at least one east African study (Henriques et al., 2014). Several strategies could minimize the impact of such newly evolving ART resistance, and these largely rely on the expectation that evolution of resistance involves a fitness cost for the parasite.

- partner drug diversification should be deployed in ACT, so that parasites less susceptible to one combination are likely to be contained if treated with a different combination.
- where indicated by falling ACT efficacy, the use of triple combinations (van der Pluijm et al., 2021) or extended regimens deploying two different ACT sequentially (Schallig et al., 2017), so that treatment comprises 6 days of artemisinin and two partner drugs in total, should be considered.
- continued development of new drugs must take place, for additional partners to combine with the artemisinins, and for potent new compounds to replace the artemisinins should their failure be catastrophic at some point in the future. In this way, the evolution of resistance in malaria parasites can be partially managed, in order to maximize the duration of the public health benefit currently enjoyed due to the effectiveness of ACT.

4. Future Challenges From Continuing *Plasmodium* Genome Adaptation

As malaria eradication programs are envisaged for implementation around the malaria endemic world, evolutionary biologists are pondering the potential of uniform, large-scale interventions to force the parasite genome through selection bottlenecks. An example of such a bottleneck is the impact of the

antimalarial chloroquine, which for decades in the 20th century exerted monolithic selection on the *P. falciparum* genome for a handful of advantageous alleles of *pfcrt* and *pfmdr1*. Some questions arise:

- *Are the potentially cataclysmic impacts of global warming on host and vector abundance and distribution a challenge that the parasite is able to meet through adaptation?* Vigilance is required to recognize emerging foci of malaria in formerly malaria-free areas, and also changes in the parasite and mosquito vector species compositions in place affected by rapid climate change.
- *Is it likely that malaria, in turn, will continue to exert selection for new mutations on the human genome, as it has in the past?* This seems unlikely, as relatively few people are prevented from reproduction by malaria, and hopes are high that efforts toward eradication will rapidly bring this number down even further. Nevertheless, there is ample evidence that already existing variants such the thalassemias, G6PD deficiency and sickle-cell Hb, are maintained at stable frequencies in human populations by offering protection against malaria.
- *Will zoonotic primate malaria infections lead to a reservoir of hit-and-run malaria cases on the fringes of forests in Africa and Asia, and might these species adapt to anthropo-centric transmission?* This fascinating and potentially dangerous scenario cannot be ruled out, and, should human malaria be eradicated at some time in the future, will require careful vigilance in human populations that frequent the forest fringes of SE Asia and central Africa.

References

Ahmed, A.M., Pinheiro, M.M., Divis, P.C., Siner, A., Zainudin, R., Wong, I.T., et al., 2014. Disease progression in *Plasmodium knowlesi* malaria is linked to variation in invasion gene family members. PLoS Neglected Trop. Dis. 8, e3086.

Alemu, A., Fuehrer, H.P., Getnet, G., Tessema, B., Noedl, H., 2013. *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri* in North-West Ethiopia. Malar. J. 12, 346.

Anderson, T.J., Roper, C., 2005. The origins and spread of antimalarial drug resistance: lessons for policy makers. Acta Trop. 94, 269−270.

Ansari, H.R., Templeton, T.J., Subudhi, A.K., Ramaprasad, A., Tang, J., Lu, F., et al., 2016. Genome-scale comparison of expanded gene families in *Plasmodium ovale wallikeri* and *Plasmodium ovale curtisi* with *Plasmodium malariae* and with other *Plasmodium* species. Int. J. Parasitol. 46, 685−696.

Anthony, T.G., Polley, S.D., Vogler, A.P., Conway, D.J., 2007. Evidence of non-neutral polymorphism in *Plasmodium falciparum* gamete surface protein genes Pfs47 and Pfs48/45. Mol. Biochem. Parasitol. 156, 117−123.

Balikagala, B., Fukuda, N., Ikeda, M., Katuro, O.T., Tachibana, S.I., Yamauchi, M., et al., 2021. Evidence of artemisinin-resistant malaria in Africa. N. Engl. J. Med. 385, 1163−1171.

Beshir, K.B., Sepúlveda, N., Bharmal, J., Robinson, A., Mwanguzi, J., Busula, A.O., et al., 2017. *Plasmodium falciparum* parasites with histidine-rich protein 2 (*pfhrp2*) and *pfhrp3* gene deletions in two endemic regions of Kenya. Sci. Rep. 7, 14718.

Beshir, K., Sutherland, C.J., Merinopoulos, I., Durrani, N., Leslie, T., Rowland, M., et al., 2010. Amodiaquine resistance in *Plasmodium falciparum* malaria is associated with the *pfcrt* 72-76 SVMNT allele in Afghanistan. Antimicrob. Agents Chemother. 54, 3714−3716.

Bruce, M.C., Macheso, A., Kelly-Hope, L.A., Nkhoma, S., McConnachie, A., Molyneux, M.E., 2008. Effect of transmission setting and mixed species infections on clinical measures of malaria in Malawi. PLoS One 3, e2775.

Brunet, M., Guy, F., Pilbeam, D., Mackaye, H.T., Likius, A., Ahounta, D., 2002. A new hominid from the upper miocene of Chad central Africa. Nature 418, 145−151.

Coatney, G.R., Collins, W.E., Warren, M., Contacos, P.G., 2003. *The Primate Malarias* Division of Parasitic Disease, Producers. CDC, Atlanta, GA, USA. CD-ROM. [originally published 1971].

Coldren, R.L., Jongsakul, K., Vayakornvichit, S., Noedl, H., Fukuda, M.M., 2007. Apparent relapse of imported *Plasmodium ovale* malaria in a pregnant woman. Am. J. Trop. Med. Hyg. 77, 992−994.

Collins, W.E., Jeffery, G.M., 2007. *Plasmodium malariae*: parasite and disease. Clin. Microbiol. Rev. 20, 579−592.

Collins, W.E., Jeffery, G.M., 2005. *Plasmodium ovale:* parasite and disease. Clin. Microbiol. Rev. 18, 570−581.

Conway, D.J., Machado, R.L., Singh, B., Dessert, P., Mikes, Z.S., Povoa, M.M., et al., 2001. Extreme geographical fixation of variation in the *Plasmodium falciparum* gamete surface protein gene Pfs48/45 compared with microsatellite loci. Mol. Biochem. Parasitol. 115, 145−156.

Cox-Singh, J., Davis, T.M., Lee, K.S., Shamsul, S.S., Matusop, A., Ratnam, S., et al., 2008. *Plasmodium knowlesi* malaria in humans is widely distributed and potentially life threatening. Clin. Infect. Dis. 46, 165−171.

Cuenca, P.R., Key, S., Jumail, A., Surendra, H., Ferguson, H.M., Drakeley, C.J., Fornace, K., 2021. Epidemiology of the zoonotic malaria Plasmodium knowlesi in changing landscapes. Adv. Parasitol. 113, 225−286.

Délicat-Loembet, L., Rougeron, V., Ollomo, B., Arnathau, C., Roche, B., Elguero, E., et al., 2015. No evidence for ape *Plasmodium* infections in humans in Gabon. PLoS One 10, e0126933.

de Oliveira, T.C., Rodrigues, P.T., Early, A.M., Duarte, A.M.R.C., Buery, J.C., Bueno, M.G., et al., 2021. *Plasmodium simium*: population genomics reveals the origin of a reverse zoonosis. J. Infect. Dis. 224, 1950−1961.

Diez Benavente, E., Florez de Sessions, P., Moon, R.W., Holder, A.A., Blackman, M.J., Roper, C., et al., 2017. Analysis of nuclear and organellar genomes of *Plasmodium knowlesi* in humans reveals ancient population structure and recent recombination among host-specific subpopulations. PLoS Genet. 13, e1007008.

Diez Benavente, E.D., Manko, E., Phelan, J., Campos, M., Nolder, D., Fernandez, D., et al., 2021. Distinctive genetic structure and selection patterns in *Plasmodium vivax* from south Asia and east Africa. Nat. Commun. 12, 3160.

Dinko, B., Oguike, M.C., Larbi, J.B., Bousema, J.T., Sutherland, C.J., 2013. Persistent detection of *Plasmodium falciparum, P. malariae, P. ovale curtisi* and *P. ovale wallikeri* after ACT treatment of asymptomatic Ghanaian school-children. Int J Parasitol DDR 3, 45−50.

Dondorp, A.M., Nosten, F., Yi, P., Das, D., Phyo, A.P., Tarning, J., et al., 2013. Artemisinin resistance in *Plasmodium falciparum* malaria. N. Engl. J. Med. 361, 455−467.

Duraisingh, M.T., Jones, P., Sambou, I., von Seidlein, L., Pinder, M., Warhurst, D.C., 2000. The tyrosine-86 allele of the pfmdr1 gene of *Plasmodium falciparum* is associated with increased sensitivity to the anti-malarials mefloquine and artemisinin. Mol. Biochem. Parasitol. 108, 13−23.

Duval, L., Fourment, M., Nerrienet, E., Rousset, D., Sadeuh, S.A., Goodman, S.M., et al., 2010. African apes as reservoirs of *Plasmodium falciparum* and the origin and diversification of the Laverania subgenus. Proc. Natl. Acad. Sci. U.S.A. 107, 10561−10566.

Duval, L., Nerrienet, E., Rousset, D., Sadeuh Mba, S.A., Houze, S., Fourment, M., et al., 2009. Chimpanzee malaria parasites related to *Plasmodium ovale* in Africa. PLoS One 4, e5520.

Duval, L., Robert, V., Csorba, G., Hassanin, A., Randrianarivelojosia, M., Walston, J., et al., 2007. Multiple host-switching of Haemosporidia parasites in bats. Malar. J. 6, 157.

Escalante, A.A., Smith, D.L., Kim, Y., 2009. The dynamics of mutations associated with anti-malarial drug resistance in *Plasmodium falciparum*. Trends Parasitol. 25, 557−563.

Fornace, K.M., Nuin, N.A., Betson, M., Grigg, M.J., William, T., Anstey, N.M., et al., 2016. Asymptomatic and submicroscopic carriage of *Plasmodium knowlesi* malaria in household and community members of clinical cases in Sabah, Malaysia. J. Infect. Dis. 213, 784−787.

Fuehrer, H.P., Campino, S., Sutherland, C.J., 2022. The primate malaria parasites *Plasmodium malariae, Plasmodium brasilianum* and *Plasmodium ovale* spp.: genomic insights into distribution, dispersal and host transitions. Malar. J. 21, 138.

Fuehrer, H.P., Habler, V.E., Fally, M.A., Harl, J., Starzengruber, P., Swoboda, P., et al., 2012. *Plasmodium ovale* in Bangladesh: genetic diversity and the first known evidence of the sympatric distribution of *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri* in southern Asia. Int. J. Parasitol. 42, 693−699.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., et al., 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419, 498−511.

Garnham, P.C.C., 1966. Malaria Parasites and Other Haemosporidia. Blackwell, Oxford.

Geiman, Q.M., Meagher, M.J., 1967. Susceptibility of a new world monkey to *Plasmodium falciparum* from man. Nature 215, 437−439.

Ghinai, I., Cook, J., Hla, T.T., Htet, H.M., Hall, T., Lubis, I.N., et al., 2017. Malaria epidemiology in central Myanmar: identification of a multi-species asymptomatic reservoir of infection. Malar. J. 16, 16.

Grignard, L., Nolder, D., Sepúlveda, N., Berhane, A., Mihreteab, S., Kaaya, R., et al., 2020. A novel multiplex qPCR assay for detection of *Plasmodium falciparum* with histidine-rich protein 2 and 3 (*pfhrp2* and *pfhrp3*) deletions in polyclonal infections. EBioMedicine 55, 102757.

Hacia, J.G., 2001. Genome of the apes. Trends Genet. 17, 637−645.

Hall, N., Karras, M., Raine, J.D., Carlton, J.M., Kooij, T.W., Berriman, M., et al., 2005. A comprehensive survey of the *Plasmodium* life cycle by genomic, transcriptomic, and proteomic analyses. Science 307, 82−86.

Hayakawa, T., Arisue, N., Udono, T., Hirai, H., Sattabongkot, J., Toyama, T., et al., 2006. Identification of *Plasmodium malariae,* a human malaria parasite, in imported chimpanzees. PLoS One 4, e7412.

Henriques, G., Hallett, R.L., Beshir, K.B., Gadalla, N.B., Johnson, R.E., Burrow, R., et al., 2014. Directional selection at the *pfmdr1, pfcrt, pfubp1*, and *pfap2mu* loci of *Plasmodium falciparum* in Kenyan children treated with ACT. J. Infect. Dis. 210, 2001−2008.

Humphreys, G.A., Merinopoulos, I., Ahmed, J., Whitty, C.J.M., Mutabingwa, T.K., Sutherland, C.J., et al., 2007. Amodiaquine and artemether-lumefantrine select distinct alleles of the *Plasmodium falciparum pfmdr1* gene in Tanzanian children treated for uncomplicated malaria. Antimicrob. Agents Chemother. 51, 991−997.

Hunt, P., Afonso, A., Creasey, A., Culleton, R., Sidhu, A.B., Logan, J., et al., 2007. Gene encoding a deubiquinating enzyme is mutated in artesunate- and chloroquine-resistant rodent malaria parasites. Mol. Microbiol. 65, 27−40.

Ibrahim, A., Diez Benavente, E., Nolder, D., Proux, S., Higgins, M., Muwanguzi, J., et al., 2020. Selective whole genome amplification of Plasmodium malariae DNA from clinical samples reveals insights into population structure. Sci. Rep. 10, 10832.

Iezhova, T.A., Valkiunas, G., Bairlein, F., 2005. Vertebrate host specificity of two avian malaria parasites of the subgenus Novyella: *Plasmodium nucleophilum* and *Plasmodium vaughani*. J. Parasitol. 91, 472−474.

Kho, S., Qotrunnada, L., Leonardo, L., Andries, B., Wardani, P.A.I., Fricot, A., et al., 2021. Hidden biomass of intact malaria parasites in the human spleen. N. Engl. J. Med. 384, 2067−2069.

Krief, S., Escalante, A.A., Pacheco, M.A., Mugisha, L., André, C., Halbwax, M., et al., 2010. On the diversity of malaria parasites in African apes and the origin of *Plasmodium falciparum* from Bonobos. PLoS Pathog. 6, e1000765.

Lal, A.A., Goldman, I.F., 1991. Circumsporozoite protein gene from *Plasmodium reichenowi* a chimpanzee malaria parasite evolutionarily related to the human malaria parasite *Plasmodium falciparum*. J. Biol. Chem. 266, 6686−6689.

Laufer, M.K., Thesing, P.C., Eddington, N.D., Masonga, R., Dzinjalamala, F.K., Takala, S.L., et al., 2006. Return of chloroquine antimalarial efficacy in Malawi. N. Engl. J. Med. 355, 1959−1966.

Lee, K.S., Cox-Singh, J., Brooke, G., Matusop, A., Singh, B., 2009. *Plasmodium knowlesi* from archival blood films: further evidence that human infections are widely distributed and not newly emergent in Malaysian Borneo. Int. J. Parasitol. 39, 1125−1128.

Lekweiry, K.M., Abdallahi, M.O., Ba, H., Arnathau, C., Durand, P., Trape, J.F., et al., 2009. Preliminary study of malaria incidence in Nouakchott, Mauritania. Malar. J. 8, 92.

Lindo, J.F., Bryce, J.H., Ducasse, M.B., Howitt, C., Barrett, D.M., Morales, J.L., et al., 2007. *Plasmodium malariae* in Haitian refugees, Jamaica. Emerg. Infect. Dis. 13, 931−933.

Liu, W., Li, Y., Learn, G.H., Rudicell, R.S., Robertson, J.D., Keele, B.F., Ndjango, J.B., et al., 2010. Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature 467, 420−425.

Liu, W., Li, Y., Shaw, K.S., Learn, G.H., Plenderleith, L.J., Malenke, J.A., et al., 2014. African origin of the malaria parasite *Plasmodium vivax*. Nat. Commun. 5, 3346.

Lubis, I.N.D., Wijaya, H., Lubis, M., Lubis, C.P., Divis, P.C.S., Beshir, K.B., Sutherland, C.J., 2017. Contribution of plasmodium knowlesi to multispecies human malaria infections in north sumatera, Indonesia. J. Infect. Dis. 215, 1148−1155.

Luchavez, J., Espino, F., Curameng, P., Espina, R., Bell, D., Chiodini, P., et al., 2008. Human infections with *Plasmodium knowlesi*, the Philippines. Emerg. Infect. Dis. 14, 811−813.

Makanga, B., Yangari, P., Rahola, N., Rougeron, V., Elguero, E., Boundenga, L., et al., 2016. Ape malaria transmission and potential for ape-to-human transfers in Africa. Proc. Natl. Acad. Sci. U. S. A. 113, 5329−5334. **pii:** 201603008.

Martin, M.J., Rayner, J.C., Gagneux, P., Barnwell, J.W., Varki, A., 2005. Evolution of human-chimpanzee differences in malaria susceptibility: relationship to human genetic loss of N-glycolylneuraminic acid. Proc. Natl. Acad. Sci. U. S. A. 102, 12819−12824.

Mehlotra, R.K., Mattera, G., Bockarie, M.J., Maguire, J.D., Baird, J.K., Sharma, Y.D., et al., 2008. Discordant patterns of genetic variation at two chloroquine resistance loci in worldwide populations of the malaria parasite *Plasmodium falciparum*. Antimicrob. Agents Chemother. 52, 2212−2222.

Miller, L.H., 1977. Hypothesis on the mechanism of erythrocyte invasion by malaria merozoites. Bull WHO 55, 157−162.

Mohring, F., Hart, M.N., Rawlinson, T.A., Henrici, R., Charleston, J.A., Diez, B., et al., 2019. Rapid and iterative genome editing in the malaria parasite *Plasmodium knowlesi* provides new tools for *P. vivax* research. Elife 8, e45829.

Mohring, F., van Schalkwyk, D.A., Henrici, R.C., Blasco, B., Leroy, D., Sutherland, C.J., Moon, R.W., 2022. Cation ATPase (ATP4) orthologue replacement in the malaria parasite *Plasmodium knowlesi* reveals species-specific responses to ATP4-targeting drugs. mBio 13, e0117822.

Moon, R.W., Hall, J., Rangkuti, F., Ho, Y.S., Almond, N., Mitchell, G.H., et al., 2013. Adaptation of the genetically tractable malaria pathogen *Plasmodium knowlesi* to continuous culture in human erythrocytes. Proc. Natl. Acad. Sci. U.S.A. 110, 531−536.

Mourier, T., de Alvarenga, D.A.M., Kaushik, A., de Pina-Costa, A., Douvropoulou, O., Guan, Q., et al., 2021. The genome of the zoonotic malaria parasite *Plasmodium simium* reveals adaptations to host switching. BMC Biol. 19, 219.

Mu, J., Joy, D.A., Duan, J., Huang, Y., Carlton, J., Walker, et al., 2005. Host switch leads to emergence of *Plasmodium vivax* malaria in humans. Mol. Biol. Evol. 22, 1686−1693.

Mueller, I., Zimmerman, P.A., Reeder, J.C., 2007. *Plasmodium malariae* and *Plasmodium ovale* - the 'bashful' malaria parasites. Trends Parasitol. 23, 278−283.

Muwanguzi, J., Henriques, G., Sawa, P., Bousema, T., Sutherland, C.J., Beshir, K.B., 2016. Lack of K13 mutations in *Plasmodium falciparum* persisting after artemisinin combination therapy treatment of Kenyan children. Malar. J. 15, 36.

Nabarro, L.E.B., Nolder, D., Broderick, C., Nadjm, B., Smith, V., Blaze, M., et al., 2018. Geographical and temporal trends and seasonal relapse in *Plasmodium ovale* spp. and *Plasmodium malariae* infections imported to the UK between 1987 and 2015. BMC Med. 16, 218.

Neveu, G., Richard, C., Dupuy, F., Behera, P., Volpe, F., Subramani, P.A., et al., 2020. *Plasmodium falciparum* sexual parasites develop in human erythroblasts and affect erythropoiesis. Blood 136, 1381−1393.

Ng, O.T., Ooi, E.E., Lee, C.C., Lee, P.J., Ng, L.C., Pei, S.W., et al., 2008. Naturally acquired human *Plasmodium knowlesi* infection, Singapore. Emerg. Infect. Dis. 14, 814−816.

Noedl, H., Se, Y., Schaecher, K., Smith, B.L., Socheat, D., Fukuda, M.M., 2008. Artemisinin resistance in Cambodia 1 (ARC1) study consortium. Evidence of artemisinin-resistant malaria in western Cambodia. N. Engl. J. Med. 359, 2619−2620.

Nolder, D., Oguike, M.C., Maxwell-Scott, H., Niyazi, H.A., Smith, V., Chiodini, P.L., et al., 2014. An observational study of malaria in British travellers: *Plasmodium ovale wallikeri* and *Plasmodium ovale curtisi* differ significantly in the duration of latency. BMJ Open 3, e002711.

Nolder, D., Stewart, L., Tucker, J., Ibrahim, A., Gray, A., Corrah, T., et al., 2021. Failure of rapid diagnostic tests in Plasmodium falciparum malaria cases among travelers to the UK and Ireland: identification and characterisation of the parasites. Int. J. Infect. Dis. 108, 137−144.

Ochola, L.I., Tetteh, K.K., Stewart, L.B., Riitho, V., Marsh, K., Conway, D.J., 2010. Allele frequency-based and polymorphism-versus-divergence indices of balancing selection in a new filtered set of polymorphic genes in *Plasmodium falciparum*. Mol. Biol. Evol. 27, 2344−2351.

Oguike, M.C., Betson, M., Burke, M., Nolder, D., Stothard, R., Kleinschmidt, I., et al., 2011. *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri* circulate simultaneously in African communities. Int. J. Parasitol. 41, 677−683.

Oguike, M.C., Sutherland, C.J., 2015. Dimorphism in genes encoding sexual-stage proteins of *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri*. Int. J. Parasitol. 45, 449−454.

Ord, R., Alexander, N., Dunyo, S., Hallett, R.L., Jawara, M., Targett, G.A.T., et al., 2007. Seasonal carriage of pfcrt and pfmdr1 alleles in Gambian *Plasmodium falciparum* imply reduced fitness of chloroquine-resistant parasites. J. Infect. Dis. 196, 1613−1619.

Ord, R.L., Tami, A., Sutherland, C.J., 2008. *ama1* genes of sympatric *Plasmodium vivax* and *P. falciparum* from Venezuela differ significantly in genetic diversity and recombination frequency. PLoS One 3, e3366.

Otto, T.D., Böhme, U., Sanders, M., Reid, A., Bruske, E.I., Duffy, C.W., et al., 2018. Long read assemblies of geographically dispersed *Plasmodium falciparum* isolates reveal highly structured subtelomeres. Wellcome Open Res. 3, 52.

Palumbi, S.R., Metz, E.C., 1991. Strong reproductive isolation between closely related tropical sea urchins (genus *Echinometra*). Mol. Biol. Evol. 8, 227−239.

Perkins, S.L., Schall, J.J., 2002. A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. J. Parasitol. 88, 972−978.

Plenderleith, L.J., Liu, W., Li, Y., Loy, D.E., Mollison, E., Connell, J., et al., 2022. Zoonotic origin of the human malaria parasite *Plasmodium malariae* from African apes. Nat. Commun. 13, 1868.

Polley, S.D., Conway, D.J., 2001. Strong diversifying selection on domains of the *Plasmodium falciparum* apical membrane antigen 1 gene. Genetics 158, 1505−1512.

Pologe, L.G., Ravetch, J.V., 1986. A chromosomal rearrangement in a *P. falciparum* histidine-rich protein gene is associated with the knobless phenotype. Nature 322, 474−477.

Prugnolle, F., Durand, P., Neel, C., Ollomo, B., Ayala, F.J., Arnathau, C., et al., 2010. African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U.S.A. 107, 1458−1463.

Putaporntip, C., Hongsrimuang, T., Seethamchai, S., Kobasa, T., Limkittikul, K., Cui, L., et al., 2009. Differential prevalence of Plasmodium infections and cryptic *Plasmodium knowlesi* malaria in humans in Thailand. J. Infect. Dis. 199, 1143−1150.

Putaporntip, C., Hughes, A.L., Jongwutiwes, S., 2013. Low level of sequence diversity at merozoite surface protein-1 locus of *Plasmodium ovale curtisi* and *P. ovale wallikeri* from Thai isolates. PLoS One 8, e58962.

Putaporntip, C., Kuamsab, N., Seethamchai, S., Pattanawong, U., Rojrung, R., Yanmanee, S., et al., 2022. Cryptic *Plasmodium inui* and *Plasmodium fieldi* infections among symptomatic malaria patients in Thailand. Clin. Infect. Dis. 75, 805−812.

Ramaprasad, A., Klaus, S., Douvropoulou, O., Culleton, R., Pain, A., 2021. *Plasmodium vinckei* genomes provide insights into the pan-genome and evolution of rodent malaria parasites. BMC Biol. 19, 69.

Ricklefs, R.E., Fallon, S.M., 2002. Diversification and host switching in avian malaria parasites. Proc. Biol. Sci. 269, 885−892.

Robinson, T., Campino, S.G., Auburn, S., Assefa, S.A., Polley, S.D., Manske, M., et al., 2011. Drug-resistant genotypes and multi-clonality in *Plasmodium falciparum* analysed by direct genome sequencing from peripheral blood of malaria patients. PLoS One 6, e23204.

Rojo-Marcos, G., Rubio-Muñoz, J.M., Angheben, A., Jaureguiberry, S., García-Bujalance, S., Tomasoni, L.R., et al., 2018. Prospective comparative multi-centre study on imported *Plasmodium ovale wallikeri* and *Plasmodium ovale curtisi* infections. Malar. J. 17, 399.

Roper, C., Pearce, R., Nair, S., Sharp, B., Nosten, F., Anderson, T., 2004. Intercontinental spread of pyrimethamine-resistant malaria. Science 305, 1124.

Rutledge, G.G., Böhme, U., Sanders, M., Reid, A.J., Cotton, J.A., Maiga-Ascofare, O., et al., 2017. *Plasmodium malariae* and *P. ovale* genomes provide insights into malaria parasite evolution. Nature 542, 101−104.

Sám, Twu, O., Hayton, K., Reyes, S., Fay, M.P., Ringwald, P., Wellems, T.E., 2009. Geographic patterns of Plasmodium falciparum drug resistance distinguished by differential responses to amodiaquine and chloroquine. Proc. Natl. Acad. Sci. U. S. A. 106, 18883−18889.

Schallig, H.D., Tinto, H., Sawa, P., Kaur, H., Duparc, S., Ishengoma, D.S., et al., 2017. Randomised controlled trial of two sequential artemisinin-based combination therapy regimens to treat uncomplicated falciparum malaria in African children: a protocol to investigate safety, efficacy and adherence. BMJ Glob. Health 2, e000371.

Scully, E.J., Liu, W., Li, Y., Ndjango, J.N., Peeters, M., Kamenya, S., et al., 2022. The ecology and epidemiology of malaria parasitism in wild chimpanzee reservoirs. Commun. Biol. 5, 1020.

Setiadi, W., Sudoyo, H., Trimarsanto, H., Sihite, B.A., Saragih, R.J., Juliawaty, R., et al., 2016. A zoonotic human infection with simian malaria, Plasmodium knowlesi, in Central Kalimantan, Indonesia. Malar. J. 15, 218.

Sharp, P.M., Plenderleith, L.J., Hahn, B.H., 2020. Ape origins of human malaria. Annu. Rev. Microbiol. 74, 39−63.

Siddiqui, W.A., Schnell, J.V., Geiman, Q.M., 1972. A model in vitro system to test the susceptibility of human malarial parasites to antimalarial drugs. Am. J. Trop. Med. Hyg. 21, 393−399.

Singh, B., Kim Sung, L., Matusop, A., Radhakrishnan, A., Shamsul, S.S., Cox-Singh, J., et al., 2004. A large focus of naturally acquired *Plasmodium knowlesi* infections in human beings. Lancet 363, 1017−1024.

Sondén, K., Wyss, K., Jovel, I., Vieira da Silva, A., Pohanka, A., Asghar, M., et al., 2017. High rate of treatment failures in nonimmune travelers treated with artemether-lumefantrine for uncomplicated *Plasmodium falciparum* malaria in Sweden: retrospective comparative analysis of effectiveness and case series. Clin. Infect. Dis. 64, 199−206.

Stephens, J.W.W., 1922. A new malaria parasite of man. Ann. Trop. Med. Parasitol. 16, 383−386.

Sundararaman, S.A., Plenderleith, L.J., Liu, W., Loy, D.E., Learn, G.H., Li, Y., et al., 2016. Genomes of cryptic chimpanzee *Plasmodium* species reveal key evolutionary events leading to human malaria. Nat. Commun. 7, 11078.

Sutherland, C.J., Lansdell, P., Sanders, M., Muwanguzi, J., van Schalkwyk, D.A., Kaur, H., et al., 2017. pfk13-independent treatment failure in four imported cases of *Plasmodium falciparum* malaria treated with artemether-lumefantrine in the United Kingdom. Antimicrob. Agents Chemother. 61, 10−1128.

Sutherland, C.J., Tanomsing, N., Nolder, D., Oguike, M., Jennison, C., Pukrittayakamee, S., et al., 2010. Two non-recombining sympatric forms of the human malaria parasite *Plasmodium ovale* occur globally. J. Infect. Dis. 201, 1544−1550.

Sutherland, C.J., 2016. Persistent parasitism: the adaptive biology of malariae and ovale malaria. Trends Parasitol. 32, 808−819.

Suwanarusk, R., Russell, B., Chavchich, M., Chalfein, F., Kenangalem, E., Kosaisavee, V., et al., 2007. Chloroquine resistant *Plasmodium vivax*: in vitro characterisation and association with molecular polymorphisms. PLoS One 2, e1089.

Tanomsing, N., Imwong, M., Sutherland, C.J., Dolecek, C., Hien, T.T., Nosten, F., et al., 2013. Genetic marker suitable for identification and genotyping of *Plasmodium ovale curtisi* and *Plasmodium ovale wallikeri*. J. Clin. Microbiol. 51, 4213−4216.

Teo, B.H., Lansdell, P., Smith, V., Blaze, M., Nolder, D., Beshir, K.B., et al., 2015. Delayed onset of symptoms and atovaquone-proguanil chemoprophylaxis breakthrough by *Plasmodium malariae* in the absence of mutation at codon 268 of *pmcytb*. PLoS Neglected Trop. Dis. 9, e0004068.

Tibúrcio, M., Silvestrini, F., Bertuccini, L., Sander, A.F., Turner, L., Lavstsen, T., et al., 2013. Early gametocytes of the malaria parasite *Plasmodium falciparum* specifically remodel the adhesive properties of infected erythrocyte surface. Cell Microbiol. 15, 647−659.

Trager, W., Jensen, J.B., 1976. Human malaria parasites in continuous culture. Science 193, 673−675.

Tsuchida, H., Yamaguchi, K., Yamamoto, S., Ebisawa, I., 1982. Quartan malaria following splenectomy 36 years after infection. Am. J. Trop. Med. Hyg. 31, 163−165.

Tumwebaze, P.K., Conrad, M.D., Okitwi, M., Orena, S., Byaruhanga, O., Katairo, T., et al., 2022. Decreased susceptibility of *Plasmodium falciparum* to both dihydroartemisinin and lumefantrine in northern Uganda. Nat. Commun. 13, 6353.

Uwimana, A., Legrand, E., Stokes, B.H., Ndikumana, J.M., Warsame, M., Umulisa, N., et al., 2020. Emergence and clonal expansion of *in vitro* artemisinin-resistant *Plasmodium falciparum* kelch13 R561H mutant parasites in Rwanda. Nat. Med. 26, 1602−1608.

Van den Eede, P., Van, H.N., Van Overmeir, C., Vythilingam, I., Duc, T.N., Hung le, X., et al., 2009. Human *Plasmodium knowlesi* infections in young children in central Vietnam. Malar. J. 8, 249.

van der Pluijm, R.W., Amaratunga, C., Dhorda, M., Dondorp, A.M., 2021. Triple artemisinin-based combination therapies for malaria - a new paradigm? Trends Parasitol. 37, 15−24.

Van Dijk, M.R., Janse, C.J., Thompson, J., Waters, A.P., Braks, J.A., Dodemont, H.J., et al., 2001. A central role for P48/45 in malaria parasite male gamete fertility. Cell 104, 153−164.

van Schalkwyk, D.A., Blasco, B., Davina Nuñez, R., Liew, J.W.K., Amir, A., Lau, Y.L., et al., 2019. *Plasmodium knowlesi* exhibits distinct *in vitro* drug susceptibility profiles from those of *Plasmodium falciparum*. Int J Parasitol Drugs Drug Resist 9, 93−99.

van Schalkwyk, D.A., Moon, R.W., Blasco, B., Sutherland, C.J., 2017. Comparison of the susceptibility of Plasmodium knowlesi and Plasmodium falciparum to antimalarial agents. J. Antimicrob. Chemother. 72, 3051−3058.

van Schalkwyk, D.A., Moon, R.W., Duffey, M., Leroy, D., Sutherland, C.J., 2021. *Ex vivo* susceptibility to new antimalarial agents differs among human-infecting *Plasmodium* species. Int. J. Parasitol. Drugs Drug Resist. 17, 5−11.

van Schalkwyk, D.A., Riscoe, M.K., Pou, S., Winter, R.W., Nilsen, A., Duffey, M., et al., 2020. Novel endochin-like quinolones exhibit potent *in vitro* activity against *Plasmodium knowlesi* but do not synergize with proguanil. Antimicrob. Agents Chemother. 64, 10−1128.

van Schalkwyk, D.A., Pratt, S., Nolder, D., Stewart, L.B., Liddy, H., Muwanguzi-Karugaba, J., et al., 2023. Treatment failure in a UK malaria patient harbouring genetically variant *Plasmodium falciparum* from

Uganda with reduced *in vitro* susceptibility to artemisinin and lumefantrine. Clin. Infect. Dis. 78, 445−452. Nov 29:ciad724. Epub ahead of print.

Vembar, S.S., Seetin, M., Lambert, C., Nattestad, M., Schatz, M.C., Baybayan, P., et al., 2016. Complete telomere-to-telomere de novo assembly of the *Plasmodium falciparum* genome through long-read (>11 kb), single molecule, real-time sequencing. DNA Res. 23, 339−351.

Vinetz, J.M., Li, J., McCutchan, T.F., Kaslow, D.C., 1998. *Plasmodium malariae* Infection in an asymptomatic 74-year-old Greek woman with splenomegaly. N. Engl. J. Med. 338, 367−371.

Weatherall, D.J., Miller, L.H., Baruch, D.I., Marsh, K., Doumbo, O.K., Casals-Pascual, C., et al., 2002. Malaria and the red cell. Hematol. Am. Soc. Hematol. Educ. Program 2002, 35−57.

Weedall, G.D., Preston, B.M., Thomas, A.W., Sutherland, C.J., Conway, D.J., 2007. Differential evidence of natural selection on two leading sporozoite stage malaria vaccine candidate antigens. Int. J. Parasitol. 37, 77−85.

Winter, G., Kawai, S., Haeggström, M., Kaneko, O., von Euler, A., Kawazu, S., et al., 2005. SURFIN is a polymorphic antigen expressed on *Plasmodium falciparum* merozoites and infected erythrocytes. J. Exp. Med. 201, 1853−1863.

Wirjanata, G., Handayuni, I., Prayoga, P., Apriyanti, D., Chalfein, F., Sebayang, B.F., et al., 2015. Quantification of Plasmodium ex vivo drug susceptibility by flow cytometry. Malar. J. 14, 417.

Ye, R., Hu, D., Zhang, Y., Huang, Y., Sun, X., Wang, J., et al., 2016. Distinctive origin of artemisinin-resistant *Plasmodium falciparum* on the China-Myanmar border. Sci. Rep. 6, 20100.

Zheng, H., Zhu, H.M., Ning, B.F., Li, X.Y., 2006. [Molecular identification of naturally acquired Plasmodium knowlesi infection in a human case]. Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi 24, 273−276 (Chinese).

Further reading

Ord, R., Polley, S., Tami, A., Sutherland, C., 2005. High sequence diversity and evidence of balancing selection in the Pvmsp3 gene of Plasmodium vivax in the Venezuelan Amazon. Mol. Biochem. Parasitol. 144, 89.

CHAPTER

26

Host Adaptation in the *Mycobacterium* Genus: An Evolutionary and Genomic Perspective

Juan Calvet Seral[a], Elena Campos Pardos[a] and Jesús Gonzalo-Asensio

Mycobacterial Genetics Group, Department of Microbiology, Faculty of Medicine, University of Zaragoza, Zaragoza, Spain

1. Overview: Tuberculosis Disease and Historic Events in Mycobacterial Genetics

The *Mycobacterium* genus belongs to the phylum Actinobacteria, and it comprises Gram-positive species with a high GC genomic content (ranging from 58% to 69%). In contrast to other Gram-positive bacteria, mycobacteria possess a multilayered cell envelope rich in uncommon lipids responsible for their distinctive Ziehl-Neelsen staining. Mycobacteria are widely distributed in either the environment or infecting different hosts and they are subdivided into rapid- and slow-growing species, based on their ability to develop colonies in less or more than 7 days, respectively. Fast-growing species include opportunistic or nonpathogenic bacteria, whereas slow-growers include major human pathogenic mycobacteria, such as *Mycobacterium tuberculosis*, *M. ulcerans*, or *M. leprae* causing tuberculosis (TB), Buruli ulcer or leprosy, respectively. This chapter will focus on the *M. tuberculosis* Complex (MTBC) comprising a group of closely related subspecies or ecotypes adapted to cause TB in different mammalian hosts including humans.

TB, historically known as phthisis, consumption, or white plague, was considered as an inherited condition until the revolutionary discovery of Robert Koch in 1882, who defined TB as an infectious disease caused by a bacterial pathogen, named *M. tuberculosis*. During more than 135 years, diverse studies have clarified microbiological and immunological aspects of the disease. In the last 20 years, the dramatic increase in genomic data has allowed unprecedented insights into the evolution of TB-causing mycobacteria and the mechanisms employed to circumvent its elimination by the immune system during infection.

TB in humans is mainly known for causing a lung pathology, but disease may affect any organ. Also, as will be detailed below, other hosts aside from humans are susceptible to TB. In this context, since our clinical and scientific knowledge comes primarily from human samples, we will predominantly refer to human pulmonary TB along this chapter, albeit readers must be conscious that other

[a]Both authors contributed equally to the writing of this chapter.

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00003-3

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

manifestations exist. The infectious cycle starts transmitting the tubercle bacilli by the respiratory route. When a patient with active pulmonary disease coughs, sneezes, or even speaks, the aerosolized mycobacteria are then inhaled by neighboring individuals and they reach the alveoli, where resident macrophages phagocyte them. The fate of most bacteria in the phagosome is often fatal but *M. tuberculosis* is able to survive, and even to escape, from the phagosome. Once engulfed, *M. tuberculosis* triggers a host−pathogen cross talk causing a specific activation of the host immune system resulting in a nucleus of necrotic infected macrophages surrounded by a mantle of NK, neutrophils, T and B cells, enclosed by a fibrous cuff. Within this structure called granuloma, *M. tuberculosis* can survive for decades in an asymptomatic state, and it is estimated that one-quarter of the human population are latent carriers of the tubercle bacilli. Thus, *M. tuberculosis* has evolved mechanisms to persist in the host cells, being able to avoid the immune system but, at the same time, chronically stimulating it. In order to transmit from person-to-person, this containment must fail, which usually happens as a consequence of immune system weakening. Then, the granuloma breaks down causing a lesion in the lung where *M. tuberculosis* multiplies extracellularly, reaching the airways where bacilli can spread to the airways and the infectious cycle continues in a new infected person. The maintenance of *M. tuberculosis* in the host population relies on the ability to tightly regulate rounds of infection, reactivation to disease, and transmission. Disease occurs in approximately 5%−10% of the infected individuals, which likely represent the optimal percentage to ensure *M. tuberculosis* survival in the human population. It is possible that higher disease rates would decimate susceptible hosts while lower rates would not ensure proper transmission rates. Indeed, as will be discussed latter, TB-causing bacteria have likely adapted their virulence and transmissibility to different population densities. Unlike other microorganisms that rely on antigenic diversity to escape from the immune system, MTBC members have emerged from a genetic bottleneck and consequently show little DNA diversity. This is particularly well documented in the repertoire on human T-cell epitopes, which are hyper-conserved among *M. tuberculosis* strains, and point at this bacterium as a professional pathogen evolved for being recognized by the host immune system.

Our genetic knowledge of TB-causing bacteria required the development of suitable genetic tools (plasmids, phages, transposons, and gene replacement strategies). These pioneering studies were initiated by two independent laboratories led by Prof. Williams R. Jacobs at Albert Einstein Institute (New York) (Jacobs et al., 1987) and Prof. Brigitte Gicquel at Pasteur Institute (Paris) (Martin et al., 1990). Subsequent genetic studies sought to restore virulence to an *M. tuberculosis* attenuated strain—namely H37Ra—through genomic complementation with DNA from its virulent counterpart—namely H37Rv—resulting in the identification of a genomic segment with a potential role in virulence (Pascopella et al., 1994). Another founding study identified in 1991 the IS*6110* transposon in *M. tuberculosis*, and demonstrated its utility as epidemiological marker based on its exclusive presence in MTBC members, which differ in the copy number and location of this transposon between different strains (Otal et al., 1991). This study paved the way for "modern" molecular epidemiology in TB, and IS*6110* has been worldwide used in *M. tuberculosis* epidemiology for more than 25 years after its original description (van Embden et al., 1993).

Pregenomic studies used subtractive hybridization of an "interrogated" genome to a reference genome to identify large deletions in the "interrogated" DNA. Using this approach, several deletions (designated Regions of Difference, RD1 to RD3) were identified in the BCG vaccine relative to its parental *M. bovis* strain or to *M. tuberculosis* (Mahairas et al., 1996; Philipp et al., 1996). RD1 was latter demonstrated to be the main contributor to BCG vaccine attenuation, and conversely to

M. tuberculosis pathogen virulence. The RD1 region contains a specialized secretion system which translocate a virulent factor—named ESAT-6—outside the bacterium (Pym et al., 2002). Subsequent studies used bacterial artificial chromosomes (BACs) to refine genome maps. Further, since this BAC library achieved almost total coverage of the *M. tuberculosis* chromosome, it was a powerful tool for the H37Rv genome-sequencing project (Brosch et al., 1998). BAC libraries also allowed to gain resolution in subtractive hybridization experiments to identify 7 novel RD deletions (RD4 to RD10) in *M. bovis* relative to *M. tuberculosis* H37Rv. Some of these RD deletions were found in *M. africanum* (RD7 to RD10), which allowed to delineate the evolutive steps suffered by MTBC members to adapt to specific hosts (Gordon et al., 1999).

In 1998, more than a century after the initial description of *M. tuberculosis* by Robert Koch, the complete genome sequence of the *M. tuberculosis* H37Rv strain was deciphered. It consists of a 4.4 Mb genome coding for 3959 ORF, 45 tRNA, and 1 rRNA operon. The 65% G + C content impacts on the biased codon usage in coding sequences. An important proportion of the *M. tuberculosis* genome is devoted to lipogenesis and lipolysis, which reflects the particular lifestyle of the bacterium. Another finding from the genome sequence is the high content of proteins with repetitive structures whose role is recently being elucidated (Cole et al., 1998).

2. Host–Pathogen Co-evolution of the Tubercle Bacillus in Humans and Other Mammals

Publication of the genome sequence of *M. tuberculosis* H37Rv (Cole et al., 1998) was a scientific landmark to decipher the biology of *M. tuberculosis* since it opened new perspectives and enabled "omic" approaches. Development of H37Rv DNA probes for microarrays were used to either compare transcriptional responses or to analyze genomes from related mycobacteria by comparative genomics. Identification of RD genomic deletions in *Mycobacterium* species and the finding that they not occur independently in the different strains of the MTBC, led to propose an evolutionary landscape for the MTBC. This phylogeny is based on the idea of irreversible loss of genetic regions that cannot be compensated by Horizontal Gene Transfer (HGT) between MTBC strains. This pattern of reductive evolution might be associated with the intracellular lifestyle (which is well reflected in the obligate pathogen *M. leprae* possessing a highly reduced genome) and also consistent with the presence of a genetic bottleneck at the split between MTBC and the progenitor population (Fig. 26.1A).

According to distribution of RD deletions, MTBC species can be classified into human-adapted (*M. tuberculosis* and *M. africanum*) and animal-adapted species (*M. bovis* and related ecotypes) (Brosch et al., 2002). This evolutionary scenario allowed to revisit the preexisting hypothesis that human-infecting mycobacteria arose through zoonotic transmission of *M. bovis* during cattle domestication. In contrast, this phylogeny revealed that *M. bovis* (and animal-adapted strains) have accumulated deletions with respect to human-adapted lineages and consequently derive from these latter. Archeological records also conflict with the zoonotic hypothesis since evidences for human TB have been found in human remains dating prior to animal domestication (Fig. 26.1A and D).

Analysis of genome sequences has served to establish and update a genome-based phylogeny that is congruent with the genomic deletion-based phylogeny (Fig. 26.1C). Latest classifications divide the MTBC in 10 different linages: L1–L9 plus the animal adapted linage. The human-adapted species include *M. tuberculosis* (L1–L4, L7, and L8) and *M. africanum* (L5, L6, and L9), while the animal-

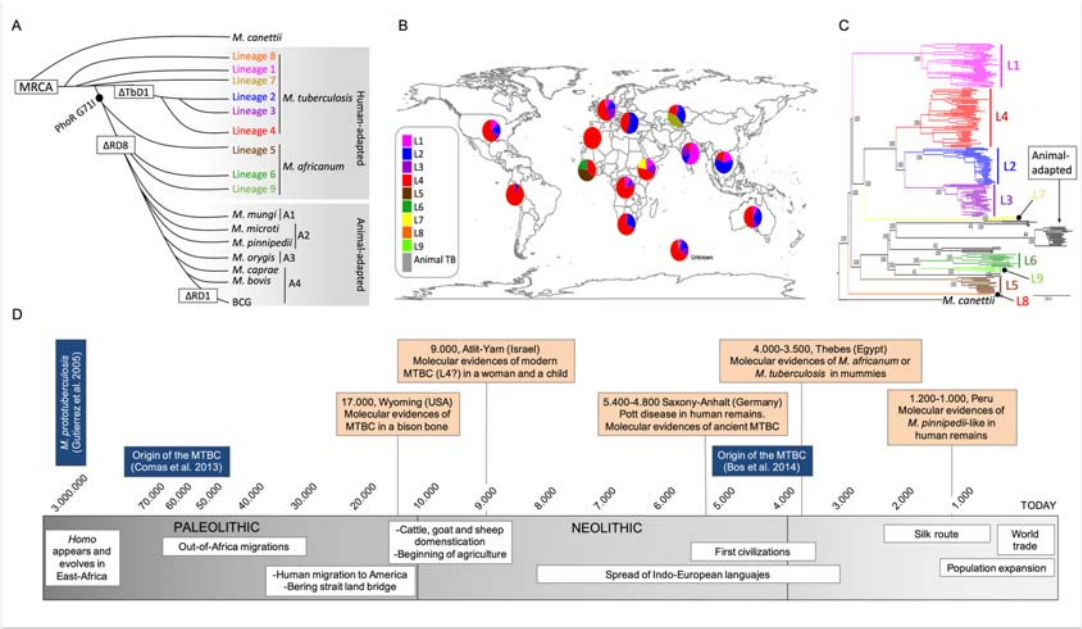


FIGURE 26.1

(A) Deletion-based phylogeny of the MTBC showing the human- and animal-adapted lineages. With the exception of RD1, RD8, TbD1, and PhoR (Gly71Ile), the remaining deletions and polymorphisms have been omitted for clarity. Please note that *M. canettii* does not belong to the MTBC. (B) Phylogeographic distribution on the main MTBC lineages causing TB in humans. L2, L3, and L4 strains are globally distributed while L1, L5, L6, L7, L8, and L9 lineages are geographically restricted. (C) Maximum-likelihood phylogeny of the MTBC. Note the similarity with deletion-based phylogeny. (D) Time-line showing the main historic events of humankind in relation to MTBC evolution. *Orange boxes* indicate relevant archeological and molecular evidences for MTBC in human and animal remains. *Dark blue boxes* indicate estimates about the origin of *M. prototuberculosis* and the MTBC. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

Images from panels A, B and C have been reproduced and/or adapted from the following Open-Access references Broset, E., Martin, C., Gonzalo-Asensio, J., 2015. Evolutionary landscape of the Mycobacterium tuberculosis complex from the viewpoint of PhoPR: implications for virulence regulation and application to vaccine development. mBio 6; Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., Otchere, I.D., Asante-Poku, A., Asare, P., Sanchez-Buso, L., Gehre, F., Sanoussi, C.N., Antonio, M., Affolabi, D., Fyfe, J., Beckert, P., Niemann, S., Alabi, A.S., Grobusch, M.P., Kobbe, R., Parkhill, J., Beisel, C., Fenner, L., Bottger, E.C., Meehan, C.J., Harris, S.R., de Jong, B.C., Yeboah-Manu, D., Brites, D., 2021. Phylogenomics of Mycobacterium africanum reveals a new lineage and a complex evolutionary history. Microb. Genom. 7; Napier, G., Campino, S., Merid, Y., Abebe, M., Woldeamanuel, Y., Aseffa, A., Hibberd, M.L., Phelan, J., Clark, T.G., 2020. Robust barcoding and identification of Mycobacterium tuberculosis lineages for epidemiological and clinical studies. Genome Med. 12, 114.

adapted ecotypes are *M. bovis*, *M. caprae*, *M. microti*, *M. pinnipedii*, *M. orygis*, and *M. mungi*, grouped into four different animal clades (A1−A4) (Brites et al., 2018). Similarly to the animal-adapted ecotypes that have evolved to infect specific mammals, human-adapted strains have evolved to cause TB in human subpopulations. Accordingly, *Mycobacterium canettii* strains (which are thought to

resemble in many aspects the putative common ancestor of the members of the MTBC) have mainly been isolated in the horn of Africa (Fabre et al., 2010); *M. africanum* L5 and L6 strains are commonly found in west African countries (de Jong et al., 2010) and *M. africanum* L9 isolates came from Somalia and Djibouti (Coscolla et al., 2021); *M. tuberculosis* L8 has been isolated in the African Great Lakes region (Ngabonziza et al., 2020), L7 is frequent in Ethiopia, L1 in the Indian-ocean rim and L2, L3, and L4, which show wider distribution, predominantly infect people from east Asia, east India and America-Europe-Africa, respectively (Comas et al., 2013; Napier et al., 2020) (Fig. 26.1B). A more detailed look at the evolution of the Beijing lineage (L2), which is associated with the massive spread of drug resistance in Eurasia, suggest that worldwide colonization with this lineage started with the onset of agriculture in China, Korea, and Japan about 6000 years ago (Merker et al., 2015).

Phylogenies of the MTBC are also useful to estimate the most recent common ancestor (MRCA) of the tubercle bacillus. *Mycobacterium canettii* (formerly known as *Mycobacterium prototuberculosis*) predated the clonal expansion of the MTBC and therefore this species is considered the progenitor of the tubercle bacillus. The origin of *M. canettii* is situated 2.6–2.8 millions years ago. Thus, TB is older than plague, typhoid fever, or malaria, and might have affected early hominids. Consistent with this hypothesis, Australopithecus and *Homo habilis* were present 3 million years ago in the horn of Africa, a region where most *M. canettii* strains have been isolated (Gutierrez et al., 2005) (Fig. 26.1D). *M. canettii* presents particular genomic characteristics with respect to MTBC members: (i) there are evidences for HGT and recombination events, which point to a putative environmental source where frequent contact with different bacteria might allow interbacteria DNA transfer, (ii) it presents larger genome sizes in contrast with the reductive evolution of the MTBC, and (iii) there is a 25-fold higher frequency of single nucleotide polymorphisms (SNPs) compared to MTBC. These features suggest that *M. canettii* diverged from the MRCA of the tubercle bacilli before the clonal expansion of the MTBC (Supply et al., 2013).

Assuming a constant mutation rate, it has been possible to date the origin of the MTBC speciation. However, there is still debate about the exact date when speciation occurred. One controversial study attributed a tuberculous leptomeningitis diagnostic to a *Homo erectus* frontal bone from Turkey (500,000 years ago) which presented characteristic lesions of TB (Kappelman et al., 2008). However, this diagnostic has been questioned by other authors (Roberts et al., 2009), and the finding is controversial as it suggests that TB predated the origin of modern humans. Some studies situate this origin less than 6000 years ago (Bos et al., 2014) while others propose a more ancient origin around 70,000 years ago (Comas et al., 2013). The oldest archeological evidence of MTBC comes from a bison bone dated back 17,000 years ago and molecular clocks should predate this irrefutable proof. Domestication of cattle began 10,500 years ago and molecular evidences for the presence of MTBC DNA in this ancient bison reinforce the idea that even if TB arose as a zoonosis, it predated animal domestication (Fig. 26.1D).

Considering the ancient origin and genomic features of *M. canettii*, the genetic bottleneck of the MTBC prior to animal domestication, and the lack of evidence for human-to-human transmission of *M. canettii*, the origin of human TB has been proposed as an opportunistic infection from an environmental reservoir (most probably somewhere in the Horn Africa) by the progenitor of the tubercle bacilli (Koeck et al., 2011). Eventually, this opportunistic bacterium would have evolved as a professional pathogen in mammalian hosts (Gagneux, 2012). One hypothesis contemplates that a key step into adaptation to mammalian hosts was the ability of the bacilli to survive into free-living protozoa, which feed from environmental bacteria, as a previous step of being able to multiply inside

mammalian macrophages (Jang et al., 2008). In this context, it has been recently demonstrated that *M. tuberculosis,* but not its ancestor *M. canettii*, has evolved to depend on vitamin B12 supplied by host cells. This B12 dependency would otherwise result in lower virulence of *M. tuberculosis* under conditions of B12 deficiency and viceversa (Campos-Pardos et al., 2024). Another theoretical study proposes that fire domestication had an impact in the human-to-human transmission of the bacilli by simultaneously promoting the social interaction and inducing lung damage by the smoke (Chisholm et al., 2016).

On another scenario, assuming that *M. bovis* and other animal-adapted strains derive from *M. tuberculosis*, this implies that the MRCA was adapted to humans and consequently humans would have given TB to animals via "reverse" zoonosis (Smith et al., 2009). Since *M. africanum* phylogenetically lies between *M. tuberculosis* and animal-adapted species, together with the high prevalence of *M. africanum* in West African countries, and the sporadic isolation of *M. africanum* from monkeys and cows in this region (de Jong et al., 2010), it is tempting to speculate that *M. africanum* represent an opportunistic spill over transmission into animals from humans.

Association of MTBC members with certain mammals represent an attractive example of host−pathogen adaptation. Considering the highly conserved genome of MTBC species, it is intriguing to think that two humans chosen at random have twice (0.1%) the sequence diversity than human-adapted *M. tuberculosis* and cattle-adapted *M. bovis* (0.05%). However, after a critical examination of these numbers, a 0.05% sequence divergence is translated into approximately 2000 polymorphisms between two MTBC strains. Considering that these species code for roughly 4000 genes, it is inferred that half of the genes are virtually affected by polymorphisms. Even if some polymorphisms represent silent mutations, it is tempting to assume that some polymorphisms might have determined the host range of MTBC. As detailed later, the profound implications in virulence for a single amino acid mutation in *M. bovis* and *M. africanum* with respect to *M. tuberculosis* strains (Gonzalo-Asensio et al., 2014) was recently demonstrated (Fig. 26.1A).

Hereafter, we will focus on the host−pathogen evolution for the most relevant MTBC species that cause TB in diverse hosts.

2.1 *M. tuberculosis*

This bacterium is the major, but not the unique, cause of TB in humans. TB caused 20% mortality between the 17th and 19th centuries and remains today the most lethal infectious disease, only surpassed by COVID-19. The close association of *M. tuberculosis* with humans is demonstrated by the striking similarity of the phylogenetic trees from 220 strains of the MTBC and 4.955 mitochondrial genomes representative of the main human haplogroups. The incidence of TB was not only markedly increased as a consequence of the Neolithic expansion but also this period triggered an increment in the population diversity of the MTBC and humans. Therefore, it is suggested that changes in human demography have shaped the evolution of *M. tuberculosis* as demonstrated by the existence of different lineages with specific geographic distribution. Further, dating estimates of the different lineages remarkably correlate with human migratory movements. Dispersal of modern humans occurred in two major waves: an initial dispersal around the Indian Ocean 62,000−75,000 years ago and a later dispersal into Eurasia 25,000−38,000 years ago. The first wave would have given rise to L1 (East Africa, Philippines and Indian Ocean rim) and the second wave to L2 (East Asia), L3 (East Africa, Central Asia) and L4 (Europe, America, Africa). Later migrations, conquests and slave trade would

have contributed to the worldwide spread of L4 strains into America and Africa. The L7 (Ethiopia) might have arisen among a human population that remained in or returned to Africa. Concerning L8, which could be more ancient of all linages studied to date, seems to be restricted also to the African Great Lakes region (Fig. 26.1B). Altogether, the African origin of modern humans and the coalescence of all MTBC branches in this continent reinforces the hypothesis of the human origin of TB (Comas et al., 2013; Gagneux, 2012; Ngabonziza et al., 2020).

2.2 *M. africanum*

M. africanum constitutes the second major cause of human TB (Fig. 26.1B). Although *M. africanum* has been identified alongside *M. tuberculosis* in human remains from Egypt dating back 4000 years ago, today lineages 5 and 6 of *M. africanum* are geographically restricted to West Africa where they show a prevalence ranging to 21%–66%. Recently, it has been described a new linage of *M. africanum*, L9, whose most likely ancestral distribution is East Africa (Coscolla et al., 2021).

Out-of-Africa cases of *M. africanum*–related TB are rare and frequently linked to African immigrants (de Jong et al., 2010). This highlights the exquisite adaptation of this species to a geographically restricted population. It has been reported that although *M. tuberculosis* and *M. africanum* L6 are equally transmitted in the Gambia, a 2-year follow-up demonstrated a fivefold reduction in *M. africanum* L6 relative to *M. tuberculosis* in progression to active disease (de Jong et al., 2008). It is also remarkable that *M africanum* did not established in the New World despite massive slave trade from 15th to 19th centuries, which likely indicates that *M. africanum* was readily outcompeted by *M. tuberculosis* in the Americas. Further, the phylogenetic relatedness of *M. africanum* with animal strains, the evidence of *M. africanum*–infected animals and the low transmission to disease in humans might point at this specie either as a spill over transmission to animals by reverse zoonosis or an animal reservoir for *M. africanum*.

2.3 *M. canettii*

It is important to remember that *M. canettii* is considered an ancestral lineage from which the MTBC emerged and consequently does not belong to this latter (Fig. 26.1A). Alongside *M. tuberculosis* and *M. africanum*, *M. canettii* cause human TB. In contrast, *M. canettii* exhibit distinctive smooth colony morphology in culture media and human isolates are geographically restricted to East African populations. It is estimated that this species emerged 2.8 million years ago, which mean that early hominids in East Africa would have been infected with a TB-like disease (Fig. 26.1D) (Gutierrez et al., 2005). At the genome level, *M. canettii* strains have distinctive features compared with the MTBC. Whole genome sequencing of five representative isolates from different subtypes (A, D, L, J, and K) showed a similar organization and synteny with MTBC genomes. However, *M. canettii* genomes are 10–115 Kb larger than those of the MTBC and they present evidences for HGT and recombination that resulted in a genetic mosaicism of *M. canettii* inherited by MTBC. In fact, given that *M. canettii* strains might have an environmental reservoir where different strains can establish direct contact, the existence of HGT mechanisms in this species is plausible. Recent studies at the Pasteur Institute (Paris) have demonstrated that *M. canettii* uptake DNA from MTBC members and integrate it in a highly resembling distributive conjugal transfer manner. In contrast, *M. tuberculosis* and *M. bovis* failed to act as conjugative strains, what agrees with the clonality observed along the MTBC (Boritsch et al.,

2016b; Madacki et al., 2021). Phenotypically, *M. canettii* strains grow faster than *M. tuberculosis* especially at low temperatures (30°C) and show reduced persistence in the mouse model of aerosol infection. These features reflecting broader environmental adaptability but lower virulence and transmission makes us to think on *M. canettii* as a missing link between the obligate pathogens of the MTBC and a putative progenitor environmental mycobacteria (Supply et al., 2013). Genomic comparison of *M. canettii* strains belonging to a pathogenic outbreak has raised the question about the eventual increased virulence of particular strains. This study suggests that current epidemic clones of *M. canettii* might mimic an earlier phase before MTBC speciation. Thus, the eventual acquisition of the appropriate mutations, either in the environmental reservoir or within an infected host, would enable person-to-person transmission of *M. canettii* (Blouin et al., 2014). In this line, an experimental study has described a role for lipooligosaccharide (LOS) synthesis in host−pathogen interaction and in *M. canettii* virulence. Smooth *M. canetii* variants producing LOS are less virulent than rough variants lacking LOS. Further insight into the biological mechanism indicates that recombination between two *pks5* genes is responsible for the abrogation of LOS synthesis, and the resulting genetic organization in this locus is similar to that of *M. tuberculosis* (Boritsch et al., 2016a). In another related study, it has been demonstrated that experimental evolution of *M. canettii* during in vivo passages in mice, results on gain of function, instead of loss of function mutations. Authors identify and validate two loci conferring enhanced in vivo persistence of *M. canettii*. Altogether, these findings reinforce the notion that *M. canettii* have the intrinsic ability to evolve from an environmental opportunist to a professionalized human-adapted pathogen (Allen et al., 2021).

2.4 Animal-Adapted Ecotypes: The Four Animal Clades

Recently, it has been proposed a new phylogenetic framework to classify these ecotypes into four animal clades: A1, A2, A3, and A4. All known animal-adapted members of the MTBC share a common ancestor at the branching point which is characterized by clade-specific deletions in RD7, RD8, RD9, and RD10 (Brites et al., 2018). Of note, the human-adapted *M. africanum* L6 and L9 shares a common ancestor with A1-A4 animal clades (Smith et al., 2006). This new proposed phylogeny would be compatible with two scenarios during the evolutionary history of the nonhuman MTBC. One scenario would present the ancestor of the group including L6, L9, and all animal-adapted clades as a generalist capable of infecting a wide group of mammals, and different host adaptations would have occurred thereafter. In another possible scenario, the ancestor of L6, L9, and animal-adapted MTBC was adapted to humans, and subsequent host jumps lead to the host specificity of the four clades (Brites et al., 2018).

The A1 clade members are close relatives of the *M. africanum* L6 and L9, and here belong *M. mungi*, *M. suricattae*, the dassie bacillus and the chimpanzee bacillus. Interestingly, *M. mungi*, *M. suricattae*, and the dassie bacillus have different RD1 deletions, whereas in chimpanzee bacillus this deletion is not present, suggesting that the RD1 region might be essential for virulence, not only in humans, but also in primates (Dippenaar et al., 2015).

The second animal clade, A2, comprises *M. pinnipedii* and *M. microti* ecotypes. *M. pinnipedii* infects marine mammals (seals and sea lions) (Cousins et al., 2003), however contemporary *M. pinnipedii* is known to infect humans occasionally (e.g., zoo keepers or seal trainers). It has been reported that sea mammals could have played a role in transmitting TB to the Americas across the ocean in the precontact era. Phylogenetic data indicate that modern strains of *M. tuberculosis* from

America are closely related to those from Europe (L4) supporting the assumption that TB in the New World was introduced during colonization. However, archeological evidences of TB in 1000 years old pre-Columbian and Peruvian human mummies are incompatible with this notion since these evidences were present before European contact. Genome sequencing revealed that an MTBC member closely related to *M. pinnipedii* was causative of human TB in a Peruvian population. This scenario is supported by zoonotic transmission of *M. pinnipedii* to humans, resulting from the handling and consumption of seal or sea lion meat at that time, and also with the transmission of *M. pinnipedii* to Australian seal colonies through transatlantic movements (Bos et al., 2014). Since *M. pinnipedii* genome has been sequenced from precontact human skeletal remains from inland Peruvian and Colombian sites where contact with infectious seals and sea lions is unlikely to have occurred, these findings suggest that, in addition to zoonotic events, a possible human-to-human transmission of *M. pinnipedii* is also possible (Vagene et al., 2022). As to *M. microti*, it was isolated from voles in 1930s, but has since then been found in cats, pigs, wild boars and llamas and it has been also isolated once from an immuno-compromised human (van Soolingen et al., 1998). *M. microti* carries an RD1 deletion, supporting its low virulence in humans (Brites et al., 2018).

In contrast to the Clades A1 and A2 comprising multiple ecotypes, clade A3 only comprises *M. orygis*. Even though *M. orygis* is primarily adapted to Cervidae species, they can also infect cattle and a substantial proportion of human isolates come from patients with zoonotic infections. Interestingly, all human isolates are born in India, Pakistan, Nepal, or South and South-East Asia. In addition, A3 clade shares its most recent ancestor with animal clade A4, so it has been hypothesized that its evolutionary success might also be due to the ability of infecting domestic animals such as cattle in South Asia, leading to zoonotic TB in humans through consumption of raw milk (Brites et al., 2018).

TB in animals is mainly known from cases in cattle and goats, caused by *M. bovis* and *M. caprae*, respectively, and together form the fourth animal clade, A4. *M. bovis* is the major source of animal TB, disease commonly referred to as bovine TB. Its genome is 99.95% identical to *M. tuberculosis*. Hence, the about 2.000 polymorphisms, insertions and deletions differential between both strains clearly bias the host tropism. *M. bovis* requires pyruvate when glycerol is the primary carbon source due to an inactive pyruvate kinase. This finding, in addition to have implications during laboratory work, will surely have functional implications in the infection of the cattle host. Bovine TB has been documented across all continents except Antarctica leading to the general assumption that bovine disease is present where cattle are found. Historical data suggest that bovine TB could have emerged in Europe, and thereafter, it would have been distributed to European colonies by exportation of infected animals. Aside from being an economic problem for farmers, bovine TB has historically represented a primary cause of human TB through consumption of unpasteurized milk. Pasteurization of milk and epidemiology control measures that implies test and slaughter of infected cows have served to substantially reduce the incidence of bovine- and related-TB in humans (Smith et al., 2006). Today, three major *M. bovis* complexes (African 1, African 2, and European 1) exist with characteristic phylogeographical distribution (Smith, 2012). These complexes are likely the result of a founder effect by which a clone is introduced in a new territory and subsequently flourishes in the absence of competition. The African 1 (Af1) subtype is characterized by the loss of RDAf1 and is frequently found in West-Central Africa. However, typing has demonstrated that each country has unique population structures indicative that transhumance movements have not substantially contributed to disperse and homogenize the Af1 population. The African 2 (Af2) subtype is characterized by the loss of RDAf2. Af2 isolates are

dominant in East Africa and similar to Af1 strains, there exist unique population structures suggesting that population mixing is infrequent. Altogether Af1 and Af2 are mutually exclusive since they do not share phylogenetic history and rarely found out of Africa. The European 1 (Eu1) subtype has a specific chromosomal deletion (RDEu1) and it is worldwide distributed in America (with the exception of Brazil), East Asia, Australia, New Zealand, South Africa, and the British Isles. The dominant presence of Eu1 in the British Isles and British colonies suggests that Eu1 subtype was distributed from the former. This hypothesis is supported by trading records of Hereford beef cattle, which was exported by United Kingdom since the early 19th century. It remains to be answered the *M. bovis* specialization to infect different cattle hosts and the contribution of social, historical, and agricultural factors to the current scenario of bovine TB (Smith, 2012).

Disease in goats caused by *M. caprae* is the second major source of animal TB and an important source of zoonosis. Although *M. caprae* was originally considered a subspecies of *M. bovis*, genetic evidences support the elevation of *M. caprae* as a separate member of the MTBC. Indeed, if we look at the deletion-based phylogeny, *M. bovis* lacks RD4 region with respect to *M. caprae*, pointing at caprine isolates as ancestors of bovine isolates (Aranaz et al., 2003) (Fig. 26.1A). Considering that goat and cattle domestication started 10,000−11,000 years ago, it remains unanswered whether bovine TB is the result of a spill-over transmission from infected goats or a scenario where *M. bovis* and *M. caprae* arose independently to infect specific hosts. Genome data from *M. bovis* and *M. caprae* are limited, but the advent of next-generation sequencing is filling the gap in the knowledge of animal-adapted strains. Whole genome sequencing of *M. bovis* and *M. caprae* field isolates have served to validate the ancestry of *M. caprae* relative to *M. bovis* (de la Fuente et al., 2015).

3. Evolution of the MTBC From a Genomic Perspective

The growing availability of mycobacterial genomes has been useful to understand which driving forces have shaped the MTBC evolution. The strict clonality and the reductive evolution by accumulating genetic deletions might be indicative of bacterial adaptation to the intracellular environment. In this context, the obligate and noncultivable intracellular pathogen *M. leprae* has a drastic gene reduction resulting in barely 1.600 ORF compared to 3900 ORF in *M. tuberculosis*. *Mycobacterium* not only offers several notorious examples of naturally reduced but also laboratory-reduced genomes (e.g., *M. bovis* BCG) (Orgeur and Brosch, 2018).

Based on the hypothesis that the ancestor of MTBC species was an environmental bacterium, HGT is likely to have occurred between this ancestor and other microorganisms living in the same environment. A study based in determination of G + C content, codon usage and genomic signatures identified genomic islands present in the *M. tuberculosis* genome but absent from closely related species not belonging to the MTBC (*M. marinum* and *M. ulcerans*) (Becq et al., 2007). Importantly, not all *M. canettii* strains carry these genomic islands, suggesting that they could have been acquired by HGT before the speciation of the MTBC. A large 13 Kb genomic island coding for 15 ORF encompasses the rv1041c-rv1055 genes, and some genes from this HGT region are involved in virulence either in cultured macrophages or immunodeficient mice, pointing at a pathogenicity island (Becq et al., 2007). Fixation of an HGT-transferred region frequently relies upon conferring a selective advantage to the recipient bacteria. HGT transfer in the MTBC ancestor might have helped this bacterium to invade and persist within environmental protozoan phagocytes, which would have

eventually evolved to survive within mammalian phagocytic cells. This notion is supported by the observation that pathogenic but not avirulent mycobacteria survive and persist into amebas (Jang et al., 2008). One example of possible region fixed through HGT is the *espACD* region that might represent a sort of pathogenicity island, since it is involved in virulence in the MTBC and, interestingly, it has been fixed also in different pathogenic slow growing mycobacterial species in different genome locations (Ates and Brosch, 2017). Further validating the assumption that the ancestor of MTBC evolved through HGT, it has been demonstrated that different subtypes of *M. canettii* vary in the presence and location of DNA regions, conferring a composite, mosaic-like structure to the ancestor of the tubercle bacillus (Gutierrez et al., 2005).

The split between MTBC and its ancestral progenitor is characterized by a genetic bottleneck that reduced to a minimum the genetic diversity. Speciation in the MTBC entails a host preference and the strict clonality in the MTBC could be either an adaptation or a consequence of pathogenicity in a restricted host population. The evolutionary scenario of the MTBC indicates a reductive evolution (Fig. 26.1A and C) suggesting the fixation of different polymorphisms during speciation (Brosch et al., 2002; Comas et al., 2013). With the technological explosion of whole genome sequencing our knowledge about the timing of the molecular clock of the MTBC is increasing. The study of 6.285 genomes with different epidemiological settings and global diversity has helped to establish a clock calibration between 0.04 and 2.2 SNPs per-genome-per-year, with considerable differences among sublineages/clades (Menardo et al., 2019). These results support the strong clonality of the MTBC population.

Combined with the acquisition of new genes through HGT, deletions of genes or genome regions also enabled the transition of the MTBC ancestor to a professional pathogen. A good example is the 2.593-bp genomic deletion of the MmpS6/MmpL6-encoding an *M. tuberculosis*-specific deletion 1 (TbD1) region, which is presented as a differential genomic characteristic of the "modern" human adapted linages L2, L3, and L4. These are the most widespread linages in human population, whereas TbD1-intact "ancestral" lineages are manly restricted to specific geographical areas, such as South India and South East Asia or East Africa. This TbD1 deletion confers increased resistance to oxidative stress and hypoxia, and correlates with enhanced virulence in different infection models. It is also though that this deletion is one of the genomic reasons in the success of these "modern" lineages for their global spread (Bottai et al., 2020).

However, despite the prevailing view that MTBC genomes have evolved by genetic decay and polymorphism fixation, recent studies have revealed large-scale duplications in *M. tuberculosis*. It was initially thought that genome duplications were exclusive of the Beijing (L2) lineage, but our cumulative genomic knowledge has allowed the identification of large genome duplications also in L4 strains. The finding that the *rv3128c-rv3644* region is prone to duplication events in *M. tuberculosis* might suggest a selective advantage for strains carrying duplications (Weiner et al., 2012). Supporting this fitness advantage, it has been demonstrated that drug resistance and virulence is associated with gene duplication events in different bacteria (Galagan, 2014). However, it has been also documented the opposite effect for gene duplication; the finding that a 350 Kb duplication in L2 strains is selected when bacteria are grown in vitro but it results in lower virulence for mice argue against an overall fitness advantage for gene duplication events (Domenech et al., 2014). Gene duplication also occurs in the H37Rv laboratory strain of *M. tuberculosis*, indicative that these polymorphisms are not exclusive of human isolated strains and consequently it is an inherent feature of the *M. tuberculosis* chromosome. Although boundaries of some duplicated regions are flanked by repetitive or transposable

elements, the precise mechanism for gene duplication remains elusive. Another mechanism contributing to genome evolution in the MTBC are deletions mediated by the mobile element IS*6110*. This transposon is exclusively present in MTBC members in variable number and localization making it an invaluable tool for molecular epidemiology of TB (van Embden et al., 1993). The presence of two adjacent IS*6110* copies can lead to recombination with the concomitant loss of the genomic region between them (Sampson et al., 2003). In order to detect or predict IS*6110*-mediated deletions it is required to know the precise insertion sites across the chromosome. This task is hardly accomplished by current next generation sequencing methodologies due to the repetitive nature of IS*6110*, being necessary to locate insertion sites by either single molecule sequencing methods as PacBio or Nanopore, or by molecular biology techniques. Recently, it has been reported that IS*6110*-mediated mutations in the *ppe38* locus block secretion of PE_PGRS proteins and this result in enhanced virulence, reinforcing the role of IS*6110* as an evolutionary force in the MTBC (Ates et al., 2018). In addition to contribute to the genome plasticity of MTBC members, the mobile element IS*6110* has a function as a mobile promoter and this effect is more prominent in an intracellular environment (Alonso et al., 2011). Further, it has been recently demonstrated that IS*6110* expression correlates with IS*6110* copy number resulting in higher transposition events in human-adapted strains, which usually carry more IS*6110* copies than animal-adapted strains. These results suggest that transposition events are dynamically adapted to the host (Gonzalo-Asensio et al., 2018). Finally, since transposition frequency in the MTBC is higher than the mutation rate in this species, it is expected that IS*6110*-mediated changes (either deletions or increased expression of flanking genes) are important driving forces in the evolution and host adaptation of the MTBC.

4. Evolution in the Laboratory Environment and In Vitro Attenuation of Bacteria From the MTBC

In terms of pathogenesis, it is tempting to think that bacterial evolution is exclusively focused on host adaptation. However, in the context of the MTBC there are two historic examples of bacterial attenuation in the laboratory that deserve attention: the H37 dissociation and the development of BCG. These scientific landmarks have provided mycobacteriologists with laboratory reference strains and a vaccine against TB.

The *M. tuberculosis* H37 parent strain was originally isolated from a patient with chronic pulmonary disease by Edward R. Baldwin in 1905. In 1934, Steenken and collaborators reported that repeated subculture of the H37 strain in solid egg media at pH = 6.2 resulted in two different colony morphologies. The passages continued until two stable variants were obtained. One of these variants was completely attenuated when inoculated in guinea pigs, a highly sensitive model of TB. Accordingly, these strains were designated H37Rv (v for virulent) and H37Ra (a for avirulent) (Steenken et al., 1934). Both strains were maintained at Trudeau Institute (New York) for many years and were later deposited in the American Type Culture Collection. Historically, H37Rv and its attenuated counterpart H37Ra have been widely used as reference strains for virulence studies of *M. tuberculosis* since 1940. The genetic basis for H37Ra attenuation remained unclear until 2008 when its genome sequence was publicly available 10 years after deciphering the genome sequence of its H37Rv counterpart. When compared with H37Rv, H37Ra acquired multiple point mutations, deletions, and/or genomic rearrangements during in vitro passage. Although the precise role for every

polymorphism in H37Ra attenuation is elusive, it has been reported that a point mutation resulting in a single amino acid substitution (Ser219Leu) in the *phoP* gene have significantly contributed to H37Ra attenuation (Lee et al., 2008). This gene is part of the PhoP-PhoR two-component virulence system, which is a well-known regulator of pathogenic phenotypes, including secretion of the virulence factor ESAT-6, biosynthesis of acyltrehalose-based lipids, and modulation of antigen export, in members of the *Mycobacterium tuberculosis* complex (MTBC) (Broset et al., 2015). As a consequence of this mutation, the PhoP-PhoR virulence system is nonfunctional in H37Ra, and accordingly, this strain lacks immunomodulatory lipids and does not secrete ESAT-6 (Chesne-Seck et al., 2007; Frigui et al., 2008). However, even if these phenotypes are key contributors to the avirulence of H37Ra, they fail to explain the complete attenuation of this strain. Thus, other polymorphisms arisen during in vitro passages (for example, those leading to loss of the phtiocerol dimycocerosate (PDIM) virulence lipid) likely contribute together with the *phoP* mutation to the avirulent phenotype of H37Ra.

The BCG vaccine, which in 2021 celebrated its 100th anniversary, receives its name from "Bacille de Calmette et Guérin" due to the scientists who developed this strain more than a century ago. In 1900, Albert Calmette and Camille Guérin at the Pasteur Institute in Lille began their research for a TB vaccine. In an attempt to avoid clumping of mycobacteria in liquid broth, they eventually discovered that addition of ox bile to the medium resulted in lowering the virulence of the culture. They applied this knowledge to attenuate an *M. bovis* strain isolated in 1902 from a cow with TB mastitis. In 1908, Calmette and Guérin started successive cultures of *M. bovis* in bile, glycerine and potato medium. By 1913, their plan to initiate a vaccination trial in cattle was interrupted by World War I and they continued *M. bovis* subcultivation throughout the German occupation of Lille. In 1919, after 230 passages, they obtained a bacterium that resulted attenuated in rabbits, guinea pigs, cattle, and horses. This paved the way for a human vaccination trial in infants who were victims of TB mainly through consumption of untreated cow milk. This trial started in Paris in 1921 with a severe disease reduction in vaccinated children compared to the unvaccinated controls (Calmette, 1931). After these encouraging results, BCG lots were distributed in 1924 to different countries where they continued being propagated in the same original way as at the Pasteur Institute. Subcultivation was the unique mean available to maintain the vaccine properties of BCG and to avoid the reversion to a virulent state. In 1960s−1970s, the availability of low temperature freezers and/or lyophilizers, made possible that master seed lots of BCG were prepared worldwide. Today, we know that during its worldwide subcultivation, BCG continued the in vitro evolution leading to accumulation of specific deletions and polymorphisms that have ultimately resulted in the presence of BCG substrains with different vaccine potential (Behr, 2002). Almost a century after its initial construction, we know the attenuation basis of BCG. Compared to *M. bovis*, BCG has lost more than 100 genes most of them grouped in RD regions (Brosch et al., 2007). Specifically, deletion of the 9.5 Kb RD1 region in BCG is thought to have substantially contributed to BCG attenuation (Fig. 26.1A). The RD1 region encoding Rv3871−Rv3879 genes includes ESAT-6, known to be a key virulence factor of virulent mycobacteria (van der Wel et al., 2007). In addition, sequencing of 14 BCG strains in conjunction with transcriptional and proteomic analysis have served to reveal the profound differences existing between BCG strains (Abdallah et al., 2015) as a consequence of a 100-year in vitro evolution.

Furthermore, it is well documented that during extended periods of in vitro culture, *M. tuberculosis* undergoes a spontaneous loss of PDIM known to be a key virulence lipid. Consequently, selection of PDIM-negative colonies might bias experimental findings (Domenech and Reed, 2009). In the same

line, another study identified a polymorphism in the promoter region of the *whiB6* gene that alters its regulation. Importantly, this polymorphism that results in a decreased secretion of ESAT-6 is exclusive of *M. tuberculosis* H37Rv being absent in the remaining members of the MTBC (Solans et al., 2014). Aside from the role of WhiB6 in regulating ESAT-6 secretion, other WhiB6-regulated functions might bias the experimental findings when using a single reference strain, making advisable the use of several strains in laboratory research.

5. Short-term Evolution of *M. tuberculosis* During Infection, Drug Treatment, and Disease

The current phylogenetic tree of the MTBC is the result of a long host—pathogen co-evolution during thousands of years. This speciation in the MTBC has likely occurred by fixation of polymorphisms conferring selective advantages to infect a specific host. Thus, it is tempting to assume that different polymorphisms arise during a single infection-disease-transmission cycle and those allowing a better bacterial survival in the host population will become fixed and maintained in a bacterial lineage. During the course of infection, members of the MTBC face numerous stresses and the ability of the bacillus to overcome these barriers will likely result in fixation of advantageous alleles. However, identification of these polymorphisms has remained hampered until recently due to two main reasons: the broad consideration that the MTBC is strictly clonal and the technical limitation to sequence different bacterial populations from a single host. Today, the use of sensitive, high density sequencing techniques has allowed the transition from a scenario where each host is represented by a single MTBC strain isolated in a pure culture to a scenario where each host is known to harbor a diverse bacterial population.

A pioneer study using the cynomolgus macaque model aimed to identify polymorphisms in *M. tuberculosis* during active, latent, or reactivated disease by sequencing 33 strains representative of these infection stages. Although no insertions or deletions were detected, 14 SNP were identified and validated by traditional Sanger sequencing. These SNPs were not present in the initial inoculum suggestive that they arose during the course of the infection. The mutational capacity, calculated as the mutation rate per generation, was comparable in active, latent, and reactivated bacteria. Within a specific lesion, SNPs were independent or shared between the sequenced strains, which indicates that some SNP accumulate within lesions over the course of the infection. In addition, the study suggests that the polymorphic pattern is the result of oxidative DNA damage during the adaptive immune response to the infection (Ford et al., 2011).

In a 2015 study, genome data from *M. tuberculosis* isolated during the course of infection in five different patients were interrogated. The experimental design included *M. tuberculosis* strains from different lineages and different susceptibility profiles to anti-TB drugs and patients from different origins. A series of statistical measures to quantify the mutation rate within a patient revealed that several SNP arose across the course of the infection. Diversity intrapatient varies dramatically, likely as a function of disease severity. When examining the SNP diversity intra- and between-patient it was observed that genes involved in the regulation, synthesis, and transport of immunomodulatory lipids were prone to SNP accumulation (O'Neill et al., 2015). Given the role of these lipids in the interplay with the host immune system, it is not surprising that genes related to their synthesis are linked to a positive selection. Although this study does not represent either the global host population or the broad

MTBC phylogeny or the entire bacterial population in the infected lung, it reinforces the notion of intrapatient diversity.

Studies from 2016 to 2017 have demonstrated that purifying selection act during infection inside patients, without the need of human-to-human transmission. *M. tuberculosis* diversifies withing the individual as infection progress and reaches different parts of the lungs or even other organs, and that this evolution occurs in presence (Trauner et al., 2017) and absence of drug treatment (Lieberman et al., 2016). Overall, this kind of studies starts to delineate the evolutionary forces that can turn a single bacterium into a widespread lineage, but future works are needed to decipher the biological significance of these polymorphisms arisen in vivo in the context of host−pathogen co-evolution.

A particular scenario of MTBC adaptation to the host is related to drug treatment of infected patients. Since anti-TB drugs represent an immediate threat for bacterial survival, polymorphisms conferring drug resistance will be rapidly fixed in the population. The evolutionary forces governing host adaptation and drug resistance operate at different timescales and they affect different alleles. A study collected a total of seven sputum samples from three patients and examined genetic mutations during different stages in the development of drug resistance. Surprisingly, in all seven samples it was found a high diversity to adapt to antibiotic stress that as many as four to five different resistant mutants were detected in a single sputum, suggesting that drug resistance results from a heterogeneous genotype. However, the study suggests that although multiple resistant mutations coexist in the host, ultimately only a single resistance mutation will be fixed in the population (Sun et al., 2012). In this context of drug-resistance-conferring mutations, a striking example of the selective sweep phenomenon (in which a positively selected locus and thereby the entire chromosome are fixed reducing genetic diversity) were reported in 2014 in an individual who developed extensively drug-resistant tuberculosis (XDR-TB) during treatment. In another investigation, analysis of genomes sequences of nine serial *M. tuberculosis* isolates from the same patient revealed that resistance mutations were acquired multiple times by individual clones. For most drugs, resistance arose through independent emergence of mutations in more than one clone, of which only one clone ultimately prevailed, displacing the other clones in the process (Eldholm et al., 2014).

Another question is why some *M. tuberculosis* strains are preferentially associated with acquisition of resistance to multiple drugs, which poses a dangerous threat against TB control programs. Epidemiological data indicate that isolates from the East-Asia lineage (L2), which include Beijing strains, are associated with increased risk of drug resistance in several countries. A set of experiments demonstrated that L2 strains acquire resistance to drugs in vitro more rapidly than strains from L4, which is the result of the higher basal mutation rate of the L2 lineage. Consequently, it is postulated that L2 strains acquire drug resistance mutations even before the contact with the antibiotic. Extrapolation of this result alongside with estimate of mutation rate in humans suggest that individuals infected with L2 strains are at a notably increased risk of acquiring multiple drug resistance before treatment compared to individuals infected with other strains (Ford et al., 2013). Considering the worldwide expansion of Beijing-L2 strains, this finding has extraordinary implications for diagnosis and treatment of these isolates. Since a higher bacterial burden is associated with a higher number of potential drug resistant bacteria, improvements in diagnostic tools and treatment regimens can help to limit the emergence of drug resistance. A recent genome-wide analysis of 6.465 *M. tuberculosis* clinical isolates not only identified candidate mutations for antibiotic resistance, but suggest the involvement of efflux pumps (DrrA and Rv2688c) in the emergence of multidrug resistance (Coll et al., 2018).

In the context of the evolution of drug resistant strains, it remains to be elucidated why these bacteria successfully survive in the host population even if the drug resistance phenotype is frequently associated with a reduced fitness compared to drug susceptible strains. However, it is observed that some *M. tuberculosis* isolates resistant to the first line drug rifampicin show no fitness reduction compared to susceptible strains. By comparing the genomes of rifampicin resistant clones with their corresponding susceptible isolates recovered from the same patient at an earlier time point, it was identified a set of potential compensatory mutations in RNA polymerase genes (the target of rifampicin) (Comas et al., 2012). Another study used 5310 genome sequences from clinical isolates, worldwide distributed, and revealed drug resistance patterns conserved globally. They have been able to demonstrate that the global epidemic of drug-resistant strains is caused by both de novo and direct transmission from infected patients. They have also been able to identify precursor mutations that often precede to multidrug resistance, enabling the redesign of molecular diagnosis, in order to start targeted treatments of patients with pre-MDR-TB (Manson et al., 2017).

An aspect frequently unnoticed when studying MTBC adaptation to their hosts is referred to transposition of IS*6110* during infection. As described above, this transposon has a role not only in shaping the chromosome of MTBC members but also it functions as a mobile promoter. Although the study of IS*6110* is constricted to molecular epidemiology, very few studies have provided insights into the biological function of IS*6110* transposition in the MTBC and/or the transposition rate within the host. However, upon examination of epidemiologic data, it has been reported that IS*6110* has an unusually high mutation rate (7.9×10^{-5} transposition events per site per generation) compared to the basal nucleotide mutation rate in the MTBC ($\approx 10^{-10}$–10^{-9} events per nucleotide per generation). Transposition rate is even higher than the mutation rate in hypermutator strains carrying a defective DNA repair machinery (10^{-7}–10^{-6} events per nucleotide per generation) (Tanaka, 2004). These results are supported by the finding that *IS6110* preferentially occurs in vivo during an experimental mouse infection with *M. tuberculosis* (Gonzalo-Asensio et al., 2018). These observations strongly suggest that transposition of IS*6110* is at positive selection within patients and highlight the importance of IS*6110* as an epidemiological marker due to its high biological variability. However, provided that IS*6110* produces a replicative transposition resulting in duplication of IS*6110* elements, it has been suggested that accumulation of IS*6110* across the chromosome might have a negative effect either by inactivating essential genes or by mediating deleterious genomic deletions. On this basis, it has been suggested that IS*6110* copy number should be somewhat regulated in order to avoid damage to the bacterial host (Tanaka et al., 2004).

6. Adaptative Cues of the MTBC as the Most Successful Pathogens

Today, *M. tuberculosis* is present in one every four humans and *M. bovis* is present everywhere where cattle is present. These data illustrate the view that bacteria from the MTBC are the most successful pathogens probably as a result of a long and intimate host–pathogen co-evolution. Further, different *M. tuberculosis* and *M. africanum* lineages exhibit characteristic phylogeographical distributions and this is also applied to *M. bovis* subtypes. One might argue that interspecies polymorphisms likely account for their specific host tropism. In favor of this hypothesis, infection of humans with *M. bovis* rarely result in transmission of the bovine bacillus and the converse is also true, *M. tuberculosis* produces less severe pathology in cattle than *M. bovis*. Thus, it is tempting to speculate that

interspecies polymorphisms have shaped the host–pathogen range of the MTBC. One study demonstrated a deleterious effect of a Gly71Ile mutation in PhoR from *M. africanum* L6 and from A4 *M. bovis* (Fig. 26.1A). Considering the key role of the *phoP-phoR* two-component system in *M. tuberculosis* virulence through the control of immunomodulatory lipids and the secretion of ESAT-6, it was observed that *M. africanum* L6 and *M. bovis* have a downregulated PhoP-PhoR regulon and consequently they are unable to produce immunomodulatory lipids. Moreover, an absence of the ESAT-6 virulence factor was observed in the culture supernatants when the mutated allele was transferred into an *M. tuberculosis* genetic background (Gonzalo-Asensio et al., 2014). In addition, this study uncovered the molecular mechanism by which a rare *M. bovis* isolate ("B strain") is able to transmit between humans due to the insertion of an IS*6110* element upstream the *phoP* gene. Since IS*6110* behave as a mobile promoter, this insertion resulted in increased expression of *phoP*. Consequently, the *M. bovis* "B strain" carries a deregulated *phoP* gene that results in the compensatory production of PhoP-dependent phenotypes despite carrying a nonfunctional PhoR (Gonzalo-Asensio et al., 2014). Altogether, these findings highlight the importance of *phoP-phoR* polymorphisms in shaping host–pathogen adaptation and encourage mycobacteriologists to profound in the implications of other interspecies polymorphisms.

MTBC bacteria persist and progress in their respective hosts despite development of immune responses. Upon engulfment of MTBC bacteria, phagocytic cells are able to present in cell-surface receptors epitopes derived from proteolysis of MTBC antigens. The subsequent recognition of these epitopes by T $CD4^+$ and $CD8^+$ lymphocytes results in antigen-specific immune responses. This can lead to the erroneous assumption that similarly to other pathogens, MTBC members are subjected to immune escape and antigenic proteins tend to be hypervariable and subjected to diversifying selection. At a first glance, clonality in the MTBC argues against this hypothesis. After, exploration of 21 representative MTBC genome sequences, it was demonstrated that the ratio of the rates of nonsynonymous and synonymous substitutions (dN/dS) was significantly lower for essential genes than for nonessential genes. This ratio decreases in T-cell antigens in comparison with essential genes. Importantly, dN/dS ratios for T-cells epitopes were even lower than in nonepitope regions in the same T-cell antigen. Overall, MTBC antigens are under purifying selection as strong as, or perhaps even stronger, than that of essential genes (Comas et al., 2010; Yruela et al., 2016). The lack of immune evasion reflects that members of the MTBC have developed different strategies for immune subversion and these pathogens probably benefit from being recognized by the host immune system. Adaptative immunity mediated by recognition of MTBC epitopes results in containment of the infection during long periods. This latent infection with subsequent reactivation to disease is characteristic of human disease and could represent an evolutive strategy to ensure the pathogen transmission to future generations of susceptible hosts.

In this regard, it has been suggested that phylogenetic diversity between MTBC lineages would be reflected in a corresponding diversity when eliciting immune responses. Examination of the macrophage inflammatory response after infection with *M. tuberculosis* or *M. africanum* representative of L1-L6 lineages resulted in lineage-specific profiles. Evolutionary modern lineages (L2, L3, L4) elicited lower inflammatory responses than evolutionary ancient lineages (L1, L5, L6) (Portevin et al., 2011). Since development of disease and the subsequent transmission to new hosts is hypothesized to correlate with reduced immune responses, this result has implications in the host–pathogen evolution of the MTBC. Ancient lineages with higher inflammatory responses are associated with a containment of infection and longer latency period. Conversely, the lower inflammatory phenotype in modern

lineages is associated with an early progressive disease. To understand the biological significance of this finding, it is necessary to critically observe which host population is preferentially targeted by each MTBC lineage. Ancient strains are associated with low-density countries and mainly infect rural populations in Africa. These nomadic and low-density settlements resemble a scenario of hunting, gathering and pastoralism characteristic of Palaeolithic and Neolithic period. By contrast, modern strains are frequently found in crowded countries of Europe, India, China, and America.

It has been proposed an "Ecological Theory" to explain the evolution of the MTBC. This theory predicts that when virulence is positively correlated with transmission, as is the case in TB, access to a larger number of susceptible hosts favors higher virulence and a shorter latency period. Most of the co-evolutionary history of MTBC with humans has occurred when human population densities were low. During this period, infection with ancient strains would have guaranteed a latent infection that after reactivation decades later would enable access to a new birth cohort of susceptible hosts. This way, ancient lineages of the MTBC are not disproportionately virulent in order to avoid decimation of the susceptible population. By contrast, living in crowded countries emerging from the 18th century, would have enabled the accessibility to a large number of susceptible hosts. This lifestyle might have selected more virulent strains without risk to decimate their hosts. We can hypothesize that modern lineages are more evolutionary advantaged than ancient lineages since they are less geographically constrained. Hence, from an ecological point of view, ancient strains might be referred to as "specialists," while modern lineages might represent "generalists" (Gagneux, 2012). Understanding the evolutive strategies of the MTBC may help to predict future trends in the spread of the disease and consequently to anticipate control measures. As an example, according to the Ecological Theory, the increasing number of the African population will probably entail a displacement of ancient strains by strains belonging to modern lineages in the future. Indeed, an epidemiologic study in the Gambia showed that strains from modern lineages were three times more likely to cause active disease than members from ancient lineages (de Jong et al., 2008).

Irrespective of the individual adaptations of different lineages to specific hosts, MTBC members share common virulence mechanisms. It has been extensively documented the role of ESAT-6 in promoting mycobacterial virulence. This protein is not only conserved between members of the MTBC, but also in other related pathogenic species such as *M. leprae* or *M. marinum* causative of human leprosy and fish-to-human zoonosis, respectively. Conversely, the BCG vaccine lacking ESAT-6, or *phoP* mutants unable to secrete ESAT-6, result attenuated. A pioneering study demonstrated that 48 hours after macrophage infection, *M. tuberculosis* and *M. leprae* translocated from the harsh phagolysosome to the gentle cytosol. This translocation is impaired in the BCG vaccine, which led to the demonstration that phagosomal escape requires a proper production and secretion of ESAT-6 (van der Wel et al., 2007). Subsequent studies went deeply into the mechanism of this phagolysosomal rupture mediated by ESAT-6 secretion and linked cytosolic escape with virulence of *M. tuberculosis* and *M. marinum* (Houben et al., 2012; Simeone et al., 2012). Another study demonstrated that in addition to mediate phagolysosomal escape, the RD1 region (containing ESAT-6) plays a role in inducing apoptosis on infected cells. This apoptotic mechanism promotes cell-to-cell spread of virulent mycobacteria and enables colonization of neighboring cells. Accordingly, the BCG vaccine, or new generation vaccines based on *phoP* mutants, unable to produce or secrete ESAT-6, respectively, is inefficient in inducing apoptosis or colonizing new cells (Aguilo et al., 2013). In sum, for MTBC members that does not produce classical toxins, we can consider ESAT-6 as the main virulence determinant.

7. Pending Questions and Concluding Remarks

Knowledge about mycobacterial diversity of the MTBC has led to propose an evolutive branding of MTBC lineages. However, we still do not know in detail which factor(s) contribute to the host preference in the MTBC. This question is hardly accomplished since the natural hosts of the MTBC are humans, cows, goats, or seals among others. This implies a handicap in animal experimentation since laboratory models (mainly mice, guinea pigs, and nonhuman primates) does not reflect this host variability. Further, these models attempt to characterize virulence traits rather than measuring the transmissibility of MTBC strains, a key phenotype of the tubercle bacillus to ensure survival in the host population. It is also likely that yet unknown host factors have also contributed to shape the current phylogeographical distribution of the MTBC. In this respect, it has been documented a correlation between the *NRAMP1* human allele (associated to TB resistance) and the duration of urban settlements (Barnes et al., 2011). Other polymorphisms (*IFNG, NOS2A, MBL, VDR*, and some *TLR*) have been associated with susceptibility to TB in case-control association studies (Moller et al., 2010). A future linkage between MTBC genomes and representative human haplogroups might infer host susceptibility factors. In a near future, these strategies will allow personalized treatments or even reconsider vaccination strategies.

The advent of next-generation sequencing techniques has proven useful to delineate inter- and between-species polymorphisms in the MTBC. However, little works have gone deeply on the biological significance of these variations. Some studies have succeeded in reconstructing the phylogenetic evolution of the MTBC based on the biological implications of polymorphisms in the *phoP-phoR* virulence system (Broset et al., 2015). Similar works are needed to study the functional consequences of the approximately 2.000 polymorphisms existing between MTBC species. We should be aware that inherent difficulties in manipulation of mycobacteria and the slow growth of MTBC strains frequently hamper these genetic approaches.

The debate to estimate the MRCA of the MTBC is a consequence of the weakness of experimental evidence. These estimates frequently assume a constant mutation rate over time and over different lineages. This observation was partly validated using a macaque model showing similar mutation rates during in vitro culture and latent and active TB. However, we cannot rule our differences in mutation rates between different lineages of the MTBC. Indeed, it was latter demonstrated that L2 strains exhibit higher mutation rates than L4 isolates in vitro. It is at present unknown whether this finding is extrapolated to the TB infection. Further, as observed with IS*6110* transposition, it might be also possible that some mutations are under positive selection during TB infection and disease. Future interrogation of these questions together with the application of next-generation sequencing to ancient DNA will surely help to reconstruct the origin of the MTBC.

TB, unlike other infectious diseases, is benefited from efficacious immune system recognition. The hyperconservation of T cell epitopes likely plays a key role in development of a latent TB infection. This latency and reactivation to active disease of the MTBC resembles the lambda phage cycle by promoting prophage transmission for many bacterial generations until a stress response triggers production of virulent phage particles. However, the environmental signal(s) that mediate TB reactivation are poorly understood. Several studies have deciphered the phagolysosomal signals encountered by *M. tuberculosis* by examining transcriptional profiles during macrophage infection. These studies have inferred that phagolysosomal environment is nitrosative, oxidative, functionally hypoxic,

carbohydrate poor, and capable of perturbing the pathogen's cell envelope (Schnappinger et al., 2003). Other approaches have addressed the time-dependent transcriptional profiles during intracellular infection (Rohde et al., 2012) and studies reported in 2013 have provided unprecedented level of detail about the transcription networks of *M. tuberculosis* (Galagan et al., 2013). In the future, combination of genomic, transcriptomic, proteomic and metabolomic data from animal infection models will hopefully delineate the host−pathogen interplay during both the latency period and reactivation.

List of Authors or Contributors

J. G.-A. wrote the original draft. J. C.-S. and E. C.-P. updated the previous version of the chapter and contributed equally to the writing of the manuscript. All authors reviewed and read critically the final version of the manuscript.

J. C.-S. is a biotechnology graduate by the University of Zaragoza, with 6-year experience in genetic manipulation and molecular microbiology of mycobacteria. His main interest is understanding virulence regulation in bacteria causative of tuberculosis.

E. C.-P. is a biotechnology graduate by the University of Zaragoza, with a 5-year experience in deciphering the host−pathogen adaptation in bacteria causing Tuberculosis in humans and animals. Her main interest is understanding host factors driving disease susceptibility.

J. G.-A. is a PhD in molecular and cellular biology and Associate Professor at the University of Zaragoza. He has >30 years' experience in mycobacterial genetics and host−pathogen adaptation in the *Mycobacterium* genus. His main interest is applying basic research to develop intervention strategies to fight tuberculosis. Specifically, J. G.-A. has contributed to the development and molecular characterization of the MTBVAC live attenuated vaccine, and he currently explores antivirulence strategies against *M. tuberculosis*.

References

Abdallah, A.M., Hill-Cawthorne, G.A., Otto, T.D., Coll, F., Guerra-Assuncao, J.A., Gao, G., Naeem, R., Ansari, H., Malas, T.B., Adroub, S.A., Verboom, T., Ummels, R., Zhang, H., Panigrahi, A.K., McNerney, R., Brosch, R., Clark, T.G., Behr, M.A., Bitter, W., Pain, A., 2015. Genomic expression catalogue of a global collection of BCG vaccine strains show evidence for highly diverged metabolic and cell-wall adaptations. Sci. Rep. 5, 15443.

Aguilo, J.I., Alonso, H., Uranga, S., Marinova, D., Arbues, A., de Martino, A., Anel, A., Monzon, M., Badiola, J., Pardo, J., Brosch, R., Martin, C., 2013. ESX-1-induced apoptosis is involved in cell-to-cell spread of Mycobacterium tuberculosis. Cell Microbiol. 15, 1994−2005.

Allen, A.C., Malaga, W., Gaudin, C., Volle, A., Moreau, F., Hassan, A., Astarie-Dequeker, C., Peixoto, A., Antoine, R., Pawlik, A., Frigui, W., Berrone, C., Brosch, R., Supply, P., Guilhot, C., 2021. Parallel in vivo experimental evolution reveals that increased stress resistance was key for the emergence of persistent tuberculosis bacilli. Nat. Microbiol. 6, 1082−1093.

Alonso, H., Aguilo, J.I., Samper, S., Caminero, J.A., Campos-Herrero, M.I., Gicquel, B., Brosch, R., Martin, C., Otal, I., 2011. Deciphering the role of IS6110 in a highly transmissible Mycobacterium tuberculosis Beijing strain, GC1237. Tuberculosis 91, 117−126.

Aranaz, A., Cousins, D., Mateos, A., Dominguez, L., 2003. Elevation of Mycobacterium tuberculosis subsp. caprae Aranaz et al. 1999 to species rank as Mycobacterium caprae comb. nov., sp. nov. Int. J. Syst. Evol. Microbiol. 53, 1785−1789.

Ates, L.S., Brosch, R., 2017. Discovery of the type VII ESX-1 secretion needle? Mol. Microbiol. 103, 7−12.

Ates, L.S., Dippenaar, A., Ummels, R., Piersma, S.R., van der Woude, A.D., van der Kuij, K., Le Chevalier, F., Mata-Espinosa, D., Barrios-Payan, J., Marquina-Castillo, B., Guapillo, C., Jimenez, C.R., Pain, A., Houben, E.N.G., Warren, R.M., Brosch, R., Hernandez-Pando, R., Bitter, W., 2018. Mutations in ppe38 block PE_PGRS secretion and increase virulence of Mycobacterium tuberculosis. Nat. Microbiol. 3, 181−188.

Barnes, I., Duda, A., Pybus, O.G., Thomas, M.G., 2011. Ancient urbanization predicts genetic resistance to tuberculosis. Evolution 65, 842−848.

Becq, J., Gutierrez, M.C., Rosas-Magallanes, V., Rauzier, J., Gicquel, B., Neyrolles, O., Deschavanne, P., 2007. Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. Mol. Biol. Evol. 24, 1861−1871.

Behr, M.A., 2002. BCG–different strains, different vaccines? Lancet Infect. Dis. 2, 86−92.

Blouin, Y., Cazajous, G., Dehan, C., Soler, C., Vong, R., Hassan, M.O., Hauck, Y., Boulais, C., Andriamanantena, D., Martinaud, C., Martin, E., Pourcel, C., Vergnaud, G., 2014. Progenitor "Mycobacterium canettii" clone responsible for lymph node tuberculosis epidemic, Djibouti. Emerg. Infect. Dis. 20, 21−28.

Boritsch, E.C., Frigui, W., Cascioferro, A., Malaga, W., Etienne, G., Laval, F., Pawlik, A., Le Chevalier, F., Orgeur, M., Ma, L., Bouchier, C., Stinear, T.P., Supply, P., Majlessi, L., Daffe, M., Guilhot, C., Brosch, R., 2016a. pks5-recombination-mediated surface remodelling in Mycobacterium tuberculosis emergence. Nat. Microbiol. 1.

Boritsch, E.C., Khanna, V., Pawlik, A., Honore, N., Navas, V.H., Ma, L., Bouchier, C., Seemann, T., Supply, P., Stinear, T.P., Brosch, R., 2016b. Key experimental evidence of chromosomal DNA transfer among selected tuberculosis-causing mycobacteria. Proc. Natl. Acad. Sci. U. S. A. 113, 9876−9881.

Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., Campbell, T.J., Majander, K., Wilbur, A.K., Guichon, R.A., Wolfe Steadman, D.L., Cook, D.C., Niemann, S., Behr, M.A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J.E., Gagneux, S., Stone, A.C., Krause, J., 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of new world human tuberculosis. Nature 514, 494−497.

Bottai, D., Frigui, W., Sayes, F., Di Luca, M., Spadoni, D., Pawlik, A., Zoppo, M., Orgeur, M., Khanna, V., Hardy, D., Mangenot, S., Barbe, V., Medigue, C., Ma, L., Bouchier, C., Tavanti, A., Larrouy-Maumus, G., Brosch, R., 2020. TbD1 deletion as a driver of the evolutionary success of modern epidemic Mycobacterium tuberculosis lineages. Nat. Commun. 11, 684.

Brites, D., Loiseau, C., Menardo, F., Borrell, S., Boniotti, M.B., Warren, R., Dippenaar, A., Parsons, S.D.C., Beisel, C., Behr, M.A., Fyfe, J.A., Coscolla, M., Gagneux, S., 2018. A new phylogenetic framework for the animal-adapted Mycobacterium tuberculosis complex. Front. Microbiol. 9, 2820.

Brosch, R., Gordon, S.V., Billault, A., Garnier, T., Eiglmeier, K., Soravito, C., Barrell, B.G., Cole, S.T., 1998. Use of a Mycobacterium tuberculosis H37Rv bacterial artificial chromosome library for genome mapping, sequencing, and comparative genomics. Infect. Immun. 66, 2221−2229.

Brosch, R., Gordon, S.V., Garnier, T., Eiglmeier, K., Frigui, W., Valenti, P., Dos Santos, S., Duthoy, S., Lacroix, C., Garcia-Pelayo, C., Inwald, J.K., Golby, P., Garcia, J.N., Hewinson, R.G., Behr, M.A., Quail, M.A., Churcher, C., Barrell, B.G., Parkhill, J., Cole, S.T., 2007. Genome plasticity of BCG and impact on vaccine efficacy. Proc. Natl. Acad. Sci. U. S. A. 104, 5596−5601.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., Cole, S.T., 2002. A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc. Natl. Acad. Sci. U. S. A. 99, 3684−3689.

Broset, E., Martin, C., Gonzalo-Asensio, J., 2015. Evolutionary landscape of the Mycobacterium tuberculosis complex from the viewpoint of PhoPR: implications for virulence regulation and application to vaccine development. mBio 6.

Calmette, A., 1931. Preventive vaccination against tuberculosis with BCG. Proc. Roy. Soc. Med. 24, 1481−1490.

Campos-Pardos, E., Uranga, S., Picó, A., Gómez, A.B., Gonzalo-Asensio, J., 2024. Dependency on host vitamin B12 has shaped Mycobacterium tuberculosis complex evolution. Nat. Commun. 15 (2161). https://doi.org/10.1038/s41467-024-46449-8.

Chesne-Seck, M.L., Barilone, N., Boudou, F., Asensio, J.G., Kolattukudy, P.E., Martin, C., Cole, S.T., Gicquel, B., Gopaul, D.N., Jackson, M., 2007. A point mutation in the two-component regulator PhoP-PhoR accounts for the absence of polyketide-derived acyltrehaloses but not that of phthiocerol dimycocerosates in Mycobacterium tuberculosis H37Ra. J. Bacteriol. 190, 1329−1334.

Chisholm, R.H., Trauer, J.M., Curnoe, D., Tanaka, M.M., 2016. Controlled fire use in early humans might have triggered the evolutionary emergence of tuberculosis. Proc. Natl. Acad. Sci. U. S. A. 113, 9051−9056.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry 3rd, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G., 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393, 537−544.

Coll, F., Phelan, J., Hill-Cawthorne, G.A., Nair, M.B., Mallard, K., Ali, S., Abdallah, A.M., Alghamdi, S., Alsomali, M., Ahmed, A.O., Portelli, S., Oppong, Y., Alves, A., Bessa, T.B., Campino, S., Caws, M., Chatterjee, A., Crampin, A.C., Dheda, K., Furnham, N., Glynn, J.R., Grandjean, L., Minh Ha, D., Hasan, R., Hasan, Z., Hibberd, M.L., Joloba, M., Jones-Lopez, E.C., Matsumoto, T., Miranda, A., Moore, D.J., Mocillo, N., Panaiotov, S., Parkhill, J., Penha, C., Perdigao, J., Portugal, I., Rchiad, Z., Robledo, J., Sheen, P., Shesha, N.T., Sirgel, F.A., Sola, C., Oliveira Sousa, E., Streicher, E.M., Helden, P.V., Viveiros, M., Warren, R.M., McNerney, R., Pain, A., Clark, T.G., 2018. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat. Genet. 50, 307−316.

Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., Galagan, J., Niemann, S., Gagneux, S., 2012. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat. Genet. 44, 106−110.

Comas, I., Chakravartti, J., Small, P.M., Galagan, J., Niemann, S., Kremer, K., Ernst, J.D., Gagneux, S., 2010. Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat. Genet. 42, 498−503.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S., 2013. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat. Genet. 45, 1176−1182.

Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., Otchere, I.D., Asante-Poku, A., Asare, P., Sanchez-Buso, L., Gehre, F., Sanoussi, C.N., Antonio, M., Affolabi, D., Fyfe, J., Beckert, P., Niemann, S., Alabi, A.S., Grobusch, M.P., Kobbe, R., Parkhill, J., Beisel, C., Fenner, L., Bottger, E.C., Meehan, C.J., Harris, S.R., de Jong, B.C., Yeboah-Manu, D., Brites, D., 2021. Phylogenomics of Mycobacterium africanum reveals a new lineage and a complex evolutionary history. Microb. Genom. 7.

Cousins, D.V., Bastida, R., Cataldi, A., Quse, V., Redrobe, S., Dow, S., Duignan, P., Murray, A., Dupont, C., Ahmed, N., Collins, D.M., Butler, W.R., Dawson, D., Rodriguez, D., Loureiro, J., Romano, M.I., Alito, A., Zumarraga, M., Bernardelli, A., 2003. Tuberculosis in seals caused by a novel member of the Mycobacterium tuberculosis complex: Mycobacterium pinnipedii sp. nov. Int. J. Syst. Evol. Microbiol. 53, 1305−1314.

de Jong, B.C., Antonio, M., Gagneux, S., 2010. Mycobacterium africanum–review of an important cause of human tuberculosis in West Africa. PLoS Neglected Trop. Dis. 4, e744.

de Jong, B.C., Hill, P.C., Aiken, A., Awine, T., Antonio, M., Adetifa, I.M., Jackson-Sillah, D.J., Fox, A., Deriemer, K., Gagneux, S., Borgdorff, M.W., McAdam, K.P., Corrah, T., Small, P.M., Adegbola, R.A., 2008. Progression to active tuberculosis, but not transmission, varies by Mycobacterium tuberculosis lineage in the Gambia. J. Infect. Dis. 198, 1037−1043.

de la Fuente, J., Diez-Delgado, I., Contreras, M., Vicente, J., Cabezas-Cruz, A., Tobes, R., Manrique, M., Lopez, V., Romero, B., Bezos, J., Dominguez, L., Sevilla, I.A., Garrido, J.M., Juste, R., Madico, G., Jones-Lopez, E., Gortazar, C., 2015. Comparative genomics of field isolates of Mycobacterium bovis and M. Caprae provides evidence for possible correlates with bacterial viability and virulence. PLoS Neglected Trop. Dis. 9, e0004232.

Dippenaar, A., Parsons, S.D.C., Sampson, S.L., van der Merwe, R.G., Drewe, J.A., Abdallah, A.M., Siame, K.K., Gey van Pittius, N.C., van Helden, P.D., Pain, A., Warren, R.M., 2015. Whole genome sequence analysis of Mycobacterium suricattae. Tuberculosis 95, 682−688.

Domenech, P., Reed, M.B., 2009. Rapid and spontaneous loss of phthiocerol dimycocerosate (PDIM) from Mycobacterium tuberculosis grown in vitro: implications for virulence studies. Microbiology 155, 3532−3543.

Domenech, P., Rog, A., Moolji, J.U., Radomski, N., Fallow, A., Leon-Solis, L., Bowes, J., Behr, M.A., Reed, M.B., 2014. Origins of a 350-kilobase genomic duplication in Mycobacterium tuberculosis and its impact on virulence. Infect. Immun. 82, 2902−2912.

Eldholm, V., Norheim, G., von der Lippe, B., Kinander, W., Dahle, U.R., Caugant, D.A., Mannsaker, T., Mengshoel, A.T., Dyrhol-Riise, A.M., Balloux, F., 2014. Evolution of extensively drug-resistant Mycobacterium tuberculosis from a susceptible ancestor in a single patient. Genome Biol. 15, 490.

Fabre, M., Hauck, Y., Soler, C., Koeck, J.L., van Ingen, J., van Soolingen, D., Vergnaud, G., Pourcel, C., 2010. Molecular characteristics of "Mycobacterium canettii" the smooth Mycobacterium tuberculosis bacilli. Infect. Genet. Evol. 10, 1165−1173.

Ford, C.B., Lin, P.L., Chase, M.R., Shah, R.R., Iartchouk, O., Galagan, J., Mohaideen, N., Ioerger, T.R., Sacchettini, J.C., Lipsitch, M., Flynn, J.L., Fortune, S.M., 2011. Use of whole genome sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection. Nat. Genet. 43, 482−486.

Ford, C.B., Shah, R.R., Maeda, M.K., Gagneux, S., Murray, M.B., Cohen, T., Johnston, J.C., Gardy, J., Lipsitch, M., Fortune, S.M., 2013. Mycobacterium tuberculosis mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat. Genet. 45, 784−790.

Frigui, W., Bottai, D., Majlessi, L., Monot, M., Josselin, E., Brodin, P., Garnier, T., Gicquel, B., Martin, C., Leclerc, C., Cole, S.T., Brosch, R., 2008. Control of M. tuberculosis ESAT-6 secretion and specific T cell recognition by PhoP. PLoS Pathog. 4, e33.

Gagneux, S., 2012. Host-pathogen coevolution in human tuberculosis. Philos. Trans. R. Soc. Lond. B Biol. Sci. 367, 850−859.

Galagan, J.E., 2014. Genomic insights into tuberculosis. Nat. Rev. Genet. 15, 307−320.

Galagan, J.E., Minch, K., Peterson, M., Lyubetskaya, A., Azizi, E., Sweet, L., Gomes, A., Rustad, T., Dolganov, G., Glotova, I., Abeel, T., Mahwinney, C., Kennedy, A.D., Allard, R., Brabant, W., Krueger, A., Jaini, S., Honda, B., Yu, W.H., Hickey, M.J., Zucker, J., Garay, C., Weiner, B., Sisk, P., Stolte, C., Winkler, J.K., Van de Peer, Y., Iazzetti, P., Camacho, D., Dreyfuss, J., Liu, Y., Dorhoi, A., Mollenkopf, H.J., Drogaris, P., Lamontagne, J., Zhou, Y., Piquenot, J., Park, S.T., Raman, S., Kaufmann, S.H., Mohney, R.P., Chelsky, D., Moody, D.B., Sherman, D.R., Schoolnik, G.K., 2013. The Mycobacterium tuberculosis regulatory network and hypoxia. Nature 499, 178−183.

Gonzalo-Asensio, J., Malaga, W., Pawlik, A., Astarie-Dequeker, C., Passemar, C., Moreau, F., Laval, F., Daffe, M., Martin, C., Brosch, R., Guilhot, C., 2014. Evolutionary history of tuberculosis shaped by conserved mutations in the PhoPR virulence regulator. Proc. Natl. Acad. Sci. U. S. A. 111, 11491−11496.

Gonzalo-Asensio, J., Perez, I., Aguilo, N., Uranga, S., Pico, A., Lampreave, C., Cebollada, A., Otal, I., Samper, S., Martin, C., 2018. New insights into the transposition mechanisms of IS6110 and its dynamic distribution between Mycobacterium tuberculosis complex lineages. PLoS Genet. 14, e1007282.

Gordon, S.V., Brosch, R., Billault, A., Garnier, T., Eiglmeier, K., Cole, S.T., 1999. Identification of variable regions in the genomes of tubercle bacilli using bacterial artificial chromosome arrays. Mol. Microbiol. 32, 643−655.

Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., Supply, P., Vincent, V., 2005. Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS Pathog. 1, e5.

Houben, D., Demangel, C., van Ingen, J., Perez, J., Baldeon, L., Abdallah, A.M., Caleechurn, L., Bottai, D., van Zon, M., de Punder, K., van der Laan, T., Kant, A., Bossers-de Vries, R., Willemsen, P., Bitter, W., van Soolingen, D., Brosch, R., van der Wel, N., Peters, P.J., 2012. ESX-1-mediated translocation to the cytosol controls virulence of mycobacteria. Cell Microbiol. 14, 1287−1298.

Jacobs Jr., W.R., Tuckman, M., Bloom, B.R., 1987. Introduction of foreign DNA into mycobacteria using a shuttle phasmid. Nature 327, 532−535.

Jang, J., Becq, J., Gicquel, B., Deschavanne, P., Neyrolles, O., 2008. Horizontally acquired genomic islands in the tubercle bacilli. Trends Microbiol. 16, 303−308.

Kappelman, J., Alcicek, M.C., Kazanci, N., Schultz, M., Ozkul, M., Sen, S., 2008. First homo erectus from Turkey and implications for migrations into temperate Eurasia. Am. J. Phys. Anthropol. 135, 110−116.

Koeck, J.L., Fabre, M., Simon, F., Daffe, M., Garnotel, E., Matan, A.B., Gerome, P., Bernatas, J.J., Buisson, Y., Pourcel, C., 2011. Clinical characteristics of the smooth tubercle bacilli 'Mycobacterium canettii' infection suggest the existence of an environmental reservoir. Clin. Microbiol. Infect. 17, 1013−1019.

Lee, J.S., Krause, R., Schreiber, J., Mollenkopf, H.J., Kowall, J., Stein, R., Jeon, B.Y., Kwak, J.Y., Song, M.K., Patron, J.P., Jorg, S., Roh, K., Cho, S.N., Kaufmann, S.H., 2008. Mutation in the transcriptional regulator PhoP contributes to avirulence of Mycobacterium tuberculosis H37Ra strain. Cell Host Microbe. 3, 97−103.

Lieberman, T.D., Wilson, D., Misra, R., Xiong, L.L., Moodley, P., Cohen, T., Kishony, R., 2016. Genomic diversity in autopsy samples reveals within-host dissemination of HIV-associated Mycobacterium tuberculosis. Nat. Med. 22, 1470−1474.

Madacki, J., Orgeur, M., Mas Fiol, G., Frigui, W., Ma, L., Brosch, R., 2021. ESX-1-independent horizontal gene transfer by Mycobacterium tuberculosis complex strains. mBio 12.

Mahairas, G.G., Sabo, P.J., Hickey, M.J., Singh, D.C., Stover, C.K., 1996. Molecular analysis of genetic differences between Mycobacterium bovis BCG and virulent M. bovis. J. Bacteriol. 178, 1274−1282.

Manson, A.L., Cohen, K.A., Abeel, T., Desjardins, C.A., Armstrong, D.T., Barry 3rd, C.E., Brand, J., Consortium, T.B.G.G., Chapman, S.B., Cho, S.N., Gabrielian, A., Gomez, J., Jodals, A.M., Joloba, M., Jureen, P., Lee, J.S., Malinga, L., Maiga, M., Nordenberg, D., Noroc, E., Romancenco, E., Salazar, A., Ssengooba, W., Velayati, A.A., Winglee, K., Zalutskaya, A., Via, L.E., Cassell, G.H., Dorman, S.E., Ellner, J., Farnia, P., Galagan, J.E., Rosenthal, A., Crudu, V., Homorodean, D., Hsueh, P.R., Narayanan, S., Pym, A.S., Skrahina, A., Swaminathan, S., Van der Walt, M., Alland, D., Bishai, W.R., Cohen, T., Hoffner, S., Birren, B.W., Earl, A.M., 2017. Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into the emergence and spread of multidrug resistance. Nat. Genet. 49, 395−402.

Martin, C., Timm, J., Rauzier, J., Gomez-Lus, R., Davies, J., Gicquel, B., 1990. Transposition of an antibiotic resistance element in mycobacteria. Nature 345, 739−743.

Menardo, F., Duchene, S., Brites, D., Gagneux, S., 2019. The molecular clock of Mycobacterium tuberculosis. PLoS Pathog. 15, e1008067.

Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., Blum, M.G., Rusch-Gerdes, S., Mokrousov, I., Aleksic, E., Allix-Beguec, C., Antierens, A., Augustynowicz-Kopec, E., Ballif, M., Barletta, F., Beck, H.P., Barry 3rd, C.E., Bonnet, M., Borroni, E., Campos-Herrero, I., Cirillo, D., Cox, H., Crowe, S., Crudu, V., Diel, R., Drobniewski, F., Fauville-Dufaux, M., Gagneux, S., Ghebremichael, S., Hanekom, M.,

Hoffner, S., Jiao, W.W., Kalon, S., Kohl, T.A., Kontsevaya, I., Lillebaek, T., Maeda, S., Nikolayevskyy, V., Rasmussen, M., Rastogi, N., Samper, S., Sanchez-Padilla, E., Savic, B., Shamputa, I.C., Shen, A., Sng, L.H., Stakenas, P., Toit, K., Varaine, F., Vukovic, D., Wahl, C., Warren, R., Supply, P., Niemann, S., Wirth, T., 2015. Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nat. Genet. 47, 242−249.

Moller, M., de Wit, E., Hoal, E.G., 2010. Past, present and future directions in human genetic susceptibility to tuberculosis. FEMS Immunol. Med. Microbiol. 58, 3−26.

Napier, G., Campino, S., Merid, Y., Abebe, M., Woldeamanuel, Y., Aseffa, A., Hibberd, M.L., Phelan, J., Clark, T.G., 2020. Robust barcoding and identification of Mycobacterium tuberculosis lineages for epidemiological and clinical studies. Genome Med. 12, 114.

Ngabonziza, J.C.S., Loiseau, C., Marceau, M., Jouet, A., Menardo, F., Tzfadia, O., Antoine, R., Niyigena, E.B., Mulders, W., Fissette, K., Diels, M., Gaudin, C., Duthoy, S., Ssengooba, W., Andre, E., Kaswa, M.K., Habimana, Y.M., Brites, D., Affolabi, D., Mazarati, J.B., de Jong, B.C., Rigouts, L., Gagneux, S., Meehan, C.J., Supply, P., 2020. A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. Nat. Commun. 11, 2917.

O'Neill, M.B., Mortimer, T.D., Pepperell, C.S., 2015. Diversity of Mycobacterium tuberculosis across evolutionary scales. PLoS Pathog. 11, e1005257.

Orgeur, M., Brosch, R., 2018. Evolution of virulence in the Mycobacterium tuberculosis complex. Curr. Opin. Microbiol. 41, 68−75.

Otal, I., Martin, C., Vincent-Levy-Frebault, V., Thierry, D., Gicquel, B., 1991. Restriction fragment length polymorphism analysis using IS6110 as an epidemiological marker in tuberculosis. J. Clin. Microbiol. 29, 1252−1254.

Pascopella, L., Collins, F.M., Martin, J.M., Lee, M.H., Hatfull, G.F., Stover, C.K., Bloom, B.R., Jacobs Jr., W.R., 1994. Use of in vivo complementation in Mycobacterium tuberculosis to identify a genomic fragment associated with virulence. Infect. Immun. 62, 1313−1319.

Philipp, W.J., Nair, S., Guglielmi, G., Lagranderie, M., Gicquel, B., Cole, S.T., 1996. Physical mapping of Mycobacterium bovis BCG pasteur reveals differences from the genome map of Mycobacterium tuberculosis H37Rv and from M. bovis. Microbiology 142 (Pt 11), 3135−3145.

Portevin, D., Gagneux, S., Comas, I., Young, D., 2011. Human macrophage responses to clinical isolates from the Mycobacterium tuberculosis complex discriminate between ancient and modern lineages. PLoS Pathog. 7, e1001307.

Pym, A.S., Brodin, P., Brosch, R., Huerre, M., Cole, S.T., 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines Mycobacterium bovis BCG and Mycobacterium microti. Mol. Microbiol. 46, 709−717.

Roberts, C.A., Pfister, L.A., Mays, S., 2009. Letter to the editor: was tuberculosis present in homo erectus in Turkey? Am. J. Phys. Anthropol. 139, 442−444.

Rohde, K.H., Veiga, D.F., Caldwell, S., Balazsi, G., Russell, D.G., 2012. Linking the transcriptional profiles and the physiological states of Mycobacterium tuberculosis during an extended intracellular infection. PLoS Pathog. 8, e1002769.

Sampson, S.L., Warren, R.M., Richardson, M., Victor, T.C., Jordaan, A.M., van der Spuy, G.D., van Helden, P.D., 2003. IS6110-mediated deletion polymorphism in the direct repeat region of clinical isolates of Mycobacterium tuberculosis. J. Bacteriol. 185, 2856−2866.

Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C., Schoolnik, G.K., 2003. Transcriptional adaptation of Mycobacterium tuberculosis within macrophages: insights into the phagosomal environment. J. Exp. Med. 198, 693−704.

Simeone, R., Bobard, A., Lippmann, J., Bitter, W., Majlessi, L., Brosch, R., Enninga, J., 2012. Phagosomal rupture by Mycobacterium tuberculosis results in toxicity and host cell death. PLoS Pathog. 8, e1002507.

Smith, N.H., 2012. The global distribution and phylogeography of Mycobacterium bovis clonal complexes. Infect. Genet. Evol. 12, 857−865.

Smith, N.H., Gordon, S.V., de la Rua-Domenech, R., Clifton-Hadley, R.S., Hewinson, R.G., 2006. Bottlenecks and broomsticks: the molecular evolution of Mycobacterium bovis. Nat. Rev. Microbiol. 4, 670−681.

Smith, N.H., Hewinson, R.G., Kremer, K., Brosch, R., Gordon, S.V., 2009. Myths and misconceptions: the origin and evolution of Mycobacterium tuberculosis. Nat. Rev. Microbiol. 7, 537−544.

Solans, L., Aguilo, N., Samper, S., Pawlik, A., Frigui, W., Martin, C., Brosch, R., Gonzalo-Asensio, J., 2014. A specific polymorphism in Mycobacterium tuberculosis H37Rv causes differential ESAT-6 expression and identifies WhiB6 as a novel ESX-1 component. Infect. Immun. 82, 3446−3456.

Steenken, W., Oatway, W.H., Petroff, S.A., 1934. Biological studies of the tubercle Bacillus: iii. Dissociation and pathogenicity of the R and S variants of the human tubercle Bacillus (H(37)). J. Exp. Med. 60, 515−540.

Sun, G., Luo, T., Yang, C., Dong, X., Li, J., Zhu, Y., Zheng, H., Tian, W., Wang, S., Barry 3rd, C.E., Mei, J., Gao, Q., 2012. Dynamic population changes in Mycobacterium tuberculosis during acquisition and fixation of drug resistance in patients. J. Infect. Dis. 206, 1724−1733.

Supply, P., Marceau, M., Mangenot, S., Roche, D., Rouanet, C., Khanna, V., Majlessi, L., Criscuolo, A., Tap, J., Pawlik, A., Fiette, L., Orgeur, M., Fabre, M., Parmentier, C., Frigui, W., Simeone, R., Boritsch, E.C., Debrie, A.S., Willery, E., Walker, D., Quail, M.A., Ma, L., Bouchier, C., Salvignol, G., Sayes, F., Cascioferro, A., Seemann, T., Barbe, V., Locht, C., Gutierrez, M.C., Leclerc, C., Bentley, S.D., Stinear, T.P., Brisse, S., Medigue, C., Parkhill, J., Cruveiller, S., Brosch, R., 2013. Genomic analysis of smooth tubercle bacilli provides insights into ancestry and pathoadaptation of Mycobacterium tuberculosis. Nat. Genet. 45, 172−179.

Tanaka, M.M., 2004. Evidence for positive selection on Mycobacterium tuberculosis within patients. BMC Evol. Biol. 4, 31.

Tanaka, M.M., Rosenberg, N.A., Small, P.M., 2004. The control of copy number of IS6110 in Mycobacterium tuberculosis. Mol. Biol. Evol. 21, 2195−2201.

Trauner, A., Liu, Q., Via, L.E., Liu, X., Ruan, X., Liang, L., Shi, H., Chen, Y., Wang, Z., Liang, R., Zhang, W., Wei, W., Gao, J., Sun, G., Brites, D., England, K., Zhang, G., Gagneux, S., Barry 3rd, C.E., Gao, Q., 2017. The within-host population dynamics of Mycobacterium tuberculosis vary with treatment efficacy. Genome Biol. 18, 71.

Vagene, A.J., Honap, T.P., Harkins, K.M., Rosenberg, M.S., Giffin, K., Cardenas-Arroyo, F., Leguizamon, L.P., Arnett, J., Buikstra, J.E., Herbig, A., Krause, J., Stone, A.C., Bos, K.I., 2022. Geographically dispersed zoonotic tuberculosis in pre-contact South American human populations. Nat. Commun. 13, 1195.

van der Wel, N., Hava, D., Houben, D., Fluitsma, D., van Zon, M., Pierson, J., Brenner, M., Peters, P.J., 2007. M. tuberculosis and M. leprae translocate from the phagolysosome to the cytosol in myeloid cells. Cell 129, 1287−1298.

van Embden, J.D., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B., Hermans, P., Martin, C., McAdam, R., Shinnick, T.M., et al., 1993. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. 31, 406−409.

van Soolingen, D., van der Zanden, A.G., de Haas, P.E., Noordhoek, G.T., Kiers, A., Foudraine, N.A., Portaels, F., Kolk, A.H., Kremer, K., van Embden, J.D., 1998. Diagnosis of Mycobacterium microti infections among humans by using novel genetic markers. J. Clin. Microbiol. 36, 1840−1845.

Weiner, B., Gomez, J., Victor, T.C., Warren, R.M., Sloutsky, A., Plikaytis, B.B., Posey, J.E., van Helden, P.D., Gey van Pittius, N.C., Koehrsen, M., Sisk, P., Stolte, C., White, J., Gagneux, S., Birren, B., Hung, D., Murray, M., Galagan, J., 2012. Independent large scale duplications in multiple M. tuberculosis lineages overlapping the same genomic region. PLoS One 7, e26038.

Yruela, I., Contreras-Moreira, B., Magalhaes, C., Osorio, N.S., Gonzalo-Asensio, J., 2016. Mycobacterium tuberculosis complex exhibits lineage-specific variations affecting protein ductility and epitope recognition. Genome Biol Evol 8, 3751−3764.

CHAPTER

Evolution, Phylogenetics, and Phylogeography of *Mycobacterium tuberculosis* complex

27

Christophe Sola[1], Igor Mokrousov[2], Muhammed Rabiu Sahal[1], Kevin La[3], Gaetan Senelle[4], Christophe Guyeux[4], Guislaine Refrégier[5] and Emmanuelle Cambau[6]

[1]*Université Paris-Saclay and IAME, INSERM-Université Paris-Cité, Paris, France;* [2]*Laboratory of Molecular Epidemiology and Evolutionary Genetics, St. Petersburg Pasteur Institute, St. Petersburg, Russia;* [3]*INSERM-Université Paris-Cité, Paris, France;* [4]*FEMTO-ST Institute, UMR 6174 CNRS, DISC Computer Science Department, University Franche-Comté (UFC), Besançon, France;* [5]*Université Paris-Saclay, CNRS, AgroParisTech, Ecologie Systématique et Evolution, Gif-Sur-Yvette, France;* [6]*Associate laboratory of the National Reference Center for Mycobacteria and Antimycobacterial Resistance, AP-HP GHU Nord, Mycobacteriology, Inserm, IAME UMR 1137, University of Paris Cité, Paris, France*

« **Et certes, il faut des objets reliés pour que les liaisons se montrent** »

And certainly, there must be connected objects for the connections to show themselves.
Antoine de Saint-Exupéry, Citadelle.

1. Introduction to *Mycobacterium tuberculosis* Complex Genetic Diversity Studies

1.1 Tuberculosis as an Animal or Human Disease: Historical, Epidemiological, Ecological, and Archeological Considerations

1.1.1 The Present Epidemiological Situation Starting From 1990

In 1997, the sequencing of the *Mycobacterium tuberculosis* complex (MTBC) bacilli by S. Cole and his team changed the history of tuberculosis that had suddenly been rocketed into modernity, as done for the rabies virus 10 years before (Tordo et al., 1988; Cole et al., 1998). In 1993, tuberculosis (TB) had been declared a global health emergency by the WHO (World Health Organization) (Anonymous, 1993). It was one of the multiple consequences of (1) the ruling period of R. Reagan that had cut all public expenses except from the ones linked to weapon industry, (2) the almost simultaneous economic and political collapse of the USSR, (3) the disinterest of Europe that arrogantly believed it had successfully managed this disease. The DOTS Implementation (Directly Observed Treatment of Short Course) in 1992−93, allowed progress in TB treatment even if in the countries with already high burden of MDR-TB (first of all, those of the Former Soviet Union), this strategy had to be revised to the more implementable and realistic DOTS-Plus in 1999. In 2020−21, given the COVID-19 pandemia,

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00025-2
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

the objectives concerning eradication of TB have been postponed, and we are back to the year 2012 in terms of objectives to eradicate the disease (Glaziou, 2020).

According to the yearly WHO-TB global report, we should consider three main types of countries, whether high TB prevalent (30 countries in 2019), high combined multi-drug resistant and TB prevalent (MDR-TB, 30 countries), or high simultaneous Human Immuno-Deficiency virus-TB co-infection (HIV-TB, 30 countries), all totaling 48 different countries where TB was still circulating at alarming rates (WHO, 2019, 2020). In 2020, the global number of TB-reported cases decreased after years of progress due to the COVID-19 crisis and because of an associated underdiagnostics, and it is now slowly recovering. There remain 18 countries with more than 100,000 incident cases, 8 in Africa (Nigeria, South Africa, Angola, Mozambique, Democratic Republic of Congo, Tanzania, Kenya, and Ethiopia) and 10 in Asia among which six with more or around 500,000 incident cases (India, Pakistan, Bangladesh, Indonesia, Philippines, China) (WHO, 2021a).

1.2 The Origin of TB Molecular Epidemiology

In 1993, American and European pioneer researchers in the field, working in particular in San Francisco and New York city but also in Europe at the Institut Pasteur or at the RIVM (*Rijksinstituut voor Volksgezondheid en Milieu*, i.e., the National Institute for Public Health and Environment of the Netherlands) have developed new molecular biology methods based on mobile genetic elements (MGE) such as the Insertion Sequence IS*6110*, or later on the—yet undesignated—CRISPR (Clustered Regularly Interspersed Short Palindromic Repeats) locus diversity, and applied it to tuberculosis patient population studies to identify potential genetic similarities between MTBC clinical isolates through DNA fingerprinting followed by clustering analysis of the banding hybridization profiles to infer tuberculosis chains of transmission (Alland et al., 1994; Small and Moss, 1993; van Embden et al., 1993). *Molecular epidemiology* of tuberculosis was defined as a set of alternate methods to traditional case contact tracking and source case finding for TB (Veen, 1992). At that time, we had almost no idea on what the global population structure of MTBC looked like (Bates and Stead, 1993).

During the last 30 years, researchers built a complex yet evolving arsenal of methods that allows to define, *in-depth*, the precise identity of all TB bacilli and allows a new class of studies on the epidemiology of tuberculosis; some of these methods however remains either costly or require bioinformatic expertise, or both. The historical "golden standard" IS*6110*-RFLP[1] method was the first one that allowed standardized fingerprints of TB genomes to be obtained, more than 30 years ago (van Soolingen et al., 1991). When we write *standardized*, one evident limitation was the choice of two different reference strains between the two sides of the Atlantic, H37Rv in the USA and Mt14323 in Europe (van Embden et al., 1993). IS*6110*-RFLP was followed by the invention of the CRISPR-typing, designated as "*spoligotyping*" (spacer oligonucleotide typing) (Kamerbeek et al., 1997; Sola, 2015), MIRU-VNTR[2] typing (Allix-Beguec et al., 2008; Supply et al., 2006), by deletion-region typing (Freeman et al., 2005; Kato-Maeda et al., 2001), by MLST[3] (Baker et al., 2004), and lastly by whole-genome sequencing (WGS) (Bryant et al., 2013; Gardy et al., 2011), not to mention other methods,

[1]Restriction Fragment Length Polymorphism.
[2]Mycobacterial Interspersed Repetitive Units-Variable Number of Tandem Repeats.
[3]Multilocus Sequence typing.

such as PFGE[4], which was very useful for other rapidly growing pathogens, however remained anecdotical for tuberculosis research since it was much too tedious, cumbersome and brings pathogenic risk whereas it had limited epidemiological interest for tuberculosis (Olson et al., 1995). Baker et al. created the first MLST scheme, by characterization of synonymous mutations in seven genes (*gyrA*, *oxyR*, *ahpC*, *katG*, *pncA*, *rpoB*, *rpsL*). Four main MTBC lineages had been described in 2004 (Baker et al., 2004) and six in 2006 (Gagneux et al., 2006).

1.3 Back to the Present Epidemiological Situation and Link With Science and Politics

Researchers rapidly demonstrated that TB transmission in large cities was not merely due to reactivation of an ancient cured disease that would have remained latent, but was thriving through recently transmitted cases in proportion of 30%−40% at least and not only around 10% as initially believed by conventional epidemiological studies. Clustering of multidrug-resistant TB cases was also demonstrated at that time (Moss et al., 1997). Such a quantitative change was of course significant: TB was back in the Western world, it was also silently prospering in Asia and Africa though unnoticed because unstudied. The European Union fortress wavered, launched a coordinated research program to promote these new tracking methods in order to definitively eradicate TB in Europe. Twelve years of concerted European action research on tuberculosis molecular epidemiology followed during which many FP6 and FP7[5] research projects (1994−2006) were launched and during which Network of Excellences (NoE) on tuberculosis research were created. However 30 years have passed and tuberculosis is still there, in Asia, in Africa, particularly in refugees, war victims, migrant populations, and incarcerated people (Eldholm et al., 2016; Ismail et al., 2018). If we did not lose the war against TB, we nevertheless lost in 2005 the opportunity to develop an integrated public health system in EU, that still rely on a combination of different national and heterogeneous public systems and an unique *integrated* (i.e., with laboratories, epidemiologists, computer scientists all at the same place) European Center for Diseases Control (ECDC) remains to be created (Tibayrenc, 2020).

1.4 The Discovery of BCG

At the end of the 19th century, the birth of microbiology triggered the discovery of new bacteria among which the human tubercle bacillus, by Robert Koch in 1882. This was rapidly followed by the discovery of a bovine variant by the American scientist Theobald Smith (dubbed as a new species by Karlson and Lessel in 1970). A subclone of a *Mycobacterium bovis* isolate gave rise to the most administered live-vaccine that was developed by Albert Calmette and Camille Guerin in 1921, the BCG, that changed the fate of tuberculosis in the world. During the next decades, maintenance of the BCG strain progenies in different countries contributed to the in vitro evolution of BCG. The BCG daughter strains (substrains) were shown to be genetically diverse with variable morphological, culture, antigenic characteristics, and residual virulence due to in vitro adaptation (Oettinger et al., 1999). From a taxonomic perspective, the human and bovine variant however define a single bacterial species,

[4]Pulse-Field Gel Electrophoresis.
[5]Sixth and seventh Framework Program of the European Union.

with around 97.5%−99% average nucleotide identity (Riojas et al., 2018). Hence, human and other mammals share a similar deadly disease, though with different physiopathological mechanisms.

1.5 Milk Pastorization and Its Impact on TB

While it was never believed that the bovine variant could have entirely been responsible for today's world-wide human tuberculosis pandemics, the link between human and bovine disease was made as early as 1905 and after slaughtering programs, around the 1950s, pasteurization of the milk increased and prevented in part pediatric tuberculosis cases (Osborne and Neale, 1966). Today milk safety through pasteurization is considered as an important step forward in food safety, and cattle tuberculosis is of great and renewed interest for evident economic reasons. The traditional belief during the 1960s was that tuberculosis must have been a mammalian food-borne disease, that infected animals through gastrointestinal invasion, and that it later switched host, to specialize as a human airborne pathogen, i.e., a pathogen transmitted through droplets in aerosols, reaching human pulmonary tissues, adhesion to endothelium and thus created the infection mechanism (Cockburn, 1963). Hence, pulmonary versus extrapulmonary (e.g., abdominal tuberculosis caused by *M. bovis*) tuberculosis diseases are likely to follow partially different pathophysiological processes, either of intestinal or respiratory origin.

1.6 A Change of Paradigm

In 2002, however, one of the consequences of the molecular biology revolution was a change in paradigm: R. Brosch et al. by using comparative genomics of human and of bovine tubercle bacilli demonstrated that today's bovine TB bacilli had a shorter genome than human TB bacilli (Brosch et al., 2002). Evidence from genetics (loss of genetic material can be considered as Unique Event Polymorphism (UEP) with no return to the former genetic state) is a fact: the human tuberculosis disease is likely to have preceded the bovine tuberculosis disease (Brosch et al., 2002) (see Fig. 27.1).

Such a change of paradigm should be considered with caution: human beings tend to adopt radical ways of thinking, being either "right" or "wrong," whereas the gathering of facts is often contradictory and the quest of the truth a long-term process.

Human ancient DNA studies may allow to reconstruct the complex story of human colonization and migrations on Earth through time, by mitochondrial (female transmitted marker) and by Y chromosomal (male transmitted marker) genomic analysis (Cavalli-Sforza, 2000; Comas et al. 2013; Silcocks and Dunstan, 2023). Recent results obtained using human genetics suggests that a homozygous mutation on the TYK2 (tyrosine kinase) gene, the P1104A polymorphism, that was found to be associated to higher risks of developing clinical forms of TB, could have emerged during the upper Paleolithic in West Eurasia 30,000 years ago (Kerner et al., 2021). This mutation would have been counterselected through Darwinian selection after a peak during the bronze age around 2000 years ago, and is still found at high frequency in British people (Kerner et al., 2021). For the authors, it is likely that the increase in frequency of this homozygous mutations coincides with the Neolithic revolution that created anthropogenic changes within the environment (Kerner et al., 2021). Such changes in turn, quite counterintuitively coincides with a worsening of human health status during the early Neolithic as compared to the Paleolithic. Indeed the Paleolithic versus Neolithic hypothesis of the origin of tuberculosis remains quite a controversial issue (Achtman, 2016; Cardona et al., 2020; Dutour 2023),

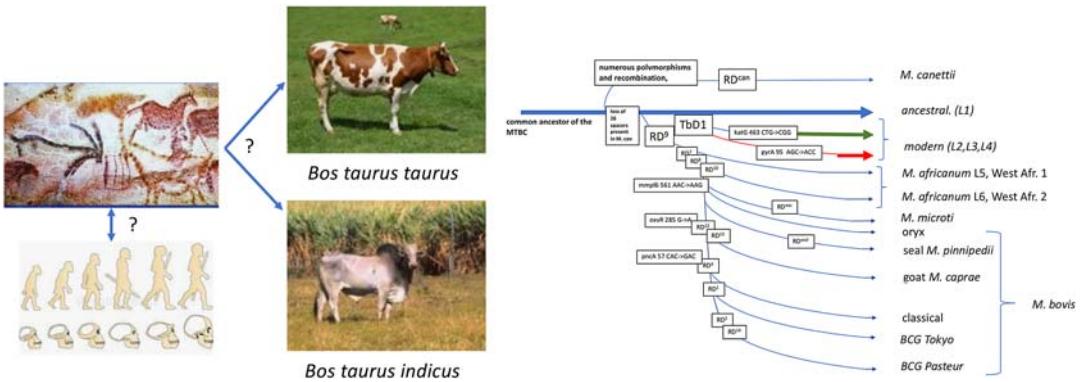


FIGURE 27.1 Global phylogeny of *Mycobacterium tuberculosis* complex.

Is tuberculosis a human disease that infected animals or is it the opposite? *left*: potential paleolithic-cattle interactions; *center*: the two domesticated subspecies of *Bos*, *right*: the phylogenetic tree of Brosch et al. (2002).

From Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., Van Soolingen, D., Cole, S.T., 2002. A new evolutionary scenario for the Mycobacterium tuberculosis *complex. Proc. Natl. Acad. Sci. U.S.A. 99 (6), 3684–3689. https://doi.org/10.1073/pnas.052548299.*

even though the gathering of clues for a paleolithic origin are growing (Coqueugniot et al., 2023; Pálfi et al., 2023a,b; Lee et al., 2023).

Moreover, the genetic diversity of human tuberculosis bacilli was never eroded as it was the case by slaughtering of cattle suffering for bovine tuberculosis bacilli, that appeared as soon as milk was acknowledged to be a potential reservoir of pediatric tuberculosis at the turn of the 20th century. Today in the western world, the combination of slaughtering, cattle breeding and selection aims to develop infectious diseases-resistant herds (Brooks-Pollock et al., 2014). Entire herds were slaughtered when TB cases appeared in domesticated cattle since the 1950s (Brooks-Pollock et al., 2014). These facts and an increasing global cattle trade, supposed to be resistant to tuberculosis in improved cattle bred, has totally changed the original bovine tuberculosis bacilli population structure hosted by cattle over time. Ancestral bovine TB strains, that were born in Africa, are very likely to have vanished (Loiseau et al., 2020). "New" ones such as *M. orygis* are emerging (van Ingen et al., 2012). We also know almost nothing about TB bacilli genetic diversity in extinct mammals or within African cattle, even though a naturally trapped bison with pathological changes suggestive of tuberculosis, and ancient DNA with *M. tuberculosis* markers, were found in Wyoming, USA (Rothschild et al., 2001). In other words, comparison between so distant genetic variants (*M. bovis* and H37Rv) of the same bacterial species may have introduced a bias in the results that have promoted the change of paradigm (Brosch et al., 2002). As we will see below, the history of *M. africanum* (L5 and L6) and the discovery of new recent sublineages (L9, L10) could help to answer the link between human-adapted and animal adapted MTBC.

1.7 Tuberculosis Archeology

In 2008, Hershkovitz et al. described the most ancient cases of Neolithic TB in Atlit-Yam, on the eastern Mediterranean shore (Hershkovitz et al., 2008). At least two more articles published since 2002 suggest that the history of the origin of animal/human TB diseases could be much more complex than believed. Both concern studies done on ancient DNA and may help to unravel the early history of tuberculosis. The first of these studies was run on DNA extracted from necropolis of the Tombs of the Nobles in Thebes West, Upper Egypt, from Middle Kingdom (2050−1650 BCE), New Kingdom (1550−1080 BCE) and from the Late Period (until 500 BCE) (Zink et al., 2003). *M. africanum*-identified DNA (today classified as L5 or L6 lineages) was found in the oldest tomb complex investigated (2050−1650 BCE) whereas *M. tuberculosis sensu stricto* strains (belonging to L4), were found in the younger material suggesting that L4 and L5−L6 may be related to each other or L4 took the place of L5−L6 either because of virulence or for other yet unknown reasons (Zink et al., 2003). These results were obtained looking at specific CRISPR spacers as detected by spoligotyping, a method that looks at passed phage-bacteria cross-talks (Bolotin et al., 2005; Kamerbeek et al., 1997; Mojica et al., 2005; Pourcel et al., 2005) (see Fig. 27.2).

A second bioarcheological study was published in 2007; five Iron Age individuals with spinal lesions recovered from the cemetery of Aymyrlyg, in South Siberia, showed the presence of *M. bovis* DNA, suggesting that animal reservoirs may have been the cause of the human disease, at least in these seminomadic iron-age pastorialist communities (Taylor et al., 2007). Since *M. bovis* and all animal-derived lineages are also sister lineages of L6, we definitively need to deepen our global understanding of L5, L6, and L4 ancient genomes.

1.8 Evolutionary History of TB: Theoretical Issues

Anthropologists suggest that a closed community of 180−440 individuals, depending on pathogen reproductive potential, could be required for an infectious disease to be maintained as endemic (McGrath, 1988). *M. bovis*, in herds of similar size, could eventually be added to these human communities to promote infection sustainability, and conditions of cross infections between human and animal could have been fulfilled in different ecosystems, although today's *M. bovis* has a limited pathogenic power for humans. Isolated groups of hunter-gatherers or of early sedentary tribes or nomadic agro-pastorialist tribes were small, of about 100−150 people maximum. If the introduction into this model of individuals from a cattle reservoir are added to the McGrawth model, human endemicity of a TB disease of animal origin could have been possible before the demographic burst of human beings. The increase of demography would have further allowed, but at a later stage, patient to patient airborne transmission.

