



Draft Genome Sequence of *Geobacillus stearothermophilus* Strain K4E3_SPR_NPP, Isolated from Kasol Hot Spring, Himachal Pradesh, India

Seema Rodge,^a  Niranjan Patil^a

^aDepartment of Microbiology, MES Abasaheb Garware College, Pune, Maharashtra, India

ABSTRACT *Geobacillus stearothermophilus* strain K4E3_SPR_NPP is a thermophilic bacterium growing optimally at 70°C. Here, we present the draft genome sequence of *Geobacillus stearothermophilus* strain K4E3_SPR_NPP, which was isolated from Kasol Hot Spring (72.3°C), Himachal Pradesh, India (32°0'46.35"N, 77°19'0.4908"E).

The organism formerly known as *Bacillus stearothermophilus* was transferred to a new genus, *Geobacillus*, by T. Nazina et al. in the year 2001 (1). Kasol Hot Spring is located in Himachal Pradesh, India (32°0'46.35"N, 77°19'0.4908"E). Water was collected from the hot spring in a sterile screw cap bottle and preserved at 4°C for further processing (2). Thermophilic bacteria were isolated and incubated at 70°C for 24 h using nutrient agar medium.

For genomic DNA extraction, a single colony was selected. DNA was extracted using the DNeasy PowerSoil Pro kit from Qiagen (catalog number 47017). The genomic DNA (gDNA) was sequenced using the Illumina NovaSeq 6000 instrument. A gDNA library was prepared as per the TruSeq DNA Sample Preparation Guide. The fastq reads were trimmed using Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) (3). The trimmed reads were assembled using Unicycler v0.5.0 (<https://github.com/rwick/Unicycler>) (4). The primary assembly was subjected to secondary assembly using CONTIGuator (<https://combo.db.e.unifi.it/contiguator>), with the closest reference genome as the reference background (5). The closest reference, based on a BLAST search of the 16S rRNA sequence similarity, was identified as *Geobacillus stearothermophilus* ATCC 12980 (GenBank accession number [JYNW000000000.1](https://www.ncbi.nlm.nih.gov/nuccore/JYNW000000000.1)). The secondary assembly was subjected to PlasmidFinder v2.1 to find the plasmids, if any (<https://cge.cbs.dtu.dk/services/PlasmidFinder/>) (6, 7). The genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.3 (8). The CRISPRCasFinder Web server was used to detect CRISPR arrays (9). The taxonomic assignment was obtained using the Genome Taxonomy Database (GTDB) toolkit (GTDB-Tk) v1.7.0, release R06-RS202 (<https://gtdb.ecogenomic.org/>) (10–16). The Genome-to-Genome Distance Calculator (GGDC) v3.0 (<https://ggdc.dsmz.de/ggdc.php#>) was used to calculate the digital DNA-DNA hybridization (dDDH) values (17). Default parameters were used for all software unless otherwise noted.

The genome of the thermophilic isolate is 2,993,965 bp long. In all, 19.8 million paired-end reads having a length of 150 bp were generated, with 1,024× genome coverage. The genome was assembled into 95 contigs with an N_{50} value of 2,865,931 bp and 52.67% GC content. One plasmid was found in the secondary assembly. NCBI PGAP was used to identify 3,223 genes, with 2,955 protein-coding sequences and 3,117 coding DNA sequences (CDSs). Furthermore, it was used to identify 86 tRNA sequences, 7 5S rRNA sequences, 1 complete 16S rRNA sequence, 6 23S rRNA sequences, and 5 noncoding RNAs (ncRNAs). A type IIC CRISPR-Cas system was identified, and the genome showed the presence of 5 CRISPR arrays and 43 spacers. The analysis using GTDB-Tk v1.7.0 identified the isolate as *Geobacillus stearothermophilus*, with the

Editor Frank J. Stewart, Montana State University

Copyright © 2022 Rodge and Patil. 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 Niranjan Patil, niranjan75@gmail.com.

The authors declare no conflict of interest.

