

Complete genome sequence of the *Sphingobacterium bambusae* type strain KACC 22910^T

Do-Hyun Kim,¹ Hyorim Choi,^{1,2} Jun Heo¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the whole genome sequence of *Sphingobacterium bambusae* KACC 22910^T. The complete genome consists of a 5.6 Mb circular chromosome with a G + C content of 44.4 % and 4,526 predicted coding genes.

KEYWORDS type strain, KACC, *Sphingobacterium bambusae*

The genus *Sphingobacterium* was first described by Yabuuchi et al. (1) and currently comprises 71 validly published species (<https://lpsn.dsmz.de/genus/sphingobacterium>). Many of these species have been isolated from soil environments. Among them, *S. bambusae* was initially isolated from a bamboo plantation by (2). However, no genomic data for this species have been registered. To address the lack of genomic data, we obtained KACC 22,910^T, which is a type strain of *S. bambusae*, and performed a complete genome analysis to establish a foundational data set for functional genomics research on soil biodiversity.

Here, we report the whole genome sequence of *S. bambusae* KACC 22910^T. The strain was obtained from the Korean Agricultural Culture Collection (KACC). It was cultured on Nutrient Agar (NA) medium (BD Difco, NJ, USA) with pH 6.0 at 28°C for 4 days. According to the manufacturer's protocol, the Qiagen MagAttract HMW DNA kit (Qiagen, Hilden, Germany) was used for genomic DNA extraction, and the same DNA products were used for genome sequence analysis on an Illumina NovaSeq 6000 (Illumina, CA, USA) and PacBio Sequel IIe (Pacific Biosciences, CA, USA), respectively. Genomic DNA libraries were prepared using the TruSeq Nano High Throughput Library Prep Kit (Illumina, CA, USA). Libraries of 7–12 kb templates were generated for sequencing with the PacBio SMRTbell prep kit 3.0. Subsequently, they were analyzed using the Sequel II Bind Kit 3.2 and Int Ctrl 3.2. Sequencing was carried out using Sequel II Sequencing Kit 2.0 and SMRT cell 8M trays. HIFI reads were obtained from the PacBio Sequel IIe system and assembled using the microbial assembly application in Flye v2.8.3 (3), based on the Hierarchical Genome Assembly Process (HGAP) (4). The Illumina raw reads of which 90% of the bases had a phred score of 30 or higher were filtered, and adapter trimming was performed using Trimmomatic 0.38 (5). Then, the assembly that was constructed using long-read data was corrected three times using the trimmed Illumina reads through Pilon v1.21 (6). Only one contig was generated as a result of the assembly with the Flye and successfully circularized with the Circlator v1.5.5 (7). Gene prediction and annotation were conducted by the NCBI Prokaryotic Genome Annotation Pipeline v6.4 (PGAP) (8). All tools were executed with default parameters unless stated otherwise.

The Illumina Novaseq 6000 and PacBio Sequel IIe sequencing generated 12,808,047 read pairs, a total of 25,408,064 reads (2 × 150 bp) and 888,234 long reads (9,015,135,467 bp; 12,352 N₅₀), respectively. A circular chromosome of *S. bambusae* KACC 22,910^T was assembled with a coverage of 200× and G + C content of 44.4% without plasmid. The chromosome is predicted to contain 4,526 total coding sequences, 18 rRNA genes, 65 tRNA genes, three non-coding RNA genes, and 11 pseudogenes.

Editor Leighton Pritchard, University of Strathclyde, Glasgow, United Kingdom

Address correspondence to Jun Heo, bioheojun@korea.kr.

The authors declare no conflict of interest.

Received 20 August 2024

Accepted 9 October 2024

Published 22 October 2024

Copyright © 2024 Kim et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

ACKNOWLEDGMENTS

This study was supported by the National Institute of Agricultural Science, Rural Development Administration, Republic of Korea (PJ017286). Genome sequencing raw data of Illumina and PacBio were supplied by DNALink (Seoul, Korea).

AUTHOR AFFILIATIONS

¹Agricultural Microbiology Division, National Institute of Agricultural Sciences, Rural Development Administration, Wanju-gun, Jeollabuk-do, South Korea

²Division of Biotechnology, Jeonbuk National University, Iksan-si, Jeollabuk-do, South Korea

AUTHOR ORCIDs

Jun Heo  <http://orcid.org/0000-0002-8412-5833>

AUTHOR CONTRIBUTIONS

Do-Hyun Kim, Investigation, Project administration | Hyorim Choi, Investigation, Methodology, Writing – original draft.

DATA AVAILABILITY

The complete genome of *Sphingobacterium bambusae* KACC 22910^T was deposited in NCBI GenBank under the BioProject accession number [PRJNA1036373](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1036373), the BioSample accession number [SAMN38124167](https://www.ncbi.nlm.nih.gov/biosample/SAMN38124167), and the GenBank accession number [CP138332](https://www.ncbi.nlm.nih.gov/genbank/CP138332). The raw Illumina and PacBio reads can access the NCBI's SRA accession number [SRR30105955](https://www.ncbi.nlm.nih.gov/sra/SRR30105955) and [SRR29094106](https://www.ncbi.nlm.nih.gov/sra/SRR29094106), respectively.

REFERENCES

1. Yabuuchi E, Kaneko T, Yano I, Moss CW, Miyoshi N. 1983. *Sphingobacterium* gen. nov., *Sphingobacterium spiritivorum* comb. nov., *Sphingobacterium multivorum* comb. nov., *Sphingobacterium mizutae* sp. nov., and *Flavobacterium indologenes* sp. nov.: Glucose-Nonfermenting Gram-Negative Rods in CDC Groups I1K-2 and I1b. *Int J Syst Bacteriol* 33:580–598. <https://doi.org/10.1099/00207713-33-3-580>
2. Duan S, Liu Z, Feng X, Zheng K, Cheng L. 2009. *Sphingobacterium bambusae* sp. nov., isolated from soil of bamboo plantation. *J Microbiol* 47:693–698. <https://doi.org/10.1007/s12275-009-0296-2>
3. 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>
4. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>
5. 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>
6. 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>
7. Hunt M, Silva ND, 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>
8. 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>