Infection, Genetics and Evolution 63 (2018) 295–306

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Mutation and recombination in pathogen evolution: Relevance, methods and controversies

Miguel Arenas[a,b,c,*], Natalia M. Araujo[d], Catarina Branco[b,c], Nadine Castelhano[b,c], Eduardo Castro-Nallar[e], Marcos Pérez-Losada[f,g]

[a] Department of Biochemistry, Genetics and Immunology, University of Vigo, Vigo, Spain
[b] Instituto de Investigação e Inovação em Saúde (i3S), University of Porto, Porto, Portugal
[c] Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP), Porto, Portugal
[d] Laboratory of Molecular Virology, Oswaldo Cruz Institute, FIOCRUZ, Rio de Janeiro, Brazil
[e] Universidad Andrés Bello, Center for Bioinformatics and Integrative Biology, Facultad de Ciencias Biológicas, Santiago, Chile
[f] Computational Biology Institute, Milken Institute School of Public Health, George Washington University, Ashburn, VA 20147, Washington, DC, United States
[g] CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, Vairão 4485-661, Portugal

ARTICLE INFO

Keywords:
Pathogen evolution
Recombination
Mutation
Evolutionary analysis
HBV evolution
HIV-1 evolution

ABSTRACT

Mutation and recombination drive the evolution of most pathogens by generating the genetic variants upon which selection operates. Those variants can, for example, confer resistance to host immune systems and drug therapies or lead to epidemic outbreaks. Given their importance, diverse evolutionary studies have investigated the abundance and consequences of mutation and recombination in pathogen populations. However, some controversies persist regarding the contribution of each evolutionary force to the development of particular phenotypic observations (e.g., drug resistance). In this study, we revise the importance of mutation and recombination in the evolution of pathogens at both intra-host and inter-host levels. We also describe state-of-the-art analytical methodologies to detect and quantify these two evolutionary forces, including biases that are often ignored in evolutionary studies. Finally, we present some of our former studies involving pathogenic taxa where mutation and recombination played crucial roles in the recovery of pathogenic fitness, the generation of interspecific genetic diversity, or the design of centralized vaccines. This review also illustrates several common controversies and pitfalls in the analysis and in the evaluation and interpretation of mutation and recombination outcomes.

1. Introduction

Mutation and recombination play a central role in the evolution of all organisms, but particularly in those that reproduce asexually. Mutation can be defined as the permanent alteration of the structure of a gene caused by the alteration of their single base units (DNA or RNA nucleotides), or the deletion, insertion, or rearrangement of larger sections of genes or chromosomes. Point mutations include the replacement of one nucleotide by another one (single nucleotide polymorphism, SNP) or the insertion or deletion of a single nucleotide. Single nucleotide mutation in a protein-coding gene may produce either a synonymous (silent) codon mutation, and thus no change in the encoded amino acid, or a nonsynonymous codon mutation, resulting in an amino acid change that may have an impact on the function of the encoded protein. A single nucleotide insertion or deletion in a protein-coding sequence will result in a frameshift such that downstream codons, including stop codons, will be translated from a different reading frame, leading to significant alteration of the encoded protein. Point mutations in non-protein-coding genes can also have functional consequences, particularly if they affect a regulatory element. Likewise, synonymous substitutions in protein-coding genes have functional consequences if they encode a low-frequency codon/anticodon pair (Sauna and Kimchi-Sarfaty, 2011; Telwatte et al., 2015). Although mutations that cause changes in protein sequences are usually harmful to organisms, on occasion, the effect may be positive in a given environment. If the latter, mutation may enable the mutant organism to withstand particular environmental stresses better than wild-type organisms, or reproduce more quickly. Mutation is the ultimate source of all genetic variation upon which other evolutionary forces such as recombination (rearrangement of genetic material), gene flow

* Corresponding author at: Department of Biochemistry, Genetics and Immunology, University of Vigo, 36310 Vigo, Spain.
E-mail addresses: marenas@uvigo.es (M. Arenas), nmaraujo@ioc.fiocruz.br (N.M. Araujo), cbranco@ipatimup.pt (C. Branco), nadines@ipatimup.pt (N. Castelhano), eduardo.castro@unab.cl (E. Castro-Nallar), mlosada@gwu.edu (M. Pérez-Losada).

http://dx.doi.org/10.1016/j.meegid.2017.09.029
Received 26 July 2017; Received in revised form 20 September 2017; Accepted 21 September 2017
Available online 22 September 2017

(migration), natural selection (differential survival and reproduction of individuals) and genetic drift (random changes in the frequency of a genetic variant in a population) can then operate. Mutations result mainly from errors during replication, which then may undergo repair or not if molecular mechanisms to check genome accuracy are lacking (e.g., single stranded RNA viruses like HIV). Consequently, mutation rates can vary several orders of magnitude across microbes. On average, bacteria show a mutation rate of 10^{-10} substitutions per base pair replicated, although some pathogenic organisms (hypermutators) can exhibit mutation frequencies up to 1000-fold greater than normal (Bridges, 2001). Double-stranded DNA (dsDNA) viruses, however, show mutation rates of 10^{-8} to 10^{-6} substitutions per nucleotide per cell infection (s/n/c), while RNA viruses show rates of up to 10^{-6} to 10^{-4} s/n/c (Pérez-Losada et al., 2015; Sanjuan et al., 2010).

"Sex" (e.g., via transduction, transformation, or conjugation in bacteria) and recombination generate or reshuffle existing variation generated via mutation. Recombination occurs when at least two genomes exchange genetic segments. Different types of recombination are recognized based on the structure of the crossover site (Austermann-Busch and Becher, 2012; Didelot and Maiden, 2010; Hanage et al., 2006; Scheel et al., 2013). Homologous recombination occurs in the same site in both parental strands, while non-homologous or illegitimate recombination (Hanage et al., 2006; Lai, 1992; Vos and Didelot, 2009) occurs at different sites of the genetic fragments involved, frequently originating aberrant (abnormal) structures (Galli and Bukh, 2014). A particular type of recombination (also known as shuffling or reassortment) occurs in some viruses with segmented genomes (e.g., influenza), which can interchange complete genome segments giving rise to new segment combinations. Recombination is a widespread phenomenon in microorganisms, particularly in viruses (Pérez-Losada et al., 2015). Indeed non-recombinant microorganisms are rare (Achtman, 2008), and recombination can also vary greatly between species (Didelot and Maiden, 2010; Pérez-Losada et al., 2015; Pérez-Losada et al., 2006; Vos and Didelot, 2009).

There is a potential interplay between mutation and recombination. Mutations provide a first source of diversity by altering specific positions and recombination shuffles those mutations by exchanging genetic fragments to further increase genetic variability. Recombination requires mutations to increase diversity but, if mutations are present, a single recombination event can provide a genetic variant that would require many mutation events to be obtained without recombination. Mutation and recombination mechanisms can thus be combined to impact microbial evolution and dynamics in different ways. They have been, for example, associated with the expansion of microbial host ranges, appearance of new species, alteration of transmission vector specificities, increase in virulence and pathogenesis, emergence of disease outbreaks, modification of tissue tropisms, vaccine escape, evasion of host immunity, and evolution of drug resistance (Chopra et al., 2003; Martin et al., 2011; Simon-Loriere and Holmes, 2011; Woodford and Ellington, 2007). Recombination can also prevent the progressive accumulation of harmful mutations in microbial genomes (i.e., preventing mutational meltdown and increasing fitness), and can also facilitate access to evolutionary innovations that would otherwise be inaccessible by mutation alone. Mutation and recombination processes also impact the structure of microbial populations. In a population in which recombination replacements are absent, diversification is slower, as it depends entirely on the accumulation of point mutations (Bachtrog and Charlesworth, 2002; Betancourt et al., 2009; Charlesworth et al., 2009); moreover, those populations tend to show high levels of linkage disequilibrium (the non-random association between alleles at different loci in a population) and are highly structured, comprising multiple independently evolving lineages. As the contribution of recombination to evolutionary change at neutral loci increases, distinct lineages become increasingly transient. At high ratios of recombination to mutation (e.g., in many RNA virus, Simon-Loriere and Holmes, 2011), independent microbial species isolates that are indistinguishable genotypically (i.e., descended from the same recent ancestor) cannot emerge because their genomes evolve too rapidly, hence producing a cluster of increasingly diverse genotypes (Feil et al., 1999; Pérez-Losada et al., 2006; Spratt et al., 2001). An interesting example of interplay between mutation and recombination is presented in Section 5.1.

Describing and understanding the population structure and diversity of microbial pathogens is of paramount importance if we aim to monitor and control drug resistance, bioterrorism, vaccination programs and emerging infectious diseases (Castro-Nallar et al., 2015; Feil and Enright, 2004; Pérez-Losada et al., 2017; Pérez-Losada et al., 2006; Pérez-Losada et al., 2013; Pérez-Losada et al., 2007b). Developing models that accurately describe pathogen evolution and the relative contributions of the evolutionary processes that generate this variation, chiefly point mutation and recombination, is inherently challenging because of the complexity of pathogens' life cycles and the difficulty in characterizing the (dynamic) fitness landscapes driving pathogen evolution (Feil et al., 2000; Feil and Spratt, 2001; Gog et al., 2015; Lloyd-Smith et al., 2015; Metcalf et al., 2015; Spratt et al., 2001).

Given these concerns, in this review we first assess, in a non-exhaustive fashion, the role of mutation and recombination at two different scales, intra-host and inter-host evolutionary dynamics (Sections 2 and 3). Several bioinformatic approaches have been developed to infer mutation and recombination occurrence in microbial genomes; some of those methodologies already take advantage of the genomic data generated using high-throughput sequencing. Here we will describe state-of-the-art analytical tools to detect and estimate mutation and recombination, including biases that are frequently ignored in evolutionary studies (Section 4). Finally, in Section 5, we discuss the role that mutation and recombination play in two well-known pathogens, HIV and HBV, in relation to the recovery of pathogenic fitness, the generation of interspecific genetic diversity and, the design of centralized vaccines. Here we present an introductory review and for further details about particular topics, we refer the reader to the specific literature cited in each section.

2. Intra-host evolutionary dynamics driven by mutation and recombination

Genetic variants resulting from the action of mutation and recombination can allow immune escape and therapy resistance. For instance, the rate of nonsynonymous substitutions was correlated with the rate of escape from neutralizing antibodies (Frost et al., 2001). This means that intra-host pathogen populations evolve differentially according to their ability to survive in environments with hard-living conditions imposed by natural selection, the immune system of the host and, if present, anti-pathogenic therapies. As a consequence, typical intra-host phylogenies are ladder-like, describing sequential replacement of pathogenic variants over time where only a few lineages coexist at any time-point but there is a rapid emergence of lineages though time (e.g., Bush et al., 1999; Lemey et al., 2006). In addition, sometimes the fate of intra-host pathogen populations can also be influenced by their effective population sizes (and their dynamics) and the generation time, especially in evolutionary scenarios driven by genetic drift (Frost et al., 2000; Lemey et al., 2006; Salemi, 2013; Shriner et al., 2004b).

The substitution rate in intra-host pathogen populations was shown to be heterogeneous over time, fitting with a relaxed molecular clock (Lemey et al., 2006), and can be explained by recurrent host-pathogen evolutionary arms races (Bliven and Maurelli, 2016; Didelot et al., 2016). This is a controversial scenario where during some periods of time molecular adaptation allows the pathogen to evolve rapidly by fixing a high proportion of mutations (Ghalambor et al., 2015; Maisnier-Patin and Andersson, 2004) and thus presenting many substitutions. However, during other periods, selection may force fixation of only a small proportion of mutations (many mutations are deleterious and thus removed from the population (Edwards et al., 2006))

leading to a few substitutions. This could explain why some studies found inverse relationships between observed mutation rate and disease progression (e.g., Delwart et al., 1997; Williamson et al., 2005) while others found the opposite (e.g., Lemey et al., 2007; Markham et al., 1998; McNearney et al., 1992). In addition, anti-pathogenic therapies can influence the observed rate of evolution since pathogens in patients treated with therapeutic drugs usually present more substitutions to develop resistance than those present in populations of untreated patients (e.g., Arenas, 2015a; Wu et al., 2003).

The importance of selection in intra-host pathogen populations has been frequently demonstrated with the estimation of the nonsynonymous/synonymous substitution rate ratio (dN/dS) (Markham et al., 1998) – the reader is referred to Section 4 for details about this estimation. This parameter can be used to investigate molecular adaptation not only for entire protein coding genes but also for site-specific positions, allowing the identification of primary and associated compensatory mutations (e.g., Arenas, 2015a; Nielsen and Yang, 1998; Yamaguchi and Gojobori, 1997). Note that a primary mutation is primarily responsible for a fundamental adaptation (i.e., maintenance of a protein function despite the drug), while compensatory mutations can yield small independent adaptive benefits of their own or adaptive benefits that offset fitness costs associated with primary mutations without compromising resistance (i.e., stabilization of a protein that suffered a primary mutation).

