

Complete genome sequence of *Sporosarcina* sp. GTC18466 isolated from a snake in Japan

Masahiro Hayashi,^{1,2,3} Jun Yonetamari,^{2,4,5} Yoshinori Muto,² Kaori Tanaka^{1,2,3,5}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT The genus *Sporosarcina* belongs to the family Caryophanaceae within the class Bacilli. Here, we report the complete genome sequence of a *Sporosarcina* species obtained from a snake skin in Japan. The genome comprised a circular chromosome measuring 3,223,418 bp in length.

KEYWORDS *Sporosarcina*, genomes

Identifying pet animal pathogens through genomic analysis is important for understanding pathogenic mechanisms and zoonotic infection risks (1, 2). *Sporosarcina* species have frequently been isolated from animals and humans (3, 4). We analyzed strain GTC18466, isolated from the surface swab of the skin of a healthy *Python regius* that was kept at a pet shop in Seki City, Gifu, Japan, in 2024. The strain was cultured on tryptic soy agar with sheep blood at 37°C for 48 h under aerobic conditions and was submitted to our laboratory as an unknown strain for further identification. The isolate was cultured and purified twice under the conditions described above. The strain was stored at –80°C after four passages during identification.

The strain was cultured on tryptic soy agar under the same conditions. Colonies were scraped, suspended in TE buffer, and pelleted by centrifugation. Genomic DNA was extracted using the NucleoBond HMW DNA kit (MACHEREY-NAGEL, Japan) according to the manufacturer's instructions.

For identification, the 16S rRNA gene was sequenced after PCR amplification using primers 16S-8F (AGAGTTTGATCMTGGCTCAG) and 16S-1492R (GGYTACCTTGTTACGACTT). BLASTN analysis (5) identified the strain as *Sporosarcina* sp., most closely related to *Sporosarcina koreensis* F73 (Type strain accession no. [DQ073393](#)), with 98.5% similarity using phylogenetic analysis.

The genome was sequenced using PromethION 2i (Oxford Nanopore Technologies [ONT], Oxford, UK) for long-read sequencing and DNBSEQ (MGI Tech Co., Ltd., Shenzhen, China) for short-read sequencing (6–8). The same DNA extract was used for both sequencing processes. A library was constructed with a Native Barcoding Kit (SQK-NBD114-24; ONT) without shearing for long-read sequencing. Small DNA fragments were removed using the Short Read Eliminator XS kit (Pacific Biosciences of California, Inc., USA). Sequencing was performed using a PromethION 2i (ONT) with a FLO-PRO114M flow cell. Base calling was performed with Dorado v.7.2.14 (ultra-high accuracy mode), and raw reads were trimmed and quality-filtered using NanoFilt v.2.7.1 (9) with parameters “-l 1,000 -q 10 –headcrop 50.” For short reads, library construction was done with the MGIEasy FS DNA Library Prep Set (MGI Tech), and 2 × 150 bp paired-end sequencing was performed on the DNBSEQ-G400 platform (MGI Tech). Reads were processed with fastp v.0.20.1 (10) using “-q 30 -n 20 -t 1 -T 1.” Short-read quality was evaluated using fastp (10), and long-read quality was scored using NanoPlot v1.32.1 (0). Short reads (with over 96.8% of bases >Q30) and long reads (mean read quality of 16.0) were assembled using Unicycler v.0.4.8 (11) with default settings. Assembly

Editor Karthik Anantharaman, University of Wisconsin-Madison, Madison, Wisconsin, USA

Address correspondence to Masahiro Hayashi, hayashi.masahiro.w8@f.gifu-u.ac.jp.

The authors declare no conflict of interest.

Received 10 October 2025

Accepted 30 December 2025

Published 20 January 2026

Copyright © 2026 Hayashi et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

TABLE 1 Information regarding the complete genome sequence of a *Sporosarcina* sp. strain isolated from a snake in Japan

	Strain name
Parameter	<i>Sporosarcina</i> sp. GTC18466
DNBSEQ sequencing ^a	
No. of reads	4,751,686
Size (kb)	693,264
Avg coverage (×)	227
DRA accession no.	DRR726413
ONT sequencing ^a	
No. of reads	128,640
Size (kb)	1,378,822
Avg read length (bp)	10,719
Avg coverage (×)	428
N ₅₀	16,091
DRA accession no.	DRR726414
Assembly	
Assembly N50 (bp) ^c	3,223,418
Estimated genome completeness (%) ^c	100.0
Estimated genome contamination (%) ^c	0.0
Genome structure	One chromosome
DDBJ/GenBank accession no.	AP044037
Genome size (bp)	3,223,418
GC content (%)	
(chromosome/plasmid name)	50.5
Average read depth	
	226.2
No. of coding sequences ^b	3,182
Number of rRNAs ^b	27
Number of tRNAs ^b	70
Number of CRISPRs ^b	1

^aDRA, DDBJ Sequence Read Archive.^bDFAST, DDBJ Fast Annotation and Submission Tool.^cDetermined with seqkit v2.9.0.

circularization and rotation were performed automatically in Unicycler, and the assembly was rotated to start with *dnaA* on the forward strand. The contig graph was visualized with Bandage v0.8.1 (12), and the taxonomic integrity was confirmed using Blobtools v1.0 with short reads (13).