Hence, several parameters could have contributed alone or jointly to disease transmission in early tuberculosis ages: (1) rate of transmission of an animal-infected compartment to a human-infected compartment through intimate contacts and vice versa; (2) reproductive rate of human or animal TB in a closed environment at equilibrium (caves, hut, primitive habitats); (3) demographical history of the human tribe and of their cattle; (4) transmission rate from human-infected population tribes to new naïve population tribes/bands through migration or social contact, since bands or tribes of hunter gatherers were never totally isolated, and (5) likelihood of long distance migration and TB transmission of one or many other potential animal reservoirs than *Bos taurus, B. indicus*, or their ancestors

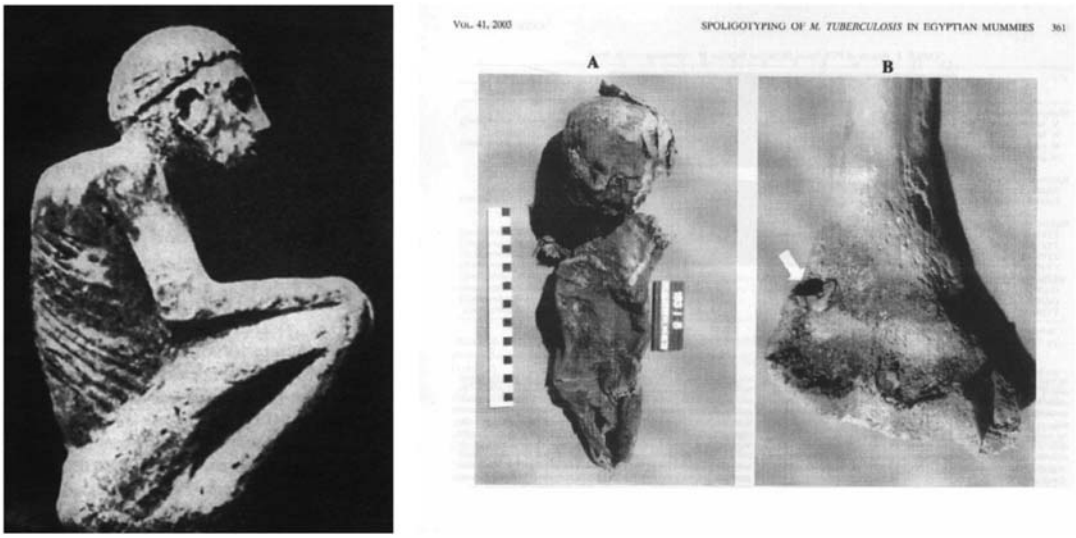


FIGURE 27.2 Archeological and bone artifacts showing signs of tuberculosis.

Left: an emaciated clay statue of a man with a classical kyphosis; *right*: Case TT183-8. The mummy of a newborn male child without any sign of evisceration. After careful opening of the chest wall, fine pleural adhesions of the lung indicated a tuberculosis infection. (B) Case TT95-PC169. The right humerus shows morphological alterations, probably due to a nonspecific inflammation. At the distal end, reactive new-bone formation and fistular defects (*arrow*) are visible.

Left figure reproduced from Haas F and Haas S.S, Origins and Spread of Mycobacterium tuberculosis in the Mediterranean basin, in Tuberculosis Past and Present, Golden Book Publisher Ltd, Tuberculosis Foundation, 1999. Right figure reproduced from Zink, A.R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H., Nerlich, A.G., 2003. Characterization of Mycobacterium tuberculosis *complex DNAs from Egyptian mummies by spoligotyping. J. Clin. Microbiol. 41 (1), 359–367. https://doi.org/10.1128/JCM.41.1.359-367.2003 by courtesy of the authors and editors.*

(Bird et al., 2019; Porco and Blower, 1998; Sanchez and Blower, 1997). Concerning the route of infection, of course the mentioned hypothesis of an airborne pathogen does not exclude an initial oro-fecal transmission and sustainability of the most recent common ancestor (MRCA) of the modern tuberculosis pathogen agent in humans before a change of pathophysiological process of infection occurred. In this sense, a link between specific microbiota and an infection with *Mycobacterium africanum* was suggested recently, that opens new ways of research (Namasivayam et al., 2020).

As seen above, such a modeling of an early infection dynamics is hard if not impossible to reconstruct since all these parameters might have been only gradually changing through time, with slowly evolving ecosystems and complex nomad pastorialists toward sedentary agricultural interactions (Vågene et al., 2022).

1.9 Development of Tuberculosis Genetics and Links to Clinics

Before the advent of molecular epidemiology, clinicians working on tuberculosis used to meet during international congresses to exchange their descriptions of the syndromes of the TB disease that they could observe in their respective countries, i.e., the various clinical expression patterns of tuberculosis diseases. "*This is not what I see in my country*" might have been a general observation of the participants. Indeed, as rapidly detected by pathologists of the former British Empire, interesting observations pointed out that clinical MTBC isolates from patients living in United Kingdom and of European origin were much more pathogenic in a guinea-pig model than MTBC isolates from Indian patients (Bhatia et al., 1961, 1963; Mitchison et al., 1960). It is known today that South Indian patients are more likely to host L1 isolates whereas UK patients of European origin are more likely to host L4 isolates (Couvin et al., 2019a; Dale et al., 1997; Evans et al., 2011). Subtle or larger deletions and small differences in genomes could explain specific clinical expression forms of the TB disease. The study of the pangenome, the full range of genes in a given species, and the differential study of conserved versus dispensable/accessory genes, is of less importance for a clonal bacterium such as MTBC, however, recent results suggest that the core genome in some lineages, could be more reduced than expected and the accessory genome larger, however these results should be taken with caution (Periwal et al., 2015; Reis and Cunha, 2021; Zakham et al., 2021). Gene presence or absence, in combination with a large diversity of mutations, could give rise to an extended landscape of variations in pathogenicity and clinical expression of the TB disease. One good example is found in some proteins of the type VII secretion system, responsible of virulence, immune modulation and nutrient uptake, that were shown to be responsible inside the L2 lineage of a hypervirulent phenotype (Ates et al., 2018b).

1.10 Genetic Diversification Mechanisms and Phage Typing

Notwithstanding the classical genetic mechanisms of mutation (increase of prevalence by positive selection, purifying selection, drift) let's focus on what may create genetic diversity within and across bacterial species: Conjugation, intra-chromosomic homologous recombination, and infection by bacteriophages are major drivers of diversification. In particular, bacteria-bacteriophages form a subtle equilibrium in ecosystems, that is sometimes broken for the exclusive benefit of one of the two partners: the phage can become *lytic* and eliminate the bacteria, or the phage can become integrated into the bacterial genome and then becomes *temperate*, i.e., a prophage integrated to bacterial genomic information. Particle (phage) counts depend on the success or failure of the attack in many systems (Salmond and Fineran, 2015). A consequence of this "arm race" between phage and bacteria had been shown to occur in mycobacteria as early as 1969 by Inge Baess in Scandinavia, Joseph Bates in the USA or Sabine Clevel-Ceres in France through the discovery of phage-typing systems (Baess, 1966; Bates and Fitzhugh, 1967; Clavel-Seres and Clement, 1984). However, phage-typing (designated as lysotyping, i.e., the ability to distinguish phage-sensitive from phage-resistant bacteria) was too difficult to standardize to allow progress of the taxonomy of MTBC (Baess, 1966; Bates and Fitzhugh, 1967; Clavel-Seres and Clement, 1984). The discovery of the bacterial CRISPR locus and its relations to previous phages encounters allowed to speculate that this locus designated as "Direct Repeat" locus in MTBC was containing both a very ancient phylogenetic information on bacteria/phage interaction, and a specific evolutionary history of this very locus due to microevolution, acting in relation with other mobile genetic elements (MGEs) such as IS*6110* (Groenen et al., 1993; Thierry et al., 1990).

1.11 Main Lineages Identified so Far

Nowadays, ten Lineages (L1 to L10) and at least 90 sublineages of MTBC have been described and characterized at the genomic level (Coll et al., 2014a; Freschi et al., 2021; Napier et al., 2020; Guyeux et al. 2024). At least 90 robust phylogenetic SNPs (one at least per hierarchical sublineage) allow to assign to a specific lineage and sublineage for most clinical isolates, wherever they come from, high or low endemic countries (Napier et al., 2020). The genetic diversity of MTBC isolates in most countries (with minor exceptions) has been characterized, and we may confidently argue that there are only few geographically isolated areas, where new MTBC variants samples could be discovered, even if this was the case recently with the L7, L8, and L9 lineages characterized from clinical isolates, respectively, found in Ethiopia, in Rwanda/Uganda and East-Africa (Blouin et al., 2012; Coscolla et al., 2021; Ngabonziza et al., 2020). New sublineages within L4 in particular in African countries are also being discovered based on WGS, e.g., the "*Congo*" or L4.7 (Malm et al., 2017). The ancestral Lineage 4 types may have been present in Africa in early human history or alternatively been introduced with European colonization between the 16th and 19th centuries. Several genetic families were identified and named according to their countries of first isolation, such as Uganda (L4.6.1), Ghana (L4.1.3), Cameroon (L4.6.2), Congo (L4.7). They were initially proposed upon distinct spoligotyping and/or Variable Number of Tandem Repeats (VNTR) or more recently based on SNPs typing patterns. Whole-genome sequencing (WGS)-based studies most of the time confirmed or even improved the definition of previously described phylogeographical specific lineages (Malm et al., 2017; Meehan et al., 2019). However we still have a dramatic under representativity of L5 and L6 in genome databases, that still limits our understanding of the population genomics of L5−L6 (Coscolla et al., 2021).

Because of the exquisite diversity and never ending entropical forces acting on MTBC bacilli, local epidemic variants might either thrive or disappear, and it seems an utopia to gather all representative isolates of all TB micro-outbreaks into a single scientific study: TB genetic bacilli are permanently evolving, clones appear and disappear, and outbreaks are often diagnosed retrospectively (Eldholm et al., 2015; Mokrousov et al., 2021). A good example is the difficulty to adequately classify L2-Beijing isolates to one of the—at least 13 up to 30—different sublineages (Comas et al., 2009; Freschi et al., 2021; Shitikov et al., 2017; Thawornwattana et al., 2021). The power of WGS and big data allows to provide exquisite characterization of isolates based on SNPs but this should also include the diversity of other polymorphic markers such as (1) the region of deletions RDs (Tsolaki et al., 2004), (2) the CRISPR-locus (van Embden et al., 2000), (3) the Mobile Genetic Elements (MGEs) such as IS*6110* and others (Roychowdhury et al., 2015), (4) the polymorphic GC-rich regions (PE-PPE genes) more prone to recombination and not user-friendly in analysis (Phelan et al., 2016).

Gathering and mathematical processing of all these variations, should finally and ideally allow to have no ambiguity left, i.e., to assign each MTBC clinical isolate to its specific evolutionary history and to assign it to its ultimate taxonomical status, i.e., a unique identity at a given time, within a given geographical setting, and at a specific point of time of the pandemics (see Fig. 27.3).

More than a perfect taxonomic identification, clinical microbiologists expect from genomics an improved patient-centered treatment. A first robust list of gene variations associated to drug resistance was released during the summer 2021 by WHO (WHO, 2021b) and this list is going to evolve with the discovery of new mutations linked to new drug regimen use for MDR-TB than includes new drugs such as bedaquiline, linezolid, pretomanid, delamanid, and clofazimine (Kadura et al., 2020). Up to 77 articles in PubMed are found using the keywords: "*Machine Learning*" and "*tuberculosis*" and "*Drug-resistance*" (search done on Nov. 16th, 2023). These results definitively show that tuberculosis

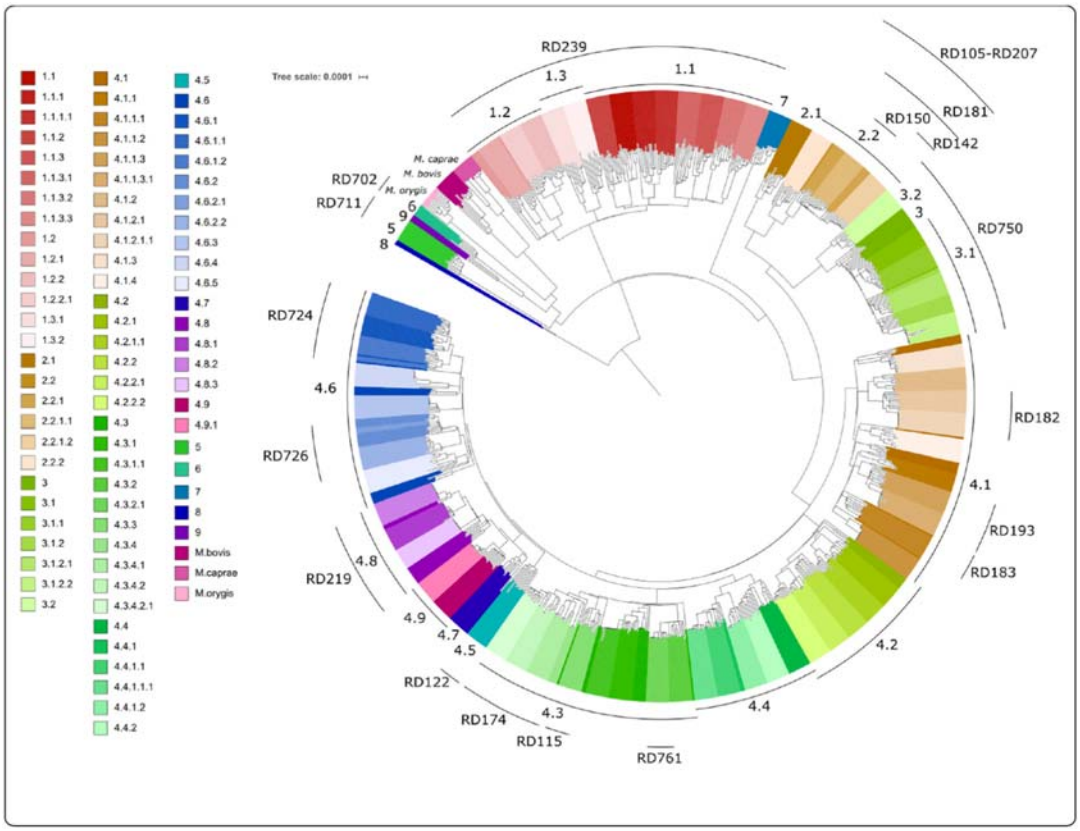


FIGURE 27.3

Latest global population structure of *Mycobacterium tuberculosis* complex based on more than 35,000 genomes analyzed.

A global phylogenetic tree describing the worldwide genetic diversity of TB bacilli in 9 lineages.

Reproduced from Napier, G., Campino, S., Merid, Y., Abebe, M., Woldeamanuel, Y., Aseffa, A., Hibberd, M.L., Phelan, J., Clark, T.G., 2020. Robust barcoding and identification of Mycobacterium tuberculosis *lineages for epidemiological and clinical studies. Genome Med. 12 (1). https://doi.org/10.1186/s13073-020-00817-3 by kind courtesy of the author.*

treatment already is and will be more and more guided by genomic results and expert-based computational systems.

2. An Introduction to Genomics and Computational Analysis Methods of Whole-Genome Sequencing Data

2.1 Basic Methodology of WGS

With the advent of WGS, tuberculosis molecular epidemiology has changed era, however, genotyping has started as early as 1965 (Fig. 27.4).

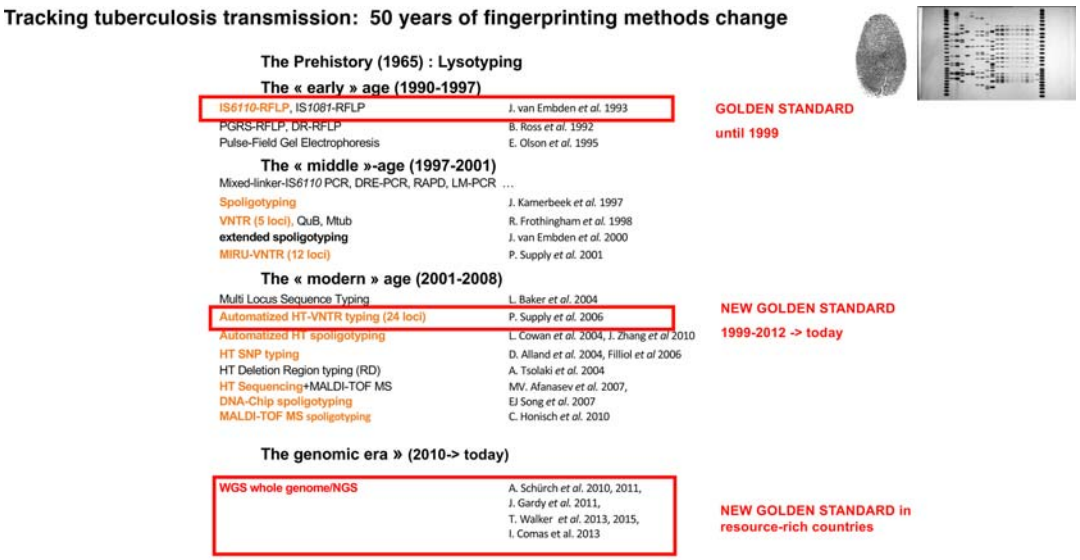


FIGURE 27.4 Chronology of methods development to assess MTBC genomic diversity.

The building of an arsenal of methods to track tuberculosis transmission, during the last 30–50 years, with the genomic era starting around 2010.

Genomic raw data, whatever the sequencing technology may be, is constituted by a sum of sequence read archives (SRA) (Leinonen et al., 2011). SRA can be produced either as paired-end (both strands) or single-end (one strand only). Paired-end SRA allows better alignment, especially for repetitive or difficult to read (high-GC) sequences (Cock et al., 2009). Many bioinformatics pipelines use paired-end preferentially. SRAs first must be cleaned before they are going to be used by other bioinformatical tools (please refer to this link for more information: https://www.bioinformatics.babraham.ac.uk/projects/). A very popular and standardized tool to do that is *fastQC* (Endrullat et al., 2016). This tool creates different graphics to check genomic parameters. For example, GC content can be calculated to verify that nothing differs from what is known for the species and this software can also detect contamination by other species if the contaminating species has a different GC content. Among the other issues, the persistence of adaptor sequences in the final sequences, a too low *phred* score, a suboptimal library, may jeopardize the results when comparing numerous sequences with bioinformatics tools. The FASTQ file format has emerged as the standardized format (Cock et al., 2009).

The *phred* score, as mentioned above, represents the error rate for each base as assigned by the sequencing process. For example, a *phred* score of 30 means an error rate at 0.1% and defines "reasonable quality" sequencing, a phred score of 20 (quite low) defined a "minimum quality" for many tools (error rate 1%), whereas a phred score of 40 defined an error rate of only 0.01%, i.e., a "high quality" sequencing. The phred score threshold must be adapted according to the theoretical depth coverage expected and mean of depth coverage obtained, depending on library quality. In fact,

for some species like *Mycobacterium leprae* (uncultivable) the sequencing will be much more difficult to obtain than for *Escherichia coli*, and for each species the number of SRA produced can be very different. When a small number of reads is produced, the theoretical depth coverage will be weak and the *phred* score can be too small for many base positions in reads. Hence, if we put a threshold at 30, a large number of reads will be removed and the final number of reads may not be sufficient to perform further genome analysis. Thus, it is very important to adapt the *phred* score threshold according to the number of reads produced. If the threshold is fixed too high, some reads are going to be cut too much and many reads will be short. Too short reads may introduce bias because they may align on the reference genome at several positions creating artifactual mutations.

Moreover, for some reads, adaptor sequences may still be present. These sequences must absolutely be removed to minimize the risk of errors during all bioinformatical downstream processes. In fact, de novo assembly (using softwares like *Spades* (Bankevich et al., 2012)), contrarily to *reference genome—based* assembly, can create mis-assembly on some reads if adaptator sequence remain present and if the reads are matching with a common region. The most classical reference genome used in WGS of tuberculosis is H37Rv (NC000962.3 version). With de novo assembly, when annotation tools (like *Prokka* (Seemann, 2014)) are used, some genes may not be detected in case of misassembly; alternatively, genes detected by de novo assembly may generate more diversity than expected into a sample. A recent paper suggested that errors in detecting traces of recombination in MTBC were likely to have been introduced due to errors in genomes alignments (Godfroid et al., 2018).

2.2 WGS Applications: Drug-Resistance Studies and Genomic Epidemiology

For an improved personalized medicine, mutation studies of specific drug-resistance or virulence genes can be directly assessed by WGS on a clinical sample (CRyPTIC et al., 2018; Ransom et al., 2020). With these methods, prediction of drug-resistance can be obtained for most important antituberculous drugs (CRyPTIC et al., 2018; Kadura et al., 2020). An alternative to WGS is targeted sequencing, such as what is done in Deeplex (Genoscreen, Lille, France) (El Achkar et al., 2019). In addition, phylogenetical trees can be computed on SNPs to detect clonal isolates, which helps in distinguishing reinfection from relapse (if the previous isolate was still available or has been sequenced) to detect outbreaks and to reconstitute the evolutionary history of a given clinical isolate or lineage (Ajawatanawong et al., 2019; Mizukoshi et al., 2017; Palittapongarnpim et al., 2018). In case of detection of MDR-TB clusters, cases of MDR-TB transmission can be differentiated from cases of patient-acquired drug-resistance (Advani et al., 2019).

2.3 WGS Allows the Most Accurate Classification of MTBC

The increasingly sophisticated methods described above have changed our understanding of tubercle bacilli diversity, and the issue of its genetic diversity has reached a great maturity level (Freschi et al., 2021; Napier et al., 2020; Thawornwattana et al., 2021). Classical methods (IS*6110*-RFLP, spoligotyping, and VNTR typing) used between 1994 and 2009 were giving—with minor variations but much with much less accuracy—similar clustering results than the ones obtained by WGS (Coll et al., 2014b; Kato-Maeda et al., 2011; Meehan et al., 2019). The reason is that MTBC is a clonal organism with strong linkage disequilibrium, no lateral gene transfer and no recombination (Supply et al., 2003).

2.4 MTBC Genomes Classification Across Time

Fig. 27.5 shows that, as early as 2006, and based on a worldwide representative database of more than 40,000 isolates, the spoligotyping-based international classification into Spoligo-International-Types (SIT) had already proposed the existence of 62 sublineages (Brudey et al., 2006). The first worldwide WGS study published in 2014, based on 1400 genomes, using 62 SNPs to describe 6 main lineages and 60 sublineages confirmed the results obtained in 2006 by spoligotyping (Coll et al., 2014a,b). The reason of this, again, is that MTBC is a highly clonal organism, with very limited or even no genetic exchanges, except for the *M. canetti* outgroup whose genome possesses a mosaic structure (Gutierrez et al., 2005). As a recent example of such premature early discoveries, Freschi et al. describe a "*new*" 4.13 L4 sublineage in their recent paper based on more than 4500 genomes (Freschi et al., 2021). However, looking closely at the spoligotyping-signature of this new L4 sublineage, we observe that they all show a nonconvergent SIT284 signature that had been defined in the early 2005 as "*East-Med1*" based on results obtained in Bulgaria (Panaiotov et al., 2005). All recently described lineages (L7−L9) have initially been detected based on their highly peculiar CRISPR locus structure (Blouin et al., 2012; Coscolla et al., 2021; Ngabonziza et al., 2020), in other words, WGS classification confirmed, extended, and showed the limits of CRISPR-typing classification, that was and still remains a powerful cost-effective technique given the clonal structure of MTBC (Fig. 27.5).

2.5 Technological Evolution of Molecular Epidemiology

With the introduction of WGS, tuberculosis genotyping went through a revolution: from repetitive sequence-based polymorphism studies, investigators from resource-rich countries who could afford doing sequencing switched to WGS (Comas and Gagneux, 2009; Gardy et al., 2011; Schürch et al., 2010). This change was subtle at the beginning however it became more evident in 2013, when a large epidemiological study performed in UK set the references in genomic epidemiology of tuberculosis (Walker et al., 2013). Local and global population genomics studies on tuberculosis as well as host-bacteria interactions studies had already started (Gagneux, 2012; Gagneux and Small, 2007; Homolka et al., 2012; Merker et al., 2015). In a couple of years, SNP typing became a method of choice and SNP discovery within all lineages a new haunt (Abadia et al., 2010; Advani et al., 2019; Coll et al., 2014b; Filliol et al., 2006). Very large datasets, setting referential frameworks for a new MTBC genomic taxonomy were built, however using sometimes quite complex local bioinformatical designations of clones and without any international consensus (Coll et al., 2014b; Freschi et al., 2021; Menardo et al., 2021; Napier et al., 2020; Palittapongarnpim et al., 2018; Thawornwattana et al., 2021).

We have reached the time where the detection of spatial and temporal phylogenetic signals in large genomic datasets will allow to precisely understand the complex historical and geographical network of genotypes emergence, of their spreading dynamics history and of their evolution (Senelle et al., 2023a,b). In brief, we are entering into an exciting period in which the global history of tuberculosis, whether in the deep past or contemporaneous, can be reconstructed using both the accumulation of genome diversity studies and the previous gathered knowledge during the last 25 years, i.e., before the genomic era, that still contains lot of information that can now be easily correlated or even reanalyzed with new eyes, thanks to the recent input of genomics (Meehan et al., 2019; Menardo et al., 2021; Rutaihwa et al., 2019). That is the reason why, we will now dive deeper into this detailed story.

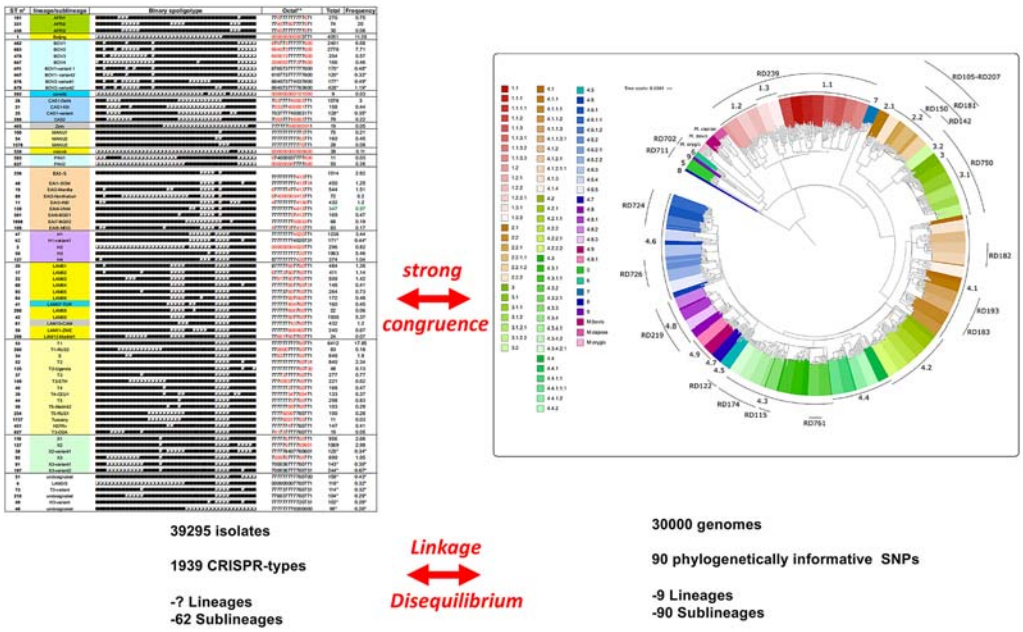


FIGURE 27.5

Congruence between spoligotyping-based and WGS-based defined global population structure of the MTBC complex is valid for many lineages and sublineages due to strong linkage disequilibrium in MTBC.

From Brudey, K., et al., 2006. Mycobacterium tuberculosis *complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics, and epidemiology. BMC Microbiol. 6, 23. doi:10.1186/1471-2180-6-23.*

3. Lineage 1: The "East-African Indian" (EAI) Lineage According to Sola et al. (2001) or "Indo-Oceanic" According to Gagneux et al. (2006)

3.1 An History of L1 Discovery and on Issues Left Unsolved

Using a recruitment of MTBC clinical isolates in Tanzania, Malaysia, and Oman, and the IS*6110*-RFLP method, a Malaysian-British team described very peculiar isolates of MTBC (Fomukong et al., 1994). These MTBC isolates showed a single-copy of IS*6110*, and RFLP patterns that were different from *M. bovis*, the MTB subspecies also known for having a low IS*6110* copy number. Other results concomitantly showed that South-East Asian countries (Vietnam in particular) were also hosting clinical isolates with low-copy number of IS*6110* (Yuen et al., 1993). Similar strains were also found in Ethiopia (E1 isolates) (Hermans et al., 1995). In 2001, putting together data from large and complementary studies, C. Sola et al. described the EA-I (East-African and Indian, now L1) and LA-M (Latin-American and Mediterranean, now L4.3) MTBC sublineages based on a strong congruence between 5 ETR loci, in particular on ETR-A, and spoligotyping results (Sola et al., 2001c). EA-I was

later found to be identical to L1, whereas LA-M was found to be identical to L4.3 (Coll et al., 2014b). Still the question on TB origin remains: is tuberculosis an Asian or African-born disease and can we link it to early human historical demography and genetic geography (Cavalli-Sforza et al., 1994) (Fig. 27.6).

Based on a genetic diversity study done in Ethiopia, Comas et al. suggested that Africa was not a virgin land of tuberculosis before colonization (Comas et al., 2015). Indeed, it is only in East-Africa that all lineages can be found (Comas et al., 2015). In the new world, another study suggested that before the discovery of America, the disease was already present on this continent (Salo et al., 1994; Bos et al., 2014). Hence, the birth of tuberculosis as a disease could be as ancient as the birth of Anatomically Modern Humans (AMH) or be as ancient as 70,000 years old, even if neolithization could have been a major driver of host change (Comas et al., 2013). Other scholars believe that tuberculosis is entirely linked to the neolithization process, i.e., be only 6000−8000 years old (Achtman, 2008, 2016; Sabin et al. 2020). Genetic drift and natural selection are the two main forces acting on tuberculosis genomes evolution (Pepperell et al., 2013). In MTBC, a strong purifying selection process is observed, together with a 25-fold expansion of the global MTBC bacilli population

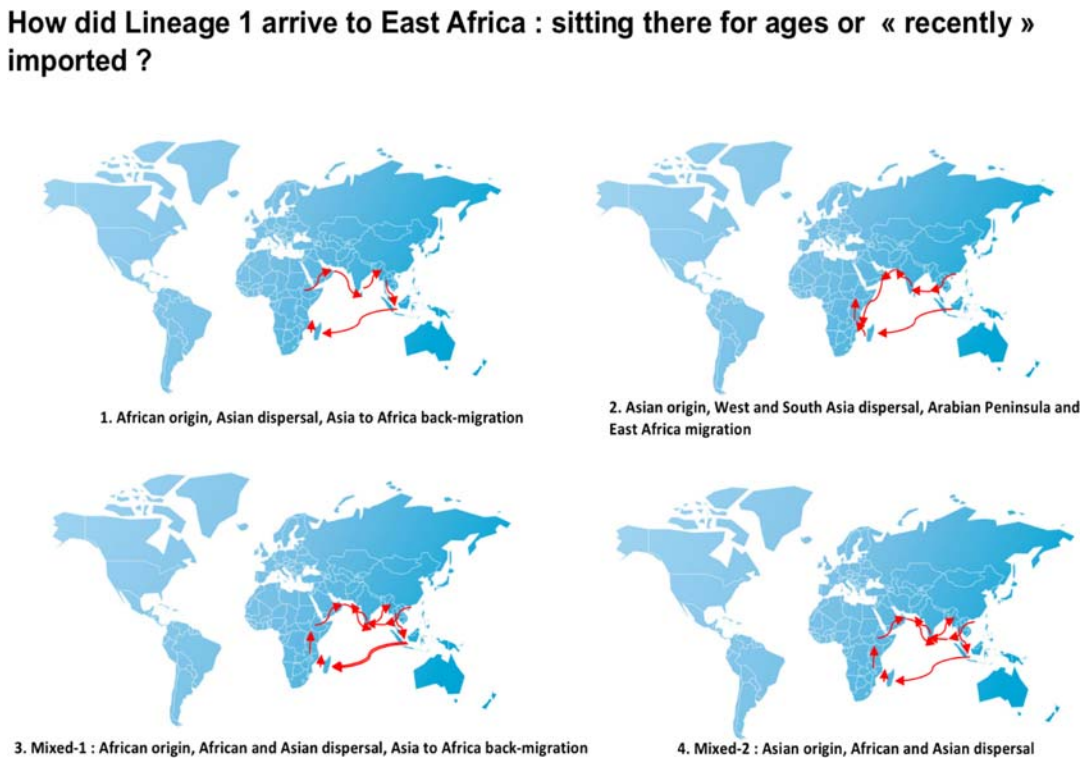


FIGURE 27.6

Suggested mixed-models of L1 tuberculosis spreading to or from Africa and Asia; did L1 emerge in Africa or South-Asia?

that likely occurred during the XVIIth century (Pepperell et al., 2013) when human population experienced strong expansion (Roser et al., 2013; Comas et al., 2015). Indeed, it is only in East-Africa that all lineages can be found (Comas et al., 2015). In the new world, another study suggested that before the discovery of America, the disease was already present on this continent (Bos et al., 2014; Salo et al., 1994). Hence, the birth of tuberculosis as a disease could be as ancient as the birth of Anatomically Modern Humans (AMH) and be as ancient as 70,000 years old, even if neolithization could have been a major driver of host change (Comas et al., 2013). Hence, it is natural to ask ourself if the selection process was similar worldwide, or different in Europe, Asia, and Africa.

3.2 Genomic Characteristics of L1

One of the main characteristics of the L1 lineage, was the presence of a specific set of genes, designated as TbD1, discovered in 2002 (Brosch et al., 2002). These genes are also present in L5, L6, and animal strains, however TbD1 is absent in many other MTBC strains (L2, L3, L4). This region encodes for the MmpL6/MmpS6 system, suggested to be an oxidative stress response system providing a selective advantage to MTBC under stress conditions (Bottai et al., 2020). The MmpL/MmpS proteins system is known to be a transport system of paramount importance that helps to achieve the correct cell envelop building of MTBC (Bottai et al., 2020) and can function in some case as efflux pumps (Yamamoto et al., 2021). The primary role of these proteins is to translocate virulence-associated envelope lipids across the plasma membrane to the periplasmic space. Bottai et al. suggested that the partial loss of this region in L2, L3, L4, from an L1 precursor contributed to their epidemic success (Bottai et al., 2020). However, the variability of other genes, particularly those involved in the synthesis of the cell wall lipids, made-up of complex molecules (lipids and lipoglycans), known to act as virulence factors, also play an important role in the pathogenicity of MTBC lineages and sublineages as shown recently (Gisch et al., 2022).

3.3 Geographical Distribution of L1

L1 is particularly prevalent in South India and South-East Asia, to a lesser extent in East-Africa, and its global population structure was recently shown to be made of at least 5 sublineages (L1.1.1, L1.1.2, L1.1.3, L1.2.1, and L1.2.2) (Menardo et al., 2021). Some of these sublineages show strong biogeographical specificities, e.g., L1.1.1, L1.1.2, L1.1.3 are more likely to be, respectively, found in the Indo-Chinese basin, South-Asia, East-Africa, and South Asia, respectively, whereas L1.2.1 and L1.2.2 are more likely to be found in East Asia and South Asia, respectively (Menardo et al., 2021). Signatures of local adaptation at the *esxH* locus (a gene coding for a protein that is part of one of the five ESX operons, encoding type VII secretion systems) that targets the human endosomal sorting complex, were also found in L1 (Menardo et al., 2021).

3.4 Importance of South-East and South Asia in Early History of TB

South-Asia undoubtedly played an important role in tuberculosis emergence and evolution and the ESX genes, and are likely to be of paramount importance in tuberculosis virulence (Orgeur and Brosch, 2018). In India, for reasons likely to relate to human population genetics, L1 is known to dominate in the South, whereas L3 is more prevalent in the North of the country (Couvin et al., 2019b).

However, the global biogeographical diversity of Indian MTBC isolates, as it is the case for African MTBC isolates, is still underrepresented in global databases with regard to the TB burden in India (Poonawalaa et al., 2020). The human genetics and ecological links that could relate L1 to specific human and/animal MTBC pathogens populations in South-India could be (1) the history of Dravidian populations, (2) the caste system and food hierarchy in Hindu's culture (Pitchappan, 2016; Sathyamala, 2019), (3) other demographical, historical and socio-cultural factors that remains to be studied. From a migratory perspective, migrations from South-Asia to the rest of the world was done through a likely maritime migratory route for L1 spreading, and, according to some authors, the Most Recent Common Ancestor (MRCA) of all L1 may not have been present before the 12th century AD, whereas it could already have been present at the fourth century BC according to other investigators (Menardo et al., 2021; O'Neill et al., 2019). The presence of the SIT109, a subbranch of L1.1.2 in Madagascar could be either explained by the fact that the Malagasy people are known to be a mixture of Borneo or Indonesian based migrants mixed with Bantu origin Africans migrants, or by more recent South-Asian immigrants (Pierron et al., 2017).

3.5 Specific Data on the Manila Lineage

Recently, the L1.2.1 sublineage, identical to the "Manilla type," discovered in the Philippines, one of the five sublineages within L1, was shown to harbor a deletion in the *whiB5* gene (Wan et al., 2014, 2017). *WhiB5* is a transcriptional regulator that contributes to virulence, acting on 58 genes, including the previously mentioned *esx* genes (Casonato et al., 2012). Extension of genomic diversity studies in South and Southeast Asia will definitely increase our knowledge of the early history of the L1 lineage, which is definitely an important lineage in the global history of tuberculosis.

4. Lineage 2: The "Beijing" or East-Asian Lineage

4.1 History of the Beijing Lineage Naming

The discovery and naming of L2 as "*Beijing*," goes back to the early TB molecular epidemiological research period. The "W" strain, a member of L2, was known since its discovery in New York and San Francisco in the early 1990s (Plikaytis et al., 1994), and all "W" strains were named at the PHRI[6] based on the IS*6110*-RFLP Database; concomitantly D. van Soolingen et al. were studying and described the same family of bacilli based on spoligotyping results, in the Beijing region in China hence the name of the family as *Beijing* family (van Soolingen et al., 1995). The Beijing lineage was characterized by its high IS*6110* copy number (16–20 copies), with a sublineage with 1 copy in the "NTF[7]" region shown to be global (Bifani et al., 2002; Kurepina et al., 1998). IS*6110* insertion in *dnaA-dnaN* is specific for all Beijing, although there are exceptions (Kremer et al., 2004). IS*6110* insertions(s) in NTF roughly subdivide within Beijing. Again as with any jumping element, IS*6110* is not an ideal defining marker, including IS*6110* in NTF (Nakanishi et al., 2013). In 2005, Tsolaki et al. pointed the importance of Regions of difference in the L2 phylogeny (RD105, RD181, RD150, RD142) (Tsolaki et al., 2005). No lineage such as Beijing has been fostering so many genetic, genomic, and molecular epidemiological

[6]Public Health Research Institute.
[7]No true acronym definition can be found, the "noise transfer region" sometimes given to explain this acronym is not justified

studies, whether for detection of multidrug resistance (a feature rapidly associated to L2), for improvement of its taxonomy through deletion region analysis, VNTR analysis, or later for SNP studies—for reviews on the L2 lineage see also (Glynn et al., 2002; Karmakar et al., 2019). Very large outbreaks of different isolates belonging to the "*modern*" L2 lineage were shown to have independently emerged in Russia and Europe, in Central Asia, in South Africa, and of course in China (Casali et al., 2014; Klotoe et al., 2019; Merker et al., 2015, 2018; Mokrousov, 2013; Niemann et al., 2009; Vyazovaya et al., 2020).

4.2 Nomenclature Inside L2 Suggests a Complex History

Initially, the L2 lineage was described as "*typical*" (modern) or "*atypical*" (ancient or ancestral) according to the IS*6110* copy number and presence of an IS*6110* copy in the NTF locus and some other markers (Mokrousov et al., 2002, 2005). Another distinction within the ancestral Beijing can also be done between *early ancient1* (RD207+) and *early ancient2* (RD181+) (Vinogradova et al., 2021). There had been many studies describing the diversity of L2, which added to the complexity of the L2 classification (Luo et al., 2015; Mestre et al., 2011; Rad et al., 2003; Tsolaki et al., 2005). This empirical nomenclature and classification, pioneered the building of a very large consortium that gathered a worldwide representative collection and thus provided a first comprehensive picture of this complex lineage, as studied by a combination of MIRU-VNTR and WGS results (Merker et al., 2015). This collaborative study indeed represented the greatest endeavor to characterize the L2 lineage so far (Merker et al., 2015). Adaptation to a lifestyle with high IS*6110* copy number and to drug-resistance by development of compensatory mutations was shown to increase the fitness of some variants (Merker et al., 2018). Recently, a useful simplification of all previous sublineages naming and L2 markers identification was done by E. Shitikov et al. that described 11 main L2 sublineages and provided a more simple picture of the L2 diversification process (Shitikov et al., 2017). Many markers useful to discriminate L2 stem from the pioneered discovery of the extended polymorphism of the 3R genes that had been found in the Pasteur Institute in Paris at a time when housekeeping gene sequencing had suggested that TB bacilli were very monomorphic, a feature that prevented the development of Multiple-Locus-Sequencing-Typing as a standardized genotyping method for tuberculosis studies, contrarily to many other bacterial models (Baker et al., 2004; Dos Vultos et al., 2008; Mestre et al., 2011; Perez-Losada et al., 2006). Conversely, region of differences typing was shown to be useful for L2 evolutionary studies (Tsolaki et al., 2005). A new small deletion named RD311, that leads to the inactivation of RD2434c was discovered recently in all modern Beijing strains (Bespiatykh et al., 2021).

4.3 Diversification History of L2

The history of the L2 lineage still contains uncertainties: when did the split L2/L3 occurred (L3 being now specific to the North of the Indian subcontinent) and when and where did L2 emerged (Luo et al., 2015). Luo et al. suggested that L2 emerged 30,000 years ago in South-East Asia, i.e., during the paleolithic. Proto-Beijing could have emerged in South-China, and later appeared the origin of RD207 in the North of China (Luo et al., 2015). Merker et al. suggested that L2 has emerged around 6600 years ago in Northeastern China, Korea and Japan (Merker et al., 2015). These two hypotheses are mutually exclusive and the most recent hypothesis would be favored if we consider archeological

evidence and that strains discovered in Korea and Japan are more deeply rooted in the TB phylogeny (Suzuki and Inoue, 2007; Suzuki et al., 2008; Guyeux et al., 2022). In addition to the classical Proto-Beijing and Asian Ancestral 1 to 3 lineages, a Thai research group recently described an "*Asia-Ancestral 4*" sublineage, and our own group described an "*Asia-Ancestral 5*" group of isolates in Japan (Ajawatanawong et al., 2019; Guyeux et al., 2022). The recent finding of the spreading of the Russian 1071-32 clone (that belongs to the *early ancient2* group) as well as circulation of a hyperlethal MDR clone 14717-15 in Far East, Russia (Vinogradova et al., 2021) demonstrate that ancestral L2 strains can still be responsible of MDR-TB cases spreading in Europe (Mokrousov et al., 2019, 2021). Other endemic L2 were also described in Korea and in Japan (Han et al., 2015; Mokrousov et al., 2019). The *Asia Ancestral 5* that is specific to Japan, could have been endemic to the island before or at the turn of Christian era (Guyeux et al., 2022).

4.4 Recent Expansion of the L2 Phylogenetical Tree

The genomic diversity of L2 is still expanding: the recent study by the aforementioned Thai group provides an exhaustive description of today's L2 global population structure (Thawornwattana et al., 2021). In this recent review, more than 30 genotypes are described based on a collection of more than 4000 L2 genomes (Thawornwattana et al., 2021). Our own research group showed that *Asia-Ancestral 5* (AAnc5) was identical to the former "G3" group described by Wada et al. in Kobe and Osaka (Wada et al., 2012), and is also identical to the L2.2A as designated in the Thai researchers review (Guyeux et al., 2022; Thawornwattana et al., 2021). In another study, Freschi et al. describe a new L2.2.1.2 (deeply branching lineage, that is indeed characterized by an SNP at position 649844; the majority of these isolates came from Korea or China, and Asian immigrants in United States and British Columbia, Canada). This branch would have emerged after AAnc5 but before AAnc2 (Fig. 27.7).

We could have spent much more lanes describing the importance of L2 for the global issue of tuberculosis control, in particular in relation to the growing threat of MDR-and XDR-TB due to L2 in China and Russia (Ding et al., 2017); the issue of L2 evolution and of its early history poses important challenges for the future (Ajawatanawong et al., 2019; Thawornwattana et al., 2021). Recently, the W148 modern L2 clone was the focus of a specific study using 720 genomes from 23 countries; it perfectly links bacterial genetics evolution with human history, suggesting that these clones emerged in the early 1960s in Central Asia, and progressively acquired compensatory mutations (*rpoA*, *rpoC*) and IS*6110*-linked deletion of important genes (*ppe38*) responsible of a hypervirulent phenotype (Merker et al., 2022).

5. Lineage 3: The Delhi or Central-Asian (CAS) Lineage According to Filliol et al. (2002) or "East-African Indian" According to Gagneux et al. (2006)

5.1 Lineage 3 Discovery History and Specific Characteristics

The L3 lineage was discovered almost simultaneously in 2002 (Filliol et al., 2002; Vijaya-Bhanu et al., 2002). As its name indicates it was first described in New Delhi by IS*6110*-RFLP clustering where it represented 75% of the isolates studied there and showed a typical spoligotyping pattern, later identified as SIT26 (Filliol et al., 2002; Vijaya-Bhanu et al., 2002). Soon after, another L3 sublineage was

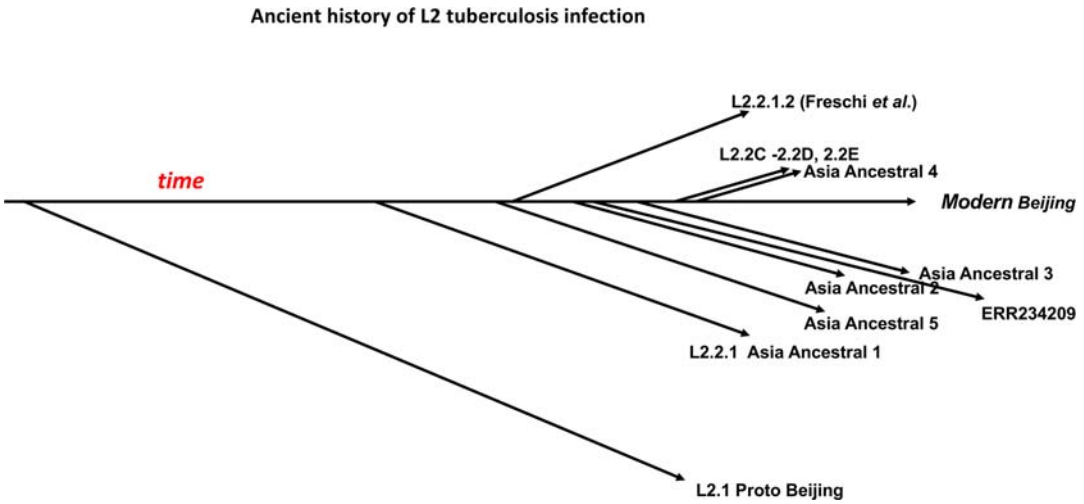


FIGURE 27.7

TB-Annotator–based evolutionary history of L2 based on published and unpublished results (Freschi et al., 2021; Guyeux et al., 2022; Luo et al., 2015; Mestre et al., 2011; Thawornwattana et al., 2021). Nomenclature is done according to Freschi et al. (2021) for L2.2.1.2, and according to Thawornwattana et al. (2021) for L2.2C, L2.2D, and L2.2.E, and according to Guyeux et al. (2022) for Asia ancestral 5.

described in Tanzania (Kilimanjaro region), associated to SIT21 (McHugh et al., 2005). All L3 strains were shown to be characterized by the absence of RD750 with loss of Rv1519 (Gagneux et al., 2006). Another recently described RD, RD316, leads to the loss of Rv3516 and Rv3517 genes (Bespiatykh et al., 2021). We confirm the loss of Rv3517 in L3 using the *TB-Annotator* v2.3 pipeline on 7200 genomes among the 102,001 recovered from the NCBI (Senelle et al., 2023c).

5.2 Description of Main L3 Sublineages

Coll et al. defined five sublineages inside L3, a definition that was extended by Napier et al. to seven sublineages (Coll et al., 2014b; Napier et al., 2020). L3.1 defines the L3 Kilimanjaro (broad sense), harboring mainly SIT21, including isolates harboring SIT288 or SIT1765 among others (Napier et al., 2020). SIT26 may also sometimes belong to L3.1. L3.1.1 is found mainly in Africa and defines the Kilimanjaro L3 sublineage (strict sense) harboring SIT21 or derived types such as SIT2267. L3.2 is a specific L3 sublineage that shows a different geographic distribution, since it is found mainly in Pakistan and India, where SIT1950 and SIT357 are prototypic signatures (Napier et al., 2020). Another study described five clonal complexes plus one basal group and presents a geographical map of their worldwide distribution, linking 24 VNTR results to WGS results (Shuaib et al., 2022).

5.3 Phylogeography of L3 and Comparison to L1

Among the most recent studies, one was performed by Menardo et al. on more than 2000 genomes; it confirms the structure of L3 and suggests a strong biogeographical specificity of local L3 sublineages

(Menardo et al., 2021). At least four independent events of introduction into East and North-East Africa are suggested (Shuaib et al., 2022). The introduction of these lineages in Europe, North America, and Australia is very likely to be linked to modern migratory movements (Shuaib et al., 2022). In India, L3 shows a strong association to the North of India, whereas L1 is linked to South-India (Couvin et al., 2019b). The full spreading history of this lineage remains to be reconstructed, however it could be associated to Islamic merchants history in the central Asian world. If older than that, it could also be linked to the Indus Civilization, which was the center of *Bos indicus* domestication and an early urban civilization (Pitt et al., 2019). L3 shares some similar and some different biogeographical spreading characteristics compared to L1 history, but it is very likely to be more recent. In addition, one striking feature of L1 is its shore-linked distribution whereas L3 appears to be more continental. This raises a question: could L1 have spread earlier and by maritime trade routes from East to West, whereas L3 would have spread later and by terrestrial trade routes mainly? It should also be pinpointed, as already mentioned, that L3 is sharing a common ancestor with L2 (Sola et al., 2003a), however this evolutionary relation should be studied in more detail with adequate genomes to understand more precisely how and when MRCA between the two lineages are found and where this change occurred (Fig. 27.8).

Due to a strong undersampling in India, central Asia, and North-East Africa, where these genomes were also found historically, e.g., in Sudan, the precise evolutionary history of L3 still remains to be written (Shuaib et al., 2018, 2020). Recently, however, a new study from the same authors, that gathers 2682 L3 clinical strains from 38 different countries, all typed by MLVA15-9, followed by a WGS study on a subset of strains, gathered with other recent results, confirms that L3 is likely to originate in South-Asia, with a single ancient population that spread to East-Africa and North-East Africa, with further evidence of at least four independent introduction into Africa (Shuaib et al., 2022).

6. Lineage 4: The By-default "T" Lineage (Sola et al. 1999) Renamed as "Euro-American" Lineage (Gagneux et al. 2006)

6.1 Lineage 4 Characterization and Discovery History

6.1.1 History of Ancient Versus Modern TB Principal Genetic Groups Naming

As early as 1997, Sreevatsan et al. split tuberculosis into three "Principal Genetic Groups" (PGG1, PGG2 and PGG3) according to a combination of two non synonymous SNPs found on two genes (Table 27.1): *katG* (ORF Rv1908c, codon 463, C/A at position 2154724) and *gyrA* (ORF Rv0006, codon 95, G/C at position 7545). PGG2 and 3 would represent "*modern*" tuberculosis (CC or GC), whereas PGG1 (CA) would represent "*ancestral*" MTBC (Sreevatsan et al., 1997). The fourth combination (GA) was not envisaged in this publication. Whereas PGG1 gathers L1, L5, L6, PGG2 gathers L2, L3, L4, and PGG3 contains only a subset of L4 (4.8 and 4.9 also designated as 4.10) Table 27.1.

6.1.2 History of the Definition of L4

The typical spoligotyping pattern of L4, consists in the simultaneous absence of the spacers number 33−36 in the 43 spacers nomenclature, or 43−44 and 46−47 in the 68 spacers nomenclature (Kamerbeek et al., 1997; van Embden et al., 2000). Such strains were first designated as "*T*" strains, by default, when the absence of these above cited spacers and only those, was observed (Sola et al., 2001b). In many isolates, other single spacer or block of spacer deletions were also detected (exemple:

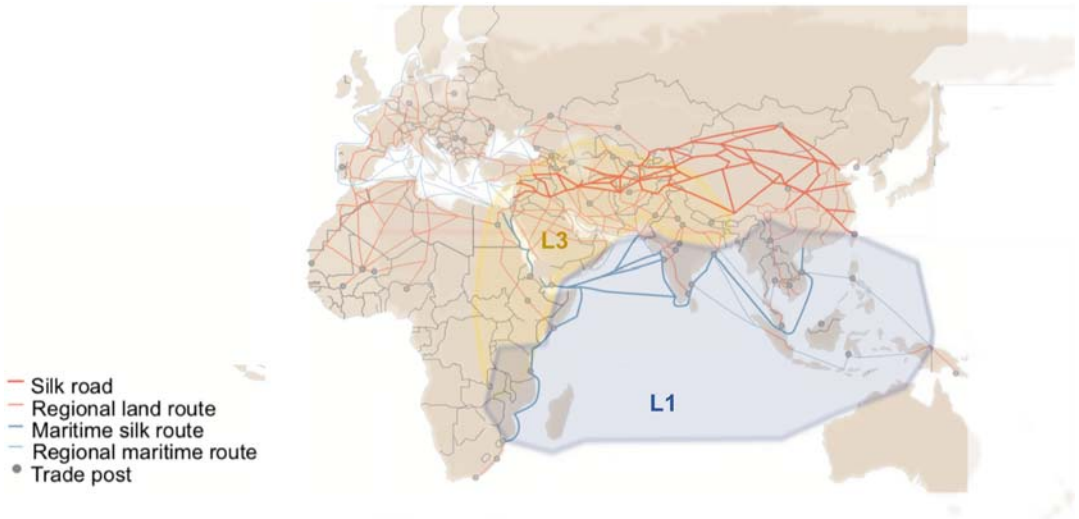


FIGURE 27.8 L1 and L3 geographical areas of distribution superimposed with historical land and see trade routes.

Superimposition of trade routes according to O'Neill et al. (2019) and of L1 and L3 geographical distribution according to Couvin et al. (2019a,b) suggests that L1 was historically spreading by see routes, whereas L3 was historically spreading by land routes.

Modified from Couvin, D., David, A., Zozio, T., Rastogi, N., 2019a. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the Mycobacterium tuberculosis *genotyping database. Infect. Genet. Evol. 72, 31–43. doi: 10.1016/j.meegid.2018.12.030; Couvin, D., Reynaud, Y., Rastogi, N., 2019b. Two tales: worldwide distribution of Central Asian (CAS) versus ancestral East-African Indian (EAI) lineages of* Mycobacterium tuberculosis *underlines a remarkable cleavage for phylogeographical, epidemiological and demographical characteristics. PLoS One 14, e0219706. doi: 10.1371/journal.pone.0219706; O'Neill, M.B., Shockey, A., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., Pepperell, C.S., 2019. Lineage specific histories of* Mycobacterium tuberculosis *dispersal in Africa and Eurasia. Mol. Ecol. 28, 3241–3256.doi: 10.1111/mec.15120.*

absence of spacer 18 for 4.1.1/X, absence of spacers 9–10 for 4.4/S, absence of spacers 21–24 for most L4.3/LAM) and these observations were linked to specific geographical settings and suggestive of phylogenetical links (Streicher et al., 2007). In 2006, the SpolDB4 database described 44 specific spoligotype signatures of phylogenetical relevance in four sublineages of *modern* strains, such as *X*/L4.1.1 (UK-IS*6110* low copy) (Dale et al., 2003), *Haarlem* (H/L4.1.2) (Kremer et al., 1999), *LAM* (Latin-American and Mediterranean/L4.3) (Sola et al., 2001c), *S*/L4.4 (Sicily) (Sola et al., 2001a), and many others (Brudey et al., 2006). All these sublineages were shown to be member of a single lineage, as guessed for "*T*," gathered and designated as "Euro-American" in a landmark paper (Gagneux et al., 2006).

6.1.3 Discovery of First SNPs to Classify L4

The first whole genome sequencing was obtained on an L4 strain, H37Rv, which remains the most famous L4 isolate, that belong to the PGG3 and was classified as L4.10 (Cole et al., 1998;

Table 27.1 Update of PGG Group Definition With Allele Combination Definition for Each Lineage (Assessed in TB-Annotator Database v2.4 Containing 113,193 Public SRAs Extracted From NCBI)

gyrA (Rv006) 7585	*katG* (Rv1908c) 2154724
C	A
G	C
CA	n = 53310 (L1,L2,L3,L5,L6) (PGG1)
GA	n = 6781 (L1,L2,L3,L5)
CC	n = 39426 (L4)
GC	n = 13626 (L4) (PGG3)

Modified from Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., Musser, J.M., 1997. Restricted structural gene polymorphism in the Mycobacterium tuberculosis *complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. U.S.A. 94, 9869–9874. doi: 10.1073/pnas.94.18.9869, based on a search on 113,193 SRAs, lineage appear by quantitative decreasing order in absolute numbers.*

Stücki et al., 2016). Then CDC1551 an L4.1.1/X strain was sequenced (Fleischmann et al., 2002). In 2006, in Texas, the group of J. Musser studied 5069 clinical isolates recovered from NYC, New Jersey, Houston, and Finland by a combination of methods (Gutacker et al., 2006). Fig. 27.9 from this chapter shows the strict link between SNP-based clustering in eight groups (I to VIII), PGG (groups 1, 2 and 3), and the spoligotyping signatures. The analysis suggested an important evolutionary transition at the point of IIA (Gutacker et al., 2006). In line with Fig. 27.9, it could be very interesting to search for isolates harboring both a SIT53 spoligotype and a PGG1 *kat*G-*gyr*A allele combination, which could represent an ancestral feature, since all extant SIT53 belong to PGG2 or to PGG3 (Fig. 27.9).

6.1.4 Current Classification of L4

In 2014, Coll et al. confirmed the existence of nine L4 sublineages (from 4.1 to 4.9) and modified the nomenclature, with up to five hierarchical splits. For example, strains originally described as SIT59/LAM11-Zimbabwe were later designated as L4.3.4.2.1 (Easterbrook et al., 2004; Heyderman et al., 1998). Some sublineages, such as 4.5 remained poorly sampled (Coll et al., 2014a,b). The L4 classification was slightly modified 2 years later and again recently (Stücki et al., 2016; Napier et al., 2020; Freschi et al., 2021).

6.1.5 Congruence and Divergence Between Spoligotypes and SNPs Results

Overall, sublineages previously defined by spoligotyping were confirmed by SNP-based analysis. However, in some cases, spoligotyping signatures were either less informative or no informative at all (such as the ubiquitous SIT53 type), convergent (homoplasic), or due to mixed infection (Mokrousov, 2021; Ouassa et al., 2012; Stavrum et al., 2009). Genetic convergence events (homoplasia) were

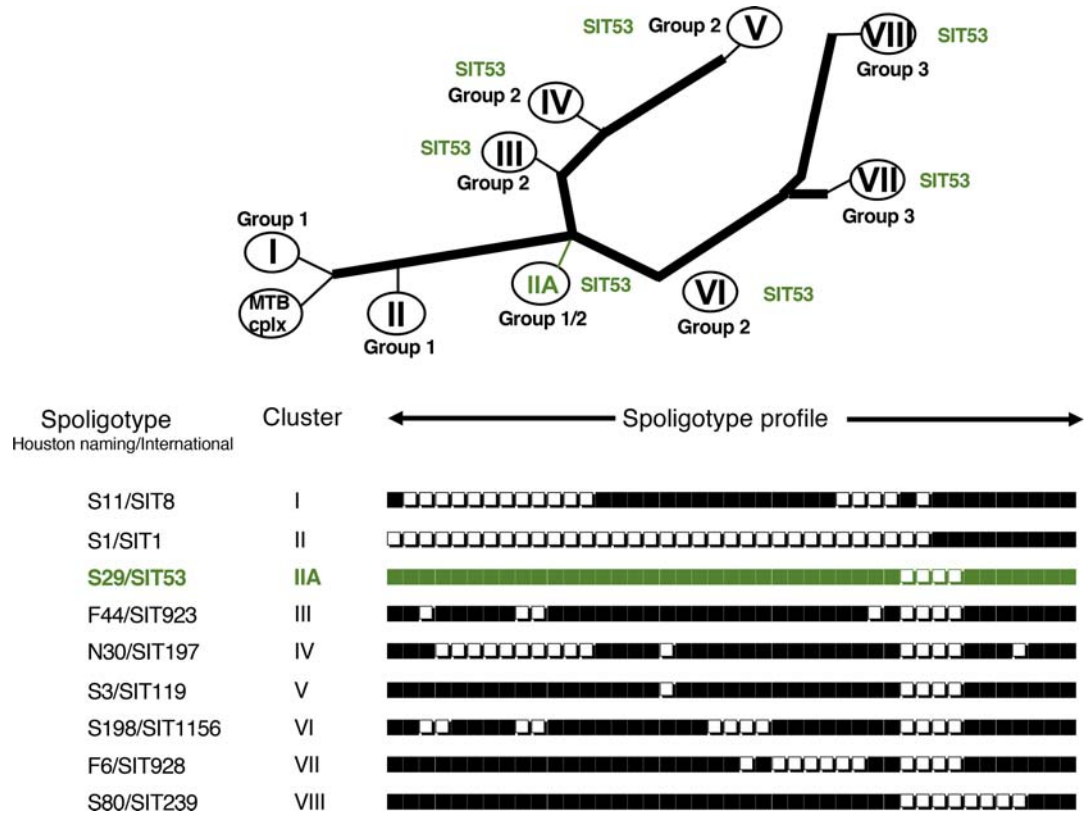


FIGURE 27.9

Phylogenetic tree of MTBC rebuilt from Gutacker et al. (2006) suggesting that SIT53 (*green*) harboring strains could be a missing-link between PGG1 and PGG2. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

From Gutacker, M.M., Mathema, B., Soini, H., Shashkina, E., Kreiswirth, B.N., Graviss, E.A., Musser, J.M., 2006. Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis *strains from 4 geographic sites. J. Infect. Dis. 193 (1), 121–128. https://doi.org/10.1086/498574.*

described for SIT4 in South-Africa; SIT4 could stem as an offspring of F11 (L4.3/LAM) or F28 (L4.4/S) lineages, and hence was revealing the ambiguity of some spoligotyping results (Warren et al., 2002a). The demonstration that the spoligotype could produce ambiguous results subsequently greatly limited the use of this technique in epidemiological studies. In these cases, studying the IS*6110* insertion sites could help in distinguishing between the F11 or the F28 lineages origin of an SIT4 isolate (Warren et al., 2002a) and combining CRISPR and IS*6110* insertion sites often allows to solve these phylogenetical issues. Spoligotyping is rarely efficient in indicating recent tuberculosis transmission but sometimes indicative of lineage (Napier et al., 2020, 2023). Often, spoligotyping signatures are linked to specific IS*6110* sites in the CRISPR (Guyeux et al., 2021; Refrégier et al., 2020) and

some clones continue to evolve on their CRISPR locus thus producing new SIT signatures (the SIT database is still maintained at the Institut Pasteur of Guadeloupe (Couvin et al., 2019a)). The pace of spoligotyping changes is slow compared to IS*6110* or VNTR changes (Meehan et al., 2018).

6.1.6 Updated Classification of L4

In 2016, based on 3366 sequenced MTBC L4 isolates from all over the world, the characterization of nine L4 sublineages defined by Coll. et al. was improved, and extended to 10 confirmed sublineages with minor changes: merging of L4.7, 4.8 and 4.9 into a single 4.10 sublineage (Principal Genetic Group 3) and the addition of two more sublineages (Stücki et al., 2016). New SNPs were added to the already long list of confirmed phylogenetic SNPs (Stücki et al., 2016). In Napier et al., new SNPs are added, and L4.7, L4.8, L4.9 were reintroduced instead of 4.10 with new SNPs-defined sub-sublineages (L4.8.1 to L4.8.3) (Napier et al., 2020). Freschi et al. recently added three more sublineages, 4.11 to 4.13 (Freschi et al., 2021) one among which, had already been described earlier with another name (e.g., 4.13 encompass the SIT284 spoligotype-previously defined *East-Med1 type*).

6.1.7 Example of Recent Input of WGS

WGS studies also allowed to cluster in the same sublineage some previously unrelated spoligotyping signatures e.g., T3-Osaka (SIT627) and TUR (SIT41) that were shown to both belong to the 4.2 lineage (Refregier et al., 2016). This is also true for the former T4_CEU1-derived (SIT39, T4 central European1 cluster) SIT1278, found in Nunavik, that belongs to the L4.1.2/Haarlem (Lee et al., 2015; Liu et al., 2018; Nguyen et al., 2003a). However, a link between the previous name, that had been given based on repetitive sequence-based markers and the new name, given based on WGS-based SNPs, is not always made, which sometimes causes a loss of knowledge.

6.1.8 Praise and Criticism of Today's Classification of L4

The classification of L4 now looks as among the most refined, the most complex and the most exhaustive to date, with recent new epidemiological clonal complex description and hierarchical numbering naming, as exemplified by a recent paper (Freschi et al., 2021). However, this current trend toward bioinformatical-based designation of clones is also creating difficulties to distinguish between local and global cluster identities, and genomic labs tend to adopt, without any prior discussion or consensus, some namings that are abstruse, even for reviewers (cf. Supplementary Material of (Freschi et al., 2021)). The bioinformatically driven designation of MTBC clones appears too complex, prone to errors (digital numbers are more difficult to copy and to memorize than alphabetical characters) and counterproductive. Indeed, when studying the global tuberculosis pandemia, we must keep in mind that a local TB outbreak is linked to a precise geographic origin within a given geographical territory and local naming if well used, i.e., without discrimination intention, could be preferred.

6.1.9 General Considerations About Clone Naming

A call for a consensus naming of epidemic MTBC clones and variants should be done, as was done for COVID-19 (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/). Keeping the current MLVA (Multiple Loci VNTR Analysis) 15-9 numbering of epidemic clones, as long as this technique remains in use, in parallel to WGS, could be a good transient alternative (Mokrousov et al., 2021), as long as reliable and full in silico MIRU-VNTR profile can be obtained, a critical bioinformatical issue

using Illumina reads for the time-being (Rajwani et al., 2018). WGS using MinION could be another helpful solution to get in silico MIRU-VNTR (Bainomugisa et al., 2018).

6.2 Description of the L4 Sublineages Characteristics (SNPs Variant and RD Content), With an Attempt to Define Their Phylogeographical Specificity

6.2.1 Sublineages Belonging to L4.1

L4.1 gathered three previously defined sublineages, L4.1.1 (also named *X*), L4.1.2 (*Haarlem, H*), and L4.1.3 (*Ghana*) (Coll et al., 2014b). In Napier et al., L4.1.1 is now split into three new sub-sublineages (L4.1.1.1 to L4.1.1.3), L4.1.2 into two new ones (L4.1.2.1 and L4.1.2.1.1). A new L4.1.4 is also added (Napier et al., 2020). This nomenclature is easier than the one used in (Freschi et al., 2021), moreover a link to in Silico spoligotype is made.

6.2.2 L4.1.1: The "X" "Anglo-Saxon" IS6110 Low-Copy Sublineage

History of discovery and naming: The "X" strains (L4.1.1) were identified simultaneously in many laboratories world-wide (Table 27.2). They were characterized by the combination of a low-IS*6110* copy number and a spoligotyping signature with the deletion of spacer 18, such as found in SIT119, SIT137, SIT92, which had been, respectively, designated as X1, X2, X3 (Brudey et al., 2006).

6.2.3 History of This Lineage and Characteristics

Whatever the geographical origin of the X family could be, geographical observations strongly suggest that a common point between South-Africa, Mexico, Barbados, and Guinea-Bissau, where these strains were first described, was the post-Columbus (after 1492 AD) history of slave trade and European colonization. Indeed, when in 2003, an exhaustive study was performed in London on more than 300 isolates, the prevalence of X strains was around 25% of all isolates (Dale et al., 2003). In this study, the X strains were shown to be associated to patients of European origin (Dale et al., 2003). The confirmation of the European ancient origin of L4, not to say on the very ancient nature of these

Table 27.2 Chronology of the Description of the "X" Spoligotype

Year of Description	SIT	No of Isolates	Country	Data Provider[a]
1992	119	1	ZAF	Streicher EM et al.
1992	119	1	GNB	Källenius G et al.
1992	119	4	MEX	Quitugua T et al.
1993	137	2	USA	Quitugua T et al.
1993	137	6	ZAF	Streicher EM et al.
1993	137	1	BRB	Rastogi N et al.
1992	92	3	ZAF	Streicher EM et al.
1992	92	1	GNB	Källenius G et al.
1993	92	9	ZAF	Streicher EM_et al.

[a]*From data extracted from SITVIT; http://www.pasteur-guadeloupe.fr:8081/SITVIT2/.*
Country given as ISO3166, data extracted from SITVIT database.

Table 27.3 L4.1.1 and Sublineage Signatures SNPs According to Napier et al. (2020), and as Found in TB-Annotator (v1; 15900 SRAs, and v2.3; 102,001 SRAs) SNP positions are given relatively to NC_000962.3 as H37Rv genome reference

Lineage Designation	SNP Position	Nt Change	AA Change	Gene Name	Rv Name	Left: (Napier's Results Into Brackets: Potential Barcoding SNPs, Robust SNP n°) n = TB-Annotator v1 Version: Number of SRAs, Into Brackets: n° of Exclusive SNPs, n° of SNP Present in 95% of Isolates), n′ = (TB-Annotator v2.3 Version)
L4.1.1	514245	C->T	359V	ctpH	Rv0425c	(30, 13) n = 253 (1,37), n′ = 2790
L4.1.1.1	1006080	G->A	161V	prrA	Rv0903c	(39, 2) n = 150 (4,50), n′ = 778
L4.1.1.2	1109535	G->A	88K	galU	Rv0993	(92, 2) n = 30 (107,30), n′ = 117
L4.1.1.3	4229087	C->T	247N	glfT1	Rv3782	(58, 3) n = 87 (45,58), n′ = 1400
L4.1.1.3.1	286300	C->A	184A	aftD	Rv0236c	(46, 3) n = 2 (115, 0), n′ = 122

From data extracted from TB-Profiler and from TB-Annotator Senelle, G., Guyeux, C., Refrégier, G., Sola, C., 2023c. TB-annotator: A Scalable Web Application That Allows In-depth Analysis of Very Large Datasets of Publicly Available Mycobacterium tuberculosis *Complex Genomes.https://biorxiv.org/cgi/content/short/2023.06.12.526393v1.*

isolates, points to the importance of these strains in the evolution and spread of tuberculosis. In 2011, Evans et al. showed a sustained transmission from 1995 till 2008 of an X1 clone, dubbed SIT490, harboring a strain with a higher IS*6110* copy number strain (7 copies); they named it the "*Mercia*" strain (Evans et al., 2011). *Mercia* was one of the kingdoms of the former Anglo-Saxon heptarchy, the collective name of the seven original kingdoms of England, that existed from the fifth to the 11th century (Davies, 1999). However, such ancient clones could even have a roman origin and may have been imported to United Kingdom earlier as suggested by (Eddy, 2015). Napier et al. defined characteristic SNPs within L4.1.1 sublineages as shown in Table 27.3.

Out of 350 SRAs tagged as L4.1.1.1 found in the drug resistant TB database from the LSHTM (accessed Sept. 9th, 2021) 221 (63%) harbored the SIT137; out of 83 SRAs tagged as L4.1.1.2, 51 (61%) harbored SIT119 and out of 555 SRAs tagged as L4.1.1.3, 167 (30%) harbored SIT92. Within the newly defined L4.1.1.3.1 sublineage, 75% were also harboring SIT119 (https://tbdr.lshtm.ac.uk/sra/browse).

6.2.4 L4.1.2: The "Haarlem" and Variant L4.1.2 Sublineages

The *Haarlem* strain (L4.1.2) is dubbed so because it created a cluster in the 90s together with other types of L4 in this town (van Soolingen et al., 1999), which is located at the West of Amsterdam in the Netherlands. Internationally acknowledged in 1999, *Haarlem* was rapidly shown to be widespread, and its spreading history is quite complex (Kremer et al., 1999). Its main prototypic spoligotype is SIT47 (notice that a variant that lost the entire left part of the CRISPR locus (SIT2) had already been described in SpolDB1 in Austria (Sola et al., 1999)). A specific SNP in the *mgtC* gene of *Haarlem* strains was described in 2006 (Alix et al., 2006). The L4.1.2.1 sublineage, also designated as the Hambourg 7199/99 strain, harbors a variant SIT45 spoligotype, that emerged in Hamburg probably around 1997, spread to Kiel and Pinneberg, and was exhaustively characterized by WGS in 2013 (Roetzer et al., 2013). Another variant of L4.1.2 is the L4.1.2.1.1 as defined by Napier et al., which is the "*M*" strain, found in Argentina, that was exhaustively characterized (Eldholm et al., 2015). The "*M*" outbreak isolates form a monophyletic clade separated by at least 80 SNPs from any other 4.1.2.1 isolates. This clone was characterized as an MDR outbreak that is found to circulate in Argentina since at least 40 years (Eldholm et al., 2015).

6.2.5 The Nunavik Sublineage

The Nunavik strain (SIT1258), is another 4.1.2 sublineage, described initially in 2003 that was found to represent a major historical cluster among the Inuits (Nguyen et al., 2003a). We previously suggested the transmission from Europe to Nunavik of SIT39, a potential common ancestor of this epidemic strain during or after the World War II (1952) based on further genotyping data contained in SITVIT2 and combined MIRU-VNTR-spoligotyping clustering results (Sola et al., 2005b). Later, based on WGS, this sublineage was found to represent 93% of all isolates in the Inuit population (Lee et al., 2015; Liu et al., 2018). The SIT1258 spoligotype is supposed to have emerged from a SIT39 ancestor, that was designated as T4-CEU (T4 central Europe) and was found to be endemic in Poland, Latvia, Czech Republic and Germany (Sola et al., 2005b). In their paper, on population genomics of TB among the Inuits, Lee et al. based on the observation that only L4 are currently circulating in Nunavik, refute the hypothesis of an introduction of tuberculosis via the Bering straits, several thousand years ago. Indeed, historical observations of previous contact with European populations point to a single introduction of an ancestral clone between the 17th and the 20th century (Lee et al., 2015). Out of the SNPs described in the Supplementary material of Lee et al., the 891756 position L4.1.2-defining SNP is present, together with the 491591 position in all "Major" Inuit population clusters; however the other SNPs mentioned in Fig. 27.10 are absent (Lee et al., 2015).

6.2.6 WGS Input on the "Haarlem" Lineage

Haarlem would deserve an entire book chapter on its own given the complexity and presumed ancient population structure. Fig. 27.10 shows some of the important SNPs that allow to define *Haarlem*. Between 30 and 50 individual historical and epidemiological clonal complex at least were described, whose history could be investigated in a near future, in order to try to reconstitute a more detailed spreading history of this lineage, in Europe and elsewhere (Fig. 27.10).

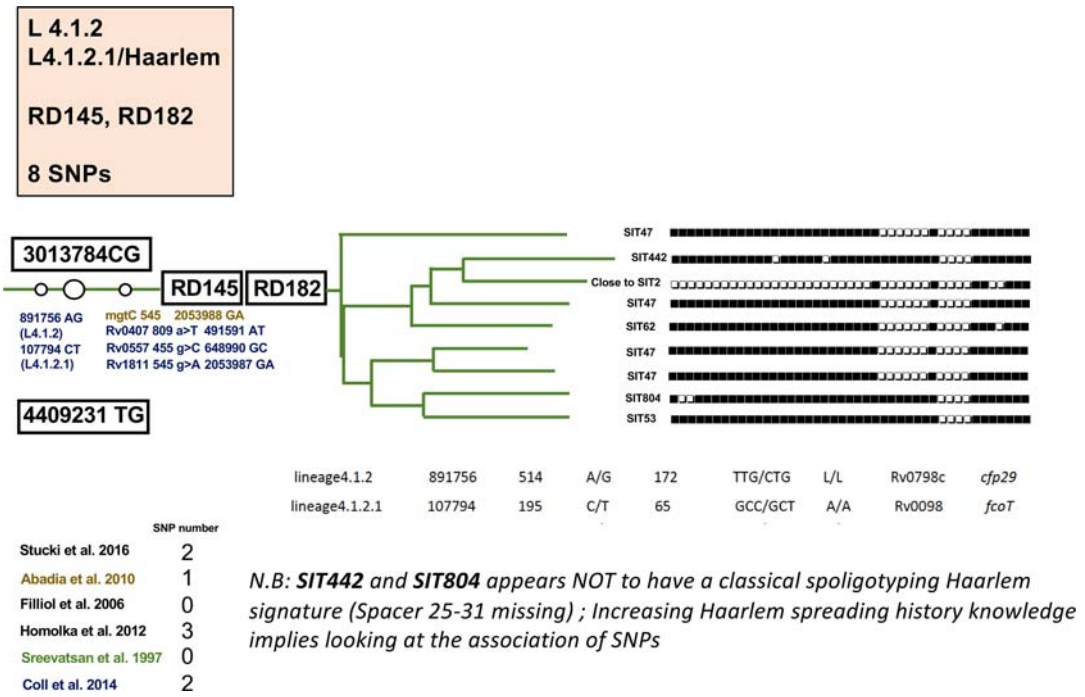


FIGURE 27.10 The L4.1.2 phylogenetical branch, as found in 2016, with some characteristic subtypes and a panel of phylogenetical SNPs with their original sources.

Modified From Stücki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., Fenner, L., Rutaihwa, L., Borrell, S., Luo, T., Gao, Q., Kato-Maeda, M., Ballif, M., Egger, M., Macedo, R., Mardassi, H., Moreno, M., Vilanova, G.T., Fyfe, J., et al., 2016. Mycobacterium tuberculosis *lineage 4 comprises globally distributed and geographically restricted sublineages. Nat. Genet. 48 (12), 1535–1543. https://doi.org/10.1038/ng.3704.*

6.2.7 L4.1.3: The "Ghana" Sublineage

The "*Ghana*" lineage was poorly defined until recently because it harbors a poorly informative SIT53 (T1) signature or poorly defined variants (SIT44, 65, 86, 118, 373, 462, 504), however this clone became more evident based on MIRU-VNTR results (Senghore et al., 2020; Yeboah-Manu et al., 2011). In 2009, a specific missense SNP was described (Rv2629, 2956731, 965 C/T) (Homolka et al., 2009, 2012). Another specific SNP, in perfect congruence with the former, was added 4 years later (Rv3921c, 4409231, T/G). In Napier et al. (2020), one more SNP was added (Rv1389, 1564799, C/T). The "Ghana" sublineage generally harbors a SIT53 (in 45 out of 58 isolates), which is known to be an ancestral ubiquitous type that was at the birth of many L4 sublineages.

6.2.8 Phylogeography and SNP Definition of L4.1.3

L4.1.3 is reported in at least 10 countries (Napier et al., 2020). L4.1.3 can be defined on the basis of the combination of 3 SNPs at positions 2956731, 4409231, 1564799. 64 SRAs belong to L4.1.3 in the *TB-Annotator* v1 version (March 2022). About 178 SNPs were found to be common in a set of five L4.1.3 isolates (Napier et al., 2020); our own recent results on 15,901 genomes confirm 117 fully exclusive SNPs and 95 exclusive in more than 95% of isolates (results not shown). One of the SNPs described as "specific to a 4.1 sublineage" and present in the drug-resistance related gene *gidB* at position 4407967, is not found in all L4.1.3, it appeared only recently and should not be considered as a phylogenetic marker of L4.1.3 (Gygli et al., 2019). It was recently shown that the L4.1.3 lineage has become multi-drug resistant and is spreading in Mali (Senghore et al., 2020).

6.2.9 The New L4.1.4 Sublineage

In their recent study, Napier et al. in 2020 described a new sublineage inside L4.1, L4.1.4. This group of isolates is characterized by a synonymous SNP in Rv0054 (at position 58786 g->c), i.e., in the *ssb* gene, a single-strand binding protein Ssb (helix-destabilizing protein) involved in the replication of the chromosome. However, the precise geographical origin of this cluster is unknown for the time-being; it was found in seven countries, including Peru and seems to be often associated to SIT450 or a derived type (SIT2511) by spoligotyping (77% in the LSHTM[8] database). All these types are characterized by the absence of spacer 18, and SIT450 could be an offspring of SIT878, found mainly in Eastern Europe (Bulgaria, Turkey, Serbia, Albania) but also found in Iraq, Egypt and Lybia according to the SITVIT database. It is out of the scope of this book chapter to describe new lineages, however it is clear that other historical and specific new lineages will be defined in the future inside L4.1.

6.2.10 L4.2: The "Ural" Sublineage

Ural refers to a mountainous geographical region, between the East European and the West Siberian plains. It is in the vicinity of the Eurasian steppe, that extends from the Arctic Ocean to the end of the Ural River. Its history is the one of the nomadic horsemen civilizations of the Mongols, and of the Russian empire expansion. The first mention of an *Ural* MTBC lineage is due to Kovalev et al. who studied the isolates present in a little town of 134,000 inhabitants, Pervouralsk (Kovalev et al., 2005). The definition was made initially thanks to MIRU-VNTR typing and not spoligotyping (Kovalev et al., 2005). Many different spoligotyping signatures worth of interest can be found in L4.2. The definition of the Ural lineage given by Mokrousov et al. in 2012 mentioned a specific spoligotyping signature with the deletion of spacers 29-31 (SIT127), that had wrongly been designated as "Haarlem 4" (H4) in SpolDB4 (Mokrousov, 2012; Brudey et al., 2006). SIT127, found in Iran, was however later shown to be part of another lineage, Lineage 4.5 that is not 4.1 nor 4.2. Conversely, SIT777 and SIT262, found in Russia, belong to L4.2. The geographic prevalence and difference between these close signatures is shown in Fig. 27.11. Spoligotype homoplasy was demonstrated on the examples of these quasi-related spoligotypes that belong to different phylogenetic branches as shown later.

SIT127 and SIT777 differ only by the presence/absence of spacer 2, however this tiny difference is important phylogenetically; hence, the *Ural* naming should only concern *Ural1*; it is not known if this

[8]London School of Hygiene and Tropical Medicine.

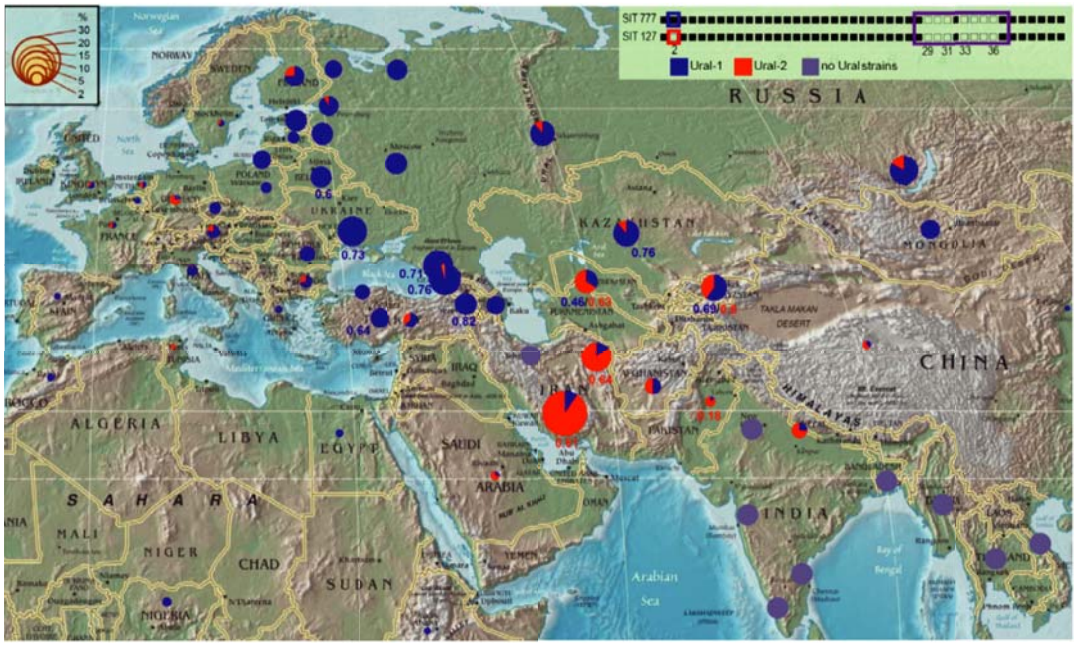


FIGURE 27.11

Different geographical spread of SIT127 in red (former "*Ural2*," part of L.4.5) as compared to SIT777 in blue (former "*Ural1*" part of L.4.2) within the Ural lineage. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

From Mokrousov, I., 2012. Infect. Genet. Evol. 12 (4), 619–629.

slight difference between these two highly similar spoligotypes appeared before or after the loss of spacer 29–31 in a potentially common ancestor (Mokrousov et al., 2017). Whatever, this branch is among the first to have diverged inside L4. Other spoligotypes with derived signatures were described by Dubiley et al. in 2007, i.e., SIT35, 1447, 1134, 262, 1292. Some of these strains are also characterized by an IS*6110* copy at position 3529109 that defines L4.2.1, in association with a number of SNPs. Hence, L4.2 encompasses the *Ural1* spoligotype family commonly found in the Pontic region north of the Black Sea. Interestingly, The TUR (*Turkish*) sublineage that is the closest relative of *Ural*, both in VNTR and WGS based trees, is phylogeographically specific to the region South of the Black Sea. This implies a common ancestor of the Ural/TUR and their origin could have happened in the Circum-Pontic area, i.e., around the Black sea littoral.

6.2.11 WGS of Ural and Structure of L4.2

In their original *barcode* paper of 2014, Coll et al. split the L4.2 sublineage in two major branches, 4.2.1 and 4.2.2; SNP position 2411730 was used to define L4.2 and SNP positions 783601 and 1487796 to define, respectively, 4.2.1 and 4.2.2. The choice of SNP 2411730 provided false-positive assignation of some isolates. In Napier et al. the definition of L4.2 is now made with SNP position

Table 27.4 SNPs Used to Define Three Hierarchical Levels Inside 4.2 With Different Branches Harboring Similar Names Depending on Authors (SRA n°* as found in TB-Annotator v1, SRA n° as found in TB-Annotator v2.3)**

Lineage	Position*	Allele Change	Mutation Type	Reference	SRA n°*	SRA n°**
4.2	1466779	C/T	Synonymous	Napier	353	2972
4.2	2411730	G/C	Synonymous	Coll	366	3022
4.2	82437	GT/G	Frameshift variant	Senelle	371	3039
4.2.1	783601	A/C	Synonymous	Coll	236	1805
4.2.1	783601	A/C	Synonymous	Napier	236	1805
4.2.2	353766	T/C	Missense	Napier	113	1155
4.2.2	1237536	C/G	Synonymous	Freschi	352	2971
4.2.2	1487796	C/A	Synonymous	Coll	114	1160
4.2.2.1	353766	T/C	Missense	Freschi	113	1155

From Napier, G., Campino, S., Merid, Y., Abebe, M., Woldeamanuel, Y., Aseffa, A., Hibberd, M.L., Phelan, J., Clark, T.G., 2020. Robust barcoding and identification of Mycobacterium tuberculosis *lineages for epidemiological and clinical studies. Genome Med. 12, 114. doi: 10.1186/s13073-020-00817-3, Freschi, L., Vargas, R., Jr., Husain, A., Kamal, S.M.M., Skrahina, A., Tahseen, S., Ismail, N., Barbova, A., Niemann, S., Cirillo, D.M., Dean, A.S., Zignol, M., Farhat, M.R., 2021. Population structure, biogeography and transmissibility of* Mycobacterium tuberculosis. *Nat. Commun. 12, 6099. doi: 10.1038/s41467-021-26248-1 and our own studies using TB-Annotator.*

1466779. SIT41/TUR now corresponds to the L4.2.2.1 *turkish* sublineage (Coll et al., 2014a). L4.2.2.2 is defined in Napier et al. (2020) and in Freschi et al. (2021); however, two different branches are described under the same name. Whereas SNP position 783601 from (Freschi et al., 2021) selects 24 isolates in *TB-Annotator*, SNP position 611463 from (Napier et al., 2020) selects only six different isolates (results not shown); these two branches do not share any common characters. Hence, these two SNPs do not allow the same hierarchical level of phylogenetical informativity, and there remains confusion on the definition of 4.2.2 as shown in Table 27.4.

L4.2 also includes the Ethiopian-SIT149 cluster (Hermans et al., 1995). Another important sublineage inside L4.2, whose position remains uncertain in the recent WGS classification, is SIT627 that defines the T3-Osaka cluster (Millet et al., 2007). This cluster was previously characterized by a specific 12-loci MIRU-VNTR signature (MIT310) identical to the one found in L4.2.2.1 (Millet et al., 2007).

The number of specific SNPs in the L4.2 sublineage seems to be sufficiently high to estimate that this family is of ancient origin (4.2, 165 clade-specific SNPs; 4.2.1, 43 clade specific SNPs; 4.2.2, 34 clade-specific SNPs; 4.2.2.1, 97 specific SNPs). The very existence of the Ethiopian L4.2 SIT149 sublineage from one hand, and of the L4.2 SIT627 "T3-Osaka" from another hand, both being low-IS*6110* copy strains, may point to a common Central Asian—Altai homeland of L4.2 (Refrégier et al., 2016). It is known that L4.2 is also present in China (Wan, 2007). In Africa, SIT149 is the dominant L4.2, and it is found almost exclusively in Ethiopia (420/3071 isolates, i.e., 14% based on spoligotype analysis according to Tulu and Ameni (2018). L4.2 could have a coalescence date that

would be consistent with its arrival in association with the major North-South human migration and origin of the Ethio-semitic languages, around 3000 years ago (Comas et al., 2015). Dating the MRCA between SIT41, SIT149 and SIT627 is heavily dependent on SNP mutation rate (Refrégier et al., 2016). For the time-being such an estimation is difficult: taking into account the lowest mutation rate (that fits the MRCA of *M. tuberculosis* tree of 67,000 years old according to congruence with human evolution, i.e., (10^{-8} mutations/site/year) or an intermediate rate (4,8 10^{-8} mutations/site/year as found in Bos et al.) or even the fastest rate (5×10^{-7} mutation/site/year in a recent molecular clock study), has important consequences on dating: the MRCA of L4.2 could be either as ancient as 12,000 years old or as young as 370 years old if we choose one of these two extreme (fast and slow) mutation rates; it is likely that the truth lies in between these two limits (Bos et al., 2014; Menardo et al., 2019; Refrégier et al., 2016).

6.2.12 L4.3: The Latin American and Mediterranean (LAM) Sublineages

LAM was defined on the basis of missing spacer 21−24 and 33−36 (SIT42), and with previous observations that such patterns were more likely to be found in Mediterranean areas and/or in Latin America, than elsewhere in the world (Sola et al., 2001c). Published spoligotyping and VNTR studies also suggested a perfect linkage disequilibrium between some VNTR and spoligotyping results, i.e., an ETR-A[9] low copy number (=2) associated either, to the prototypic SIT42 spoligotyping signature or to a derived variant, confirmed by the spatial distribution of such genotypes (Kremer et al., 1999; Soini et al., 2000; Sola et al., 2001c).

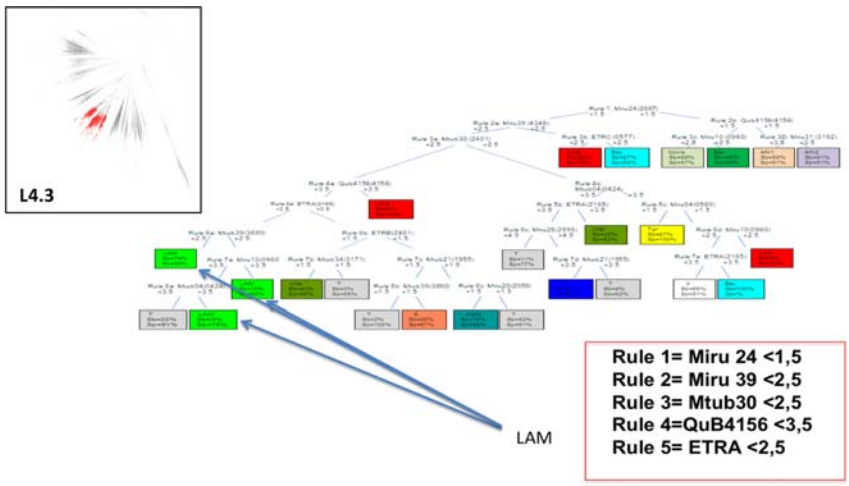
6.2.13 Characterization of LAM

Before the description of LAM, spoligotype patterns harboring the 21−24 spacer deletion in the CRISPR and other deletions, such as the "*Zimbabwean*" SIT59 found in the Harare population, or the SIT33 found in Brazil and in South Africa, had been described. This suggested a high diversity in LAM MTBC isolates (Filliol et al., 2002; Heyderman et al., 1998). In the SpolDB4 published in 2006, 12 LAM signatures were suggested however some were shown later to have been assigned wrongly to LAM due to homoplasia on spacer deletion 21−24; among those wrongly defined lineages, were LAM7-TUR (L4.2) or LAM10-CAM (L4.6). The absence of spacer 20−25 in LAM7-TUR or the absence of 23−25 in LAM10-CAM should be seen as an independent though convergent genetic event on loss of spacers 23−24, i.e., without any phylogenetic relationship, whereas the absence of spacer 21−24 was genuine in the other LAM strains. Other rules to define LAM could also be obtained using MIRU-VNTR loci such as the ones defined in Fig. 27.12.

6.2.14 Confirmation of LAM Lineage Identity Using SNPs

Using a pioneering molecular beacon SNP-based typing assay, Filliol et al. had assigned the LAM lineage to be part of the SNP Cluster Group (SCG) 5 (Filliol et al., 2006). However, the SCG5 also gathered unknown "T" spoligotypes (Filliol et al., 2006). In L4.3, one strain became famous, the F15/LAM4/KZN strain, which was responsible of an XDR outbreak in HIV patients in KwaZulu-Natal, South Africa, an outbreak that killed 52 out of 53 patients (Pillay and Sturm, 2007). Two specific LAM-defining deletion regions (RDs) had also been described in 2001 (RD115, RD174) and a third one RD761 was added later (Gagneux et al., 2006; Tsolaki et al., 2004). These deletions are important

[9]Exact Tandem Repeat, locus A, one of the VNTR loci formerly described.



Induction tree allowing assignation of isolates according to the SpolDB4 classification and based on their 24-VNTR profile, and built using the 2880 isolates with full 24-VNTR genotype and non-U SpolDB4 annotation, top left : L4.3 in TB-annotator, 16000 genomes

FIGURE 27.12 Machine learning–based induction tree and decision rules to assign L4.3 isolates based on 24 VNTR results.

Induction tree (obtained by Sipina, https://sipina-arbres-de-decision.blogspot.com) allowing LAM assignation of isolates according to the SpolDB4 classification, however, based on their 24-VNTR profile, using a collection of 2880 isolates with full 24-VNTR/spoligotyping-based SpolDB4 assignment. The five rules show how to establish hierarchically the identity of L4.3 based on VNTR results (these rules were not confronted to the WGS SNPs-based defining rules). *Top left*: L4.3 position in "TB-Annotator" RaXML phylogenetic tree, (TB-Annotator, 15901 SRAs, version 1, unpublished).

Data from a previous unpublished study, built using Sipina_W. (Kohavi et al., 1997).

markers from an evolutionary perspective, since they are Unique Event Polymorphisms (UEP). RDs allow discrimination within the "*Euro-American*" L4 all strains previously designated as "T" and classified either on the basis of absence of spacers 33–36 or on the basis of the *pks15/1* seven base pair deletion found in all the L4 strains (Gagneux et al., 2006). Rindi et al. confirmed that RD115 and RD174 deletions were mutually exclusive and were helpful to lineage (Rindi et al., 2012). These RDs were particularly helpful for uninterpretable spoligotyping signatures of LAM. This concerned many types among which the SIT1737 "Tuscany" type, the "LAM-RUS" strains, SIT222, which would not have been assigned to LAM without RD results (Dubiley et al., 2007; Rindi et al., 2012, 2014; Stücki et al., 2016).

6.2.15 More Precise Definition of LAM Using SNPs

SNPs were described quite early in L4.3 (Musser et al., 2000). In particular, two SNPs were found in *fbpC* (Rv0129c), including one in the promoter region. *FbpC* is one of the three genes that encode the Ag85C, a physiologically important enzyme that catalyzes the transfer of mycolic acid (MA) from

trehalose monomycolate to produce the mycolyl arabinogalactan (mAG) or trehalose dimycolate (TDM), which are complex lipids that define the protective mycomembrane of mycobacteria (Musser et al., 2000). Many investigators further confirmed the L4.3 specificity of the sSNP (G->A) at position 157292 in the gene *fbpC* (Chuang et al., 2010; Gibson et al., 2008; Homolka et al., 2012). A further study on 3R genes, allowed to discover the *ligB* (Rv3062, 1212C->G) polymorphism at position 3426795, a new L4.3 specific SNP that was confirmed on a larger sample (Abadia et al., 2010; Dos Vultos et al., 2008). The latest study allowed to reclassify into LAM two spoligotype signatures, SIT159 and SIT254 (previously designated as T5-RUS1) (Abadia et al., 2010). Another L4.3-specific SNP was found in the *mgtC* gene (Homolka et al., 2012). In 2016, Stücki et al. provided the first detailed picture of the diversity of the L4.3 Lineage as shown in Fig. 27.13 (Stücki et al., 2016). However, the definition of L4.3.1 and L4.3.2 was only achieved with a new set of 10 reference SNPs that now better discriminate within L4.3 (Napier et al., 2020) (see also Fig. 27.13).

An interesting *in-depth* study on the evolution of LAM and of LAM-RUS was published in 2016 based on a very large collection and analysis of 24 VNTR polymorphic data. This study anticipated some of the results obtained by Stücki et al. showing that the correlation of VNTR clustering and other evolutionary markers allows to define two main monophyletic lineages inside 4.3, an RD174/RD-Rio lineage (4.3.4) and an RD115 lineage (4.3.3) (Mokrousov et al., 2016).

6.2.16 A Particular Focus on L4.3.3 (RD115, LAM-RUS, or LAM-Russia)

The LAM-RUS was found to represent one-third of all TB cases in Tuva, Russia, in inmates (Ignatova et al., 2006; Shemyakin et al., 2004). LAM-RUS is linked to RD115 deletions and isolates found in Northern Eurasia (Mokrousov et al., 2016). However, the geographic distribution of RD115—L4.3 strains also extends to Oceania, and South and Central America (Mokrousov et al., 2016). It was shown that SIT251, 252, 254, 266, and 267 were *bona fide* LAM strains (Dubiley et al., 2007). One characteristic of the LAM-RUS is the presence of an IS*6110* in the *plcA* gene at position 2630571. Recently Gygli et al. suggested that a major phylogenetical marker in drug-resistance genes in a subset pf L4.3.3 was a G->T mutation at position 4242182 of the *embC* (Rv3793) gene (Gygli et al., 2019). Out of 32 isolates classified as L4.3.3 according to Napier et al., 17 harbored this SNP (Gygli et al., 2019). In *TB-Annotator* we confirm that the *embC* SNP allows to characterize a subset of L4.3.3 (n = 66 out of 427 L4.3.3), that is also found to be present in Brazil, Japan, Canada, Congo, and Uganda (Gygli et al., 2019; Napier et al., 2020).

6.2.17 A Particular Focus on L4.3.4 (RD174, LAM RD-Rio)

The second well-known L4.3 sublineage is the so-called "RD-Rio" described in 2007 (Lazzarini et al., 2007). Brazil used to perform spoligotyping at a high throughput during the 2000s; the brazilian investigators described isolates with previously defined LAM signatures (LAM9/SIT42, LAM6/SIT64, LAM5/SIT93) with and without RD-Rio deletion (Lazzarini et al., 2007). Some isolates were more specifically linked to one or to the other SIT, e.g., LAM3 were 100% wild-type (i.e., without RD-Rio), whereas LAM2 and LAM1 were 100% RD-Rio (Lazzarini et al., 2007). The RD-Rio deletion spans the RD174 (>23,6 kb) and results in the deletion of 10 genes, among which are some *pe-ppe* genes (Lazzarini et al., 2007). In a cohort from Belo Horizonte, it was further suggested, that patients infected by an RD-Rio strain had a more "complicated" TB case with higher cavitary TB and higher bacillary loads (Lazzarini et al., 2008). L4.3.4 were not only found in Rio but also in Russia and in Tibet (as sporadic isolates, see for example (Mokrousov et al., 2014)) or even in Kazakhstan where

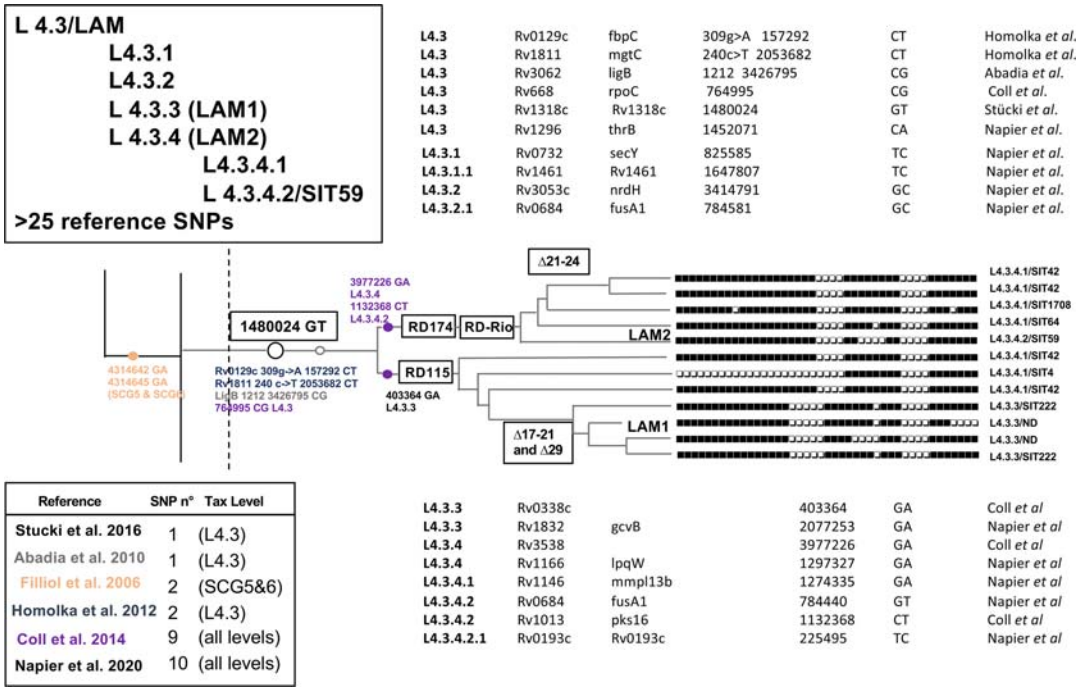


FIGURE 27.13

An updated classification of L4.3 showing the two main branches (L4.3.3 and L4.3.4) with SNP authors and SNP list (nonexhaustive) from left to right in the SNP table: SNP informativity level (lineage, sublineage) Rv number, gene name, SNP position relative to H37Rv (NC_000962.3), nucleotide change (wt/mut), author name. In the box: author name, year of publication, number of defining SNPs, hierarchical level of informativity of the SNPs described. Two important sublineages within LAM are the LAM-RUS and the LAM-Rio, since these two sublineages were studied *in depth* and are the current dominant ones.

Modified from Stücki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., Fenner, L., Rutaihwa, L., Borrell, S., Luo, T., Gao, Q., Kato-Maeda, M., Ballif, M., Egger, M., Macedo, R., Mardassi, H., Moreno, M., Vilanova, G.T., Fyfe, J., Globan, M., Thomas, J., Jamieson, F., Guthrie, J.L., Asante-Poku, A., Yeboah-Manu, D., Wampande, E., Ssengooba, W., Joloba, M., Boom, W.H., Basu, I., Bower, J., Saraiva, M., Vasconcellos, S.E., Suffys, P., Koch, A., Wilkinson, R., Gail-Bekker, L., Malla, B., Ley, S.D., Beck, H.P., de Jong, B.C., Toit, K., Sanchez-Padilla, E., Bonnet, M., Gil-Brusola, A., Frank, M., Penlap Beng, V.N., Eisenach, K., Alani, I., Ndung'u, P.W., Revathi, G., Gehre, F., Akter, S., Ntoumi, F., Stewart-Isherwood, L., Ntinginya, N.E., Rachow, A., Hoelscher, M., Cirillo, D.M., Skenders, G., Hoffner, S., Bakonyte, D., Stakenas, P., Diel, R., Crudu, V., Moldovan, O., Al-Hajoj, S., Otero, L., Barletta, F., Carter, E.J., Diero, L., Supply, P., Comas, I., Niemann, S., Gagneux, S., 2016. Mycobacterium tuberculosis *lineage 4 comprises globally distributed and geographically restricted sublineages. Nat. Genet. 48, 1535–1543. doi: 10.1038/ng.3704.*

they were detected recently (Skiba et al., 2019). Initially, such isolates were described in Portugal and South America (SIT20 or SIT17). The widespread "Lisboa3" and "Q1" strains as an example, belong to the L4.3.4 sublineage (Perdigão et al., 2019, 2020).

6.2.18 L4.4: The "S" Lineage and Its Sublineages

The "S" lineage was dubbed so because of Sicily where it was first described (Sola et al., 2001a). An original 106 Sicilian MTBC DNA collection spanning 5 years patient recruitment (1994−2000) and with only six non-italian citizens had been characterized by IS*6110*-RFLP (Nastasi and Mammina, 1999). Up to 10 percent of the samples were harboring a specific spoligotyping motif with deletion of spacers 9−10 and 33−36 that were designated by the SIT34, and tentatively designated as a new lineage (Sola et al., 1999). In a 2005 study performed in Sardinia, SIT34 isolates were also found with a high prevalence (11 out of 60, i.e., 18%) and a new Sardinian-specific variant of "S" was described, SIT1242 (Sola et al., 2005a). To the best of our knowledge, this variant is not yet sequenced. Independently, MTBC clonal clinical isolates harboring this SIT34 signature were reported in TB patients born in Quebec, Canada (143 out of 302 patients, i.e., 47%); a characteristic of this SIT34 subclade was a genomic deletion, called the $DS6^{Québec}$ (Nguyen et al., 2004). The $DS6^{Québec}$ deletion is approximately 11 kb long, from *plcD* (Rv1755c) to Rv1765c (Nguyen et al., 2003b, 2004). This deletion was found in 25% of all isolates characterized in Québec, however it is not present in strains from Europe suggesting that the expansion of this sublineage in Québec is post-Columbian. The spread of this strain to other provinces in Canada could be linked to the fur trade ongoing at the 18th century (Pepperell et al., 2011). The high prevalence of SIT34 strains in Québec could also be due to the amazing demographical history of Québec, which raised from 8500 settlers between the 17th and 18th century to six million of French-speaking Canadians (Laberge et al., 2005; Scriver, 2001). An Italian origin of these genotypes cannot be excluded since the MRCA of these isolates were first described in Sicily (Bonura et al., 2014). A recent paper suggests the introduction of $DS6^{Québec}$ MRCA in Canada in 1788 (95% HPD, 1739 to 1827) (Brynildsrud et al., 2018).

6.2.19 Finding of "S" in Africa and Oceania

In South Africa, the already mentioned « F28 » family, with an identical SIT34 spoligotyping signature was found in the 2000s (Warren et al., 2002b, 2002c). The finding of identical SIT34 signatures in America, Europe, Africa, and Oceania suggests an historical rooting of this family in Europe, with a spread history that could be linked to European colonization (Mulholland et al., 2019). At the end of the 18th century, tuberculosis was the cause of 1/5 to 1/4 deaths in London (Matossian, 1985). Another intriguing result is the finding of "S" isolates in Oceania, e.g., in the Kiribati islands and in New Zealand (Aleksic et al., 2013; Mulholland et al., 2019; Yen et al., 2013), the so-called "colonial-S types." The coincidence of James Cook's travels around the world and to Pacific, and the presence on board of his three expeditions of tuberculosis patients, could be the first event of introduction of tuberculosis from the crew into naïve populations from Oceania, and could explain how similar genotypes are found in America, South Africa, and Oceania (https://www.gutenberg.org/browse/authors/c#a2644). The "S" family was also unexpectedly detected in Yakutia, Russia, in visible prevalence in one study (Zhdanova et al., 2013) based on VNTR typing and confirmed by detection of the S-family specific SNP and spoligotyping. SIT102, another spoligotype, was shown to be part of L4.4 (Stücki et al., 2016).

6.2.20 Presence of L4.4 in China

L4.4 is also found in almost all regions of China, though at relative lower frequencies as compared with L4.2 and L4.5 (Liu et al., 2018). Globally, notwithstanding the spoligotype signatures that do include

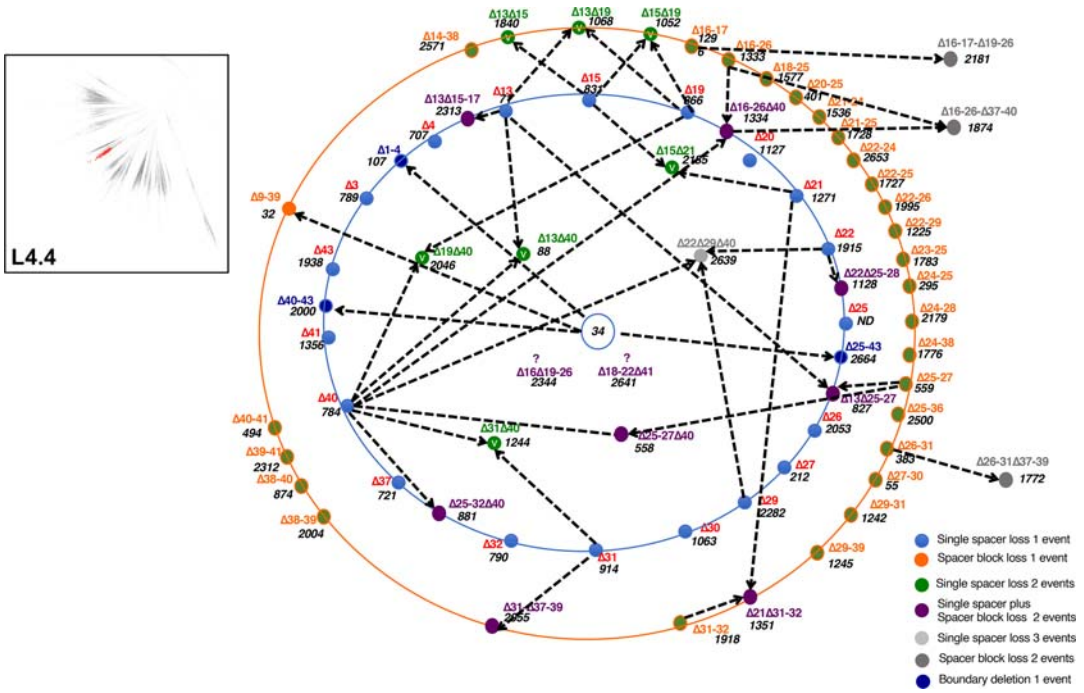


FIGURE 27.14 Variants of L4.4.1 spoligotypes (SITn°) built in a network around SIT34.

A landscape of L4.4.1 spoligotypes diversity and their potential evolution (C. Sola, unpublished results). SIT numbers are written in italicized characters. Evolutionary changes are suggested by *dotted arrows*; the first and second *circles* represent spoligotype types with single spacer loss (*blue*) or block of spacers loss (*orange* characters), which are presumed to be obtained by a single event change or multievent changes starting from SIT34 (*top left*: L4.4 branch location in *TB-Annotator* RaXML phylogenetic tree, version 1, built on 15901 SRAs). (For interpretation of the references to color in this figure legend, please refer online version of this title.)

absence of other spacers than spacer 9−10 (SIT102, SIT53, SIT52) the spoligotype subpattern landscape found within L4.4/S signatures is very large as shown in Fig. 27.14.

6.2.21 SNP Characterization in L4.4

At least 58 SNPs including the Rv0557 (*mgtA*, 648992) 457 C > G and Rv0291 (*mycP3*, 355181) G > A were found in the sublineage, L4.4.1.1 (Coll et al., 2014a; Homolka et al., 2012). L4.4 is efficiently defined by the SNP position 4307886 in Rv3834c (Coll et al., 2014b; Napier et al., 2020). Two sublineages, L4.4.1 and L4.4.2, are, respectively, defined by SNP position 2905505 and/or SNP 4151558 for L4.4.1.1 and SNP positions 985287 and SNP 4246508 for L4.4.2 (Coll et al., 2014b; Napier et al., 2020). There is no perfect congruence within this SNP set, for example, some genomes seem to harbor one but not the second (e.g., in L4.4.2, DRR034437 and DRR034483 from Japan harbor one but not the second defining SNP) which shows that improvements in SNP-based lineage definition are needed. L4.4.1.1 is a well-defined sublineage with SNP position 648992 being more robust than position 355181. It harbors SIT34 derived patterns.

L4.4.1.2 is less frequent but seems to harbor specific SIT, such as SIT1916 and SIT3481. This sublineage can be further identified based on 9 SNPs including a perfect congruence between SNP position 2694560 and SNP position 342201 (Coll et al., 2014b; Napier et al., 2020). It encompasses isolates that harbor SIT53 (e.g., that did not yet lost the SIT34 defining specific spacers). L4.4.2 is rare too, and can be defined by 10 SNPs; it is likely to represent more ancestral L4.4 strains, harboring the meaningless SIT53 or SIT52 spoligotype signatures. More studies are needed within L4.4 for better defining the previse correlation between SNPs and CRISPR markers.

6.2.22 Further Recent Results on the Colonial S Type

A recent paper described the "CS" lineages (Colonial-S-lineage) and tried to retrace the history of L4.4 in a New Zealand database of 236 genomes, that are de facto not representative of today's Europe. They suggested that two clonal complexes, respectively designated as "*Rangipo*" and "*Otara*," both belonging to the L4.4.1.1 lineage, could have emerged from European colonizers ("*likely French*") during the 18th−19th centuries (Mulholland et al., 2019). Whereas the "*Rangipo*" clone looks as a recently epidemiologically significant emerged strain (0−12 SNPs distance variation), the "*Otara*" clone looks more diverse (0−102 SNPs distance variation) and could include reactivation cases of a historically endemic and imported disease. Again, it is indeed remarkable that in Tahiti and in the Kiribati islands too, L4.4 is found to be very prevalent (Aleksic et al., 2013; Osman et al., 2017). Future studies on comparative genomics should allow retracing the spreading history of L4.4 more globally by correlating CRISPR-based and SNP-based evolutionary history, and may look for its origin.

6.2.23 L4.5: The "NEW 1" Lineage and Its Sublineages

SIT127 strains, that are now part of the L4.5 sublineage, were first described in Iran, where they represent up to 35% of all TB genotypes and initially thought to be similar to the Ural1 (L4.2.1) sublineage because these isolates share the absence of spacer 29−31 and differ only on spacer 2 (presence of spacer 2 in SIT777, absence in SIT127), and were initially wrongly designated as Ural2 as mentioned above in the L4.2 paragraph (Hadifar et al., 2019a; Haeili et al., 2013; Ramazanzadeh et al., 2006). This subtle difference on spacer 2 is phylogenetically very significant as already mentioned; by MIRU-VNTR typing these two group of strains were shown to harbor different signatures. After this initial confusion, SIT127 was renamed as "NEW 1" and was shown, in a phylogenetic tree built on 24 VNTR, to be close to L4.6 (Cameroon, Uganda) and distinct from L4.4 (Allix-Beguec et al., 2008). However, no renaming was done after this first naming, hence no geographical significatively was associated to "NEW-1" even if ancient papers showed this lineage was found to be prevalent in Afghanistan, Iran, and Pakistan (Hadifar et al., 2019a,b).

6.2.24 Molecular Characterization of "NEW-1"

An ETR-B copy number equal to 1 is found in L4.5, whereas it is quite rare in L4 and only found in some Uganda strains (Mokrousov et al., 2017). In 2004, a specific deletion region, RD122 has also been identified to describe these isolates (Tsolaki et al., 2004). According to Mokrousov et al., who reviewed 13 independent WGS studies totaling 5715 patients and studied 66 L4.5 isolates, the SNPs found in the evolution of this lineage, apart from the basic L4.5 defining SNPs (n = 8 at least), can be classified into three groups: those present in the "pre-NEW SIT53 node" (n = 15), those present in the "strict pre-NEW1 node" (n = 10) and those present in the "NEW1-node" (n = 3) (cf. Supplementary Table S3 of (Mokrousov et al., 2017)).

6.2.25 SNPs Definition of L4.5 or "NEW-1"

L4.5 is efficiently defined by SNP position 620029 (C/T), but also by SNP Position 1719757 (G/T) (Coll et al., 2014b; Napier et al., 2020). With minor variations, L4.5 can also be defined by combination of a C/T change in *fadE17* and a C/G variation in *vapC19* at position 2868976 (Senelle et al., 2023c). L4.5 is designated as "L4.2.1.1.2" in Freschi et al. and based on SNP position 4152890 it includes more strains that the other SNP-defining markers (Coll et al., 2014b; Freschi et al., 2021; Napier et al., 2020), in particular a set of four isolates from Russia and Finland that harbor a set of other specific SNPs (results not shown). Another SNP, SNP 4410301, is also shared by a subset (n = 101) of L4.5 in our current database (15901 genomes for TB-Annotator—version March 2022 version (Senelle et al., 2023c)).

We confirmed all the SNPs described by Mokrousov et al. in 2017except for one at position 2123146 on Rv1872c (results not shown), on, respectively, 90−93 genomes for the "pre-NEW SIT53" node, on 73−76 genomes for the "strict pre-NEW1 node" and on 57−58 genomes for the "NEW1-node." However, many genomes do not fit with L4.5.1 suggesting many other sublineages to be identified further (Senelle et al., 2023c).

6.2.26 Presumed Geographic Origin of L4.5

Based on a gradient frequency of L4.5 in Asia, Mokrousov et al. suggested that this sublineage originated in China 1000−1300 years ago, with migration from Xinjiang/Uyghur and arrival to Iran 800 years ago (Mokrousov et al., 2017). L4.5 lineage could be the only L4 lineage endemic to China. Recently, we confirmed what others had been observing for a couple of years: due to the Afghanistan war, to refugees migration, and to the more recent Syrian war, the 4.5 lineage also emerged in Lebanon and Syria, and could also emerge as a multidrug resistant clone (Massoud et al., 2020). In China, L4.5 is the almost exclusive L4 lineage found in Xinjiang/Uyghur, but it is also frequent in Yunnan, Sichuan and Shanxi (Liu et al., 2018). A recent study done in Kashgar prefecture in Xinjiang confirms these results since L4.5 represent up to 77% of all L4 in Kashgar (Xu et al., 2022) and Xinjiang is bordering Kyrgyzstan and Pakistan. A genomic study of DNA extracted from skeletons, found in 1996 from the Xinjiang/Uyghur town of Jiaoche, and showing bone lesions of tuberculosis, could provide clues on the presence of the L4.5 lineage in China 2000 years ago (Fusegawa et al., 2003). In Kyrgyzstan, that shares a border with China, L4.5 was described in a previous study (Mokrousov et al., 2013). It is likely that, with the increase in production of L4.5 genomes, the population structure and history of this lineage is going to unveiled in the future. The historical links between the silk road, the Islamic world and the presence of L3 and L4.5 in this region of the world remains to be studied in more detail.

6.2.27 L4.6: The Uganda (L4.6.1) and Cameroon (L4.6.2) Lineages

The Uganda (4.6.1) and Cameroon (4.6.2) strains were, respectively, designated and characterized 20 years ago in specific studies performed in Uganda and Cameroon, respectively, based on IS*6110*-RFLP and spoligotype clustering. Uganda I and Uganda II and Cameroon were shown to define epidemiologically important clones, respectively, in Uganda and in Cameroon (Niemann et al., 2002; Niobe-Eyangoh et al., 2003). They were later shown to belong to a common African sublineage of L4, designated as 4.6, as shown by Coll et al., with at least 16 diagnostic SNPs such as the G3466426A in the *lipY* gene, and the C3175335T in *vapC23* (see below in Table 27.5). These two sublineages were also rapidly shown to be prevalent beyond what their initial naming may have suggested, such as L4.6.1 in East Africa (Burundi, Kenya, Rwanda, Tanzania, Congo, Malawi, Central African Republic) and L4.6.2 in West Africa (Cameroon, Ghana, Nigeria). cf. Table 27.5.

Table 27.5 L4.6 SNPs and Sublineages as Defined by Coll et al. (2014); Napier et al. (2020) (C = Coll, N = Napier)

Lineage name	SNP position (NC000962-3)	nt	Rf/ mut	Codon	Triplet change	AA change	Rv	Gene name	Genome n° in TB-Annotator
lineage4.6-C	3466426	666	G/A	222	GTC/ GTT	V/V	Rv3097c	*lipY*	441
lineage4.6-N	18091		G/A	224		T	Rv0015c	*pknA*	436
lineage4.6.1-CN	4260268	879	G/C	293	GCC/ GCG	A/A	Rv3800c	*pks13*	70
lineage4.6.1.1-C	874787	555	G/A	185	CCG/ CCA	P/P	Rv0781	*ptrBa*	15
lineage4.6.1.1-N	4406749		G/A	261		L	Rv3918c	*parA*	16
lineage4.6.1.2-C	1501468	543	G/C	181	CCG/ CCC	P/P	Rv1332	–	54
lineage4.6.1.2-N	1098698		C/G	397		G	Rv0982	*mprB*	54
lineage4.6.2-C	4125058	642	G/C	214	CGG/ CGC	R/R	Rv3683	–	318
lineage4.6.2-N	4260742		G/A	135		P	Rv3800c	*pks13*	298
lineage4.6.2.1-C	3570528	684	C/G	228	CGG/ CGC	R/R	Rv3198c	*uvrD2*	9
lineage4.6.2.1-N	896119		C/T	100		F	Rv0803	*purL*	3
lineage4.6.2.2-CN	2875883	603	C/T	201	CTG/ CTA	L/L	Rv2555c	*alaS*	286
Lineage 4.6.3-N	734562		G/A	103		K	Rv0639	*nusG*	5
Lineage 4.6.4-N	4236903		G/A	375		A	Rv3790	*dprE1*	29
Lineage 4.6.5-N	17665		G/A	366		N	Rv0015c	*pknA*	17
lineage4.6-C	3466426	666	G/A	222	GTC/ GTT	V/V	Rv3097c	*lipY*	441

Data from Coll, F., McNerney, R., Guerra-Assuncao, J.A., Glynn, J.R., Perdigão, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., Clark, T.G., 2014a. A robust SNP barcode for typing Mycobacterium tuberculosis *complex strains. Nat. Commun. 5, 4812. doi: 10.1038/ncomms5812; Napier, G., Campino, S., Merid, Y., Abebe, M., Woldeamanuel, Y., Aseffa, A., Hibberd, M.L., Phelan, J., Clark, T.G., 2020. Robust barcoding and identification of* Mycobacterium tuberculosis *lineages for epidemiological and clinical studies. Genome Med. 12, 114. doi: 10.1186/s13073-020-00817-3 (C, N, or CN).*

The two first main sublineages, L4.6.1 (Uganda) and L4.6.2 (Cameroon), were shown to be substructured in many sub-sublineages and the structural landscape of L4.6 remains in progress with the recent addition of L4.6.3, L.4.6.4, and L4.6.5, found, among other in Ghana and Niger (Napier et al., 2020).

6.2.28 A Focus on the L4.6.1 "Uganda" Lineage

The "Uganda" lineage was first classified as *M. africanum* and designated as *M. africanum* II (Niemann et al., 2002). However, with further genetic evidence, it became obvious that this lineage should be included in *M. tuberculosis sensu stricto* (Sola et al., 2003b). RD724 was found to be specific to the L4.6.1 lineage and RD726 to the L4.6.2 Lineage (Gagneux et al., 2006). Uganda I (L4.6.1.1) and Uganda II (L4.6.1.2) can be further defined with different and non-overlapping spoligotyping signatures, e.g., SIT420 and SIT135 for L4.6.1.1 and SIT52, SIT317, and SIT73 for L4.6.1.2 (see also https://tbdr.lshtm.ac.uk/sra/browse). SIT316 was also found to be included in 4.6.1; all these strains have around 13−15 IS*6110* copies (Niemann et al., 2002). Specific MIRU-VNTR typing schemes were developed locally to try to detect recent transmission cases.

6.2.29 SNP Definition Within L4.6.1 Sublineages, Drug Resistance of L4.6.1

The SNP on *recR* 4160407 was found to be an event that appeared before the split between 4.6.1.1 and 4.6.1.2, whereas SNPs at position 990626 and 4260268 are also specific to L.4.6.1 (Stücki et al., 2016). These strains are dominant in Uganda and not widespread as compared to L4.6.2. Many of these isolates have become resistant to isoniazid (Asiimwe et al., 2008; Bazira et al., 2010; Kigozi et al., 2018) and L4.6.1 MDR strains are spreading in East-Africa (Kidenya et al., 2014). L4.6.1 is the most important MDR lineage in Uganda (Ssengooba et al., 2016).

6.2.30 Focus on the L4.6.2 Cameroon Lineage

In 2003, based on a combination of IS*6110*-RFLP and spoligotyping results on 455 clinical isolates, a group of similar strains representing up to 40% of all tuberculosis cases in Cameroon was described (Niobe-Eyangoh et al., 2003). A specific SIT61 signature was found with many variants and some (n°16, 26, 31 and 40) variable MIRU-VNTR loci (Niobe-Eyangoh et al., 2004). A specific deletion region, RD726 (1749bp) was also described by Gagneux et al. in 2006 and was verified as a reliable phylogenetical marker for the L4.6.2 (Gagneux et al., 2006). Specific SNPs were described by many investigators (Coll et al., 2014a; Homolka et al., 2012; Stücki et al., 2016). Since its initial description, new sublineages inside L4.6 have been described, as mentioned above. 4.6.3, 4.6.4, and 4.6.5 (cf. Table 27.5) (Napier et al., 2020).

6.2.31 Geographical Distribution of L4.6.2

Even if the L4.6.2 lineage was initially described in Cameroon, it was shown recently to be the most prevalent L4.6 sublineage in Nigeria (Molina-Moya et al., 2018). The history of L4.6 could be quite ancient and related to the history of agriculture-linked Bantu expansion toward the south of Africa with a cradle that could be found in Cameroon or in Nigeria, and an expansion time that could be in between 1000 BCE and 500 AD (Comas et al., 2015). However these authors also pointed another different hypothesis following the MTBC-70 "*out of Africa*" model (Comas et al., 2013): this sublineage, could be linked to previous human migration, a link provided by studying the Eurasian mitochondrial DNA

since the Haplogroups U6 and M1 were already present at the paleolithic in Africa. Their hypothesis is that L4.6 could have been in Africa since 20,000 years, which could be the coalescent point of this lineage (Comas et al., 2013). Given the rather low SNP number specific to that lineage, the former hypothesis (between 1000 BCE and 500 AD) would better fit with agriculture history in Africa as well as with Bantu expansion traces, however, the global paucity of these genomes in databases requires more investigation to argue more in favor of one or the other evolutionary scenario.

6.2.32 L4.7: The Congo Lineage

SIT144 was known since a longtime (already described in SpolDB2) with missing spacers 7 to 25 (in the classical 43 spacers nomenclature), however, L4.7, as such, was described only recently, as a new "*Congo*" lineage based on WGS (Fig. 27.15) (Malm et al., 2017). In this paper, no supplementary material allowed to do the correspondence between WGS and SIT signature, however, in Silico reconstruction using SpoTyping (Xia et al., 2016) on the SRAs provide as expected the SIT144 pattern, the specific spoligotype signature that had been described as early as in 1997 (isolate n°49 in Fig. 27.1 of one of the first publication on spoligotyping in 1997) (Goguet de la Salmoniere et al., 1997; Malm et al., 2017). This pattern had been found since the end of the nineties in immigrants from central Africa (former Zaire, today's democratic republic of Congo, the French Brazzaville Congo, and Gambia and already described in the SpolDB2,3,4 and SITVIT databases (Brudey et al., 2006; Demay et al., 2012; Filliol et al., 2002; Sola et al., 2001a). RD231 also allows to define L4.7 (Stücki et al., 2016) and superimpose to SIT144. RD231 represents the loss of 891 base pairs within Rv3448 (1404 bp) which is coding for a part of the ESX-4 type VII secretion system (Bespiatykh et al., 2021). Today, the definition of L4.7 according to Coll et al. (SNP position 4249732) or according to Napier et al. (SNP position 716918) provides different results that the ones obtained by Malm et al. (2017) (Fig. 27.15). Looking into the 16000 SRAs version of the *TB-Annotator* database, whereas the set of 23 L4.7 specific SNPs described by Malm is found in 48 isolates, the two specific SNPs mentioned by Coll et al. or by Napier et al. are found in 113 isolates, and only 12 SNPs are shared by these 113 isolates (results not shown). A more detailed structure of this lineage was published recently (Senelle et al., 2023a).

6.2.33 L4.8-4.9-4.10: The "H37Rv-Like" Lineages (Also Designated as 4.10)

The definition of these lineages corresponds to the former Principal Genetic Group 3 (based on katG-gyrA polymorphism) (Stücki et al., 2016). Coll et al. initially distinguished two sublineages, 4.8 and 4.9, whereas Stücki et al. gathered these two sublineages into a single new 4.10 lineage using RD219 as a definition criterium (Rindi et al., 2012; Stücki et al., 2016). The Coll et al. definition of 4.8 (SNP position 3836739 in Coll et al.) is now completed by SNP position 1130526 in Napier et al. (Coll et al., 2014b; Napier et al., 2020). This last one is the same SNP position that allows to initially define a hierarchical "L4.2.1.1.1.1.1.1.i1" in Freschi et al. (now renamed 4.10.i1) (Coll et al., 2014b; Freschi et al., 2021; Napier et al., 2020). However, these two SNPs are not optimal, a few isolates miss one or the other one and SNP 1227217 C/T missence mutation in the *fum* gene is better to define the L4.8 lineage (Senelle et al., 2023c).

There are no specific spoligotype signatures for this sublineage even if SIT53 is highly prevalent (1216/2310 in the *TB profiler* database). L4.8 is identical to Freschi et al. L4.10.i1 (Napier et al., 2020); the defining SNP is exclusive and incompatible with SNP 1130526 and defines L4.10.i2 (results not

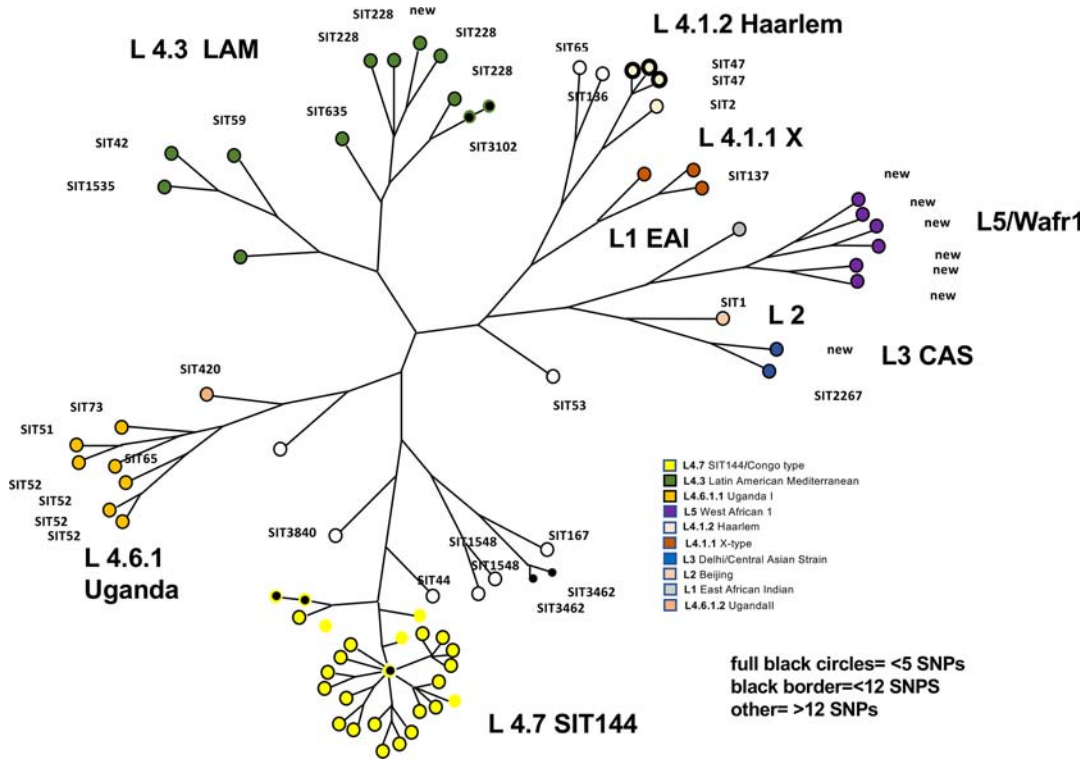


FIGURE 27.15 Discovery of L4.7.

WGS built phylogenetic tree redrawn from Malm et al. (2017) showing the description of the MTBC population structure in Congo, with emphasis on SIT144 (L4.7).

Modified from Malm, S., Linguissi, L.S., Tekwu, E.M., Vouvoungui, J.C., Kohl, T.A., Beckert, P., Sidibe, A., Rusch-Gerdes, S., Madzou-Laboum, I.K., Kwedi, S., Penlap Beng, V., Frank, M., Ntoumi, F., Niemann, S., 2017. New Mycobacterium tuberculosis *complex sublineage, Brazzaville, Congo. Emerg. Infect. Dis. 23, 423–429. doi: 10.3201/eid2303.160679.*

shown). In Napier et al. three more sublineages were described within L4.8, L4.8.1 to L4.8.3 based on SNP position 2914906 (L4.8.1 in United Kingdom, Malawi, and South Africa), 2417281 (L4.8.2, UK) and 616408 (L4.8.3, no more precision). None of these positions or of any of these new sublineages are mentioned in Freschi et al. (2021). Our own L4.8 collection suggests that there are new sublineages to be described, however, all these isolates represent recently emerged clones during new outbreaks, and as such are not deeply rooted in the MTBC phylogenetical tree (results not shown). Among the numerous L4.8 sublineages is the "C2" clone that emerged in Denmark and spread to Greenland (Folkvardsen et al., 2018).

The sublineage L4.10.i2, that can be defined by the absence of SNP position 1759252 in Freschi et al., (2021), corresponds to L4.9 when defined by Coll et al. (Coll et al., 2014b; Freschi et al., 2021). The same sublineage is defined by the absence of SNP 420008 with almost the same specificity (results not shown). The synonymous SNP 119600 in Rv102 allows to define a subfamily of these highly

Table 27.6 Lineage 4.12 In silico Reconstructed Spoligotype

run n°	Binary Spoligotyping	Octal
ERR161019		777777777740341
ERR176550		777777777740341
ERR181704		777777777740341
ERR181955		777777777540341

From Freschi, L., Vargas, R., Jr., Husain, A., Kamal, S.M.M., Skrahina, A., Tahseen, S., Ismail, N., Barbova, A., Niemann, S., Cirillo, D.M., Dean, A.S., Zignol, M., Farhat, M.R., 2021. Population structure, biogeography and transmissibility of Mycobacterium tuberculosis. *Nat. Commun. 12, 6099. doi: 10.1038/s41467-021-26248-1 with addition from the SITVIT database.*

related genomes (mean symmetric difference of 44 ± 34 SNPs) that are defined as L4.9.1 in Napier et al. and represent a single cluster.

6.2.34 L4.11 to 4.13 According to Freschi et al., (2021)

6.2.34.1 Brief Description of Lineage 4.11 (Named: 4.2.1.1.1.1.2 in Freschi et al., 2021)

This sublineage concerns a high number of isolates coming almost exclusively from Peru in the cited publication; these isolates do not harbor a specific spoligotyping signature (many harbor the meaningless SIT53). However, they can be defined based on the synonymous SNP A->G at position 801009 in *rplC*. This SNP could actually be specific to a sublineage within L4.11, since our own results on 113.193 SRAs suggest the existence of another phylogenetically close branch within this branch that is found in isolates from Bangladesh (results not shown).

6.2.34.2 Brief Description of Lineage 4.12 (Named: 4.2.1.1.1.1.2 in Freschi et al., 2021)

This sublineage concerns a very limited number of isolates (n = 6) from Malawi and South-Africa, belonging to Principal Genetic Group 2, that seems to harbor an unknown spoligotype signature as shown below in Table 27.6.

This sublineage can be efficiently described on the basis of one synonymous mutation at SNP position 1797555 G/A in the *nadC* gene. In the *TB-Profiler* informatics platform (https://tbdr.lshtm.ac.uk/sra/browse) these isolates are simply identified as L4, whereas they are described as a new sublineage in Freschi et al. (2021). In the latest version of *TB-Annotator* (V2.3, 102,001 SRAs), 12 isolates only from Malawi, India and the United States are found within this lineage.

6.2.34.3 Brief Description of Lineage 4.13 (Named: 4.2.1.2.1.2 in (Freschi et al., 2021)

This L4 sublineage was hierarchically defined as "4.2.1.2.1.2" and is now renamed as "4.13" (Freschi et al., 2021). All these isolates share the common SIT284, previously described and named as "*East-Med1*" in a study performed in Eastern Europe and Middle-East in 2005 (Panaiotov et al., 2005). This

confirms the uniqueness and phylogenetical significance of this spoligotyping signature pattern, and suggests that 4.13 encompasses SIT284 if not totally identical. In silico results using SpoTyping applied to the 4.13 genomes described in Freschi et al. indeed show that they all share SIT284 signature (03763777760771) or a derived one. These isolates can also be efficiently described using the SNP A/G at position 1063250, a synonymous variant in the *sucD* gene. This sublineage is also characterized by the "CC" combination that defines Principal Genetic Groups (Table 27.1).

7. Lineages 5 and 6: *Mycobacterium africanum* (Maf) West African 1 and 2

7.1 Focus on Lineage 5: *Mycobacterium africanum*, West African I

7.1.1 Introduction

The genomic undercharacterization of MTBC isolates originating from African TB patients, likely prone to belong to L5−L6 (*M. africanum*), is important when compared to that of patients originating from Asia, America, or Europe, which shows the poor interest of researchers for genomic research in Africa. Notwithstanding this limitation there are potential reasons that may explain the geographic restriction of *Maf* to West Africa. One practical explanation is the low recovery rate of *Maf* that is easily explained by difficulties in growing and isolating this species in routine clinical microbiology practice due to metabolic specificities (Gomgnimbou et al., 2012; Keating et al., 2005). Another parameter is the lower virulence of L6 in an animal model of infection (Ca et al., 2019); a third parameter could be linked to a process of co-evolution or co-divergence between *Maf* and some specific ethnic groups in the West African peninsula (Asante-Poku et al., 2016). *Maf* was first described by M. Castets in Dakar, as a subspecies within the MTBC, showing intermediate biochemical characters, i.e., similar to those of *M. tuberculosis* for some isolates and to those of *M. bovis* for some others (Castets et al., 1968). This phenotypic *Maf* peculiarity was well described in a French review (David et al., 1978) and more recently in de Jong et al. (2010) (Table 27.7). Meissner described also *Maf* isolates in Ghana in 1969 (Meissner and Schröder, 1969). The dysgonic colony phenotype appearing on culture plates was shown to be caused by a deficiency in glycolysis (in pyruvate metabolism) common to *M. bovis* and *M. africanum* and linked to a nonsynonymous mutation in L5 and L6 the *pykA* gene (position 1816848) (Keating et al., 2005). Genomes of L6 harbor a specific second mutation at position 1816587 of the same gene.

Although some *Maf* isolates have been described also in Eastern Africa, these new isolates were reclassified as a different Lineage, Lineage 9 by genomic analysis (Coscolla et al., 2021). Today, *Maf* is strictly applicable to isolates of West Africa. *Maf* of the Lineage 5 are named as West African I described by (Gagneux et al., 2006) and are found in countries around the Gulf of Guinea (Gehre et al., 2013), whereas Lineage 6 is more present in Mali, Senegal, Gambia, and Guinea. L5 represents up to 28% of all tuberculosis cases in Benin (Affolabi et al., 2009a) and may be the same in Nigeria (Molina-Moya et al., 2018) (Table 27.7).

7.1.2 Classification of Maf Using RDs and Spoligotyping

Unique event polymorphisms (UEP) can be used as robust phylogenetic markers to distinguish L5 (RD711) and L6 (RD702) (de Jong et al., 2010; Gagneux and Small, 2007). Members of L5 typically

Table 27.7 Comparative Characters of L5−L6 and other MTBC Lineages

Characteristic	*M. africanum* Type I		*M. tuberculosis*	*M. bovis*
	West African I	West African 2		
Morphological				
Colony appearance	Dysgonic	Dysgonic	Eugonic	Dysgonic
Depth of growth	Microaerophilic	Microaerophilic	Aerophilic	Microaerophilic
Biochemical				
Nitrate reductase	Negative to weekly positive	Negative to weekly positive	Present	Absent
Niacin production	Negative to weekly positive	Negative to weekly positive	Present	Absent
Drug susceptibility				
Pyrazinamide (PZA)	Sensitive	Sensitive	Sensitive	Resistant
Thiophene-2-carboxylic acid	Sensitive	Sensitive	Resistant	Sensitive
Molecular				
Large sequence polymorphisms	RD711	RD702	Variable	
Spoligotyping signature	Afri 2 and 3: Absence of sp8−12 and 37−39	Afri 1: Absence of sp7−9 and 39	Absence of sp34	Absence of sp 39−43
IS*6110*-RFLP	Average 10 bands	Average 6 bands	Variable	Low number

Adapted from de Jong, B.C., Antonio, M., Gagneux, S., 2010. Mycobacterium africanum*–review of an important cause of human tuberculosis in West Africa. PLoS Neglected Trop. Dis. 4, e744. doi: 10.1371/journal.pntd.0000744.*

show more than five IS*6110* copies and a spoligotype patterns lacking spacers 8−12 and 37−39 contrarily to L6 where only spacer 7−9 and 39 are absent (Brudey et al., 2004).

7.1.3 WGS Definition of Maf

L5 and L6 were first distinguished from the other lineages by the RD9 deletion spanning from 2330074 to 2332102 (2028 bp) with deletions from Rv2072c (CobL) to Rv2075c (Brosch et al., 2002). L5 is also clearly distinct from L6 through conservation of RD7, 8, 10 that are deleted in L6, together with animal strains (Brosch et al., 2002; Gagneux et al., 2006). *An in-depth* study into *M. africanum* L5 and L6 revealed 12 other regions of deletions in addition to RD702 and RD711 (Mostowy et al., 2004a). Ates et al. described an unexpected genomic diversity of L5 with RD711 being specific to the sublineage L5.1 (Ates et al., 2018a). Sublineages 5.1 and 5.2 can also be differentiated based on spoligotype signatures as shown in Table 27.8.

7.1.4 Importance of the Chronology of Loss of the Various RD

Apart from regions of difference (deletions), which are useful genomic markers to distinguish L5 and L6 from other lineages, another RD, TbD1 is an excellent phylogenetical marker that was lost in L2, L3, L4 but is still present in the other lineages. The TbD1, as mentioned in the L1 paragraph, is a 2153-bp genomic segment (Brosch et al., 2002; Smith et al., 2009; Bottai et al., 2020) first described by comparative genomic analyses using bacterial artificial chromosomal library (BAC) as well as hybridization and PCR screening (Brosch et al., 2002). Since L5 and L6 harbor TbD1, they belong to the group of "ancestral lineages" in the evolutionary scenario of MTBC, contrarily to lineages L2, L3, L4 which are characterized by the TbD1 deletion. The chronology of TbD1 loss relatively to RD9 loss is difficult to assess. On one hand, L1 and L7, L8 are likely to share a common ancestor, and on the other hand, L5 and L6 also share a common ancestor. However, when considering the four sublineages that all have an intact TbD1 (L1, L5, L6, L7), it remains difficult to precisely infer their evolutionary history Fig. 27.16.

7.1.5 Association of TbD1 to Virulence

Inspection of the TbD1 region showed that it corresponds to the *mmpS6* and *mmpL6* genes (Bottai et al., 2020). The MmpL is part of a large group of mycobacterial protein family belonging to the RND (Resistance, Nodulation, and Cell division) superfamily, whose members are involved in the diffusion through the mycobacterial cell wall of large molecules (such as various lipids and glycolipids). These proteins may play a role in virulence; isolates that have an intact TbD1 would be less resistant to oxidative stress and less virulent. By functional genomics in guinea pigs model, isolates harboring intact TbD1 have indeed been proven to possess a low virulence (Bottai et al., 2020; Domenech et al., 2005; Melly and Purdy, 2019; Székely and Cole, 2016). Hence, the success of MTBC as a more efficient pathogen could be linked to the loss of these sequences (Bottai et al., 2020) (Fig. 27.17).

There is a great interest to study L5 and L6 strain diversity from all over the African continent since the L5−L6 lineages evolution could be linked to the history of agriculture and the neolithization process in Guinea Gulf and West Africa and to the origin of L4, if we admit a 20,000 years old hypothesis of coalescence of L4.6 with a common MRCA to L1, L5, L6 (Comas et al., 2013). The L5 and L6 lineages appear together with L1/L7 regarding to TbD1 (Fig. 27.17). Ates et al. distinguish at least three L5 sublineages (L5.1.1, L5.1.2, and L5.1.3), and recently two more sublineages (L5.1.4 and L5.1.5) were added (Napier et al., 2020). In another ancient paper, based on spoligotyping signatures,

Table 27.8 Specific Signatures for L5.1 and L5.2 With SIT Numbers (Built Using Supplementary Table 1 From (Ates et al. 2018a)

Binary Spoligotyping	Isolate	lineage	SIT
	MAF3069	L5	NEW
	MAF3074	L5.1	101
	MAF3076	L5.2	NEW_USA_Ates
	MAF3079	L5.1	331
	NRC1	L5	NEW_USA_Ates
	NRC2	L5.2	new2
	NRC4	L5.2	NEW Close_1465
	NRC5	L5.1	331
	NRC6	L5.2	NEW Close_1465
	NRC7	L5.2	NEW Close_1465
	MAF30001	L6	181
	MAF30065	L5.1	NEW Close NGA Orp53
	MAF30066	L5.1	101
	MAF30068	L5.2	NEW Close_856
	MAF30072	L6	181
	MAF30077	L5.2	NEW Close_856

**Spoligo-International-Type , NEW: SIT unnamed; USA or NGA: indicates the geographical origin of the sample (United States of America, Nigeria), country name is in ISO-Code 3166−3, Orp:Orphan pattern, Close: similar to SIT*

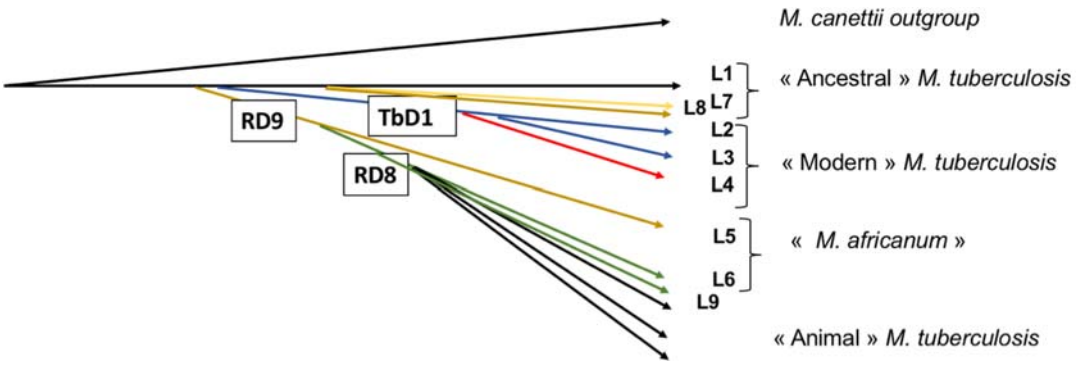


FIGURE 27.16

Current phylogeny of MTBC using region of differences.

Redrawn from Bottai, D., Frigui, W., Sayes, F., Di Luca, M., Spadoni, D., Pawlik, A., Zoppo, M., Orgeur, M., Khanna, V., Hardy, D., Mangenot, S., Barbe, V., Medigue, C., Ma, L., Bouchier, C., Tavanti, A., Larrouy-Maumus, G., Brosch, R., 2020. TbD1 deletion as a driver of the evolutionary success of modern epidemic Mycobacterium tuberculosis *lineages. Nat. Commun. 11, 684. doi: 10.1038/s41467-020-14508-5.*

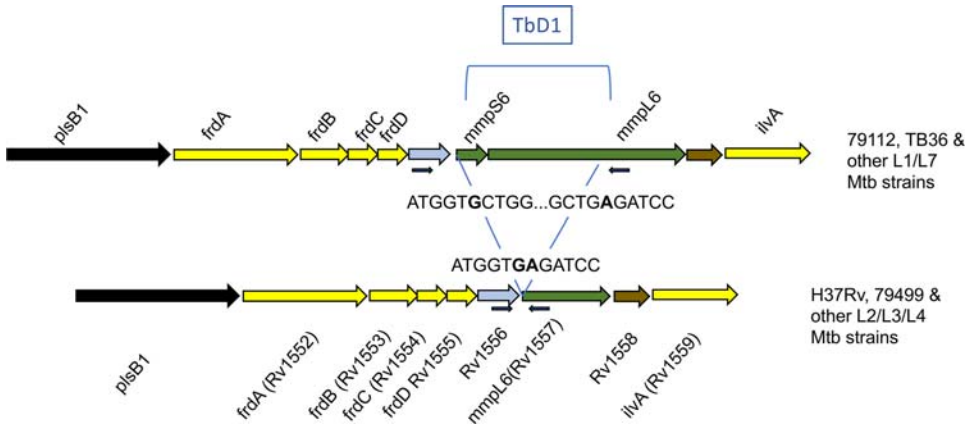


FIGURE 27.17

Intact and truncated region in *M. tuberculosis* H37Rv defining modern and ancestral lineages (Bottai et al., 2020).

