Microbiology Resource Announcements

| Environmental Microbiology | Announcement

Whole genome sequence of the denitrifying thermophile *Geobacillus thermodenitrificans* subsp. *calidus* DSM 22629^T

Mary Snook,[1] Stephan Ritter,[2] Rüdiger Pukall,[3] Markus Göker,[3] Rekha Seshadri,[2] Nathaniel K. Jue[4]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT We present a full genome sequence for the thermophilic denitrifier *Geobacillus thermodenitrificans* subsp. *calidus* DSM 22629^T (3,408,575 bp, 48.94% GC). This genome includes 3,615 predicted genes, including those needed to reduce nitrate to nitrogen gas. This organism and genome sequence provide valuable resources for future phylogenetic, genomic, and biotechnological research.

KEYWORDS denitrification, *Geobacillus thermodenitrificans*, alpha-glucosidase

Geobacillus thermodenitrificans subsp. *calidus* DSM 22629^T is a rod-shaped, thermophilic bacterium with potential traits for biotechnical use. It was isolated by Coleri et al. (1) from high-temperature (98°C) well pipeline sediment in Ankara, Turkey, demonstrating thermophily. This isolate produces thermostable alpha-glucosidase (1), which can be used in food and biofuel production. It has also demonstrated the ability to reduce nitrate (2), an important process in nutrient pollution remediation. These abilities make this strain a candidate for further study of denitrification and heat tolerance.

This type strain was cultured at the Leibniz Institute in nutrient broth (3) at 60°C for 1 day, and genomic DNA was isolated using the MasterPure Gram-positive DNA Extraction Kit (Epicentre). An Illumina DNA library was prepared with the PerkinElmer Sciclone system using a Roche KAPA Biosystems library preparation kit. DNA (200 ng) was sheared to 300 bp using a Covaris LE220, doubly SPRI size-selected, end-repaired, A-tailed, and ligated with sequencing adaptors containing unique barcodes. The library was quantified using KAPA Biosystems' qPCR Kit and run on a Roche LightCycler 480 real-time PCR instrument. The library was multiplexed with other libraries and sequenced on an Illumina NovaSeq 6000 using NovaSeq XP v1 reagent kits (Illumina), S4 flow cell, following a 2 × 150 indexed run protocol. This generated 2,152,122,366 bp of sequence in 14,087,198 150-bp paired-end reads, which were filtered for quality using BBTools version 38.86 per JGI standard operating practice protocol 1061 (4) and subsampled to 10,000,000 reads. These 10,000,000 reads were assembled using SPAdes version 3.13.0 (5) careful mode, with a PHRED quality offset of 33 and an automatically detected conservative coverage cutoff value. This resulted in a draft assembly of 20 contigs in 16 scaffolds, an estimated genome size of 3,408,575 bp, 100% completeness, and 1.1% contamination (assessed using CheckM version 2) (6). The genome was annotated by the IMG/M annotation pipeline (default parameters [7]), using GeneMark.hmm-2 version 1.05 (8) and Prodigal version 2.6.3 (9) to identify coding sequences, tRNAscan-SE version 2.0.3 (10) to predict tRNA genes, and INFERNAL 1.1.2 (11) to identify regulatory mRNAs (Table 1).

Annotations predicted seven genes involved in the KEGG (Kyoto Encyclopedia of Genes and Genomes [12]) nitrate reductase pathway. These included one nitrous oxide-producing nitrate reductase gene (IMG gene ID 2829863217), two nitrate reductase alpha subunit genes (2829860959 and 2829863220), two nitrate reductase gamma subunit genes (2829860962 and 2829863223), one gene encoding nitrous

Editor Frank J. Stewart, Montana State University, Bozeman, Montana, USA

Address correspondence to Nathaniel K. Jue, njue@csumb.edu, or Rekha Seshadri, reseshadri@lbl.gov.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 9 July 2024
Accepted 20 October 2024
Published 12 November 2024

TABLE 1 Genome assembly and annotation statistics for *G. thermodenitrificans* subsp. *calidus* DSM 22629

Statistic	Value
Total scaffold length	3,408,575 bp
Total number of scaffolds	16
Scaffold N_{50}	1,984,393 bp
Average fold coverage	631×
GC content	48.94%
Number of genes	3,615
Number of protein-coding genes	3,466
Number of rRNA genes	13
Number of tRNA genes	80
Other RNA genes	8
Regulatory and miscellaneous features	48
Coding density	86.07%

oxide reductase (2829860982), and one encoding nitric oxide reductase subunit B (2829863208). These genes encompass all necessary steps for nitrate reduction (Fig. 1a) and could be involved in DSM 22629^T's observed ability to do so (2). Annotation also predicted the presence of a gene coding for alpha-glucosidase (gene ID 2829863078). Taxonomic analysis using default parameters on the Type Strain Genome Server (13) showed high support (92.9%) for placement within the clade of other *G. thermodenitrificans* strains (Fig. 1b). Taxonomic placement and the presence of relevant genes provide