In addition to immune system responses, therapies can generate fitness losses in pathogen populations (Foll et al., 2014; Menendez-Arias et al., 2003) that favor the fixation of favorable (resistant) pathogenic variants (Poon et al., 2007a; Poon et al., 2007b). Drug resistance mutations constitute a beneficial escape mechanism to the inhibition of the protein function (i.e., altering the functional region of proteins to reduce the stability of the complex protein-drug) but also present fitness costs (i.e., generation of less stable proteins that may require additional mutations to maintain protein stability and function) (Melnyk et al., 2015). Next, these selective pressures can be observed by an excess of nonsynonymous substitutions (Arenas, 2015a; Pan et al., 2007; Poon et al., 2007b), which usually appear at specific sites related to molecular adaptation to the environment, and ultimately may describe complex mutational interaction networks (John et al., 2005; Poon et al., 2007b).

Recombination has been described as a relevant evolutionary force to increase genetic diversity in intra-host pathogen populations (Pérez-Losada et al., 2015; Poon et al., 2007b). As indicated above, it shuffles existing mutations to generate new variants that can then be better or worse adapted to the environment. The specific influence of recombination on the generation of drug resistance variants is still controversial. One can expect that recombination is important to generate drug resistance, as suggested by a computer simulation study (Carvajal-Rodriguez et al., 2007) and also by some other studies that found resistance in pathogenic variants derived from recombination events (Althaus and Bonhoeffer, 2005; Shi et al., 2010). However, other authors suggested that although recombination generates a wide diversity of variants, only a minority of recombination events are relevant for the evolution of the pathogen population (Archer et al., 2008). Indeed, some studies showed that recombination slows down the generation of resistance strains during therapy (Bretscher et al., 2004), which can be explained by the genetic barrier caused by recombination, where most of recombinant variants present low fitness and reduce the overall fitness of the population during treatment. However, if a recombinant variant is favored, its resistance mutations may be fixed in the viral population (Fraser, 2005) leading to an increase of observed nonsynonymous substitutions and speeding up adaptation (Fisher-Muller effect) (Kim and Orr, 2005). In conclusion, the role of recombination on the generation of drug resistance is not straightforward because other aspects such as the capacity to generate high fitness variants against a particular drug (fitness landscape) through the exchange of genetic material are crucial (Shi et al., 2010).

Recombination is also common in nonhuman pathogens (Croll et al., 2015; Le Gall-Recule et al., 2013; Yan et al., 2008). An interesting example is the recombination between an American bat virus and a carnivore virus (Ding et al., 2017). This recombination allowed the bat virus to acquire a protein domain (head of the G-protein ectodomain) of the carnivore virus that is involved in the response to changes in pH, trimerization of G proteins and cell-to-cell transmission during the infection. That event improved the adaptability of the virus to carnivores, leading to important epidemics in striped skunk (Ding et al., 2017).

3. Inter-host evolutionary dynamics driven by mutation and recombination

While similar evolutionary patterns may arise from intra- and inter-host observations, specific evolutionary dynamics rely on ecological and intrinsic biological properties of microorganisms, which to some extent determine their evolutionary trajectories (Castro-Nallar et al., 2012a; Rambaut et al., 2008). For instance, HIV-1 and HBV intra-host evolutionary dynamics are characterized by ladder-like phylogenies (previous section) that greatly resemble inter-host phylogenies of other pathogens such as Influenza A (Rambaut et al., 2008). In this regard, the study of inter-host evolutionary dynamics provides insights into epidemiological/ecological processes from local to global scales (Faria et al., 2017; Grubaugh et al., 2017). Phylodynamics is particularly applicable to microorganisms that exhibit high de novo evolutionary change among time-stamped sequences, i.e., Measurably Evolving Populations (Drummond et al., 2003). Usually this type of data includes a handful of genes in fast-evolving RNA viruses or genome-wide SNPs in bacterial species (Biek et al., 2015).

HIV-1 is probably one of the most heavily studied pathogens at the inter-host level, thus we have some knowledge about the relative contributions of mutation and recombination to the associated pandemic processes. However, in the mid-1990s recombination in HIV-1 was regarded as rare to non-existent due to the belief that multiple infections within the same individual, which were deemed unlikely, were needed for the emergence of recombinant virions. Recombination in retroviruses was experimentally demonstrated as early as 1991 in feline and murine species (Golovkina et al., 1994; Temin, 1991; Tumas et al., 1993). The first evidence of HIV recombination came from phylogenetic analysis (Diaz et al., 1995; Robertson et al., 1995). Nowadays, recombination in HIV is considered pervasive (Shriner et al., 2004a). Epidemiologically-relevant recombinant forms have been readily identified in areas of historically high HIV genetic diversity such as Sub-Saharan Africa, where Circulating Recombinant Forms (CRFs) have been shown to be stable over time (Lihana et al., 2012). In addition, stable CRFs have also been reported, and seem to persist and co-exist alongside other HIV lineages. For instance, researchers have reported in Angola rare HIV-1 subtypes and CRFs from archival 1993 samples that suggest early extensive diversity possibly due to migration from other recognized centers of HIV diversity such as the Democratic Republic of Congo and the Republic of Congo (Bartolo et al., 2016). Similarly, complex CRF forms (CRF_cpx) that result from the recombination of more than two subtypes, have been identified primarily in Africa, Greece, Cyprus, and Cuba. Recently, researchers traced the time-scale and spread of several CRFs_cpx and found that they have been circulating since the 1950s and, in the case of CRF11_cpx, migrated from Cameroon to other African countries (Delatorre and Bello, 2016). These results suggest that the contribution of mutation and recombination has been important since the beginning of the epidemic (in contrast to the first thoughts in the mid-1990s) and that these evolutionary forces provide a great diversity of subtypes and CRFs that persist over time.

Mutation and recombination have shaped the inter-host evolutionary history of HIV in a way that is consistent with a selectively neutral process, where several lineages persist over time. This suggests that HIV inter-host evolution is rather governed by demographic processes instead of by natural selection (Lemey et al., 2006). The

immediate consequence of these evolutionary processes has been the worldwide emergence of viral subtypes and CRFs (Castro-Nallar et al., 2012b), whose contribution to the development of a yet elusive protective vaccine cannot be ignored (Barouch et al., 2015; Barouch et al., 2013).

High recombination and substitution rates in HIV have also helped to determine its global origin, time of emergence and local epidemic patterns, both through phylogenetic and distance network approaches (Gao et al., 1999; Wertheim et al., 2014; Worobey et al., 2016). While the date of emergence of HIV gets pushed back as more archival sequences are sequenced, the current estimated time for the most recent common ancestor is 1920 (95% CI 1909–1930), a date that matches colonial ruling in West Africa, the sexual revolution and changes in health care practices (Ash, 2014; Faria et al., 2014). Thus, in retroviruses such as HIV, mutation and recombination underpin viral global distribution and pathogenesis, as well as facilitate characterization of infections in space and time.

Reassortment is a unique type of recombination that occurs in viruses whose genome is segmented. The genome of viral pathogens such as Influenza A is segmented into eight molecules, each one coding for a major protein. In those cases, the independent shuffling of segments during co-infection is thought to lessen the frequency of recombination because the genome is not constrained by intramolecular covalent bonds. Within the cell, genetic segments that possess the same packaging signal can generate hybrid viral particles. Consequently, the genetic effects of reassortment are like those of recombination, i.e., purging deleterious haplotypes from a population and increasing population-level genetic diversity for those recombinant forms that exhibit increased fitness (for a review on reassortment in segmented RNA viruses the reader is referred to McDonald et al., 2016), see also the reviews (Bernstein et al., 1985; Felsenstein, 1974). However, whether reassortment is a by-product or a driver of evolution is still controversial since fast-evolving viruses do not rely upon reassortment to overcome selective pressures (Ojosnegros et al., 2011). Potential advantages of segmented genomes might be related to better control of gene expression and/or increased coding-potential of the genome (genome economy), or even increased stability of viral particles (Ojosnegros et al., 2011). One key difference between recombination and reassortment is that the former can occur anywhere on the genome and thus can produce chimeric genes, while reassortment is restricted to the swapping of complete segments with functional genes. Recombination is not readily detected in segmented negative sense RNA viruses and has been rarely reported in the literature (Boni et al., 2012; Boni et al., 2008). In principle, this observation would suggest that reassortment can complement the selective advantage of recombination. However, it is unclear whether these processes are directly favored by natural selection or reflect viral life histories (Simon-Loriere and Holmes, 2011). Nevertheless, there is evidence of non-homologous recombination in non-human segmented RNA viruses, e.g., the fish pathogen Infectious Salmon Anemia Virus (ISAV; *Orthomyxoviridae*) (Castro-Nallar et al., 2011; Devold et al., 2006; Kibenge et al., 2009). In ISAV, non-homologous recombination can provide new protease cleavage sites that might be involved in increasing virulence, which has been observed in more than one outbreak (Godoy et al., 2014).

In the case of Influenza viruses, reassortment contributes to hosts jumps and drug resistance, and is a primary force behind the emergence of seasonal Influenza outbreaks (Vijaykrishna et al., 2015). The different combinations of segments (antigenic shift) and novel point mutations (antigenic drift) are primarily responsible for the low effectiveness of the annual flu vaccine, which can range from 13 to 53% depending on age group (Flannery et al., 2017). High substitution rates have also helped to reveal global phylogeographic patterns in Influenza A H3N2 (seasonal strain), which shows high interconnectedness in Asia, while in the Western Hemisphere, the USA appears as a major hub for global migration (Bedford et al., 2010). This analysis and others have led to the development of models of global epidemiology for Influenza A viruses, where genomic evolution is characterized by frequent reassortment and selective sweeps (Rambaut et al., 2008). This model is congruent with observations relating high Influenza diversity among wild aquatic birds (their natural reservoir) that migrate from the tropics to temperate regions of the planet (Vijaykrishna et al., 2015). The factors that dictate the observed patterns remain poorly understood, however, simulation studies have weighed the relative contribution of seasonality, local basic reproductive number, and population size. These studies suggest that local basic reproductive number strongly affects antigenic evolution and, in turn, is related to the probability of a strain spreading and fixing globally (Wen et al., 2016). Through the phylogeographic analysis of different strains and types of Influenza, we have discovered that global circulation patterns are not universal but highly dependent on strain type. For example, A/H3N2 strains do not persist between epidemics, while A/H1N1 and B strains persist over several seasons and exhibit complex global dynamics (Bedford et al., 2015).

High substitution and recombination rates are not unique to viruses. At genome-wide scales, bacterial pathogens can also accumulate variation over their larger genomes at time-scales that allow for the inference of epidemiological/ecological processes, i.e., Measurably Evolving Pathogens (Biek et al., 2015). Substitution rates in bacteria can vary widely from $\sim 10^5$ to $\sim 10^9$ nucleotide substitutions per site per year, as exemplified by *Neisseria gonorrhea* and *Mycobacterium tuberculosis*, respectively (Comas et al., 2013; Pérez-Losada et al., 2007a). As in viruses, bacterial mutation rates vary across the genome by up to an order of magnitude and are skewed towards certain types of misincorporations, which generate hot and cold spots (Jee et al., 2016). When systematically studied, substitution rates in bacteria were found to be temporally structured, varying negatively with time scales. Over short-time scales, one would expect rates to be higher due to the presence of deleterious mutations yet to be purged from the population and vice versa (Duchene et al., 2016).

Researchers have taken advantage of the amount of variation in bacteria to uncover phylogeographic patterns related to global migration and origins, epidemics, and clinical and public health aspects of outbreaks (Hasan et al., 2012; Holt et al., 2012; Njamkepo et al., 2016; Rohde et al., 2011). Likewise, recombination in bacteria can be associated with the acquisition of novel traits, e.g., antibiotic resistance (see previous section). As with nucleotide substitutions, recombination has been detected in genes with cell surface/membrane functions causing virulence (e.g., *Escherichia coli*). Conversely, “cold regions” tend to be associated with housekeeping functions, e.g., ribosomal proteins (Yahara et al., 2016). Homologous recombination has been associated with the spread of mutations that confer antibiotic resistance; for example the spread of 23S alleles among species of *Streptococcus* (Billal et al., 2011) and of penicillin-binding protein (*pbp*) genes in *Neisseria gonorrhea* (Unemo et al., 2012). Notably in bacteria, mobile genetic elements like transposons and integrons can capture and spread antibiotic resistance genes across species boundaries. Integrons possess a site-specific homologous recombination system; they have been linked to clinical *E. coli* strains as well as industrial contamination (Vinue et al., 2011; Wright et al., 2008) and are currently proposed as a potential marker for anthropogenic contamination (Gillings et al., 2015).

4. Models, methods and computational frameworks for estimating mutation and recombination

Several statistical approaches have been considered for the estimation of mutation and recombination, although currently, the most commonly applied are the maximum likelihood (ML) and Bayesian approaches. In this section, we revise models, methods and frameworks to analyze mutation and recombination and also acknowledge important biases that should be considered when carrying out such analyses.

Concerning mutation, although some programs (e.g., *BEAST*

Table 1
Evolutionary methods and frameworks commonly used to estimate substitution rates. The table describes some key capabilities for each program: approach used to perform the estimation (ML and ABC refers to maximum-likelihood and approximate Bayesian computation, respectively), type of input genetic data, estimation assuming a substitution model (+I and +G refers to proportion of invariable sites and heterogeneous substitution along the sequence according to a gamma distribution, respectively), operating systems and references.