10 L5 sublinages (L5.1 to L5.10) were suggested in a study done in Benin, Nigeria, and Sierra Leone (Gehre et al., 2013). Recent results to be reported elsewhere suggest that we are far from a comprehensive description of the population structure of L5 and L6 genomes.

7.1.6 SNP Characteristics of L5

A C/A mutation at position 1799921 was first described by Coll et al. to characterize L5, and confirmed recently by Napier et al. (Coll et al., 2014b; Napier et al., 2020). The C/T 2101921 mutation selects two more genomes (n = 361 instead of 359) and the combination of these two SNPs filter one more genome (n = 358). Using our own bioinformatic platform "*TB-Annotator*" we found that a C/T

synonymous variant in the *gyrA* gene at position 9566 is a better defining marker for L5 (n = 360 isolates out of 15901 sharing 133 exclusive markers). A detailed yet in progress analysis on the L5 diversity was published last year (Coscolla et al., 2021). Our own research group has recently reported new results on the complex taxonomy of L5, suggesting that a much higher number of sublineages can be found in the global population structure of L5 (Sahal et al., 2022, 2023).

7.2 Focus on Lineage 6: *Mycobacterium africanum*, West-African II

7.2.1 The Ecology of L6

M. africanum lineages (L5−L6) may be responsible for up to 40% of the TB cases in Benin (Gehre et al., 2013). *M. africanum* West African II (L6) shares partial identical features with L5 (loss of RD9) however specifically lost other RDs (RD7,8,10) (Brosch et al., 2002). The countries in which L6 is the most prevalent are Senegal, Gambia, and Guinea-Bissau (51% of L6 in 1999) (Källenius et al., 1999). Recently, in a study performed on 253 *Maf* genomes from Ghana, it was also suggested that L6 diversity was superior to the L5 diversity and that it was likely that these two lineages were sharing distinct ecological niches (Otchere et al., 2018). However, no animal reservoir for L6 has been discovered yet. The African origin (likely East Africa) of *M. bovis*, which is one of the closest MTBC to L6, was also recently confirmed (Loiseau et al., 2020). Similarly, another human pathogen, *Helicobacter pylori*, was recently shown to have an African origin (Linz et al. 2007).

In a 2012 paper, Bentley et al. suggested that the RD900 gene region could have been deleted twice in MTBC evolution: once leading to the modern MTBC lineages and once leading to the *M. bovis* lineage (Bentley et al., 2012). The number of pseudogenes is high in L6, confirming an ongoing genomic erosion process in MTBC (Bentley et al., 2012). Moreover, whether regulatory proteins seem to evolve neutrally in L5, they may evolve under purifying selection in L6 (Otchere et al., 2018). Other specific features were associated to L6, such as the presence of specific IS*6110* copy within the *moaX* gene, or other RDs described recently (RD702), and a mutation in the *mutT2* promoter (Comin et al., 2021) (see also Fig. 27.18).

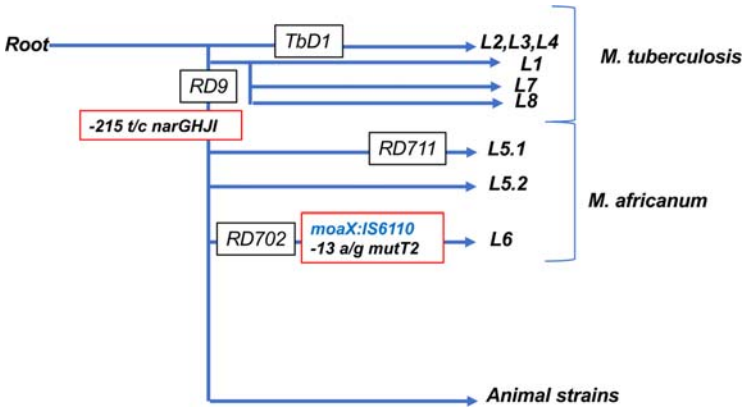


FIGURE 27.18

Current scheme of evolution of ancient and modern *Mycobacterium tuberculosis* in Africa (Comin et al., 2021).

7.2.2 Some Data on L5 and L6 Virulence and Adaptation to Oxygen Limitation

L6 isolates showed a low number of IS6*110* copies (≤6) (Comin et al., 2021). In Mali, a study with patients infected with L6 suggested that they presented a slower sputum smear conversion than L4 patients (Diarra et al., 2018). It had been shown earlier that L6 elicits an attenuated T cell response in TB patients (De Jong et al., 2006). Nonsynonymous mutations in the major *DosR* regulon genes were found in many L6 genomes of TB patients from the Gambia and Ghana. These genes are useful for fast adaptation to hypoxic stress and these mutations could contribute to a significant underregulation during dormancy in L6 (Ofori-Anyinam et al., 2017).

7.2.3 L6 Defining SNPs

The L6-defining SNP at position 1816587 as found by F. Coll et al. also includes the L9 sublineage as a "false positive," whereas some L6 seems not to have this SNP (e.g., ERR234150 or ERR1215466). Hence another SNP such as the C/T mutation in an intergenic region at position 183575, that does not select the L9 sublineage would be better to define L6.

7.2.4 Subclassification of L6 in at Least Three Sublineages

In a recent article, Coscolla et al. by performing a large WGS study on 320 L6 isolates, found that at least three sublineages, 6.1, 6.2, and 6.3 could be defined within L6. Each of this sublineage can be split into at least three sub-sublineages (Coscolla et al., 2021).

7.2.5 Spoligotyping Signatures Within L6 and Their Link to WGS-Defined Sublineage

It was shown in previous studies that more discrimination could be obtained on L5 and L6 typing by using the 68-spacer format spoligotyping instead of the classical 43-spacers format (Ates et al., 2018a; Brudey et al., 2004). Using the *CRISPR-builder TB* application and adequate L6 SRA-based in Silico CRISPR-locus reconstruction, the following observations are made (Guyeux et al., 2021): SIT181 is found in almost all sublineages if not all with the exception of L6.2.3 and L6.3.1 for the time-being; SIT2157 is found only in various L6.1 sublineages, whereas SIT2324 and SIT2441 are found only in various L6.2 sublineages. SIT321, SIT326, and SIT3130 are found only in L6.3 lineages. It is too early, given the paucity of data, to definitively link these diverse genotypes to specific geographical origins.

8. Focus on Lineage 7

8.1 Chronology of Discovery of Lineage 7

The lineage 7 was discovered independently by two teams in Djibouti and Ethiopia, respectively (Blouin et al., 2012; Firdessa et al., 2013). L7 could have been detected earlier in spoligotyping databases based on its spoligotyping signature (SIT910, octal code 700000007177771) (Brudey et al., 2006). L7 belongs to PGG1 since it shows the typical gyrA95-katG483 "CA" allele combination (Table 27.1). This sublineage did not expand out of North—Ethiopia, where it is highly prevalent (Tulu and Ameni, 2018). We found that this sublineage is not homogeneously spread in Ethiopia, a country with high specific human diversity and that has resisted to western colonization, and was the site of many civil wars between Ethiopians of different regions. If L7 did not expand, is it because this sublineage is less virulent or just because it was geographically constrained in isolated populations? It was recently suggested that the finding of this genotype was linked to patient delay in seeking

treatment, which means that this genotype could be endemic to some remote rural regions of Ethiopia, either to be found in the North-East or in the North West of Ethiopia (Comas et al., 2015; Yimer et al., 2015).

A good SNP associated with L7 can be found at position 1137518 (with a false-positive sample ERR2513690), and we suggest that this lineage is better defined using SNP position 1960 G/C in an intergenic region (Senelle et al., 2023c). On a set of 35 SRAs, the mean distance between two isolates in a symmetric distance matrix gives a total of 360 ± 128 SNPs (standard deviation) with a minimum of 165 and maximum of 575 SNPs, and a minimal time of divergence with their MRCA (assuming a maximal mean rate of 1SNP acquisition every 3 years) that could be between 495 and 1725 years ago (Schürch et al., 2010).

9. The Lineage 8 (Uganda, Rwanda)

9.1 A Sister Lineage of MTBC Found in the Great Lakes Region

Lineage 8 was found in two isolates so far, one in Rwanda and one in Uganda, close to the great lakes region (Ngabonziza et al., 2020). L8 also belongs to PGG1. L8 harbors an intact TbD1 and RD9 region (Ngabonziza et al., 2020). The two isolates were drug-resistant isolates with rare mutations on *rpoB* and *inhA* (Ngabonziza et al., 2020). They also share the presence of a 4.4 kb region common to *M. canetii* and coding for *cobF* and an intact *pks8* gene that is interrupted by a frameshift in H37Rv (Ngabonziza et al., 2020). Alternatively these strains miss five nonrepetitive genes included in two segments of 3,4 and 4,4 kb that are present in reference MTBC genomes (Ngabonziza et al., 2020). Comparison with other genomes suggested that the divergence of L8 preceded the loss of the *cobF* region. this region is involved in cobalamin/VitB12 synthesis, and other gene interruptions (such as the frameshift in *pks8*) that are common to all other MTBC lineages can be found (Ngabonziza et al., 2020). Together with the discovery of *Mycobacterium canetti*, also known as *Mycobacterium prototuberculosis* that bears a mosaic genome, of L7 (limited to Ethiopia), and now the discovery of L8, all these discoveries argues in favor of a remote East-African origin of the MTBC (Gutierrez et al., 2005; Nebenzahl-Guimaraes et al., 2016; Ngabonziza et al., 2020; van Soolingen et al., 1997). L7 and L8 show very long evolutionary branches (877 specific SNPs in L8 vs. 823 in L7). A synonymous SNP at position 4386210 in Rv3900c is shared by L7 and L8 suggesting that L7 and L8 may either share a common ancestor or have independently acquired this mutation as an homoplasic event (results not shown).

10. The Lineage 9 and 10: Newly Discovered *M. africanum* Sublineages

10.1 Discovery of New *Mycobacterium africanum* Lineages

The last group of African MTBC isolates with unclassified genomes was recently shown to belong to two new sublineages proposed to be named MTBC lineage 9 (L9) and Lineage 10 (Coscolla et al., 2021; Guyeux et al. 2024) (Fig. 27.19). The first result was obtained after analyzing a total of 675 *Maf* genomes distributed in 350 L5 and 320 L6 genomes, as well as five related genomes that could not be classified into any of the known human- or animal-associated MTBC lineages (Coscolla et al., 2021). The investigations done on these five genomes had suggested that these latter genomes are part of a

sister clade of L6, branching after the separation between all other L6 and the animal isolates, but before the split of all other L6 sublineages (L6.1, L6.2, and L6.3). The second discovery, L10, was obtained using our own pipeline "*TB-Annotator*" (Senelle et al. 2023c). We also confirmed the phylogenetic position of L9. We could identify two monophyletic genomes branching further up between animal lineage A1 and L9 (Coscolla et al., 2021; Zwyer et al. 2021) and names these ones L10 (Guyeux et al., 2024). Interestingly while L9 samples were for now only isolated from the Horn of Africa, L10 samples all seem to come from Central Africa (Democratic Republic of the Congo or Congo Republic). These L6 sister lineages showed similar deletions as L6 such as the region of difference RD702, but L9 did not show other deletions that are present in animal-associated lineages such as RD1 and RD5 (Fig. 27.19); L10 samples did not harbor the RD702 deletion typical of the L6/L9 branch, but harbor-specific characteristics.

10.2 Geographical Origin of L9 Is Likely to Be Found in East Africa

The phylogeographic origin of the five L9 genomes however differs from all other L5−L6 genomes included, as they were the only ones with an origin to be found in East Africa (one from Djibouti, three from Somalia and one isolated in Europe but the patient origin was unknown) (see also Fig. 27.20) (Coscolla et al., 2021). Some of these isolates had already been described as far back as 2012 by Blouin et al. as a sister clade to L6 (Blouin et al., 2012). The mean distance between these five genomes to L5, L6 and the animal-adapted strain is 1632, 1191, and 1491 SNPs, respectively, to L5, L6 and animal-adapted strains. This distance is greater than to the intra-cluster distance between members of these lineages. The taxonomy of African lineages of MTBC remains an important issue for better defining the early evolutionary history of MTBC. A search of the spoligotyping motif "772.0000….6.1" in SITVIT recovered six isolates, among which one was defined as SIT1153, all belonging to the L9 sublineage, as shown in Table 27.9. Of note, similarly Lineage 10 shares a specific signature with absence of spacers 7,9 and all spacers starting at spacer 26.

11. Animal Lineages and *Mycobacterium canettii*

11.1 Animal Lineages

Our review is not focused on the most recent development of livestock-associated MTBC phylogenomics and the reader will refer to these two studies for more details (Zimpel et al., 2020; Zwyer et al., 2021).

11.1.1 The Phenotypic Continuum in L6 Between Human and Animal-Adapted Lineages

It was well known before genetic studies, that *M. africanum* (L5-L6) was fulfilling a continuum of phenotypic patterns from strictly human to strictly bovine (David et al., 1978). From an epidemiological standpoint, wild-life associated MTBC infections could have been either independent, predating, or posterior to human MTBC infections; as seen in the introduction, demographic parameters may have played a key role, with a recent transmission direction inversion (from human to animals instead of from animals to human) contemporaneous to a change to the *anthropocenosis* era. Reservoirs of zoonotic tuberculosis diseases, in particular of *M. bovis*, are numerous (Larsen et al., 2020). Alternatively, local ecological parameters, in relation to food regimen or cultural practices may have

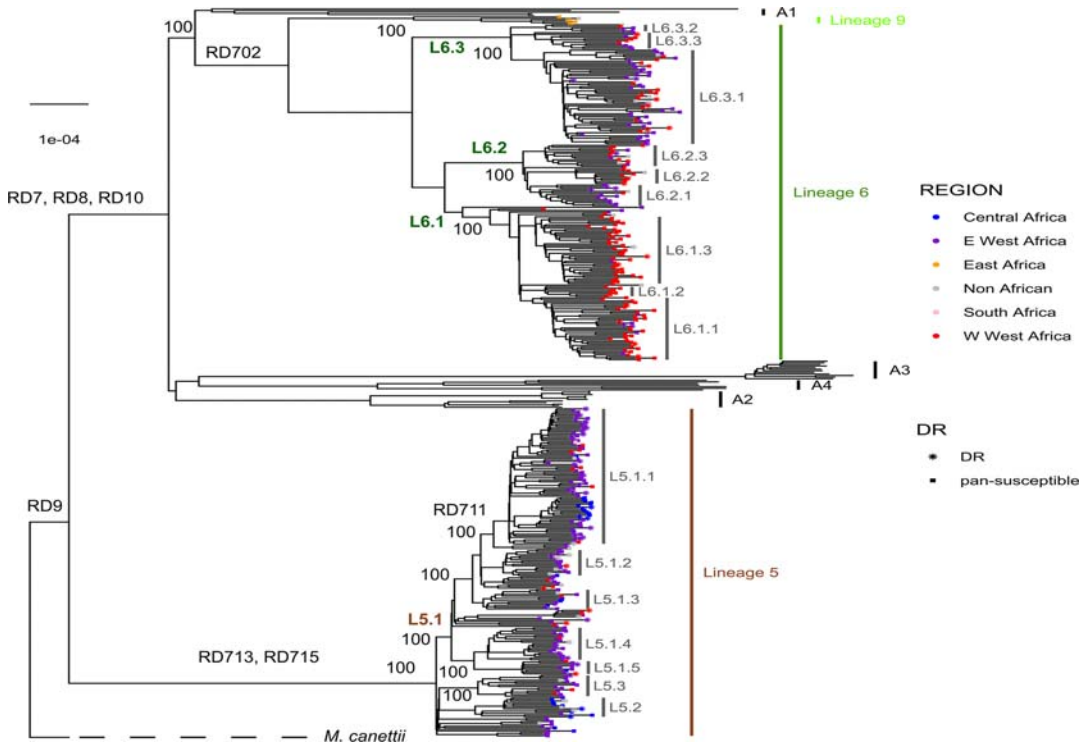


FIGURE 27.19

Phylogenetics of African L5–L6 clinical isolates of MTBC.

Reprint from Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., Otchere, I.D., Asante-Poku, A., Asare, P., Sanchez-Buso, L., Gehre, F., Sanoussi, C.N., Antonio, M., Affolabi, D., Fyfe, J., Beckert, P., Niemann, S., Alabi, A.S., Grobusch, M.P., Kobbe, R., Parkhill, J., Beisel, C., Fenner, L., Bottger, E.C., Meehan, C.J., Harris, S.R., de Jong, B.C., Yeboah-Manu, D., Brites, D., 2021. Phylogenomics of Mycobacterium africanum *reveals a new lineage and a complex evolutionary history. Microb. Genom. 7. doi: 10.1099/mgen.0.000477 by courtesy of the authors.*

played a key role in TB evolution. An example was given recently by a study that characterized in Peruvian human skeletons, mycobacterial genomes found to be close to those found in sea mammals (*M. pinnipedii*), which may implicate them in first transmitting the disease on Peruvian shores at the Holocene, i.e., less than 6000 years ago (Bos et al., 2014). However, more strangely, TB caused by this pathogen, *M. pinnipedii*, would also have been able to spread off the coast in mainland, thus inferring the likelihood of a passed human to human transmission or animal-mediated route of a zoonotic TB disease in precolonial America (Vagene et al., 2022).

11.1.2 Ecotypes Among Animal MTBC Lineages

Animal MTBC lineages or ecotypes were regularly discovered in mycobacteriology starting with *M. bovis* and *M. microti* described in the 1930s (the vole bacillus) (Pym et al., 2002; van Soolingen et al., 1998), *M. caprae* (Aranaz et al., 1999), the seal bacillus *M. pinnipedi* (Zumarraga et al., 1999),

Table 27.9 Lineage 9 and Lineage 10 specific Spoligotypes

country code	spoligotype	SIT	Year	Data source of SRA n°
NLD		1153	NA	Van der Zanden A
CHE		1153	2002	Driscoll J
USA		1153	2002	Driscoll J
FXX		Orphan	1999	Gutierrez MC
NLD		Orphan	NA	Kremer K
BEL		Orphan	2003	Allix C
COG		Orphan	ND	Fauville M
RDC		Orphan	ND	ERR2707158
Unkn.		Orphan	ND	ERR2516384

From Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., Otchere, I.D., Asante-Poku, A., Asare, P., Sanchez-Buso, L., Gehre, F., Sanoussi, C.N., Antonio, M., Affolabi, D., Fyfe, J., Beckert, P., Niemann, S., Alabi, A.S., Grobusch, M.P., Kobbe, R., Parkhill, J., Beisel, C., Fenner, L., Bottger, E.C., Meehan, C.J., Harris, S.R., de Jong, B.C., Yeboah-Manu, D., Brites, D., 2021. Phylogenomics of Mycobacterium africanum *reveals a new lineage and a complex evolutionary history. Microb. Genom. 7. doi: 10.1099/mgen.0.000477; Guyeux, C., Senelle, S., Le Meur, A., Supply, P., Gaudin, C., Phelan J. E., Clark, T. G., Rigouts, L., de Jong, B., Sola, C., Refrégier, G, 2024. Newly identified* Mycobacterium africanum *lineage 10 Central Africa. Emerging Infectious Diseases. 30, doi:10.3201/eid3003.231466 and SITVIT database*

the Dassie Bacillus (Mostowy et al., 2004b), *M. surricatae* (Dippenaar et al., 2015), *M. chimpanzee* (Coscolla et al., 2013), and *M. mungi* (Alexander et al., 2010, 2016). The ecotypes paradigm in MTBC was very well pioneered by N. Smith (Smith et al., 2006) and this complex ecological transmission could start to be unraveled in some cases such as the mentioned pre-Columbian America for *M. pinnipedii* infection (Vågene et al., 2022). From these discoveries emerges the general idea that tuberculosis in mammals was and remains an ancient story. Hence, it could be interesting to perform more studies on ancient animal DNA (aDNA), i.e., on ancient mammalian bones (or teeth) from mummified animals, to try to retrace the potentially long-lasting interactions between mammalians and mycobacteria. In our former *TB-annotator project* (n = 6009 genomes in the v2.1 version), a total of 787 public genomes of animal strains (that also includes human L6 genomes) shared a total of 15 exclusive SNPs that are listed below in Table 27.10.

When analyzing the tree shown in Fig. 27.20, orange/brown spots represent L5, L6, L9, or L10 and emerald blue spots represent animal isolates. It is out of the scope of this review to focus exhaustively on the differential genomic characteristics of these ecotypes, however, a number of remarks can be

Table 27.10 List of 15 SNPs specific to L6 and Animal Lineages Found in the *TB-Annotator* Project

Position	Wt/mut	Gene	Rv	Nature of Mutation
147812	GC/G			Intergenic
275367	G/T	php		Missense
301687	C/T			Intergenic
382489	C/T			Intergenic
405854	C/G			Intergenic
658923	C/G		Rv0567	Missense
733820	C/A	mma2		Missense
851562	G/A			Intergenic
1225198	C/T		Rv1096	Missense
1885481	C/G		pks8	Missense
2438094	T/G		pknL	Missense
2447426	G/T		TB16.3	Synonymous
2772741	C/A		Rv2469c	Missense
2968468	G/C			Intergenic
3893290	G/A			Intragenic

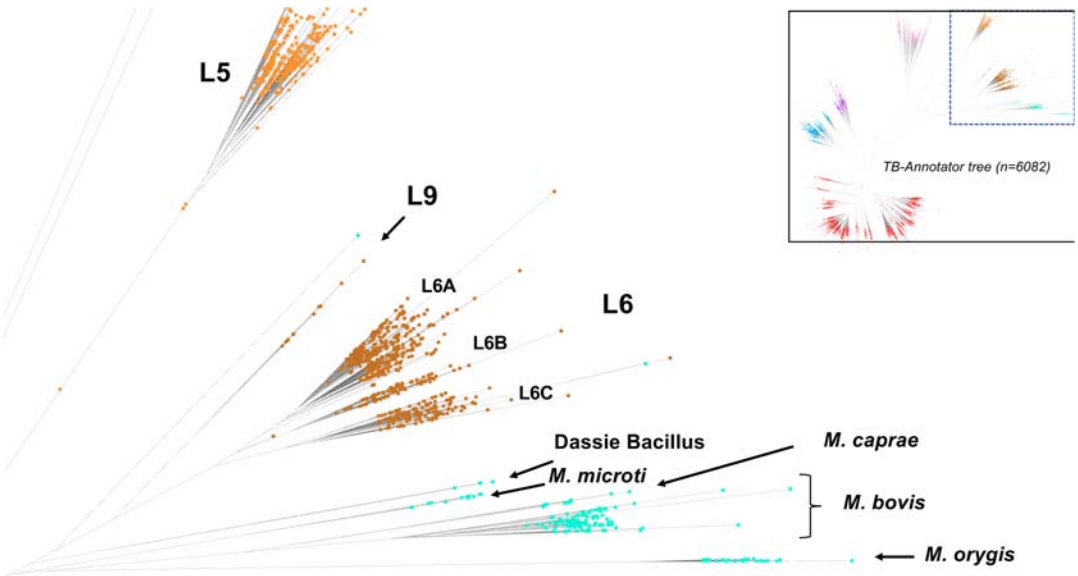


FIGURE 27.20 Main phylogenetic branches of animal MTBC in TB-Annotator (version November 2021, 6082 SRAs).

Main phylogenetic branches of animal MTBC in TB-Annotator (Version November 2021, 6082 SRAs).

done. First, it is clear that *M. bovis* shares a deeply rooted MRCA common to all L6 and animal strains with a common MRCA with L5 ancestor.

Secondly, there is also a deeply rooted branching of L9 and presumably of L10, relatively to the emergence of all other L6 branches (Fig. 27.20). There is globally a high number of specific SNP variants in all individualized animal branches that suggests a long history of evolution of each of these phylogenetical branches (Coscolla et al., 2021) (Table 27.11).

11.2 *Mycobacterium canettii*

11.2.1 General Characteristics of M. canettii

A recent review describes in detail the biology and epidemiology of *M. canettii* and interested readers are referred to this review for more details (Supply and Brosch, 2017). The description of the smooth tubercle bacillus, *Mycobacterium canettii*, ends our world tour of MTBC diversity and brings us back to the Horn of Africa (van Soolingen et al., 1997). This taxon harbors a different "smooth" colony phenotype, in opposition to the classical "rough" MTBC (van Soolingen et al., 1997). *M. canettii* should be considered a potential outgroup from an evolutionary perspective, as compared to MTBC. It has a totally different CRISPR locus, that suggests a specific ecological and geographical niche (van Embden et al., 2000). It can be identified by a C/T mutation in *hsp65* (Goh et al., 2001), a specific region of difference (RD12 can) and other mutations (Fabre et al., 2010). *M. canettii* is showing much

Table 27.11 Analysis of SNPs Characteristics All L5, L6, L9, L10, and Animal MTC Strains (n = 1104) Among 6082 genomes Database (TB-Annotator)

L5-L6-L9 and animal Strains	n	Exclusive SNPs	95%	Symmetric Difference Matrix (SNP Difference Number, Mean and Standard Deviation	Shared Missing Region of Differences	Subspecies
L5	327	143	546	ND	ND	*M africanum*
L9	11	209	0	605 ± 234	ND	*M africanum*
L10	2	283	0	432	ND	*M. africanum*
L6-A	364	45	98	ND	ND	*M africanum*
L6-B	81	110	44	460 ± 111	ND	*M africanum*
L6-C	149	93	87	537 ± 129	ND	*M africanum*
	4	62	0	1038 ± 415	N-RD25das (25%)	Dassie bac.
	13	237	0	716 ± 326	RD1mic (25%)	*M. microti*
	29	791	109	435 ± 105	220BWa, MiD2, RDoryx_1	*M. orygis*
	12	95	0	1028 ± 466	RDcap	*M. caprae*
	114	163	101	692 ± 222	N-RD17	*M. bovis*
SUM	1104	1948	985			

more genetic diversity than MTBC complex and possesses a mosaic genome that can perform homologous recombination and accepts lateral gene transfer, a feature that was lost in MTBC (Gutierrez et al., 2005). This difference was undoubtedly very important for the evolution of the species. In a recent study, it was demonstrated that horizontal gene transfer in MTBC does not process through an ESX-dependent process, as is done in *M. smegmatis* (Madacki et al., 2021). An interesting evolutionary hypothesis is that the ancestor of today's MTBC acquired less polar lipids in its outer membrane to increase its hydrophobicity, a potential virulence factor that may have increased likelihood of aerosol transmission and increased protection (Jankute et al., 2017).

11.2.2 Is M. canettii *an Environmental Bacteria?*

Although *M. canettii* can be considered as a potential progenitor species of MTBC, a potential ancestor of MTBC that could be as ancient as three millions years (Gutierrez et al., 2005), there remains missing links between *M. canettii* and the MRCA of ancestral MTBC (Ngabonziza et al., 2020; Yenew et al., 2023). A possible environmental reservoir of *M. canettii* was searched for in East Africa, but remains undiscovered. Even if *M. canetti* can be transmitted by aerosol, it appears to be impaired in patient-to-patient transmission (Osman et al., 2015). *M. kansasii*, which is also an opportunistic pathogen agent, is a good model to study the change from an environmental opportunistic to a strictly animal (and human) pathogen (Jankute et al., 2017; Wang et al., 2015). Adequate Carbon and Nitrogen sources required for such a favorable environment to host *M. canettii* could be found in local endemic plants of the horn of Africa, as suggested by Loukil et al. (2019).

12. Variability of the Clinical Expression of the Tuberculosis Disease and of Virulence of MTBC as a Function of the Lineage

Without trying to be exhaustive on the complex and numerous metabolic and phenotypic consequences of all the nucleotidic or genomic variations that we described in this MTBC genome biodiversity review, we aim at providing a broad overlook of the impact of diversity. We therefore propose here to highlight some of the recent results of this research field. The reader will refer to some excellent reviews on the virulence and virulence emergence mechanisms of MTBC (Ehrt et al., 2018; Forrellad et al., 2013; Orgeur and Brosch, 2018; Tiwari et al., 2019), and specifically in this book on the Chapter 26 on *Patho-adaptive evolution of the Mycobacterium tuberculosis complex to its hosts* by Jesús Gonzalo Asensio. Among these reviews, one describes in detail the importance of type VII secretions systems (Tiwari et al., 2019). The role of the mycobacterial cell wall in MTBC virulence has also been known for a long time and an exhaustive and detailed knowledge on the chemical components of the cell wall and on their metabolism has been acquired with time (Grabowska et al., 2020; Singh et al., 2018).

12.1 The Issue of Adaptation of the Pathogen to Its Host

The genomic diversity described in this chapter points to numerous challenges either for tuberculosis control and eradication, for development of vaccines and diagnostics or for the understanding of the pathoadaptation of MTBC (Comas and Gagneux, 2009). In parallel to these issues, we observed a strong sympatric association between humans and tubercle bacilli that was progressively unveiled

(Gagneux et al., 2006; Hershberg et al., 2008). This is true even in large cosmopolitan cities, thus underlying the strong and long-term specific interactions that have had time to co-develop between the host and the pathogen agent (Gagneux et al., 2006). These cross-talks have created a large range of diverse clinical expression and transmission specificities, that are concomitantly linked to, on the one hand, the human population demography, history, immunological status of individuals and of populations and to, on the other end, the never-ending adaptation and escape ability of the bacilli (Gagneux et al., 2006; Hershberg et al., 2008). Indeed, the variability of the clinical expression of the disease is the result of the intimate interaction of human and bacilli genotypes, whether at the population level or at the individual cell level.

At the population level, it was well known for example that the colonial troops from Senegal and Madagascar were dying of an acute and generalized tuberculosis when coming to the front during World War I (Fargettas, 2012). Even though this remains to be proven, it is very likely that the African soldiers were infected in the trenches by their unfortunate European TB bacilli-carrying companions, that rapidly contaminated them, in harsh living conditions, and triggered an acute and generalized TB disease in them, and could be considered as "naïve," even though they might have been primo-infected locally by *M. africanum* (L6 in Senegal) or by L1 isolates (in Madagascar) in their childhood. This historical susceptibility of African soldiers to TB is now better explained, combined with the knowledge that the T cell response to ESAT6, when patients were infected by L6 isolates, was less important than when they were infected with other MTBC (non-L6), and it reinforces the notion that L6 is less virulent than "modern" TB clinical isolates (de Jong et al., 2006). If we could perform WGS studies on ancient DNA samples, which may have been preserved in Ziehl-Neelsen stained archives from WWI from soldiers from Madagascar or Senegal who died of fulminant TB, this might validate this hypothesis.

At the individual cell level, new technological approaches to study host−pathogen interactions through single-cell analysis and differential RNA expression cross-talks between stress genes in the bacillus and of the pro-or antiinflammatory host−response genes in the host cells suggests that macrophages are not equal in controlling MTBC. Interstitial macrophages tend to be more successful at controlling the bacterial infection than are alveolar macrophages (Pisu et al., 2021).

12.2 Differential Virulence Inside MTBC

As early as in the sixties and seventies, it was observed that MTBC isolates of South-Asian born patient (patients likely hosting L1 isolates) were less virulent in a Guinea-Pig model than MTBC isolates from UK-born patients (likely hosting L4 isolates). This observation could have been potentially explained by the presence of a different lipid the "*attenuation indicator lipid*" in L1 cell wall isolates (Goren et al., 1974). An association between a shorter sputum conversion to negative and TB due to an L1 strain was also suggested (Click et al., 2013). In Ethiopia, patient that hosted L7 strains, were taking more time to seek for treatment than patients suffering from a tuberculosis due to a bacillus of the L3 lineage, either suggesting that L7 was less virulent than L3 or that patients with L7 were more likely to be geographically isolated than patients hosting an L3 isolate (Yimer et al., 2015). In an Australian cohort of 212 cases with Lymph Nodes tuberculosis (LNTB) and 424 controls with Pulmonary tuberculosis (PTB), a risk factor analysis suggested that patients infected with L3 were more likely to develop an LNTB than patients with PTB (Khandkar et al., 2015). Lineage 2 and Lineage 4 isolates, in particular L2 that is known to have quite early acquired resistance to Rifampin

and Isoniazid, are believed to be more virulent as well as more transmissible than other strains (Sibandze et al., 2020).

12.3 Importance of Polyketide Synthases

Such observations can be put in parallel with the observations that the cell envelope lipid profile of MTBC varies with regards to the lineage, and that the lineage could influence early interactions with the host and ultimately determine the clinical consequences of the infection (Krishnan et al., 2011). At least 16 polyketide synthase enzymes (*pks*) were described in *M. tuberculosis* H37Rv (https://mycobrowser.epfl.ch). These enzymes are necessary to build the complex cell wall of MTBC, which is made-up of mycolic acids, phthiocerol dimycocerosates, and phenolic glycolipids among other complex molecules. Such polyketides are synthesized by megaenzymes, such as the polyketide synthases, and variations in their structure, activities, and genetic recombination may promote the building of a very diverse set of mycobacterial cell wall structures, in particular if some lineages have abilities to recombine on the *pks* genes (Boritsch et al., 2016; Grabowska et al., 2020). Similarly, the PE-PPE genes, that represent up to 10% of the genomes are highly polymorphic and this has functional consequences (Gey van Pittius et al., 2006). This also creates favorable recombination conditions for specific physiopathological and clinical consequences to develop (Phelan et al., 2016).

12.4 Differential drug resistance according to lineages and sublineages:

In fine, it is highly likely that the penetration and action of the various antimycobacterial agents that we know will not act similarly on all TB bacilli, but may vary according to their lineage belonging (Grabowska et al., 2020). The link between the development of resistance to drugs and modifications of drug-target genes is known for a long time for MTBC (Musser, 1995). It is beyond the scope of this review to describe these variations and how they may act on drug-resistance mechanisms. Associations between nucleotide variations, drug resistance, and lineages are now better investigated thanks to the combination of WGS studies and statistics, through Genome-wide association studies (Coll et al., 2018; Ford et al., 2013). With the introduction of new antituberculous drugs, such as bedaquiline, linezolid, delamanid, pretomanid, and clofazimine, new results are obtained by systematic approaches and WGS databases (Kadura et al., 2020). Advances in genomics have paved the way for a favorable switch from phenotypic to predictive genomic-based drug susceptibility testing (CRyPTIC, Consortium, 2018; CRyPTIC, Consortium, 2022). Such examples of recent studies using new technological and theoretical approaches, suggest that in the future, genomic tuberculosis research could allow to investigate with more details the clinical consequences of molecular mutations, looking for epistasis, for metabolic and regulatory networks variations, in relation to phylogenetic lineages, using a combination of new and combined experimental techniques (Single-Cell Genomics and host-interaction studies) and improved data-analysis methods with the generalization of machine learning approaches.

12.5 Quantitative Genetics and GWAS

Even though the link between gene diversity, MTBC lineage and specific clinical expression forms of tuberculosis remains difficult to perform, at least two recent studies have tried to look for association between specific MTBC gene variation and meningitis, one of the most threatening clinical form of the

tuberculosis disease (Faksri et al., 2018; Ruesen et al., 2018). The first study was based on 322 isolates from 106 to 216 pulmonary and meningitis patients, respectively. The second one was based on 293 patients from 73 to 220 pulmonary and meningitis patients, respectively. The first study identified that genetic variation in Rv0218, the absence of Rv3343c and *nanK* could be associated to this specific TB phenotype. The second study identified 242 variants, linked to meningitis, among which many in *pks* and in PE-PPE genes. These two studies, which give discrepant results, show how difficult it remains to create strong association between specific genetic variants and clinical expression of the disease, given the complex quantitative multiloci nature of tuberculosis host-pathogen interactions (Faksri et al., 2018; Ruesen et al., 2018).

13. Conclusion

In this chapter, we have attempted to build an exhaustive picture of the knowledge that was gained on MTBC genomic variation, from one hand at the nucleotide level through large-scale WGS studies, and from another hand in combination with previous acquired knowledge that had been gained using classical repeated-sequence based markers during the last 25−30 years. We show that the studies on genetic diversity of MTBC have reached a great level of maturity and depth of knowledge. We hope that this review creates a useful bridge between an ancient and a new era for researchers working on MTBC genomes, whether for applied or academic purposes.

Some evolutionary important lineages such as L5 and L6 still remain poorly sampled and the unraveling of their population structure is unfinished. Alternatively, we pointed in this review on the extended knowledge that was acquired on L2 and L4. We showed that CRISPR-typing remains very useful when fine local or historical identity of epidemic clones is described and SNP-based classification would benefit from an internationally agreed (consensus) naming system that remains to be put in place. The spread of genomics correlated to computational biology now allows to decipher long-term infectious disease history from recent epidemiology by comparing genomes at different time-depths and by hypothesizing on pace of genome changes. This short-waves versus long-waves history can be assimilated to Braudel's global history. The WGS methods allow to predict the risk of new drug-resistance emergence locally in a given setting given the fast link that can be done within SNPs markers that are physically linked in the MTBC model, thus ascertaining the importance of genotyping and cluster surveillance.

Unfortunately, 30 years after the declaration of TB as a worldwide emergency by WHO, we will not be successful in reaching the UN defined targets of the Millennium goals, to eradicate TB by 2030. The COVID-19 crisis has pushed us back in terms of achievement and financing of TB eradication programs by at least 10 years (Kirby, 2021). Meanwhile, social inequities and food-supply issues have worsened in many countries due to COVID-19 crisis, to global climate change and to poor access to diagnostic and treatment in resource-limited countries. The recent Russia-Ukraine and Israel-Gaza wars, just after the COVID-19 sanitary crisis also adds a lot of suffering and will no doubt trigger an increase in TB in countries at war and beyond through food shortages and an increase in the number of refugees. It is paradoxical to see that Russia, in the WHO European Region had seen a 6% per year fall in TB incidence between 2010 and 2020, and is now directly and indirectly promoting an increase in TB the world (Kirby, 2021). It is well known that a strong public sector is needed to control pandemia and the return to normal TB diagnostics may need time for resource-limited countries; the decrease in

TB notification rate fell from 7.1 million in 2019 to 5.8 million in 2020 and this fall was also observed in 2021 and with progressive recovery in 2022 (Kirby, 2021). If the COVID-19 crisis could trigger a moral shock worldwide for more investment from private and public stakeholders to invest in public goods, such as public and individual health, and for a sustained future of all soils, animal, plants, and human ecosystems, then this recent phase of humankind's history could have been a blessing in disguise, and TB could ultimately be eradicated by around 2040 to 2050.

References

Abadia, E., Zhang, J., dos Vultos, T., Ritacco, V., Kremer, K., Aktas, E., Matsumoto, T., Refrégier, G., van Soolingen, D., Gicquel, B., Sola, C., 2010. Resolving lineage assignation on *Mycobacterium tuberculosis* clinical isolates classified by spoligotyping with a new high-throughput 3R SNPs based method. Infect. Genet. Evol. 10, 1066−1074. https://doi.org/10.1016/j.meegid.2010.07.006.

Achtman, M., 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu. Rev. Microbiol. 62, 53−70. https://doi.org/10.1146/annurev.micro.62.081307.162832.

Achtman, M., 2016. How old are bacterial pathogens? Proc. Biol. Sci. R. Soc. 283 (1836), 20160990. https://doi.org/10.1098/rspb.2016.0990.

Advani, J., Verma, R., Chatterjee, O., Pachouri, P.K., Upadhyay, P., Singh, R., Yadav, J., Naaz, F., Ravikumar, R., Buggi, S., Suar, M., Gupta, U.D., Pandey, A., Chauhan, D.S., Tripathy, S.P., Gowda, H., Prasad, T.S.K., 2019. Whole genome sequencing of *Mycobacterium tuberculosis* clinical isolates from India reveals genetic heterogeneity and region-specific variations that might affect drug susceptibility. Front. Microbiol. 10, 309. https://doi.org/10.3389/fmicb.2019.00309.

Affolabi, D., Anyo, G., Faihun, F., Sanoussi, N., Shamputa, I.C., Rigouts, L., Kestens, L., Anagonou, S., Portaels, F., 2009. First molecular epidemiological study of tuberculosis in Benin. Int. J. Tuberc. Lung Dis. 13, 317−322.

Ajawatanawong, P., Yanai, H., Smittipat, N., Disratthakit, A., Yamada, N., Miyahara, R., Nedsuwan, S., Imasanguan, W., Kantipong, P., Chaiyasirinroje, B., Wongyai, J., Plitphonganphim, S., Tantivitayakul, P., Phelan, J., Parkhill, J., Clark, T.G., Hibberd, M.L., Ruangchai, W., Palittapongarnpim, P., Juthayothin, T., Thawornwattana, Y., Viratyosin, W., Tongsima, S., Mahasirimongkol, S., Tokunaga, K., Palittapongarnpim, P., 2019. A novel Ancestral Beijing sublineage of *Mycobacterium tuberculosis* suggests the transition site to modern Beijing sublineages. Sci. Rep. 9, 13718. https://doi.org/10.1038/s41598-019-50078-3.

Aleksic, E., Merker, M., Cox, H., Reiher, B., Sekawi, Z., Hearps, A.C., Ryan, C.E., Lee, A.V., Goursaud, R., Malau, C., O'Connor, J., Cherry, C.L., Niemann, S., Crowe, S.M., 2013. First molecular epidemiology study of *Mycobacterium tuberculosis* in Kiribati. PLoS One 8, e55423. https://doi.org/10.1371/journal.pone.0055423.

Alexander, K.A., Larsen, M.H., Robbe-Austerman, S., Stuber, T.P., Camp, P.M., 2016. Draft genome sequence of the *Mycobacterium tuberculosis* complex pathogen M. Mungi, identified in a banded mongoose (*Mungos mungo*) in northern Botswana. Genome Announc. 4 (4), e00471−e516. https://doi.org/10.1128/genomeA.00471-16.

Alexander, K.A., Laver, P.N., Michel, A.L., Williams, M., van Helden, P.D., Warren, R.M., Gey van Pittius, N.C., 2010. Novel *Mycobacterium tuberculosis* complex pathogen, M. mungi. Emerg. Infect. Dis. 16, 1296−1299. https://doi.org/10.3201/eid1608.100314.

Alix, E., Godreuil, S., Blanc-Potard, A.B., 2006. Identification of a Haarlem genotype-specific single nucleotide polymorphism in the mgtC virulence gene of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 44, 2093−2098. https://doi.org/10.1128/JCM.00278-06.

Alland, D., Kalkut, G.E., Moss, A.R., Mc Adam, R.A., Hahn, J.A., Bosworth, B., Drucker, E., Bloom, B.R., 1994. Transmission of tuberculosis in New York City: an analysis by DNA fingerprinting and conventional epidemiologic methods. N. Eng. J. Med. 330, 1710−1716. https://doi.org/10.1056/NEJM199406163302403.

Allix-Beguec, C., Harmsen, D., Weniger, T., Supply, P., Niemann, S., 2008. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. J. Clin. Microbiol. 46, 2692−2699. https://doi.org/10.1128/JCM.00540-08.

Anonymous, 1993. Tuberculosis: a global emergency. World Health Forum 14, 438.

Aranaz, A., Liebana, E., Gomez-Mampaso, E., Galan, J.C., Cousins, D., Ortega, A., Blazquez, J., Baquero, F., Mateos, A., Suarez, G., Dominguez, L., 1999. *Mycobacterium tuberculosis* subsp. caprae subsp. nov.: a taxonomic study of a new member of the *Mycobacterium tuberculosis* complex isolated from goats in Spain. Int. J. Syst. Bacteriol. 49, 1263−1273. https://doi.org/10.1099/00207713-49-3-1263.

Asante-Poku, A., Otchere, I.D., Osei-Wusu, S., Sarpong, E., Baddoo, A., Forson, A., Laryea, C., Borrell, S., Bonsu, F., Hattendorf, J., Ahorlu, C., Koram, K.A., Gagneux, S., Yeboah-Manu, D., 2016. Molecular epidemiology of *Mycobacterium africanum* in Ghana. BMC Infect. Dis. 16, 385. https://doi.org/10.1186/s12879-016-1725-6.

Asiimwe, B.B., Koivula, T., Kallenius, G., Huard, R.C., Ghebremichael, S., Asiimwe, J., Joloba, M.L., 2008. *Mycobacterium tuberculosis* Uganda genotype is the predominant cause of TB in Kampala, Uganda. Int. J. Tuberc. Lung Dis. 12, 386−391.

Ates, L.S., Dippenaar, A., Sayes, F., Pawlik, A., Bouchier, C., Ma, L., Warren, R.M., Sougakoff, W., Majlessi, L., van Heijst, J.W.J., Brossier, F., Brosch, R., 2018a. Unexpected genomic and phenotypic diversity of Mycobacterium africanum lineage 5 affects drug resistance, protein secretion, and immunogenicity. Genome Biol. Evol. 10, 1858−1874. https://doi.org/10.1093/gbe/evy145.

Ates, L.S., Dippenaar, A., Ummels, R., Piersma, S.R., van der Woude, A.D., van der Kuij, K., Le Chevalier, F., Mata-Espinosa, D., Barrios-Payan, J., Marquina-Castillo, B., Guapillo, C., Jimenez, C.R., Pain, A., Houben, E.N.G., Warren, R.M., Brosch, R., Hernandez-Pando, R., Bitter, W., 2018b. Mutations in ppe38 block PE_PGRS secretion and increase virulence of *Mycobacterium tuberculosis*. Nat. Microbiol. 3, 181−188. https://doi.org/10.1038/s41564-017-0090-6.

Baess, I., 1966. A bacteriophage for subdividing the species *M. tuberculosis*. Am. Rev. Res. Dis. 93, 622−623.

Bainomugisa, A., Duarte, T., Lavu, E., Pandey, S., Coulter, C., Marais, B.J., Coin, L.M., 2018. A complete high-quality MinION nanopore assembly of an extensively drug-resistant *Mycobacterium tuberculosis* Beijing lineage strain identifies novel variation in repetitive PE/PPE gene regions. Microb. Genom. 4 (7), e000188. https://doi.org/10.1099/mgen.0.000188.

Baker, L., Brown, T., Maiden, M.C., Drobniewski, F., 2004. Silent nucleotide polymorphisms and a phylogeny for *Mycobacterium tuberculosis*. Emerg. Infect. Dis. 10, 1568−1577. https://doi.org/10.3201/eid1009.040046.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455−477. https://doi.org/10.1089/cmb.2012.0021.

Bates, J.H., Fitzhugh, J.K., 1967. Subdivision of the species *M. tuberculosis* by bacteriophage typing. Am. Rev. Res. Dis. 96, 7−10.

Bates, J.H., Stead, W.W., 1993. The history of tuberculosis as a global epidemic. Med. Clin. N. Am. 77, 1205−1217. https://doi.org/10.1016/s0025-7125(16)30188-2.

Bazira, J., Matte, M., Asiimwe, B.B., Joloba, L.M., 2010. Genetic diversity of *Mycobacterium tuberculosis* in Mbarara, south western Uganda. Afr. Health Sci. 10, 306−311.

Bentley, S.D., Comas, I., Bryant, J.M., Walker, D., Smith, N.H., Harris, S.R., Thurston, S., Gagneux, S., Wood, J., Antonio, M., Quail, M.A., Gehre, F., Adegbola, R.A., Parkhill, J., de Jong, B.C., 2012. The genome of

Mycobacterium africanum West African 2 reveals a lineage-specific locus and genome erosion common to the *M. tuberculosis* complex. PLoS Neglected Trop. Dis. 6, e1552. https://doi.org/10.1371/journal.pntd.0001552.

Bespiatykh, D., Bespyatykh, J., Mokrousov, I., Shitikov, E., 2021. A comprehensive map of *Mycobacterium tuberculosis* complex regions of difference. mSphere 6, e0053521. https://doi.org/10.1128/mSphere.00535-21.

Bhatia, A.L., Csillag, A., Mitchison, D.A., Selkon, J.B., Somasundaram, P.R., Subbaiah, T.V., 1961. The virulence in the Guinea-pig of tubercle bacilli isolated before treatment from South Indian patients with pulmonary tuberculosis. 2. Comparison with virulence of tubercle bacilli from British patients. Bull. World Health Organ. 25, 313−322.

Bhatia, A.L., Jacob, C.V., Hitze, K.L., Ramachandran, K., Selkon, J.B., 1963. A comparison of the virulence in the Guinea-pig of tubercle bacilli from Thai and British patients. Bull. World Health Organ. 29, 483−490.

Bifani, P.J., Mathema, B., Kurepina, N.E., Kreiswirth, B.N., 2002. Global dissemination of the *Mycobacterium tuberculosis* W-Beijing family strains. Trends Microbiol. 10, 45−52. https://doi.org/10.1016/s0966-842x(01)02277-6.

Bird, D.W., Bird, R.B., Codding, B.F., Zeanah, D.W., 2019. Variability in the organization and size of hunter-gatherer groups: foragers do not live in small-scale societies. J. Hum. Evol. 131, 96−108. https://doi.org/10.1016/j.jhevol.2019.03.005.

Blouin, Y., Hauck, Y., Soler, C., Fabre, M., Vong, R., Dehan, C., Cazajous, G., Massoure, P.L., Kraemer, P., Jenkins, A., Garnotel, E., Pourcel, C., Vergnaud, G., 2012. Significance of the identification in the horn of Africa of an exceptionally deep branching *Mycobacterium tuberculosis* clade. PLoS One 7, e52841. https://doi.org/10.1371/journal.pone.0052841.

Bolotin, A., Quinquis, B., Sorokin, A., Ehrlich, S.D., 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151, 2551−2561. https://doi.org/10.1099/mic.0.28048-0.

Bonura, C., Gomgnimbou, M.K., Refrégier, G., Aleo, A., Fasciana, T., Giammanco, A., Sola, C., Mammina, C., 2014. Molecular epidemiology of tuberculosis in Sicily, Italy: what has changed after a decade? BMC Infect. Dis. 14, 602. https://doi.org/10.1186/s12879-014-0602-4.

Boritsch, E.C., Frigui, W., Cascioferro, A., Malaga, W., Etienne, G., Laval, F., Pawlik, A., Le Chevalier, F., Orgeur, M., Ma, L., Bouchier, C., Stinear, T.P., Supply, P., Majlessi, L., Daffe, M., Guilhot, C., Brosch, R., 2016. pks5-recombination-mediated surface remodelling in *Mycobacterium tuberculosis* emergence. Nat. Microbiol. 1, 15019. https://doi.org/10.1038/nmicrobiol.2015.19.

Bos, K.I., Harkins, K.M., Herbig, A., Coscolla, M., Weber, N., Comas, I., Forrest, S.A., Bryant, J.M., Harris, S.R., Schuenemann, V.J., Campbell, T.J., Majander, K., Wilbur, A.K., Guichon, R.A., Wolfe Steadman, D.L., Cook, D.C., Niemann, S., Behr, M.A., Zumarraga, M., Bastida, R., Huson, D., Nieselt, K., Young, D., Parkhill, J., Buikstra, J.E., Gagneux, S., Stone, A.C., Krause, J., 2014. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature 514, 494−497. https://doi.org/10.1038/nature13591.

Bottai, D., Frigui, W., Sayes, F., Di Luca, M., Spadoni, D., Pawlik, A., Zoppo, M., Orgeur, M., Khanna, V., Hardy, D., Mangenot, S., Barbe, V., Medigue, C., Ma, L., Bouchier, C., Tavanti, A., Larrouy-Maumus, G., Brosch, R., 2020. TbD1 deletion as a driver of the evolutionary success of modern epidemic *Mycobacterium tuberculosis* lineages. Nat. Commun. 11, 684. https://doi.org/10.1038/s41467-020-14508-5.

Brooks-Pollock, E., Roberts, G.O., Keeling, M.J., 2014. A dynamic model of bovine tuberculosis spread and control in Great Britain. Nature 511, 228−231. https://doi.org/10.1038/nature13529.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., Cole, S.T., 2002. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. Proc. Nat. Acad. Sci. U.S.A. 99, 3684−3689. https://doi.org/10.1073/pnas.052548299.

Brudey, K., Driscoll, J., Rigouts, L., Prodinger, W.M., Gori, A., Al-Hajoj, S.A.M., Allix, C., Aristimuno, L., Arora, J., Baumanis, V., Binder, L., Cafrune, P., Cataldi, A., Cheong, S., Diel, R., Ellermeier, C., Evans, J.T., Fauville-Dufaux, M., Ferdinand, S., Garcia de Viedma, D., Garzelli, C., Gazzola, L., Gomes, H.M., Gutierrez, M.C., Hawkey, P.M., van Helden, P.D., Kadival, G.V., Kreiswirth, B.N., Kremer, K., Kubin, M., Kulkarni, S.P., Liens, B., Lillebaek, T., Ly, H.M., Martin, C., Martin, C., Mokrousov, I., Narvskaia, O., Ngeow, Y.F., Naumann, L., Niemann, S., Parwati, I., Rahim, M.Z., Rasolofo-Razanamparany, V., Rasolonavalona, T., Rossetti, M.L., Rüsch-Gerdes, S., Sajduda, A., Samper, S., Shemyakin, I.G., Singh, U.B., Somoskovi, A., Skuce, R., Van Soolingen, D., Streicher, E.M., Suffys, P.N., Tortoli, E., Tracevska, T., Vincent, V., Victor, T.C., Warren, R., Yap, S.F., Zaman, K., Portaels, F., Rastogi, N., Sola, C., 2006. *Mycobacterium tuberculosis* complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics, and epidemiology. BMC Microbiol. 6 (doi), 23. https://doi.org/10.1186/1471-2180-6-23.

Brudey, K., Gutierrez, M.C., Vincent, V., Parsons, L.M., Salfinger, M., Rastogi, N., Sola, C., 2004. Mycobacterium africanum genotyping using novel spacer oligonucleotides in the direct repeat locus. J. Clin. Microbiol. 42, 5053−5057. https://doi.org/10.1128/JCM.42.11.5053-5057.2004.

Bryant, J.M., Schürch, A.C., van Deutekom, H., Harris, S.R., de Beer, J.L., de Jager, V., Kremer, K., van Hijum, S.A., Siezen, R.J., Borgdorff, M., Bentley, S.D., Parkhill, J., van Soolingen, D., 2013. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. BMC Infect. Dis. 13, 110. https://doi.org/10.1186/1471-2334-13-110.

Brynildsrud, O.B., Pepperell, C.S., Suffys, P., Grandjean, L., Monteserin, J., Debech, N., Bohlin, J., Alfsnes, K., Pettersson, J.O., Kirkeleite, I., Fandinho, F., da Silva, M.A., Perdigão, J., Portugal, I., Viveiros, M., Clark, T., Caws, M., Dunstan, S., Thai, P.V.K., Lopez, B., Ritacco, V., Kitchen, A., Brown, T.S., van Soolingen, D., O'Neill, M.B., Holt, K.E., Feil, E.J., Mathema, B., Balloux, F., Eldholm, V., 2018. Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. Sci. Adv. 4, eaat5869. https://doi.org/10.1126/sciadv.aat5869.

Ca, B., Fonseca, K.L., Sousa, J., Maceiras, A.R., Machado, D., Sanca, L., Rabna, P., Rodrigues, P.N.S., Viveiros, M., Saraiva, M., 2019. Experimental evidence for limited in vivo virulence of *Mycobacterium africanum*. Front. Microbiol. 10, 2102. https://doi.org/10.3389/fmicb.2019.02102.

Cardona, P.J., Català, M., Prats, C., 2020. Origin of tuberculosis in the Paleolithic predicts unprecedented population growth and female resistance. Sci. Rep. 10, 42. https://doi.org/10.1038/s41598-019-56769-1.

Casali, N., Nikolayevskyy, V., Balabanova, Y., Harris, S.R., Ignatyeva, O., Kontsevaya, I., Corander, J., Bryant, J., Parkhill, J., Nejentsev, S., Horstmann, R.D., Brown, T., Drobniewski, F., 2014. Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nat. Genet. 46, 279−286. https://doi.org/10.1038/ng.2878.

Casonato, S., Cervantes Sanchez, A., Haruki, H., Rengifo Gonzalez, M., Provvedi, R., Dainese, E., Jaouen, T., Gola, S., Bini, E., Vicente, M., Johnsson, K., Ghisotti, D., Palu, G., Hernandez-Pando, R., Manganelli, R., 2012. WhiB5, a transcriptional regulator that contributes to *Mycobacterium tuberculosis* virulence and reactivation. Infect. Immun. 80, 3132−3144. https://doi.org/10.1128/IAI.06328-11.

Castets, M., Boisvert, H., Grumbach, F., Brunel, M., Rist, N., 1968. Les bacilles tuberculeux de type africain: note préliminaire. Rev. Tuberc. Pneumol. 32, 179−184.

Cavalli-Sforza, L.L., 2000. Genes, Peoples and Languages. Farrar, Straus and Giroux, New York.

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. Chapter 2: genetic history of world populations, 2.5: comparison with archaeological data. In: Luca Cavalli-Sforza, L., Menozzi, P., Piazza, A. (Eds.), The History and Geography of Human Genes. Princeton University Press, Princeton, New Jersey, pp. 93−96.

Chuang, P.C., Chen, Y.M., Chen, H.Y., Jou, R., 2010. Single nucleotide polymorphisms in cell wall biosynthesis-associated genes and phylogeny of Mycobacterium tuberculosis lineages. Infect. Genet. Evol. 10 (4), 459−466. https://doi.org/10.1016/j.meegid.2010.03.003. PMID: 20223296.

Clavel-Seres, S., Clement, F., 1984. Distribution of lysotypes of *Mycobacterium tuberculosis* in relation to the country of origin of the patient. Ann. Microbiol. (Paris) 135B, 35−44.

Click, E.S., Winston, C.A., Oeltmann, J.E., Moonan, P.K., Mac Kenzie, W.R., 2013. Association between *Mycobacterium tuberculosis* lineage and time to sputum culture conversion. Int. J. Tuberc. Lung Dis. 17, 878−884. https://doi.org/10.5588/ijtld.12.0732.

Cock, P.J.A., Fields, C.J., Goto, N., Heuer, M.L., Rice, P.M., 2009. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 38, 1767−1771. https://doi.org/10.1093/nar/gkp1137.

Cockburn, A., 1963. The Evolution and Eradication of Infectious Diseases. John Hopkins Press, Baltimore.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry 3rd, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393, 537−544. https://doi.org/10.1038/31159.

Coll, F., McNerney, R., Guerra-Assuncao, J.A., Glynn, J.R., Perdigão, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., Clark, T.G., 2014a. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nat. Commun. 5, 4812. https://doi.org/10.1038/ncomms5812.

Coll, F., Preston, M., Guerra-Assuncao, J.A., Hill-Cawthorn, G., Harris, D., Perdigão, J., Viveiros, M., Portugal, I., Drobniewski, F., Gagneux, S., Glynn, J.R., Pain, A., Parkhill, J., McNerney, R., Martin, N., Clark, T.G., 2014b. PolyTB: a genomic variation map for *Mycobacterium tuberculosis*. Tuberculosis 94 (3), 346−354. https://doi.org/10.1016/j.tube.2014.02.005.

Coll, F., Phelan, J., Hill-Cawthorne, G.A., Nair, M.B., Mallard, K., Ali, S., Abdallah, A.M., Alghamdi, S., Alsomali, M., Ahmed, A.O., Portelli, S., Oppong, Y., Alves, A., Bessa, T.B., Campino, S., Caws, M., Chatterjee, A., Crampin, A.C., Dheda, K., Furnham, N., Glynn, J.R., Grandjean, L., Minh Ha, D., Hasan, R., Hasan, Z., Hibberd, M.L., Joloba, M., Jones-Lopez, E.C., Matsumoto, T., Miranda, A., Moore, D.J., Mocillo, N., Panaiotov, S., Parkhill, J., Penha, C., Perdigão, J., Portugal, I., Rchiad, Z., Robledo, J., Sheen, P., Shesha, N.T., Sirgel, F.A., Sola, C., Oliveira Sousa, E., Streicher, E.M., Helden, P.V., Viveiros, M., Warren, R.M., McNerney, R., Pain, A., Clark, T.G., 2018. Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. Nat. Genet. 50, 307−316. https://doi.org/10.1038/s41588-017-0029-0.

Comas, I., Gagneux, S., 2009. The past and future of tuberculosis research. PLoS Pathog. 5, e1000600. https://doi.org/10.1371/journal.ppat.1000600.

Comas, I., Homolka, S., Niemann, S., Gagneux, S., 2009. Genotyping of genetically monomorphic bacteria: DNA sequencing in *Mycobacterium tuberculosis* highlights the limitations of current methodologies. PLoS One 4, e7815. https://doi.org/10.1371/journal.pone.0007815.

Comas, I., Coscolla, M., Luo, T., Borrell, S., Holt, K.E., Kato-Maeda, M., Parkhill, J., Malla, B., Berg, S., Thwaites, G., Yeboah-Manu, D., Bothamley, G., Mei, J., Wei, L., Bentley, S., Harris, S.R., Niemann, S., Diel, R., Aseffa, A., Gao, Q., Young, D., Gagneux, S., 2013. Out-of-Africa migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. Nat. Genet. 45, 1176−1182. https://doi.org/10.1038/ng.2744.

Comas, I., Hailu, E., Kiros, T., Bekele, S., Mekonnen, W., Gumi, B., Tschopp, R., Ameni, G., Hewinson, R.G., Robertson, B.D., Goig, G.A., Stücki, D., Gagneux, S., Aseffa, A., Young, D., Berg, S., 2015. Population genomics of *Mycobacterium tuberculosis* in Ethiopia contradicts the virgin soil hypothesis for human tuberculosis in sub-Saharan Africa. Curr. Biol. 25, 3260−3266. https://doi.org/10.1016/j.cub.2015.10.061.

Comin, J., Monforte, M.L., Samper, S., , Aragonese Working Group on Molecular Epidemiology of Tuberculosis (EPIMOLA), Otal, I., 2021. Analysis of *Mycobacterium africanum* in the last 17 years in Aragon identifies a specific location of IS6110 in Lineage 6. Sci. Rep. 11, 10359. https://doi.org/10.1038/s41598-021-89511-x.

Coqueugniot, H., Pálfi, G., Gely, B., Dutour, O., 2023. Upper Paleolithic tuberculosis: a probable case illustrated by paleoimaging methods (Azilian site of Les Iboussieres, France). Tuberculosis 143S, 102373.

Coscolla, M., Lewin, A., Metzger, S., Maetz-Rennsing, K., Calvignac-Spencer, S., Nitsche, A., Dabrowski, P.W., Radonic, A., Niemann, S., Parkhill, J., Couacy-Hymann, E., Feldman, J., Comas, I., Boesch, C., Gagneux, S., Leendertz, F.H., 2013. Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. Emerg. Infect. Dis. 19, 969−976. https://doi.org/10.3201/eid1906.121012.

Coscolla, M., Gagneux, S., Menardo, F., Loiseau, C., Ruiz-Rodriguez, P., Borrell, S., Otchere, I.D., Asante-Poku, A., Asare, P., Sanchez-Buso, L., Gehre, F., Sanoussi, C.N., Antonio, M., Affolabi, D., Fyfe, J., Beckert, P., Niemann, S., Alabi, A.S., Grobusch, M.P., Kobbe, R., Parkhill, J., Beisel, C., Fenner, L., Bottger, E.C., Meehan, C.J., Harris, S.R., de Jong, B.C., Yeboah-Manu, D., Brites, D., 2021. Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. Microb. Genom. 7. https://doi.org/10.1099/mgen.0.000477.

Couvin, D., David, A., Zozio, T., Rastogi, N., 2019a. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. Infect. Genet. Evol. 72, 31−43. https://doi.org/10.1016/j.meegid.2018.12.030.

Couvin, D., Reynaud, Y., Rastogi, N., 2019b. Two tales: worldwide distribution of Central Asian (CAS) versus ancestral East-African Indian (EAI) lineages of *Mycobacterium tuberculosis* underlines a remarkable cleavage for phylogeographical, epidemiological and demographical characteristics. PLoS One 14, e0219706. https://doi.org/10.1371/journal.pone.0219706.

Allix-Beguec, C., Arandjelovic, I., Bi, L., Beckert, P., Bonnet, M., Bradley, P., Cabibbe, A.M., Cancino-Munoz, I., Caulfield, M.J., Chaiprasert, A., Cirillo, D.M., Clifton, D.A., Comas, I., Crook, D.W., De Filippo, M.R., de Neeling, H., Diel, R., Drobniewski, F.A., Faksri, K., Farhat, M.R., Fleming, J., Fowler, P., Fowler, T.A., Gao, Q., Gardy, J., Gascoyne-Binzi, D., Gibertoni-Cruz, A.L., Gil-Brusola, A., Golubchik, T., Gonzalo, X., Grandjean, L., He, G., Guthrie, J.L., Hoosdally, S., Hunt, M., Iqbal, Z., Ismail, N., Johnston, J., Khanzada, F.M., Khor, C.C., Kohl, T.A., Kong, C., Lipworth, S., Liu, Q., Maphalala, G., Martinez, E., Mathys, V., Merker, M., Miotto, P., Mistry, N., Moore, D.A.J., Murray, M., Niemann, S., Omar, S.V., Ong, R.T., Peto, T.E.A., Posey, J.E., Prammananan, T., Pym, A., Rodrigues, C., Rodrigues, M., Rodwell, T., Rossolini, G.M., Sanchez Padilla, E., Schito, M., Shen, X., Shendure, J., Sintchenko, V., Sloutsky, A., Smith, E.G., Snyder, M., Soetaert, K., Starks, A.M., Supply, P., Suriyapol, P., Tahseen, S., Tang, P., Teo, Y.Y., Thuong, T.N.T., Thwaites, G., Tortoli, E., van Soolingen, D., Walker, A.S., Walker, T.M., Wilcox, M., Wilson, D.J., Wyllie, D., Yang, Y., Zhang, H., Zhao, Y., Zhu, B., CRyPTIC Consortium and the 100,000 Genomes Project, 2018. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. N. Engl. J. Med. 379, 1403−1415. https://doi.org/10.1056/NEJMoa1800474.

CRyPTIC, Consortium, 2022. Genome-wide association studies of global *Mycobacterium tuberculosis* resistance to 13 antimicrobials in 10,228 genomes identify new resistance mechanisms. PLoS Biol. 20, e3001755. https://doi.org/10.1371/journal.pbio.3001755.

Dale, J.W., Al-Ghusein, H., Al-Hashmi, S., Butcher, P., Dickens, A.L., Drobniewski, F., Forbes, K.J., Gillespie, S.H., Lamprecht, D., McHugh, T.D., Pitman, R., Rastogi, N., Smith, A.T., Sola, C., Yesilkaya, H., 2003. Evolutionary relationships among strains of *Mycobacterium tuberculosis* with few copies of IS6110. J. Bacteriol. 185, 2555−2562. https://doi.org/10.1128/JB.185.8.2555-2562.2003.

Dale, J.W., Banerjee, D., Gillespie, S.H., Maguire, H., 1997. Molecular Epidemiology of Tuberculosis in London, Combined Meeting of the Concerted Actions on Tuberculosis.

David, H.L., Jahan, M.T., Jumin, A., Grandry, J., Lehmann, E.H., 1978. Numerical taxonomy of *Mycobacterium africanum*. Int. J. System. Bacteriol. 28, 467−472.

Davies, N., 1999. The Isles: A History. Oxford University Press, New York.
de Jong, B.C., Hill, P.C., Brookes, R.H., Gagneux, S., Jeffries, D.J., Otu, J.K., Donkor, S.A., Fox, A., McAdam, K.P., Small, P.M., Adegbola, R.A., 2006. Mycobacterium africanum elicits an attenuated T cell response to early secreted antigenic target, 6 kDa, in patients with tuberculosis and their household contacts. J. Infect. Dis. 193, 1279−1286. https://doi.org/10.1086/502977.
de Jong, B.C., Antonio, M., Gagneux, S., 2010. Mycobacterium africanum–review of an important cause of human tuberculosis in West Africa. PLoS Neglected Trop. Dis. 4, e744. https://doi.org/10.1371/journal.pntd.0000744.
Demay, C., Liens, B., Burguière, T., Hill, V., Couvin, D., Millet, J., Mokrousov, I., Sola, C., Zozio, T., Rastogi, N., 2012. SITVITWEB − a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Infect. Genet. Evol. 12, 755−766. https://doi.org/10.1016/j.meegid.2012.02.004.
Diarra, B., Kone, M., Togo, A.C.G., Sarro, Y.D.S., Cisse, A.B., Somboro, A., Degoga, B., Tolofoudie, M., Kone, B., Sanogo, M., Baya, B., Kodio, O., Maiga, M., Belson, M., Orsega, S., Krit, M., Dao, S., Maiga II, Murphy, R.L., Rigouts, L., Doumbia, S., Diallo, S., de Jong, B.C., 2018. Mycobacterium africanum (Lineage 6) shows slower sputum smear conversion on tuberculosis treatment than *Mycobacterium tuberculosis* (Lineage 4) in Bamako, Mali. PLoS One 13, e0208603. https://doi.org/10.1371/journal.pone.0208603.
Ding, P., Li, X., Jia, Z., Lu, Z., 2017. Multidrug-resistant tuberculosis (MDR-TB) disease burden in China: a systematic review and spatio-temporal analysis. BMC Infect. Dis. 17, 57. https://doi.org/10.1186/s12879-016-2151-5.
Dippenaar, A., Parsons, S.D.C., Sampson, S.L., van der Merwe, R.G., Drewe, J.A., Abdallah, A.M., Siame, K.K., Gey van Pittius, N.C., van Helden, P.D., Pain, A., Warren, R.M., 2015. Whole genome sequence analysis of *Mycobacterium suricattae*. Tuberculosis 95, 682−688. https://doi.org/10.1016/j.tube.2015.10.001.
Domenech, P., Reed, M.B., Barry 3rd, C.E., 2005. Contribution of the *Mycobacterium tuberculosis* MmpL protein family to virulence and drug resistance. Infect. Immun. 73, 3492−3501. https://doi.org/10.1128/IAI.73.6.3492-3501.2005.
Dos Vultos, T., Mestre, O., Rauzier, J., Golec, M., Rastogi, N., Rasolofo, V., Tonjum, T., Sola, C., Matic, I., Gicquel, B., 2008. Evolution and diversity of clonal bacteria: the paradigm of *Mycobacterium tuberculosis*. PLoS One 3, e1538. https://doi.org/10.1371/journal.pone.0001538.
Dubiley, S., Kirillov, E., Ignatova, A., Stepanshina, V., Shemyakin, I., 2007. Molecular characteristics of the *Mycobacterium tuberculosis* LAM-RUS family prevalent in Central Russia. J. Clin. Microbiol. 45, 4036−4038. https://doi.org/10.1128/JCM.01217-07.
Dutour, O., 2023. The paleopathology and paleoepidemiology of upper paleolithic tuberculosis: review of evidence and hypotheses. Tuberculosis 143S, 102348.
Easterbrook, P.J., Gibson, A., Murad, S., Lamprecht, D., Ives, N., Ferguson, A., Lowe, O., Mason, P., Ndudzo, A., Taziwa, A., Makombe, R., Mbengeranwa, L., Sola, C., Rastogi, N., Drobniewski, F., 2004. High rates of clustering of strains causing tuberculosis in Harare, Zimbabwe: a molecular epidemiological study. J. Clin. Microbiol. 42, 4536−4544. https://doi.org/10.1128/JCM.42.10.4536-4544.2004.
Eddy, J.J., 2015. The ancient city of Rome, its empire, and the spread of tuberculosis in Europe. Tuberculosis 95 (Suppl. 1), S23−S28. https://doi.org/10.1016/j.tube.2015.02.005.
Ehrt, S., Schnappinger, D., Rhee, K.Y., 2018. Metabolic principles of persistence and pathogenicity in *Mycobacterium tuberculosis*. Nat. Rev. Microbiol. 16, 496−507. https://doi.org/10.1038/s41579-018-0013-4.
El Achkar, S., Demanche, C., Osman, M., Rafei, R., Ismail, M.B., Yaacoub, H., Pincon, C., Duthoy, S., De Matos, F., Gaudin, C., Trovato, A., Cirillo, D.M., Hamze, M., Supply, P., 2019. Drug-resistant tuberculosis, Lebanon, 2016 - 2017. Emerg. Infect. Dis. 25, 564−568. https://doi.org/10.3201/eid2503.181375.

Eldholm, V., Monteserin, J., Rieux, A., Lopez, B., Sobkowiak, B., Ritacco, V., Balloux, F., 2015. Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. Nat. Commun. 6, 7119. https://doi.org/10.1038/ncomms8119.

Eldholm, V., Pettersson, J.H., Brynildsrud, O.B., Kitchen, A., Rasmussen, E.M., Lillebaek, T., Ronning, J.O., Crudu, V., Mengshoel, A.T., Debech, N., Alfsnes, K., Bohlin, J., Pepperell, C.S., Balloux, F., 2016. Armed conflict and population displacement as drivers of the evolution and dispersal of *Mycobacterium tuberculosis*. Proc. Nat. Acad. Sci. U.S.A. 113, 13881−13886. https://doi.org/10.1073/pnas.1611283113.

Endrullat, C., Glokler, J., Franke, P., Frohme, M., 2016. Standardization and quality management in next-generation sequencing. Appl. Transl. Genom. 10, 2−9. https://doi.org/10.1016/j.atg.2016.06.001.

Evans, J.T., Serafino Wani, R.L., Anderson, L., Gibson, A.L., Smith, E.G., Wood, A., Olowokure, B., Abubakar, I., Mann, J.S., Gardiner, S., Jones, H., Sonnenberg, P., Hawkey, P.M., 2011. A geographically-restricted but prevalent *Mycobacterium tuberculosis* strain identified in the West Midlands Region of the UK between 1995 and 2008. PLoS One 6, e17930. https://doi.org/10.1371/journal.pone.0017930.

Fabre, M., Hauck, Y., Soler, C., Koeck, J.L., van Ingen, J., van Soolingen, D., Vergnaud, G., Pourcel, C., 2010. Molecular characteristics of "*Mycobacterium canettii*" the smooth *Mycobacterium tuberculosis* bacilli. Infect. Genet. Evol. 10, 1165−1173. https://doi.org/10.1016/j.meegid.2010.07.016.

Faksri, K., Xia, E., Ong, R.T., Tan, J.H., Nonghanphithak, D., Makhao, N., Thamnongdee, N., Thanormchat, A., Phurattanakornkul, A., Rattanarangsee, S., Ratanajaraya, C., Suriyaphol, P., Prammananan, T., Teo, Y.Y., Chaiprasert, A., 2018. Comparative whole-genome sequence analysis of *Mycobacterium tuberculosis* isolated from tuberculous meningitis and pulmonary tuberculosis patients. Sci. Rep. 8, 4910.

Fargettas, J., 2012. Les Tirailleurs sénégalais. Les soldats noirs entre légendes et réalités. 1939−1945. Tallandier, Paris.

Filliol, I., Driscoll, J.R., Van Soolingen, D., Kreiswirth, B.N., Kremer, K., Valétudie, G., Anh, D.D., Barlow, R., Banerjee, D., Bifani, P.J., Brudey, K., Cataldi, A., Cooksey, R.C., Cousins, D.V., Dale, J.W., Dellagostin, O.A., Drobniewski, F., Engelmann, G., Ferdinand, S., Gascoyne-Binzi, D., Gordon, M., Gutierrez, M.C., Haas, W., Heersma, H., Källenius, G., Kassa-Kelembho, E., Koivula, T., Ly, H.M., Makristathis, A., Mammina, C., Martin, G., Moström, P., Mokrousov, I., Narbonne, V., Narvskaya, O., Nastasi, A., Ngo Niobe Eyangoh, S., Pape, J.W., Rasolofo-Razanamparany, V., Ridell, M., Rossetti, M.L., Stauffer, F., Suffys, P.N., Takiff, H., Texier-Maugein, J., Vincent, V., de Waard, J., Sola, C., Rastogi, N., 2002. Global distribution of *Mycobacterium tuberculosis* spoligotypes. Emerg. Inf. Dis. 8, 1347−1350. https://doi.org/10.3201/eid0811.020125.

Filliol, I., Motiwala, A.S., Cavatore, M., Qi, W., Hazbon, M.H., Bobadilla del Valle, M., Fyfe, J., Garcia-Garcia, L., Rastogi, N., Sola, C., Zozio, T., Guerrero, M.I., Leon, C.I., Crabtree, J., Angiuoli, S., Eisenach, K.D., Durmaz, R., Joloba, M.L., Rendon, A., Sifuentes-Osornio, J., Ponce de Leon, A., Cave, M.D., Fleischmann, R., Whittam, T.S., Alland, D., 2006. Global phylogeny of *Mycobacterium tuberculosis* based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J. Bacteriol. 188, 759−772. https://doi.org/10.1128/JB.188.2.759-772.2006.

Firdessa, R., Berg, S., Hailu, E., Schelling, E., Gumi, B., Erenso, G., Gadisa, E., Kiros, T., Habtamu, M., Hussein, J., Zinsstag, J., Robertson, B.D., Ameni, G., Lohan, A.J., Loftus, B., Comas, I., Gagneux, S., Tschopp, R., Yamuah, L., Hewinson, G., Gordon, S.V., Young, D.B., Aseffa, A., 2013. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. Emerg. Infect. Dis. 19, 460−463.

Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs Jr., W.R., Venter, J.C., Fraser, C.M., 2002. Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. J. Bacteriol. 184, 5479−5490. https://doi.org/10.1128/JB.184.19.5479-5490.2002.

Folkvardsen, D.B., Norman, A., Andersen, A.B., Rasmussen, E.M., Lillebaek, T., Jelsbak, L., 2018. A major *Mycobacterium tuberculosis* outbreak caused by one specific genotype in a low-incidence country: exploring gene profile virulence explanations. Sci. Rep. 8, 11869. https://doi.org/10.1038/s41598-018-30363-3.

Fomukong, N.G., Tang, T.H., Al-Maamary, S., Ibrahim, W.A., Ramayah, S., Yates, M., Zainuddin, Z.F., Dale, J.W., 1994. Insertion sequence typing of *Mycobacterium tuberculosis*: characterization of a widespread subtype with a single copy of IS6110. Tuber. Lung Dis. 75, 435−440. https://doi.org/10.1016/0962-8479(94)90117-1.

Ford, C.B., Shah, R.R., Maeda, M.K., Gagneux, S., Murray, M.B., Cohen, T., Johnston, J.C., Gardy, J., Lipsitch, M., Fortune, S.M., 2013. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. Nat. Genet. 45, 784−790. https://doi.org/10.1038/ng.2656.

Forrellad, M.A., Klepp, L.I., Gioffre, A., Sabio y Garcia, J., Morbidoni, H.R., de la Paz Santangelo, M., Cataldi, A.A., Bigi, F., 2013. Virulence factors of the *Mycobacterium tuberculosis* complex. Virulence 4, 3−66. https://doi.org/10.4161/viru.22329.

Freeman, R., Kato-Maeda, M., Hauge, K.A., Horan, K.L., Oren, E., Narita, M., Wallis, C.K., Cave, D., Nolan, C.M., Small, P.M., Cangelosi, G.A., 2005. Use of rapid genomic deletion typing to monitor a tuberculosis outbreak within an urban homeless population. J. Clin. Microbiol. 43, 5550−5554. https://doi.org/10.1128/JCM.43.11.5550-5554.2005.

Freschi, L., Vargas Jr., R., Husain, A., Kamal, S.M.M., Skrahina, A., Tahseen, S., Ismail, N., Barbova, A., Niemann, S., Cirillo, D.M., Dean, A.S., Zignol, M., Farhat, M.R., 2021. Population structure, biogeography and transmissibility of *Mycobacterium tuberculosis*. Nat. Commun. 12, 6099. https://doi.org/10.1038/s41467-021-26248-1.

Fusegawa, H., Wang, B.H., Sakurai, K., Nagasawa, K., Okauchi, M., Nagakura, K., 2003. Outbreak of tuberculosis in a 2000-year-old Chinese population. Kansenshogaku zasshi. J. Jpn. Assoc. Infect. Dis. 77, 146−149.

Gagneux, S., DeRiemer, K., Van, T., Kato-Maeda, M., de Jong, B.C., Narayanan, S., Nicol, M., Niemann, S., Kremer, K., Gutierrez, M.C., Hilty, M., Hopewell, P.C., Small, P.M., 2006. Variable host-pathogen compatibility in *Mycobacterium tuberculosis*. Proc. Natl. Acad. Sci. U.S.A. 103, 2869−2873. https://doi.org/10.1073/pnas.0511240103.

Gagneux, S., Small, P.M., 2007. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. Lancet Infect. Dis. 7, 328−337. https://doi.org/10.1016/S1473-3099(07)70108-1.

Gagneux, S., 2012. Host-pathogen coevolution in human tuberculosis. Philos. Trans. R. Soc. B Biol. Sci. 367, 850−859. https://doi.org/10.1098/rstb.2011.0316.

Gardy, J.L., Johnston, J.C., Ho Sui, S.J., Cook, V.J., Shah, L., Brodkin, E., Rempel, S., Moore, R., Zhao, Y., Holt, R., Varhol, R., Birol, I., Lem, M., Sharma, M.K., Elwood, K., Jones, S.J., Brinkman, F.S., Brunham, R.C., Tang, P., 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N. Engl. J. Med. 364, 730−739. https://doi.org/10.1056/NEJMoa1003176.

Gehre, F., Antonio, M., Faihun, F., Odoun, M., Uwizeye, C., de Rijk, P., de Jong, B.C., Affolabi, D., 2013. The first phylogeographic population structure and analysis of transmission dynamics of *M. africanum* West African 1–combining molecular data from Benin, Nigeria and Sierra Leone. PLoS One 8, e77000. https://doi.org/10.1371/journal.pone.0077000.

Gey van Pittius, N.C., Sampson, S.L., Lee, H., Kim, Y., van Helden, P.D., Warren, R.M., 2006. Evolution and expansion of the *Mycobacterium tuberculosis* PE and PPE multigene families and their association with the duplication of the ESAT-6 (esx) gene cluster regions. BMC Evol. Biol. 6, 95. https://doi.org/10.1186/1471-2148-6-95.

Gibson, A.L., Huard, R.C., Gey van Pittius, N.C., Lazzarini, L.C., Driscoll, J., Kurepina, N., Zozio, T., Sola, C., Spindola, S.M., Kritski, A.L., Fitzgerald, D., Kremer, K., Mardassi, H., Chitale, P., Brinkworth, J., Garcia de

Viedma, D., Gicquel, B., Pape, J.W., van Soolingen, D., Kreiswirth, B.N., Warren, R.M., van Helden, P.D., Rastogi, N., Suffys, P.N., Lapa e Silva, J., Ho, J.L., 2008. Application of sensitive and specific molecular methods to uncover global dissemination of the major RDRio Sublineage of the Latin American-Mediterranean Mycobacterium tuberculosis spoligotype family. J. Clin. Microbiol. 46 (4), 1259−1267. https://doi.org/10.1128/JCM.02231-07. PMID: 18234868.

Gisch, N., Utpatel, C., Gronbach, L.M., Kohl, T.A., Schombel, U., Malm, S., Dobos, K.M., Hesser, D.C., Diel, R., Gotsch, U., Gerdes, S., Shuaib, Y.A., Ntinginya, N.E., Khosa, C., Viegas, S., Kerubo, G., Ali, S., Al-Hajoj, S.A., Ndung'u, P.W., Rachow, A., Hoelscher, M., Maurer, F.P., Schwudke, D., Niemann, S., Reiling, N., Homolka, S., 2022. Sub-lineage specific phenolic glycolipid patterns in the *Mycobacterium tuberculosis* complex lineage 1. Front. Microbiol. 13, 832054. https://doi.org/10.3389/fmicb.2022.832054.

Glaziou, P., 2020. Predicted impact of the COVID-19 pandemic on global tuberculosis deaths in 2020. medRxiv. https://doi.org/10.1101/2020.04.28.20079582.

Glynn, J.R., Whiteley, J., Bifani, P.J., Kremer, K., Van Soolingen, D., 2002. Worldwide occurrence of Beijing/W strains of *Mycobacterium tuberculosis*: a systematic review. Emerg. Infect. Dis. 8, 843−849. https://doi.org/10.3201/eid0805.020002.

Godfroid, M., Dagan, T., Kupczok, A., 2018. Recombination signal in *Mycobacterium tuberculosis* stems from reference-guided assemblies and alignment artefacts. Genome Biol. Evol. 10, 1920−1926. https://doi.org/10.1093/gbe/evy143.

Goguet de la Salmoniere, Y.O., Li, H.M., Torrea, G., Bunschoten, A., van Embden, J., Gicquel, B., 1997. Evaluation of spoligotyping in a study of the transmission of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 35, 2210−2214. https://doi.org/10.1128/jcm.35.9.2210-2214.1997.

Goh, K.S., Legrand, E., Sola, C., Rastogi, N., 2001. Rapid differentiation of "Mycobacterium canettii" from other *Mycobacterium tuberculosis* complex organisms by PCR-restriction analysis of the hsp65 gene. J. Clin. Microbiol. 39, 3705−3708. https://doi.org/10.1128/JCM.39.10.3705-3708.2001.

Gomgnimbou, M.K., Refrégier, G., Diagbouga, S.P., Adama, S., Kabore, A., Ouiminga, A., Sola, C., 2012. Spoligotyping of *Mycobacterium africanum*, Burkina Faso. Emerg. Infect. Dis. 18, 117−119. https://doi.org/10.3201/eid1801.110275.

Goren, M.B., Brokl, O., Schaefer, W.B., 1974. Lipids of putative relevance to virulence in *Mycobacterium tuberculosis*: phthiocerol dimycocerosate and the attenuation indicator lipid. Infect. Immun. 9, 150−158. https://doi.org/10.1128/iai.9.1.150-158.1974.

Grabowska, A.D., Brison, Y., Maveyraud, L., Gavalda, S., Faille, A., Nahoum, V., Bon, C., Guilhot, C., Pedelacq, J.D., Chalut, C., Mourey, L., 2020. Molecular basis for extender unit specificity of mycobacterial polyketide synthases. ACS Chem. Biol. 15, 3206−3216. https://doi.org/10.1021/acschembio.0c00772.

Groenen, P.M., Bunschoten, A.E., van Soolingen, D., van Embden, J.D., 1993. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. Mol. Microbiol. 10, 1057−1065. https://doi.org/10.1111/j.1365-2958.1993.tb00976.x.

Gutacker, M.M., Mathema, B., Soini, H., Shashkina, E., Kreiswirth, B.N., Graviss, E.A., Musser, J.M., 2006. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. J. Infect. Dis. 193, 121−128. https://doi.org/10.1086/498574.

Gutierrez, M.C., Brisse, S., Brosch, R., Fabre, M., Omais, B., Marmiesse, M., Supply, P., Vincent, V., 2005. Ancient origin and gene mosaicism of the progenitor of *Mycobacterium tuberculosis*. PLoS Pathog. 1, e5. https://doi.org/10.1371/journal.ppat.0010005.

Guyeux, C., Senelle, G., Refrégier, G., Bretelle-Establet, F., Cambau, E., Sola, C., 2022. Connection between two historical tuberculosis outbreak sites in Japan, Honshu, by a new ancestral *Mycobacterium tuberculosis* L2 sublineage. Epidemiol. Infect. 150, 1−10. https://doi.org/10.1017/S0950268822000048.

Guyeux, C., Sola, C., Nous, C., Refrégier, G., 2021. CRISPRbuilder-TB: "CRISPR-builder for tuberculosis". Exhaustive reconstruction of the CRISPR locus in *Mycobacterium tuberculosis* complex using SRA. PLoS Comput. Biol. 17, e1008500. https://doi.org/10.1371/journal.pcbi.1008500.

Guyeux, C., Senelle, S., Le Meur, A., Supply, P., Gaudin, C., Phelan, J.E., Clark, T.G., Rigouts, L., de Jong, B., Sola, C., Refrégier, G., 2024. *Mycobacterium africanum* evolutionary history: insights from a new lineage identified in Central Africa. In: Emerging Infectious Diseases. March Issue (in press).

Gygli, S.M., Keller, P.M., Ballif, M., Blochliger, N., Homke, R., Reinhard, M., Loiseau, C., Ritter, C., Sander, P., Borrell, S., Collantes Loo, J., Avihingsanon, A., Gnokoro, J., Yotebieng, M., Egger, M., Gagneux, S., Bottger, E.C., 2019. Whole-genome sequencing for drug resistance profile prediction in *Mycobacterium tuberculosis*. Antimicrob. Agents Chemother. 63 (4), e02175−e2218. https://doi.org/10.1128/AAC.02175-18.

Hadifar, S., Kamakoli, M.K., Fateh, A., Siadat, S.D., Vaziri, F., 2019a. Enhancing the differentiation of specific genotypes in *Mycobacterium tuberculosis* population. Sci. Rep. 9, 17946. https://doi.org/10.1038/s41598-019-54393-7.

Hadifar, S., Shamkhali, L., Kargarpour Kamakoli, M., Mostafaei, S., Khanipour, S., Mansoori, N., Fateh, A., Davar Siadat, S., Vaziri, F., 2019b. Genetic diversity of *Mycobacterium tuberculosis* isolates causing pulmonary and extrapulmonary tuberculosis in the capital of Iran. Mol. Phylogenet. Evol. 132, 46−52. https://doi.org/10.1016/j.ympev.2018.11.019.

Haeili, M., Darban-Sarokhalil, D., Fooladi, A.A., Javadpour, S., Hashemi, A., Siavoshi, F., Feizabadi, M.M., 2013. Spoligotyping and drug resistance patterns of isolates from five provinces of Iran. MicrobiologyOpen 2, 988−996. https://doi.org/10.1002/mbo3.139.

Han, S.J., Song, T., Cho, Y.J., Kim, J.S., Choi, S.Y., Bang, H.E., Chun, J., Bai, G.H., Cho, S.N., Shin, S.J., 2015. Complete genome sequence of K from a Korean high school outbreak, belonging to the Beijing family. Stand. Genomic Sci. 10, 78. https://doi.org/10.1186/s40793-015-0071-4.

Hermans, P.W., Messadi, F., Guebrexabher, H., van Soolingen, D., de Haas, P.E., Heersma, H., de Neeling, H., Ayoub, A., Portaels, F., Frommel, D., et al., 1995. Analysis of the population structure of *Mycobacterium tuberculosis* in Ethiopia, Tunisia, and The Netherlands: usefulness of DNA typing for global tuberculosis epidemiology. J. Infect. Dis. 171, 1504−1513. https://doi.org/10.1093/infdis/171.6.1504.

Hershberg, R., Lipatov, M., Small, P.M., Sheffer, H., Niemann, S., Homolka, S., Roach, J.C., Kremer, K., Petrov, D.A., Feldman, M.W., Gagneux, S., 2008. High functional diversity in *Mycobacterium tuberculosis* driven by genetic drift and human demography. PLoS Biol. 6, e311. https://doi.org/10.1371/journal.pbio.0060311.

Hershkovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y., Gernaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E., Bar-Gal, G.K., Spigelman, M., 2008. Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a neolithic settlement in the eastern Mediterranean. PLoS One 3, e3426. https://doi.org/10.1371/journal.pone.0003426.

Heyderman, R.S., Goyal, M., Roberts, P., Ushewokunze, S., Zizhou, S., Marshall, B.G., Makombe, R., Van Embden, J.D., Mason, P.R., Shaw, R.J., 1998. Pulmonary tuberculosis in Harare, Zimbabwe: analysis by spoligotyping. Thorax 53, 346−350. https://doi.org/10.1136/thx.53.5.346.

Homolka, S., Koser, C., Archer, J., Rusch-Gerdes, S., Niemann, S., 2009. Single-nucleotide polymorphisms in Rv2629 are specific for *Mycobacterium tuberculosis* genotypes Beijing and Ghana but not associated with rifampin resistance. J. Clin. Microbiol. 47, 223−226. https://doi.org/10.1128/JCM.01237-08.

Homolka, S., Projahn, M., Feuerriegel, S., Ubben, T., Diel, R., Nubel, U., Niemann, S., 2012. High resolution discrimination of clinical *Mycobacterium tuberculosis* complex strains based on single nucleotide polymorphisms. PLoS One 7, e39855. https://doi.org/10.1371/journal.pone.0039855.

Ignatova, A., Dubiley, S., Stepanshina, V., Shemyakin, I., 2006. Predominance of multi-drug-resistant LAM and Beijing family strains among *Mycobacterium tuberculosis* isolates recovered from prison inmates in Tula Region, Russia. J. Med. Microbiol. 55, 1413−1418. https://doi.org/10.1099/jmm.0.46575-0.

Ismail, M.B., Rafei, R., Dabboussi, F., Hamze, M., 2018. Tuberculosis, war, and refugees: spotlight on the Syrian humanitarian crisis. PLoS Pathog. 14, e1007014. https://doi.org/10.1371/journal.ppat.1007014.

Jankute, M., Nataraj, V., Lee, O.Y., Wu, H.H.T., Ridell, M., Garton, N.J., Barer, M.R., Minnikin, D.E., Bhatt, A., Besra, G.S., 2017. The role of hydrophobicity in tuberculosis evolution and pathogenicity. Sci. Rep. 7, 1315. https://doi.org/10.1038/s41598-017-01501-0.

Kadura, S., King, N., Nakhoul, M., Zhu, H., Theron, G., Koser, C.U., Farhat, M., 2020. Systematic review of mutations associated with resistance to the new and repurposed *Mycobacterium tuberculosis* drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid. J. Antimicrob. Chemother. 75, 2031−2043. https://doi.org/10.1093/jac/dkaa136.

Källenius, G., Koivula, T., Ghebremichael, S., Hoffner, S.E., Norberg, R., Svensson, E., Dias, F., Marklund, B.I., Svenson, S.B., 1999. Evolution and clonal traits of *Mycobacterium tuberculosis* complex in Guinea-Bissau. J. Clin. Microbiol. 37, 3872−3878. https://doi.org/10.1128/JCM.37.12.3872-3878.1999.

Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., van Embden, J., 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. 35, 907−914. https://doi.org/10.1128/jcm.35.4.907-914.1997.

Karlson, A., Lessel, E., 1970. *Mycobacterium bovis* nom. nov. Int. J. Syst. Bacteriol. 20.

Karmakar, M., Trauer, J.M., Ascher, D.B., Denholm, J.T., 2019. Hyper transmission of Beijing lineage *Mycobacterium tuberculosis*: systematic review and meta-analysis. J. Infect. 79, 572−581. https://doi.org/10.1016/j.jinf.2019.09.016.

Kato-Maeda, M., Gagneux, S., Flores, L.L., Kim, E.Y., Small, P.M., Desmond, E.P., Hopewell, P.C., 2011. Strain classification of *Mycobacterium tuberculosis*: congruence between large sequence polymorphisms and spoligotypes. Int. J. Tuberc. Lung Dis. 15, 131−133.

Kato-Maeda, M., Rhee, J.T., Gingeras, T.R., Salamon, H., Drenkow, J., Smittipat, N., Small, P.M., 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res. 11, 547−554. https://doi.org/10.1101/gr.166401.

Keating, L.A., Wheeler, P.R., Mansoor, H., Inwald, J.K., Dale, J., Hewinson, R.G., Gordon, S.V., 2005. The pyruvate requirement of some members of the *Mycobacterium tuberculosis* complex is due to an inactive pyruvate kinase: implications for in vivo growth. Mol. Microbiol. 56, 163−174. https://doi.org/10.1111/j.1365-2958.2005.04524.x.

Kerner, G., Laval, G., Patin, E., Boisson-Dupuis, S., Abel, L., Casanova, J.L., Quintana-Murci, L., 2021. Human ancient DNA analyses reveal the high burden of tuberculosis in Europeans over the last 2,000 years. Am. J. Hum. Genet. 108, 517−524. https://doi.org/10.1016/j.ajhg.2021.02.009.

Khandkar, C., Harrington, Z., Jelfs, P.J., Sintchenko, V., Dobler, C.C., 2015. Epidemiology of peripheral Lymph node tuberculosis and genotyping of *M. tuberculosis* strains: a case-control study. PLoS One 10, e0132400.

Kidenya, B.R., Webster, L.E., Behan, S., Kabangila, R., Peck, R.N., Mshana, S.E., Ocheretina, O., Fitzgerald, D.W., 2014. Epidemiology and genetic diversity of multidrug-resistant tuberculosis in East Africa. Tuberculosis 94, 1−7. https://doi.org/10.1371/journal.pone.0132400.

Kigozi, E., Kasule, G.W., Musisi, K., Lukoye, D., Kyobe, S., Katabazi, F.A., Wampande, E.M., Joloba, M.L., Kateete, D.P., 2018. Prevalence and patterns of rifampicin and isoniazid resistance conferring mutations in *Mycobacterium tuberculosis* isolates from Uganda. PLoS One 13, e0198091. https://doi.org/10.1371/journal.pone.0198091.

Kirby, T., 2021. Global tuberculosis progress reversed by COVID-19 pandemic. Lancet Respir. Med. 9, e118−e119. https://doi.org/10.1016/S2213-2600(21)00496-3.

Klotoe, B.J., Kacimi, S., Costa-Conceicao, E., Gomes, H.M., Barcellos, R.B., Panaiotov, S., Haj Slimene, D., Sikhayeva, N., Sengstake, S., Schuitema, A.R., Akhalaia, M., Alenova, A., Zholdybayeva, E., Tarlykov, P., Anthony, R., Refrégier, G., Sola, C., 2019. Genomic characterization of MDR/XDR-TB in Kazakhstan by a combination of high-throughput methods predominantly shows the ongoing transmission of L2/Beijing 94-32 central Asian/Russian clusters. BMC Infect. Dis. 19, 553. https://doi.org/10.1186/s12879-019-4201-2.

Kohavi, R., Sommerfield, D., Dougherty, J., 1997. Data mining using MLC++: a machine learning library in C++. Int. J. Artif. Intell. Tools 6 (4), 537–566.

Kovalev, S.Y., Kamaev, E.Y., Kravchenko, M.A., Kurepina, N.E., Skorniakov, S.N., 2005. Genetic analysis of *Mycobacterium tuberculosis* strains isolated in Ural region, Russian Federation, by MIRU-VNTR genotyping. Int. J. Tuberc. Lung Dis. 9, 746–752.

Kremer, K., van Soolingen, D., Frothingham, R., Haas, W.H., Hermans, P.W., Martin, C., Palittapongarnpim, P., Plikaytis, B.B., Riley, L.W., Yakrus, M.A., Musser, J.M., van Embden, J.D., 1999. Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. J. Clin. Microbiol. 37, 2607–2618. https://doi.org/10.1128/JCM.37.8.2607-2618.1999.

Kremer, K., Glynn, J.R., Lillebaek, T., Niemann, S., Kurepina, N.E., Kreiswirth, B.N., Bifani, P.J., van Soolingen, D., 2004. Definition of the Beijing/W lineage of *Mycobacterium tuberculosis* on the basis of genetic markers. J. Clin. Microbiol. 42, 4040–4049. https://doi.org/10.1128/JCM.42.9.4040-4049.2004.

Krishnan, N., Malaga, W., Constant, P., Caws, M., Tran, T.H., Salmons, J., Nguyen, T.N., Nguyen, D.B., Daffe, M., Young, D.B., Robertson, B.D., Guilhot, C., Thwaites, G.E., 2011. *Mycobacterium tuberculosis* lineage influences innate immune response and virulence and is associated with distinct cell envelope lipid profiles. PLoS One 6, e23870. https://doi.org/10.1371/journal.pone.0023870.

Kurepina, N.E., Sreevatsan, S., Plikaytis, B.B., Bifani, P.B., Connell, N.D., Donneelly, R.J., vanSoolingen, D., Musser, J.M., Kreiswirth, B.N., 1998. Characterization of the phylogenetic distribution and chromosomal insertion sites of five IS*6110* elements in *Mycobacterium tuberculosis*: non random integration in the dnaA-dnaN region. Tubercle Lung Dis. 79, 31–42. https://doi.org/10.1054/tuld.1998.0003.

Laberge, A.M., Michaud, J., Richter, A., Lemyre, E., Lambert, M., Brais, B., Mitchell, G.A., 2005. Population history and its impact on medical genetics in Quebec. Clin. Genet. 68, 287–301. https://doi.org/10.1111/j.1399-0004.2005.00497.x.

Larsen, M.H., Lacourciere, K., Parker, T.M., Kraigsley, A., Achkar, J.M., Adams, L.B., Dupnik, K.M., Hall-Stoodley, L., Hartman, T., Kanipe, C., Kurtz, S.L., Miller, M.A., Salvador, L.C.M., Spencer, J.S., Robinson, R.T., 2020. The many hosts of mycobacteria 8 (MHM8): a conference report. Tuberculosis 121, 101914. https://doi.org/10.1016/j.tube.2020.101914.

Lazzarini, L.C., Huard, R.C., Boechat, N.L., Gomes, H.M., Oelemann, M.C., Kurepina, N., Shashkina, E., Mello, F.C., Gibson, A.L., Virginio, M.J., Marsico, A.G., Butler, W.R., Kreiswirth, B.N., Suffys, P.N., Lapa, E.S.J.R., Ho, J.L., 2007. Discovery of a novel *Mycobacterium tuberculosis* lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. J. Clin. Microbiol. 45, 3891–3902. https://doi.org/10.1128/JCM.01394-07.

Lazzarini, L.C., Spindola, S.M., Bang, H., Gibson, A.L., Weisenberg, S., da Silva Carvalho, W., Augusto, C.J., Huard, R.C., Kritski, A.L., Ho, J.L., 2008. RDRio *Mycobacterium tuberculosis* infection is associated with a higher frequency of cavitary pulmonary disease. J. Clin. Microbiol. 46, 2175–2183. https://doi.org/10.1128/JCM.00065-08.

Lee, R.S., Radomski, N., Proulx, J.F., Levade, I., Shapiro, B.J., McIntosh, F., Soualhine, H., Menzies, D., Behr, M.A., 2015. Population genomics of *Mycobacterium tuberculosis* in the Inuit. Proc. Natl. Acad. Sci. U.S.A. 112, 13609–13614. https://doi.org/10.1073/pnas.1507071112.

Lee, O.Y., Wu, H.H.T., Besra, G.S., Minnikin, D.E., Jaeger, H.Y., Maixner, F., Zink, A., Gasparik, M., Pap, I., Bereczki, Z., Pálfi, G., 2023. Sensitive lipid biomarker detection for tuberculosis in late Neanderthal skeletons from Subalyuk Cave, Hungary. Tuberculosis 143S, 102420.

Leinonen, R., Sugawara, H., Shumway, M., International Nucleotide Sequence Database Collaboration, 2011. The sequence read archive. Nucleic Acids Res. 39, D19–D21. https://doi.org/10.1093/nar/gkq1019.

Liu, Q., Ma, A., Wei, L., Pang, Y., Wu, B., Luo, T., Zhou, Y., Zheng, H.X., Jiang, Q., Gan, M., Zuo, T., Liu, M., Yang, C., Jin, L., Comas, I., Gagneux, S., Zhao, Y., Pepperell, C.S., Gao, Q., 2018. China's tuberculosis

epidemic stems from historical expansion of four strains of *Mycobacterium tuberculosis*. Nat. Ecol. Evol. 2, 1982−1992. https://doi.org/10.1038/s41559-018-0680-6.

Linz, B., Balloux, F., Moodley, Y., Manica, A., Liu, H., Roumagnac, P., Falush, D., Stamer, C., Prugnolle, F., van der Merwe, S.W., Yamaoka, Y., Graham, D.Y., Perez-Trallero, E., Wadstrom, T., Suerbaum, S., Achtman, M., 2007. An African origin for the intimate association between humans and *Helicobacter pylori*. Nature 445, 915−918.

Loiseau, C., Menardo, F., Aseffa, A., Hailu, E., Gumi, B., Ameni, G., Berg, S., Rigouts, L., Robbe-Austerman, S., Zinsstag, J., Gagneux, S., Brites, D., 2020. An African origin for *Mycobacterium bovis*. Evol. Med. Public Health 2020, 49−59. https://doi.org/10.1093/emph/eoaa005.

Loukil, A., Bouzid, F., Osman, D.A., Drancourt, M., 2019. Decrypting the environmental sources of *Mycobacterium canettii* by high-throughput biochemical profiling. PLoS One 14, e0222078.

Luo, T., Comas, I., Luo, D., Lu, B., Wu, J., Wei, L., Yang, C., Liu, Q., Gan, M., Sun, G., Shen, X., Liu, F., Gagneux, S., Mei, J., Lan, R., Wan, K., Gao, Q., 2015. Southern East asian origin and coexpansion of *Mycobacterium tuberculosis* Beijing family with han Chinese. Proc. Natl. Acad. Sci. U.S.A. 112, 8136−8141. https://doi.org/10.1073/pnas.1424063112. Epub 2015 Jun 15.

Madacki, J., Orgeur, M., Mas Fiol, G., Frigui, W., Ma, L., Brosch, R., 2021. ESX-1-Independent horizontal gene transfer by *Mycobacterium tuberculosis* complex strains. mBio 12 (3), e00965−e1021. https://doi.org/10.1128/mBio.00965-21.

Malm, S., Linguissi, L.S., Tekwu, E.M., Vouvoungui, J.C., Kohl, T.A., Beckert, P., Sidibe, A., Rusch-Gerdes, S., Madzou-Laboum, I.K., Kwedi, S., Penlap Beng, V., Frank, M., Ntoumi, F., Niemann, S., 2017. New *Mycobacterium tuberculosis* complex sublineage, Brazzaville, Congo. Emerg. Infect. Dis. 23, 423−429. https://doi.org/10.3201/eid2303.160679.

Massoud, K., Araj, G., Reslan, L., Fadlallah, S., Wehbe, M., Itani, L., Avedissian, A., Dbaibo, G., Saade, A., Refrégier, G., Matar, G., Sola, C., 2020. Spoligotyping of *Mycobacterium tuberculosis* isolates using Luminex®-based method in Lebanon: *Mycobacterium tuberculosis* genotyping in Lebanon. J. Infect. Dev. Countr. 14 (8), 878−885. https://doi.org/10.3855/jidc.12072.

Matossian, M.K., 1985. Death in London, 1750-1909. J. Interdiscipl. Hist. 16, 183−197.

McGrath, J.W., 1988. Social networks of disease spread in the lower Illinois valley: a simulation approach. Am. J. Phys. Anthropol. 77, 483−496. https://doi.org/10.1002/ajpa.1330770409.

McHugh, T.D., Batt, S.L., Shorten, R.J., Gosling, R.D., Uiso, L., Gillespie, S.H., 2005. *Mycobacterium tuberculosis* lineage: a naming of the parts. Tuberculosis 85, 127−136. https://doi.org/10.1016/j.tube.2004.06.002.

Meehan, C.J., Goig, G.A., Kohl, T.A., Verboven, L., Dippenaar, A., Ezewudo, M., Farhat, M.R., Guthrie, J.L., Laukens, K., Miotto, P., Ofori-Anyinam, B., Dreyer, V., Supply, P., Suresh, A., Utpatel, C., van Soolingen, D., Zhou, Y., Ashton, P.M., Brites, D., Cabibbe, A.M., de Jong, B.C., de Vos, M., Menardo, F., Gagneux, S., Gao, Q., Heupink, T.H., Liu, Q., Loiseau, C., Rigouts, L., Rodwell, T.C., Tagliani, E., Walker, T.M., Warren, R.M., Zhao, Y., Zignol, M., Schito, M., Gardy, J., Cirillo, D.M., Niemann, S., Comas, I., Van Rie, A., 2019. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open issues. Nat. Rev. Microbiol. 17, 533−545. https://doi.org/10.1038/s41579-019-0214-5.

Meehan, C.J., Moris, P., Kohl, T.A., Pecerska, J., Akter, S., Merker, M., Utpatel, C., Beckert, P., Gehre, F., Lempens, P., Stadler, T., Kaswa, M.K., Kuhnert, D., Niemann, S., de Jong, B.C., 2018. The relationship between transmission time and clustering methods in *Mycobacterium tuberculosis* epidemiology. EBioMedicine 37, 410−416. https://doi.org/10.1016/j.ebiom.2018.10.013.

Meissner, G., Schröder, K., 1969. The so-called African mycobacteria strains from the tropical West Africa. Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. 1. Abt. Medizinisch-hygienische Bakteriologie, Virusforschung und Parasitologie. Originale 211, 69−81.

Melly, G., Purdy, G.E., 2019. MmpL proteins in physiology and pathogenesis of *M. tuberculosis*. Microorganisms 7. https://doi.org/10.3390/microorganisms7030070.

Menardo, F., Duchene, S., Brites, D., Gagneux, S., 2019. The molecular clock of *Mycobacterium tuberculosis*. PLoS Pathog. 15, e1008067. https://doi.org/10.1371/journal.ppat.1008067.

Menardo, F., Rutaihwa, L., Zwyer, M., Borrell, S., Comas, I., Conceiçao, E., Coscolla, M., Cox, H., Joloba, M., Dou, H., Feldmann, J., Fenner, L., Fyfe, J., Gao, Q., GarcÌa de Viedma, D., Garcia-Basteiro, A., Gygli, S., Hella, J., Hiza, H., Jugheli, L., Kamwela, L., Kato-Maeda, M., Liu, Q., Ley, S., Loiseau, C., Mahasirimongkol, S., Malla, B., Palittapongarnpim, P., Rakotosamimanana, N., Rasolofo, V., Reinhard, M., Reither, K., Sasamalo, M., Silva Duarte, R., Sola, C., Suffys, P., Batista Lima, K., Yeboah-Manu, D., Beisel, C., Brites, D., Gagneux, S., 2021. Local adaptation in populations of *Mycobacterium tuberculosis* endemic to the Indian Ocean Rim [version 2; peer review: 2 approved]. F1000Research 10, 60. https://doi.org/10.12688/f1000research.28318.2.

Merker, M., Blin, C., Mona, S., Duforet-Frebourg, N., Lecher, S., Willery, E., Blum, M.G., Rusch-Gerdes, S., Mokrousov, I., Aleksic, E., Allix-Beguec, C., Antierens, A., Augustynowicz-Kopec, E., Ballif, M., Barletta, F., Beck, H.P., Barry 3rd, C.E., Bonnet, M., Borroni, E., Campos-Herrero, I., Cirillo, D., Cox, H., Crowe, S., Crudu, V., Diel, R., Drobniewski, F., Fauville-Dufaux, M., Gagneux, S., Ghebremichael, S., Hanekom, M., Hoffner, S., Jiao, W.W., Kalon, S., Kohl, T.A., Kontsevaya, I., Lillebaek, T., Maeda, S., Nikolayevskyy, V., Rasmussen, M., Rastogi, N., Samper, S., Sanchez-Padilla, E., Savic, B., Shamputa, I.C., Shen, A., Sng, L.H., Stakenas, P., Toit, K., Varaine, F., Vukovic, D., Wahl, C., Warren, R., Supply, P., Niemann, S., Wirth, T., 2015. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. Nat. Genet. 47, 242−249. https://doi.org/10.1038/ng.3195.

Merker, M., Barbier, M., Cox, H., Rasigade, J.P., Feuerriegel, S., Kohl, T.A., Diel, R., Borrell, S., Gagneux, S., Nikolayevskyy, V., Andres, S., Nubel, U., Supply, P., Wirth, T., Niemann, S., 2018. Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. Elife 7, e38200. https://doi.org/10.7554/eLife.38200.

Merker, M., Rasigade, J.P., Barbier, M., Cox, H., Feuerriegel, S., Kohl, T.A., Shitikov, E., Klaos, K., Gaudin, C., Antoine, R., Diel, R., Borrell, S., Gagneux, S., Nikolayevskyy, V., Andres, S., Crudu, V., Supply, P., Niemann, S., Wirth, T., 2022. Transcontinental spread and evolution of *Mycobacterium tuberculosis* W148 European/Russian clade toward extensively drug resistant tuberculosis. Nat. Commun. 13, 5105. https://doi.org/10.1038/s41467-022-32455-1.

Mestre, O., Luo, T., Dos Vultos, T., Kremer, K., Murray, A., Namouchi, A., Jackson, C., Rauzier, J., Bifani, P., Warren, R., Rasolofo, V., Mei, J., Gao, Q., Gicquel, B., 2011. Phylogeny of *Mycobacterium tuberculosis* Beijing strains constructed from polymorphisms in genes involved in DNA replication, recombination and repair. PLoS One 6, e16020. https://doi.org/10.1371/journal.pone.0016020.

Millet, J., Miyagi-Shiohira, C., Yamane, N., Sola, C., Rastogi, N., 2007. Assessment of mycobacterial interspersed repetitive unit-QUB markers to further discriminate the Beijing genotype in a population-based study of the genetic diversity of *Mycobacterium tuberculosis* clinical isolates from Okinawa, Ryukyu Islands, Japan. J. Clin. Microbiol. 45, 3606−3615. https://doi.org/10.1128/JCM.00348-07.

Mitchison, D.A., Wallace, J.G., Bhatia, A.L., Selkon, J.B., Subbaiah, T.V., Lancaster, M.C., 1960. A comparison of the virulence in Guinea-pigs of South Indian and British tubercle bacilli. Tubercle 41, 1−22. https://doi.org/10.1016/s0041-3879(60)80019-0.

Mizukoshi, F., Miyoshi-Akiyama, T., Iwai, H., Suzuki, T., Kiritani, R., Kirikae, T., Funatogawa, K., 2017. Genetic diversity of *Mycobacterium tuberculosis* isolates from Tochigi prefecture, a local region of Japan. BMC Infect. Dis. 17, 365. https://doi.org/10.1186/s12879-017-2457-y.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., Soria, E., 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60, 174−182. https://doi.org/10.1007/s00239-004-0046-3.

Mokrousov, I., Narvskaya, O., Otten, T., Vyazovaya, A., Limeschenko, E., Steklova, L., Vyshnevskyi, B., 2002. Phylogenetic reconstruction within *Mycobacterium tuberculosis* Beijing genotype in northwestern Russia. Res. Microbiol. 153, 629−637. https://doi.org/10.1016/s0923-2508(02)01374-8.

Mokrousov, I., Ly, H.M., Otten, T., Lan, N.N., Vyshnevskyi, B., Hoffner, S., Narvskaya, O., 2005. Origin and primary dispersal of the *Mycobacterium tuberculosis* Beijing genotype: clues from human phylogeography. Genome Res. 15, 1357−1364. https://doi.org/10.1101/gr.3840605.

Mokrousov, I., 2013. Insights into the origin, emergence, and current spread of a successful Russian clone of *Mycobacterium tuberculosis*. Clin. Microbiol. Rev. 26, 342−360. https://doi.org/10.1128/CMR.00087-12.

Mokrousov, I., Isakova, J., Valcheva, V., Aldashev, A., Rastogi, N., 2013. Molecular snapshot of *Mycobacterium tuberculosis* population structure and drug-resistance in Kyrgyzstan. Tuberculosis 93, 501−507. https://doi.org/10.1016/j.tube.2013.05.008.

Mokrousov, I., Jiao, W.W., Wan, K., Shen, A., 2014. Stranger in a strange land: Ibero-American strain of *Mycobacterium tuberculosis* in Tibet, China. Infect. Genet. Evol. 26, 323−326. https://doi.org/10.1016/j.meegid.2014.06.011.

Mokrousov, I., Vyazovaya, A., Iwamoto, T., Skiba, Y., Pole, I., Zhdanova, S., Arikawa, K., Sinkov, V., Umpeleva, T., Valcheva, V., Alvarez Figueroa, M., Ranka, R., Jansone, I., Ogarkov, O., Zhuravlev, V., Narvskaya, O., 2016. Latin-American-Mediterranean lineage of *Mycobacterium tuberculosis*: human traces across pathogen's phylogeography. Mol. Phylogenet. Evol. 99, 133−143. https://doi.org/10.1016/j.ympev.2016.03.020.

Mokrousov, I., Shitikov, E., Skiba, Y., Kolchenko, S., Chernyaeva, E., Vyazovaya, A., 2017. Emerging peak on the phylogeographic landscape of *Mycobacterium tuberculosis* in West Asia: definitely smoke, likely fire. Mol. Phylogenet. Evol. 116, 202−212. https://doi.org/10.1016/j.ympev.2017.09.002.

Mokrousov, I., Vyazovaya, A., Pasechnik, O., Gerasimova, A., Dymova, M., Chernyaeva, E., Tatarintseva, M., Stasenko, V., 2019. Early ancient sublineages of *Mycobacterium tuberculosis* Beijing genotype: unexpected clues from phylogenomics of the pathogen and human history. Clin. Microbiol. Infection 25, 1039 e1031−e1039 e1036. https://doi.org/10.1016/j.cmi.2018.11.024.

Mokrousov, I., 2012. The quiet and controversial: Ural family of *Mycobacterium tuberculosis*. Infect. Genet. Evol. 12, 619−629.

Mokrousov, I., 2021. Ubiquitous and multifaceted: SIT53 spoligotype does not correlate with any particular family of *Mycobacterium tuberculosis*. Tuberculosis 126, 102024. https://doi.org/10.1016/j.tube.2020.102024.

Mokrousov, I., Vyazovaya, A., Sinkov, V., Gerasimova, A., Ioannidis, P., Jiao, W., Khromova, P., Papaventsis, D., Pasechnik, O., Perdigão, J., Rastogi, N., Shen, A., Skiba, Y., Solovieva, N., Suffys, P., Tafaj, S., Umpeleva, T., Vakhrusheva, D., Yarusova, I., Zhdanova, S., Zhuravlev, V., Ogarkov, O., 2021. Practical approach to detection and surveillance of emerging highly resistant *Mycobacterium tuberculosis* Beijing 1071-32-cluster. Sci. Rep. 11, 21392. https://doi.org/10.1038/s41598-021-00890-7.

Molina-Moya, B., Gomgnimbou, M.K., Spinasse, L., Obasanya, J., Oladimeji, O., Dacombe, R., Edwards, T., Daragon, X.O., Lawson, L., Abdurrahman, S.T., Cuevas, L.E., Dominguez, J., Sola, C., 2018. *Mycobacterium tuberculosis* complex genotypes circulating in Nigeria based on spoligotyping obtained from Ziehl-Neelsen stained slides extracted DNA. PLoS Neglected Trop. Dis. 12, e0006242. https://doi.org/10.1371/journal.pntd.0006242.

Moss, A.R., Alland, D., Telzak, E., Hewlett Jr., D., Sharp, V., Chiliade, P., LaBombardi, V., Kabus, D., Hanna, B., Palumbo, L., Brudney, K., Weltman, A., Stoeckle, K., Chirgwin, K., Simberkoff, M., Moghazeh, S., Eisner, W., Lutfey, M., Kreiswirth, B., 1997. A city-wide outbreak of a multiple-drug-resistant strain of *Mycobacterium tuberculosis* in New-York. Int. J. Tuberc. Lung Dis. 1, 115−121.

Mostowy, S., Onipede, A., Gagneux, S., Niemann, S., Kremer, K., Desmond, E.P., Kato-Maeda, M., Behr, M., 2004a. Genomic analysis distinguishes *Mycobacterium africanum*. J. Clin. Microbiol. 42, 3594−3599.

Mostowy, S., Cousins, D., Behr, M.A., 2004b. Genomic interrogation of the dassie bacillus reveals it as a unique RD1 mutant within the *Mycobacterium tuberculosis* complex. J. Bacteriol. 186, 104−109. https://doi.org/10.1128/JB.186.1.104-109.2003.

Mulholland, C.V., Shockey, A.C., Aung, H.L., Cursons, R.T., O'Toole, R.F., Gautam, S.S., Brites, D., Gagneux, S., Roberts, S.A., Karalus, N., Cook, G.M., Pepperell, C.S., Arcus, V.L., 2019. Dispersal of *Mycobacterium tuberculosis* driven by historical european trade in the South Pacific. Front. Microbiol. 2778. https://doi.org/10.3389/fmicb.2019.02778.

Musser, J.M., 1995. Antimicrobial agent resistance in Mycobacteria: molecular genetic insight. Clin. Microbiol. Rev. 8, 496−514. https://doi.org/10.1128/CMR.8.4.496.

Musser, J.M., Amin, A., Ramaswamy, S., 2000. Negligible genetic diversity of mycobacterium tuberculosis host immune system protein targets: evidence of limited selective pressure. Genetics 155 (1), 7−16. https://doi.org/10.1093/genetics/155.1.7. PMID: 10790380.

Nakanishi, N., Wada, T., Arikawa, K., Millet, J., Rastogi, N., Iwamoto, T., 2013. Evolutionary robust SNPs reveal the misclassification of *Mycobacterium tuberculosis* Beijing family strains into sublineages. Infect. Genet. Evol. 16C, 174−177. https://doi.org/10.1016/j.meegid.2013.02.007.

Namasivayam, S., Diarra, B., Diabate, S., Sarro, Y.D.S., Kone, A., Kone, B., Tolofoudie, M., Baya, B., Diakite, M.T., Kodio, O., Cohen, K., Holl, J., Achenbach, C.J., Chatterjee, S., Murphy, R.L., Bishai, W., Diallo, S., Sher, A., Maiga, M., 2020. Patients infected with *Mycobacterium africanum* versus *Mycobacterium tuberculosis* possess distinct intestinal microbiota. PLoS Neglected Trop. Dis. 14, e0008230. https://doi.org/10.1371/journal.pntd.0008230.

Napier, G., Campino, S., Merid, Y., Abebe, M., Woldeamanuel, Y., Aseffa, A., Hibberd, M.L., Phelan, J., Clark, T.G., 2020. Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. Genome Med. 12, 114. https://doi.org/10.1186/s13073-020-00817-3.

Napier, G., Couvin, D., Refrégier, G., Guyeux, C., Meehan, C.J., Sola, C., Campino, S., Phelan, J., Clark, T.G., 2023. Comparison of in silico predicted *Mycobacterium tuberculosis* spoligotypes and lineages from whole genome sequencing data. Sci. Rep. 13, 11368.

Nastasi, A., Mammina, C., 1999. Epidemiological study of tuberculosis in Palermo, Italy: IS6110 fingerprinting of *Mycobacterium tuberculosis* strains isolated in the years 1994-1998. Infection 27, 318−322. https://doi.org/10.1007/s150100050036.

Nebenzahl-Guimaraes, H., Yimer, S.A., Holm-Hansen, C., de Beer, J., Brosch, R., van Soolingen, D., 2016. Genomic characterization of *Mycobacterium tuberculosis* lineage 7 and a proposed name: 'Aethiops vetus'. Microb. Genom. 2, e000063. https://doi.org/10.1099/mgen.0.000063.

Ngabonziza, J.C.S., Loiseau, C., Marceau, M., Jouet, A., Menardo, F., Tzfadia, O., Antoine, R., Niyigena, E.B., Mulders, W., Fissette, K., Diels, M., Gaudin, C., Duthoy, S., Ssengooba, W., André, E., Kaswa, M.K., Habimana, Y.M., Brites, D., Affolabi, D., Mazarati, J.B., de Jong, B.C., Rigouts, L., Gagneux, S., Meehan, C.J., Supply, P., 2020. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. Nat. Commun. 11 (1), 2917. https://doi.org/10.1038/s41467-020-16626-6.

Nguyen, D., Proulx, J.F., Westley, J., Thibert, L., Dery, S., Behr, M.A., 2003a. Tuberculosis in the Inuit community of Quebec, Canada. Am. J. Respir. Crit. Care Med. 168 (11), 1353−1357. https://doi.org/10.1164/rccm.200307-910OC.

Nguyen, D., Brassard, P., Westley, J., Thibert, L., Proulx, M., Henry, K., Schwartzman, K., Menzies, D., Behr, M.A., 2003b. Widespread pyrazinamide-resistant *Mycobacterium tuberculosis* family in a low-incidence setting. J. Clin. Microbiol. 41, 2878−2883. https://doi.org/10.1128/JCM.41.7.2878-2883.2003.

Nguyen, D., Brassard, P., Menzies, D., Thibert, L., Warren, R., Mostowy, S., Behr, M., 2004. Genomic characterization of an endemic *Mycobacterium tuberculosis* strain: evolutionary and epidemiologic implications. J. Clin. Microbiol. 42, 2573−2580. https://doi.org/10.1128/JCM.42.6.2573-2580.2004.

Niemann, S., Rusch-Gerdes, S., Joloba, M.L., Whalen, C.C., Guwatudde, D., Ellner, J.J., Eisenach, K., Fumokong, N., Johnson, J.L., Aisu, T., Mugerwa, R.D., Okwera, A., Schwander, S.K., 2002. *Mycobacterium africanum* subtype II is associated with two distinct genotypes and is a major cause of human tuberculosis in Kampala, Uganda. J. Clin. Microbiol. 40, 3398−3405. https://doi.org/10.1128/JCM.40.9.3398-3405.2002.

Niemann, S., Koser, C.U., Gagneux, S., Plinke, C., Homolka, S., Bignell, H., Carter, R.J., Cheetham, R.K., Cox, A., Gormley, N.A., Kokko-Gonzales, P., Murray, L.J., Rigatti, R., Smith, V.P., Arends, F.P., Cox, H.S., Smith, G., Archer, J.A., 2009. Genomic diversity among drug sensitive and multidrug resistant isolates of *Mycobacterium tuberculosis* with identical DNA fingerprints. PLoS One 4, e7407. https://doi.org/10.1371/journal.pone.0007407.

Niobe-Eyangoh, S.N., Kuaban, C., Sorlin, P., Cunin, P., Thonnon, J., Sola, C., Rastogi, N., Vincent, V., Gutierrez, M.C., 2003. Genetic biodiversity of *Mycobacterium tuberculosis* complex strains from patients with pulmonary tuberculosis in Cameroon. J. Clin. Microbiol. 41, 2547−2553. https://doi.org/10.1128/JCM.41.6.2547-2553.2003.

Niobe-Eyangoh, S.N., Kuaban, C., Sorlin, P., Thonnon, J., Vincent, V., Gutierrez, M.C., 2004. Molecular characteristics of strains of the Cameroon family, the major group of *Mycobacterium tuberculosis* in a country with a high prevalence of tuberculosis. J. Clin. Microbiol. 42, 5029−5035. https://doi.org/10.1128/JCM.42.11.5029-5035.2004.

O'Neill, M.B., Shockey, A., Zarley, A., Aylward, W., Eldholm, V., Kitchen, A., Pepperell, C.S., 2019. Lineage specific histories of *Mycobacterium tuberculosis* dispersal in Africa and Eurasia. Mol. Ecol. 28, 3241−3256. https://doi.org/10.1111/mec.15120.

Oettinger, T., Jorgensen, M., Ladefoged, A., Haslov, K., Andersen, P., 1999. Development of the *Mycobacterium bovis* BCG vaccine: review of the historical and biochemical evidence for a genealogical tree. Tuber. Lung Dis. 79, 243−250. https://doi.org/10.1054/tuld.1999.0206.

Ofori-Anyinam, B., Dolganov, G., Van, T., Davis, J.L., Walter, N.D., Garcia, B.J., Voskuil, M., Fissette, K., Diels, M., Driesen, M., Meehan, C.J., Yeboah-Manu, D., Coscolla, M., Gagneux, S., Antonio, M., Schoolnik, G., Gehre, F., de Jong, B.C., 2017. Significant under expression of the DosR regulon in *M. tuberculosis* complex lineage 6 in sputum. Tuberculosis 104, 58−64.

Olson, E.S., Forbes, K.J., Watt, B., Pennington, T.H., 1995. Population genetics of *Mycobacterium tuberculosis* complex in Scotland analysed by pulse-field gel electrophoresis. Epidemiol. Infect. 114, 153−160. https://doi.org/10.1017/s0950268800052006.

Orgeur, M., Brosch, R., 2018. Evolution of virulence in the *Mycobacterium tuberculosis* complex. Curr. Opin. Microbiol. 41, 68−75. https://doi.org/10.1016/j.mib.2017.11.021.

Osborne, A.D., Neale, A.V., 1966. Milk and tuberculosis. Bristol Medico-Chirurgical J. 81, 12−14.

Osman, D.A., Bouzid, F., Canaan, S., Drancourt, M., 2015. Smooth tubercle bacilli: neglected opportunistic tropical pathogens. Front. Public Health 3, 283. https://doi.org/10.3389/fpubh.2015.00283.

Osman, D.A., Phelippeau, M., Drancourt, M., Musso, D., 2017. Diversity of *Mycobacterium tuberculosis* lineages in French polynesia. J. Microbiol. Immunol. Infect. 50, 199−206. https://doi.org/10.1016/j.jmii.2015.05.018.

Otchere, I.D., Coscollá, M., Sánchez-Busó, L., Asante-Poku, A., Brites, D., Loiseau, C., Meehan, C., Osei-Wusu, S., Forson, A., Laryea, C., Yahayah, A.I., Baddoo, A., Ansa, G.A., Aboagye, S.Y., Asare, P., Borrell, S., Gehre, F., Beckert, P., Kohl, T.A., N'Dira, S., Beisel, C., Antonio, M., Niemann, S., de Jong, B.C., Parkhill, J., Harris, S.R., Gagneux, S., Yeboah-Manu, D., 2018. Comparative genomics of *Mycobacterium africanum* Lineage 5 and Lineage 6 from Ghana suggests distinct ecological niches. Sci. Rep. 8, 11269. https://doi.org/10.1038/s41598-018-29620-2.

Ouassa, T., Borroni, E., Loukou, G.Y., Faye-Kette, H., Kouakou, J., Menan, H., Cirillo, D.M., 2012. High prevalence of shared international type 53 among *Mycobacterium tuberculosis* complex strains in retreated patients from Cote d'Ivoire. PLoS One 7, e45363. https://doi.org/10.1371/journal.pone.0045363.

Pálfi, G., Dutour, O., Hajdu, T., Sola, C., Zink, A., 2023a. Paleopathology and evolution of tuberculosis editorial. Tuberculosis 143S, 102428.

Pálfi, G., Molnar, E., Bereczki, Z., Coqueugniot, H., Dutour, O., Tillier, A.M., Rosendahl, W., Sklanitz, A., Mester, Z., Gasparik, M., Maixner, F., Zink, A., Minnikin, D.E., Pap, I., 2023b. Re-examination of the

Subalyuk Neanderthal remains uncovers signs of probable TB infection (Subalyuk Cave, Hungary). Tuberculosis 143S, 102419.

Palittapongarnpim, P., Ajawatanawong, P., Viratyosin, W., Smittipat, N., Disratthakit, A., Mahasirimongkol, S., Yanai, H., Yamada, N., Nedsuwan, S., Imasanguan, W., Kantipong, P., Chaiyasirinroje, B., Wongyai, J., Toyo-Oka, L., Phelan, J., Parkhill, J., Clark, T.G., Hibberd, M.L., Ruengchai, W., Palittapongarnpim, P., Juthayothin, T., Tongsima, S., Tokunaga, K., 2018. Evidence for host-bacterial co-evolution via genome sequence analysis of 480 Thai *Mycobacterium tuberculosis* lineage 1 isolates. Sci. Rep. 8, 11597. https://doi.org/10.1038/s41598-018-29986-3.

Panaiotov, S., Zozio, T., Evans, J.T., Brankova, N., Karcheva, A., Levterova, V., Kantardjiev, T., Hawkey, P., Rastogi, N., Sola, C., 2005. Biodiversity of *Mycobacterium tuberculosis* in Bulgaria assessed by spoligo- and MIRU-typing. In: 26th Annual Congress of the European Society for Mycobacteriology June 26th-29th 2005. Abstract Book, O-7, p. 37. Istanbul, Turkey, 26-29 Jun 2005.

Pepperell, C.S., Granka, J.M., Alexander, D.C., Behr, M.A., Chui, L., Gordon, J., Guthrie, J.L., Jamieson, F.B., Langlois-Klassen, D., Long, R., Nguyen, D., Wobeser, W., Feldman, M.W., 2011. Dispersal of *Mycobacterium tuberculosis* via the Canadian Fur trade. Proc. Natl. Acad. Sci. U.S.A. 108, 6526−6531. https://doi.org/10.1073/pnas.1016708108.

Pepperell, C.S., Casto, A.M., Kitchen, A., Granka, J.M., Cornejo, O.E., Holmes, E.C., Birren, B., Galagan, J., Feldman, M.W., 2013. The role of selection in shaping diversity of natural *M. tuberculosis* populations. PLoS Pathog. 9, e1003543. https://doi.org/10.1371/journal.ppat.1003543.

Perdigão, J., Silva, C., Diniz, J., Pereira, C., Machado, D., Ramos, J., Silva, H., Abilleira, F., Brum, C., Reis, A.J., Macedo, M., Scaini, J.L., Silva, A.B., Esteves, L., Macedo, R., Maltez, F., Clemente, S., Coelho, E., Viegas, S., Rabna, P., Rodrigues, A., Taveira, N., Jordao, L., Kritski, A., Lapa, E.S.J.R., Mokrousov, I., Couvin, D., Rastogi, N., Couto, I., Pain, A., McNerney, R., Clark, T.G., von Groll, A., Dalla-Costa, E.R., Rossetti, M.L., Silva, P.E.A., Viveiros, M., Portugal, I., 2019. Clonal expansion across the seas as seen through CPLP-TB database: a joint effort in cataloguing *Mycobacterium tuberculosis* genetic diversity in Portuguese-speaking countries. Infect. Genet. Evol. 72, 44−58. https://doi.org/10.1016/j.meegid.2018.03.011.

Perdigão, J., Gomes, P., Miranda, A., Maltez, F., Machado, D., Silva, C., Phelan, J.E., Brum, L., Campino, S., Couto, I., Viveiros, M., Clark, T.G., Portugal, I., 2020. Using genomics to understand the origin and dispersion of multidrug and extensively drug resistant tuberculosis in Portugal. Sci. Rep. 10, 2600. https://doi.org/10.1038/s41598-020-59558-3.

Perez-Losada, M., Browne, E.B., Madsen, A., Wirth, T., Viscidi, R.P., Crandall, K.A., 2006. Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. Infect. Genet. Evol. 6, 97−112. https://doi.org/10.1016/j.meegid.2005.02.003.

Periwal, V., Patowary, A., Vellarikkal, S.K., Gupta, A., Singh, M., Mittal, A., Jeyapaul, S., Chauhan, R.K., Singh, A.V., Singh, P.K., Garg, P., Katoch, V.M., Katoch, K., Chauhan, D.S., Sivasubbu, S., Scaria, V., 2015. Comparative whole-genome analysis of clinical isolates reveals characteristic architecture of *Mycobacterium tuberculosis* pangenome. PLoS One 10, e0122979. https://doi.org/10.1371/journal.pone.0122979.

Phelan, J.E., Coll, F., Bergval, I., Anthony, R.M., Warren, R., Sampson, S.L., Gey van Pittius, N.C., Glynn, J.R., Crampin, A.C., Alves, A., Bessa, T.B., Campino, S., Dheda, K., Grandjean, L., Hasan, R., Hasan, Z., Miranda, A., Moore, D., Panaiotov, S., Perdigão, J., Portugal, I., Sheen, P., de Oliveira Sousa, E., Streicher, E.M., van Helden, P.D., Viveiros, M., Hibberd, M.L., Pain, A., McNerney, R., Clark, T.G., 2016. Recombination in pe/ppe genes contributes to genetic variation in *Mycobacterium tuberculosis* lineages. BMC Genom. 17, 151. https://doi.org/10.1186/s12864-016-2467-y.

Pierron, D., Heiske, M., Razafindrazaka, H., Rakoto, I., Rabetokotany, N., Ravololomanga, B., Rakotozafy, L.M., Rakotomalala, M.M., Razafiarivony, M., Rasoarifetra, B., Raharijesy, M.A., Razafindralambo, L., Ramilisonina, Fanony, F., Lejamble, S., Thomas, O., Mohamed Abdallah, A., Rocher, C., Arachiche, A., Tonaso, L., Pereda-Loth, V., Schiavinato, S., Brucato, N., Ricaut, F.X., Kusuma, P., Sudoyo, H., Ni, S., Boland, A., Deleuze, J.F., Beaujard, P., Grange, P., Adelaar, S., Stoneking, M., Rakotoarisoa, J.A.,

Radimilahy, C., Letellier, T., 2017. Genomic landscape of human diversity across Madagascar. Proc. Natl. Acad. Sci. U.S.A. 114, E6498−E6506. https://doi.org/10.1073/pnas.1704906114.

Pillay, M., Sturm, A.W., 2007. Evolution of the extensively drug-resistant F15/LAM4/KZN strain of *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa. Clin. Infect. Dis. 45, 1409−1414. https://doi.org/10.1086/522987.

Pisu, D., Huang, L., Narang, V., Theriault, M., Le-Bury, G., Lee, B., Lakudzala, A.E., Mzinza, D.T., Mhango, D.V., Mitini-Nkhoma, S.C., Jambo, K.C., Singhal, A., Mwandumba, H.C., Russell, D.G., 2021. Single cell analysis of *M. tuberculosis* phenotype and macrophage lineages in the infected lung. J. Exp. Med. 218 (9), e20210615. https://doi.org/10.1084/jem.20210615.

Pitchappan, R.M., 2016. Not all the infected develop the disease - a "Lotus and Cactus" model. Infect. Genet. Evol. 40, 303−309. https://doi.org/10.1016/j.meegid.2015.11.007.

Pitt, D., Sevane, N., Nicolazzi, E.L., MacHugh, D.E., Park, S.D.E., Colli, L., Martinez, R., Bruford, M.W., Orozco-terWengel, P., 2019. Domestication of cattle: two or three events? Evol. Appl. 12, 123−136. https://doi.org/10.1111/eva.12674.

Plikaytis, B.B., Marden, J.L., Crawford, J.T., Woodley, C.L., Butler, W.R., Shinnick, T.M., 1994. Multiplex PCR assay specific for the multidrug-resistant strain W of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 32, 1542−1546. https://doi.org/10.1128/jcm.32.6.1542-1546.1994.

Poonawalaa, H., Kumar, N., Peacock, S.J., 2020. A review of published spoligotype data indicates the diversity of *Mycobacterium tuberculosis* from India is under-represented in global databases. Infect. Genet. Evol. 78, 104072. https://doi.org/10.1016/j.meegid.2019.104072.

Porco, T.C., Blower, S.M., 1998. Quantifying the intrinsic transmission dynamics of tuberculosis. Theor. Popul. Biol. 54, 117−132. https://doi.org/10.1006/tpbi.1998.1366.

Pourcel, C., Salvignol, G., Vergnaud, G., 2005. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151, 653−663. https://doi.org/10.1099/mic.0.27437-0.

Pym, A.S., Brodin, P., Brosch, R., Huerre, M., Cole, S.T., 2002. Loss of RD1 contributed to the attenuation of the live tuberculosis vaccines *Mycobacterium bovis* BCG and *Mycobacterium microti*. Mol. Microbiol. 46, 709−717. https://doi.org/10.1046/j.1365-2958.2002.03237.x.

Rad, M.E., Bifani, P., Martin, C., Kremer, K., Samper, S., Rauzier, J., Kreiswirth, B., Blazquez, J., Jouan, M., van Soolingen, D., Gicquel, B., 2003. Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. Emerg. Infect. Dis. 9, 838−845. https://doi.org/10.3201/eid0907.020803.

Rajwani, R., Shehzad, S., Siu, G.K.H., 2018. MIRU-profiler: a rapid tool for determination of 24-loci MIRU-VNTR profiles from assembled genomes of *Mycobacterium tuberculosis*. PeerJ 6, e5090. https://doi.org/10.7717/peerj.5090.

Ramazanzadeh, R., Farnia, P., Amirmozafari, N., Ghazi, F., Ghadertotonchi, Z., Kamran, J., Mohammadi, F., Mirsaedi, M., Masjedi, M., 2006. Comparison between molecular epidemiology, geographical regions and drug resistance in *Mycobacterium tuberculosis* strains isolated from Iranian and Afghan patients. Chemotherapy 52, 316−320. https://doi.org/10.1159/000095971.

Ransom, E.M., Potter, R.F., Dantas, G., Burnham, C.D., 2020. Genomic prediction of antimicrobial resistance: ready or not, here it comes. Clin. Chem. 66, 1278−1289. https://doi.org/10.1093/clinchem/hvaa172.

Refrégier, G., Abadia, E., Matsumoto, T., Ano, H., Takashima, T., Tsuyuguchi, I., Aktas, E., Comert, F., Gomgnimbou, M.K., Panaiotov, S., Phelan, J., Coll, F., McNerney, R., Pain, A., Clark, T.G., Sola, C., 2016. Turkish and Japanese *Mycobacterium tuberculosis* sublineages share a remote common ancestor. Infect. Genet. Evol. 45, 461−473. https://doi.org/10.1016/j.meegid.2016.10.009.

Refrégier, G., Sola, C., Guyeux, C., 2020. Unexpected diversity of CRISPR unveils some evolutionary patterns of repeated sequences in *Mycobacterium tuberculosis*. BMC Genom. 21, 841. https://doi.org/10.1186/s12864-020-07178-6.

Reis, A.C., Cunha, M.V., 2021. The open pan-genome architecture and virulence landscape of *Mycobacterium bovis*. Microb. Genom. 7. https://doi.org/10.1099/mgen.0.000664.

Rindi, L., Lari, N., Garzelli, C., 2012. Large Sequence Polymorphisms of the Euro-American lineage of *Mycobacterium tuberculosis*: a phylogenetic reconstruction and evidence for convergent evolution in the DR locus. Infect. Genet. Evol. 12, 1551−1557. https://doi.org/10.1016/j.meegid.2012.06.008.

Rindi, L., Medici, C., Bimbi, N., Buzzigoli, A., Lari, N., Garzelli, C., 2014. Genomic variability of *Mycobacterium tuberculosis* strains of the Euro-American lineage based on large sequence deletions and 15-locus MIRU-VNTR polymorphism. PLoS One 9, e107150. https://doi.org/10.1371/journal.pone.0107150.

Riojas, M.A., McGough, K.J., Rider-Riojas, C.J., Rastogi, N., Hazbon, M.H., 2018. Phylogenomic analysis of the species of the *Mycobacterium tuberculosis* complex demonstrates that *Mycobacterium africanum, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti* and *Mycobacterium pinnipedii* are later heterotypic synonyms of *Mycobacterium tuberculosis*. Int. J. Syst. Evol. Microbiol. 68, 324−332. https://doi.org/10.1099/ijsem.0.002507.

Roetzer, A., Diel, R., Kohl, T.A., Ruckert, C., Nubel, U., Blom, J., Wirth, T., Jaenicke, S., Schuback, S., Rusch-Gerdes, S., Supply, P., Kalinowski, J., Niemann, S., 2013. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. PLoS Med. 10, e1001387. https://doi.org/10.1371/journal.pmed.1001387.

Roser, M., Ritchie, H., Ortiz-Espina, E.,A., 2013. World Population Growth. Published online at. OurWorldInData.org.

Rothschild, B.M., Martin, L.D., Lev, G., Bercovier, H., Bar-Gal, G.K., Greenblatt, C., Donoghue, H., Spigelman, M., Brittain, D., 2001. *Mycobacterium tuberculosis* complex DNA from an extinct bison dated 17,000 years before the present. Clin. Infect. Dis. 33, 305−311. https://doi.org/10.1086/321886.

Roychowdhury, T., Mandal, S., Bhattacharya, A., 2015. Analysis of IS6110 insertion sites provide a glimpse into genome evolution of *Mycobacterium tuberculosis*. Sci. Rep. 5, 12567. https://doi.org/10.1038/srep12567.

Ruesen, C., Chaidir, L., van Laarhoven, A., Dian, S., Ganiem, A.R., Nebenzahl-Guimaraes, H., Huynen, M.A., Alisjahbana, B., Dutilh, B.E., van Crevel, R., 2018. Large-scale genomic analysis shows association between homoplastic genetic variation in *Mycobacterium tuberculosis* genes and meningeal or pulmonary tuberculosis. BMC Genom. 19, 122.

Rutaihwa, L.K., Menardo, F., Stücki, D., Gygli, S.M., Ley, S.D., Malla, B., Feldmann, J., Borrell, S., Beisel, C., Middelkoop, K., Carter, E.J., Diero, L., Ballif, M., Jugheli, L., Reither, K., Fenner, L., Brites, D., Gagneux, S., 2019. Multiple introductions of *Mycobacterium tuberculosis* lineage 2−Beijing into Africa over centuries. Front. Ecol. Evol. 7. https://doi.org/10.3389/fevo.2019.00112.

Sabin, S., Herbig, A., Vågene, Å.J., Ahlström, T., Bozovic, G., Arcini, C., Kühnert, D., Bos, K.I., 2020. A seventeenth-century Mycobacterium tuberculosis genome supports a Neolithic emergence of the Mycobacterium tuberculosis complex. Genome Biol 21 (1), 201. https://doi.org/10.1186/s13059-020-02112-1.

Sahal, M.R., Senelle, G., La, K., Molina-Moya, B., Dominguez, J., Panda, T., Cambau, E., Refrégier, G., Sola, C., Guyeux, C., 2022. An Updated Evolutionary History and Taxonomy of *Mycobacterium tuberculosis* Lineage 5, Also Called *M. africanum*. bioRxiv, CSHL. https://doi.org/10.1101/2022.11.21.517336.

Sahal, M.R., Senelle, G., La, K., Panda, T.W., Taura, D.W., Guyeux, C., Cambau, E., Sola, C., 2023. *Mycobacterium tuberculosis* complex drug-resistance, phylogenetics, and evolution in Nigeria: comparison with Ghana and Cameroon. PLoS Neglected Trop. Dis. 17, e0011619. https://doi.org/10.1371/journal.pntd.0011619.

Salmond, G.P., Fineran, P.C., 2015. A century of the phage: past, present and future. Nat. Rev. Microbiol. 13, 777−786. https://doi.org/10.1038/nrmicro3564.

Salo, W.L., Aufderheide, A.C., Buikstra, J., Holcomb, T.A., 1994. Identification of *Mycobacterium tuberculosis* DNA in a pre-Columbian Peruvian mummy. Proc. Natl. Acad. Sci. U.S.A. 91, 2091−2094. https://doi.org/10.1073/pnas.91.6.2091.

Sanchez, M.A., Blower, S.M., 1997. Uncertainty and sensitivity analysis of the basic reproductive rate. Tuberculosis as an example. Am. J. Epidemiol. 145, 1127−1137. https://doi.org/10.1093/oxfordjournals.aje.a009076.

Sathyamala, C., 2019. Meat-eating in India: whose food, whose politics, and whose rights? Pol. Futures Educ. Internet 17, 878−891. https://doi.org/10.1177/1478210318780553.

Schürch, A.C., Kremer, K., Kiers, A., Daviena, O., Boeree, M.J., Siezen, R.J., Smith, N.H., van Soolingen, D., 2010. The tempo and mode of molecular evolution of *Mycobacterium tuberculosis* at patient-to-patient scale. Infect. Genet. Evol. 10, 108−114. https://doi.org/10.1016/j.meegid.2009.10.002.

Scriver, C.R., 2001. Human genetics: lessons from Quebec populations. Annu. Rev. Genomics Hum. Genet. 2, 69−101. https://doi.org/10.1146/annurev.genom.2.1.69.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068−2069. https://doi.org/10.1093/bioinformatics/btu153.

Senelle, G., Guyeux, C., Refrégier, G., Sola, C., 2023a. Towards a better understanding of the long-lasting evolutionary history of *Mycobacterium tuberculosis*. Tuberculosis 143S, 102374.

Senelle, G., Sahal, M., La, K., Billard-Pomares, T., Marin, J., Mougari, F., Bridier-Nahmias, A., Carbonnelle, E., Cambau, E., Refrégier, G., Guyeux, C., Sola, C., 2023b. Towards the reconstruction of a global TB history using a new pipeline "TB-Annotator". Tuberculosis 143S, 102376.

Senelle, G., Guyeux, C., Refrégier, G., Sola, C., 2023c. TB-annotator: A Scalable Web Application that Allows In-Depth Analysis of Very Large Datasets of Publicly Available *Mycobacterium tuberculosis* Complex Genomes. https://biorxiv.org/cgi/content/short/2023.06.12.526393v1.

Senghore, M., Diarra, B., Gehre, F., Otu, J., Worwui, A., Muhammad, A.K., Kwambana-Adams, B., Kay, G.L., Sanogo, M., Baya, B., Orsega, S., Doumbia, S., Diallo, S., de Jong, B.C., Pallen, M.J., Antonio, M., 2020. Evolution of *Mycobacterium tuberculosis* complex lineages and their role in an emerging threat of multidrug resistant tuberculosis in Bamako, Mali. Sci. Rep. 10, 327. https://doi.org/10.1038/s41598-019-56001-0.

Shemyakin, I.G., Stepanshina, V.N., Ivanov, I.Y., Lipin, M.Y., Anisimova, V.A., Onasenko, A.G., Korobova, O.V., Shinnick, T.M., 2004. Characterization of drug-resistant isolates of *Mycobacterium tuberculosis* derived from Russian inmates. Int. J. Tuberc. Lung Dis. 8, 1194−1203.

Shitikov, E., Kolchenko, S., Mokrousov, I., Bespyatykh, J., Ischenko, D., Ilina, E., Govorun, V., 2017. Evolutionary pathway analysis and unified classification of East Asian lineage of *Mycobacterium tuberculosis*. Sci. Rep. 7, 9227. https://doi.org/10.1038/s41598-017-10018-5.

Shuaib, Y.A., Khalil, E.A.G., Schaible, U.E., Wieler, L.H., Bakheit, M.A.M., Mohamed-Noor, S.E., Abdalla, M.A., Homolka, S., Andres, S., Hillemann, D., Lonnroth, K., Richter, E., Niemann, S., Kranzer, K., 2018. Smear microscopy for diagnosis of pulmonary tuberculosis in eastern Sudan. Tuberc. Res. Treat. 2018, 8038137. https://doi.org/10.1155/2018/8038137.

Shuaib, Y.A., Khalil, E.A.G., Wieler, L.H., Schaible, U.E., Bakheit, M.A., Mohamed-Noor, S.E., Abdalla, M.A., Kerubo, G., Andres, S., Hillemann, D., Richter, E., Kranzer, K., Niemann, S., Merker, M., 2020. *Mycobacterium tuberculosis* complex lineage 3 as causative agent of pulmonary tuberculosis, eastern Sudan(1). Emerg. Infect. Dis. 26, 427−436. https://doi.org/10.3201/eid2603.191145.

