Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

Complete genome sequence of the marine *Rhodococcus* sp. H-CA8f isolated from Comau fjord in Northern Patagonia, Chile

Undabarrena A.[a], Salvà-Serra F.[b,c,d], Jaén-Luchoro D.[b,c,d], Castro-Nallar E.[e], Mendez K.N.[e], Valencia R.[a], Ugalde J.A.[e,f], Moore E.R.B.[b,c,g], Seeger M.[a], Cámara B.[a,*]

[a] *Laboratorio de Microbiología Molecular y Biotecnología Ambiental, Departamento de Química & Centro de Biotecnología Daniel Alkalay Lowitt, Universidad Técnica Federico Santa María, Valparaíso, Chile*
[b] *Department of Infectious Diseases, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden*
[c] *Centre for Antibiotic Resistance Research (CARe), University of Gothenburg, Gothenburg, Sweden*
[d] *Microbiology, Department of Biology, University of the Balearic Islands, Palma de Mallorca, Spain*
[e] *Center for Bioinformatics and Integrative Biology, Facultad de Ciencias Biológicas, Universidad Andrés Bello, Santiago, Chile*
[f] *uBiome, Inc., San Francisco, CA, United States*
[g] *Culture Collection University of Gothenburg (CCUG), Gothenburg, Sweden*

ARTICLE INFO

Keywords:
Rhodococcus
Marine actinobacteria
Comau fjord
Novel mega-plasmid
PacBio sequencing

ABSTRACT

Rhodococcus sp. H-CA8f was isolated from marine sediments obtained from the Comau fjord, located in Northern Chilean Patagonia. Whole-genome sequencing was achieved using PacBio RS II platform, comprising one closed, complete chromosome of 6,19 Mbp with a 62.45% G + C content. The chromosome harbours several metabolic pathways providing a wide catabolic potential, where the upper biphenyl route is described. Also, *Rhodococcus* sp. H-CA8f bears one linear mega-plasmid of 301 Kbp and 62.34% of G + C content, where genomic analyses demonstrated that it is constituted mostly by putative ORFs with unknown functions, representing a novel genetic feature. These genetic characteristics provide relevant insights regarding Chilean marine actinobacterial strains.

1. Introduction

The *Rhodococcus* genus, belonging to the *Nocardiaceae* family of the *phylum Actinobacteria*, was first described in 1891 (Zopf, 1891). Members of this genus have been isolated from diverse niches, such as soils; fresh water; seawater; plants and animals. They are characterized as Gram- and catalase-positive, aerobic, non-motile and non-sporulating, high G + C mycolic-acid-containing rods (Barka et al., 2016). *Rhodococcus* are well-known for their broad catabolic diversity, as they possess the ability to degrade numerous recalcitrant and toxic pollutants (Alvarez et al., 1991; Wang et al., 2010) representing a genus of considerable biotechnological interest (van der Geize and Dijkhuizen, 2004; Larkin et al., 2005, 2006). Whole-genome sequencing has opened new opportunities for elucidating pathways that provide understanding of the metabolic potential of *Rhodococcus*. For instance, the genome of *Rhodococcus jostii* RHA1, a polychlorinated biphenyl-degrader, is arranged in a large linear chromosome of 7,8 Mbp with three other linear plasmids (1,1 Mbp, 442 Kbp and 332 Kbp, respectively). Its sequencing revealed insights into its exceptional catabolic richness, comprising numerous ligases and oxygenases, the latter present in aromatic compounds degradation pathways (McLeod et al., 2006). Also, genome analyses have shown that *Rhodococcus* degrading-abilities are based on a hyper-recombination strategy associated with large genomes for broad-range substrate utilization (Larkin et al., 2005). The hyper-recombination strategy relies upon the acquisition and storage of many genes to deploy for recombination, promoting dispersal of newly acquired DNA without the help of mobile genetic elements (Larkin et al., 2005). In addition to their large genomes, *Rhodococcus* harbours small circular and/or large linear plasmids which overall contributes to explaining their rich repertoire of catabolic genes (Larkin et al., 2010).

2. Data description

The Chilean Patagonia is one of the most extended fjord regions in the world, with a complex coastline and topography (Pantoja et al., 2011). Within this environment, Comau fjord was selected for bioprospection for being part of a pristine Marine Protected Area. The geographic uniqueness of this remote fjord proved to be a promising source of novel actinobacteria with biotechnological potential, specifically for producing antimicrobial activities (Undabarrena et al., 2016).

