AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Rhodococcus* sp. Strain W8901, a Human Clinical Specimen, Assembled Using MiSeq and MinION Sequence Data

Christopher A. Gulvik,[a] Dhwani Batra,[b] Dakota T. Howard,[b] Milli Sheth,[b] Ben W. Humrighouse,[a] Justin Lee,[b] John R. McQuiston,[a] Brent A. Lasker[a]

[a]Bacterial Special Pathogens Branch, Division of High-Consequence Pathogens and Pathology, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[b]Biotechnology Core Facility Branch, Division of Scientific Resources, National Center for Emerging and Zoonotic Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

ABSTRACT *Rhodococcus* sp. strain W8901 is a Gram-positive, aerobic, mycolic acid-containing coccobacillus obtained from a patient with acute lymphocytic leukemia. Here, we report on the complete, circular genome sequence obtained using Illumina MiSeq and Oxford Nanopore Technologies MinION reads in order to better resolve the phylogeny of a rare pathogen.

The genus *Rhodococcus* (1) presently includes >50 validly named species (https://www.bacterio.net/genus/rhodococcus). Members of the genus *Rhodococcus* are characterized as Gram-positive, nonmotile, aerobic, catalase-positive, and non-endospore-producing, with their cell wall comprising chemotype IV peptidoglycan and mycolic acids composed of 34 to 64 carbon atoms (2–4). *Rhodococcus* species are noted for their catabolic and biodegradation abilities but, other than *Rhodococcus equi*, are rare opportunistic pathogens (3). Isolate 86-07, here designated strain W8901, was obtained by culturing central venous catheter blood from a patient diagnosed with acute lymphocytic leukemia. Isolate W8901 was streaked on Trypticase soy agar supplemented with 5% sheep blood (TSAB) and grown for 3 days at 35°C. As described by Langer et al. (5), genomic DNA from strain W8901 was purified and amplified by PCR, and subsequent multilocus sequence typing, and 16S rRNA gene sequence analysis showed close sequence similarity to *R. equi* DSM 20307^T. The complete genome of strain W8901 was sequenced to better understand the taxonomic assignment of this rare human pathogen.

Strain W8901 was streaked onto TSAB from stock maintained at the Centers for Disease Control and Prevention, and a colony was cultured in Trypticase soy broth for 3 days at 35°C at 200 rpm. Genomic DNA used for both libraries was purified according to the manufacturer's protocol using the MasterPure DNA purification kit (Epicentre, Madison, WI). Genomic DNA libraries were made with the NEBNext Ultra DNA library prep kit (New England BioLabs, Ipswich, MA, USA) and the rapid barcoding kit (Oxford Nanopore Technologies [ONT]). Libraries sequenced with a 2 × 250-bp MiSeq (Illumina, San Diego, CA, USA) and a MinION (ONT) instrument using R9.4.1 flow cells produced 12,151,790 and 167,065 reads, respectively. Default parameters were used for all software unless otherwise specified. The long-read base calling with Guppy v3.2.8 generated an N_{50} value of 14,417 bp. Flye v2.7-b1585 (6) was used to assemble a single circular 5,713,496-Mbp chromosome with 191× coverage. Visualization with Bandage v0.8.1 suggested that the genome is complete with a single circular chromosome (7). The MinION reads were then used to correct assembly errors with three sequential rounds of read mapping using Minimap v2.17-r941 with the "-x map-ont" setting and correction with Racon v1.3.2 (8, 9). Medaka's consensus v1.0.1 (https://github.com/nanoporetech/medaka) function was used as a final long-read correction round

Citation Gulvik CA, Batra D, Howard DT, Sheth M, Humrighouse BW, Lee J, McQuiston JR, Lasker BA. 2021. Complete genome sequence of *Rhodococcus* sp. strain W8901, a human clinical specimen, assembled using MiSeq and MinION sequence data. Microbiol Resour Announc 10:e00613-21. https://doi.org/10.1128/MRA.00613-21.