Shuaib, Y.A., Utpatel, C., Kohl, T.A., Barilar, I., Diricks, M., Ashraf, N., Wieler, L.H., Kerubo, G., Mesfin, E.A., Diallo, A.B., Al-Hajoj, S., Ndung'u, P., Fitzgibbon, M.M., Vaziri, F., Sintchenko, V., Martinez, E., Viegas, S.O., Zhou, Y., Azmy, A., Al-Amry, K., Godreuil, S., Varma-Basil, M., Narang, A., Ali, S., Beckert, P., Dreyer, V., Kabwe, M., Bates, M., Hoelscher, M., Rachow, A., Gori, A., Tekwu, E.M., Sidze, L.K., Jean-Paul, A.A., Beng, V.P., Ntoumi, F., Frank, M., Diallo, A.G., Mboup, S., Tessema, B., Beyene, D., Khan, S.N., Diel, R., Supply, P., Maurer, F.P., Hoffmann, H., Niemann, S., Merker, M., 2022. Origin and global expansion of *Mycobacterium tuberculosis* complex lineage 3. Genes 13. https://doi.org/10.3390/genes13060990.

Sibandze, D.B., Magazi, B.T., Malinga, L.A., Maningi, N.E., Shey, B.A., Pasipanodya, J.G., Mbelle, N.N., 2020. Machine learning reveals that *Mycobacterium tuberculosis* genotypes and anatomic disease site impacts drug resistance and disease transmission among patients with proven extra-pulmonary tuberculosis. BMC Infect. Dis. 20, 556. https://doi.org/10.1186/s12879-020-05256-4.

Silcocks, M., Dunstan, S.J., 2023. Parallel signatures of *Mycobacterium tuberculosis* and human Y-chromosome phylogeography support the Two Layer model of East Asian population history. Commun. Biol. 6, 1037.

Singh, P., Rameshwaram, N.R., Ghosh, S., Mukhopadhyay, S., 2018. Cell envelope lipids in the pathophysiology of *Mycobacterium tuberculosis*. Future Microbiol. 13, 689−710. https://doi.org/10.2217/fmb-2017-0135.

Skiba, Y., Mokrousov, I., Nabirova, D., Vyazovaya, A., Maltseva, E., Malakhova, N., Ismagulova, G., Pole, I., Ranka, R., Sapiyeva, Z., Ismailov, S., Moffett, D., 2019. *Mycobacterium tuberculosis* RD-Rio strain in Kazakhstan. Emerg. Infect. Dis. 25, 604−606. https://doi.org/10.3201/eid2503.181179.

Small, P.M., Moss, A., 1993. Molecular epidemiology and the new tuberculosis. Infect. Agent. Dis. 2, 132−138.

Smith, N.H., Hewinson, R.G., Kremer, K., Brosch, R., Gordon, S.V., 2009. Myths and misconceptions: the origin and evolution of *Mycobacterium tuberculosis*. Nat. Rev. Microbiol. 7, 537−544. https://doi.org/10.1038/nrmicro2165. Epub 2009 Jun 1.

Smith, N.H., Kremer, K., Inwald, J., Dale, J., Driscoll, J.R., Gordon, S.V., van Soolingen, D., Glyn Hewinson, R., Maynard Smith, J., 2006. Ecotypes of the *Mycobacterium tuberculosis* complex. J. Theor. Biol. 239, 220−225. https://doi.org/10.1016/j.jtbi.2005.08.036.

Soini, H., Pan, X., Amin, A., Graviss, E.A., Siddiqui, A., Musser, J.M., 2000. Characterization of *Mycobacterium tuberculosis* isolates from patients in Houston, Texas, by spoligotyping. J. Clin. Microbiol. 38, 669−676. https://doi.org/10.1128/JCM.38.2.669-676.2000.

Sola, C., Devallois, A., Horgen, L., Maisetti, J., Filliol, I., Legrand, E., Rastogi, N., 1999. Tuberculosis in the Caribbean: using spacer oligonucleotide typing to understand strain origin and transmission. Emerg. Infect. Dis. 5, 404−414. https://doi.org/10.3201/eid0503.990311.

Sola, C., Ferdinand, S., Mammina, C., Nastasi, A., Rastogi, N., 2001a. Genetic diversity of *Mycobacterium tuberculosis* in Sicily based on spoligotyping and variable number of tandem DNA repeats and comparison with a spoligotyping database for population-based analysis. J. Clin. Microbiol. 39, 1559−1565. https://doi.org/10.1128/JCM.39.4.1559-1565.2001.

Sola, C., Filliol, I., Gutierrez, M.C., Mokrousov, I., Vincent, V., Rastogi, N., 2001b. Spoligotype database of *Mycobacterium tuberculosis*: biogeographic distribution of shared types and epidemiologic and phylogenetic perspectives. Emerg. Infect. Dis. 7, 390−396. https://doi.org/10.3201/eid0703.010304.

Sola, C., Filliol, I., Legrand, E., Mokrousov, I., Rastogi, N., 2001c. *Mycobacterium tuberculosis* phylogeny reconstruction based on combined numerical analysis with IS1081, IS6110, VNTR, and DR-based spoligotyping suggests the existence of two new phylogeographical clades. J. Mol. Evol. 53, 680−689. https://doi.org/10.1007/s002390010255.

Sola, C., Filliol, I., Legrand, E., Lesjean, S., Locht, C., Supply, P., Rastogi, N., 2003a. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. Infect. Genet. Evol. 3, 125−133. https://doi.org/10.1016/s1567-1348(03)00011-x.

Sola, C., Rastogi, N., Gutierrez, M.C., Vincent, V., Brosch, R., Parsons, L.M., Niemann, S., Rüsch-Gerdes, S., Schwander, S.K., 2003b. Is *Mycobacterium africanum* subtype II (Uganda I and Uganda II) a genetically well-defined subspecies of the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. 41, 1345−1348. https://doi.org/10.1128/JCM.41.3.1345-1348.2003.

Sola, C., Ferdinand, S., Sechi, L.A., Zanetti, S., Martial, D., Mammina, C., Nastasi, A., Fadda, G., Rastogi, N., 2005a. *Mycobacterium tuberculosis* molecular evolution in western Mediterranean islands of Sicily and Sardinia. Infect. Genet. Evol. 5, 145−156. https://doi.org/10.1016/j.meegid.2004.08.003.

Sola, C., Zozio, T., Ellermeier, C., Sajduda, A., Naumann, L., Nguyen, D., Behr, M., de Haas, P., van Hest, R., van Soolingen, D., Prodinger, W.M., Allix, C., Fauville-Dufaux, M., Rastogi, N., 2005b. The presumed origin of a recent tuberculosis outbreak among the Inuit community of Nunavik. In: 26th Annual Congress of the European Society of Mycobacteriology. Abstract Book, P-27 Poster. *presenting author, p. 91, Istanbul, Turkey, June 25-29th 2005.

Sola, C., 2015. Clustured regularly interspersed short palindromic repeats (CRISPR) genetic diversity studies as a mean to reconstruct the evolution of the *Mycobacterium tuberculosis* complex. Tuberculosis 95. https://doi.org/10.1016/j.tube.2015.02.029. Epub 2015 Feb 14.

Sreevatsan, S., Pan, X., Stockbauer, K.E., Connell, N.D., Kreiswirth, B.N., Whittam, T.S., Musser, J.M., 1997. Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. Proc. Natl. Acad. Sci. U.S.A. 94, 9869–9874. https://doi.org/10.1073/pnas.94.18.9869.

Ssengooba, W., Meehan, C.J., Lukoye, D., Kasule, G.W., Musisi, K., Joloba, M.L., Cobelens, F.G., de Jong, B.C., 2016. Whole genome sequencing to complement tuberculosis drug resistance surveys in Uganda. Infect. Genet. Evol. 40, 8–16. https://doi.org/10.1016/j.meegid.2016.02.019.

Stavrum, R., Mphahlele, M., Ovreas, K., Muthivhi, T., Fourie, P.B., Weyer, K., Grewal, H.M., 2009. High diversity of *Mycobacterium tuberculosis* genotypes in South Africa and preponderance of mixed infections among ST53 isolates. J. Clin. Microbiol. 47, 1848–1856. https://doi.org/10.1128/JCM.02167-08.

Streicher, E.M., Victor, T.C., van der Spuy, G., Sola, C., Rastogi, N., van Helden, P.D., Warren, R.M., 2007. Spoligotype signatures in the *Mycobacterium tuberculosis* complex. J. Clin. Microbiol. 45, 237–240. https://doi.org/10.1128/JCM.01429-06.

Stücki, D., Brites, D., Jeljeli, L., Coscolla, M., Liu, Q., Trauner, A., Fenner, L., Rutaihwa, L., Borrell, S., Luo, T., Gao, Q., Kato-Maeda, M., Ballif, M., Egger, M., Macedo, R., Mardassi, H., Moreno, M., Vilanova, G.T., Fyfe, J., Globan, M., Thomas, J., Jamieson, F., Guthrie, J.L., Asante-Poku, A., Yeboah-Manu, D., Wampande, E., Ssengooba, W., Joloba, M., Boom, W.H., Basu, I., Bower, J., Saraiva, M., Vasconcellos, S.E., Suffys, P., Koch, A., Wilkinson, R., Gail-Bekker, L., Malla, B., Ley, S.D., Beck, H.P., de Jong, B.C., Toit, K., Sanchez-Padilla, E., Bonnet, M., Gil-Brusola, A., Frank, M., Penlap Beng, V.N., Eisenach, K., Alani, I., Ndung'u, P.W., Revathi, G., Gehre, F., Akter, S., Ntoumi, F., Stewart-Isherwood, L., Ntinginya, N.E., Rachow, A., Hoelscher, M., Cirillo, D.M., Skenders, G., Hoffner, S., Bakonyte, D., Stakenas, P., Diel, R., Crudu, V., Moldovan, O., Al-Hajoj, S., Otero, L., Barletta, F., Carter, E.J., Diero, L., Supply, P., Comas, I., Niemann, S., Gagneux, S., 2016. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. Nat. Genet. 48, 1535–1543. https://doi.org/10.1038/ng.3704.

Supply, P., Warren, R.M., Banuls, A.L., Lesjean, S., Van Der Spuy, G.D., Lewis, L.A., Tibayrenc, M., Van Helden, P.D., Locht, C., 2003. Linkage disequilibrium between minisatellite loci supports clonal evolution of *Mycobacterium tuberculosis* in a high tuberculosis incidence area. Mol. Microbiol. 47, 529–538. https://doi.org/10.1046/j.1365-2958.2003.03315.x.

Supply, P., Allix, C., Lesjean, S., Cardoso-Oelemann, M., Rusch-Gerdes, S., Willery, E., Savine, E., de Haas, P., van Deutekom, H., Roring, S., Bifani, P., Kurepina, N., Kreiswirth, B., Sola, C., Rastogi, N., Vatin, V., Gutierrez, M.C., Fauville, M., Niemann, S., Skuce, R., Kremer, K., Locht, C., van Soolingen, D., 2006. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 44, 4498–4510. https://doi.org/10.1128/JCM.01392-06.

Supply, P., Brosch, R., 2017. The biology and epidemiology of *Mycobacterium canettii*. Adv. Exp. Med. Biol. 1019, 27–41. https://doi.org/10.1007/978-3-319-64371-7_2.

Suzuki, T., Fujita, H., Choi, J.G., 2008. Brief communication: new evidence of tuberculosis from prehistoric Korea-Population movement and early evidence of tuberculosis in far East Asia. Am. J. Phys. Anthropol. 136, 357–360. https://doi.org/10.1002/ajpa.20811.

Suzuki, T., Inoue, T., 2007. Earliest evidence of spinal tuberculosis from the Aneolithic Yayoi period in Japan. Int. J. Osteoarchaeol. 17, 392−492.

Székely, R., Cole, S.T., 2016. Mechanistic insight into mycobacterial MmpL protein function. Mol. Microbiol. 99, 831−834. https://doi.org/10.1111/mmi.13306.

Taylor, G.M., Murphy, E., Hopkins, R., Rutland, P., Chistov, Y., 2007. First report of *Mycobacterium bovis* DNA in human remains from the Iron Age. Microbiology 153, 1243−1249. https://doi.org/10.1099/mic.0.2006/002154-0.

Thawornwattana, Y., Mahasirimongkol, S., Yanai, H., Myat Win Maung, H., Cui, Z., Chongsuvivatwong, V., Palittapongarnpim, P., 2021. Revised nomenclature and SNP barcode for *Mycobacterium tuberculosis* lineage 2. Microb. Genom. https://doi.org/10.1099/mgen.0.000697.

Thierry, D., Cave, M., Eisenach, K., Crawford, J., Bates, J., Gicquel, B., Guesdon, J., 1990. IS*6110*, an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Res. 18, 188. https://doi.org/10.1093/nar/18.1.188.

Tibayrenc, M., 2020. COVID-19 and the project of "European center for disease control" (ECDC). Infect. Genet. Evol. 86, 104604. https://doi.org/10.1016/j.meegid.2020.104604.

Tiwari, S., Casey, R., Goulding, C.W., Hingley-Wilson, S., Jacobs Jr., W.R., 2019. Infect and inject: how *Mycobacterium tuberculosis* exploits its major virulence-associated type VII secretion system, ESX-1. Microbiol. Spectr. 7 (3), 10−1128. https://doi.org/10.1128/microbiolspec.BAI-0024-2019.

Tordo, N., Poch, O., Ermine, A., Keith, G., Rougeon, F., 1988. Completion of the rabies virus genome sequence determination: highly conserved domains among the L (polymerase) proteins of unsegmented negative-strand RNA viruses. Virology 165, 565−576. https://doi.org/10.1016/0042-6822(88)90600-9.

Tsolaki, A.G., Hirsh, A.E., DeRiemer, K., Enciso, J.A., Wong, M.Z., Hannan, M., Goguet de la Salmoniere, Y.O., Aman, K., Kato-Maeda, M., Small, P.M., 2004. Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. Proc. Natl. Acad. Sci. U.S.A. 101, 4865−4870. https://doi.org/10.1073/pnas.0305634101.

Tsolaki, A.G., Gagneux, S., Pym, A.S., Goguet de la Salmoniere, Y.O., Kreiswirth, B.N., Van Soolingen, D., Small, P.M., 2005. Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 43, 3185−3191. https://doi.org/10.1128/JCM.43.7.3185-3191.2005.

Tulu, B., Ameni, G., 2018. Spoligotyping based genetic diversity of *Mycobacterium tuberculosis* in Ethiopia: a systematic review. BMC Infect. Dis. 18, 140. https://doi.org/10.1186/s12879-018-3046-4.

Vagene, A.J., Honap, T.P., Harkins, K.M., Rosenberg, M.S., Giffin, K., Cardenas-Arroyo, F., Leguizamon, L.P., Arnett, J., Buikstra, J.E., Herbig, A., Krause, J., Stone, A.C., Bos, K.I., 2022. Geographically dispersed zoonotic tuberculosis in pre-contact South American human populations. Nat. Commun. 13, 1195. https://doi.org/10.1038/s41467-022-28562-8.

van Embden, J.D., Cave, M.D., Crawford, J.T., Dale, J.W., Eisenach, K.D., Gicquel, B., Hermans, P., Martin, C., McAdam, R., Shinnick, T.M., et al., 1993. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J. Clin. Microbiol. 31, 406−409. https://doi.org/10.1128/jcm.31.2.406-409.1993.

van Embden, J.D.A., van Gorkom, T., Kremer, K., Jansen, R., van der Zeijst, B.A.M., Schouls, L.M., 2000. Genetic variation and evolutionary origin of the Direct repeat locus of *Mycobacterium tuberculosis* complex bacteria. J. Bacteriol. 182, 2393−2401. https://doi.org/10.1128/JB.182.9.2393-2401.2000.

van Ingen, J., Rahim, Z., Mulder, A., Boeree, M.J., Simeone, R., Brosch, R., van Soolingen, D., 2012. Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. Emerg. Infect. Dis. 18, 653−655.

van Soolingen, D., Hermans, P.W.M., de Haas, P.E.W., Sool, D.R., van Embden, J.D.A., 1991. The occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: evaluation of an insertion

sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. J. Clin. Microbiol. 29, 2578−2586. https://doi.org/10.1128/jcm.29.11.2578-2586.1991.

van Soolingen, D., Qian, L., de Haas, P.E., Douglas, J.T., Traore, H., Portaels, F., Qing, H.Z., Enkhsaikan, D., Nymadawa, P., van Embden, J.D., 1995. Predominance of a single genotype of *Mycobacterium tuberculosis* in countries of east Asia. J. Clin. Microbiol. 33, 3234−3238. https://doi.org/10.1128/jcm.33.12.3234-3238.1995.

van Soolingen, D., Hoogenboezem, T., de Haas, P.E., Hermans, P.W., Koedam, M.A., Teppema, K.S., Brennan, P.J., Besra, G.S., Portaels, F., Top, J., Schouls, L.M., van Embden, J.D., 1997. A novel pathogenic taxon of the *Mycobacterium tuberculosis* complex, Canetti: characterization of an exceptional isolate from Africa. Int. J. Syst. Bacteriol. 47, 1236−1245. https://doi.org/10.1099/00207713-47-4-1236.

van Soolingen, D., van der Zanden, A.G., de Haas, P.E., Noordhoek, G.T., Kiers, A., Foudraine, N.A., Portaels, F., Kolk, A.H., Kremer, K., van Embden, J.D., 1998. Diagnosis of *Mycobacterium microti* infections among humans by using novel genetic markers. J. Clin. Microbiol. 36, 1840−1845. https://doi.org/10.1128/JCM.36.7.1840-1845.1998.

van Soolingen, D., Borgdorff, M.W., de Haas, P.E., Sebek, M.M., Veen, J., Dessens, M., Kremer, K., van Embden, J.D., 1999. Molecular epidemiology of tuberculosis in The Netherlands: a nationwide study from 1993 through 1997. J. Infect. Dis. 180, 726−736. https://doi.org/10.1086/314930.

Veen, J., 1992. Microepidemics of tuberculosis: the stone-in-the-pond principle. Tuberc. Lung Dis. 73, 73−76. https://doi.org/10.1016/0962-8479(92)90058-R.

Vijaya-Bhanu, N., van Soolingen, D., van Embden, J.D.A., Dar, L., Pandey, R.M., Seth, P., 2002. Predominance of a novel *Mycobacterium tuberculosis* genotype in the Delhi region of India. Tuberculosis 82, 105−112. https://doi.org/10.1054/tube.2002.0332.

Vinogradova, T., Dogonadze, M., Zabolotnykh, N., Badleeva, M., Yarusova, I., Vyazovaya, A., Gerasimova, A., Zhdanova, S., Vitovskaya, M., Solovieva, N., Pasechnik, O., Ogarkov, O., Mokrousov, I., 2021. Extremely lethal and hypervirulent *Mycobacterium tuberculosis* strain cluster emerging in Far East, Russia. Emerg. Microbes Infect. 10, 1691−1701. https://doi.org/10.1080/22221751.2021.1967704.

Vyazovaya, A., Proshina, E., Gerasimova, A., Avadenii, I., Solovieva, N., Zhuravlev, V., Narvskaya, O., Mokrousov, I., 2020. Increased transmissibility of Russian successful strain Beijing B0/W148 of *Mycobacterium tuberculosis*: indirect clues from history and demographics. Tuberculosis 122, 101937. https://doi.org/10.1016/j.tube.2020.101937.

Wada, T., Iwamoto, T., Hase, A., Maeda, S., 2012. Scanning of genetic diversity of evolutionarily sequential *Mycobacterium tuberculosis* Beijing family strains based on genome wide analysis. Infect. Genet. Evol. 12, 1392−1396. https://doi.org/10.1016/j.meegid.2012.04.029.

Walker, T.M., Ip, C.L., Harrell, R.H., Evans, J.T., Kapatai, G., Dedicoat, M.J., Eyre, D.W., Wilson, D.J., Hawkey, P.M., Crook, D.W., Parkhill, J., Harris, D., Walker, A.S., Bowden, R., Monk, P., Smith, E.G., Peto, T.E., 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect. Dis. 13, 137−146. https://doi.org/10.1016/S1473-3099(12)70277-3.

Wan, K., 2007. Study on Genetic Diversity and Evolution of *Mycobacterium tuberculosis* the Agent of Tuberculosis in China. University Paris-Sud, Paris.

Wan, X., Qian, L., Hou, S., Drees, K.P., Foster, J.T., Douglas, J.T., 2014. Complete genome sequences of Beijing and Manila family strains of *Mycobacterium tuberculosis*. Genome Announc. 2. https://doi.org/10.1128/genomeA.01135-14.

Wan, X., Koster, K., Qian, L., Desmond, E., Brostrom, R., Hou, S., Douglas, J.T., 2017. Genomic analyses of the ancestral Manila family of *Mycobacterium tuberculosis*. PLoS One 12, e0175330. https://doi.org/10.1371/journal.pone.0175330.

Wang, J., McIntosh, F., Radomski, N., Dewar, K., Simeone, R., Enninga, J., Brosch, R., Rocha, E.P., Veyrier, F.J., Behr, M.A., 2015. Insights on the emergence of *Mycobacterium tuberculosis* from the analysis of *Mycobacterium kansasii*. Genome Biol. Evol. 7, 856−870. https://doi.org/10.1093/gbe/evv035.

Warren, R.M., Streicher, E.M., Sampson, S.L., van der Spuy, G.D., Richardson, M., Nguyen, D., Behr, M.A., Victor, T.C., van Helden, P.D., 2002a. Microevolution of the direct repeat region of *Mycobacterium tuberculosis*: implications for interpretation of spoligotyping data. J. Clin. Microbiol. 40, 4457−4465. https://doi.org/10.1128/JCM.40.12.4457-4465.2002.

Warren, R.M., Streicher, E.M., Charalambous, S., Churchyard, G., van der Spuy, G.D., Grant, A.D., van Helden, P.D., Victor, T.C., 2002b. Use of spoligotyping for accurate classification of recurrent tuberculosis. J. Clin. Microbiol. 40, 3851−3853. https://doi.org/10.1128/JCM.40.10.3851-3853.2002.

Warren, R.M., van der Spuy, G.D., Richardson, M., Beyers, N., Booysen, C., Behr, M.A., van Helden, P.D., 2002c. Evolution of the IS6110-based restriction fragment length polymorphism pattern during the transmission of *Mycobacterium tuberculosis*. J. Clin. Microbiol. 40, 1277−1282.

WHO, 2019. Global Tuberculosis Report 2019. WHO, Geneva, Switzerland. https://www.who.int/publications/i/item/9789241565714.

WHO, 2020. Global Tuberculosis Report 2020. WHO, Geneva, Switzerland. https://www.who.int/publications/i/item/9789240013131.

WHO, 2021a. Global Tuberculosis Report 2021. WHO, Geneva, Switzerland. https://www.who.int/publications/i/item/9789240037021.

WHO, 2021b. Catalogue of Mutations in *Mycobacterium tuberculosis* Complex and Their Association with Drug Resistance. WHO, Geneva, p. 96. https://www.who.int/publications/i/item/9789240028173.

Xia, E., Teo, Y.Y., Ong, R.T., 2016. SpoTyping: fast and accurate in silico *Mycobacterium* spoligotyping from sequence reads. Genome Med. 8, 19. https://doi.org/10.1186/s13073-016-0270-7.

Xu, A.M., He, C.J., Cheng, X., Abuduaini, A., Tuerxun, Z., Sha, Y.Z., Kaisaier, A., Peng, H.M., Zhen, Y.H., Zhang, S.J., Xu, J.R., Li, L., Zou, X.G., 2022. Distribution and identification of *Mycobacterium tuberculosis* lineage in Kashgar prefecture. BMC Infect. Dis. 22, 312. https://doi.org/10.1186/s12879-022-07307-4.

Yamamoto, K., Nakata, N., Mukai, T., Kawagishi, I., Ato, M., 2021. Coexpression of MmpS5 and MmpL5 contributes to both efflux transporter MmpL5 trimerization and drug resistance in *Mycobacterium tuberculosis*. mSphere 6 (1), e00518−e00520. https://doi.org/10.1128/mSphere.00518-20.

Yeboah-Manu, D., Asante-Poku, A., Bodmer, T., Stücki, D., Koram, K., Bonsu, F., Pluschke, G., Gagneux, S., 2011. Genotypic diversity and drug susceptibility patterns among *M. tuberculosis* complex isolates from South-Western Ghana. PLoS One 6, e21906. https://doi.org/10.1371/journal.pone.0021906.

Yen, S., Bower, J.E., Freeman, J.T., Basu, I., O'Toole, R.F., 2013. Phylogenetic lineages of tuberculosis isolates in New Zealand and their association with patient demographics. Int. J. Tuberc. Lung Dis. 17, 892−897. https://doi.org/10.5588/ijtld.12.0795.

Yenew, B., Ghodousi, A., Diriba, G., Tesfaye, E., Cabibbe, A.M., Amare, M., Moga, S., Alemu, A., Dagne, B., Sinshaw, W., Mollalign, H., Meaza, A., Tadesse, M., Gamtesa, D.F., Abebaw, Y., Seid, G., Zerihun, B., Getu, M., Chiacchiaretta, M., Gaudin, C., Marceau, M., Didelot, X., Tolera, G., Abdella, S., Kebede, A., Getahun, M., Mehammed, Z., Supply, P., Cirillo, D.M., 2023. A smooth tubercle bacillus from Ethiopia phylogenetically close to the *Mycobacterium tuberculosis* complex. Nat. Commun. 14, 7519.

Yimer, S.A., Norheim, G., Namouchi, A., Zegeye, E.D., Kinander, W., Tonjum, T., Bekele, S., Mannsaker, T., Bjune, G., Aseffa, A., Holm-Hansen, C., 2015. *Mycobacterium tuberculosis* lineage 7 strains are associated with prolonged patient delay in seeking treatment for pulmonary tuberculosis in Amhara Region, Ethiopia. J. Clin. Microbiol. 53, 1301−1309. https://doi.org/10.1128/JCM.03566-14.

Yuen, L.K., Ross, B.C., Jackson, K.M., Dwyer, B., 1993. Characterization of *Mycobacterium tuberculosis* strains from Vietnamese patients by Southern blot hybridization. J. Clin. Microbiol. 31, 1615−1618. https://doi.org/10.1128/jcm.31.6.1615-1618.1993.

Zakham, F., Sironen, T., Vapalahti, O., Kant, R., 2021. Pan and core genome analysis of 183 *Mycobacterium tuberculosis* strains revealed a high inter-species diversity among the human adapted strains. Antibiotics (Basel) 10 (5), 500. https://doi.org/10.3390/antibiotics10050500.

Zhdanova, S., Heysell, S.K., Ogarkov, O., Boyarinova, G., Alexeeva, G., Pholwat, S., Zorkaltseva, E., Houpt, E.R., Savilov, E., 2013. Primary multidrug-resistant *Mycobacterium tuberculosis* in 2 regions, eastern Siberia, Russian Federation. Emerg. Infect. Dis. 19, 1649−1652. https://doi.org/10.3201/eid1910.121108.

Zimpel, C.K., Patane, J.S.L., Guedes, A.C.P., de Souza, R.F., Silva-Pereira, T.T., Camargo, N.C.S., de Souza Filho, A.F., Ikuta, C.Y., Neto, J.S.F., Setubal, J.C., Heinemann, M.B., Guimaraes, A.M.S., 2020. Global distribution and evolution of *Mycobacterium bovis* lineages. Front. Microbiol. 11, 843. https://doi.org/10.3389/fmicb.2020.00843.

Zink, A.R., Sola, C., Reischl, U., Grabner, W., Rastogi, N., Wolf, H., Nerlich, A.G., 2003. Characterization of *Mycobacterium tuberculosis* complex DNAs from Egyptian mummies by spoligotyping. J. Clin. Microbiol. 41, 359−367. https://doi.org/10.1128/JCM.41.1.359-367.2003.

Zumarraga, M.J., Bernadelli, A., Bastida, R., Quse, V., Loureiro, J., Cataldi, A., Bigi, F., Alito, A., Castro-Ramos, M., Samper, S., Otal, I., Martin, C., Romano, M.I., 1999. Molecular characterization of mycobacterial isolates from seals. Microbiology (Read.) 145, 2519−2526. https://doi.org/10.1099/00221287-145-9-2519.

Zwyer, M., Çavusoglu, C., Ghielmetti, G., et al., 2021. A new nomenclature for the livestock-associated *Mycobacterium tuberculosis* complex based on phylogenomics. Open Res. Eur. 1, 100. https://doi.org/10.12688/openreseurope.14029.2, 1.

Further Reading

Gagneux, S., 2017. Strain variation in the Mycobacterium tuberculosis complex: its role in biology, epidemiology and control. In: Springer.

CHAPTER

The Evolution of *Staphylococcus aureus*

28

JoAnn McClure[1] and Kunyan Zhang[1,2]

[1]*Centre for Antimicrobial Resistance, Alberta Health Services/Alberta Precision Laboratories /University of Calgary, Calgary, AB, Canada;* [2]*University of Calgary, Pathology and Laboratory Medicine, Calgary, AB, Canada*

1. Introduction

Staphylococcus aureus, first isolated in 1880 by Alexander Ogston from a surgical wound infection (Ogston, 1881), is a Gram-positive, spherical bacterium that forms grape-like clusters. While the organism is known to asymptomatically colonize the nasal passages up to 30% of people (Gorwitz et al., 2008; Kluytmans et al., 1997), it is also associated with mild to moderate diseases such as skin soft tissue infections, and more severe infections like endocarditis, osteomyelitis, and sepsis (Diekema et al., 2001). *S. aureus* is equipped with an extensive array of virulence factors and toxins, responsible for diseases such as toxic shock syndrome and scalded skin syndrome, as well as allowing the bacteria to modulate and/or evade the human immune response. Of clinical importance is the ability of *S. aureus* to acquire resistance to multiple antibiotic classes, complicating treatment. Within 2 years following the introduction of penicillin, penicillin-resistance *S. aureus* was observed, and shortly after the introduction of methicillin (the semisynthetic penicillin), methicillin-resistant *Staphylococcus aureus* (MRSA) was identified. Infections with MRSA are associated with higher mortality rates, longer hospital stays and higher healthcare costs than infections with methicillin-susceptible strains (Antonanzas et al., 2015; Thampi et al., 2015; Whitby et al., 2001). As such, the CDC considers methicillin-resistant *Staphylococcus aureus* (MRSA) an urgent threat to public health, and the Public Health Agency of Canada (PHAC) regards it as one of the top-priority antimicrobial-resistant pathogens (Garner et al., 2015).

The epidemiology and prevalence of *S. aureus*, particularly MRSA, is constantly changing, both over time and within/between geographic locations. MRSA was initially associated with health-care settings, but in the 1980s emerged as a community-associated (CA) pathogen, having acquires a unique molecular profile, distinct from the previously described hospital-associated (HA) MRSA (Boyle-Vavra and Daum, 2007; Chambers and DeLeo, 2009; Herold et al., 1998). CA-MRSA was initially limited to populations outside of healthcare settings, and responsible primarily for mild infections, forming the basis for their differentiation from HA-MRSA. The epidemiological and molecular distinction between HA-MRSA and CA-MRSA has, however, become blurred, as CA-MRSA have become a common etiological agent of nosocomial outbreaks as well (Bratu et al., 2005; Maltezou et al., 2009; Saunders et al., 2007; Teare et al., 2010). In recent years livestock-associated (LA-)

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00014-8

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

MRSA, genetically distinct from both HA- and CA-MRSA, have also emerged as pathogens of concern, initially associated with farmers, but now also seen in people lacking livestock association (Benito et al., 2014; Stein, 2009; Van Rijen et al., 2008; Voss et al., 2005).

Understanding the molecular epidemiology of *S. aureus* is essential for controlling the spread of this pathogen with wide reaching effects on human health. In this chapter, the evolutionary nature of *S. aureus* is discussed.

2. Evolution of *Staphylococcus aureus*

A commonly held theory is that *S. aureus* was initially susceptible to newly developed antibiotics, however, due to horizontal gene transfer and chromosomal mutations, over time it has acquired resistance to many classes of antibiotics. While these resistance genes may have preexisted undetected in *S. aureus* populations, the species is in fact remarkable in its ability to acquire resistance and virulence determinants. Not only has this acquisition presented a major challenge to the treatment of staphylococcal infections but it has also been a driving force in its global evolution, as antibiotics represent the strongest selective pressure faced by the species. A detailed timeline, presenting *S. aureus*' major evolutionary milestones is shown in Fig. 28.1.

Resistance to penicillin is the first notable change for the species, mediated by acquisition of a plasmid-encoded penicillinase, capable of hydrolyzing the antibiotic's beta-lactam ring and inactivating it (Bondi and Dietz, 1945; Lowy, 2003). Prior to the discovery of penicillin in the early 1940s, the mortality rate from serious *S. aureus* infections was as high as 80% (Skinner and Keefer, 1941). Discovery of penicillin in the early 1940s dramatically increased survival for patients, instrumental in decreasing the death toll due to bacterial pneumonia and meningitis among soldiers in World War II. As early as 1942, however, penicillin-resistant *S. aureus* were identified, rapidly becoming the predominant strains in hospitals by the mid-1940s, and community by the 1950s (Kirby, 1944; Lowy, 2003; Rammelkamp and Maxon, 1942). In the 1950s, the penicillin-resistant *S. aureus* clone known as phage-type 80/81 (belonging to CC30) was responsible for epidemics in Australia, North America, and the United Kingdom, first emerging in hospitals, then spreading to the community (Bynoe et al., 1956; Rountree and Freeman, 1955; Shaffer et al., 1957). Phage-type 80/81 *S. aureus*, however, largely disappeared following the introduction of methicillin, the first semisynthetic antibiotic, in the late 1950s (Jevons and Parker, 1964). *S. aureus* strains affecting humans were later shown to fall primarily into one of 10 clonal complexes, CC1, CC5, CC8, CC12, CC15, CC22, CC25, CC30, CC45, and CC151/121, with CC30 accounting for the largest percentage (Enright et al., 2000, 2002; Feil et al., 2003, 2004; Feil and Enright, 2004; Feng et al., 2008; Goering et al., 2008; Hallin et al., 2007; Tenover et al., 2008). These clones are represented in Fig. 28.2, where a whole genome phylogenetic dendrogram shows that they are genetically distinct linages.

Shortly after the introduction of methicillin in 1959, the first methicillin-resistant *S. aureus* was identified in a hospital in the United Kingdom (Jevons, 1961). Throughout the 1960s and 1970s, the "archaic" clone (ST250-MRSA-I) of MRSA circulated primarily in European hospitals, however, by the 1980s it had mostly disappeared and was replaced by five prevalent clonal lineages, which rapidly spread and became a global pandemic (Crisóstomo et al., 2001; Enright et al., 2002; Robinson and Enright, 2003). These include CC5, CC8 (including the CC8-ST239 subgroup), CC22, CC30, and CC45 (Enright et al., 2002), each of which is shown in Fig. 28.2, co-localized with an associated

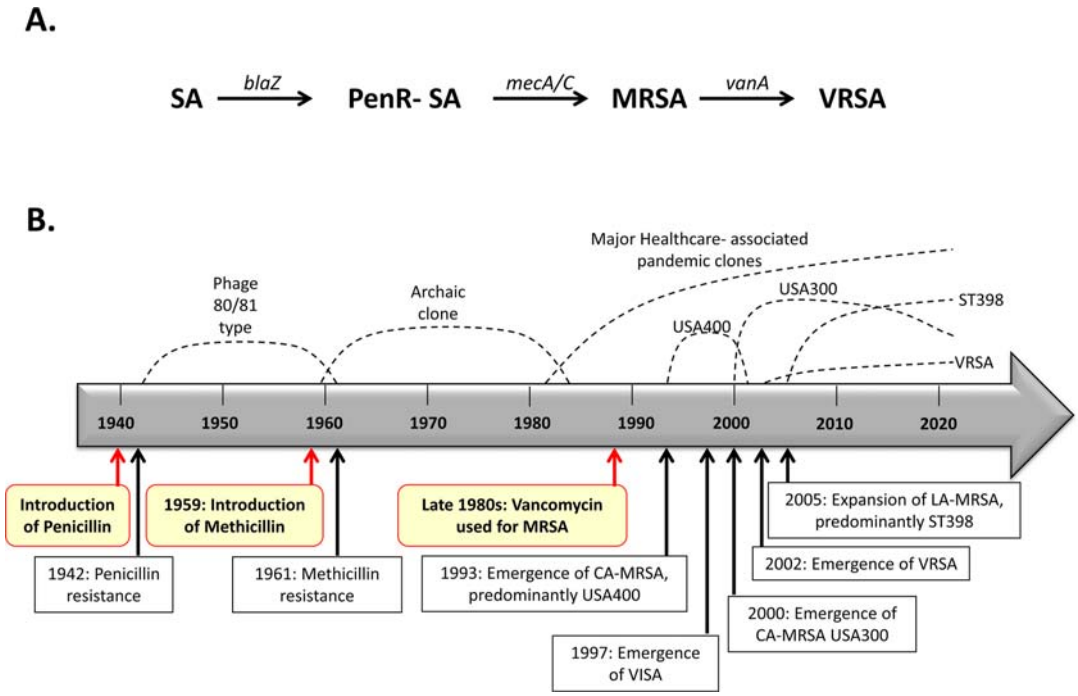


FIGURE 28.1 ***Staphylococcus aureus*** **(SA) evolution.**

(A) The evolution of antibiotic resistance in *S. aureus* is depicted, along with the major genes associated with the resistance. (B) Evolutionary timeline of *S. aureus* showing the major advancements below the line, and predominant clones above.

blaZ, beta-lactamase resistance gene; *mecA/C*, methicillin-resistance gene A or C; *PenR*, penicillin-resistant; *SA*, Staphylococcus aureus; *vanA*, vancomycin resistance gene A.

MSSA lineage. Outbreaks of MRSA were subsequently reported in the United States, Australia, the United Kingdom, Ireland, and Asia during the 1980s and 1990s, however, remained primarily linked to healthcare-associated facilities (Bradley et al., 1985; Cafferkey et al., 1985; Faoagali et al., 1992; Hanifah et al., 1992; Thompson et al., 1982). In recent decades, other MRSA lineages have emerged as pandemic clones, such as CC59, CC398, while strains such CC80, CC72, and CC152 are also increasing in frequency and may represent future epidemic lineages. In fact, MRSA strains have continued to emerge and decline in various geographic locations, and have continued to accumulate virulence and resistance determinants, increasing the burden of MRSA infections both from a health and economic standpoint.

The next major development in the evolution of *S. aureus* was the emergence of community-associated MRSA, which rapidly increased in incidence and was associated with serious skin and soft tissue infections, often in young and healthy individuals. First identified in 1993 in an aboriginal population in Western Australia, strain WA-1 was distinct from the circulating MRSA and lacked multidrug resistance (Udo et al., 1993). Strain MW2, a close relative of WA-1, was next reported in the

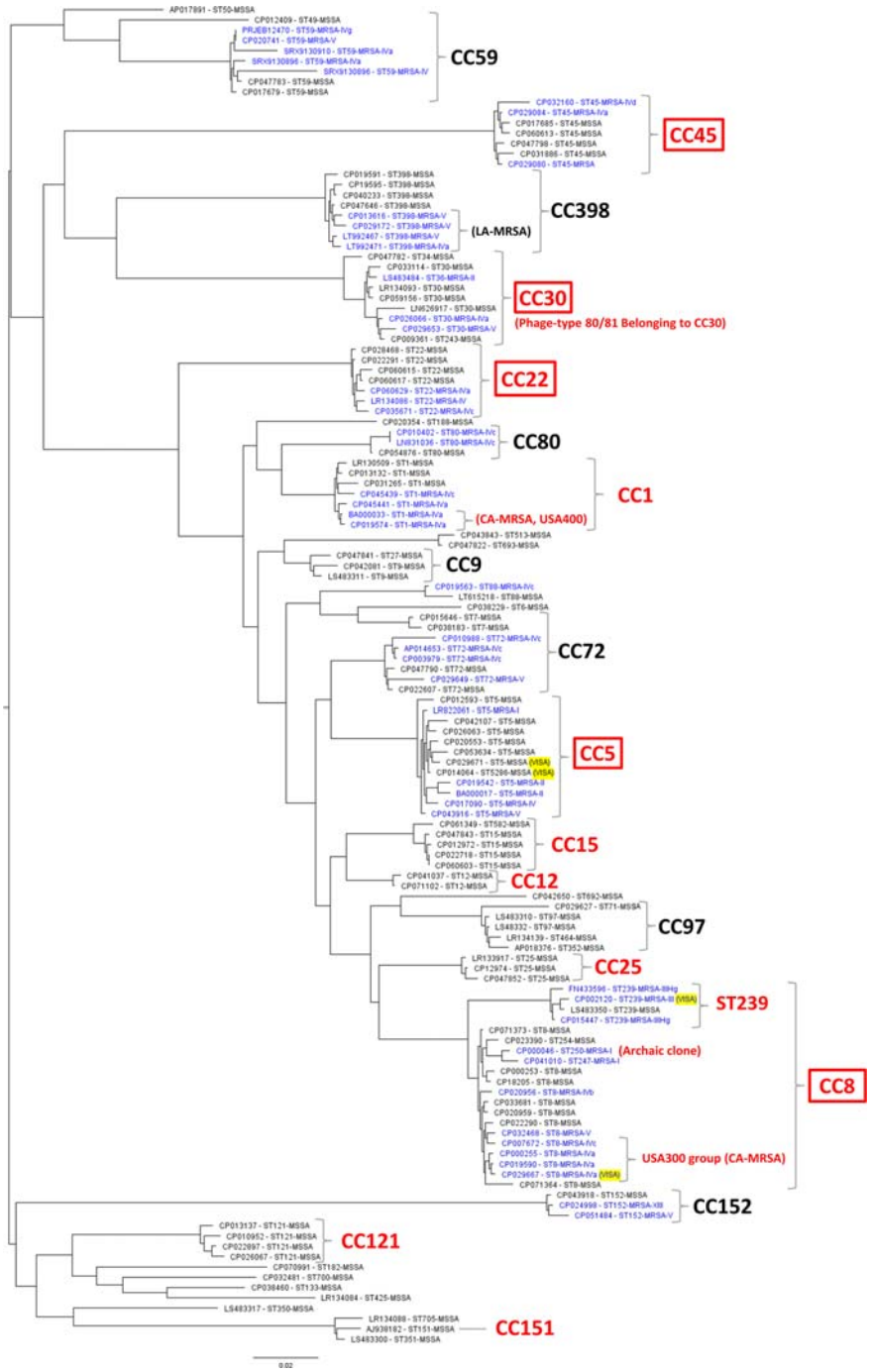


FIGURE 28.2 Single nucleotide polymorphism (SNP)-whole genome sequence (WGS) phylogenetic dendrogram showing representatives from significant lineages in *S. aureus* evolution.

United States between 1997−1999, causing sepsis and necrotizing pneumonia in otherwise healthy children (Centers for Disease Control and Prevention, 1999; David and Daum, 2010). In both situations, patients lacked healthcare association and had no identifiable risk factors for MRSA infection. These strains belonged to pulsed-field type USA400 (ST1-MRSA-IV), representing the most common CA-MRSA in the United States before 2001. However, starting in 2000 they were replaced by an unrelated CA-MRSA strain, USA300 (ST8-MRSA-IVa) (Como-Sabetti et al., 2009; McDougal et al., 2003; Stemper et al., 2006). As with the previous strain, USA300 was associated with infections of apparently healthy individuals lacking hospital association (Adcock et al., 1998; Aiello et al., 2006; Begier et al., 2004; Stemper et al., 2006). USA300 appears to be declining in frequency in North America (Glaser et al., 2016), while different CA-MRSA are emerging as epidemic lineages in other parts of the world (such as CC59 in Asia). These strains have now blurred ecological boundaries appearing in hospitals as nosocomial pathogens, complicating the definition of community-associated (Patel et al., 2008; Seybold et al., 2006). As such, the group is now identified based on a combination of their epidemiology, genetic profiles, and antibiotic resistance patterns (David et al., 2008; Otter and French, 2012).

Livestock-associated MRSA represents another big jump in staphylococcal evolution. First identified in the early 1970s, LA-MRSA are capable of colonizing and infecting economically important animals such as sheep, cows, goats, poultry, rabbits, and horses, as well as domestic animals such as dogs and cats (Cuny et al., 2010; Weese, 2010). LA-MRSA, however, gained notoriety when they were responsible for high rates of colonization and infection among Dutch pig farmers in 2005 (Armand-Lefevre et al., 2005; Huijsdens et al., 2006; Voss et al., 2005). Many reports of infection in humans working in close contact with livestock followed, but of greater concern were the reports of infections in humans lacking any livestock associated contact (Armand-Lefevre et al., 2005; Benito et al., 2014; Brunel et al., 2014; Graveland et al., 2011; Grisold et al., 2010; Huijsdens et al., 2006; Uhlemann et al., 2012, 2013, 2017; Van Belkum et al., 2008; Van Rijen et al., 2008; Voss et al., 2005; Wulf et al., 2012). LA-MRSA is genetically distinct from human isolates, with CC398 (ST398) representing the most commonly encountered clone.

In recent years, the increased use of vancomycin, the only antibiotic for which MRSA were reliably sensitive, has led to emergence of vancomycin-intermediate (VISA) and vancomycin-resistant *S. aureus* (VRSA), raising alarms that they may become completely untreatable infections. While vancomycin was approved for human use in 1958, it wasn't used for the treatment of MRSA until the late 1980s (Levine, 2006; Sorrell et al., 1982). In 1997, a clinical *S. aureus* isolate from Japan was found to have reduced susceptibility to vancomycin (VISA), however, retrospective studies suggest that VISA dated back to 1987 in the United Sates (Hiramatsu et al.,

A high diversity of MSSA is seen, while the archaic clones, major hospital MRSA CCs, CA-MRSA CCs, and LA-MRSA CCs are seen as unique branches on the tree, genetically distinct from each other. The tree was generated with CSI phylogeny (https://cge.food.dtu.dk/services/CSIPhylogeny/) using ATCC 12228 (AE015929) as the reference/outgroup strain, and midpoint rooted. MSSA is indicated in black font and MRSA in *blue* font. Predominant MSSA CCs are indicated with *red* font, while the predominant MRSA CCs are also boxed in *red*. VISA strains are highlighted in *yellow*. (For interpretation of the references to color in this figure legend, please refer online version of this title.)
CA-, community-associated; *CC*, clonal complex; *LA-*, livestock-associated; *MRSA*, methicillin-resistant Staphylococcus aureus; *MSSA*, methicillin-susceptible Staphylococcus aureus; *VISA*, vancomycin-intermediate Staphylococcus aureus.

1997; Jackson and Hicks, 1987). VISA strains show resistance to vancomycin concentrations of 4−8 μg/mL and are typically associated with prolonged vancomycin treatment in hospitalized patients, with mutations in several genes known to contribute to development of the phenotype. Complete vancomycin resistance (MIC ≥16 μg/mL) was first discovered in the 1980s in enterococci, conferred by the *vanA* operon, which encodes an altered peptidoglycan that is no longer susceptible to vancomycin (Bugg et al., 1991; Murray, 2000). The *vanA* operon is carried on transposon Tn*1546*, which is part of an enterococcal conjugative plasmid (Arthur et al., 1993). High level Vancomycin resistance was reported in *S. aureus* in 2002, resulting from carriage of the *vanA* gene on Tn*1546*, believed to have transposed from the enterococcal plasmid to a staphylococcal plasmid (Centers for Disease Control and Prevention, 2002; Chang et al., 2003; Périchon and Courvalin, 2009; Zhu et al., 2010). Even though VRSA was first reported in the United States (Michigan), it has remained relatively limited in its diversity, spread and geographic distribution (Centers for Disease Control and Prevention, 2002). In the United States, 15 VRSA isolates have been identified, all belonging to CC5 (ST5) (Kos et al., 2012). Subsequent studies have identified VRSA carrying the *vanA* gene in India, Iran, Pakistan, Brazil, and Portugal, and a recent review indicates that VRSA have been isolated from at least 15 countries (Aligholi et al., 2008; Azhar et al., 2017; Azimian et al., 2012; Dezfulian et al., 2012; Melo-Cristino et al., 2013; Mirani and Jamil, 2011; Panesso et al., 2015; Rossi et al., 2014; Saha et al., 2008; Thati et al., 2011; Tiwari and Sen, 2006; Wu et al., 2021). Despite this increase, the incidence of VRE remains low, however, adherence to stringent infection control practices will be needed to prevent this threat from becoming a global calamity.

The prevalence and epidemiology of *S. aureus* is constantly changing, influenced by the acquisition of antibiotic resistance determinants and ecological niche shifts. Acquisition of penicillin resistance was followed by acquisition of methicillin resistance then vancomycin resistance, while *S. aureus* moved from being primarily a hospital-associated pathogen to also being prevalent in the community and associated with livestock. While much of the focus over the years has been given to MRSA, MSSA is also responsible for causing significant diseases, but are often overlooked in terms of tracking and characterization. As such, the true diversity of epidemic staphylococcal lineages may be significantly underestimated.

3. Methicillin Resistance and the Staphylococcal Cassette Chromosome *mec*

Resistance to methicillin results from acquisition of the *mecA* gene, which codes for an altered penicillin-binding protein (PBP2a or PP2′) that has low affinity for semisynthetic penicillins (Hartman and Tomasz, 1984; Reynolds and Brown, 1985; Utsui and Yokota, 1985). While the origin of the *mecA* gene has not been definitively identified, researchers have proposed that it originated in *S. sciuri* or *S. fleurettii*, where close homologs were found (Tsubakishita et al., 2010; Wu et al., 2001). For many years, *mecA* was the sole methicillin-resistance gene described in *S. aureus*. However, in recent years, homologs *mecC* and *mecB* have also been identified, though less frequently encountered (Baba et al., 2009; Becker et al., 2018; García-Álvarez et al., 2011). *mecA* is widely disseminated among staphylococcal species and was found to be carried on a mobile genetic element (MGE), termed the

staphylococcal chromosome cassette *mec* (SCC*mec*) (Katayama et al., 2000). SCC*mec* is a 20−60 kb genetic element inserted at an *attB* site in the *orfX* gene of the staphylococcal chromosome. Other genetic components can be incorporated into SCC*mec*, including plasmids (pT181, pUB110) and transposons (Tn554), which can contain resistance determinants to other classes of antibiotics and/or heavy metals (International Working Group on the Classification of Staphylococcal Cassette Chromosome, Elements, 2009). The structure and content of SCC*mec* elements are highly varied, however, they share common components including a *mec* gene complex, a *ccr* gene complex, and joining regions.

3.1 *mec* Gene Complex

The *mec* gene complex is composed of the methicillin resistance gene along with its regulatory genes (*mecI* and *mecR1*) and associated insertion sequences (IS431 and IS1272). Based on differences in these components, the *mec* gene complex has been classified into five major classes (A-E), as depicted in Fig. 28.3 (Ito and Hiramatsu, 1998; Katayama et al., 2001; Shore et al., 2011). Variants of the major classes, with different length truncations or containing other insertion sequences or transposons, are specified by a numerical string following the class.

The class A *mec* gene complex contains a complete *mecR1* and *mecI* upstream of *mecA*, followed by the hypervariable region (HVR) and IS431. The class B *mec* gene complex carries a IS1272 and truncated mecR1 (*ΔmecR1*) upstream of *mecA*, followed by the HVR and IS431. There are two distinct versions of the class C *mec* gene complex, termed C1 and C2. In both cases they are composed of a truncated *mecR1* upstream of *mecA* and the HVR downstream, with IS431 on either end of the complex. In class C1 both the upstream and downstream IS431 share the same orientation, while in class C2 the upstream IS431 is in the reverse orientation. The class D *mec* gene complex carries a truncated *mecR1*, *mecA* and IS431, however does not have an IS adjacent to the *ΔmecR1*. Finally, the class E *mec* gene complex is unique in carrying the β-lactamase gene (*blaZ*) and *mecC*, along with complete *mecI* and *mecR1* genes.

First reported in 2011, *mecC* is a homolog of *mecA*, sharing 69% identity at the sequence level and 63% identity at the amino acid level (García-Álvarez et al., 2011). It was identified in *S. aureus* strain LGA251, isolated from a milk sample in southwest England. The strain was resistant to oxacillin and cefoxitin but *mecA* PCR and PBP2a slide agglutination negative, and sequencing of the genome revealed the novel *mec* homolog. There is speculation that *mecC* could have originated in coagulase-negative staphylococci, and two allotypes of *mecC* (*mecC1* and *mecC2*) were identified in *S. xylosus* and *S. saprophyticus*, however, the issue remains unresolved (Harrison et al., 2013; Malyszko et al., 2014).

Other homologs of *mecA* have also been described. In 2009, *mecB* was identified in the closely related species, *Macrococcus caseolyticus,* and in 2018 it was detected on a plasmid in *S. aureus* (Baba et al., 2009; Becker et al., 2018). The *mecB* gene has subsequently been identified in SCC*mec* elements in *M. caseolyticus* (SCC*mec*KM45013), and *M. bohemicus* (SCC*mec* $_{H889678/16/1}$) (Foster and Paterson, 2020; Gómez-Sanz et al., 2015). Another homolog, *mecD*, was also described on a genomic island in *M. caseolyticus*, but has not been detected in *S. aureus* to date (Schwendener et al., 2017). Recently, *mecD* was described as occurring in a composite SCC*mec* element (SCC*mec* $_{19Msa422}$) in *M. bohemicus* (Schwendener et al., 2021).

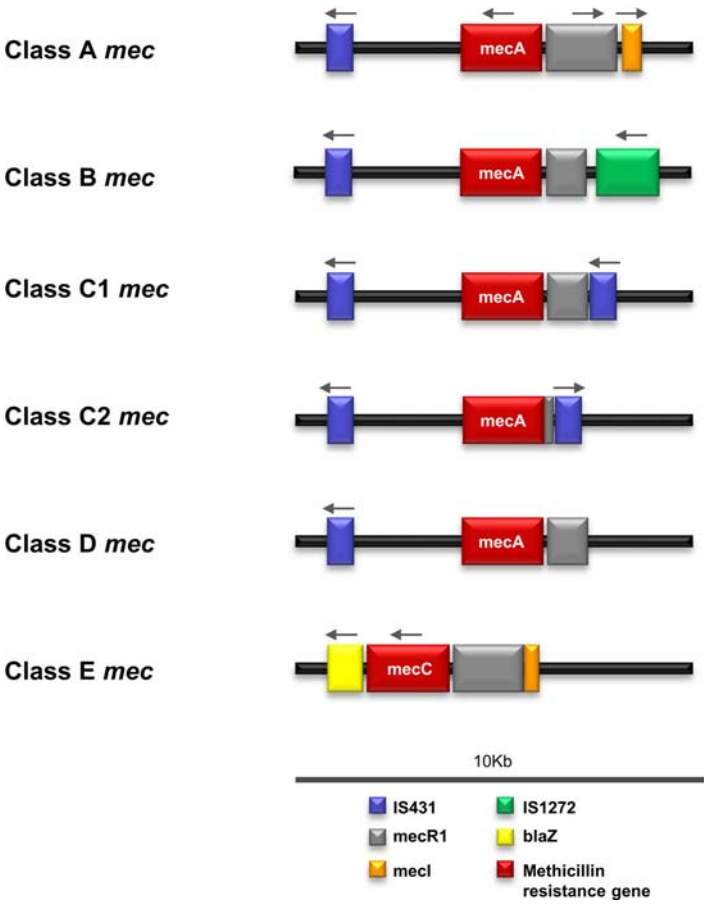


FIGURE 28.3 Structure of the five classes (A to E) of *mec* gene complex.

Class A from SCC*mec* II (D86934) shows the typical arrangement of IS*431-mecA-mecR1-mecI*. Class B from SCC*mec* I (AB033763) shows the typical structure of IS*431-mecA*-truncated *mecR1*-IS*1272*. Class C1 from SCC*mec* VII (AB373032) has the IS*431-mecA*-truncated *mecR1*-IS*431* structure with the IS*431* elements coding in the same direction, while class C2 from SCC*mec* V (AB121219) has the same genes, however, the IS*431* elements are coding in opposite directions. The class D *mec* complex is represented in *S. caprae* (strain JA186) and contains IS*431-mecA*-truncated *mecR1* but no downstream IS, while the class E complex from SCC*mec* XI (FR821779) is composed of *blaZ-mecC-mecR1-mecI*. *Purple* boxes, IS*431*; *gray* boxes, *mecR1*; *orange* boxes, *mecI*; *green* boxes, IS*1272*; *yellow* boxes, *blaZ*; *red* boxes, methicillin resistance gene (*mecA* or *mecC*). *Arrows* above the genes represent the orientation of the coding sequence. A 10 kb scale bar is shown. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

3.2 *ccr* Gene Complex

Present in the *ccr* gene complex are the cassette chromosome recombinase genes (*ccrAB* and/or *ccrC*), responsible for site specific integration and excision of SCC*mec* from the chromosome. Initially identified were the *ccrA* and *ccrB* genes, which function together and are both needed to

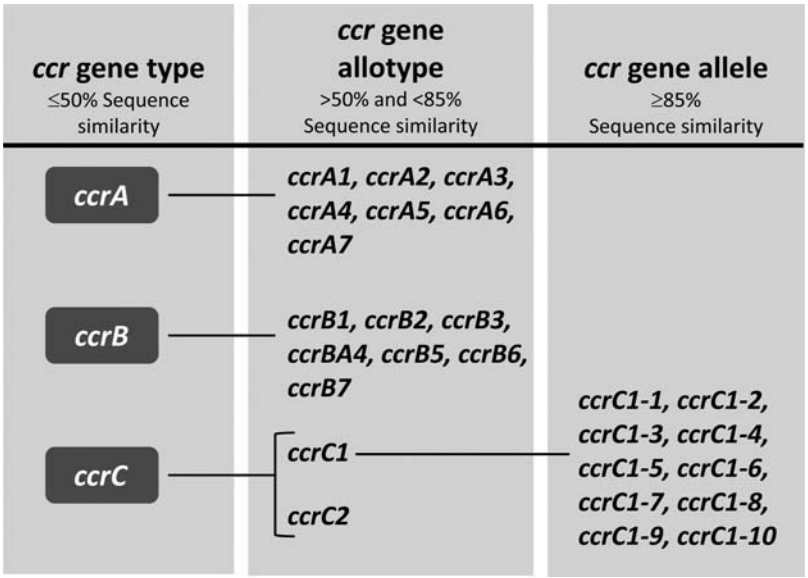


FIGURE 28.4 Classification of *ccr* gene homologues.

Three types of *ccr* genes have been described (*ccrA*, *ccrB*, and *ccrC*), sharing ≤50% sequence similarity. The types are divided into allotypes based on sharing >50% to ≤85% sequence similarity. *ccrA* and *ccrB* each have seven allotype (1−7), while *ccrC* has 2 (*ccrC1* and *ccrC2*). The *ccrC1* allotype is further divided into 10 alleles (*ccrC*1-1 to *ccrC*1-10), each sharing ≥85% sequence similarity. Representative strains and corresponding GenBank accession numbers are as follows: *ccrA1* and *ccrB1*, NCTC10442 (AB033763); *ccrA2* and *ccrB2*, N315 (D86934); *ccrA3* and *ccrB3*, 85/2082 (AB014436); *ccrA4* and *ccrB4*, ATCC 12228 (AE015929); *ccrA5* and *ccrB5*, KM241 (AM904731); *ccrA6*, SHP6P021P (MN220716); *ccrB6*, ATCC 15305 (AP008934); *ccrA7* and *ccrB7*, TSU33 (AB353724); *ccrC1-1*, WIS (AB121219); *ccrC1-2*, TSGH17 (AY894416); *ccrC1-3*, 85/2082 (AB037671); *ccrC1-4*, M (U10927); *ccrC1-5*, JCSC1435 (AP006716); *ccrC1-6*, 25−60 (EF190467); *ccrC1-7*, 13−48 (EF190468); *ccrC1-8*, P1 (AB462393); *ccrC1-9*, ATCC 15305 (AP008934); *ccrC1-10*, M06/0171 (HE980450); *ccrC2*, BA01611 (KR187111).

mobilize the SCC*mec* cassette, however, in 2004, Ito et al. identified a single gene coding for a novel cassette chromosome recombinase (*ccrC*), which was capable of mobilizing the SCC*mec* cassette (Ito et al., 2004; Wang and Archer, 2010). The *ccr* genes are separated based on sharing less than 50% sequence identity, and further divided into allotypes based on sharing 50%−85% identity, and different alleles sharing ≥85% sequence identity, as shown in Fig. 28.4 (International Working Group on the Classification of Staphylococcal Cassette Chromosome, Elements, 2009). Currently seven allotypes of *ccrA* and *ccrB* (*ccrA1* to *7* and *ccrB1* to *7*) have been named, as have two allotypes of *ccrC* (*ccrC1* and *ccrC2*). *ccrC1* is further divided into 10 alleles (1−10). Different combinations of *ccr* allotypes have resulted in nine different *ccr* gene complexes types; 1 (A1B1), 2 (A2B2), 3 (A3B3), 4 (A4B4), 5 (C1), 6 (A1B6), 7 (A1B6), 8 (A1B3), and 9 (C2) (International Working Group on the Classification of Staphylococcal Cassette Chromosome, Elements, 2009).

3.3 J Regions

Joining (J) regions are nonessential regions within the SCC*mec* element that are not part of the *mec* or *ccr* gene complex. The J1 regions (formerly referred to as L-C region) is located between the right chromosomal junction and the *ccr* gene complex, while the J2 region (formerly referred to as C-M region) is located between the *ccr* and *mec* gene complex, and the J3 (formerly referred to as I-R region) is located between the *mec* gene complex and left chromosomal region. The joining regions can contain resistance determinants for other antimicrobials and heavy metals, such as tetracycline, aminoglycosides, mercury, etc.

3.4 SCC*mec* Types

The specific combination of *ccr* and *mec* gene complex is used to define the SCC*mec* type of a strain, with variations in the J regions used to define subtypes. As per naming rules set out by the International Working group on the Classification of Staphylococcal Cassette Chromosome (IDW-SCC), there are currently 14 major types of SCC*mec* element described in MRSA, as depicted in Fig. 28.5 (International Working Group on the Classification of Staphylococcal Cassette Chromosome, Elements, 2009). They include SCC*mec* I (type 1 *ccr*, class B *mec*; 1B), II (type 2 *ccr,* class A *mec*; 2A), III (type 3 *ccr,* class A *mec*; 3A), IV (type 2 *ccr,* class B *mec*; 2B), V (type 5 *ccr,* class C2 *mec*; 5C2), VI (type 4 *ccr,* class B *mec*; 4B), VII (type 5 *ccr,* class C1 *mec*; 5C1), VIII (type 4 *ccr,* class A *mec*; 4A), IX (type 1 *ccr,* class C2 *mec*; 1C2), X (type 7 *ccr,* class C1 *mec*; 7C1), XI (type 8 *ccr,* class E *mec*; 8E), XII (type 9 *ccr,* class C2 *mec*; 9C2), XIII (type 9 *ccr,* class A *mec*; 9A), and XIV (type 5 ccr, class A *mec*; 5A) (Baig et al., 2018; García-Álvarez et al., 2011; International Working Group on the Classification of Staphylococcal Cassette Chromosome, Elements, 2009; Li et al., 2011; Urushibara et al., 2020; Wu et al., 2015). As mentioned, polymorphisms in the J regions give rise to SCC*mec* subtypes, often expressed as upper or lower case letters following the type (ex. IIb, IIIA).

SCC elements have also been described carrying genes other than *mecA*, though still carrying the *ccr* genes and integrated at the specific site in the chromosome with corresponding repeat elements. Examples are SCC*Hg* with mercury resistance and SCC*fur* with fusidic acid resistance. Also characterized are SCC-like regions that do not carry the *ccr* gene, such as pseudo-SCC (ΨSCC) (International Working Group on the Classification of Staphylococcal Cassette Chromosome, Elements, 2009). Elements have also been described carrying two or more *ccr* gene complexes, such as the one found in *S. aureus* strains PM1 or TSGH17 (Boyle-Vavra et al., 2005; Takano et al., 2008).

Although unable to be assigned a name, according to official criteria set out by IWG-SCC, studies have indicated that coagulase-negative Staphylococci (CoNS) carry a larger diversity of SCC elements with unique types or subtypes (Chen et al., 2017; Hanssen and Sollid, 2007). One such example is the newly described SCC*mec*-C34847 (type 1 *ccr*, class A *mec*, type 4 *ccr*) from *S. hominis*, which has a complex composite structure indicative of a high degree of recombination in the element (McClure et al., 2021b). This combination of type 1 *ccr* with class A *mec* (1A) represents a novel combination and new SCC*mec* type. It, however as mentioned, cannot be assigned an official number since it is present in CoNS rather than *S. aureus*. SCC*mec* elements in CoNS are of importance since studies have suggested that CoNS are capable of transferring SCC elements to *S. aureus*, and may thus represent a source of new elements for MRSA (Hanssen and Sollid, 2007; Wielders et al., 2001; Wisplinghoff et al., 2003).

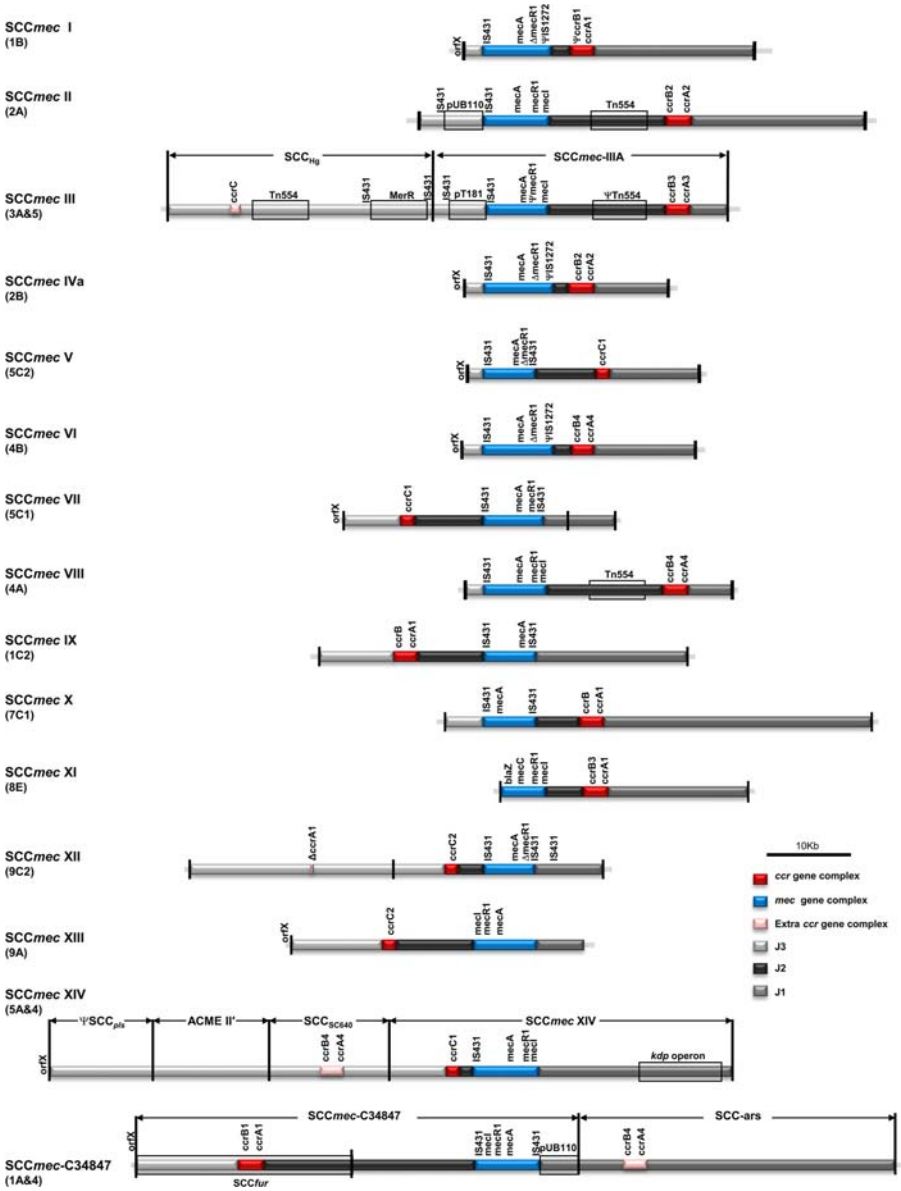


FIGURE 28.5 Illustration of the SCC*mec* types reported to date.

Fourteen SCC*mec* types (I to XIV) have been identified in *S. aureus*, as well as countless other types that are not assigned a number in CoNS. Scaled illustrations of representatives from each SCC*mec* type are shown here, along with a complex SCC*mec* representing a new type identified in *S. hominis*. The *ccr* and *mec* gene complexes are indicated, by *red* and *blue* color respectively, with corresponding genes identified. Extra *ccr* complexes, not part of the core SCC*mec* element, are indicated in *light pink*. Boundaries and/or direct/

4. Typing Methods for *S. aureus* Characterization

A thorough understanding of the epidemiology and evolution of infectious disease relies on typing methods for the characterization and discrimination of isolates. From this information, clonal relationships among strains can be deduced, allowing the temporal and geographical propagation of strains to be understood. While phenotypic methods have been used to classify strains in the past, nowadays typing schemes that rely on molecular methods are used and provide greater discriminatory power, as described below.

4.1 Pulsed Field Gel Electrophoresis (PFGE)

PFGE has long been considered the gold standard for *S. aureus* typing. First described in 1984, the technique relies on digestion of the genome into large fragments using a restriction enzyme, followed by separation of the fragments by gel electrophoresis (Goering, 2010; Schwartz and Cantor, 1984). In PFGE, the voltage direction is periodically switched during electrophoresis, resulting in a DNA fingerprint which can be used to determine the relatedness of strains. For *S. aureus*, the restriction enzyme smaI was found to be the most reproducible for typing nearly all strains (Ichiyama et al., 1991; Prevost et al., 1992). The ST398 clone is an exception, with the enzyme consensus sequence modified by a DNA methyltransferase. The neoshizomer Cfr9I is, however, able to cleave this strain at the same recognition site as smaI, producing a fingerprint for typing this lineage (Argudín et al., 2010). PFGE has a high discriminatory power and, with standardization of the method, can produce results that are reproducible at both the intra- and interlaboratory levels (Bannerman et al., 1995; Cookson et al., 2007). PFGE is limited by a long turn-around time, as well as high cost for equipment and analysis software, plus data interpretation can be problematic without high standardization. More importantly, the technique separates DNA based on size rather than sequence, meaning small changes like mutations that affect the sequence type of a strain cannot the detected, while insertions or deletions of mobile genetic elements can alter the fingerprint.

4.2 Multilocus Sequence Typing (MLST)

MLST relies on sequencing a 402−516 bp fragment from seven conserved, essential housekeeping genes (*arcC*, *aroE*, *glpF*, *gmk*, *pta*, *tpi*, and *yqiL*). These genes are stable and evolve slowly, yet provided the best resolving power for identifying *S. aureus* lineages (Enright et al., 2000). The genes for each locus are assigned an allele number based on point mutations within the sequenced region, and

inverted repeat regions are indicated by *black* bars. The J1 region is colored light gray, the J2 region in *dark gray*, and the J3 region in medium *gray*. Notable elements in the J regions are indicated by *black* boxes. Representative strains and corresponding GenBank accession numbers are as follows: SCC*mec* I (IB), NCTC10442 (AB033763); SCC*mec* II (2A), N315 (D86934); SCC*mec* III (3A), 85/2082 (AB014436); SCC*mec* IV (2B), CA05 (AB063172); SCC*mec* V (5C2), WIS (AB121219); SCC*mec* VI (4B), HDE288 (AF411935); SCC*mec* VII (5C1), JCSC6082 (AB373032); SCC*mec* VIII (4A), C10682 (FJ390057); SCC*mec* IX (1C2), JCSC6943 (AB505628); SCC*mec* X (7C1), JCSC6945 (AB505630); SCC*mec* XI (8E), LGA251 (FR821779); SCC*mec* XII (9C2), BA01611 (KR187111); SCC*mec* XIII (9A), 55-99-44 (CP024999); SCC*mec* XIV (5A), SC792 (LC440647); SCC*mec*-C34847 (1A), C34837 (KU936053). A 10 kb scale bar is shown. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

the series of seven numbers arranged in the order *arcC-aroE-glpF-gmk-pta-tip-yqiL* defines the sequence type of a strain. The database at PubMLST (https://pubmlst.org/saureus/) is used to facilitate sequence analysis, while the eBURST (Based Upon Related Sequence Type) algorithm permits analysis of the evolutionary relationships between isolates (Feil et al., 2004; Spratt et al., 2004). MLST has high discriminatory power and is highly reproducible, being unambiguous and portable between laboratories and around the world. Drawbacks, however, include the cost associated with generating high quality sequences for seven loci, as well as the intolerance for sequencing errors.

4.3 Staphylococcal Protein A (*spa*) Typing

The *spa* gene is conserved in *S. aureus* and has been used as an effective method of single locus sequence typing. The gene contains a variable X region, comprised of polymorphic variable-number of tandem repeats (VNTRs), which form the basis for *spa* typing (Frénay et al., 1996; Koreen et al., 2004; Shopsin et al., 1999). Each of the repetitive sequences is given a numerical (Ridom nomenclature system) or letter (Kreiswirth nomenclature system) code identifier, and the order of repeats defines the type (Harmsen et al., 2003; Koreen et al., 2004). Ridom SpaServer (http://spaserver.ridom.de) permits repeat assignation, harmonizing data from around the world and permitting reproducibility between laboratories. The BURP (Based Upon Repeat Pattern) software allows clonal relatedness of isolates to be predicted based on *spa* repeat patterns (Mellmann et al., 2007; Strommenger et al., 2006). *spa* typing has high discriminatory power and is cost-effective, however, risks misclassifying strains due to recombination and/or homoplasy (Nübel et al., 2008). As such, *spa* typing is best when combined with other typing methods.

4.4 SCC*mec* Typing

SCC*mec* typing schemes initially involved cloning and sequencing, however, in 2001, Ito et al. introduced a method that used long range PCR amplification with multiple sets of primers to identify the *mec* and *ccr* gene complexes (Hiramatsu et al., 1992; Ito et al., 1999, 2001). Today the most widely used methods for SCC*mec* typing involve multiplex PCR. In 2002, Oliveira et al. developed a multiplex assay targeting the *mecA* gene, as well as several other loci on SCC*mec* for quick identification of types, however, results were difficult to interpret and the assay was limited to known types (Oliveira and De Lencastre, 2002). Multiplex assays subsequently developed by Zhang et al. (Zhang et al., 2012; Zhang et al., 2005) and Milheirico et al. (Milheiriço et al., 2007) are currently the most commonly used assays, both being more simple to interpret, but are both limited to detection of type I−V. Unfortunately, there is currently no single PCR assay able to detect all SCC*mec* types and subtypes I-XIV. Targeted DNA microarray is another option for SCC*mec* typing, but only known types can be identified with the method, which also suffers from the need for trained personnel and specialized equipment (Monecke et al., 2008). In the future, development of a multiplex assay capable of identifying individual *ccr* and *mec* complex types would permit identification of both known and new combinations of the complexes, and thus identify all existing and most new SCC*mec* types.

4.5 Whole Genome Sequence (WGS) Typing

Next Generation Sequencing (NGS) is a cost-effective way to identify genetic variations on a genome-wide scale. WGS is the ultimate tool with high discriminatory power for studying the diversity within a

species. However, it does suffer in that significant computing and bioinformatics resources are required to be able to extract that information. Online tools, such as those from The Center for Genomic Epidemiology (Lyndgby, Denmark; available at https://www.genomicepidemiology.org), are available for extracting the MLST, *spa* and SCC*mec* type of a strain from genome sequences, as well as for identifying virulence and resistance determinants. WGS is also useful for phylogenetic analysis (Bouckaert et al., 2019; Stamatakis, 2014; Suchard et al., 2018), as well as for the identification of mobile genetic elements, such as pathogenicity islands and/or phages (Arndt et al., 2016; Zhou et al., 2011). In the future, WGS may be used for extended MLST (eMLST) in which a subset of, or all genes in the genome, would be included rather than only the seven housekeeping genes currently used. Variations include core genome MLST (cgMLST; including all core genes not subject to selection pressure), whole genome MLST (wgMLST; also including genes subject to selection pressure), ribosomal MLST (rMLST; including ribosomal genes), and the pan-genome approach (including all genes within the species). Given the huge amount of data that can be extracted, WGS will certainly become the primary tool for staphylococcal epidemiological and evolutionary analysis.

5. Molecular Dynamics in MRSA Evolution

Large-scale typing studies have indicated that, prior to the acquisition of SCC*mec*, *S. aureus* populations were diverse, but only contained a few epidemic lineages (Oliveira et al., 2012). Under antibiotic selective pressures, SCC*mec* integrated into select accommodating strains and spread worldwide through clonal expansion. Over time MRSA have evolved into phylogenetically distinct lineages, with most diseases caused by a small number of pandemic clones (Enright et al., 2002). In Fig. 28.2 whole genome sequencing illustrates these distinct pandemic lineages, while also showing a larger variety of MSSA clones. As mentioned, the major difference between MRSA and methicillin-susceptible *S. aureus* (MSSA) is acquisition of the SCC*mec* element, which analysis suggest may have occurred on at least 20 different occasions into different MSSA lineages (Robinson and Enright, 2003). This "multi-clone theory" of MRSA evolution proposes that SCC*mec* was introduced multiple times into different genetic lineages of *S. aureus* and is supported by several studies (Enright et al., 2002; Fitzgerald et al., 2001; Musser and Kapur, 1992; Robinson and Enright, 2003). Acquisition of SCC*mec* by MSSA was, in fact, found to be four times more common than replacement of one SCC*mec* by another within a strain (Robinson and Enright, 2003). DNA microarray technology has shown that the *mecA* gene was horizontally transferred at least five times into different genetic backgrounds of *S. aureus*, supporting the notion that MRSA evolved multiple independent times (Fitzgerald et al., 2001). An alternate theory, the single-clone theory, proposes that all MRSA have a mutual ancestor, with SCC*mec* introduced once into *S. aureus* followed by expansion and diversification, but has limited support (Kreiswirth et al., 1993).

Staphylococcal clonal complexes (CC) are defined using multilocus sequence data to group strains, and can be visualized with the eBURST algorithm (Feil et al., 2004). Strains within a clone differ by single nucleotide polymorphisms at fewer than three of the loci assessed, are generally isolated from more than one country, and share similar resistance profiles and SCC*mec* elements. The clonal complexes are made up of founder STs along with single or double locus variants of the founder, with a low degree of diversity observed within each lineage (Feil et al., 2003; Nübel et al., 2008). This low diversity may be affected by purifying selection, which has been described for the housekeeping genes

used for MLST typing, as well as for other core genes of the *S. aureus* chromosome (Kuhn et al., 2006). As mentioned above, human-associated *S. aureus* fall primarily into 10 clonal complexes (CC1, CC5, CC8, CC12, CC15, CC22, CC25, CC30, CC45, and CC151/121) (Enright et al., 2000, 2002; Feil et al., 2003, 2004; Feil and Enright, 2004; Feng et al., 2008; Goering et al., 2008; Hallin et al., 2007; Tenover et al., 2008). Global MRSA are more limited in their diversity, primarily belonging to five major clonal complexes (CC5, CC8 [including the CC8-ST239 subgroup], CC22, CC30, and CC45). However, new lineages such as ST59 and ST398 have emerged as epidemic clones (Enright et al., 2002). The five major MRSA lineage were shown to be derived from epidemic MSSA lineages present in the 1960s, suggesting that the selective factors responsible for superior epidemicity of a clone could also play a role in the adaptation to antibiotic selective pressures and, hence, acquisition of methicillin resistance (Gomes et al., 2006). The fact that these lineages are able to grow in frequency and cause epidemics highlights the importance of understanding the processes underlying new strain emergence.

Both mutation and recombination represent possible mechanisms of clonal divergence, essential for species diversification and adaptation to new ecological niches. Using MLST data to estimating the rate of recombination within clonal complexes, Feil et al. concluded that specific nucleotides and alleles were more prone to mutation, and single locus variants were 15 times more likely to differ from their ancestral strain due to point mutations than recombination (Feil et al., 2003). Twelve of the 13 single-locus variants in the study appeared to result from point mutation, while only one appeared to result from recombination. While recombination appears to have negligible impact on the core genome of staphylococci, recombination in *S. aureus* might occur more frequently between closely related strains than between phylogenetically distant ones (Kuhn et al., 2006). Waldron and Lindsay showed that the type I restriction modification system, SauI, blocks horizontal transfers between *S. aureus* strains of different CCs, possibly explaining why recombination is greater within a CC than between them (Waldron and Lindsay, 2006). When combined with selection mechanisms, recombination might favor the rapid creation of new clonal complexes with increased fitness (Basic-Hammer et al., 2010; Hughes and Friedman, 2005).

5.1 The Contribution of Mobile Genetic Elements to MRSA Evolution

S. aureus is generally considered a highly clonal species, with a conserved core genome that evolves primarily through mutation. Diversity in the species occurs mainly due to mobile genetic elements, acquired from other lineages or species through horizontal transfer. Up to 25% of the *S. aureus* genome is made up MGEs such as plasmids, transposons, pathogenicity islands, bacteriophages, and staphylococcal cassette chromosomes (Alibayov et al., 2014; Lindsay and Holden, 2004). These MGEs comprise the accessory genome, distinct from the core genome whose genes are present in 95% of strains (El Garch et al., 2009; Holden et al., 2004; Kuroda et al., 2001; Ohta et al., 2004). While the core genome carries genes associated with cell survival and metabolism, the accessory genome encodes proteins related to adaptation to environmental conditions, such as virulence factors and/or antibiotic resistance determinants. As such, the transfer of MGEs between strains can rapidly change the resistance, pathogenic potential, and epidemiologic spread of strains.

Staphylococcal plasmids are known to encode resistance to antibiotics as well as organic and inorganic ions, and toxin genes. Following acquisition of plasmid encoded penicillinase, penicillin resistant strains spread rapidly in hospitals during the 1940s, responsible for the first epidemic wave of

antibiotic resistant *S. aureus* (Chambers and DeLeo, 2009; Kirby, 1944). Transposable elements, such as insertion sequences (IS) and transposons (Tn) are also known to carry antibiotic resistance determinants (including penicillin resistance), as well as to alter chromosomal gene expression and possibly mediate molecular rearrangements of the genome (Alibayov et al., 2014; Lyon et al., 1987; Needham et al., 1995; Rouch et al., 1987). In addition to antibiotic resistance determinants, virulence factors can also be found on plasmids, pathogenicity islands, genomic islands, and prophages. Many adhesion, toxin, and superantigen genes are carried on these MGEs, modulating the way strains interact with their human or animal hosts, possibly affecting the success of a strain within a geographic location. Bacteriophages also represent a means by which *S. aureus* acquires virulence genes. Immune evasion cluster (IEC) genes such as staphylokinase (*sak*), chemotaxis inhibitory protein (*chp*), and staphylococcal inhibitor of complement (*scn*), modulate and resist the innate immune response, are commonly found on φSA3 phages. Panton-Valentine leukocidin (PVL), linked to the emergence and rapid spread of the USA300 lineage, is carried on φSA2 phages. Exfoliative toxin (*eta*), which can cause bullous impetigo and scalded-skin syndrome and was associated with outbreaks in Japan and the Czech Republic, is carried on φSA1 phages (Růžičková et al., 2012; Shi et al., 2011). While phage encoded virulence genes are expressed when the phage is in the lysogenic phase, their expression is upregulate when the phage is induced (Goerke et al., 2006; Sumby and Waldor, 2003). In this way, phages can positively or negatively impact virulence gene expression and influence the emergence of new virulent *S. aureus* lineages. Also described are helper phages, which can increase the mobility of staphylococcal pathogenicity islands, facilitating their transfer to other strains (Mir-Sanchis et al., 2012).

Studies have suggested that there is an association between carriage of some MGEs (with their associated virulence genes) and *S. aureus* linage (Booth et al., 2001; Moore and Lindsay, 2001; Peacock et al., 2002; Van Belkum et al., 2006). Peacock et al. demonstrated that the gene coding for toxic shock syndrome (*tst*), which is located on the pathogenicity island SaPI, was highly associated with clonal complexes CC30 and CC39, yet appeared infrequently in the other lineages tested (Peacock et al., 2002; Ruzin et al., 2001). Similarly, Holtfreter et al. showed that genes in the enterotoxin gene cluster (*egc*), located on genomic island vSAβ, were present in CC5, CC22, and CC45, as well as a subcluster of CC25, but were absent from the other CCs (Holtfreter et al., 2007). They also showed that *sec-sel* on SaPI3 were detected primarily in clonal complex CC45, while plasmid bound genes *sed-sej-ser* were mainly found in CC8. The varied distribution of some MGEs suggests that horizontal mobility is limited, by forces such as phage incompatibility or restriction modification systems, among others (Kuroda et al., 2001; Moller et al., 2019; Waldron and Lindsay, 2006). Other MGEs are more widely distributed among staphylococcal lineages. The bacteriophage φSA3, for example, is present in many human associated staphylococcal lineages such as CC59, CC8 (including the ST239 subgroup), CC398, CC22, CC25, CC30, CC45, and CC5, however, its structure varies between and within a lineage (Chen et al., 2013; Goerke et al., 2009; Kashif et al., 2019; McClure et al., 2018). Kashif et al. demonstrated that not only did the phage structure and virulence gene content vary between sublineages within the ST398 clone, but those differences were associated with varied virulence in a *Caenorhabditis elegans* infection model (Kashif et al., 2019). Another study, however, demonstrated that acquisition of φSA3 was not a major factor leading to the re-adaptation of LA-MRSA clone ST398 back to the human host, nor was it necessary for their ability to colonize and infect human hosts (Van Alen et al., 2018). As such, the biological significance of MGE variability, and its effect on the expansion and success of *S. aureus* lineages still needs further investigation.

6. Global MRSA Lineages

Through mutation, recombination, and the acquisition of mobile genetic elements, we see the constant emergence of new staphylococcal strains, often with rapid epidemiological spread. In each geographical location, however, only one or two MRSA lineages predominate at a time, and these successful lineages are seen to peak in prevalence before declining and being replaced by new strains (Wyllie et al., 2011). As such, the epidemiology of major MRSA lineages can be described as waves of expansion followed by population equilibrium and decline, and occurs on a global scale (Chambers and DeLeo, 2009). For example, in the mid-1980s Portuguese hospitals were dominated by the Portuguese clone (ST239-MRSA-III variant), but in 1992−1993 it was replaced by the Iberian clone (ST247-MRSA-IA), and eventually by the Brazilian clone (ST239-MRSA-III/IIIA) (Aires De Sousa and De Lencastre, 2004). German hospitals were dominated by the northern German clone (ST247-MRSA-I) and Hannover clones (ST254-MRSA-IV) between 1994 and 2002, but were replaced by the Berlin (ST45-MRSA-IV), southern German, Barnim (ST22-MRSA-IV), and Rhine-Hesse (ST5-MRSA-II) clones by the 2000s (Strommenger et al., 2004). In Asia, a CA-MRSA clone (ST254-MRSA-IV) was the major clone in the early 1990s, but in the late 1990s was replaced by ST239, and eventually ST5 and ST59 (Aires de Sousa et al., 2003; Chen et al., 2009). A CA-MRSA strain, USA300 (ST8-MRSA-IV) is currently the dominant strain in North America, initially recognized in the community, but increasingly isolated from hospital settings as well (Davis et al., 2007; Klevens et al., 2006). At the turn of the century, USA300 replaced another CA-MRSA strain, USA400 (ST1-MRSA-IV), and is replacing USA100 (ST5-MRSA-II), USA800 (ST5-MRSA-IV), and USA700 (ST72-MRSA-IV), which were the predominant nosocomial strains in the United States (Baba et al., 2002; Tenover et al., 2006, 2008). USA300, itself, is now declining in frequency in North America (Glaser et al., 2016). While the reasons why MRSA clones become dominant in an area and are subsequently replaced by other clones remains unclear, longitudinal studies have shown replacement of the predominant clone within a decade (Aires De Sousa and De Lencastre, 2004; Pérez-Roth et al., 2004).

As mentioned, the majority of global MRSA falls into one of five major clonal complexes (CC5, CC8, CC22, CC30, and CC45), however, among them CC5 and CC8 (including the CC8-ST239 subgroup) are the most prevalent ones encountered. CC5 and CC8 each contain various STs and occur in many regions of the world, while CC22 is likewise globally widespread (Chambers and DeLeo, 2009; Hsu et al., 2005; Monecke et al., 2011). CC30-ST36 is common in the United States and the United Kingdom, CC45 is common in the United States and Europe, and the CC8-ST239 subgroup is the most common clone reported in Asian countries (Hsu et al., 2005; Ko et al., 2005; Monecke et al., 2011). The distribution and predominant clones from each geographic location are listed in Table 28.1.

6.1 HA-MRSA

MRSA is one of the major causes of multidrug-resistant healthcare-associated infections. HA-MRSA strain is defined as being specifically successful in hospital settings, possibly being endemic in some locations and causing increased levels of infection. The ability to persist in the environment, to colonize the host, the repertoire of virulence determinants, and the high levels of drug resistance have all contributed to the success of HA-MRSA. Not all MRSA clones, however, have successfully established themselves in hospital environments, and the distribution of successful clones varies by

Table 28.1 Prevalent Hospital, Community, and Livestock-Associated MRSA Clones From Different Regions of the World

Continent/Region	HA-MRSA	CA-MRSA	LA-MRSA
Europe	ST36-MRSA-II (EMRSA-16,USA200, CMRSA4/8/9)	ST80-MRSA-IV (European clone)	ST398-MRSA-V
	ST5-MRSA-I (EMRSA-3)	ST22-MRSA-IV (EMRSA-15)	ST9
	ST239-MRSA-III (EMRSA-1/4/11, Vienna/Brazilian/ Hungarian clone)	ST1-MRSA-IV (USA400)	ST97-MRSA-V
	ST247-MRSA-I/II (EMRSA-5/7, Iberian clone)	ST30-MRSA-IV (Southwest Pacific clone)	ST1379-MRSA-V
	ST247-MRSA-I/II (EMRSA-5/7, Iberian clone)	ST59-MRSA-V	ST1-MRSA-IV (USA400)
	ST241-MRSA-III (UK−Finland clone)	ST5-MRSA-IV (Pediatric clone, USA800)	ST130-MRSA-XI
	ST8-MRSA-II	ST8-MRSA-IV (USA300)	
	ST105-MRSA-II	ST377-MRSA-V	
	ST228-MRSA-I	ST88-MRSA-IV (African clone)	
	ST5-MRSA-II (New York/Japan clone, USA100, CMRSA2)	ST93-MRSA-IV (Queensland clone)	
	ST239-MRSA-III variant (Portuguese clone)	ST772-MRSA-V	
	ST225-MRSA-II	ST45-MRSA-IV/V	
United States/Canada	ST5-MRSA-II (USA100, CMRSA2)	ST8-MRSA-IV (USA300, CMRSA10)	ST398-MRSA-V
	ST45-MRSA-II (USA600, CMRSA1)	ST1-MRSA-IV (USA400, CMRSA7)	ST5-MRSA-IV
	ST36-MRSA-II (USA200, CMRSA4/ 8/9)	ST5-MRSA-IV (Pediatric clone, USA800)	
	ST8-MRSA-II	ST22-MRSA-IV (EMRSA-15)	
	ST241-MRSA-II		
	ST239-MRSA-III		

Table 28.1 Prevalent Hospital, Community, and Livestock-Associated MRSA Clones From Different Regions of the World—cont'd

Continent/Region	HA-MRSA	CA-MRSA	LA-MRSA
Mexico/South America	ST5-MRSA-I (Cordobes/Chilean clone)	ST8-MRSA-IV (USA300 LA variant)	
	ST239-MRSA-III (Brazilian clone)	ST5-MRSA-IV (Pediatric clone, USA800)	
	ST5-MRSA-II (New York/Japan clone, USA100, CMRSA2)	ST30-MRSA-IVc (Southwest Pacific clone)	
		ST1-MRSA-IV	
Asia	ST239-MRSA-III	ST59-MRSA-IV/V	ST9-MRSA
	ST241-MRSA-III	ST338-MRSA-V_T	ST398-MRSA-V
	ST5-MRSA-II (New York/Japan clone, USA100)	ST910-MRSA-IVa	
	ST254-MRSA-I	ST1-MRSA-IV	
	ST30-MRSA-I	ST89-MRSA-IV	
	ST2590-MRSA-II	ST8-MRSA-IV (CA-MRSA/J, USA300)	
		ST91-MRSA-IV	
		ST30-MRSA-IV (Southwest Pacific clone)	
		ST5-MRSA-IV (New York/Japan clone)	
		ST72-MRSA-IVa (USA700)	
		ST88-MRSA-IV	
		ST80-MRSA-IV	
		ST834-MRSA-IV	
		ST121-MRSA-V	
		ST22-MRSA-IV	
		ST772-MRSA-V	
Australia/New Zealand	ST239-MRSA-III	ST93-MRSA-IV (Queensland clone)	
		ST30-MRSA_IV (Southwest Pacific clone)	

Continued

Table 28.1 Prevalent Hospital, Community, and Livestock-Associated MRSA Clones From Different Regions of the World—cont'd

Continent/Region	HA-MRSA	CA-MRSA	LA-MRSA
		ST8-MRSA-IV (USA300)	
		ST1-MRSA-IV (WA-MRSA-1)	
		ST5-MRSA-IV	
		ST45-MRSA-IV	
		ST78-MRSA-IV	
		ST22-MRSA-IV	
Africa	ST239-MRSA-III (Brazilian/Hungarian clone)	ST88-MRSA-IV (African clone)	ST398-MRSA-IV
	ST241-MRSA-III	ST80-MRSA-IV (European clone)	
	ST250-MRSA-I	ST30-MRSA-IV (Southwest Pacific clone)	
	ST5-MRSA-I	ST121-MRSA-V	
	ST36-MRSA-II (USA200, CMRSA4/8/9)	ST152-MRSA-V	
	ST5-MRSA-III (USA100, CMRSA2)	ST15-MRSA-IV	
	ST247-MRSA-I	ST8-MRSA-IV (USA300)	
	ST1819-MRSA-I	ST72-MRSA-V	
		ST789-MRSA-IV	
		ST2021-MRSA-V	
		ST612-MRSA-IV	
Middle East	ST239-MRSA-III	ST80-MRSA-IV (European clone)	
		ST22-MRSA-IVa	
		ST30-MRSA-IV (Southwest Pacific clone)	

region (Stefani et al., 2012). The incidence of HA-MRSA also varies by region, being higher in Eastern Asia (greater than 70%), and lowest in the Nordic region (Iceland, Denmark, Norway, Finland, and Sweden) and Netherlands (at less than 5%). There is evidence that some HA-MRSA clones evolved multiple times, like CC5 (ST5), for example, which SNP analysis suggests has evolved in several countries through acquisition of SCC*mec* in a local ST5 MSSA strain (Nübel et al., 2008). Other HA-MRSA clones appear to have dispersed internationally, like the CC8-ST239 subgroup, for which

phylogenetic analysis has shown dissemination through North and South America and Asia, as well as dispersion from South America to Europe and from Thailand to China (Gray et al., 2011; Harris et al., 2010). Analysis of HA-MRSA from Europe, the United States and South America has revealed that they fall into a limited number of pandemic clones, including the Iberian or UK-EMRSA5 (ST247-MRSA-IA), Brazilian (ST239-MRSA-IIIA), Hungarian (ST239-MRSA-III), New York/Japan (ST5-MRSA-II), pediatric (ST5-MRSA-IV), EMRSA-15 (ST22-MRSA-IV), EMRSA-16 (ST36-MRSA-II), and Berlin (ST45-MRSA-IV) clones (Enright et al., 2002; Nübel et al., 2008). These clones are believed to be highly transmissible and well adapted to hospital environments. However, molecular analysis has failed to provide a strong link between genotype and epidemic phenotype (Hetem et al., 2012; Kuhn et al., 2006, 2010).

6.2 CA-MRSA

Community-associated MRSA has become a major cause of endemic hospital-acquired infections, taking over from HA-MRSA lineages. Some mathematical models, in fact, predict the eventual replacement of conventional HA-MRSA strains with CA-MRSA (Boyce, 2008; D'Agata et al., 2009; Otter and French, 2006, 2011; Skov and Jensen, 2009; Tickler et al., 2017). Of greater concern is the exposure of CA-MRSA strains, usually limited to β-lactam resistance, to the antibiotic selective pressures found in hospitals. Decreased susceptibility to ciprofloxacin, vancomycin and daptomycin has already been detected in some CA-MRSA strains (Graber et al., 2007; Huang et al., 2006).

CA-MRSA were initially represented by the USA400 (ST1-MRSA-IVa) lineage, then replaced by USA300 (ST8-MRSA-IVa), which has grown to become a dominant linage in many parts of the world. What factors are responsible for the global emergence, success and rapid spread of CA-MRSA lineages still need to be clarified. Several mobile genetic elements have, however, been linked. CA-MRSA generally carry the smaller SCC*mec* types IV or V, believed to impart a smaller fitness cost in community settings where antibiotics are not present (Chambers and DeLeo, 2009). The smaller SCC*mec* types are also believed to transfer more efficiently to recipient clones, resulting in the emergence of unique-fit MRSA clones (Berglund and Söderquist, 2008; Ma et al., 2002). CA-MRSA are also frequently found to carry the phage borne PVL toxin genes (*lukS-PV* and *lukF-PV*), which causes tissue necrosis and leukocyte destruction (Panton and Valentine, 1932; Vandenesch et al., 2003). Vandenesch et al. noted a strong association between CA-MRSA strains and PVL, and suggested that acquisition of PVL may have played a role in the origin and spread of CA-MRSA (Vandenesch et al., 2003). In support of this, plasmid encoded PVL was shown to cause necrotizing pneumonia in BALB/c mice (Labandeira-Rey et al., 2007). Another study found that in a rabbit model of CA-MRSA bacteremia, PVL had a small and transient role in the acute phase of infection (Diep et al., 2008). Other groups, however, have challenged the role of PVL in CA-MRSA infection. In one study, USA300 and USA400 PVL positive and negative strains were found to cause similar skin disease and have similar virulence and toxicity (Voyich et al., 2006). In a retrospective study of CA-MRSA in a large Canadian healthcare region, Zhang et al. identified coexisting PVL positive and negative USA400 strains, noting that the strains showed similar clinical characteristics, and concluding that genetic background of the strain played a greater role in their success (Zhang et al., 2008). They went on to show that PVL positive and negative sister strains of USA400 showed no significant differences in virulence and cytotoxicity when tested in various in vitro cell culture an animal models (Wu et al., 2008). As such, the role of PVL in the emergence and worldwide

dissemination of CA-MRSA remains unclear. The Arginine Catabolic Mobile Element (ACME), carrying genes for the arginine deiminase pathway and oligopeptide permease system, is another factor hypothesized to enhance USA300's success. It is associated with resistance to clearance by the host and is associates with increased fitness in a rabbit bacteremia model (Diep et al., 2008; Thurlow et al., 2013). Up to 15% of USA300, however, do not carry ACME (Uhlemann et al., 2011), and a study by Wu et al. showed that both USA300 wild type and ACME deletion mutants had similar in vitro and in vivo survival and virulence, suggesting that ACME alone is not a significant force driving the success of USA300 (Wu et al., 2020). Mobile genetic elements are also implicated in the evolution and success of the CA-MRSA ST59 lineage, which is the most commonly isolated sequence type in Eastern Asia. SCC*mec*, PVL, IEC, and the mobile element structure (MES) are believed to have influenced expansion and success of the lineage in Asia, and driven diversification of North American and East Asian sublineages (Chen et al., 2013; Huang et al., 2008; Hung et al., 2012; McClure et al., 2021a; Pang et al., 2020).

It is unlikely that any single factor is responsible for the emergence and expansion of CA-MRSA lineages, but rather that multiple factors are implicated, influenced by local selective pressures.

6.3 LA-MRSA

LA-MRSA is most commonly represented by isolates belonging to CC398 (ST398), however, other notable LA-MRSA lineages include CC9 (Chuang and Huang, 2015; Feβler et al., 2011; Meemken et al., 2010) and CC97 (Luini et al., 2015; Luzzago et al., 2014; Wang et al., 2015). On top of these, human-associated lineages such as (but not limited to) CC1 (Luini et al., 2015; Normanno et al., 2015), CC8-ST239 (Nemeghaire et al., 2013, 2014), CC5 (Buyukcangaz et al., 2013; Molla et al., 2012), CC30, and CC45 are also known to infect livestock (Agersø et al., 2012; Luzzago et al., 2014; Nemeghaire et al., 2014). While LA-MRSA were initially identified in livestock, they were later associated with humans in close contact with animals, then isolated from humans lacking any livestock exposure. LA-MRSA can also spread to humans via the environment, through manure or airborne transmission (Schulz et al., 2012; Vanderhaeghen et al., 2010), and via contaminated meat and meat products (van Rijen et al., 2013).

The evolution of LA-MRSA appears to be intricately tied to the gain and loss of mobile genetic elements. Despite first being identified in animals, WGS and detailed phylogenetic analysis has shown that LA-MRSA ST398 most likely originated from a human MSSA strain, which acquired methicillin and tetracycline resistance to favor survival amidst the large amount of antibiotics used in animal husbandry (Price et al., 2012). The jump to animal host was also accompanied by loss of φSA3, and the human innate immune modulators carried within. Additionally, a highly virulent human-adapted CA-MRSA ST398 was shown to be accompanied by acquisition of SCC*mec*, without the gain of virulence markers typically associated with CA-MRSA (He et al., 2018). Similarly, carriage of multiple prophages by ST398 has been shown to expand its virulence and increase its adaptation to humans (Diene et al., 2017). Other lineages of LA-MRSA have MGEs influencing their evolution as well, such as MRSA ST5, for which phylogenetic analysis shows a human origin with the hop to poultry accompanied by gain of avian-specific MGEs (Lowder et al., 2009). Likewise, in CC1 no animal-associated isolates are found to carry PVL, while it is common in human associated strains (Luini et al., 2015; Normanno et al., 2015). In contrast to the human to animal jump proposed for other LA-MRSA linages, human MRSA CC97 are believed to have originated from an MSSA CC97 that

jumped from livestock to humans approximately 40 years ago, followed by acquisition of SCC*mec* (Spoor et al., 2013).

The continued use of antimicrobials in livestock animals (often in food as a growth stimulant) will continue to influence the transfer of mobile genetic elements between these *S. aureus* strains, increasing resistance and favoring adaptation to their animal or human hosts.

7. Concluding Remarks

S. aureus genomes have evolved over time, predominantly due to mutations, small genomic rearrangements, and the acquisition/loss of mobile genetic elements. The species is extraordinary in its ability to develop antibiotic resistance, which has played a pivotal role in driving its evolution. Initially only encountered in hospitals, MRSA has expanded to occupy community and animal niches, with new lineages continually appearing, often becoming epidemic, then being replaced. Understanding the molecular forces that direct *S. aureus* evolution is important for tracking and controlling the spread of this superbug. However, the continued overuse of antibiotics pushes the species to acquire greater and greater resistance, and threatens to push us to the point where no viable treatments remain. As such MRSA will remain an important area of research, with a focus on nonantibiotic approaches to prevention and treatment needed to stay ahead of the threat. Whole genome sequencing will greatly expand our ability to understand the course of epidemics, and permit the identification of bacterial determinants subject to diversifying selection, which can be targeted for treatment.

References

Adcock, P.M., et al., 1998. Methicillin-resistant *Staphylococcus aureus* in two child care centers. J. Infect. Dis. 178 (2), 577−580. https://doi.org/10.1086/517478.

Agersø, Y., et al., 2012. Study of methicillin resistant *Staphylococcus aureus* (MRSA) in Danish pigs at slaughter and in imported retail meat reveals a novel MRSA type in slaughter pigs. Vet. Microbiol. 157 (1−2), 246−250. https://doi.org/10.1016/j.vetmic.2011.12.023.

Aiello, A.E., et al., 2006. Methicillin-resistant *Staphylococcus aureus* among US prisoners and military personnel: review and recommendations for future studies. Lancet Infect. Dis. 6 (6), 335−341. https://doi.org/10.1016/S1473-3099(06)70491-1.

Aires De Sousa, M., De Lencastre, H., 2004. Bridges from hospitals to the laboratory: genetic portraits of methicillin-resistant *Staphylococcus aureus* clones. FEMS Immunol. Med. Microbiol. 40 (2), 101−111. https://doi.org/10.1016/S0928-8244(03)00370-5.

Aires de Sousa, M., et al., 2003. Frequent recovery of a single clonal type of multidrug-resistant *Staphylococcus aureus* from patients in two hospitals in Taiwan and China. J. Clin. Microbiol. 41 (1), 159−163. https://doi.org/10.1128/JCM.41.1.159-163.2003.

Alibayov, B., et al., 2014. *Staphylococcus aureus* mobile genetic elements. Mol. Biol. Rep. 41 (8), 5005−5018. https://doi.org/10.1007/s11033-014-3367-3.

Aligholi, M., et al., 2008. Emergence of high-level vancomycin-resistant Staphylococcus aureus in the Imam Khomeini Hospital in Tehran. Med. Princ. Pract. 17 (5), 432−434. https://doi.org/10.1159/000141513.

Antonanzas, F., Lozano, C., Torres, C., 2015. Economic features of antibiotic resistance: the case of methicillin-resistant *Staphylococcus aureus*. Pharmacoeconomics 33 (4), 285−325. https://doi.org/10.1007/s40273-014-0242-y.

Argudín, M.A., Rodicio, M.R., Guerra, B., 2010. The emerging methicillin-resistant *Staphylococcus aureus* ST398 clone can easily be typed using the Cfr9I SmaI-neoschizomer. Lett. Appl. Microbiol. 50 (1), 127−130. https://doi.org/10.1111/j.1472-765x.2009.02742.x.

Armand-Lefevre, L., Ruimy, R., Andremont, A., 2005. Clonal comparison of *Staphylococcus aureus* isolates from healthy pig farmers, human controls, and pigs. Emerg. Infect. Dis. 11 (5), 711−714. https://doi.org/10.3201/eid1105.040866.

Arndt, D., et al., 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44 (1), W16−W21. https://doi.org/10.1093/nar/gkw387.

Arthur, M., et al., 1993. Characterization of Tn1546, a Tn3-related transposon conferring glycopeptide resistance by synthesis of depsipeptide peptidoglycan precursors in *Enterococcus faecium* BM4147. J. Bacteriol. 175 (1), 117−127. https://doi.org/10.1128/jb.175.1.117-127.1993.

Azhar, A., et al., 2017. Detection of high levels of resistance to linezolid and vancomycin in *Staphylococcus aureus*. J. Med. Microbiol. 66 (9), 1328−1331. https://doi.org/10.1099/jmm.0.000566.

Azimian, A., et al., 2012. Genetic characterization of a vancomycin-resistant *Staphylococcus aureus* isolate from the respiratory tract of a patient in a University Hospital in Northeastern Iran. J. Clin. Microbiol. 50 (11), 3581−3585. https://doi.org/10.1128/JCM.01727-12.

Baba, T., et al., 2002. Genome and virulence determinants of high virulence community-acquired MRSA. Lancet 359 (9320), 1819−1827. https://doi.org/10.1016/s0140-6736(02)08713-5.

Baba, T., et al., 2009. Complete genome sequence of *Macrococcus caseolyticus* strain JSCS5402, reflecting the ancestral genome of the human-pathogenic staphylococci. J. Bacteriol. 191 (4), 1180−1190. https://doi.org/10.1128/jb.01058-08.

Baig, S., et al., 2018. Novel SCC*mec* type XIII (9A) identified in an ST152 methicillin-resistant *Staphylococcus aureus*. Infect. Genet. Evol. 61, 74−76. https://doi.org/10.1016/j.meegid.2018.03.013.

Bannerman, T.L., et al., 1995. Pulsed-field gel electrophoresis as a replacement for bacteriophage typing of *Staphylococcus aureus*. J. Clin. Microbiol. 33 (3), 551−555. https://doi.org/10.1128/jcm.33.3.551-555.1995.

Basic-Hammer, N., et al., 2010. Impact of recombination on genetic variability within *Staphylococcus aureus* clonal complexes. Infect. Genet. Evol. 10 (7), 1117−1123. https://doi.org/10.1016/j.meegid.2010.07.013.

Becker, K., et al., 2018. Plasmid-encoded transferable *mecB*-mediated methicillin resistance in *Staphylococcus aureus*. Emerg. Infect. Dis. 24 (2), 242−248. https://doi.org/10.3201/eid2402.171074.

Begier, E.M., et al., 2004. A high-morbidity outbreak of methicillin-resistant *Staphylococcus aureus* among players on a college football team, facilitated by cosmetic body shaving and turf burns. Clin. Infect. Dis. 39 (10), 1446−1453. https://doi.org/10.1086/425313.

Benito, D., et al., 2014. Characterization of tetracycline and methicillin resistant *Staphylococcus aureus* strains in a Spanish hospital: is livestock-contact a risk factor in infections caused by MRSA CC398? Int. J. Med. Microbiol. 304 (8), 1226−1232. https://doi.org/10.1016/j.ijmm.2014.09.004.

Berglund, C., Söderquist, B., 2008. The origin of a methicillin-resistant *Staphylococcus aureus* isolate at a neonatal ward in Sweden—possible horizontal transfer of a staphylococcal cassette chromosome *mec* between methicillin-resistant *Staphylococcus haemolyticus* and *Staphylococcus aureus*. Clin. Microbiol. Infect. 14 (11), 1048−1056. https://doi.org/10.1111/j.1469-0691.2008.02090.x.

Bondi, A., Dietz, C.C., 1945. Penicillin resistant staphylococci. Exp. Biol. Med. 60 (1), 55−58. https://doi.org/10.3181/00379727-60-15089.

Booth, M.C., et al., 2001. Clonal associations among *Staphylococcus aureus* isolates from various sites of infection. Infect. Immun. 69 (1), 345−352. https://doi.org/10.1128/IAI.69.1.345-352.2001.

Bouckaert, R., et al., 2019. Beast 2.5: an advanced software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 15 (4). https://doi.org/10.1371/journal.pcbi.1006650.

Boyce, J.M., 2008. Community-associated methicillin-resistant *Staphylococcus aureus* as a cause of health care-associated infection. Clin. Infect. Dis. 46 (6), 795−798. https://doi.org/10.1086/528717.

Boyle-Vavra, S., Daum, R.S., 2007. Community-acquired methicillin-resistant *Staphylococcus aureus*: the role of Panton-Valentine leukocidin. Lab. Invest. 87 (1), 3–9. https://doi.org/10.1038/labinvest.3700501.

Boyle-Vavra, S., et al., 2005. Successful multiresistant community-associated methicillin-resistant *Staphylococcus aureus* lineage from Taipei, Taiwan, that carries either the novel staphylococcal chromosome cassette *mec* (SCC*mec*) type VT or SCC*mec* type IV. J. Clin. Microbiol. 43 (9), 4719–4730. https://doi.org/10.1128/JCM.43.9.4719-4730.2005.

Bradley, J.M., et al., 1985. Methicillin-resistant *Staphylococcus aureus* in a London hospital. Lancet 325 (8444), 1493–1495. https://doi.org/10.1016/S0140-6736(85)92263-9.

Bratu, S., et al., 2005. Community-associated methicillin-resistant *Staphylococcus aureus* in hospital nursery and maternity units. Emerg. Infect. Dis. 11 (6), 808–813. https://doi.org/10.3201/eid1106.040885.

Brunel, A.S., et al., 2014. Methicillin-sensitive *Staphylococcus aureus* CC398 in intensive care unit, France. Emerg. Infect. Dis. 20 (9), 1511–1515. https://doi.org/10.3201/eid2009.130225.

Bugg, T.D.H., et al., 1991. Molecular basis for vancomycin resistance in *Enterococcus faecium* BM4147: biosynthesis of a depsipeptide peptidoglycan precursor by vancomycin resistance proteins VanH and VanA. Biochemistry 30 (43), 10408–10415. https://doi.org/10.1021/bi00107a007.

Buyukcangaz, E., et al., 2013. Molecular typing of *Staphylococcus aureus* and methicillin-resistant *S. aureus* (MRSA) isolated from animals and retail meat in North Dakota, United States. Foodborne Pathog. Dis. 10 (7), 608–617. https://doi.org/10.1089/fpd.2012.1427.

Bynoe, E.T., Elder, R.H., Comtois, R.D., 1956. Phage-typing and antibiotic-resistance of staphylococci isolated in a general hospital. Can. J. Microbiol. 2 (3), 346–358. https://doi.org/10.1139/m56-041.

Cafferkey, M.T., et al., 1985. Methicillin-resistant *Staphylococcus aureus* in Dublin 1971-84. Lancet 326 (8457), 705–708. https://doi.org/10.1016/S0140-6736(85)92942-3.

Centers for Disease Control and Prevention, 1999. Four pediatric deaths from community-acquired methicillin-resistant *Staphylococcus aureus* - Minnesota and North Dakota, 1997-1999. MMWR Morb. Mortal. Wkly. Rep. 48 (32), 707–710.

Centers for Disease Control and Prevention, 2002. *Staphylococcus aureus* resistant to vancomycin–United States. MMWR Morb. Mortal. Wkly. Rep. 51, 565–567.

Chambers, H.F., DeLeo, F.R., 2009. Waves of resistance: *Staphylococcus aureus* in the antibiotic era. Nat. Rev. Microbiol. 7 (9), 629–641. https://doi.org/10.1038/nrmicro2200.

Chang, S., et al., 2003. Infection with vancomycin-resistant *Staphylococcus aureus* containing the *vanA* resistance gene. N. Engl. J. Med. 348 (14), 1342–1347. https://doi.org/10.1056/NEJMoa025025.

Chen, C.J., et al., 2009. Change in the molecular epidemiology of methicillin-resistant *Staphylococcus aureus* bloodstream infections in Taiwan. Diagn. Microbiol. Infect. Dis. 65 (2), 199–201. https://doi.org/10.1016/j.diagmicrobio.2009.05.020.

Chen, C.J., et al., 2013. Characterization and comparison of 2 distinct epidemic community-associated methicillin-resistant *Staphylococcus aureus* clones of ST59 lineage. PLoS One 8 (9). https://doi.org/10.1371/journal.pone.0063210.

Chen, X.P., et al., 2017. Extreme diversity and multiple SCC*mec* elements in coagulase-negative *Staphylococcus* found in the clinic and community in Beijing, China. Ann. Clin. Microbiol. Antimicrob. 16 (1). https://doi.org/10.1186/s12941-017-0231-z.

Chuang, Y.Y., Huang, Y.C., 2015. Livestock-associated methicillin-resistant *Staphylococcus aureus* in Asia: an emerging issue? Int. J. Antimicrob. Agents 45 (4), 334–340. https://doi.org/10.1016/j.ijantimicag.2014.12.007.

Como-Sabetti, K., et al., 2009. Community-associated methicillin-resistant *Staphylococcus aureus*: trends in case and isolate characteristics from six years of prospective surveillance. Publ. Health Rep. 124 (3), 427–435. https://doi.org/10.1177/003335490912400312.

Cookson, B.D., et al., 2007. Evaluation of molecular typing methods in characterizing a European collection of epidemic methicillin-resistant *Staphylococcus aureus* strains: the HARMONY collection. J. Clin. Microbiol. 45 (6), 1830−1837. https://doi.org/10.1128/JCM.02402-06.

Crisóstomo, M.I., et al., 2001. The evolution of methicillin resistance in *Staphylococcus aureus*: similarity of genetic backgrounds in historically early methicillin-susceptible and -resistant isolates and contemporary epidemic clones. Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9865−9870. https://doi.org/10.1073/pnas.161272898.

Cuny, C., et al., 2010. Emergence of methicillin-resistant *Staphylococcus aureus* (MRSA) in different animal species. Int. J. Med. Microbiol. 300 (2−3), 109−117. https://doi.org/10.1016/j.ijmm.2009.11.002.

David, M.Z., Daum, R.S., 2010. Community-associated methicillin-resistant *Staphylococcus aureus*: epidemiology and clinical consequences of an emerging epidemic. Clin. Microbiol. Rev. 23 (3), 616−687. https://doi.org/10.1128/CMR.00081-09.

David, M.Z., et al., 2008. What is community-associated methicillin-resistant *Staphylococcus aureus*? JID J. Infect. Dis. 197 (9), 1235−1243. https://doi.org/10.1086/533502.

Davis, S.L., et al., 2007. Epidemiology and outcomes of community-associated methicillin-resistant *Staphylococcus aureus* infection. J. Clin. Microbiol. 45 (6), 1705−1711. https://doi.org/10.1128/jcm.02311-06.

Dezfulian, A., et al., 2012. Identification and characterization of a high vancomycin-resistant *Staphylococcus aureus* harboring VanA gene cluster isolated from diabetic foot ulcer. Iran. J. Basic Med. Sci. 15 (2), 803−806.

Diekema, D.J., et al., 2001. Survey of infections due to *Staphylococcus* species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific region for the SENTRY Antimicrobial Surveillance Program, 1997−1999. Clin. Infect. Dis. 32 (s2), S114−S132. https://doi.org/10.1086/320184.

Diene, S.M., et al., 2017. Prophages and adaptation of *Staphylococcus aureus* ST398 to the human clinic. BMC Genom. 18 (1). https://doi.org/10.1186/s12864-017-3516-x.

Diep, B.A., Palazzolo-Ballance, A.M., et al., 2008a. Contribution of Panton-Valentine leukocidin in community-associated methicillin-resistant *Staphylococcus aureus* pathogenesis. PLoS One 3 (9). https://doi.org/10.1371/journal.pone.0003198.

Diep, B.A., Stone, G.G., et al., 2008b. The arginine catabolic mobile element and staphylococcal chromosomal cassette *mec* linkage: convergence of virulence and resistance in the USA300 clone of methicillin-resistant *Staphylococcus aureus*. JID J. Infect. Dis. 197 (11), 1523−1530. https://doi.org/10.1086/587907.

D'Agata, E.M.C., et al., 2009. Modeling the invasion of community-acquired methicillin-resistant *Staphylococcus aureus* into hospitals. Clin. Infect. Dis. 48 (3), 274−284. https://doi.org/10.1086/595844.

El Garch, F., et al., 2009. StaphVar-DNA microarray analysis of accessory genome elements of community-acquired methicillin-resistant *Staphylococcus aureus*. J. Antimicrob. Chemother. 63 (5), 877−885. https://doi.org/10.1093/jac/dkp089.

Enright, M.C., et al., 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J. Clin. Microbiol. 38 (3), 1008−1015. https://doi.org/10.1128/jcm.38.3.1008-1015.2000.

Enright, M.C., et al., 2002. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc. Natl. Acad. Sci. U.S.A. 99 (11), 7687−7692. https://doi.org/10.1073/pnas.122108599.

Faoagali, J.L., Thong, M.L., Grant, D., 1992. Ten years' experience with methicillin-resistant *Staphylococcus aureus* in a large Australian hospital. J. Hosp. Infect. 20 (2), 113−119. https://doi.org/10.1016/0195-6701(92)90113-z.

Feil, E., Enright, M., 2004. Analyses of clonality and the evolution of bacterial pathogens. Curr. Opin. Microbiol. 7 (3), 308−313. https://doi.org/10.1016/j.mib.2004.04.002.

Feil, E.J., et al., 2003. How clonal is *Staphylococcus aureus*? J. Bacteriol. 185 (11), 3307−3316. https://doi.org/10.1128/JB.185.11.3307-3316.2003.

Feil, E.J., et al., 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 186 (5), 1518–1530. https://doi.org/10.1128/JB.186.5.1518-1530.2004.

Feng, Y., et al., 2008. Evolution and pathogenesis of *Staphylococcus aureus:* lessons learned from genotyping and comparative genomics. FEMS Fed. Eur. Microbiol. Soc. Microbiol. Rev. 32 (1), 23–37. https://doi.org/10.1111/j.1574-6976.2007.00086.x.

Feßler, A.T., et al., 2011. Characterization of methicillin-resistant *Staphylococcus aureus* isolates from food and food products of poultry origin in Germany. Appl. Environ. Microbiol. 77 (20), 7151–7157. https://doi.org/10.1128/aem.00561-11.

Fitzgerald, J.R., et al., 2001. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc. Natl. Acad. Sci. U.S.A. 98 (15), 8821–8826. https://doi.org/10.1073/pnas.161098098.

Foster, G., Paterson, G.K., 2020. Methicillin-resistant *Macrococcus bohemicus* encoding a divergent SCC*mecB* element. Antibiotics 9 (9). https://doi.org/10.3390/antibiotics9090590.

Frénay, H.M.E., et al., 1996. Molecular typing of methicillin-resistant *Staphylococcus aureus* on the basis of protein a gene polymorphism. Eur. J. Clin. Microbiol. Infect. Dis. 15 (1), 60–64. https://doi.org/10.1007/BF01586186.

García-Álvarez, L., et al., 2011. Methicillin-resistant *Staphylococcus aureus* with a novel *mecA* homologue in human and bovine populations in the UK and Denmark: a descriptive study. Lancet Infect. Dis. 11 (8), 595–603. https://doi.org/10.1016/S1473-3099(11)70126-8.

Garner, M.J., et al., 2015. An assessment of antimicrobial resistant disease threats in Canada. PLoS One 10 (4). https://doi.org/10.1371/journal.pone.0125155.

Glaser, P., et al., 2016. Demography and intercontinental spread of the USA300 community-acquired methicillin-resistant *Staphylococcus aureus* lineage. mBio 7 (1). https://doi.org/10.1128/mBio.02183-15.

Goering, R.V., 2010. Pulsed field gel electrophoresis: a review of application and interpretation in the molecular epidemiology of infectious disease. Infect. Genet. Evol. 10 (7), 866–875. https://doi.org/10.1016/j.meegid.2010.07.023.

Goering, R.V., et al., 2008. Molecular epidemiology of methicillin-resistant and methicillin-susceptible *Staphylococcus aureus* isolates from global clinical trials. J. Clin. Microbiol. 46 (9), 2842–2847. https://doi.org/10.1128/JCM.00521-08.

Goerke, C., Koller, J., Wolz, C., 2006. Ciprofloxacin and trimethoprim cause phage induction and virulence modulation in *Staphylococcus aureus*. Antimicrob. Agents Chemother. 50 (1), 171–177. https://doi.org/10.1128/AAC.50.1.171-177.2006.

Goerke, C., et al., 2009. Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. J. Bacteriol. 191 (11), 3462–3468. https://doi.org/10.1128/JB.01804-08.

Gomes, A.R., Westh, H., de Lencastre, H., 2006. Origins and evolution of methicillin-resistant *Staphylococcus aureus* clonal lineages. Antimicrob. Agents Chemother. 50 (10), 3237–3244. https://doi.org/10.1128/aac.00521-06.

Gómez-Sanz, E., et al., 2015. First staphylococcal cassette chromosome *mec* containing a *mecB* -carrying gene complex independent of transposon Tn 6045 in a *Macrococcus caseolyticus* isolate from a canine infection. Antimicrob. Agents Chemother. 59 (8), 4577–4583. https://doi.org/10.1128/aac.05064-14.

Gorwitz, R.J., et al., 2008. Changes in the prevalence of nasal colonization with *Staphylococcus aureus* in the United States, 2001-2004. JID J. Infect. Dis. 197 (9), 1226–1234. https://doi.org/10.1086/533494.

Graber, C.J., et al., 2007. Intermediate vancomycin susceptibility in a community-associated MRSA clone. Emerg. Infect. Dis. 13 (3), 491–493. https://doi.org/10.3201/eid1303.060960.

Graveland, H., et al., 2011. Livestock-associated methicillin-resistant *Staphylococcus aureus* in animals and humans. Int. J. Med. Microbiol. 301 (8), 630–634. https://doi.org/10.1016/j.ijmm.2011.09.004.

Gray, R.R., et al., 2011. Testing spatiotemporal hypothesis of bacterial evolution using methicillin-resistant *Staphylococcus aureus* ST239 genome-wide data within a Bayesian framework. Mol. Biol. Evol. 28 (5), 1593−1603. https://doi.org/10.1093/molbev/msq319.

Grisold, A.J., et al., 2010. Occurrence and genotyping using automated repetitive-sequence-based PCR of methicillin-resistant *Staphylococcus aureus* ST398 in Southeast Austria. Diagn. Microbiol. Infect. Dis. 66 (2), 217−221. https://doi.org/10.1016/j.diagmicrobio.2009.09.006.

Hallin, M., et al., 2007. Genetic relatedness between methicillin-susceptible and methicillin-resistant *Staphylococcus aureus*: results of a national survey. J. Antimicrob. Chemother. 59 (3), 465−472. https://doi.org/10.1093/jac/dkl535.

Hanifah, Y.A., Hiramatsu, K., Yokota, T., 1992. Characterization of methicillin-resistant *Staphylococcus aureus* associated with nosocomial infections in the University Hospital, Kuala Lumpur. J. Hosp. Infect. 21 (1), 15−28. https://doi.org/10.1016/0195-6701(92)90150-k.

Hanssen, A.-M., Sollid, J.U.E., 2007. Multiple staphylococcal cassette chromosomes and allelic variants of cassette chromosome recombinases in *Staphylococcus aureus* and coagulase-negative staphylococci from Norway. Antimicrob. Agents Chemother. 51 (5), 1671−1677. https://doi.org/10.1128/aac.00978-06.

Harmsen, D., et al., 2003. Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for *spa* repeat determination and database management. J. Clin. Microbiol. 41 (12), 5442−5448. https://doi.org/10.1128/JCM.41.12.5442-5448.2003.

Harris, S.R., et al., 2010. Evolution of MRSA during hospital transmission and intercontinental spread. Science 327 (5964), 469−474. https://doi.org/10.1126/science.1182395.

Harrison, E.M., et al., 2013. A *Staphylococcus xylosus* isolate with a new *mecC* allotype. Antimicrob. Agents Chemother. 57 (3), 1524−1528. https://doi.org/10.1128/AAC.01882-12.

Hartman, B.J., Tomasz, A., 1984. Low-affinity penicillin-binding protein associated with beta-lactam resistance in *Staphylococcus aureus*. J. Bacteriol. 158 (2), 513−516. https://doi.org/10.1128/jb.158.2.513-516.1984.

He, L., et al., 2018. Detection and analysis of methicillin-resistant human-adapted sequence type 398 allows insight into community-associated methicillin-resistant *Staphylococcus aureus* evolution. Genome Med. 10 (1). https://doi.org/10.1186/s13073-018-0514-9.

Herold, B.C., et al., 1998. Community-acquired methicillin-resistant *Staphylococcus aureus* in children with no identified predisposing risk. J. Am. Med. Assoc. 279 (8), 593−598. https://doi.org/10.1001/jama.279.8.593.

Hetem, D.J., et al., 2012. Nosocomial transmission of community-associated methicillin-resistant *Staphylococcus aureus* in Danish hospitals. J. Antimicrob. Chemother. 67 (7), 1775−1780. https://doi.org/10.1093/jac/dks125.

Hiramatsu, K., et al., 1992. Molecular cloning and nucleotide sequence determination of the regulator region of *mecA* gene in methicillin-resistant *Staphylococcus aureus* (MRSA). FEBS Fed. Eur. Biochem. Soc. Lett. 298 (2−3), 133−136. https://doi.org/10.1016/0014-5793(92)80039-j.

Hiramatsu, K., et al., 1997. Methicillin-resistant *Staphylococcus aureus* clinical strain with reduced vancomycin susceptibility. J. Antimicrob. Chemother. 40 (1), 135−136. https://doi.org/10.1093/jac/40.1.135.

Holden, M.T.G., et al., 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the evolution of virulence and drug resistance. Proc. Natl. Acad. Sci. U.S.A. 101 (26), 9786−9791. https://doi.org/10.1073/pnas.0402521101.

Holtfreter, S., et al., 2007. Clonal distribution of superantigen genes in clinical *Staphylococcus aureus* isolates. J. Clin. Microbiol. 45 (8), 2669−2680. https://doi.org/10.1128/JCM.00204-07.

Hsu, L.Y., et al., 2005. Dissemination of multi-susceptible methicillin-resistant *Staphylococcus aureus* in Singapore. J. Clin. Microbiol. 43 (6), 2923−2925. https://doi.org/10.1128/JCM.43.6.2923-2925.2005.

Huang, H., et al., 2006. Comparisons of community-associated methicillin-resistant *Staphylococcus aureus* (MRSA) and hospital-associated MSRA infections in Sacramento, California. J. Clin. Microbiol. 44 (7), 2423−2427. https://doi.org/10.1128/JCM.00254-06.

Huang, Y.-C., et al., 2008. Comparative molecular analysis of community-associated and healthcare-associated methicillin-resistant *Staphylococcus aureus* isolates from children in northern Taiwan. Clin. Microbiol. Infect. 14 (12), 1167−1172. https://doi.org/10.1111/j.1469-0691.2008.02115.x.

Hughes, A.L., Friedman, R., 2005. Nucleotide substitution and recombination at orthologous loci in *Staphylococcus aureus*. J. Bacteriol. 187 (8), 2698−2704. https://doi.org/10.1128/JB.187.8.2698-2704.2005.

Huijsdens, X.W., et al., 2006. Community-acquired MRSA and pig-farming. Ann. Clin. Microbiol. Antimicrob. 5. https://doi.org/10.1186/1476-0711-5-26.

Hung, W.C., et al., 2012. Comparative genomics of community-acquired ST59 methicillin-resistant *Staphylococcus aureus* in Taiwan: novel mobile resistance structures with IS1216V. PLoS One 7 (10). https://doi.org/10.1371/journal.pone.0046987.

Ichiyama, S., et al., 1991. Genomic DNA fingerprinting by pulsed-field gel electrophoresis as an epidemiological marker for study of nosocomial infections caused by methicillin-resistant *Staphylococcus aureus*. J. Clin. Microbiol. 29 (12), 2690−2695. https://doi.org/10.1128/jcm.29.12.2690-2695.1991.

International Working Group on the Classification of Staphylococcal Cassette Chromosome, Elements, 2009. Classification of staphylococcal cassette chromosome *mec* (SCC*mec*): guidelines for reporting novel SCC*mec* elements. Antimicrob. Agents Chemother. 53 (12), 4961−4967. https://doi.org/10.1128/aac.00579-09.

Ito, T., Hiramatsu, K., 1998. Acquisition of methicillin resistance and progression of multiantibiotic resistance in methicillin-resistant *Staphylococcus aureus*. Yonsei Med. J. 39 (6), 526−533. https://doi.org/10.3349/ymj.1998.39.6.526.

Ito, T., Katayama, Y., Hiramatsu, K., 1999. Cloning and nucleotide sequence determination of the entire *mec* DNA of pre-methicillin-resistant *Staphylococcus aureus* N315. Antimicrob. Agents Chemother. 43 (6), 1449−1458. https://doi.org/10.1128/aac.43.6.1449.

Ito, T., et al., 2001. Structural comparison of three types of staphylococcal cassette chromosome *mec* integrated in the chromosome in methicillin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. 45 (5), 1323−1336. https://doi.org/10.1128/aac.45.5.1323-1336.2001.

Ito, T., et al., 2004. Novel type V staphylococcal cassette chromosome *mec* driven by a novel cassette chromosome recombinase, *ccrC*. Antimicrob. Agents Chemother. 48 (7), 2637−2651. https://doi.org/10.1128/AAC.48.7.2637-2651.2004.

Jackson, M.A., Hicks, R.A., 1987. Vancomycin failure in staphylococcal endocarditis. Pediatr. Infect. Dis. J. 6 (8), 750−752. https://doi.org/10.1097/00006454-198708000-00011.

Jevons, M.P., 1961. 'Celbenin' -resistant staphylococci. Br. Med. J. 1 (5219), 124−125. https://doi.org/10.1136/bmj.1.5219.124-a.

Jevons, M.P., Parker, M.T., 1964. The evolution of new hospital strains of *Staphylococcus aureus*. J. Clin. Pathol. 17 (3), 243−250. https://doi.org/10.1136/jcp.17.3.243.

Kashif, A., et al., 2019. *Staphylococcus aureus* ST398 virulence is associated with factors carried on prophage φSa3. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.02219.

Katayama, Y., Ito, T., Hiramatsu, K., 2000. A new class of genetic element, staphylococcus cassette chromosome *mec*, encodes methicillin resistance in *Staphylococcus aureus*. Antimicrob. Agents Chemother. 44 (6), 1549−1555. https://doi.org/10.1128/aac.44.6.1549-1555.2000.

Katayama, Y., Ito, T., Hiramatsu, K., 2001. Genetic organization of the chromosome region surrounding *mecA* in clinical staphylococcal strains: role of IS431-mediated *mecI* deletion in expression of resistance in *mecA*-carrying, low-level methicillin-resistant *Staphylococcus haemolyticus*. Antimicrob. Agents Chemother. 45 (7), 1955−1963. https://doi.org/10.1128/aac.45.7.1955-1963.2001.

Kirby, W.M.M., 1944. Extraction of a highly potent penicillin inactivator from penicillin-resistant staphylococci. Science 99 (2579), 452−453. https://doi.org/10.1126/science.99.2579.452.

Klevens, R.M., et al., 2006. Community-associated methicillin-resistant *Staphylococcus aureus* and healthcare risk factors. Emerg. Infect. Dis. 12 (12), 1991−1993. https://doi.org/10.3201/eid1212.060505.

Kluytmans, J., van Belkum, A., Verbrugh, H., 1997. Nasal carriage of *Staphylococcus aureus*: epidemiology, underlying mechanisms, and associated risks. Clin. Microbiol. Rev. 10 (3), 505–520. https://doi.org/10.1128/cmr.10.3.505.

Ko, K.S., et al., 2005. Distribution of major genotypes among methicillin-resistant *Staphylococcus aureus* clones in Asian Countries. J. Clin. Microbiol. 43 (1), 421–426. https://doi.org/10.1128/jcm.43.1.421-426.2005.

Koreen, L., et al., 2004. *Spa* typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. J. Clin. Microbiol. 42 (2), 792–799. https://doi.org/10.1128/JCM.42.2.792-799.2004.

Kos, V.N., et al., 2012. Comparative genomics of vancomycin-resistant *Staphylococcus aureus* strains and their positions within the clade most commonly associated with methicillin-resistant *S. aureus* hospital-acquired infection in the United States. mBio 3 (3). https://doi.org/10.1128/mBio.00112-12.

Kreiswirth, B., et al., 1993. Evidence for a clonal origin of methicillin resistance in *Staphylococcus aureus*. Science 259 (5092), 227–230. https://doi.org/10.1126/science.8093647.

Kuhn, G., Francioli, P., Blanc, D.S., 2006. Evidence for clonal evolution among highly polymorphic genes in methicillin-resistant *Staphylococcus aureus*. J. Bacteriol. 188 (1), 169–178. https://doi.org/10.1128/jb.188.1.169-178.2006.

Kuhn, G., et al., 2010. Comparative genomics of epidemic versus sporadic *Staphylococcus aureus* strains does not reveal molecular markers for epidemicity. Infect. Genet. Evol. 10 (1), 89–96. https://doi.org/10.1016/j.meegid.2009.10.011.

Kuroda, M., et al., 2001. Whole genome sequencing of meticillin-resistant *Staphylococcus aureus*. Lancet 357 (9264), 1225–1240. https://doi.org/10.1016/S0140-6736(00)04403-2.

Labandeira-Rey, M., et al., 2007. *Staphylococcus aureus* Panton-Valentine leukocidin causes necrotizing pneumonia. Science 315 (5815), 1130–1133. https://doi.org/10.1126/science.1137165.

Levine, D.P., 2006. Vancomycin: a history. Clin. Infect. Dis. 42 (1), S5–S12. https://doi.org/10.1086/491709.

Li, S., et al., 2011. Novel types of staphylococcal cassette chromosome *mec* elements identified in clonal complex 398 methicillin-resistant *Staphylococcus aureus* strains. Antimicrob. Agents Chemother. 55 (6), 3046–3050. https://doi.org/10.1128/AAC.01475-10.

Lindsay, J.A., Holden, M.T.G., 2004. *Staphylococcus aureus*: superbug, super genome? Trends Microbiol. 12 (8), 378–385. https://doi.org/10.1016/j.tim.2004.06.004.

Lowder, B.V., et al., 2009. Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. Proc. Natl. Acad. Sci. U.S.A. 106 (46), 19545–19550. https://doi.org/10.1073/pnas.0909285106.

Lowy, F.D., 2003. Antimicrobial resistance: the example of *Staphylococcus aureus*. J. Clin. Invest. 111 (9), 1265–1273. https://doi.org/10.1172/JCI18535.

Luini, M., et al., 2015. Methicillin-resistant *Staphylococcus aureus* (MRSA) is associated with low within-herd prevalence of intra-mammary infections in dairy cows: genotyping of isolates. Vet. Microbiol. 178 (3–4), 270–274. https://doi.org/10.1016/j.vetmic.2015.05.010.

Luzzago, C., et al., 2014. Clonal diversity, virulence-associated genes and antimicrobial resistance profile of *Staphylococcus aureus* isolates from nasal cavities and soft tissue infections in wild ruminants in Italian Alps. Vet. Microbiol. 170 (1–2), 157–161. https://doi.org/10.1016/j.vetmic.2014.01.016.

Lyon, B.R., Gillespie, M.T., Skurray, R.A., 1987. Detection and characterization of IS256, an insertion sequence in *Staphylococcus aureus*. Microbiology 133 (11), 3031–3038. https://doi.org/10.1099/00221287-133-11-3031.

Ma, X.X., et al., 2002. Novel type of staphylococcal cassette chromosome *mec* identified in community-acquired methicillin-resistant *Staphylococcus aureus* strains. Antimicrob. Agents Chemother. 46 (4), 1147–1152. https://doi.org/10.1128/aac.46.4.1147-1152.2002.

Maltezou, H.C., et al., 2009. Panton−Valentine leukocidin-positive methicillin-resistant *Staphylococcus aureus* outbreak among healthcare workers in a long-term care facility. Int. J. Infect. Dis. 13 (6), e401−e406. https://doi.org/10.1016/j.ijid.2009.02.004.

Malyszko, I., Schwarz, S., Hauschild, T., 2014. Detection of a new mecC allotype, *mecC2*, in methicillin-resistant *Staphylococcus saprophyticus*. J. Antimicrob. Chemother. 69 (7), 2003−2005. https://doi.org/10.1093/jac/dku043.

McClure, J.A.M., et al., 2018. Genomic comparison of highly virulent, moderately virulent, and avirulent strains from a genetically closely-related MRSA ST239 sub-lineage provides insights into pathogenesis. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.01531.

McClure, J.A., et al., 2021a. *Staphylococcus aureus* ST59: concurrent but separate evolution of North American and East Asian lineages. Front. Microbiol. 12. https://doi.org/10.3389/fmicb.2021.631845.

McClure, J.A., Conly, J.M., Zhang, K., 2021b. Characterizing a novel staphylococcal cassette chromosome *mec* with a composite structure from a clinical strain of *Staphylococcus hominis*, C34847. Antimicrob. Agents Chemother. 65 (11). https://doi.org/10.1128/AAC.00777-21.

McDougal, L.K., et al., 2003. Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the United States: establishing a national database. J. Clin. Microbiol. 41 (11), 5113−5120. https://doi.org/10.1128/JCM.41.11.5113-5120.2003.

Meemken, D., et al., 2010. Livestock associated methicillin-resistant *Staphylococcus aureus* (LaMRSA) isolated from lesions of pigs at necropsy in northwest Germany between 2004 and 2007. Zoonoses Publ. Health 57 (7−8). https://doi.org/10.1111/j.1863-2378.2009.01313.x.

Mellmann, A., et al., 2007. Based Upon Repeat Pattern (BURP): an algorithm to characterize the long-term evolution of *Staphylococcus aureus* populations based on *spa* polymorphisms. BMC Microbiol. 7. https://doi.org/10.1186/1471-2180-7-98.

Melo-Cristino, J., et al., 2013. First case of infection with vancomycin-resistant *Staphylococcus aureus* in Europe. Lancet 382 (9888). https://doi.org/10.1016/s0140-6736(13)61219-2.

Milheiriço, C., Oliveira, D.C., De Lencastre, H., 2007. Update to the multiplex PCR strategy for assignment of *mec* element types in *Staphylococcus aureus*. Antimicrob. Agents Chemother. 51 (9), 3374−3377. https://doi.org/10.1128/AAC.00275-07.

Mir-Sanchis, I., et al., 2012. Control of *Staphylococcus aureus* pathogenicity island excision. Mol. Microbiol. 85 (5), 833−845. https://doi.org/10.1111/j.1365-2958.2012.08145.x.

Mirani, Z.A., Jamil, N., 2011. Effect of sub-lethal doses of vancomycin and oxacillin on biofilm formation by vancomycin intermediate resistant *Staphylococcus aureus*. J. Basic Microbiol. 51 (2), 191−195. https://doi.org/10.1002/jobm.201000221.

Molla, B., et al., 2012. Epidemiology and genotypic characteristics of methicillin-resistant *Staphylococcus aureus* strains of porcine origin. J. Clin. Microbiol. 50 (11), 3687−3693. https://doi.org/10.1128/JCM.01971-12.

Moller, A.G., Lindsay, J.A., Read, T.D., 2019. Determinants of phage host range in *Staphylococcus* species. Appl. Environ. Microbiol. 85 (11). https://doi.org/10.1128/AEM.00209-19.

Monecke, S., Slickers, P., Ehricht, R., 2008. Assignment of *Staphylococcus aureus* isolates to clonal complexes based on microarray analysis and pattern recognition. FEMS Immunol. Med. Microbiol. 53 (2), 237−251. https://doi.org/10.1111/j.1574-695x.2008.00426.x.

Monecke, S., et al., 2011. A field guide to pandemic, epidemic and sporadic clones of methicillin-resistant *Staphylococcus aureus*. PLoS One 6 (4). https://doi.org/10.1371/journal.pone.0017936.

Moore, P.C.L., Lindsay, J.A., 2001. Genetic variation among hospital isolates of methicillin-sensitive *Staphylococcus aureus*: evidence for horizontal transfer of virulence genes. J. Clin. Microbiol. 39 (8), 2760−2767. https://doi.org/10.1128/JCM.39.8.2760-2767.2001.

Murray, B.E., 2000. Vancomycin-resistant enterococcal infections. N. Engl. J. Med. 342 (10), 710−721. https://doi.org/10.1056/NEJM200003093421007.

Musser, J.M., Kapur, V., 1992. Clonal analysis of methicillin-resistant *Staphylococcus aureus* strains from intercontinental sources: association of the *mec* gene with divergent phylogenetic lineages implies dissemination by horizontal transfer and recombination. J. Clin. Microbiol. 30 (8), 2058−2063. https://doi.org/10.1128/jcm.30.8.2058-2063.1992.

Needham, C., Noble, W.C., Dyke, K.G.H., 1995. The staphylococcal insertion sequence IS257 is active. Plasmid 34 (3), 198−205. https://doi.org/10.1006/plas.1995.0005.

Nemeghaire, S., et al., 2013. Characterization of methicillin-resistant *Staphylococcus aureus* from healthy carrier chickens. Avian Pathol. 42 (4), 342−346. https://doi.org/10.1080/03079457.2013.805183.

Nemeghaire, S., et al., 2014. Epidemiology and molecular characterization of methicillin-resistant *Staphylococcus aureus* nasal carriage isolates from bovines. BMC Vet. Res. 10 (1). https://doi.org/10.1186/1746-6148-10-153.

Normanno, G., et al., 2015. Methicillin-resistant *Staphylococcus aureus* (MRSA) in slaughtered pigs and abattoir workers in Italy. Food Microbiol. 51, 51−56. https://doi.org/10.1016/j.fm.2015.04.007.

Nübel, U., et al., 2008. Frequent emergence and limited geographic dispersal of methicillin-resistant *Staphylococcus aureus*. Proc. Natl. Acad. Sci. U.S.A. 105 (37), 14130−14135. https://doi.org/10.1073/pnas.0804178105.

Ogston, A., 1881. Report upon micro-organisms in surgical diseases. BMJ 1 (1054), 369−375. https://doi.org/10.1136/bmj.1.1054.369.

Ohta, T., et al., 2004. Nucleotide substitutions in *Staphylococcus aureus* strains, Mu50, Mu3, and N315. DNA Res. 11 (1), 51−56. https://doi.org/10.1093/dnares/11.1.51.

Oliveira, D.C., De Lencastre, H., 2002. Multiplex PCR strategy for rapid identification of structural types and variants of the *mec* element in methicillin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. 46 (7), 2155−2161. https://doi.org/10.1128/AAC.46.7.2155-2161.2002.

Oliveira, D.C., De Lencastre, H., Tomasz, A., 2012. Evolution of molecular techniques for the characterization of MRSA clones. In: Antibiotic Discovery and Development. Springer US, Portugal, pp. 571−592. https://doi.org/10.1007/978-1-4614-1400-1_17.

Otter, J.A., French, G.L., 2006. Nosocomial transmission of community-associated methicillin-resistant *Staphylococcus aureus*: an emerging threat. Lancet Infect. Dis. 6 (12), 753−755. https://doi.org/10.1016/S1473-3099(06)70636-3.

Otter, J.A., French, G.L., 2011. Community-associated meticillin-resistant *Staphylococcus aureus* strains as a cause of healthcare-associated infection. J. Hosp. Infect. 79 (3), 189−193. https://doi.org/10.1016/j.jhin.2011.04.028.

Otter, J.A., French, G.L., 2012. Community-associated methicillin-resistant *Staphylococcus aureus*: the case for a genotypic definition. J. Hosp. Infect. 81 (3), 143−148. https://doi.org/10.1016/j.jhin.2012.04.009.

Panesso, D., et al., 2015. Methicillin-susceptible, vancomycin-resistant *Staphylococcus aureus*, Brazil. Emerg. Infect. Dis. 21 (10), 1844−1848. https://doi.org/10.3201/eid2110.141914.

Pang, R., et al., 2020. The genomic context for the evolution and transmission of community-associated *Staphylococcus aureus* ST59 through the food chain. Front. Microbiol. 11. https://doi.org/10.3389/fmicb.2020.00422.

Panton, P.N., Valentine, F.C.O., 1932. Staphylococcal toxin. Lancet, 24468-7. https://doi.org/10.1016/S0140-6736(01).

Patel, M., et al., 2008. Emergence of USA300 MRSA in a tertiary medical centre: implications for epidemiological studies. J. Hosp. Infect. 68 (3), 208−213. https://doi.org/10.1016/j.jhin.2007.12.010.

Peacock, S.J., et al., 2002. Virulent combinations of adhesin and toxin genes in natural populations of *Staphylococcus aureus*. Infect. Immun. 70 (9), 4987−4996. https://doi.org/10.1128/IAI.70.9.4987-4996.2002.

Pérez-Roth, E., et al., 2004. Tracking methicillin-resistant *Staphylococcus aureus* clones during a 5-year period (1998 to 2002) in a Spanish hospital. J. Clin. Microbiol. 42 (10), 4649−4656. https://doi.org/10.1128/JCM.42.10.4649-4656.2004.

Périchon, B., Courvalin, P., 2009. VanA-type vancomycin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. 53 (11), 4580–4587. https://doi.org/10.1128/aac.00346-09.

Prevost, G., Jaulhac, B., Piemont, Y., 1992. DNA fingerprinting by pulsed-field gel electrophoresis is more effective than ribotyping in distinguishing among methicillin-resistant *Staphylococcus aureus* isolates. J. Clin. Microbiol. 30 (4), 967–973. https://doi.org/10.1128/jcm.30.4.967-973.1992.

Price, L.B., et al., 2012. *Staphylococcus aureus* CC398: host adaptation and emergence of methicillin resistance in livestock. mBio 3 (1), 1–6. https://doi.org/10.1128/mBio.00305-11.

Rammelkamp, C.H., Maxon, T., 1942. Resistance of *Staphylococcus aureus* to the action of penicillin. Exp. Biol. Med. 51 (3), 386–389. https://doi.org/10.3181/00379727-51-13986.

Reynolds, P.E., Brown, D.F., 1985. Penicillin-binding proteins of beta-lactam-resistant strains of *Staphylococcus aureus*. Effect of growth conditions. FEBS Lett. 192, 28–32. https://doi.org/10.1016/0014-5793(85.

Robinson, D.A., Enright, M.C., 2003. Evolutionary models of the emergence of methicillin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. 47 (12), 3926–3934. https://doi.org/10.1128/AAC.47.12.3926-3934.2003.

Rossi, F., et al., 2014. Transferable vancomycin resistance in a community-associated MRSA lineage. N. Engl. J. Med. 370 (16), 1524–1531. https://doi.org/10.1056/NEJMoa1303359.

Rouch, D.A., et al., 1987. The aacA-aphD gentamicin and kanamycin resistance determinant of Tn4001 from *Staphylococcus aureus*: expression and nucleotide sequence analysis. Microbiology 133 (11), 3039–3052. https://doi.org/10.1099/00221287-133-11-3039.

Rountree, P.M., Freeman, B.M., 1955. Infections caused by a particular phage type of *Staphylococcus aureus*. Med. J. Aust. 42 (5), 157–161.

Růžičková, V., et al., 2012. Major clonal lineages in impetigo *Staphylococcus aureus* strains isolated in Czech and Slovak maternity hospitals. Int. J. Med. Microbiol. 302 (6), 237–241. https://doi.org/10.1016/j.ijmm.2012.04.001.

Ruzin, A., Lindsay, J., Novick, R.P., 2001. Molecular genetics of SaPI1 - a mobile pathogenicity island in *Staphylococcus aureus*. Mol. Microbiol. 41 (2), 365–377. https://doi.org/10.1046/j.1365-2958.2001.02488.x.

Saha, B., et al., 2008. Identification and characterization of a vancomycin-resistant *Staphylococcus aureus* isolated from Kolkata (South Asia). J. Med. Microbiol. 57 (1), 72–79. https://doi.org/10.1099/jmm.0.47144-0.

Saunders, A., et al., 2007. A nosocomial outbreak of community-associated methicillin-resistant *Staphylococcus aureus* among healthy newborns and postpartum mothers. Can. J. Infect Dis. Med. Microbiol. 18 (2), 128–132. https://doi.org/10.1155/2007/617526.

