



Methylome and Complete Genome Sequence of *Parageobacillus toebii* DSM 14590^T, a Thermophilic Bacterium

E. Anne Hatmaker,^a Kaela B. O'Dell,^{a,b} Lauren A. Riley,^{a,b} Ireneec C. Payne,^a  Adam M. Guss^{a,b}

^aBiosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

^bBredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee—Knoxville, Knoxville, Tennessee, USA

ABSTRACT Here, we present the first complete genome assembly of the thermophilic bacterium *Parageobacillus toebii* DSM 14590^T. The *P. toebii* DSM 14590^T genome consists of a 3,270,071-bp circular chromosome and a 52,989-bp native plasmid.

Parageobacillus is a recently defined genus of Gram-positive, facultative thermophiles; many *Parageobacillus* species were formerly classified as *Geobacillus* species (1). Members of both *Geobacillus* and *Parageobacillus* have potential for biotechnological applications and can produce thermostable enzymes. The aerobic bacterium *Parageobacillus toebii* DSM 14590^T was isolated from hay compost in South Korea (2). The genome sequence and methylome of *P. toebii* DSM 14590^T will inform approaches for genetic engineering of the strain and provide resources for studying *Parageobacillus* species in general.

P. toebii DSM 14590^T was acquired from the German Collection of Microorganisms and Cell Cultures (DSMZ) and was grown in LB liquid medium with 3 g/liter beef extract at 55°C. We used the 100/G Genomic tip extraction kit and bacterial protocol (Qiagen, Valencia, CA, USA) to isolate genomic DNA from *P. toebii* DSM 14590^T. The DNA was not sheared or size selected. Long-read sequencing of *P. toebii* DSM 14590^T was generated at the DOE Joint Genome Institute (JGI). A PacBio SMRTbell library was constructed and sequenced on the PacBio RS II platform (Menlo Park, CA, USA) (3). Sequencing generated 379,771 filtered subreads totaling 769,802,073 bp; PacBio filtering removes reads if the quality score is <0.75 or the length is <50 bp and trims hairpin adapters from sequences, splitting them into subreads. The reads were assembled using the Hierarchical Genome Assembly Process 3 (HGAP3) v2.3.0 (4) with default parameters.

The genome of *P. toebii* DSM 14590^T contains a circular chromosome 3,270,071 bp long and a circular plasmid of 52,989 bp. Long-read sequencing data provided 350× coverage when mapped back to the assembled genome and over 1,800× coverage for the plasmid, suggesting an average plasmid copy number of 5. The genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (5), which predicted 3,375 chromosomal genes, including 3,253 protein-coding, 28 rRNA, and 90 tRNA genes. Additionally, the chromosome is predicted to encode four noncoding RNAs. The *P. toebii* plasmid, pDSM14590, is predicted to encode 61 genes, all of which are considered protein-coding genes.

JGI also performed DNA modification detection and motif analysis using the PacBio SMRT analysis platform v5.0.1.9585 with default parameters. Modified sites were then identified and grouped into motifs using MotifFinder (6), with motifs representing recognition sequences of active methyltransferase genes (7). One methylated motif, garaAtt, was found (100% of occurrences modified). No restriction-modification (RM) systems were predicted within the *P. toebii* DSM 14590^T genome, but all the necessary genes involved in the bacteriophage exclusion (BREX) system were

Citation Hatmaker EA, O'Dell KB, Riley LA, Payne IC, Guss AM. 2020. Methylome and complete genome sequence of *Parageobacillus toebii* DSM 14590^T, a thermophilic bacterium. *Microbiol Resour Announc* 9:e00589-20. <https://doi.org/10.1128/MRA.00589-20>.

Editor Frank J. Stewart, Georgia Institute of Technology

Copyright © 2020 Hatmaker 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/).

Address correspondence to Adam M. Guss, gussam@ornl.gov.

Received 20 May 2020

Accepted 22 May 2020

Published 18 June 2020

predicted (DER53_11070 to DER53_11085, DER53_12455, and DER53_16290). Like RM systems, BREX is a methylation-based bacterial defense system wherein a nonpalindromic sequence is methylated on the genome and infection by unmethylated phage DNA is hindered, though the mechanism is not fully elucidated (8). No prophage regions were predicted within the *P. toebii* DSM 14590^T genome using VirSorter (9) with default parameters.

Data availability. The accession number for the *P. toebii* DSM 14590^T chromosome is CP049703, and the plasmid pDSM14590 is available under the accession number CP049704. The BioProject accession number is PRJNA455457, the BioSample accession number is SAMN09062732, and the reads have been deposited in the NCBI SRA under accession numbers SRX4823098 and SRX4823099.

ACKNOWLEDGMENTS

This manuscript has been authored by UT-Battelle, LLC, under contract number DE-AC05-00OR22725 with the U.S. Department of Energy.

This work was supported by the Center for Bioenergy Innovation, U.S. DOE Bioenergy Research Center, supported by the Office of Biological and Environmental Research in the DOE Office of Science. The Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. DOE under contract DE-AC05-00OR22725. The PacBio DNA sequencing work was conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231. The data were generated for JGI proposal 502982. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Aliyu H, Lebre P, Blom J, Cowan D, De Maayer P. 2016. Phylogenomic re-assessment of the thermophilic genus *Geobacillus*. *Syst Appl Microbiol* 39:527–533. <https://doi.org/10.1016/j.syapm.2016.09.004>.
2. Sung MH, Kim H, Bae JW, Rhee SK, Jeon CO, Kim K, Kim JJ, Hong SP, Lee SG, Yoon JH, Park YH, Baek DH. 2002. *Geobacillus toebii* sp. nov., a novel thermophilic bacterium isolated from hay compost. *Int J Syst Evol Microbiol* 52:2251–2255. <https://doi.org/10.1099/00207713-52-6-2251>.
3. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, DeWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138. <https://doi.org/10.1126/science.1162986>.
4. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
5. 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>.
6. Leung WHP, Tam WC, Chang BCH, Halgamuge SK. 2003. Effects of search pattern variations in motif discovery algorithm: MotifFinder. *IFAC Proc Vol* 33:501–506. [https://doi.org/10.1016/S1474-6670\(17\)33554-1](https://doi.org/10.1016/S1474-6670(17)33554-1).
7. Clark TA, Murray IA, Morgan RD, Kislyuk AO, Spittle KE, Boitano M, Fomenkov A, Roberts RJ, Korlach J. 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. *Nucleic Acids Res* 40:e29. <https://doi.org/10.1093/nar/gkr1146>.
8. Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y, Afik S, Ofir G, Sorek R. 2015. BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J* 34:169–183. <https://doi.org/10.15252/embj.201489455>.
9. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. <https://doi.org/10.7717/peerj.985>.