Received 25 February 2022

Accepted 29 April 2022

Published 19 May 2022

following classification: domain *Bacteria*, phylum *Firmicutes*, class *Bacilli*, order *Bacillales*, family *Anoxybacillaceae*. Using GGDC v3.0, the generalized linear model (GLM)-based DDH estimate was calculated as 74.0% and the distance as 0.0308 with *Geobacillus stearothermophilus* ATCC 12980 (GenBank accession number [JYNW000000001](https://www.ncbi.nlm.nih.gov/nuccore/JYNW000000001)). As the DDH is not greater than 79%, there is a possibility that the isolate belongs to a new subspecies (18).

Data availability. The whole-genome sequencing project for *Geobacillus stearothermophilus* strain K4E3_SPR_NPP has been deposited at DDBJ/ENA/GenBank under the accession number [JAKLOQ000000000](https://www.ncbi.nlm.nih.gov/nuccore/JAKLOQ000000000). The version described in this paper is version [JAKLOQ020000000](https://www.ncbi.nlm.nih.gov/nuccore/JAKLOQ020000000). The raw sequencing reads have been deposited at GenBank under the SRA accession number [SRX14002899](https://www.ncbi.nlm.nih.gov/sra/SRX14002899). The BioSample accession number is [SAMN25211115](https://www.ncbi.nlm.nih.gov/biosample/SAMN25211115), and the BioProject accession number is [PRJNA799873](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA799873).

ACKNOWLEDGMENTS

We thank P. B. Buchade, principal of MES Abasaheb Garware College, Pune, Maharashtra, India, and R. A. Walhe, head of the Research Center, Department of Microbiology, MES Abasaheb Garware College, Pune, Maharashtra, India, for extending support to carry out the research work.

We declare no competing interests.

REFERENCES

- Nazina TN, Tourova TP, Poltarau AB, Novikova EV, Grigoryan AA, Ivanova AE, Lysenko AM, Petrunyaka VV, Osipov GA, Belyaev SS, Ivanov MV. 2001. Taxonomic study of aerobic thermophilic bacilli: descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzonensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, G. th. *Int J Syst Evol Microbiol* 51:433–446. <https://doi.org/10.1099/00207713-51-2-433>.
- Baird RB, Eaton AD, Rice EW (ed). 2017. Standard methods for the examination of water and wastewater, 23rd ed. American Public Health Association, Washington, DC.
- Krueger F. 2015. Trim Galore (RRID: SCR_011847). <https://github.com/FelixKrueger/TrimGalore>.
- 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>.
- Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med* 6:11. <https://doi.org/10.1186/1751-0473-6-11>.
- Carattoli A, Zankari E, García-Fernández A, Voldby Larsen M, Lund O, Villa L, Møller Aarestrup F, Hasman H. 2014. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother* 58:3895–3903. <https://doi.org/10.1128/AAC.02412-14>.
- Clausen PTLC, Aarestrup FM, Lund O. 2018. Rapid and precise alignment of raw reads against redundant databases with KMA. *BMC Bioinformatics* 19:307. <https://doi.org/10.1186/s12859-018-2336-6>.
- 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>.
- Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res* 46:W246–W251. <https://doi.org/10.1093/nar/gky425>.
- Chaumeil PA, 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>.
- Parks DH, Chuvochina M, Chaumeil PA, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 38:1079–1086. <https://doi.org/10.1038/s41587-020-0501-8>.
- Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. <https://doi.org/10.1038/nbt.4229>.
- Matsen FA, Kodner RB, Armbrust EV. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 11:538. <https://doi.org/10.1186/1471-2105-11-538>.
- Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>.
- Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.
- Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M. 2022. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res* 50:D801–D807. <https://doi.org/10.1093/nar/gkab902>.
- Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C, Rohde M, Fartmann B, Goodwin LA, Chertkov O, Reddy T, Pati A, Ivanova NN, Markowitz V, Kyrpides NC, Woyke T, Göker M, Klenk HP. 2014. Complete genome sequence of DSM 30083^T, the type strain (U5/41^T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci* 9:2. <https://doi.org/10.1186/1944-3277-9-2>.