Schulz, J., et al., 2012. Longitudinal study of the contamination of air and of soil surfaces in the vicinity of pig barns by livestock-associated methicillin-resistant *Staphylococcus aureus*. Appl. Environ. Microbiol. 78 (16), 5666–5671. https://doi.org/10.1128/AEM.00550-12.

Schwartz, D.C., Cantor, C.R., 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell 37 (1), 67–75. https://doi.org/10.1016/0092-8674(84)90301-5.

Schwendener, S., Cotting, K., Perreten, V., 2017. Novel methicillin resistance gene mecD in clinical *Macrococcus caseolyticus* strains from bovine and canine sources. Sci. Rep. 7 (1). https://doi.org/10.1038/srep43797.

Schwendener, S., et al., 2021. Novel SCC*mec* element containing the methicillin resistance gene *mecD* in *Macrococcus bohemicus*. J. Global Antimicrob. Resist. 24, 360–362. https://doi.org/10.1016/j.jgar.2021.02.001.

Seybold, U., et al., 2006. Emergence of community-associated methicillin-resistant *Staphylococcus aureus* USA300 genotype as a major cause of health care-associated blood stream infections. Clin. Infect. Dis. 42 (5), 647–656. https://doi.org/10.1086/499815.

Shaffer, T.E., et al., 1957. Staphylococcal Infections in newborn infants. Report of 19 epidemics caused by an identical strain of *Staphylococcus pyogenes*. Am. J. Public Health Nation's Health 47 (8), 990–994. https://doi.org/10.2105/ajph.47.8.990.

Shi, D., et al., 2011. Bullous impetigo in children infected with methicillin-resistant *Staphylococcus aureus* alone or in combination with methicillin-susceptible *S. aureus*: analysis of genetic characteristics, including assessment of exfoliative toxin gene carriage. J. Clin. Microbiol. 49 (5), 1972−1974. https://doi.org/10.1128/jcm.01742-10.

Shopsin, B., et al., 1999. Evaluation of protein a gene polymorphic region DNA sequencing for typing of *Staphylococcus aureus* strains. J. Clin. Microbiol. 37 (11), 3556−3563. https://doi.org/10.1128/jcm.37.11.3556-3563.1999.

Shore, A.C., et al., 2011. Detection of staphylococcal cassette chromosome *mec* type XI carrying highly divergent *mecA, mecI, mecR1, blaZ*, and *ccr* genes in human clinical isolates of clonal complex 130 methicillin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. 55 (8), 3765−3773. https://doi.org/10.1128/AAC.00187-11.

Skinner, D., Keefer, C.S., 1941. Significance of bacteremia caused by *Staphylococcus aureus*: a study of one hundred and twenty-two cases and a review of the literature concerned with experimental infection in animals. Arch. Intern. Med. 68 (5), 851−875. https://doi.org/10.1001/archinte.1941.00200110003001.

Skov, R.L., Jensen, K.S., 2009. Community-associated methicillin-resistant *Staphylococcus aureus* as a cause of hospital-acquired infections. J. Hosp. Infect. 73 (4), 364−370. https://doi.org/10.1016/j.jhin.2009.07.004.

Sorrell, T.C., et al., 1982. Vancomycin therapy for methicillin-resistant *Staphylococcus aureus*. Ann. Intern. Med. 97 (3), 344−351. https://doi.org/10.7326/0003-4819-97-3-344.

Spoor, L.E., et al., 2013. Livestock origin for a human pandemic clone of community-associated methicillin-resistant *Staphylococcus aureus*. mBio 4 (4). https://doi.org/10.1128/mBio.00356-13.

Spratt, B.G., et al., 2004. Displaying the relatedness among isolates of bacterial species - the eBURST approach. FEMS Fed. Eur. Microbiol. Soc. Microbiol. Lett. 241 (2), 129−134. https://doi.org/10.1016/j.femsle.2004.11.015.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (9), 1312−1313. https://doi.org/10.1093/bioinformatics/btu033.

Stefani, S., et al., 2012. Methicillin-resistant *Staphylococcus aureus* (MRSA): global epidemiology and harmonisation of typing methods. Int. J. Antimicrob. Agents 39 (4), 273−282. https://doi.org/10.1016/j.ijantimicag.2011.09.030.

Stein, R.A., 2009. Methicillin-resistant *Staphylococcus aureus*-the new zoonosis. Int. J. Infect. Dis. 13 (3), 299−301. https://doi.org/10.1016/j.ijid.2008.09.008.

Stemper, M.E., et al., 2006. Shift in *Staphylococcus aureus* clone linked to an infected tattoo. Emerg. Infect. Dis. 12 (9), 1444−1446. https://doi.org/10.3201/eid1209.051634.

Strommenger, B., et al., 2004. Obvious lack of association between dynamics of epidemic methicillin-resistant *Staphylococcus aureus* in central Europe and *agr* specificity groups. Eur. J. Clin. Microbiol. Infect. Dis. 23 (1), 15−19. https://doi.org/10.1007/s10096-003-1046-8.

Strommenger, B., et al., 2006. Assignment of *Staphylococcus* isolates to groups by spa typing, SmaI macrorestriction analysis, and multilocus sequence typing. J. Clin. Microbiol. 44 (7), 2533−2540. https://doi.org/10.1128/jcm.00420-06.

Suchard, M.A., et al., 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 4 (1). https://doi.org/10.1093/ve/vey016.

Sumby, P., Waldor, M.K., 2003. Transcription of the toxin genes present within the staphylococcal phage φSa3ms is intimately linked with the phage's life cycle. J. Bacteriol. 185 (23), 6841−6851. https://doi.org/10.1128/JB.185.23.6841-6851.2003.

Takano, T., et al., 2008. Novel characteristics of community-acquired methicillin-resistant *Staphylococcus aureus* strains belonging to multilocus sequence type 59 in Taiwan. Antimicrob. Agents Chemother. 52 (3), 837−845. https://doi.org/10.1128/AAC.01001-07.

Teare, L., et al., 2010. Outbreak of Panton−Valentine leucocidin-positive methicillin-resistant *Staphylococcus aureus* in a regional burns unit. J. Hosp. Infect. 76 (3), 220−224. https://doi.org/10.1016/j.jhin.2010.04.023.

Tenover, F.C., et al., 2006. Characterization of a strain of community-associated methicillin-resistant *Staphylococcus aureus* widely disseminated in the United States. J. Clin. Microbiol. 44 (1), 108−118. https://doi.org/10.1128/JCM.44.1.108-118.2006.

Tenover, F.C., et al., 2008. Characterization of *Staphylococcus aureus* isolates from nasal cultures collected from individuals in the United States in 2001 to 2004. J. Clin. Microbiol. 46 (9), 2837−2841. https://doi.org/10.1128/JCM.00480-08.

Thampi, N., et al., 2015. Multicenter study of health care cost of patients admitted to hospital with *Staphylococcus aureus* bacteremia: impact of length of stay and intensity of care. Am. J. Infect. Control 43 (7), 739−744. https://doi.org/10.1016/j.ajic.2015.01.031.

Thati, V., Shivannavar, C.T., Gaddad, S.M., 2011. Vancomycin resistance among methicillin resistant *Staphylococcus aureus* isolates from intensive care units of tertiary care hospitals in Hyderabad. Indian J. Med. Res. 134 (11), 704−708. https://doi.org/10.4103/0971-5916.91001.

Thompson, R.L., Cabezudo, I., Wenzel, R.P., 1982. Epidemiology of nosocomial infections caused by methicillin-resistant *Staphylococcus aureus*. Ann. Intern. Med. 97 (3), 309−317. https://doi.org/10.7326/0003-4819-97-3-309.

Thurlow, L.R., et al., 2013. Functional modularity of the arginine catabolic mobile element contributes to the success of USA300 methicillin-resistant *Staphylococcus aureus*. Cell Host Microbe 13 (1), 100−107. https://doi.org/10.1016/j.chom.2012.11.012.

Tickler, I.A., et al., 2017. Continued expansion of USA300-like methicillin-resistant *Staphylococcus aureus* (MRSA) among hospitalized patients in the United States. Diagn. Microbiol. Infect. Dis. 88 (4), 342−347. https://doi.org/10.1016/j.diagmicrobio.2017.04.016.

Tiwari, H.K., Sen, M.R., 2006. Emergence of vancomycin resistant *Staphylococcus aureus* (VRSA) from a tertiary care hospital from northern part of India. BMC Infect. Dis. 6. https://doi.org/10.1186/1471-2334-6-156.

Tsubakishita, S., et al., 2010. Origin and molecular evolution of the determinant of methicillin resistance in staphylococci. Antimicrob. Agents Chemother. 54 (10), 4352−4359. https://doi.org/10.1128/aac.00356-10.

Udo, E.E., Pearman, J.W., Grubb, W.B., 1993. Genetic analysis of community isolates of methicillin-resistant *Staphylococcus aureus* in Western Australia. J. Hosp. Infect. 25 (2), 97−108. https://doi.org/10.1016/0195-6701(93)90100-E.

Uhlemann, A.C., et al., 2011. The environment as an unrecognized reservoir for community-associated methicillin resistant *Staphylococcus aureus* USA300: a case-control study. PLoS One 6 (7). https://doi.org/10.1371/journal.pone.0022407.

Uhlemann, A.-C., et al., 2012. Molecular characterization of *Staphylococcus aureus* from outpatients in the Caribbean reveals the presence of pandemic clones. Eur. J. Clin. Microbiol. Infect. Dis. 31 (4), 505−511. https://doi.org/10.1007/s10096-011-1339-2.

Uhlemann, A.C., et al., 2013. Emergence of sequence type 398 as a community- and healthcare-associated methicillin-susceptible *Staphylococcus aureus* in Northern Manhattan. Clin. Infect. Dis. 57 (5), 700−703. https://doi.org/10.1093/cid/cit375.

Uhlemann, A.C., et al., 2017. Evolutionary dynamics of pandemic methicillin-sensitive *Staphylococcus aureus* ST398 and its international spread via routes of human migration. mBio 8 (1). https://doi.org/10.1128/mBio.01375-16.

Urushibara, N., et al., 2020. Novel staphylococcal cassette chromosome *mec* (SCC*mec*) type XIV (5A) and a truncated SCC*mec* element in SCC composite islands carrying *speG* in ST5 MRSA in Japan. J. Antimicrob. Chemother. 75 (1), 46−50. https://doi.org/10.1093/jac/dkz406.

Utsui, Y., Yokota, T., 1985. Role of an altered penicillin-binding protein in methicillin- and cephem-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. 28 (3), 397–403. https://doi.org/10.1128/aac.28.3.397.

Van Alen, S., et al., 2018. Prevalence and genomic structure of bacteriophage phi3 in human-derived livestock-associated methicillin-resistant *Staphylococcus aureus* isolates from 2000 to 2015. J. Clin. Microbiol. 56 (9). https://doi.org/10.1128/JCM.00140-18.

Van Belkum, A., et al., 2006. Clonal distribution and differential occurrence of the enterotoxin gene cluster, egc, in carriage- versus bacteremia-associated isolates of *Staphylococcus aureus*. J. Clin. Microbiol. 44 (4), 1555–1557. https://doi.org/10.1128/JCM.44.4.1555-1557.2006.

Van Belkum, A., et al., 2008. Methicillin -resistant and -susceptible *Staphylococcus aureus* sequence type 398 in pigs and humans. Emerg. Infect. Dis. 14 (3), 479–483. https://doi.org/10.3201/eid1403.070760.

Van Rijen, M.M.L., Van Keulen, P.H., Kluytmans, J.A., 2008. Increase in a Dutch hospital of methicillin-resistant *Staphylococcus aureus* related to animal farming. Clin. Infect. Dis. 46 (2), 261–263. https://doi.org/10.1086/524672.

van Rijen, M.M.L., et al., 2013. Lifestyle-associated risk factors for community-acquired methicillin-resistant *Staphylococcus aureus* carriage in the Netherlands: an exploratory hospital-based case-control study. PLoS One 8 (6). https://doi.org/10.1371/journal.pone.0065594.

Vandenesch, F., et al., 2003. Community-acquired methicillin-resistant *Staphylococcus aureus* carrying panton-valentine leukocidin genes: worldwide emergence. Emerg. Infect. Dis. 9 (8), 978–984. https://doi.org/10.3201/eid0908.030089.

Vanderhaeghen, W., et al., 2010. Methicillin-resistant *Staphylococcus aureus* (MRSA) in food production animals. Epidemiol. Infect. 138 (5), 606–625. https://doi.org/10.1017/s0950268809991567.

Voss, A., et al., 2005. Methicillin-resistant *Staphylococcus aureus* in pig farming. Emerg. Infect. Dis. 11 (12), 1965–1966. https://doi.org/10.3201/eid1112.050428.

Voyich, J.M., et al., 2006. Is Panton-Valentine leukocidin the major virulence determinant in community-associated methicillin-resistant *Staphylococcus aureus* disease? JID J. Infect. Dis. 194 (12), 1761–1770. https://doi.org/10.1086/509506.

Waldron, D.E., Lindsay, J.A., 2006. Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into *Staphylococcus aureus* and between *S. aureus* isolates of different lineages. J. Bacteriol. 188 (15), 5578–5585. https://doi.org/10.1128/JB.00418-06.

Wang, L., Archer, G.L., 2010. Roles of CcrA and CcrB in excision and integration of staphylococcal cassette chromosome *mec*, a *Staphylococcus aureus* genomic island. J. Bacteriol. 192 (12), 3204–3212. https://doi.org/10.1128/JB.01520-09.

Wang, D., et al., 2015. Bovine mastitis *Staphylococcus aureus*: antibiotic susceptibility profile, resistance genes and molecular typing of methicillin-resistant and methicillin-sensitive strains in China. Infect. Genet. Evol. 31, 9–16. https://doi.org/10.1016/j.meegid.2014.12.039.

Weese, J.S., 2010. Methicillin-resistant *Staphylococcus aureus* in animals. ILAR J. 51 (3), 233–244. https://doi.org/10.1093/ilar.51.3.233.

Whitby, M., McLaws, M.L., Berry, G., 2001. Risk of death from methicillin-resistant *Staphylococcus aureus* bacteraemia: a meta-analysis. Med. J. Aust. 175 (5), 264–267. https://doi.org/10.5694/j.1326-5377.2001.tb143562.x.

Wielders, C.L., et al., 2001. In-vivo transfer of *mecA* DNA to *Staphylococcus aureus*. Lancet 357, 4832–4837. https://doi.org/10.1016/s0140-6736(00.

Wisplinghoff, H., et al., 2003. Related clones containing SCC*mec* type IV predominate among clinically significant *Staphylococcus epidermidis* isolates. Antimicrob. Agents Chemother. 47 (11), 3574–3579. https://doi.org/10.1128/AAC.47.11.3574-3579.2003.

Wu, S.W., Lencastre, H.de, Tomasz, A., 2001. Recruitment of the *mecA* gene homologue of *Staphylococcus sciuri* into a resistance determinant and expression of the resistant phenotype in *Staphylococcus aureus*. J. Bacteriol. 183 (8), 2417−2424. https://doi.org/10.1128/jb.183.8.2417-2424.2001.

Wu, K., et al., 2008. Panton-Valentine leukocidin (PVL)-positive and -negative community-associated MRSA (CA-MRSA) USA400 sibling strains exhibit similar virulence characteristics in in vitro cell culture and animal models. In: 13th International Symposium of Staphylococci Staphylococcal Infectinon.

Wu, Z., et al., 2015. Novel type XII staphylococcal cassette chromosome *mec* harboring a new cassette chromosome recombinase, CcrC2. Antimicrob. Agents Chemother. 59 (12), 7597−7601. https://doi.org/10.1128/aac.01692-15.

Wu, K., et al., 2020. Arginine catabolic mobile element in evolution and pathogenicity of the community-associated methicillin-resistant *Staphylococcus aureus* strain USA300. Microorganisms 8 (2). https://doi.org/10.3390/microorganisms8020275.

Wu, Q., et al., 2021. Systematic review and meta-analysis of the epidemiology of vancomycin-resistance Staphylococcus aureus isolates. Antimicrob. Resist. Infect. Control 10 (1). https://doi.org/10.1186/s13756-021-00967-y.

Wulf, M.W.H., et al., 2012. Infection and colonization with methicillin resistant *Staphylococcus aureus* ST398 versus other MRSA in an area with a high density of pig farms. Eur. J. Clin. Microbiol. Infect. Dis. 31 (1), 61−65. https://doi.org/10.1007/s10096-011-1269-z.

Wyllie, D., Paul, J., Crook, D., 2011. Waves of trouble: MRSA strain dynamics and assessment of the impact of infection control. J. Antimicrob. Chemother. 66 (12), 2685−2688. https://doi.org/10.1093/jac/dkr392.

Zhang, K., et al., 2005. Novel multiplex PCR assay for characterization and concomitant subtyping of staphylococcal cassette chromosome *mec* types I to V in methicillin-resistant *Staphylococcus aureus*. J. Clin. Microbiol. 43 (10), 5026−5033. https://doi.org/10.1128/JCM.43.10.5026-5033.2005.

Zhang, K., et al., 2008. Coexistence of Panton-Valentine leukocidin-positive and -negative community-associated methicillin-resistant *Staphylococcus aureus* USA400 sibling strains in a large Canadian health-care region. JID J. Infect. Dis. 197 (2), 195−204. https://doi.org/10.1086/523763.

Zhang, K., McClure, J.A., Conly, J.M., 2012. Enhanced multiplex PCR assay for typing of staphylococcal cassette chromosome *mec* types I to V in methicillin-resistant *Staphylococcus aureus*. Mol. Cell. Probes 26 (5), 218−221. https://doi.org/10.1016/j.mcp.2012.04.002.

Zhou, Y., et al., 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39 (Suppl. l), W347−W352. https://doi.org/10.1093/nar/gkr485.

Zhu, W., et al., 2010. Dissemination of an *Enterococcus* Inc18-like *vanA* plasmid associated with vancomycin-resistant *Staphylococcus aureus*. Antimicrob. Agents Chemother. 54 (10), 4314−4320. https://doi.org/10.1128/AAC.00185-10.

This page intentionally left blank

CHAPTER

Origin and Emergence of HIV/AIDS 29

Martine Peeters[1], Mirela D'arc[1,2], Lucie Etienne[1,3], Eric Delaporte[1] and Ahidjo Ayouba[1]

[1]*TransVIHMI, University of Montpellier, Institut de Recherche pour le Développement (IRD), INSERM, Montpellier, France;* [2]*Department of Genetics, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil;* [3]*International Center for Infectiology Research, INSERM U1111 and Ecole Normale Supérieure de Lyon and Université Lyon 1 and CNRS UMR Lyon, France*

1. History of AIDS

AIDS or the "Acquired Immune Deficiency Syndrome" was first recognized between 1979 and 1981 among Men having Sex with Men (MSM) in New York, Los Angeles, and San Francisco who presented with pneumonia caused by *Pneumocystis carinii* and/or with symptoms of Kaposi sarcoma (CDC, 1981) (Fig. 29.1). Subsequently, patients with similar symptoms were seen among Intravenous Drug Users (IDUs), hemophiliacs, Haitians, and Africans in Europe. In May 1983, the etiologic agent of AIDS, the Human Immunodeficiency Virus (HIV), was identified (Barré-Sinoussi et al., 1983). In 1984, several authors reported AIDS cases in women and men in hospitals from sub-Saharan Africa, suggesting the existence of a heterosexual epidemic and confirming presence of the disease on the African continent (Ellrodt et al., 1984; Piot et al., 1984; Van de Perre et al., 1984). Sero-epidemiological studies showed subsequently that a significant proportion of the population in certain regions of Africa was infected with HIV in the 1980s and even before. In the early 1990s, the epidemic exploded in Southern and Eastern Africa, where in certain urban areas 25% of pregnant women were HIV-positive (Buvé et al., 2002).

Molecular epidemiological studies revealed that the epicenter of the HIV pandemic is situated in Central Africa, and more precisely the area of Kinshasa, the capital city of the Democratic Republic of Congo (DRC) (Vidal et al., 2000; Worobey et al., 2008; Faria et al., 2014). The virus has been introduced from Africa in Haiti in the 1960s before it started to circulate in North America (Worobey et al., 2016) about a decade before the discovery and description of the first AIDS cases (Gilbert et al., 2007; Low-Beer, 2001). The last estimates show that since the start of the epidemic more than 84 million individuals have been infected with the virus, causing more than 40 million deaths. At the end of 2021, it is estimated that around 38.4 million people were living with HIV (Fig. 29.1), and more than 70% of HIV-infected persons live in sub-Saharan Africa. Increasing and earlier access to Antiretroviral Treatment (ART) improved life quality of HIV-infected individuals but reduced also the spread of HIV-1 (UNAIDS/WHO, 2014). As a consequence, the number of new HIV infections dropped by 32% since 2001, nevertheless, around 1.5 million people became infected in 2021 and 650,000 died of the disease. HIV/AIDS was one

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00019-7
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

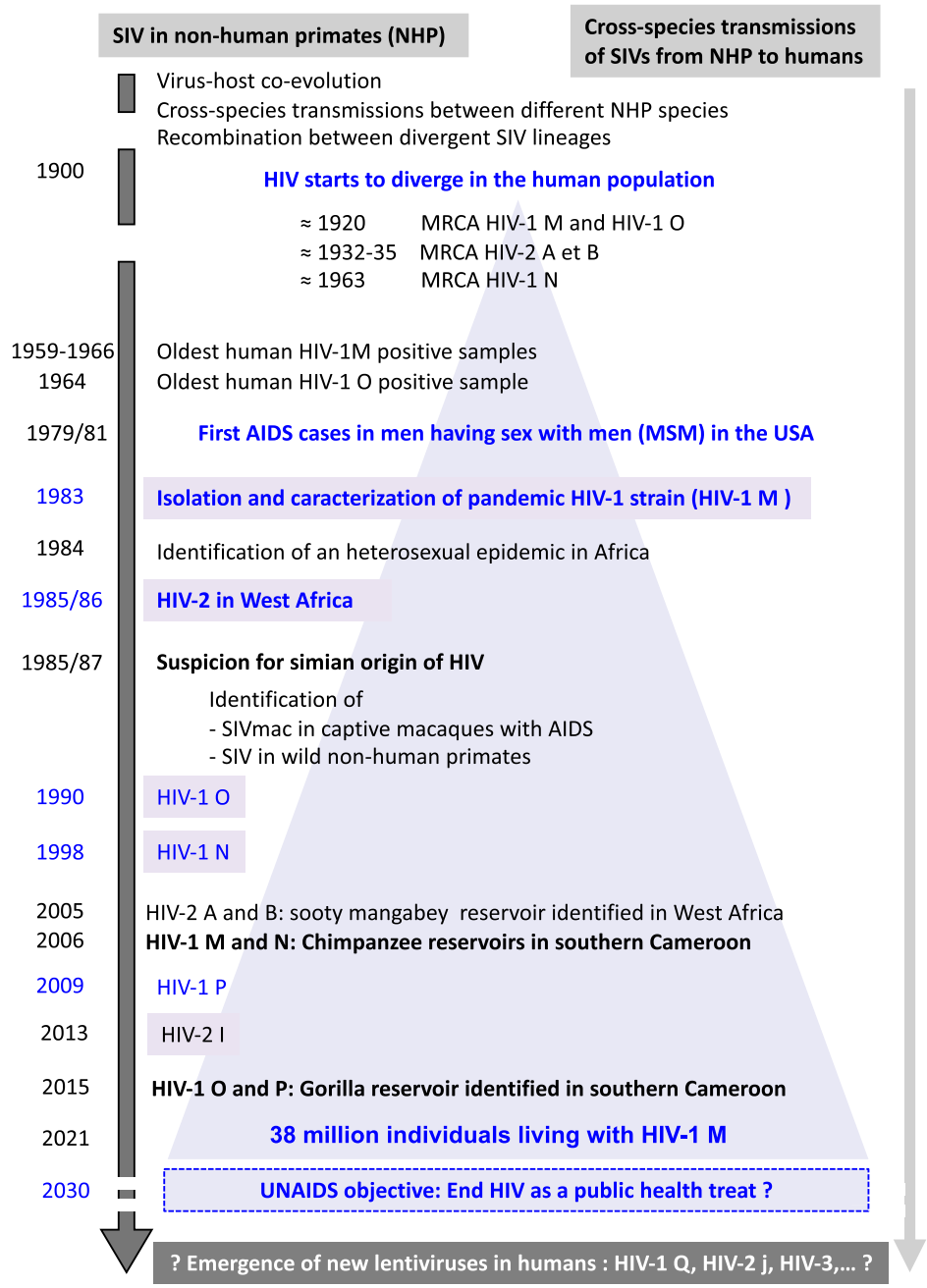


FIGURE 29.1 History of the AIDS epidemic: Past and future events.

Dates referring to events in the history of the HIV epidemic in humans are shown at the *left*, major events are highlighted in *blue*, boxes indicate subsequent discoveries of the different HIV-1 groups. The number of persons infected with HIV increases over time as illustrated by the *blue triangle*. *Gray arrows* represent a schematic time scale of the different events. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

of the most important infectious diseases that emerged in the 20th century and many efforts are still necessary to end HIV/AIDS as a public health threat in the 21st century as aimed by UNAIDS. Like for other emerging infectious diseases, it is thus important to understand where this virus came from, how it has been introduced into the human population and which factors were associated with host adaptation and epidemic spread to evaluate risk for potential emergence or re-emergence of retroviruses.

2. Human Immunodeficiency Viruses are Closely Related to Simian Immunodeficiency Virus From Nonhuman Primates

Whereas HIV-1 has been identified in 1983, another closely related virus, HIV-2, has been described in 1985 in West Africa and among West Africans with AIDS in France (Barin et al., 1985; Clavel et al., 1986; Kanki et al., 1986). HIV-1 and HIV-2 share the same morphology and $CD4^+$ T cell tropism but are genetically and antigenically distinct. HIV-1 and HIV-2 belong to the *Lentivirus* genus of the Retroviridae family where five serogroups are recognized, each reflecting the vertebrate hosts with which they are associated (primates, sheep and goats, horses, cats, and cattle). Both, HIV-1 and HIV-2, are most closely related to the lentiviruses from Nonhuman Primates (NHP), called Simian Immunodeficiency Virus (SIV), and are thus most likely the result of cross-species transmissions of SIVs from African NHP.

2.1 Discovery of the First SIV

Shortly after the identification of HIV-1 as the cause of AIDS in 1983, the first SIV, SIVmac, was isolated from rhesus macaques (*Macaca mulatta*) with immune deficiency and clinical symptoms similar to AIDS at the New England Regional Primate Research Center (NERPRC) in the United States (Daniel et al., 1985; Henrickson et al., 1983). Retrospective studies revealed that SIVmac was introduced at NERPRC by other rhesus monkeys, previously housed at the California National Primate Research Center (CNPRC), where they survived an earlier disease outbreak (late 1960s), characterized by immune suppression and opportunistic infections (Mansfield et al., 1995). A decade after the first outbreak, stump-tailed macaques (*Macaca arctoides*) developed a similar disease in the same settings and a lentivirus, called SIVstm was isolated from frozen tissue from one of these monkeys (Lowenstine et al., 1986). In both cases, the infected macaques had been in contact with healthy, but retrospectively shown, SIVsmm positive sooty mangabeys (*Cercocebys atys*) from west Africa at the CNPRC (Apetrei et al., 2005; Lowenstine et al., 1986). The close phylogenetic relationship between SIVmac, SIVstm, and SIVsmm identified sooty mangabeys as the plausible source of SIV in captive macaques (Murphey-Corb et al., 1986).

Since SIVmac induced a disease in rhesus macaques with remarkable similarity to AIDS in humans, a simian origin of HIV was soon suspected. The discovery of HIV-2, the agent of AIDS in West Africa, and the remarkable relatedness of HIV-2 with SIVsmm, naturally occurring in sooty mangabeys in West Africa, reinforced this hypothesis (Gao et al., 1992; Hirsch et al., 1989).

2.2 SIVs in African NHP

Currently, SIV infection has been documented in more than 45 different NHP species by serological and/or sequence analysis (Table 29.1 and Fig. 29.2) (Peeters et al., 2014). A high genetic diversity is

Table 29.1 SIV Infections in Old World Monkeys and Apes in Africa. For Each Species, the Genus, Species, and Subspecies (if Applicable) are Given. Species Representing a Reservoir for HIV-1 and 2 are Highlighted in Bold

Genus	Species/Subspecies	Common Name	SIV
Pan	***troglodytes troglodytes***	**Central chimpanzee**	**SIVcpz*Ptt***
	troglodytes schweinfurthii	Eastern chimpanzee	SIVcpz*Pts*
Gorilla	***gorilla gorilla***	**Western lowland gorilla**	**SIVgor**
Colobus	guereza	*Mantled guereza*	*SIVcol-1, col-2*
Piliocolobus	*badius badius*	Western red colobus	SIVwrc*Pbb*
	badius temminckii	Temminck's red colobus	SIVwrc*Pbt*
	tholloni	Tshuapa red colobus	SIVtrc[a]
	rufomitratus tephrosceles	Ugandan red colobus	SIVkrc
Procolobus	*verus*	Olive colobus	SIVolc
Lophocebus	*albigena*	Gray-cheeked managabey	b
	aterrimus	Black crested mangabey	SIVbkm[a]
Papio	*anubis*	Olive baboon	b
	cynocephalus	Yellow baboon	SIVagm-Ver[a]
	ursinus	Chacma baboon	SIVagm-Ver[a]
Cercocebus	***atys***	**Sooty mangabey**	**SIVsmm**
	torquatus	Red-capped mangabey	SIVrcm
	agilis	Agile mangabey	SIVagi
Mandrillus	*sphinx*	Mandrill	SIVmnd-1, mnd-2
	leucophaeus	Drill	SIVdrl
Allenopithecus	*nigroviridis*	Allen's swamp monkey	SIVasm[a,c]
Miopithecus	*talapoin*	Angolan talapoin	SIVtal[a]
	ogouensis	Gabon talapoin	SIVtal
Erythrocebus	*patas*	Patas monkey	SIVagm-sab[a]
Chlorocebus	*sabaeus*	Green monkey	SIVagm-Sab
	aethiops	Grivet	SIVagm-Gri
	tantalus	Tantalus monkey	SIVagm-Tan
	pygerythrus	Vervet monkey	SIVagm-Ver
Cercopithecus	*diana*	Diana monkey	b
	nictitans	Greater spot-nosed monkey	SIVgsn
	mitis	Blue monkey	SIVblu[a]
	albogularis	Sykes's monkey	SIVsyk
	mona	Mona monkey	SIVmon
	campbelli	Campbell's mona	b
	pogonias	Crested mona	b

Table 29.1 SIV Infections in Old World Monkeys and Apes in Africa. For Each Species, the Genus, Species, and Subspecies (if Applicable) are Given. Species Representing a Reservoir for HIV-1 and 2 are Highlighted in Bold—cont'd

Genus	Species/Subspecies	Common Name	SIV
	denti	Dent's mona	SIVden
	cephus	Mustached guenon	SIVmus-1,mus-2, mus-3
	erythrotis	Red-eared monkey	SIVery[a]
	ascanius	Red-tailed monkey	SIVasc
	lhoest	l'Hoest monkey	SIVlho
	solatus	Sun-tailed monkey	SIVsun
	preussi	Preuss's monkey	SIVpre[a]
	hamlyni	Owl-faced monkey	b
	neglectus	de Brazza's monkey	SIVdeb
	wolfi	Wolf's monkey	SIVwol[a]

[a]*only partial SIV sequences are available*
[b]*only serological evidence for SIV.*
[c]*unpublished Ahuka-Mundeke S et al.*

observed among the different SIVs, but generally, each NHP species is infected with a species-specific virus, which forms monophyletic lineages in phylogenetic trees. These species-specific SIVs are identified by a lower-case-three-letter code, which corresponds to the initial letters of the common species name; e.g., SIVagm for African green monkeys, SIVrcm for red-capped mangabeys. When different subspecies of the same species are infected, an abbreviation referring to the name of the subspecies is added to the virus designation, i.e., SIVcpz*Ptt* and SIVcpz*Pts* to differentiate between the two chimpanzee subspecies, *Pan troglodytes troglodytes* and *P. t. schweinfurthii*, respectively.

Importantly, the number of African NHP species infected with SIV is most probably underestimated, since around 30 species of the 73 actually recognized Old World NHP species in sub-Saharan Africa have not been tested yet or only very few individuals. Knowing that the vast majority (90%) of NHP species tested is infected with SIV, many of the remaining species can be expected to harbor SIV infections. In addition, the majority of SIV lineages known to date were discovered by HIV screening or confirmatory assays, which are based on cross-reactivity with HIV-1 and/or HIV-2 antigens. Because the extent of this cross-reactivity is not known, SIV infection is most likely underestimated. To increase sensitivity and chances to uncover new SIV lineages, Elisa's using SIV-specific antigens were developed and the Multiple Analyte Profiling technology (xMAP) has also been adapted to allow a single sample to be tested simultaneously against antigens from a wide diversity of SIV lineages (Simon et al., 2001; Aghokeng et al., 2006, 2010; Ahuka-Mundeke et al., 2011). This latter approach reduces workload and limits the amount of biological material which is often only available in very low quantities (Ahuka-Mundeke et al., 2011). Using these SIV lineage-specific antibody assays new SIV lineages have been identified and it has been shown that prevalence rates can differ significantly among species (0% to >40%), but also within species according to the sampling site (Aghokeng et al., 2010; Ahuka-Mundeke et al., 2011).

Interestingly, only Old World primates are infected with SIVs, and only those from the African continent. No SIVs have been identified in Asian primates species but it has to be noted that studies on Asian and New World primates are scarce (Ayouba et al., 2013a; Locatelli and Peeters, 2012).

The widespread presence of SIVs in numerous African NHP suggest that SIV is very old. A recent report studied SIVs in NHP from Bioko Island, Equatorial Guinea, and related species on the continent. This study suggests that SIVs evolved independently since Bioko became isolated from the African mainland 10,000 to 12,000 years ago with the elevation of the sea levels. Molecular clock analysis which used the date of the separation of Bioko Island to calibrate showed that SIVs have been present in African primates for more than 32,000 years (Worobey et al., 2010). More recently, a study on SIVagmVer diversity in vervets on the two sides of the Drakensberg Mountain in South Africa estimated that the origin of SIVs could be retraced to 800,000−2,500,000 years (Ma et al., 2013). Moreover, the presence of endogenous "prosimian" viruses in lemurs suggests that they have been infecting primates for millions of years but most likely after the divergence of African and Asian monkeys (Gifford et al., 2008, 2009).

2.3 Pathogenicity of SIVs in Their Natural Hosts

Although SIVs are called immune deficiency viruses, these viruses generally do not induce an AIDS-like disease in their natural hosts, suggesting that they evolved with their hosts over an extended period of time (Silvestri et al., 2007). This absence of pathogenicity has been extensively studied in captive but naturally infected sooty mangabeys and African green monkeys. Their life span, as well as their immunological system, do not seem to be affected by SIV, despite the presence of high viral loads in blood and tissues (Keele et al., 2009; Liovat et al., 2009; Silvestri et al., 2003). However, some cases of AIDS have been described in captive monkeys, but mainly at an age which is not reached in their natural habitats, for examples, two mandrills developed AIDS, after 17 years of SIVmnd infection and a sooty mangabey progressed to AIDS after 18 years (Ling et al., 2004). An extensive analysis of the natural history of SIVagm infection in wild populations of sabaeus monkeys, the African green monkey species endemic to West Africa, and on vervet monkeys, the species endemic to south Africa, showed also that SIV infection has no impact on the immune system in their natural hosts, confirming that SIVagm is not pathogenic for Agrican green monkeys in their natural environment (Ma et al., 2013, 2014).

Nevertheless, the paradigm of nonpathogenicity has been challenged recently by observations on wild chimpanzees (*P. t. schweinfurthii*) from the Gombe National Park in Tanzania (Keele et al., 2009). SIVcpz*Pts* infection was associated with a 10- to 16-fold age-corrected increased risk of mortality and reduced fertility in terms of average number of births and survival of newborns among infected females. Moreover, retrospective analysis of tissues conserved from dead SIVcpz*Pts*-infected animals showed that they have also symptoms of immune deficiency. Similarly, a young naturally SIV-infected *P. t. troglodytes* chimpanzee confiscated in Cameroon in 2003, subspecies in which the ancestors of HIV-1 have been documented, developed an AIDS-like disease. Clinical follow-up and biological analysis over a 7 years period showed a significant decline of CD4 counts, a severe thrombocytopenia, a weight loss and frequent periods of infections with diverse pathogens. Finally, the animal died at age of 10 years (Etienne et al., 2011).

An in-depth demographic and epidemiologic study of the habituated chimpanzee populations (13% SIVcpz infected) from Gombe National Park together with those from a neighboring unhabituated

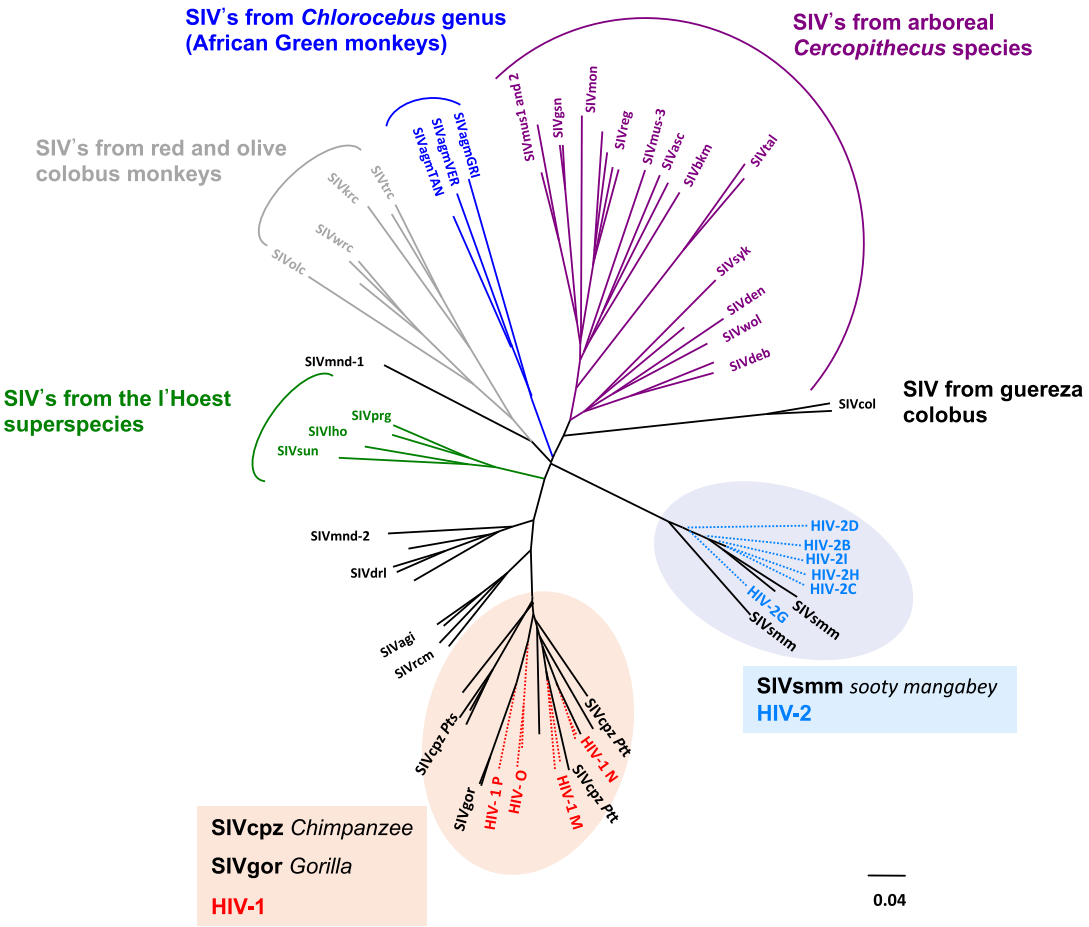


FIGURE 29.2 Genetic diversity and evolutionary history of the different HIV/SIV lineages.

Phylogenetic tree analysis using the neighbor-joining method on a 294 bp fragment from the *pol* gene of different SIVs infecting nonhuman primates and HIVs infecting humans. Branch lengths are drawn to scale (the scale bar indicates 0.04 substitutions per site). The different colors are used for clarity to discriminate the different SIV lineages. The different HIV-1 (*red*) and HIV-2 (*blue*) lineages are interspersed with the SIVcpz/SIVgor (*red circle*) lineages from chimpanzees and gorillas, and SIVsmm (*blue circle*) lineages from sooty mangabeys, respectively. The correspondence between the SIV lineages and their natural hosts are shown in Table 29.1. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

community (Kalande) showed a catastrophic population decline in the highly SIVcpz infected (46%) Kalande chimpanzee population (Rudicell et al., 2010). Mathematical models applied on these chimpanzee populations showed that SIVcpz infection can cause serious population decline. Models showed also that depending on the population structure and transmission dynamics, SIVcpz may be more likely to go extinct than its chimpanzee host and intercommunity migration increases survival of

infected populations. These results are in line with the unequal distribution of SIVcpz in Africa and may explain how chimpanzees as a species have survived this pathogen.

The fact that such a major effect of SIVcpz infection went undetected for decades supports the need for more studies on the natural history of SIV infection in African NHP in their native habitat, and not in captive environments where health status is controlled, nutrition is monitored and with limited exposure to infectious agents.

2.4 Evolution and Phylogeny of SIVs

The genetic diversity among NHP lentiviruses is extremely complex. There are many examples of co-evolution between viruses and their hosts, but recombination between distant SIVs is not exceptional. Phylogenies for some SIVs and their hosts suggest co-evolution over long time periods (Fig. 29.2). This is the case for the four different African green monkey species from the *Chlorocebus* genus that live in geographically separate and nonoverlapping areas across Africa. Each *Chlorocebus* species is infected with a specific SIV, e.g., SIVagmVER in vervets, SIVagmGRI in grivets, and SIVagmTAN and SIVagmSAB in tantalus and sabaeus monkeys, respectively (Wertheim and Worobey, 2007). The SIVs from the l'hoesti superspecies (e.g., SIVlho from *C. lhoesti*, SIVsun from *C. solatus* and SIVprg from *C. preussi*) (Beer et al., 1999) and SIVs form arboreal *Cercopithecus* species form also each separate clusters in the phylogenetic tree of SIVs (Bibollet-Ruche et al., 2004). However, the recent characterization of a new SIVagmTAN strain form a tantalus monkey in Cameroon showed a mosaic structure between SIVagmSAB from sabaeus monkeys from West Africa and SIVagmTAN from tantalus monkeys from Central Africa suggesting that the evolution of SIVagm in the *Chlorocebus* genus is more complex than previously thought (Ayouba et al., 2015; Augier et al., 2020).

Nevertheless, cross-species transmissions could also give erroneous impressions of co-evolution, especially when chances for efficient host switch are higher among genetically closely related species (Charleston and Robertson, 2002). There are indeed numerous examples of cross-species transmissions of SIVs between primates living in the same habitats or in polyspecific associations. For example, SIVagm from African green monkeys has been transmitted to Patas monkeys in West Africa (Bibollet-Ruche et al., 1996) and to yellow and chacma baboons in South Africa (Jin et al., 1994; van Rensburg et al., 1998). There are also more complex examples of cross-species transmissions of SIVs between greater spot-nosed monkeys (SIVgsn) and mustached monkeys (SIVmus), followed by recombination as seen for SIVmus-2 in mustached monkeys in Cameroon (Aghokeng et al., 2007) or SIVmus-3 in Gabon (Liégeois et al., 2014) (Table 29.1). One of the most striking examples of cross-species transmission, followed by recombination is SIVcpz in chimpanzees. The 5′ region of SIVcpz is most similar to SIVrcm from red-capped mangabeys, and the 3′ region is found to be closely related to SIVgsn from greater spot-nosed monkeys (Bailes et al., 2003). However, additional analyses including more SIV genomes from a larger range of NHP species, suggests that part of 5′ region cannot be clearly classified to any of the currently known SIVs lineages and that maybe three SIV lineages are involved in the genetic structure of SIVcpz (Bell and Bedford, 2017). Chimpanzees hunt various monkey species for food and have most likely been exposed to SIVs at multiple occasions, but only a single hybrid virus has spread among chimpanzees, which was subsequently transmitted to gorillas and humans (D'arc et al., 2015; Takehisa et al., 2009; Van Heuverswyn et al., 2006). Some NHP are infected with more than one SIV lineage, often as a result of cross-species transmission and recombination, e.g., SIVmnd-1 in mandrills from Southern Gabon, and SIVmnd-2 in animals living in

Northern Gabon and Cameroon or SIVmus in mustached monkeys in which 3 different variants have been described (Souquière et al., 2001; Liégeois et al., 2014).

As numerous cross-species transmissions among different primate species occurred, both HIV-1 and HIV-2 in humans are also the results of cross-species transmissions of SIVs from African primates (Hahn et al., 2000). The closest simian relatives of HIV-1 are SIVcpz and SIVgor, in chimpanzees (*Pan troglodytes troglodytes*) and gorillas (*Gorilla gorilla gorilla*), respectively, from West Central Africa, and SIVsmm in sooty mangabeys (*Cercocebus atys*) from West Africa are the closest relatives of HIV-2.

3. HIV-1 Is Derived From SIVs Circulating Among African Apes

Based on phylogenetic analyses of numerous isolates obtained from diverse geographic origins, HIV-1 is classified into four groups, M, N, O, and P (Fig. 29.1). Group M (for Major), discovered in 1983 (Barré-Sinoussi et al., 1983), represents the vast majority of HIV-1 strains found worldwide and is responsible for the global pandemic. Group O (for Outlier), described in 1990 (De Leys et al., 1990), remained restricted to West Central Africa, and represents currently less than 1% of HIV-1 infections in Cameroon and is estimated to have infected a cumulative number of 100,000 individuals (Ayouba et al., 2001; Kluge et al., 2014; Villabona-Arenas et al., 2015). Group N and group P, described in 1998 and 2009, respectively, have only been observed in a handful of patients, all from Cameroon except one HIV-1 N case that was identified in a patient from Togo (Delaugerre et al., 2011; Mourez et al., 2013). Each HIV-1 group corresponds to an independent cross-species transmission of SIVs from apes to humans.

3.1 SIVs From Chimpanzees and Gorillas are the Ancestors of HIV-1 in Humans

The first SIVcpz strains have been isolated from two captive but wild-born chimpanzees in Gabon more than 30 years ago, SIVcpzGab1 and SIVcpzGab2 (Peeters et al., 1989). Genetic analysis of the SIVcpzGab1 genome revealed the presence of the accessory gene, *vpu*, also identified in HIV-1. Furthermore, phylogenetic analysis showed that SIVcpzGab1 was more closely related to HIV-1 than to any other SIV (Huet et al., 1990). Characterization of a third SIVcpz, SIVcpzANT, showed an unexpected high degree of divergence among the chimpanzee viruses (Peeters et al., 1992; Vanden Haesevelde et al., 1996). Subsequent subspecies identification of the chimpanzee hosts revealed that the SIVcpzANT strain was isolated from a member of the *P. t. schweinfurtii* subspecies, whereas the other chimpanzees belonged to the *P. t. troglodytes* subspecies (Gao et al., 1999). These findings suggested two distinct SIVcpz lineages according to the host subspecies: SIVcpz*Ptt* and SIVcpz*Pts* from central (*P. t. troglodytes*) and eastern chimpanzees (*P. t. schweinfurtii*), respectively. All HIV-1 groups were more closely related to SIVcpz*Ptt* than to SIVcpz*Pts* (Fig. 29.3). Although these data pointed already to the West Central African chimpanzees (*P. t. troglodytes*) as the natural reservoir of the ancestors of HIV-1, the SIVcpz reservoirs in wild-living apes which are at the origin of HIV in humans still needed to be identified. The major issue in studying SIVcpz infection in wild chimpanzees is their endangered status and the fact that they live in isolated forest regions. In addition, all previously studied chimpanzees were wild-caught, but they were captured as infants, which do not reflect true prevalences among wild living adult animals. The development of noninvasive methods to detect and

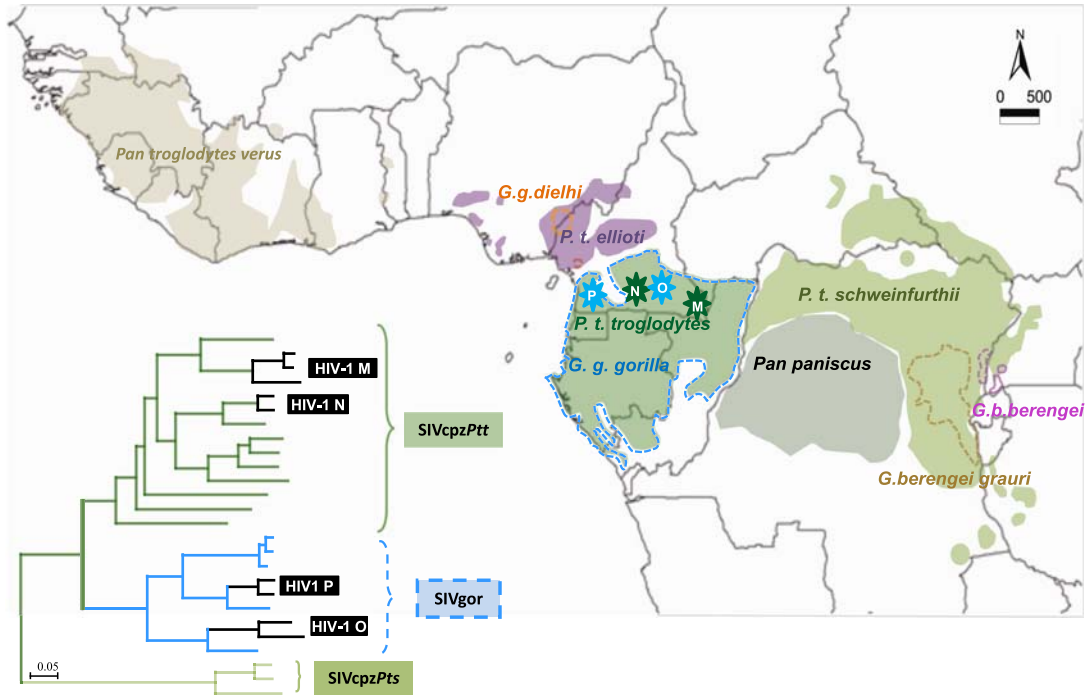


FIGURE 29.3 HIV-1 is derived from SIVs circulating in chimpanzees and gorillas from West Central Africa.

Evolutionary relationship of SIVcpz*Pts* (*olive green*), SIVcpz*Ptt* (*green*), SIVgor (*blue*), and HIV-1 group M, N, O, and P (*black*) strains based on maximum likelihood phylogenetic analysis of partial *env* (gp41) sequences. Horizontal branch lengths are drawn to scale (the scale bar indicates 0.05 substitutions per site). The map represents the geographical range of the four gorilla and the four chimpanzee subspecies, and bonobos. Dots on the map indicate the ape populations infected with the ancestors of each HIV-1 group with the corresponding color of the chimpanzee or gorilla strains in the phylogenetic tree. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

characterize SIVcpz in fecal and urine samples boosted the search for new SIVcpz strains in wild ape populations in Africa (Santiago et al., 2002, 2003a). The first full-length SIVcpz sequence from a wild-infected chimpanzee, SIVcpzTan1, was obtained from a fecal sample from a *P. t. schweinfurtii* chimpanzee in Gombe National Park, Tanzania (Santiago et al., 2003b).

Subsequently, large-scale studies have been conducted across West, Central, and East Africa and today thousands of fecal samples have been collected from the four different chimpanzee subspecies. These studies showed that only the two subspecies from Central Africa are infected and confirmed also the presence of subspecies-specific SIVcpz*Ptt* and SIVcpz*Pts* lineages in *P. t. troglodytes* and *P. t. schweinfurthii*, respectively. Interestingly, phylogeographic clusters are also observed within the SIVcpz*Ptt* and SIVcpz*Pts* lineages. As such, the reservoirs of the ancestors of the pandemic HIV-1 group M were identified in chimpanzee communities in South-East Cameroon. Similarly, the ancestors of HIV-1 group N, have been identified in chimpanzees living in South-Central Cameroon (Fig. 29.3). The overall SIVcpz prevalence ranges between 10% and 13% in both chimpanzee

subspecies, but they are very heterogeneous and can vary from 0% to >40% in certain chimpanzee communities (Boué et al., 2015; Keele et al., 2006; Li et al., 2012; Van Heuverswyn et al., 2007; Worobey et al., 2004). Only SIVcpz*Ptt* strains have been transmitted to humans, despite the fact that both chimpanzee subspecies represent significant SIVcpz reservoirs. Although numerous samples have been tested, no SIV infection has been detected in the two other chimpanzee subspecies, *P. t. elioti* (previously *P. t. vellerosus*) and *P. t. verus* (Prince et al., 2002; Switzer et al., 2005; Van Heuverswyn et al., 2007).

While the reservoirs of the ancestors of HIV-1 M and N were identified in 2006, the reservoirs of group O and P remained unknown. In 2006, SIV infection was described for the first time in western gorillas (*Gorilla gorilla gorilla*) in Cameroon (Van Heuverswyn et al., 2006). Surprisingly, the newly characterized gorilla viruses, termed SIVgor, formed a monophyletic group within the HIV-1/SIVcpz*Ptt* radiation, but in contrast to SIVcpz*Ptt*, they were most closely related to HIV-1 group O and P (Fig. 29.3) (Takehisa et al., 2009; Van Heuverswyn et al., 2006). Moreover, phylogenetic relationships between SIVcpz, SIVgor, and HIV-1 indicate that chimpanzees represent the original reservoir of SIVs now found in chimpanzees, gorillas, and humans (Fig. 29.3). Initially it was not clear how the virus has been transmitted from chimpanzees to gorillas because physical encounters between the two species, like biting or another contact with infected blood or body fluids, have not been reported. Importantly, in 2021, two lethal coalitionary attacks of chimpanzees (*Pan troglodytes troglodytes*) on gorillas (*Gorilla gorilla gorilla*) have been reported for the first time in the Loango National Park, Gabon (Southern et al., 2021). In addition, chimpanzees and gorillas have overlapping habitats and often feed in the same fruit trees (Head et al., 2011; Morgan and Sanz, 2006; Stanford and Nkurunungi, 2003), which leads to direct and indirect contacts, that resulted in the cross-species transmission of other pathogens, e.g., Ebola and hepatitis B (Lyons et al., 2012; Walsh et al., 2007).

Today thousands of fecal samples have been studied from all gorilla species and subspecies, e.g., western lowland (*G. gorilla gorilla*) and cross river (*G. g. diehli*) gorillas in West-Central Africa, and eastern lowland (*G. beringei graueri*) and mountain (*G. b. beringei*) gorillas in East-Central Africa (D'arc et al., 2015; Neel et al., 2010). Among the 55 different sites studied across Central Africa, covering a large proportion of the geographic range of gorillas, SIVgor infection was only identified in six sites all located in Southern Cameroon and only in western lowland gorillas (D'arc et al., 2015; Neel et al., 2010). The global prevalence of SIVgor (<2%) is lower than the prevalence observed in SIVcpz infection from chimpanzees. Similarly as for chimpanzees, an unequal distribution is observed, ranging from 0% to >30% (D'arc et al., 2015; Neel et al., 2010). Despite the low prevalence, the genetic characterization of SIVgor strains from different field locations showed a high genetic diversity and phylogeographic clustering. As such, HIV-1 group P viruses fall within the radiation of the SIVgor strains from a gorilla community in South-West Cameroon, strongly suggesting the origin of HIV-1 P in this region of Western Cameroon. Finally, sequence analysis of an SIVgor strain from South-Central Cameroon resolved the origin of HIV-1 group O (D'arc et al., 2015). Like chimpanzees, gorillas became infected more recently with SIV than monkeys, and it is thus important to study the impact of SIVgor infections on survival and pathogenicity. A pilot study using noninvasive surveillance of SIVgor in western lowland gorilla's in southern Cameroon over more than a decade showed inter and intragroup viral transmission and association of specific mammalian viral families and SIVgor in a dysbiosis context (Villabona-Arenas et al., 2018; D'Arc et al., 2018).

It is now clear that central chimpanzees (*P. t. troglodytes*) are the reservoirs for the pandemic HIV-1 group M strain and HIV-1 N, and western lowland gorillas (*G. g. gorilla*) are the reservoirs for HIV-1

groups O and P. The origin of all HIV-1 groups has now been solved and today we know that two ape species are involved in the origin of HIV-1 in humans, each species transmitted SIV to humans on at least two occasions.

Importantly, the finding that SIVcpz strains from East African chimpanzees, including those from Kisangani in DRC, are more distantly related to HIV-1 provides also evidence that the Oral Polio Vaccine (OPV), which was largely distributed in this part of Africa at the end of the 1950s, is not at the origin of the HIV-1 epidemic (Worobey et al., 2004). It has been suggested that tissues derived from SIVcpz-infected chimpanzees, captured in the Northeastern part of DRC were used for the OPV production. However, this geographical region is situated in the middle of the *P. t. schweinfurtii* range and the characterization of SIVcpz*Pts* from wild chimpanzees in DRC proved once more the inconsistency of the OPV theory (Li et al., 2012; Worobey et al., 2004) (Fig. 29.3).

3.2 The Cross-Species Transmissions Resulting in HIV-1 Viruses in Humans Occurred in West Central Africa

Since the four groups of HIV-1 fall within the HIV-1/SIVcpz*Ptt*/SIVgor radiation, the cross-species transmissions giving rise to HIV-1 occurred in Western Equatorial Africa, the home of central chimpanzees (*P. t. troglodytes*) and western lowland gorillas (*G. g. gorilla*). Furthermore, no human counterpart is found for SIVcpz*Pts* from *P. t. schweinfurtii*. The recent studies in wild chimpanzee and gorilla communities in Cameroon, not only strengthen the evidence of the West Central African origin of HIV-1 but also indicate that the four groups of HIV-1 arose from geographically distinct chimpanzee and gorilla populations in Southern Cameroon (Fig. 29.3). This coincides with the geographical area of group N, O, and P infections, which remain actually mainly restricted to Cameroon (Brennan et al., 2008; Mourez et al., 2013; Vessière et al., 2010).

The four HIV-1 groups have thus their origin in West Central Africa, but only one, HIV-1 group M, has spread across Africa and all the other continents. Moreover, the reservoir of the ancestors of HIV-1 M has been identified at almost 1000 km distance from the epicenter of the HIV-1 epidemic in DRC (Vidal et al., 2000; Worobey et al., 2008). A combination of several factors (viral, host, socioeconomic, demographic, etc.) are thus most likely involved in the subsequent efficient spread of HIV-1 M.

3.3 HIV-1 Started to Diverge in the Human Population at the Beginning of the 20th Century

As mentioned above, the SIVcpz*Ptt* ancestors of group M have been identified in Cameroon, but the highest genetic diversity of HIV-1 M, in number of co-circulating subtypes and intrasubtype diversity, has been observed in the Western part of DRC, suggesting this region may be the epicenter of HIV-1 group M (Vidal et al., 2000). Retrospective studies showed that 20 years before the AIDS epidemic was recognized in the United States, HIV-1 M (subtype A, C and D) circulated already in humans in Kinshasa; e.g., HIV-1 subtype D was identified in a serum from 1959 (Zhu et al., 1998), subtype A in a biopsy from 1960 (Worobey et al., 2008), and subtype C in a biopsy from 1966 (Gryseels et al., 2020). Similarly, HIV-1 group O was identified in a Norwegian sailor who became infected in the 1960s (Jonassen et al., 1997). The most recent common ancestors of HIV-1 groups M and O are both estimated around 1920 (Faria et al., 2014; Wertheim and Worobey, 2009; Gryseels et al., 2020). The low spread and lower levels of diversity seen in group N and especially in group P indicate that they most

likely emerged more recently. The most recent common ancestor of HIV-1 group N is estimated around 1963 with a confidence interval of 1948–77 (Wertheim and Worobey, 2009). Only two HIV-1 P infected patients have been identified today and estimates on dates are thus uncertain. Coalescent studies suggest that groups O and M underwent similar rates of exponential growth until about 1960, after which group M undergoes accelerated epidemic growth (Faria et al., 2014).

3.4 SIVs are Transmitted to Humans by Exposure to Infected Primates

Although the conditions and circumstances of cross-species from SIVs from primates to humans remain unknown, human exposure to blood or other secretions of infected primates, through hunting and butchering of primate bushmeat, represents the most plausible source for human infection. Also, bites and other injuries caused by primates kept as pet animals can favor a possible viral transmission. However, factors associated with single cross-species transmission have to be differentiated from those associated with the epidemic spread, the latter being a combination of viral, host, and environmental factors.

4. Origin of HIV-2: An Other Emergence, an Other Epidemic

Two independent studies in 1989 and 1992 confirmed the homologies between HIV-2 and SIVsmm infecting sooty mangabeys in West Africa (Gao et al., 1992; Hirsch et al., 1989). Sooty mangabeys are home to West Africa, from Senegal to Ivory Coast, coinciding with the endemic region of HIV-2 (Fig. 29.4). In contrast to HIV-1, HIV-2 remained restricted to West Africa where it infects one to two million people. Today, HIV-2 prevalences are decreasing, since HIV-1 M is now also predominating in West Africa (de Silva et al., 2008; Tienen et al., 2010; Visseaux et al., 2016). The highest HIV-2 prevalences have been observed in Guinea-Bissau and Southern Senegal (Casamance area). Overall, HIV-2 is less pathogenic, less transmissible with almost absence of mother to child transmission, and less efficient sexual transmission most likely related to lower viral loads (Gottlieb et al., 2006; Hawes et al., 2008). However, a high genetic diversity is seen among HIV-2 strains, nine groups (A–I) of HIV-2 have been described so far (Ayouba et al., 2013b; Damond et al., 2004) (Fig. 29.4), each corresponding to a cross-species transmission. Only groups A and B are largely represented in the HIV-2 epidemic, with group A circulating in the Western part of West Africa (Senegal, Guinea-Bissau) and group B being predominant in Ivory Coast. To date, no subtypes have been formally described but preliminary data suggest that HIV-2 group A may be divided in two distinct subtypes with distinct geographical origins (Visseaux et al., 2016), one subtype is associated to HIV-2 epidemics in Senegal, Gambia, Guinea-Bissau, and Guinea and the other is associated to Mali and Ivory Coast (Visseaux et al., 2016). The MRCA of the remaining groups cannot be estimated by phylogenetic approaches as only one sequence is available. The ancestors of the HIV-2 group A and B viruses have been identified in wild sooty mangabey populations in the Tai forest in Ivory Coast, close to the border with Liberia (Santiago et al., 2005). The other HIV-2 groups have been documented in one or two (group F) individuals only. Except for group G and H, group C, D, E, F, and I were isolated in rural areas in Sierra Leone, Liberia, and Ivory Coast, and these viruses are in general more closely related to the SIVsmm strains obtained from sooty mangabeys found in the same area than to any other HIV-2 strain. Molecular clock analysis traced the origin of the epidemic HIV-2 groups A and B to be around 1932 (1906–55) and 1935 (1907–61), respectively (Wertheim and Worobey, 2009). These dates seem to

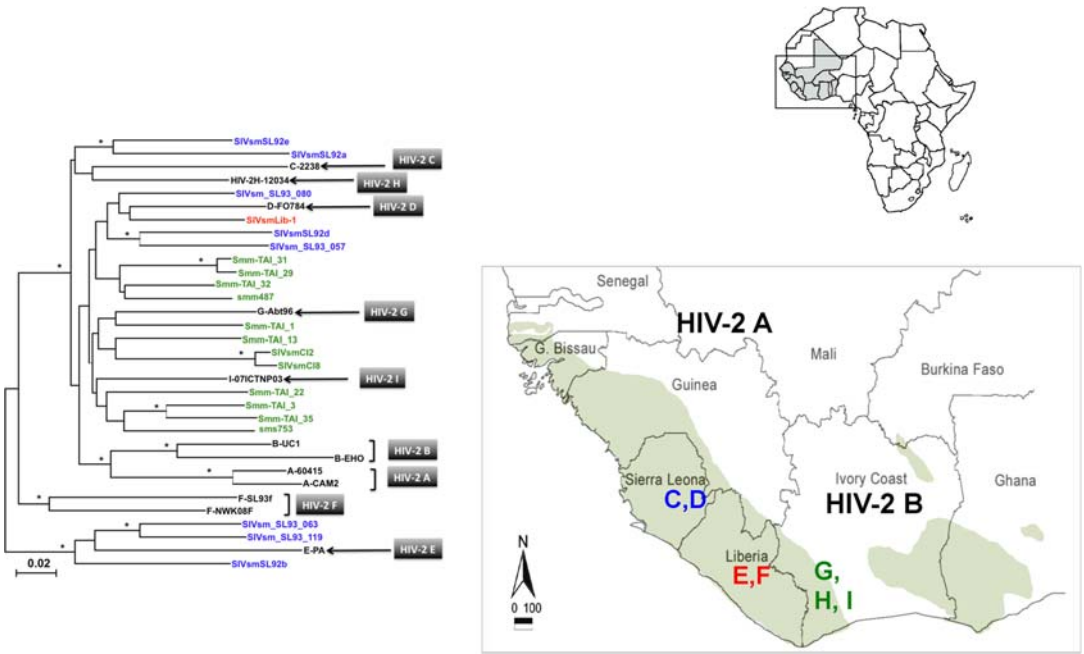


FIGURE 29.4 HIV-2 is derived from SIVs circulating in sooty mangabeys from West Central Africa.

Evolutionary relationship of HIV-2 groups A to I (*black*) and SIVsmm. SIVsmm strains obtained from mangabeys in different regions are colored: Green for Ivory Coast, blue for Sierra Leone, and red for Liberia. Letters on the map indicate HIV-2 groups, HIV-2 A and B are epidemic in the western and eastern part of West Africa, respectively, the other groups have been identified in single individuals and are colored according to the country of origin. The phylogenetic tree was derived by neighbor-joining analysis of partial (692bp) *gag* nucleotide sequences from HIV-2 and SIVsmm sequences from West Africa. Asterisks represent nonparametric bootstrap branch support ≥70%. Horizontal branch lengths are drawn to scale. The geographic range of sooty mangabeys (*Cercocebus atys*) is highlighted in *green*. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

coincide with a political unstable period in Guinea-Bissau, and it has been suggested that civil wars at that time amplified the rapid spread of HIV-2 into the human population (Lemey et al., 2003).

5. Ongoing Exposure of Humans to a Large Diversity of SIVs: Risk for a Novel HIV?

5.1 Exposure to SIV Infected NHP is Still Ongoing

Bushmeat hunting, as a source of animal proteins and income for the family, is a longstanding common component of rural households in the Congo Basin, and more generally throughout sub-Saharan Africa. However the bushmeat trade has increased significantly during the last decades, due to the expanding logging industry and the increasing demand of bushmeat in cities (Laporte et al., 2007). Indeed, a recent study showed that in Northern Congo, the bushmeat trade increased together with the

increasing presence of logging concessions in the same area (Poulsen et al., 2009). Commercial logging has led to road constructions penetrating remote forest areas, to human migration, and the development of social and economic networks (including those of sex workers), which support this industry. Furthermore, villages around logging concessions have grown from a few hundred to several thousand inhabitants in just a few years, and the number of people entering previously inaccessible forest areas increased significantly over the last decades. Importantly, the HIV prevalences in these rural settings are also increasing, for example around 20% of women aged between 15 and 35 years are HIV infected in rural forest areas in Southern Cameroon (Chia et al., 2014; Laurent et al., 2004). A high number of individuals with immune deficiency are thus potentially exposed to new viruses, and recombination between newly introduced SIVs and circulating HIVs can pose an additional risk for the outbreak of a novel HIV epidemic (Chia et al., 2014; Laurent et al., 2004; LeBreton et al., 2007). In addition, to high HIV prevalences in rural forest areas in Cameroon, where chimpanzee and gorilla populations infected with the ancestors of the different HIV-1 groups have been identified and transmitted to humans during the 20th century, almost 20% of HIV-1 strains form recent transmission chains composed of two or three sequences but one cluster included 10 HIV-1 strains from women living in 4 different villages on a major road for logging concessions in the south-east (60 km distance) (Edoul et al., 2020). Many factors allowing rapid establishment in the human population and subsequent rapid spread to urban areas of a new retrovirus or other pathogens of zoonotic origin are now present.

Studies on primate bushmeat in West and Central Africa showed that bushmeat hunting is not limited to chimpanzees and mangabeys, but the majority of hunted primates is represented by multiple *Cercopithecus* species, colobus monkeys, mandrills, drills, etc. Moreover, these data revealed ongoing exposure of humans to a plethora of different SIVs (Aghokeng et al., 2010; Ahuka-Mundeke et al., 2011; Liégeois et al., 2012; Mossoun et al., 2015; Peeters et al., 2002). Cross-species transmissions with SIV strains from other primates to humans should thus be considered. It has also to be noted that apes are not only hunted for bushmeat but also medicinal uses (Meder, 1999). The socioeconomic changes, which go together with the presence of logging or other industries in remote areas, combined with the SIV prevalence and genetic complexity in wild living primates suggest that the magnitude of human exposure to SIV has increased, as have the social and environmental conditions that support the emergence and spread of new zoonotic infections.

5.2 SIV Prevalences and Cross-Species Transmissions

The chances for cross-species transmissions most likely increase when the frequency of exposure and SIV prevalences are high. A recent study in pet monkeys in Cameroon revealed that SIV prevalence in mandrill pets could reach 23% (Ndembi et al., 2009). We recently showed among more than 2500 samples in Cameroon and by using SIV lineage−specific Elisas that about 3% of primate bushmeat is SIV infected, but prevalences can vary from 0% to 10% depending on geographic localities, or from 0% to 40% according to species (Aghokeng et al., 2010). Interestingly, the lowest prevalences (0%−1%) were observed among the most frequently hunted species, reducing thus the risk for cross-species transmissions. However, this situation can be different in other geographic areas, for example, in the DRC a pilot study showed that 20% of the primate bushmeat are infected with the highest prevalences among the most frequently hunted monkeys (Ahuka-Mundeke et al., 2011). Studies in Cameroon showed also that SIVcpz and SIVgor prevalences in chimpanzee communities infected with

the ancestors of HIV-1 M and N and gorillas communities infected with the ancestor of HIV-1 P are among the highest, i.e., around 30% (D'arc et al., 2015; Keele et al., 2006; Van Heuverswyn et al., 2007).

Moreover, in West Africa, the SIVsmm prevalences of wild mangabeys are around 50%, and at least nine cross-species from mangabeys to humans occurred. In this same region, 50%−80% of western red colobus monkeys are also infected with SIVs (Leendertz et al., 2010; Locatelli et al., 2008) and are together with mangabeys highly represented among primate bushmeat, but no SIVwrc cross-species transmission to humans has been documented yet. Western red colobus monkeys represent also 80% of animal proteins among chimpanzees from the *P. t. verus* subspecies in the Tai forest in Ivory Coast, and again no SIVwrc infection could be identified in this chimpanzee subspecies (Leendertz et al., 2011).

5.3 Host Restriction Factors

The initial adaptation of SIV to chimpanzees and gorillas most likely played a role in the successful virus transmission to humans, for example, the Vif proteins of SIVcpz and SIVgor are effective antagonists of human APOBEC3D, F, and G proteins (Etienne et al., 2013; Nakano et al., 2020; Zhang et al., 2017). Several retroviral restriction factors have been identified in humans, APOBEC3G induces lethal hypermutations in the retroviral genome; Trim5alpha proteins restrict the incoming viral capsid; tetherin inhibits the release of viral particles (Kirchhoff, 2010); and SAMHD1 is an antiretroviral protein expressed in cells of the myeloid lineage that inhibits an early step of the viral life cycle (Laguette et al., 2011). A recent study showed the role of the APOBEC3G host restriction factor to explain the absence of SIVwrc from red colobus monkeys in chimpanzees, as mentioned above (Etienne et al., 2015). Similarly, SIVgor resulted from a single introduction of SIVcpz from sympatric chimpanzees and functional analyses identified APOBEC3G as a barrier for virus transmission from chimpanzees to gorillas (D'arc et al., 2015).

A deletion in the cytoplasmic domain of human tetherin leads to resistance of SIVcpz and SIVgor Nef protein to counteract tetherin (Jia et al., 2009; Sauter et al., 2009; Zhang et al., 2009). Interestingly, the four HIV-1 groups of HIV-1 used different ways to deal with tetherin restriction. HIV-1M strains switched from Nef to Vpu, Vpu of HIV-1 N acquired modest anti-tetherin activity but lost the CD4 degradation function and HIV-1 O Nef proteins target a region adjacent to the protective deletion in human tetherin (Sauter et al., 2009, 2012; Kluge et al., 2014). HIV-1P evolved not yet an effective tetherin antagonist in humans (Sauter et al., 2011; Yang et al., 2011). Finally, recent findings showed that, in addition to these factors, chimpanzees' CD4 molecule diversity protects these apes against SIV infection (Bibollet-Ruche et al., 2019; Russell et al., 2021).

5.4 Viral Adaptation

The majority of the viruses infecting wild animals are not able to cross the species barrier, adapt to a new host, and spread into the new species. The viral and molecular characteristics that allowed the ancestors of HIV-1 and HIV-2 to cross and adapt to humans are not yet completely identified. More studies are needed to find out why, for example, SIVs from mangabeys, chimpanzees, and gorillas have been transmitted on multiple occasions and not those of other monkeys. Moreover, cross-species transmissions are not always followed by efficient spread into the new populations, as illustrated by

HIV-1 versus HIV-2 and among the different HIV-1 and HIV-2 groups, and depend thus not only on the virus but also on the host and the environment (Chastel, 2000; Lloyd-Smith et al., 2009).

Actually, only limited studies showed some viral adaptation, for example, at the Gag-30 position in the p17 region of the *gag* gene, methionine or leucine is present among SIVcpz/SIVgor, but in humans all HIV-1 strains harbor an arginine at this position (Wain et al., 2007) suggesting an adaptation of the virus to its new host. HIV-1 groups M and O have undergone independent adaptations to acquire resistance to the potent restriction factor tetherin; in group M viruses, Vpu adapted to acquire anti-tetherin activity (Sauter et al., 2009); and in group O, the Nef protein evolved to use a different target within tetherin (Kluge et al., 2014). Nef allows viral persistence in the host but controls also for superactivation of the immune system. However, this latter function is lost in certain SIV lineages and more precisely in the ancestors of the HIV-1/SIVcpz lineage which could thus have resulted in higher pathogenicity in humans, in contrast to SIVsmm/HIV-2 where this adaptation is not observed and which are less pathogenic (Schindler et al., 2006). It has been shown recently that passage of SIVs from monkeys through chimpanzees facilitated the subsequent adaptation of HIV-1 to humans, a series of gene loss and adaptation events that generated the chimpanzee precursor of HIV-1 and lowered the species barrier to human infection (Etienne et al., 2013). Another study showed a lower viral fitness for HIV-2 compared to HIV-1, and for HIV-1 O versus M, which could also partially explain the lower prevalence and limited spread of HIV-2 and HIV-1 O (Ariën et al., 2005). Despite progress in our knowledge on viral and host factors, more studies are needed to understand the global picture.

5.5 Human Factors

Human factors play also a major role in the epidemic spread of a new viruses, especially among viruses that are transmitted by blood or sexual contacts, as it is the case for HIV. The difference in the localizations of the origin of HIV-1 M (South Cameroon) and the origin of the epidemic (DRC, for HIV-1 M) illustrates the importance of human and environmental factors in the epidemic spread. The main factors involved in the human-to-human spread for HIV are sexual risk behavior, high prevalences of sexually transmitted infections, absence of circumcision, and transmission through unsafe blood transfusions and nonsterile needles. These factors associated with socio-demographic factors like human density in forest areas, increasing transport between urban and rural areas, human migration, urbanization, and increase in commercial sex were in favor of an epidemic spread of the virus.

The early spread of HIV-1 in the human population has been reconstructed recently with statistical phylodynamic models on large collections of sequence data from Central Africa (Faria et al., 2014). The study confirmed that the HIV-1 group M pandemic started in Kinshasa around the mid-1920s. The ancestor of HIV-1 M arrived most likely from Southern Cameroon in Kinshasa via the Sangha and Congo rivers during the early colonial period (Faria et al., 2014; Sharp and Hahn, 2008). The subsequent spatial expansion in Central Africa is associated with transportation networks, railway and fluvial. Iatrogenic interventions and/or changes in sexual behavior were critical for the pandemic growth of group M. A unique combination of circumstances during a particular spatial and socio-historical window allowed thus the establishment, spatial expansion and epidemic growth of HIV-1 group M to pandemic proportions. The fact that group O viruses have not spread even more widely in the human population could thus reflect the absence of epidemiological opportunity during the early stages of the pandemic expansion of AIDS 50 years ago (Faria et al., 2014).

5.6 Ongoing Cross-Species Transmissions From Other Retroviruses From Primates to Humans

Simian Foamy Virus (SFV) are infecting primates at high levels: 70% of primates in captivity (Hussain et al., 2003), 97% of wild western red colobus monkeys (Goldberg et al., 2009), and between 44% and 100% of wild chimpanzees (Liu et al., 2008). SFVs have been documented in primate care workers in the United States, hunters in Cameroon and Gabon and in women living in rural DRC (Betsem et al., 2011; Mouinga-Ondémé et al., 2012; Switzer et al., 2004, 2012; Wolfe et al., 2004). Moreover, SFV prevalence can reach 25% in persons who reported ape bites (Calattini et al., 2007). Recent epidemiological studies in New World Primates (NWP) also showed a wide distribution of distinct SFVs, highlighting the risk of potential zoonotic transmission in this area which is still poorly studied (Muniz et al., 2013, 2015). However, human-to-human transmission or an SFV epidemic has not been documented yet, and no pathogenicity associated with SFV infection has been observed in humans. In this example, cross-species transmissions were thus possible, but the viral adaptation was insufficient to allow a spread of the virus into the new host. Nevertheless, a recent case−control study compared 24 Cameroonian hunters infected with gorilla SFV and 24 controls matched for age and ethnicity. The results showed that, showed significant differences in some blood markers (Buseyne et al., 2018). Hence, more studies are still necessary. Cases and controls had similar frequencies of general, cutaneous, gastrointestinal, neurological, and cardiorespiratory signs.

Exposure to nonhuman primates allowed also the emergence of four types of human T-lymphotropic virus (HTLV) type 1 to 4 in humans, all of them with the simian counterparts (STLV-1−4) already identified (LeBreton et al., 2014; Switzer et al., 2009; Wolfe et al., 2005). Furthermore, several studies among hunters in Central Africa showed ongoing cross-species transmissions of STLVs (Calvignac-Spencer et al., 2012; Sintasath et al., 2009; Zheng et al., 2010). These T-lymphotropic viruses are thus able to cross the species barrier, and certain variants were able to spread into the human population leading to HTLV-1 and HTLV-2, which are endemic in certain parts of the world.

The ongoing transmissions of SFV and STLV highlight the risk for potential emergence of a new SIV into the human population. The recent discoveries in 2009 of a new HIV-1, group P, and in 2013 of a new HIV-2, group I, are additional demonstrations of ongoing zoonotic transmissions (Ayouba et al., 2013b; Plantier et al., 2009; Vallari et al., 2011).

6. Conclusion

Today, we have a clear picture of the origin of HIV and the seeds of the AIDS pandemic. The current HIV-1 epidemic provides evidence for the extraordinary impact that can result from a single primate lentiviral zoonotic transmission event. Despite the fact that the first AIDS cases have been observed around 1980 in the United States, the virus circulated already early in the 20th century in the human population in Central Africa (Fig. 29.1). Currently, at least, 13 SIV cross-species transmissions occurred, nine for HIV-2 and four for HIV-1, but most likely several others occurred in the past which remained unrecognized since some viruses were not able to adapt to the new host or the environment was not suitable for epidemic spread. Because humans are still exposed to a plethora of primate lentiviruses through hunting and handling of primate bushmeat, the possibility of additional cross-species transfers of primate lentiviruses from other primate species in addition to chimpanzees, gorillas, or sooty mangabeys has to be considered. One major public health implication is that these SIVs

strains are not always recognized by commercial HIV-1/HIV-2 screening assays. As a consequence, due to the long incubation period, human infection with such variants can go unrecognized for several years and lead to another epidemic. Also, social and environmental factors associated with epidemic spread are now present in rural forest areas. Identification of SIVs in wild primates can serve as sentinels by signaling which pathogens may be a risk for humans and allow the development of serological and molecular assays to detect transmissions with others SIVs in humans.

References

Aghokeng, A.F., Liu, W., Bibollet-Ruche, F., Loul, S., Mpoudi-Ngole, E., Laurent, C., Mwenda, J.M., Langat, D.K., Chege, G.K., McClure, H.M., Delaporte, E., Shaw, G.M., Hahn, B.H., Peeters, M., 2006. Widely varying SIV prevalence rates in naturally infected primate species from Cameroon. Virology 345, 174−189.

Aghokeng, A.F., Bailes, E., Loul, S., Courgnaud, V., Mpoudi-Ngolle, E., Sharp, P.M., Delaporte, E., Peeters, M., 2007. Full-length sequence analysis of SIVmus in wild populations of mustached monkeys (*Cercopithecus cephus*) from Cameroon provides evidence for two co-circulating SIVmus lineages. Virology 360, 407−418.

Aghokeng, A.F., Ayouba, A., Mpoudi-Ngole, E., Loul, S., Liegeois, F., Delaporte, E., Peeters, M., 2010. Extensive survey on the prevalence and genetic diversity of SIVs in primate bushmeat provides insights into risks for potential new cross-species transmissions. Infect. Genet. Evol. 10, 386−396.

Augier, C., Beyne, E., Villabona-Arenas, C.J., Mpoudi Ngole, E., Peeters, M., Ayouba, A., 2020. Identification of a novel Simian Immunodeficiency Virus-infected African Green Monkey (*Chlorocebus tantalus*) confirms that tantalus monkeys in Cameroon are infected with a mosaic SIVagm lineage. AIDS Res. Hum. Retrovir. 36, 167−170.

Ahuka-Mundeke, S., Ayouba, A., Mbala-Kingebeni, P., Liegeois, F., Esteban, A., Lunguya-Metila, O., Demba, D., Bilulu, G., Mbenzo-Abokome, V., Inogwabini, B.-I., Muyembe-Tamfum, J.-J., Delaporte, E., Peeters, M., 2011. Novel multiplexed HIV/simian immunodeficiency virus antibody detection assay. Emerg. Infect. Dis. 17, 2277−2286.

Apetrei, C., Kaur, A., Lerche, N.W., Metzger, M., Pandrea, I., Hardcastle, J., Falkenstein, S., Bohm, R., Koehler, J., Traina-Dorge, V., Williams, T., Staprans, S., Plauche, G., Veazey, R.S., McClure, H., Lackner, A.A., Gormus, B., Robertson, D.L., Marx, P.A., 2005. Molecular epidemiology of simian immunodeficiency virus SIVsm in U.S. primate centers unravels the origin of SIVmac and SIVstm. J. Virol. 79, 8991−9005.

Ariën, K.K., Abraha, A., Quiñones-Mateu, M.E., Kestens, L., Vanham, G., Arts, E.J., 2005. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. J. Virol. 79, 8979−8990.

Ayouba, A., Mauclère, P., Martin, P.M., Cunin, P., Mfoupouendoun, J., Njinku, B., Souquières, S., Simon, F., 2001. HIV-1 group O infection in Cameroon, 1986 to 1998. Emerg. Infect. Dis. 7, 466−467.

Ayouba, A., Duval, L., Liégeois, F., Ngin, S., Ahuka-Mundeke, S., Switzer, W.M., Delaporte, E., Ariey, F., Peeters, M., Nerrienet, E., 2013a. Nonhuman primate retroviruses from Cambodia: high simian foamy virus prevalence, identification of divergent STLV-1 strains and no evidence of SIV infection. Infect. Genet. Evol. 18, 325−334.

Ayouba, A., Akoua-Koffi, C., Calvignac-Spencer, S., Esteban, A., Locatelli, S., Li, H., Li, Y., Hahn, B.H., Delaporte, E., Leendertz, F.H., Peeters, M., 2013b. Evidence for continuing cross-species transmission of SIVsmm to humans: characterization of a new HIV-2 lineage in rural Côte d'Ivoire. AIDS 27, 2488−2491.

Ayouba, A., Njouom, R., Chia, J.E., Ahuka-Mundeke, S., Kfutwah, A., Ngole, E.M., Nerrienet, E., Delaporte, E., Peeters, M., 2015. Molecular characterization of a new mosaic Simian Immunodeficiency Virus in a naturally

infected tantalus monkey (*Chlorocebus tantalus*) from Cameroon: a challenge to the virus-host co-evolution of SIVagm in African green monkeys. Infect. Genet. Evol. 30, 65−73.

Bailes, E., Gao, F., Bibollet-Ruche, F., Courgnaud, V., Peeters, M., Marx, P.A., Hahn, B.H., Sharp, P.M., 2003. Hybrid origin of SIV in chimpanzees. Science 300, 1713.

Barin, F., M'Boup, S., Denis, F., Kanki, P., Allan, J.S., Lee, T.H., Essex, M., 1985. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of west Africa. Lancet 2, 1387−1389.

Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., Montagnier, L., 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220, 868−871.

Beer, B.E., Bailes, E., Goeken, R., Dapolito, G., Coulibaly, C., Norley, S.G., Kurth, R., Gautier, J.P., Gautier-Hion, A., Vallet, D., Sharp, P.M., Hirsch, V.M., 1999. Simian immunodeficiency virus (SIV) from sun-tailed monkeys (*Cercopithecus solatus*): evidence for host-dependent evolution of SIV within the *C. lhoesti* superspecies. J. Virol. 73, 7734−7744.

Bell, S.M., Bedford, T., 2017. Modern-day SIV viral diversity generated byextensive recombination and cross-species transmission. PLoS Pathog. 13, e1006466.

Betsem, E., Rua, R., Tortevoye, P., Froment, A., Gessain, A., 2011. Frequent and recent human acquisition of simian foamy viruses through apes' bites in central Africa. PLoS Pathog. 7, e1002306.

Bibollet-Ruche, F., Bailes, E., Gao, F., Pourrut, X., Barlow, K.L., Clewley, J.P., Mwenda, J.M., Langat, D.K., Chege, G.K., McClure, H.M., Mpoudi-Ngole, E., Delaporte, E., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H., 2004. New simian immunodeficiency virus infecting De Brazza's monkeys (*Cercopithecus neglectus*): evidence for a cercopithecus monkey virus clade. J. Virol. 78, 7748−7762.

Bibollet-Ruche, F., Galat-Luong, A., Cuny, G., Sarni-Manchado, P., Galat, G., Durand, J.P., Pourrut, X., Veas, F., 1996. Simian immunodeficiency virus infection in a patas monkey (*Erythrocebus patas*): evidence for cross-species transmission from African green monkeys (*Cercopithecus aethiops sabaeus*) in the wild. J. Gen. Virol. 77 (Pt 4), 773−781.

Bibollet-Ruche, F., Russell, R.M., Liu, W., Stewart-Jones, G.B.E., Sherrill-Mix, S., Li, Y., Learn, G.H., Smith, A.G., Gondim, M.V.P., Plenderleith, L.J., Decker, J.M., Easlick, J.L., Wetzel, K.S., Collman, R.G., Ding, S., Finzi, A., Ayouba, A., Peeters, M., Leendertz, F.H., van Schijndel, J., Goedmakers, A., Ton, E., Boesch, C., Kuehl, H., Arandjelovic, M., Dieguez, P., Murai, M., Colin, C., Koops, K., Speede, S., Gonder, M.K., Muller, M.N., Sanz, C.M., Morgan, D.B., Atencia, R., Cox, D., Piel, A.K., Stewart, F.A., Ndjango, J.N., Mjungu, D., Lonsdorf, E.V., Pusey, A.E., Kwong, P.D., Sharp, P.M., Shaw, G.M., Hahn, B.H., 2019. CD4 receptor diversity in chimpanzees protects against SIV infection. Proc. Natl. Acad. Sci. U.S.A. 116, 3229−3238.

Boué, V., Locatelli, S., Boucher, F., Ayouba, A., Butel, C., Esteban, A., Okouga, A.-P., Ndoungouet, A., Motsch, P., Le Flohic, G., Ngari, P., Prugnolle, F., Ollomo, B., Rouet, F., Liégeois, F., 2015. High rate of Simian Immunodeficiency Virus (SIV) infections in wild chimpanzees in northeastern Gabon. Viruses 7, 4997−5015.

Brennan, C.A., Bodelle, P., Coffey, R., Devare, S.G., Golden, A., Hackett Jr., J., Harris, B., Holzmayer, V., Luk, K.C., Schochetman, G., Swanson, P., Yamaguchi, J., Vallari, A., Ndembi, N., Ngansop, C., Makamche, F., Mbanya, D., Gurtler, L.G., Zekeng, L., Kaptue, L., 2008. The prevalence of diverse HIV-1 strains was stable in Cameroonian blood donors from 1996 to 2004. J. Acquir. Immune Defic. Syndr. 49, 432−439.

Buseyne, F., Betsem, E., Montange, T., Njouom, R., Bilounga Ndongo, C., Hermine, O., Gessain, A., 2018. Clinical signs and blood test results among humans infected with zoonotic simian foamy virus: a case-control study. J. Infect. Dis. 218, 144−151.

Buvé, A., Bishikwabo-Nsarhaza, K., Mutangadura, G., 2002. The spread and effect of HIV-1 infection in sub-Saharan Africa. Lancet 359, 2011−2017.

Calattini, S., Betsem, E.B.A., Froment, A., Mauclère, P., Tortevoye, P., Schmitt, C., Njouom, R., Saib, A., Gessain, A., 2007. Simian foamy virus transmission from apes to humans, rural Cameroon. Emerg. Infect. Dis. 13, 1314−1320.

Calvignac-Spencer, S., Adjogoua, E.V., Akoua-Koffi, C., Hedemann, C., Schubert, G., Ellerbrok, H., Leendertz, S.A.J., Pauli, G., Leendertz, F.H., 2012. Origin of human T-lymphotropic virus type 1 in rural Côte d'Ivoire. Emerg. Infect. Dis. 18, 830−833.

CDC, 1981. Pneumocystis Pneumonia — Los Angeles, Epidemiologic Notes and Reports. Centers for Disease Control and Prevention of USA.

Charleston, M.A., Robertson, D.L., 2002. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. Syst. Biol. 51, 528−535.

Chastel, C., 2000. Emergential success: a new concept for a better appraisal of viral emergences and reemergences. Acta Virol. 44, 375−376.

Chia, J.E., Aghokeng, A., Guichet, E., Ayouba, A., Ahuka-Mundeke, S., Vidal, N., Switzer, W.M., Delaporte, E., Mpoudi Ngole, E., Peeters, M., 2014. Ongoing cross-species transmission of simian retroviruses and high HIV prevalence in Cameroon. In: CROI 2014 - Program and Abstracts. Presented at the Conference on Retroviruses and Opportunistic Infections 2014, p. 222.

Clavel, F., Guétard, D., Brun-Vézinet, F., Chamaret, S., Rey, M.A., Santos-Ferreira, M.O., Laurent, A.G., Dauguet, C., Katlama, C., Rouzioux, C., 1986. Isolation of a new human retrovirus from West African patients with AIDS. Science 233, 343−346.

Damond, F., Worobey, M., Campa, P., Farfara, I., Colin, G., Matheron, S., Brun-Vézinet, F., Robertson, D.L., Simon, F., 2004. Identification of a highly divergent HIV type 2 and proposal for a change in HIV type 2 classification. AIDS Res. Hum. Retroviruses 20, 666−672.

Daniel, M.D., Letvin, N.L., King, N.W., Kannagi, M., Sehgal, P.K., Hunt, R.D., Kanki, P.J., Essex, M., Desrosiers, R.C., 1985. Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. Science 228, 1201−1204.

D'arc, M., Ayouba, A., Esteban, A., Learn, G.H., Boué, V., Liegeois, F., Etienne, L., Tagg, N., Leendertz, F.H., Boesch, C., Madinda, N.F., Robbins, M.M., Gray, M., Cournil, A., Ooms, M., Letko, M., Simon, V.A., Sharp, P.M., Hahn, B.H., Delaporte, E., Mpoudi Ngole, E., Peeters, M., 2015. Origin of the HIV-1 group O epidemic in western lowland gorillas. Proc. Natl. Acad. Sci. U. S. A. 112, E1343−E1352.

D'arc, M., Furtado, C., Siqueira, J.D., Seuánez, H.N., Ayouba, A., Peeters, M., Soares, M.A., 2018. Assessment of the gorilla gut virome in association with natural simian immunodeficiency virus infection. Retrovirology 15, 19. https://doi.org/10.1186/s12977-018-0402-9.

Delaugerre, C., De Oliveira, F., Lascoux-Combe, C., Plantier, J.-C., Simon, F., 2011. HIV-1 group N: travelling beyond Cameroon. Lancet 378, 1894.

De Leys, R., Vanderborght, B., Vanden Haesevelde, M., Heyndrickx, L., van Geel, A., Wauters, C., Bernaerts, R., Saman, E., Nijs, P., Willems, B., 1990. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of west-central African origin. J. Virol. 64, 1207−1216.

De Silva, T.I., Cotten, M., Rowland-Jones, S.L., 2008. HIV-2: the forgotten AIDS virus. Trends Microbiol. 16, 588−595.

Edoul, G., Chia, J.E., Vidal, N., Guichet, E., Montavon, C., Delaporte, E., Mpoudi Ngole, E., Ayouba, A., Peeters, M., 2020. High HIV burden and recent transmission chains in rural forest areas in southern Cameroon, where ancestors of HIV-1 have been identified in ape populations. Infect. Genet. Evol. 84, 104358.

Ellrodt, A., Barre-Sinoussi, F., Le Bras, P., Nugeyre, M.T., Palazzo, L., Rey, F., Brun-Vezinet, F., Rouzioux, C., Segond, P., Caquet, R., 1984. Isolation of human T-lymphotropic retrovirus (LAV) from Zairian married couple, one with AIDS, one with prodromes. Lancet 1, 1383−1385.

Etienne, L., Nerrienet, E., LeBreton, M., Bibila, G.T., Foupouapouognigni, Y., Rousset, D., Nana, A., Djoko, C.F., Tamoufe, U., Aghokeng, A.F., Mpoudi-Ngole, E., Delaporte, E., Peeters, M., Wolfe, N.D., Ayouba, A., 2011.

Characterization of a new simian immunodeficiency virus strain in a naturally infected *Pan troglodytes troglodytes* chimpanzee with AIDS related symptoms. Retrovirology 13 (8), 4.

Etienne, L., Hahn, B.H., Sharp, P.M., Matsen, F.A., Emerman, M., 2013. Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. Cell Host Microb. 14, 85−92.

Etienne, L., Bibollet-Ruche, F., Sudmant, P.H., Wu, L.I., Hahn, B.H., Emerman, M., 2015. The role of the antiviral APOBEC3 gene family in protecting chimpanzees against lentiviruses from monkeys. PLoS Pathog. 11, e1005149.

Faria, N.R., Rambaut, A., Suchard, M.A., Baele, G., Bedford, T., Ward, M.J., Tatem, A.J., Sousa, J.D., Arinaminpathy, N., Pépin, J., Posada, D., Peeters, M., Pybus, O.G., Lemey, P., 2014. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science 346, 56−61.

Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Sharp, P.M., Shaw, G.M., Hahn, B.H., 1992. Human infection by genetically diverse SIVsm-related HIV-2 in west Africa. Nature 358, 495−499.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.B., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H., 1999. Origin of HIV-1 in the chimpanzee *Pan troglodytes troglodytes*. Nature 397, 436−441.

Gifford, R.J., Katzourakis, A., Tristem, M., Pybus, O.G., Winters, M., Shafer, R.W., 2008. A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. Proc. Natl. Acad. Sci. U.S.A. 105, 20362−20367.

Gilbert, M.T.P., Rambaut, A., Wlasiuk, G., Spira, T.J., Pitchenik, A.E., Worobey, M., 2007. The emergence of HIV/AIDS in the Americas and beyond. Proc. Natl. Acad. Sci. U. S. A. 104, 18566−18570.

Gilbert, C., Maxfield, D.G., Goodman, S.M., Feschotte, C., 2009. Parallel germline infiltration of a lentivirus in two Malagasy lemurs. PLoS Genet. 5, e1000425.

Goldberg, T.L., Sintasath, D.M., Chapman, C.A., Cameron, K.M., Karesh, W.B., Tang, S., Wolfe, N.D., Rwego, I.B., Ting, N., Switzer, W.M., 2009. Coinfection of Ugandan red colobus (*Procolobus [Piliocolobus] rufomitratus tephrosceles*) with novel, divergent delta-, lenti-, and spumaretroviruses. J. Virol. 83, 11318−11329.

Gottlieb, G.S., Hawes, S.E., Agne, H.D., Stern, J.E., Critchlow, C.W., Kiviat, N.B., Sow, P.S., 2006. Lower levels of HIV RNA in semen in HIV-2 compared with HIV-1 infection: implications for differences in transmission. AIDS 20, 895−900.

Gryseels, S., Watts, T.D., Kabongo Mpolesha, J.M., Larsen, B.B., Lemey, P., Muyembe-Tamfum, J.J., Teuwen, D.E., Worobey, M., 2020. A near full-length HIV-1 genome from 1966 recovered from formalin-fixed paraffin-embedded tissue. Proc. Natl. Acad. Sci. U.S.A. 117, 12222−12229.

Hahn, B.H., Shaw, G.M., De Cock, K.M., Sharp, P.M., 2000. AIDS as a zoonosis: scientific and public health implications. Science 287, 607−614.

Hawes, S.E., Sow, P.S., Stern, J.E., Critchlow, C.W., Gottlieb, G.S., Kiviat, N.B., 2008. Lower levels of HIV-2 than HIV-1 in the female genital tract: correlates and longitudinal assessment of viral shedding. AIDS 22, 2517−2525.

Head, J.S., Boesch, C., Makaga, L., Robbins, M.M., 2011. Sympatric chimpanzees (*Pan troglodytes troglodytes*) and gorillas (*Gorilla gorilla gorilla*) in Loango national Park, Gabon: dietary composition, seasonality, and intersite comparisons. Int. J. Primatol. 32, 755−775.

Henrickson, R.V., Maul, D.H., Osborn, K.G., Sever, J.L., Madden, D.L., Ellingsworth, L.R., Anderson, J.H., Lowenstine, L.J., Gardner, M.B., 1983. Epidemic of acquired immunodeficiency in rhesus monkeys. Lancet 1, 388−390.

Hirsch, V.M., Olmsted, R.A., Murphey-Corb, M., Purcell, R.H., Johnson, P.R., 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature 339, 389−392.

Huet, T., Cheynier, R., Meyerhans, A., Roelants, G., Wain-Hobson, S., 1990. Genetic organization of a chimpanzee lentivirus related to HIV-1. Nature 345, 356−359.

Hussain, A.I., Shanmugam, V., Bhullar, V.B., Beer, B.E., Vallet, D., Gautier-Hion, A., Wolfe, N.D., Karesh, W.B., Kilbourn, A.M., Tooze, Z., Heneine, W., Switzer, W.M., 2003. Screening for simian foamy virus infection by using a combined antigen Western blot assay: evidence for a wide distribution among Old World primates and identification of four new divergent viruses. Virology 309, 248−257.

Jia, B., Serra-Moreno, R., Neidermyer, W., Rahmberg, A., Mackey, J., Fofana, I.B., Johnson, W.E., Westmoreland, S., Evans, D.T., 2009. Species-specific activity of SIV nef and HIV-1 Vpu in overcoming restriction by Tetherin/BST2. PLoS Pathog. 5, e1000429.

Jin, M.J., Rogers, J., Phillips-Conroy, J.E., Allan, J.S., Desrosiers, R.C., Shaw, G.M., Sharp, P.M., Hahn, B.H., 1994. Infection of a yellow baboon with simian immunodeficiency virus from African green monkeys: evidence for cross-species transmission in the wild. J. Virol. 68, 8454−8460.

Jonassen, T.O., Stene-Johansen, K., Berg, E.S., Hungnes, O., Lindboe, C.F., Frøland, S.S., Grinde, B., 1997. Sequence analysis of HIV-1 group O from Norwegian patients infected in the 1960s. Virology 231, 43−47.

Kanki, P.J., Barin, F., M'Boup, S., Allan, J.S., Romet-Lemonne, J.L., Marlink, R., McLane, M.F., Lee, T.H., Arbeille, B., Denis, F., 1986. New human T-lymphotropic retrovirus related to simian T-lymphotropic virus type III (STLV-IIIAGM). Science 232, 238−243.

Keele, B.F., Van Heuverswyn, F., Li, Y., Bailes, E., Takehisa, J., Santiago, M.L., Bibollet-Ruche, F., Chen, Y., Wain, L.V., Liegeois, F., Loul, S., Ngole, E.M., Bienvenue, Y., Delaporte, E., Brookfield, J.F.Y., Sharp, P.M., Shaw, G.M., Peeters, M., Hahn, B.H., 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313, 523−526.

Keele, B.F., Jones, J.H., Terio, K.A., Estes, J.D., Rudicell, R.S., Wilson, M.L., Li, Y., Learn, G.H., Beasley, T.M., Schumacher-Stankey, J., Wroblewski, E., Mosser, A., Raphael, J., Kamenya, S., Lonsdorf, E.V., Travis, D.A., Mlengeya, T., Kinsel, M.J., Else, J.G., Silvestri, G., Goodall, J., Sharp, P.M., Shaw, G.M., Pusey, A.E., Hahn, B.H., 2009. Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature 460, 515−519.

Kirchhoff, F., 2010. Immune evasion and counteraction of restriction factors by HIV-1 and other primate lentiviruses. Cell Host Microb. 8, 55−67.

Kluge, S.F., Mack, K., Iyer, S.S., Pujol, F.M., Heigele, A., Learn, G.H., Usmani, S.M., Sauter, D., Joas, S., Hotter, D., Bibollet-Ruche, F., Plenderleith, L.J., Peeters, M., Geyer, M., Sharp, P.M., Fackler, O.T., Hahn, B.H., Kirchhoff, F., 2014. Nef proteins of epidemic HIV-1 group O strains antagonize human tetherin. Cell Host Microb. 16, 639−650.

Laguette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Ségéral, E., Yatim, A., Emiliani, S., Schwartz, O., Benkirane, M., 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. Nature 474, 654−657.

Laporte, N.T., Stabach, J.A., Grosch, R., Lin, T.S., Goetz, S.J., 2007. Expansion of industrial logging in central Africa. Science 316, 1451.

Laurent, C., Bourgeois, A., Mpoudi, M., Butel, C., Peeters, M., Mpoudi-Ngolé, E., Delaporte, E., 2004. Commercial logging and HIV epidemic, rural Equatorial Africa. Emerg. Infect. Dis. 10, 1953−1956.

LeBreton, M., Yang, O., Tamoufe, U., Mpoudi-Ngole, E., Torimiro, J.N., Djoko, C.F., Carr, J.K., Tassy Prosser, A., Rimoin, A.W., Birx, D.L., Burke, D.S., Wolfe, N.D., 2007. Exposure to wild primates among HIV-infected persons. Emerg. Infect. Dis. 13, 1579−1582.

LeBreton, M., Switzer, W.M., Djoko, C.F., Gillis, A., Jia, H., Sturgeon, M.M., Shankar, A., Zheng, H., Nkeunen, G., Tamoufe, U., Nana, A., Le Doux Diffo, J., Tafon, B., Kiyang, J., Schneider, B.S., Burke, D.S., Wolfe, N.D., 2014. A gorilla reservoir for human T-lymphotropic virus type 4. Emerg. Microb. Infect. 3, e7.

Leendertz, S.A.J., Junglen, S., Hedemann, C., Goffe, A., Calvignac, S., Boesch, C., Leendertz, F.H., 2010. High prevalence, coinfection rate, and genetic diversity of retroviruses in wild red colobus monkeys (*Piliocolobus badius badius*) in Tai National Park, Cote d'Ivoire. J. Virol. 84, 7427−7436.

Leendertz, S.A.J., Locatelli, S., Boesch, C., Kücherer, C., Formenty, P., Liegeois, F., Ayouba, A., Peeters, M., Leendertz, F.H., 2011. No evidence for transmission of SIVwrc from western red colobus monkeys (*Piliocolobus badius badius*) to wild West African chimpanzees (*Pan troglodytes verus*) despite high exposure through hunting. BMC Microbiol. 11, 24.

Lemey, P., Pybus, O.G., Wang, B., Saksena, N.K., Salemi, M., Vandamme, A.-M., 2003. Tracing the origin and history of the HIV-2 epidemic. Proc. Natl. Acad. Sci. U. S. A. 100, 6588−6592.

Li, Y., Ndjango, J.-B., Learn, G.H., Ramirez, M.A., Keele, B.F., Bibollet-Ruche, F., Liu, W., Easlick, J.L., Decker, J.M., Rudicell, R.S., Inogwabini, B.-I., Ahuka-Mundeke, S., Leendertz, F.H., Reynolds, V., Muller, M.N., Chancellor, R.L., Rundus, A.S., Simmons, N., Worobey, M., Shaw, G.M., Peeters, M., Sharp, P.M., Hahn, B.H., 2012. Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir. J. Virol. 86, 10776−10791.

Liégeois, F., Boué, V., Mouacha, F., Butel, C., Ondo, B.M., Pourrut, X., Leroy, E., Peeters, M., Rouet, F., 2012. New STLV-3 strains and a divergent SIVmus strain identified in non-human primate bushmeat in Gabon. Retrovirology 9, 28.

Liégeois, F., Schmidt, F., Boué, V., Butel, C., Mouacha, F., Ngari, P., Ondo, B.M., Leroy, E., Heeney, J.L., Delaporte, E., Peeters, M., Rouet, F., 2014. Full-length genome analyses of two new simian immunodeficiency virus (SIV) strains from mustached monkeys (*C. cephus*) in Gabon illustrate a complex evolutionary history among the SIVmus/Mon/gsn lineage. Viruses 6, 2880−2898.

Ling, B., Apetrei, C., Pandrea, I., Veazey, R.S., Lackner, A.A., Gormus, B., Marx, P.A., 2004. Classic AIDS in a sooty mangabey after an 18-year natural infection. J. Virol. 78, 8902−8908.

Liovat, A.-S., Jacquelin, B., Ploquin, M.J., Barré-Sinoussi, F., Müller-Trutwin, M.C., 2009. African non human primates infected by SIV - why don't they get sick? Lessons from studies on the early phase of non-pathogenic SIV infection. Curr. HIV Res. 7, 39−50.

Liu, W., Worobey, M., Li, Y., Keele, B.F., Bibollet-Ruche, F., Guo, Y., Goepfert, P.A., Santiago, M.L., Ndjango, J.-B.N., Neel, C., Clifford, S.L., Sanz, C., Kamenya, S., Wilson, M.L., Pusey, A.E., Gross-Camp, N., Boesch, C., Smith, V., Zamma, K., Huffman, M.A., Mitani, J.C., Watts, D.P., Peeters, M., Shaw, G.M., Switzer, W.M., Sharp, P.M., Hahn, B.H., 2008. Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. PLoS Pathog. 4, e1000097.

Lloyd-Smith, J.O., George, D., Pepin, K.M., Pitzer, V.E., Pulliam, J.R.C., Dobson, A.P., Hudson, P.J., Grenfell, B.T., 2009. Epidemic dynamics at the human-animal interface. Science 326, 1362−1367.

Locatelli, S., Liegeois, F., Lafay, B., Roeder, A.D., Bruford, M.W., Formenty, P., Noë, R., Delaporte, E., Peeters, M., 2008. Prevalence and genetic diversity of simian immunodeficiency virus infection in wild-living red colobus monkeys (*Piliocolobus badius badius*) from the Taï forest, Côte d'Ivoire SIVwrc in wild-living western red colobus monkeys. Infect. Genet. Evol. 8, 1−14.

Locatelli, S., Peeters, M., 2012. Cross-species transmission of simian retroviruses: how and why they could lead to the emergence of new diseases in the human population. AIDS 26, 659−673.

Low-Beer, D., 2001. The distribution of early acquired immune deficiency syndrome cases and conditions for the establishment of new epidemics. Philos. Trans. R. Soc. Lond. B Biol. Sci. 356, 927−931.

Lowenstine, L.J., Pedersen, N.C., Higgins, J., Pallis, K.C., Uyeda, A., Marx, P., Lerche, N.W., Munn, R.J., Gardner, M.B., 1986. Seroepidemiologic survey of captive Old-World primates for antibodies to human and simian retroviruses, and isolation of a lentivirus from sooty mangabeys (*Cercocebus atys*). Int. J. Cancer 38, 563−574.

Lyons, S., Sharp, C., LeBreton, M., Djoko, C.F., Kiyang, J.A., Lankester, F., Bibila, T.G., Tamoufé, U., Fair, J., Wolfe, N.D., Simmonds, P., 2012. Species association of hepatitis B virus (HBV) in non-human apes; evidence for recombination between gorilla and chimpanzee variants. PLoS One 7, e33430.

Ma, D., Jasinska, A., Kristoff, J., Grobler, J.P., Turner, T., Jung, Y., Schmitt, C., Raehtz, K., Feyertag, F., Martinez Sosa, N., Wijewardana, V., Burke, D.S., Robertson, D.L., Tracy, R., Pandrea, I., Freimer, N., Apetrei, C., The

International Vervet Research Consortium, 2013. SIVagm infection in wild African green monkeys from South Africa: epidemiology, natural history, and evolutionary Considerations. PLoS Pathog. 9, e1003011.

Ma, D., Jasinska, A.J., Feyertag, F., Wijewardana, V., Kristoff, J., He, T., Raehtz, K., Schmitt, C.A., Jung, Y., Cramer, J.D., Dione, M., Antonio, M., Tracy, R., Turner, T., Robertson, D.L., Pandrea, I., Freimer, N., Apetrei, C., International Vervet Research Consortium, 2014. Factors associated with siman immunodeficiency virus transmission in a natural African nonhuman primate host in the wild. J. Virol. 88, 5687−5705.

Mansfield, K.G., Lerch, N.W., Gardner, M.B., Lackner, A.A., 1995. Origins of simian immunodeficiency virus infection in macaques at the new England regional primate Research center. J. Med. Primatol. 24, 116−122.

Meder, A., 1999. Gorillas in African Culture and medicine. Gorilla J. 18, 11−15.

Morgan, D., Sanz, C., 2006. Chimpanzee feeding ecology and comparisons with sympatric gorillas in the Goualougo Triangle, Republic of Congo. In: Feeding Ecology in Apes and Other Primates. Cambridge University Press, Cambridge, U.K., pp. 97−122

Mossoun, A., Pauly, M., Akoua-Koffi, C., Couacy-Hymann, E., Leendertz, S.A.J., Anoh, A.E., Gnoukpoho, A.H., Leendertz, F.H., Schubert, G., 2015. Contact to non-human primates and risk factors for zoonotic disease emergence in the Taï region. Côte d'Ivoire. EcoHealth 12, 580−591.

Mouinga-Ondémé, A., Caron, M., Nkoghé, D., Telfer, P., Marx, P., Saïb, A., Leroy, E., Gonzalez, J.-P., Gessain, A., Kazanji, M., 2012. Cross-species transmission of simian foamy virus to humans in rural Gabon, Central Africa. J. Virol. 86, 1255−1260.

Mourez, T., Simon, F., Plantier, J.-C., 2013. Non-M variants of human immunodeficiency virus type 1. Clin. Microbiol. Rev. 26, 448−461.

Muniz, C.P., Troncoso, L.L., Moreira, M.A., Soares, E.A., Pissinatti, A., Bonvicino, C.R., Seuánez, H.N., Sharma, B., Jia, H., Shankar, A., Switzer, W.M., Santos, A.F., Soares, M.A., 2013. Identification and characterization of highly divergent simian foamy viruses in a wide range of new world primates from Brazil. PLoS One 8, e67568.

Muniz, C.P., Jia, H., Shankar, A., Troncoso, L.L., Augusto, A.M., Farias, E., Pissinatti, A., Fedullo, L.P., Santos, A.F., Soares, M.A., Switzer, W.M., 2015. An expanded search for simian foamy viruses (SFV) in Brazilian New World primates identifies novel SFV lineages and host age-related infections. Retrovirology 12, 94.

Murphey-Corb, M., Martin, L.N., Rangan, S.R.S., Baskin, G.B., Gormus, B.J., Wolf, R.H., Andes, W.A., West, M., Montelaro, R.C., 1986. Isolation of an HTLV-III-related retrovirus from macaques with simian AIDS and its possible origin in asymptomatic mangabeys. Nature 321, 435−437.

Nakano, Y., Yamamoto, K., Ueda, M.T., Soper, A., Konno, Y., Kimura, I., Uriu, K., Kumata, R., Aso, H., Misawa, N., Nagaoka, S., Shimizu, S., Mitsumune, K., Kosugi, Y., Juarez-Fernandez, G., Ito, J., Nakagawa, S., Ikeda, T., Koyanagi, Y., Harris, R.S., Sato, K., 2020. A role for gorilla APOBEC3G in shaping lentivirus evolution including transmission to humans. PLoS Pathog. 16, e1008812.

Ndembi, N., Kaptue, L., Ido, E., 2009. Exposure to SIVmnd-2 in southern Cameroon: public health implications. AIDS Rev. 11, 135−139.

Neel, C., Etienne, L., Li, Y., Takehisa, J., Rudicell, R.S., Bass, I.N., Moudindo, J., Mebenga, A., Esteban, A., Van Heuverswyn, F., Liegeois, F., Kranzusch, P.J., Walsh, P.D., Sanz, C.M., Morgan, D.B., Ndjango, J.-B.N., Plantier, J.-C., Locatelli, S., Gonder, M.K., Leendertz, F.H., Boesch, C., Todd, A., Delaporte, E., Mpoudi-Ngole, E., Hahn, B.H., Peeters, M., 2010. Molecular epidemiology of simian immunodeficiency virus infection in wild-living gorillas. J. Virol. 84, 1464−1476.

Peeters, M., Honoré, C., Huet, T., Bedjabaga, L., Ossari, S., Bussi, P., Cooper, R.W., Delaporte, E., 1989. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. AIDS 3, 625−630.

Peeters, M., Fransen, K., Delaporte, E., Van den Haesevelde, M., Gershy-Damet, G.M., Kestens, L., van der Groen, G., Piot, P., 1992. Isolation and characterization of a new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-captured chimpanzee. AIDS 6, 447−451.

Peeters, M., Courgnaud, V., Abela, B., Auzel, P., Pourrut, X., Bibollet-Ruche, F., Loul, S., Liegeois, F., Butel, C., Koulagna, D., Mpoudi-Ngole, E., Shaw, G.M., Hahn, B.H., Delaporte, E., 2002. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. Emerg. Infect. Dis. 8, 451−457.

Peeters, M., D'Arc, M., Delaporte, E., 2014. Origin and diversity of human retroviruses. AIDS Rev. 16, 23−34.

Piot, P., Quinn, T.C., Taelman, H., Feinsod, F.M., Minlangu, K.B., Wobin, O., Mbendi, N., Mazebo, P., Ndangi, K., Stevens, W., 1984. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. Lancet 2, 65−69.

Plantier, J.-C., Leoz, M., Dickerson, J.E., De Oliveira, F., Cordonnier, F., Lemée, V., Damond, F., Robertson, D.L., Simon, F., 2009. A new human immunodeficiency virus derived from gorillas. Nat. Med. 15, 871−872.

Poulsen, J.R., Clark, C.J., Mavah, G., Elkan, P.W., 2009. Bushmeat supply and consumption in a tropical logging concession in northern Congo. Conserv. Biol. 23, 1597−1608.

Prince, A.M., Brotman, B., Lee, D.-H., Andrus, L., Valinsky, J., Marx, P., 2002. Lack of evidence for HIV type 1-related SIVcpz infection in captive and wild chimpanzees (*Pan troglodytes verus*) in West Africa. AIDS Res. Hum. Retroviruses 18, 657−660.

Rudicell, R.S., Holland Jones, J., Wroblewski, E.E., Learn, G.H., Li, Y., Robertson, J.D., Greengrass, E., Grossmann, F., Kamenya, S., Pintea, L., Mjungu, D.C., Lonsdorf, E.V., Mosser, A., Lehman, C., Collins, D.A., Keele, B.F., Goodall, J., Hahn, B.H., Pusey, A.E., Wilson, M.L., 2010. Impact of simian immunodeficiency virus infection on chimpanzee population dynamics. PLoS Pathog. 6, e1001116.

Russell, R.M., Bibollet-Ruche, F., Liu, W., Sherrill-Mix, S., Li, Y., Connell, J., Loy, D.E., Trimboli, S., Smith, A.G., Avitto, A.N., Gondim, M.V.P., Plenderleith, L.J., Wetzel, K.S., Collman, R.G., Ayouba, A., Esteban, A., Peeters, M., Kohler, W.J., Miller, R.A., François-Souquiere, S., Switzer, W.M., Hirsch, V.M., Marx, P.A., Piel, A.K., Stewart, F.A., Georgiev, A.V., Sommer, V., Bertolani, P., Hart, J.A., Hart, T.B., Shaw, G.M., Sharp, P.M., Hahn, B.H., 2021. CD4 receptor diversity represents an ancient protection mechanism against primate lentiviruses. Proc. Natl. Acad. Sci. U.S.A. 118, e2025914118.

Santiago, M.L., Rodenburg, C.M., Kamenya, S., Bibollet-Ruche, F., Gao, F., Bailes, E., Meleth, S., Soong, S.-J., Kilby, J.M., Moldoveanu, Z., Fahey, B., Muller, M.N., Ayouba, A., Nerrienet, E., McClure, H.M., Heeney, J.L., Pusey, A.E., Collins, D.A., Boesch, C., Wrangham, R.W., Goodall, J., Sharp, P.M., Shaw, G.M., Hahn, B.H., 2002. SIVcpz in wild chimpanzees. Science 295, 465.

Santiago, M.L., Lukasik, M., Kamenya, S., Li, Y., Bibollet-Ruche, F., Bailes, E., Muller, M.N., Emery, M., Goldenberg, D.A., Lwanga, J.S., Ayouba, A., Nerrienet, E., McClure, H.M., Heeney, J.L., Watts, D.P., Pusey, A.E., Collins, D.A., Wrangham, R.W., Goodall, J., Brookfield, J.F.Y., Sharp, P.M., Shaw, G.M., Hahn, B.H., 2003a. Foci of endemic simian immunodeficiency virus infection in wild-living eastern chimpanzees (*Pan troglodytes schweinfurthii*). J. Virol. 77, 7545−7562.

Santiago, M.L., Bibollet-Ruche, F., Bailes, E., Kamenya, S., Muller, M.N., Lukasik, M., Pusey, A.E., Collins, D.A., Wrangham, R.W., Goodall, J., Shaw, G.M., Sharp, P.M., Hahn, B.H., 2003b. Amplification of a complete simian immunodeficiency virus genome from fecal RNA of a wild chimpanzee. J. Virol. 77, 2233−2242.

Santiago, M.L., Range, F., Keele, B.F., Li, Y., Bailes, E., Bibollet-Ruche, F., Fruteau, C., Noë, R., Peeters, M., Brookfield, J.F.Y., Shaw, G.M., Sharp, P.M., Hahn, B.H., 2005. Simian immunodeficiency virus infection in free-ranging sooty mangabeys (*Cercocebus atys atys*) from the Taï Forest, Côte d'Ivoire: implications for the origin of epidemic human immunodeficiency virus type 2. J. Virol. 79, 12515−12527.

Sauter, D., Schindler, M., Specht, A., Landford, W.N., Münch, J., Kim, K.-A., Votteler, J., Schubert, U., Bibollet-Ruche, F., Keele, B.F., Takehisa, J., Ogando, Y., Ochsenbauer, C., Kappes, J.C., Ayouba, A., Peeters, M., Learn, G.H., Shaw, G., Sharp, P.M., Bieniasz, P., Hahn, B.H., Hatziioannou, T., Kirchhoff, F., 2009. Tetherin-driven adaptation of vpu and nef function and the evolution of pandemic and nonpandemic HIV-1 strains. Cell Host Microb. 6, 409−421.

Sauter, D., Hue, S., Petit, S.J., Plantier, J.-C., Towers, G.J., Kirchhoff, F., Gupta, R.K., 2011. HIV-1 Group P is unable to antagonize human tetherin by Vpu, Env or Nef. Retrovirology 8, 103.

Sauter, D., Unterweger, D., Vogl, M., Usmani, S.M., Heigele, A., Kluge, S.F., Hermkes, E., Moll, M., Barker, E., Peeters, M., et al., 2012. Human tetherin exerts strong selection pressure on the HIV-1 group N Vpu protein. PLoS Pathog. 8, e1003093.

Schindler, M., Münch, J., Kutsch, O., Li, H., Santiago, M.L., Bibollet-Ruche, F., Müller-Trutwin, M.C., Novembre, F.J., Peeters, M., Courgnaud, V., Bailes, E., Roques, P., Sodora, D.L., Silvestri, G., Sharp, P.M., Hahn, B.H., Kirchhoff, F., 2006. Nef-mediated suppression of T cell activation was lost in a lentiviral lineage that gave rise to HIV-1. Cell 125, 1055−1067.

Sharp, P.M., Hahn, B.H., 2008. AIDS: prehistory of HIV-1. Nature 455, 605−606.

Silvestri, G., Sodora, D.L., Koup, R.A., Paiardini, M., O'Neil, S.P., McClure, H.M., Staprans, S.I., Feinberg, M.B., 2003. Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia. Immunity 18, 441−452.

Silvestri, G., Paiardini, M., Pandrea, I., Lederman, M.M., Sodora, D.L., 2007. Understanding the benign nature of SIV infection in natural hosts. J. Clin. Invest. 117, 3148−3154.

Simon, F., Souquière, S., Damond, F., Kfutwah, A., Makuwa, M., Leroy, E., Rouquet, P., Berthier, J.L., Rigoulet, J., Lecu, A., Telfer, P.T., Pandrea, I., Plantier, J.C., Barré-Sinoussi, F., Roques, P., Müller-Trutwin, M.C., Apetrei, C., 2001. Synthetic peptide strategy for the detection of and discrimination among highly divergent primate lentiviruses. AIDS Res. Hum. Retroviruses 17, 937−952.

Sintasath, D.M., Wolfe, N.D., Zheng, H.Q., LeBreton, M., Peeters, M., Tamoufe, U., Djoko, C.F., Diffo, J.L.D., Mpoudi-Ngole, E., Heneine, W., Switzer, W.M., 2009. Genetic characterization of the complete genome of a highly divergent simian T-lymphotropic virus (STLV) type 3 from a wild *Cercopithecus mona* monkey. Retrovirology 6, 97.

Souquière, S., Bibollet-Ruche, F., Robertson, D.L., Makuwa, M., Apetrei, C., Onanga, R., Kornfeld, C., Plantier, J.C., Gao, F., Abernethy, K., White, L.J., Karesh, W., Telfer, P., Wickings, E.J., Mauclère, P., Marx, P.A., Barré-Sinoussi, F., Hahn, B.H., Müller-Trutwin, M.C., Simon, F., 2001. Wild Mandrillus sphinx are carriers of two types of lentivirus. J. Virol. 75, 7086−7096.

Southern, L.M., Deschner, T., Pika, S., 2021. Lethal coalitionary attacks of chimpanzees (*Pan troglodytes troglodytes*) on gorillas (*Gorilla gorilla gorilla*) in the wild. Sci. Rep. 11, 14673.

Stanford, C.B., Nkurunungi, J.B., 2003. Behavioral ecology of sympatric chimpanzees and gorillas in Bwindi Impenetrable National Park, Uganda: diet. Int. J. Primatol. 24, 901−918.

Switzer, W.M., Bhullar, V., Shanmugam, V., Cong, M.-E., Parekh, B., Lerche, N.W., Yee, J.L., Ely, J.J., Boneva, R., Chapman, L.E., Folks, T.M., Heneine, W., 2004. Frequent simian foamy virus infection in persons occupationally exposed to nonhuman primates. J. Virol. 78, 2780−2789.

Switzer, W.M., Parekh, B., Shanmugam, V., Bhullar, V., Phillips, S., Ely, J.J., Heneine, W., 2005. The epidemiology of simian immunodeficiency virus infection in a large number of wild- and captive-born chimpanzees: evidence for a recent introduction following chimpanzee divergence. AIDS Res. Hum. Retroviruses 21, 335−342.

Switzer, W.M., Salemi, M., Qari, S.H., Jia, H., Gray, R.R., Katzourakis, A., Marriott, S.J., Pryor, K.N., Wolfe, N.D., Burke, D.S., Folks, T.M., Heneine, W., 2009. Ancient, independent evolution and distinct molecular features of the novel human T-lymphotropic virus type 4. Retrovirology 6, 9.

Switzer, W.M., Tang, S., Ahuka-Mundeke, S., Shankar, A., Hanson, D.L., Zheng, H., Ayouba, A., Wolfe, N.D., LeBreton, M., Djoko, C.F., Tamoufe, U., Esteban, A., Heneine, W., Peeters, M., Wright, L.L., Muyembe-Tamfum, J.J., Wemakoy, E.O., Mulembakani, P., Hoff, N.A., Rimoin, A.W., 2012. Novel simian foamy virus infections from multiple monkey species in women from the Democratic Republic of Congo. Retrovirology 9, 100.

Takehisa, J., Kraus, M.H., Ayouba, A., Bailes, E., Van Heuverswyn, F., Decker, J.M., Li, Y., Rudicell, R.S., Learn, G.H., Neel, C., Ngole, E.M., Shaw, G.M., Peeters, M., Sharp, P.M., Hahn, B.H., 2009. Origin and biology of simian immunodeficiency virus in wild-living western gorillas. J. Virol. 83, 1635−1648.

Tienen, C. van, van der Loeff, M.S., Zaman, S.M.A., Vincent, T., Sarge-Njie, R., Peterson, I., Leligdowicz, A., Jaye, A., Rowland-Jones, S., Aaby, P., Whittle, H., 2010. Two distinct epidemics: the rise of HIV-1 and decline of HIV-2 infection between 1990 and 2007 in rural Guinea-Bissau. J. Acquir. Immune Defic. Syndr. 53, 640−647, 1999.

UNAIDS/WHO, 2014. The Gap Report, UNAIDS/JC2656. United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO).

Vallari, A., Holzmayer, V., Harris, B., Yamaguchi, J., Ngansop, C., Makamche, F., Mbanya, D., Kaptué, L., Ndembi, N., Gürtler, L., Devare, S., Brennan, C.A., 2011. Confirmation of putative HIV-1 group P in Cameroon. J. Virol. 85, 1403−1407.

Vanden Haesevelde, M.M., Peeters, M., Jannes, G., Janssens, W., van der Groen, G., Sharp, P.M., Saman, E., 1996. Sequence analysis of a highly divergent HIV-1-related lentivirus isolated from a wild captured chimpanzee. Virology 221, 346−350.

Van de Perre, P., Rouvroy, D., Lepage, P., Bogaerts, J., Kestelyn, P., Kayihigi, J., Hekker, A.C., Butzler, J.P., Clumeck, N., 1984. Acquired immunodeficiency syndrome in Rwanda. Lancet 2, 62−65.

Van Heuverswyn, F., Li, Y., Neel, C., Bailes, E., Keele, B.F., Liu, W., Loul, S., Butel, C., Liegeois, F., Bienvenue, Y., Ngolle, E.M., Sharp, P.M., Shaw, G.M., Delaporte, E., Hahn, B.H., Peeters, M., 2006. Human immunodeficiency viruses: SIV infection in wild gorillas. Nature 444, 164.

Van Heuverswyn, F., Li, Y., Bailes, E., Neel, C., Lafay, B., Keele, B.F., Shaw, K.S., Takehisa, J., Kraus, M.H., Loul, S., Butel, C., Liegeois, F., Yangda, B., Sharp, P.M., Mpoudi-Ngole, E., Delaporte, E., Hahn, B.H., Peeters, M., 2007. Genetic diversity and phylogeographic clustering of SIVcpz*Ptt* in wild chimpanzees in Cameroon. Virology 368, 155−171.

Van Rensburg, E.J., Engelbrecht, S., Mwenda, J., Laten, J.D., Robson, B.A., Stander, T., Chege, G.K., 1998. Simian immunodeficiency viruses (SIVs) from eastern and southern Africa: detection of a SIVagm variant from a chacma baboon. J. Gen. Virol. 79 (Pt 7), 1809−1814.

Vessière, A., Rousset, D., Kfutwah, A., Leoz, M., Depatureaux, A., Simon, F., Plantier, J.C., 2010. Diagnosis and monitoring of HIV-1 group O-infected patients in Cameroun. J. Acquir. Immune Defic. Syndr. 53, 107−110.

Vidal, N., Peeters, M., Mulanga-Kabeya, C., Nzilambi, N., Robertson, D., Ilunga, W., Sema, H., Tshimanga, K., Bongo, B., Delaporte, E., 2000. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. J. Virol. 74, 10498−10507.

Villabona-Arenas, C.J., Domyeum, J., Mouacha, F., Butel, C., Delaporte, E., Peeters, M., Mpoudi-Ngole, E., Aghokeng, A.F., 2015. HIV-1 group O infection in Cameroon from 2006 to 2013: prevalence, genetic diversity, evolution and public health challenges. Infect. Genet. Evol. 36, 210−216.

Villabona-Arenas, C.J., Ayouba, A., Esteban, A., D'arc, M., Mpoudi Ngole, E., Peeters, M., 2018. Noninvasive western lowland gorilla's health monitoring: a decade of simian immunodeficiency virus surveillance in southern Cameroon. Ecol. Evol. 8, 10698−10710.

Visseaux, B., Damond, F., Matheron, S., Descamps, D., Charpentier, C., 2016. HIV-2 molecular epidemiology. Infect. Genet. Evol. 46, 233−240.

Wain, L.V., Bailes, E., Bibollet-Ruche, F., Decker, J.M., Keele, B.F., Van Heuverswyn, F., Li, Y., Takehisa, J., Ngole, E.M., Shaw, G.M., Peeters, M., Hahn, B.H., Sharp, P.M., 2007. Adaptation of HIV-1 to its human host. Mol. Biol. Evol. 24, 1853−1860.

Walsh, P.D., Breuer, T., Sanz, C., Morgan, D., Doran-Sheehy, D., 2007. Potential for Ebola transmission between gorilla and chimpanzee social groups. Am. Nat. 169, 684−689.

Wertheim, J.O., Worobey, M., 2009. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. PLoS Comput. Biol. 5, e1000377.

Wertheim, J.O., Worobey, M., 2007. A challenge to the ancient origin of SIVagm based on African green monkey mitochondrial genomes. PLoS Pathog. 3, e95.

Wolfe, N.D., Switzer, W.M., Carr, J.K., Bhullar, V.B., Shanmugam, V., Tamoufe, U., Prosser, A.T., Torimiro, J.N., Wright, A., Mpoudi-Ngole, E., McCutchan, F.E., Birx, D.L., Folks, T.M., Burke, D.S., Heneine, W., 2004. Naturally acquired simian retrovirus infections in central African hunters. Lancet 363, 932−937.

Wolfe, N.D., Heneine, W., Carr, J.K., Garcia, A.D., Shanmugam, V., Tamoufe, U., Torimiro, J.N., Prosser, A.T., Lebreton, M., Mpoudi-Ngole, E., McCutchan, F.E., Birx, D.L., Folks, T.M., Burke, D.S., Switzer, W.M., 2005. Emergence of unique primate T-lymphotropic viruses among central African bushmeat hunters. Proc. Natl. Acad. Sci. U.S.A. 102, 7994−7999.

Worobey, M., Santiago, M.L., Keele, B.F., Ndjango, J.-B.N., Joy, J.B., Labama, B.L., Dhed'A, B.D., Rambaut, A., Sharp, P.M., Shaw, G.M., Hahn, B.H., 2004. Origin of AIDS: contaminated polio vaccine theory refuted. Nature 428, 820.

Worobey, M., Gemmel, M., Teuwen, D.E., Haselkorn, T., Kunstman, K., Bunce, M., Muyembe, J.-J., Kabongo, J.-M.M., Kalengayi, R.M., Van Marck, E., Gilbert, M.T.P., Wolinsky, S.M., 2008. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. Nature 455, 661−664.

Worobey, M., Telfer, P., Souquière, S., Hunter, M., Coleman, C.A., Metzger, M.J., Reed, P., Makuwa, M., Hearn, G., Honarvar, S., Roques, P., Apetrei, C., Kazanji, M., Marx, P.A., 2010. Island biogeography reveals the deep history of SIV. Science 329, 1487.

Worobey, M., Watts, T.D., McKay, R.A., Suchard, M.A., Granade, T., Teuwen, D.E., Koblin, B.A., Heneine, W., Lemey, P., Jaffe, H.W., 2016. 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. Nature 593, 98−101.

Yang, S.J., Lopez, L.A., Exline, C.M., Haworth, K.G., Cannon, P.M., 2011. Lack of adaptation to human tetherin in HIV-1 group O and P. Retrovirology 8, 78.

Zhang, F., Wilson, S.J., Landford, W.C., Virgen, B., Gregory, D., Johnson, M.C., Munch, J., Kirchhoff, F., Bieniasz, P.D., Hatziioannou, T., 2009. Nef proteins from simian immunodeficiency viruses are tetherin antagonists. Cell Host Microb. 6, 54−67.

Zhang, Z., Gu, Q., de Manuel Montero, M., Bravo, I.G., Marques-Bonet, T., Haussinger, D., Munk, C., 2017. Stably expressed APOBEC3H forms a barrier for cross-species transmission of simian immunodeficiency virus of chimpanzee to humans. PLoS Pathog. 13, e1006746.

Zheng, H., Wolfe, N.D., Sintasath, D.M., Tamoufe, U., Lebreton, M., Djoko, C.F., Diffo, J.L.D., Pike, B.L., Heneine, W., Switzer, W.M., 2010. Emergence of a novel and highly divergent HTLV-3 in a primate hunter in Cameroon. Virology 401, 137−145.

Zhu, T., Korber, B.T., Nahmias, A.J., Hooper, E., Sharp, P.M., Ho, D.D., 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 391, 594−597.

This page intentionally left blank

CHAPTER

Zoonotic Origin and Evolution of SARS Coronavirus

30

Zhengli Shi[1], Ben Hu[1] and Lin-Fa Wang[2]

[1]*CAS Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China;* [2]*Program in Emerging Infectious Disease, Duke-NUS Medical School, Singapore, Singapore*

1. A Brief History of SARS

The first reported case of "atypical pneumonia," now known as severe acute respiratory syndrome or SARS, occurred in Foshan, Guangdong province, China, on November 16, 2002. Before the end of February 2003, a total of 11 index cases occurred independently in 9 cities of Guangdong Province, which forms the early phase of the SARS epidemic (Table 30.1) (Chinese SMEC, 2004). These index cases spread the virus to their close relatives and hospital staffs and provided the early demonstration of the respiratory transmission mode of the disease. The clinical symptoms of SARS are nonspecific. The index cases all began to have fever higher than 38°C and displayed common respiratory symptoms such as cough, headache, and shortness of breath.

The dynamics of the outbreak was largely shaped by the presence of the so-called super spread event (SSE), in which a single patient was shown to spread the virus to a large number of contacts (Chinese SMEC, 2004). It is the SSEs that triggered the large scale of SARS pandemic in China. The first SSE patient is a businessman specialized in fishery wholesale. He was treated in three hospitals from January 30, 2003 to February 10, 2003 and along the way infected at least 78 other individuals including hospital staffs, patients, and close relatives and friends (Chinese SMEC, 2004). The second SSE individual, who caused the major spread of the disease out of Guangdong, was a business lady, native of Shanxi province. She went to Guangdong for business in late February and became sick while traveling. She went back to her home province and infected eight family members as well as five hospital staffs. The spread continued to Beijing when she decided to seek better treatment in Beijing (Table 30.1) (Chinese SMEC, 2004; Zhao, 2007).

The beginning of the global transmission occurred in Metropole Hotel of Hong Kong where a professor of nephrology from a Guangdong hospital stayed during a private visit. Without knowing, the urologist was infected with SARS-CoV a few days before he traveled to Hong Kong. It is later found that he spread the virus to at least 15 other persons in the hotel and in the hospital where he was treated. Among them, five of the hotel contacts continued their international journeys and further transmitted the disease to Vietnam, Singapore, Canada, and other countries (Table 30.1). This marks the true beginning of a disastrous worldwide pandemic (http://www.who.int/csr/sars/en/).

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00010-0
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

Table 30.1 Chronology of Major SARS Events

Date	Event
November 16, 2002	The first recognized SARS patient, in Foshan, Guangdong province, China
November 16, 2002–March 10, 2003	11 independent index cases in Foshan, Heyuan, Jiangmen, Zhongshan, Guanzhou, Zhaoqing, Shenzhen, Dongguan, China, more than 50 secondary infections
January 22, 2003	SARS spreading in Guangdong province
March 22, 2003	SARS spreading to Shanxi, Beijing
February 21, 2003	SARS spreading to Hong Kong, the beginning of global pandemic
February 28, 2003	SARS spreading to Viet Nam
March 12, 2003	Global alert of SARS pandemic by WHO
March 14, 2003	SARS spreading to Canada
March 6, 2003	SARS spreading to Singapore
March 15, 2003	WHO issue a rare emergency travel advisory as a global alert to international travelers, health care professionals, and health authorities
March 24, 2003	Coronavirus was isolated from SARS patient
April 4, 2003	SARS spreading to Philippines
April 12, 2003	Full-length genome of SARS virus was determined
April 17, 2003	The laboratory network announced conclusive identification of the SARS causative agent: a new coronavirus unlike any other known human or animal virus in its family
May 23, 2003	SARS coronavirus was detected in market animals
August 7, 2003	WHO reported 8096 cases and 774 death in the earth
September 2003–May 2004	1 in Singapore, 1 in Taiwan and 1 in Beijing through laboratory accident
December 16, 2003–January 8, 2004	4 cases in Guangdong, no death

WHO played a key role in the investigation and control of the SARS outbreak from the very beginning. For the first time in history, WHO issued a global travel alert on March 12, 2003 which greatly reduced the rate of long-distance transmission of the disease (Table 30.1). On March 17, 2003, WHO established a 9-nation/11-institute SARS network which played a major role in the rapid identification of the causative agent and development of diagnostic tests. Thanks to the international effort co-coordinated by WHO, the SARS outbreaks were effectively under control by July 5, 2003.

Following the major SARS outbreaks of 2003–2004, there were several minor outbreaks with much smaller impacts. Between December 2003 and January 2004, four independent SARS cases were reported in Guangdong, and none of them led to fatal infection or wide-spread transmission (Table 30.1). Subsequent epidemiological tracing revealed that all cases could be linked to civet trading activities (Wang et al., 2005). In addition, there were three laboratory outbreaks in September 2003, December 2003, and April 2004 in Singapore, Taiwan, and Beijing, respectively (Table 30.1). The most severe outbreak was associated with the incident in Beijing which resulted in a total of nine infection cases with one death. None of the other two laboratory infections resulted in further spread of the virus (Lim et al., 2006).

2. SARS Coronavirus (SARS-CoV)

Rapid identification of the causative agent in an outbreak caused by an unknown pathogen is the key for an effective response. However, in the case of SARS, this was not the case. Due to the association of nonspecific clinical symptoms associated with SARS patients, several pathogens were initially "identified" as the potential causes of SARS, which included chlamydia, influenza virus, and paramyxovirus (WHO, 2003). The confusion continued until March 2003 when three laboratories independently confirmed that a previously unknown coronavirus was the most likely etiological agent of SARS (Ksiazek et al., 2003; Peiris et al., 2003; Drosten et al., 2003).

Coronaviruses are enveloped viruses with large single-stranded, positive-sense RNA genomes. Coronaviruses can infect and cause disease in a broad array of avian and mammal species, including humans. The name "coronavirus" is derived from the Greek word, meaning crown, as the virus envelope appears under electron microscopy to be crowned by a characteristic ring of small bulbous structures. Within the virion, the ssRNA genome is encased in a helical nucleocapsid composed of many copies of the nucleocapsid (N) protein. The lipid bilayer envelope contains three proteins, envelope (E), and membrane (M), which coordinate virion assembly and release, and the large spike protein (S), which confers the virus's characteristic corona shape as well as serves as the principal mediator of host cell attachment and entry via virus- and host-specific cell receptors. The size of the SARS-CoV viral particle is approximately 80−90 nm and the genomic size is around 29.7 kb (Rota et al., 2003). Its genome contains 14 open reading frames (ORFs) flanked by 5′- and 3′-untranslated regions of 265 and 342 nucleotides, respectively (Marra et al., 2003). While CoVs carry species-specific accessory proteins encoded by their downstream ORFs, the order of essential genes—the replicase/transcriptase gene, S gene, E gene, M gene and N is highly conserved (Graham and Baric, 2010). Similar to other known coronaviruses, the SARS-CoV genome expression starts with two long open reading frames (ORFs), ORF1a and ORF1b which accounts for two-thirds of the genomic capacity, followed by ORFs encoding S, E, M and N proteins (Fig. 30.1). In addition to these conserved core genes in coronaviruses, the SARS-CoV genome contains several accessory genes which are specific to SARS-CoV and their encoded products have no homology to known proteins (Snijder et al., 2003). Taxonomically, SARS-CoV belongs to the species *Severe acute respiratory syndrome-related coronavirus*, the only species of the subgenus *Sarbecovirus* in the genus *Betacoronavirus* (previously known as group 2 coronavirus) (International Committee on Taxonomy of Viruses Executive Committee, 2020). The etiological agent of COVID-19, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which shares 79% genome sequence identity with SARS-CoV, is another member of this species.

3. The Animal Link of SARS-CoV

Due to the rapid spread of the disease and the delay in identification of the causative agent, there was no detailed epidemiological tracing done at the beginning of the outbreak. However, through retrospective investigation, it emerged that the majority of the early index cases were limited in several cities of Guangdong province and most of them have history of direct or indirect contact with wildlife, including handling, killing and selling wild animals as well as preparing and serving wildlife meat in restaurants (Guan et al., 2003; Xu et al., 2004).

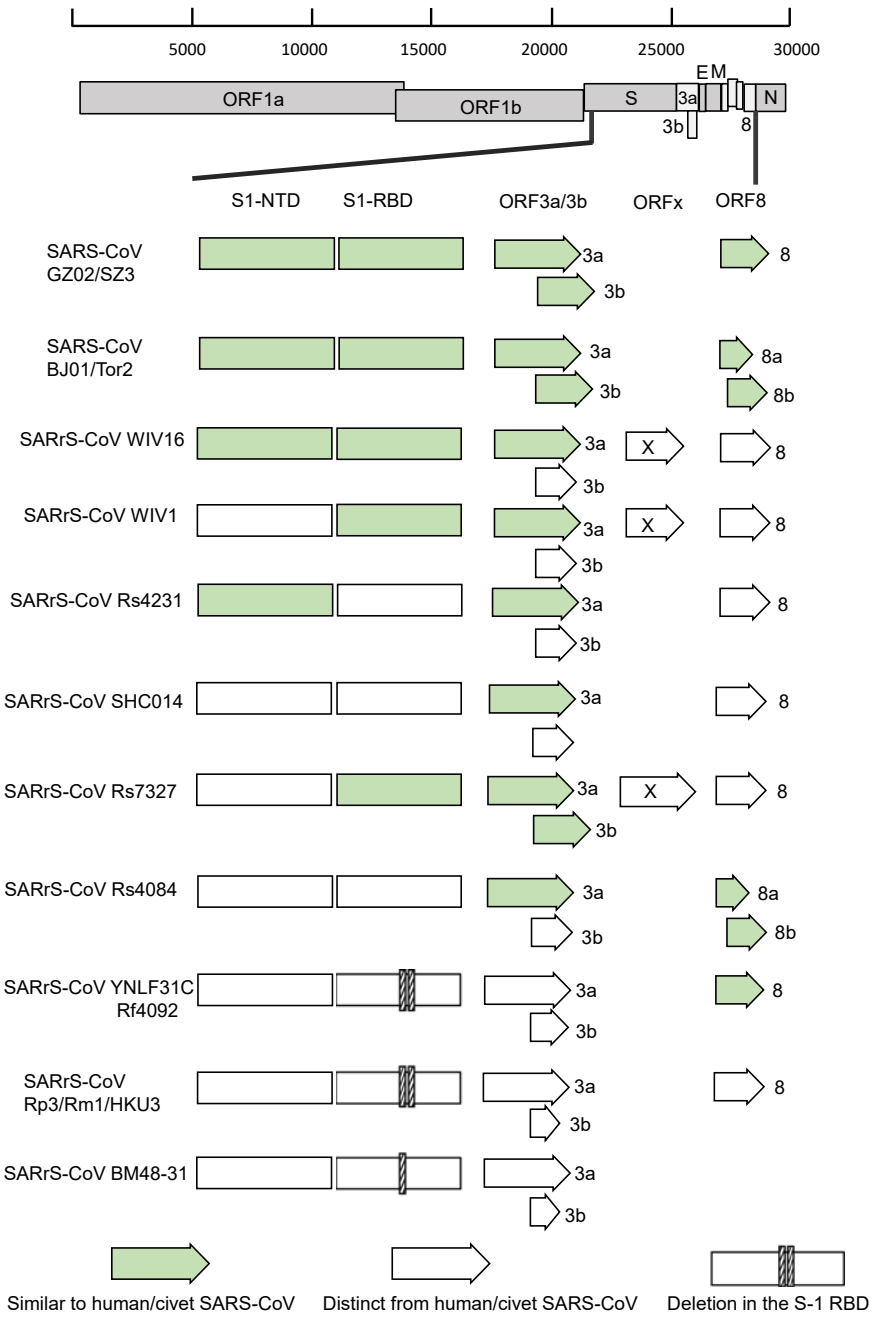


FIGURE 30.1 Genomic structure and variable genomic regions of different SARS-CoV and bat SARSr-CoV strains.

WIV1, WIV16, SHC014, Rs4231, Rs7327, and Rs4084 were identified from *Rhinolophus sinicus* in China; YNLF31C and Rf4092 from *Rhinolophus ferrumequinum* in China; Rm1 from *Rhinolohphus macrotis* in

Immediately after SARS-CoV was identified as the etiological agent of SARS, studies were conducted in those markets for evidence of SARS-CoV infection in market animals. One of the earliest and most important studies was conducted by a joint team from Hong Kong and Shenzhen in mainland China (Guan et al., 2003). Out of 25 market animals sampled in this investigation, SARS-CoV-like viruses were isolated from four out of six masked palm civets (*Paguma larvata*) and one raccoon dog (*Nyctereutes procyonoides*). Antibodies against SARS-CoV were detected in masked palm civets, raccoon dog as well as Chinese ferret-badgers (*Melogale moschata*). Genome sequencing indicated that the viruses isolated from civets were almost identical to those from human, suggesting a highly possible zoonotic transmission of SARS-CoV (Guan et al., 2003). These data indicated that at least three different animal species were infected by a coronavirus closely related to SARS-CoV. This study provided the first direct evidence that SARS-CoV existed in animals, pointing to an animal link of the SARS outbreak.

Although three animals were identified to be susceptible to SARS-CoV infection, the larger quantity of civets for sale in comparison to other animals in the market made them the target animals of subsequent surveillance studies. The role of civets as a major carrier of SARS-CoV in the markets was further confirmed by serological studies involving more samples (Kan et al., 2005; Tu et al., 2004). During the investigation of the second wave of SARS outbreaks during December 2003 to January 2004, there were two lines of evidences suggesting a direct transmission from civets to humans. First, all four independent cases had the history of direct or indirect contact with civets. Second, sequencing analysis indicated that SARS-CoV sequences derived from human samples in this wave showed higher similarity to those in the civets during that period than to the human SARS-CoVs during the major 2002–2003 outbreak (Wang et al., 2005).

In summary, based on the findings of above studies, it was concluded that the civet-to-human transmission is a major, if not the only, source of the SARS-CoV introduction into human population (Wang et al., 2006; Wang and Eaton, 2007; Shi and Hu, 2008).

4. Natural Reservoirs of SARS-CoV

Natural reservoir refers to the long-term host of a pathogen in the nature. It is often the case that such hosts, though infected with the pathogen, do not develop disease or the infection is subclinical, asymptomatic, and nonlethal. The discovery of natural reservoirs is crucial for elucidating the

China; BM48-31 from *Rhinolophus blasii* in Europe; BJ01 and Tor2 from patients during the middle and late phase of the 2002–2003 SARS outbreak; GZ02 from a patient during the early phase of the 2002–2003 SARS outbreak; SZ from civet during the 2002–2003 outbreak. The bat SARS-CoV-related viruses without deletions in the S1-RBD can be further divided into four genotypes: Genotype 1, represented by WIV16, is highly similar to SARS-CoV in both the NTD and the RBD; genotype 2, represented by WIV1 and Rs7327, differs in NTD from SARS-CoV; genotype 3, represented by Rs4231, differs in RBD from SARS-CoV; and genotype 4, represented by SHC014 and Rs4084, differs in both NTD and RBD from SARS-CoV. The Rs7327 genome encodes a nontruncated ORF3b with identical size and highly similar sequence compared with SARS-CoV. WIV1, WIV16 or Rs7327 contains an extra ORFX between ORF6 and ORF7. Rf4092/YNLF31C has a single ORF8 similar to that of GZ02/SZ3. Rs4084 has split ORF8a/ORF8b with high sequence similarity to those of BJ01/Tor2. The ORF8 is absent in BM48-31.

complete life cycle of infectious diseases, which in turn will help to provide effective prevention and control strategies. It is clear that civets played a pivotal role in the SARS outbreaks in southern China. However, the role of civets as a potential natural reservoir host was less evident and eventually ruled out by several studies. Serological and molecular studies indicated that only civets in the markets were infected with SARS-CoV, whereas the populations of civets in the wild or in the farms are free of infections (Tu et al., 2004; Poon et al., 2005; Lan et al., 2005). Civets presented obvious clinical symptoms when experimentally infected with SARS-CoV (Wu et al., 2005). Comparative genome sequence analysis indicated that SARS-CoVs in civets experienced rapid mutation, suggesting that the viruses were still adapting to the host rather than persisting in equilibrium expected for viruses in their natural reservoir species (Kan et al., 2005; Song et al., 2005).