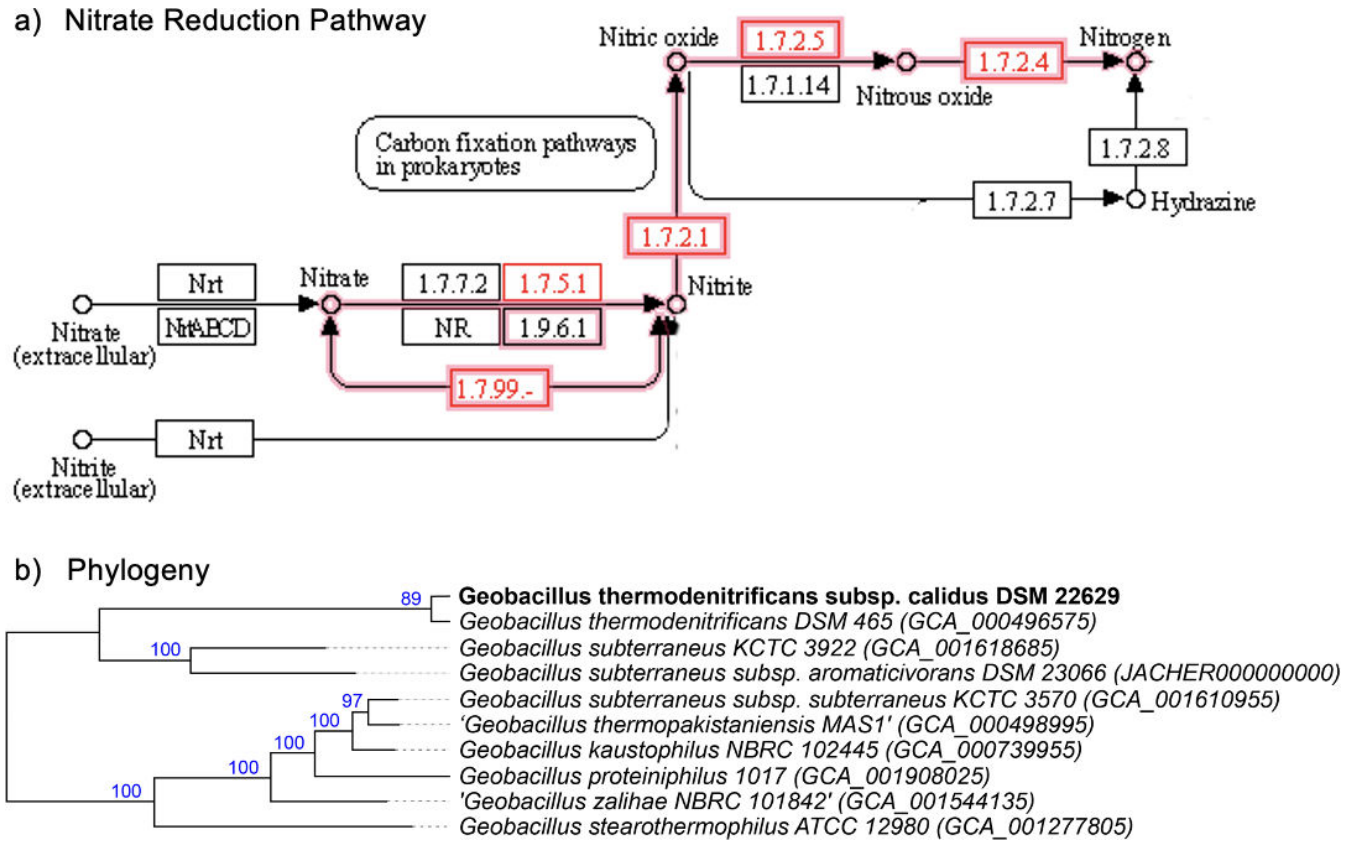

FIG 1 (a) KEGG metabolic pathway for nitrate reduction. Key enzymes are highlighted in pink, and protein domains present in *G. thermodenitrificans* subsp. *calidus* DSM 22629 are shown in red text. (b) Cladogram showing the phylogeny of *G. thermodenitrificans* subsp. *calidus* DSM 22629, created using whole genome comparisons with other type strains in the Type Strain Genome Server. The closest related non-identical taxon is *G. thermodenitrificans* DSM 465, and the closest related taxon of a different species is *Geobacillus subterraneus*. Numbers indicate support for each branch (%).

support for its previous assessments as a denitrifier (2) and its value as a type strain for future study.

