

Complete genome sequence of *Pediococcus pentosaceus* TOKAI 759m, a promising soy milk yogurt starter isolated from rice in Japan

Tomoyuki Hibi,^{1,2} Takahiro Kawanabe,^{2,3} Yuki Nakashima,^{3,4} Shin Yasuda,^{1,2,3} Masashi Hirano,^{1,2,3} Hideki Kinoshita^{1,2,3,5}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT *Pediococcus pentosaceus* TOKAI 759m is a Gram-positive bacterium isolated from rice in Japan. Our previous study demonstrated that this strain is a promising starter culture for soy milk yogurt due to its various beneficial properties. The strain has been commercialized. In this study, we present the complete genome sequence of this strain.

KEYWORDS complete genome, lactic acid bacteria, *Pediococcus pentosaceus*, soy milk yogurt starter, Illumina, Oxford Nanopore Technologies

Our previous study demonstrated that the soy milk yogurt fermented with *Pediococcus pentosaceus* TOKAI 759m (TK 759m) can modulate the gut microbiota and potentially suppress the production of pro-inflammatory cytokines (1) and neuroinflammation (2) *in vivo*. TK 759m was originally isolated from rice in Japan (3) and has recently been commercialized with a product name “Soypedio”. Briefly, washed rice water was plated on De Man, Rogosa, and Sharpe (MRS) (Difco Laboratories, Detroit, MI, USA) agar at 37°C for 2 days under anaerobic conditions. A single colony was then inoculated in MRS broth and cultured at 37°C for 24 h. After several passages, the strain was stored at –80°C in glycerol.

TK 759m was cultured overnight under the same conditions described above. Following incubation, the bacterial cells were harvested by centrifugation. First, pellets intended for short-read sequencing were sent to MySkin Corporation (Tokyo, Japan). Genomic DNA was then extracted from the pellet using DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer’s protocol. Sequencing libraries were prepared using MiSeq Reagent Kit v3 (Illumina, San Diego, CA, USA), and 300 bp paired-end reads were generated on a MiSeq platform (Illumina). The raw short reads were quality-checked and trimmed using fastp v0.24.0 (4, 5) with the parameters “-q 30 -n 20 -t 1 -T 1.” Before trimming, a total of 3,657,140 reads were obtained, with a combined total sequencing of 611,318,165 bp. Meanwhile, genomic DNA for long-read sequencing was extracted from the pellet using Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA). No size selection of DNA was performed before library preparation. Libraries were prepared using Rapid Sequencing Kit V14 (SQK-RAD114, Oxford Nanopore Technologies [ONT], Oxford, UK) and sequenced on R10.4.1 Flow Cell (FLO-FLG114, ONT) on a MinION Mk1B (ONT) under the control of MinKNOW v.24.11.10 with default settings. The sequence data were then basecalled with Dorado v.7.6.8 (High-accuracy model v.4.3.0, 400 bps), which is built into the MinKNOW software. The raw long reads were assessed using NanoPlot v1.44.1 (6). The read N50 was 17,625 bp, and the total number of raw reads was 18,109 reads. Subsequently, the raw long reads were trimmed using NanoFilt v.2.8.0 (7) with the parameters “-q 10 --headcrop 50 -l 1000”. Trimmed long and short reads were assembled using Unicycler v0.5.1 (8) with default settings. Circularization of each contig was determined using Bandage v0.8.1 (Fig. 1; 9).

Editor Daria Van Tyne, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA

Address correspondence to Hideki Kinoshita, kinoshita@tokai.ac.jp.

Hideki Kinoshita is the manager of Probio Co., Ltd., a startup from Tokai University, which sells TK 759m. Yuki Nakashima holds more than 5% of the shares in the company. The other authors declare no conflicts of interest.

Received 26 June 2025

Accepted 20 September 2025

Published 29 October 2025

Copyright © 2025 Hibi 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/).

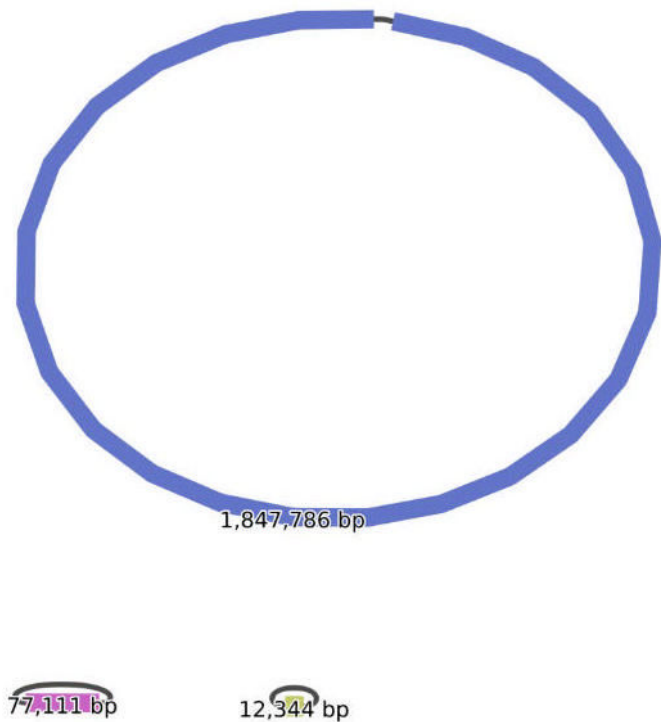


FIG 1 Bandage plot of complete genome assembly.

