

# Complete genome sequences of three *Geobacillus stearothermophilus* strains isolated from a Korean hot spring

Dariimaa Ganbat,<sup>1</sup> Jae-Yoon Sung,<sup>2</sup> Seong-Bo Kim,<sup>3</sup> Sondor Ganbat,<sup>1</sup> Mi Hwa Park,<sup>4</sup> Dong-Woo Lee,<sup>2</sup> Sang-Jae Lee<sup>1</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 3.

**ABSTRACT** We present the complete genome sequences of three *Geobacillus stearothermophilus* strains—EF60063, EF60133, and EF60134—isolated from a Korean hot spring. S8 family serine peptidases and M4 family metallopeptidases were identified in all strains. Strain EF60063 harbors three plasmids. These findings advance understanding of the genus *Geobacillus*.

**KEYWORDS** *Geobacillus stearothermophilus*, thermophile, genome, plasmid, hot spring

*Geobacillus* species are recognized for their metabolic versatility and enzymatic capabilities, widely utilized in various industrial and biotechnological applications, and remain promising candidates for research (1–3). We isolated multiple thermophilic strains from Korean hot springs (4), assessed their enzymatic activities, and identified several *Geobacillus* isolates exhibiting proteolytic activity; among these, three strains were selected for whole-genome sequencing (Fig. 1). These strains, EF60063, EF60133, and EF60134, were isolated from a Mungang Hot Spring (26.4–28.3°C, pH 8.9–9.1) water sample that was collected on 18 January 2017 in South Korea (36°53′15.4″ N 127°57′20.9″ E). 1 mL of water was serially diluted in 0.85% saline, and was spread on marine agar plates (BD Difco 2216). Plates were incubated at 60°C for 1 week, and the obtained isolates were cryopreserved in 5% (vol/vol) DMSO at –80°C. Strains were identified by amplifying the 16S rRNA gene using BioFACT S-Taq DNA polymerase with universal primers 27F and 1492R, and taxonomic assignment was confirmed via NCBI BLASTN v2.8.0+ (Table 1).

Following the culture conditions detailed in our previous study (5), strains were grown at 60°C in mLB medium, and genomic DNA (gDNA) was extracted using the Wizard Genomic DNA Purification Kit (Promega).

The Oxford Nanopore Technologies (ONT) MinION platform was used to generate long reads for EF60133. gDNA was sequenced on an ONT MinION (R10.4, FLO-MIN112) using the SQK-LSK112.24 ligation kit with NEBNext Companion Module for library preparation, without DNA shearing and size selection. Base calling was performed with Guppy v6.0.6 and read quality assessed using NanoStat v1.4.0 (6). Adapters auto-trimmed during base calling were verified using FastQC v0.12.0 (7). Simultaneously, Illumina sequencing was performed on a NovaSeq 6000 (paired-end 2 × 150 bp), with libraries prepared using the TruSeq Nano DNA Sample Prep Kit. Reads were trimmed using Trimmomatic v0.39 (8) and assessed with FastQC v0.12.0. The genome was assembled with Flye v2.9.2 (9), and the consensus sequence was polished using bcftools v1.20 (10) (parameters “–minDP 5 –minQ 30”) based on variant calling results.

For EF60063, sequencing was performed on the PacBio Sequel platform using the Sequel Sequencing Kit 3.0, following the standard workflow. The sequencing library was prepared using the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences). DNA was sheared by g-TUBE (Covaris), size-selected (15–20 kb) with AMPure PB beads, and reads were quality-filtered with a Q20 threshold. Additionally, Illumina sequencing was

**Editor** Elinne Becket, California State University San Marcos, San Marcos, California, USA

Address correspondence to Dong-Woo Lee, leehicam@yonsei.ac.kr, or Sang-Jae Lee, sans76@gmail.com.

Dariimaa Ganbat and Jae-Yoon Sung contributed equally to this article. Author order was determined by drawing straws.

The authors declare no conflict of interest.

