genomeA
Journals.ASM.org

Genome Sequence of *Rhodococcus* sp. Strain PML026, a Trehalolipid Biosurfactant Producer and Biodegrader of Oil and Alkanes

C. M. Sambles,[a] D. A. White[b]

School of Biosciences, Geoffrey Pope Building, University of Exeter, Exeter, United Kingdom[a]; Plymouth Marine Laboratory, Prospect Place, Plymouth, Devon, United Kingdom[b]

***Rhodococcus* sp. strain PML026 produces an array of trehalolipid biosurfactant compounds in order to utilize hydrophobic carbon sources, such as oils and alkanes. Here, we report the high-quality draft genome sequence of this strain, which has a total length of 5,168,404 bp containing 4,835 protein-coding sequences, 12 rRNAs, and 45 tRNAs.**

Received 27 March 2015 **Accepted** 30 March 2015 **Published** 7 May 2015

Citation Sambles CM, White DA. 2015. Genome sequence of *Rhodococcus* sp. strain PML026, a trehalolipid biosurfactant producer and biodegrader of oil and alkanes. Genome Announc 3(3):e00433-15. doi:10.1128/genomeA.00433-15.

Address correspondence to D. A. White, dani@pml.ac.uk.

Members of the *Rhodococcus* genus are well known for their ability to produce cell-bound and extracellular trehalolipid biosurfactant compounds in the presence of hydrophobic substrates (1–3). Trehalolipids exhibit a range of potential bioactivities (4–9) and are excellent emulsifying compounds with applications in microbe-enhanced oil recovery and oil spill treatment (10, 11). *Rhodococcus* sp. strain PML026 is a novel marine bacterium that was recently isolated and shown to produce a range of trehalolipid compounds in order to assimilate oil and alkanes (12). To better understand the trehalolipid production and other abilities of this strain, a genome sequence analysis of *Rhodococcus* sp. PML026 was carried out.

Using the Illumina HiSeq 2500 platform, two paired-end libraries (insert sizes, 250 bp and 500 bp) and two mate-pair libraries (insert sizes, 9 kbp and 11 kbp) were sequenced, generating 146,696,910 pairs of sequence reads (100 bp) from the two paired-end libraries and 73,808,154 read pairs (100 bp) from the two mate-pair libraries. After duplicate removal using FastUniq (13) and filtering and trimming using Trim Galore, 93% of the read pairs (136,458,888) from the paired-end libraries were retained, and 78% of the read pairs (57,494,134) from the mate-pair libraries remained. Using subsets of the two paired-end libraries, ~5 million trimmed and filtered reads per library were used for genome assembly with SPAdes (version 3.1.1) (14) using the parameter --careful with *k*-mers of 21, 33, 55, 77, 99, and 127. Further scaffolding was performed with SSPACE (version 3) (15) using all trimmed and filtered paired-end reads and mate-pair reads. Finally, GapFiller (version 1-10) (16) was used to close 19 gaps, and incorrect scaffolds were split using REAPR (version 1.0.17). The resulting assembly consisted of 37 contigs in 16 scaffolds with a total length of 5,168,404 bp. Annotation was performed using Prokka (version 1.10) (17) using a genus database generated from four *Rhodococcus* sp. genomes, *R. erythropolis* PR4 (GenBank accession no. NC_012490), *R. jostii* RHA1 (GenBank accession no. NC_008268), *R. opacus* B4 (GenBank accession no. NC_012522), and *R. pyridinivorans* SB3094 (GenBank accession no. NC_023150). tRNAs and transfer-messenger RNAs (tmRNAs) were predicted using Aragorn (version 1.2), ribosomal RNAs with Barrnap (version 0.5), and coding sequences with Prodigal (version 2.60). The genome contains 4,835 protein-coding sequences, 12 rRNAs, and 45 tRNAs.

Nucleotide sequence accession number. The genome sequence and annotation data for *Rhodococcus* sp. PML026 have been submitted to GenBank under the accession no. JZIS00000000.

ACKNOWLEDGMENTS

This work was supported in full by internal funding (PML-Exeter collaboration fund).

We are grateful to Audrey Farbos, Karen Moore, and Konrad Paszkiewicz of the Exeter Sequencing Service facility.

