AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete and Circularized Bacterial Genome Sequence of *Gordonia* sp. Strain X0973

Christopher A. Gulvik,[a] Dhwani Batra,[b] Lori A. Rowe,[b] Milli Sheth,[b] Sarah Nobles,[b,c] Justin S. 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 Disease, 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 Disease, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

[c]Oak Ridge Institute for Science and Education, Oak Ridge, Tennessee, USA

ABSTRACT *Gordonia* sp. strain X0973 is a Gram-positive, weakly acid-fast, aerobic actinomycete obtained from a human abscess with *Gordonia araii* NBRC 100433^T as its closest phylogenetic neighbor. Here, we report using Illumina MiSeq and PacBio reads to assemble the complete and circular genome sequence of 3.75 Mbp with 3,601 predicted coding sequences.

In 1971, the genus *Gordonia* was proposed by Tsukamura to describe a group of aerobic, Gram-positive, weakly acid-fast, nonmotile, catalase-positive, arylsulfatase-negative, coccobacillary actinomycetes with an oxidative carbohydrate metabolism (1). Members of the genus *Gordonia* have been isolated from a variety of environments, including soil, water, plants, animals, wastewater, and activated sludge (2, 3). *Gordonia* species are noted for their biodegradative and bioremediation capabilities and their ability to synthesize novel secondary metabolites and industrially relevant enzymes (2–4). A few members are opportunistic pathogens and have been isolated from human sources, including sputum, sternal wound, lung, and ear infections (5). Isolate X0973 was acquired by the Centers for Disease Control and Prevention (CDC) for identification in 2012 from the hand abscess of a patient living in Missouri. Isolate X0973 was grown aerobically on Trypticase soy agar supplemented with 5% sheep blood (TSAB) at 35°C and then identified as a member of the genus *Gordonia* by 16S rRNA gene BLAST analysis (GenBank accession number KR259249) of the most similar 16S rRNA gene sequence to a validated type strain, *Gordonia amarae* ATCC 27808. In this investigation, we report the complete genome sequence of a rare human pathogen, *Gordonia* sp. strain X0973.

Gordonia sp. X0973 was obtained from the Special Bacteriology Reference Laboratory at the CDC and was cultured in a flask of 20 ml of Trypticase soy broth for 4 days at 35°C and 200 rpm from a single colony grown on TSAB. Cells were lysed and genomic DNA was purified using the Power Microbial DNA isolation kit. Genomic DNA libraries were prepared using the NEBNext Ultra DNA library prep kit. An Illumina MiSeq system using a v2 reagent kit generated 2 × 250-bp reads. Sequence reads (1,955,100 total) were filtered for read quality, base-called, and demultiplexed using bcl2fastq v2.20. Needle-sheared and Blue Pippin size-selected (20 kbp) DNA libraries were created with the SMRTbell template prep kit v1.0. The DNA/polymerase binding kit P6v2 and C4v2 chemistry were used for sequencing on an RS II instrument, which generated 202,665 total reads (N_{50}, 11,603 bp). Default parameters were used for all software unless otherwise specified. Bash5tools v0.8.0 was used to extract subreads of ≥500 bp. These 101,000 long reads (876 Mbp) were scrubbed as previously described (6, 7). Long-read scrubbing removed 214.5 Mbp (24.5%) due to low-quality

Citation Gulvik CA, Batra D, Rowe LA, Sheth M, Nobles S, Lee JS, McQuiston JR, Lasker BA. 2021. Complete and circularized bacterial genome sequence of *Gordonia* sp. strain X0973. Microbiol Resour Announc 10:e01479-20. https://doi.org/10.1128/MRA.01479-20.

Editor David Rasko, University of Maryland School of Medicine

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

Received 23 December 2020
Accepted 10 February 2021
Published 4 March 2021

scores, repaired 35.5 Mbp (4.1%) of low-quality nucleotides, and discarded 212.6 Mbp (25.2%) of chimeras, and 6.0 Mbp (0.7%) of adaptamers were clipped off. The 89,627 scrubbed reads (N_{50}, 9,878 bp; 650.0 Mbp) were assembled in Flye v2.7-b1585 with the "-g 3.8m" setting (8), which produced a 3,747,033-Mbp single circularized contig with 172× coverage. The Unicycler polish function in Unicycler v0.4.9b was used to correct assembly errors and depended on Bowtie v2.3.4.3, Pilon v1.23, and ALE v20180904 (9–12). Sequential rounds of Illumina read polishing were performed until the assembly likelihood score no longer improved and 927 variants were corrected (1 single nucleotide polymorphism [SNP] and 926 single base insertions of C or G). The chromosome was reoriented with *dnaA* being the start of the contig, and the GC content was calculated as 68.8% using Biopython v1.74 (13). CheckM v1.0.13 suggested that the assembly is 100% complete (14). Annotation was performed with PGAP v4.11, which predicted 3,601 coding sequences, 2 rRNA gene operons, and 46 pseudogenes (15). The relatedness of the strain X0973 genome to *Gordonia* type strains was determined based on average nucleotide identity (ANI) (80.9% ± 5.4%) and digital DNA-DNA hybridization (dDDH) (22.6% [20.3% to 25.0%]) with *Gordonia araii* NBRC 100433^T (GCF_000241265.1) as the closest neighbor (16, 17).

Data availability. The whole-genome sequence of *Gordonia* sp. X0973 has been deposited in the DDBJ/ENA/GenBank database under the accession number CP054691. The version described in this paper is the first version, CP054691.1. The raw sequence data were deposited in the SRA under accession numbers SRR11951410 and SRR11951409.

ACKNOWLEDGMENT

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. Tsukamura M. 1971. Proposal of a new genus, *Gordona*, for slightly acid-fast organisms occurring in sputa of patients with pulmonary disease and in soil. J Gen Microbiol 68:15–26. https://doi.org/10.1099/00221287-68-1-15.
2. Arenskötter M, Bröker D, Steinbüchel A. 2004. Biology of the metabolically diverse genus *Gordonia*. Appl Environ Microbiol 70:3195–3204. https://doi.org/10.1128/AEM.70.6.3195-3204.2004.
3. Sowani H, Kulkarni M, Zinjarde S. 2018. An insight into the ecology, diversity and adaptations of *Gordonia* species. Crit Rev Microbiol 44:393–413. https://doi.org/10.1080/1040841X.2017.1418286.
4. Drzyzga O. 2012. The strengths and weaknesses of *Gordonia*: a review of an emerging genus with increasing biotechnological potential. Crit Rev Microbiol 38:300–316. https://doi.org/10.3109/1040841X.2012.668134.
5. Lasker BA, Moser B, Brown JM. 2011. *Gordonia* p 95–110. *In* Liu D (ed), Molecular detection of human pathogens. CRC Press, Boca Raton, FL.
6. Gulvik CA, Arthur RA, Humrighouse BW, Batra D, Rowe LA, Lasker BA, McQuiston JR. 2019. Complete genome sequence of *Nocardia farcinica* W6977^T obtained by combining Illumina and PacBio reads. Microbiol Resour Announc 8:e01373-18. https://doi.org/10.1128/MRA.01373-18.
7. Myers EW. 2020. DASCRUBBER. https://github.com/thegenemyers.
8. 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.
9. 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.
10. Langmead B, Salzberg S. 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. Clark SC, Egan R, Frazier PI, Wang Z. 2013. ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. Bioinformatics 29:435–443. https://doi.org/10.1093/bioinformatics/bts723.
13. 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.
14. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
15. 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.
16. 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.
17. 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.