**Received** 21 August 2025

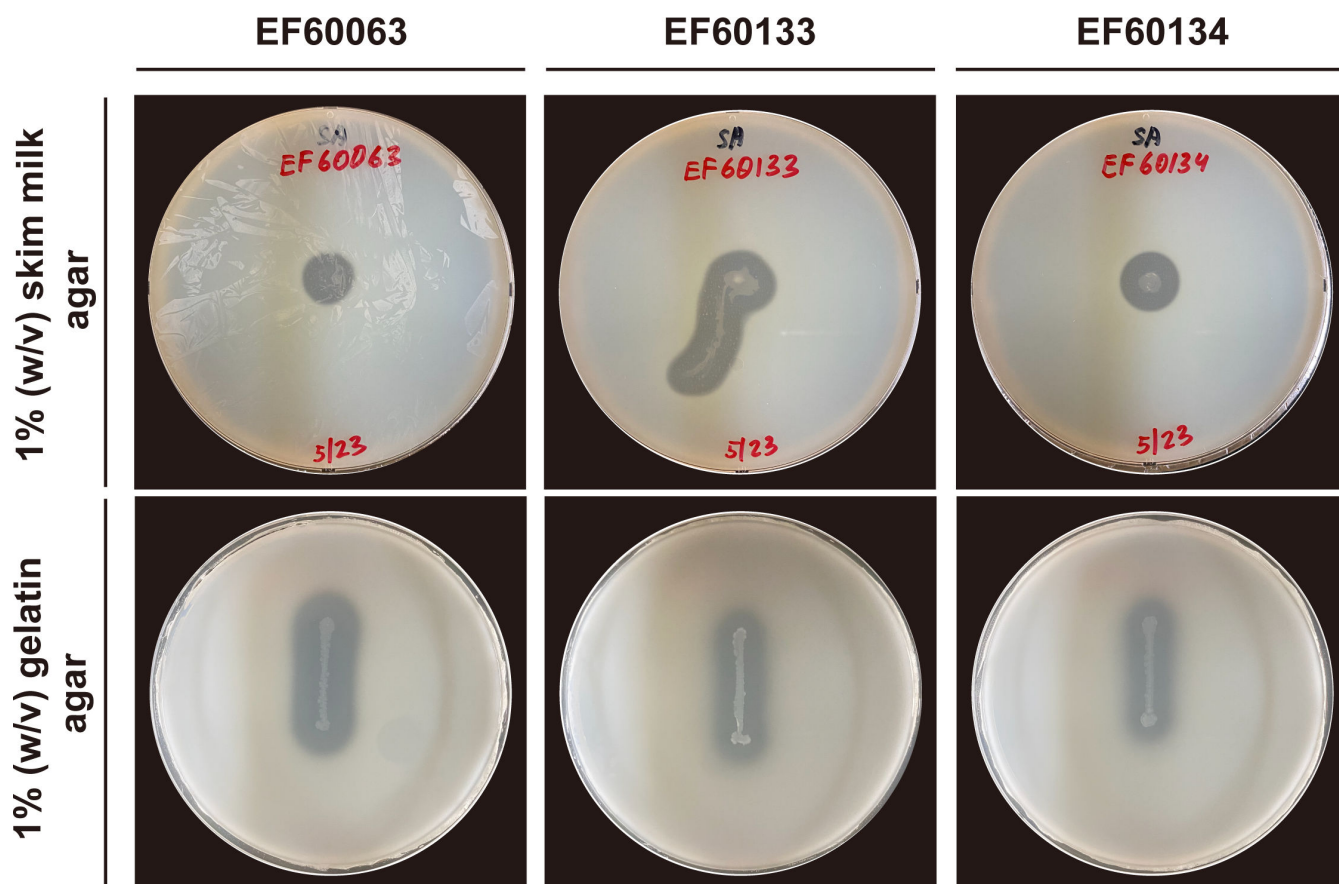
**Accepted** 3 December 2025

**Published** 15 December 2025

Copyright © 2025 Ganbat 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 Summary of quality assessments for bacterial strains sequenced in this study

| Isolate | Closely related species                                    | Identity (%) | BioSample accession number   | SRA accession number (O/P: ONT/PacBio, I: Illumina)              | Total number of raw reads | Read <i>N</i> <sub>50</sub> (bp) | GenBank accession number                                                                                     | No. of contigs | Type of contig                                       | Genome size (bp)                       | GC content (%)               | Coverage (x) | Total genes | Protein-coding genes |
|---------|------------------------------------------------------------|--------------|------------------------------|------------------------------------------------------------------|---------------------------|----------------------------------|--------------------------------------------------------------------------------------------------------------|----------------|------------------------------------------------------|----------------------------------------|------------------------------|--------------|-------------|----------------------|
| EF60063 | <i>Geobacillus stearothermophilus</i> BGSC 9A20 (AY608928) | 99.60        | <a href="#">SAMN35768708</a> | P: <a href="#">SRR35022333</a><br>I: <a href="#">SRR35022332</a> | 212,884<br>10,348,816     | 6,075                            | <a href="#">CP128454</a><br><a href="#">CP128455</a><br><a href="#">CP128456</a><br><a href="#">CP128457</a> | 4              | Chromosome<br>pEF60063-1<br>pEF60063-2<br>pEF60063-3 | 3,435,864<br>65,200<br>37,039<br>2,180 | 52.5<br>40.5<br>43.0<br>48.0 | 251.0        | 3,677       | 3,353                |
| EF60133 | <i>Geobacillus stearothermophilus</i> BGSC 9A20 (AY608928) | 99.87        | <a href="#">SAMN35768736</a> | O: <a href="#">SRR35048013</a><br>I: <a href="#">SRR35030277</a> | 1,091,881<br>14,697,890   | 4,622                            | <a href="#">CP128458</a>                                                                                     | 1              | Chromosome                                           | 3,654,319                              | 52.0                         | 732.0        | 3,714       | 3,426                |
| EF60134 | <i>Geobacillus stearothermophilus</i> BGSC 9A20 (AY608928) | 99.93        | <a href="#">SAMN35768739</a> | P: <a href="#">SRR35022364</a>                                   | 321,844                   | 8,152                            | <a href="#">CP128459</a>                                                                                     | 1              | Chromosome                                           | 3,602,581                              | 52.0                         | 529.9        | 3,744       | 3,409                |