REFERENCES

1. **Lang S, Philp JC.** 1998. Surface-active lipids in rhodococci. Antonie Van Leeuwenhoek **74:**59–70. http://dx.doi.org/10.1023/A:1001799711799.
2. **Shul'ga AN, Karpenko EV, Eliseev SA, Turovskii AA, Koronelli TV.** 1990. Extracellular lipids and surface active properties of *Rhodococcus erythropolis* depending on the source of carbon nutrition. Mikrobiologiya **59:**443–447.
3. **Yakimov MM, Giuliano L, Bruni V, Scarfi S, Golyshin PN.** 1999. Characterization of Antarctic hydrocarbon-degrading bacteria capable of producing bioemulsifiers. New Microbiol **22:**249–256.
4. **Isoda H, Shinmoto H, Matsumura M, Nakahara T.** 1995–1996. Succinoyl trehalose lipid induced differentiation of human monocytoid leukemic cell line U937 into monocyte-macrophages. Cytotechnology **19:**79–88. http://dx.doi.org/10.1007/BF00749758.
5. **Ortiz A, Teruel JA, Espuny MJ, Marqués A, Manresa A, Aranda FJ.** 2008. Interactions of a *Rhodococcus* sp. biosurfactant trehalose lipid with phosphatidylethanolamine membranes. Biochim Biophys Acta **1778:**2806–2813. http://dx.doi.org/10.1016/j.bbamem.2008.07.016.
6. **Ortiz A, Teruel JA, Manresa Á, Espuny MJ, Marqués A, Aranda FJ.** 2011. Effects of a bacterial trehalose lipid on phosphatidylglycerol membranes. Biochim Biophys Acta **1808:**2067–2072. http://dx.doi.org/10.1016/j.bbamem.2011.05.003.
7. **Zaragoza A, Aranda FJ, Espuny MJ, Teruel JA, Marqués A, Manresa A, Ortiz A.** 2009. Mechanism of membrane permeabilization by a bacterial trehalose lipid biosurfactant produced by *Rhodococcus* sp. Langmuir **25:**7892–7898. http://dx.doi.org/10.1021/la900480q.
8. **Zaragoza A, Aranda FJ, Espuny MJ, Teruel JA, Marqués A, Manresa A, Ortiz A.** 2010. Hemolytic activity of a bacterial trehalose lipid biosurfac-

tant produced by *Rhodococcus* sp.: evidence for a colloid-osmotic mechanism. Langmuir **26:**8567–8572. http://dx.doi.org/10.1021/la904637k.
9. **Zaragoza A, Teruel JA, Aranda FJ, Marqués A, Espuny MJ, Manresa Á, Ortiz A**. 2012. Interaction of a *Rhodococcus* sp. trehalose lipid biosurfactant with model proteins: thermodynamic and structural changes. Langmuir **28:**1381–1390. http://dx.doi.org/10.1021/la203879t.
10. **Liu C, Liu H**. 2011. *Rhodococcus erythropolis* strain NTU-1 efficiently degrades and traps diesel and crude oil in batch and fed-batch bioreactors. Proc Biochem **46:**202–209. http://dx.doi.org/10.1016/j.procbio.2010.08.008.
11. **Pacheco GJ, Ciapina EMP, Gomes EDB, Pereira Junior N**. 2010. Biosurfactant production by *Rhodococcus erythropolis* and its application to oil removal. Braz J Microbiol **41:**685–693. http://dx.doi.org/10.1590/S1517-83822010000300019.
12. **White DA, Hird LC, Ali ST**. 2012. Production and characterisation of a trehalolipid biosurfactant produced by the novel marine bacterium *Rhodococcus* sp., strain PML026. J Appl Microbiol **115:**744–755. http://dx.doi.org/10.1111/jam.12287.
13. **Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S**. 2012. FastUniq: a fast *de novo* duplicates removal tool for paired short reads. PLoS One **7:**e52249. http://dx.doi.org/10.1371/journal.pone.0052249.
14. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA**. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.
15. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W**. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27:**578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.
16. **Nadalin F, Vezzi F, Policriti A**. 2012. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. BMC Bioinformatics **13**(Suppl 1):S8. http://dx.doi.org/10.1186/1471-2105-13-S14-S8.
17. **Seemann T**. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.