Table 1 summarizes the genomic information. Unicycler and Bandage indicated a complete circular genome. Quality and genome statistics were computed using CheckM (v1.2.2) (14) and seqkit (v2.9.0) (15). CheckM confirmed that the genome was 100% complete, with 0% contamination. The DDBJ Fast Annotation and Submission Tool v1.2.0 (16) predicted 3,182 coding sequences, 27 rRNAs, 70 tRNAs, and 1 CRISPR sequence. Species identification using an average nucleotide identity value of the genome sequence with GTDB-tk (v2.3.2) was unsuccessful (17). Furthermore, the taxonomic identification of the strain GTC18466 was performed using the Type Strain Genome Server (18). Genome- and rRNA-based comparisons identified *Sporosarcina* as the most closely related genus; however, no species exceeded the species delineation threshold in the digital DNA–DNA hybridization value.

ACKNOWLEDGMENTS

We thank Kyoko Hatazaki, Akiko Katano, and Ayako Nagasawa for their technical support.

The authors declare that they have no competing financial or non-financial interests.

AUTHOR AFFILIATIONS

¹Institute for Glyco-core Research iGCORE, Gifu University, Gifu, Japan

²Division of Anaerobe Research, Life Science Research Center, Gifu University, Gifu, Japan

³Gifu University Center for Conservation of Microbial Genetic Resource, Gifu, Japan

⁴Division of Clinical Laboratory, Gifu University Hospital, Gifu, Japan

⁵United Graduate School of Drug Discovery and Medical Information Sciences, Gifu University, Gifu, Japan

AUTHOR ORCID*s*

Masahiro Hayashi  <http://orcid.org/0000-0001-6317-9077>

AUTHOR CONTRIBUTIONS

Masahiro Hayashi, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Project administration, Resources, Supervision, Validation, Writing – original draft | Jun Yonetamari, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – review and editing | Yoshinori Muto, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Writing – review and editing | Kaori Tanaka, Project administration, Supervision, Writing – review and editing

DATA AVAILABILITY

The genome sequences of the *Sporosarcina* sp. GTC18466 strain from the snake have been deposited in DDBJ (<https://www.ddbj.nig.ac.jp/index.html>) under the accession number [AP044037](#). Raw sequence data for GTC18466 have been deposited in the DDBJ/Sequence Read Archive under accession numbers [DRR726413](#) and [DRR726414](#).

ETHICS APPROVAL

This study was exempted from institutional ethics committee approval.

REFERENCES

- Occhiogrosso L, Capozza P, Buonavoglia A, Decaro N, Trotta A, Marin C, Corrente M. 2023. Bacterial periodontitis in horses: an epidemiological study in southern Italy. *Animals (Basel)* 13:1814. <https://doi.org/10.3390/ani13111814>
- Wylensek D, Hitch TCA, Riedel T, Afrizal A, Kumar N, Wortmann E, Liu T, Devendran S, Lesker TR, Hernández SB, et al. 2020. A collection of bacterial isolates from the pig intestine reveals functional and taxonomic diversity. *Nat Commun* 11:6389. <https://doi.org/10.1038/s41467-020-19929-w>
- Wolfgang WJ, Coorevits A, Cole JA, De Vos P, Dickinson MC, Hannett GE, Jose R, Nazarian EJ, Schumann P, Van Landschoot A, Wirth SE, Musser KA. 2012. *Sporosarcina newyorkensis* sp. nov. from clinical specimens and raw cow's milk. *Int J Syst Evol Microbiol* 62:322–329. <https://doi.org/10.1099/ijs.0.030080-0>
- Tsuda K, Nagano H, Ando A, Shima J, Ogawa J. 2015. Isolation and characterization of psychrotolerant endospore-forming *Sporosarcina* species associated with minced fish meat (surimi). *Int J Food Microbiol* 199:15–22. <https://doi.org/10.1016/j.jfoodmicro.2014.12.031>
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Sydenham TV, Overballe-Petersen S, Hasman H, Wexler H, Kemp M, Justesen US. 2019. Complete hybrid genome assembly of clinical multidrug-resistant *Bacteroides fragilis* isolates enables comprehensive identification of antimicrobial-resistance genes and plasmids. *Microb Genom* 5:e000312. <https://doi.org/10.1099/mgen.0.000312>
- Hayashi M, Niwa A, Yonetamari J, Muto Y, Yokoyama S, Nakamura M, Yokobori Y, Ogawa M, Ichioka R, Kikuchi R, Okura H, Ogura S, Tetsuka N, Baba H, Tanaka K. 2024. Complete genome sequence of *Pigmentibacter ruber* isolated from a human patient in Japan. *Microbiol Resour Announc* 13:e0013324. <https://doi.org/10.1128/mra.00133-24>
- Hayashi M, Yonetamari J, Muto Y, Tanaka K. 2024. Complete genome sequence of *Peptostreptococcus porci* isolated from porcine endocarditis in Japan. *Microbiol Resour Announc* 13:e0020124. <https://doi.org/10.1128/mra.00201-24>
- De Coster W, D'Hert S, Schultz DT, Cruys M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>
- Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- 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>
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Res* 6:1287. <https://doi.org/10.12688/f1000research.12232.1>
- 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>
- 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>

16. 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>
17. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-TK: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>
18. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>