**FIG 1** Extracellular protease activity of *Geobacillus stearothermophilus* strains EF60063, EF60133, and EF60134 on 1% (wt/vol) skim milk agar (upper) and 1% (wt/vol) gelatin agar (lower) plates after incubation at 55°C for 72 h.

performed as described above. The genome was assembled with the HGAP v4.0 (11), followed by polishing to refine the consensus.

Finally, EF60134 was sequenced on the PacBio Sequel platform as described above and assembled using HGAP v4.0.

Circular assemblies were processed with Circlator v1.5.5 and oriented to begin at the *dnaA* gene (12). Assemblies were annotated using the NCBI PGAP v.6.5 (13). S8 family serine peptidases and M4 family metallopeptidases were identified in all strains. Default parameters were used unless noted otherwise.

#### ACKNOWLEDGMENTS

This work was partly supported by the Technology Innovation Program (grant number 20015807) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea); the High Seas Bioresources Program of Korea Institute of the Marine Science and Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (grant number KIMST-20210646); the High Value-Added Food Technology Development Project of the Ministry of Agriculture, Food and Rural Affairs, Republic of Korea (grant number RS-2024-00407212); and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health and Welfare, Republic of Korea (grant number RS-2023-KH141233).

#### AUTHOR AFFILIATIONS

<sup>1</sup>Department of Bioscience and Research Center for Extremophiles and Marine Microbiology, Silla University, Busan, South Korea

<sup>2</sup>Department of Biotechnology, Yonsei University, Seoul, South Korea

<sup>3</sup>Bio-Living Engineering Major, Global Leaders College, Yonsei University, Seoul, South Korea

<sup>4</sup>Department of Food and Nutrition, College of Medical and Life Science, Silla University, Busan, South Korea

## PRESENT ADDRESS

Dariimaa Ganbat, Department of Applied Biosciences, Kyungpook National University, Daegu, South Korea

## AUTHOR ORCID*s*

Dariimaa Ganbat  <http://orcid.org/0000-0003-4244-5465>

Jae-Yoon Sung  <http://orcid.org/0000-0002-2475-8743>

Dong-Woo Lee  <http://orcid.org/0000-0002-2272-8321>

Sang-Jae Lee  <http://orcid.org/0000-0002-9516-3165>

## DATA AVAILABILITY

The WGS projects for the strains EF60063, EF60133, and EF60134 have been deposited in GenBank under BioProject number [PRJNA983507](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA983507). Further details and accession numbers are listed in Table 1.

## REFERENCES

1. Najar IN, Thakur N. 2020. A systematic review of the genera *Geobacillus* and *Parageobacillus*: their evolution, current taxonomic status and major applications. *Microbiology (Reading)* 166:800–816. <https://doi.org/10.1099/mic.0.000945>
2. Kananavičiūtė R, Čitavičius D. 2015. Genetic engineering of *Geobacillus* spp. *J Microbiol Methods* 111:31–39. <https://doi.org/10.1016/j.mimet.2015.02.002>
3. Khaswal A, Chaturvedi N, Mishra SK, Kumar PR, Paul PK. 2022. Current status and applications of genus *Geobacillus* in the production of industrially important products—a review. *Folia Microbiol* 67:389–404. <https://doi.org/10.1007/s12223-022-00961-w>
4. Lee YJ, Ganbat D, Oh D, Kim H, Jeong GE, Cha IT, Kim SB, Nam G, Jung YJ, Lee SJ. 2022. Isolation and characterization of thermophilic bacteria from hot springs in Republic of Korea. *Microorganisms* 10:2375. <https://doi.org/10.3390/microorganisms10122375>
5. Sung JY, Ganbat D, Kim SB, Lee SJ, Lee DW. 2024. Complete genome sequences of *Geobacillus stearothermophilus* strains EF60045 and SJEF4-2 from Korean hot springs. *Microbiol Resour Announc* 13:e0057324. <https://doi.org/10.1128/mra.00573-24>
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. Simon A. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
8. 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>
9. 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>
10. Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27:2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>
11. Chin C-S, 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>
12. 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>
13. 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>