AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

# Draft Genome Sequence of Hyperthermophilic, Halotolerant *Parageobacillus toebii* PW12, Isolated from the Tattapani Hot Spring, Northwest Himalayas

**Parul Sharma,[a] Sonika Gupta,[a] Anuradha Sourirajan,[a] David J. Baumler,[b,c,d] Kamal Dev[a,b]**

[a]Faculty of Applied Sciences and Biotechnology, Shoolini University, Solan, Himachal Pradesh, India
[b]Department of Food Science and Nutrition, University of Minnesota—Twin Cities, St. Paul, Minnesota, USA
[c]Microbial and Plant Genomic Institute, University of Minnesota—Twin Cities, St. Paul, Minnesota, USA
[d]Biotechnology Institute, University of Minnesota—Twin Cities, St. Paul, Minnesota, USA

**ABSTRACT** Here, we report the genome sequence of hyperthermophilic and halophilic *Parageobacillus toebii* PW12, isolated from the Tattapani hot spring in the northwest Himalayas. The genome size of *Parageobacillus toebii* PW12 is 3,210,377 bp. The G+C content is 42.05%, and 3,382 coding sequences (CDS), 80 tRNAs, 5 noncoding RNAs (ncRNAs), and 4 CRISPR arrays were predicted.

Thermophilic microorganisms belonging to the *Archaea* and *Bacteria* inhabit hot springs, thermal pools, and fumaroles that are found throughout the world (1). Since the discovery of thermophilic microorganisms and novel enzymes, such as *Taq* polymerase from *Thermus aquaticus*, a large number of thermophiles have found industrial applications, including their importance as sources of thermostable enzymes (proteases, amylase, lipase, xylanase, cellulase, and DNA restriction enzymes) and other products of industrial utility (2–4). *Geobacillus* spp. were first classified by Nazina et al. (5) and constitute the most abundant thermophiles isolated from a wide range of environments (6), and they have huge biotechnological potential (7). Recently, polygenomic studies showed that the genus *Geobacillus* consists of two distinct genera, *viz.*, major *Geobacillus* and minor *Parageobacillus* (8). The objective of this study was to explore thermophilic bacteria from the Tattapani hot spring in the Himalayas that can be used for decomposition of cellulosic biomass in different polluted environments.

We isolated *Parageobacillus toebii* strain PW12 from a Tattapani hot spring water sample by spreading the water sample on nutrient agar medium and incubating it at 70°C (9). Isolated colonies were purified by the streaking method. The bacterial strain was identified by amplification and sequencing of 16S rRNA genes. The nucleotide sequence was subjected to a BLASTN (http://blast.ncbi.nlm.nih.gov) search of the NCBI database for taxonomic identity and submitted to NCBI under GenBank accession number KJ509869. The bacterial strain showed optimum cellulase enzyme activity at 80 to 90°C and pH 6 to 8.0, and cellulase activity was tolerant to metal ions ($Mn^{2+}$, $Co^{2+}$, $Fe^{2+}$, $Cd^{2+}$, and $Hg^{2+}$), detergents (SDS, Triton X-100), solvents (toluene, cyclohexane, $H_2O_2$, *n*-butanol, ethanol), and EDTA (9). For genome sequencing, genomic DNA was extracted by a standard method for bacterial DNA isolation (10) and separated on 1% agarose gel by using TAE buffer (40 mM Tris base, 1 mM EDTA, and 20 mM glacial acetic acid) (10) and visualized by using a gel documentation unit (Alpha Innotech). A Qubit 2.0 fluorometer was used for determining DNA concentration. The paired-end (PE) sequencing library was prepared using an Illumina TruSeq Nano DNA high-throughput (HT) library preparation kit, and the PCR-amplified library was analyzed with a Bioanalyzer 2100 (Agilent Technologies) instrument using a high-sensitivity (HS) DNA chip, per the manufacturer's instructions, and loaded onto an Illumina HiSeq 2500 platform for

**Citation** Sharma P, Gupta S, Sourirajan A, Baumler DJ, Dev K. 2019. Draft genome sequence of hyperthermophilic, halotolerant *Parageobacillus toebii* PW12, isolated from the Tattapani hot spring, northwest Himalayas. Microbiol Resour Announc 8:e01163-18. https://doi.org/10.1128/MRA.01163-18.

**Editor** Frank J. Stewart, Georgia Institute of Technology



Address correspondence to David J. Baumler, dbaumler@umn.edu, or Kamal Dev, kamaldevbhardwaj1969@gmail.com.