Editor J. Cameron Thrash, University of Southern California

Address correspondence to Brent A. Lasker, blasker@cdc.gov.

Received 14 June 2021
Accepted 23 July 2021
Published 2 September 2021

prior to short-read (single nucleotide polymorphism [SNP] and indel) polishing. Paired short reads (trimmed to a Phred score of at least 30 with Trimmomatic v0.35 [10]) were mapped to the chromosome for sequential polishing stages until no more errors were identified. Unicycler's polish function along with Bowtie v2.3.5.1, SAMtools v1.10, and Pilon v1.23 corrected 120 SNPs and 2,165 indels (11–14). Of the indels, 1,838 were homopolymers (98% were C or G), and the other remaining corrections were 37 insertions and 290 deletions. The *dnaA* sequence coordinates were located with Prokka v1.14.0, and the chromosome's start was reoriented to it with Biopython v1.74 (15, 16). The polished chromosome contains 68.13% GC and was annotated with Prokaryotic Genome Annotation Pipeline (PGAP) v4.11 (17), which identified 5 rRNA gene operons and 5,062 putative protein-encoding genes. The current closest phylogenetically related genome is that of *Rhodococcus defluvii* Ca11^T with an ANI of 89.2% and digital DNA-DNA hybridization (dDDH) of 36.1%, which suggests that strain W8901 is within the radiation of the *Rhodococcus* genus (18, 19).

Data availability. The whole-genome sequence of *Rhodococcus* sp. W8901 has been deposited at the DDBJ/ENA/GenBank database under the accession number CP054690. The version described in this paper is the first version, CP054690.1. The sequencing read data have been deposited in the NCBI SRA under accession numbers SRR11951435 and SRR11951436.

ACKNOWLEDGMENTS

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention (CDC). The mention of company names or products does not constitute endorsement by the CDC.

REFERENCES

1. Zopf W. 1891. Über Ausscheidung von Fettfarbstoffen (Lipochromen) seitens gewisser Spaltpilze. Ber Dtsch Bot Ges 9:22–28.
2. Jones AL, Goodfellow ML. 2012. Genus IV. *Rhodococcus* (Zopf 1891) emend Goodfellow, Alderson and Chun 1998a, p 437–477. *In* Goodfellow M, Kampfer P, Busse HJ, Trujillo ME, Suzuki K, Ludwig W, Whitman WBV (ed), Bergey's manual of systematic bacteriology, 2nd ed, vol 5. The Actinobacteria, part A. Springer, New York, NY.
3. Majidzadeh M, Fatahi-Bafghi M. 2018. Current taxonomy of *Rhodococcus* species and their role in infections. Eur J Clin Microbiol Infect Dis 37:2045–2062. https://doi.org/10.1007/s10096-018-3364-x.
4. Sangal V, Goodfellow M, Jones AL, Schwalbe EC, Blom J, Hoskisson PA, Sutcliffe IC. 2016. Next-generation systematics: an innovative approach to resolve the structure of complex prokaryotic taxa. Sci Rep 6:38392. https://doi.org/10.1038/srep38392.
5. Langer AJ, Feja K, Lasker BA, Hinrikson HP, Morey RE, Pellegrini GJ, Smith TL, Robertson C. 2010. Investigation of an apparent outbreak of *Rhodococcus equi* bacteremia. Diagn Microbiol Infect Dis 67:95–100. https://doi.org/10.1016/j.diagmicrobio.2010.01.006.
6. Kolmogorov M, Yuan J, Lin Y, Pevzner P. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
7. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
8. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
9. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.
10. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
11. 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.
12. Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
14. 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.
15. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
16. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423. https://doi.org/10.1093/bioinformatics/btp163.
17. 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.
18. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. https://doi.org/10.1099/ijs.0.64483-0.
19. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60. https://doi.org/10.1186/1471-2105-14-60.