Trimmed short reads were then mapped to assembled sequences with BWA-MEM2 v2.2.1 (10), indexed using SAMtools v1.21 (11). Finally, the assembled sequences were polished using indexed short reads with Pilon v1.24 (12). This polishing workflow was repeated once.

Completeness check, taxonomy check, and annotation of polished sequences were performed using DDBJ Fast Annotation and Submission Tool (DFAST) v1.3.6 (13). In this annotation, the genome was rotated to start with the *dnaA* gene. Final genome coverage was calculated by mapping the trimmed short and long reads to the annotated FASTA file using minimap2 v2.29 (14, 15) and SAMtools v1.21 (11). Detailed information on the annotated genome is shown in Table 1.

TABLE 1 Accession numbers and sequencing statistics for *P. pentosaceus* TOKAI 759m

Assembly statistics	Value
Total length (bp)	1,937,242
No. of sequences	3
Chromosome mean depth (x)	338.572
Plasmid 1 mean depth (x)	327.538
Plasmid 2 mean depth (x)	767.818
GC content (%)	37.4
No. of CDSs	1916
No. of rRNAs	15
No. of tRNAs	56
No. of CRISPRs	1
Completeness check (%)	93.67
Taxonomy check (%)	98.92
(ANI to <i>Pediococcus pentosaceus</i> NBRC 107768 [GCA_007992275.1])	

ACKNOWLEDGMENTS

This study was supported by funding from the Research Institute of Agriculture at Tokai University, as well as by separate research funding from Tokai University.

AUTHOR AFFILIATIONS

¹Graduate School of Bioscience, Tokai University, Mashiki-machi, Kumamoto, Japan

²Research Institute of Agriculture, Tokai University, Mashiki-machi, Kumamoto, Japan

³Undergraduate School of Agriculture, Tokai University, Mashiki-machi, Kumamoto, Japan

⁴JSPS Research Fellowship for Young Scientists, Tokai University, Mashiki-machi, Kumamoto, Japan

⁵Probio Co. Ltd., Nishihara-mura, Kumamoto, Japan

AUTHOR ORCIDS

Tomoyuki Hibi  <http://orcid.org/0009-0009-9719-5564>

Takahiro Kawanabe  <http://orcid.org/0009-0006-7245-7802>

Hideki Kinoshita  <http://orcid.org/0000-0002-7240-3832>

AUTHOR CONTRIBUTIONS

Tomoyuki Hibi, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Writing – original draft, Writing – review and editing | Takahiro Kawanabe, Conceptualization, Methodology, Software, Supervision, Validation | Yuki Nakashima, Investigation, Methodology, Writing – review and editing | Shin Yasuda, Methodology, Software, Writing – review and editing | Masashi Hirano, Project administration, Supervision, Writing – review and editing | Hideki Kinoshita, Conceptualization, Funding acquisition, Supervision, Writing – review and editing

DATA AVAILABILITY

All data sets have been deposited in DDBJ/ENA/GenBank under the following accession numbers: BioProject [PRJDB35442](https://www.ncbi.nlm.nih.gov/bioproject/PRJDB35442), BioSample [SAMD01046883](https://www.ncbi.nlm.nih.gov/biosample/SAMD01046883), raw short-read data [DRR698283](https://www.ncbi.nlm.nih.gov/sra/DRR698283), raw long-read data [DRR698284](https://www.ncbi.nlm.nih.gov/sra/DRR698284), and assembled and annotated read data [AP041211](https://www.ncbi.nlm.nih.gov/sra/AP041211), [AP041212](https://www.ncbi.nlm.nih.gov/sra/AP041212), and [AP041213](https://www.ncbi.nlm.nih.gov/sra/AP041213).

REFERENCES

- Nakashima Y, Onuki K, Hibi T, Ohno R, Sugawa H, Tominaga Y, Yasuda S, Kinoshita H. 2024. Soymilk yogurt fermented using *Pediococcus pentosaceus* TOKAI 759 m improves mice gut microbiota and reduces pro-inflammatory cytokine production. *Biosci Biotechnol Biochem* 88:1349–1361. <https://doi.org/10.1093/bbb/zbae112>
- Nakashima Y, Hibi T, Urakami M, Hoshino M, Morii T, Sugawa H, Katsuta N, Tominaga Y, Takahashi H, Otomo A, Hadano S, Yasuda S, Hokamura A, Imai S, Kinoshita H. 2025. Soymilk yogurt prepared using *Pediococcus pentosaceus* TOKAI 759m ameliorates cognitive function through gut microbiota modulation in high-fat diet mice. *Curr Res Food Sci* 10:100993. <https://doi.org/10.1016/j.crfs.2025.100993>
- Yamamoto N, Shoji M, Hoshigami H, Watanabe K, Watanabe K, Takatsuzu T, Yasuda S, Igoshi K, Kinoshita H. 2019. Antioxidant capacity of soymilk yogurt and exopolysaccharides produced by lactic acid bacteria. *Biosci Microbiota Food Health* 38:97–104. <https://doi.org/10.12938/bmfh.18-017>
- 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>
- Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* 2:e107. <https://doi.org/10.1002/imt.2.107>
- De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 39:btad311. <https://doi.org/10.1093/bioinformatics/btad311>
- De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>
- 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>
- Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS); Rio de Janeiro, Brazil; p 314–324. <https://doi.org/10.1109/IPDPS.2019.00041>
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
- 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>
- 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>

14. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
15. Li H. 2021. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 37:4572–4574. <https://doi.org/10.1093/bioinformatics/ab705>