**Received** 20 August 2018
**Accepted** 20 December 2018
**Published** 24 January 2019

**TABLE 1** Global statistics of the genome sequence of *Parageobacillus toebii* PW12

| Parameter | Statistic |
|---|---|
| Total sequence length (bp) | 3,238,931 |
| No. of genes (total) | 3,462 |
| No. of CDS (total) | 3,382 |
| No. of genes (coding) | 3,006 |
| No. of CDS (coding) | 3,006 |
| No. of RNA genes | 80 |
| Total no. (type) of rRNAs | 2 (5S), 5 (16S), 5 (23S) |
| No. (type) of complete rRNAs | 2 (5S) |
| No. (type) of partial rRNAs | 5 (16S), 5 (23S) |
| No. of tRNAs | 63 |
| No. of ncRNAs | 5 |
| No. of CRISPR arrays | 4 |
| No. of scaffolds | 232 |
| Scaffold $N_{50}$ (bp) | 26,951 |
| Scaffold $L_{50}$ | 35 |
| No. of contigs | 570 |
| Contig $N_{50}$ (bp) | 16,823 |
| Contig $L_{50}$ | 53 |

cluster generation and sequencing. A total number of 1,707,006 paired-end reads with 512,101,800 bp was obtained using 2 × 150-bp chemistry on the Illumina platform. The *de novo* genome assembly of high-quality PE reads and scaffolding was accomplished using SOAPdenovo v.2 (11), with a genome coverage of 150.0×. In the Illumina sequencing, each base in a read with a Phred score of 30 was assigned a quality score by a Phred-like algorithm (12). The assembled genome sequence of *Parageobacillus toebii* PW12 consists of 3,238,931 bp, arranged into 232 scaffolds. The G+C content was 42.05%. Totals of 3,382 coding sequences (CDS), 80 RNAs, 5 noncoding RNAs (ncRNAs), and 4 CRISPR arrays were predicted using the National Center for Biological Information (NCBI) Prokaryotic Genome Annotation Pipeline and the best-placed reference protein set of GeneMarkS+ (Annotation Software v.4.6), as described (13, 14). Three trinucleotide [$(CGC)_5$, $(ATA)_5$, and $(CTT)_5$] and two dinucleotide [$(TA)_6$ and $(CT)_6$] simple sequence repeats (SSRs) were identified using the MicroSatellite identification tool (MISA), as described (15). Genome statistics are summarized in Table 1.

**Data availability.** This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number QREZ00000000. The version described in this paper is version QREZ01000000. Raw sequence reads are available at NCBI under BioProject accession number PRJNA484024, BioSample accession number SAMN09756839 (Tattapani hot spring [Himachal Pradesh] isolate), and SRA accession numbers SRX5163167 (*Parageobacillus toebii* PW12 experiment) and SRR8352206 (Parageobacillus toebii_PW12_R1.fq.gz run).

## ACKNOWLEDGMENTS

We acknowledge Prem Kumar Khosla, Vice Chancellor, and the members of the Yeast Biology Laboratory (YBL), Shoolini University, for supporting this project.

This work was funded by Shoolini University (to K.D.) to promote research under the Center for Omics and Biodiversity Research.

## REFERENCES

1. Waring GA. 1965. Thermal springs of the United States and other countries of the world: a summary. Geological Survey Professional paper 492, revised by R. R. Blankenship and R. Bentall. U.S. Government Printing Office, Washington, DC. https://pubs.usgs.gov/pp/0492/report.pdf.
2. Bergquist PC, Morgan HW. 1992. The molecular genetics and biotechnological application of enzyme from extremely thermophilic eubacteria, pp. 44–45. *In* Herbert RA, Sharp RJ (ed), Molecular biology and biotechnology of extremophiles. Chapman and Hall, New York, NY.
3. Lama L, Calandrelli V, Gambacorta A, Nicolaus B. 2004. Purification and characterization of thermostable xylanase and beta-xylosidase by the thermophilic bacterium *Bacillus thermantarcticus*. Res Microbiol 155: 283–289. https://doi.org/10.1016/j.resmic.2004.02.001.

4. Schallmey M, Singh A, Ward OP. 2004. Developments in the use of *Bacillus* species for industrial production. Can J Microbiol 50:1–17. https://doi.org/10.1139/w03-076.
5. Nazina TN, Tourova TP, Poltaraus 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 uzenensis* sp. nov. from petroleum reservoirs and transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus*, *G. thermocatenulatus*, *G. thermoleovorans*, *G. kaustophilus*, *G. thermoglucosidasius* and *G. thermodenitrificans*. Int J Syst Evol Microbiol 51:433–446. https://doi.org/10.1099/00207713-51-2-433.
6. Daniel DR. 2014. The *Geobacillus* paradox: why is a thermophilic bacterial genus so prevalent on a mesophilic planet? Microbiology 160:1–11. https://doi.org/10.1099/mic.0.071696-0.
7. Hussein AH, Lisowska BK, Leak DJ. 2015. The genus *Geobacillus* and their biotechnological potential. Adv Appl Microbiol 92:1–48. https://doi.org/10.1016/bs.aambs.2015.03.001.
8. 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.
9. Sharma P, Gupta S, Sourirajan A, Dev K. 2015. Characterization of extracellular thermophilic cellulase from *Geobacillus* sp. isolated from Tattapani hot spring of Himachal Pradesh, India. Int J Adv Biotechnol Res 6:433–442.
10. Sambrook J, Fritsch FF, Maniatis T. 1989. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
11. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience 1:18. https://doi.org/10.1186/2047-217X-1-18.
12. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res 8:175–185. https://doi.org/10.1101/gr.8.3.175.
13. 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.
14. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. Nucleic Acids Res 46:D851–D860. https://doi.org/10.1093/nar/gkx1068.
15. Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: a Web server for microsatellite prediction. Bioinformatics 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198.