



Complete Genome Sequence of *Bacillus* sp. Strain TM2, Isolated from Ground *Tetragnatha* Spider Extract

 Masahiro Ito,^{a,b,c} Satomu Hasunuma^a

^aFaculty of Life Sciences, Toyo University, Itakura-machi, Gunma, Japan

^bBio-Resilience Research Project, Toyo University, Itakura-machi, Gunma, Japan

^cBio-Nano Electronics Research Centre, Toyo University, Kawagoe, Saitama, Japan

ABSTRACT *Bacillus* sp. strain TM2, an aerobic Gram-positive bacterium, was isolated from a solution of ground *Tetragnatha maxillosa* spider in saline solution in Japan. Here, we report the complete genome sequence of this bacterium, which has a 3.67-Mbp genome containing 3,702 protein-coding sequences, 24 rRNA-coding sequences, and 81 tRNA-coding sequences.

We report the whole-genome sequence of *Bacillus* sp. strain TM2, which was isolated from a solution of ground *Tetragnatha maxillosa* spider isolated in Japan (36.2231N, 139.6347E). The method for isolating strain TM2 from the spider was the same as that reported previously (1). TM2 is a high-concentration-cesium-resistant bacterium that can grow in the presence of 900 mM cesium chloride. This bacterium appeared to be most closely related to *Bacillus altitudinis* 41KF2b^T, based on 16S rRNA gene sequence identity (2).

The method for preparing chromosomal DNA was the same as that reported previously (1). The exact same DNA extraction was used for both Oxford Nanopore Technologies (ONT) and Illumina libraries. For long-read sequencing, a DNA library with barcodes added to a sample using the native barcoding expansion kit (ONT, UK) was prepared using a ligation sequencing kit and sequenced with a GridION sequencer (ONT) and an R9.4.1 flow cell. In the library preparation, the genomic DNA was repaired and terminally modified without shearing of the genome, and then the adapters were ligated. The size selection process used AMPure XP beads (Beckman Coulter, USA) and the ligation sequencing kit to remove genomic DNA of less than 3,000 bp. Raw sequence data were base called using Guppy (v4.0.11+f1071ce) (3). The adapter sequences were removed using Porechop (v0.2.3) (4), and reads of 1,000 bases or less were removed using FilTlong (v0.2.0) (5), which yielded 163,106 high-quality paired-end reads. The average length of the long reads was 13,069 bp, and the total number of bases was 2,289,871,481 bp; the N_{50} value was 3,674,367 bp. For short-read sequencing, a DNA library was prepared using the MGIEasy FS DNA library preparation set (MGI Tech, China) according to the manufacturer's instructions, and the library's quality was confirmed using a Fragment Analyzer system and the double-stranded DNA (dsDNA) 915 reagent kit (Advanced Analytical Technologies Inc., USA). The reaction time for enzyme cleavage was 4 min, and an average fragment length of 442 bp was prepared. Circularized DNA was prepared using the prepared library and the MGIEasy circularization kit according to the manual. DNBSEQ 2 × 200-bp paired-end sequencing was performed using the DNBSEQ-G400 sequencer according to the manufacturer's instructions. The adapter sequences were removed using Cutadapt (v2.7) (6). About 3.5 million read pairs (1.05 Gbp) were sampled from the sequence, from which the adapter sequences were removed using Seqkit (v0.11.0) (7); Sickle (v1.33) (8) was used to remove bases with quality scores of less than 20, and reads with fewer than 127 bases and their paired reads were discarded, yielding 6,265,860 high-quality paired-end reads. The average length of the short reads was 200 bp, and the total number of bases was 2,348,640,800 bp.

Editor Vincent Bruno, University of Maryland School of Medicine

Copyright © 2022 Ito and Hasunuma. 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/).

Address correspondence to Masahiro Ito, masahiro.ito@toyo.jp.

The authors declare no conflict of interest.

Received 10 February 2022

Accepted 27 April 2022

Published 10 May 2022

High-quality short-read and long-read sequence data were assembled under the default conditions of Unicycler (v0.4.7) (9), and the genome was circularized. The results of the coding graph assembled using Bandage (v0.8.1) (10) were confirmed, and the integrity of the assembled genomic data was confirmed using CheckM (v1.1.2) (11). The contamination estimated by CheckM1 was 0%. The coverage calculated from the total bases used for assembly was 1,260 \times . Default parameters were used for all software except where otherwise noted. The final chromosome sequence was 3,674,367 bp (G+C content, 41.4%), with an extrachromosomal plasmid of 7,196 bp (G+C content, 35.9%). Automatic annotation was performed using Annotated at DFAST (12), which predicted 3,702 coding sequences, 24 rRNA genes, and 81 tRNA genes.

Data availability. The GenBank accession number for the whole-genome sequence of *Bacillus* sp. strain TM2 is [AP025262](https://doi.org/10.1128/mra.01110-21), and that for the extrachromosomal plasmid is [AP025263](https://doi.org/10.1128/mra.01110-21). Raw sequencing data were deposited in the SRA under accession number [DRA013015](https://doi.org/10.1128/mra.01110-21).

ACKNOWLEDGMENTS

This work was supported by a grant for the Toyo University Top Priority Research Promotion program and the Toyo University Intellectual Property Practical Application Promotion program (M.I.).

REFERENCES

- Ito M, Hasunuma S. 2022. Complete genome sequence of *Bacillus* sp. strain NC3, isolated from *Trichonephila* spider ground extract. *Microbiol Resour Announc* 11:e01110-21. <https://doi.org/10.1128/mra.01110-21>.
- Shivaji S, Chaturvedi P, Suresh K, Reddy GSN, Dutt CBS, Wainwright M, Narlikar JV, Bhargava PM. 2006. *Bacillus aerius* sp. nov., *Bacillus aerophilus* sp. nov., *Bacillus stratosphericus* sp. nov. and *Bacillus altitudinis* sp. nov., isolated from cryogenic tubes used for collecting air samples from high altitudes. *Int J Syst Evol Microbiol* 56:1465–1473. <https://doi.org/10.1099/ijs.0.64029-0>.
- Wick RR, Judd LM, Holt KE. 2019. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol* 20:129. <https://doi.org/10.1186/s13059-019-1727-y>.
- Wick RR. 2018. Porechop. <https://github.com/rwick/Porechop>.
- Wick RR. 2018. Filtlong. <https://github.com/rwick/Filtlong>.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12. <https://doi.org/10.14806/ej.17.1.200>.
- Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11:e0163962. <https://doi.org/10.1371/journal.pone.0163962>.
- Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33). <https://github.com/najoshi/sickle>.
- 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>.
- 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>.
- 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>.
- Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 34:1037–1039. <https://doi.org/10.1093/bioinformatics/btx713>.