* Corresponding author.
E-mail address: beatriz.camara@usm.cl (B. Cámara).

https://doi.org/10.1016/j.margen.2018.01.004
Received 24 October 2017; Received in revised form 18 January 2018; Accepted 29 January 2018

Previously, one of these strains has been sequenced (Undabarrena et al., 2017a) revealing insights into the metabolic pathways involved in antimicrobial synthesis (Undabarrena et al., 2017b). *Rhodococcus* sp. H-CA8f was also obtained from this sampling campaign, which was isolated from 22-m-deep marine sediments retrieved from Lilihuapi Island, Comau fjord (Undabarrena et al., 2016). Phenotypic characterization, phylogenetic analysis and antimicrobial activity from *Rhodococcus* sp. H-CA8f's crude extracts were previously explored (Undabarrena et al., 2016), supporting its selection as an interesting candidate for whole-genome sequencing. Therefore in this study, we report the complete genome sequence of *Rhodococcus* sp. H-CA8f, which may facilitate the understanding of the genetic determinants involved in their metabolic versatility.

DNA extraction of *Rhodococcus* sp. H-CA8f was obtained using Genomic-tip 500/G kit (Qiagen, Germany) and subsequently purified using a DNA Clean & Concentrator™-100 kit (Zymo Research, USA). Quality and purity of DNA were verified by NanoDrop 2000 Spectrophotometer; Qubit dsDNA BR Assay kit (Thermo Fisher Scientific, USA); 0.8% agarose gel electrophoresis and PCR-amplification for subsequent 16S rRNA gene Sanger-sequencing. PacBio sequencing was achieved using a Single Molecule Real Time RS II platform (Uppsala Genome Center, National Genomics Intrastructure, SciLifeLab, Sweden) with one SMRT cell and a 20-kb insert library. The sequencing yielded 632,156,818 bp distributed in 56,549 reads with an average read length of 11,178 bp. PacBio reads were applied for self-correction and genome assembly using HGAP.3; which consists of a pre-assembly step followed by a *de novo* assembly step with PacBio's Assemble Unitig and a final polishing with Quiver (Chin et al., 2013). As a result, one gap-free contig of 6,196,004 bp was obtained comprising one entire closed chromosome, together with a second gap-free contig of 301,603 bp comprising a putative linear mega-plasmid. The chromosome (*i.e.*, unitig 0) presented a G + C content of 62.45% with approximately 80-fold coverage, whereas the plasmid (*i.e.*, unitig 1) bears a G + C content of 62.34% with approximately 100-fold coverage. Genome sequencing project and MIGS mandatory fields for *Rhodococcus* sp. H-CA8f are presented in Table 1, along with full genomic features which are detailed in Table 2.

Table 2
Genomic features of the *Rhodococcus* sp. H-CA8f genome.

MIGS ID	Features	*Rhodococcus* sp. H-CA8f
MIGS-32	GenBank ID	CP023720 (unitig 0); CP023721 (unitig 1)
MIGS-32	BIOPROJECT	PRJNA411856
MIGS-28	Finishing quality	closed complete genome
MIGS-29	Sequencing platforms	PacBio
MIGS-28	Libraries used	1 SMRT Cell
MIGS-31.2	Fold coverage	80.35 × (unitig 0); 100. 45 × (unitig 1)
MIGS-30	Assemblers	HGAP.3
MIGS-32	Gene calling method	PGAP (NCBI)
	Length (bp)	6,196,004 (unitig 0); 301,603 (unitig 1)
	Contigs	1 (unitig 0); 1 (unitig 1)
	G + C content (%)	62.45% (unitig 0); 62.34% (unitig 1)
	ORFs	6118
	CDS	6047
	rRNAs	15
	tRNAs	53
	ncRNA	3

To aid the characterization of *Rhodococcus* sp. H-CA8f genomic features, a complementary annotation through several algorithms was performed. First, both replicative units were annotated using the Prokaryotic Genome Annotation Pipeline (PGAP) at NCBI version 4.2 (Tatusova et al., 2016) resulting in a total of 6118 putative Open Reading Frames (ORFs) assigned as 6047 CDS; 53 tRNAs and 15 complete rRNAs. Analysis from this annotation revealed a total of 17.3% and 83.5% of hypothetical proteins for the chromosome and the plasmid, respectively. To reinforce the annotation concerning putative products, ORFs coding sequences were annotated against the bacterial eggNOG 4.5 database (Huerta-Cepas et al., 2016) using eggNOG mapper (Huerta-Cepas et al., 2017) with default parameters. Similarly, Prokka annotation (Seeman, 2014) indicated that 34.4% and 83.5% of putative genes encode for hypothetical proteins for the chromosome

Table 1
General features of the *Rhodococcus* sp. H-CA8f genome according to MIGS mandatory recommendations.