Program	Approach	Type of data	Substitution model	Operating systems	References
BEAST	Bayesian	DNA	K2P, HKY85, GTR, JC, TN93, F81, +I +G	Linux, Mac, Windows	(Bouckaert et al., 2014)
		Coding DNA	GY94, +I +G		
		Amino acid	Blosum62, CPREV, Dayhoff, JTT, MTREV, WAG, +I +G		
OmegaMap	Bayesian	Coding DNA	NY98	Linux, Mac, Windows	(Wilson and McVean, 2006)
Lamarc	Bayesian, ML	DNA	JC, K2P, F84, GTR, +G	Linux, Mac, Windows	(Kuhner, 2006)
MEGA	ML	DNA	JC, K2P, F81, HKY85, TN93, T3P GTR, +I +G	Linux, Mac, Windows	(Kumar et al., 2016)
mlRho	ML	DNA[a]	–	Linux	(Haubold et al., 2010)
CodABC	ABC	Coding DNA	GY94 +I +G	Linux	(Arenas et al., 2015)

[a] Multiple contigs derived from a multiple alignment of DNA sequences (assembled reads).

(Bouckaert et al., 2014)) implement the estimation of the mutation rate per generation per site (μ), most computer programs estimate mutation rates at the population level as $\Theta = 2nN\mu L$, where $n = 1$ or 2 for haploid or diploid populations, respectively, N is the effective population size and L is the sequence length. Thus, the estimation of Θ may not be straightforward due to its dependence on N (population size fluctuations could bias the estimation). Several computer programs exist to perform this estimation from genetic data; Table 1 shows a list of these programs and their capabilities. The choice of the appropriate software usually depends on the genetic marker under analysis, previous knowledge of the parameter (i.e., prior distributions in Bayesian and approximate Bayesian computation approaches) and computational time (Bayesian computation generally requires longer times of analysis).

The ratio of nonsynonymous to synonymous substitutions (*dN/dS* or ω) can provide an interesting measure of selection in protein-coding data. The estimation of this parameter, at both global (entire sequence) and local (codon specific) levels, has been implemented in a variety of evolutionary frameworks based on different statistical approaches (Table 2). Current frameworks are adapting traditional approaches to estimate *dN/dS* from allelic mixtures (i.e., derived from next-generation sequencing (NGS)) (e.g., Gonzales et al., 2002) and to remove biases that other evolutionary processes may induce (i.e., recombination) (e.g., Arenas et al., 2015). Concerning the latter, the estimation of *dN/dS* can be biased towards the generation of false positively selected sites (PSS) if recombination is ignored (Anisimova et al., 2003; Arenas and Posada, 2010a, 2014), which is relevant since many pathogens show high levels of recombination. This bias is caused by the different evolutionary histories that recombination fragments may present (Schierup and Hein, 2000), depicting an ancestral recombination graph (Arenas, 2013; Griffiths and Marjoram, 1997) that must be considered for posterior analyses (Arenas and Posada, 2010c; Mallo et al., 2016). Under recombination, one can either estimate *dN/dS* for each particular recombinant fragment (in this case note that sequence fragments involved in recombination events must be previously identified) (e.g., Pérez-Losada et al., 2011; Pérez-Losada et al., 2009) (Fig. 1), or use a method that co-estimates recombination and *dN/dS* (see Table 2).

Current analytical tools to analyze recombination can detect (recombination tests), map (position of recombination breakpoints) and quantify (recombination rate) this evolutionary force. Several frameworks have been developed to analyze recombination; a comprehensive list is kindly provided by David Robertson's lab http://bioinf.man.ac.uk/robertson/recombination/programs.shtml. In Table 3 we describe the capabilities of the most commonly used frameworks to estimate recombination rates and breakpoints.

A variety of tests exist to detect the presence/absence of recombination (Posada, 2002; Wiuf et al., 2001), which generally are more accurate with high genetic diversity (Posada, 2002). Importantly,

Table 2
Evolutionary methods and frameworks commonly used to estimate molecular adaptation through the nonsynonymous/synonymous substitution rate ratio. The table describes some key capabilities for each program: approach used to perform the estimation (ML and ABC refers to maximum-likelihood and approximate Bayesian computation, respectively), codon substitution model in terms of exchangeability matrix (+I and +G refers to proportion of invariable sites and heterogeneous substitution along the sequence according to a gamma distribution, respectively) and variation of *dN/dS* along sequences, consideration of recombination rate in the estimation of *dN/dS*, operating systems and references.

Program	Approach	Codon substitution model		Consideration of recombination	Operating systems	References
		Exchangeability matrix	Heterogeneous dN/dS			
OmegaMap	Bayesian	NY98	Yes[a]	Yes	Linux, Mac, Windows	(Wilson and McVean, 2006)
Hyphy/Datamonkey	ML	GY94, MG94, +I +G	Yes	No[b]	Linux, Mac, online, Windows	(Kosakovsky Pond and Frost, 2005; Kosakovsky Pond et al., 2005)
PAML	ML	GY94, NG86, LPB93, LWL85, YN00, +G	Yes	No	Linux, Mac, Windows	(Xu and Yang, 2013; Yang, 2007)
KaKs_Calculator	ML, Approximate methods (counting rule)	NG, LWL, LPB, YN, +G	Yes[c]	No	Linux, Windows	(Wang et al., 2010)
CodABC	ABC	GY94, +I +G	No	Yes	Linux	(Arenas et al., 2015)
SNAP	Approximate (counting rule)	NG86	No	No	Linux, online	(Korber, 2000; Nei and Gojobori, 1986)
START	Approximate (counting rule)	NG86	No	No	Linux, Mac, Windows	(Jolley et al., 2001)
Syn-SCAN	Approximate (counting rule)	NG86 accounting for ambiguous nucleotides	No	No	Linux, Windows	(Gonzales et al., 2002)

[a] The program estimates *dN/dS* for each user-specified partition.
[b] Recombination can be considered if the user provides a tree for each recombinant fragment into the input file (nexus MSA).
[c] The program can estimate variable *dN/dS* through a sliding window.

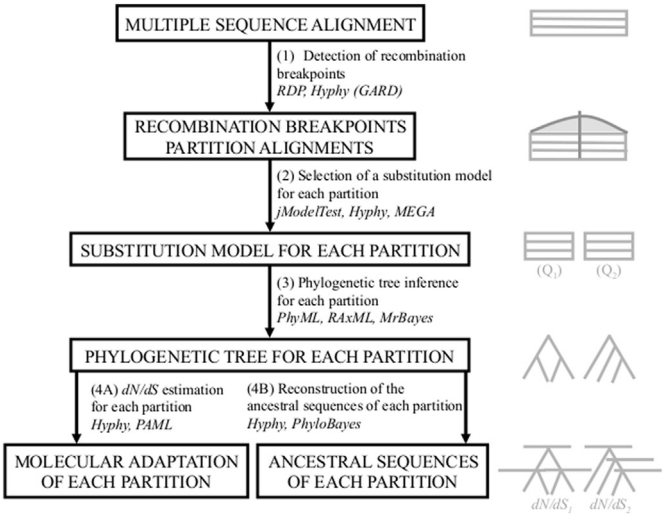

Fig. 1. Pipeline for estimating molecular adaptation and ancestral sequences accounting for recombination. (1) Recombination breakpoints can be estimated from a multiple sequence alignment (MSA), which can then be split into a set of smaller MSAs (partitions) according to the breakpoints. Note that the number of MSAs is equal to the number of breakpoints plus one. In the illustrative example (on the right), a single breakpoint is depicted, leading to two partitions. (2) A substitution model of molecular evolution (*Q*) can be identified for each partition. (3) A phylogenetic tree for each partition can be inferred according to the corresponding substitution model. (4A) Molecular adaptation (*dN/dS*) can be estimated for each partition. (4B) Ancestral sequences for each partition can be inferred. For nodes that are common in both partitions (i.e., the root node), the ancestral sequences of each partition can be concatenated to obtain the entire ancestral sequence (further details in Arenas and Posada, 2010c).

rate variation among sites, which is frequent in many pathogens (Worobey, 2001), can bias the detection of recombination by generating false positives (Posada, 2002; Worobey, 2001), although some methods account for this effect (Husmeier and Wright, 2001). Tests based on homoplasy performs better under low levels of genetic diversity (Bruen et al., 2006), whereas other methods such as chi-square tests perform better when genetic diversity is high (Smith, 1992). Altogether, the use of several recombination tests is recommended to achieve a well-supported identification (Posada, 2002).

Recombination is usually not homogeneous along sequences (e.g., Archer et al., 2008) and thus breakpoint identification can be useful for a variety of purposes such as the characterization of recombination forms (e.g., Archer et al., 2008; Minin et al., 2007), investigating adaptive recombination (e.g., Arenas et al., 2016; Monjane et al., 2011) and molecular mechanisms of recombination (Pérez-Losada et al., 2015), and to properly carry out other evolutionary analyses such as the estimation of *dN/dS* or ancestral sequences (Arenas and Posada, 2010a, 2010b, 2010c; Pérez-Losada et al., 2011; Pérez-Losada et al., 2009) (Fig. 1). Despite the large number of programs available to detect recombination breakpoints, some of them (e.g., *RDP* (Martin et al., 2010)) are well-established mainly due to the implementation of a graphical user interface (GUI). Other programs were designed to analyze large datasets in short computing times by implementing algorithms that run in parallel on a multiprocessor system (e.g., the *GARD* method (Kosakovsky Pond et al., 2006) implemented in the *Hyphy* package (Kosakovsky Pond et al., 2005)).

Few programs exist that can estimate the recombination rate per generation per site (*r*), instead most programs estimate the population recombination rate $\rho = 2nNrL$. The accurate estimation of ρ may also not be straightforward due to its dependence on *N*. Several computer programs implementing different strategies have been developed to estimate recombination rates (Table 3). We highlight the widely used packages *LDHat* (McVean et al., 2004), *Lamarc* (Kuhner, 2006), *RDP* (Martin et al., 2015) and *OmegaMap* (Wilson and McVean, 2006). *LDHat* is one of the most relevant tools for estimating recombination rates and is based on the implementation of the coalescent composite likelihood method of Hudson (Hudson, 2001) into a Bayesian approach. *Lamarc* implements both ML and Bayesian approaches and, in addition to the recombination rate, it can estimate other population genetics parameters such as population growth and migration rates. From our experience, convergence between independent runs can be difficult to

Table 3
Evolutionary methods and frameworks commonly used to estimate recombination rate and recombination breakpoints. The table describes some key capabilities for each program: approach used to perform the estimation (ML and ABC refers to maximum-likelihood and approximate Bayesian computation, respectively), type of input genetic data, estimation of homogeneous recombination rate (recombination rate is assumed constant along the sequence) and heterogeneous recombination rate (recombination rate can vary along the sequence), detection of recombination breakpoints, operating systems and references.

Program	Approach	Type of data	Recombination rate		Recombination breakpoints	Operating systems	References
			Homogeneous	Heterogeneous			
RDP	Bayesian	DNA	Yes	Yes	Yes	Windows	(Martin et al., 2015)
LDhat	Bayesian	DNA, SNP	Yes	Yes	No	Linux	(McVean et al., 2004)
OmegaMap	Bayesian	Coding DNA	Yes	Yes[a]	No	Linux, Mac, Windows	(Wilson and McVean, 2006)
LDhelmet	Bayesian	DNA, SNP	Yes	Yes	No	Linux	(Chan et al., 2012)
DualBrothers	Bayesian	DNA	No	No	Yes	Linux	(Minin et al., 2005)
PhyML_Multi	Bayesian[b]	DNA	No	No	Yes	Linux, Mac	(Boussau et al., 2009)
jpHMM	Bayesian[b]	DNA[c]	No	No	Yes	Linux, online	(Schultz et al., 2009)
Lamarc	Bayesian, ML	DNA	Yes	No	No	Linux	(Kuhner, 2006)
mlRho	ML	DNA[d]	Yes	No	No	Linux	(Haubold et al., 2010)
Hyphy/Datamonkey (GARD)	ML	DNA	No	No	Yes	Linux, Mac, online, Windows	(Kosakovsky Pond and Frost, 2005; Kosakovsky Pond et al., 2005; Kosakovsky Pond et al., 2006)
FastEPRR	Machine Learning	DNA	Yes	Yes[e]	No	Linux, Windows	(Gao et al., 2016)
Recco	Cost optimization	DNA	No	No	Yes	Linux, Mac, Windows	(Maydt and Lengauer, 2006)
CodABC	ABC	Coding DNA	Yes	No	No	Linux	(Arenas et al., 2015)

[a] The program estimates a recombination rate for each user-specified partition.
[b] Hidden Markov Model (HMM).
[c] Can only be applied to HBV and HIV-1 data.
[d] Multiple contigs derived from a multiple alignment of DNA sequences (assembled reads).
[e] Estimation of recombination rate in user-specified windows.

reach in *LDHat* and *Lamarc*, especially for large datasets. The package *RDP* implements an adapted version of the Bayesian approach of *LDhat* into a user-friendly GUI. From our experience, *RDP* is robust in reaching convergence under the settings specified by default. The program *OmegaMap* estimates recombination rate, as well as *dN/dS* and Θ, in user-specified blocks (partitions) with a Bayesian approach based on the product of approximating conditionals (PAC) likelihood (Li and Stephens, 2003). Interestingly, approximate Bayesian computation (ABC) (Beaumont et al., 2002) has been presented as an alternative approach to provide more accurate estimations of the recombination rate (and other evolutionary parameters such as *dN/dS* and Θ) (Lopes et al., 2014; Wilson et al., 2009). The advantage of ABC is the co-estimation of different evolutionary parameters, thus accounting for the effects of every parameter on the others (Lopes et al., 2014). By contrast, this approach requires extensive and realistic computer simulations, which may lead to long computer times. The program *CodABC* (Arenas et al., 2015) implements this ABC approach and can be conveniently run in parallel on a multiprocessor machine.

