Complete Genome Sequence of *Rhodococcus* sp. Strain 9, Isolated from Contaminated Australian Groundwater

Varsha Naidu,[a,b] Lachlan Hyde,[a*] Bart A. Eijkelkamp,[c] Mallavarapu Megharaj,[d] Suresh Subashchandrabose,[d]§ Karl A. Hassan[a,b]

aSchool of Environmental and Life Sciences, University of Newcastle, Callaghan, New South Wales, Australia
bARC Centre of Excellence in Synthetic Biology, Macquarie University, North Ryde, New South Wales, Australia
cMolecular Sciences and Technology, College of Science and Engineering, Flinders University, Bedford Park, South Australia, Australia
dGlobal Centre for Environmental Remediation (GCER), University of Newcastle, Callaghan, New South Wales, Australia

ABSTRACT   Here, we report the 7.7-Mbp genome sequence of *Rhodococcus* sp. strain 9, which was isolated from Australian groundwater contaminated with phenols and trichloroethylene. This strain has previously been shown to efficiently degrade *p*-nitrophenol and high-molecular-weight polycyclic aromatic hydrocarbons (PAHs).

*R*hodococcus is a Gram-positive genus classified in the phylum *Actinomycetota* (1). Species within this genus have an extraordinary catabolic versatility, allowing them to degrade several recalcitrant pollutants (1). Here, we present the genomic data of *Rhodococcus* sp. strain 9, which was isolated from Australian groundwater contaminated with phenols and trichloroethylene (2). Strain 9 has been shown to degrade and utilize *p*-nitrophenol (2) and low- and high-molecular-weight polycyclic aromatic hydrocarbons (PAHs), including phenanthrene, pyrene, and benzo[*a*]pyrene, as sources of carbon and energy (3, 4).

Strain 9 was grown in Luria-Bertani broth at 25°C with shaking (120 rpm) for 40 h. Total genomic DNA was extracted using the DNeasy UltraClean microbial kit (Qiagen), following the manufacturer's instructions. The same extracted DNA was used for Illumina and Nanopore library preparation and sequencing at the Microbial Genome Sequencing Center (Pittsburgh, USA). Illumina sequencing libraries were prepared using the Illumina DNA prep kit and Integrated DNA Technologies 10-bp unique dual indexes; the libraries were sequenced on an Illumina NextSeq 2000 platform (2 × 151 bp), demultiplexed, and adaptor trimmed using bcl2fastq v2.20.0422. Oxford Nanopore sequencing libraries were prepared using Oxford Nanopore's "Genomic DNA by Ligation" protocol and sequenced on a Nanopore R9 flow cell (R9.4.1) with no size selection, followed by signal processing, base calling, and adapter trimming using Guppy v4.2.2 (GPU mode). The Illumina reads were processed using fastp v0.22.0 (5) to remove adapter and barcode sequences, correct mismatched bases in overlaps, and filter low-quality reads. Nanopore long reads of <1 kb were removed using Filtlong v0.2.0 (https://github.com/rrwick/Filtlong). *De novo* hybrid assembly was performed with 166,391 filtered reads (N_{50}, 11,298 bp) using Trycycler v0.3.3 (6). Briefly, 12 assemblies were generated using Flye v2.9-b1768 (7), Raven v1.2 (8), and Minipolish v0.1.2 (9), which predicted 3 contigs. The Trycycler v0.3.3 cluster function was used to group similar contigs from different assemblies into clusters (6). The Trycycler v0.3.3 reconcile function was used to reconcile each contig in each cluster, including rotating, trimming, and circularization, which resulted in two circular contigs and one linear contig (6). The Trycycler v0.3.3 functions msa, partition, and consensus were used to generate a consensus assembly, which was polished using Medaka v1.0.3, Bowtie2 v2.4.5 (10), and Pilon v1.24 (11). The linear contig was reassembled using Unicycler (Galaxy v0.4.8.0) (12) to recover the ends, which resulted in 691,102 bp. A BLASTn search of NCBI was performed (28 July 2022), using as the query 1 kbp of the right and left ends of the linear contig, which matched (query coverage, >95%; identity, >87%) the *Rhodococcus jostii* RHA1 linear plasmids pRHL2 (GenBank accession no. CP000433.1) and pRHL3

(CP000434.1). Collectively, this assembly predicted a circular 7,770,219-bp chromosome, a potentially linear 691,102-bp plasmid, and a circular 139,990-bp plasmid. Default parameters were used for all computational tools unless specified otherwise. Functional annotation of the genome was carried out using NCBI's Prokaryotic Genome Annotation Pipeline (PGAP) v6.2 (13). Genome annotation predicted 8,095 genes in strain 9, of which 8,031 are coding DNA sequences and 64 genes encode RNA, including 12 rRNAs, 49 tRNAs, 2 noncoding RNAs (ncRNAs), and 1 transfer-messenger RNA (tmRNA).

The genome of strain 9 was compared with other *Rhodococcus* species. FastANI v1.3 (14) was used to compute the average nucleotide identity (ANI) against 99 *Rhodococcus* genomes in the RefSeq database (downloaded on 7 April 2022) which had "complete genome assembly" status. The organism with the highest ANI (98.0843%) to strain 9 was *Rhodococcus opacus* 1CP (BioProject accession no. PRJNA253567).

Data availability. The draft genome sequence is available at NCBI GenBank under the accession no. CP095403.1, and the raw sequencing reads are available at the NCBI Sequence Read Archive (SRA) under the BioProject accession no. PRJNA813342. The version described in this article is the first version.

REFERENCES

1. Zampolli J, Zeaiter Z, Di Canito A, Di Gennaro P. 2019. Genome analysis and -omics approaches provide new insights into the biodegradation potential of Rhodococcus. Appl Microbiol Biotechnol 103:1069–1080. https://doi.org/10.1007/s00253-018-9539-7.
2. Subashchandrabose SR, Venkateswarlu K, Krishnan K, Naidu R, Lockington R, Megharaj M. 2018. Rhodococcus wratislaviensis strain 9: an efficient p-nitrophenol degrader with a great potential for bioremediation. J Hazard Mater 347:176–183. https://doi.org/10.1016/j.jhazmat.2017.12.063.
3. Subashchandrabose SR, Venkateswarlu K, Naidu R, Megharaj M. 2019. Biodegradation of high-molecular weight PAHs by Rhodococcus wratislaviensis strain 9: overexpression of amidohydrolase induced by pyrene and BaP. Sci Total Environ 651:813–821. https://doi.org/10.1016/j.scitotenv.2018.09.192.
4. Subashchandrabose SR, Venkateswarlu K, Venkidusamy K, Palanisami T, Naidu R, Megharaj M. 2019. Bioremediation of soil long-term contaminated with PAHs by algal-bacterial synergy of Chlorella sp. MM3 and Rhodococcus wratislaviensis strain 9 in slurry phase. Sci Total Environ 659:724–731. https://doi.org/10.1016/j.scitotenv.2018.12.453.
5. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
6. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Meric G, Vezina B, Wyres KL, Holt KE. 2021. Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biol 22:266. https://doi.org/10.1186/s13059-021-02483-z.
7. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
8. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci 1:332–336. https://doi.org/10.1038/s43588-021-00073-4.
9. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152.
10. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
11. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
12. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.
13. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
14. Jain C, Rodriguez RL, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.