



Complete genome sequence of *Streptococcus thermophilus* ST250, a probiotic strain from the NORDBIOTIC collection

Agnieszka K. Szczepankowska,¹ Bożena Cukrowska,² Tamara Aleksandrak-Piekarczyk¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the complete genome sequence of *Streptococcus thermophilus* ST250 from the NORDBIOTIC collection, consisting of a single circular chromosome of 1,829,878 bp.

KEYWORDS *Streptococcus thermophilus*, probiotic strain, complete genome sequence, circular chromosome, hybrid genome assembly, Oxford Nanopore sequencing, Illumina sequencing, food isolate

Streptococcus thermophilus includes strains with probiotic properties (1). Strain ST250 degrades immunoreactive gliadin peptides, a trait not previously reported for other *S. thermophilus* strains (2). It was obtained from the NORDBIOTIC collection (Nordic Biotic Ltd., Warsaw, Poland), described as a cheese isolate, and is also deposited in the DSMZ collection (DSMZ33808). The lyophilized culture was reconstituted in M17 broth (Oxoid) supplemented with 0.5% (wt/vol) glucose (GM17), plated on GM17 agar, and incubated overnight at 37°C under anaerobic conditions.

For DNA extraction, a single colony was inoculated into GM17 broth and grown overnight under the same conditions. Cells were harvested by centrifugation and pretreated with lysozyme (20 mg/mL; Sigma-Aldrich) and mutanolysin (5 U/mL; A&A Biotechnology) for 30 min at 37°C. Genomic DNA was then isolated using a genomic DNA extraction procedure described previously (3), with sodium dodecyl sulfate lysis, proteinase K digestion, and organic extraction. DNA concentration and quality were assessed using a Qubit Fluorometer (Thermo Fisher Scientific), and the sample was subjected to high-quality sequencing using a combined Illumina and Nanopore approach, which enabled accurate base-level correction and resolution of the complete circular chromosome.

Illumina libraries were prepared with the NEBNext Ultra II FS DNA Library Prep Kit (New England Biolabs) and sequenced on a MiSeq instrument (2 × 300 bp, v3 chemistry). Read quality was assessed with FastQC v0.11 and SeqKit v2.8.1, and reads were trimmed and quality filtered with fastp v0.23.4 (4).

For Oxford Nanopore Technologies (ONT) sequencing, high-molecular-weight DNA was sheared using a 26G needle and size-selected to remove fragments of <10 kb using the Short Read Eliminator Kit (Circulomics). A ligation-based library prepared with the Native Barcoding Kit V14 (ONT) was sequenced on a PromethION P2 Solo instrument using an R10.4.1 flow cell. Reads were basecalled with Dorado v7.6.8 (<https://github.com/nanoporetech/dorado>), trimmed with Porechop_ABI v0.5, filtered with Chopper v0.12.0b, and quality-checked using NanoPlot v1.42 (5, 6).

Long reads were assembled and checked for circularity with Autocycler v0.5.2 (7) using Flye v2.9, Canu v2.3, Raven v1.8.1, Miniasm v0.3-r179, metaDBG v1.2, NECAT v0.0.1, and NextDenovo v2.5.2 (8–14). The resulting assembly was polished with Illumina reads using PyPOLCA v0.4 and Polypolish v0.6 (15, 16). Default parameters were used for all software. The final sequence was rotated relative to *dnaA* using dnaapler v1.3.0

Editor David Rasko, University of Maryland Baltimore, Baltimore, Maryland, USA

Address correspondence to Tamara Aleksandrak-Piekarczyk, tamara@ibb.waw.pl.

B.C. is the chief of the Research and Development Department at Nordic Biotic Ltd., has conducted research financed by Danone and Biomed S.A., and has conducted lectures sponsored by Nutricia, Danone, Mead Johnson, Bayer, Aurovitas, Polpharma, and Biomed S.A. The funders had no role in the study design; data collection, analyses, or interpretation; manuscript preparation; or the decision to publish. The other authors report no disclosures.

Received 13 April 2026

Accepted 20 May 2026

Published 11 June 2026

Copyright © 2026 Szczepankowska 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/).

TABLE 1 Statistics of post-processed Illumina and Oxford Nanopore data sets for genome assembly polishing and coverage estimation of *S. thermophilus* ST250^a

Platform	Illumina	ONT
Data type	Read pairs	Reads
Number	4,131,861	70,254
Post-processed bases (bp)	1,229,588,551	320,445,406
N50		6,283

^aIllumina reads were adapter-trimmed and quality-filtered before polishing and coverage estimation. Mean mapped coverage was calculated after alignment to the final genome assembly.

(17). Final assembly quality was evaluated using QUAST v5.0.2 and Bandage v0.9 (18, 19). The final mean mapped coverage of the chromosome was 294×. Illumina and ONT sequencing statistics are in Table 1.

The final assembly consisted of a circular chromosome of 1,829,878 bp with GC content 39.0%. Annotation with the National Center for Biotechnology Information Prokaryotic Genome Annotation Pipeline v6.10 (20) identified 1,889 genes, including 1,574 protein-coding genes, 226 pseudogenes, 67 transfer RNAs, 18 ribosomal RNAs, 4 non-coding RNAs, and 3 CRISPR arrays.