Some care must be taken when estimating recombination in sequences obtained via NGS. The assembly of reads or fragments may generate artifacts leading to false recombination signatures (Prosperi et al., 2011; Zanini et al., 2016). Under this scenario, some complex methods have been proposed to estimate recombination in NGS data through experimental (Beerenwinkel et al., 2012; Laureau et al., 2016) and computational (Gao et al., 2016; Johnson and Slatkin, 2009; Marttinen et al., 2012; Zhang, 2013) approaches.

5. Exemplar studies: mutation and recombination at play

This section describes some studies performed in our laboratories that describe the roles of mutation and recombination in different areas of pathogen evolution, including controversies related with the interpretation of some results.

5.1. HIV-1 in vitro fitness recovery

Population dynamics and molecular mechanisms can affect viral fitness and, consequently, viral evolution. Concerning the evolutionary mechanisms at the molecular level, an effective mutational process is clearly fundamental to increase fitness (e.g., Lorenzo-Redondo et al., 2011; Poon et al., 2007b). However, the specific contribution of recombination to fitness is controversial. Although recombination can generate better adapted variants, it could also break down existing good ones (Hadany and Beker, 2003). Some studies showed that most recombinant forms present low fitness, especially when evolving under strong selective pressures (Bretscher et al., 2004; Nijhuis et al., 1998). By contrast, other studies indicated that recombination can increase fitness through particular mutations and epistatic interactions (Carobene et al., 2009; Moradigaravand et al., 2014). Next, the influence of recombination on viral fitness is probably affected by factors such as the mutation-selection balance (equilibrium between deleterious mutation and selection to remove deleterious variants) (Hadany and Beker, 2003; Moradigaravand and Engelstadter, 2012). Therefore, to investigate the specific impact of recombination on viral fitness, in vitro studies can be useful since they avoid some of the in vivo processes (i.e., compartmentalization, transmission between individuals or antiviral therapies) that may confound the detection of recombination.

We have recently investigated the influence of mutation and recombination on HIV-1 in vitro fitness recovery (Arenas et al., 2016). This study was based on plaque-to-plaque passages of several HIV-1 biological clones (Ebert, 1998) that generated drastic fitness losses (Yuste et al., 1999), followed by a process of fitness recovery during a total of 30 large population passages (Arenas et al., 2016). An advantage of working with in vitro populations is the possibility of monitoring the evolutionary process. In that study we evaluated the amount of observed mutation and recombination in the population each 10 population passages. We found an increase of non-synonymous substitutions with the number of passages that correlated with a fitness increase, suggesting that HIV-1 viruses evolved towards new and more adapted molecular variants (Arenas et al., 2016).

Concerning recombination, its amount was also correlated with fitness increase, although not for all the studied clones. Interestingly, for many clones, recombination increased with each passage, while fitness increased only in the consecutive passage. This finding may be explained by two processes: (*i*) some mutations must be fixed after the recombination event in order to optimize the new recombinant form and thus increase fitness, and/or (*ii*) the new recombinant form needs to reach a given frequency in the viral population to present a detectable influence on fitness (Galli et al., 2010; Iglesias-Sanchez and Lopez-Galindez, 2002). A few clones did not require recombination to increase fitness, suggesting that the mutation process was sufficient and new recombination forms could be deleterious and so, not observable (Archer et al., 2008; Iglesias-Sanchez and Lopez-Galindez, 2002). Finally, one clone presented recombination but not a fitness increase, which could be explained by the generation of neutral recombination forms (Archer et al., 2008; Bretscher et al., 2004; Nijhuis et al., 1998). Altogether, only those recombination events that generate advantageous recombinant forms, which are usually a minority (see Archer et al., 2008), can generate a fitness increase.

In conclusion, mutation and recombination seem to allow viral populations to move on a complex fitness landscape (Lorenzo-Redondo et al., 2014) where a recombination event might generate a “big jump”, leading to a new form that could present a totally different fitness level, and a mutation event might generate a “small jump”, leading to a new form that could present a small fitness variation. Thus, the need for recombination or mutation events to increase fitness can be related to the shape of the fitness landscape of the studied viral population and the position of the viral population in such a landscape. These landscapes suggest nonrandom evolutionary trajectories to recover fitness.

5.2. Genetic diversity in HBV populations from Argentina and Brazil

Hepatitis B virus (HBV) infection is a major public health problem worldwide. An estimated 257 million people are chronically infected with HBV, and > 800,000 people die every year due to complications of hepatitis B, including cirrhosis and hepatocellular carcinoma (WHO, 2017). The HBV genome is a partially dsDNA molecule, ~3.2 kb in length, which has a highly compact coding structure with most genes located in overlapping reading frames. Despite the severe constraints imposed by this genomic organization, there is great diversity among HBV isolates. Based on a sequence divergence of > 8% in the entire genome, HBV is currently grouped into 10 genotypes (A–J) and a growing number of subgenotypes spread worldwide (Tong and Revill, 2016). This large genetic variability results from accumulating mutations due to an error-prone HBV reverse transcriptase. The estimated nucleotide substitution rate ranges between 10^{-4} and 10^{-6} substitutions per site per year (Osiowy et al., 2006; Paraskevis et al., 2015; Zhou and Holmes, 2007), which is higher than that observed in most dsDNA viruses.

In addition, recombination events between HBV variants over the years played a major role in the emergence and establishment of the currently classified HBV genotypes (Castelhano et al., 2017; Simmonds and Midgley, 2005). Novel HBV hybrids generated by both inter- and intra-genotype recombination have been increasingly documented worldwide (e.g., Araujo, 2015; Bollyky et al., 1996; Morozov et al., 2000; Simmonds and Midgley, 2005; Su et al., 2014). Moreover, some recombinant strains have become dominant variants in certain geographic regions, demonstrating potential for spreading and developing their own epidemiology. This is the case of subgenotype B2 (B/C recombinant), which became prevalent in mainland East Asia (Sugauchi et al., 2002), and the C/D recombinant, which is the most common circulating strain in Western China (Tibet) (Cui et al., 2002).

In a previous review, our group performed a large-scale data retrieval and analysis on HBV complete intergenotypic genomes (Araujo, 2015). We identified 436 recombinants that included all but one (H) HBV genotypes. Most recombinant forms (60%) were either B/C or A/D hybrids. The other combinations were A/B/C, A/C, A/C/G, A/D, A/E, A/G, B/C/U (U = unknown genotype), C/F, C/G, C/J, D/E, D/F and F/G. In agreement with previous reports (Hino et al., 1991; Pineau et al., 1998; Simmonds and Midgley, 2005), several regions with recombination hotspots (i.e., nt 1700–2000, and nt 2100–2300) were identified and showed a statistically significant increase of recombination breakpoints in comparison with the rest of the genome. Recombination events involving these genomic regions (pre-C/C and X genes) can transfer several relevant mutations, such as those related to viral replication, HBeAg expression, and hepatocarcinogenesis. Next, the worldwide distribution of HBV inter-genotypes showed that their circulation patterns are similar to those of their original genotypes. For instance, recombinants of the globally spread genotypes A and D are also found worldwide, whereas hybrids of other genotypes are restricted to specific regions. Indeed, recombinant forms between genotype F and other HBV genotypes have been identified almost exclusively in South American countries such as Argentina (D/F and F/G), Brazil (F/G) and Bolivia (C/F) (Araujo, 2015). As an additional example, in 2013, we provided the first full-length genomic sequences of F/G hybrids (Araujo et al., 2013). HBV genomes were found in two HIV-positive patients from Argentina and Brazil, both having high serum HBV loads ($> 10^8$ copies/ml). Beside "pure" strains of genotypes F (subgenotype F4) and G, several F/G recombinants comprising genomes with totally different arrangements were characterized. This finding indicates that different recombination events occurred in a single patient. It is noteworthy that a single recombination event can transfer several mutations from one strain to another, thus generating hybrids with an enhanced virulence. In that case, both recombinant and non-recombinant genomes carried several clinically relevant mutations, some of them previously associated with hepatocarcinogenesis (Araujo et al., 2013). All this evidence indicates that recombination is crucial for generating novel HBV inter-genotypes and viral forms with increased virulence.

5.3. Centralized vaccines design

Among the large variety of strategies to design vaccines against pathogens with high mutation and recombination rates, an interesting strategy from the evolutionary perspective is based on the usage of centralized genes to obtain centralized vaccines (Arenas and Posada, 2010b; Doria-Rose et al., 2005; McBurney and Ross, 2007). Centralized genes maintain the immunogenic properties of the population while minimizing the genetic distance to the descendant target strains (Gao et al., 2003). The first step to design centralized vaccines consists in the inference of centralized sequences such as: (*i*) the consensus sequence (CON) (e.g., Ellenberger et al., 2002; Novitsky et al., 2002), based on the abundance of states, (*ii*) the ancestral sequence (ANC) (e.g., Doria-Rose et al., 2005; Kothe et al., 2006), the sequence of the most recent common ancestor (MRCA) and, (*iii*) the center-of-tree sequence (COT) (Rolland et al., 2007), the sequence of the point in the phylogeny with the smallest evolutionary distance to each sample node. Centralized vaccines have been mainly applied to HIV-1 populations and, unfortunately, have only produced partial T-cell responses (Frahm et al., 2008). Several factors could potentially lead to the low protective efficacy of these vaccines, including: the large genetic diversity existing in the population (Vidal et al., 2000), low sample size (i.e., underestimated population diversity), or analytical issues (inaccurate or lack of estimation of recombination or substitution rates) in the computational inference of the centralized genes (Arenas and Posada, 2010b; Ross et al., 2006). Concerning the latter, it has been demonstrated that the best fitting substitution model of evolution must be considered for probabilistic evolutionary analysis of genetic data (Posada and Crandall, 2001). Although model selection is traditionally considered, recent progress in this field suggest that different genetic regions evolve under different substitution models (Arbiza et al., 2011; Arenas, 2015b), leading to the need for frameworks modeling heterogeneous substitution along the sequence (Lartillot and Philippe, 2004). The consideration of more complex substitution models (i.e., non-reversible, non-stationary or accounting for molecular constraints) (Arenas, 2015b; Wilke, 2012) could also improve the inference of centralized genes (Arenas and Posada, 2010b).

Recombination has been often ignored in the inference of centralized sequences, although both ANC and COT sequences are based on phylogenetic tree inferences that can be biased if recombination is ignored (Arenas and Posada, 2010c). Moreover, sequence errors derived from ignoring recombination can affect the inference of epitopes and *N*-glycosylation sites (Arenas and Posada, 2010c), which can be crucial for the protein function. Future approaches to estimate centralized sequences need to take recombination into account when applying standard phylogenetic methodologies by splitting recombinant gene sections according to recombination breakpoints or by using network approaches (Arenas and Posada, 2010c; Mallo et al., 2016).

6. Conclusions

Mutation and recombination are essential evolutionary processes for most pathogens to adapt to rapidly changing environments such as host-microbe immune interactions or drug therapies. At the population level, these forces can also generate the genetic diversity needed to cause, maintain, and expand epidemics. Consequently, epidemiological studies of pathogens often include analysis of the magnitude and relevance of these evolutionary forces to characterize the past dynamics of epidemics (i.e., microbial population growth or decline), transmission chains (e.g., chains of infection in forensic cases) (Budowle and Williamson, 2011) or even to make predictions about the future of pathogen populations (i.e., forecasting the future of genetic variants for vaccine selection) (Luksza and Lassig, 2014). However, there are still many studies ignoring fundamental methodological aspects of pathogen evolution that can result in severe biases. First, the consideration of a properly fit substitution model is well-established (Posada, 2003), although most evolutionary studies still apply traditional substitution models that can be far from the "real" substitution process (Arenas, 2015b; Wilke, 2012). Next, ignoring recombination in recombinant pathogens can heavily bias phylogenetic and population parameter inferences (Anisimova et al., 2003; Arenas and Posada, 2010a; Posada and Crandall, 2002; Schierup and Hein, 2000). Detecting mutation and recombination is relevant for investigating pathogen evolution, but also their amount and type needs to be considered and accurately estimated to avoid biases in evolutionary analyses. Finally, the specific roles of mutation and recombination on certain phenomena, such as drug resistance or fitness recovery, should be carefully interpreted because of the compound effect of other population processes such as population size fluctuations and selection.