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain *Bacteria*	TAS (Woese et al., 1990)
		Phylum *Actinobacteria*	TAS (Garrity et al., 2001)
		Class *Actinobacteria*	TAS (Stackebrandt et al., 1997)
		Order *Actinomycetales*	TAS (Stackebrandt et al., 1997)
		Suborder *Corynebacterineae*	TAS (Stackebrandt et al., 1997)
		Family *Nocardiaceae*	TAS (Stackebrandt et al., 1997)
		Genus *Rhodococcus*	TAS (Undabarrena et al., 2016)
		Species *Rhodococcus* sp.	TAS (Undabarrena et al., 2016)
		Strain H-CA8f	TAS (Undabarrena et al., 2016)
	Gram stain	Positive	NAS
	Cell shape	Rods	TAS (Undabarrena et al., 2016)
	Motility	Non-motile	IDA
	Sporulation	Non-sporulating	TAS (Undabarrena et al., 2016)
	Temperature range	4–30 °C	TAS (Undabarrena et al., 2016)
	Optimum temperature	30 °C	TAS (Undabarrena et al., 2016)
	pH range; optimum	7	IDA
	Carbon source	D-Glucose	IDA
MIGS-6	Habitat	Marine Sediments	
MIGS-6.3	Salinity	3.5%	TAS (Undabarrena et al., 2016)
MIGS-22	Oxygen requirement	Aerobic	TAS (Undabarrena et al., 2016)
MIGS-15	Biotic relationship	Free-living	TAS (Undabarrena et al., 2016)
MIGS-14	Pathogenicity	Not reported	NAS
MIGS-4	Geographic location	Lilihuapi Island, Comau fjord	TAS (Undabarrena et al., 2016)
MIGS-5	Sample collection	Jan-2013	TAS (Undabarrena et al., 2016)
MIGS-4.1	Latitude	42° 20,634′S	TAS (Undabarrena et al., 2016)
MIGS-4.2	Longitude	72° 27, 429′W	TAS (Undabarrena et al., 2016)
MIGS-4.4	Altitude	−22.9 m depth	TAS (Undabarrena et al., 2016)

[a] Evidence codes – IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (*i.e.*, a direct reports exists in the literature); NAS: Non-traceable Author Statement (*i.e.*, not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are the Gene Ontology project (Ashburner et al., 2000).