Continuing searching for potential reservoir host of SARS-CoV resulted in the simultaneous discovery of related coronaviruses in bats by two independent teams in 2005. Using serological and PCR surveillance, both groups discovered coronaviruses similar to SARS-CoV in different horseshoe bats in the genus *Rhinolophus* from China (Li et al., 2005b; Lau et al., 2005). Complete genome sequence analysis revealed that the bat viruses have an identical genome organization and a nucleotide sequence identity of 88%−92% to SARS-CoV (Fig. 30.1) (Ren et al., 2006; Lau et al., 2010). Their nonstructural protein ORF1a/1b and all structural proteins except the S protein have a sequence identity above 93% with those of human/civet SARS-CoVs (Table 30.2). These bat coronaviruses belong to the same species as SARS-CoV and are termed SARS-related coronaviruses (SARSr-CoV).

In the following years more than 1 decade, a dramatically increasing number of novel SARSr-CoVs have been identified in a great diversity of horseshoe bat species. The bat SARSr-CoVs, which are also known as bat sarbecoviruses as they represent members of this single subgenus, consist of two lineages with closer phylogenetic relationship to SARS-CoV and SARS-CoV-2, respectively (Fig. 30.2). Different strains of viruses in the SARS-CoV-related lineage have been found not only in Chinese bats but also in European and African bats, suggesting a wide geographic distribution and high genetic diversity (Ge et al., 2013; Drexler et al., 2010; Tao and Tong, 2019; He et al., 2014; Yang et al., 2013; Wu et al., 2016a,b; Lau et al., 2015; Lin et al., 2017; Ruiz-Aravena et al., 2022). Analysis of nonsynonymous and synonymous substitution rates in bat SARSr-CoVs suggests that these viruses are not experiencing a positive selection pressure that would be expected if horseshoe bats are new host to these viruses, but they may have had a long evolutionary history in their bat hosts. These findings would support the notion that horseshoe bats may serve as natural reservoirs of viruses with close evolutionary relationship to SARS-CoV (Lau et al., 2010; Tang et al., 2009).

During a 5-year longitudinal surveillance of SARSr-CoV in a bat colony conducted in China, three live SARSr-CoVs have been successfully isolated from bat samples (Ge et al., 2013; Yang et al., 2015a; Hu et al., 2017). These isolates, namely WIV1, WIV16, and Rs4874, exhibited higher homology to SARS-CoV than previously known bat SARSr-CoVs, sharing 95.6%−96% genome sequence identity with SARS-CoV. They are close to SARS-CoV even in the highly variable receptor-binding domain (RBD) of the S protein, and thus are able to use the same cellular receptor, angiotensin converting enzyme 2 (ACE2), as SARS-CoV does. The isolation and characterization of bat SARSr-CoVs closely related to SARS-CoV not only in genome sequence but also in receptor usage provided robust evidence for the bat origin of SARS-CoV (Ge et al., 2013; Yang et al., 2015a).

Based on the full-length RNA-dependent RNA polymerase (*RdRp*) gene sequences, the SARS-CoV-related bat sarbecoviruses can be divided into four sublineages (Fig. 30.2). Among them, strains of sublineage 1 showed closer phylogenetic relationship to SARS-CoV, including WIV1 and

Table 30.2 Comparison of Gene Products Between SARS-CoV and Bat SARSr-CoVs

	Gene Product Size (aa)												Amino Acid Sequence Identity with Tor2/SZ3 (%)[a]									
Gene/ ORF	**Tor2**	**SZ3**	**Rf1**	**Rp3**	**Rm1**	**HK U3-1**	**Rs672**	**BM 48-31**	**WIV1**	**WIV16**	**RsSH C014**	**Rs40 84**	**Rf1**	**Rp3**	**Rm1**	**HK U3-1**	**Rs6 72**	**BM 48-31**	**WI V1**	**WIV 16**	**RsSH C014**	**Rs4 084**
P1a	4382	4382	4377	4380	4388	4376	4189	4369	4378	4382	4378	4382	94.3	96.6	93.6	94.3	98.1	81.6	98.0	98.1	98.1	98.0
P1b	2628	2628	2628	2628	2628	2628	2628	2628	2628	2628	2628	2628	98.3	99.1	98.9	98.6	99.3	96.0	99.4	99.1	99.3	98.9
nsp3[b]	1922	1922	1917	1920	1928	1916	1729	1912	1922	1922	1922	1922	92.4	95.4	91.0	92.1	96.9	76.8	97.0	97.1	96.9	96.9
S	1255	1255	1241	1241	1241	1242	1241	1259	1256	1255	1256	1256	78.4	80.2	80.6	79.4	80.1	76.0	92.4	97.3	90.0	89.9
S1	680	680	666	666	666	667	666	684	681	680	681	681	66.1	66.7	67.4	67.2	67.0	64.5	86.5	95.4	81.9	81.8
S2	575	575	575	575	575	575	575	575	575	575	575	575	92.7	96.2	96.0	93.9	95.5	89.6	99.5	99.5	99.5	99.5
ORF3a	274	274	274	274	274	274	274	271	274	274	274	274	86.9	84.3	84.3	82.5	90.8	71.5	97.8	98.2	97.1	97.1
ORF3b	154	154	113	56	56	39	114	NA	114	114	114	114	89	NA	NA	NA	97.4	NA	98.2	98.2	99.1	99.1
E	76	76	76	76	76	76	76	76	76	76	76	76	96.1	100	98.7	100	100	92.1	100	100	98.7	98.7
M	221	221	221	221	221	221	221	227	221	221	221	221	97.7	97.3	97.3	98.6	98.6	91.4	98.2	98.2	98.2	98.2
ORF6	63	63	63	63	63	63	63	62	63	63	63	63	93.7	92.1	92.1	93.7	98.4	50.0	93.7	92.1	96.8	96.8
ORF7a	122	122	122	122	122	122	122	118	122	122	122	122	91.8	95.1	93.4	94.3	95.9	58.5	95.1	95.1	95.9	95.9
ORF7b	44	44	44	44	44	44	44	40	44	44	44	44	90.9	93.2	93.2	93.2	93.2	70.0	93.2	93.2	93.2	93.2
ORF8a	39	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	47	NA	NA	NA	NA	NA	NA	NA	NA	NA	97.4
ORF8b	84	NP	NP	NP	NP	NP	NP	NP	NP	NP	NP	92	NA	NA	NA	NA	NA	NA	NA	NA	NA	96.4
ORF8	NP	122	122	121	121	121	121	NA	121	121	121	NP	81.8	39.5	39.5	37.7	39.5	NA	39.5	38.6	40.4	NA
N	422	422	421	421	420	421	422	417	422	422	422	422	95.7	98.1	97.9	96.7	98.6	88.2	99.8	99.5	99.5	99.5
ORF9b	98	98	96	97	97	97	98	96	98	98	98	98	82.5	86.6	91.8	88.7	91.8	47.9	98.0	98.0	98.0	95.9
ORF9c	70	70	70	70	70	70	70	69	70	70	70	70	85.7	91.4	91.4	88.6	94.3	47.8	94.3	95.7	95.7	95.7

NA, *not applicable;* NP, *not present.*

[a]*Tor2 was used for all homology calculations with the exception of ORF8, which absent in Tor2, SZ3 was used instead.*

[b]*The region of nsp3 is high variable and was calculated alone.*

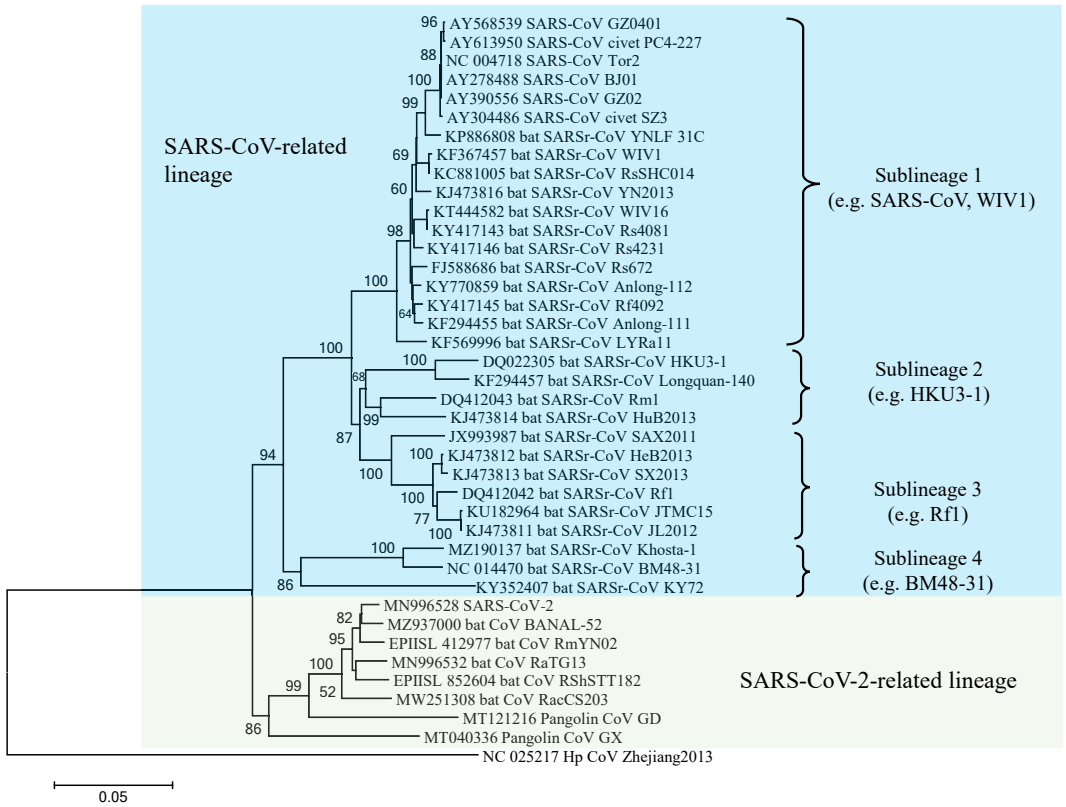


FIGURE 30.2 Phylogenetic tree of sarbecoviruses.

The phylogenetic tree is generated based on full-length RdRp sequences of selected SARS-CoVs and bat SARSr-CoVs using the Neighbor-Joining algorithm in the MEGA6 program with a bootstrap of 1000 replicates. A *Hipposideros* bat coronavirus Hp CoV Zhejiang2013 is used as an outgroup. Numbers above branches indicate bootstrap values from 1000 replicates. Scale bar, 0.5 substitutions per site. GZ02: SARS-CoV isolate from patients during the early phase of the 2002–2003 SARS outbreak; BJ01, Tor2: SARS-CoV isolates from patients during the middle and late phase of the 2002–2003 SARS outbreak; SZ3: SARS-CoV isolate from a civet during the 2002–2003 SARS outbreak; GZ0401: SARS-CoV isolate from patients during the 2003–2004 SARS outbreak; PC4-227: SARS-CoV isolate from a civet during the 2003–2004 SARS outbreak. Sarbecoviruses belonging to the SARS-CoV-related lineage and SARS-CoV-2-related lineage are indicated with *blue* and *green* background, respectively. The SARS-CoV-related lineage can be further divided into four sublineages, represented by WIV1, HKU3-1, Rf1, and BM48-31, respectively. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

WIV16. Interestingly, currently known bat SARSr-CoVs that fall into this sublineage were almost all discovered in southwestern China. Other two sublineages, represented by HKU3 and Rf1, are less closely related to SARS-CoV, and mainly circulate in South/central China and northern China, respectively. The last sublineage is composed of SARSr-CoVs identified in bats from Europe and

Africa. These SARSr-CoVs are more distant from SARS-CoV, with genome sequence identities lower than 80%, and they have some characteristic genomic features such as the loss of the entire ORF8 (Alkhovsky et al., 2022; Drexler et al., 2010; Tao and Tong, 2019). It is suggested that the evolution of bat SARS-CoV-related viruses are strongly correlated with their geographic origin.

5. Molecular Evolution of SARS-CoV in Humans and Animals

5.1 Rapid Adaptation of SARS-CoVs in Humans

On the basis of epidemiological data, the Chinese SARS molecular epidemiology consortium divided the course of the 2002−2003 outbreak into three stages, the early, middle, and late phases, respectively (Chinese SMEC, 2004). The early phase is defined as the period from the first emergence of SARS to the first documented SSE. The middle phase refers to the ensuing events up to the first cluster of SARS cases in a hotel (Hotel M) in Hong Kong, while cases following this cluster fall into the late phase.

Analysis of all the viral sequences available from human patients and animals revealed two major hallmarks of rapid virus evolution during the initial stages of the 2002−2003 outbreaks: (1) All isolates from early patients and market animals contained a 29-nucleotide (nt) sequence in ORF8 that is absent in most of the publicly available human SARS-CoV sequences derived from later phases of the outbreaks; (2) A characteristic motif of single-nucleotide variations (SNVs) were identified in SARS-CoVs of different phases and all these SNVs were located in the S gene which codes for the spike protein responsible for attachment to host cellular receptor (Song et al., 2005). All SARS-CoV isolates from epidemic countries and regions outside the mainland China could be traced to Guangdong or Hong Kong based on the S-gene SNV motif (Lan et al., 2005; Tang et al., 2007).

During the second sporadic outbreaks of 2003−04, it was shown that the SARS-CoV sequences from index patients were almost identical to that from civets collected in the same period and all retained the 29-nt sequence in the ORF8 gene. The mild disease symptoms associated with these viruses and the lack of rapid human-to-human transmission provided further evidence that the rapid adaptation of the SARS-CoV in the first major outbreak of 2002−03 was essential for its establishment and pathogenesis in humans.

With the available data on genomic variation and sampling time, it is now possible to calculate the neutral mutation rate and to estimate the date for the most recent common ancestors (MRCAs) of SARS-CoV. The estimate obtained is around 8.00×10^{-6} nt^{-1}/day, suggesting that SARS-CoV evolved at a relatively constant and neutral rate both in humans and palm civets. From these calculations, it was estimated that the MRCAs for different transmission lineages in palm civets and humans occurred in mid-November 2002. This estimate was consistent with the observation of the first SARS case around November 16, 2002 in Foshan, Guangdong (Chinese SMEC, 2004; Song et al., 2005; Zhao, 2007).

5.2 Receptor Usage and Evolutionary Selection

The S protein of coronavirus is responsible for attachment to cellular receptor to initiate the first step of virus infection. ACE2 was identified as a main functional receptor for SARS-CoV (Li et al., 2003). Further analysis demonstrated that the region covering aa 318−510 of S protein is the RBD, which is both essential and sufficient to bind the human ACE2 molecule in vitro (Wong et al., 2004). Detailed

analysis of the crystal structure of the RBD-ACE2 complex revealed that 19 key residues have close contact with the receptor molecule, which are located from aa 424 to 494, and this region is termed the receptor binding motif (RBM) (Li et al., 2005a).

When the epidemiological data was analyzed in combination with the data on infectivity of SARS-CoVs isolated in humans and civets at the different phases of the outbreaks, a clear correlation could be established between the evolution of the S proteins and virus infectivity. It was observed that the S protein is the fastest evolving protein of SARS-CoV during interspecies transmission from animals to humans and during the following phase of human-to-human transmission. The majority of mutations are located in the S1 domain (31 out of a total of 48 SNVs), particularly in the RBD (Chinese SMEC, 2004; Wong et al., 2004). The interaction analysis between the S proteins of different isolates and the ACE2 molecules demonstrated that two aa residues in the S protein, aa 479 and aa 487, played a key role in virus infectivity (Qu et al., 2005; Li et al., 2005c). For aa residue 479, all 2002−2003 human isolates contain asparagine (N). However, the palm civet isolates seem to have variable aa residues at this position. All 2002−2003 and some 2003−2004 civet isolates have lysine (K), while other 2003−2004 isolates have either asparagine (N) or arginine (R). For aa residue 487, all isolates from civets of the 2002−2003 and 2003−2004 outbreaks and the isolates from patients of the 2003−2004 outbreak have a codon for serine (S), whereas all isolates from human patients of the 2002−2003 outbreak have a codon for threonine (T) (Fig. 30.3). When examined using an HIV-based pseudovirus infection assay, S proteins with all combinations of residues 487/479 could efficiently use the civet ACE2 as entry receptor but showed different infectivity in human ACE2-mediated infection (Li et al., 2005c; Qu et al., 2005). The combination of N479/T487 had the highest infectivity, followed by N479/S487, and K479/S487 were the lowest, which almost aborted the infection. These results demonstrated at the molecular interface that the rapid evolution of the S protein, especially at the aa positions important for host receptor engagement, was essential for the adaptation and establishment of an effective and productive human infection.

Sequences of the S-RBD region are highly diverse among different sarbecoviruses. Based on the RBM sequences and the capability of ACE2 usage, the bat SARS-CoV-related viruses can be generally grouped into three clades. In one clade, most of strains, such as WIV1, WIV16, and SHC014, exhibit higher sequence identity to SARS-CoV in the RBD and their size of RBD is identical to that of SARS-CoV (Fig. 30.3). This group of SARSr-CoVs have been shown to be able to use human ACE2 for cellular entry (Ge et al., 2013; He et al., 2014; Hu et al., 2017; Yang et al., 2015a; Menachery et al., 2015). The strain SHC014 can replicate well in transgenic mice expressing human ACE2 and caused tissue damage in tested animals (Menachery et al., 2015). Both in vitro and in vivo data have suggested the risk of cross-species infection of these bat SARSr-CoVs. In addition, SARSr-CoVs circulating in bats from Europe, for example, BM48-31 from Bulgaria and Khosta-1/2 from Russia, contained a 4-aa deletion in the RBM and their RBD sequences are distinct from SARS-CoV (Fig. 30.3) (Drexler et al., 2010; Alkhovsky et al., 2022). Regardless of this, the Khosta strains also possess the ability to bind to bat or human ACE2 (Starr et al., 2022). In contrast, strains in the other clade, which have been found extensively distributed in China, have two deletions (5 aa and 12−13 aa, respectively) and have a major sequence difference in the RBM if compared with SARS-CoV (Fig. 30.3). We have demonstrated experimentally that these SARSr-CoV are unable to use ACE2 as a functional receptor.

Among the horseshoe bat reservoirs of SARSr-CoV, the species *Rhinolophus sinicus* harbored SARS-CoV-related viruses with a rich diversity in the S gene, which is in accordance with the high polymorphism of ACE2 among *Rhinolophus sinicus* populations. These ACE2 variants support

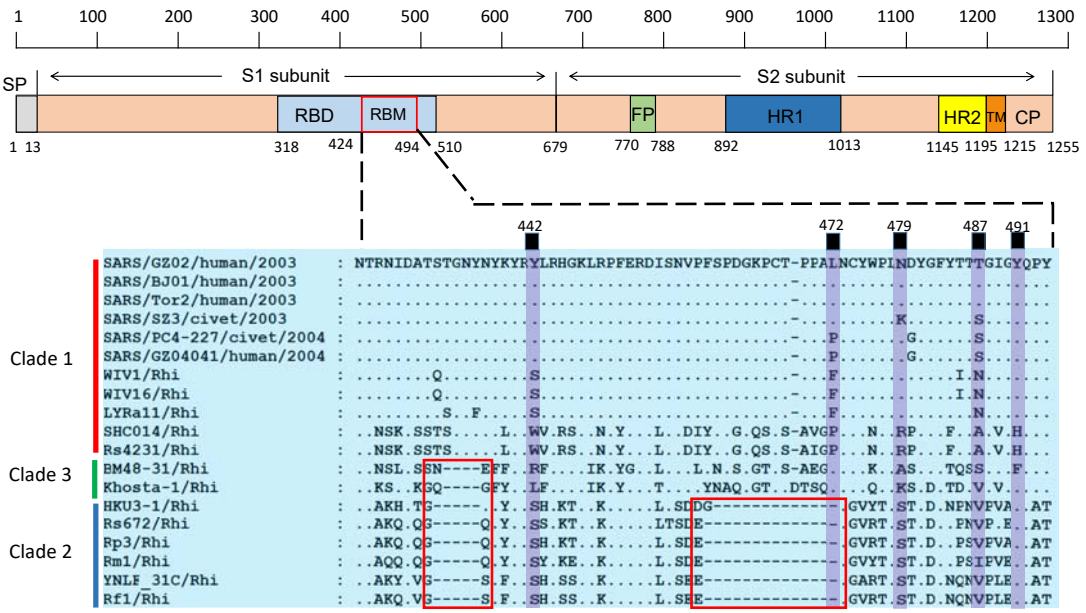


FIGURE 30.3 Alignment of amino acid sequences of the receptor binding motif in the S protein of different SARS-CoV-related coronaviruses.

GZ02: SARS-CoV isolate from patients during the early phase of the 2002–2003 SARS outbreak; BJ01, Tor2: SARS-CoV isolates from patients during the middle and late phase of the 2002–2003 SARS outbreak; SZ3: SARS-CoV isolate from a civet during the 2002–2003 SARS outbreak; GZ0401: SARS-CoV isolate from patients during the 2003–2004 SARS outbreak; PC4-227: SARS-CoV isolate from a civet during the 2003–2004 SARS outbreak. The *red* line and *green* line indicate the SARS-CoV-related viruses without deletions or with only one 4-aa deletion in the RBD, respectively (clade 1 and clade 3). The *blue* line indicates the SARS-CoV-related viruses with two deletions in the RBD (clade 2). The bars in *violet* indicate the location of the five key amino acid residues involved in the interaction of the SARS-CoV S protein with human ACE2. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

SARS-CoV and SARSr-CoV infection but with varying binding affinities to spike proteins of SARS-CoV and different SARSr-CoVs (Guo et al., 2020). Positive selection is detected in the residues at the interface between ACE2 and SARSr-CoV S protein, suggesting the long-term evolutionary arm race dynamics between the viruses and their bat hosts (Guo et al., 2020).

5.3 Potential Recombination Origin of SARS-CoV

Besides the S gene, the other two regions highly variable among SARS-CoV and different bat SARSr-CoVs are the ORF8 and ORF3 (Hu et al., 2017). Like the civet SARS-CoVs and human SARS-CoVs in the early phase of the 2002–2003 outbreak, most bat SARSr-CoVs have an intact ORF8, but the sequences are quite distinct from SARS-CoV, sharing lower than 40% aa sequence identities (Table 30.2). By contrast, the ORF8 of several SARSr-CoV strains have much higher similarity to SARS-CoV, with the aa sequence identities from 82% to 98%. In the European and African bat

SARSr-CoVs such as BM48-31, the ORF8 is completely lost (Fig. 30.1). The ORF3b of SARS-CoV is truncated to different extents at the C-terminal in almost all bat SARSr-CoVs, but the nontruncated ORF3b highly identical to that of SARS-CoV has also been detected in a bat SARSr-CoV (Hu et al., 2017).

Although the direct progenitor of SARS-CoV has not been found in bats, we have identified a rich gene pool of SARS-CoV in a longitudinal surveillance study. SARSr-CoV closely related to SARS-CoV in the hypervariable N-terminal domain of the S1 protein (S1-NTD), S1-RBD, ORF3, and ORF8, respectively, were discovered in a single bat colony (Fig. 30.1). As evidence of recombination are shown in the genomes of various SARSr-CoVs from there, it can be hypothesized that the progenitor of SARS-CoV may have originated from bats through a series of recombination events among the bat SARSr-CoVs (Hu et al., 2017).

6. Preparedness of Future SARS-like Disease Emergence

6.1 Virus Surveillance in Wildlife

At the end of 2019, a novel sarbecovirus named SARS-CoV-2 emerged and caused the largest pandemic in the recent 100 years (Chen et al., 2020; Li et al., 2020b; Wu et al., 2020; Zhou et al., 2020b; Zhu et al., 2020; Hu et al., 2021). This virus shares around 79.6% net identity to SARS-CoV and is likely to have originated from bats as numerous coronaviruses related to SARS-CoV-2 were discovered in bats (Zhou et al., 2020a,b; Delaune et al., 2021; Temmam et al., 2022; Wacharapluesadee et al., 2021). Nevertheless, in contrast to the rapid identification of the intermediate host of SARS-CoV, the intermediate hosts of SARS-CoV-2 remain elusive.

Zoonoses account for the majority of emerging diseases in the last 40 years, and many of them originated from wildlife (Jones et al., 2008; Chomel et al., 2007; Bengis et al., 2004; Woolhouse and Gowtage-Sequeria, 2005). The story of SARS is just one of such examples which spectacularly demonstrated the evolution of a bat virus into a human pathogen responsible for a severe pandemic in modern human history. In general, pathogens carried by wildlife reservoirs usually do not cause clinical symptoms and they are dormant until spillover into domestic animals or humans. The traditional outbreak response measures which are deployed after the disease emerged are still useful, but are no longer sufficient for early surveillance and precaution of emerging infectious disease outbreaks. With the recent demonstration of an increasing number of spillover events which led to outbreaks of severe diseases in humans and domestic animals, we believe it is vital that from now on we include active surveillance among wildlife as part of an integrated strategy for infectious disease prevention and control.

Surveillance of wildlife animals has been made more feasible and productive owing to the advance in modern molecular techniques including PCR with family or genus-specific primers, the next generation high-throughput sequencing technologies and high-density virus microarrays (Yanai-Balser et al., 2010; Gaynor et al., 2007; Ge et al., 2012; Lipkin and Anthony, 2015; Briese et al., 2015; Li et al., 2020a; He et al., 2021). Since the SARS outbreaks, especially after the discovery of SARSr-CoVs in bats, there is a significant surge in international effort for coronavirus surveillance in wildlife (Ruiz-Aravena et al., 2022; Drexler et al., 2014; Woo et al., 2009a,b, 2012; Alekseev et al., 2008; Corman et al., 2015; Fan et al., 2019; Pomorska-Mol et al., 2022; Wang et al., 2017; Li et al., 2021a). It is speculated that the majority of known coronaviruses that cause diseases in humans or livestock have derived from bat coronaviruses (Hu et al., 2015; Cui et al., 2019). This hypothesis was unfortunately

proved by the outbreaks of two more recently emerging coronavirus diseases, Middle East Respiratory Syndrome (MERS) (Zaki et al., 2012) and COVID-19. Even though the MERS-CoV-like viruses found in bats are not the direct progenitor of the MERS-CoV, the evolutionary ancestor of MERS-CoV is supposed to have originated from bats (Wang et al., 2014; Yang et al., 2014, 2015b; Anthony et al., 2017). The association between bats and SARS as well as MERS highlights the necessity of active surveillance in bat reservoirs for precaution and preparedness of coronavirus outbreaks that may happen in the future.

6.2 Development of Pan-Coronavirus Therapeutics and Vaccines

Concerning the cross-species transmission risk of some bat SARSr-CoVs, it is necessary to prepare in advance prophylactic and therapeutic technologies against future occurrence of this group of viruses. Due to the genetic variability of SARSr-CoVs in wildlife, development of broad-spectrum vaccines and antivirals targeting variants of sarbecoviruses should be strengthened.

The SARSr-CoVs, though diverse in the S1 protein, are highly conserved in the domains responsible for membrane fusion, which may serve as a target for designing pan-vaccines (Ma et al., 2020). A recent study revealed that the inactivated SARS-CoV-2 vaccine provided partial to full protection against bat SARSr-CoV infection in human ACE2 transgenic mice, indicating the feasibility of developing pan-sarbecovirus vaccines (Liu et al., 2022a). The similar conformational structure of some SARSr-CoVs suggests that there are potential epitopes which may induce cross-protection antibodies (Liu et al., 2022b). Different research groups have identified monoclonal antibodies neutralizing not only the SARS-CoV-2 variants but also bat SARSr-CoVs (Wec et al., 2020; Tortorici et al., 2021; Rappazzo et al., 2021).

The pan-coronavirus antivirals can be mainly categorized as viral replication inhibitors and entry inhibitors (Lu et al., 2021; Li et al., 2021b). As the RdRp is relatively conserved among coronaviruses, it is one of the common targets for pan-drug development. A novel nucleotide analog named molnupiravir can broadly inhibit propagation of SARS-CoV-2, SARS-CoV, and MERS-CoV, and shows resistance to the proofreading exoribonuclease encoded by coronaviruses. It appears to be a promising candidate of pan-coronavirus drug (Malone and Campbell, 2021). Some peptides have been developed as broad-spectrum coronavirus fusion inhibitors. For example, A cholesterol-conjugated lipopeptide called EK1C4, which was derived from the HR2 domain of human coronavirus OC43, has shown potent inhibitory activity against a variety of human coronaviruses as well as bat SARSr-CoVs (Wang et al., 2021; Xia et al., 2020). Stock of the broad-spectrum therapeutic agents for emerging coronaviruses would be an important line of defense in the case of a novel epidemic outbreak.

7. Concluding Remarks

The emergence of SARS-CoV, MERS-CoV, and SARS-CoV-2 has had a giant impact on the global health and economy. The experiences gained from the SARS and COVID-19 outbreaks and the following in-depth studies on SARSr-CoVs has provided invaluable knowledge and guidelines to our future fighting against new emerging infectious diseases. One of the major lessons is that we need to pay much more attention to the reservoir species for a better understanding of the genetic diversity and evolution of the viruses, the subtle interplay at the virus–host interface, and the potential factors contributing to the disturbance of virus–host equilibrium which may in turn triggers spillover.

References

Alekseev, K.P., Vlasova, A.N., Jung, K., Hasoksuz, M., Zhang, X., Halpin, R., Wang, S., Ghedin, E., Spiro, D., Saif, L.J., 2008. Bovine-like coronaviruses isolated from four species of captive wild ruminants are homologous to bovine coronaviruses, based on complete genomic sequences. J. Virol. 82, 12422−12431. https://doi.org/10.1128/JVI.01586-08.

Alkhovsky, S., Lenshin, S., Romashin, A., Vishnevskaya, T., Vyshemirsky, O., Bulycheva, Y., Lvov, D., Gitelman, A., 2022. SARS-like coronaviruses in horseshoe bats (Rhinolophus spp.) in Russia, 2020. Viruses 14. https://doi.org/10.3390/v14010113.

Anthony, S.J., Gilardi, K., Menachery, V.D., Goldstein, T., Ssebide, B., Mbabazi, R., Navarrete-Macias, I., Liang, E., Wells, H., Hicks, A., Petrosov, A., Byarugaba, D.K., Debbink, K., Dinnon, K.H., Scobey, T., Randell, S.H., Yount, B.L., Cranfield, M., Johnson, C.K., Baric, R.S., Lipkin, W.I., Mazet, J.A., 2017. Further evidence for bats as the evolutionary source of middle east respiratory syndrome coronavirus. mBio 8. https://doi.org/10.1128/mBio.00373-17.

Bengis, R.G., Leighton, F.A., Fischer, J.R., Artois, M., Morner, T., Tate, C.M., 2004. The role of wildlife in emerging and re-emerging zoonoses. Rev. Sci. Tech. 23, 497−511.

Briese, T., Kapoor, A., Mishra, N., Jain, K., Kumar, A., Jabado, O.J., Lipkin, W.I., 2015. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. mBio 6, e01491. https://doi.org/10.1128/mBio.01491-15.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T., Zhang, X., Zhang, L., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395, 507−513. https://doi.org/10.1016/S0140-6736(20)30211-7.

Chinese, S.M.E.C., 2004. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 303, 1666−1669. https://doi.org/10.1126/science.1092002.

Chomel, B.B., Belotto, A., Meslin, F.X., 2007. Wildlife, exotic pets, and emerging zoonoses. Emerg. Infect. Dis. 13, 6−11. https://doi.org/10.3201/eid1301.060480.

Corman, V.M., Baldwin, H.J., Tateno, A.F., Zerbinati, R.M., Annan, A., Owusu, M., Nkrumah, E.E., Maganga, G.D., Oppong, S., Adu-Sarkodie, Y., Vallo, P., Da Silva Filho, L.V., Leroy, E.M., Thiel, V., Van Der Hoek, L., Poon, L.L., Tschapka, M., Drosten, C., Drexler, J.F., 2015. Evidence for an ancestral association of human coronavirus 229E with bats. J. Virol. 89, 11858−11870. https://doi.org/10.1128/JVI.01755-15.

Cui, J., Li, F., Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17, 181−192. https://doi.org/10.1038/s41579-018-0118-9.

Delaune, D., Hul, V., Karlsson, E.A., Hassanin, A., Ou, T.P., Baidaliuk, A., Gambaro, F., Prot, M., Tu, V.T., Chea, S., Keatts, L., Mazet, J., Johnson, C.K., Buchy, P., Dussart, P., Goldstein, T., Simon-Loriere, E., Duong, V., 2021. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. Nat. Commun. 12, 6563. https://doi.org/10.1038/s41467-021-26809-4.

Drexler, J.F., Corman, V.M., Drosten, C., 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antiviral. Res. 101, 45−56. https://doi.org/10.1016/j.antiviral.2013.10.013.

Drexler, J.F., Gloza-Rausch, F., Glende, J., Corman, V.M., Muth, D., Goettsche, M., Seebens, A., Niedrig, M., Pfefferle, S., Yordanov, S., Zhelyazkov, L., Hermanns, U., Vallo, P., Lukashev, A., Muller, M.A., Deng, H., Herrler, G., Drosten, C., 2010. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. J. Virol. 84, 11336−11349. https://doi.org/10.1128/JVI.00650-10.

Drosten, C., Gunther, S., Preiser, W., Van Der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A., Berger, A., Burguiere, A.M., Cinatl, J., Eickmann, M., Escriou, N.,

Grywna, K., Kramme, S., Manuguerra, J.C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H.D., Osterhaus, A.D., Schmitz, H., Doerr, H.W., 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. 348, 1967−1976. https://doi.org/10.1056/NEJMoa030747.

Fan, Y., Zhao, K., Shi, Z.L., Zhou, P., 2019. Bat coronaviruses in China. Viruses 11. https://doi.org/10.3390/v11030210.

Gaynor, A.M., Nissen, M.D., Whiley, D.M., Mackay, I.M., Lambert, S.B., Wu, G., Brennan, D.C., Storch, G.A., Sloots, T.P., Wang, D., 2007. Identification of a novel polyomavirus from patients with acute respiratory tract infections. PLoS Pathog. 3, e64. https://doi.org/10.1371/journal.ppat.0030064.

Ge, X., Li, Y., Yang, X., Zhang, H., Zhou, P., Zhang, Y., Shi, Z., 2012. Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. J. Virol. 86, 4620−4630. https://doi.org/10.1128/JVI.06671-11.

Ge, X.Y., Li, J.L., Yang, X.L., Chmura, A.A., Zhu, G., Epstein, J.H., Mazet, J.K., Hu, B., Zhang, W., Peng, C., Zhang, Y.J., Luo, C.M., Tan, B., Wang, N., Zhu, Y., Crameri, G., Zhang, S.Y., Wang, L.F., Daszak, P., Shi, Z.L., 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503, 535−538. https://doi.org/10.1038/nature12711.

Graham, R.L., Baric, R.S., 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. 84, 3134−3146. https://doi.org/10.1128/JVI.01394-09.

Guan, Y., Zheng, B.J., He, Y.Q., Liu, X.L., Zhuang, Z.X., Cheung, C.L., Luo, S.W., Li, P.H., Zhang, L.J., Guan, Y.J., Butt, K.M., Wong, K.L., Chan, K.W., Lim, W., Shortridge, K.F., Yuen, K.Y., Peiris, J.S., Poon, L.L., 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302, 276−278. https://doi.org/10.1126/science.1087139.

Guo, H., Hu, B.J., Yang, X.L., Zeng, L.P., Li, B., Ouyang, S., Shi, Z.L., 2020. Evolutionary arms race between virus and host drives genetic diversity in bat severe acute respiratory syndrome-related coronavirus spike genes. J. Virol. 94. https://doi.org/10.1128/JVI.00902-20.

He, B., Gong, W., Yan, X., Zhao, Z., Yang, L., Tan, Z., Xu, L., Zhu, A., Zhang, J., Rao, J., Yu, X., Jiang, J., Lu, Z., Zhang, Y., Wu, J., Li, Y., Shi, Y., Jiang, Q., Chen, X., Tu, C., 2021. Viral metagenome-based precision surveillance of pig population at large scale reveals viromic signatures of sample types and influence of farming management on pig virome. mSystems 6, e0042021. https://doi.org/10.1128/mSystems.00420-21.

He, B., Zhang, Y., Xu, L., Yang, W., Yang, F., Feng, Y., Xia, L., Zhou, J., Zhen, W., Feng, Y., Guo, H., Zhang, H., Tu, C., 2014. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. J. Virol. 88, 7070−7082. https://doi.org/10.1128/JVI.00631-14.

Hu, B., Ge, X., Wang, L.F., Shi, Z., 2015. Bat origin of human coronaviruses. Virol. J. 12, 221. https://doi.org/10.1186/s12985-015-0422-1.

Hu, B., Guo, H., Zhou, P., Shi, Z.L., 2021. Characteristics of SARS-CoV-2 and COVID-19. Nat. Rev. Microbiol. 19, 141−154. https://doi.org/10.1038/s41579-020-00459-7.

Hu, B., Zeng, L.P., Yang, X.L., Ge, X.Y., Zhang, W., Li, B., Xie, J.Z., Shen, X.R., Zhang, Y.Z., Wang, N., Luo, D.S., Zheng, X.S., Wang, M.N., Daszak, P., Wang, L.F., Cui, J., Shi, Z.L., 2017. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathog. 13, e1006698. https://doi.org/10.1371/journal.ppat.1006698.

International Committee on Taxonomy of Viruses Executive Committee, 2020. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. Nat. Microbiol. 5, 668−674. https://doi.org/10.1038/s41564-020-0709-x.

Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P., 2008. Global trends in emerging infectious diseases. Nature 451, 990−993. https://doi.org/10.1038/nature06536.

Kan, B., Wang, M., Jing, H., Xu, H., Jiang, X., Yan, M., Liang, W., Zheng, H., Wan, K., Liu, Q., Cui, B., Xu, Y., Zhang, E., Wang, H., Ye, J., Li, G., Li, M., Cui, Z., Qi, X., Chen, K., Du, L., Gao, K., Zhao, Y.T., Zou, X.Z., Feng, Y.J., Gao, Y.F., Hai, R., Yu, D., Guan, Y., Xu, J., 2005. Molecular evolution analysis and geographic

investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. J. Virol. 79, 11892−11900. https://doi.org/10.1128/JVI.79.18.11892-11900.2005.

Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., Rollin, P.E., Dowell, S.F., Ling, A.E., Humphrey, C.D., Shieh, W.J., Guarner, J., Paddock, C.D., Rota, P., Fields, B., Derisi, J., Yang, J.Y., Cox, N., Hughes, J.M., Leduc, J.W., Bellini, W.J., Anderson, L.J., Group, S.W., 2003. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. 348, 1953−1966. https://doi.org/10.1056/NEJMoa030781.

Lan, Y.C., Liu, T.T., Yang, J.Y., Lee, C.M., Chen, Y.J., Chan, Y.J., Lu, J.J., Liu, H.F., Hsiung, C.A., Ho, M.S., Hsiao, K.J., Chen, H.Y., Chen, Y.M., 2005. Molecular epidemiology of severe acute respiratory syndrome-associated coronavirus infections in Taiwan. J. Infect. Dis. 191, 1478−1489. https://doi.org/10.1086/428591.

Lau, S.K., Feng, Y., Chen, H., Luk, H.K., Yang, W.H., Li, K.S., Zhang, Y.Z., Huang, Y., Song, Z.Z., Chow, W.N., Fan, R.Y., Ahmed, S.S., Yeung, H.C., Lam, C.S., Cai, J.P., Wong, S.S., Chan, J.F., Yuen, K.Y., Zhang, H.L., Woo, P.C., 2015. Severe acute respiratory syndrome (SARS) coronavirus ORF8 protein is acquired from SARS-related coronavirus from greater horseshoe bats through recombination. J. Virol. 89, 10532−10547. https://doi.org/10.1128/JVI.01048-15.

Lau, S.K., Li, K.S., Huang, Y., Shek, C.T., Tse, H., Wang, M., Choi, G.K., Xu, H., Lam, C.S., Guo, R., Chan, K.H., Zheng, B.J., Woo, P.C., Yuen, K.Y., 2010. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. J. Virol. 84, 2808−2819. https://doi.org/10.1128/JVI.02219-09.

Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Tsoi, H.W., Wong, B.H., Wong, S.S., Leung, S.Y., Chan, K.H., Yuen, K.Y., 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc. Natl. Acad. Sci. U. S. A. 102, 14040−14045. https://doi.org/10.1073/pnas.0506735102.

Li, B., Si, H.R., Zhu, Y., Yang, X.L., Anderson, D.E., Shi, Z.L., Wang, L.F., Zhou, P., 2020a. Discovery of bat coronaviruses through surveillance and probe capture-based next-generation sequencing. mSphere 5. https://doi.org/10.1128/mSphere.00807-19.

Li, F., Li, W., Farzan, M., Harrison, S.C., 2005a. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309, 1864−1868. https://doi.org/10.1126/science.1116480.

Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., Tu, W., Chen, C., Jin, L., Yang, R., Wang, Q., Zhou, S., Wang, R., Liu, H., Luo, Y., Liu, Y., Shao, G., Li, H., Tao, Z., Yang, Y., Deng, Z., Liu, B., Ma, Z., Zhang, Y., Shi, G., Lam, T.T.Y., Wu, J.T., Gao, G.F., Cowling, B.J., Yang, B., Leung, G.M., Feng, Z., 2020b. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med. 382, 1199−1207. https://doi.org/10.1056/NEJMoa2001316.

Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., Choe, H., Farzan, M., 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature 426, 450−454. https://doi.org/10.1038/nature02145.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., Mceachern, J., Field, H., Daszak, P., Eaton, B.T., Zhang, S., Wang, L.F., 2005b. Bats are natural reservoirs of SARS-like coronaviruses. Science 310, 676−679. https://doi.org/10.1126/science.1118391.

Li, W., Zhang, C., Sui, J., Kuhn, J.H., Moore, M.J., Luo, S., Wong, S.K., Huang, I.C., Xu, K., Vasilieva, N., Murakami, A., He, Y., Marasco, W.A., Guan, Y., Choe, H., Farzan, M., 2005c. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. EMBO J. 24, 1634−1643. https://doi.org/10.1038/sj.emboj.7600640.

Li, X., Wang, L., Liu, P., Li, H., Huo, S., Zong, K., Zhu, S., Guo, Y., Zhang, L., Hu, B., Lan, Y., Chmura, A., Wu, G., Daszak, P., Liu, W.J., Gao, G.F., 2021a. A novel potentially recombinant rodent coronavirus with a polybasic cleavage site in the spike protein. J. Virol. 95, e0117321. https://doi.org/10.1128/JVI.01173-21.

Li, X., Zhang, L., Chen, S., Ouyang, H., Ren, L., 2021b. Possible targets of pan-coronavirus antiviral strategies for emerging or re-emerging coronaviruses. Microorganisms 9. https://doi.org/10.3390/microorganisms9071479.

Lim, W., Ng, K.C., Tsang, D.N., 2006. Laboratory containment of SARS virus. Ann. Acad. Med. Singap. 35, 354−360.

Lin, X.D., Wang, W., Hao, Z.Y., Wang, Z.X., Guo, W.P., Guan, X.Q., Wang, M.R., Wang, H.W., Zhou, R.H., Li, M.H., Tang, G.P., Wu, J., Holmes, E.C., Zhang, Y.Z., 2017. Extensive diversity of coronaviruses in bats from China. Virology 507, 1−10. https://doi.org/10.1016/j.virol.2017.03.019.

Lipkin, W.I., Anthony, S.J., 2015. Virus hunting. Virology 479−480, 194−199. https://doi.org/10.1016/j.virol.2015.02.006.

Liu, M.Q., Jiang, R.D., Guo, J., Chen, Y., Yang, D.S., Wang, X., Lin, H.F., Li, A., Li, B., Hu, B., Wang, Z.J., Yang, X.L., Shi, Z.L., 2022a. Inactivated SARS-CoV-2 vaccine shows cross-protection against bat SARS-related coronaviruses in human ACE2 transgenic mice. J. Virol. 96, e0016922. https://doi.org/10.1128/jvi.00169-22.

Liu, Z., Xu, W., Chen, Z., Fu, W., Zhan, W., Gao, Y., Zhou, J., Zhou, Y., Wu, J., Wang, Q., Zhang, X., Hao, A., Wu, W., Zhang, Q., Li, Y., Fan, K., Chen, R., Jiang, Q., Mayer, C.T., Schoofs, T., Xie, Y., Jiang, S., Wen, Y., Yuan, Z., Wang, K., Lu, L., Sun, L., Wang, Q., 2022b. An ultrapotent pan-beta-coronavirus lineage B (beta-CoV-B) neutralizing antibody locks the receptor-binding domain in closed conformation by targeting its conserved epitope. Protein Cell 13, 655−675. https://doi.org/10.1007/s13238-021-00871-6.

Lu, L., Su, S., Yang, H., Jiang, S., 2021. Antivirals with common targets against highly pathogenic viruses. Cell 184, 1604−1620. https://doi.org/10.1016/j.cell.2021.02.013.

Ma, C., Su, S., Wang, J., Wei, L., Du, L., Jiang, S., 2020. From SARS-CoV to SARS-CoV-2: safety and broad-spectrum are important for coronavirus vaccine development. Microbes. Infect. 22, 245−253. https://doi.org/10.1016/j.micinf.2020.05.004.

Malone, B., Campbell, E.A., 2021. Molnupiravir: coding for catastrophe. Nat. Struct. Mol. Biol. 28, 706−708. https://doi.org/10.1038/s41594-021-00657-8.

Marra, M.A., Jones, S.J., Astell, C.R., Holt, R.A., Brooks-Wilson, A., Butterfield, Y.S., Khattra, J., Asano, J.K., Barber, S.A., Chan, S.Y., Cloutier, A., Coughlin, S.M., Freeman, D., Girn, N., Griffith, O.L., Leach, S.R., Mayo, M., Mcdonald, H., Montgomery, S.B., Pandoh, P.K., Petrescu, A.S., Robertson, A.G., Schein, J.E., Siddiqui, A., Smailus, D.E., Stott, J.M., Yang, G.S., Plummer, F., Andonov, A., Artsob, H., Bastien, N., Bernard, K., Booth, T.F., Bowness, D., Czub, M., Drebot, M., Fernando, L., Flick, R., Garbutt, M., Gray, M., Grolla, A., Jones, S., Feldmann, H., Meyers, A., Kabani, A., Li, Y., Normand, S., Stroher, U., Tipples, G.A., Tyler, S., Vogrig, R., Ward, D., Watson, B., Brunham, R.C., Krajden, M., Petric, M., Skowronski, D.M., Upton, C., Roper, R.L., 2003. The Genome sequence of the SARS-associated coronavirus. Science 300, 1399−1404. https://doi.org/10.1126/science.1085953.

Menachery, V.D., Yount Jr., B.L., Debbink, K., Agnihothram, S., Gralinski, L.E., Plante, J.A., Graham, R.L., Scobey, T., Ge, X.Y., Donaldson, E.F., Randell, S.H., Lanzavecchia, A., Marasco, W.A., Shi, Z.L., Baric, R.S., 2015. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. Nat. Med. 21, 1508−1513. https://doi.org/10.1038/nm.3985.

Peiris, J.S., Lai, S.T., Poon, L.L., Guan, Y., Yam, L.Y., Lim, W., Nicholls, J., Yee, W.K., Yan, W.W., Cheung, M.T., Cheng, V.C., Chan, K.H., Tsang, D.N., Yung, R.W., Ng, T.K., Yuen, K.Y., Group, S.S., 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet 361, 1319−1325. https://doi.org/10.1016/s0140-6736(03)13077-2.

Pomorska-Mol, M., Ruszkowski, J.J., Gogulski, M., Domanska-Blicharz, K., 2022. First detection of Hedgehog coronavirus 1 in Poland. Sci. Rep. 12, 2386. https://doi.org/10.1038/s41598-022-06432-z.

Poon, L.L., Chu, D.K., Chan, K.H., Wong, O.K., Ellis, T.M., Leung, Y.H., Lau, S.K., Woo, P.C., Suen, K.Y., Yuen, K.Y., Guan, Y., Peiris, J.S., 2005. Identification of a novel coronavirus in bats. J. Virol. 79, 2001−2009. https://doi.org/10.1128/JVI.79.4.2001-2009.2005.

Qu, X.X., Hao, P., Song, X.J., Jiang, S.M., Liu, Y.X., Wang, P.G., Rao, X., Song, H.D., Wang, S.Y., Zuo, Y., Zheng, A.H., Luo, M., Wang, H.L., Deng, F., Wang, H.Z., Hu, Z.H., Ding, M.X., Zhao, G.P., Deng, H.K., 2005. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. J. Biol. Chem. 280, 29588−29595. https://doi.org/10.1074/jbc.M500662200.

Rappazzo, C.G., Tse, L.V., Kaku, C.I., Wrapp, D., Sakharkar, M., Huang, D., Deveau, L.M., Yockachonis, T.J., Herbert, A.S., Battles, M.B., O'brien, C.M., Brown, M.E., Geoghegan, J.C., Belk, J., Peng, L., Yang, L., Hou, Y., Scobey, T.D., Burton, D.R., Nemazee, D., Dye, J.M., Voss, J.E., Gunn, B.M., Mclellan, J.S., Baric, R.S., Gralinski, L.E., Walker, L.M., 2021. Broad and potent activity against SARS-like viruses by an engineered human monoclonal antibody. Science 371, 823−829. https://doi.org/10.1126/science.abf4830.

Ren, W., Li, W., Yu, M., Hao, P., Zhang, Y., Zhou, P., Zhang, S., Zhao, G., Zhong, Y., Wang, S., Wang, L.F., Shi, Z., 2006. Full-length genome sequences of two SARS-like coronaviruses in horseshoe bats and genetic variation analysis. J. Gen. Virol. 87, 3355−3359. https://doi.org/10.1099/vir.0.82220-0.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Penaranda, S., Bankamp, B., Maher, K., Chen, M.H., Tong, S., Tamin, A., Lowe, L., Frace, M., Derisi, J.L., Chen, Q., Wang, D., Erdman, D.D., Peret, T.C., Burns, C., Ksiazek, T.G., Rollin, P.E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., Mccaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A.D., Drosten, C., Pallansch, M.A., Anderson, L.J., Bellini, W.J., 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300, 1394−1399. https://doi.org/10.1126/science.1085952.

Ruiz-Aravena, M., Mckee, C., Gamble, A., Lunn, T., Morris, A., Snedden, C.E., Yinda, C.K., Port, J.R., Buchholz, D.W., Yeo, Y.Y., Faust, C., Jax, E., Dee, L., Jones, D.N., Kessler, M.K., Falvo, C., Crowley, D., Bharti, N., Brook, C.E., Aguilar, H.C., Peel, A.J., Restif, O., Schountz, T., Parrish, C.R., Gurley, E.S., Lloyd-Smith, J.O., Hudson, P.J., Munster, V.J., Plowright, R.K., 2022. Ecology, evolution and spillover of coronaviruses from bats. Nat. Rev. Microbiol. 20, 299−314. https://doi.org/10.1038/s41579-021-00652-2.

Shi, Z., Hu, Z., 2008. A review of studies on animal reservoirs of the SARS coronavirus. Virus Res. 133, 74−87. https://doi.org/10.1016/j.virusres.2007.03.012.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., Gorbalenya, A.E., 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. 331, 991−1004. https://doi.org/10.1016/s0022-2836(03)00865-9.

Song, H.D., Tu, C.C., Zhang, G.W., Wang, S.Y., Zheng, K., Lei, L.C., Chen, Q.X., Gao, Y.W., Zhou, H.Q., Xiang, H., Zheng, H.J., Chern, S.W., Cheng, F., Pan, C.M., Xuan, H., Chen, S.J., Luo, H.M., Zhou, D.H., Liu, Y.F., He, J.F., Qin, P.Z., Li, L.H., Ren, Y.Q., Liang, W.J., Yu, Y.D., Anderson, L., Wang, M., Xu, R.H., Wu, X.W., Zheng, H.Y., Chen, J.D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L.Y., Li, H., Chen, F., Di, B., He, L.J., Lin, J.Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X.J., Spiga, O., Guo, Z.M., Pan, H.Y., He, W.Z., Manuguerra, J.C., Fontanet, A., Danchin, A., Niccolai, N., Li, Y.X., Wu, C.I., Zhao, G.P., 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc. Natl. Acad. Sci. U. S. A. 102, 2430−2435. https://doi.org/10.1073/pnas.0409608102.

Starr, T.N., Zepeda, S.K., Walls, A.C., Greaney, A.J., Alkhovsky, S., Veesler, D., Bloom, J.D., 2022. ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. Nature 603, 913−918. https://doi.org/10.1038/s41586-022-04464-z.

Tang, J.W., Cheung, J.L., Chu, I.M., Ip, M., Hui, M., Peiris, M., Chan, P.K., 2007. Characterizing 56 complete SARS-CoV S-gene sequences from Hong Kong. J. Clin. Virol. 38, 19−26. https://doi.org/10.1016/j.jcv.2006.10.001.

Tang, X., Li, G., Vasilakis, N., Zhang, Y., Shi, Z., Zhong, Y., Wang, L.F., Zhang, S., 2009. Differential stepwise evolution of SARS coronavirus functional proteins in different host species. BMC Evol. Biol. 9, 52. https://doi.org/10.1186/1471-2148-9-52.

Tao, Y., Tong, S., 2019. Complete genome sequence of a severe acute respiratory syndrome-related coronavirus from Kenyan bats. Microbiol. Resour. Announc. 8. https://doi.org/10.1128/MRA.00548-19.

Temmam, S., Vongphayloth, K., Baquero, E., Munier, S., Bonomi, M., Regnault, B., Douangboubpha, B., Karami, Y., Chretien, D., Sanamxay, D., Xayaphet, V., Paphaphanh, P., Lacoste, V., Somlor, S., Lakeomany, K., Phommavanh, N., Perot, P., Dehan, O., Amara, F., Donati, F., Bigot, T., Nilges, M., Rey, F.A., Van Der Werf, S., Brey, P.T., Eloit, M., 2022. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature 604, 330−336. https://doi.org/10.1038/s41586-022-04532-4.

Tortorici, M.A., Czudnochowski, N., Starr, T.N., Marzi, R., Walls, A.C., Zatta, F., Bowen, J.E., Jaconi, S., Di Iulio, J., Wang, Z., De Marco, A., Zepeda, S.K., Pinto, D., Liu, Z., Beltramello, M., Bartha, I., Housley, M.P., Lempp, F.A., Rosen, L.E., Dellota Jr., E., Kaiser, H., Montiel-Ruiz, M., Zhou, J., Addetia, A., Guarino, B., Culap, K., Sprugasci, N., Saliba, C., Vetti, E., Giacchetto-Sasselli, I., Fregni, C.S., Abdelnabi, R., Foo, S.C., Havenar-Daughton, C., Schmid, M.A., Benigni, F., Cameroni, E., Neyts, J., Telenti, A., Virgin, H.W., Whelan, S.P.J., Snell, G., Bloom, J.D., Corti, D., Veesler, D., Pizzuto, M.S., 2021. Broad sarbecovirus neutralization by a human monoclonal antibody. Nature 597, 103−108. https://doi.org/10.1038/s41586-021-03817-4.

Tu, C., Crameri, G., Kong, X., Chen, J., Sun, Y., Yu, M., Xiang, H., Xia, X., Liu, S., Ren, T., Yu, Y., Eaton, B.T., Xuan, H., Wang, L.F., 2004. Antibodies to SARS coronavirus in civets. Emerg. Infect. Dis. 10, 2244−2248. https://doi.org/10.3201/eid1012.040520.

Wacharapluesadee, S., Tan, C.W., Maneeorn, P., Duengkae, P., Zhu, F., Joyjinda, Y., Kaewpom, T., Chia, W.N., Ampoot, W., Lim, B.L., Worachotsueptrakun, K., Chen, V.C., Sirichan, N., Ruchisrisarod, C., Rodpan, A., Noradechanon, K., Phaichana, T., Jantarat, N., Thongnumchaima, B., Tu, C., Crameri, G., Stokes, M.M., Hemachudha, T., Wang, L.F., 2021. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. Nat. Commun. 12, 972. https://doi.org/10.1038/s41467-021-21240-1.

Wang, L.F., Eaton, B.T., 2007. Bats, civets and the emergence of SARS. Curr. Top. Microbiol. Immunol. 315, 325−344. https://doi.org/10.1007/978-3-540-70962-6_13.

Wang, L.F., Shi, Z., Zhang, S., Field, H., Daszak, P., Eaton, B.T., 2006. Review of bats and SARS. Emerg. Infect. Dis. 12, 1834−1840. https://doi.org/10.3201/eid1212.060401.

Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Chen, H., Zheng, H., Xu, Y., Zhang, E., Wang, H., Ye, J., Li, G., Li, M., Cui, Z., Liu, Y.F., Guo, R.T., Liu, X.N., Zhan, L.H., Zhou, D.H., Zhao, A., Hai, R., Yu, D., Guan, Y., Xu, J., 2005. SARS-CoV infection in a restaurant from palm civet. Emerg. Infect. Dis. 11, 1860−1865. https://doi.org/10.3201/eid1112.041293.

Wang, Q., Qi, J., Yuan, Y., Xuan, Y., Han, P., Wan, Y., Ji, W., Li, Y., Wu, Y., Wang, J., Iwamoto, A., Woo, P.C., Yuen, K.Y., Yan, J., Lu, G., Gao, G.F., 2014. Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. Cell Host Microbe 16, 328−337. https://doi.org/10.1016/j.chom.2014.08.009.

Wang, W., Lin, X.D., Liao, Y., Guan, X.Q., Guo, W.P., Xing, J.G., Holmes, E.C., Zhang, Y.Z., 2017. Discovery of a highly divergent coronavirus in the asian house shrew from China illuminates the origin of the alphacoronaviruses. J. Virol. 91. https://doi.org/10.1128/JVI.00764-17.

Wang, X., Xia, S., Zhu, Y., Lu, L., Jiang, S., 2021. Pan-coronavirus fusion inhibitors as the hope for today and tomorrow. Protein. Cell 12, 84−88. https://doi.org/10.1007/s13238-020-00806-7.

Wec, A.Z., Wrapp, D., Herbert, A.S., Maurer, D.P., Haslwanter, D., Sakharkar, M., Jangra, R.K., Dieterle, M.E., Lilov, A., Huang, D.L., Tse, L.V., Johnson, N.V., Hsieh, C.L., Wang, N.S., Nett, J.H., Champney, E., Burnina, I., Brown, M., Lin, S., Sinclair, M., Johnson, C., Pudi, S., Bortz, R., Wirchnianski, A.S., Laudermilch, E., Florez, C., Fels, J.M., O'brien, C.M., Graham, B.S., Nemazee, D., Burton, D.R., Baric, R.S., Voss, J.E., Chandran, K., Dye, J.M., Mclellan, J.S., Walker, L.M., 2020. Broad neutralization of SARS-related viruses by human monoclonal antibodies. Science 369, 731. https://doi.org/10.1126/science.abc7424.

WHO, 2003. Severe acute respiratory syndrome (SARS). Wkly. Epidemiol. Rec. 78, 81−88.

Wong, S.K., Li, W., Moore, M.J., Choe, H., Farzan, M., 2004. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. J. Biol. Chem. 279, 3197−3201. https://doi.org/10.1074/jbc.C300520200.

Woo, P.C., Lau, S.K., Huang, Y., Yuen, K.Y., 2009a. Coronavirus diversity, phylogeny and interspecies jumping. Exp. Biol. Med. 234, 1117−1127. https://doi.org/10.3181/0903-MR-94.

Woo, P.C., Lau, S.K., Lam, C.S., Lai, K.K., Huang, Y., Lee, P., Luk, G.S., Dyrting, K.C., Chan, K.H., Yuen, K.Y., 2009b. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J. Virol. 83, 908−917. https://doi.org/10.1128/JVI.01977-08.

Woo, P.C., Lau, S.K., Lam, C.S., Lau, C.C., Tsang, A.K., Lau, J.H., Bai, R., Teng, J.L., Tsang, C.C., Wang, M., Zheng, B.J., Chan, K.H., Yuen, K.Y., 2012. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J. Virol. 86, 3995−4008. https://doi.org/10.1128/JVI.06540-11.

Woolhouse, M.E., Gowtage-Sequeria, S., 2005. Host range and emerging and reemerging pathogens. Emerg. Infect. Dis. 11, 1842−1847. https://doi.org/10.3201/eid1112.050997.

Wu, D., Tu, C., Xin, C., Xuan, H., Meng, Q., Liu, Y., Yu, Y., Guan, Y., Jiang, Y., Yin, X., Crameri, G., Wang, M., Li, C., Liu, S., Liao, M., Feng, L., Xiang, H., Sun, J., Chen, J., Sun, Y., Gu, S., Liu, N., Fu, D., Eaton, B.T., Wang, L.F., Kong, X., 2005. Civets are equally susceptible to experimental infection by two different severe acute respiratory syndrome coronavirus isolates. J. Virol. 79, 2620−2625. https://doi.org/10.1128/JVI.79.4.2620-2625.2005.

Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579, 265−269. https://doi.org/10.1038/s41586-020-2008-3.

Wu, Z., Yang, L., Ren, X., He, G., Zhang, J., Yang, J., Qian, Z., Dong, J., Sun, L., Zhu, Y., Du, J., Yang, F., Zhang, S., Jin, Q., 2016a. Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. ISME J. 10, 609−620. https://doi.org/10.1038/ismej.2015.138.

Wu, Z., Yang, L., Ren, X., Zhang, J., Yang, F., Zhang, S., Jin, Q., 2016b. ORF8-related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. J. Infect. Dis. 213, 579−583. https://doi.org/10.1093/infdis/jiv476.

Xia, S., Liu, M., Wang, C., Xu, W., Lan, Q., Feng, S., Qi, F., Bao, L., Du, L., Liu, S., Qin, C., Sun, F., Shi, Z., Zhu, Y., Jiang, S., Lu, L., 2020. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. Cell Res. 30, 343−355. https://doi.org/10.1038/s41422-020-0305-x.

Xu, H.F., Wang, M., Zhang, Z.B., Zou, X.Z., Gao, Y., Liu, X.N., Lu, E.J., Pan, B.Y., Wu, S.J., Yu, S.Y., 2004. [An epidemiologic investigation on infection with severe acute respiratory syndrome coronavirus in wild animals traders in Guangzhou]. Zhonghua Yufang Yixue Zazhi 38, 81−83.

Yanai-Balser, G.M., Duncan, G.A., Eudy, J.D., Wang, D., Li, X., Agarkova, I.V., Dunigan, D.D., Van Etten, J.L., 2010. Microarray analysis of Paramecium bursaria chlorella virus 1 transcription. J. Virol. 84, 532−542. https://doi.org/10.1128/JVI.01698-09.

Yang, L., Wu, Z., Ren, X., Yang, F., He, G., Zhang, J., Dong, J., Sun, L., Zhu, Y., Du, J., Zhang, S., Jin, Q., 2013. Novel SARS-like betacoronaviruses in bats, China, 2011. Emerg. Infect. Dis. 19, 989−991. https://doi.org/10.3201/eid1906.121648.

Yang, L., Wu, Z., Ren, X., Yang, F., Zhang, J., He, G., Dong, J., Sun, L., Zhu, Y., Zhang, S., Jin, Q., 2014. MERS-related betacoronavirus in Vespertilio superans bats, China. Emerg. Infect. Dis. 20, 1260−1262. https://doi.org/10.3201/eid2007.140318.

Yang, X.L., Hu, B., Wang, B., Wang, M.N., Zhang, Q., Zhang, W., Wu, L.J., Ge, X.Y., Zhang, Y.Z., Daszak, P., Wang, L.F., Shi, Z.L., 2015a. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. J. Virol. 90, 3253−3256. https://doi.org/10.1128/JVI.02582-15.

Yang, Y., Liu, C., Du, L., Jiang, S., Shi, Z., Baric, R.S., Li, F., 2015b. Two mutations were critical for bat-to-human transmission of middle east respiratory syndrome coronavirus. J. Virol. 89, 9119−9123. https://doi.org/10.1128/JVI.01279-15.

Zaki, A.M., Van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. 367, 1814−1820. https://doi.org/10.1056/NEJMoa1211721.

Zhao, G.P., 2007. SARS molecular epidemiology: a Chinese fairy tale of controlling an emerging zoonotic disease in the genomics era. Philos. Trans. R. Soc. Lond. B Biol. Sci. 362, 1063−1081. https://doi.org/10.1098/rstb.2007.2034.

Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E.C., Hughes, A.C., Bi, Y., Shi, W., 2020a. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. Curr. Biol. 30, 2196−2203. https://doi.org/10.1016/j.cub.2020.05.023.

Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., Chen, H.D., Chen, J., Luo, Y., Guo, H., Jiang, R.D., Liu, M.Q., Chen, Y., Shen, X.R., Wang, X., Zheng, X.S., Zhao, K., Chen, Q.J., Deng, F., Liu, L.L., Yan, B., Zhan, F.X., Wang, Y.Y., Xiao, G.F., Shi, Z.L., 2020b. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270−273. https://doi.org/10.1038/s41586-020-2012-7.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., China Novel Coronavirus, I., Research, T., 2020. A novel coronavirus from patients with pneumonia in China, 2019. N. Engl. J. Med. 382, 727−733. https://doi.org/10.1056/NEJMoa2001017.

This page intentionally left blank

CHAPTER

Ecology and Evolution of Avian Influenza A Viruses in Wild Birds

31

Michelle Wille[1,2]

[1]*Centre for Pathogen Genomics, Department of Microbiology and Immunology, The University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia;* [2]*WHO Collaborating Centre for Reference and Research on Influenza, at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia*

1. Role of Wild Birds as a Reservoir for a Diversity of Viruses

Human pandemics, including the ongoing pandemic of SARS-CoV-2, have brought a global focus on wildlife as reservoirs for a variety of zoonotic diseases (Allocati et al., 2016; Brook and Dobson, 2015; Hayman et al., 2013; Lytras et al., 2021). While zoonotic transmission from wild birds to humans remain rare events, birds play a central role as reservoirs or vectors for zoonotic viruses such as influenza A virus (Shi et al., 2018), *West Nile Virus* (Murray et al., 2010), *Crimean-Congo hemorrhagic fever virus* (Lindeborg et al., 2012), *Tick-borne encephalitis virus* (Waldenstrom et al., 2007), *Louping-ill virus* (Jeffries et al., 2014), Newcastle disease virus (*Orthoavulavirus javaense*) (Alexander, 2000a), *Japanese encephalitis virus* (Nemeth et al., 2012), and *St. Louis encephalitis virus* (Gruwell et al., 2000).

In addition to zoonotic viruses, wild birds are reservoirs for viruses which have significant socioeconomic implications for the food production industry. Key examples include influenza A virus (FAO, 2005; Stamoulis, 2017), Newcastle disease virus (Alexander, 1995), and infectious bronchitis virus (*Avian coronavirus* and *Avian coronavirus 9203*) (Ignjatovic and Sapats, 2000). As discussed in more detail below, viruses such as influenza A virus circulate in wild birds, often without any sign of disease, but some of these viruses have enormous consequences following introduction into poultry flocks. For example, outbreaks of highly pathogenic avian influenza H5N1 between 2003 and 2004 cost billions of dollars and resulted in the culling of hundreds of millions of birds (FAO, 2005), with current outbreaks similarly having an enormous impact. Despite the key role that wild birds play as reservoirs for zoonotic and socioeconomically important viruses, rather little is known about the diversity of viruses found in wild birds. The first viruses which were described in birds include Newcastle disease virus and infectious bronchitis virus, which were described almost a century ago. Since then, most viruses described in birds have followed the same formula: a decrease in egg production in layer hens and/or notable morbidity or mortality of food production birds resulted in disease investigations and subsequent virus discovery. With the recognition that wild birds may harbor viruses of concern for the food production industry, in the last decades there has

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00005-7
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

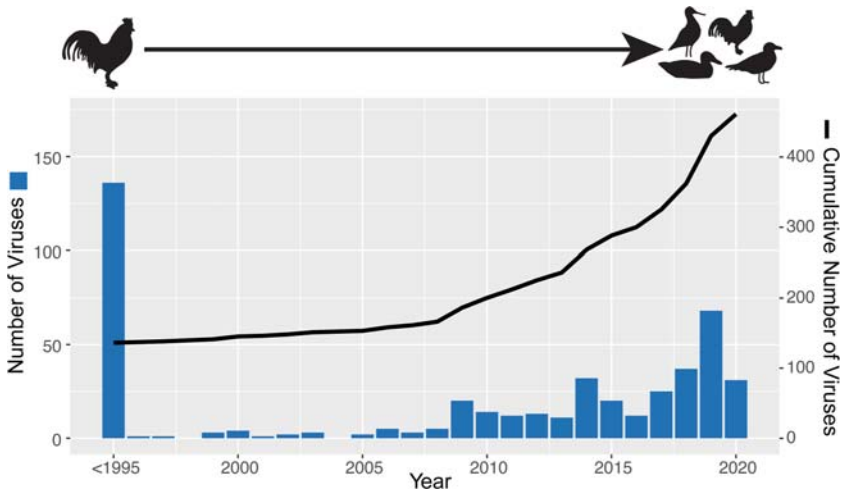


FIGURE 31.1

Increase in viral diversity described in birds over time. With the recognition that wild birds are hosts for numerous viruses infecting poultry and humans, combined with advances in sequencing, metagenomics, and metatranscriptomics, we have seen a dramatic increase in the number of viruses described in wild birds in the last 15 years. Bars indicate the number of new viruses described per year. Line represents the cumulative number of viruses described. All viruses described before 1995 have been presented as a single bar and thus the year is indicated as "<1995."

Figure modified from Wille, M., et al., 2021a. How accurately can we assess zoonotic risk? PLoS Pathog 19, e3001135. https://doi.org/10.1371/journal.pbio.3001135 and distributed under a CC-BY license.

been a shift to not only intensive surveillance of wild birds for known viruses, but also the discovery of novel viruses from apparently healthy wild birds. As of May 2020, 463 different viruses have been described in birds (Fig. 31.1) (Wille et al., 2021a), and novel viruses are continually being described. Indeed, between 2018 and 2021, 83 novel viruses were described in *Charadriiformes* and *Anseriformes* in Australia (Vibin et al., 2018, 2020a,b; Wille et al., 2018a, 2019b, 2021b), and given that in all studies approximately 80% of viruses described were novel, it is likely that a huge diversity of viruses will still be found in wild birds.

Wild birds are unique as potential vectors for emerging infectious diseases given the large species diversity, long life span (in some taxa), seasonal aggregations, spatial population structure, and capacity for long-distance migration. It is through migration that wild birds interconnect all parts of the globe—they transport nutrients, energy, and other organisms between disparate locations, on local, regional, or global scales (Altizer et al., 2011; Bauer and Hoye, 2014; Hubalek, 2004). For example, millions of individual birds utilizing the East Asian Australasian flyway migrate from Oceania, through south-east Asia to Siberia and Alaska each year to breed. However, the major flyways are simplifications, and there are numerous exceptions where species or populations may deviate from the common patterns (Fig. 31.2). These migratory birds connect distinct avian communities in space and time throughout their migratory cycle—breeding areas, moulting sites, stopover sites, or at

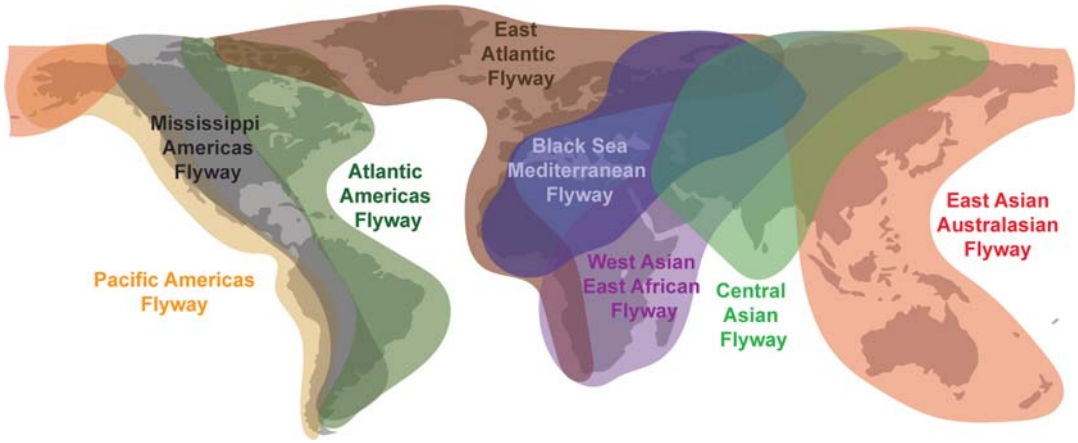


FIGURE 31.2

Migratory flyways of wild bird populations.

Adapted from information collected and analyzed by Wetlands International.

shared overwintering sites (Altizer et al., 2011; Bauer and Hoye, 2014; Hubalek, 2004). As such, the transmission of viruses and their geographical spread is dependent on the ecology of the migrating hosts.

Influenza A virus is a key example of a virus that has a reservoir in wild birds, has enormous consequences if introduced to food production birds, and is zoonotic. The ecology of this virus is defined by its avian hosts, and through migration, novel strains have rapidly spread across the globe.

2. Avian Influenza A Virus

2.1 Introduction

Avian influenza virus is a member of the *Alphainfluenzavirus influenzae* species, in the family *Orthomyxoviridae*. Influenza A virus is a multihost virus, infecting humans (Vijaykrishna et al., 2015), pigs (Vijaykrishna et al., 2011), horses (Daly et al., 2011), dogs (Voorhees et al., 2018), seals (Zohari et al., 2014), bats (Tong et al., 2012, 2013), and an enormous array of wild and domestic bird species (Fig. 31.3 and Table 31.1). Birds are considered the natural reservoir for all influenza A viruses as they maintain a huge diversity of viral subtypes. Further, to date, all pandemic human strains of influenza A virus have comprised at least one segment of avian origin (Guan et al., 2010; Lindstrom et al., 2004; Scholtissek et al., 1978; Smith et al., 2009).

In the avian reservoir, avian influenza A viruses (AIV) are classified by two schemes: subtype and pathotype. First, viruses are classified based on their two surface proteins: hemagglutinin (HA) and neuraminidase (NA), of which there are 18 and 11 subtypes, respectively. Sixteen HA (H1−H16) and nine NA subtypes (N1−N9) have been described in birds (Olsen et al., 2006, 2014); H17−H18 and N10−N11 have thus far only been described in bats (Tong et al., 2012, 2013). The total number of HA-NA combinations described in wild birds is unclear, and Olson et al. (2014) report 102 of the possible

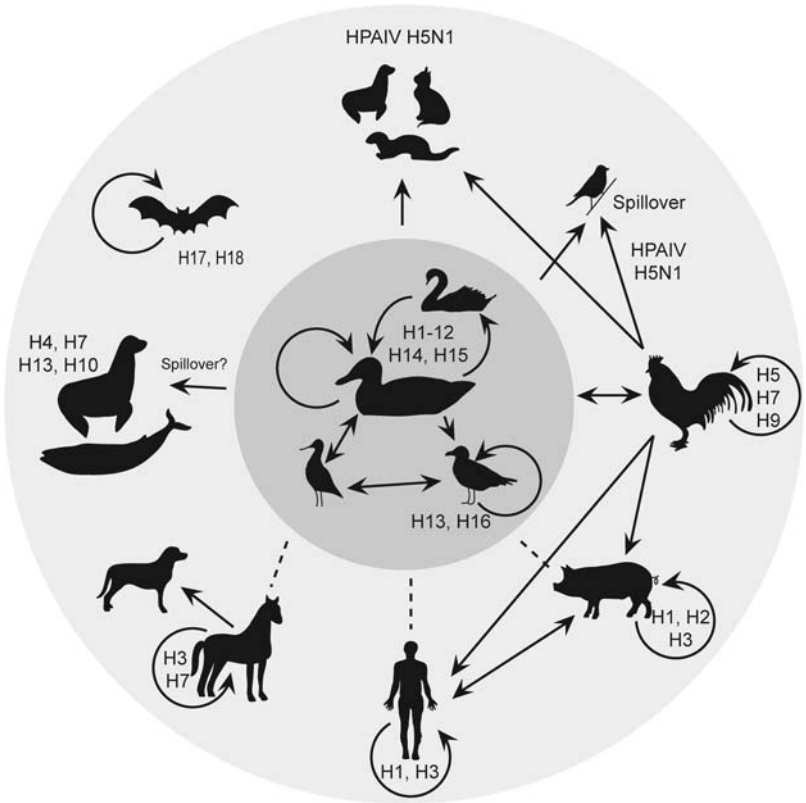


FIGURE 31.3

Host range of influenza A virus. Wild birds of the members Anseriformes (ducks, geese, swans) and Charadriiformes (gulls, shorebirds) are considered the natural reservoir with detections comprising subtypes H1—H16. Humans, pigs, and poultry are important reservoirs for select influenza A virus subtypes, and transmission between these hosts are central to the emergence of human pandemic influenza A viruses. HPAIV H5N1, which emerged and are amplified in poultry, are being detected in a huge variety of wild birds not normally reservoirs for LPAIV, with detections also occurring in a large diversity of mammals including humans. Solid lines are established pathways, dashed lines are potential or infrequently observed pathways.

144 HA-NA avian subtype combinations have been detected in wild birds, globally (Olson et al., 2014) (Fig. 31.4). Mallards (*Anas platyrhynchos*) are hosts for an enormous diversity of HA-NA subtypes, with more than 100 subtype combinations detected in this species alone. Within select subtypes, genetic lineages may be allocated clade names. For example, within H5 viruses, there are numerous clades and subclades, with 2.3.4.4b being highly prominent, globally, as outlined at the end of this chapter.

Second, viruses are subdivided into pathotypes comprising low (LPAIV) and high pathogenic avian influenza virus (HPAIV) (may be referred to as "high pathogenic," "highly pathogenic," or "high

Table 31.1 List of Wild Bird Orders and Families in Which Avian Influenza has Been Detected[a]

Order	Family	Group	No. Host Species in Which LPAIV has Been Detected
Accipitriformes	Accipitridae	Hawks, Eagles, Kites	14
Anseriformes	Anatidae	Ducks, Geese, Swans	91
Casuariiformes	Casuariidae	Cassowaries and Emu	1
Charadriiformes	Alcidae	Auks, Murres, Puffins	6[b]
	Charadriidae	Plovers, Lapwings	3[c]
	Haematopodidae	Oystercatchers	3
	Laridae (including Sternidae)	Gulls, Terns, Skimmers	37
	Recurvirostridae	Stilt, Avocet	2
	Scolopacidae	Sandpipers	18
	Stercorariidae	Skuas, Jaegars	1
Ciconiiformes	Ciconiidae	Storks	3
Columbiformes	Columbidae	Pigeons, Doves	6
Coraciiformes	Meropidae	Bee-eaters	1
Falconiformes	Falconidae	Falcons, Caracaras	5
Galliformes	Cracidae	Guans, Chachalacas, Curassows	1
	Numididae	Guineafowl	1
	Phasianidae	Pheasants, Grouse	11
Gruiformes	Gruidae	Cranes	6
	Rallidae	Rails, Gallinules, Coots	6
Otidiformes	Otididae	Bustards	3
Passeriformes	Aegithinidae	Ioras	1
	Alaudidae	Larks	1
	Corvidae	Crows, Jays, Magpies	6
	Estrilidae	Waxbill, Munia, Firefinches	3
	Fringillidae	Finches	2
	Hirundinidae	Swallows	1
	Irenidae	Fairy-bluebirds	1
	LAniidae	Shrikes	1
	Leiothrichidae	Laughingthrushes	3
	Motacillidae	Wagtails, Pipits	1
	Muscicapidae	Old World Flycatchers	4
	Paridae	Tits, Chikadee	1
	Passeridae	Sparrows	3
	Pycnonotidae	Bulbuls	1
	Sturnidae	Starlings	4
	Turdidae	Thrushes	1

Continued

Table 31.1 List of Wild Bird Orders and Families in Which Avian Influenza has Been Detected[a]—cont'd

Order	Family	Group	No. Host Species in Which LPAIV has Been Detected
	Zosteropidae	White-eyes	1
Pelecaniformes	Ardeidae	Herons, Egrets	9
	Pelecanidae	Pelicans	2
	Threskiornithidae	Ibises, Spoonbills	2
Piciformes	Megalamidae	Barbets	1
Podicipediiformes	Podicipedidae	Grebes	5
Procesllariiformes	Procellariidae	Petrels, Shearwaters	1
Psittaciformes	Psittacidae	New World and African Parrots	2
	Psittaculidae	Old World Parrots	1
Pterocliformes	Pteroclidae	Sandgrouse	1
Spenisciformes	Speniscidae	Penguins	3
Strifiiformes	Strigidae	Owls	4
Suliformes	Phalacrocoracidae	Cormorants	5
Tinamiformes	Tinamidae	Tinamous	1

[a]*Reproduced with minor modification (indicated with an asterisk) from Verhagen et al. (2021).*
[b]*Alcidae have been updated from Verhagen et al. (2021) to include all members identified in Lang et al. (2016).*
[c]*Charadriidae have been updated from Verhagen et al. (2021) to include those species identified in Australia from Wille et al., 2022b.*

pathogenicity"). Birds infected with LPAIV generally show no clinical signs of infection (Olsen et al., 2006; Webster et al., 1992). Two subtypes, H5 and H7, have the potential to occur as HPAIV. HPAIV viruses are classified as those that cause more than 75% mortality in chickens following intravenous infection. The switch from LPAIV to HPAIV occurs following the introduction of LPIAV H5 or H7 into poultry (Alexander, 2000b; Byrne et al., 2021; Richard et al., 2017; Seekings et al., 2018); this transition never been documented in wild bird hosts (Alexander, 2000b, 2007). HPAIV preferentially infects the epithelium of the respiratory tract, including the trachea, lungs, and air sacs (Bröjer et al., 2009; Keawcharoen et al., 2008; Worobey et al., 2014). However, lesions associated with HPAIV have been found in an array of organs following systemic infections (Bröjer et al., 2009). Indeed, neurotropism is widely reported in wild birds and mammals infected with the current panzootic strain of HPAIV (Bauer et al., 2023; Caliendo et al., 2022a). Mechanistically, the switch from LPAIV to HPAIV follows changes in the HA protein. The HA protein of influenza A virus is initially synthesized as a single polypeptide precursor (HA0), which is cleaved into HA1 and HA2 subunits by trypsin-like proteases in the host cell. The switch from the LPAIV to the HPAIV phenotype for H5 and H7 viruses is achieved by the introduction of basic amino acid residues into the HA0 cleavage site, facilitating cleavage by ubiquitously expressed proteases (Webster and Rott, 1987). It is thought that this property facilitates systemic virus replication and a mortality of up to 100% in poultry.

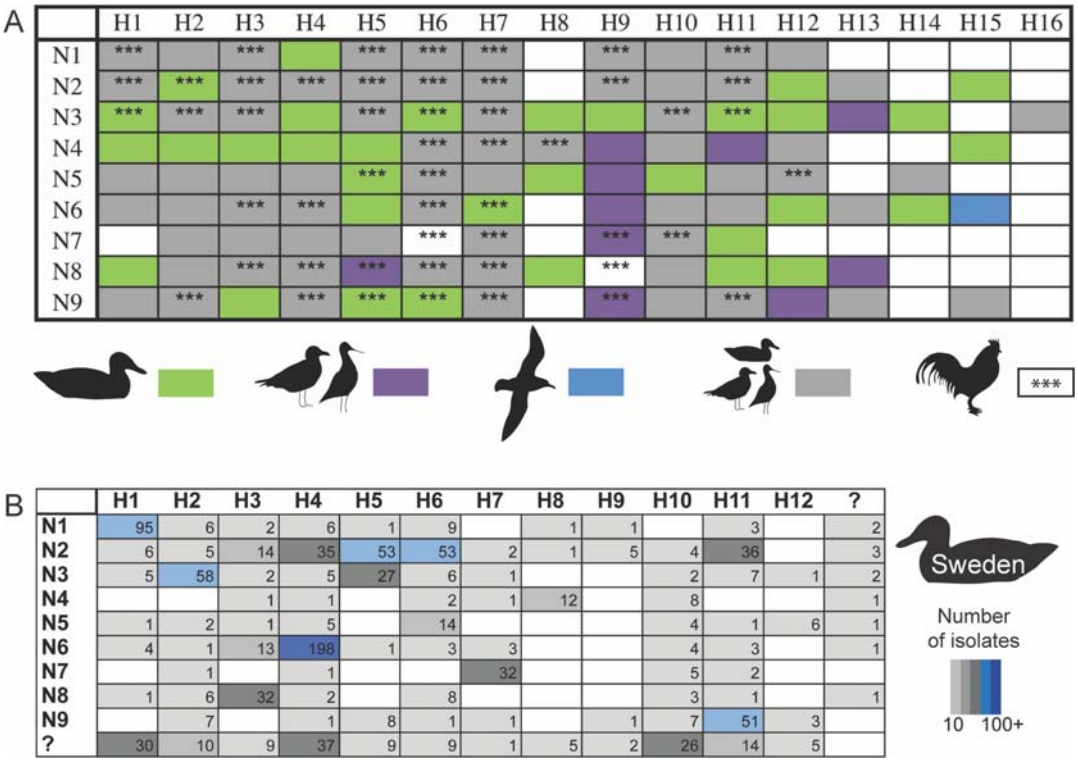


FIGURE 31.4

Subtype diversity of LPAIV. (A) Global subtype distribution and detection across different avian host groups: Anseriformes (*green*), Charadriiformes (*purple*), Procellariiformes (*blue*), more than one order (*gray*), and poultry (*asterisk*). (B) Subtype distribution and frequency in Mallards utilizing Ottenby Bird Observatory in Sweden from 2002 to 2009. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

(A) Modified from Olson, S.H., et al., 2014. Sampling strategies and biodiversity of influenza A subtypes in wild birds. PLoS One 9, e90826. https://doi.org/10.1371/journal.pone.0090826 and distributed under a CC-BY licence. (B) Modified from Latorre-Margalef, N., et al. 2014. Long-term variation in influenza A virus prevalence and subtype diversity in a migratory mallards in Northern Europe. Proc. R. Soc. B 281. https://doi.org/10.1098/rspb.2014.0098

I have devoted this chapter to LPAIV, with the exception of the final section wherein I outline the role of wild birds in the current HPAIV panzootic. The vast majority of avian influenza viruses are LPAIV and LPAIVs have been isolated from >280 wild bird species, belonging to more than 50 families (Verhagen et al., 2021), with different species playing different roles in the ecology. Considering wild birds, not all orders play the same role in LPAIV maintenance. Members of the *Anseriformes* (ducks, geese, and swans) and select members of the *Charadriiformes* (gulls, shorebirds, seabirds) are central to the ecology of this virus (Olsen et al., 2006). Bird taxa such as the *Passeriiformes* (songbirds), which comprises the most species of any avian order, are poor hosts for LPAIV, with many studies recording no viral detection and no anti-NP antibodies across thousands of samples

collected (Slusher et al., 2014). The role of other waterbirds, such as grebes, herons, cormorants, and pelicans, is less clear. A number of studies have detected viruses in these avian groups, although in large global studies they are either undersampled or prevalence is low (Cross et al., 2013; Lebarbenchon et al., 2015b; Simulundu et al., 2009; Wang et al., 2012).

2.2 Subtype Diversity in Birds

While a large diversity of subtypes are found in birds, these subtypes and subtype combinations are not detected at equal frequencies. Within Mallards, some subtypes are very common, or overrepresented in datasets, while others are rare, or entirely absent. Common subtypes include H3, H4, and H6, and comprise the vast majority of viruses isolated at study sites in North America and Europe (Bahl et al., 2009, 2013; Latorre-Margalef et al., 2014; Verhagen et al., 2014b, 2017; Wilcox et al., 2011). In contrast, there are a number of HA subtypes that are rare or uncommon in Mallard dominated surveillance systems. These subtypes may occur annually at low numbers, such as H11, may be isolated only sporadically, e.g., H8, H12, or may be entirely absent e.g., H13−H16 (Wille et al., 2018b). One hypothesis stipulates that subtypes rarely found in Mallard-dominated study sites may have reservoirs in other hosts, such as shorebirds (Latorre-Margalef et al., 2014), seabirds, infrequently sampled waterfowl such as diving ducks or geese, or other waterbirds such as cormorants (Wille et al., 2018b). Specifically, due to the focus of surveillance systems on waterfowl, which are critical in early detection of HPAIV, there has been limited surveillance of other wild birds which may potentially be reservoirs for LPAIV.

In addition to broad patterns of HA presence or absence, some HA-NA combinations are overrepresented, such as H4N6, H6N2, H3N8, which are frequently isolated. While H6N2 is one of the more common subtype combinations at Ottenby Bird Observatory, Sweden ($n = 53$), the isolation of an H6 virus with an NA subtype other than N2 is very uncommon (Fig. 31.4). This strongly suggests HA-NA linkage with fitness differences between HA-NA subtypes (Dugan et al., 2008; Latorre-Margalef et al., 2014). Surprisingly, when challenging ducks with combinations that are common (e.g., H3N8) compared to uncommon (e.g., H4N8), there appear to be no fitness differences, such as virus load or duration of shedding (Lebarbenchon et al., 2012); thus the mechanisms that drives subtype abundance patterns are still unknown.

Furthermore, some subtypes or groups of subtypes appear early in the season, and others appear late in the season. This is attributed to subtype specific immunity against HA subtypes (homosubtypic immunity), or closely related subtypes (heterosubtypic immunity) (Latorre-Margalef et al., 2013). For example, in a long-term monitoring of migratory Mallards at Ottenby Bird Observatory, Sweden, the first viruses to appear in the season are normally H3 class viruses, which include phylogenetically related H3, H4, H7, and H10 subtypes. This is in contrast to the later arriving H1 class viruses (H1 clade: H1, H2, H5, and H6; and H9 clade: H8, H9, and H12) (Latorre-Margalef et al., 2014). Beyond Ottenby, an interrogation into subtype dynamics at the HA clade level the order in which HA subtypes appear has not been undertaken at other study sites.

3. AIV Ecology in Waterfowl

3.1 Mallards as Model Species for Studies of LPAIV Ecology

The detailed ecology of LPAIV is best described in Mallard, a species believed to be the central reservoir in the Northern Hemisphere. Mallards comprise the dominant species in a number of LPAIV

(and HPAIV) surveillance systems, particularly in Europe (Latorre-Margalef et al., 2014; van Dijk et al., 2014) as this species has consistently high LPAIV prevalence across years and sampling sites. A review by Olsen et al. (2006) demonstrated that nearly 50% of all waterfowl samples analyzed for LPAIV were from Mallard, with a global prevalence of 12.9%. It is unlikely that the status of Mallard has changed since, given ongoing intensive surveillance of this species. A search of the Influenza Research Database (IRD; http://www.fludb.org; Sept 2021) indicates 110,312 samples have been collected from Mallards, compared to 24,826 in Blue-winged Teal (*Anas discors*), 12,089 in Northern Pintails (*Anas acuta*), or 7388 in Northern Shoveler (*Anas clypeata*), which comprise other intensively sampled *Anas* species. Further, the vast majority of wild bird LPAIV isolates are from Mallards, and as raised previously, this species hosts an enormous diversity of LPAIV subtypes (Olsen et al., 2006; Olson et al., 2014). For example, at Ottenby Bird Observatory (Sweden) 74 different HA-NA subtype combinations across 1081 isolates in Mallards sampled between 2002 and 2009 were recorded (Fig. 31.4) (Latorre-Margalef et al., 2014). A search of IRD (Sept 2021) suggests 105 different HA-NA subtype combinations have now been detected in Mallards, globally.

3.1.1 Drivers of the Seasonal Cycle of LPAIV Prevalence in Mallards

LPAIV prevalence in Mallard in the Northern Hemisphere follows a seasonal pattern, whereby the prevalence of LPAIV is low during the late winter, spring, and summer, followed by a peak in viral prevalence during the autumn (Hatchette et al., 2004; Ito et al., 1995; Latorre-Margalef et al., 2014; Olsen et al., 2006; Wilcox et al., 2011). This pattern has been repeatedly observed at a number of study sites across the Northern Hemisphere, including Sweden (Latorre-Margalef et al., 2014), The Netherlands (Munster et al., 2007; van Dijk et al., 2014), Alberta Canada (Hatchette et al., 2004; Sharp et al., 1997), United States (Minnesota (Wilcox et al., 2011), California (Hill et al., 2012), Alaska (Ip et al., 2008; Runstadler et al., 2007)), and prevalence can be up to 30% during the autumnal peak of prevalence (Fig. 31.5). Generally, the prevalence of LPAIV peaks at a single time during the autumn and the timing, shape, and intensity of the infection curve shows strong interannual heterogeneity (Avril et al., 2016; Latorre-Margalef et al., 2014; Verhagen et al., 2014b).

This seasonal prevalence pattern is largely driven by the ecology of Mallards (and other *Anas* ducks). During the summer months, Mallards disperse across the landscape to breed, leading to high levels of juvenile recruitment into the population. Critically, these juveniles are immunologically naïve, and one study suggested that in the late summer/early autumn, up to 30% of the Mallard population may comprise juveniles resulting in a substantial susceptible proportion of the population (Bentz, 1985; Munster and Fouchier, 2009). As such, the vast majority of infections are in juvenile birds, with statistical differences in age classes being consistently reported (Avril et al., 2016; Ip et al., 2008; Latorre-Margalef et al., 2014; van Dijk et al., 2014; Webster et al., 1992; Wilcox et al., 2011). Beyond basic statistical differences, a capture-mark recapture study comprising 32,500 individuals illustrated that juveniles experienced twice the risk of becoming infected as compared to adults, whereas recovery rates were similar across different age categories (Avril et al., 2016). In the autumn months, widely dispersed breeding individuals and their offspring congregate during migration, which increases host density. As a result, there is a dramatic increase in LPAIV prevalence during autumn migration. Migration plays a key role in LPAIV prevalence, with two studies linking an increase in LPAIV prevalence with the arrival of migratory birds (Gaidet et al., 2012a; van Dijk et al., 2014). Whether LPAIV infection has an effect on migration remains unclear, with some studies reporting no effects (Bengtsson et al., 2016), and others reporting minor negative effects on migration (van Dijk et al., 2015) in Mallards. Indeed, by modeling data collected from 2004 to 2008, van Toor et al. (2018)

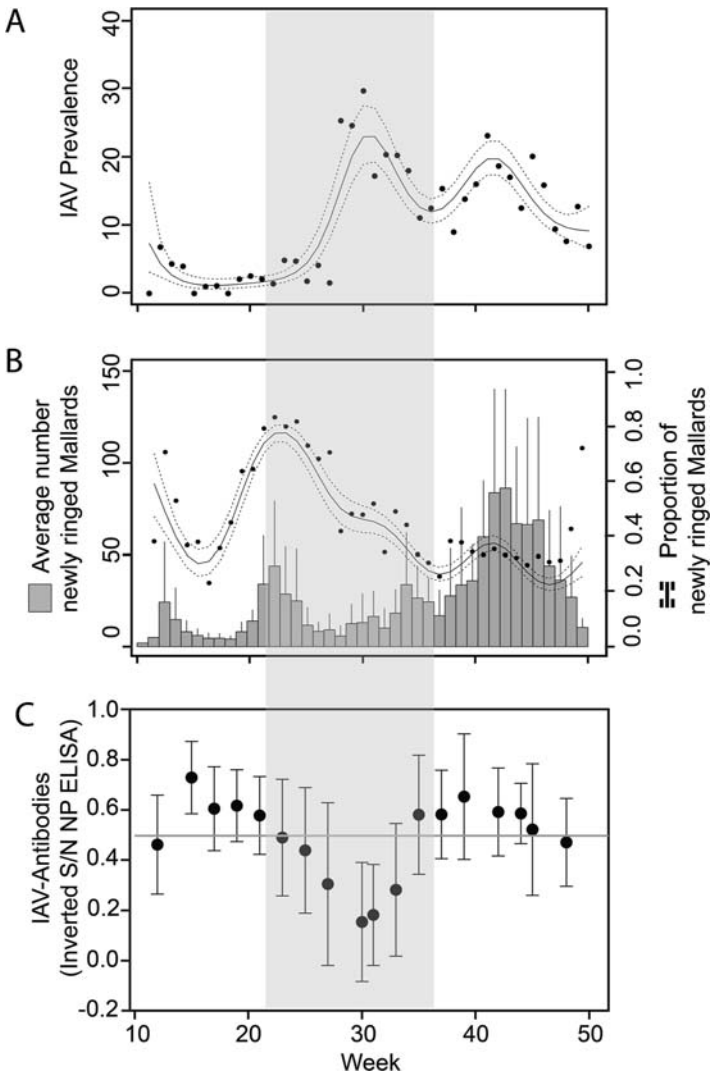


FIGURE 31.5

Seasonal dynamics of LPAIV in Mallards. (A) Prevalence of LPAIV across an annual cycle. Modeled prevalence (*line*), 95% confidence interval of the model (*dashed lines*) and prevalence estimates from 8 years (*points*). (B) Juvenile recruitment into the population. Gray bars comprise the average number of newly ringed/banded Mallards, and the line is the modeled proportion of newly ringed Mallards across all Mallards captured at Ottenby Bird Observatory. (C) S/N values of anti-NP antibodies in 10 Mallards, repeatedly sampled, occupying the duck trap at Ottenby. Gray horizontal line indicates the manufacturers recommended cut off, where values >0.5 are positive for anti-NP antibodies.

Panel A and B reproduced from Latorre-Margalef, N., et al., 2014. Long-term variation in influenza A virus prevalence and subtype diversity in a migratory Mallards in Northern Europe. Proc. R. Soc. B 281. https://doi.org/10.1098/rspb.2014.0098. Panel C modified from Tolf, C., et al., 2013. Individual variation in influenza A virus infection histories and long-term immune responses in mallards. PLoS One 8, e61201. https://doi.org/10.61371/journal.pone.0061201 and distributed under a CC-BY attribution.

predicted 36 virus dispersal events, and critically, the vast majority of these were conducted by juvenile birds ($n = 27$) as compared to adult ($n = 1$) or unaged birds ($n = 8$). Further virus dispersal events did not occur randomly throughout the season. Rather dispersal events were clustered in time; i.e., they were most likely to occur when a period of a large population size at Ottenby was followed by a pronounced emigration event (van Toor et al., 2018). In overwintering populations of Mallards, peak prevalence of LPAIV infection in the autumn corresponds with the arrival of migrants. Indeed, both local and long-distance migrants were more often infected with LPAIV than residents in a duck population in the Netherlands (van Dijk et al., 2014). This prevalence peak may be driven by, not only the introduction of new LPAIV subtypes (as described below), but a replacement of recovered residents with susceptible migrants (Lisovski et al., 2018). Higher susceptibility in migrants may be because either they have not yet been exposed to LPAIV, or their immune system is depressed due to their migration (Buehler et al., 2008; Lisovski et al., 2018). As such migrants play a key role in amplification of LPAIV.

Given high LPAIV prevalence in the autumn, by winter, population immunity should be very high with few susceptible individuals remaining. In studies following anti-NP antibodies in Mallards across the year, anti-NP antibodies are high in the winter and spring, but a dramatic decline in seroprevalence is detected in the summer months, corresponding to when ducks begin to breed (Fig. 31.5). This suggests that the susceptible portion of the population in late summer/early autumn is comprised not only of immunologically naïve birds recruited into the population following breeding but is also comprised of adults with waning immunity (Tolf et al., 2013; van Dijk et al., 2014). Taken together, this suggests that a substantial proportion of the population is susceptible to LPAIV infection by the early autumn. Thus, it is a combination of Mallard phenology, ecology, and biology that are drivers for the seasonal pattern of LPAIV prevalence (Fig. 31.5).

3.1.2 From Populations to Individuals

Most field studies conducted on LPAIV in Mallards focus on describing population level parameters, such as virus prevalence or seroprevalence. Although important, population level data can mask processes occurring at the individual level and fail to acknowledge individual variation in infection patterns. The biggest hurdle for conducting individual-level disease ecology studies is to follow individuals and their disease states over time. Sentinel ducks provide an excellent opportunity to do so. These birds are often placed in duck traps or cages and therefore become naturally infected by their wild conspecifics as they share water and are separated only by some mesh.

Through daily sampling of 10 sentinel ducks, Tolf et al. (2013) provides the best insight into LPAIV infection at the individual level (Fig. 31.6). First, while population level data illustrate high prevalence through the autumn, with one general peak, the individuals monitored in Tolf et al. (2013) demonstrated that juvenile ducks may have three to four discrete infections during their first autumn, with four to six different HA-NA subtype combinations detected in each individual. At this high temporal sampling resolution, ducks had discrete periods of infections and periods of no shedding. Different infections ranged in duration, with infection late in the year lasting 1−2 days in contrast to primary infections which lasted ∼6 days, and in some cases prolonged shedding of secondary infections (e.g., 14 days in the case of one ducks intermittently shedding H4N6) (Tolf et al., 2013) (Fig. 31.6). These individuals also had different patterns of immune responses. For example, while most ducks rapidly developed anti-NP antibodies following infection, one individual did not have detectable antibodies until November (3 months after being placed in the trap). There were also large

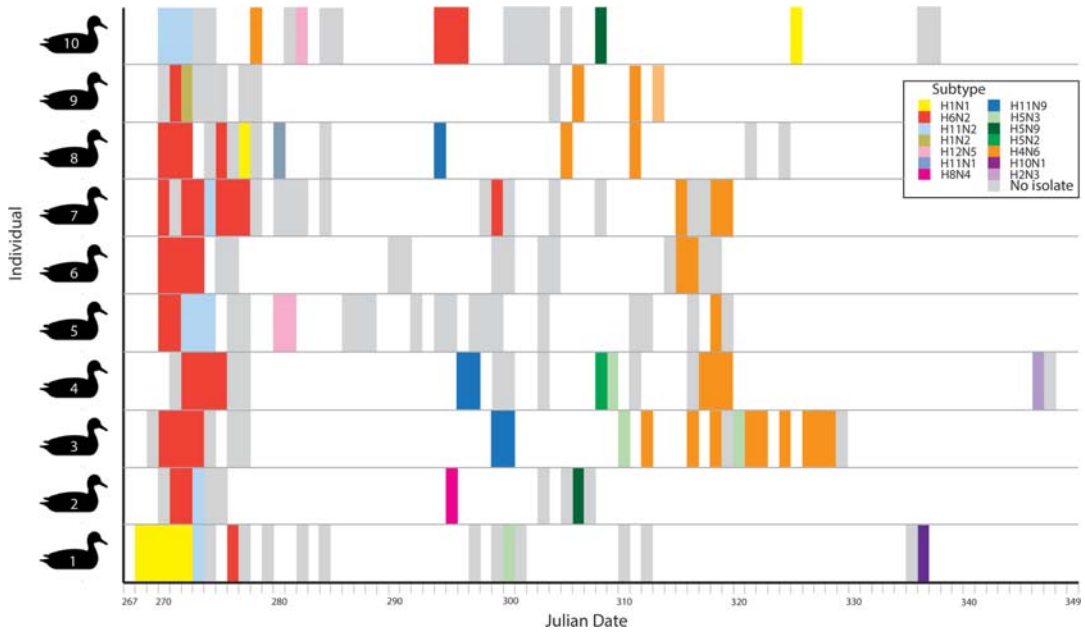


FIGURE 31.6

Ten Mallards residing in a duck trap were sampled daily across autumn 2009. Individuals are on the Y-axis, and day of the year on the X-axis. White space indicates the birds were sampled, but PCR negative. Gray indicates that the birds were sampled, positive by PCR, but we were unable to recover an isolate. All other colors indicate birds were sampled, positive by PCR, viruses were isolated and successfully isolated, and different colors correspond to different HA-NA subtypes.

Figure modified from Wille, M., et al., 2013. Frequency and patterns of reassortment in natural influenza A virus infection in a reservoir host. Virology 443, 150–60.

difference in longevity of these anti-NP antibodies, with one individual maintaining anti-NP antibodies through the duration of the study (1.5 years), and three ducks losing detectable anti-NP antibodies by December (Tolf et al., 2013). There is certainly more scope to better uncover individual variation in natural systems (as compared to experimental systems, which are not discussed here).

3.1.3 Mallard Immune Response Against LPAIV

Beyond poultry, the only species for which we have some understanding of the immune responses against LPAIV is in the Mallard.

Following infection, it has been demonstrated that *RIG-I* (retinoic acid inducible gene 1) and *Mx* (myxovirus resistance gene) are highly upregulated at the site of infection (Barber et al., 2008; Helin et al., 2018; Vanderven et al., 2012). The *RIG-I* gene is absent in chickens, and may explain why chickens display severe morbidity and mortality following infections (particularly with HPAIV), whereas Mallards may display no clinical signs of disease (Barber et al., 2008; Vanderven et al., 2012). A number of other important innate immune genes that are upregulated are effectors of the interferon (IFN) pathways (Vanderven et al., 2012). Interestingly, the major histocompatibility complex (MHC),

a part of the acquired immune response, appears to be important both early and late in IAV infection (Vanderven et al., 2012).

Tightly regulated expression of innate immune genes in the early stages of infection appear to allow ducks to control infection, mitigating the need for sustained antibody production. Indeed, ducks are purported to have poor long-term immune memory (Magor, 2011). Compared to chickens, Mallards have a translocation of the IgA (Magor, 2011; Magor et al., 1998, 1999) and a truncation of the IgY (Warr et al., 1995) which appear to affect antibody functionality. Truncated IgY is able to neutralize viruses, but is not involved in agglutination, complement fixation or opsonization (Lundqvist et al., 2006). In chickens, protection against avian influenza virus infection is primarily through antibodies directed at the HA (Kapczynski and Swayne, 2009), which is likely also true in Mallards. Antibodies directed against NA, or other proteins such as NP may contribute of clearance of infection (Nayak et al., 2010), but it is unclear.

Most studies of the immune response of Mallards have been done in laboratory settings using Pekin strain Mallards as opposed to genetically diverse wild Mallards. Using lysis, agglutination, haptoglobin as indicators of innate humoral immune response, Van Dijk et al. (2020) found significant differences in the responses of domestic Mallards as compared to wild Mallards, suggesting the results of previous studies may need to be reassessed. Similarly, there may be differences in experimentally infected Mallards (e.g., high inoculum doses, and routes of infection that are not consistent with the known transmission route) as compared to naturally infected birds.

3.1.4 LPAIV in Other Waterfowl

Given the perceived role of Mallards as central reservoirs for LPAIV, they are arguably the only species for which we have disentangled LPAIV ecology. Indeed, it is often assumed that LPAIV ecology in Mallard is representative of other waterfowl, particularly other *Anas* ducks.

A long-term study (1975−2015) from North America implicates a substantial number of waterfowl species as key reservoirs for LPAIV, including Blue-winged Teals, Northern Pintails, Northern Shovelers, American Wigeons (*Anas americana*), Cinnamon Teals (*Anas cyanoptera*), and to a lesser extend diving ducks and seaducks like Redhead (*Aythya americana*), Bufflehead (*Bucephala albeola*), and Surf scoter (*Melanitta perspicillata*) (Diskin et al., 2020). These data conform with findings of other large, multi-species surveillance studies performed in the United States (e.g., Ip et al., 2008; Stallknecht and Brown, 2007). In order to disentangle the roles of different waterfowl species, Hill et al. (2010) assessed a number of host traits and found that the foraging strategy played a key role in dictating viral prevalence in different waterfowl species, as did the density of lamellae, which are small structures found on the bills of waterfowl (Hill et al., 2010).

As such, the heavy focus on Mallard may provide a very simplistic model of LPAIV dynamics across duck hosts, particularly of the related *Anas* ducks. Blue-winged Teals have emerged as an important waterfowl host due to their role in LPAIV dispersal from North America into Central America (e.g., Guatemala) (Gonzalez-Reiche et al., 2012, 2017). However, based on limited studies, teals move viruses south, with no evidence of viral movement north. Indeed, there is no evidence of teals dispersing South American lineage-viruses into the United States in the spring (Ramey et al., 2014b). Northern Pintails have similarly been identified as important species for the dispersal of LPAIV as pintails move between Alaska and Asia, which is reflected in LPAIV genomes comprising genes of both North American and Eurasian origins (so called “intercontinental virus reassortants”) having been isolated from this species in Alaska (Pearce et al., 2011).

While geese have largely been neglected, a number of studies have been undertaken in swans (*Cygnus* sp.), and in general, patterns of LPAIV ecology in swans diverge from the traditional duck narrative. First, there are observable fitness consequences of LPAIV infections in swans such as lower rates of foraging and body fat accumulation, delayed migration onset and shorter migration flights of infected birds compared to uninfected birds (van Gils et al., 2007). Hoye et al. (2016) clarified that most fitness effects were only significant in birds experiencing their first LPAIV infection (Hoye et al., 2016). Furthermore, birds experiencing their first LPAIV infection were unlikely to be resighted 1 year after infection, suggesting a survival cost of LPAIV which is not present in ducks (Hoye et al., 2016). Further, unlike *Anas* ducks which had limited signs of morbidity or mortality to HPAIV (Sturm-Ramirez et al., 2004, 2005), the mortality of swans infected with this virus was high, with up to 40% of juvenile swans dying following infection (Hill et al., 2019).

In terms of immune response, as outlined above, in ducks, anti-NP-antibodies may wane each year (Tolf et al., 2013), and ducks are reported to have poor immune memory (Magor, 2011). This is in contrast to swans, where the immune repertoire (neutralizing anti-HA antibodies) increases with age (Hill et al., 2016). A key explanation of why we see different patterns of ecology in swans as compared to ducks may be in their life histories. Critically, unlike ducks, swans are long-lived species and may live 20–30 years, whereas the average lifespan of a Mallard is 3 years (although they may live to ~10 years). This difference may dictate how birds resource their immune response against LPAIV infection, although this is something that has not been well explored in this context.

4. Beyond the Duck-Centric, Temperate Northern Hemisphere Model of LPAIV Ecology

It is important to recognize that our current understanding of LPAIV ecology is described from a mostly duck-centric, temperate and Northern Hemisphere perspective (Gaidet et al., 2012a). This model ignores data from the Southern Hemisphere, including many tropical and subtropical study sites, and the role of other wild bird taxa, such as the *Charadriiformes* in the ecology of this virus.

4.1 AIV Ecology in the Southern Hemisphere

A continental-scale study of LPAIV dynamics across North America demonstrated that infectious disease dynamics may vary geographically due to differences in climate, seasonality, and host ecology (Lisovski et al., 2017). In particular, bird populations in low-latitude environments wherein seasons are less pronounced had lower LPAIV prevalence with limited seasonal variation (Diskin et al., 2020; Lisovski et al., 2017). Studies from Australia and Africa reinforce these findings, whereby most studies report LPAIV prevalence of less than 5% with no obvious seasonal peak linked to recruitment of juveniles into the population (although, in Australia, this may be affected by the highly aggregated nature of sampling strategies employed) (Curran et al., 2015; Ferenczi et al., 2016; Grillo et al., 2015; Hansbro et al., 2010; Haynes et al., 2009; Vijaykrishna et al., 2013).

In both Australia and Afrotropical regions, rainfall is a key predictor of LPAIV, although with different process on these two continents. In Australia, profound interannual variation in the timing and quantity of rainfall, which is strongly linked with El Niño–Southern Oscillation (ENSO) and

Indian Ocean Dipole (IOD), is thought to drive duck population breeding ecology and therewith LPAIV dynamics on this continent (Briggs, 1992; Ferenczi et al., 2016; Halse and Jaensch, 1989; Norman and Nicholls, 1991; Stuecker et al., 2017; Wille et al., 2022a). Despite many parts of Australia being defined as temperate, annual recruitment of immunologically naïve juvenile waterfowl into avian populations is irregular due to highly variable climatic conditions which impact breeding cycles. In some years waterfowl may not breed, or breed in small numbers (Roshier et al., 2002). In contrast, during years with higher rainfall, some species may have elongated breeding times (5−7 months), or breeding may be competently opportunistic and take place at any time of the year with multiple broods (Briggs, 1992; Halse and Jaensch, 1989; Norman and Nicholls, 1991). As a result, LPAIV prevalence is highly modulated by patterns of rainfall across years, whereby LPAIV prevalence is higher in years following rain, as compared to dry years where prevalence may be $<2\%$ regardless of the season (Ferenczi et al., 2016; Wille et al., 2022b). In Afrotropical regions, high intraannual rainfall variations drive seasons (Gaidet, 2016; Gaidet et al., 2012a). Most Afrotropical wetlands experience extreme seasonal variations: seasonal rainfall and river flooding inundate a vast network of temporary ponds and floodplains that dry out during the dry season through high evaporation and human extraction. This strong seasonality in wetland habitat distribution and availability impacts the temporal dynamics of breeding and aggregation of the wildfowl species. The lower and seasonally less variable but continuous LPAIV prevalence in Afrotropical regions may result from a slower turnover of susceptible birds in the wildfowl community compared with temperate regions, due to a more continuous recruitment of juvenile birds and a more gradual pace of first infection (Gaidet, 2016; Gaidet et al., 2012). LPAIV prevalence was further positively correlated with the arrival of migratory waterfowl from Eurasia, whereby prevalence progressively increased from the time when Eurasian migrants arrive in Afrotropical regions (September) to peak at the end of their wintering period (February−March), and decreased after their departure (Gaidet, 2016; Gaidet et al., 2012).

4.2 AIV in Shorebirds

Due to limited waterfowl migration, particularly connecting Australia and South America, the movement of LPAIVs from the Nearctic/Palearctic to the Neotropics/Afrotropics/Australalian realms is mostly likely carried by long-distance shorebird migrants (Araujo et al., 2018; De Araujo et al., 2014; Gaidet et al., 2012b; Hoye et al., 2021; Hurt et al., 2006; Rimondi et al., 2018; Wille et al., 2019a, 2022a). LPAIVs from Australian shorebirds are often intercontinental reassortants, with viruses comprising gene segments from across the East-Asian Australia Flyway: North America, Eurasia, and Australia (Hoye et al., 2021; Hurt et al., 2006; Wille et al., 2022a). An evolutionary study similarly showed the importance of shorebirds in introducing viruses to South America, as many of the recent characterized virus detections belonged to the main North American shorebird associated lineages rather than divergent South American lineages (Rimondi et al., 2018). The role of shorebirds, rather than waterfowl, for the introduction of LPAIV to South America may explain the absence of H14 in South America despite the introduction and rapid spread across North America and into central America (Latorre-Margalef et al., 2015; Ramey et al., 2014a).

Beyond their putative role in long distance virus dispersal, the role of shorebirds in LPAIV ecology remains enigmatic. Across study sites in Europe, Alaska, Australia, and Africa, viral prevalence was

low with very limited circulation (De Araujo et al., 2014; Gaidet et al., 2012b; Hanson et al., 2008; Haynes et al., 2009; Hurt et al., 2006; Iverson et al., 2008; Munster et al., 2007; Winker et al., 2008). However, occasional "outbreaks" may be detected, wherein following years of limited circulation of LPAIV in a population, LPAIV prevalence may increase substantially (Hoye et al., 2021). The exception is Delaware Bay, wherein LPAIV prevalence is ~14% shorebirds during spring migration, annually (Krauss et al., 2010). As such, LPAIV circulation in shorebirds appears to be localized, species-specific, and highly variable.

Delaware Bay, USA, hosts one of the largest seasonal aggregations of shorebirds in the world, with over a million shorebirds congregating to forage on the eggs of Horseshoe crabs (*Linudus polyphemus*) (Krauss et al., 2010). This site acts as the final stopover site prior to the arctic breeding grounds of shorebirds utilizing the Atlantic-Americas Flyway. The high density or shorebirds, often foraging in shallow water, facilitates efficient circulation and transmission of LPAIV (Hanson et al., 2008). Critically, shorebirds at Delaware Bay act as an amplifier of LPAIV likely circulating in duck or gull populations that also use the area; that is, rather than amplifying viruses imported by shorebirds from their nonbreeding areas, viruses are annually introduced into the shorebirds that use Delaware Bay as a stopover site from local ducks (Dugan et al., 2008; Maxted et al., 2012, 2016). Delaware Bay is not the only location where shorebirds congregate in large numbers—Banc d'Arguin in Mauritania hosts 2.3 million overwintering shorebirds (Gaidet et al., 2012b), Copper River Delta in Alaska, USA hosts one to four million shorebirds per year (the largest aggregation in the North America) (Bishop et al., 2000), Wadden Sea in the Netherlands may host up to one million Dunlin (*Calidis alpina*) (Blew et al., 2017), and an estimated two million shorebirds utilize the Yellow Sea in China (Melville et al., 2016). Despite numerous locations playing the same role as stopover sites for large numbers of shorebirds, Delaware Bay is unique in the role it plays as an LPAIV hotspot, suggesting some key factors are at play which we have yet to ascertain. These may include a combination of host- and virus-specific ecological factors such as foraging behavior, migratory behavior, habitat preference, susceptibility, local bird density, and species composition, aggregation of host species, environmental persistence of the virus, receptor specificity, tissue tropism, and replication and transmission kinetics. This may also reflect differences in effort, whereby little work on avian influenza is performed at key stopover sites with the exception of Delaware Bay.

As with waterfowl, not all shorebirds are equal. Long distance, migratory shorebirds in the family Scolopacidae (*e.g.*, stints, sandpipers, knots) play a key role of LPAIV dynamics as LPAIV are more frequently detected in this group, globally. Of all the shorebird species that congregate at Delaware Bay, Ruddy Turnstone (*Arenaria interpres*) appear to have the highest viral prevalence and seroprevalence, in contrast to closely related species such as Sanderling (*Calidris alba*) which are consistently negative and have negligible LPAIV seroprevalence (Maxted et al., 2012). A recent study from Australia, which includes nine different members of the Scolopacidae, show significant heterogeneity in both viral prevalence and seroprevalence across members, however Red-necked Stint (*Calidris ruficolis*), Sharp-tailed Sandpiper (*Calidris acuminata*), Ruddy Turnstone, and Red Knot had significantly higher viral prevalence and seroprevalence in comparison to other shorebird species. Further, this study found not only a statistically significant species effect but also a statistically significant phylogenetic effect such that avian phylogeny is an important variable in dictating LPAIV host competence (Wille et al., 2022b). In contrast to members of the Scolopacidae, LPAIV has only been detected in three members of the Charadriidae (plovers) to date (Table 31.1), and prevalence is generally much lower than members of the Scolopacidae.

4.3 AIV in Gulls and Other Seabirds

Seabirds play a key role in the history of avian influenza. The first record of avian influenza in wild birds is of a mortality event in which HPAIV H5 viruses were isolated from the Common Terns (*Sterna hirundo*) in 1961 in Southern Africa (Becker, 1966). The first LPAIV isolated from healthy wild birds, in which no signs of disease were observed, were Wedge-tailed Shearwaters (*Ardenna pacifica*) sampled on the Barrier Reef Islands of Australia (Downie et al., 1977; Downie and Laver, 1973; Laver and Webster, 1972). Despite this important role in the early history of avian influenza research, little effort is dedicated to gull and seabird hosts.

Gulls (Laridae) comprise unique hosts for LPAIV as they are the only avian reservoirs for H13 and H16 (Fouchier et al., 2005; Huang et al., 2014b; Munster et al., 2007; Verhagen et al., 2014a, 2020; Wille et al., 2011a,b). These subtypes are rarely found in other avian hosts (e.g., terns or shorebirds) and have poor attachment to duck tissues in receptor binding studies (Verhagen et al., 2021). Beyond H13/H16, gulls are occasionally infected by other subtypes (Huang et al., 2014b; Wille et al., 2011a). Data from colonies of "small" gulls, such as Black-headed Gull (*Chroicocephalus ridibundus*) in the Netherlands and Ring-billed Gulls (*Larus delawarensis*) in Canada recorded the highest prevalence of LPAIV (H13/H16) in summer and early autumn. Most gull species breed in dense colonies, with adults and juveniles crowded in a small space, potentially creating good opportunities for virus spread especially among immunologically naïve juveniles. Virus prevalence appeared to peak at the end of the breeding season when juveniles fledged (Velarde et al., 2010; Verhagen et al., 2014a). This is in contrast to data from "large white-headed gulls" in Canada, such as Herring Gulls (*Larus smithsonianus*) and Great Black-backed Gulls (*Larus marinus*), for from which prevalence was higher in autumn and winter months (Huang et al., 2014b; Wille et al., 2011b). Coincidentally, viruses from Glaucous-winged Gulls (*Larus hyperboreus*) in Alaska were also collected in the autumn, although repeated sampling was not undertaken in these birds (Wille et al., 2011a). Virus genomes from gulls, particularly H13 and H16 viruses have a strong signal of geographic mosaicism, *i.e.*, intercontinental reassortment (Fig. 31.7). That is, rather than having viral segments falling into classical geographically defined lineages, they instead have genome segments from both the American and Eurasian gene pools (Verhagen et al., 2020). This is likely driven by the movement patterns of gulls, wherein, rather than following flyway patterns defined above, gulls may move across oceans. For example, ringed/banded Great Black-backed Gulls in Canada have been recorded in Europe, and gulls from both Europe and Canada may mix along the continental shelf (Wille et al., 2011b) (Fig. 31.7).

A thorough review of other seabirds as hosts for LPAIV was undertaken in 2016, and a few key avian taxa were identified as hosts (Lang et al., 2016), including terns (*Sternidae*), murres/guillemots (*Uria* spp.) and penguins (*Sphensciformes*). LPAIV were identified in terns and noddies in some of the earliest studies of LPAIV in healthy wild birds in Australia (Mackenzie et al., 1984), and a more recent study in the Indian Ocean revealed that seroprevalence varied between species of terns on the island colonies they occupied. Critically, anti-HA antibodies were detected against 11 different HA subtypes, including H16 (Lebarbenchon et al., 2015a). LPAIV genomes from terns in Australia have showed similar features of intercontinental reassortment (Kishida et al., 2008), likely due to long distance migrations. Common Murres (Common Guillemot; *Uria algae*) and Thick-billed Murres (Brünnich's Guillemot; *Uria lomvia*) are pelagic seabird species that only come to shore to breed in dense colonies. The first indication that murres were reservoirs for LPAIV came from a viral isolate from Common

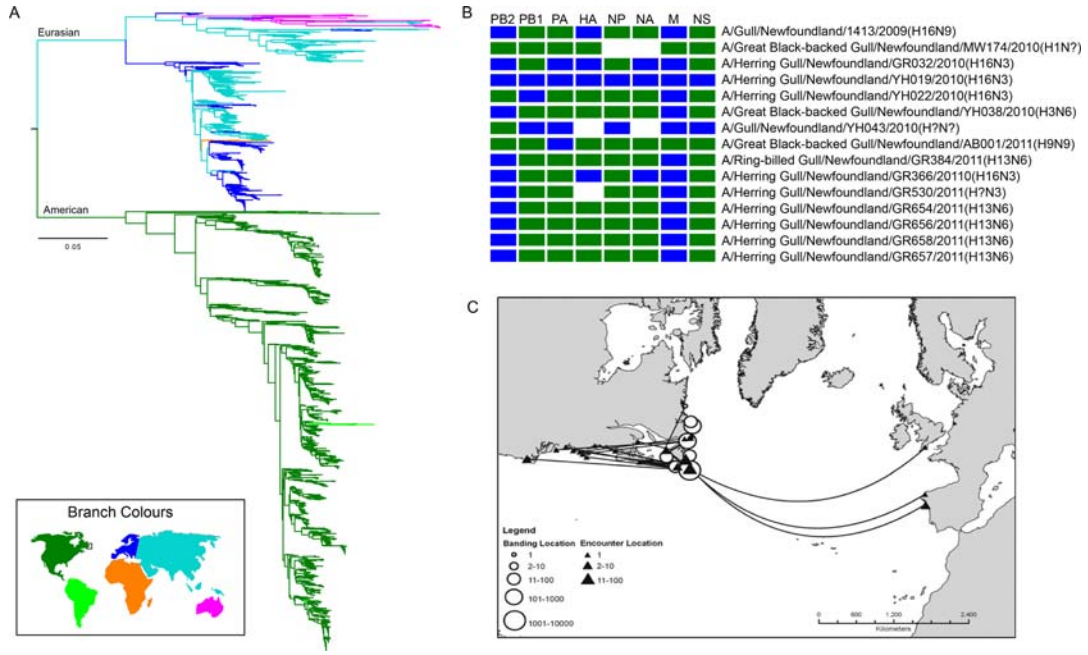


FIGURE 31.7

(A) The phylogenetic structure of LPAIV is influenced by migratory flyways. Maximum likelihood phylogenetic tree of 2218 H4 sequences downloaded from IRD (https://www.fludb.org/) and branches colored by the sample collection location, illustrating deep phylogenetic division and consistent continental scale geographic structure. (B, C) Viruses isolated from gulls and seabirds occupying continental fringes may have genomes comprising intercontinental reassortants, i.e., contain segments with origins in both Eurasia and North America. (B) LPAIV genomes from gulls isolated in Newfoundland (eastern Canada). Tiles represent segments, with each column indicating a different segment (ordered by size from left to right). In cases where tiles are green, they comprise segments falling into the "Americas" clade and blue from the "Eurasia" clade. In cases where tiles are white, no sequence is available. (C) Map of band recoveries of Great Black-backed Gulls (*Larus marinus*) banded in Newfoundland demonstrate movement to Eurasia, and thus independent of flyways. (For interpretation of the references to color in this figure legend, please refer online version of this title.)

(B) Data from Huang, Y., et al., 2014b. Perpetuation and reassortment of gull influenza A viruses in Atlantic North America. Virology 456, 353–63. Figure reproduced from Wille, M., et al., 2011b. Reassortment of American and Eurasian genes in an influenza A virus isolated from a great black-backed gull (Larus marinus), a species demonstrated to move between these regions. Arch. Virol. 156, 107–15.

Murres in the Russian Pacific (Sazonov et al., 1977), and LPAIV have now been repeatedly detected birds in north Pacific Ocean (Ramey et al., 2010). In the north Atlantic, the first detection of LPAIV in murres were from Thick-billed Murres that died in a wreck in the winter months (Granter et al., 2010) on the Atlantic coast of Canada. Subsequent studies in eastern Canada have revealed occasional high prevalence on breeding colonies, however detailed ecology is still to be resolved (Huang et al., 2014a; Wille et al., 2014).

A key unknown is the ecology of LPAIV in Antarctica. Swabs, feces, and serum have been collected from Antarctic penguins since the 1970s, however, it was only in 2014 that the first viruses were isolated (Hurt et al., 2014, 2016; Lang et al., 2016; Wallensten et al., 2006). To date, H11N2 viruses have been detected in Adélie penguins (*Pygoscelis adeliae*) and Snowy Sheathbill (*Chiona alba*), H5N5 viruses have been detected in Chinstrap Penguins (*Pygoscelis antarcticus*) on the Antarctic Peninsula (Hurt et al., 2014, 2016) and H4N7 has been detected in Southern Giant Petrels (*Macronectes giganteus*) on the South Shetland Islands (de Sousa Peterson et al., 2017). Viral genomes of H11N2 viruses detected in Antarctica have unique genome constellations, comprising not only segments of the American lineages, but the PB1 and PA genes comprised the H3N8 equine lineages, potentially suggesting long term isolated circulation in Antarctica. Beyond these scant detections and reporting of anti-NP antibodies, we have no grasp of the ecology of these viruses, whether they are maintained in Antarctic or represent incursions from South America, and the role that penguins or other Antarctic seabirds play as reservoirs or spill-over hosts. This shortcoming is having profound consequences for our ability to assess the introduction and spread of HPAIV into this region (Dewar et al., 2023).

5. Effect of Infection on Birds

Whether infection with LPAIV has any effects on the host, particularly dabbling duck hosts, remains contentious (Kuiken, 2013). Currently, it is widely reported that wild birds, particularly dabbling ducks, infected with LPAIV exhibit no clinical disease signs, despite being infected and reinfected with a virus, and shedding these viruses at high viral loads. Despite this, there is some evidence for subclinical effects and potentially fitness effects of infection.

Overall, there are limited physiological responses of individuals against infection. LPAIV replicate in the surface epithelium of the respiratory tract and gastrointestinal tract, however, gross lesions are absent at the site of infection in natural infections (Kuiken, 2013). As discussed above, there is evidence of a dampened immune response in Mallards against LPAIV infection.

Studies of body condition, refueling and migration reveal subtle effects of LPAIV infection. There is some suggestion that LPAIV infection may have a negative impact on body mass, however, it is unclear whether it is that LPAIV has an effect on body mass or whether birds in poor condition are more susceptible to acquiring infection (Flint and Franson, 2009; Latorre-Margalef et al., 2009a,b). A number of studies have since reported that body mass plays a role in viral shedding and susceptibility. In an experimental setting, food limitation in Mallards resulted in limited viral shedding compared to individuals in good condition (Arsnoe et al., 2011). In a study of White-fronted Geese (*Anser albifrons*), individuals with a lower body weight had a higher probability of infection, but only for one of 4 years (Kleijn et al., 2010). In turn, a study on Bewick's Swans (*Cygnus bewickii*) demonstrated that birds experiencing their first infection (naïve-infected) had a reduced foraging rate but had similar body stores to reinfected and uninfected individuals (Hoye et al., 2016) (as discussed above). Body mass may be influenced directly by the cost of infection, but sickness induced behavior may result in reduced refueling and feeding rates (van Gils et al., 2007).

Whether LPAIV impacts bird migration is unclear. Studies in Mallards have found no effects of overall staging time, or the speed and distance of subsequent migration (Latorre-Margalef

et al., 2009b). Tracking studies have found either no effect, or a weak effect of LPAIV infection on Mallards (Bengtsson et al., 2016; van Dijk et al., 2015). van Dijk et al. (2015) found a weak negative association between LPAIV infection in Mallard and regional movements (>100 m) on the final days of tracking, being exacerbated by poor weather conditions. Bengtsson et al. (2016) found no differences in activity or movement between infected and uninfected Mallards during stopover in autumn. Regardless of whether there is no effect or a weak effect, Mallards are competent vectors for dispersal, and on average dispersed viruses over 160 km in just 3 hours (van Toor et al., 2018).

The impact of infection seems to be more detrimental to swans. Naturally infected Bewick's Swans (*Cygnus columbianus bewickii)* had reduced refueling and feeding rates, and delayed and protracted migration distances as when infected with LPAIV, however the sample size of this study was only two infected birds (van Gils et al., 2007). When swans were experimentally infected and released, this trend was not observed (Hoye, 2011). A follow up study on Bewick's Swans detected a potential difference in survival, where naïve-infected swans were unlikely to be resighted 1 year after infection compared to uninfected or reinfected individuals (Hoye et al., 2016). This study also illustrates a difference in response between individuals infected for the first time and those uninfected and reinfected, whereby birds that have been infected and reinfected have similar responses to birds that have never been infected (Hoye et al., 2016).

6. Evolution and Genetics

Influenza A viruses evolve through two main mechanisms: antigenic drift and antigenic shift. Antigenic drift is the fixation, by natural selection, of mutations in the HA and NA that enable the virus to evade the immune system. These mutations are generated by an error prone RNA-dependant RNA polymerase that does not have proof-reading capability. LPAIV have been demonstrated to have rapid rates of evolution change, characterized by accumulations of synonymous and nonsynonymous mutations. The rate of mutations varies across segments, with an average rate of 3.41×10^{-3} subs/site/year (Bahl et al., 2009; Chen and Holmes, 2006; Worobey et al., 2014). To put this in context, the rate of RNA virus mutation is a million times greater than that of vertebrates, and this allows for the rapid adaptation of these viruses to new environments (Pybus and Rambaut, 2009). The ratio of synonymous to nonsynonymous mutations (dN/dS) varies enormously across proteins and is used as an indicator of selection pressure on genes. Higher dN/dS rations are suggestive of positive selection and would be expected for the surface glycoprotein genes hemagglutinin and neuraminidase, which interact with the host immune system. Lower dN/dS ratios are suggestive of negative or purifying selection, which would be expected for the matrix and polymerase genes as these segments are more functionally constrained, and therefore nonsynonymous mutations may have a deleterious effect. Thus, the highest dN/dS ratios observed were in NS1 (0.147), NA (0.206), and HA (0.130) (Chen and Holmes, 2006, 2010; Worobey et al., 2014).

Genomic reassortment is central to the biology of all influenza viruses and occurs frequently in all hosts studied (Lowen, 2017, 2018; Steel and Lowen, 2014). It is the segmented nature of the viral genome that enables evolution through reassortment. The consequence of reassortment is that, following coinfection, viral progeny contains various gene segment combinations from the different parental viruses. Indeed, given a coinfection with two AIVs, each with 8 segments, 256 different

genetic progeny are possible, generating significant viral diversity. Some of the novel genome constellations generated may facilitate adaptations to new hosts, help evade immune responses, and in the case of human influenzas, may assist in the generation of antiviral resistance. Critically, all human pandemic influenza viruses have been the result of reassortment. Reassortment has long been described in LPAIV in waterfowl. For example, Dugan et al. (2008) showed that five different H4N6 isolates from Mallards collected at the same location and on the same day comprised four different genome constellations (Dugan et al., 2008). Through full genome sequencing of 96 isolates from Mallards isolated in 2011, 56% were identified as reassortants (Wille et al., 2013). This frequent reassortment rapidly unlinks genome constellations, such that LPAIV in wild birds do not circulate as "fixed" genome constellations, but rather the continuous production of "transient" genome constellations. Reassortment in the wild bird reservoir is most easily observed when the viruses comprise a mosaic of Eurasian and North American origin gene segments, and have been most often detected in gulls and seabirds that move across continental margins (Ramey et al., 2010; Wille et al., 2011a).

Through phylogenetic analysis of viral genomes, avian influenza viruses can be divided into two main phylogenetic lineages: the Eurasian and American lineage (Olsen et al., 2006; Webster et al., 1992) (here illustrated by H4, Fig. 31.7). This is driven by the migration patterns of wild bird hosts, wherein waterfowl migrate between North and South America, and between Asia, Europe, and Africa. This allopatric separation has resulted in a major phylogenetic split between the Eurasian and American genetic lineages of LPAIV. However, there is evidence of virus exchange between these regions. For example, H6 from Eurasian ancestors were detected in North American wild birds and poultry and subsequently outcompeted the ancestral North American lineage of H6 (Bahl et al., 2009; Webby et al., 2003). This phenomenon is driven by avian taxa that migrate beyond the traditional flyway model. For example, whole-genome analyses of influenza isolates obtained from Northern Pintails in Alaska, a species that migrates between Alaska and Asia, suggests intercontinental virus exchange at a relatively high frequency. Of the analyzed LPAIV isolated in Alaska, 44% had at least one gene segment originating from the Eurasian lineage, whereas all other gene segments were of the American lineage (Pearce et al., 2011). This phenomenon is often found in H13 and H16 viruses, which are limited to gulls. Wille et al. (2011a) found that all H13 and H16 viruses isolated from gulls in Alaska and eastern North America had at least one segment of the Eurasian lineage (Wille et al., 2011a). To date, there is no evidence for cross-hemisphere movement of entire low pathogenic LPAIV genomes, rather only the introduction of single gene segments. There has, however, now been multiple introductions of HPAIV from Eurasia to North America (Alkie et al., 2022, 2023; Caliendo et al., 2022b). The partial geographic isolation of influenza virus hosts seems therefore sufficient to facilitate divergent evolution and continue the existence of separate viral gene pools. In addition to two hemispheric lineages, there is also evidence for host specific lineages, particularly in H13 and H16 subtypes found only in gulls. These viruses belong to a group of distinct LPAIVs based on genetic, functional, and ecological properties and have thus evolved into separate genetic lineages as compared to viruses circulating in *Anseriformes* and *Charadriiformes*. Gene segments of gull viruses are genetically distinct from those circulating in other wild birds, suggesting that they have been separated for a sufficient amount of time to allow genetic differentiation by sympatric speciation (Fouchier et al., 2005; Wille et al., 2011a).

7. The Role of Wild Birds in the HPAIV H5N1 Panzootic

As wild birds are the reservoir for LPAIV, most of this chapter has been dedicated to the topic. Wild birds are affected by HPAIV, however, prior to ~2016 had not been considered a reservoir. Following a genetic change in the virus in 2021, with the emergence of lineage 2.3.4.4b HPAIV H5N1, cases of HPAIV have exploded in both wild birds and poultry, with a dramatic geographic expansion of the virus, as well as substantial increase in the number of wild bird species affected. The situation is highly dynamic and evolving rapidly, and this section generally pertains to events and outbreaks prior to August 2023.

The lineage of HPAIV H5 circulating today, referred to as the goose/Guandong lineage (gs/GD), emerged in Asia in 1996, and was believed to originate from precursor LPAIV H5 viruses in migratory waterbirds (Chen et al., 2004; Duan et al., 2008). While HPAIV are normally eradicated through stamping out, the gs/GD lineage has been enduring, and has diversified dramatically in the last ~25 years. For clarity I have divided the epidemiology of gs/GD HPAIV H5Nx into three intervals: 1996–2014, 2014–2020, and 2021–present. The period between 1996 and 2014 comprised entrenchment into poultry populations, viral diversification, and a substantial spill over event into wild birds. The next phase, coinciding with the emergence of lineage 2.3.4.4 was associated with an increase in wild bird outbreaks, and the first indications of amplification in wild birds. The final interval coincides with the current panzootic.

Between 1997 and 2003 gs/GD viruses became widespread in poultry (Chan, 2002; Guan et al., 2002; Shortridge et al., 1998), and due to endemic circulation, evolved, resulting in many different sublineages. In addition to the significant mortality in the poultry industry, gs/GD HPAIV H5 viruses spilled over into wild birds from poultry in Asia on a number of occasions. The most significant event during this period occurred in 2005, where HPAIV H5N1 caused a substantial number of mass mortality events in wild birds. For example, over 6000 birds died at Lake Qinghai, China in 2005; Lake Qinghai plays a role as key migratory stopover site for wild birds (Chen et al., 2005; Liu et al., 2005). It was hypothesized that wild birds played a role in introducing the virus to Europe through long distance migration, although this was debated given that human activities linked to the poultry trade may have contributed to viral spread (Feare, 2007, 2010; Kilpatrick et al., 2006; Lycett et al., 2019; Normile, 2006; Starick et al., 2008).

In 2014, a new clade of gs/GD HPAIV H5, referred to as 2.3.4.4, emerged and spilled over into wild birds. Unlike previous gs/GD H5 viruses, 2.3.4.4 HPAIV H5Nx virus was mainly associated with an N8 subtype. This clade was first reported in a duck farm in China in 2010 (Wu et al., 2014) and was detected in both poultry and wild birds in Korea following an outbreak in 2014 (Lee et al., 2014). With their migrations, wild birds moved this novel virus into both North America and Europe (Global Consortium for H5N8 and Related Influenza Viruses, 2016; Verhagen et al., 2015). Between 2014 and 2020, 2.3.4.4 HPAIV H5Nx was repeatedly introduced into Europe from Asia, with significant incursions in 2014 (H5N8), in 2016 (H5N6), and 2020 (H5N8) (Xie et al., 2023) (Fig. 31.8). While there was initially limited indication of local amplification of these viruses in local wild birds, data from 2016/2017 demonstrated a shift wherein this virus was maintained and amplified by wild birds. Local amplification was associated with wild bird mortality events, but 2.3.4.4 H5Nx was also detected in live, apparently healthy Mallards. There was evidence for reassortment of H5N8 viruses and LPAIV viruses in the wild bird reservoir (Poen et al., 2016, 2018), and reassortment led to the emergence of

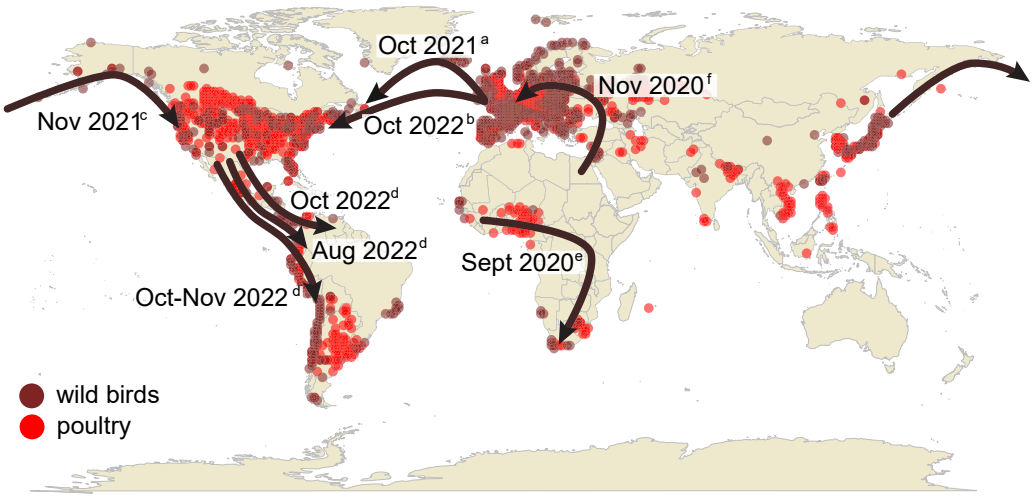


FIGURE 31.8

Notifications of HPAI to the World Organization for Animal Health (WOAH) between October 2021—May 2023. Notifications comprising poultry and wild birds are shown in different colors. *Arrows* are stylized to indicate notable viral incursions, with the estimate date of incursion from time structured phylogeny shown, and is extracted from (a) Caliendo et al. (2022), (b) Alkie et al. (2023), (c) Alkie et al. (2022), (d) Ariyama et al. (2023), and (e) Abolnik et al. (2023), and (f) Xie et al. (2023).

clade 2.3.4.4 HPAIV H5N6 viruses in European wild birds in 2017 (Poen et al., 2019). Following sporadic detections from 2018 to 2020, more than 200 outbreaks per month were reported in 2020/2021. Unlike the current panzootic, these outbreaks continued to comprise HPAIV H5N8.

An incursion of HPAIV H5N8 into North America in 2014 comprised the first time gs/GD viruses were detected in North America. Following the arrival of the virus with wild birds, spill-over (or rather, spill-back) of viruses into poultry systems led to outbreaks across the United States and Canada (Pasick et al., 2015; Stoute et al., 2016). Intriguingly, following the expansion of HPAIV H5N8 across the continent, this virus disappeared from North America (Krauss et al., 2016). The virus was introduced to Africa in 2016, and became entrenched in poultry populations of West Africa, and caused outbreaks in poultry and wild birds in southern African (Fusaro et al., 2019; Molini et al., 2020).

The current HPAIV panzootic commenced in 2021 following a genetic change in the virus. While the outbreaks in 2014/15 and 2016/17 had their origins in China, the current panzootic virus was first detected in Egyptian poultry (ca. 2016−17), but emerged from H5N8 viruses in Europe in 2020 (Xie et al., 2023). Specifically, this panzootic coincided with the replacement of all circulating 2.3.4.4 subclades with clade 2.3.4.4b, and replacement of the N8 with the N1 neuraminidase subtype. As a result, the number of outbreaks in wild birds (particularly waterbirds) and poultry exploded (European Food Safety Authority et al., 2023). Since August 2021, there have been tens of thousands of outbreak notifications affecting poultry and wild birds, globally (FAO, 2022). As of early 2023, 320 species

belonging to 21 orders have been affected, the vast majority of which are waterbirds and seabirds (e.g., orders *Anseriformes*, *Charadriiformes*, *Pelecaniformes*, *Podicipediformes*, *Suliformes*) (Klaassen and Wille, 2023). Mass mortality events in wild birds have been substantial and in some instances had population or species level effects (Wille and Waldenström, 2023). To date (November 2023), the only continent free from HPAIV is Oceania (Wille and Klaassen, 2023).

HPAIV has caused numerous mortality events, with key examples presented in Wille and Waldenström (2023), and summarized here. HPAIV has been causing outbreaks in wild birds of Southern Africa since ∼2017, and of great concern is the impact on African Penguins (*Spheniscus demersus*). Between December 2018 and February 2019 more than 350 African Penguins (∼5% of the colony) died in Namibia (Molini et al., 2020), and in 2022 a further 230 penguins died in South Africa (Reuters, 2022). Arguably, one of the first reported outbreaks in the panzootic occurred in wintering Barnacle Geese (*Branta leucopsis*) in Scotland, where approximately one-third of the Svalbard breeding population died (Promed mail, 2021a; RSPB, 2023; Wille and Barr, 2022). An additional 5000 Greenland breeding Barnacle Geese died in Scotland in the following winter, comprising ∼10% of the Greenland population. Toward the end of 2021 an outbreak occurred in Common Cranes (*Grus grus*) in Israel (Promed mail, 2021b; Wille and Barr, 2022), with 5200 birds dying. Commencing in 2022, an outbreak decimated colony nesting birds. Approximately 40% of the south-east European population of Dalmatian Pelicans (*Pelecanus crispus*) died in March 2022 in Greece (Alexandrou et al., 2022). Approximately 17% of the total northwestern European breeding population of Sandwich Terns (*Thalasseus sandvicensis*) were found dead across Europe (Knief et al., 2023). Unusual mortality was recorded at nearly all Northern Gannet (*Morus bassanus*) breeding colonies, and at Bass Rock in Scotland the number of occupied nests decreased by 71% (Lane et al., 2023). In North America, 62% of Caspian Terns (*Hydroprogne caspia*) breeding on Lake Michigan died in 2022 (Harvey et al., 2023). Predatory and scavenging birds have also been affected, with a number of satellite tracked Griffon Vultures (*Griffon fulvus*) dying or displaying immobility due to HPAIV in Europe (Duriez et al., 2023). In early 2023, 15 endangered California Condors (*Gymnogyps californianus*) died due to HPAIV, triggering an initiative to vaccinate the population to protect them (Kozlov, 2023). By 2023, substantial outbreaks began occurring in Black-headed Gulls (*Croicephalus ridibundus*) in Europe (EFSA). Reports have suggested more than 2000 dead gulls reported in the Netherlands (DWHC, 2023), >10,000 have died in the UK (BirdGuides, 2023), and outbreaks have also been reported in Sweden, France, Belgium, Germany and Lithuania (Wille and Waldenström, 2023). Ongoing outbreaks (August 2023) include substantial outbreaks in a large variety of seabird species, including Black-legged Kittiwakes (Rissa trydactyla) and Murres (Uria sp). More than 500,000 wild birds have died in South America since the emergence of HPAIV (Breed et al., 2023). Early outbreaks in South America were heavily skewed to pelicans and cormorants, rather than waterfowl. Approximately 41% of all Peruvian Pelicans (*Pelecanus thagus*) are believed to have died in a matter of weeks in early 2023. Guanay Cormorant (*Leucocarbo bougainvillei*) and Peruvian Boobie (*Sula variegata*) were also heavily affected (Ariyama et al., 2023; Leguia et al., 2023). Following the rapid southbound movement, outbreaks are now occurring across the continent.

In October 2023, the first detection of HPAIV occurred in the sub-Antarctic Islands of South Georgia. Shortly afterward, HPAIV was detected in on the Falkland (Malvinas) Islands. Viral

genomics has confirmed that the detections on South Georgia and the Falkland (Malvinas) Islands are not linked, that is, two independent incursions from South America occurred (Bennison et al., 2023). To date, infections have been confirmed in Brown Skuas (*Stercorarius antarcticus*), Kelp Gulls (*Larus dominicanus*), Southern Fulmar (*Fulmarus glacialoides*), Black-browed Albatross (*Thalassarche melanophris*). Suspected cases have also occurred in Southern Giant Petrel (*Macronectes giganteus*), Arctic Tern (*Sterna paradisaea*), and Southern Elephant Seals (*Mirounga leonine*) (SCAR, 2023). The situation is rapidly evolving.

While wild birds are undoubtedly being heavily affected by HPAIV, they are also contributing to virus spread, both on local and long-distance scales. Indeed, the virus was introduced to North America on three occasions, twice over the Atlantic Ocean and once over the Pacific Ocean (Alkie et al., 2022, 2023; Caliendo et al., 2022b). Given the large distances, wild birds must be able to not only survive infection, but critically, be able to migrate while infected. There is some emerging data around infection survival, the vast majority of which has centered on Mallards. First, there is evidence that infected Mallards do not show disease signs and survive infection. For example, apparently healthy *Anas* ducks sampled in the Netherlands were positive for HPAIV H5Nx (Poen et al., 2018). In experimental studies 58% of Mallards not only survived infection, but had no observable disease signs (Spackman et al., 2023). There is also evidence for Mallards to migrate despite HPAIV infection. For example, apparently healthy Mallards were fitted with satellite trackers to reveal their wintering and stop-over sites. Two of these birds were positive for HPAIV H5N8 (Lv et al., 2022), and undertook migration without delays. In North America, wintering mallards infected with highly pathogenic avian influenza show few signs of altered local or migratory movements (Teitelbaum et al., 2023). Beyond Mallards, there is less evidence. In Northern Gannets, individuals that survived infection had substantial change in iris color (Lane et al., 2023), and there was some evidence from satellite tracking that infected birds undertook long distance movement (Jeglinski et al., 2023). This is in contrast to vultures which exhibited strong indications of sickness behavior, with a dramatic decrease in movement (Duriez et al., 2023).

While HPAIV is heavily affecting both wild birds and poultry, the number of mammalian cases have dramatically increased since 2021. In general, mammals infected are those scavenging or predating on sick and dead birds, and are thus being infected through consuming infected prey. The diversity of mammals infected substantial, but concerning also include farmed minks and foxes. Wild mammals infected members of the Canidae (dogs, foxes, etc.), Felidae (cats, lions etc.), Mephitidae (skunks), Multelidae (minks), Otariidae (otters), Procyonidae (racoons), Ursidae (bears) (European Food Safety Authority et al., 2023). More concerning is the impact on marine mammals including members of the Delphinidae (dolphins), Phocoenidae (porpoises), and Phocidae (seals and seal lions). Indeed, in South America more than 30,000 South American Sea Lions have died since the emergence of HPAIV in the region (Plaza et al., 2023). Most recently, an entire generation of South Elephant Seals juveniles is believed to have died. Finally, HPAIV has caused a number of human cases. Generally the risk to humans remains low, although elevated to low to medium for occupationally exposed people (e.g., poultry workers) (Wille and Barr, 2022).

HPAIV is clearly a OneHealth problem, and it is likely that intensive outbreaks of HPAIV H5Nx will continue to pose challenges in the future unless a OneHealth response is prioritized.

References

Abolnik, C., et al., 2023. The molecular epidemiology of clade 2.3.4.4b H5N1 high pathogenicity avian influenza in Southern Africa, 2021-2022. Viruses 15, 1382. https://doi.org/10.3390/v15061383.