Acknowledgements

The authors want to thank Michel Tibayrenc, Editor-in-Chief of the journal *Infection, Genetics and Evolution*, for the invitation to contribute with this review article on mutation and recombination in pathogen evolution for a special issue of the journal.

Financial & competing interests disclosure

The authors have declared that no competing interests exist. MA was supported by the fellowship "Ramón y Cajal" RYC-2015-18241 from the Spanish Government. NMA was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), grant number 444071/2014-8. NC was supported by NORTE2020 (Portugal),

grant number NORTE–01–0145–FEDER–000029. IPATIMUP integrates the i3S Research Unit, which is partially supported by FCT. ECN was supported by “CONICYT-FONDECYT de iniciación en la investigación 11160905”. MP-L was partially supported by a DC D-CFAR pilot award, University Facilitating Fund Award from George Washington University, USA NIH CCNS214443F and NIH UL1TR000075 awards from the NIH National Center for Advancing Translational Sciences, and a 2015 HIV Phylodynamics Supplement award from the District of Columbia Center for AIDS Research, an NIH funded program (P30AI117970), which is supported by the following NIH Co-Funding and Participating Institutes and Centers: NIAID, NCI, NICHD, NHLBI, NIDA, NIMH, NIA, FIC, NIGMS, NIDDK, and OAR. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

References

Achtman, M., 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu. Rev. Microbiol. 62, 53–70.

Althaus, C.L., Bonhoeffer, S., 2005. Stochastic interplay between mutation and recombination during the acquisition of drug resistance mutations in human immunodeficiency virus type 1. J. Virol. 79, 13572–13578.

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, 1229–1236.

Araujo, N.M., 2015. Hepatitis B virus intergenotypic recombinants worldwide: an overview. Infect. Genet. Evol. 36, 500–510.

Araujo, N.M., Araujo, O.C., Silva, E.M., Villela-Nogueira, C.A., Nabuco, L.C., Parana, R., Bessone, F., Gomes, S.A., Trepo, C., Kay, A., 2013. Identification of novel recombinants of hepatitis B virus genotypes F and G in human immunodeficiency virus-positive patients from Argentina and Brazil. J. Gen. Virol. 94, 150–158.

Arbiza, L., Patricio, M., Dopazo, H., Posada, D., 2011. Genome-wide heterogeneity of nucleotide substitution model fit. Genome Biol. Evol. 3, 896–908.

Archer, J., Pinney, J.W., Fan, J., Simon-Loriere, E., Arts, E.J., Negroni, M., Robertson, D.L., 2008. Identifying the important HIV-1 recombination breakpoints. PLoS Comput. Biol. 4, e1000178.

Arenas, M., 2013. The importance and application of the ancestral recombination graph. Front. Genet. 4, 206.

Arenas, M., 2015a. Genetic consequences of antiviral therapy on HIV-1. Comput. Math. Methods Med. 2015, 9.

Arenas, M., 2015b. Trends in substitution models of molecular evolution. Front. Genet. 6, 319.

Arenas, M., Posada, D., 2010a. Coalescent simulation of intracodon recombination. Genetics 184, 429–437.

Arenas, M., Posada, D., 2010b. Computational design of centralized HIV-1 genes. Curr. HIV Res. 8, 613–621.

Arenas, M., Posada, D., 2010c. The effect of recombination on the reconstruction of ancestral sequences. Genetics 184, 1133–1139.

Arenas, M., Posada, D., 2014. The influence of recombination on the estimation of selection from coding sequence alignments. In: Fares, M.A. (Ed.), Natural Selection: Methods and Applications. CRC Press/Taylor & Francis, Boca Raton, pp. 112–125.

Arenas, M., Lopes, J.S., Beaumont, M.A., Posada, D., 2015. CodABC: a computational framework to coestimate recombination, substitution, and molecular adaptation rates by approximate bayesian computation. Mol. Biol. Evol. 32, 1109–1112.

Arenas, M., Lorenzo-Redondo, R., Lopez-Galindez, C., 2016. Influence of mutation and recombination on HIV-1 in vitro fitness recovery. Mol. Phylogenet. Evol. 94, 264–270.

Ash, C., 2014. The hidden history of the HIV pandemic. Science 346, 50.

Austermann-Busch, S., Becher, P., 2012. RNA structural elements determine frequency and sites of nonhomologous recombination in an animal plus-strand RNA virus. J. Virol. 86, 7393–7402.

Bachtrog, D., Charlesworth, B., 2002. Reduced adaptation of a non-recombining neo-Y chromosome. Nature 416, 323–326.

Barouch, D.H., Stephenson, K.E., Borducchi, E.N., Smith, K., Stanley, K., McNally, A.G., Liu, J., Abbink, P., Maxfield, L.F., Seaman, M.S., Dugast, A.S., Alter, G., Ferguson, M., Li, W., Earl, P.L., Moss, B., Giorgi, E.E., Szinger, J.J., Eller, L.A., Billings, E.A., Rao, M., Tovanabutra, S., Sanders-Buell, E., Weijtens, M., Pau, M.G., Schuitemaker, H., Robb, M.L., Kim, J.H., Korber, B.T., Michael, N.L., 2013. Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. Cell 155, 531–539.

Barouch, D.H., Alter, G., Broge, T., Linde, C., Ackerman, M.E., Brown, E.P., Borducchi, E.N., Smith, K.M., Nkolola, J.P., Liu, J., Shields, J., Parenteau, L., Whitney, J.B., Abbink, P., Ng'ang'a, D.M., Seaman, M.S., Lavine, C.L., Perry, J.R., Li, W., Colantonio, A.D., Lewis, M.G., Chen, B., Wenschuh, H., Reimer, U., Piatak, M., Lifson, J.D., Handley, S.A., Virgin, H.W., Koutsoukos, M., Lorin, C., Voss, G., Weijtens, M., Pau, M.G., Schuitemaker, H., 2015. Protective efficacy of adenovirus/protein vaccines against SIV challenges in rhesus monkeys. Science 349, 320–324.

Bartolo, I., Calado, R., Borrego, P., Leitner, T., Taveira, N., 2016. Rare HIV-1 subtype J genomes and a new H/U/CRF02_AG recombinant genome suggests an ancient origin of HIV-1 in Angola. AIDS Res. Hum. Retrovir. 32, 822–828.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162, 2025–2035.

Bedford, T., Cobey, S., Beerli, P., Pascual, M., 2010. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). PLoS Pathog. 6, e1000918.

Bedford, T., Riley, S., Barr, I.G., Broor, S., Chadha, M., Cox, N.J., Daniels, R.S., Gunasekaran, C.P., Hurt, A.C., Kelso, A., Klimov, A., Lewis, N.S., Li, X., McCauley, J.W., Odagiri, T., Potdar, V., Rambaut, A., Shu, Y., Skepner, E., Smith, D.J., Suchard, M.A., Tashiro, M., Wang, D., Xu, X., Lemey, P., Russell, C.A., 2015. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. Nature 523, 217–220.

Beerenwinkel, N., Gunthard, H.F., Roth, V., Metzner, K.J., 2012. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front. Microbiol. 3, 329.

Bernstein, H., Byerly, H.C., Hopf, F.A., Michod, R.E., 1985. Genetic damage, mutation, and the evolution of sex. Science 229, 1277–1281.

Betancourt, A.J., Welch, J.J., Charlesworth, B., 2009. Reduced effectiveness of selection caused by a lack of recombination. Curr. Biol. 19, 655–660.

Biek, R., Pybus, O.G., Lloyd-Smith, J.O., Didelot, X., 2015. Measurably evolving pathogens in the genomic era. Trends Ecol. Evol. 30, 306–313.

Billal, D.S., Feng, J., Leprohon, P., Legare, D., Ouellette, M., 2011. Whole genome analysis of linezolid resistance in Streptococcus pneumoniae reveals resistance and compensatory mutations. BMC Genomics 12, 512.

Bliven, K.A., Maurelli, A.T., 2016. Evolution of bacterial pathogens within the human host. Microbiol Spectr. 4.

Bollyky, P.L., Rambaut, A., Harvey, P.H., Holmes, E.C., 1996. Recombination between sequences of hepatitis B virus from different genotypes. J. Mol. Evol. 42, 97–102.

Boni, M.F., Zhou, Y., Taubenberger, J.K., Holmes, E.C., 2008. Homologous recombination is very rare or absent in human influenza A virus. J. Virol. 82, 4807–4811.

Boni, M.F., Smith, G.J., Holmes, E.C., Vijaykrishna, D., 2012. No evidence for intra-segment recombination of 2009 H1N1 influenza virus in swine. Gene 494, 242–245.

Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537.

Boussau, B., Gueguen, L., Gouy, M., 2009. A mixture model and a hidden markov model to simultaneously detect recombination breakpoints and reconstruct phylogenies. Evol. Bioinformatics Online 5, 67–79.

Bretscher, M.T., Althaus, C.L., Muller, V., Bonhoeffer, S., 2004. Recombination in HIV and the evolution of drug resistance: for better or for worse? BioEssays 26, 180–188.

Bridges, B.A., 2001. Hypermutation in bacteria and other cellular systems. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 356, 29–39.

Bruen, T.C., Philippe, H., Bryant, D., 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172, 2665–2681.

Budowle, B., Williamson, P.C., 2011. Microbial Forensics. Wiley Encyclopedia of Forensic Science. John Wiley & Sons, Ltd.

Bush, R.M., Bender, C.A., Subbarao, K., Cox, N.J., Fitch, W.M., 1999. Predicting the evolution of human influenza A. Science 286, 1921–1925.

Carobene, M.G., Rodriguez Rodrigues, C., De Candia, C.A., Turk, G., Salomon, H., 2009. In vitro dynamics of HIV-1 BF intersubtype recombinants genomic regions involved in the regulation of gene expression. Virol. J. 6, 107.

Carvajal-Rodriguez, A., Crandall, K.A., Posada, D., 2007. Recombination favors the evolution of drug resistance in HIV-1 during antiretroviral therapy. Infect. Genet. Evol. 7, 476–483.

Castelhano, N., Araujo, N.M., Arenas, M., 2017. Heterogeneous recombination among hepatitis B virus genotypes. Infect. Genet. Evol. 54, 486–490.

Castro-Nallar, E., Cortez-San Martin, M., Mascayano, C., Molina, C., Crandall, K.A., 2011. Molecular phylodynamics and protein modeling of infectious salmon anemia virus (ISAV). BMC Evol. Biol. 11, 349.

Castro-Nallar, E., Crandall, K., Pérez-Losada, M., 2012a. Genetic diversity and molecular epidemiology of HIV transmission. Futur. Virol. 7, 239–252.

Castro-Nallar, E., Pérez-Losada, M., Burton, G.F., Crandall, K.A., 2012b. The evolution of HIV: inferences using phylogenetics. Mol. Phylogenet. Evol. 62, 777–792.

Castro-Nallar, E., Hasan, N.A., Cebula, T.A., Colwell, R.R., Robison, R.A., Johnson, W.E., Crandall, K.A., 2015. Concordance and discordance of sequence survey methods for molecular epidemiology. PeerJ 3, e761.

Chan, A.H., Jenkins, P.A., Song, Y.S., 2012. Genome-wide fine-scale recombination rate variation in drosophila melanogaster. PLoS Genet. 8, e1003090.

Charlesworth, B., Betancourt, A.J., Kaiser, V.B., Gordo, I., 2009. Genetic recombination and molecular evolution. Cold Spring Harb. Symp. Quant. Biol. 74, 177–186.

Chopra, I., O'Neill, A.J., Miller, K., 2003. The role of mutators in the emergence of antibiotic-resistant bacteria. Drug Resist. Updat. 6, 137–145.

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.

Croll, D., Lendenmann, M.H., Stewart, E., McDonald, B.A., 2015. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics 201, 1213–1228.

Cui, C., Shi, J., Hui, L., Xi, H., Zhuoma, Quni, Tsedan, Hu, G., 2002. The dominant hepatitis B virus genotype identified in Tibet is a C/D hybrid. J. Gen. Virol. 83, 2773–2777.

Delatorre, E., Bello, G., 2016. Time-scale of minor HIV-1 complex circulating recombinant forms from Central and West Africa. BMC Evol. Biol. 16, 249.

Delwart, E.L., Pan, H., Sheppard, H.W., Wolpert, D., Neumann, A.U., Korber, B., Mullins, J.I., 1997. Slower evolution of human immunodeficiency virus type 1 quasispecies during progression to AIDS. J. Virol. 71, 7498–7508.