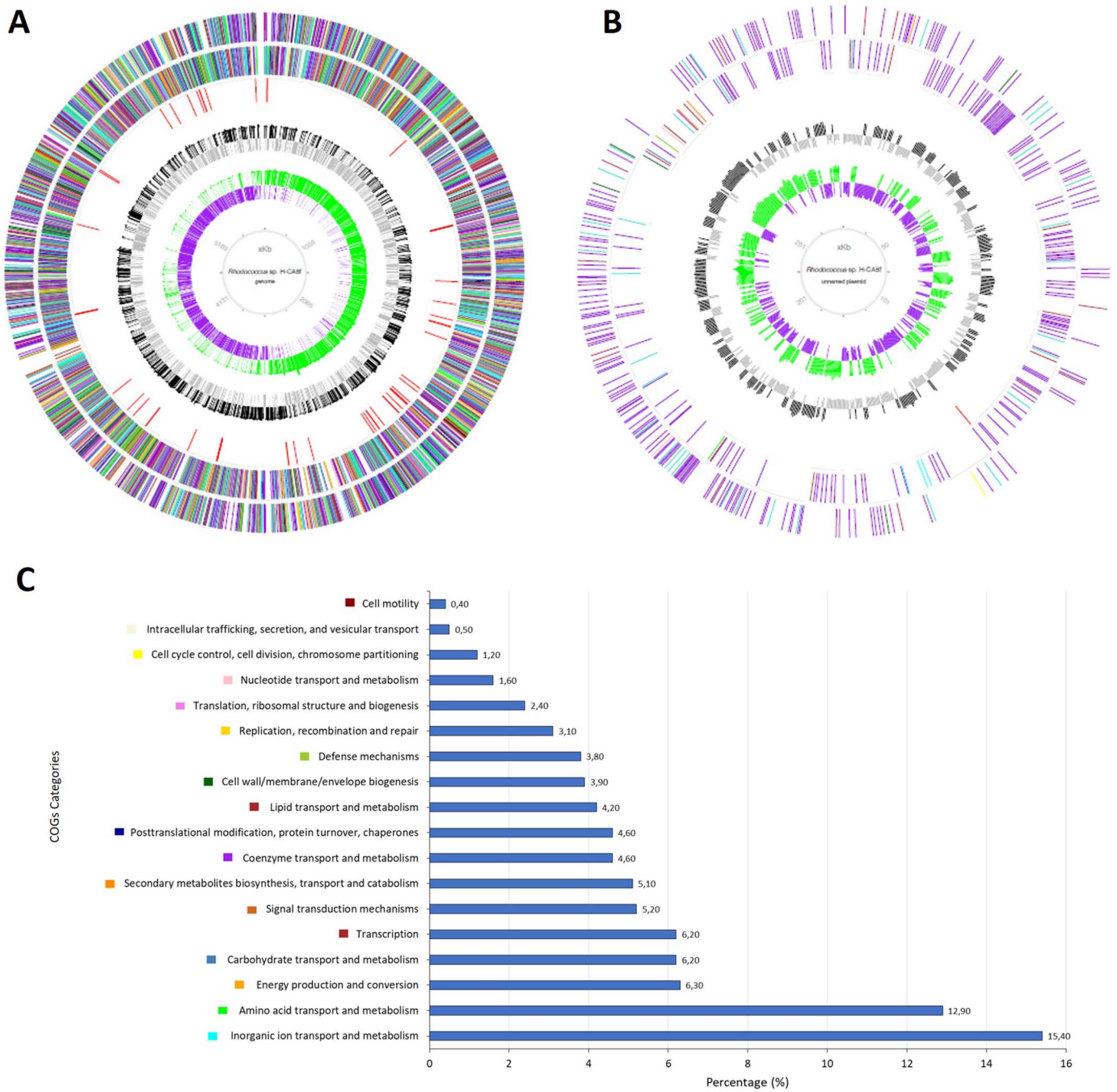

Fig. 1. Circular map of both replicative units of *Rhodococcus* sp. H-CA8f genome with coding sequences coloured by COGs categories (eggNOG): A) Complete chromosome; B) Linear mega- plasmid. Labeling from outside to the inside: ORFs on the forward strand; ORFs on the reverse strand; RNA genes (tRNAs red, rRNAs blue); G + C content (black/gray peaks indicate values higher/ lower than the average G + C content, respectively); GC skew (calculated as (G − C)/(G + C), green/purple peaks indicates values higher/lower than 1, respectively). G + C content and GC skew was calculated using a sliding window approach (3000 bp window size and 1500 bp step size); C) Detailed percentage of COGs categories for *Rhodococcus* sp. H-CA8f chromosome. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and the plasmid, respectively. Furthermore, ORFs with coding sequences were grouped into Clusters of Orthologous Group (COGs) categories, through annotation of protein sequences against the Conserved Domain Database (CDD) using the RPS-BLAST algorithm (Galperin et al., 2014; Marchler-Bauer et al., 2003). Fig. 1 depicts the circular map obtained using the plotMyGBK wrapper script (https://github.com/microgenomics/plotMyGBK) for both the chromosome (Fig. 1A) and the plasmid (Fig. 1B) along with their gene's functional classification into COGs categories (eggNOG) represented by different colours. Comparatively, it can be easily visualized that only a few genes in the plasmid were able to be classified, represented mostly by putative genes. In contrast, it was possible to classify CDS according to COGs functionality for the chromosome, where the percentage of each category is detailed in Fig. 1C. Most abundant categories are: *inorganic ion transport and metabolism* (15%) and *amino acid transport and metabolism* (12%); with 25% of poorly characterized genes (*i.e.*, *general function prediction* or *function unknown*). As for the plasmid, the majority of coding sequences belong to unknown or poorly characterized functions (*i.e.*, 31% ORFs are grouped into *general function prediction* or *function unknown* categories). These results demonstrate that a considerable fraction of genes discovered in the *Rhodococcus* sp. H-CA8f's plasmid encode for proteins that have no assigned function, remaining