Alexander, D.J., 1995. The epidemiology and control of avian influenza and Newcastle disease. J. Comp. Pathol. 112, 105−126.

Alexander, D.J., 2000a. Newcastle disease and other avian paramyxoviruses. Revue scientifique et technique (International Office of Epizootics) 19, 443−462.

Alexander, D.J., 2000b. A review of avian influenza in different bird species. Vet. Microbiol. 74, 3−13.

Alexander, D.J., 2007. An overview of the epidemiology of avian influenza. Vaccine 25, 5637−5644.

Alexandrou, O., Malakou, M., Catsadorakis, G., 2022. The impact of avian influenza 2022 on Dalmatian pelicans was the worst ever wildlife disaster in Greece. Oryx 56, 813−813.

Alkie, T.N., et al., 2023. Recurring trans-Atlantic incursion of clade 2.3.4.4b H5N1 viruses by long distance migratory birds from Northern Europe to Canada in 2022/2023. Viruses 15, 1836. https://doi.org/10.3390/v15091836.

Alkie, T.N., et al., 2022. A threat from both sides: multiple introductions of genetically distinct H5 HPAI viruses into Canada via both East Asia-Australasia/Pacific and Atlantic flyways. Virus Evol. 8, veac077.

Allocati, N., et al., 2016. Bat-man disease transmission: zoonotic pathogens from wildlife reservoirs to human populations. Cell Death Discov. 2, 16048.

Altizer, S., Bartel, R., Han, A., 2011. Animal migration and infectious disease risk. Science 331, 296−302.

Araujo, J., et al., 2018. Migratory birds in southern Brazil are a source of multiple avian influenza virus subtypes. Influenza Other Respir. Viruses 12, 220−231.

Ariyama, N., et al., 2023. Highly pathogenic avian influenza A(H5N1) clade 2.3.4.4b virus in wild birds, Chile. Emerg. Infect. Dis. 29, 1842−1845.

Arsnoe, D.M., et al., 2011. Influence of body condition on influenza A virus infection in mallard ducks: experimental infection data. PLoS One 6, e22633. https://doi.org/10.1371/journal.pone.0022633.

Avril, A., et al., 2016. Capturing individual-level parameters of influenza A virus dynamics in wild ducks using multistate models. J. Appl. Ecol. 53, 1289−1297. https://doi.org/10.1111/1365-2664.12699, 12.

Bahl, J., et al., 2013. Influenza A virus migration and persistence in North American wild birds. PLoS Pathog. 9, e1003570. https://doi.org/10.1371/journal.ppat.1003570.

Bahl, J., et al., 2009. Gene flow and competitive exclusion of avian influenza A virus in natual resevoir hosts. Virology 390, 289−297.

Barber, M.R.W., et al., 2008. Association of RIG-I with innate immunity of ducks to influenza. Proc. Natl. Acad. Sci. U.S.A. 107, 5913−5918.

Bauer, L., et al., 2023. The neuropathogenesis of highly pathogenic avian influenza H5Nx viruses in mammalian species including humans. Trends Neurosci. 46, 953−970.

Bauer, S., Hoye, B.J., 2014. Migratory animals couple biodiversity and ecosystem functioning worldwide. Science 344, 54. https://doi.org/10.1126/Science.1242552.

Becker, W.B., 1966. Isolation and classification of tern virus: influenza virus A/Tern/South Africa/1961. J. Hyg. 64, 309−320.

Bengtsson, D., et al., 2016. Does influenza A virus infection affect movement behaviour during stopover in its wild reservoir host? R. Soc. Open Sci. https://doi.org/10.1098/rsos.150633.

Bennison, A., et al., 2023. Detection and spread of high pathogenicity avian influenza virus H5N1 in the Antarctic Region. bioRxiv. https://www.biorxiv.org/content/10.1101/2023.1111.1123.568045.

Bentz, P.-G., 1985. Studies on some urban Mallard Anas platyrhynchos populations in Scandinavia. Part I: causes of death, mortality and longevity among Malmö Mallards as shown by ringing recoveries. Fauna Norv. Ser. C Cinclus 8, 44−56.

BirdGuides, 2023. BTO Provides Grim Update on Bird Flu Spread. https://www.birdguides.com/news/bto-provides-grim-update-on-bird-flu-spread/. (Accessed 1 August 2023).
Bishop, M.A., et al., 2000. A method to estimate migrant shorebird numbers on the Copper River Delta, Alaska. J. Field Ornithol. 71, 627−637.
Blew, J., Günther, K., Hälterlein, B., Kleefstra, R., Laursen, K., Ludwig, J., Scheiffarth, G., 2017. Migratory birds. In: al, K.S.e (Ed.), Wadden Sea Quality Status Report 2017. Common Wadden Sea Secretariat, Wilhelmshaven, Germany.
Breed, A.C., et al., 2023. Southward Expansion of High Pathogenicity Avian Influenza H5 in Wildlife in South America: Estimated Impact on Wildife Populations, and Risk of Incursion into Antarctica. Offflu. https://www.offlu.org/wp-content/uploads/2023/2008/OFFLU-statement-HPAI-wildlife-South-America-20230823.pdf.
Briggs, S.V., 1992. Movement patterns and breeding characteristics of arid zone ducks. Corella 16, 15−22.
Bröjer, C., et al., 2009. Pathology of natural highly pathogenic avian influenza H5N1 infection in wild tufted ducks (*Aythya fuligula*). J. Vet. Diagn. Invest. 21, 579−587.
Brook, C.E., Dobson, A.P., 2015. Bats as 'special' reservoirs for emerging zoonotic pathogens. Trends Microbiol. 23, 172−180.
Buehler, D.M., et al., 2008. Seasonal redistribution of immune function in a migrant shorebird: annual-cycle effects override adjustments to thermal regime. Am. Nat. 172, 783−796.
Byrne, A.M.P., et al., 2021. H7N7 avian influenza virus mutation from low to high pathogenicity on a layer chicken farm in the UK. Viruses 13, 259. https://doi.org/10.3390/v13020259, 2.
Caliendo, V., et al., 2022a. Tropism of highly pathogenic avian influenza H5 viruses from the 2020/2021 epizootic in wild ducks and geese. Viruses 14, 289. https://doi.org/10.3390/v14020280.
Caliendo, V., et al., 2022b. Transatlantic spread of highly pathogenic avian influenza H5N1 by wild birds from Europe to North America in 2021. Sci. Rep. 12, 11729.
Chan, P.K., 2002. Outbreak of avian influenza A(H5N1) virus infection in Hong Kong in 1997. Clin. Infect. Dis. 34 (Suppl. 2), S58−S64.
Chen, H., et al., 2004. The evolution of H5N1 influenza viruses in ducks in southern China. Proc. Natl. Acad. Sci. U. S. A. 101, 10452−10457.
Chen, H., et al., 2005. Avian flu: H5N1 virus outbreak in migratory waterfowl. Nature 436, 191−192.
Chen, R., Holmes, E.C., 2006. Avian influenza virus exhibits rapid evolutionary dynamics. Mol. Biol. Evol. 23, 2336−2341.
Chen, R., Holmes, E.C., 2010. Hitchhiking and the population genetic structure of avian influenza virus. J. Mol. Evol. 70, 98−105.
Cross, T.A., et al., 2013. Prevalence of avian paramyxovirus 1 and avian influenza virus in double-crested Cormorants (*Phalacrocorax auritus*) in eastern North America. J. Wildl. Dis. 49, 965−977.
Curran, J.M., Ellis, T.M., Robertson, I.D., 2015. Serological surveillance of wild waterfowl in northern Australia for avian influenza virus shows variations in prevalence and a cyclical periodicity of infection. Avian Dis. 59, 492−497.
Daly, J.M., MacRae, S., Newton, J.R., Wattrang, E., Elton, D.M., 2011. Equine influenza: a review of an unpredictable virus. Vet. J. 189, 7−14.
De Araujo, J., et al., 2014. Avian influenza virus (H11N9) in migratory shorebirds wintering in the Amazon Region, Brazil. PLoS One 9, e110141. https://doi.org/10.1371/journal.pone.0110141.
de Sousa Peterson, E., et al., 2017. First detection of avian infuenza virus (H4N7) in Giant Petrel monitored by geolocators in the Antarctic region. Marine Biology 164, 62. https://doi.org/10.1007/s00227-00017-03086-00220.
Dewar, M., et al., 2023. Biological risk assessment of highly pathogenic avian influenza in the Southern Ocean. In: Risk Assessment Prepared for the Scientific Committee on Antarctic Science, 77 pages. https://www.scar.org/

library/science-74/life-sciences/antarctic-wildlife-health-network-awhn/5973-risk-assessment-avian-influenza/file/.
Diskin, E.R., et al., 2020. Subtype diversity of influenza A virus in North American waterfowl: a multidecade study. J. Virol. 94. https://doi.org/10.1128/JVI.02022-19 e02022−e02019.
Downie, J.C., Hinshaw, V.S., Laver, W.G., 1977. Ecology of influenza - isolation of type A influenza viruses from Australian pelagic birds. Aust. J. Exp. Biol. Med. Sci. 55, 635−643.
Downie, J.C., Laver, W.G., 1973. Isolation of a type A influenza virus from an Australian pelagic bird. Virology 51, 259−269.
Duan, L., et al., 2008. The development and genetic diversity of H5N1 influenza virus in China, 1996-2006. Virology 380, 243−254.
Dugan, V.G., et al., 2008. The evolutionary genetics and emergence of avian influenza A viruses in wild birds. PLoS Pathog. 4, e1000076. https://doi.org/10.1371/journal.ppat.1000076.
Duriez, O., et al., 2023. Highly pathogenic avian influenza affects vultures' movements and breeding output. Curr. Biol. 33, 3766−3774 e3763.
DWHC, 2023. Keep a Look-Out for Avian Influenza in Black-Headed Gulls. https://dwhc.nl/en/keep-a-look-out-for-avian-influenza-in-black-headed-gulls/.
European Food Safety Authority, Control, et al., 2023. Avian influenza overview March - April 2023. EFSA J. 21, 8039. https://doi.org/10.2903/j.efsa.2023.8039.
FAO, 2005. Economic and Social Impacts of Avian Influenza. FAO Emergency Centre for Transboundary Animal Diseases Operations, Geneva, Switzerland.
FAO, 2022. Global AIV with Zoonotic Potential Situation Update (22 December 2022. https://www.fao.org/ag/againfo/programmes/en/empres/Global_AIV_Zoonotic_Update/situation_update.html.
Feare, C.J., 2007. The role of wild birds in the spread of HPAI H5N1. Avian Dis. 51, 440−447.
Feare, C.J., 2010. Role of wild birds in the spread of highly pathogenic avian influenza virus H5N1 and implications for global surveillance. Avian Dis. 54, 201−212.
Ferenczi, M., et al., 2016. Avian influenza infection dynamics under variable climatic conditions, viral prevalence is rainfall driven in waterfowl from temperate, south-east Australia. Vet. Res. 47, 23. https://doi.org/10.1186/s13567-13016-10308-13562.
Flint, P.L., Franson, J.C., 2009. Does influenza A affect body condition of wild mallard ducks, or vice versa? Proc. R. Soc. B 276, 2345−2346.
Fouchier, R.A.M., et al., 2005. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from Black-headed Gulls. J. Virol. 79, 2814−2822.
Fusaro, A., et al., 2019. Disentangling the role of Africa in the global spread of H5 highly pathogenic avian influenza. Nat. Commun. 10, 5310. https://doi.org/10.1038/s41467-019-13287-y.
Gaidet, N., 2016. Ecology of avian influenza virus in wild birds in tropical Africa. Avian Dis. 60, 296−301.
Gaidet, N., et al., 2012a. Understanding the ecological drivers of avian influenza virus infection in wildfowl: a continental-scale study across Africa. Proc. Biol. Sci. 279, 1131−1141.
Gaidet, N., et al., 2012b. Investigating avian influenza infection hotspots in old-world shorebirds. PLoS One 7, e46049. https://doi.org/10.41371/journal.pone.0046049, 460.
Global Consortium for H5N8, Related Influenza Viruses, 2016. Role for migratory wild birds in the global spread of avian influenza H5N8. Science 354, 213−217.
Gonzalez-Reiche, A.S., et al., 2012. Influenza A viruses from wild birds in Guatemala belong to the North American lineage. PLoS One 7, e32873. https://doi.org/10.1371/journal.pone.0032873.
Gonzalez-Reiche, A.S., et al., 2017. Evidence of intercontinental spread and uncommon variants of low-pathogenicity avian influenza viruses in ducks overwintering in Guatemala. mSphere 2. https://doi.org/10.1128/mSphere.00362-00316.

Granter, A., et al., 2010. The genome sequence of an H11N2 avian influenza virus from a Thick-billed Murre (*Uria lomvia*) shows marine-specific and regional patterns of relationships to other viruses. Virus Gene. 41, 224–230.
Grillo, V.L., et al., 2015. Avian influenza in Australia: a summary of 5 years of wild bird surveillance. Aust. Vet. J. 93, 387–393.
Gruwell, J.A., et al., 2000. Role of peridomestic birds in the transmission of St. Louis encephalitis virus in southern California. J. Wildl. Dis. 36, 13–34.
Guan, Y., et al., 2002. Emergence of multiple genotypes of H5N1 avian influenza viruses in Hong Kong SAR. Proc. Natl. Acad. Sci. U. S. A. 99, 8950–8955.
Guan, Y., et al., 2010. The emergence of pandemic influenza viruses. Protein Cell 1, 9–13.
Halse, S.A., Jaensch, R.P., 1989. Breeding seasons of waterbirds in South-Western Australia - the importance of rainfall. Emu 89, 232–249.
Hansbro, P.M., et al., 2010. Surveillance and analysis of avian influenza viruses, Australia. Emerg. Infect. Dis. 16, 1896–1904.
Hanson, B.A., et al., 2008. Is the occurenence of avian influenza virus in *Charadriiformes* species and location dependant? J. Wildlife Dis. 44, 351–361.
Harvey, J.A., et al., 2023. The changing dynamics of highly pathogenic avian influenza H5N1: next steps for management & science in North America. Biol. Conserv. 282.
Hatchette, T.F., et al., 2004. Influenza A viruses in feral Canadian ducks: extensive reassortment in nature. J. Gen. Virol. 85, 2327–2337.
Hayman, D.T.S., et al., 2013. Ecology of zoonotic infectious diseases in bats: current knowledge and future directions. Zoonoses Public Health 60, 2–21.
Haynes, L., et al., 2009. Australian surveillance for avian influenza viruses in wild birds between July 2005 and June 2007. Aust. Vet. J. 87, 266–272.
Helin, A.S., et al., 2018. A rapid and transient innate immune response to avian influenza infection in mallards. Mol. Immunol. 95, 64–72.
Hill, N.J., et al., 2012. Migration strategy affects avian influenza dynamics in mallards (*Anas platyrhynchos*). Mol. Ecol. 21, 5986–5999.
Hill, N.J., et al., 2010. Waterfowl ecology and avian influenza in California: do host traits inform us about viral occurrence? Avian Dis. 54, 426–432.
Hill, S.C., et al., 2019. Comparative micro-epidemiology of pathogenic avian influenza virus outbreaks in a wild bird population. Philos. Trans. R. Soc. Lond. B Biol. Sci. 374, e20180259. https://doi.org/10.1098/rstb.2018.0259.
Hill, S.C., et al., 2016. Antibody responses to avian influenza viruses in wild birds broaden with age. Proc. Biol. Sci. 283. https://doi.org/10.1098/rspb.2016.2159.
Hoye, B., 2011. Host–Pathogen Interactions on the Move: Migratory Waterfowl and Avian Influenza Viruses. Utrecht University, The Netherlands.
Hoye, B., et al., 2021. Reassortment and persistence of influenza A viruses from diverse geographic origins within Australian wild birds: evidence from a small, isolated population of Ruddy turnstones. J. Virol. https://doi.org/10.1128/JVI.02193-20 e02193–e02120.
Hoye, B., et al., 2016. Hampered performance of migratory swans: intra- and inter-seasonal effects of avian influenza virus. Integr. Comp. Biol. 56, 317–329.
Huang, Y., et al., 2014a. Diverse inter-continental and host lineage reassortant avian influenza A viruses in pelagic seabirds. Infect. Genet. Evol. 22, 103–111.
Huang, Y., et al., 2014b. Perpetuation and reassortment of gull influenza A viruses in Atlantic North America. Virology 456, 353–363.

Hubalek, Z., 2004. An annotated checklist of pathogenic microorganisms associated with migratory birds. J. Wildl. Dis. 40, 639–659.

Hurt, A.C., et al., 2006. Isolation of avian influenza viruses from two different transhemispheric migratory shorebird species in Australia. Arch. Virol. 151, 2301–2309.

Hurt, A.C., et al., 2016. Evidence for the introduction, reassortment, and persistence of diverse influenza A viruses in Antarctica. J. Virol. 90, 9674–9682.

Hurt, A.C., et al., 2014. Detection of evolutionarily distinct avian influenza A viruses in Antarctica. mBio 5. https://doi.org/10.1128/mBio.01098-14 e01098–e01014.

Ignjatovic, J., Sapats, S., 2000. Avian infectious bronchitis virus. Rev. Sci. Tech. 19, 493–508.

Ip, H.S., et al., 2008. Prevalence of influenza A viruses in wild migratory birds in Alaska: patterns of variation in detection at a crossroads of intercontinental flyways. Virol. J 5, 71–81.

Ito, T., et al., 1995. Perpetuation of influenza A viruses in Alaskan waterfowl reservoirs. Arch. Virol. 140, 1163–1172.

Iverson, S.A., et al., 2008. Low prevalence of avian influenza virus in shorebirds on the pacific coast of North America. Waterbirds 31, 602–610.

Jeffries, C.L., et al., 2014. Louping ill virus: an endemic tick-borne disease of Great Britain. J. Gen. Virol. 95, 1005–1014.

Jeglinski, J., et al., 2023. HPAIV Outbreak Triggers Long-Distance Movements in Breeding Northern Gannets–Implications for Disease Spread. Authorea preprint. https://doi.org/10.22541/au.168173494.141037735/v168173492.

Kapczynski, D.R., Swayne, D.E., 2009. Influenza vaccines for avian species. Curr. Top. Microbiol. Immunol. 333, 133–152.

Keawcharoen, J., et al., 2008. Wild ducks as long-distance vectors of highly pathogenic avian influenza virus (H5N1). Emerg. Infect. Dis. 14, 600–606.

Kilpatrick, A.M., et al., 2006. Predicting the global spread of H5N1 avian influenza. Proc. Natl. Acad. Sci. U S A 103, 19368–19373.

Kishida, N., et al., 2008. H2N5 influenza virus isolates from terns in Australia: genetic reassortments between those of the Eurasian and American lineages. Virus Gene. 37, 16–21.

Klaassen, M., Wille, M., 2023. The plight and role of wild birds in the current bird flu panzootic. Nat. Ecol. Evol. 7, 1541–1542.

Kleijn, D., et al., 2010. Dynamics and ecological consequences of avian influenza virus infection in greater white-fronted geese in thier winter staging areas. Proc. Biol. Sci. 277, 2041–2048.

Knief, U., et al., 2023. Highly pathogenic avian influenza causes mass mortality in Sandwich tern (*Thalasseus sandvicensis*) breeding colonies across northwestern Europe. bioRxiv. https://www.biorxiv.org/content/10.1101/2023.1105.1112.540367v540361.

Kozlov, M., 2023. US will vaccinate birds against avian flu for first time — what researchers think. Nature 618, 220–221. https://doi.org/10.1038/d41586-023-01760-0.

Krauss, S., et al., 2010. Coincident ruddy turnstone migration and horseshoe crab spawning creates an ecological 'hot spot' for influenza viruses. Proc. Biol. Sci. 277, 3373–3379.

Krauss, S., et al., 2016. The enigma of the apparent disappearance of Eurasian highly pathogenic H5 clade 2.3.4.4 influenza A viruses in North American waterfowl. Proc. Natl. Acad. Sci. U S A 113, 9033–9038.

Kuiken, T., 2013. Is low pathogenic avian influenza virus virulent for wild waterbirds? Proc. Biol. Sci. 280, 20130990. https://doi.org/10.1098/rspb.2013.0990.

Lane, J.V., et al., 2023. High pathogenicity avian influenza (H5N1) in Northern Gannets: global spread, clinical signs, and demographic consequences. bioRxiv. https://doi.org/10.1111/ibi.13275.

Lang, A.S., et al., 2016. Assessing the role of seabirds in the ecology of influenza A viruses. Avian Dis. 60, 378–386.

Latorre-Margalef, N., et al., 2013. Heterosubtypic immunity to influenza A virus infections in Mallards may explain existence of multpile virus subtypes. PLoS Pathog. 9 (6), e1003443. https://doi.org/10.1371/journal.ppat.1003443.
Latorre-Margalef, N., et al., 2009a. Does influenza A affect body condition of wild mallard ducks, or vice versa? A reply to Flint and Franson. Proc. R. Soc. B 276, 2347−2349.
Latorre-Margalef, N., et al., 2009b. Effects of influenza A virus infection on migrating mallard ducks. Proc. R. Soc. A B 276, 1029−1036.
Latorre-Margalef, N., et al., 2015. Serologic evidence of influenza A (H14) virus introduction into North America. Emerg. Infect. Dis. 21, 2257−2259.
Latorre-Margalef, N., et al., 2014. Long-term variation in influenza A virus prevalence and subtype diversity in a migratory Mallards in Northern Europe. Proc. R. Soc. B 281. https://doi.org/10.1098/rspb.2014.0098.
Laver, W.G., Webster, R.G., 1972. Antibodies to human influenza virus neuraminidase (the A-Asian-57 H2N2 strain) in sera from Australian pelagic birds. Bull. World Health Organ. 47, 535−541.
Lebarbenchon, C., et al., 2015a. Influenza A virus on oceanic islands: host and viral diversity in seabirds in the Western Indian Ocean. PLoS Pathog. 11, e1004925. https://doi.org/10.1371/journal.ppat.1004925.
Lebarbenchon, C., et al., 2012. Reassortment influenza A viruses in wild duck populations: effects on viral shedding and persistence in water. Proc. Biol. Sci. 279, 3967−3975. https://doi.org/10.1098/rspb.2012.1271.
Lebarbenchon, et al., 2015b. Isolation of type A influenza viruses from Red-necked Grebes (*Podiceps grisegena*). J. Wildl. Dis. 51, 290−293.
Lee, Y.J., et al., 2014. Novel reassortant influenza A(H5N8) viruses, South Korea, 2014. Emerg. Infect. Dis. 20, 1087−1089.
Leguia, M., et al., 2023. Highly pathogenic avian influenza A (H5N1) in marine mammals and seabirds in Peru. bioRxiv. https://doi.org/10.1038/s41467-023-41182-0.
Lindeborg, M., et al., 2012. Migratory birds, ticks, and crimean-Congo hemorrhagic fever virus. Emerg. Infect. Dis. 18, 2095−2097.
Lindstrom, S.E., et al., 2004. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957-1972: evidence for genetic divergence and multiple reassortment events. Virology 328, 101−119.
Lisovski, S., Hoye, B.J., Klaassen, M., 2017. Geographic variation in seasonality and its influence on the dynamics of an infectious disease. Oikos 126, 931−936.
Lisovski, S., et al., 2018. The roles of migratory and resident birds in local avian influenza infection dynamics. J. Appl. Ecol. 55, 2963−2975.
Liu, J., et al., 2005. Highly pathogenic H5N1 influenza virus infection in migratory birds. Science 309, 1206.
Lowen, A.C., 2017. Constraints, drivers, and implications of Influenza A virus reassortment. Ann. Rev. Virol. 4, 105−121.
Lowen, A.C., 2018. It's in the mix: reassortment of segmented viral genomes. PLoS Pathog. 14, e1007200. https://doi.org/10.1371/journal.ppat.1007200.
Lundqvist, M.L., et al., 2006. Immunoglobulins of the non-galliform birds: antibody expression and repertoire in the duck. Dev. Comp. Immunol. 30, 93−100.
Lv, X., et al., 2022. Highly pathogenic avian influenza A(H5N8) clade 2.3.4.4b viruses in satellite-tracked wild ducks, Ningxia, China, 2020. Emerg. Infect. Dis. 28, 1039−1042.
Lycett, S.J., Duchatel, F., Digard, P., 2019. A brief history of bird flu. Philos. Trans. R. Soc. B 374, 20180257. https://doi.org/10.1098/rstb.2018.0257.
Lytras, S., et al., 2021. The animal origin of SARS-CoV-2. Science 373, 968−970.
Mackenzie, J.S., et al., 1984. Isolation of otho- viruses and paramyxoviruses from wild birds in Western Australia and the characterization of novel influenza A viruses. Aust. J. Exp. Biol. Med. Sci. 62, 89−99.
Magor, K.E., 2011. Immunoglobulin genetics and antibody responses to influenza in ducks. Dev. Comp. Immunol. 35, 1008−1016.

Magor, K.E., et al., 1999. Opposite orientation of the alpha- and upsilon-chain constant region genes in the immunoglobulin heavy chain locus of the duck. Immunogenetics 49, 692−695.
Magor, K.E., et al., 1998. Secretory immune system of the duck (*Anas platyrhynchos*). Identification and expression of the genes encoding IgA and IgM heavy chains. Eur. J. Immunol. 28, 1063−1068.
Maxted, A.M., et al., 2012. Avian influenza virus infection dynamics in shorebird hosts. J. Wildl. Dis. 48, 322−334.
Maxted, A.M., et al., 2016. Spring migration stopover ecology of avian influenza virus shorebird hosts at Delaware Bay. Avian Dis. 60, 394−405.
Melville, D.S., et al., 2016. Shorebirds along the Yellow Sea coast of China face an uncertain future − a review of threats. Emu 116, 100−110. https://doi.org/10.1071/MU15045.
Molini, U., et al., 2020. Avian influenza H5N8 outbreak in African penguins (*Spheniscus demersus*), Namibia, 2019. J. Wildlfe Dis. 56, 214−218.
Munster, V., Fouchier, R.A.M., 2009. Avian influenza virus: of virus and bird ecology. Vaccine 27, 6340−6344.
Munster, V.J., et al., 2007. Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. PLoS Pathog. 3, e61. https://doi.org/10.1371/journal.ppat.0030061.
Murray, K.O., et al., 2010. West Nile virus and its emergence in the United States of America. Vet Res 41, 67. https://doi.org/10.1051/vetres/2010039.
Nayak, B., et al., 2010. Contributions of the avian influenza virus HA, NA, and M2 surface proteins to the induction of neutralizing antibodies and protective immunity. J. Virol. 84, 2408−2420.
Nemeth, N., et al., 2012. North American birds as potential amplifying hosts of Japanese Encephalitis Virus. Am. J. Trop. Med. Hyg. 87, 760−767.
Norman, F.I., Nicholls, N., 1991. The Southern Oscillation and variations in waterfowl abundance in southeastern Australia. Aust. J. Ecol. 16, 485−490.
Normile, D., 2006. Avian influenza. Evidence points to migratory birds in H5N1 spread. Science 311, 1225.
Olsen, B., et al., 2006. Global patterns of influenza A virus in wild birds. Science 312, 384−388.
Olson, S.H., et al., 2014. Sampling strategies and biodiversity of influenza A subtypes in wild birds. PLoS One 9, e90826. https://doi.org/10.1371/journal.pone.0090826.
Pasick, J., et al., 2015. Reassortant highly pathogenic influenza A H5N2 virus containing gene segments related to Eurasian H5N8 in British Columbia, Canada, 2014. Sci. Rep. 5, 9484.
Pearce, J.M., et al., 2011. Interspecific exchange of avian influenza virus genes in Alaska: the influenza of trans-hemispheric migratory tendency and breeding ground sympatry. Mol. Ecol. 20, 1015−1025.
Plaza, P., et al., 2023. Pacific and Atlantic Sea Lion Mortality Caused by Highly Pathogenic Avian Influenza A(H5n1) in South America. SSRN preprint. https://doi.org/10.2139/ssrn.4611782.
Poen, M.J., et al., 2018. Local amplification of highly pathogenic avian influenza H5N8 viruses in wild birds in The Netherlands, 2016 to 2017. Euro Surveill. 23, 17−00449. https://doi.org/10.2807/1560-7917.ES.2018.23.4.17-00449.
Poen, M.J., et al., 2019. Co-circulation of genetically distinct highly pathogenic avian influenza A clade 2.3.4.4 (H5N6) viruses in wild waterfowl and poultry in Europe and East Asia, 2017-18. Virus Evol. 5, vez004. https://doi.org/10.1093/ve/vez1004.
Poen, M.J., et al., 2016. Lack of virological and serological evidence for continued circulation of highly pathogenic avian influenza H5N8 virus in wild birds in The Netherlands, 14 November 2014 to 31 January 2016. Euro Surveill. 21, 11−21.
Promed mail, 2021a. Avian Influenza (09): Europe (UK) Wild Bird, HPAI, Conservation Concern. Archive Number: 20220106.8700696. https://promedmail.org/promed-post/?id=8700696.
Promed mail, 2021b. Avian Influenza (17): Asia (Israel) HPAI H5N1, Wild Bird, Poultry, Epidemiology. https://promedmail.org/promed-post/?id=8700909.

Pybus, O.G., Rambaut, A., 2009. Evolutionary analysis of the dynamics of viral infectious disease. Nat. Rev. Genet. 10, 540−550.

Ramey, A.M., et al., 2010. Transmission and reassortment of avian influenza viruses at the Asian-North American interface. Virology 406, 352−359.

Ramey, A.M., et al., 2014a. Genomic characterization of H14 subtype Influenza A viruses in new world waterfowl and experimental infectivity in mallards (*Anas platyrhynchos*). PLoS One 9, e95620. https://doi.org/10.1371/journal.pone.0095620.

Ramey, A.M., et al., 2014b. Optimizing surveillance for South American origin influenza A viruses along the United States gulf coast through genomic characterization of isolates from Blue-winged Teal (*Anas discors*). Transbound Emerg Dis 63, 194−202.

Reuters, 2022. More Penguins Dying from Avian Flu at Cape Town's Boulders Beach Colony. https://www.reuters.com/business/environment/more-penguins-dying-avian-flu-cape-towns-boulders-beach-colony-2022-2010-2001/.

Richard, M., et al., 2017. Mechanisms and risk factors for mutation from low to highly pathogenic avian influenza virus. EFSA Support. Publ. https://doi.org/10.2903/sp.efsa.2017.EN-1287. EN-1287, 26pp.

Rimondi, A., et al., 2018. Evidence of a fixed internal gene constellation in influenza A viruses isolated from wild birds in Argentina (2006-2016). Emerg. Microbes Infect. 7, 194. https://doi.org/10.1038/s41426-018-0190-2.

Roshier, D., Robertson, A.I., Kingsford, R.T., 2002. Responses of waterbirds to flooding in an arid region of Australia and implications for conservation. Biol. Conserv. 106, 399−441.

RSPB, 2023. Avian Flu (Bird Flu). https://www.rspb.org.uk/birds-and-wildlife/advice/how-you-can-help-birds/disease-and-garden-wildlife/avian-influenza-updates/.

Runstadler, J., et al., 2007. Using RRT-PCR analysis and virus isolation to determine the prevalence of avian influenza virus infections in ducks at Minto Flats Refuge, Alaska, during August 2005. Arch. Virol. 152, 1901−1910.

Sazonov, S.S., et al., 1977. Isolation of an influenza virus, similar to A/Port Chalmers/1/73(H3N2) from a common murre at Sakhalin island in U.S.S.R (strain A/CommonMurre/Sakhalin/1/74). Arch. Virol. 53, 1−7.

SCAR, 2023. Sub-Antarctic and Antarctic Highly Pathogenic Avian Influenza H5N1 Monitoring Project. https://www.scar.org/resources/hpai-monitoring-project/.

Scholtissek, C., et al., 1978. On the origin of human influenza virus subtypes H2N2 and H3N2. Virology 87, 13−20.

Seekings, A.H., et al., 2018. Direct evidence of H7N7 avian influenza virus mutation from low to high virulence on a single poultry premises during an outbreak in free range chickens in the UK, 2008. Infect. Genet. Evol. 64, 13−31.

Sharp, G.B., et al., 1997. Coinfection of wild ducks by influenza A viruses: distribution patterns and biological significance. J. Virol. 71, 6128−6135.

Shi, J.Z., et al., 2018. Rapid evolution of H7N9 highly pathogenic viruses that emerged in China in 2017. Cell Host Microbe 24, 558−568.

Shortridge, K.F., et al., 1998. Characterization of avian H5N1 influenza viruses from poultry in Hong Kong. Virology 252, 331−342.

Simulundu, E., et al., 2009. Characterization of H3N6 avian influenza virus isolated from a wild white pelican in Zambia. Arch. Virol. 154, 1517−1522.

Slusher, M.J., et al., 2014. Are passerine birds reservoirs for influenza a viruses? J. Wildl. Dis. https://doi.org/10.7589/2014-7502-7043.

Smith, G.J., et al., 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature 459, 1122−1125.

Spackman, E., et al., 2023. The pathogenesis of a 2022 North American highly pathogenic clade 2.3.4.4b H5N1 avian influenza virus in mallards (Anas platyrhynchos). Avian Pathol. 52, 219−228.

Stallknecht, D.E., Brown, J.D., 2007. Wild birds and the epidemiology of avian influenza. J. Wildl. Dis. 43, S15−S20.

Stamoulis, K., 2017. The economic impact of avian influenza, Avian Influenza. In: A Global Thread. In the Framework of the G7 CVOs Meeting. Rome, Greece. https://www.izsvenezie.it/documenti/formazione/corsi-convegni/2017/2017-10-04-avian-influenza-global-threat/stamoulis.pdf.

Starick, E., et al., 2008. Phylogenetic analyses of highly pathogenic avian influenza virus isolates from Germany in 2006 and 2007 suggest at least three separate introductions of H5N1 virus. Vet. Microbiol. 128, 243−252.

Steel, J., Lowen, A.C., 2014. Influenza a virus reassortment. Curr. Top. Microbiol. Immunol. 385, 377−401.

Stoute, S., et al., 2016. Highly pathogenic Eurasian H5N8 avian influenza outbreaks in two commercial poultry flocks in California. Avian Dis. 60, 688−693.

Stuecker, M.F., et al., 2017. Revisiting ENSO/Indian Ocean Dipole phase relationships. Geophys. Res. Lett. 44, 2481−2492.

Sturm-Ramirez, K., et al., 2005. Are ducks contributing to the endemicity of highly pathogenic H5N1 influenza virus in Asia? J. Virol. 79, 11269−11279.

Sturm-Ramirez, K.M., et al., 2004. Reemerging H5N1 influenza viruses in Hong Kong in 2002 are highly pathogenic to ducks. Virology 78, 4892−4901.

Teitelbaum, C.S., et al., 2023. North American wintering mallards infected with highly pathogenic avian influenza show few signs of altered local or migratory movements. Sci. Rep. 13, 14473. https://doi.org/10.1038/s41598-023-40921-z.

Tolf, C., et al., 2013. Individual variation in influenza A virus infection histories and long-term immune responses in mallards. PLoS One 8, e61201. https://doi.org/10.61371/journal.pone.0061201, 612.

Tong, S., et al., 2012. A distinct lineage of influenza A virus from bats. Proc. Natl. Acad. Sci. U S A 109, 4269−4274.

Tong, S.X., et al., 2013. New world bats harbor diverse influenza A viruses. PLoS Pathog. 9, e1003657. https://doi.org/10.1371/journal.ppat.1003657.

van Dijk, J., et al., 2015. Weak negative associations between avian influenza virus infection and movement behaviour in a key host species, the mallard Anas platyrhynchos. Oikos 10, 1293−1303. https://doi.org/10.1111/oik.01836.

van Dijk, J.G.B., et al., 2014. Juveniles and migrants as drivers for seasonal epizootics of avian influenza virus. J. Anim. Ecol. 83, 266−275.

Van Dijk, J.G., et al., 2020. A comparative study of the innate humoral immune response to avian influenza virus in wild and domestic mallards. Front 11, 608274.

van Gils, J.A., et al., 2007. Hampered foraging and migratory performance in swans infected with low-pathogenic avian influenza A virus. PLoS One 2, e184. https://doi.org/10.1371/journal.pone.0000184.

van Toor, M.L., et al., 2018. As the duck flies - estimating the dispersal of low pathogenic avian influenza viruses by migrating mallards. Front. Ecol. Evol. 6, 1−15. https://doi.org/10.3389/fevo.2018.00208.

Vanderven, H.A., et al., 2012. Avian influenza rapidly induces antiviral genes in duck lung and intestine. Mol. Immunol. 51, 316−324.

Velarde, R., et al., 2010. Avian influenza virus H13 circulating in Ring-billed gulls (*Larus delawarensis*) in southern Ontario, Canada. Avian Dis. 54, 411−419.

Verhagen, J.H., et al., 2021. Host range of influenza A virus H1 to H16 in Eurasian ducks based on tissue and receptor binding studies. J. Virol. 95. https://doi.org/10.1128/JVI.01873-20 e01873−e01820.

Verhagen, J.H., et al., 2015. How a virus travels the world. Science 347, 616−617.

Verhagen, J.H., et al., 2017. Discordant detection of avian influenza virus subtypes in time and space between poultry and wild birds; towards improvement of surveillance programs. PLoS One 12. https://doi.org/10.1371/journal.pone.0173470.

Verhagen, J.H., et al., 2014a. Epidemiology of influenza A virus among black-headed gulls, The Netherlands, 2006-2010. Emerg. Infect. Dis. 20, 138−141.
Verhagen, J.H., et al., 2020. Phylogeography and antigenic diversity of low pathogenic avian influenza H13 and H16 viruses. J. Virol. 94. https://doi.org/10.1128/JVI.00537-20 e00537−e00520.
Verhagen, J.H., et al., 2014b. Migratory birds reinforce local circulation of avian influenza viruses. PLoS One 9, e112366. https://doi.org/10.1371/journal.pone.0112366.
Vibin, J., et al., 2018. Metagenomics detection and characterisation of viruses in faecal samples from Australian wild birds. Sci. Rep. 8, e8686. https://doi.org/10.1038/s41598-018-26851-1.
Vibin, J., et al., 2020a. Metagenomic characterisation of additional and novel avian viruses from Australian wild ducks. Sci. Rep. 10, 22284. https://doi.org/10.1038/s41598-020-79413-9.
Vibin, J., et al., 2020b. Metagenomic characterisation of avian parvoviruses and picornaviruses from Australian wild ducks. Sci. Rep. 10, 12800. https://doi.org/10.1038/s41598-020-69557-z.
Vijaykrishna, D., et al., 2013. The recent establishment of North American H10 lineage influenza viruses in Australian wild waterfowl and the evolution of Australian avian influenza viruses. J. Virol. 87, 10182−10189.
Vijaykrishna, D., et al., 2015. The contrasting phylodynamics of human influenza B viruses. Elife 4, e05055. https://doi.org/10.7554/eLife.05055.
Vijaykrishna, D., et al., 2011. Long-term evolution and transmission dynamics of swine influenza A virus. Nature 473, 519−U263.
Voorhees, I.E.H., et al., 2018. Multiple incursions and recurrent epidemic fade-out of H3N2 canine influenza A virus in the United States. J. Virol. 92.
Waldenstrom, J., et al., 2007. Migrating birds and tickborne encephalitis virus. Emerg. Infect. Dis. 13, 1215−1218.
Wallensten, A., et al., 2006. Mounting evidence for the presence of influenza A virus in the avifauna of the Antarctic region. Antarct. Sci. 18, 353−356.
Wang, B., et al., 2012. Complete genome sequence of an H9N2 avian influenza virus isolated from egret in Lake Dongting wetland. J. Virol. 86, 11939. https://doi.org/10.1128/JVI.02042-12.
Warr, G.W., et al., 1995. IgY: clues to the origins of modern antibodies. Immunol. Today 16, 392−398.
Webby, R.J., et al., 2003. Multiple genotypes of nonpathogenic H6N2 influenza viruses isolated from chickens in California. Avian Dis. 47, 905−910.
Webster, R.G., et al., 1992. Evolution and ecology of influenza A viruses. Microbiol. Rev. 56, 152−179.
Webster, R.G., Rott, R., 1987. Influenza virus A pathogenicity: the pivotal role of hemagglutinin. Cell 50, 665−666.
Wilcox, B.R., et al., 2011. Influenza A viruses in ducks in Northwestern Minnesota: fine scale spatial and temporal variation in prevalence and subtype diversity. PLoS One 6, e24010. https://doi.org/10.1371/journal.pone.0024010.
Wille, M., et al., 2011a. Extensive geographic mosaicism in avian influenza viruses from gulls in the northern hemisphere. PLoS One 6, e20664. https://doi.org/10.1371/journal.pone.0020664.
Wille, M., et al., 2011b. Reassortment of American and Eurasian genes in an influenza A virus isolated from a great black-backed gull (*Larus marinus*), a species demonstrated to move between these regions. Arch. Virol. 156, 107−115.
Wille, M., et al., 2013. Frequency and patterns of reassortment in natural influenza A virus infection in a reservoir host. Virology 443, 150−160.
Wille, M., et al., 2014. Evaluation of seabirds in Newfoundland and Labrador, Canada, as hosts of influenza A viruses. J. Wildl. Dis. 50, 98−103.
Wille, M., et al., 2018a. Virus-virus interactions and host ecology are associated with RNA virome structure in wild birds. Mol. Ecol. 27, 5263−5278. https://doi.org/10.1111/mec.14918, 52.

Wille, M., et al., 2018b. Where do all the subtypes go? Temporal dynamics of H8−H12 influenza A viruses in waterfowl. Virus Evol. 4. https://doi.org/10.1093/ve/vey1025.
Wille, M., et al., 2019a. Serologic evidence of exposure to highly pathogenic avian influenza H5 viruses in migratory shorebirds, Australia. Emerg. Infect. Dis. 25, 1903−1910.
Wille, M., et al., 2019b. Virome heterogeneity and connectivity in waterfowl and shorebird communities. ISME J. 13, 2603−2616.
Wille, M., et al., 2021a. How accurately can we assess zoonotic risk? PLoS Pathog. 19, e3001135. https://doi.org/10.1371/journal.pbio.3001135.
Wille, M., et al., 2021b. RNA virome abundance and diversity is associated with host age in a bird species. Virology 561, 98−106.
Wille, M., et al., 2022a. Australia as a global sink for the genetic diversity of avian influenza A virus. PLoS Pathog. 18, e1010150. https://doi.org/10.1371/journal.ppat.1010150.
Wille, M., et al., 2022b. Strong phylogenetic and ecological effects on host competency for avian influenza in Australian wild birds. Proc. R. Soc. B 290, e29020222237. https://doi.org/10.1098/rspb.2022.2237.
Wille, M., Barr, I.G., 2022. Resurgence of avian influenza virus. Science 376, 459−460.
Wille, M., Klaassen, M., 2023. No evidence for HPAI H5N1 2.3.4.4b incursion into Australia in 2022. Influenza Other Respir. Viruses 17, e13118.
Wille, M., Waldenström, J., 2023. Weathering the storm of high pathogenicity avian influenza in waterbirds. Waterbirds 46, 100−109.
Winker, K., et al., 2008. Rarity of influenza A virus in spring shorebirds, Southern Alaska. Emerg. Infect. Dis. 14, 1314−1316.
Worobey, M., et al., 2014. A synchronized global sweep of the internal genes of modern avian influenza virus. Nature 508, 254−257.
Wu, H.B., et al., 2014. Novel reassortant influenza A(H5N8) viruses in domestic ducks, Eastern China. Emerg. Infect. Dis. 20, 1315−1318.
Xie, R., et al., 2023. The episodic resurgence of highly pathogenic avian influenza H5 virus. Nature 622, 810−817.
Zohari, S., et al., 2014. Avian influenza A(H10N7) virus involvement in mass mortality of harbour seals (*Phoca vitulina*) in Sweden, March through October 2014. Euro Surveill. 19, 20967. https://doi.org/10.2807/1560-7917.es2014.19.46.20967.

CHAPTER

Evolution of SARS-COV-2

32

Manuela Sironi[1] **and Fernando González-Candelas**[2,3,4]

[1]*Scientific Institute IRCCS E.MEDEA, Bioinformatics, Bosisio Parini, Italy;* [2]*University of Valencia, Institute for Integrative Systems Biology (I2SysBio, UV-CSIC), Valencia, Spain;* [3]*CIBER in Epidemiology and Public Health, Valencia, Spain;* [4]*FISABIO Foundation, Joint Research Unit Infection and Public Health, Valencia, Spain*

1. Introduction

The COVID-19 pandemic started as a small outbreak of pneumonia-like illness in Wuhan, China, during December 2019. Since then, it has escalated to become the most devastating pandemic of the 21st century. SARS-CoV-2, the causative agent, is a coronavirus and, being newly emerged, jumped into a fully susceptible population. The lack of preexisting immunity, its high human-to-human transmissibility, and the ability to be transmitted by asymptomatic/pre-symptomatic individuals have contributed to rapid viral spread. Thus, SARS-CoV-2 has reached every corner of the world creating unprecedented public health emergencies and concerns.

The unfolding of the pandemic was accompanied by synergistic efforts of the global scientific community to study SARS-CoV-2 biology, epidemiology, and evolution, as well as to develop diagnostic tools and treatment strategies. The licensing of effective vaccines against COVID-19 was achieved with unprecedented, unimaginable speed.

Also, like never before, the spread of the virus has been tracked nearly in real-time by genome surveillance programs throughout the world. This generated an enormous wealth of data, allowing researchers to study the evolutionary trajectories of SARS-CoV-2, map the occurrence of new viral mutations, and pinpoint the emergence of novel lineages. Despite such enormous advances, some pieces of information are still missing. Indeed, sequencing efforts in China and Southeastern Asia have provided information about the huge diversity of coronaviruses hosted by bats and other animals, but the events associated with the initial spillover of SARS-CoV-2 remain unclear.

2. Coronaviruses

2.1 Classification and Phylogenetic Relationships

Coronaviruses (CoVs, family Coronaviridae, order *Nidovirales*, kingdom *Orthornavirae*, realm *Riboviria*) are a large family of enveloped, nonsegmented, positive-sense RNA viruses. Within the Coronaviridae family, the *Orthocoronavirinae* subfamily comprises a variety of viruses that infect mammals

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00012-4

Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

and birds (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/222/coronaviridae). This subfamily also includes nine CoVs that have been shown to infect humans (Forni et al., 2017; Zhou et al., 2020; Cui et al., 2019). Three of such human CoVs are highly pathogenic and are often referred to as "epidemic" (SARS-CoV-2, SARS-CoV, and MERS-CoV). Other four human CoVs (HCoV-OC43, HcoV-NL63, HcoV-229E, and HcoV-HKU1) usually cause mild symptoms and are known as "endemic" or "common cold" CoVs (Forni et al., 2017; Zhou et al., 2020; Cui et al., 2019). The last two CoVs have been known for years but were only recently reported to infect humans: in Malaysia, a canine alphacoronavirus was detected in patients with pneumonia, whereas in Haiti, porcine deltacoronavirus was found in the plasma samples of three children with febrile illness (Vlasova et al., 2022; Lednicky et al., 2021). The pathogenic potential and transmissibility of these two CoVs in humans will need to be evaluated further.

The *Orthocoronavirinae* subfamily is divided into four genera *(Alphacoronavirus, Betacoronavirus, Gammacoronavirus*, and *Deltacoronavirus)* and several subgenera (https://talk.ictvonline.org/taxonomy/). Individual viruses are classified based on pairwise patristic distances calculated on conserved genome regions and relative thresholds, which demarcate virus clusters at different ranks. The approach is implemented in the DivErsity pArtitioning by hieRarchical Clustering (DEmARC) software (Lauber and Gorbalenya, 2012). Human-infecting coronaviruses are classified in different genera. Specifically, SARS-CoV-2, as well as SARS-CoV, MERS-CoV, HCoV-OC43, and HCoV-HKU1 belong to the *Betacoronavirus* genus. Within this genus, SARS-CoV-2 and SARS-CoV belong to the *Sarbecovirus* subgenus, which is populated by a number of bat viruses. These two human viruses are classified within the same species *Severe acute respiratory syndrome-related coronavirus* (Lauber and Gorbalenya, 2012; Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020).

2.2 Genomic Features

CoV genomes are unusually long and complex compared to those of other RNA viruses, ranging in length from ~27 to ~31 kb (Forni et al., 2017; Zhou et al., 2020; Cui et al., 2019; Brian and Baric, 2005). In general, all RNA viruses share a common feature: an error-prone RNA-dependent RNA polymerase (RdRp), which determines high mutation rates and fast adaptation to new hosts or environments (Holmes, 2003). Such high error rates are also thought to limit genome size, because, as genome size expands, the number of deleterious mutations also increases to a point that affects fitness (Holmes, 2003). Thus, human RNA virus genomes have an average length of 9−10 kb, with only few families, including coronaviruses, deviating sensibly from this size (Belshaw et al., 2007; Simon-Loriere et al., 2013). Indeed, it is widely believed that increased replication fidelity allowed genome expansion in coronaviruses: because they encode enzymes with proofreading ability (i.e., an RNA 3′-to-5′ exoribonuclease and an endoribonuclease) (Eckerle et al., 2007, 2010; Smith et al., 2013), coronaviruses typically have lower mutation rates compared to other RNA viruses (Cui et al., 2019; Forni et al., 2017; Zhou et al., 2020). Nonetheless, features distinct from replication fidelity, such as a peculiar genome organization and a processive replication complex, may have also contributed to genome expansion in CoVs (Subissi et al., 2014).

All coronaviruses have linear genomes with similar organization and encode about 20 proteins common to all family members (Forni et al., 2017; Zhou et al., 2020; Cui et al., 2019). Invariably, about two-thirds of the genome consist of two large open reading frames (ORF1a and ORF1b), which overlap by few nucleotides at their termini. Both ORF1a and ORF1b are translated into polyproteins

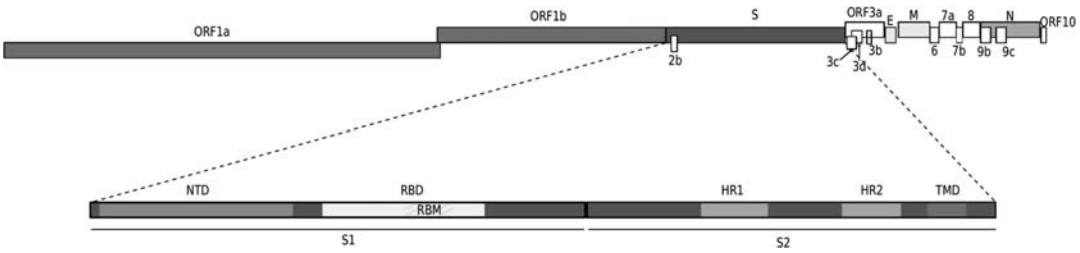


FIGURE 32.1 Schematic representation of the coding regions of the SARS-CoV-2 genome

ORFs that are present in all CoV genomes are shaded in gray. ORFs encoding accessory proteins are in white. The enlargement shows the domain structure of the spike protein. NTD, N-terminal domain; RBD, receptor binding domain; RBM, receptor binding motif, HR1 and HR2, heptad repeats 1 and 2; TMD, transmembrane domain.

(Fig. 32.1). These latter are processed to generate 16 nonstructural proteins (nsp1 to nsp16) that carry out different, essential functions during viral replication. The remaining portion of the genome includes ORFs for the structural proteins, namely spike (S), envelope (E), membrane (M), and nucleoprotein (N) (Forni et al., 2017; Zhou et al., 2020; Cui et al., 2019) (Fig. 32.1). The spike protein, whose trimers decorate the CoV surface, mediates receptor binding and cell entry. It thus represents a major determinant of host range and includes two functionally distinct units: the S1 region (which contains an N-terminal domain, NTD, and the receptor binding domain, RBD) and the S2 region (which includes the fusion peptide, two heptad repeats and the transmembrane region) (Graham and Baric, 2010; Subissi et al., 2014) (Fig. 32.1). S proteins show remarkable diversity across CoVs, whereas M, E, and N tend to be more conserved in sequence.

CoVs also encode a variable number of accessory proteins, which differ in sequence and number even among closely related viruses (Forni et al., 2022). Recent analysis of accessory ORFs in the four CoV genera indicated that a minority of them (e.g., orthologs of SARS-CoV-2 ORF3a) are conserved in several coronavirus genomes, even in different genera. The majority, though, are present in one or a few coronavirus genomes. The origin of most of these accessory proteins is unknown although some were clearly captured from host genomes or from other viruses (Forni et al., 2022). Recombination among CoVs also contributed to diversifying the complement of accessory proteins, clearly reflecting the highly dynamic nature of CoV genome evolution (Forni et al., 2022).

Although dispensable for virus replication, accessory proteins often play a role at the host–virus interface and represent virulence factors (Forni et al., 2017, 2022; Zhou et al., 2020; Cui et al., 2019). In the case of SARS-CoV-2, six accessory proteins (ORF3a, 6, 7a, 7b, 8, and 10) are annotated in the reference genome (GenBank ID: NC_045512.2) (Fig. 32.1). Moreover, using different approaches, several unannotated accessory ORFs were described, including some within ORFs S (ORF2b), N (ORF9b, ORF9c), and ORF3a (ORF3b, ORF3c, ORF3d). These internal ORFs are predicted to be translated using different reading frames (Finkel et al., 2021) (Fig. 32.1).

The functional characterization of accessory proteins has been an area of intense research because these ORFs might contribute to the pathogenetic mechanisms of SARS-CoV-2. For instance, ORF8 was shown to affect different aspects of the human immune response (Beaudoin-Bussières et al., 2022; Li et al., 2020; Rashid et al., 2021; Wu et al., 2022; Geng et al., 2021; Zhang et al., 2021), and the ORF9b

protein antagonizes type I and III interferons by targeting multiple innate antiviral signaling pathways (Han et al., 2021). Likewise, ORF10 inhibits the cell innate immune response by the induction of MAVS degradation (Xingyu Li et al., 2022). Also, ORF3a and ORF7a were reported to block autophagy (Hayn et al., 2021; Hou et al., 2023; Xingyu Li et al., 2022; Miao et al., 2021; Chen et al., 2021).

2.3 Host Range and Ecology

The largest diversity of alpha and betaCoVs is thought to be hosted by bats (Ghai et al., 2021). Across different geographic locations, bats from several families have been found positive for one or more CoVs. In a survey of CoV diversity in multiple host taxa from 20 countries, sequences representing 100 discrete phylogenetic clusters were detected, 91 of which were found in bats. Extrapolation from these data led to the conclusion that bats host more than 3000 coronaviruses, most of which remain undescribed (Anthony et al., 2017). Conversely, the largest diversity of gammaCoVs and deltaCoVs is hosted by wild birds (Wille and Holmes, 2020). Rodents also host a diversity of CoVs. Whereas prior to 2015 the only known rodent CoV was mouse hepatitis virus (MHV), field studies have identified a number of CoVs in wild rodents, as well as in shrews. Rodent CoVs mainly cluster within the *Embecovirus* subgenus (*Betacoronavirus* genus) and the *Luchacovirus* subgenus (*Alphacoronavirus* genus) (Tsoleridis et al., 2016; Wang et al., 2015, 2020; Wu et al., 2018; Ge et al., 2017). As for shrew CoVs, to date they were all sampled in China and represent a divergent group of alphacoronaviruses (Wang et al., 2017).

As far as domestic animals are concerned, CoVs that infect livestock and companion animals are also common. At least six CoVs can cause disease in pigs. These include three alphacoronaviruses (transmissible gastroenteritis virus, porcine respiratory coronavirus, porcine epidemic diarrhea virus, and swine acute diarrhea syndrome-coronavirus), one betacoronavirus (betacoronavirus 1, also referred to as porcine hemagglutinating encephalomyelitis virus), and a porcine deltacoronavirus (Wang et al., 2019). Betacoronavirus 1 also infects other mammals. In fact, unique among CoVs, this virus is a species complex with distinct host-specific strains. The best studied of these is bovine coronavirus (BCoV), which can be associated with a suite of clinical diseases in calves and cattle. BCoV also infects several other livestock species, including horses, sheep, and camels (Ghai et al., 2021). Other betacoronavirus 1 strains were instead associated with infection in rabbits and dogs (Ghai et al., 2021). Thus, canine respiratory coronavirus, which is prevalent in many countries, can cause mild to severe respiratory signs (Priestnall, 2020). Dogs (and cats) also suffer from infection with alphacoronaviruses. Canine enteric coronavirus (CCoV) is often associated with mild enteritis in puppies and dogs, especially in kennels. However, a recently emerged pantropic variant identified in 2005 was reported to result in a very high mortality rate in puppies (Buonavoglia et al., 2006). For this reason, CCoV is considered an emerging pathogen of dogs (Licitra et al., 2014). CCoV is closely related to TGEV (transmissible gastroenteritis virus) and to feline coronavirus (FCoV), with which it can recombine (Licitra et al., 2014). Indeed, there are 2 FCoV serotypes and type II is hypothesized to have originated from a recombination event between FCoV type I and CCoV (Licitra et al., 2014). Notably, clinical disease caused by CCoV was recently reported in humans, as the virus was detected in the nasopharyngeal swabs of children with pneumonia hospitalized in Malaysia, in 2017−18. The complete genome sequence analysis demonstrated that the isolated strain, CCoV-human pneumonia (HuPn)-2018 (CCoV-HuPn-2018), was a novel canine-feline-like recombinant virus with a unique nucleoprotein sequence (Vlasova et al., 2022). The prevalence of CCoV-HuPn-2018 and its significance as a human pathogen will need further investigation to assess the human-to-human transmission

potential of this newly emerged strain. Nonetheless, together with the blatant cases of the SARS, MERS, and COVID-19 epidemics, it testifies the remarkable propensity of coronaviruses for host shifts and their ability to rapidly adapt to diverse species.

Whereas the circumstances of SARS-CoV-2 emergence are still unclear, as detailed below, the events associated with the spillovers of SARS-CoV and MERS-CoV have been investigated in great detail. It is now widely accepted that a virus related to SARS-CoV was transmitted from bats to palm civets, which served as an intermediate host for human infection (Cui et al., 2019; Licitra et al., 2014). Close interaction between humans and civets sold in wildlife markets probably created the ecological opportunities for transmission to humans. Passage in civets is also thought to have played a role in increasing affinity for the human receptor (Song et al., 2005). Likewise, dromedary camels acted as an intermediate host for the transmission of MERS-CoV to humans. Whereas MERS-CoV has an ultimate origin in bats, camels are common hosts for the virus in the Middle East, Africa, and Asia. Fortunately, MERS-CoV has a limited ability of human-to-human transmission, resulting in relatively small outbreaks initiated by exposure to camels (Cui et al., 2019).

Livestock have also been intermediate hosts in the emergence of two human endemic coronaviruses, namely HCoV-OC43 and HCoV-229E. The former is closely related to BCoV, suggesting an origin from cattle, whereas the closest relative of HCoV-22E was found, again, in camelids (Cui et al., 2019). Thus, all human CoVs have a zoonotic origin, but the host-switching ability of these viruses also holds in reservoir hosts and in other animals. In this respect, it is worth noting that spillover events can also occur among hosts that are not closely related. One notable example is porcine deltacoronavirus, which is thought to have emerged from the cross-species transmission from birds to pigs (Wang et al., 2019). Importantly, porcine deltacoronavirus strains were recently detected in the plasma samples of three febrile children. Genome analyses suggested independent zoonoses of distinct viral lineages with changes in the spike S1 subunit possibly increasing affinity for the human receptor (Wang et al., 2019).

The fact that at least some CoVs, including SARS-CoV-2, tend to have a wide host range is also testified by the several reverse zoonoses that occurred during the COVID-19 pandemic. The first confirmed case in an animal occurred in February 2020, when a dog in Hong Kong tested positive for the virus shortly after its owner was diagnosed with COVID-19. Since then, cases of infections with SARS-CoV-2 were reported in pets (cats and hamsters in addition to dogs), zoo animals (e.g., gorillas, lions, tigers, mountain lions, and otters), free-ranging wild animals (e.g., deers, leopards), and farmed animals (minks) (Delahay et al., 2021; Nerpel et al., 2022). In particular, the largest animal outbreaks were reported in mink farms. Human transmission to minks was first reported in the Netherlands in April 2020, and subsequently in Denmark in June 2020 (Larsen et al., 2021; Oreshkova et al., 2020). Other countries subsequently reported infection in farmed minks. The rapid spread of the virus in these animals raised concerns, as mutations associated to mink adaptation were described (Ren et al., 2021). Notably, instances of SARS-CoV-2 back-transmissions from minks to humans were also reported (Oude Munnink et al., 2021; Ren et al., 2021; Oude Munnink et al., 2021).

3. Origin of SARS-CoV-2

3.1 Phylogenetic Relationships of SARS-CoV-2 to Other CoVs

Since the beginning of the pandemic, the search for the ancestor of SARS-CoV-2 has been actively pursued. As a result, retrospective and prospective metagenomic sequencing efforts have uncovered a

number of viruses showing close or distant relationships to SARS-CoV-2. All of them have been detected in China or other locations in Southeast Asia (Hu et al., 2018; Delaune et al., 2021; Zhou et al., 2020, 2021; Li et al., 2021; Temmam et al., 2022; Wacharapluesadee et al., 2021). Analysis of these viral sequences indicated that no known virus can be considered the direct ancestor of SARS-CoV-2 and that recombination is widespread in sarbecoviruses. As a consequence, sarbecovirus genomes have a mosaic structure (Lytras et al., 2022; Li et al., 2021; Temmam et al., 2022; Wacharapluesadee et al., 2021; Boni et al., 2020; MacLean et al., 2021). This poses challenges in the reconstruction of phylogenetic relationships, as recombination leads to the violation of the assumption that a single evolutionary history can be reconstructed. Indeed, early after SARS-CoV-2 emergence, Zhou and co-workers (Zhou et al., 2020) published the sequence of a *Rhinolophus affinis* virus (RaTG13) showing >96% identity with SARS-CoV-2. RaTG13 was sampled in 2013 in Yunnan, China. Despite its overall high sequence similarity to SARS-CoV-2, the RBDs of the two viruses display remarkable divergence, suggestive of recombination (Fig. 32.2). The release of the RaTG13 sequence was followed by the identification of two sarbecovirus lineages in Malayan pangolins (*Manis javanica*) (Lam et al., 2020; Xiao et al., 2020). The two lineages reflect the geographic origin of the samples from Chinese provinces (Guangdong and Guangxi). Despite the overall lower similarity across the genome compared to RaTG13, the pangolin viruses from Guangdong possess identical amino acids as SARS-CoV-2 at six critical RBD residues (Lam et al., 2020; Xiao et al., 2020) (Fig. 32.2). Subsequently, other viruses related to SARS-CoV-2 were identified. These include RmYN02, sampled from *Rhinolophus malayanus* in 2019 in China (Yunnan) (Zhou et al., 2020). RmYN02 is closely related to SARS-CoV-2 in the long ORF1ab gene, whereas it shows low sequence identity (61.3%) in the RBD (Fig. 32.2). Notably, though, this virus resembles SARS-CoV-2 in its

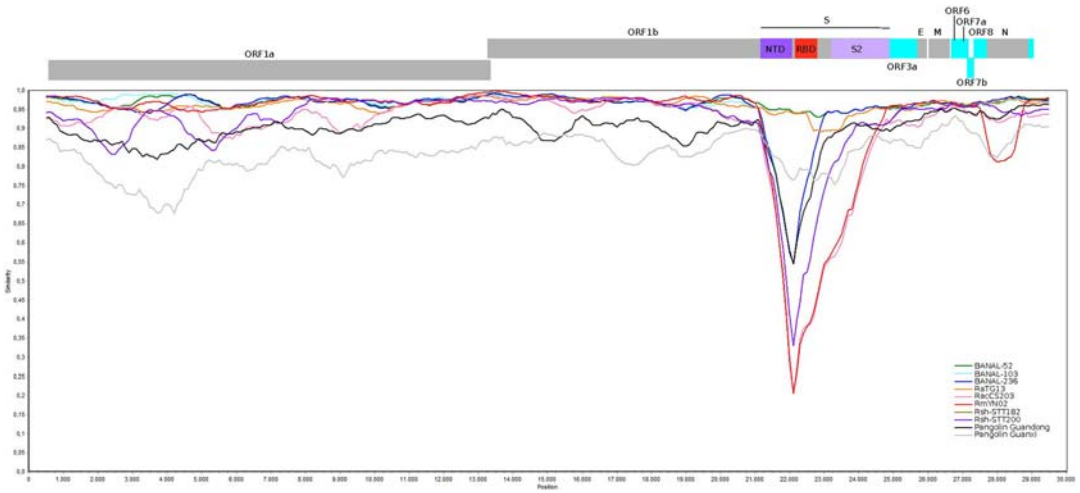


FIGURE 32.2 Similarity plot of representative betacoronaviruses to the SARS-CoV-2 reference genome

Similarity plot (generated with SimPlot version 3.5.1) (Lole et al., 1999) of representative betacoronaviruses to the SARS-CoV-2 reference genome (Wuhan-Hu-1). Each curve represents the comparison between a genome and the reference. Similarity (Kimura distance) was calculated within sliding windows of 1000 bp moving with a step of 100 bp. A schematic representation of the SARS-CoV-2 genome is also shown.

possessing an insertion of multiple amino acids at the junction site of the S1 and S2 subunits of the spike protein. Such an insertion represents a poly-basic (furin) cleavage site and it is absent in other sarbecoviruses (Zhou et al., 2020).

Viruses more distantly related to SARS-CoV-2 were also detected outside China. Thus, two betacoronaviruses (STT182 and STT200) were described in *R. shameli* bats sampled from Cambodia in 2010 and share 92.6% nucleotide identity with SARS-CoV-2 (as well as five of the six critical RBD sites) (Delaune et al., 2021) (Fig. 32.2). In addition, a bat (*R. acuminatus*) coronavirus isolated in Thailand (RacCS203) in June 2020 was found to be closely related to RmYN02 (Wacharapluesadee et al., 2021).

Finally, surveys in Laos recently identified viruses closely related to SARS-CoV-2 (*R. malayanus* BANAL-52, *R. pusillus* BANAL-103, and *R. marshalli* BANAL-236) (Fig. 32.2). The RBDs of these viruses differ from that of SARS-CoV-2 by only one or two residues. Notably, their S proteins can bind the human cellular receptor of SARS-CoV-2 (angiotensin-converting enzyme 2, ACE2) and BANAL-236 can infect human cells (Li et al., 2021; Temmam et al., 2022; Wacharapluesadee et al., 2021). These viruses are among the closest relatives of SARS-CoV-2 identified to date.

3.2 Initial Emergence and Spread

On December 31, 2019, the Chinese government notified the World Health Organization (WHO) of an outbreak of severe pneumonia of unknown etiology in Wuhan, Hubei province. Initial cases were diagnosed independently in several Wuhan hospitals between 18 and December 29, 2019. Deep sequencing from lower respiratory tract samples indicated infection with a novel betacoronavirus, which was initially named 2019 novel coronavirus (2019-nCoV) (Wu et al., 2020; Zhu et al., 2019). The first genomic sequence of the virus was released on January 10, 2020 (https://virological.org/t/novel-2019-coronavirus-genome/319). Of the initial 41 people hospitalized with unknown pneumonia, 27 had direct epidemiological link to the Huanan wholesale seafood market (Huang et al., 2020), which was closed and sanitized on January 1, 2020, after the collection of environmental samples (Huang et al., 2020; World Health Organization, 2021). A subsequent systematic review of all cases notified to China's National Notifiable Disease Reporting System by hospitals in Wuhan showed that 55 of 168 of the earliest known COVID-19 cases were associated with this market (World Health Organization, 2021) (see below).

The Chinese authorities imposed a lock-down in Wuhan and other municipalities on January 23, 2020. But the virus was already on its way. The first case of the disease outside China was reported in Thailand on January 13, 2020 (https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON234). On January 30, when the number of confirmed cases reached 98 and no deaths were reported outside of China, the WHO declared the outbreak of coronavirus as a public health emergency of international concern. The pandemic status was declared on March 11, 2020, with over 118,000 cases in 110 countries (https://apps.who.int/iris/handle/10665/331475).

Early in the epidemic, the WHO decided that, following best practices for naming infectious diseases (https://www.who.int/publications/i/item/WHO-HSE-FOS-15.1), the severe pneumonia caused by the coronavirus should be named COVID-19 (coronavirus disease 2019) (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200211-sitrep-22-ncov.pdf?sfvrsn=fb6d49b1_2). Later on, the International Committee on Taxonomy of Viruses (ICTV) indicated that the novel coronavirus should be named SARS-CoV-2 and should be classified within the species *Severe acute respiratory syndrome-related coronavirus* (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020).

3.3 Hypotheses About the Origin of SARS-CoV-2

The origin of SARS-CoV-2 has been the topic of unprecedented scientific debate and public speculation. The still missing information about a potential intermediate host and the presence in Wuhan of a Virology Laboratory (Wuhan Institute of Virology (WIV)) with active projects on coronaviruses have fueled debate about the possibility that the SARS-CoV-2 spillover might have been caused by a research accident (lab leak hypothesis) (Dyer, 2021; Harrison and Sachs, 2022). In May 2020, the WHO decided that an international and multidisciplinary team of experts should join Chinese experts and authorities with the aim to identify the zoonotic source of the virus and the route of introduction to the human population. The resulting report "WHO-convened Global Study of Origins of SARS-CoV-2: China Part, Joint WHO-China Study, 14 January−10 February 2021" indicated the direct introduction from a bat or intermediate host as likely pathways and the lab leak hypothesis as unlikely (Huang et al., 2020; World Health Organization, 2021). Besides these conclusions, which have been criticized and debated (Bloom et al., 2021; Harrison and Sachs, 2022; Dyer, 2021), the report provided a wealth of epidemiological data that were used in independent works to gain insight into the early events associated with SARS-CoV-2 spread. For instance, through analysis of cases reported to the National Notifiable Disease Reporting System (NNDRS) in China, the WHO mission identified 174 COVID-19 cases in Hubei province in December 2019. These early cases are most likely free from ascertainment bias because they were based on identification of clinical symptoms before the Huanan wholesale seafood market was suggested as a risk factor (Huang et al., 2020; World Health Organization, 2021). Using the geographic coordinates of 155 such cases, Worobey and coworkers showed that the majority clustered in central Wuhan, near the Huanan market (Worobey et al., 2022). Such distribution of cases could not be expected based on Wuhan population density or demographic patterns. Notably, early cases with no apparent epidemiological link to the market were also found to reside significantly closer to it than patients with a direct link to the market. This is very important as the presence of un-linked cases was regarded as an indication against a major role of the Huanan market in the initial spread of the virus. The authors thus concluded that the Huanan wholesale seafood market was the early epicenter of the pandemic (Worobey et al., 2022). Indeed, data from the WHO report and other sources, including business registries from the Huanan market, also revealed that SARS-CoV-2-positive environmental samples were mainly clustered in areas where live wild animals were on sale (Worobey et al., 2022; Huang et al., 2020; World Health Organization, 2021).

The Huanan seafood wholesale market was one of the four markets in Wuhan known to sell wild or farmed live animals (Worobey et al., 2022). The fact that early links to the COVID-19 pandemic lead to it are strongly supportive of a scenario whereby, in analogy to SARS-CoV, the SARS-CoV-2 spillover was initiated by animals for sale. As noted by Worobey and coworkers, in a city with 11 million inhabitants, many other locations (e.g., shopping malls, hospitals, places of worship) might in principle be considered ideal venues to spread a respiratory pathogen. The fact that the market was the epicenter of the pandemic thus suggests that animals were the source of infection as opposed to the possibility that the crowded vendors' stalls functioned as superspreader sites (Worobey et al., 2022).

In line with these data is the analysis of early viral lineages. Two lineages, denoted A and B, were detected since the beginning of the epidemic. Lineage A differs from lineage B by two nucleotide substitutions (C8782T and T28144C), and it is more closely related to bat coronaviruses hosted by *Rhinolophus* spp (Pekar et al., 2022). Lineage B has been the most common both in the initial samples from the market and throughout the pandemic. It includes early genomes sequenced from patients

associated with the market, as well as the reference genome (Wuhan/Hu-1/2019). The earliest lineage A genomes from humans lack a direct epidemiological connection to the Huanan market, but were sampled from individuals who lived or had recently stayed close to the market (Worobey, 2021). Through reconstruction of the SARS-CoV-2 ancestral haplotype, Pekar and coworkers showed that the most likely scenario envisages that the A and B lineages were independently transmitted to humans in separate zoonotic events. The first zoonotic transmission possibly involved lineage B and occurred around the mid of November 2019. Lineage A was transmitted a week later (Pekar et al., 2022). Again, the fact that both lineages were associated with the market is not consistent with the idea that the market was simply a superspreading site, as in this case a lineage-specific superspreading event would be expected (Worobey et al., 2022). Indeed, the most likely hypothesis is that both lineages diverged and were circulating in nonhuman mammals prior to their spillover to humans (Pekar et al., 2022).

Unfortunately, little information is available concerning the animals that were sold in the market and their provenance. The market was supplied with farmed wildlife from within and outside the Hubei Province. Between 1st January and March 2, 2020, more than 450 animal samples were collected. These included goods from the market or warehouses and refrigerators related to it, as well as stray animals in the market. All of them tested negative for SARS-CoV-2 genetic material (Huang et al., 2020; World Health Organization, 2021). It was however noticed that the sampled species were not representative of those sold in the market, and that raccoon dogs, considered as a likely intermediate host due to their susceptibility to SARS-CoV-2, were not included (Holmes et al., 2021). Testing was also extended to animal farms in Hubei Province, and surveillance programs in wild animals were launched. However, almost 2500 samples turned out to be negative for SARS-CoV-2 nucleic acids. Likewise, retrospective antibody testing of livestock and poultry samples collected across China revealed no evidence of exposure to SARS-CoV-2 (Huang et al., 2020; World Health Organization, 2021). Thus, the animal origin of SARS-CoV-2 remains cloaked in mystery. Other possible sources of infection, such as cold chain products, which of course were present in the Huanan wholesale seafood market and that were supplied from diverse regions, cannot be discarded. However, they are presently considered less likely than the live/freshly butchered animal hypothesis (Huang et al., 2020; World Health Organization, 2021).

Concerning the lab-leak scenario, one of the strongest arguments in its support is the presence, in the spike protein of SARS-CoV-2, of a furin cleavage site that is absent in most sarbecoviruses sampled to date and instead closely resembles an eight−amino-acid sequence of the human ENaC alpha subunit (Anand et al., 2020; Harrison and Sachs, 2022). However, the occurrence of this sequence might well be a coincidence and derive from natural selection rather than genetic manipulation (Holmes et al., 2021; Anand et al., 2020; Harrison and Sachs, 2022). In fact, many mammalian proteins possess furin cleavage sites and several coronavirus (although not most sarbecovirus) S proteins also do (Stout et al., 2021). Thus, at present, the strongest evidence is in favor of a natural, possibly wildlife origin of SARS-CoV-2 and a zoonotic transmission associated with animal trade.

4. Genetic Diversity of SARS-CoV-2

4.1 Mutation and Substitution Rates

The mutation rate of every organism is essential to understand the processes involved in determining its ecology and evolution. A mutation is any change in the sequence of the genetic material that can be

transmitted to the next generation. For most biological processes, with the notable exceptions of cancer cells and tissues and some genes in immune cells that code for antigens recognition and immunological memory, the only relevant mutations are those transmitted to the offspring. Cellular organisms, which include archaea, bacteria, and eukaryotes, have a genome based on DNA and the main molecular processes originating mutations occur during the replication of this essential molecule. Organisms in these three domains share DNA polymerases with very similar characteristics, among which an error-correction mechanism is very relevant. As a consequence, they have very similar mutation rates for single base substitutions during the replication of their genetic material, at around 10^{-9}–10^{-10} errors/site/replication. Nevertheless, apart from point errors during replication, this process is prone to other types of mutations that affect several, even many, nucleotides at a time, such as expansion/contraction of short- or large-tandem repeats, duplications, inversions, and so on.

RNA viruses, as indicated previously, have much simpler genomes than those of cellular organisms. Most of them encode their own enzyme to replicate their genomes because as these are based on RNA they cannot depend on a usually nonexisting enzyme (RNA-dependent RNA-polymerase, RdRp) in the cells that they infect. A critical difference between RdRps and DNA polymerases is the general lack of those error-correction domains that lower the mutation rate of cellular organisms. As a consequence, the mutation rate for point substitutions of RNA viruses is orders of magnitude higher than that of cellular organisms, in the order of 10^{-3}–10^{-5} e/s/r. This difference has profound consequences in many biological features of RNA viruses, two of which have already been commented on: a limited genome size, usually below 11 kb, and a fast evolutionary rate.

At this point, it is convenient to remark on the differences between similar concepts. **Mutation rates**, as defined, correspond to the frequency or probability in which any nucleotide of the viral genome is changed during the replication process. **Mutation frequency** is the observed relative frequency of any particular mutation in a population, that is, the ratio between the number of genomes carrying that specific mutation and the total number of genomes in the population. These two parameters are almost always different because in any population there will act different processes, mainly selection and genetic drift, that alter the observed frequency of the mutations produced during replication of its individuals. For most mutations in viral genomes, we never observe any representative in a population, as purifying selection and/or drift will generally purge most of them. Occasionally, some of the mutations may provide a slight, rarely substantial, advantage to their carriers and their frequency in the population will rise, temporarily or permanently. In the latter case, the pace at which a new mutation replaces a previous variant in the population corresponds to the **substitution rate**.

The mutation rate of SARS-CoV-2 is estimated to be similar to that of other CoVs (Jungreis et al., 2021). In addition, CoVs encode an RNA proofreading exoribonuclease (nsp14-ExoN) that is different from the polymerase and responsible for high-fidelity RNA synthesis, resistance to nucleoside analogs, immune evasion, and virulence (Gribble et al., 2021). The reduction in mutation rate at the genome level allows a more efficient and reliable replication of the hereditary message in these viruses, which is also reflected in their larger genome sizes (around 30 kb). For many, during the early days of the COVID-19 pandemic, the lower mutation rate of SARS-CoV-2 was considered to be good news, as the new virus would not be able to evolve at such a fast pace as other infamous RNA viruses (HIV, HCV, influenza A, to name just a few). However, evolution does not depend only on mutation rates and many other factors have contributed to a rapid, and clinically and epidemiologically relevant, fast diversification of SARS-CoV-2 in the human population, as we will consider later.

The initial estimates of the rate of substitution for SARS-CoV-2 were around 24−26 substitutions/genome/year. That is, if we compare the genome sequence of any virus and that of its ancestor 1 year ago, we expect to find about 24−26 differences between them. These estimates were based on comparisons of sequences separated at much shorter intervals (a few weeks or months) and, consequently, were prone to errors. More recent studies (Neher, 2022) have reduced this estimate to changes occurring **within** lineages although the overall rate across lineages is similar. There are interesting differences between the types of mutations responsible for these evolutionary rates: while the rate of synonymous mutations is the same within and among lineages, approximately six substitutions per year, those of nonsynonymous mutations differ substantially. Within lineages, the initial estimates of 12−16 substitutions per year have now reduced to 6−9 substitutions. These differences have likely arisen as a result of the different evolutionary pressures exerted on the virus in the initial stages of its spread, with less selection acting on the virus, than at present, in which different lineages attain distinct evolutionary peaks, under adaptive pressure, but this is less intense once they have gained one of those peaks.

4.2 Recombination

Recombination occurs frequently in many virus families, especially those with positive single stranded RNA as genetic material, and coronaviruses are no exception (Müller et al., 2022). Recombination has been known to contribute significantly to the diversification and adaptation of coronaviruses. At the molecular level, the same enzyme responsible for the reduced mutation rate of CoVs, the exoribonuclease nsp14-ExoN, is also responsible for the increased rate of recombination in these viruses (Gribble et al., 2021).

Recombination events involving the spike protein and some of its most relevant domains, such as the receptor binding domain (RBD), have been characterized before the emergence of SARS-CoV-2. Furthermore, it has been postulated that recombination in this gene has played a major role in shifting and emergence in new hosts (Graham and Baric, 2010). This was the case in the emergence of SARS-CoV using civets as intermediary hosts. The spike RBD of the viruses infecting civets and those producing early infections in humans presented signs of adaptive evolution, mediated by positive selection, in residues that facilitated viral recognition of the ACE2 receptors (Kan et al., 2005; Li et al., 2005; Chinese SARS Molecular Epidemiology Consortium, 2004).

Early analyses of the SARS-CoV-2 genome already revealed that recombination had intervened in the evolution of this lineage (Boni et al., 2020; Li et al., 2020). These authors identified three regions in the genomes of 68 sarbecoviruses included in their study that were free from recombination (BFRs, breakpoint free regions) and, thus, were appropriate for dating evolutionary events in their phylogenetic history (Fig. 32.3). These regions spanned a total of 13,081 nt, less than 44% of the total SARS-CoV-2 genome length, and were all included in the large *orf1ab* gene. Of note, although a possible involvement of recombination in critical sites and regions of the spike gene, such as the already mentioned RBD and RBM and the furin cleavage site S1/S2, also critical for the conformational changes allowing the fusion of viral and cellular membranes and subsequent infection, has been suggested (Patiño-Galindo et al., 2021), no clear evidence including the putative parental donors has been found yet, thus adding some uncertainty in the explanation for the immediate origin of SARS-CoV-2 and its adaptation to human hosts.

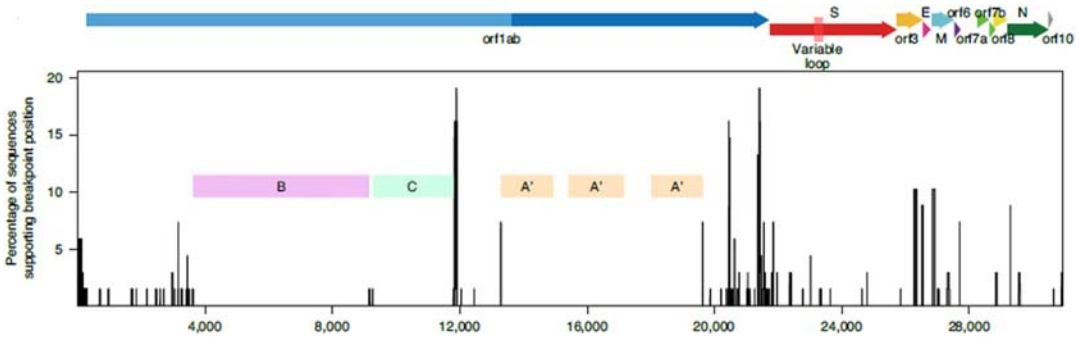


FIGURE 32.3 Recombination-free regions in sarbecovirus genomes

Regions free from recombination (A, B, C, marked in orange, pink, and green, respectively) found in an analysis of 68 sarbecovirus genomes. Region A is further subdivided into three subregions (labeled A′). (For interpretation of the references to color in this figure legend, please refer online version of this title.)

From Boni, M.F., Lemey, P., Jiang, X., Lam, T.T.Y., Perry, B.W., Castoe, T.A., Rambaut, A., Robertson, D.L., 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat. Microbiol. 5 (11), 1408–1417. https://doi.org/10.1038/s41564-020-0771-4

Recombination has continued after the emergence of SARS-CoV-2 as a new human pathogen (Gutierrez et al., 2022; Turakhia et al., 2022). The necessary conditions for recombination to occur are quite easily fulfilled during infection of this virus (Pérez-Losada et al., 2015) but the detection of such events requires the combination of quite distinct lineages in the recombinant and nonrecombinant portions of the new genome. This is much less frequent given the low diversity of SARS-CoV-2 genomes, especially at the within lineage level. In some cases, homoplasic mutations and sequencing errors may mimic the patterns expected under recombination. Consequently, only those recombinant genomes encompassing large fragments of sufficiently divergent lineages will be detected with statistical confidence as such. This will lead to underestimating the extent of recombination although only those events that would likely represent a dramatic change in the genomic sequence from that of the corresponding parentals. This implies that most recombination events in SARS-CoV-2 detected during the pandemic correspond to interlineage recombinants (Gutierrez et al., 2022) and, more specifically, to intervariant exchanges (Wertheim et al., 2022). In this particular study, recombination between variants alpha and epsilon was detected in an individual who was simultaneously coinfected with the two viral strains. Retrospective analyses of sequencing reads revealed that super- and co-infection of the same person with different strains or lineages during periods of epidemiological coexistence, leading to increased intrahost diversity, is not an usual phenomenon in the case of SARS-CoV-2 (Dezordi et al., 2022). A consequence of the dynamics of diversity is that the detection of recombinant sequences will experience fluctuations through time, with few detections in periods of low genetic diversity (VanInsberghe et al., 2021) and increased detection as more diversity builds up.

As of this writing (February 2023), there are 71 recombinant lineages included among the Pango designated lineages (https://github.com/cov-lineages/pango-designation/blob/master/lineage_notes.txt). However, additional putative recombinants have appeared without reaching the status of recognized lineage (an updated list is maintained at https://docs.google.com/spreadsheets/d/1cQILRxXD756gJoRsaqMdJkxZm7sEjhV7ceY398Iz7gI/). Pango denotes recombinant lineages by

starting the assigned name with an "X" (see Section 4.3). Only lineage XBB.1 has reached more than 10,000 sequences deposited in GISAID so far, making it the most successful recombinant lineage up to now (Fig. 32.4).

Some of the first recombinant genomes were detected in the United Kingdom shortly after the emergence of lineage B.1.1.7 (VOC Alpha). In the first 2 months after the identification of this new variant, 16 recombinant genomes with different previously circulating lineages were found in an analysis of over 279,000 sequences (Jackson et al., 2021). The most common second parental strains corresponded to variant B.1.177, the previously dominant variant in the United Kingdom until B.1.1.7 started to spread.

A requisite for recombination to occur is that a person is simultaneously infected with two different viral strains (Wertheim et al., 2022). The chances for coinfection increase with the time of co-circulation of the two variants and with the absolute number of infections produced during that time. The larger the number of cases the higher the chances that a person is coinfected with two strains. This is a prerequisite for recombination that was reflected in the appearance of several Delta-Omicron recombinant lineages (Bolze et al., 2022; Colson et al., 2022; Combes et al., 2022; Lacek et al., 2022). Despite the alarm derived from the detection of some of these "Deltacrom" variants, no major changes in the transmissibility, virulence, or response to treatment were observed and a few weeks after their initial detection these variants disappeared from the radar of genomic surveillance.

The recent diversification of SARS-CoV-2 into several sublineages within the Omicron (B.1.1.529) VOC has provided new opportunities for recombination. Previous VOCs were replaced by new dominant lineages too early for them to diversify widely, and detectable, within-lineage recombinants were rare (Dezordi et al., 2022) and did not attain even relatively high frequencies. However, most of the currently identified recombinants have originated from recombination events involving different

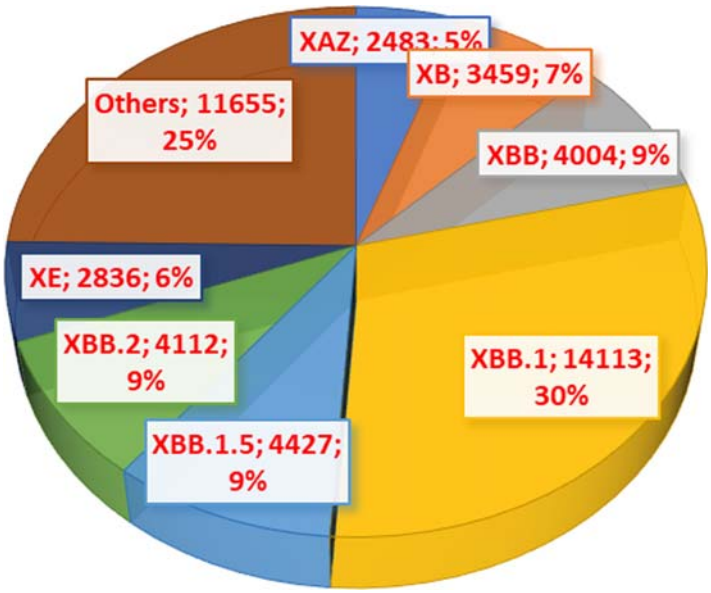


FIGURE 32.4 Frequency distribution of SARS-CoV-2 recombinant lineages

Only lineages with frequency >5% are indicated, along with the total number of sequences and their proportion among the 44,253 recombinant sequences deposited in GISAID (January 31, 2023).

sublineages of Omicron (Gu et al., 2022; Lindh et al., 2022; Sekizuka et al., 2022). The genetic diversity generated within Omicron is similar to that generated among previous lineages of SARS-CoV-2 (Fig. 32.5). As a consequence, occasional recombination events involving genetically distinct lineages, especially BA.1, BA.2, or BA.5, will result in distinguishable recombinant genomes. For instance, XE is the result of recombination between BA.1 and BA.2: the former lineage contributed NSP1-6 and the rest of the genome contains BA.2-derived information (SARS-CoV-2 Variants of Concern as of August 25, 2022 by ECDC).

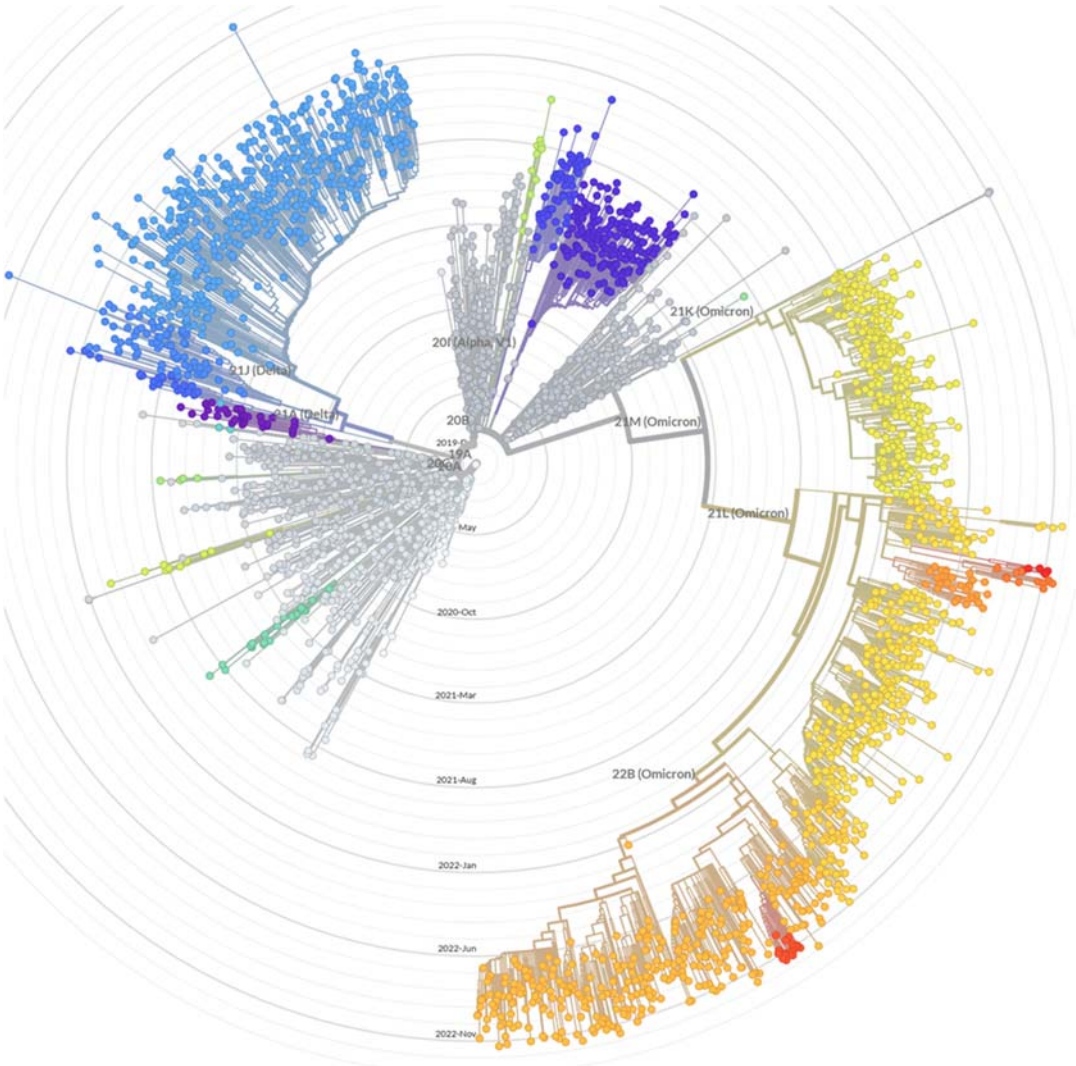


FIGURE 32.5

Phylogenetic tree of SARS-CoV-2 showing representative sequences of the main variants and lineages.

From https://nextstrain.org/ncov/gisaid/global/all-time (Analysis and images reprinted from Nextstrain.org [nextstrain.org/ncov/gisaid/global] under an Attribution 4.0 International (CC-BY-4.0) license.)

Until now, recombination within or among SARS-CoV-2 lineages has not represented a major factor in the generation of variation nor in the phenotypic features of the dominant variants, although two Omicron lineages, BA.2.10.1 and BA.2.75, have recombined resulting in XBB.1, the parental lineage of the most successful recombinant lineages until now, including XBB.1.5 (Fig. 32.5). This recombinant lineage has increased transmissibility, infectivity, and binding ability to ACE2 compared to its ancestor (Yue et al., 2023; Uriu et al., 2023), probably derived from the acquisition of mutation S: F486P. However, this is not necessarily the outcome of future recombinants. Recombination events can bring together genome fragments that have evolved to adapt to different conditions such as increased transmission or reduced neutralization by antibodies (Wang et al., 2022), as suspected for lineage XBB.1.5, currently spreading in the United States with increased transmission rate than its predecessors. This might accelerate the adaptive evolution of SARS-CoV-2 to face increased immune pressure from previous infections, vaccination or treatment with monoclonal antibodies, risking the efficiency of therapeutic treatments based on single or dual combinations of antibodies (Arora et al., 2023).

4.3 Nomenclature of SARS-CoV-2 Lineages

The ICTV (International Committee on the Taxonomy of Viruses) classifies viruses at the species level and higher taxonomic ranks, such as genera, orders, and classes (Lefkowitz et al., 2018), but there is no universal system for designing viral groups below the species level. This has resulted in a variety of nomenclature systems that vary from one species to another. The main criteria used by virologists to denote subspecific clusters are the viral immunogenic features, such as serotypes in Dengue or Bluetongue viruses, combinations of variants in specific proteins, such as hemagglutinin and neuraminidase in influenza A virus, or stable genetic differences, such as genotypes in hepatitis C virus.

Soon after the discovery of SARS-CoV-2, two divergent lineages, denoted as A and B, were observed (Pekar et al., 2022). Subsequent evolution and divergence within and between these two original lineages led to different naming systems depending on where the information was stored and/or analyzed. Complete genome sequences started to be deposited and shared at GISAID (www.gisaid.org), with a naming system based on marker variants (amino acid substitutions) with respect to the reference genome (WIV04, GenBank accession MN996528). This scheme was very useful in the first stages of the pandemic, but the huge number of sequences deposited made it inconsistent soon after. Homoplasies, resulting from convergent and backward mutations, introduce inconsistencies in the naming of clades based on a single or a few amino acid sites, leading to polyphyletic groups that do not reflect the genealogical/evolutionary relationships of the corresponding isolates.

A popular site for the analysis of SARS-CoV-2 sequences, nextstrain.org, introduced a different naming system, based on the phylogenetic relationships among them. In this platform, a consistent pipeline of phylogenetic and phylodynamic analysis (Hadfield et al., 2018) is applied to specific subsets of sequences. This system assigns names on the basis of phylogenetic groupings, using the number of the year of isolation of the first representative of a clade and letters to differentiate the different clades of the same year. For instance, the first two lineages of SARS-CoV-2 were denoted as 19A and 19B, respectively. Occasionally, some lineages receive additional identifiers, such as 20E.EU1 or 20I/501Y.V1, but with no fixed rule. Apart from this point, the problem with this system is the lack of clear criteria to define and identify the well-established lineages that get a particular name.

These problems were addressed by a new naming system for viral clusters below the species level proposed by Rambaut et al. (2020) and known as PANGO (Phylogenetic Assignment of Named Global Outbreak Lineages). The main goal of this system is to provide a consistent nomenclature for epidemiologically relevant lineages. The scheme is dynamic and puts a limit on the number and depth of hierarchical levels. The names start with a letter (A and B, for the earliest lineages) and up to four hierarchical levels, each defined as descendant from a preceding level given four conditions: (a) one or more shared nucleotide differences from the ancestor lineage; (b) comprising at least five genomes with >95% of the genome sequenced; (c) genomes within a lineage exhibit at least one shared nucleotide change among them; and (d) a bootstrap support >70% for the lineage-defining node. The resulting names (e.g., A, B.1, B.2.5, etc.) are informative, consistent, and coherent and facilitate the use of neutral tags, with no negative connotations. The list of PANGO lineages can be accessed at https://github.com/cov-lineages/pango-designation. In fact, this system has already been applied during the recent pandemic expansion of Mpox virus to prevent using toponymic clade names. Active lineages are regularly updated and made accessible by tools such as Pangolin (https://github.com/hCoV-2019/pangolin).

Unfortunately, there is no equivalence rule among the three naming systems (Alm et al., 2020; González-Candelas et al., 2021) that would help "translating" one clade/lineage name in one system to the others. The different names that a particular genome can receive are provided in the Nextstrain builds (https://nextstrain.org/ncov). Ideally, it would be desirable to have a "rule of equivalences" that would "translate" the name of a clade/lineage in a nomenclature system to the corresponding name in any other, but this is not possible. The PANGO system has been updated to accommodate the recombinant lineages (Pybus, 2021). The general rules for designation of PANGO lineages still apply, but new recombinant lineages of the highest level are preceded by an X, so we will have XA.1, XB.1.1, etc. These names do not contain information about the putative parental lineages.

These nomenclature systems became too technical to be followed by the media and public health authorities around the world when different lineages were designated as variants of concern (VOCs) or variants of interest (VOIs) since early 2021. Following similar patterns as every other SARS-CoV-2 lineage, the initial VOCs and VOIs started to diversify into new lineages which, following the PANGO naming rules, were identified with different names. In most cases, the new lineages inherited the main features of their ancestors and, consequently, were also considered VOCs or VOIs. To facilitate communication and easier identification of these variants, in May 2021, the WHO adopted a different, much simpler naming system to be applied to variants of concern or under investigation (Konings et al., 2021). This system follows the Greek alphabet, albeit with some exceptions, and assigns a new letter to each VOC or VOI declared by WHO. The most prevalent variants since the adoption of this system correspond to VOCs Alpha (PANGO lineage B.1.1.7 and its derivatives), Delta (B.1.617), and Omicron (B.1.1.529). To be considered as VOC or VOI, a variant must meet several requisites relative to its transmissibility, virulence, or response to public health or social measures and be distinct from other variants (https://www.who.int/activities/tracking-SARS-CoV-2-variants). The many subvariants of Omicron are all classified as VOCs and none has been denoted with a different Greek letter so far. WHO experts do not consider that the observed differences among them warrant the assignment of a new VOC.

5. Evolution of SARS-CoV-2

5.1 VOC/VOI Emergence

Since the end of 2020, several variants have emerged and spread considerably fast, rapidly displacing previous variants present in the populations of their origin and ultimately the whole world. The rapid displacement can be due to several causes, including epidemiological factors, as was the case of a previous variant (20E-EU1 or B.1.177) that became dominant in many European countries in the summer-fall of 2020 (Hodcroft et al., 2021), or because of intrinsic viral characteristics, such as increased transmissibility, evasion of prior immunity in the population, or combinations of these and other factors.

The first such variant to call the attention of public health and virus evolution specialists appeared in South East England by the end of November 2020 (Public Health England, 2020), although the first sequence deposited in GISAID was dated in September of the same year. Its emergence was coincident with a dramatic surge of infections in the area, later displaced to neighboring counties, and ultimately to the whole country and many other countries. The variant presented an unprecedented accumulation of seven amino acid substitutions at the spike protein compared to its closest phylogenetic relatives (Hill et al., 2022), although the complete genome presented 23 mutations, including 3 deletions and 6 synonymous substitutions. One of the nonsynonymous mutations in the RBD, N501Y, affected ACE2 binding, whereas a two amino acid deletion in the same protein, positions H69−V70, interfered with some diagnostic tests and probably also contributed to the higher transmissibility of this variant (Public Health England, 2020).

The emergence of Delta (lineage B.1.617.2) in India in the early spring of 2021 represented a dramatic change from the previous situation, in which Alpha and Kappa were the dominating variants (McCrone et al., 2022; Dhar et al., 2021). Delta was more transmissible than Alpha, with estimates ranging between 63% and 167% in different US states (Earnest et al., 2022), and also produced more viral copies. These features, along with enhanced immune evasion, host receptor avidity, reinfection, and vaccination breakthrough (Dhar et al., 2021) led to a rapid replacement worldwide. The global spread of Delta was contemporary to vaccination campaigns, especially in high-income countries, and, as a result of its differences with the original Wuhan strain used in the different vaccine formulations, it showed a relatively lower capacity to be neutralized by vaccine-induced antibodies (Planas et al., 2021). One of the subvariants of Delta, lineage AY.4.2, appeared in the United Kingdom and spread rapidly there and in some other countries thanks to its increased transmissibility (Eales et al., 2022). This led to its designation as Delta-plus subvariant, in fear that it might lead to another wave of infections. This proved to be partially wrong as a new, much more transmissible VOC rapidly replaced Delta and Delta-plus.

Omicron (B.1.1.529) was detected at the end of November 2021 in South Africa and Botswana, associated with a rapid resurge of cases in the South African province of Guateng (Viana et al., 2022). It rapidly caught the attention of public health authorities, because of its much faster transmission in a non-naïve population, and of virologists and evolutionary biologists, because of the very large genetic distance to other circulating variants in the area at that time: it differed from its closest genetic ancestor by as many as 53 mutations, 32 of which were found in the spike gene. In a matter of days (November 26, 2021), the new lineage was denoted as VOC and designated as Omicron.

One common feature of the basal sequences defining VOCs is the many accumulated substitutions observed when compared to their most closely related sequences. Those substitutions are distributed throughout the viral genome but a relatively high number of them are found in the spike gene, with a higher proportion of nonsynonymous substitutions than synonymous ones. Several hypotheses have been advanced to explain their rapid evolution: (i) evolution within an immunocompromised patient with prolonged infection, (ii) evolution in wild animals and reentry into humans, (iii) continued evolution in a population passing undetected to the genomic surveillance systems, and (iv) accelerated evolution by increased mutation rates as a result of treatment with mutagenic drugs. We will consider these alternatives next.

The most widely accepted hypothesis for the origin of VOCs is that the initial sequences evolved in an immunocompromised person during a prolonged, uncontrolled infectious process (Harari et al., 2022; Weigang et al., 2021). So far, no direct evidence of that case directly linked to VOC has been found. However, the indirect evidence is very strong. Several studies have analyzed the evolution of SARS-CoV-2 through the course of prolonged infections (several weeks, even months) in immunocompromised patients (Chaguza et al., 2022; Weigang et al., 2021; Maponga et al., 2023). Notably, many of the mutations found in these sequences after weeks of infection are identical and/or occur in the same positions of the spike protein than those found in the initial sequences of VOCs. These mutations have been associated to escape from weakened immune responses and are also coincident with mutations found in response to treatments with monoclonal antibodies (Scherer et al., 2022). Reduced neutralization to some antibodies has been a common feature of VOCs.

SARS-CoV-2 has been found to be capable of infecting a wide range of domestic, peridomestic, and free-ranging animals (Hale et al., 2022; Li et al., 2023), apart from its presumed original host, possibly a bat species. For several of those new species, there have been natural infections linked to contacts with humans: our species has been an intermediate host between bats and these new hosts (Sharun et al., 2021; Mahdy et al., 2020) and there are also reports in which humans have been infected from these animals (Kuchipudi et al., 2022; Oude Munnink et al., 2021). The spillover events are usually accompanied by adaptive evolution to the new host, which is accomplished through changes in the spike. We have already commented about the important roles of this protein in binding to cell receptors and entry, which are different among species, and eliciting immune responses. Correspondingly, accelerated evolution in this protein has been found in some of these species, such as white-tailed deer or minks, that have been a source of infection to humans (Oude Munnink et al., 2021; Pickering et al., 2022). One of these animal-to-human transmission events was associated with a very divergent lineage (B.1.641) that included 76 mutations with respect to the ancestral Wuhan Hu-1 sequence. This divergence is similar to that found in some VOCs, although this particular case has not spread substantially. This same process has been invoked to explain the emergence of Omicron in mice (Wei et al., 2021).

Genomic surveillance has provided a great deal of information about the spread and evolution of SARS-CoV-2 since the very start of the pandemic. However, the distribution of surveillance efforts and data is far from being homogeneous, with many countries providing sequence data for only a small fraction of the infected population (Brito et al., 2022). Then, it is possible that new variants remain undetected for long periods of time if they occur in those areas with little migration and low transmission. Under these circumstances, a new lineage might evolve, gradually accumulating mutations until it will eventually spread enough to be detected. This scenario was proposed for the evolution of

Omicron in several African countries months before its detection in South Africa and Botswana in November 2021, but there is not support for its acceptance (Mallapaty, 2022).

RNA viruses rapidly adapt to environmental changes thanks to their high mutation rates. But high mutation rates also place viral populations close to the error threshold, where mutations can be so frequent that lead to population extinction through error catastrophe (Eigen, 1971; Anderson et al., 2004). This possibility has been leveraged by some antiviral treatments aimed at increasing the mutation rate and thus induce the rapid accumulation of detrimental mutations with a deleterious effect on the viability of viral populations (Eigen, 2002; Crotty et al., 2001). Treatment of COVID-19 with some of these mutagenic drugs might lead to the accumulation of mutations in relatively short periods. Some cases have already been observed, with up to 120 mutations being detected after only 46 days of treatment with molupiravir (https://twitter.com/LongDesertTrain/status/1600559000155979776). Again, there is no evidence that this mechanism has actually been responsible for the evolution of a new VOC, but this is a possibility to be considered in future scenarios.

5.2 Evolution of Transmissibility

Transmission is the key factor driving the evolution of pathogens and SARS-CoV-2 is no exception. The rapid replacement by dominant variants in the pandemic has been driven by increased transmissibility, which can be achieved by different means. Many viral traits contribute to transmissibility, but there are also important factors from the host that contribute to it, thus making transmissibility a very plastic property. At the epidemiological level, transmissibility is usually evaluated by R_x, with different values of x representing several meanings or concepts. For instance, R_0, the basic reproduction number, is used to represent intrinsic transmissibility, the average number of secondary infections produced by a person infected with the virus in a completely naïve population. R_e represents the effective reproduction number, which is dependent on the specific population where it is being measured and, consequently, also involves the particular features of each viral and host populations. R_t is usually employed to represent the changing value of R_e through time in a particular population.

The estimates of R_0 for the initial pandemic virus ranged from 1.4 up to 6.49, with a median of 3.28 and an average of 2.79 for the 14 studies in Chinese populations during the first month of the pandemic reviewed by Liu et al. (2020).

The Delta variant had a substantially higher R_e than the R_0 estimates for the original virus. Its R_e was estimated with an average value of 5.08 in a review of five studies by Liu and Rocklöv (2021). This estimate is also much higher than the corresponding reproductive number estimates for other viral infections such as SARS, MERS, smallpox, Ebola, seasonal influenza, and pandemic influenza (Biggerstaff et al., 2014; Chowell and Nishiura, 2015; Petersen et al., 2020).

Estimates of basic and effective reproductive numbers for Omicron easily explain the rapid displacement of the Delta VOC by the new variant (Liu and Rocklöv, 2022). This study compiled information on 18 reports, 8 estimating R_0 and 10 estimating R_e. The average estimate for R_0 was 3.8 times higher than that of Delta, with a range from 5.5 to 24 (median 10, mean 9.5), whereas the corresponding values for the effective reproductive number were 2.5 times higher than Delta, with average R_e of 3.4, median 2.8, and a range from 0.88 to 9.4. One study (Bartha et al., 2021) obtained an estimate for R_0 of 24 in South Africa, and it represents a theoretical ceiling assuming no immune evasion (Liu and Rocklöv, 2022).

Transmissibility is a complex trait because it compiles the result of many different processes acting at different levels (Anderson and May 1992). A higher transmissibility can be attained through changes in the affinity to cell receptors (Hong et al., 2022; Omotuyi et al., 2022), increased virion productivity, shorter incubation period (Wu et al., 2022), escape from immunity, binding to new receptors, longer infection or shedding time (Keske et al., 2023), longer persistence in the environment, etc (Anderson et al., 2021). The different variants that have shown increased transmissibility have explored some of these possibilities, although not all have been documented convincingly. For instance, the Omicron variant has acquired many mutations in the RBD and other regions of the spike protein leading to a different preferential cell entry pathway, endosomal fusion, whereas Delta and previous variants used cell surface fusion after proteolysis by TMPRSS2 (Willett et al., 2022). Apart from changes in the ability to infect different cell types, preferentially in the upper respiratory tract, Omicron had a reduced capacity to form cell syncytia, resulting not only in changes in transmissibility but also possibly in pathogenicity of this variant (Willett et al., 2022). In contrast, Delta has a shorter generation time but higher viral production than previous variants (Li et al., 2022).

5.3 Evasion of Host Immunity

As noted previously, one of the features of new SARS-CoV-2 variants causing more concern is the accumulation of mutations that facilitate escape from recognition by the immune system. There are two main areas of concern: reductions in the efficiency of vaccines and non-neutralization by monoclonal antibodies used as therapy. This interest has fueled a large number of investigations and surveillance analysis on the effects of new variants, subvariants, and mutations on immune responses and neutralization capability of antibodies, assays in a range of systems including cell cultures, in vitro neutralization assays, and animal experiments that are almost impossible to summarize here. In consequence, we will only mention the most relevant aspects of changes in the immunology of the host and the mutations in viral epitopes in relation to the evolution of the virus.

The large majority of the vaccines administered so far are based on the spike protein, which is the antigen richest protein of the virus, but not the only including regions recognized by the human immune system. Protection against SARS-CoV-2 (as well as against most other viruses) is mediated by different components of the immune system: humoral responses (antibodies), T-cell—mediated responses, and innate immunity (a network of molecules that sense the presence of the virus and trigger the production of interferons and cytokines).

The spike protein is the main target of neutralizing antibodies and, before late 2022, all COVID-19 vaccines were based on the spike protein of the original virus isolated in 2019. As a consequence, the humoral response is a major determinant of vaccine efficacy (Gilbert et al., 2022).

Antibodies recognize epitopes located in several SARS-CoV-2 proteins, but most of them target epitopes in the spike: of the 12,004 antibodies registered at the Coronavirus Antibody Database (https://opig.stats.ox.ac.uk/webapps/covabdab/, accessed on February 12, 2023), 11,900 recognize different epitopes in the S protein. Early analyses identified three hot-regions for epitopes in this protein (Li et al., 2020) in addition to others, including the RBD, for which the correlation in production of antibodies by different individual sera was not high. Epitopes in this protein have received special attention because of its role in most anti-COVID-19 vaccines developed and administered so far (Liu et al., 2020; Dejnirattisai et al., 2021). In addition, the involvement of this protein in several major functions related to infection and transmission has made it a preferential target of monoclonal antibodies-based therapies.

There are two major problems at the population level with the protection provided by antibodies to SARS-CoV-2. One is their relatively fast waning; the other results from natural selection favoring epitope mutations that evade recognition by antibodies while keeping, if not improving, their functionality. Antibody levels measured in infection-naïve persons receiving successive doses of vaccines, in different combinations and timings, helped define the recommended course of vaccination for the different formulations and regimes (Polack et al., 2020; Sadoff et al., 2021). In addition, follow-up studies of these cohorts revealed a consistent trend of waning immunity through time (Goldberg et al., 2021; Tartof et al., 2021) which, along with the continuous evolution of the virus and the emergence of variants with different epitope configuration in the spike (e.g., Omicron), have led to recommending the administration of booster shots (Sette and Crotty, 2022), now including hybrid formulations with the original Wuhan and the new Omicron variants to restore the protective effect of antibodies to prevent systemic infection.

The second issue derived from the extensive vaccination campaigns and natural infections is the elevated frequency of antibodies and the selective pressure they may exert upon the virus. As previously mentioned, the major fraction of antibodies against SARS-CoV-2 are directed against epitopes in the spike. In consequence, a faster evolution in this protein, compared with others in the viral proteome, is to be expected, as mutations in epitopes that provide an escape from recognition by antibodies will be favored by natural selection (Jungreis et al., 2021; Telenti et al., 2022). Although this has indeed been observed, the selective pressure is not as high as it might be expected. The reason is that current vaccines do not provide protection against transmission of the virus, and this is the main trait driving its evolution. As long as the virus can find new hosts on which it may replicate, although less intensively and for a shorter time than when producing a complete, systemic infection, and be transmitted, the pressure to escape from the hosts' antibodies will not be very high. Nevertheless, signs of convergent evolution in the spike protein have been detected in the recent evolution of Omicron subvariants, with mutations accumulating in epitopes, possibly as a consequence of selection to escape from antibodies (Focosi et al., 2023). An undesirable byproduct of this selective evolution is the increasing resistance to some monoclonal antibodies used in the treatment of severe infections, such as Evushled and Bebtelomivab.

In the case of T-cell responses, the pattern of immunodominance is much broader than for neutralizing antibodies. In a study on convalescent patients, Tarke and coworkers (Tarke et al., 2021) found that nine viral proteins (NSP3, NSP4, NSP12, NSP13, S, ORF3a, M, ORF8, and N) accounted for more than 80% of the total $CD4^+$ T-cell response. For $CD8^+$ T-cell responses, eight viral proteins (NSP3, NSP4, NSP6, NSP12, S, ORF3a, M, and N) contributed 81% of the total response. Also, at the population level, there is more heterogeneity for T-cell than B-cell responses, as a consequence of the highly polymorphic nature of human HLA genes. For these reasons, the signatures of T-cell escape are expected to be more distributed across the viral genome and to confer less drastic phenotypes than changes associated with antibody escape.

In general, T-cell responses and the selective pressure they impose on the virus have been analyzed in comparatively little detail. A few studies have however documented intra-host evolution within T-cell epitopes. In the context of acute infections, mutations within $CD8^+$ epitopes in the N, M, and spike proteins were detected and associated with the loss of epitope-specific responses (Agerer et al., 2021). But it is in the case of prolonged infections, especially in immunocompromised hosts, that the pressure on T-cell epitopes is highest. For instance, the NSP3 T504P mutation, which determines the loss of a $CD8^+$ epitope, was detected in multiple individuals with impaired humoral immune deficiency but

active T-cell responses (Stanevich et al., 2023; Wilkinson et al., 2022). Specific studies have also addressed the T-cell responses toward VOCs. Variants that affect $CD8^+$ T-cell epitopes were detected in the different viral proteins, including N and S. The N protein P13L, present in Omicron, disrupts a $B^*27:05$ restricted epitope, whereas L452R found in the spike proteins of the Delta, Epsilon, and BA.4/BA.5 variants results in the loss of an $A^*24:02$-restricted $CD8^+$ epitope (Motozono et al., 2021). Spike mutations that affect $CD4^+$ T-cell responses were also identified in some VOCs (e.g., D1118H in Alpha; L18F, D80A, and D215G in Beta) (Reynolds et al., 2021; Riou et al., 2022). However, it remains to be evaluated whether the selective pressure driving the emergence of these mutations is indeed caused by T-cell responses or by other effects (e.g., antibody evasion, affinity for ACE2 binding). Also, and reassuringly, the bulk of T-cell responses induced by infection and vaccination was shown to be preserved against most VOCs. For instance, Keeton and co-workers analyzed vaccinated and convalescent subjects and measured the amount of T-cell responses against the Omicron spike protein (Keeton et al., 2022). They found that 70% and 80% of the $CD4^+$ and $CD8^+$ T-cell response was maintained in both study groups. Also, the magnitude of Omicron-reactive T cells was similar to that of the Beta and Delta variants. Innate immunity is another essential component of the response to viral infections. In analogy to most pathogens, SARS-CoV-2 can hijack different innate immunity pathways. As anticipated above, this role is mainly played by accessory proteins, but the N protein (as well as a recently described truncated form of N) has also been shown to participate to immune suppression by sequestering viral RNA to prevent activation of cytosolic sensors (Mears et al., 2022; Thorne et al., 2022). Mutations in these proteins might thus contribute to SARS-CoV-2 adaptation. Indeed, the Alpha variant was found to suppress innate immune responses more effectively than earlier lineages. This is achieved by increasing the expression of the immune antagonists N, ORF9b, and ORF6 proteins (Thorne et al., 2022). Notably, BA.4 and BA.5 also show increased expression of ORF6 (Reuschl et al., 2023). The specific genetic changes that underlie increased expression of immune antagonists have not been identified, but may reside in regulatory regions that modulate subgenomic RNA synthesis and/or protein expression.

Some VOCs, including Alpha and Omicron, were also shown to be more resistant to the antiviral effects of interferon (Thorne et al., 2022; Lista et al., 2022). The mechanisms responsible for this effect are unknown, although some studies have suggested that changes in the spike proteins make the virus less susceptible to restriction by IFITM proteins (Mesner et al., 2023; Winstone et al., 2021).

5.4 Ongoing Evolution

SARS-CoV-2 is no exception to a basic biological rule: as long as it reproduces it will keep evolving. As other RNA viruses, it evolves at a fast pace at the molecular level, although these changes do not necessarily represent relevant differences at the phenotypic level. However, the different processes that drive the evolution of every species may change through time and this, as described above, has already occurred in SARS-CoV-2.

As of this writing (February 2023), the COVID-19 pandemic is dominated by the VOC Omicron (lineage B.1.1.529), which has evolved into a large number of subvariants. This lineage has been expanding for more than a year. During this time, no new similar variant has emerged with a similar success, even at a local scale. There is nothing that prevents this from occurring, but that is the result from the stochastic nature of mutations and the advantage(s) conferred by them. We have witnessed how VOC Alpha replaced previously circulating lineages, all characterized by the D614G mutation in

the spike protein. The same story was repeated a few months later, when VOC Delta replaced Alpha, which had become the dominant variant in most countries, and again when Omicron replaced Delta. However, since Omicron became dominant the pattern of evolution has changed in different aspects. Firstly, there has been **competition between subvariants** of Omicron, especially BA.2 and BA.4/BA.5, with no clear winner. The two subvariants have coexisted in some countries but in others one or the other has prevailed, usually gaining a stronghold earlier, with some contribution from stochastic factors (early arrival, spread in a highly efficient transmission network, association to a few super-spreading events, and so on). The long duration of Omicron dominance has allowed the evolution of many subvariants within the previous lineages. This is also reflected in a proliferation of PANGO lineages and an increasing difficulty in keeping track of them. It is not possible to intensify surveillance efforts on one or a few variants, as most of the newly arisen subvariants have very similar characteristics. Another difference with previous evolution periods of SARS-CoV-2 is the emergence of "**second generation**" variants (Roemer et al., 2022). These are defined as sufficiently distinct variants, with a number of mutations that differentiate them from their known ancestors, from which they do not represent a radical change of properties. These second-generation variants (examples are BA.2.75, BA.2.10.4, BJ.1, BS.1(9), BA.2.3.20, BA.2.83, BP.1, and DD.1) may have arisen in immunocompromised persons with persistent infections of SARS-CoV-2 but they may also derive from gradual accumulation of mutations in unsampled transmission chains. They differ in the immunogenic properties but much less so than Delta differed from Alpha or Omicron from Delta. Thirdly, another potentially important difference in the current evolution of SARS-CoV-2 is the observation of **convergent evolution**. Different subvariants of the major Omicron sublineages (BA.2, BA.4/BA.5) are incorporating the same mutations in a few key positions of the RBD. These positions, especially R346, K444, L452, N460, and F486, are involved in epitopes targeted by some approved drugs for treatment of Covid-19 that are monoclonal antibodies (Evusheld and Bebtelovimab). The increased number of mutations in these positions reduces the neutralization of these variants by antibodies, with the potential risk of lowering vaccine efficiency and increasing transmissibility. Finally, two of the second-generation variants, BJ.1 and BA.2.75, have recombined originating a new subvariant, XBB. The difference is that this variant has increased significantly in frequency in some areas, and it has been characterized as the "most immunity evading" variant found so far.

It is not possible to predict how these variants will evolve in the near future. Competition among variants may be overcome by convergent evolution, with some of them becoming established in different populations and with distinct results in terms of relative prevalence. It is also possible that a new variant appears displacing all Omicron subvariants in a short time. From a public health perspective, in light of different alternative scenarios, the recommended course of action is essentially the same: maintain genomic surveillance to keep track of the circulating variants and their dynamics and adjust vaccine composition accordingly.

6. Conclusion

For the first time in history, modern science is witnessing the emergence, spread, and early evolution of a new pathogen for humans from almost the very moment when the first infections of humans occurred. Through these years, the virus has evolved as a result of different factors and processes. Some of these are intrinsic, such as mutation, recombination, or response to immune pressure, but

others have depended on epidemiological or political factors. As a result, the virus has accumulated changes at different paces, alternating moments of cumulative incorporation of mutations with some moments in which radically different variants appeared in a particular population and spread all over the world a few weeks later. This has resulted in the evolution of new variants that have usually been associated with new waves of infections. One obvious consequence of this type of evolution is its intrinsic unpredictability. New variants may appear anywhere at any time, but the chances of this occurring will be higher the more people are infected, especially immunocompromised persons capable of sustaining long lasting infections. Current vaccines cannot fully stop transmission and the waning of the immune response adds to the chances for the emergence of new variants in the future. Extending vaccination to cover as many persons as possible, especially those in high-risk groups, and maintaining an active surveillance of the genomic variation of the virus will help to control the spread of new high-risk variants while SARS-CoV-2 keeps evolving.

Acknowledgments

MS was supported by the Italian Ministry of Health ("Ricerca Corrente 2022" Program). FGC was supported by project PID2021-127010OB-I00, from the MICIN (Spanish Government), project CIPROM-2021-053, from Generalitat Valenciana, and CIBER in Epidemiology and Public Health.

References

Agerer, B., et al., 2021. SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8+ T cell responses. Sci. Immunol. 6 (57), eabg6461. https://www.science.org/doi/full/10.1126/sciimmunol.abg6461.

Alm, E., et al., 2020. Geographical and temporal distribution of SARS-CoV-2 clades in the WHO European Region, January to June 2020. Euro Surveill. 25, 2001410. https://doi.org/10.2807/1560-7917.ES.2020.25.32.2001410.

Anand, P., et al., 2020. SARS-CoV-2 strategically mimics proteolytic activation of human ENaC. Elife 9. https://doi.org/10.7554/eLife.58603. India: eLife Sciences Publications Ltd.

Anderson, J.P., Daifuku, R., Loeb, L.A., 2004. Viral error catastrophe by mutagenic nucleosides. Annu. Rev. Microbiol. 58, 183−205. https://doi.org/10.1146/annurev.micro.58.030603.123649. United States.

Anderson, R.M., May, R.M., 1992. Infectious Diseases of Humans: Dynamics and Control. OUP.

Anderson, R.M., et al., 2021. The SARS-CoV-2 pandemic: remaining uncertainties in our understanding of the epidemiology and transmission dynamics of the virus, and challenges to be overcome. Interface Focus 11 (6), 20210008. https://doi.org/10.1098/rsfs.2021.0008. United Kingdom: Royal Society Publishing.

Anthony, S.J., et al., 2017. Global patterns in coronavirus diversity. Virus Evolution 3 (1), vex012. https://doi.org/10.1093/ve/vex012. United States: Oxford University Press.

Arora, P., et al., 2023. Omicron sublineage BQ.1.1 resistance to monoclonal antibodies. Lancet Infect. Dis. 23 (1), 22−23. https://doi.org/10.1016/S1473-3099(22)00733-2. Germany: Elsevier Ltd.

Bartha, F.A., et al., 2021. Potential severity, mitigation, and control of Omicron waves depending on pre-existing immunity and immune evasion. medRxiv. https://doi.org/10.1101/2021.12.15.21267884. Hungary: medRxiv.

Beaudoin-Bussières, G., et al., 2022. SARS-CoV-2 accessory protein ORF8 decreases antibody-dependent cellular cytotoxicity. Viruses 14 (6), 1237. https://doi.org/10.3390/v14061237. MDPI AG.

Belshaw, R., Pybus, O.G., Rambaut, A., 2007. The evolution of genome compression and genomic novelty in RNA viruses. Genome Res. 17 (10), 1496−1504. https://doi.org/10.1101/gr.6305707. United Kingdom.

Biggerstaff, M., et al., 2014. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. BMC Infect. Dis. 14 (1), 480. https://doi.org/10.1186/1471-2334-14-480. Springer Science and Business Media LLC.

Bloom, J.D., et al., 2021. Investigate the origins of COVID-19. Science 372, 694.

Bolze, A., et al., 2022. Evidence for SARS-CoV-2 Delta and Omicron co-infections and recombination. Med 3, 848−859. https://doi.org/10.1016/j.medj.2022.10.002.

Boni, M.F., et al., 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat. Microbiol. 5 (11), 1408−1417. https://doi.org/10.1038/s41564-020-0771-4. United States: Nature Research.

Brian, D.A., Baric, R.S., 2005. Coronavirus genome structure and replication. Curr. Top. Microbiol. Immunol. 287, 1−30. https://doi.org/10.1007/3-540-26765-4_1. United States: Springer Verlag.

Brito, A.F., et al., 2022. Global disparities in SARS-CoV-2 genomic surveillance. Nat. Commun. 13, 7003. https://doi.org/10.1038/s41467-022-33713-y.

Buonavoglia, C., et al., 2006. Canine coronavirus highly pathogenic for dogs. Emerg. Infect. Dis. 12 (3), 492−494. https://doi.org/10.3201/eid1203.050839. Centers for Disease Control and Prevention (CDC).

Chaguza, C., et al., 2022. Accelerated SARS-CoV-2 intrahost evolution leading to distinct genotypes during chronic infection. medRxiv. https://doi.org/10.1101/2022.06.29.22276868. United States: medRxiv.

Chen, Di, et al., 2021. ORF3a of SARS-CoV-2 promotes lysosomal exocytosis-mediated viral egress. Dev. Cell 56 (23), 3250−3263.e5. https://doi.org/10.1016/j.devcel.2021.10.006. Elsevier BV.

Chinese SARS Molecular Epidemiology Consortium, 2004. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 303 (5664), 1666−1669. https://doi.org/10.1126/science.1092002. American Association for the Advancement of Science (AAAS).

Chowell, G., Nishiura, H., 2015. Characterizing the transmission dynamics and control of Ebola virus disease. PLoS Biol. 13 (1), e1002057. https://doi.org/10.1371/journal.pbio.1002057. Public Library of Science (PLoS).

Colson, P., et al., 2022. Culture and identification of a 'Deltamicron' SARS-CoV-2 in a three cases cluster in southern France. J. Med. Virol. 94 (8), 3739−3749. https://doi.org/10.1002/jmv.27789. France: John Wiley and Sons Inc.

Combes, P., et al., 2022. Evidence of co-infections during Delta and Omicron SARS-CoV-2 variants co-circulation through prospective screening and sequencing. Clin. Microbiol. Infection 28 (11), 1503.e5−1503.e8. https://doi.org/10.1016/j.cmi.2022.06.030. Elsevier BV.

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat. Microbiol. 5, 536−544.

Crotty, S., Cameron, C.E., Andino, R., 2001. RNA virus error catastrophe: direct molecular test by using ribavirin. Proc. Natl. Acad. Sci. U. S. A. 98 (12), 6895−6900. https://doi.org/10.1073/pnas.111085598.

Cui, J., Li, F., Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17 (3), 181−192. https://doi.org/10.1038/s41579-018-0118-9. China: Nature Publishing Group.

Dejnirattisai, W., et al., 2021. The antigenic anatomy of SARS-CoV-2 receptor binding domain. Cell 184, 2183−2200. https://doi.org/10.1016/j.cell.2021.02.032.

Delahay, R.J., et al., 2021. Assessing the risks of SARS-CoV-2 in wildlife. One Health Outlook 3 (1), 7. https://doi.org/10.1186/s42522-021-00039-6. Springer Science and Business Media LLC.

Delaune, D., et al., 2021. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. Nat. Commun. 12 (1), 6563. https://doi.org/10.1038/s41467-021-26809-4. Springer Science and Business Media LLC.

Dezordi, F.Z., et al., 2022. Unusual SARS-CoV-2 intrahost diversity reveals lineage superinfection. Microb. Genom. 8 (3), 751. https://doi.org/10.1099/mgen.0.000751.

Dhar, M.S., et al., 2021. Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. Science 374, 995−999.

Dyer, O., 2021. Covid-19: WHO says laboratory escape theory is 'extremely unlikely' after mission to China. BMJ (Clinical research ed.) 372, n428. https://doi.org/10.1136/bmj.n428. France: NLM (Medline).

Eales, O., et al., 2022. SARS-CoV-2 lineage dynamics in England from September to November 2021: high diversity of Delta sub-lineages and increased transmissibility of AY.4.2. BMC Infect. Dis. 22 (1), 647. https://doi.org/10.1186/s12879-022-07628-4. Springer Science and Business Media LLC.

Earnest, R., et al., 2022. Comparative transmissibility of SARS-CoV-2 variants delta and alpha in New England, USA. Cell Rep. Med. 399, 100583. https://doi.org/10.1016/j.xcrm.2022.100583.

Eckerle, L.D., et al., 2007. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. J. Virol. 81 (22), 12135−12144. https://doi.org/10.1128/JVI.01296-07. United States.

Eckerle, L.D., et al., 2010. Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. PLoS Pathog. 6 (5), 1−15. https://doi.org/10.1371/journal.ppat.1000896. United States.

Eigen, M., 2002. Error catastrophe and antiviral strategy. Proc. Natl. Acad. Sci. U. S. A. 99 (21), 13374−13376. https://doi.org/10.1073/pnas.212514799.

Eigen, M., 1971. Selforganization of matter and the evolution of biological macromolecules. Die Naturwissenschaften 58 (10), 465−523. https://doi.org/10.1007/bf00623322. Springer Science and Business Media LLC.

Finkel, Y., et al., 2021. The coding capacity of SARS-CoV-2. Nature 589 (7840), 125−130. https://doi.org/10.1038/s41586-020-2739-1. Springer Science and Business Media LLC.

Focosi, D., et al., 2023. Convergent evolution in SARS-CoV-2 spike creates a variant soup from which new COVID-19 waves emerge. Int. J. Mol. Sci. 24 (3), 2264. https://doi.org/10.3390/ijms24032264. Italy: MDPI.

Forni, D., et al., 2022. Homology-based classification of accessory proteins in coronavirus genomes uncovers extremely dynamic evolution of gene content. Mol. Ecol. 31 (13), 3672−3692. https://doi.org/10.1111/mec.16531. Wiley.

Forni, D., et al., 2017. Molecular evolution of human coronavirus genomes. Trends Microbiol. 25 (1), 35−48. https://doi.org/10.1016/j.tim.2016.09.001. Elsevier BV.

Ge, X.Y., et al., 2017. Detection of alpha- and betacoronaviruses in rodents from Yunnan, China. Virol. J. 14 (1), 98. https://doi.org/10.1186/s12985-017-0766-9. China: BioMed Central Ltd.

Geng, H., et al., 2021. SARS-CoV-2 ORF8 forms intracellular aggregates and inhibits IFNγ-induced antiviral gene expression in human lung epithelial cells. Front. Immunol. 12, 679482. https://doi.org/10.3389/fimmu.2021.679482. United States: NLM (Medline).

Ghai, R.R., et al., 2021. Animal reservoirs and hosts for emerging alphacoronaviruses and betacoronaviruses. Emerg. Infect. Dis. 27 (4), 1015−1022. https://doi.org/10.3201/eid2704.203945. United States: Centers for Disease Control and Prevention (CDC).

Gilbert, P.B., et al., 2022. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. Science 375 (6576), 43−50. https://doi.org/10.1126/science.abm3425. Moderna, Inc. Team§.

Goldberg, Y., et al., 2021. Waning immunity after the BNT162B2 vaccine in Israel. N. Engl. J. Med. 385 (24), e85. https://doi.org/10.1056/NEJMoa2114228. Israel: Massachussetts Medical Society.

González-Candelas, F., et al., 2021. One year into the pandemic: short-term evolution of SARS-CoV-2 and emergence of new lineages. Infect. Genet. Evol. 92, 104869. https://doi.org/10.1016/j.meegid.2021.104869. Elsevier BV.

Graham, R.L., Baric, R.S., 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. J. Virol. 84 (7), 3134−3146. https://doi.org/10.1128/JVI.01394-09. United States.

Gribble, J., et al., 2021. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. PLoS Pathog. 17 (1), e1009226. https://doi.org/10.1371/journal.ppat.1009226. United States: Public Library of Science.

Gu, H., et al., 2022. Detection of a BA.1/BA.2 recombinant in travelers arriving in Hong Kong, February 2022. Emerg. Infect. Dis. 28 (6), 1276−1278. https://doi.org/10.3201/eid2806.220523. Hong Kong: medRxiv.

Gutierrez, B., et al., 2022. Emergence and widespread circulation of a recombinant SARS-CoV-2 lineage in North America. Cell Host Microbe 30 (8), 1112−1123.e3. https://doi.org/10.1016/j.chom.2022.06.010. United Kingdom: Cell Press.

Hadfield, J., et al., 2018. NextStrain: real-time tracking of pathogen evolution. Bioinformatics 34 (23), 4121−4123. https://doi.org/10.1093/bioinformatics/bty407. United States: Oxford University Press.

Hale, V.L., et al., 2022. SARS-CoV-2 infection in free-ranging white-tailed deer. Nature 602 (7897), 481−486. https://doi.org/10.1038/s41586-021-04353-x. United States: Nature Research.

Han, L., et al., 2021. SARS-CoV-2 ORF9b antagonizes type I and III interferons by targeting multiple components of the RIG-I/MDA-5−MAVS, TLR3−TRIF, and cGAS−STING signaling pathways. J. Med. Virol. 93 (9), 5376−5389. https://doi.org/10.1002/jmv.27050. China: John Wiley and Sons Inc.

Harari, S., et al., 2022. Drivers of adaptive evolution during chronic SARS-CoV-2 infections. Nat. Med. 28 (7), 1501−1508. https://doi.org/10.1038/s41591-022-01882-4. Springer Science and Business Media LLC.

Harrison, N.L., Sachs, J.D., 2022. A call for an independent inquiry into the origin of the SARS-CoV-2 virus. Proc. Natl. Acad. Sci. U. S. A. 119 (21), e2202769119. https://doi.org/10.1073/pnas.2202769119. United States: National Academy of Sciences.

Hayn, M., et al., 2021. Systematic functional analysis of SARS-CoV-2 proteins uncovers viral innate immune antagonists and remaining vulnerabilities. Cell Rep. 35 (7), 109126. https://doi.org/10.1016/j.celrep.2021.109126. Germany: Elsevier B.V.

Hill, V., et al., 2022. The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. Virus Evol. 8 (2), veac080. https://doi.org/10.1093/ve/veac080. United Kingdom: bioRxiv.

Hodcroft, E.B., et al., 2021. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. Nature 595 (7869), 707−712. https://doi.org/10.1038/s41586-021-03677-y. Switzerland: Nature Research.

Holmes, E.C., 2003. Error thresholds and the constraints to RNA virus evolution. Trends Microbiol. 11 (12), 543−546. https://doi.org/10.1016/j.tim.2003.10.006. United Kingdom: Elsevier Ltd.

Holmes, E.C., et al., 2021. The origins of SARS-CoV-2: a critical review. Cell 184 (19), 4848−4856. https://doi.org/10.1016/j.cell.2021.08.017. Australia: Elsevier B.V.

Hong, Q., et al., 2022. Molecular basis of receptor binding and antibody neutralization of Omicron. Nature 604 (7906), 546−552. https://doi.org/10.1038/s41586-022-04581-9. Springer Science and Business Media LLC.

Hou, P., et al., 2023. The ORF7a protein of SARS-CoV-2 initiates autophagy and limits autophagosome-lysosome fusion via degradation of SNAP29 to promote virus replication. Autophagy 19 (2), 551−569. https://doi.org/10.1080/15548627.2022.2084686. China: Taylor and Francis Ltd.

Hu, D., et al., 2018. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. Emerg. Microb. Infect. 7 (1), 1−10. https://doi.org/10.1038/s41426-018-0155-5. Informa UK Limited.

Huang, C., et al., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395 (10223), 497−506. https://doi.org/10.1016/s0140-6736(20)30183-5. Elsevier BV.

Jackson, B., et al., 2021. Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. Cell 184 (20), 5179−5188.e8. https://doi.org/10.1016/j.cell.2021.08.014. United Kingdom: Elsevier B.V.

Jungreis, I., Sealfon, R., Kellis, M., 2021. SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. Nat. Commun. 12 (1), 2642. https://doi.org/10.1038/s41467-021-22905-7. Springer Science and Business Media LLC.

Kan, B., et al., 2005. Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. J. Virol. 79 (18), 11892−11900. https://doi.org/10.1128/JVI.79.18.11892-11900.2005. China.

Keeton, R., et al., 2022. T cell responses to SARS-CoV-2 spike cross-recognize Omicron. Nature 603, 488−492. https://doi.org/10.1038/s41586-022-04460-3.

Keske, Ş., et al., 2023. Duration of infectious shedding of SARS-CoV-2 Omicron variant and its relation with symptoms. Clin. Microbiol. Infect. 29 (2), 221−224. https://doi.org/10.1016/j.cmi.2022.07.009.

Konings, F., et al., 2021. SARS-CoV-2 Variants of Interest and Concern naming scheme conducive for global discourse. Nat. Microbiol. 6 (7), 821−823. https://doi.org/10.1038/s41564-021-00932-w. Switzerland: Nature Research.

Kuchipudi, S.V., et al., 2022. Multiple spillovers from humans and onward transmission of SARS-CoV-2 in white-tailed deer. Proc. Natl. Acad. Sci. U. S. A. 119 (6). https://doi.org/10.1073/pnas.2121644119. United States: National Academy of Sciences.

Lacek, K.A., et al., 2022. SARS-CoV-2 delta-omicron recombinant viruses, United States. Emerg. Infect. Dis. 28 (7), 1442−1445. https://doi.org/10.3201/eid2807.220526. United States: Centers for Disease Control and Prevention (CDC).

Lam, T.T.Y., et al., 2020. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature 583 (7815), 282−285. https://doi.org/10.1038/s41586-020-2169-0. China: Nature Research.

Larsen, H.D., et al., 2021. Preliminary report of an outbreak of SARS-CoV-2 in mink and mink farmers associated with community spread, Denmark, June to November 2020. Euro Surveill. 26 (5), 2100009. https://doi.org/10.2807/1560-7917.ES.2021.26.5.210009. Denmark: European Centre for Disease Prevention and Control (ECDC).

Lauber, C., Gorbalenya, A.E., 2012. Partitioning the genetic diversity of a virus family: approach and evaluation through a case study of picornaviruses. J. Virol. 86 (7), 3890−3904. https://doi.org/10.1128/jvi.07173-11. American Society for Microbiology.

Lednicky, J.A., et al., 2021. Independent infections of porcine deltacoronavirus among Haitian children. Nature 600 (7887), 133−137. https://doi.org/10.1038/s41586-021-04111-z. United States: Nature Research.

Lefkowitz, E.J., et al., 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res. 46 (1), D708−D717. https://doi.org/10.1093/nar/gkx932. United States: Oxford University Press.

Li, F., et al., 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309 (5742), 1864−1868. https://doi.org/10.1126/science.1116480. American Association for the Advancement of Science (AAAS).

Li, L.L., et al., 2021. A novel SARS-CoV-2 related coronavirus with complex recombination isolated from bats in Yunnan province, China. Emerg. Microb. Infect. 10 (1), 1683−1690. https://doi.org/10.1080/22221751.2021.1964925. China: Taylor and Francis Ltd.

Li, M., et al., 2023. Comparative susceptibility of SARS-CoV-2, SARS-CoV, and MERS-CoV across mammals. ISME J. 17 (4), 549−560. https://doi.org/10.1038/s41396-023-01368-2. Springer Science and Business Media LLC.

Li, X., et al., 2022. SARS-CoV-2 ORF10 suppresses the antiviral innate immune response by degrading MAVS through mitophagy. Cell. Mol. Immunol. 19 (1), 67−78. https://doi.org/10.1038/s41423-021-00807-4. Springer Science and Business Media LLC.

Li, Y., et al., 2020. Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. Cell. Mol. Immunol. 17 (10), 1095−1097. https://doi.org/10.1038/s41423-020-00523-5. China: Springer Nature.

Li, B., et al., 2022. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. Nat. Commun. 13 (1), 460. https://doi.org/10.1038/s41467-022-28089-y. China: Nature Research.

Li, J.Y., et al., 2020. The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway. Virus Res. 286, 198074. https://doi.org/10.1016/j.virusres.2020.198074. China: Elsevier B.V.

Li, X., et al., 2020. Emergence of SARS-CoV-2 through recombination and strong purifying selection. Sci. Adv. 6 (27), eabb9153. https://doi.org/10.1126/sciadv.abb9153. American Association for the Advancement of Science (AAAS).

Licitra, B.N., Duhamel, G.E., Whittaker, G.R., 2014. Canine enteric coronaviruses: emerging viral pathogens with distinct recombinant spike proteins. Viruses 6 (8), 3363−3376. https://doi.org/10.3390/v6083363. United States.

Lindh, E., et al., 2022. Genomic and epidemiological report of the recombinant XJ lineage SARS-CoV-2 variant, detected in northern Finland, January 2022. Euro Surveill. 27 (16), 2200257. https://doi.org/10.2807/1560-7917.es.2022.27.16.2200257. European Centre for Disease Control and Prevention (ECDC).

Lista, M.J., et al., 2022. The P681H mutation in the spike glycoprotein of the alpha variant of SARS-CoV-2 escapes IFITM restriction and is necessary for type I interferon resistance. J. Virol. 96, e0125022. https://doi.org/10.1128/jvi.01250-22.

Liu, L., et al., 2020. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. Nature 584 (7821), 450−456. https://doi.org/10.1038/s41586-020-2571-7. United States: Nature Research.

Liu, Y., et al., 2020. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J. Trav. Med. 27 (2), taaa021. https://doi.org/10.1093/jtm/taaa021. China: Oxford University Press.

Liu, Y., Rocklöv, J., 2022. The effective reproductive number of the Omicron variant of SARS-CoV-2 is several times relative to Delta. J. Trav. Med. 29 (3), taac037. https://doi.org/10.1093/jtm/taac037. Oxford University Press (OUP).

Liu, Y., Rocklöv, J., 2021. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. J. Trav. Med. 28 (7), taab124. https://doi.org/10.1093/jtm/taab124. Oxford University Press (OUP).

Lole, K.S., et al., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C- infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73 (1), 152−160. https://doi.org/10.1128/jvi.73.1.152-160.1999. India: American Society for Microbiology.

Lytras, S., et al., 2022. Exploring the natural origins of SARS-CoV-2 in the light of recombination. Genome Biol. Evol. 14 (2), evac018. https://doi.org/10.1093/gbe/evac018. United Kingdom: Oxford University Press.

MacLean, O.A., et al., 2021. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. PLoS Biol. 19 (3), e3001115. https://doi.org/10.1371/journal.-pbio.3001115. United Kingdom: Public Library of Science.

Mahdy, M.A.A., Younis, W., Ewaida, Z., 2020. An overview of SARS-CoV-2 and animal infection. Front. Vet. Sci. 7, 596391. https://doi.org/10.3389/fvets.2020.596391. Egypt: Frontiers Media S.A.

Mallapaty, S., 2022. Surprising Omicron origins study comes under scrutiny. Nature 612 (7940), 387−388. https://doi.org/10.1038/d41586-022-04357-1. Springer Science and Business Media LLC.

Maponga, T.G., et al., 2023. Persistent SARS-CoV-2 infection with accumulation of mutations in a patient with poorly controlled HIV infection. Clin. Infect. Dis. 76 (3), ciac548. https://doi.org/10.1093/cid/ciac548.

McCrone, J.T., et al., 2022. Context-specific emergence and growth of the SARS-CoV-2 Delta variant. Nature 610 (7930), 154−160. https://doi.org/10.1038/s41586-022-05200-3. United Kingdom: Nature Research.

Mears, H.V., et al., 2022. Emergence of new subgenomic mRNAs in SARS-CoV-2. bioRxiv. https://doi.org/10.1101/2022.04.20.488895. United Kingdom: bioRxiv.

Mesner, D., et al., 2023. SARS-CoV-2 evolution influences GBP and IFITM sensitivity. Proc. Natl. Acad. Sci. U. S. A. 120 (5), e2212577120. https://doi.org/10.1073/pnas.2212577120. United Kingdom: National Academy of Sciences.

Miao, G., et al., 2021. ORF3a of the COVID-19 virus SARS-CoV-2 blocks HOPS complex-mediated assembly of the SNARE complex required for autolysosome formation. Dev. Cell 56 (4), 427−442.e5. https://doi.org/10.1016/j.devcel.2020.12.010. Elsevier BV.

Motozono, C., et al., 2021. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. Cell Host Microbe 29 (7), 1124−1136.e11. https://doi.org/10.1016/j.chom.2021.06.006. Japan: Cell Press.

Müller, N.F., Kistler, K.E., Bedford, T., 2022. A Bayesian approach to infer recombination patterns in coronaviruses. Nat. Commun. 13 (1), 4186. https://doi.org/10.1038/s41467-022-31749-8. United States: Nature Research.

Neher, R.A., 2022. Contributions of adaptation and purifying selection to SARS-CoV-2 evolution. Virus Evol. 8 (2), veac113. https://doi.org/10.1093/ve/veac113. Switzerland: bioRxiv.

Nerpel, A., et al., 2022. SARS-ANI: a global open access dataset of reported SARS-CoV-2 events in animals. Sci. Data 9 (1), 438. https://doi.org/10.1038/s41597-022-01543-8. Springer Science and Business Media LLC.

Omotuyi, O., et al., 2022. SARS-CoV-2 Omicron spike glycoprotein receptor binding domain exhibits super-binder ability with ACE2 but not convalescent monoclonal antibody. Comput. Biol. Med. 142, 105226. https://doi.org/10.1016/j.compbiomed.2022.105226. Elsevier BV.

Oreshkova, N., et al., 2020. SARS-CoV-2 infection in farmed minks, The Netherlands, April and May 2020. Euro Surveill. 25 (23), 2001005. https://doi.org/10.2807/1560-7917.ES.2020.25.23.2001005. Netherlands: European Centre for Disease Prevention and Control (ECDC).

Oude Munnink, B.B., et al., 2021. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. Science 371 (6525), 172−177. https://doi.org/10.1126/science.abe5901. Netherlands: American Association for the Advancement of Science.

Patiño-Galindo, J.Á., et al., 2021. Recombination and lineage-specific mutations linked to the emergence of SARS-CoV-2. Genome Med. 13 (1), 124. https://doi.org/10.1186/s13073-021-00943-6. United States: BioMed Central Ltd.

Pekar, J.E., et al., 2022. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. Science 377 (6609), 960−966. https://doi.org/10.1126/science.abp8337. United States: American Association for the Advancement of Science.

Pérez-Losada, M., et al., 2015. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. Infect. Genet. Evol. 30, 296−307. https://doi.org/10.1016/j.meegid.2014.12.022. Portugal: Elsevier.

Petersen, E., et al., 2020. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. Lancet Infect. Dis. 20 (9), e238−e244. https://doi.org/10.1016/S1473-3099(20)30484-9. Switzerland: Lancet Publishing Group.

Pickering, B., et al., 2022. Divergent SARS-CoV-2 variant emerges in white-tailed deer with deer-to-human transmission. Nat. Microbiol. 7 (12), 2011−2024. https://doi.org/10.1038/s41564-022-01268-9. Canada: Nature Research.

Planas, D., et al., 2021. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature 596 (7871), 276−280. https://doi.org/10.1038/s41586-021-03777-9. France: Nature Research.

Polack, F.P., et al., 2020. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. N. Engl. J. Med. 383 (27), 2603−2615. https://doi.org/10.1056/NEJMoa2034577. Argentina: Massachussetts Medical Society.

Priestnall, S.L., 2020. Canine respiratory coronavirus: a naturally occurring model of COVID-19? Vet. Pathol. 57 (4), 467−471. https://doi.org/10.1177/0300985820926485. United Kingdom: SAGE Publications Inc.

Public Health England, 2020. Public Health England Investigation of Novel SARS-COV-2 Variant 202012/01: Technical Briefing 1.

Pybus, O.G., 2021. Pango Lineage Nomenclature: Provisional Rules for Naming Recombinant Lineages. Available at: https://virological.org/t/pango-lineage-nomenclature-provisional-rules-for-naming-recombinant-lineages/657. (Accessed 7 October 2023).

Rambaut, A., et al., 2020. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat. Microbiol. 5 (11), 1403–1407. https://doi.org/10.1038/s41564-020-0770-5. United Kingdom: Nature Research.

Rashid, F., et al., 2021. The ORF8 protein of SARS-CoV-2 induced endoplasmic reticulum stress and mediated immune evasion by antagonizing production of interferon beta. Virus Res. 296, 198350. https://doi.org/10.1016/j.virusres.2021.198350. Elsevier BV.

Ren, W., et al., 2021. Mutation Y453F in the spike protein of SARS-CoV-2 enhances interaction with the mink ACE2 receptor for host adaption. PLoS Pathog. 17 (11), 1010053. https://doi.org/10.1371/journal.ppat.1010053. Public Library of Science (PLoS).

Reuschl, A.K., et al., 2023. Evolution of enhanced innate immune suppression by SARS-CoV-2 Omicron subvariants. bioRxiv. https://doi.org/10.1101/2022.07.12.499603. United Kingdom: bioRxiv.

Reynolds, C.J., et al., 2021. Prior SARS-CoV-2 infection rescues B and T cell responses to variants after first vaccine dose. Science 372 (6549), 1418–1423. https://doi.org/10.1126/science.abh1282. United Kingdom: American Association for the Advancement of Science.