Devold, M., Karlsen, M., Nylund, A., 2006. Sequence analysis of the fusion protein gene from infectious salmon anemia virus isolates: evidence of recombination and reassortment. J. Gen. Virol. 87, 2031–2040.
Diaz, R.S., Sabino, E.C., Mayer, A., Mosley, J.W., Busch, M.P., 1995. Dual human immunodeficiency virus type 1 infection and recombination in a dually exposed transfusion recipient. The Transfusion Safety Study Group. J. Virol. 69, 3273–3281.
Didelot, X., Maiden, M.C., 2010. Impact of recombination on bacterial evolution. Trends Microbiol. 18, 315–322.
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, 150–162.
Ding, N.Z., Xu, D.S., Sun, Y.Y., He, H.B., He, C.Q., 2017. A permanent host shift of rabies virus from *Chiroptera* to *Carnivora* associated with recombination. Sci Rep 7, 289.
Doria-Rose, N.A., Learn, G.H., Rodrigo, A.G., Nickle, D.C., Li, F., Mahalanabis, M., Hensel, M.T., McLaughlin, S., Edmonson, P.F., Montefiori, D., Barnett, S.W., Haigwood, N.L., Mullins, J.I., 2005. Human immunodeficiency virus type 1 subtype B ancestral envelope protein is functional and elicits neutralizing antibodies in rabbits similar to those elicited by a circulating subtype B envelope. J. Virol. 79, 11214–11224.
Drummond, A.J., Pybus, O.G., Rambaut, A., Forsberg, R., Rodrigo, A.G., 2003. Measurably evolving populations. Trends Ecol. Evol. 18, 481–488.
Duchene, S., Di Giallonardo, F., Holmes, E.C., 2016. Substitution model adequacy and assessing the reliability of estimates of virus evolutionary rates and time scales. Mol. Biol. Evol. 33, 255–267.
Ebert, D., 1998. Experimental evolution of parasites. Science 282, 1432–1435.
Edwards, C.T., Holmes, E.C., Pybus, O.G., Wilson, D.J., Viscidi, R.P., Abrams, E.J., Phillips, R.E., Drummond, A.J., 2006. Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection. Genetics 174, 1441–1453.
Ellenberger, D.L., Li, B., Lupo, L.D., Owen, S.M., Nkengasong, J., Kadio-Morokro, M.S., Smith, J., Robinson, H., Ackers, M., Greenberg, A., Folks, T., Butera, S., 2002. Generation of a consensus sequence from prevalent and incident HIV-1 infections in West Africa to guide AIDS vaccine development. Virology 302, 155–163.
Faria, N.R., Rambaut, A., Suchard, M.A., Baele, G., Bedford, T., Ward, M.J., Tatem, A.J., Sousa, J.D., Arinaminpathy, N., Pepin, 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.
Faria, N.R., Quick, J., Claro, I.M., Theze, J., de Jesus, J.G., Giovanetti, M., Kraemer, M.U.G., Hill, S.C., Black, A., da Costa, A.C., Franco, L.C., Silva, S.P., Wu, C.H., Raghwani, J., Cauchemez, S., du Plessis, L., Verotti, M.P., de Oliveira, W.K., Carmo, E.H., Coelho, G.E., Santelli, A., Vinhal, L.C., Henriques, C.M., Simpson, J.T., Loose, M., Andersen, K.G., Grubaugh, N.D., Somasekar, S., Chiu, C.Y., Munoz-Medina, J.E., Gonzalez-Bonilla, C.R., Arias, C.F., Lewis-Ximenez, L.L., Baylis, S.A., Chieppe, A.O., Aguiar, S.F., Fernandes, C.A., Lemos, P.S., Nascimento, B.L.S., Monteiro, H.A.O., Siqueira, I.C., de Queiroz, M.G., de Souza, T.R., Bezerra, J.F., Lemos, M.R., Pereira, G.F., Loudal, D., Moura, L.C., Dhalia, R., Franca, R.F., Magalhaes, T., Marques Jr., E.T., Jaenisch, T., Wallau, G.L., de Lima, M.C., Nascimento, V., de Cerqueira, E.M., de Lima, M.M., Mascarenhas, D.L., Neto, J.P.M., Levin, A.S., Tozetto-Mendoza, T.R., Fonseca, S.N., Mendes-Correa, M.C., Milagres, F.P., Segurado, A., Holmes, E.C., Rambaut, A., Bedford, T., Nunes, M.R.T., Sabino, E.C., Alcantara, L.C.J., Loman, N.J., Pybus, O.G., 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. Nature 546, 406–410.
Feil, E.J., Enright, M.C., 2004. Analyses of clonality and the evolution of bacterial pathogens. Curr. Opin. Microbiol. 7, 308–313.
Feil, E.J., Spratt, B.G., 2001. Recombination and the population structures of bacterial pathogens. Annu. Rev. Microbiol. 55, 561–590.
Feil, E.J., Maiden, M.C., Achtman, M., Spratt, B.G., 1999. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. Mol. Biol. Evol. 16, 1496–1502.
Feil, E.J., Enright, M.C., Spratt, B.G., 2000. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. Res. Microbiol. 151, 465–469.
Felsenstein, J., 1974. The evolutionary advantage of recombination. Genetics 78, 737–756.
Flannery, B., Chung, J.R., Thaker, S.N., Monto, A.S., Martin, E.T., Belongia, E.A., McLean, H.Q., Gaglani, M., Murthy, K., Zimmermann, R.K., Nowalk, M.P., Jackson, M.L., Jackson, L.A., Foust, A., Sessions, W., Berman, L., Spencer, S., Fry, A.M., 2017. Interim estimates of 2016–17 seasonal influenza vaccine effectiveness — United States, February 2017. In: Morbidity and Mortality Weekly Report (MMWR). 66. pp. 167–171.
Foll, M., Poh, Y.P., Renzette, N., Ferrer-Admetlla, A., Bank, C., Shim, H., Malaspinas, A.S., Ewing, G., Liu, P., Wegmann, D., Caffrey, D.R., Zeldovich, K.B., Bolon, D.N., Wang, J.P., Kowalik, T.F., Schiffer, C.A., Finberg, R.W., Jensen, J.D., 2014. Influenza virus drug resistance: a time-sampled population genetics perspective. PLoS Genet. 10, e1004185.
Frahm, N., Nickle, D.C., Linde, C.H., Cohen, D.E., Zuniga, R., Lucchetti, A., Roach, T., Walker, B.D., Allen, T.M., Korber, B.T., Mullins, J.I., Brander, C., 2008. Increased detection of HIV-specific T cell responses by combination of central sequences with comparable immunogenicity. AIDS 22, 447–456.
Fraser, C., 2005. HIV recombination: what is the impact on antiretroviral therapy? J. R. Soc. Interface 2, 489–503.
Frost, S.D., Nijhuis, M., Schuurman, R., Boucher, C.A., Brown, A.J., 2000. Evolution of lamivudine resistance in human immunodeficiency virus type 1-infected individuals: the relative roles of drift and selection. J. Virol. 74, 6262–6268.
Frost, S.D., Dumaurier, M.J., Wain-Hobson, S., Brown, A.J., 2001. Genetic drift and within-host metapopulation dynamics of HIV-1 infection. Proc. Natl. Acad. Sci. U. S. A. 98, 6975–6980.
Galli, A., Bukh, J., 2014. Comparative analysis of the molecular mechanisms of recombination in hepatitis C virus. Trends Microbiol. 22, 354–364.
Galli, A., Kearney, M., Nikolaitchik, O.A., Yu, S., Chin, M.P., Maldarelli, F., Coffin, J.M., Pathak, V.K., Hu, W.S., 2010. Patterns of Human Immunodeficiency Virus type 1 recombination ex vivo provide evidence for coadaptation of distant sites, resulting in purifying selection for intersubtype recombinants during replication. J. Virol. 84, 7651–7661.
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.
Gao, F., Bhattacharya, T., Gaschen, B., Taylor, J., Moore, J.P., Novitsky, V., Yusim, K., Lang, D., Foley, B., Beddows, S., Alam, M., Haynes, B., Hahn, B.H., Korber, B., 2003. Consensus and ancestral state HIV vaccines. Science 299, 1515–1518.
Gao, F., Ming, C., Hu, W., Li, H., 2016. New software for the fast estimation of population recombination rates (FastEPRR) in the genomic era. G3 (Bethesda) 6, 1563–1571.
Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N., Hughes, K.A., 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. Nature 525, 372–375.
Gillings, M.R., Gaze, W.H., Pruden, A., Smalla, K., Tiedje, J.M., Zhu, Y.G., 2015. Using the class 1 integron-integrase gene as a proxy for anthropogenic pollution. ISME J. 9, 1269–1279.
Godoy, M.G., Suarez, R., Lazo, E.S., Llegues, K.O., Kibenge, M.J., Wang, Y., Kibenge, F.S., 2014. Genetic analysis and comparative virulence of infectious salmon anemia virus (ISAV) types HPR7a and HPR7b from recent field outbreaks in Chile. Virol. J. 11, 204.
Gog, J.R., Pellis, L., Wood, J.L., McLean, A.R., Arinaminpathy, N., Lloyd-Smith, J.O., 2015. Seven challenges in modeling pathogen dynamics within-host and across scales. Epidemics 10, 45–48.
Golovkina, T.V., Jaffe, A.B., Ross, S.R., 1994. Coexpression of exogenous and endogenous mouse mammary tumor virus RNA in vivo results in viral recombination and broadens the virus host range. J. Virol. 68, 5019–5026.
Gonzales, M.J., Dugan, J.M., Shafer, R.W., 2002. Synonymous-non-synonymous mutation rates between sequences containing ambiguous nucleotides (Syn-SCAN). Bioinformatics 18, 886–887.
Griffiths, R.C., Marjoram, P., 1997. An ancestral recombination graph. In: Donelly, P., Tavaré, S. (Eds.), Progress in Population Genetics And Human Evolution. Springer-Verlag, Berlin, pp. 257–270.
Grubaugh, N.D., Ladner, J.T., Kraemer, M.U.G., Dudas, G., Tan, A.L., Gangavarapu, K., Wiley, M.R., White, S., Theze, J., Magnani, D.M., Prieto, K., Reyes, D., Bingham, A.M., Paul, L.M., Robles-Sikisaka, R., Oliveira, G., Pronty, D., Barcellona, C.M., Metsky, H.C., Baniecki, M.L., Barnes, K.G., Chak, B., Freije, C.A., Gladden-Young, A., Gnirke, A., Luo, C., MacInnis, B., Matranga, C.B., Park, D.J., Qu, J., Schaffner, S.F., Tomkins-Tinch, C., West, K.L., Winnicki, S.M., Wohl, S., Yozwiak, N.L., Quick, J., Fauver, J.R., Khan, K., Brent, S.E., Reiner Jr., R.C., Lichtenberger, P.N., Ricciardi, M.J., Bailey, V.K., Watkins, D.I., Cone, M.R., Kopp, E.W.t., Hogan, K.N., Cannons, A.C., Jean, R., Monaghan, A.J., Garry, R.F., Loman, N.J., Faria, N.R., Porcelli, M.C., Vasquez, C., Nagle, E.R., Cummings, D.A.T., Stanek, D., Rambaut, A., Sanchez-Lockhart, M., Sabeti, P.C., Gillis, L.D., Michael, S.F., Bedford, T., Pybus, O.G., Isern, S., Palacios, G., Andersen, K.G., 2017. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature 546, 401–405.
Hadany, L., Beker, T., 2003. On the evolutionary advantage of fitness-associated recombination. Genetics 165, 2167–2179.
Hanage, W.P., Fraser, C., Spratt, B.G., 2006. The impact of homologous recombination on the generation of diversity in bacteria. J. Theor. Biol. 239, 210–219.
Hasan, N.A., Choi, S.Y., Eppinger, M., Clark, P.W., Chen, A., Alam, M., Haley, B.J., Taviani, E., Hine, E., Su, Q., Tallon, L.J., Prosper, J.B., Furth, K., Hoq, M.M., Li, H., Fraser-Liggett, C.M., Cravioto, A., Huq, A., Ravel, J., Cebula, T.A., Colwell, R.R., 2012. Genomic diversity of 2010 Haitian cholera outbreak strains. Proc. Natl. Acad. Sci. U. S. A. 109, E2010–2017.
Haubold, B., Pfaffelhuber, P., Lynch, M., 2010. mlRho - a program for estimating the population mutation and recombination rates from shotgun-sequenced diploid genomes. Mol. Ecol. 19 (Suppl. 1), 277–284.
Hino, O., Tabata, S., Hotta, Y., 1991. Evidence for increased in vitro recombination with insertion of human hepatitis B virus DNA. Proc. Natl. Acad. Sci. U. S. A. 88, 9248–9252.
Holt, K.E., Baker, S., Weill, F.X., Holmes, E.C., Kitchen, A., Yu, J., Sangal, V., Brown, D.J., Coia, J.E., Kim, D.W., Choi, S.Y., Kim, S.H., da Silveira, W.D., Pickard, D.J., Farrar, J.J., Parkhill, J., Dougan, G., Thomson, N.R., 2012. *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat. Genet. 44, 1056–1059.
Hudson, R.R., 2001. Two-locus sampling distributions and their application. Genetics 159, 1805–1817.
Husmeier, D., Wright, F., 2001. Detection of recombination in DNA alignments with hidden Markov models. J. Comput. Biol. 8, 401–427.
Iglesias-Sanchez, M.J., Lopez-Galindez, C., 2002. Analysis, quantification, and evolutionary consequences of HIV-1 in vitro recombination. Virology 304, 392–402.
Jee, J., Rasouly, A., Shamovsky, I., Akivis, Y., Steinman, S.R., Mishra, B., Nudler, E., 2016. Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. Nature 534, 693–696.
John, M., Moore, C.B., James, I.R., Mallal, S.A., 2005. Interactive selective pressures of HLA-restricted immune responses and antiretroviral drugs on HIV-1. Antivir. Ther. 10, 551–555.
Johnson, P.L., Slatkin, M., 2009. Inference of microbial recombination rates from metagenomic data. PLoS Genet. 5, e1000674.
Jolley, K.A., Feil, E.J., Chan, M.S., Maiden, M.C., 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17, 1230–1231.
Kibenge, F.S., Godoy, M.G., Wang, Y., Kibenge, M.J., Gherardelli, V., Mansilla, S., Lisperger, A., Jarpa, M., Larroquete, G., Avendano, F., Lara, M., Gallardo, A., 2009.

