Complete Genome Sequence of *Paradevosia shaoguanensis* Type Strain J5-3, Obtained Using Nanopore and Illumina Sequencing Technologies

Yang Wang,[a] Songxue Wang[a]

[a]Academy of National Food and Strategic Reserves Administration, Beijing, People's Republic of China

ABSTRACT The complete genome sequence of *Paradevosia shaoguanensis* J5-3^T (China General Microbiological Culture Collection Center [CGMCC] 1.12430^T) is presented here. The complete genome sequence of *P. shaoguanensis* J5-3^T will provide valuable references for classification and comparative genome analysis.

The genus *Paradevosia*, which was first found in 2015, belongs to the family *Hyphomicrobiaceae* and is a close relative of the genera *Devosia*, *Youhaiella*, and *Pelagibacterium* (1–4). *Paradevosia shaoguanensis* J5-3^T is the type species of the genus *Paradevosia*. However, there has been no genome sequence reported for this microorganism so far. In the present study, we determined the complete genome sequence of *Paradevosia shaoguanensis* J5-3^T to gain a greater understanding of this strain.

P. shaoguanensis strain J5-3^T (i.e., CGMCC 1.12430^T or LMG 27409^T) was purchased from the China General Microbiological Culture Collection Center (CGMCC). The strain was grown aerobically at 30°C in LB for 24 h, and then its total genomic DNA was extracted using the Wizard genomic DNA purification kit (Promega) following the manufacturer's instructions. The whole genome was sequenced at Biomaker Technology, Inc. (Beijing, People's Republic of China) with a combination of the Nanopore PromethION 48 system and the Illumina NovaSeq 6000 platform. For long-read sequencing, library preparation used the ligation sequencing kit 1D (SQK-LSK109) and sequencing used R9.4.1 ONT flow cells on the PromethION 48 device (Oxford Nanopore Technologies [ONT], Oxford, UK). Raw sequence data (fast5 format) were base-called using the Albacore sequencing pipeline version 2.0.2 software, and a total of 130,089 raw reads (N_{50}, 24,483 bp) were generated. Short (<2,000 bp) and/or low-quality (quality scores of <6) reads were filtered using Filtlong software (https://github.com/rrwick/Filtlong). After elimination of adaptor sequences and low-quality reads, the remaining reads (103,940 reads, with an average read length of 16,423 bp and N_{50} value of 24,805 bp) were *de novo* assembled into an initial assembly (4,629,545 bp) with Canu version 1.8 (5). The assembly was circularized using the Minimus2 circularization pipeline (https://github.com/sanger-pathogens/circlator/wiki/Minimus2-circularization-pipeline), and the overlapping sequences were trimmed at the assembly ends (6). For short-read sequencing, a library with an insert size of 350 bp was constructed using a TrueLib DNA library rapid prep kit and sequenced using an Illumina NovaSeq 6000 system. The quality of Illumina reads was assessed using FastQC version 0.11.5, and then these reads (7,881,476 reads) were mapped to the assembly with the Burrows-Wheeler Aligner version 0.7.17 (7) for sequence and assembly error correction with Pilon version 1.2.3 (8). The assembly quality and completeness were assessed using BUSCO version 5.1.2 with the alphaproteobacteria_odb10 database, which showed that the genome sequence is >99.4% complete (9). Default settings were used for all software unless otherwise noted. The genome sequence of *P. shaoguanensis* J5-3^T was assembled into a single circular chromosomal contig of 4,629,545 bp (368.74× coverage) with a mean GC content of 63.91%.

The assembled sequence was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (10, 11). The annotation revealed 4,420 protein-coding genes from 4,450 coding DNA sequences (CDSs) in 4,505 genes, with 2 rRNA operons comprising 5S, 16S, and 23S rRNAs, 50 tRNAs, 4 noncoding RNAs, and 25 pseudogenes.

Data availability. The full genomic sequence of *P. shaoguanensis* J5-3^T has been deposited in NCBI/GenBank under BioProject number PRJNA694805 with GenBank accession number CP068983, BioSample number SAMN17574782, and SRA numbers SRR13528957 (PromethION) and SRR13528956 (Illumina).

REFERENCES

1. Geng S, Pan X-C, Mei R, Wang Y-N, Sun J-Q, Liu X-Y, Tang Y-Q, Wu X-L. 2015. *Paradevosia shaoguanensis* gen. nov., sp. nov., isolated from a coking wastewater. Curr Microbiol 70:110–118. https://doi.org/10.1007/s00284-014-0689-2.
2. Wang YX, Huang FQ, Nogi Y, Pang SJ, Wang PK, Lv J. 2015. *Youhaiella tibetensis* gen. nov., sp. nov., isolated from subsurface sediment. Int J Syst Evol Microbiol 65:2048–2055. https://doi.org/10.1099/ijs.0.000219.
3. Xu XW, Huo YY, Wang CS, Oren A, Cui HL, Vedler E, Wu M. 2011. *Pelagibacterium halotolerans* gen. nov., sp. nov. and *Pelagibacterium luteolum* sp. nov., novel members of the family *Hyphomicrobiaceae*. Int J Syst Evol Microbiol 61:1817–1822. https://doi.org/10.1099/ijs.0.023325-0.
4. Nakagawa Y, Sakane T, Yokota A. 1996. Transfer of "*Pseudomonas riboflavina*" (Foster 1944), a Gram-negative, motile rod with long-chain 3-hydroxy fatty acids, to *Devosia riboflavina* gen. nov., sp. nov., nom. rev. Int J Syst Bacteriol 46:16–22. https://doi.org/10.1099/00207713-46-1-16.
5. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
6. Hunt M, Silva N, De Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.
7. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
8. 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.
9. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
10. 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.
11. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res 49:D1020–D1028. https://doi.org/10.1093/nar/gkaa1105.