unknown, and may represent novel features so far not described. In order to further address the novelty of the plasmid, PANNZER2 was used, a protein function annotation framework very effective according to the latest CAFA contest (Jiang et al., 2016). GeneMarkS protein prediction of the plasmid sequence was used as input against PANNZER2 (Koskinen et al., 2015) and SANSparallel (Somervuo and Holm, 2015) remote servers, revealing no matches within these databases and, therefore, with no possibility of assignment of Gene Orthology (GO) for any of the sequences. Only few genetic determinants were possible to be identified through the RAST annotation server, which are involved in the partitioning proteins ParA/ParB; CRISPR-Cas systems and phage-related genes. Accordingly, this striking feature has been reported for *Rhodococcus* linear plasmids, constituted by a substantial number of putative genes likely to been acquired by Horizontal Gene Transfer (HGT) and probably involved in catabolic functions (Warren et al., 2004). Also, circular plasmids are considerably smaller in size in comparison to the linear plasmids (Larkin et al., 2010). Therefore for *Rhodococcus* sp. H-CA8f's plasmid linearity confirmation, a BLAST analysis against itself was conducted and, as hypothesized, the ends of the unitig 1 do not overlap with each other. In addition, a BLASTn analysis confirmed a hit of more than 40 bp (42/48 bp) between the beginning and the end of the unitig 1, which may represent Terminal Inverted Repeats (TIRs). TIRs are typically reported at the ends of many *Streptomyces* linear genomes (Hinnebusch and Tilly, 1993). Similarly, studies on *Rhodococcus* telomers of linear plasmids showed similar features regarding linear plasmids from *Streptomyces* (Larkin et al., 2010). Therefore, we propose that unitig 1 may represent one linear mega-plasmid of 301 Kbp which harbours mostly putative ORFs encoding for hypothetical proteins or with unknown functions, representing a novel genetic feature of *Rhodococcus* sp. H-CA8f.

Alongside the plasmid, to aid the catabolic potential of *Rhodococcus* sp. H-CA8f, a detailed bioinformatics analysis of the chromosome was conducted. The *Rhodococcus* genus is recognized for metabolizing some of the most difficult substrates due to a high diversity of different types of mono- and di-oxygenases. In this sense, a great diversity of these enzymes was found in *Rhodococcus* sp. H-CA8f, comprising at least 104 genes encoding for monooxygenases (including 10 genes encoding for cytochrome P450 and 10 genes for alkanes or alkanesulfonate monooxygenases); along with 75 genes encoding for dioxygenases (including two genes encoding for catechol 1,2-dioxygenases, four genes encoding for protocatechuate 3,4- and 4,5-dioxygenases and one gene encoding for a homogentisate 1,2-dioxygenase and for a biphenyl 2,3-dioxygenase, respectively). One of the degradation pathways detected, the upper biphenyl catabolic pathway is represented within the *Rhodococcus* sp. H-CA8f's chromosome by one *bphA* gene encoding a biphenyl 2,3-dioxygenase (NCBI ID CPI83_03450); three copies of *bphC* gene, one annotated as 2,3-dihydroxybiphenyl 1,2-dioxygenase (NCBI ID CPI83_25055) and two as biphenyl-2,3-diol 1,2-dioxygenase (NCBI IDs CPI83_04510 and CPI83_22290); and one *bphD* gene encoding for 2-hydroxy-6-oxo-6-phenylhexa-2,4-dienoate hydrolase (NCBI ID CPI83_27250). However, no *bphB* gene (2,3-dihydro-2,3-dihydroxybiphenyl-2,3 dehydrogenase) was found in the NCBI or Prokka annotation. To address the functionality of this metabolic pathway, *Rhodococcus* sp. H-CA8f was grown on a minimum-medium plate with biphenyl as sole carbon source as previously described (Agulló et al., 2007), and colonies were observed after 4 days of incubation (data not shown). As the metabolic pathway seems to be active, a deeper search of the *bphB* gene was aimed. A comparison using DIAMOND (Buchfink et al., 2015) with an e-value threshold of 1e-20, was performed using dehydrogenases from recognized *Rhodococcus* biphenyl-degrader strains (*R. jostii* RHA1, *R. erythropolis* TA421, *R. globerulus* P6, and *Rhodococcus* sp. strain R04). A survey for the presence of Short-Chain Dehydrogenase/Reductase (SDR) domain (Pfam name: adh_short, accession: PF00106) and bphB/todD domain (TIGR name: BphB_TodD, accession: TIGR03325) were performed for the candidates queries using HMMER, giving positive results for both domains in all cases. Thus, a total of 8 putative *bphB* candidate genes were found in the *Rhodococcus* sp. H-CA8f's chromosome. Among them, the NCBI annotated gene as ID CPI83_22360 with a 49% BLAST identity to a trans-carveol dehydrogenase (annotated as a SDR family oxidoreductase and as a trans-carveol dehydrogenase with NCBI and Prokka, respectively) showed identity with all queries mentioned above (*R. erythropolis* TA421 and *R. globerulus* P6, 35.80% identity; *R. jostii bphB1*, 32.80% and *bphB2*, 32.10% identity; *Rhodococcus* sp. R04, 39.20% identity). However, as the value obtained is not enough for determining the presence of *bphB* gene, these results suggests that probably the *bphB* gene may have adopted a divergent strategy of evolution.