Riou, C., et al., 2022. Escape from recognition of SARS-CoV-2 variant spike epitopes but overall preservation of T cell immunity. Sci. Transl. Med. 14 (631), eabj6824. https://doi.org/10.1126/scitranslmed.abj6824. South Africa: American Association for the Advancement of Science.

Roemer, C., et al., 2022. SARS-CoV-2 evolution, post-Omicron. Available at: https://virological.org/t/sars-cov-2-evolution-post-omicron/911. (Accessed 12 February 2023).

Sadoff, J., et al., 2021. Safety and efficacy of single-dose Ad26.CoV2.s vaccine against covid-19. N. Engl. J. Med. 384 (23), 2187–2201. https://doi.org/10.1056/NEJMoa2101544. Netherlands: Massachussetts Medical Society.

Scherer, E.M., et al., 2022. SARS-CoV-2 evolution and immune escape in immunocompromised patients. N. Engl. J. Med. 386 (25), 2436–2438. https://doi.org/10.1056/NEJMc2202861. United States: Massachussetts Medical Society.

Sekizuka, T., et al., 2022. Recombination between SARS-CoV-2 Omicron BA.1 and BA.2 variants identified in a traveller from Nepal at the airport quarantine facility in Japan. J. Trav. Med. 29 (6), taac051. https://doi.org/10.1093/jtm/taac051. Oxford University Press (OUP).

Sette, A., Crotty, S., 2022. Immunological memory to SARS-CoV-2 infection and COVID-19 vaccines. Immunol. Rev. 310 (1), 27–46. https://doi.org/10.1111/imr.13089. Wiley.

Sharun, K., et al., 2021. SARS-CoV-2 in animals: potential for unknown reservoir hosts and public health implications. Vet. Q. 41 (1), 181–201. https://doi.org/10.1080/01652176.2021.1921311. India: Taylor and Francis Ltd.

Simon-Loriere, E., Holmes, E.C., Pagán, I., 2013. The Effect of gene overlapping on the rate of RNA virus evolution. Mol. Biol. Evol. 30 (8), 1916–1928. https://doi.org/10.1093/molbev/mst094. France: Oxford University Press.

Smith, E.C., et al., 2013. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. PLoS Pathog. 9 (8), 1003565. https://doi.org/10.1371/journal.ppat.1003565. United States.

Song, H.D., et al., 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc. Natl. Acad. Sci. U. S. A 102 (7), 2430–2435. https://doi.org/10.1073/pnas.0409608102. China.

Stanevich, O.V., et al., 2023. SARS-CoV-2 escape from cytotoxic T cells during long-term COVID-19. Nat. Comm. 14, 149. https://doi.org/10.1038/s41467-022-34033-x.

Stout, A.E., et al., 2021. Furin cleavage sites in the spike proteins of bat and rodent coronaviruses: implications for virus evolution and zoonotic transfer from rodent species. One Health 13, 100282. https://doi.org/10.1016/j.onehlt.2021.100282. United States: Elsevier B.V.

Subissi, L., et al., 2014. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc. Natl. Acad. Sci. U. S. A 111 (37), E3900−E3909. https://doi.org/10.1073/pnas.1323705111. France: National Academy of Sciences.

Tarke, A., et al., 2021. Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. Cell Rep. Med. 2 (2), 100204. https://doi.org/10.1016/j.xcrm.2021.100204. United States: Cell Press.

Tartof, S.Y., et al., 2021. Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. Lancet 398 (10309), 1407−1416. https://doi.org/10.1016/S0140-6736(21)02183-8. United States: Elsevier B.V.

Telenti, A., Hodcroft, E.B., Robertson, D.L., 2022. The evolution and biology of SARS-CoV-2 variants. Cold Spring Harb. Perspect. Med. 12 (5), a041390. https://doi.org/10.1101/cshperspect.a041390. United States: Cold Spring Harbor Laboratory Press.

Temmam, S., et al., 2022. Bat coronaviruses related to SARS-CoV-2 and infectious for human cells. Nature 604 (7905), 330−336. https://doi.org/10.1038/s41586-022-04532-4. France: Nature Research.

Thorne, L.G., et al., 2022. Evolution of enhanced innate immune evasion by SARS-CoV-2. Nature 602, 487−495.

Tsoleridis, T., et al., 2016. Discovery of novel alphacoronaviruses in European rodents and shrews. Viruses 8 (3), 84. https://doi.org/10.3390/v8030084. United Kingdom: MDPI AG.

Turakhia, Y., et al., 2022. Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape. Nature 609 (7929), 994−997. https://doi.org/10.1038/s41586-022-05189-9. Springer Science and Business Media LLC.

Uriu, K., et al., 2023. The genotype to phenotype Japan (G2P-Japan) Consortium, Sato, K., 2023. Enhanced transmissibility. Lancet Infect. Dis. 23 (3), 280−281. https://doi.org/10.1016/S1473-3099(23)00051-8.

VanInsberghe, D., et al., 2021. Recombinant SARS-CoV-2 genomes circulated at low levels over the first year of the pandemic. Virus Evol. 7 (2), veab059. https://doi.org/10.1093/ve/veab059. United States: Oxford University Press.

Viana, R., et al., 2022. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature 603, 679−686. https://doi.org/10.1038/s41586-022-04411-y.

Vlasova, A.N., et al., 2022. Novel canine coronavirus isolated from a hospitalized patient with pneumonia in East Malaysia. Clin. Infect. Dis. 74 (3), 446−454. https://doi.org/10.1093/cid/ciab456. United States: Oxford University Press.

Wacharapluesadee, S., et al., 2021. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. Nat. Commun. 12 (1), 1430. https://doi.org/10.1038/s41467-021-21240-1. Thailand: Nature Research.

Wang, Q., et al., 2022. Alarming antibody evasion properties of rising SARS-CoV-2 BQ and XBB subvariants. bioRxiv. https://doi.org/10.1101/2022.11.23.517532. United States: bioRxiv.

Wang, Q., et al., 2019. Emerging and re-emerging coronaviruses in pigs. Curr. Opin. Virol. 34, 39−49. https://doi.org/10.1016/j.coviro.2018.12.001. United States: Elsevier B.V.

Wang, W., et al., 2015. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. Virology 474, 19−27. https://doi.org/10.1016/j.virol.2014.10.017. China: Academic Press Inc.

Wang, W., et al., 2017. Discovery of a highly divergent coronavirus in the Asian house shrew from China illuminates the origin of the alphacoronaviruses. J. Virol. 91 (17). https://doi.org/10.1128/JVI.00764-17. China: American Society for Microbiology.

Wang, W., et al., 2020. Extensive genetic diversity and host range of rodent-borne coronaviruses. Virus Evol. 6 (2), veaa078. https://doi.org/10.1093/ve/veaa078. China: Oxford University Press.

Wei, C., et al., 2021. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. J. Genet. Genom. 48 (12), 1111−1121. https://doi.org/10.1016/j.jgg.2021.12.003. China: Institute of Genetics and Developmental Biology.

Weigang, S., et al., 2021. Within-host evolution of SARS-CoV-2 in an immunosuppressed COVID-19 patient as a source of immune escape variants. Nat. Commun. 12 (1), 6405. https://doi.org/10.1038/s41467-021-26602-3. Springer Science and Business Media LLC.

Wertheim, J.O., et al., 2022. Detection of SARS-CoV-2 intra-host recombination during superinfection with Alpha and Epsilon variants in New York City. Nat. Commun. 13 (1), 3645. https://doi.org/10.1038/s41467-022-31247-x. United States: Nature Research.

Wilkinson, S.A.J., et al., 2022. Recurrent SARS-CoV-2 mutations in immunodeficient patients. Virus Evol. 8 (2), veac050. https://doi.org/10.1093/ve/veac050. United Kingdom: medRxiv.

Wille, M., Holmes, E.C., 2020. Wild birds as reservoirs for diverse and abundant gamma- and deltacoronaviruses. FEMS (Fed. Eur. Microbiol. Soc.) Microbiol. Rev. 44 (5), 631−644. https://doi.org/10.1093/femsre/fuaa026. Australia: Oxford University Press.

Willett, B.J., et al., 2022. ARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. Nat. Microbiol. 7, 1161−1179. https://doi.org/10.1038/s41564-022-01143-7.

Winstone, H., et al., 2021. The polybasic cleavage site in SARS-CoV-2 spike modulates viral sensitivity to type I interferon and IFITM2. J. Virol. 95 (9). https://doi.org/10.1128/JVI.02422-20. United Kingdom: American Society for Microbiology.

World Health Organization, 2021. WHO-Convened Global Study of Origins of SARS-CoV-2: China Part: Joint WHO-China Study.

Worobey, M., et al., 2022. The huanan seafood wholesale market in Wuhan was the early epicenter of the COVID-19 pandemic. Science 377 (6609), 951−959. https://doi.org/10.1126/science.abp8715. United States: American Association for the Advancement of Science.

Worobey, M., 2021. Dissecting the early COVID-19 cases in Wuhan. Science 374 (6572), 1202−1204. https://doi.org/10.1126/science.abm4454. American Association for the Advancement of Science (AAAS).

Wu, F., et al., 2020. A new coronavirus associated with human respiratory disease in China. Nature 579 (7798), 265−269. https://doi.org/10.1038/s41586-020-2008-3. China: Nature Research.

Wu, X., et al., 2022. Viral mimicry of interleukin-17A by SARS-CoV-2 ORF8. mBio 13 (2), e0040222. https://doi.org/10.1128/mbio.00402-22. United States: American Society for Microbiology.

Wu, Y., et al., 2022. Incubation period of COVID-19 caused by unique SARS-CoV-2 strains: a systematic review and meta-analysis. JAMA Netw. Open 5 (8), e2228008. https://doi.org/10.1001/jamanetworkopen.2022.28008. China: NLM (Medline).

Wu, Z., et al., 2018. Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. Microbiome 6 (1), 178. https://doi.org/10.1186/s40168-018-0554-9. Springer Science and Business Media LLC.

Xiao, K., et al., 2020. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature 583 (7815), 286−289. https://doi.org/10.1038/s41586-020-2313-x. China: Nature Research.

Yue, C., et al., 2023. ACE2 binding and antibody evasion in enhanced transmissibility of XBB.1.5. Lancet Infect. Dis. 23 (3), 278−280. https://doi.org/10.1016/S1473-3099(23)00010-5. China: bioRxiv.

Zhang, Y., et al., 2021. The ORF8 protein of SARS-CoV-2 mediates immune evasion through down-regulating MHC-I. Proc. Natl. Acad. Sci. U. S. A. 118 (23), e2024202118. https://doi.org/10.1073/pnas.2024202118. Proceedings of the National Academy of Sciences.

Zhou, H., et al., 2020. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. Curr. Biol. 30 (11), 2196−2203.e3. https://doi.org/10.1016/j.cub.2020.05.023. China: Cell Press.

Zhou, H., et al., 2021. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. Cell 184 (17), 4380−4391.e14. https://doi.org/10.1016/j.cell.2021.06.008. Elsevier BV.

Zhou, P., et al., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579 (7798), 270−273. https://doi.org/10.1038/s41586-020-2012-7. China: Nature Research.

Zhu, N., et al., 2019. A novel coronavirus from patients with pneumonia in China. N. Engl. J. Med. 382, 727−733. https://doi.org/10.1056/NEJMoa2001017.

CHAPTER

33 Molecular Tools for Systematic and Epidemiological Investigations of Socioeconomically Important Parasitic Nematodes of Animals and Humans

Robin B. Gasser[1], **Ian Beveridge**[1] **and Neil B. Chilton**[2]

[1]*Department of Veterinary Biosciences, Melbourne Veterinary School, The University of Melbourne, Melbourne, VIC, Australia;* [2]*Department of Biology, University of Saskatchewan, Saskatoon, SK, Canada*

1. Background and Historic Account

The accurate identification of parasites to species or genotype is central to many fundamental (taxonomy, phylogeny, and epidemiology) and applied (diagnosis of infection and disease) areas of the discipline of Parasitology. Traditionally, parasites have been identified based on their morphological features and measurements. However, specific identification is frequently not possible using morphology and morphometrics alone, because often there are very few characters that allow these worms to be identified or distinguished, such that complementary approaches need to be employed. Classical biochemical (e.g., multilocus enzyme electrophoresis; Andrews and Chilton, 1999) and/or DNA-based methods (e.g., Southern blot hybridization; Southern, 2006) were used, but the advent of the polymerase chain reaction (PCR) technique (Saiki et al., 1985) changed the way molecular investigations were undertaken. PCR, conducted using thermostable *Taq* polymerase (Saiki et al., 1988), led to a revolution in all areas of the Life Sciences, including Parasitology. Our team was among the first to apply PCR to studies of parasitic nematodes.

From 1990, our research was focused on identifying genetic markers that would allow the specific identification of parasitic nematodes of veterinary, agricultural, and human health importance, particularly members of the order Strongylida. In the absence of genomic information for all of the parasites that we were working with, searching for specific markers was "like looking for a needle in a haystack." Initially, we were challenged by the lack of genetic data and publications on the topic. Nonetheless, there was some published information on nuclear genetic markers that would allow the characterization of flatworms to a species or genotype level, employing PCR-coupled sequencing and restriction fragment length polymorphism (RFLP) approaches (e.g., Adlard et al., 1993; Bowles and McManus, 1993; Bowles et al., 1993). Some of those studies employed the internal transcribed spacers (ITS) of nuclear ribosomal DNA as genetic markers, and promoted their use for systematic studies.

Genetics and Evolution of Infectious Diseases. https://doi.org/10.1016/B978-0-443-28818-0.00029-X
Copyright © 2024 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

However, at the time (1990–91), no work had been done on ITS in nematodes; essentially, these spacers were completely unknown.

In 1991, using a newly purchased Macintosh computer, linked to a modem, we gained first access to an early version of the GenBank database. As no DNA or RNA sequence data were available for any of the parasitic nematodes that we were studying, or planning to explore, we elected to search this early database for ITS sequences of the free-living rhabditid nematode, *Caenorhabditis elegans*, which was being extensively studied as a model organism for metazoans, and for which genetic data sets were already being deposited in GenBank. We succeeded in finding sequences for the first and second internal transcribed spacers (ITS-1 and ITS-2) and the three ribosomal rRNA genes (18S, 5.8S and 28S) of *C. elegans*, which was the "starting point" for the definition of species-specific genetic markers of strongylid nematodes.

Subsequently, we identified in GenBank a range of nuclear ribosomal rRNA gene sequences as well as ITS-1 and ITS-2 sequences from a range of eukaryotic organisms, and downloaded them for alignment and comparison. Our goal was to find conserved regions in the small subunit (*SSU* or 18S) and large subunit (*LSU* = 5.8S + 28S) genes of nuclear ribosomal DNA, flanking ITS-1 and ITS-2, respectively, for the design of oligonucleotide primers for subsequent use in PCR. At the time, our main focus was on ITS-2, because it was shorter in length and was more variable in sequence among the range of taxa that we had examined, and that the precursor RNA (pre-RNA) of ITS-2 is involved in regulating ribosome biogenesis (van der Sande et al., 1992; van Nues et al., 1995), which we thought might be unique to individual species.

This work led to the initial design of primers NC1 (forward; 5′-ACG TCT GGT TCA GGG TTG TT-3′) at the 3′-end of the 5.8S rRNA gene and NC2 (reverse; 5′-TTA GTT TCT TTT CCT CCG CT-3′) at the 5′ end of 28S rRNA gene—these NC primers were named after Neil Chilton, who was intimately involved in their selection and design. NC1 was not conserved across all taxa, but seemed to be shared by some chromadorean nematodes; on the other hand, NC2 was very conserved across animals and plants. Thus, we reasoned that NC1 might provide the specificity that we needed to amplify strongylid and perhaps other nematode groups, and that NC2 would provide us with an anchor-primer that would work on a very broad range of nematodes and, possibly, other helminths. Subsequently, we also identified other regions in the *SSU* and 5.8S gene for the design of oligonucleotide primers for the amplification and analysis of ITS-1 (Chilton, 2004).

In this chapter, (i) we describe the evaluation of these primers and the sequences of ITS regions that they amplify for the specific identification of parasitic nematodes by DNA sequencing and other methods; (ii) review the applications of PCR-coupled sequencing and associated analyses to investigate the taxonomy, phylogeny, and epidemiology of parasitic nematodes, and for the diagnosis of nematode infections. We also discuss (iv) the utility of mitochondrial genetic markers to complement systematic studies of nematodes conducted using ITS sequence data, and (vi) the exciting opportunities that genome-wide explorations offer for future, large-scale studies of the epidemiology, population genetics, and evolution of parasitic nematodes.

2. Critical Evaluation of Internal Transcribed Spacers of Nuclear Ribosomal DNA as Specific Markers for Parasitic Nematodes

In the early 1990s, we first set out to assess the utility of ITS-2 as a species marker for strongylid nematodes. To do this, we isolated total genomic DNAs from individual morphologically identified

adult male specimens of selected species, PCR-amplified ITS-2 from these DNA samples using primers NC1 and NC2, and then directly cycle-sequenced the resultant PCR products using each of these same primers. In this way, we were able to critically assess sequence variation within and among individuals of a species (e.g., Campbell et al., 1995; Gasser et al., 1993; Hoste et al., 1993, 1995; Stevenson et al., 1995, 1996). We also evaluated variation within and among individual nematode eggs (e.g., Gasser et al., 1993). Throughout this work, we consistently demonstrated low levels ($\leq$1%) of sequence variability among individual worms and eggs of the same species, with some microheterogeneity in sequence within individuals for some species, and higher levels (>1.3%) of sequence differences between species than within species (Gasser, 1999). Despite sequence microheterogeneity being detectable within individuals of some species (e.g., *Trichostrongylus vitrinus*), the overall "consensus" sequence within an individual was consistently representative of the species (Gasser, 1999, 2006).

We were intrigued by this microheterogeneity and sought an explanation for its occurrence. A review of key literature (Gasser, 1999) revealed that (tandem-repetitive) ribosomal DNA undergoes a process called "concerted evolution" (Elder and Turner, 1995), in which there is a strong tendency/pressure for such repeat elements to maintain sequence homogeneity within a species. Nonetheless, we observed that there were differences in the extent of sequence microheterogeneity in ITS-2 among species; in most cases, it was minor (cf. Gasser, 2006), but in lungworms, this heterogeneity was conspicuous (Conole et al., 1999, 2001; Johnson et al., 2004). Irrespective of this finding, ITS-2 sequences determined directly from PCR products (i.e., not via molecular cloning) were shown to be representative of species within the orders Strongylida and Ascaridida. In cases where intraindividual sequence heterogeneity occurred, we speculated that ITS-2 paralogs related to variation in microsatellite numbers and lengths and/or multiple loci on distinct chromosomes, although these proposals require verification for individual species. It is important to emphasize at this point that most evaluation of ITSs focused on ITS-2 of the orders Strongylida and Ascaridida, not on other nematode groups, and not on ITS-1, although evidence accumulated over the years that ITS-1 is useful as a species marker (Gasser, 2001, 2006, 2013).

For nematode groups other than these two orders, there has been no rigorous evaluation of ITS-1 or ITS-2 as species markers, although some researchers have (often incorrectly) assumed that these regions are useful or reliable for all members of the phylum Nematoda and have, thus, employed them for genetic analyses or systematic studies. Indeed, some studies have revealed complexity in ITS, reflected in sequence heterogeneity within some taxa, similar to that seen within *Metastrongylus* species (e.g., Conole et al., 1999, 2001). To attempt to circumvent this challenge, PCR products were often cloned and then sequenced, rather than being sequenced directly, leading to erroneous or incomplete sequence data sets being produced. In some published works, the use of sequence data (obtained via molecular cloning) in phylogenetic and/or other analyses will have led to erroneous results and likely misinterpretations.

In conclusion, our critical evaluations have shown that ITS-1 and ITS-2 are reliable species-specific markers for members of the orders Strongylida and Ascaridida (Chilton, 2004; Gasser, 2006, 2013; Gasser et al., 2005; Zhu et al., 2001b). While our focus was consistently on parasitic nematodes of animals including humans, other research groups employed ITS sequences for studying plant-parasitic nematodes (e.g., Powers et al., 1997; Powers, 2004). This work, initiated in the 1990s, laid the foundation for a wide range of molecular systematic and epidemiological studies of parasitic nematodes as well as diagnostic applications.

Currently, there are >800 articles in public databases describing the application of ITS to nematodes. The intent of this chapter is not to review all literature, but rather to provide a perspective on some of the progress made since 1990, cite relevant reviews and selected original articles, and look ahead at the new opportunities that arise because of the availability of advanced nucleic acid–based sequencing, diagnostic and analytical technologies, combined with major progress in the area of informatics. We provide a synthesis that combines past information and experiences with the latest advances.

3. Taxonomic Investigations

In the early 1990s, the advent of PCR and the availability of reliable genetic markers (i.e., the regions ITS-2 and subsequently, ITS-1) for the specific identification of nematodes of the orders Strongylida and Ascaridida provided a foundation to genetically characterize a range of socioeconomically important parasite species at any stage of development (egg, larva or adult) and to assess the genetic diversity among known morphospecies—using methods including conventional PCR-based sequencing, restriction fragment length polymorphism (RFLP) and/or mutation scanning, including single strand confirmation polymorphism (SSCP) and denaturing gel electrophoresis (DGGE) (Gasser, 1997, 1998, 2001, 2013; Gasser and Monti, 1997; Gasser et al., 1997, 1998a,c, 2002, 2006; Gasser and Zhu, 1999; Gasser and Chilton, 2001).

Our initial focus was on important strongylid nematodes of livestock animals (including sheep, cattle, and pigs) as well as companion animals (horses, dogs, and cats) and humans. Within the order Strongylida, gastrointestinal nematodes (Trichostrongyloidea and Strongyloidea) cause major production losses in ruminant livestock, including sheep, goats, and cattle. Initial studies characterized species of *Trichostrongylus*, *Haemonchus, Teladorsagia*, *Ostertagia*, *Cooperia*, *Nematodirus*, *Oesophagostomum*, and *Chabertia* (Gasser et al., 1993, 1994, 1996a,b, 1999a,c; Gasser and Hoste, 1995; Hoste et al., 1993, 1995, 1998; Newton et al., 1998a,b,c,d; Stevenson et al., 1995, 1996). For instance, Newton et al. (1998) studied genetic differences among *Nematodirus spathiger*, *Nematodirus filicollis*, *Nematodirus helvetianus*, and *Nematodirus battus* in ITS-2. Pairwise comparisons of ITS-2 sequences revealed differences ranging from 3.9% to 24.7%, and that *N. spathiger* and *N. helvetianus* were most similar genetically. *Nematodirus battus* was the most genetically distinct species, differing in sequence by 22.8%–24.7% with respect to the other three species. Some nucleotide differences among the four species related to distinct endonuclease restriction sites that could be used in a diagnostic PCR-coupled RFLP. This work supported a range of other studies of the taxonomy of trichostrongylid nematodes.

Nematodes of the subfamilies Strongylinae and Cyathostominae (order Strongylida) are of major clinical importance in equids. The widespread use of drugs against these parasites has led to problems of resistance within the Cyathostominae, and an increase in their prevalence and the intensity of cyathostomin infections (Herd, 1990; Klei and French, 1998; Kaplan, 2002; Matthews, 2014). Limitations in the understanding of fundamental biological aspects and the systematics of these parasites related largely to limitations in their specific identification using traditional, morphological approaches. From 1996 to 2000, a range of studies was undertaken, culminating in the characterization of 26 species by their ITS-1 and/or ITS-2 sequences (Hung et al., 1996, 1997, 1999a,c, 2000), including the identification of potential cryptic (i.e., genetically distinct but morphologically similar) species

(Hung et al., 1999a), which provided a foundation for their classification and diagnosis of infections, as well as epidemiological investigations (Chilton, 2004; Gasser et al., 2004).

Extending studies on the genetic characterization of *Oesophagostomum* species of pigs (Gasser et al., 1998b; Newton et al., 1997), extensive work was undertaken to characterise *Oesophagostomum bifurcum* from humans and related species from nonhuman primates in parts of Africa (Gasser et al., 1999a,b, 2006, 2009b; de Gruijter et al., 2002, 2004, 2006; Polderman et al., 2010; Romstad et al., 1997a; Schindler et al., 2005; Verweij et al., 2000, 2001) as well as a number of hookworms including *Necator americanus* from people, and *Ancylostoma caninum*, *Ancylostoma ceylanicum, Ancylostoma tubaeforme, Ancylostoma braziliense*, *Uncinaria stenocephala*, and *Bunostomum trigonocephalum* from respective canine, feline or ruminant host species (Chilton and Gasser, 1999; Gasser et al., 1996b, 1998a, 2008a, 2009a; de Gruijter et al., 2005; Monti et al., 1998; Romstad et al., 1998; Verweij et al., 2001).

Over the decades, much research has been focused on the very distinctive and diverse array of helminth parasites of fauna native to Australia, including marsupials and monotremes, particularly the strongyloid nematodes of kangaroos and wallabies (for a summary, see Beveridge and Smales, 2022). Although these parasites have no apparent adverse socioeconomic impact on their hosts, they have been neglected, in terms of research. As the body plans of these worms are simple, and morphological characterization does not reflect their true diversity, PCR-based sequencing and analyses of ITS-1 and/or ITS-2 have been useful to characterize particular morphotypes (species) (e.g., Chilton et al., 2002, 2016a; Fazenda et al., 2009, 2010; Beveridge et al., 2014) and to evaluate their genetic identity and diversity (e.g., Chilton et al., 1997b, 2009b; Gasser and Chilton, 2001; Huby-Chilton et al., 2001; Shuttleworth et al., 2014, 2016). A range of studies has shown extensive diversity in the nematodes of marsupials and hidden (cryptic) species have been commonly discovered (e.g., Chilton et al., 1995, 2009a, 2012, 2016b; Gasser et al., 2001; Huby-Chilton et al., 2002; Sukee et al., 2020c, 2021; Tan et al., 2012). In 2014, Beveridge and Gasser reviewed salient information in this area, discussed the mechanisms (co-speciation and host switching or colonization (cf. Paterson and Banks, 2001)), likely leading to extensive genetic diversity in nematodes (Strongylida), and recommended avenues for future research employing advanced molecular (genomic) and informatic approaches.

Members of the superfamily Ascaridoidea infect all major vertebrate groups (Bruce et al., 1994; Gibson, 1983; Hartwich, 1974; Sprent, 1983), some of which are of major socioeconomic significance. Representatives of the genus *Toxocara* can cause clinical disease and are transmissible from animals to humans (are zoonotic; Bowman, 2020; Ma et al., 2018). For instance, larvae of *Toxocara canis* can invade human tissues and induce neurotoxocarosis (including eosinophilic meningoencephalitis), ocular larva migrans (OLM), visceral larva migrans (VLM), and/or covert toxocariasis (CT) (Ma et al., 2018). Larvae of *Toxocara cati*, *Toxocara vitulorum*, and *Toxascaris leonina* can invade the tissues of animals, but the extent to which they are implicated in human disease has been unclear (Bowman, 2020; Holland and Smith, 2006). Nonetheless, *T. cati* is likely an underrecognized zoonotic worm (Bowman, 2020; Fisher, 2003; Maciag et al., 2022). Larvae of a range of ascaridoids can invade human tissues, induce pathological changes or disease and are extremely challenging to identify to genus or species using morphological criteria, given their small size, such that molecular tools are needed for diagnosis.

Initially, we focused on characterizing ITS-1 and ITS-2 sequences of species of *Toxocara* and related nematodes (e.g., Jacobs et al., 1997; Le et al., 2016; Li et al., 2006; Zhu et al., 1998a, 1999,

2000a,b, 2001a,b, 2002; Peng et al., 2003). Jacobs et al. (1997) first showed that adults of *T. canis*, *T. cati* and *Ts. leonina* could be identified and differentiated by their ITS-2 sequence. The sequence differences (26%−50%) between these species were markedly greater than the variation (0%−0.6%) within each species, indicating that ITS-2 provides specific genetic markers, and allowing the development of PCR assays for the differentiation of *T. cati* from *T. canis* (larvae) and/or other ascaridoids which can occur in human tissues (Wu and Bowman, 2020). This initial work showed the utility of ITS-1 and ITS-2 for specific identification, and provided an avenue for the detection of cryptic (or hidden) species. Zhu et al. (1998a) conducted a molecular investigation of *Toxocara* adults (called *Toxocara* sp. cf. *canis*) from cats in Malaysia which, superficially, appeared to be morphologically similar to *T. canis*—which is specific to canids. Genetic characterization of *Toxocara* sp. cf. *canis* revealed substantially higher differences (11%−24%) in the ITS-1 and ITS-2 sequences from *T. canis* and *T. cati* than within individual taxa, leading to the description of a new species, *T. malaysiensis*, by Gibbons et al. (2001).

This work paved the way for genetic studies of a broad range of ascaridoids, including members of the family Anisakidae, known for its taxonomic complexity and "hidden" species (cf. Mattiucci and Nascetti, 2008; Mattiucci et al., 2018; Shamsi, 2014). Early work characterized selected species of anisakids by their ITS-1 and/or ITS-2 sequences (Zhu et al., 1998b, 2000), allowing the specific identification and also the discovery of marked genetic variation within some taxa (Zhu et al., 2001a, 2002, 2007), leading to an expansion of the application of PCR-based SSCP and/or direct sequencing analysis to a wider range of anisakid nematodes (e.g., Du et al., 2010; Jabbar et al., 2012a,b; Shamsi et al., 2008, 2009, 2011a,b, 2012, 2013; Zhang et al., 2007, 2013).

4. Phylogenetic Investigations

Given that numerous taxonomic studies had collectively demonstrated the reliability and utility of ITS-1 and ITS-2 for the identification of species and the identification/characterization of sibling species (within the orders Strongylida and Ascaridida) (reviewed by Gasser et al., 1999, 2006; Gasser, 2013), there was confidence that these sequences would be useful for phylogenetic analyses to test hypotheses regarding the systematic relationships, originally formulated based on morphological features and biological information.

Obviously, the benefit of using ITS-1 and/or ITS-2 as markers was that these regions allowed specific identification, but, sometimes sequence difference between/among taxa from distinct families were high, making it challenging to achieve reliable sequence alignments for subsequent phylogenetic analyses. For this reason, Chilton et al. (1998) and Chilton and Gasser (1999) created secondary structure models to optimize positional homology in sequence alignments for meaningful phylogenetic analyses. Concurrently, Hung et al. (2000) used a similar approach to undertake a comprehensive phylogenetic investigation of equine strongylid nematodes. These authors used ITS-1 and ITS-2 data sets to reconstruct systematic relationships of 30 species of equine strongyles. The findings provided support for the hypothesis that the genera with large subglobular buccal capsules are ancestral to those with small cylindrical buccal capsules, but did not support the division of the subfamilies Strongylinae and Cyathostominae based on morphological features. These results provided a sound framework for expanded investigations of equine strongyles, supported through the use of improved, phylogenetically informative morphological data sets.

Extending this work, Chilton and colleagues undertook a range of phylogenetic studies of distinct groups of parasitic nematodes (order Strongylida) (e.g., Chilton et al., 1997a, 2001, 2011, 2015, 2016a, 2019; Sukee et al., 2020a). For instance, Chilton et al. (2019) explored the phylogenetic relationships of 42 species of nematodes from Australian macropodid marsupials (belonging to the tribes Zoniolaiminea, Macropostrongylinea, and Coronostrongylinea) using ITS-1 and ITS-2 sequence data. The findings showed that none of these tribes was monophyletic, and paraphyly was evident in three genera (*Wallabinema*, *Papillostrongylus*, and *Monilonema*). Species of key genera (*Zoniolaimus*, *Wallabinema*, *Popovastrongylus*, *Macropostrongylus*, and *Coronostrongylus*) were shown to occur in a number of host genera, whereas *Thallostonema* species were limited to hosts belonging to *Thylogale* spp. and *Notamacropus rufogriseus*. Based on these findings, as well as extensive studies of the related nematodes in the tribe Labiostrongylinea (see Chilton et al., 2011) and the genus *Cloacina* (Chilton et al., 2017), modes of evolution of these nematodes were proposed.

5. Diagnostic and Epidemiological Investigations

The accurate diagnosis of parasitic nematode infections of animals and humans underpins many areas, particularly the management and control of the diseases, to reduce the production losses that they cause, and investigations of parasite prevalence, geographical distribution, and transmission patterns. As some conventional coprological methods can have limitations, in terms of specificity and/or sensitivity (Gasser et al., 2008b; Gasser, 2013), there has been a need for improved molecular methods. The critical evaluation of the utility of ITS-1 and/or ITS-2 for the identification of species of strongylid and ascaridioid nematodes (Sections 2 and 3) provided a solid foundation for the development of specific and sensitive diagnostic PCR assays.

Given the importance of strongylid nematodes of livestock and problems linked to anthelmintic resistance, there was a significant focus on developing molecular-diagnostic assays using markers in ITS. Building on work conducted using conventional PCR, quantitative and robotic PCR-based platforms for high-throughput diagnosis were developed, with implications for epidemiological investigations and for use in control programs (reviewed by Gasser, 2013; Roeber et al., 2013a,b). For example, Roeber et al. (2011) critically assessed the performance of a qPCR method (Bott et al., 2009), established for the diagnosis of gastrointestinal nematode infections in sheep, using well-defined control samples. This assessment revealed that the qPCR assay achieved high diagnostic sensitivity (98%) and specificity (100%), and test results were in "good" agreement (Kappa = 0.95) with coprological testing, although complementary evidence showed that the qPCR assay performed better than quantitative fecal flotation. This molecular assay was employed to estimate the prevalences of key parasites in in a winter rainfall region of Australia, and revealed that the nematodes *Teladorsagia circumcincta* (80%) and *Trichostrongylus* spp. (66%) were most prevalent, followed by *Chabertia ovina* (34%), *Oesophagostomum venulosum* (29%), and *H. contortus* (1%). The conclusion from this study was that this qPCR tool had the potential to replace conventional methods of fecal flotation and larval culture (Roeber et al., 2012a). A subsequent study showed that this qPCR assay could be used effectively to support the diagnosis of anthelmintic resistance in nematodes, in conjunction with conventional fecal egg count reduction testing (FECRT).

Subsequently, Roeber et al. (2012b) developed an enhanced diagnostic platform for the routine testing of large numbers of samples at relatively low cost and with limited manual intervention. This

multiplexed-tandem PCR (MT-PCR) was critically assessed for its performance relative to qPCR as well as conventional larval culture using fecal samples from different flocks of sheep from a broad geographical range across Australia, and achieved favorable diagnostic specificity (i.e., 87.5%) and sensitivity (100%). This platform was then validated in different environments and contexts in Europe (Scotland and Belgium), and was demonstrated to perform well as a routine high throughput diagnostic assay for gastrointestinal nematodes of sheep, with potential to be applied to a range of other parasites, host species (such as alpacas) and/or geographic regions (e.g., Gebrekidan et al., 2015, 2018; Perera et al., 2015a,b; Rashid et al., 2018a,b, 2019; Roeber et al., 2015, 2017; Stracke et al., 2019).

Other PCR assays using specific markers in ITS have been developed for parasitic nematodes of humans and companion animals (including dogs or cats) which can be reservoirs for transmission of (e.g., hookworm) infections to humans (Azzopardi et al., 2021; Campbell et al., 2016, 2017a,b; Clarke et al., 2018; Colella et al., 2021; Hii et al., 2018; Kamani et al., 2021; Massetti et al., 2020, 2021, 2022; Mulinge et al., 2019; Papaiakovou et al., 2017, 2021, 2022; Vaz Nery et al., 2019; Zendejas-Heredia et al., 2021). For instance, Azzopardi et al. (2021) established an efficient qPCR workflow system for the detection of six clinically important soil-transmitted helminths, including the hookworms, *Ascaris* sp., *Ancylostoma ceylanicum, Ancylostoma duodenale*, and *Necator americanus* as well as *Trichuris trichiura* and *Strongyloides stercoralis*. For the latter four species, the assay relies on markers in ITS-1 or ITS-2, and this assay can be readily implemented in any qPCR-capable laboratory. Another example is the study by Massetti et al. (2020) who established and evaluated high-throughput multiplex qPCR assays for the detection and surveillance of zoonotic species of canine hookworms including *Ancylostoma braziliense*, *Ancylostoma ceylanicum*, and *Ancylostoma caninum*. These qPCR assays were developed to support large epidemiological surveillance investigations, evaluate the role of canids in the transmission of hookworms to people and to assess the efficacy of anthelminthic treatment programs in canid populations.

6. Nemabiome Investigations

The exquisite specificity of ITS-2 for the molecular identification of strongylid nematode species was the foundation for the development of a deep sequencing approach to explore and/or quantitate the species composition of nematode populations ("nemabiome"; https://www.nemabiome.ca/; cf. Avramenko et al., 2015), akin to that used to investigate the make-up of bacterial populations (microbiome) employing the small subunit of the rRNA gene (e.g., Thompson et al., 2017). The major strength or advantage is that the oligonucleotide primer NC1 is entirely specific for strongylid nematode species, such that amplicons produced do not contain background amplicons originating from fecal or environmental contaminants (e.g., microbes, plants, or host).

The principle of the nemabiome method is that ITS-2 is PCR-amplified from genomic DNAs isolated directly from fecal or environmental samples containing strongylid nematodes, or pooled nematodes, and barcoded amplicons are then subjected to deep sequencing (e.g., Illumina) (Avramenko et al., 2015). The ITS-2 sequence data obtained are then mapped and/or compared to reference data sets in an ITS-2 database (Workentine et al., 2020), which has been continually improved, expanded, and curated over time.

Avramenko et al. (2015) were the first to optimize the nemabiome methodology and to conduct a nemabiome investigation of cattle in Canada, to characterize and quantitate the species composition of

parasitic nematode communities. Logically extending this work, Avramenko et al. (2017) studied the species diversity and impact of anthelmintics on nematodes in beef calves. A conspicuous feature was low treatment efficacy and an overall increase in the proportion of *Cooperia* species relative to *Ostertagia ostertagi* following treatment, demonstrating the utility of nemabiome investigations for the surveillance of treatment success/failure. Subsequently, Avramenko et al. (2018) explored the composition of strongylid nematode communities within and between commercial and conservation bison herds (58 production groups) in Western Canada. They discovered markedly higher infection intensities and parasite species diversity in commercial bison herds than those reported previously in beef cattle herds grazing in similar latitudes, with *Os. ostertagi* being the most abundant species, followed by *Cooperia oncophora*, consistent with clinical and production losses seen in bison herds. Unexpectedly, *Haemonchus placei* was identified as the most abundant species in five of the production groups; this blood-feeding nematode is a pathogenic parasite, not previously reported at such high abundance in any livestock species in such northerly latitudes; also the abundance of *Trichostrongylus axei* was unusually high with reference to cattle.

Overall, these and numerous earlier studies (reviewed by Chilton, 2004; Gasser, 2006; Gasser et al., 2008b; Beveridge and Gasser, 2014) demonstrate clearly that comprehensive qualitative and quantitative insights can be gained of species composition of strongylid parasite communities in various species of domesticated and wild animals, and that the "nemabiome" concept and methodology have broad and important implications for studying nematode populations in animals including humans. Such studies could include explorations of host−parasite and parasite−parasite relationships in live animals, parasite epidemiology (prevalence and distribution), parasite ecology, and the response of parasite populations to anthelmintic treatment and control campaigns.

Facilitated by the specificity of the NC1 primer and the exquisite amplification efficiency of the ITS-2 region from strongylid nematodes and the analytical sensitivity of PCR, the main focus of nemabiome investigations has been on strongylid nematodes, but there is essentially no limit to broadening such studies to other nematode or parasite groups (more generally) if suitable primer sets are available and have been assessed for specific amplification from genomic DNA of a particular species or group(s) of species. For example, for ascaridoid nematodes, oligonucleotide primers NC13 (forward; 5′-ATC GAT GAA GAA CGC AGC-3′) and NC2 (reverse; 5′-TTA GTT TCT TTT CCT CCG CT-3′) are usually well-suited (cf. Gasser, 1999; Zhu and Gasser, 1998; Zhu et al., 2000a,b, 2001a,b). The key to the application of the deep sequencing approach for nemabiome investigations is the relative conservation and consistent, reproducible performance of primers (for amplification and/or sequencing) across target species within particular orders, superfamilies and/or families, depending on the research aim(s) or biological question(s) being addressed. It is recommended that time and effort is spent on rigorously evaluating the suitability and performance of primers and primer sets for particular parasite groups, the reagents and protocols, the informatic pipeline, and then the performance of the entire workflow to be used, prior to embarking on large projects.

7. Considerations and Extensions

Studies have consistently demonstrated that, particularly, ITS-2 provides robust species markers for strongylid and ascaridoid nematodes, which has been exceptionally useful for both fundamental and applied investigations using PCR-based tools (reviewed by Chilton, 2004; Gasser, 2006, 2013). The

main advantages have been that this rDNA region is short (usually 250 to 450 bp), is tandemly repetitive and undergoes concerted evolution to achieve sequence homogenization within species; because of its repetitive nature, ITS can be readily amplified from minute amounts of template (usually from as little as 0.1−1.0 pg of DNA) isolated directly from nematode specimens (or parts thereof), or from fecal samples. Thus, the analytical sensitivity of PCR is exceptionally high. The forward primer that was originally designed for ITS-2 of strongylid nematodes (NC1) provided high specificity for nematodes of the order Strongylida, with the exception of some metastrongylids, in which case, the primer NC5 (forward; 5′-GTA GGT GAA CCT GCG GAA GGA TCA TT-3′) or NC13 might each be used together with the conserved anchor primer NC2 (reverse). On the other hand, primers NC13 and NC2 are useful for members of the Ascaridida, provided that DNA samples are derived directly from the nematode (or parts thereof) without heterologous host DNA, as, like NC2, NC13 is quite conserved among taxa from different kingdoms of life, potentially allowing for some background amplification from host, microbial and/or environmental contaminants.

Advantages of ITS sequences are that they are relatively informative (in terms of interspecific differences), and, thus, useful for phylogenies of related species, groups, or families of nematodes, but not for large phylogenies (Gaucher et al., 2010) of genetically diverse taxa, as frequently the nucleotide sequences among species from distinct families or superfamilies can be too divergent to allow reliable sequence alignment/s, even employing secondary structure models (Chilton et al., 1998; Hung et al., 1999b) to guide alignments.

As indicated, sequence differences in ITS-1 and ITS-2 sequences can be too variable among species to allow comprehensive phylogenetic analyses across families, superfamilies, and orders, such that more conserved, phylogenetically informative DNA regions have been used. For comprehensive phylogenies (e.g., across superfamilies), the challenge is to have sufficient nucleotide variation across a region that does not vary markedly in length (i.e., limited insertion/deletion events, with enough phylogenetic signal over relatively long sequences) to resolve both shallow and deep branches. In some cases, large (*LSU*) and/or small (*SSU*) subunits of ribosomal RNA genes have been useful (Chilton et al., 2003, 2006, 2015); for others, mitochondrial data sets have been informative and effective in phylogenetic and population genetic analyses.

8. Mitochondrial Genetic Markers

Mitochondrial gene markers have been utilized for population genetic investigations of a range of parasitic nematode species (reviewed by Hu et al., 2004; Hu and Gasser, 2006; Jex et al., 2010a,b, 2015), because they likely evolve rapidly and, thus, display higher levels of sequence variation within species than do ITS-1 and/or ITS-2, although the mode of inheritance of mitochondrial markers in nematodes is unclear (i.e., whether inheritance is simply maternal, or maternal with some paternal leakage; cf. Hoolahan et al., 2011), potentially raising questions about the validity of genetic structures inferred. However, mitochondrial genomic data sets, and particularly the proteomes inferred from them, are considered useful for assessing phylogenetic/genetic relationships. The use of amino acid (instead of nucleotide) sequences overcomes the challenge associated with (the proposed) excessive mutation rates in mitochondrial DNA of nematodes, as only nonsynonymous base changes lead to amino acid changes, and the proteins inferred from all genes encoded in the mitochondrial genome would be expected to be representative of a species.

Using this rationale, we employed mitochondrial sequence data sets to study genetic and phylogenetic relationships. Initially, a long PCR-based approach, combining Sanger sequencing, assembly and annotation—with or without molecular cloning—was used (Gasser et al., 2012a,b; Hu and Gasser, 2006; Hu et al., 2002, 2003a,b,c, 2004, 2007; Liu et al., 2016), followed by high-throughput sequencing, assembly, and comparative analyses (Jex et al., 2008a,b, 2009, 2010a,b, 2015; Jabbar et al., 2013a,b,c, 2014a,b; Mohandas et al., 2014a,b).

Although such methods have been widely used, we were astonished to learn that the mitochondrial genomes of some helminth species (reported to be complete) were actually incomplete. This incompleteness related to the absence of noncoding regions. To illustrate this point, recent work of selected parasitic flatworm species (Kinkar et al., 2019, 2020, 2021) showed that Nanopore-based sequencing can discover long noncoding elements that consistently remain undetected by conventional sequencing methods, likely explained by assembly problems across repetitive regions. Thus, caution is warranted when concluding that a mitochondrial genome assembled from short-read (e.g., Ilumina) data is complete.

Despite this challenge, proteomic data inferred from mitochondrial genomes have been suitable for systematic studies and also to infer phylogenetic/genetic relationships of nematodes (Hu and Gasser, 2006). This approach has been particularly useful when ITS or other nuclear ribosomal DNA regions exhibit sequence heterogeneity within a species or genotype (e.g., *Trichinella*). In one study, for example, we used concatenated amino acid sequence data sets inferred from mitochondrial genomes—assembled from Illumina-derived, short-read data sets—to construct a robust phylogeny for all 12 currently recognized *Trichinella* species and genotypes (Mohandas et al., 2014b), which was essentially consistent with that inferred subsequently using nuclear genomic data sets (Zarlenga et al., 2006; Korhonen et al., 2016). A similar mitogenomic sequencing−informatic approach has been applied recently to investigate other groups of parasitic nematodes from Australian macropodid and vombatid marsupials (e.g., Sukee et al., 2020b, 2021), demonstrating clearly its broader applicability.

9. Technological Advances and Prospects

The majority of studies of parasitic nematodes has relied on the use of a small number of nuclear ribosomal and mitochondrial DNA regions for PCR-based analyses (Sections 1−8). Most investigations have employed the short ITS-1 and ITS-2 regions in nuclear DNA (each usually <500 bp), and some have used parts of the mitochondrial cytochrome *c* oxidase subunit 1 (*cox*-1) and/or NADH dehydrogenase 1 (*nad*-1) genes (<450 bp). The choice of these genetic markers has depended mainly on the species or group of nematodes under investigation, and the taxonomic, phylogenetic, or population genetic question being tackled/addressed.

The mitochondrial and nuclear genomes are unknown for the vast majority of helminths. Clearly, improvements to genomic sequencing (e.g., Illumina and Nanopore) and bioinformatic technologies provide unique prospects for large-scale investigations of their diversity using genome-wide data sets. On one hand, it is possible to sequence mitochondrial genomes (15−20 kb) from genomic DNA from single worms, and workflows have been constructed that allow rapid in silico assembly, annotation, and comparative analyses of mitochondrial genomes (e.g., Jex et al., 2010b; Mohandas et al., 2014b). On the other hand, nuclear genomes (50−800 Mb) can now be sequenced from genomic DNA from single worms with or without whole-genome amplification. As genomic sequencing is now

commonplace and accessible as a commercial service, with no investments required in equipment, the main bottleneck, at this stage, relates to bioinformatic processing and analyses of data sets.

Clearly, the use of advanced sequencing technologies has led to a revolution in genomics and an explosion in the number of genome-wide sequence data sets for a wide range of organisms including nematodes. The major expansion in genomic and associated proteomic and transcriptomic data sets for socioeconomically important parasitic nematodes of animals and plants (Kikuchi et al., 2017; Ma et al., 2020; Wit and Gilleard, 2017) has meant that the focus is now turning to large-scale analyses of such data sets, which requires novel, efficient, and reliable tools that deliver repeatable/reproducible results. Although a number of bioinformatic tools, such as CloudMap (Minevich et al., 2012), Churchill (Kelly et al., 2015), TOGGLE (Monat et al., 2015), and Sarek (Garcia et al., 2020), have been developed to handle and process large data sets and to carry out genetic analyses of them, there has been a delay in the development of specialized workflows and pipelines for genetic and statistical analyses.

For most parasitic nematodes, nuclear draft genome assemblies are fragmented, nucleotide variability within species (variome) is not known, and the specimens or samples used for sequencing can contain extraneous DNA (e.g., host DNA or potentially microbial contamination). For these reasons, we recently created a workflow, called *Escalibur*, for the *de novo*-analysis of nucleotide variation in eukaryotic species (Korhonen et al., 2016). This workflow has the following key steps: (i) trimming and mapping of data; (ii) selection of a reference genome; (iii) removal of contaminating reads; (iv) recalibration of base calls; and (v) variant-calling. This pipeline employs the well-established GATK variant caller (https://gatk.broadinstitute.org/hc/en-us) and workflow description language, and is both scalable and efficient at identifying nucleotide variability in individual worms on a genome-wide scale. Thus, this and similar informatic workflows are likely to accelerate the analyses of genetic variation and assessments of genetic diversity within and among populations of particular species.

10. Concluding Remarks

This chapter shows the significant impact that advanced molecular tools (employing suitably validated and specific genetic markers) have had, and continue to have, in various subdisciplines of nematodology, including taxonomy, phylogeny, and epidemiology, and for the diagnosis of nematode infections/diseases. Looking ahead, the massive, recent expansion and improvement of genomic, transcriptomic, proteomic and metabolomic technologies, combined with substantially enhanced biological and biotechnological information, databases, and resources as well as bioinformatic tools, will provide tremendous opportunities to support the multidisciplinary efforts of scientists to understand the fascinating "world of worms."

Acknowledgments

We are grateful to the many colleagues, staff, and students who contributed to original research referred to in the present chapter, and to the funding bodies (Australian Research Council, the National Health and Medical Research Council of Australia and Wellcome Trust) and industry partners (including Melbourne Water, Yourgene Biosciences, and Yourgene Health) who have supported our research programs over the past decades.

References

Avramenko, R.W., Bras, A., Redman, E.M., Woodbury, M.R., Wagner, B., Shury, T., Liccioli, S., Windeyer, M.C., Gilleard, J.S., 2018. High species diversity of trichostrongyle parasite communities within and between Western Canadian commercial and conservation bison herds revealed by nemabiome metabarcoding. Parasites Vectors 11, 299.

Avramenko, R.W., Redman, E.M., Lewis, R., Bichuette, M.A., Palmeira, B.M., Yazwinski, T.A., Gilleard, J.S., 2017. The use of nemabiome metabarcoding to explore gastro-intestinal nematode species diversity and anthelmintic treatment effectiveness in beef calves. Int. J. Parasitol. 47, 893−902.

Avramenko, R.W., Redman, E.M., Lewis, R., Yazwinski, T.A., Wasmuth, J.D., Gilleard, J.S., 2015. Exploring the gastrointestinal "nemabiome": deep amplicon sequencing to quantify the species composition of parasitic nematode communities. PLoS One 10, e0143559.

Adlard, R.D., Barker, S.C., Blair, D., Cribb, T.H., 1993. Comparison of the second internal transcribed spacer (ribosomal DNA) from populations and species of Fasciolidae (Digenea). Int. J. Parasitol. 23, 423−425.

Andrews, R.H., Chilton, N.B., 1999. Multilocus enzyme electrophoresis: a valuable technique for providing answers to problems in parasite systematics. Int. J. Parasitol. 29, 213−253.

Azzopardi, K.I., Hardy, M., Baker, C., Bonnici, R., Llewellyn, S., McCarthy, J.S., Traub, R.J., Steer, A.C., 2021. Detection of six soil-transmitted helminths in human stool by qPCR - a systematic workflow. PLoS One 16, e0258039.

Beveridge, I., Gasser, R.B., 2014. Diversity in parasitic helminths of Australasian marsupials and monotremes: a molecular perspective. Int. J. Parasitol. 44, 859−864.

Beveridge, I., Smales, R.L., 2022. Review of the cloacininae stossich (Nemata: Strongyloidea) from Australasian marsupials (Marsupialia: Macropodoidea). Man (Lond.): J. Paras. Biodiv. 19, 1−197.

Beveridge, I., Nguyen, H., Nyein, S., Cheng, A., Koehler, A.V., Shuttleworth, M.E., Gasser, R.B., Jabbar, A., 2014. Description of *Cloacina atthis* sp. nov. from the stomach of the euro (*Macropus robustus*) (Marsupialia: Macropodidae) from Western Australia based on morphological and molecular criteria. Parasitol. Res. 113, 3485−3493.

Bott, N.J., Campbell, B.E., Beveridge, I., Chilton, N.B., Rees, D., Hunt, P.W., Gasser, R.B., 2009. A combined microscopic-molecular method for the diagnosis of strongylid infections in sheep. Int. J. Parasitol. 39, 1277−1287.

Bowles, J., McManus, D.P., 1993. Rapid discrimination of *Echinococcus* species and strains using a polymerase chain reaction-based RFLP method. Mol. Biochem. Parasitol. 57, 231−240.

Bowles, J., Hope, M., Wu, T., Liu, X., McManus, D.P., 1993. Nuclear and mitochondrial genetic markers highly conserved between Chinese and Philippine *Schistosoma japonicum*. Acta Trop. 55, 217−229.

Bowman, D.D., 2020. *Toxocara* and toxocariasis. In: Advances in Parasitology, vol. 109, pp. 2−883. Hardcover ISBN: 9780128209585; eBook ISBN: 9780128209592.

Bruce, N.L., Adlard, R.D., Cannon, L.R.G., 1994. Synoptic checklist of ascaridoid parasites (Nematoda) from fish hosts. Invertebr. Taxon. 8, 583−674.

Campbell, A.J.D., Gasser, R.B., Chilton, N.B., 1995. Differences in a ribosomal sequence of *Strongylus* species allows identification of single eggs. Int. J. Parasitol. 25, 359−365.

Campbell, S.J., Vaz Nery, S., D'Este, C.A., Gray, D.J., McCarthy, J.S., Traub, R.J., Andrews, R.M., Llewellyn, S., Vallely, A.J., Williams, G.M., Amaral, S., Clements, A.C., 2016. Water, sanitation and hygiene related risk factors for soil-transmitted helminth and *Giardia duodenalis* infections in rural communities in Timor-Leste. Int. J. Parasitol. 46, 771−779.

Campbell, S.J., Vaz Nery, S., D'Este, C.A., Gray, D.J., McCarthy, J.S., Traub, R.J., Andrews, R.M., Llewellyn, S., Vallely, A.J., Williams, G.M., Clements, A.C., 2017a. Investigations into the association between soil-transmitted helminth infections, haemoglobin and child development indices in Manufahi District, Timor-Leste. Parasites Vectors 10, 192.

Campbell, S.J., Vaz Nery, S., Wardell, R., D'Este, C.A., Gray, D.J., McCarthy, J.S., Traub, R.J., Andrews, R.M., Llewellyn, S., Vallely, A.J., Williams, G.M., Clements, A.C., 2017b. Water, sanitation and hygiene (WASH) and environmental risk factors for soil-transmitted helminth intensity of infection in Timor-Leste, using real time PCR. PLoS Neglected Trop. Dis. 11, e0005393.

Chilton, N.B., 2004. The use of nuclear ribosomal DNA markers for the identification of bursate nematodes (order Strongylida) and for the diagnosis of infections. Anim. Health Res. Rev. 5, 173−187.

Chilton, N.B., Gasser, R.B., Beveridge, I., 1995. Differences in a ribosomal DNA sequence of morphologically indistinguishable species within the *Hypodontus macropi* complex (Nematoda: Strongyloidea). Int. J. Parasitol. 25, 647−651.

Chilton, N.B., Gasser, R.B., Beveridge, I., 1997a. Phylogenetic relationships of Australian strongyloid nematodes inferred from ribosomal DNA sequence data. Int. J. Parasitol. 27, 1481−1494.

Chilton, N.B., Beveridge, I., Hoste, H., Gasser, R.B., 1997b. Evidence for hybridisation between *Paramacropostrongylus iugalis* and *P. typicus* (Nematoda: Strongyloidea) in grey kangaroos, *Macropus fuliginosus* and *M. giganteus* in eastern Australia. Int. J. Parasitol. 27, 475−482.

Chilton, N.B., Hoste, H., Newton, L.A., Beveridge, I., Gasser, R.B., 1998. Common secondary structures for the second internal transcribed spacer pre-rRNA of two subfamilies of trichostrongylid nematodes. Int. J. Parasitol. 28, 1765−1773.

Chilton, N.B., Gasser, R.B., 1999. Sequence differences in the internal transcribed spacer rDNA among four species of hookworm (Ancylostomatoidea: *Ancylostoma*). Int. J. Parasitol. 29, 1971−1977.

Chilton, N.B., Newton, L.A., Beveridge, I., Gasser, R.B., 2001. Evolutionary relationships of trichostrongyloid nematodes (Strongylida) inferred from ribosomal DNA sequence data. Mol. Phylogenet. Evol. 19, 367−386.

Chilton, N.B., Huby-Chilton, F., Gasser, R.B., Beveridge, I., 2002. Review of *Papillostrongylus* Johnston & Mawson, 1939 (Nematoda: Strongyloidea) from wallabies and kangaroos (Marsupialia: Macropodidae) using morphological and molecular techniques, with the description of *P. barbatus* n. sp. Syst. Parasitol. 51, 81−93.

Chilton, N.B., Huby-Chilton, F., Gasser, R.B., 2003. First complete large subunit rRNA sequence and secondary structure for a parasitic nematode - phylogenetic and diagnostic implications. Mol. Cell. Probes 17, 33−39.

Chilton, N.B., Huby-Chilton, F., Gasser, R.B., Beveridge, I., 2006. The evolutionary origins of nematodes within the order Strongylida are related to predilection sites within hosts. Mol. Phylogenet. Evol. 40, 118−128.

Chilton, N.B., Huby-Chilton, F., Johnson, P.M., Beveridge, I., Gasser, R.B., 2009a. Genetic variation within species of the nematode genus *Cloacina* (Strongyloidea: cloacininae) parasitic in the stomachs of rock wallabies, *Petrogale* spp. (Marsupialia: macropodidae) in Queensland and implications for host phylogeny. Aust. J. Zool. 57, 1−10.

Chilton, N.B., Huby-Chilton, F., Smales, L.R., Gasser, R.B., Beveridge, I., 2009b. Genetic divergence between island and continental populations of the parasitic nematode *Labiosimplex australis* in Australia. Parasitol. Res. 104, 229−236.

Chilton, N.B., Huby-Chilton, F., Beveridge, I., Smales, L.R., Gasser, R.B., Andrews, R.H., 2011. Phylogenetic relationships of species within the tribe Labiostrongylinea (Nematoda: cloacinidae) based on ribosomal DNA spacer sequence data. Parasitol. Int. 60, 381−387.

Chilton, N.B., Jabbar, A., Huby-Chilton, F., Jex, A.R., Gasser, R.B., Beveridge, I., 2012. Genetic variation within the *Hypodontus macropi* (Nematoda: Strongyloidea) complex from macropodid hosts in Australia. Electrophoresis 33, 3544−3554.

Chilton, N.B., Huby-Chilton, F., Koehler, A.V., Gasser, R.B., Beveridge, I., 2015. The phylogenetic relationships of endemic Australasian trichostrongylin families (Nematoda: Strongylida) parasitic in marsupials and monotremes. Parasitol. Res. 114, 3665−3673.

Chilton, N.B., Huby-Chilton, F., Koehler, A.V., Gasser, R.B., Beveridge, I., 2016a. Phylogenetic relationships of species of the oesophageal parasitic nematode genera *Cyclostrongylus* and *Spirostrongylus* (Strongyloidea:

chabertiidae: Cloacininae) with their wallaby hosts (Marsupialia: macropodidae). Mol. Cell. Probes 30, 93–99.

Chilton, N.B., Huby-Chilton, F., Koehler, A.V., Gasser, R.B., Beveridge, I., 2016b. Detection of cryptic species of *Rugopharynx* (Nematoda: Strongylida) from the stomachs of Australian macropodid marsupials. Int. J. Parasitol. Parasites Wildl. 5, 124–133.

Chilton, N.B., Shuttleworth, M.A., Huby-Chilton, F., Koehler, A.V., Jabbar, A., Gasser, R.B., Beveridge, I., 2017. Speciation within the genus *Cloacina* (Nematoda: Strongylida) from macropodid marsupials: species flocks and intra-host speciation. Parasitology 144, 1828–1840.

Chilton, N.B., Huby-Chilton, F., Koehler, A.V., Gasser, R.B., Beveridge, I., 2019. Phylogenetic relationships of three tribes of cloacinine nematodes (Strongylida: chabertiidae) from macropodid marsupials. J. Helminthol. 93, 486–493.

Clarke, N.E., Llewellyn, S., Traub, R.J., McCarthy, J., Richardson, A., Vaz Nery, S., 2018. Quantitative polymerase chain reaction for diagnosis of soil-transmitted helminth infections: a comparison with a flotation-based technique and an investigation of variability in DNA detection. Am. J. Trop. Med. Hyg. 99, 1033–1040.

Colella, V., Khieu, V., Worsley, A., Senevirathna, D., Muth, S., Huy, R., Odermatt, P., Traub, R.J., 2021. Risk profiling and efficacy of albendazole against the hookworms *Necator americanus* and *Ancylostoma ceylanicum* in Cambodia to support control programs in Southeast Asia and the Western Pacific. Lancet Reg. Health West Pac. 16, 100258.

Conole, J., Chilton, N.B., Jarvis, T., Gasser, R.B., 1999. Intraspecific and interspecific variation in the second internal transcribed spacer sequence for *Metastrongylus* (Nematoda: metastrongyloidea) detected by high resolution PCR-RFLP. Int. J. Parasitol. 29, 1941–1947.

Conole, J., Chilton, N.B., Jarvis, T., Gasser, R.B., 2001. Mutation scanning analysis of microsatellite variability in the ITS-2 (precursor rRNA) for three species of *Metastrongylus* (Nematoda: metastrongyloidea). Parasitology 122, 195–206.

de Gruijter, J.M., Polderman, A.M., Zhu, X.Q., Gasser, R.B., 2002. Screening for haplotypic variability within *Oesophagostomum bifurcum* (Nematoda) employing a single-strand conformation polymorphism approach. Mol. Cell. Probes 16, 185–190.

de Gruijter, J.M., Ziem, J., Verweij, J.J., Polderman, A.M., Gasser, R.B., 2004. Genetic substructuring within *Oesophagostomum bifurcum* (Nematoda) from human and non-human primates from Ghana based on random amplification of polymorphic DNA analysis. Am. J. Trop. Med. Hyg. 71, 227–233.

de Gruijter, J.M., van Lieshout, L., Gasser, R.B., Verweij, J.J., Brienen, E.A.T., Ziem, J.B., Yelifari, L., Polderman, A.M., 2005. PCR-based differential diagnosis of *Ancylostoma duodenale* and *Necator americanus* in humans from northern Ghana. Trop. Med. Int. Health 10, 574–580.

de Gruijter, J.M., Blotkamp, J., Gasser, R.B., Amponsah, S., Polderman, A.M., 2006. Morphological variability within *Oesophagostomum bifurcum* among different primate species from Ghana. J. Helminthol. 80, 357–361.

Du, C., Zhang, L., Shi, M., Ming, Z., Hu, M., Gasser, R.B., 2010. Elucidating the identity of *Anisakis* larvae from a broad range of marine fishes from the Yellow Sea, China, using a combined electrophoretic-sequencing approach. Electrophoresis 31, 654–658.

Elder, J.F., Turner, B.J., 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. QRB (Q. Rev. Biol.) 70, 297–320.

Fazenda, I.P., Beveridge, I., Chilton, N.B., Jex, A.R., Pangasa, A., Campbell, B.E., Huby-Chilton, F., Carvalho, L.M., Gasser, R.B., 2009. Analysis of genetic variation in *Globocephaloides* populations from macropodid marsupials using a mutation scanning-based approach. Electrophoresis 30, 2758–2764.

Fazenda, I.P., Gasser, R.B., Carvalho, L.M., Beveridge, I., 2010. Resurrection and redescription of *Globocephaloides wallabiae* Johnston et Mawson, 1939 (Nematoda, Trichostrongyloidea) from macropodid marsupials in north-eastern Australia. Acta Parasitol. 55, 138–143.

Fisher, M., 2003. *Toxocara cati:* an underestimated zoonotic agent. Trends Parasitol. 19, 167–170.

Garcia, M., Juhos, S., Larsson, M., Olason, P.I., Martin, M., Eisfeldt, J., DiLorenzo, S., Sandgren, J., Díaz De Ståhl, T., Ewels, P., Wirta, V., Nistér, M., Käller, M., Nystedt, B., 2020. Sarek: a portable workflow for whole-genome sequencing analysis of germline and somatic variants. F1000research 9, 63.

Gasser, R.B., 1997. Mutation scanning methods for the analysis of parasite genes. Int. J. Parasitol. 27, 1449−1463.

Gasser, R.B., 1998. What's in that band? Int. J. Parasitol. 28, 989−996.

Gasser, R.B., 1999. PCR-based technology in veterinary parasitology. Vet. Parasitol. 84, 229−258.

Gasser, R.B., 2006. Molecular tools − advances, opportunities and prospects. Vet. Parasitol. 136, 69−89.

Gasser, R.B., 2013. A perfect time to harness advanced advanced molecular technologies to explore *Toxocara*. Vet. Parasitol. 193, 353−364.

Gasser, R.B., Chilton, N.B., Hoste, H., Beveridge, I., 1993. Rapid sequencing of rDNA from single worms and eggs of parasitic helminths. Nucleic Acids Res. 21, 2525−2526.

Gasser, R.B., Chilton, N.B., Hoste, H., Stevenson, L.A., 1994. Species identification of trichostrongyle nematodes by PCR-linked RFLP. Int. J. Parasitol. 24, 291−293.

Gasser, R.B., Hoste, H., 1995. Genetic markers for closely-related parasitic nematodes. Mol. Cell. Probes 9, 315−320.

Gasser, R.B., Stevenson, L.A., Chilton, N.B., Nansen, P., Bucknell, D.G., Beveridge, I., 1996a. Species markers for equine strongyles detected in intergenic spacer rDNA by PCR-RFLP. Mol. Cell. Probes 10, 371−378.

Gasser, R.B., Stewart, L.E., Speare, R., 1996b. Genetic markers in ribosomal DNA for hookworm identification. Acta Trop. 62, 15−21.

Gasser, R.B., Monti, J.R., 1997. Identification of parasitic nematodes by PCR-SSCP. Mol. Cell. Probes 11, 201−209.

Gasser, R.B., Monti, J.R., Zhu, X.Q., Chilton, N.B., Hung, G.C., Guldberg, P., 1997. Polymerase chain reaction-linked single-strand conformation polymorphism of ribosomal DNA to fingerprint parasites. Electrophoresis 18, 1564−1566.

Gasser, R.B., Zhu, X.Q., 1999. Sequence-based analysis of DNA fragments by mutation detection techniques. Parasitol. Today 15, 462−465.

Gasser, R.B., 2001. Identification of parasitic nematodes and study of genetic variability using PCR approaches. (Chapter 4. In: Kennedy, M.W., Harnett, W. (Eds.), Parasitic Nematodes: Molecular Biology, Biochemistry and Immunology. CAB Publishers, pp. 53−82.

Gasser, R.B., Monti, J.R., Qian, B.-Z., Polderman, A.M., Nansen, P., Chilton, N.B., 1998a. A mutation scanning approach for the identification of hookworm species and analysis of population variation. Mol. Biochem. Parasitol. 92, 303−312.

Gasser, R.B., Woods, W.G., Bjørn, H., 1998b. Distinguishing *Oesophagostomum dentatum* from *Oesophagostomum quadrispinulatum* developmental stages by a single-strand conformation polymorphism method. Int. J. Parasitol. 28, 1903−1909.

Gasser, R.B., Zhu, X.Q., Chilton, N.B., Newton, L.A., Nedergaard, T., Guldberg, P., 1998c. Analysis of sequence homogenisation in rDNA arrays of *Haemonchus contortus* by DGGE. Electrophoresis 19, 2391−2395.

Gasser, R.B., Woods, W.G., Huffman, M.A., Blotkamp, J., Polderman, A.M., 1999a. Molecular separation of *Oesophagostomum stephanostomum* and *Oesophagostomum bifurcum* (Nematoda: Strongyloidea) from non-human primates. Int. J. Parasitol. 29, 1087−1091.

Gasser, R.B., Woods, W.G., Blotkamp, J., Verweij, J., Storey, P.A., Polderman, A.M., 1999b. Screening for nucleotide variations in ribosomal DNA arrays of *Oesophagostomum bifurcum* by PCR-coupled single-strand conformation polymorphism. Electrophoresis 20, 1486−1491.

Gasser, R.B., Rossi, L., Zhu, X.Q., 1999c. Identification of *Nematodirus* species (Nematoda: molineidae) from wild ruminants in Italy using genetic markers in ribosomal DNA. Int. J. Parasitol. 29, 1809−1817.

Gasser, R.B., Zhu, X.Q., Beveridge, I., Chilton, N.B., 2001. Mutation scanning analysis of sequence heterogeneity in the second internal transcribed spacer (rDNA) within some members of the *Hypodontus macropi* (Nematoda: Strongyloidea) complex. Electrophoresis 22, 1076−1085.

Gasser, R.B., Chilton, N.B., 2001. Applications of single-strand conformation polymorphism (SSCP) to taxonomy, diagnosis, population genetics and molecular evolution of parasitic nematodes. Vet. Parasitol. 101, 201−213.

Gasser, R.B., Zhu, X.Q., Chilton, N.B., 2002. The value of mutation scanning approaches for detecting genetic variation - implications for studying intestinal nematodes of humans. (Chapter 13. Series Editors. In: Holland, C.V., Kennedy, M.W. (Eds.), World Class Parasites. In: Black, S., Seed, J.R. (Eds.), The Geohelminths: *Ascaris, Trichuris* and Hookworm, vol. 2. Kluwer Academic Press, Boston, pp. 219−233.

Gasser, R.B., Hung, G.-C., Chilton, N.B., Beveridge, I., 2004. Advances in developing molecular-diagnostic tools for strongyloid nematodes of equids: fundamental and applied implications. Mol. Cell. Probes 18, 3−16.

Gasser, R.B., Zhu, X.Q., Hu, M., Jacobs, D.E., Chilton, N.B., 2005. Molecular genetic characterisation of members of the genus *Toxocara* (Nematoda: Ascaridoidea) − systematic, population genetic and epidemiological considerations. CABI book. In: Holland, C., Smith, H. (Eds.), *Toxocara*: The Enigmatic Parasite. CABI Press, Wallingford, UK, pp. 18−31.

Gasser, R.B., de Gruijter, J.M., Polderman, A.M., 2006a. Insights into the epidemiology and genetic make-up of *Oesophagostomum bifurcum* from human and non-human primates using molecular tools. Parasitology 132, 453−460.

Gasser, R.B., Hu, M., Chilton, N.B., Campbell, B.E., Jex, A.J., Otranto, D., Cafarchia, C., Beveridge, I., Zhu, X.Q., 2006b. Single-strand conformation polymorphism (SSCP) for the analysis of genetic variation. Nat. Protoc. 1, 3121−3128.

Gasser, R.B., Cantacessi, C., Loukas, A., 2008a. DNA technological progress toward advanced diagnostic tools to support human hookworm control. Biotechnol. Adv. 26, 35−45.

Gasser, R.B., Bott, N.J., Chilton, N.B., Hunt, P.R., Beveridge, I., 2008b. Toward practical, DNA-based diagnostic methods for parasitic nematodes of livestock − bionomic and biotechnological implications. Biotechnol. Adv. 26, 325−334.

Gasser, R.B., Cantacessi, C., Campbell, B.E., 2009a. Improved molecular-diagnostic tools for human hookworms. Expert Rev. Mol. Diagn. 9, 17−21.

Gasser, R.B., de Gruijter, J.M., Polderman, A.M., 2009b. The utility of molecular methods for elucidating primate-pathogen relationships − the *Oesophagostomum bifurcum* example. In: Huffman, M.A., Chapman, C.A. (Eds.), Primate Parasite Ecology: The Dynamics and Study of Host-Parasite Relationships, Cambridge Studies in Biological and Evolutionary Anthropology, vol. 13. Cambridge University Press, Cambridge.

Gasser, R.B., Jabbar, A., Mohandas, N., Höglund, J., Hall, R.S., Littlewood, D.T., Jex, A.R., 2012a. Assessment of the genetic relationship between *Dictyocaulus* species from *Bos taurus* and *Cervus elaphus* using complete mitochondrial genomic datasets. Parasites Vectors 5, 241.

Gasser, R.B., Jabbar, A., Mohandas, N., Schnyder, M., Deplazes, P., Littlewood, D.T., Jex, A.R., 2012b. Mitochondrial genome of *Angiostrongylus vasorum*: comparison with congeners and implications for studying the population genetics and epidemiology of this parasite. Infect. Genet. Evol. 12, 1884−1891.

Gaucher, E.A., Kratzer, J.T., Randall, R.N., 2010. Deep phylogeny–how a tree can help characterize early life on Earth. Cold Spring Harbor Perspect. Biol. 2, a002238.

Gebrekidan, H., Gasser, R.B., Perera, P.K., McGrath, S., McGrath, S., Stevenson, M.A., Jabbar, A., 2015. Investigating the first outbreak of oriental theileriosis in cattle in South Australia using multiplexed tandem PCR (MT-PCR). Ticks Tick Borne Dis 6, 574−578.

Gebrekidan, H., Gasser, R.B., Stevenson, M.A., Jabbar, A., 2018. Multiplexed tandem PCR (MT-PCR) assay using the major piroplasm surface protein gene for the diagnosis of *Theileria orientalis* infection in cattle. J. Clin. Microbiol. 56, 10−1128.

Gibbons, L.M., Jacobs, D.E., Sani, R.A., 2001. *Toxocara malayensis* n. sp. (Nematoda: Ascaridoidea) from the domestic cat (*Felis catus* Linnaeus, 1758). J. Parasitol. 87, 660−665.

Gibson, D.I., 1983. The systematics of ascaridoid nematodes — a current assessment. In: Stone, A.R., Platt, H.M., Khalil, L.F. (Eds.), Concepts in Nematode Systematics. Systematics Association Special Volume No. 22. Academic Press, London, pp. 321−338.

Hartwich, G., 1974. Keys to genera of the Ascaridoidea. In: Anderson, R.C., Chabaud, A.G., Willmott, S. (Eds.), CIH Keys to the Nematode Parasites of Vertebrates. No. 2. Commonwealth Agricultural Bureaux. Farnham Royal, UK, pp. 1−15.

Herd, R.P., 1990. The changing world of worms: the rise of the cyathostomes and the decline of *Strongylus vulgaris*. Compend. Continuing Educ. Pract. Vet. 12, 732−734.

Hii, S.F., Senevirathna, D., Llewellyn, S., Inpankaew, T., Odermatt, P., Khieu, V., Muth, S., McCarthy, J., Traub, R.J., 2018. Development and evaluation of a multiplex quantitative real-time polymerase chain reaction for hookworm species in human stool. Am. J. Trop. Med. Hyg. 99, 1186−1193.

Holland, C.V., Smith, H.V., 2006. *Toxocara:* The Enigmatic Parasite. CABI Publishing − CAB International, Wallingford, p. 302.

Hoolahan, A.H., Blok, V.C., Gibson, T., Dowton, M., 2011. Paternal leakage of mitochondrial DNA in experimental crosses of populations of the potato cyst nematode *Globodera pallida*. Genetica 139, 1509−1519.

Hoste, H., Gasser, R.B., Chilton, N.B., Mallet, S., Beveridge, I., 1993. Lack of intraspecific variation in the second internal transcribed spacer (ITS-2) of *Trichostrongylus colubriformis* ribosomal DNA. Int. J. Parasitol. 23, 1069−1071.

Hoste, H., Chilton, N.B., Gasser, R.B., Beveridge, I., 1995. Differences in the second internal transcribed spacer (ribosomal DNA) between five species of *Trichostrongylus* (Nematoda: trichostrongylidae). Int. J. Parasitol. 25, 75−80.

Hoste, H., Chilton, N.B., Beveridge, I., Gasser, R.B., 1998. A comparison of the first internal transcribed spacer of ribosomal DNA in seven species of *Trichostrongylus* (Nematoda: trichostrongylidae). Int. J. Parasitol. 28, 1251−1260.

Hu, M., Chilton, N.B., Gasser, R.B., 2002. Long PCR-based amplification and sequencing of the entire mitochondrial genome from parasitic nematodes. Mol. Cell. Probes 16, 261−267.

Hu, M., Gasser, R.B., Abs EL-Osta, Y.G., Chilton, N.B., 2003a. Structure and organisation of the mitochondrial genome of *Dirofilaria immiti*s. Parasitology 127, 37−51.

Hu, M., Chilton, N.B., Abs EL-Osta, Y.G., Gasser, R.B., 2003b. Comparative analysis of mitochondrial genome data for *Necator americanus* from two endemic regions reveals substantial genetic variation. Int. J. Parasitol. 33, 955−963.

Hu, M., Chilton, N.B., Gasser, R.B., 2003c. The mitochondrial genome of *Strongyloides stercoralis* (Rhabditida) − idiosyncratic gene order, and evolutionary implications. Int. J. Parasitol. 33, 1393−1408.

Hu, M., Chilton, N.B., Gasser, R.B., 2004. The mitochondrial genomics of parasitic nematodes of socio-economic importance: recent progress, and implications for population genetics and systematics studies. Adv. Parasitol. 56, 133−212.

Hu, M., Gasser, R.B., 2006. Mitochondrial genomes of parasitic nematodes − progress and perspectives. Trends Parasitol. 22, 78−84.

Hu, M., Jex, A.R., Campbell, B.E., Gasser, R.B., 2007. Long PCR amplification of the entire mitochondrial genome from individual helminths for direct sequencing. Nat. Protoc. 2, 2339−2344.

Huby-Chilton, F., Beveridge, I., Gasser, R.B., Chilton, N.B., 2001. Single-strand conformation polymorphism analysis of genetic variation in *Labiostrongylus longispicularis* from kangaroos. Electrophoresis 22, 1925−1929.

Huby-Chilton, F., Beveridge, I., Gasser, R.B., Chilton, N.B., 2002. Redescription of *Zoniolaimus mawsonae* Beveridge, 1983 (Nematoda: Strongyloidea) and the description of *Z. latebrosus* n. sp. from the red kangaroo,

Macropus rufus (Marsupialia: macropodidae) based on morphological and molecular data. Syst. Parasitol. 51, 135−147.

Hung, G.C., Jacobs, D.E., Krecek, R.S., Gasser, R.B., Chilton, N.B., 1996. *Strongylus asini* (Nematoda, Strongyloidea): genetic relationships with other *Strongylus* species determined by ribosomal DNA. Int. J. Parasitol. 26, 1407−1411.

Hung, G.C., Chilton, N.B., Beveridge, I., McDonnell, A., Lichtenfels, J.R., Gasser, R.B., 1997. Molecular delineation of *Cylicocyclus nassatus* and *C. ashworthi* (Nematoda: strongylidae). Int. J. Parasitol. 27, 601−605.

Hung, G.C., Chilton, N.B., Beveridge, I., Gasser, R.B., 1999a. Molecular evidence for cryptic species within *Cylicostephanus minutus* (Nematoda: strongylidae). Int. J. Parasitol. 29, 285−291.

Hung, G.C., Chilton, N.B., Beveridge, I., Gasser, R.B., 1999b. Secondary structure model for the ITS-2 precursor rRNA of strongyloid nematodes of equids: implications for phylogenetic inference. Int. J. Parasitol. 29, 1949−1964.

Hung, G.C., Gasser, R.B., Beveridge, I., Chilton, N.B., 1999c. Species-specific amplification by PCR of ribosomal DNA from some equine strongyles. Parasitology 119, 69−80.

Hung, G.C., Chilton, N.B., Beveridge, I., Gasser, R.B., 2000. A molecular systematic framework for equine strongyles based on DNA sequence data. Int. J. Parasitol. 30, 95−103.

Jabbar, A., Abdouslam, A., Norbury, L.J., Eisenbarth, A., Shamsi, S., Gasser, R.B., Lopata, A.L., Beveridge, I., 2012a. Larval anisakid nematodes from teleost fishes from lizard island, northern great barrier reef, Australia. Mar. Freshw. Res. 63, 1283−1299.

Jabbar, A., Khoon, A.T., Hui, T.X., Schaeffner, B.C., Jex, A.R., Nolan, M.J., Lopata, A., Gasser, R.B., Beveridge, I., 2012b. Mutation scanning-based analysis of anisakid larvae from *Sillago flindersi* from Bass Strait, Australia. Electrophoresis 33, 499−505.

Jabbar, A., Jex, A.R., Mohandas, N., Hall, R.S., Littlewood, D.T., Gasser, R.B., 2013a. The mitochondrial genome of *Aelurostrongylus abstrusus* − diagnostic, epidemiological and systematic implications. Gene 516, 294−300.

Jabbar, A., Beveridge, I., Mohandas, N., Chilton, N.B., Littlewood, D.T., Jex, A.R., Gasser, R.B., 2013b. Analyses of mitochondrial amino acid sequence datasets support the proposal that specimens of *Hypodontus macropi* from three species of macropodid hosts represent distinct species. BMC Evol. Biol. 13, 259.

Jabbar, A., Mohandas, N., Jex, A.R., Gasser, R.B., 2013c. The mitochondrial genome of *Protostrongylus rufescens* − implications for population and systematic studies. Parasites Vectors 6, 263.

Jabbar, A., Littlewood, D.T.J., Mohandas, N., Briscoe, A.G., Foster, P.G., Müller, F., von Samson-Himmelstjerna, G., Jex, A.R., Gasser, R.B., 2014a. The mitochondrial genome of *Parascaris univalens* - implications for a "forgotten" parasite. Parasites Vectors 7, 428.

Jabbar, A., Mohandas, N., Gasser, R.B., 2014b. Characterisation of the mitochondrial genome of *Parafilaroides normani* (lungworm) of *Arctocephalus pusillus doriferus* (Australian Fur seal). Parasitol. Res. 113, 3049−3055.

Jacobs, D.E., Zhu, X.Q., Gasser, R.B., Chilton, N.B., 1997. PCR-based methods for identification of potentially zoonotic ascaridoid parasites of the dog, fox and cat. Acta Trop. 68, 191−200.

Jex, A.R., Hu, M., Littlewood, D.T.J., Waeschenbach, A., Gasser, R.B., 2008a. Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). BMC Genom. 9, 11.

Jex, A.R., Waeschenbach, A., Littlewood, D.T.J., Hu, M., Gasser, R.B., 2008b. The mitochondrial genome of *Toxocara canis*. PLoS Neglected Trop. Dis. 2, e273.

Jex, A.R., Waeschenbach, A., Hu, M., van Wyk, J.A., Beveridge, I., Littlewood, D.T.J., Gasser, R.B., 2009. The mitochondrial genomes of *Ancylostoma caninum* and *Bunostomum phlebotomum* − two hookworms of animal health and zoonotic importance. BMC Genom. 10, 1−10.

Jex, A.R., Hall, R.S., Littlewood, D.T.J., Gasser, R.B., 2010a. An integrated pipeline for next-generation sequencing and annotation of mitochondrial genomes. Nucleic Acids Res. 38, 522−533.

Jex, A.R., Littlewood, D.T.J., Gasser, R.B., 2010b. Toward next-generation sequencing of mitochondrial genomes− focus on parasitic worms of animals and biotechnological implications. Biotechnol. Adv. 28, 151−159.

Jex, A.R., Littlewood, D.T.J., Gasser, R.B., 2015. Sequencing and annotation of mitochondrial genomes from individual parasitic helminths. Methods Mol. Biol. 1201, 51−63.

Johnson, M., Abs EL-Osta, Y.G., Hu, M., Gasser, R.B., 2004. An electrophoretic tool for the genetic characterisation and delineation of lungworms. Mol. Cell. Probes 197−203.

Kamani, J., Massetti, L., Olubade, T., Balami, J.A., Samdi, K.M., Traub, R.J., Colella, V., González-Miguel, J., 2021. Canine gastrointestinal parasites as a potential source of zoonotic infections in Nigeria: a nationwide survey. Prev. Vet. Med. 192, 105385.

Kaplan, R.M., 2002. Anthelmintic resistance in nematodes of horses. Vet. Res. 33, 491−507.

Kelly, B.J., Fitch, J.R., Hu, Y., Corsmeier, D.J., Zhong, H., Wetzel, A.N., Nordquist, R.D., Newsom, D.L., White, P., 2015. Churchill: an ultra-fast, deterministic, highly scalable and balanced parallelization strategy for the discovery of human genetic variation in clinical and population-scale genomics. Genome Biol. 16, 1−14.

Kikuchi, T., Eves-van den Akker, S., Jones, J.T., 2017. Genome evolution of plant-parasitic nematodes. Annu. Rev. Phytopathol. 55, 333−354.

Kinkar, L., Korhonen, P.K., Cai, H., Gauci, C.G., Lightowlers, M.W., Saarma, U., Jenkins, D.J., Li, J., Young, N.D., Gasser, R.B., 2019. Long-read sequencing reveals a 4.4 kb tandem repeat region in the mitogenome of *Echinococcus granulosus* (*sensu stricto*) genotype G1. Parasites Vectors 12, 238.

Kinkar, L., Young, N.D., Sohn, W.-M., Stroehlein, A.J., Korhonen, P.K., Gasser, R.B., 2020. First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach. PLoS Neglected Trop. Dis. 14, e0008552.

Kinkar, L., Gasser, R.B., Webster, B.L., Rollinson, D., Littlewood, D.T.J., Stroehlein, A.J., Korhonen, P.K., Young, N.D., 2021. Nanopore sequencing resolves elusive long tandem-repeat regions in mitochondrial genomes. Int. J. Mol. Sci. 22, 1811.

Klei, T.R., French, D.D., 1998. Small strongyles: an emerging parasite problem for horses. Equine Pract. 20, 26−30.

Korhonen, P.K., Pozio, E., La Rosa, G., Chang, B., Koehler, A.V., Jex, A.R., Hofmann, A., Boag, P.R., Sternberg, P.W., Young, N.D., Gasser, R.B., 2016. Phylogenomic and biogeographic reconstruction of the *Trichinella* complex. Nat. Commun. 7, 10513.

Le, T.H., Nguyen, T.L.A., Nguyen, K.T., Nguyen, N.T.B., Thi, D., Thuy, T., Gasser, R.B., 2016. *Toxocara malaysiensis* infection in domestic cats in Vietnam - an emerging zoonotic issue? Infect. Genet. Evol. 37, 94−98.

Li, M.W., Zhu, X.Q., Gasser, R.B., Lin, R.Q., Sani, R.A., Lun, Z.R., Jacobs, D.E., 2006. The occurrence of *Toxocara malaysiensis* in cats in China, confirmed by sequence-based analyses of ribosomal DNA. Parasitol. Res. 99, 554−557.

Liu, G.H., Nadler, S.A., Liu, S.S., Zhao, Y., Podolska, M., D'Amelio, S., Gasser, R.B., Zhu, X.Q., 2016. Mitochondrial phylogenomics yields strongly supported hypotheses for ascaridomorph nematodes. Sci. Rep. 6, 39248.

Ma, G., Holland, C.V., Wang, T., Hofmann, A., Fan, C.-K., Maizels, R.M., Hotez, P.J., Gasser, R.B., 2018. Human toxocariasis. Lancet Infect. Dis. 18, e14−e24.

Ma, G., Wang, T., Korhonen, P.K., Hofmann, A., Sternberg, P.W., Young, N.D., Gasser, R.B., 2020. Elucidating the molecular and developmental biology of parasitic nematodes: moving to a multiomics paradigm. Adv. Parasitol. 108, 175−229.

Maciag, L., Morgan, E.R., Holland, C., 2022. *Toxocara:* time to let *cati* "out of the bag". Trends Parasitol. 38, 280−289.

Massetti, L., Colella, V., Zendejas, P.A., Ng-Nguyen, D., Harriott, L., Marwedel, L., Wiethoelter, A., Traub, R.J., 2020. High-throughput multiplex qPCRs for the surveillance of zoonotic species of canine hookworms. PLoS Neglected Trop. Dis. 14, e0008392.

Massetti, L., Kamani, J., Wiethoelter, A., McDonagh, P., Colella, V., Traub, R.J., 2021. Field application of a novel multiplex qPCR assay reveals the occurrence of the zoonotic hookworm Ancylostoma braziliense in Nigerian dogs. Acta Trop. 213, 105758.

Massetti, L., Wiethoelter, A., McDonagh, P., Rae, L., Marwedel, L., Beugnet, F., Colella, V., Traub, R.J., 2022. Faecal prevalence, distribution and risk factors associated with canine soil-transmitted helminths contaminating urban parks across Australia. Int. J. Parasitol. 22. S0020-7519(22)00105-9 (ahead of print).

Matthews, J.B., 2014. Anthelmintic resistance in equine nematodes. Int. J. Parasitol. Drugs Drug Resist. 4, 310−315.

Mattiucci, S., Nascetti, G., 2008. Advances and trends in the molecular systematics of anisakid nematodes, with implications for their evolutionary ecology and host−parasite co-evolutionary processes. Adv. Parasitol. 66, 47−148.

Mattiucci, S., Cipriani, P., Levsen, A., Paoletti, M., Nascetti, G., 2018. Molecular epidemiology of *Anisakis* and anisakiasis: an ecological and evolutionary road map. Adv. Parasitol. 99, 93−263.

Minevich, G., Park, D.S., Blankenberg, D., Poole, R.J., Hobert, O., 2012. CloudMap: a cloud-based pipeline for analysis of mutant genome sequences. Genetics 192, 1249−1269.

Mohandas, N., Jabbar, A., Podolska, M., Zhu, X.Q., Littlewood, D.T., Jex, A.R., Gasser, R.B., 2014a. Mitochondrial genomes of *Anisakis simplex* and *Contracaecum osculatum* (*sensu stricto*) - comparisons with selected chromadorean nematodes. Infect. Genet. Evol. 21, 452−462.

Mohandas, N., Pozio, E., La Rosa, G., Korhonen, P.K., Young, N.D., Koehler, A.V., Hall, R.S., Sternberg, P.W., Boag, P.R., Jex, A.R., Chang, B., Gasser, R.B., 2014b. Mitochondrial genomes of *Trichinella* species and genotypes - a basis for diagnosis, and systematic and epidemiological explorations. Int. J. Parasitol. 44, 1073−1080.

Monat, C., Tranchant-Dubreuil, C., Kougbeadjo, A., Farcy, C., Ortega-Abboud, E., Amanzougarene, S., Ravel, S., Agbessi, M., Orjuela-Bouniol, J., Summo, M., Sabot, F., 2015. TOGGLE: toolbox for generic NGS analyses. BMC Bioinf. 16, 1−8.

Monti, J.R., Chilton, N.B., Qian, B.-Z., Gasser, R.B., 1998. Specific amplification of *Necator americanus* or *Ancylostoma duodenale* DNA by PCR using markers in ITS-1 rDNA, and its implications. Mol. Cell. Probes 12, 71−78.

Mulinge, E., Njenga, S.M., Odongo, D., Magambo, J., Zeyhle, E., Mbae, C., Kagendo, D., Kanyi, H., Traub, R.J., Wassermann, M., Kern, P., Romig, T., 2019. J. Helminthol. 94, e43.

Newton, L.A., Chilton, N.B., Monti, J.R., Bjørn, H., Varady, M., Christensen, C.M., Gasser, R.B., 1997. Rapid PCR-based delineation of the porcine nodular worms, *Oesophagostomum dentatum* and *O. quadrispinulatum*. Mol. Cell. Probes 11, 149−153.

Newton, L.A., Chilton, N.B., Beveridge, I., Hoste, H., Nansen, P., Gasser, R.B., 1998a. Genetic markers for strongylid nematodes of livestock defined by PCR-based restriction analysis of spacer rDNA. Acta Trop. 69, 1−15.

Newton, L.A., Chilton, N.B., Beveridge, I., Gasser, R.B., 1998b. Genetic evidence indicating that *Cooperia surnabada* and *Cooperia oncophora* are one species. Int. J. Parasitol. 28, 331−336.

Newton, L.A., Chilton, N.B., Beveridge, I., Gasser, R.B., 1998c. Differences in the second internal transcribed spacer of four *Nematodirus* species (Nematoda: molineidae). Int. J. Parasitol. 28, 337−341.

Newton, L.A., Chilton, N.B., Beveridge, I., Gasser, R.B., 1998d. Systematic relationships of some members of the genera *Oesophagostomum* and *Chabertia* (Nematoda: chabertiidae) based on ribosomal DNA sequence data. Int. J. Parasitol. 28, 1781−1789.

Papaiakovou, M., Pilotte, N., Grant, J.R., Traub, R.J., Llewellyn, S., McCarthy, J.S., Krolewiecki, A.J., Cimino, R., Mejia, R., Williams, S.A., 2017. A novel, species-specific, real-time PCR assay for the detection of the emerging zoonotic parasite Ancylostoma ceylanicum in human stool. PLoS Neglected Trop. Dis. 11, e0005734.

Papaiakovou, M., Littlewood, D.T.J., Gasser, R.B., Anderson, R.M., 2021. How qPCR complements the WHO roadmap (2021-2030) for soil-transmitted helminths. Trends Parasitol. 37, 698−708.

Papaiakovou, M., Littlewood, D.T.J., Doyle, S.R., Gasser, R.B., Cantacessi, C., 2022. Worms and bugs of the gut − the search for diagnostic signatures using barcoding and metagenomics. Parasites Vectors 15, 118.

Paterson, A.M., Banks, J., 2001. Analytical approaches to measuring cospeciation of host and parasites: through a glass darkly. Int. J. Parasitol. 31, 1012−1022.

Peng, W., Yuan, K., Zhou, X., Hu, M., Gasser, R.B., 2003. Molecular-epidemiological study of *Ascaris* genotypes in China based on single strand conformation polymorphism analysis of ribosomal DNA. Electrophoresis 24, 2308−2315.

Perera, P.K., Gasser, R.B., Firestone, S.M., Smith, L., Roeber, F., Jabbar, A., 2015a. Semi-quantitative multiplexed-tandem PCR for the detection and differentiation of four *Theileria orientalis* genotypes in cattle. J. Clin. Microbiol. 53, 79−87.

Perera, P.K., Gasser, R.B., Read, E., Malmo, J., Nguyen, H., Nyein, S., Cheng, C., Jex, A.R., Rawlin, G., Spithill, T.W., Jabbar, A., 2015b. Use of multiplexed tandem PCR to estimate the prevalence and intensity of *Theileria orientalis* infections in cattle. Infect. Genet. Evol. 28 (32), 68−73.

Polderman, A.M., Eberhard, M., Baeta, S., Gasser, R.B., van Lieshout, L., Magnussen, P., Olsen, A., Spannbrucker, N., Ziem, J., Horton, J., 2010. The rise and fall of human oesophagostomiasis. Adv. Parasitol. 71, 93−155.

Powers, T.O., Todd, T.C., Burnell, A.M., Murray, P.C., Fleming, C.C., Szalanski, A.L., Adams, B.A., Harris, T.S., 1997. The rDNA internal transcribed spacer region as a taxonomic marker for nematodes. J. Nematol. 29, 441−450.

Powers, T., 2004. Nematode molecular diagnostics: from bands to barcodes. Annu. Rev. Phytopathol. 42, 367−383.

Rashid, M.H., Gebrekidan, H., Jabbar, A., 2018a. Multiplexed-tandem PCR (MT-PCR) assay to detect and differentiate gastrointestinal nematodes of alpacas. Parasites Vectors 11, 370.

Rashid, M.H., Vaughan, J.L., Stevenson, M.A., Campbell, A.J.D., Beveridge, I., Jabbar, A., 2018b. Anthelmintic resistance in gastrointestinal nematodes of alpacas (*Vicugna pacos*) in Australia. Parasites Vectors 11, 388.

Rashid, M.H., Beveridge, I., Vaughan, J.L., Jabbar, A., 2019. Worm burdens and associated histopathological changes caused by gastrointestinal nematodes in alpacas from Australia. Parasitol. Res. 118, 1031−1038.

Roeber, F., Jex, A.R., Campbell, A.J., Campbell, B.E., Anderson, G.A., Gasser, R.B., 2011. Evaluation and application of a molecular method to assess the composition of strongylid nematode populations in sheep with naturally acquired infections. Infect. Genet. Evol. 11, 849−854.

Roeber, F., Larsen, J.W., Anderson, N., Campbell, A.J., Anderson, G.A., Gasser, R.B., Jex, A.R., 2012a. A molecular diagnostic tool to replace larval culture in conventional faecal egg count reduction testing in sheep. PLoS One 7, e37327.

Roeber, F., Jex, A.R., Campbell, A.J., Nielsen, R., Anderson, G.A., Stanley, K.K., Gasser, R.B., 2012b. Establishment of a robotic, high-throughput platform for the specific diagnosis of gastrointestinal nematode infections in sheep. Int. J. Parasitol. 42, 1151−1158.

Roeber, F., Jex, A.R., Gasser, R.B., 2013a. Advances in the diagnosis of key gastrointestinal nematode infections of livestock, with an emphasis on small ruminants. Biotechnol. Adv. 31, 1135−1152.

Roeber, F., Jex, A.R., Gasser, R.B., 2013b. Next-generation molecular-diagnostic tools for gastrointestinal nematodes of livestock, with an emphasis on small ruminants – a turning point? Adv. Parasitol. 83, 267–333.

Roeber, F., Jex, A.R., Gasser, R.B., 2015. A real-time PCR for the diagnosis of gastrointestinal nematode infections of small ruminants. Methods Mol. Biol. 1247, 145–152.

Roeber, F., Morrison, A., Casaert, S., Smith, L., Claerebout, E., Skuce, P., 2017. Multiplexed-tandem PCR for the specific diagnosis of gastrointestinal nematode infections in sheep: an European validation study. Parasites Vectors 10, 226.

Romstad, A., Gasser, R.B., Nansen, P., Polderman, A.M., Monti, J.R., Chilton, N.B., 1997a. Characterization of *Oesophagostomum bifurcum* and *Necator americanus* by PCR-RFLP of rDNA. J. Parasitol. 83, 963–966.

Romstad, A., Gasser, R.B., Nansen, P., Polderman, A.M., Chilton, N.B., 1998. *Necator americanus* (Strongylida: ancylostomatidae) from Africa and Malaysia have different ITS-2 rDNA sequences. Int. J. Parasitol. 28, 611–615.

Saiki, R.K., Scharf, S.J., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., Arnheim, N., 1985. Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230, 1350–1354.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A., 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239, 487–491.

van der Sande, C.A., Kwa, M., van Nues, R.W., van Heerikhuizen, H., Raué, H.A., Planta, R.J., 1992. Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. J. Mol. Biol. 223, 899–910.

Schindler, A.R., de Gruijter, J.M., Polderman, A.M., Gasser, R.B., 2005. Definition of genetic markers in nuclear ribosomal DNA for a neglected parasite of primates, *Ternidens deminutus* (Nematoda: Strongylida) – diagnostic and epidemiological implications. Parasitology 131, 539–546.

Shamsi, S., 2014. Recent advances in our knowledge of Australian anisakid nematodes. Int. J. Parasitol. Parasites Wildl. 3, 178–187.

Shamsi, S., Eisenbarth, A., Saptarshi, S., Beveridge, I., Gasser, R.B., Lopata, A.L., 2011a. Occurrence and abundance of anisakid nematode larvae in five species of fish from southern Australian waters. Parasitol. Res. 108, 927–934.

Shamsi, S., Gasser, R.B., Beveridge, I., 2011b. Mutation scanning-coupled sequencing of nuclear ribosomal DNA spacers as a tool for the specific identification of different *Contracaecum* (Nematoda: Anisakidae) larval types. Mol. Cell. Probes 25, 13–18.

Shamsi, S., Gasser, R.B., Beveridge, I., 2012. Genetic characterisation and taxonomy of species of *Anisakis* (Nematoda : Anisakidae) parasitic in Australian marine mammals. Invertebr. Systemat. 26, 204–212.

Shamsi, S., Gasser, R.B., Beveridge, I., 2013. Description and genetic characterisation of *Hysterothylacium* (Nematoda: raphidascarididae) larvae parasitic in Australian marine fishes. Parasitol. Int. 62, 320–328.

Shamsi, S., Gasser, R.B., Beveridge, I., Shabani, A.A., 2008. *Contracaecum pyripapillatum* n. sp. (Nematoda: Anisakidae) and a description of *C. multipapillatum* (von Drasche, 1882) from the Australian pelican, *Pelecanus conspicillatus*. Parasitol. Res. 103, 1031–1039.

Shamsi, S., Norman, R., Gasser, R.B., Beveridge, I., 2009. Genetic and morphological evidences for the existence of sibling species within *Contracaecum rudolphi* (Hatwich, 1964) (Nematoda: Anisakidae) in Australia. Parasitol. Res. 105, 529–538.

Shuttleworth, M.A., Beveridge, I., Chilton, N.B., Koehler, A.V., Gasser, R.B., Jabbar, A., 2014. Genetic variation within and among species of *Cloacina* (Strongyloidea: cloacinine) from the swamp wallaby, *Wallabia bicolor* (marsupialia: macropodidae). Infect. Genet. Evol. 28, 261–269.

Shuttleworth, M.A., Beveridge, I., Koehler, A.V., Gasser, R.B., Jabbar, A., 2016. Molecular characterization of species of *Cloacina* (Strongyloidea: cloacinine) from the common wallaroo, *Macropus robustus* (marsupialia: macropodidae) in Australia. Infect. Genet. Evol. 44, 245−253.

Sprent, J.F.A., 1983. Observations on the systematics of ascaridoid nematodes. In: Stone, A.R., Platt, H.M., Khalil, L.F. (Eds.), Concepts in Nematode Systematics, Systematics Association Special, vol. 22. Academic Press, London, pp. 303−319.

Sukee, T., Huynh, T., Beveridge, I., Jabbar, A., 2020c. Genetic variation in *Austrostrongylus thylogale* Johnston & Mawson, 1940 (Nematoda: Trichostrongylida) from the tammar wallaby, *Notamacropus eugenii* (gray), and the quokka, *Setonix brachyurus* (Quoy & Gaimard) (marsupialia: macropodidae) in Australia. Parasites Vectors 13, 135.

Sukee, T., Beveridge, I., Sabir, A.J., Jabbar, A., 2020a. Phylogenetic relationships within the nematode subfamily Phascolostrongylinae (Nematoda: Strongyloidea) from Australian macropodid and vombatid marsupials. Microorganisms 9, 9.

Sukee, T., Koehler, A.V., Hall, R.S., Beveridge, I., Gasser, R.B., Jabbar, A., 2020b. Phylogenetic analysis of mitogenomic data sets resolves the relationship of seven *Macropostrongyloides* species from Australian macropodid and vombatid marsupials. Pathogens 9, 1042.

Sukee, T., Beveridge, I., Koehler, A.V., Hall, R., Gasser, R.B., Jabbar, A., 2021. Phylogenetic relationships of the nematode subfamily Phascolostrongylinae from macropodid and vombatid marsupials inferred using mitochondrial protein sequence data. Parasites Vectors 14, 523.

Southern, E., 2006. Southern blotting. Nat. Protoc. 1, 518−525.

Stevenson, L.A., Chilton, N.B., Gasser, R.B., 1995. Differentiation of *Haemonchus placei* from *H. contortus* (Nematoda: trichostrongylidae) by the ribosomal second internal transcribed spacer. Int. J. Parasitol. 25, 483−488.

Stevenson, L.A., Gasser, R.B., Chilton, N.B., 1996. The ITS2 rDNA of *Teladorsagia circumcincta, T. trifurcata* and *T. davtiani* (Nematoda: trichostrongylidae) indicates that these taxa are one species. Int. J. Parasitol. 26, 1123−1126.

Stracke, K., Clarke, N., Awburn, C.V., Vaz Nery, S., Khieu, V., Traub, R.J., Jex, A.R., 2019. Development and validation of a multiplexed-tandem qPCR tool for diagnostics of human soil-transmitted helminth infections. PLoS Neglected Trop. Dis. 13, e0007363.

Tan, N., Chilton, N.B., Huby-Chilton, F., Jex, A.R., Gasser, R.B., Beveridge, I., 2012. Molecular evidence for a cryptic species within the parasitic nematode *Macroponema comani* (Strongyloidea: cloacininae). Mol. Cell. Probes 26, 170−174.

Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., Tripathi, A., Gibbons, S.M., Ackermann, G., Navas-Molina, J.A., Janssen, S., Kopylova, E., Vázquez-Baeza, Y., González, A., Morton, J.T., Mirarab, S., Zech Xu, Z., Jiang, L., Haroon, M.F., Kanbar, J., Zhu, Q., Jin Song, S., Kosciolek, T., Bokulich, N.A., Lefler, J., Brislawn, C.J., Humphrey, G., Owens, S.M., Hampton-Marcell, J., Berg-Lyons, D., McKenzie, V., Fierer, N., Fuhrman, J.A., Clauset, A., Stevens, R.L., Shade, A., Pollard, K.S., Goodwin, K.D., Jansson, J.K., Gilbert, J.A., Knight, R., Earth Microbiome Project Consortium, 2017. A communal catalogue reveals Earth's multiscale microbial diversity. Nature 551, 457−463.

van Nues, R.W., Rientjes, J.M., Morré, S.A., Mollee, E., Planta, R.J., Venema, J., Raué, H.A., 1995. Evolutionarily conserved structural elements are critical for processing of internal transcribed spacer 2 from *Saccharomyces cerevisiae* precursor ribosomal RNA. J. Mol. Biol. 250, 24−36.

Vaz Nery, S., Traub, R.J., McCarthy, J.S., Clarke, N.E., Amaral, S., Llewellyn, S., Weking, E., Richardson, A., Campbell, S.J., Gray, D.J., Vallely, A.J., Williams, G.M., Andrews, R.M., Clements, A.C., 2019. WASH for WORMS: a cluster-randomized controlled trial of the impact of a community integrated water, sanitation, and hygiene and deworming intervention on soil-transmitted helminth infections. Am. J. Trop. Med. Hyg. 100, 750−761.

Verweij, J., Polderman, A.M., Wimmenhove, M.C., Gasser, R.B., 2000. PCR assay for the specific amplification of *Oesophagostomum bifurcum* DNA from human faeces. Int. J. Parasitol. 30, 137−142.

Verweij, J.J., Pit, D.S.S., van Lieshout, L., Baeta, S.M., Dery, G.D., Gasser, R.B., Polderman, A.M., 2001. Determining the prevalence of *Oesophagostomum bifurcum* and *Necator americanus* infections using specific PCR amplification of DNA from faecal samples. Trop. Med. Int. Health 6, 726−731.

Wit, J., Gilleard, J.S., 2017. Resequencing helminth genomes for population and genetic studies. Trends Parasitol. 33 (5), 388−399.

Workentine, M.L., Chen, R., Zhu, S., Gavriliuc, S., Shaw, N., Rijke, J., Redman, E.M., Avramenko, R.W., Wit, J., Poissant, J., Gilleard, J.S., 2020. A database for ITS2 sequences from nematodes. BMC Genet. 21, 74.

Wu, T., Bowman, D.D., 2020. Visceral larval migrans of *Toxocara canis* and *Toxocara cati* in non-canid and non-felid hosts. Adv. Parasitol. 109, 63−88.

Zarlenga, D.S., Rosenthal, B.M., La Rosa, G., Pozio, E., Hoberg, E.P., 2006. Post-Miocene expansion, colonization, and host switching drove speciation among extant nematodes of the archaic genus *Trichinella*. Proc. Natl. Acad. Sci. U.S.A. 103, 7354−7359.

Zendejas-Heredia, P.A., Colella, V., Hii, S.F., Traub, R.J., 2021. Comparison of the egg recovery rates and limit of detection for soil-transmitted helminths using the Kato-Katz thick smear, faecal flotation and quantitative real-time PCR in human stool. PLoS Neglected Trop. Dis. 15, e0009395.

Zhang, L., Du, X., An, R., Li, L., Gasser, R.B., 2013. Identification and genetic characterization of *Anisakis* larvae from marine fishes in the South China Sea using an electrophoretic-guided approach. Electrophoresis 34, 888−894.

Zhang, L., Hu, M., Shamsi, S., Beveridge, I., Li, H., Xu, Z., Li, L., Cantacessi, C., Gasser, R.B., 2007. The specific identification of anisakid larvae from fishes from the Yellow Sea, China, using mutation scanning-coupled sequence analysis of nuclear ribosomal DNA. Mol. Cell. Probes 21, 386−390.

Zhu, X.Q., Gasser, R.B., 1998. Single-strand conformation polymorphism (SSCP)-based mutation scanning approaches to fingerprint sequence variation in ribosomal DNA of ascaridoid nematodes. Electrophoresis 19, 1366−1373.

Zhu, X.Q., Jacobs, D.E., Chilton, N.B., Sani, R.A., Cheng, N.A.B.Y., Gasser, R.B., 1998a. Molecular characterization of a *Toxocara* variant from cats in Kuala Lumpur, Malaysia. Parasitology 117, 155−164.

Zhu, X.Q., Gasser, R.B., Podolska, M., Chilton, N.B., 1998b. Characterisation of anisakid nematodes with zoonotic potential by nuclear ribosomal DNA sequences. Int. J. Parasitol. 28, 1911−1921.

Zhu, X.Q., Chilton, N.B., Jacobs, D.E., Boes, J., Gasser, R.B., 1999. Characterisation of *Ascaris* from human and pig hosts by nuclear ribosomal DNA sequences. Int. J. Parasitol. 29, 469−478.

Zhu, X.Q., D'Amelio, S., Paggi, L., Gasser, R.B., 2000a. Assessing sequence variation in the internal transcribed spacers of ribosomal DNA within and among members of the *Contracaecum osculatum* complex (Nematoda: Ascaridoidea: Anisakidae). Parasitol. Res. 86, 677−683.

Zhu, X.Q., Gasser, R.B., Jacobs, D.E., Hung, G.-C., Chilton, N.B., 2000b. Relationships among some ascaridoid nematodes based on ribosomal DNA sequence data. Parasitol. Res. 86, 738−744.

Zhu, X.Q., D'Amelio, S., Hu, M., Paggi, L., Gasser, R.B., 2001a. Electrophoretic detection of population variation within *Contracaecum ogmorhini* (Nematoda: Ascaridoidea: Anisakidae). Electrophoresis 22, 1930−1934.

Zhu, X.Q., Gasser, R.B., Chilton, N.B., Jacobs, D.E., 2001b. Molecular approaches for studying ascaridoid nematodes with zoonotic potential, with an emphasis on *Toxocara* species. J. Helminthol. 75, 101−108.

Zhu, X.Q., D'Amelio, S., Palm, H., Paggi, L., George-Nascimento, M., Gasser, R.B., 2002. SSCP-based identification of members within the *Pseudoterranova decipiens* complex (Nematoda: Ascaridoidea: Anisakidae) using genetic markers in the internal transcribed spacers of ribosomal DNA. Parasitology 124, 615−623.

Zhu, X.Q., D'Amelio, S., Gasser, R.B., Yang, T.B., Paggi, L., He, F., Lin, R.Q., Song, H.Q., Ai, L., Li, A.X., 2007. Practical PCR tools for the delineation of *Contracaecum rudolphii* A and *Contracaecum rudolphii* B (Ascaridoidea: Anisakidae) using genetic markers in nuclear ribosomal DNA. Mol. Cell. Probes 21, 97−102.

Further reading

Chilton, N.B., Huby-Chilton, F., Gasser, R.B., Koehler, A.V., Beveridge, I., 2016. *Pharyngostrongylus thylogale* n. sp. (Nematoda: Strongylida) from the stomachs of macropodid marsupials defined by morphological and molecular criteria. Syst. Parasitol. 93, 749−760.

Li, A.X., D'Amelio, S., Paggi, L., He, F., Gasser, R.B., Lun, Z.R., Abollo, E., Turchetto, M., Zhu, X.Q., 2005. Genetic evidence for the existence of sibling species within *Contracaecum rudolphii* (Hartwich, 1964) and the validity of *Contracaecum septentrionale* (Kreis, 1955) (Nematoda: Anisakidae). Parasitol. Res. 96, 361−366.

Roeber, F., Jex, A.R., Gasser, R.B., 2013c. Impact of gastrointestinal parasitic nematodes of sheep and the role of advanced molecular tools for exploring epidemiology and drug resistance - an Australian perspective. Parasites Vectors 6, 153.

Romstad, A., Gasser, R.B., Polderman, A.M., Nansen, P., Pit, D.S.S., Chilton, N.B., 1997b. Differentiation of *Oesophagostomum bifurcum* from *Necator americanus* by PCR using genetic markers in spacer ribosomal DNA. Mol. Cell. Probes 11, 169−176.

Zarlenga, D.S., Chute, M.B., Martin, A., Kapel, C.M., 1999. A multiplex PCR for unequivocal differentiation of all encapsulated and non-encapsulated genotypes of *Trichinella*. Int. J. Parasitol. 29, 1859−1867.

Index

'*Note:* Page numbers followed by "f" indicate figures, "t" indicates tables'.

A

Accidental virulence, 89
Acetylcholinesterase, 330–331
Acidic nuclear phosphoprotein 32 family member A (ANP32A), 587
Acinetobacter baumannii, 431
Acquired Immune Deficiency Syndrome (AIDS). *See* Human immunodeficiency virus (HIV)
Actinomycetes, 231
"Adapt Globally Adapt Locally" model, 30
Adenosine Triphosphate (ATP), 323–324
Aedes aegypti, 257–258, 387
Aedes albopictus, 317–318
Agriculture, 216–217
Allele-based analytical methods, 434–435
Allele-based approaches, 433f
AlphaPept, 471
AlphaPeptDeep, 471
Alternative insecticides, 322
American Academy of Pediatrics (AAP), 168–169
Anaplasma phagocytophilum, 38
Anatomically Modern Humans (AMH), 697–698
"Ancestral lineages", 730
Ancestry informative markers, 168
Ancient *versus* modern TB naming, 703
Anglicized non-Latinized binomial names (NLBNs), 8
Anglicized Non-Latinized Virus Species, 8
Animal-adapted ecotypes, 664–666
Animal-adapted lineages, 736–737
Animal lineages, 736–741
Animal MTBC lineages, 737–740
Animal pathogens, 81–82
Anopheles gambiae, 233–234, 330–331
Anopheles mosquito, 150
Anopheles stephensi, 322
Antifungals, 91–92
  resistance, 92
Antimicrobial peptides, 539
Antimicrobial resistance (AMR), 69, 291–292, 303–304
  genetic mechanisms, 299–300
  genomics, 294–295
    epidemiology, 292–293
  genotypic prediction, 293–296
  surveillance, 292–293
Antiretroviral Treatment (ART), 811–813
ApicoPLAST Neural Network (PlastNN), 478–479
Arbovirus vectors, 340
Archaic Adaptive Introgression (AAI), 173
Arginine Catabolic Mobile Element (ACME), 793–794
Artemisinin-based combination therapy (ACT), 232, 648
Arthropod-pathogen interactions, 257
Ascorbate peroxidase (APEX), 469
Aspergillus, 83–85
  A. fumigatus, 89, 93
Assembly-based methods, 295
Asymptomatic gonorrhea infections, 298–299
ATP-binding cassette (ABC) transporters, 91
Aureobasidium, 96
Austenina, 359
Average Nucleotide Identity (ANI), 22
Avian influenza A viruses (AIV), 585–587, 865–870, 867t–868t
  ecology, in waterfowl, 870–876
  gulls, 879–881
  seabirds, 879–881
  shorebirds, 877–878

B

Bacillus spizizenii, 35, 39
Bacillus thurengiensis, 331–332
Bacterial-human protein-protein interactions, 478–479
Bacterial phylogeography, 67–68
Bacterial populations, 58–66
  structure, 64–69
Bacterial species, 116–123, 143–144, 185–192
Bacterial systematics, 17–19
Batrachochytrium dendrobatidis, 81, 94–95
Batrachochytrium salamandrivorans, 94
Bayesian coalescent approach, 442
Bayesian framework, 435
Bayesian inference (BI), 436–437
Bayesian model, 478
BCG vaccine, 669, 685–686
Beijing Genome Institute (BGI), 503–504
"Beijing" or East-Asian Lineage, 699–701
Bidirectional human-animal interface, 549–551
Bioassays, 332–334
Biochemical assays, 335
Biogeographical islands, 407–408
Bioinformatics, 476–479
  tools in proteomics, 470–471
Biological factors, 337

Bioterrorism, 272
β-lactam antibiotics, 303
β-lactamase genes, 303
Blastomyces, 85–86
"Block and lock" strategy, 537
Borrelia burgdorferi, 62
Bottlenecks shape transmission dynamics, 64–65
Bovine coronavirus (BCoV), 902–903
Brain-derived neurotrophic factor (BDNF), 524
Burkholderia pseudomallei, 58–59

C

Caenorhabditis elegans, 934
Cameroon lineages, 722–724
Campylobacter jejuni, 45, 58–59
Canalization, 396–397
Candida, 83–85
  C. auris, 92
  C. glabrata, 92–93
  C. krusei, 93
Candidate vaccine viruses (CVV), 553
Canonical variate analysis (CVA), 400, 404
Carboxylesterases, 326–327
ccr gene complex, 780–781
Ceftriaxone monotherapy, 299
Center for Disease Control (CDC) bottle bioassay, 318
Central nervous system (CNS), 84
cgMLST, 431–432
Chagas disease, 369
  vectors, 369–371
Chikungunya virus, 590
Chlorfenapyr, 323–324
Chromatin immunoprecipitation (ChIP), 534
Chromosomal rearrangements, 639
Chytridiomycosis, 81
Clone naming, 707–708
Clostridioides difficile, 496
Coccidioidomycosis, 85–86
CodABC, 438
Cohesive Recombination model, 36
Collisionally activated dissociation (CAD), 253
Colonial sciences, 513–514
Colonial S type, 721
Columbian exchange, 219–220
Commercial sequencing facilities, 494
Communicable diseases, 173–174
Community-associated MRSA, 793–794
Community-associated (CA) pathogen, 773–774
Compensatory mutations role, 298
Completionist systematics
  ecotype level, 38–44
  species level, 19–24
Comprehensive Antibiotic Resistance Database (CARD), 295–296
Congo lineage, 725
Conventional Fourier analysis, 390
Conventional virulence factors, 89–90
Convention on Biological Diversity (CBD), 515
Convergent evolution, 920–921
Coronavirus Antibody Database, 918
CRISPR/Cas9-based epigenome editing, 540
Cross-resistance, 335–336
Cross-species transmissions, 588–592, 828
  determinant, 583–585
  molecular determinants, 585–588
Cryphonectria parasitica, 94–95
"Cryptic species", 398
Cryptic zoonoses, 588–589
Cryptococcus, 83–85
Cryptococcus gattii, 127–128
Cryptococcus neoformans, 84, 88–89, 127–130, 472
Cynomolgus macaque model, 670
Cytochrome P450 monooxygenases (P450), 325–326

D

Deep phylogenies, 114
Deep sequencing, 64–65
Delhi or Central-Asian (CAS) Lineage, 701–703
Demography, 215–216
Developmental stability, 397
Diachasmimorpha longicaudata, 394
Diagnostic markers, 477–478
"Dilution effect", 410
Dipeptidyl Peptidase 4 (DPP4), 583
Discriminant analysis (DA), 405
Disease vectors, insecticide resistance in, 339–341
Divergent biochemical adaptation, 644
DNA-based typing analyses, 432–443
DNA-based typing approaches, 428–432
DNA methyltransferase (DNMT) inhibitors, 540
Drosophila neotestacea, 234
Drosophila pseudoobscura, 26
Drug-resistance studies, 694
Drug susceptibilities, 644
  prediction, 297

E

East-African Indian (EAI) lineage, 696–699
  genomic characteristics, 698
  geographical distribution, 698
  Manila lineage, 699
  TB, 698–699
Ecological divergence, 27–31
Ecological health, 517b
"Ecological Theory", 673

Ecosystem health, 515–516, 517b
Ecosystem services, 515–516
Ecotypes, 363–364
Ectoparasites, 144
Electron transfer dissociation (ETD), 253
Elliptic Fourier algorithm (EFA), 390
Emergent viruses, 573–574
Emerging infectious diseases (EIDs), 80, 149–150
Emerging viruses
  longer-term evolution of, 591–592
  transmissibility in, 589–591
Endoparasites, 143
Enterococcus faecium, 67–68
Environmental health, 517b
Environmental pollution, 322
Environment and host microorganisms interactions, 272–273
Ephemerovirus fiumeadelaidense, 8
Epigenetic alterations, infectious diseases in, 529–532
Epigenetic biomarkers, 533
Epigenetic changes, long-term effects of, 532
Epigenetic host response to infections, 526–527
Epigenetic inheritance, 527–528
  tradeoff, 529
Epigenetic lab tests, 534
Epigenetic modifications, 525, 538–539
Epigenetic research, 540
Epigenetics
  definition, 523
  disease development, 525
  epigenetic modifications, 524–525
  infectious diseases, 534–537
  infectious diseases role, 525–529
  signatures, 533–535
Escherichia coli, 21, 24, 116–118
"Euro-American" Lineage, 703–728
European Gonococcal Antimicrobial Surveillance Program (Euro-GASP), 301–302
Evolutionary arms race, 144–145
Exploratory data analyses (EDA), 399
Extended-spectrumcephalosporins (ESCs) cefixime, 299
Extremotolerance, 90–91

F

Facultative pathogenesis, 87–89
Feline coronavirus (FCoV), 902–903
"Female farming system", 222–224
Fitness cost, 338
Fluctuating asymmetry (FA), 397
Fluxomics, 270–271
Formalin-fixed paraffin-embedded (FFPE) tissues, 533
Fungi, 143
  pathogens, 80–82, 86–87

G

GABA receptors, 328
Gammaherpesvirus genomes, 182–183
Gene expressions, 534–535
  profiling, 534–535
Gene flow, 406–408
Gene interactions, 337–338
Generalized Procrustes Analysis (GPA), 388
Genetically modified mosquitoes (GMMs), 233–234
"Genetic assimilation", 393
Genetics
  diversification mechanisms, 690
  eradication/control, 364–365
  exchange, 607
  factors, 337–339
  inequality, 169–171
  mutations, 147
  stratifications, 167
  tools, 173
  variation, 165
Genexus Purification system, 503
Genomes, 147
  mining, 231
  sequencing, 22, 30, 173
Genome-wide association studies (GWAS), 23, 168, 232–233, 743–744
Genomic antimicrobial susceptibility testing, 497
Genomic deletion-basedphylogeny, 659–661
Genomic epidemiology, 694
Genomic features, 900–902
Genomic reassortment, 882–883
Genomic signals, 637–638
Genomics Service Companies (GSC), 494–495
Genomic studies of genus, 638–639
Genomic surveillance, 916–917
Genomic variation, 493–494
Genotyping, 446–448
Geography, 67–69
Geotype plus Boeing model, 36
"*Ghana*" sublineage, 711
Giardia duodenalis, 131–132
GlaxoSmithKline, 232
Global health, 517b
Global MRSA lineages, 789–795
Global Prokaryotic Census, 19–20
Glossina, 359
Glossina morsitans centralis, 362–363
Glossina palpalis gambiensis, 364
Glossina tachinoides, 364
Glutathione S-transferases (GSTs), 324–325
Gonococcal Resistance to Antimicrobials Surveillance Program (GRASP), 301

Greater Mekong subregion (GMSR), 232
Green fluorescent protein (GFP), 609
GWS-based analyses, 441–443

H

"H37Rv-Like" lineages, 725–727
"*Haarlem*" lineage, 710
"Haarlem" sublineages, 710
Haemonchus placei, 940–941
HA-MRSA, 789–793
Health factor, environment as, 512
Helicobacter pylori, 184–187, 261
Hendra virus, 581
Hepatitis B virus (HBV) HBx protein functions, 531
Hepatitis C Virus (HCV), 252
Heritability, 392–393
Herpes simplex virus (HSV), 535
Heterogeneity, 59–61
Heuristic need, 408–409
Hidden genetic variability, 393
Hierarchical agglomerative clustering (HAC), 399
High pathogenic avian influenza virus (HPAIV), 866–868
High-throughput MLST (HiMLST), 431–432
High-Throughput Sequencing (HTS) technologies, 293
Histone deacetylase (HDAC) inhibitors, 539–540
Histonedeacetylase inhibitors (HDACi), 538
Histoplasma, 85–86
Homogenization, 167
Homo Sapiens, 164, 181–182, 635–637
Horizontal gene transfer (HGT), 130, 659
Hormonal regulation mechanisms, 331
Hortaea werneckii, 96
Host behavior, parasite manipulation of, 152–153
Host defense mechanisms, 87–88
Host epigenome, 528
Host immunity, 637–638
  evasion, 918–920
Host immunogenetics adaptations, 151–152
Host-pathogen co-evolution, 659–666
Host-pathogen crosstalks, 270–271
Host-pathogen interactions, 252–270, 539
Host-pathogen interactomes, 260–261
Host proteome responses, 250–251
Host restriction factors, 826
Host specificity, 643–644
Host susceptibility, 525–526, 528
Host transitions, in speciation events, 639–640
Host-vector interactions, 256–257
HPAIV, 887
HPAIV H5N1 panzootic, 884–887
Huanan seafood wholesale market, 906
Human-animal interface, 557–559
Human factors, 827
Human genetic variability and susceptibility, 173–174
Human Genome Project (HGP), 229
Human herpesviruses, 195–197
Human herpesviruses 6A (HHV6A), 196–197
Human immunodeficiency virus (HIV), 97, 527
  cross-species transmissions, 822
  diverge, in human population, 822–823
  history, 811–813
  HIV-2 origin, 823–824
  novel, 824–828
  Simian Immunodeficiency Virus (SIV)
    African NHP, 813–816
    chimpanzees, 819–822
    evolution, 818–819
    gorillas, 819–822
    pathogenicity, 816–818
    phylogeny, 818–819
    transmittion, to humans, 823
Human pathogens, 82–87
Human polyomaviruses, 193–195
Human population calibration points, 187
Human population proteomics, 264–270
Human-to-human transmission, 828
Human tuberculosis bacilli, 687
Hybridism, 393–395
Hygienism, 513

I

IAV hemagglutinin (HA) glycoprotein, 554
Immune deficiency viruses, 816
Immune evasion cluster (IEC) genes, 787–788
Immune response, 538
Indispensable fatty acid synthase (FAS) pathway, 231–232
Infectious disease genomics, 229–230
  clinical application, 234–236
  drug resistance, 232–233
  drug target, 231–232
  new drug discovery, 231
  therapeutic response, 232–233
  vaccine target, 230–231
  vector control, 233–234
Infectious diseases, 530–537
  epigenetic alterations, 529–532
  epigenetic factors against, 539–540
  epigenetics, 534–537
  prevention, 538–539
  proteomics, 466–469
  transmission, 525–526
  treatment, 538–539
Infectious pathogens, 526–528
  resistance, 527–528
  susceptibility, 528–529

Influenza A virus (IAV), 554, 555t, 865, 882
evolution, 551–554
host switches, 554–556
Influenza virus, 585–588
Insecticide classification, 319–324
Insecticide resistance, 318–319, 335–340
disease vectors, 339–341
management, 341–342
mosquitoes, 332–335
Integrated Vector Management (IVM), 342
Interferons (IFNs), 530
Interferon-stimulated genes (ISGs), 538
Internal transcribedspacers (ITS), 933–934
International Code for Nomenclature of Prokaryotes, 20–21
International Code of Virus Classification and Nomenclature (ICVCN), 9
International Committee on Nomenclature of Viruses (ICNV), 3
International Committee on Taxonomy of Viruses (ICTV), 3, 8–9, 905, 913
definition, of virus species, 9
latinized binomial nomenclature, 9–11
International Nucleotide Sequence Database Collaboration (INSDC), 494
"Inter-user effect", 401–402
Intraclonal mating, 612
Intralineage genetic exchange, 615–616
In vitro attenuation of bacteria, 668–670
In vitro diagnostic companies, 493–495
applications, 496–498
commercial sequencing instruments, 498–505
customers, 496–498
genomics service companies, 506
infectious diseases, 506
library preparation, 505–506
reagents, 498–505
software, 498–505
In vitro laboratory-based approach, 232–233
Ion AmpliSeq panels, 503
Ion Chef system, 503
Ion GeneStudio S5 system, 503
Ion Personal Genome Machine (ion PGM), 503
Isoenzyme typing, 111

J

JC polyomavirus (JCPyV), 193–194
Joining (J) regions, 782

K

Kinetoplast DNA, 613
Klebsiella, 304–305
K. pneumoniae, 302–305, 444, 448

L

L4.1.4 sublineage, 712
L4 sublineages characteristics, 708–728
Laboratory environment, 668–670
LAMARC, 438
LA-MRSA, 794–795
Landmark-based geometric morphometry, 386–389
Larval tests, 332–333
Lassa virus, 588
Latin American and Mediterranean (LAM) sublineages, 715
characterization, 715
lineage identity using SNPs, 715–716
using SNPs, 716–717
Latinized Linnaean binomial format (LLBNs), 8
LDhat, 438
Leishmania
genetic crosses
in sand flies, 623
in vitro, 623–624
genetic exchange, in natural populations, 621–622
Leishmania donovani/infantum, 123–125
Leishmania spp., 624
Lineage classification, 725–727
Lipooligosaccharide (LOS) synthesis, 663–664
Lipo-polysaccharides (LPS), 257–258
Listeria monocytogenes, 60, 431, 532
Longer-term evolution, of emerging viruses, 591–592
Low- and middle-income countries (LMIC), 296
Low-frequency and rare variants, 166
LPAIV, 869f
ecology, 870–881
mallard immune response against, 874–875
mallards, 871–873
other waterfowl, 875–876
Lymph Nodes tuberculosis (LNTB), 742–743

M

Machine learning, 479
approach, 405
models, 477–478
Macroparasites, 143
Magnaporthe oryzae, 94–95
Major histocompatibility complex (MHC), 151, 216
Malaria, 217–219
vectors, 340
Markov chain MonteCarlo (MCMC) method, 436–437
Mass spectrometric immunoassay (MSIA), 264, 265f
Mass spectrometry-based proteomics, 465–469, 471–472
Measles morbillivirus, 8
Measles virus, 9
Measurement error, 400
mec gene complex, 779

Men having Sex with Men (MSM), 811
Merkel cell carcinoma (MCC), 236
Merkel cell polyomavirus (MCV), 236
Messenger RNAs (mRNAs), 534–535
Metabolic resistance, 324–327
Metabolomics, 270–271
Metagenome-assembled genome sequence (MAG), 21
Metagenomics NGS (mNGS), 497
MetaMLST, 431–432
Methicillin-resistant *Staphylococcus aureus* (MRSA), 59, 66
  mobile genetic elements, 787–788
  molecular dynamics, 786–788
Metric change, genetics of, 391–395
Microbial genome program (MGP), 229–230
Microbial sequence typing applications, 443–448
Microbial typing techniques, 427
Microparasites, 142–143
MicroRNA (miRNA) modulators, 540
Migrants flow, 406–408
Milk pastorization, 686
Millennium Ecosystem Assessment, 515
Mitochondrial genetic markers, 942–943
Mobile genetic elements (MGE), 684
Modern morphometrics, 385
Molecular assays, 334–335
Molecular clock analysis, 66
Molecular epidemiology, 111–113, 444–446
Monoextremotolerance, 90–91
Mononucleosis, 528–529
"Morphometric revolution", 385
Morphometrics, taxonomic use of, 399–403
Mosquitoes
  eight genera, 403–404
  insecticide resistance, 332–335
  thirteen species, 404–406
Most recent common ancestors (MRCAs), 189, 661, 698–699, 823–824, 849
Mucocutaneous leishmaniasis (MCL), 622
Multidrug resistance (MDR), 445
Multigene bifurcating trees, 114
Multilocus enzyme electrophoresis (MLEE), 432–434, 614
Multilocus microsatellite typing (MLMT), 619
Multilocus sequence typing (MLST), 60, 185, 427–431, 448, 784–785
Multilocus Variable number of tandem repeats Analysis (MLVA), 428–431
Multiomics integration, 475–476
Multiple-Locus-Sequencing-Typing, 700
Multiple resistance, 336
Multiple sequence alignment (MSA), 435
Mutant clouds/quasispecies, 147
Mutation frequency, 908
Mutation rates, 908
Mycobacterial genetics, 657–659
Mycobacterial interspersed repetitive units (MIRUs), 188
Mycobacterium abscessus, 60
Mycobacterium africanum, 663, 672–673
  sublineages, 735–736
Mycobacterium africanum (Maf), 688–689, 728–734
  chronology, of loss, 730
  ecology, 733
  oxygen limitation, 734
  RDs, 728–730
  SNPs, 732–734
  spoligotyping, 728–730
  spoligotyping signatures, 734
  TbD1 to virulence, 730–731
  WGS definition, 730
Mycobacterium bovis, 665–666, 672–673, 685–686
Mycobacterium canettii, 188, 663–664, 736–741
  characteristics, 740–741
  environmental bacteria, 741
Mycobacterium caprae, 665–666
Mycobacterium leprae, 171–172, 189–192, 666, 674, 693–694
Mycobacterium marinum, 674
Mycobacterium microti, 664–665
Mycobacterium pinnipedii, 664–665
Mycobacterium tuberculosis (MTB), 33, 58, 171–172, 187–189, 296–298, 528, 662–663, 666–668, 670–672, 683–692, 733f
Mycobacterium tuberculosis complex (MTBC), 297, 657, 666–670, 672–674, 735, 737f, 741–744
  classification, 694
  differential virulence, 742–743
  genomes classification, 695
Mycobacterium tuberculosis Drug Resistance tests (MTBDR), 493–494
MyxomaResistance Protein 1 (MX1), 585–587

N

Nakaseomyces glabrata, 86
Nano-Niche ecotypes, 34–35
Nano-Niche model, 34–35
Nanopore-based sequencing, 504–505, 943
National Human Genome Research Institute (NHGRI), 229
Natural habitats, 361
Natural insecticides, 319–320
Negative frequency–dependent selection (NFDS), 62–64
Neisseria gonorrhea, 68–69, 298–302, 445
Neisseria gonorrhoeae, 299–302
Neisseria lactamica, 59
Neisseria meningitidis, 59, 118–123
Nemabiome investigations, 940–941

Nematodirusbattus, 936
Nemorhina, 359
Neonicotinoids, 323
Network of Excellences (NoE), 685
Neutral and metric characters, 407
"NEW 1" lineage
  and its sublineages, 721
  molecular characterization, 721–722
New fungal pathogens, 92–97
Next generation sequencing (NGS), 236, 427–428, 494, 499–500, 499t–501t, 534
Next-Generation sequencing MLST (NGMLST) approach, 431–432
NGS-based approaches, 431–432
Nipah virus, 585
Nonhuman primates (NHP), 824–825
Nonrecombining portions of the Y chromosome (NRY), 181
Normalized elliptic Fourier coefficients (NEF), 399
Nuclear ribosomal DNA, 934–936
Nucleotide-based analytical methods, 435–439
Nucleotide-based approaches, 433f
Nunavik Sublineage, 710

O

Oesophagostomum bifurcum, 937
OmegaMap, 438
Onchocerciasis Control Program (OCP), 342
One Health, 517b, 518f, 559–560
  definition, 516–518
  origin, 516
"Ontogenetic" allometry, 391
Open reading frames (ORFs), 230–231
Operational factors, 339
Ophiostoma ulmi, 94–95
Organophosphates (OPs), 321
Outer-membrane vesicles (OMV), 230
Outer surface protein C (ospC), 62
Overall Genome Relatedness Indices (OGRIs), 22, 24
Oxford Nanopore Technologies (ONT), 293–294

P

Pacific Biosciences (PacBio) platforms, 293–294
Pairwise genome sequence comparisons, 44
Pan-coronavirus therapeutics and vaccines, 853
Pan-genome, of species and populations, 61–62
Panstrongylus geniculatus, 371
Paracoccidioides, 85–86
Parasite-absent world, 153–154
Parasite genomes, 638
Parasite–host relationships
  animal infectious diseases, 150–151
  human infectious diseases, 149–150
Parasite infection, 250–251
"Parasite-proteomics" surveys, 253
Parasites, 149–150, 154, 215–216
  nematodes, 934–936
  species, 123–128
"Parasito-proteomics", 253–256
Partial warps scores (PW), 388
Pathogeno-proteomics, 253–256
Pathogens, 145–146
  genomics, 182–184, 496
  immune evasion, 246–250
  population structure, 116
  replication, 539
  strains, 172
Pattern recognition receptors (PRRs), 530
Phage typing, 690
Phenotype regulation, 396–397
Phenotypic continuum, 736–737
Phenotypic plasticity (PP), 395–396
Phenotypic variation, 395–397
Phlebotomus duboscqi, 623
PhoP-PhoR virulence system, 668–669
Phylogenetic Assignment of Named Global Outbreak Lineages (PANGO), 914
Phylogenetic maximum likelihood (ML) inference, 436
Phylogenetics, 148–149
  inference, 446–448
  investigations, 938–939
  relatedness, 436–437, 440–443
Phylogeography, 712
Phytophthora infestans, 80
Phytophthora ramorum, 80
Pichia kudriavzevii, 93
Piperonyl butoxide (PBO), 325–326, 334
Planetary health, 517b
Plant pathogens, 80–81
Plasmids role, 304
Plasmodium, 639–644
  ecological transitions, 645–646
  emergence, 646–647
  habitat destruction, 645–646
Plasmodium falciparum, 218, 234, 261, 478–479, 504, 640–641
Plasmodium Ovale Curtisi, 641–643
Plasmodium Ovale Wallikeri, 641–643
Plasmodium phylogeny, 635–637
Plasmodium Reichenowi, 640–641
Pneumocystis, 85–86, 93
Polyextremotolerance, 90–91
Polygenic scores, 167
Polyketide synthases, 743
Polymerase Chain Reaction (PCR), 534

Population dynamics, 437–439, 441, 443
Population genetics, 92–95, 164–168, 359–362, 637
Population genomics, 92–95
Population structure studies, 406–411
Porcinedeltacoronavirus strains, 903
Posttranslational modification (PTMs), 469
Potential taxonomical conflicts, 725–727
Predominant clonal evolution (PCE)
  model, 113–116
  traits, 115
Preindustrial societies, 220–221
Preventive medicine, epigenetic factors in, 537–540
PREZODE program, 519
Principal component analysis (PCA), 404
"Principal Genetic Groups", 703
Prognostic markers, 477–478
Protective polymorphisms, 217
ProteinEvolverABC, 438
Proteogenomics, 258–260
Proteomics, 244–245, 261–270, 272, 476–479
  approaches, 257–258
  arthropod vectors, 256–257
  bioinformatics tools, 470–471
  infectious diseases, 466–469
  studies, 245–252
  tools, 248t–249t
Protozoa, 143
Pseudolandmark-based geometric morphometry, 389–390
Pseudomonas aeruginosa, 33, 62, 472
Public health interventions, 637
Public health policy, 540
Pulsed Field Gel Electrophoresis (PFGE), 784
Pyriproxyfen, 323
Pyrroles, 323–324

Q

Quantitative genetics, 743–744
Quantitative trait loci (QTL), 391–392

R

Random amplification of polymorphic DNA (RAPD), 619
Range expansion, 68–69
Rare variants (RVs), 166
RDs, 728–730
Reactive oxygen species (ROS), 531
Receptor binding domain (RBD), 584, 909
Recurrent Niche Invasion model, 35
Red queen hypothesis, 144–145
Reservoir hosts, viral diversification in, 579–580
Resistance evolvement, 335
Resistance mechanisms, 324–332, 334–335
Restriction fragment length polymorphism (RFLP) approaches, 933–934
Restriction-modification systems, 121
"Reverse vaccinology" approach, 230–231
Rhinolophus sinicus, 850–851
Rhizobium leguminosarum, 35
Rhodesian sleeping sickness, 357
Rhodnius ecuadoriensis, 371
Rhodnius prolixus, 379–380
Ribosomal Multilocus Sequence Typing method (rMLST), 431–432
Riverine species, 361–362
RNA viruses, 145–146
Ruminococcus gnavus, 40
Russian doll model, 115
Russian doll pattern (RDP), 115f, 166

S

Saccharomyces cerevisiae, 83
Salmonella typhimurium, 171–172, 473–474
Sample-to-Insight GeneReader NGS System, 504
Sanger sequencing, 429, 498
SARS Coronavirus (SARS-CoV), 843
  animal link, 843–845
  animals, 849–852
  history, 841–842
  humans, 849–852
  like disease emergence, 852–853
  natural reservoirs, 845–849
  potential recombination origin, 851–852
SARS Coronavirus-2 (SARS-COV-2), 557–559, 899
  coronaviruses, 899–903
  evolution, 915–921
  genetic diversity, 907–914
  origin, 903–907
S. aureus characterization, 784–786
SCC*mec* types, 782, 785
Secretomes, 258–260
SELDI-TOF-MS Technology, 251–252
Self-reported ancestry (SRA), 168
Semilandmarks, 389
Sequence clusters, 58–59
Sequence read archives (SRA), 693
Sequence Type Analysis and Recombinational Tests (START2), 434
Sequence Typing Analysis and Retrieval System (STARS), 434
Sequencing-by-synthesis (SBS) approach, 498
Seven-valent pneumococcal conjugate vaccine (PCV7), 63
Sexual isolation, 27–31
Sexually transmitted infections (STIs), 298–299
Shape, as polygenic character, 391–392

"Shock and kill" approach, 537
"Shock and kill" strategy, 537
Silencing RNA (sRNA), 536
Simian Immunodeficiency Virus (SIV)
  African NHP, 813–816
  chimpanzees, 819–822
  cross-species transmissions, 825–826
  diversity, 824–828
  evolution, 818–819
  gorillas, 819–822
  pathogenicity, 816–818
  phylogeny, 818–819
  prevalences, 825–826
  transmittion, to humans, 823
Simulium squamosum, 341
Single Amplified Genome (SAG), 21
Single-cell proteomics, 476
Single Molecule Real-Time (SMRT) sequencing technology, 504
Single molecule sequencing, 504
Single nucleotide polymorphisms (SNPs), 431–432, 612, 705–707, 712, 720–722, 724
"S" lineage and its sublineages, 719
SNPs-based analyses, 439–441
Speciation events, 639
Species complexes, 363–364
Species-Less model, 33
Species richness, 410–411
Species taxa, 24–27
Speedy Speciation model, 33
Spinochordodes tellinii, 152–153
Spoligotypes, 705–707, 728–730
Sporothrix brasiliensis, 86–87
Stable Ecotype model, 31
Standard typing approaches, 428–431
Staphylococcal chromosome cassette *mec* (SCC*mec*), 778–779
Staphylococcal clonal complexes (CC), 786–787
Staphylococcal plasmids, 787–788
Staphylococcal Protein A *(spa)* typing, 785
Staphylococcus aureus (SA), 143–144
  evolution, 774–778, 775f
  methicillin resistance, 778–782
  staphylococcal cassette chromosome *mec*, 778–782
Staphylococcus epidermidis, 67–68
Sterile Insect Technique (SIT) signatures, 725
Streptococcus agalactiae, 232
Streptococcus pneumoniae, 498
Streptococcus pyogenes, 31
Streptomyces platensis, 232
Subtype diversity in birds, 870
Support vector machine (SVM), 406
Surface-enhanced laser desorption time of flight mass spectrometry (SELDI-TOF-MS) technology, 251–252
Symbiotic host–parasite relationship, 141
Synthetic insecticides, 320–322

T

Tacaribe mammarenavirus, 583
Target-site modification, 327–331
Taxonomic investigations, 936–938
Taxonomic questions, 398–406
Taxonomic signal of shape, 399
Teladorsagia circumcincta, 939
Theory-based systematics, of bacteria, 44–46
Thermotolerance, 90
"T" Lineage, 703–728
Total adjusted accuracy (TAU), 404
Toxin receptors, 331–332
Toxocara canis, 143
Toxoplasma gondii, 152–153
Traditional flyway model, 883
Traditional morphometrics, 385
Traditional phylogenetic methods, 434–435
Tradition-based systematics of bacteria, 44–46
Transferrin Receptor (TFRC), 583
Transmissible diseases, 169–171
  epidemiology, 172–173
  and history, 173
Triatoma brasiliensis, 371
Triatoma dimidiata, 378
Triatoma infestans, 371, 378, 621
Triatoma longipennis, 371
Triatoma rubrofasciata, 371
Triatominae
  cytogenetics, 371–377
  evolution, 371
  genetic diversity, 377–378
  population structure, 377–378
  taxonomy, 371
Trichophyton mentagrophytes, 86–87
Trypanosoma brucei, 213–214
  genetic crosses, 608–609
  genetic exchange location, 609–610
  meiosis, 610–613
  mendelian inheritance, 610–613
Trypanosoma cruzi, 125–128, 127f, 369
  behavior, of experimental hybrids, 618
  diversity, 614–616
  genetic crosses, 616–617
  genetic exchange
    experimental, 619–620
    location, 617–618
  natural hybrids, 619–620

Tsetse bacteriome, 362
Tsetse flies, 358–359
  populations genetics, 362–365
Tsetse species and subspecies, 358t
Tube assay, 333
Tubercle bacillus, 659–666
Tuberculosis (TB), 683–684, 688–689
  archeology, 688
  disease, 657–659, 741–744
  genetics, 690
  molecular epidemiology, 684–685

U

Uganda lineage, 722–724
Ultraconserved elements (UCE), 371
Unesco Conference on biosphere, 514–515
Unique Event Polymorphism (UEP), 686, 715–716, 728–730
United Nations Environment Program (UNEP), 517–518
"*Ural*" sublineage, 712–713
Urinary tract infections (UTIs), 477
Ustilago maydis, 94–95

V

Vaccination strategies, 539
Vancomycin-intermediate (VISA), 777–778
Vancomycin-resistant *S. aureus* (VRSA), 777–778
Variable Number of Tandem Repeats (VNTRs), 188, 691, 785
Variant L4.1.2 sublineages, 710
Variants of concern (VOCs), 914–917
Variants of interest (VOIs), 914–917
Variant surface glycoprotein (VSG), 614
Varicella-zoster virus (VZV), 195
Vector Biology Network (VBN), 233
Vector-borne diseases (VBDs), 317
Vector control programs, 340
Vibrio cholerae, 66, 448
Vibrio vulnificus, 62
Viral adaptation, 826–827
Viral cross-species transmission, 582
Viral diversification in reservoir hosts, 579–580
Viral emergence, 574f, 577
  ecological drivers, 575–577
  genetic determinants, 582–588
  molecular determinants, 582–588
  societal drivers, 575–577
  sources, 578–581
Viral evolution, 146–147
Viral genetics, 579–580
Viral genomics, 886–887
Viral metagenomics, 573
Viral zoonosis, intermediate hosts in, 580–581
Virological features, 582–583
Virosphere, 573–574
Virulence factors, 89–92, 591–592
Viruses, 144, 193–197
  classification, 3–5
  human herpesviruses, 195–197
  human polyomaviruses, 193–195
  living organisms, 12
  surveillance, in wildlife, 852–853
Voltage-gated sodium channels (VGSCs), 328–330

W

White-nose syndrome (WNS), 81
WHO Bottle assay, 333
Whole-genome sequencing (WGS), 61, 125, 296, 300–301, 431–432, 448–449, 684–685, 707, 713–715
  applications, 694
  methodology, 692–694
  molecular epidemiology, 695
  MTBC classification, 694
  typing, 785–786
Wild birds role, 863–865, 884–887
Within-host diversity, 64–65
World Conservation Society (WCS), 516

X

"X" "Anglo-Saxon" IS*6110* low-copy sublineage, 708

Y

Yeasts, 128–130
Yellow fluorescent protein (YFP), 611

Z

Zoonoses, 518–519, 646–647
Zoonotic hantaviruses, 575–576
Zoonotic viruses, 578–579
Zymoseptoria tritici, 80–81, 94–95

GENETICS AND EVOLUTION OF INFECTIOUS DISEASES

THIRD EDITION

Edited by **MICHEL TIBAYRENC**

Genetics and Evolution of Infectious Diseases, Third Edition, discusses the constantly evolving field of infectious diseases and their continued impact on the health of populations, especially in resource-limited areas of the world where they must confront the dual burden of death and disability due to infectious and chronic illnesses. Although substantial gains have been made in public health interventions for the treatment, prevention, and control of infectious diseases during the last century, in recent decades, the world has witnessed the emergence of the human immunodeficiency virus (HIV) and the COVID-19 pandemic, increasing antimicrobial resistance, and the emergence of many new bacterial, fungal, parasitic, and viral pathogens.

Fully updated and revised, the Third Edition presents the consequences of such diseases, the evolution of infectious diseases, the genetics of host–pathogen relationship, and the control and prevention strategies that are or can be developed. This book offers valuable information to biomedical researchers, clinicians, public health practitioners, decisions-makers, and students and postgraduates studying infectious diseases, microbiology, medicine, and public health that is relevant to the control and prevention of neglected and emerging worldwide diseases that are a major cause of global morbidity, disability, and mortality.

Key Features

- Takes an integrated approach to infectious diseases.
- Provides the latest developments in the field of infectious diseases.
- Focuses on the contribution of evolutionary and genomic studies for the study and control of transmissible diseases.
- Includes updated and revised contributions from leading authorities as well as six new chapters.

About the Editor

Michel Tibayrenc has worked on the evolution of infectious diseases for more than 35 years. He is a Director of Research Emeritus and was head of the research unit, Genetics and Evolution of Infectious Diseases, for 20 years at the French Institut de Recherche pour le Développement (IRD). He is also the Founder and Editor-in-Chief (now emeritus) of *Infection, Genetics and Evolution* (Elsevier) and the Founder and principal organizer (until 2021) of the international congresses for MEEGID (molecular epidemiology and evolutionary genetics of infectious diseases). He has authored more than 200 international papers. With his collaborator Jenny Telleria, he is the Founder and Scientific Adviser of the Bolivian Society of Human Genetics (2012). Dr. Tibayrenc has been awarded the Belgian Society of Tropical Medicine (1985), the medal of the Instituto Oswaldo Cruz, Rio de Janeiro (2000) for his work on Chagas disease, and is a fellow of the American Association for the Advancement of Science (1993).

ISBN 978-0-443-28818-0

9 780443 288180

ELSEVIER

elsevier.com/books-and-journals