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).