As exposure to toxic compounds may represent a stress condition to the cells, a search for genetic determinants involved in stress response in *Rhodococcus* sp. H-CA8f's chromosome was additionally aimed. Interestingly, a rich repertoire of genes involved in stress response were found, comprising: peroxidases & peroxiredoxines (34 genes) (including 8 genes encoding bromoperoxidases & chloroperoxidases and 6 genes encoding alkyl hydroperoxide reductases); thioredoxin & thioredoxin reductases (12 genes); and universal stress proteins (11 genes), among others. These genes may be involved in the general maintenance of *Rhodococcus* sp. H-CA8f metabolism when facing degradation of recalcitrant substrates or growth on toxic pollutants (*i.e.*, biphenyl) (Agulló et al., 2007), contributing to the wide tolerance against environmental abiotic stressors.

Finally, an antiSMASH version 4.0.2 (Medema et al., 2011) search was conducted to gain insights into the potential to produce secondary specialized metabolites, and 17 Biosynthetic Gene Clusters (BGCs) were identified for the *Rhodococcus* sp. H-CA8f chromosome. Among them, 6 Non-Ribosomal Peptides Synthetases (NRPSs) and 2 Polyketide Synthases (PKSs) were detected, which may be involved in the antimicrobial activity previously observed (Undabarrena et al., 2016).

To our knowledge, this is the first report of a closed, whole-genome sequence of a marine *Rhodococcus* isolate from a Chilean remote fjord. Interestingly, *Rhodococcus* sp. H-CA8f harbours a novel plasmid, mostly without functional characterization, which may deepen future comparative genomics studies. PacBio genome sequencing was crucial for closing the complete genome of *Rhodococcus* sp. H-CA8f, unveiling genomic features such as the biphenyl degradation pathway.

3. Accession number

Complete genome sequence of *Rhodococcus* sp. H-CA8f is available in NCBI under the accession number CP023720 (chromosome) and CP023721 (plasmid). Bioproject number is PRJNA411856 and Biosample number is SAMN07689502. The strain *Rhodococcus* sp. H-CA8f is deposited in the CCUG Culture Collection (Culture Collection University of Gothenburg, Sweden) under the number CCUG 69070.

Acknowledgements

Authors thank the scuba divers and staff from Huinay Scientific Field Station. This is publication N° 148 of the Huinay Scientific Field Station. Thanks to Kaisa Throrell, Lars Engstrand and staff of the sequencing facilities at SciLifeLab, Stockholm; Olga Vinnere Pettersson and staff of the Uppsala Genome Centre sequencing facilities, both from Sweden. This work was supported by the 'Comisión Nacional de Investigación Científica y Tecnológica' (FONDECYT) N° 1171555 (B.C), N° 1110991 (M.S), N° 11160905 (E.C.N) and N° 11140666 (J.A.U); and the Swedish Research Links Programme of the Swedish Research Council (Vetenskapsrådet Project N° 2013-6713) (E.R.B.M and B.C). A.U acknowledges CONICYT PhD fellowship and CONICYT Gastos Operacionales N° 21120621. D.J.L. and E.R.B.M. acknowledge support by the Joint Programme Initiative - Anti-Microbial Resistance (JPI-AMR) (Vetenskapsrådet Project N° 2016-06504). F.S.S., D.J.L. and E.R.B.M. were supported by the Culture Collection University of Gothenburg (CCUG), Department of Clinical Microbiology, Sahlgrenska

University Hospital, Gothenburg, Sweden.

References

Agulló, L., Cámara, B., Martinez, P., Latorre, V., Seeger, M., 2007. Response to (chloro) biphenyls of the polychlorobiphenyl-degrader *Burkholderia xenovorans* LB400 involves stress proteins also induced by heat shock and oxidative stress. FEMS Microbiol. Lett. 267 (2), 167–175.

Alvarez, P.J., Anid, P.J., Vogel, T.M., 1991. Kinetics of aerobic biodegradation of benzene and toluene in sandy aquifer material. Biodegradation 2 (1), 43–51.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene ontology: tool for the identification of biology. Nat. Genet. 25, 25–29.

Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.P., Clément, C., Ouhdouch, Y., van Wezel, G.P., 2016. Taxonomy, physiology, and natural products of *Actinobacteria*. Microbiol. Mol. Biol. Rev. 25 (80 (1)), 1–43.

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12 (1), 59–60.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., Turner, S.W., Korlach, J., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10 (06), 563–569.

Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2014. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43 (Database issue), D261–D269.

van der Geize, R., Dijkhuizen, L., 2004. Arnessing the catabolic diversity of rhodococci for environmental and biotechnological applications. Curr. Opin. Microbiol. 7 (3), 255–261.

Hinnebusch, J., Tilly, K., 1993. Linear plasmids and chromosomes in bacteria. Mol. Microbiol. 10 (5), 917–922.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., alter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., von Mering, C., Bork, P., 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44 (D1), D286–93.

Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., Bork, P., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34 (8), 2115–2122.

Jiang, Y., Oron, T.R., Clark, W.T., Bankapur, A.R., D'Andrea, D., Lepore, R., Funk, C.S., Kahanda, I., Verspoor, K.M., Ben-Hur, A., Koo Da, C.E., Penfold-Brown, D., Shasha, D., Youngs, N., Bonneau, R., Lin, A., Sahraeian, S.M., Martelli, P.L., Profiti, G., Casadio, R., Cao, R., Zhong, Z., Cheng, J., Altenhoff, A., Skunca, N., Dessimoz, C., Dogan, T., Hakala, K., Kaewphan, S., Mehryary, F., Salakoski, T., Ginter, F., Fang, H., Smithers, B., Oates, M., Gough, J., Törönen, P., Koskinen, P., Holm, L., Chen, C.T., Hsu, W.L., Bryson, K., Cozzetto, D., Minneci, F., Jones, D.T., Chapman, S., Bkc, D., Khan, I.K., Kihara, D., Ofer, D., Rappoport, N., Stern, A., Cibrian-Uhalte, E., Denny, P., Foulger, R.E., Hieta, R., Legge, D., Lovering, R.C., Magrane, M., Melidoni, A.N., Mutowo-Meullenet, P., Pichler, K., Shypitsyna, A., Li, B., Zakeri, P., ElShal, S., Tranchevent, L.C., Das, S., Dawson, N.L., Lee, D., Lees, J.G., Sillitoe, I., Bhat, P., Nepusz, T., Romero, A.E., Sasidharan, R., Yang, H., Paccanaro, A., Gillis, J., Sedeño-Cortés, A.E., Pavlidis, P., Feng, S., Cejuela, J.M., Goldberg, T., Hamp, T., Richter, L., Salamov, A., Gabaldon, T., Marcet-Houben, M., Supek, F., Gong, Q., Ning, W., Zhou, Y., Tian, W., Falda, M., Fontana, P., Lavezzo, E., Toppo, S., Ferrari, C., Giollo, M., Piovesan, D., Tosatto, S.C., Del Pozo, A., Fernández, J.M., Maietta, P., Valencia, A., Tress, M.L., Benso, A., Di Carlo, S., Politano, G., Savino, A., Rehman, H.U., Re, M., Mesiti, M., Valentini, G., Bargsten, J.W., van Dijk, A.D., Gemovic, B., Glisic, S., Perovic, V., Veljkovic, V., Veljkovic, N., Almeida-E-Silva, D.C., Vencio, R.Z., Sharan, M., Vogel, J., Kansakar, L., Zhang, S., Vucetic, S., Wang, Z., Sternberg, M.J., Wass, M.N., Huntley, R.P., Martin, M.J., O'Donovan, C., Robinson, P.N., Moreau, Y., Tramontano, A., Babbitt, P.C., Brenner, S.E., Linial, M., Orengo, C.A., Rost, B., Greene, C.S., Mooney, S.D., Friedberg, I., Radivojac, P., 2016. An expanded evaluation of protein function prediction methods shows an improvement in accuracy. Genome Biol. 17 (1), 184.

Koskinen, P., Törönen, P., Nokso-Koivisto, J., Holm, L., 2015. PANNZER: high-throughput functional annotation of uncharacterized proteins in a error – prone environment. Bioinformatics 31 (10), 1544–1552.

Larkin, M.J., Kulakov, L.A., Allen, C.C.R., 2005. Biodegradation and *Rhodococcus* –masters of catabolic versatility. Curr. Opin. Biotechnol. 16, 282–290.