An *in silico* biosafety screen did not detect antimicrobial resistance determinants, plasmid replicons, or virulence factors listed in the Virulence Factor Database. No canonical genes associated with the production of food-relevant biogenic amines were detected.

ACKNOWLEDGMENTS

Sequencing was performed in the DNA Sequencing and Synthesis Facility of the Institute of Biochemistry and Biophysics, Polish Academy of Sciences.

AUTHOR AFFILIATIONS

¹Institute of Biochemistry and Biophysics, Polish Academy of Sciences (IBB PAS), Warsaw, Poland

²Department of Research and Development, Nordic Biotic Ltd., Warsaw, Poland

AUTHOR ORCIDs

Agnieszka K. Szczepankowska  <http://orcid.org/0000-0002-8733-3283>

Tamara Aleksandrak-Piekarczyk  <http://orcid.org/0000-0002-4725-760X>

AUTHOR CONTRIBUTIONS

Agnieszka K. Szczepankowska, Writing – original draft, Writing – review and editing | Bożena Cukrowska, Writing – review and editing | Tamara Aleksandrak-Piekarczyk, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

The complete chromosome sequence has been deposited in GenBank under accession no. [JBWDW010000001.1](https://www.ncbi.nlm.nih.gov/nuccore/JBWDW010000001.1). The associated project and sample are available under BioProject accession no. [PRJNA1437409](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1437409) and BioSample accession no. [SAMN56509525](https://www.ncbi.nlm.nih.gov/biosample/SAMN56509525), respectively. Raw sequencing reads have been deposited in the Sequence Read Archive under accession nos. [SRR37693796](https://www.ncbi.nlm.nih.gov/sra/SRR37693796) (Illumina) and [SRR37693800](https://www.ncbi.nlm.nih.gov/sra/SRR37693800) (Oxford Nanopore Technologies).

REFERENCES

1. Iyer R, Tomar SK, Uma Maheswari T, Singh R. 2010. *Streptococcus thermophilus* strains: multifunctional lactic acid bacteria. *Int Dairy J* 20:133–141. <https://doi.org/10.1016/j.idairyj.2009.10.005>
2. Leszczyńska J, Szczepankowska AK, Majak I, Mańkowska D, Smolińska B, Ścieszka S, Diowski A, Cukrowska B, Aleksandrak-Piekarczyk T. 2024. Reducing immunoreactivity of gluten peptides by probiotic lactic ACID bacteria for dietary management of gluten-related diseases. *Nutrients* 16:976. <https://doi.org/10.3390/nu16070976>
3. Vasileuskaya A, Szczepankowska AK, Gawor J, Chmielewska-Jeznach M, Aleksandrak-Piekarczyk T. 2025. Draft genome sequence of *Lactococcus lactis* IBB2, a candidate probiotic strain isolated from bovine milk. *Microbiol Resour Announc* 14:e0084625. <https://doi.org/10.1128/mra.00846-25>
4. 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>
5. Bonenfant Q, Noé L, Touzet H. 2023. Porechop_ABI: discovering unknown adapters in Oxford Nanopore Technology sequencing reads for downstream trimming. *Bioinform Adv* 3:vbac085. <https://doi.org/10.1093/bioadv/vbac085>
6. 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>
7. Wick RR, Howden BP, Stinear TP. 2025. Autocycler: long-read consensus assembly for bacterial genomes. *Bioinformatics* 41:btaf474. <https://doi.org/10.1093/bioinformatics/btaf474>
8. 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>
9. 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>
10. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. *Nat Comput Sci* 1:332–336. <https://doi.org/10.1038/s43588-021-00073-4>
11. Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32:2103–2110. <https://doi.org/10.1093/bioinformatics/btw152>
12. Benoit G, Raguideau S, James R, Phillippy AM, Chikhi R, Quince C. 2024. High-quality metagenome assembly from long accurate reads with metaDBG. *Nat Biotechnol* 42:1378–1383. <https://doi.org/10.1038/s41587-023-01983-6>
13. Chen Y, Nie F, Xie SQ, Zheng YF, Dai Q, Bray T, Wang YX, Xing JF, Huang ZJ, Wang DP, et al. 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun* 12:60. <https://doi.org/10.1038/s41467-020-20236-7>
14. Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, Li J, Sandoval JR, Cooper DN, Ye K, et al. 2024. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol* 25:107. <https://doi.org/10.1186/s13059-024-03252-4>
15. Bouras G, Judd LM, Edwards RA, Vreugde S, Stinear TP, Wick RR. 2024. How low can you go? short-read polishing of Oxford Nanopore bacterial genome assemblies. *Microb Genom* 10:001254. <https://doi.org/10.1099/mgen.0.001254>
16. Wick RR, Holt KE. 2022. Polypolish: short-read polishing of long-read bacterial genome assemblies. *PLoS Comput Biol* 18:e1009802. <https://doi.org/10.1371/journal.pcbi.1009802>
17. Bouras G, Grigson SR, Papudeshi B, Mallawaarachchi V, Roach MJ. 2024. Dnaapler: a tool to reorient circular microbial genomes. *JOSS* 9:5968. <https://doi.org/10.21105/joss.05968>
18. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
19. 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>
20. 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>