Infectious salmon anaemia virus (ISAV) isolated from the ISA disease outbreaks in Chile diverged from ISAV isolates from Norway around 1996 and was disseminated around 2005, based on surface glycoprotein gene sequences. Virol. J. 6, 88.
Kim, Y., Orr, H.A., 2005. Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. Genetics 171, 1377–1386.
Korber, B., 2000. HIV signature and sequence variation analysis. In: Rodrigo, A.G., Learn, G.H. (Eds.), Computational Analysis of HIV Molecular Sequences. Kluwer Academic Publishers, Dordrecht, Netherlands, pp. 55–72.
Kosakovsky Pond, S.L., Frost, S.D., 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533.
Kosakovsky Pond, S.L., Frost, S.D., Muse, S.V., 2005. HYPHY: hypothesis testing using phylogenies. Bioinformatics 21, 676–679.
Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D., 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics 22, 3096–3098.
Kothe, D.L., Li, Y., Decker, J.M., Bibollet-Ruche, F., Zammit, K.P., Salazar, M.G., Chen, Y., Weng, Z., Weaver, E.A., Gao, F., Haynes, B.F., Shaw, G.M., Korber, B.T., Hahn, B.H., 2006. Ancestral and consensus envelope immunogens for HIV-1 subtype C. Virology 352, 438–449.
Kuhner, M.K., 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22, 768–770.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Lai, M.M., 1992. RNA recombination in animal and plant viruses. Microbiol. Rev. 56, 61–79.
Lartillot, N., Philippe, H., 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. 21, 1095–1109.
Laureau, R., Loeillet, S., Salinas, F., Bergstrom, A., Legoix-Ne, P., Liti, G., Nicolas, A., 2016. Extensive recombination of a yeast diploid hybrid through meiotic reversion. PLoS Genet. 12, e1005781.
Le Gall-Recule, G., Lavazza, A., Marchandeau, S., Bertagnoli, S., Zwingelstein, F., Cavadini, P., Martinelli, N., Lombardi, G., Guerin, J.L., Lemaitre, E., Decors, A., Boucher, S., Le Normand, B., Capucci, L., 2013. Emergence of a new lagovirus related to rabbit haemorrhagic disease virus. Vet. Res. 44, 81.
Lemey, P., Rambaut, A., Pybus, O.G., 2006. HIV evolutionary dynamics within and among hosts. AIDS Rev. 8, 125–140.
Lemey, P., Kosakovsky Pond, S.L., Drummond, A.J., Pybus, O.G., Shapiro, B., Barroso, H., Taveira, N., Rambaut, A., 2007. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS Comput. Biol. 3, e29.
Li, N., Stephens, M., 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165, 2213–2233.
Lihana, R.W., Ssemwanga, D., Abimiku, A., Ndembi, N., 2012. Update on HIV-1 diversity in Africa: a decade in review. AIDS Rev. 14, 83–100.
Lloyd-Smith, J.O., Funk, S., McLean, A.R., Riley, S., Wood, J.L., 2015. Nine challenges in modelling the emergence of novel pathogens. Epidemics 10, 35–39.
Lopes, J.S., Arenas, M., Posada, D., Beaumont, M.A., 2014. Coestimation of recombination, substitution and molecular adaptation rates by approximate Bayesian computation. Heredity 112, 255–264.
Lorenzo-Redondo, R., Borderia, A.V., Lopez-Galindez, C., 2011. Dynamics of in vitro fitness recovery of HIV-1. J. Virol. 85, 1861–1870.
Lorenzo-Redondo, R., Delgado, S., Moran, F., Lopez-Galindez, C., 2014. Realistic three dimensional fitness landscapes generated by self organizing maps for the analysis of experimental HIV-1 evolution. PLoS One 9, e88579.
Luksza, M., Lassig, M., 2014. A predictive fitness model for influenza. Nature 507, 57–61.
Maisnier-Patin, S., Andersson, D.I., 2004. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res. Microbiol. 155, 360–369.
Mallo, D., Sánchez-Cobos, A., Arenas, M., 2016. Diverse considerations for Successful phylogenetic tree reconstruction: impacts from model misspecification, recombination, homoplasy, and pattern recognition. In: Elloumi, M., Iliopoulos, C., Wang, J., Zomaya, A. (Eds.), Pattern Recognition in Computational Molecular Biology. John Wiley & Sons, Inc, pp. 439–456.
Markham, R.B., Wang, W.C., Weisstein, A.E., Wang, Z., Munoz, A., Templeton, A., Margolick, J., Vlahov, D., Quinn, T., Farzadegan, H., Yu, X.F., 1998. Patterns of HIV-1 evolution in individuals with differing rates of CD4 T cell decline. Proc. Natl. Acad. Sci. U. S. A. 95, 12568–12573.
Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.
Martin, D.P., Biagini, P., Lefeuvre, P., Golden, M., Roumagnac, P., Varsani, A., 2011. Recombination in eukaryotic single stranded DNA viruses. Viruses 3, 1699–1738.
Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, vev003.
Marttinen, P., Hanage, W.P., Croucher, N.J., Connor, T.R., Harris, S.R., Bentley, S.D., Corander, J., 2012. Detection of recombination events in bacterial genomes from large population samples. Nucleic Acids Res. 40, e6.
Maydt, J., Lengauer, T., 2006. Recco: recombination analysis using cost optimization. Bioinformatics 22, 1064–1071.
McBurney, S.P., Ross, T.M., 2007. Developing broadly reactive HIV-1/AIDS vaccines: a review of polyvalent and centralized HIV-1 vaccines. Curr. Pharm. Des. 13, 1957–1964.
McDonald, S.M., Nelson, M.I., Turner, P.E., Patton, J.T., 2016. Reassortment in segmented RNA viruses: mechanisms and outcomes. Nat. Rev. Microbiol. 14, 448–460.
McNearney, T., Hornickova, Z., Markham, R., Birdwell, A., Arens, M., Saah, A., Ratner, L., 1992. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc. Natl. Acad. Sci. U. S. A. 89, 10247–10251.
McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., Donnelly, P., 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304, 581–584.
Melnyk, A.H., Wong, A., Kassen, R., 2015. The fitness costs of antibiotic resistance mutations. Evol. Appl. 8, 273–283.
Menendez-Arias, L., Martinez, M.A., Quinones-Mateu, M.E., Martinez-Picado, J., 2003. Fitness variations and their impact on the evolution of antiretroviral drug resistance. Curr. Drug Targets Infect. Disord. 3, 355–371.
Metcalf, C.J., Birger, R.B., Funk, S., Kouyos, R.D., Lloyd-Smith, J.O., Jansen, V.A., 2015. Five challenges in evolution and infectious diseases. Epidemics 10, 40–44.
Minin, V.N., Dorman, K.S., Fang, F., Suchard, M.A., 2005. Dual multiple change-point model leads to more accurate recombination detection. Bioinformatics 21, 3034–3042.
Minin, V.N., Dorman, K.S., Fang, F., Suchard, M.A., 2007. Phylogenetic mapping of recombination hotspots in human immunodeficiency virus via spatially smoothed change-point processes. Genetics 175, 1773–1785.
Monjane, A.L., van der Walt, E., Varsani, A., Rybicki, E.P., Martin, D.P., 2011. Recombination hotspots and host susceptibility modulate the adaptive value of recombination during maize streak virus evolution. BMC Evol. Biol. 11, 350.
Moradigaravand, D., Engelstadter, J., 2012. The effect of bacterial recombination on adaptation on fitness landscapes with limited peak accessibility. PLoS Comput. Biol. 8, e1002735.
Moradigaravand, D., Kouyos, R., Hinkley, T., Haddad, M., Petropoulos, C.J., Engelstadter, J., Bonhoeffer, S., 2014. Recombination accelerates adaptation on a large-scale empirical fitness landscape in HIV-1. PLoS Genet. 10, e1004439.
Morozov, V., Pisareva, M., Groudinin, M., 2000. Homologous recombination between different genotypes of hepatitis B virus. Gene 260, 55–65.
Nei, M., Gojobori, T., 1986. Simple method for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.
Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.
Nijhuis, M., Boucher, C.A., Schipper, P., Leitner, T., Schuurman, R., Albert, J., 1998. Stochastic processes strongly influence HIV-1 evolution during suboptimal protease-inhibitor therapy. Proc. Natl. Acad. Sci. U. S. A. 95, 14441–14446.
Njamkepo, E., Fawal, N., Tran-Dien, A., Hawkey, J., Strockbine, N., Jenkins, C., Talukder, K.A., Bercion, R., Kuleshov, K., Kolinska, R., Russell, J.E., Kaftyreva, L., Accou-Demartin, M., Karas, A., Vandenberg, O., Mather, A.E., Mason, C.J., Page, A.J., Ramamurthy, T., Bizet, C., Gamian, A., Carle, I., Sow, A.G., Bouchier, C., Wester, A.L., Lejay-Collin, M., Fonkoua, M.C., Hello, S.L., Blaser, M.J., Jernberg, C., Ruckly, C., Merens, A., Page, A.L., Aslett, M., Roggentin, P., Fruth, A., Denamur, E., Venkatesan, M., Bercovier, H., Bodhidatta, L., Chiou, C.S., Clermont, D., Colonna, B., Egorova, S., Pazhani, G.P., Ezernitchi, A.V., Guigon, G., Harris, S.R., Izumiya, H., Korzeniowska-Kowal, A., Lutynska, A., Gouali, M., Grimont, F., Langendorf, C., Marejkova, M., Peterson, L.A., Perez-Perez, G., Ngandjio, A., Podkolzin, A., Souche, E., Makarova, M., Shipulin, G.A., Ye, C., Zemlickova, H., Herpay, M., Grimont, P.A., Parkhill, J., Sansonetti, P., Holt, K.E., Brisse, S., Thomson, N.R., Weill, F.X., 2016. Global phylogeography and evolutionary history of *Shigella dysenteriae* type 1. Nat. Microbiol. 1, 16027.
Novitsky, V., Smith, U.R., Gilbert, P., McLane, M.F., Chigwedere, P., Williamson, C., Ndung'u, T., Klein, I., Chang, S.Y., Peter, T., Thior, I., Foley, B.T., Gaolekwe, S., Rybak, N., Gaseitsiwe, S., Vannberg, F., Marlink, R., Lee, T.H., Essex, M., 2002. Human immunodeficiency virus type 1 subtype C molecular phylogeny: consensus sequence for an AIDS vaccine design? J. Virol. 76, 5435–5451.
Ojosnegros, S., Garcia-Arriaza, J., Escarmis, C., Manrubia, S.C., Perales, C., Arias, A., Mateu, M.G., Domingo, E., 2011. Viral genome segmentation can result from a trade-off between genetic content and particle stability. PLoS Genet. 7, e1001344.
Osiowy, C., Giles, E., Tanaka, Y., Mizokami, M., Minuk, G.Y., 2006. Molecular evolution of hepatitis B virus over 25 years. J. Virol. 80, 10307–10314.
Pan, C., Kim, J., Chen, L., Wang, Q., Lee, C., 2007. The HIV positive selection mutation database. Nucleic Acids Res. 35, D371–375.
Paraskevis, D., Angelis, K., Magiorkinis, G., Kostaki, E., Ho, S.Y., Hatzakis, A., 2015. Dating the origin of hepatitis B virus reveals higher substitution rate and adaptation on the branch leading to F/H genotypes. Mol. Phylogenet. Evol. 93, 44–54.
Pérez-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.
Pérez-Losada, M., Crandall, K.A., Zenilman, J., Viscidi, R.P., 2007a. Temporal trends in gonococcal population genetics in a high prevalence urban community. Infect. Genet. Evol. 7, 271–278.
Pérez-Losada, M., Porter, M.L., Tazi, L., Crandall, K.A., 2007b. New methods for inferring population dynamics from microbial sequences. Infect. Genet. Evol. 7, 24–43.
Pérez-Losada, M., Posada, D., Arenas, M., Jobes, D.V., Sinangil, F., Berman, P.W., Crandall, K.A., 2009. Ethnic differences in the adaptation rate of HIV gp120 from a vaccine trial. Retrovirology 6, 67.
Pérez-Losada, M., Jobes, D.V., Sinangil, F., Crandall, K.A., Arenas, M., Posada, D., Berman, P.W., 2011. Phylodynamics of HIV-1 from a phase III AIDS vaccine trial in Bangkok, Thailand. PLoS One 6, e16902.
Pérez-Losada, M., Cabezas, P., Castro-Nallar, E., Crandall, K.A., 2013. Pathogen typing in the genomics era: MLST and the future of molecular epidemiology. Infect. Genet. Evol. 16, 38–53.
Pérez-Losada, M., Arenas, M., Galan, J.C., Palero, F., Gonzalez-Candelas, F., 2015. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. Infect. Genet. Evol. 30C, 296–307.
Pérez-Losada, M., Arenas, M., Castro-Nallar, E., 2017. Multilocus sequence typing of pathogens: methods, analyses, and applications. In: Tibayrenc, M. (Ed.), Genetics and Evolution of Infectious Diseases, Second Edition. Elsevier, London, pp. 383–404.
Pineau, P., Marchio, A., Mattei, M.G., Kim, W.H., Youn, J.K., Tiollais, P., Dejean, A., 1998.

