

Draft Genome Sequence of Lignocellulose-Degrading Thermophilic Bacterium *Geobacillus* sp. Strain WSUCF1

Aditya Bhalla, Amoldeep Singh Kainth, Rajesh K. Sani

Department of Chemical and Biological Engineering, South Dakota School of Mines and Technology, Rapid City, South Dakota, USA

***Geobacillus* sp. strain WSUCF1 is a thermophilic spore-forming member of the phylum *Firmicutes*, isolated from a soil sample collected from the compost facility. We report the draft genome sequence of this isolate with an estimated genome size of 3.4 Mb. The genome sequence of this isolate revealed several genes encoding glycoside hydrolases, making it a potential candidate for plant biomass degradation.**

Received 4 July 2013 Accepted 22 July 2013 Published 15 August 2013

Citation Bhalla A, Kainth AS, Sani RK. 2013. Draft genome sequence of lignocellulose-degrading thermophilic bacterium *Geobacillus* sp. strain WSUCF1. *Genome Announc.* 1(4):e00595-13. doi:10.1128/genomeA.00595-13.

Copyright © 2013 Bhalla et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Rajesh K. Sani, Rajesh.Sani@sdsmt.edu.

Processes utilizing thermophilic lignocellulose-degrading microorganisms have a great potential for the conversion of lignocellulose to biofuels (1, 2). *Geobacillus* sp. strain WSUCF1 (hereafter referred to as WSUCF1) is a thermophilic microbe isolated from a soil sample collected from the compost facility at Washington State University, Pullman, WA, after enrichment on cellulose as a source of carbon and energy. It produces highly thermostable lignocellulose deconstruction enzymes when grown on lignocellulosic substrates, such as corn stover and prairie cord grass (1). In order to reveal the complete gene repertoire of WSUCF1, the genome was sequenced at Integrated Genomics on an Ion Torrent personal genome machine (3). The genome of WSUCF1 is 3,402,383 bp, with a G+C content of 52.21%. The genome was assembled from a total of 2,207,232 sequence reads, with an average read length of 113.65 bp. The reads were assembled into 346 contigs, with an average contig length of about 9,833 bp. The genome contains 4,191 total open reading frames (ORFs) with 47 tRNAs and 3 rRNA operons. The annotations were performed using the ERGO genome analysis suite (4). The results revealed that 17.3% of the total ORFs were assigned for carbohydrate metabolism. Out of 865 ORFs for carbohydrate metabolism, 70 ORFs were found to be involved in polysaccharide degradation. Thirteen predicted open reading frames were annotated as xylan-degrading enzymes and 3 ORFs for cellulose degradation. These results suggest that WSUCF1 holds the potential to be exploited for biofuel production processes. We also identified genes for other industrially important enzymes, such as α -amylases and lipases. The genomic composition of WSUCF1

can be compared to those of other thermophilic and mesophilic organisms to delineate the molecular basis of the thermostability of enzymes.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [ATCO00000000](https://www.ncbi.nlm.nih.gov/nuccore/ATCO00000000). The version described in this paper is version ATCO01000000.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation–Industry/University Cooperative Research Center (NSF-I/UCRC) (grant no. 441087) and the office of the Vice President for Research. We also acknowledge support from the Department of Chemical and Biological Engineering at the South Dakota School of Mines and Technology.

REFERENCES

1. Rastogi G, Bhalla A, Adhikari A, Bischoff KM, Hughes SR, Christopher LP, Sani RK. 2010. Characterization of thermostable cellulases produced by *Bacillus* and *Geobacillus* strains. *Bioresour. Technol.* 101:8798–8806.
2. Bhalla A, Bansal N, Kumar S, Bischoff KM, Sani RK. 2013. Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresour. Technol.* 128:751–759.
3. Vogel U, Szczepanowski R, Claus H, Jünemann S, Prior K, Harmsen D. 2012. Ion torrent personal genome machine sequencing for genomic typing of *Neisseria meningitidis* for rapid determination of multiple layers of typing information. *J. Clin. Microbiol.* 50:1889–1894.
4. Overbeek R, Larsen N, Walunas T, D'Souza M, Pusch G, Selkov E, Liolios K, Joukov V, Kaznadzey D, Anderson I, Bhattacharyya A, Burd H, Gardner W, Hanke P, Kapatral V, Mikhailova N, Vasieva O, Osterman A, Vonstein V, Fonstein M, Ivanova N, Kyrpides N. 2003. The ERGO genome analysis and discovery system. *Nucleic Acids Res.* 31: 164–171.