ACKNOWLEDGMENTS

This genome was sequenced as a part of the One Thousand Microbial Genomes Phase 4 (KMG IV) project (proposal DOI: https://doi.org/10.46936/10.25585/60001087). The work conducted by the US Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the US Department of Energy operated under contract no. DE-AC02-05CH11231. This announcement was largely prepared by university students as an educational project (the "Adopt-a-genome" Project) supported by the JGI.

The authors thank the Leibniz Institute for culturing and creating cell paste for genomic DNA extraction and the JGI staff for performing library preparation, sequencing, assembly, and annotation.

AUTHOR AFFILIATIONS

[1]Department of Applied Environmental Science, California State University Monterey Bay, Seaside, California, USA
[2]DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[3]Leibniz Institute DSMZ, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany
[4]Department of Biology and Chemistry, California State University Monterey Bay, Seaside, California, USA

AUTHOR ORCIDs

Rekha Seshadri http://orcid.org/0000-0003-3219-2900
Nathaniel K. Jue http://orcid.org/0000-0003-3769-5638

FUNDING

Funder	Grant(s)	Author(s)
U.S. Department of Energy (DOE)	DE-AC02-05CH11231	Rekha Seshadri

AUTHOR CONTRIBUTIONS

Mary Snook, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review and editing | Stephan Ritter, Data curation, Formal analysis, Investigation, Writing – review and editing | Rüdiger Pukall, Conceptualization, Data curation, Investigation, Methodology, Writing – review and editing | Markus Göker, Conceptualization, Data curation, Investigation, Methodology, Writing – review and editing | Rekha Seshadri, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review and editing | Nathaniel K. Jue, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

Single Read Archive whole genome shotgun data (SRX7542789) and reference assembly (GCA_035984515.1) are available in GenBank (PRJNA583236). The annotated genome (IMG ID 2829860437) and comparative genomics tools are available in the Integrated Microbial Genomes with Microbiomes (IMG/M) system (14).

REFERENCES

1. Coleri A, Cokmus C, Ozcan B, Akkoc N, Akcelik M. 2009. Isolation of α-glucosidase-producing thermophilic bacilli from hot springs of Turkey. Mikrobiologiia 78:68–78. https://doi.org/10.1134/S0026261709010081
2. Cihan AC, Ozcan B, Tekin N, Cokmus C. 2011. *Geobacillus thermodenitrificans* subsp. *calidus*, subsp. nov., a thermophilic and α-glucosidase producing bacterium isolated from Kizilcahamam, Turkey. J Gen Appl Microbiol 57:83–92. https://doi.org/10.2323/jgam.57.83
3. Koblitz J. 1: Nutrient agar. Media | MediaDive. Available from: https://mediadive.dsmz.de. Retrieved 23 Sep 2024.
4. Bushnell B, Rood J, Singer E. 2017. BBMerge – accurate paired shotgun read merging via overlap. PLoS ONE 12:e0185056. https://doi.org/10.1371/journal.pone.0185056
5. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, McLean J, Lasken R, Clingenpeel SR, Woyke T, Tesler G, Alekseyev MA, Pevzner PA. 2013. Assembling genomes and mini-metagenomes from highly chimeric reads, p 158–170. In Deng M, Jiang R, Sun F, Zhang X (ed), Research in computational molecular biology. Springer Berlin Heidelberg, Berlin, Heidelberg.
6. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2014. 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
7. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-M, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). Stand Genomic Sci 10:86. https://doi.org/10.1186/s40793-015-0077-y
8. Besemer J. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res 29:2607–2618. https://doi.org/10.1093/nar/29.12.2607
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
10. Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1
11. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29:2933–2935. https://doi.org/10.1093/bioinformatics/btt509
12. Kanehisa M. 2000. Post-genome Informatics. Oxford University Press.
13. 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
14. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, Hajek P, Ritter SJ, Webb C, Wu D, Varghese NJ, Reddy TBK, Mukherjee S, Ovchinnikova G, Nolan M, Seshadri R, Roux S, Visel A, Woyke T, Eloe-Fadrosh EA, Kyrpides NC, Ivanova NN. 2023. The IMG/M data management and analysis system v.7: content updates and new features. Nucleic Acids Res 51:D723–D732. https://doi.org/10.1093/nar/gkac976