Extensive analysis of duplicated-inverted hepatitis B virus integrations in human hepatocellular carcinoma. J. Gen. Virol. 79 (Pt 3), 591–600.
Poon, A.F., Kosakovsky Pond, S.L., Bennett, P., Richman, D.D., Leigh Brown, A.J., Frost, S.D., 2007a. Adaptation to human populations is revealed by within-host polymorphisms in HIV-1 and hepatitis C virus. PLoS Pathog. 3, e45.
Poon, A.F., Kosakovsky Pond, S.L., Richman, D.D., Frost, S.D., 2007b. Mapping protease inhibitor resistance to human immunodeficiency virus type 1 sequence polymorphisms within patients. J. Virol. 81, 13598–13607.
Posada, D., 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. Mol. Biol. Evol. 19, 708–717.
Posada, D., 2003. Selecting models of evolution. In: Vandemme, A., Salemi, M. (Eds.), The Phylogenetic Handbook. Cambridge University Press, Cambridge, UK, pp. 256–282.
Posada, D., Crandall, K.A., 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50, 580–601.
Posada, D., Crandall, K.A., 2002. The effect of recombination on the accuracy of phylogeny estimation. J. Mol. Evol. 54, 396–402.
Prosperi, M.C., Prosperi, L., Bruselles, A., Abbate, I., Rozera, G., Vincenti, D., Solmone, M.C., Capobianchi, M.R., Ulivi, G., 2011. Combinatorial analysis and algorithms for quasispecies reconstruction using next-generation sequencing. BMC Bioinf. 12, 5.
Rambaut, A., Pybus, O.G., Nelson, M.I., Viboud, C., Taubenberger, J.K., Holmes, E.C., 2008. The genomic and epidemiological dynamics of human influenza A virus. Nature 453, 615–619.
Robertson, D.L., Sharp, P.M., McCutchan, F.E., Hahn, B.H., 1995. Recombination in HIV-1. Nature 374, 124–126.
Rohde, H., Qin, J., Cui, Y., Li, D., Loman, N.J., Hentschke, M., Chen, W., Pu, F., Peng, Y., Li, J., Xi, F., Li, S., Li, Y., Zhang, Z., Yang, X., Zhao, M., Wang, P., Guan, Y., Cen, Z., Zhao, X., Christner, M., Kobbe, R., Loos, S., Oh, J., Yang, L., Danchin, A., Gao, G.F., Song, Y., Li, Y., Yang, H., Wang, J., Xu, J., Pallen, M.J., Wang, J., Aepfelbacher, M., Yang, R., Consortium, E.c.O.H.G.A.C.-S, 2011. Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. N. Engl. J. Med. 365, 718–724.
Rolland, M., Jensen, M.A., Nickle, D.C., Yan, J., Learn, G.H., Heath, L., Weiner, D., Mullins, J.I., 2007. Reconstruction and function of ancestral center-of-tree human immunodeficiency virus type 1 proteins. J. Virol. 81, 8507–8514.
Ross, H.A., Nickle, D.C., Liu, Y., Heath, L., Jensen, M.A., Rodrigo, A.G., Mullins, J.I., 2006. Sources of variation in ancestral sequence reconstruction for HIV-1 envelope genes. Evol. Bioinformatics Online 2, 18–41.
Salemi, M., 2013. The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective. Infect. Dis. Rep. 5, e3.
Sanjuan, R., Nebot, M.R., Chirico, N., Mansky, L.M., Belshaw, R., 2010. Viral mutation rates. J. Virol. 84, 9733–9748.
Sauna, Z.E., Kimchi-Sarfaty, C., 2011. Understanding the contribution of synonymous mutations to human disease. Nat. Rev. Genet. 12, 683–691.
Scheel, T.K., Galli, A., Li, Y.P., Mikkelsen, L.S., Gottwein, J.M., Bukh, J., 2013. Productive homologous and non-homologous recombination of hepatitis C virus in cell culture. PLoS Pathog. 9, e1003228.
Schierup, M.H., Hein, J., 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics 156, 879–891.
Schultz, A.K., Zhang, M., Bulla, I., Leitner, T., Korber, B., Morgenstern, B., Stanke, M., 2009. jpHMM: improving the reliability of recombination prediction in HIV-1. Nucleic Acids Res. 37, W647–651.
Shi, B., Kitchen, C., Weiser, B., Mayers, D., Foley, B., Kemal, K., Anastos, K., Suchard, M., Parker, M., Brunner, C., Burger, H., 2010. Evolution and recombination of genes encoding HIV-1 drug resistance and tropism during antiretroviral therapy. Virology 404, 5–20.
Shriner, D., Rodrigo, A.G., Nickle, D.C., Mullins, J.I., 2004a. Pervasive genomic recombination of HIV-1 in vivo. Genetics 167, 1573–1583.
Shriner, D., Shankarappa, R., Jensen, M.A., Nickle, D.C., Mittler, J.E., Margolick, J.B., Mullins, J.I., 2004b. Influence of random genetic drift on human immunodeficiency virus type 1 env evolution during chronic infection. Genetics 166, 1155–1164.
Simmonds, P., Midgley, S., 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J. Virol. 79, 15467–15476.
Simon-Loriere, E., Holmes, E.C., 2011. Why do RNA viruses recombine? Nat. Rev. Microbiol. 9, 617–626.
Smith, J.M., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.
Spratt, B.G., Hanage, W.P., Feil, E.J., 2001. The relative contributions of recombination and point mutation to the diversification of bacterial clones. Curr. Opin. Microbiol. 4, 602–606.
Su, H., Liu, Y., Xu, Z., Cheng, S., Ye, H., Xu, Q., Liu, Q., Tan, S., Xu, D., Liu, Y., 2014. A novel complex A/C/G intergenotypic recombinant of hepatitis B virus isolated in southern China. PLoS One 9, e84005.
Sugauchi, F., Orito, E., Ichida, T., Kato, H., Sakugawa, H., Kakumu, S., Ishida, T., Chutaputti, A., Lai, C.L., Ueda, R., Miyakawa, Y., Mizokami, M., 2002. Hepatitis B virus of genotype B with or without recombination with genotype C over the precore region plus the core gene. J. Virol. 76, 5985–5992.
Telwatte, S., Hearps, A.C., Johnson, A., Latham, C.F., Moore, K., Agius, P., Tachedjian, M., Sonza, S., Sluis-Cremer, N., Harrigan, P.R., Tachedjian, G., 2015. Silent mutations at codons 65 and 66 in reverse transcriptase alleviate indel formation and restore fitness in subtype B HIV-1 containing D67N and K70R drug resistance mutations. Nucleic Acids Res. 43, 3256–3271.
Temin, H.M., 1991. Sex and recombination in retroviruses. Trends Genet. 7, 71–74.
Tong, S., Revill, P., 2016. Overview of hepatitis B viral replication and genetic variability. J. Hepatol. 64, S4–16.
Tumas, K.M., Poszgay, J.M., Avidan, N., Ksiazek, S.J., Overmoyer, B., Blank, K.J., Prystowsky, M.B., 1993. Loss of antigenic epitopes as the result of env gene recombination in retrovirus-induced leukemia in immunocompetent mice. Virology 192, 587–595.
Unemo, M., Golparian, D., Nicholas, R., Ohnishi, M., Gallay, A., Sednaoui, P., 2012. High-level cefixime- and ceftriaxone-resistant *Neisseria gonorrhoeae* in France: novel penA mosaic allele in a successful international clone causes treatment failure. Antimicrob. Agents Chemother. 56, 1273–1280.
Vidal, N., Mulanga-Kabeya, C., Nzilambi, N., Delaporte, E., Peeters, M., 2000. Identification of a complex env subtype E HIV type 1 virus from the democratic republic of congo, recombinant with A, G, H, J, K, and unknown subtypes. AIDS Res. Hum. Retrovir. 16, 2059–2064.
Vijaykrishna, D., Mukerji, R., Smith, G.J., 2015. RNA virus reassortment: an evolutionary mechanism for host jumps and immune evasion. PLoS Pathog. 11, e1004902.
Vinue, L., Jove, T., Torres, C., Ploy, M.C., 2011. Diversity of class 1 integron gene cassette Pc promoter variants in clinical Escherichia coli strains and description of a new P2 promoter variant. Int. J. Antimicrob. Agents 38, 526–529.
Vos, M., Didelot, X., 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J. 3, 199–208.
Wang, D., Zhang, Y., Zhang, Z., Zhu, J., Yu, J., 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics 8, 77–80.
Wen, F., Bedford, T., Cobey, S., 2016. Explaining the geographical origins of seasonal influenza A (H3N2). Proc. Biol. Sci. 283.
Wertheim, J.O., Leigh Brown, A.J., Hepler, N.L., Mehta, S.R., Richman, D.D., Smith, D.M., Kosakovsky Pond, S.L., 2014. The global transmission network of HIV-1. J Infect Dis 209, 304–313.
WHO, 2017. Hepatitis B. 204 World Health Organization Fact Sheet.
Wilke, C.O., 2012. Bringing molecules back into molecular evolution. PLoS Comput. Biol. 8, e1002572.
Williamson, S., Perry, S.M., Bustamante, C.D., Orive, M.E., Stearns, M.N., Kelly, J.K., 2005. A statistical characterization of consistent patterns of human immunodeficiency virus evolution within infected patients. Mol. Biol. Evol. 22, 456–468.
Wilson, D.J., McVean, G., 2006. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics 172, 1411–1425.
Wilson, D.J., Gabriel, E., Leatherbarrow, A.J., Cheesbrough, J., Gee, S., Bolton, E., Fox, A., Hart, C.A., Diggle, P.J., Fearnhead, P., 2009. Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. Mol. Biol. Evol. 26, 385–397.
Wiuf, C., Christensen, T., Hein, J., 2001. A simulation study of the reliability of recombination detection methods. Mol. Biol. Evol. 18, 1929–1939.
Woodford, N., Ellington, M.J., 2007. The emergence of antibiotic resistance by mutation. Clin. Microbiol. Infect. 13, 5–18.
Worobey, M., 2001. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. Mol. Biol. Evol. 18, 1425–1434.
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.
Wright, M.S., Baker-Austin, C., Lindell, A.H., Stepanauskas, R., Stokes, H.W., McArthur, J.V., 2008. Influence of industrial contamination on mobile genetic elements: class 1 integron abundance and gene cassette structure in aquatic bacterial communities. ISME J. 2, 417–428.
Wu, T.D., Schiffer, C.A., Gonzales, M.J., Taylor, J., Kantor, R., Chou, S., Israelski, D., Zolopa, A.R., Fessel, W.J., Shafer, R.W., 2003. Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. J. Virol. 77, 4836–4847.
Xu, B., Yang, Z., 2013. PAMLX: a graphical user interface for PAML. Mol. Biol. Evol. 30, 2723–2724.
Yahara, K., Didelot, X., Jolley, K.A., Kobayashi, I., Maiden, M.C., Sheppard, S.K., Falush, D., 2016. The landscape of realized homologous recombination in pathogenic bacteria. Mol. Biol. Evol. 33, 456–471.
Yamaguchi, Y., Gojobori, T., 1997. Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts. Proc. Natl. Acad. Sci. U. S. A. 94, 1264–1269.
Yan, S., Liu, H., Mohr, T.J., Jenrette, J., Chiodini, R., Zaccardelli, M., Setubal, J.C., Vinatzer, B.A., 2008. Role of recombination in the evolution of the model plant pathogen *Pseudomonas syringae* pv. tomato DC3000, a very atypical tomato strain. Appl. Environ. Microbiol. 74, 3171–3181.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yuste, E., Sanchez-Palomino, S., Casado, C., Domingo, E., Lopez-Galindez, C., 1999. Drastic fitness loss in human immunodeficiency virus type 1 upon serial bottleneck events. J. Virol. 73, 2745–2751.
Zanini, F., Brodin, J., Albert, J., Neher, R.A., 2016. Error rates, PCR recombination, and sampling depth in HIV-1 whole genome deep sequencing. Virus Res. 239, 106–114.
Zhang, Y., 2013. A dynamic Bayesian Markov model for phasing and characterizing haplotypes in next-generation sequencing. Bioinformatics 29, 878–885.
Zhou, Y., Holmes, E.C., 2007. Bayesian estimates of the evolutionary rate and age of hepatitis B virus. J. Mol. Evol. 65, 197–205.