Larkin, M.J., Kulakov, L.A., Allen, C.C.R., 2006. Biodegradation by members of the genus *Rhodococcus*: biochemistry, physiology, and genetic adaptation. Adv. Appl. Microbiol. 59, 1–29.

Larkin, M.J., Kulakov, L.A., Allen, C.C.R., 2010. Genomes and plasmids in *Rhodococcus*. In: Alvarez, H. (Ed.), Biology of *Rhodococcus*. Microbiology Monographs, vol. 16 Springer, Berlin, Heidelberg.

Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., Liebert, C.A., Liu, C., Madej, T., Marchler, G.H., Mazumder, R., Nikolskaya, A.N., Panchenko, A.R., Rao, B.S., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Vasudevan, S., Wang, Y., Yamashita, R.A., Yin, J.J., Bryant, S.H., 2003. CDD: a curated Entrez database of conserved domain alignments. Nucleic Acids Res. 31 (1), 383–387.

McLeod, M.P., Warren, R.L., Hsiao, W.W., Araki, N., Myhre, M., Fernandes, C., Miyazawa, D., Wong, W., Lillquist, A.L., Wang, D., Dosanjh, M., Hara, H., Petrescu, A., Morin, R.D., Yang, G., Stott, J.M., Schein, J.E., Shin, H., Smailus, D., Siddiqui, A.S., Marra, M.A., Jones, S.J., Holt, R., Brinkman, F.S., Miyauchi, K., Fukuda, M., Davies, J.E., Mohn, W.W., Eltis, L.D., 2006. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. Proc. Natl. Acad. Sci. 103, 15582–15587.

Medema, M.H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M.A., Weber, T., Takano, E., Breitling, R., 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 39, W339–46.

Pantoja, S., Iriarte, J.L., Daneri, G., 2011. Oceanography of the Chilean Patagonia. Cont. Shelf Res. 31, 149–153.

Seeman, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30 (14), 2068–2069.

Somervuo, P., Holm, L., 2015. SANSparallel: interactive homology search against Uniprot. Nucleic Acids Res. 43 (W1), W24–29.

Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., Ostell, J., 2016. NCBI prokaryotic genome annotation pipeline. Nucleic Acids Res. 44 (14), 6614–6624.

Undabarrena, A., Beltrametti, F., Claverías, F.P., González, M., Moore, E.R.B., Seeger, M., Cámara, B., 2016. Exploring the diversity and antimicrobial potential of marine *Actinobacteria* from the Comau fjord in Northern Patagonia, Chile. Front. Microbiol. 7, 1135.

Undabarrena, A., Ugalde, J.A., Castro-Nallar, E., Seeger, M., Cámara, B., 2017a. Genome sequence of *Streptomyces* sp. H-KF8 a marine actinobacterium isolated from a Northern Chilean Patagonian fjord. Genome Announc 5 (6) (pii:e01645–16).

Undabarrena, A., Ugalde, J.A., Seeger, M., Cámara, B., 2017b. Genomic data mining of the marine actinobacteria Streptomyces sp. H-KF8 unveils insights into multi-stress related genes and metabolic pathways involved in antimicrobial synthesis. PeerJ. 5, e2912.

Wang, Z., Xu, J., Li, Y., Wang, K., Wang, Y., Hong, Q., Li, W.J., Li, S.P., 2010. *Rhodococcus jianlingiae* sp. nov., an actinobacterium isolated from sludge of a carbendazim wastewater treatment facility. Int. Syst. Evol. Microbiol. 60 (Pt 2), 378–381.

Warren, R., Hsiao, W.W., Kudo, H., Myhre, M., Dosanjh, M., Petrescu, A., Kobayashi, H., Shimizu, S., Miyauchi, K., Masai, E., Yang, G., Stott, J.M., Schein, J.E., Shin, H., Khattra, J., Smailus, D., Butterfield, Y.S., Siddiqui, A., Holt, R., Marra, M.A., Jones, S.J., Mohn, W.W., Brinkman, F.S., Fukuda, M., Davies, J., Eltis, L.D., 2004. Functional characterization of a catabolic plasmid from polychlorinated- biphenyl-degrading *Rhodococcus* sp. strain RHA1. J. Bacteriol. 186 (22), 7783–7795.

Zopf, W., 1891. Über Ausscheidung von Fettfarbstoffen (Lipochromen) seitens gewisser Spaltpilze Ber. Dtsch. Bot. Ges. 9, 22–28 (in German).