genomeA
Journals.ASM.org

Complete Genome Sequence of *Geobacillus* sp. Strain GHH01, a Thermophilic Lipase-Secreting Bacterium

Sandra Wiegand,[a] Ulrich Rabausch,[b] Jennifer Chow,[b] Rolf Daniel,[a] Wolfgang R. Streit,[b] Heiko Liesegang[a]

Department of Genomic and Applied Microbiology, Göttingen Genomics Laboratory, Institut für Mikrobiologie und Genetik, Norddeutsches Zentrum für Mikrobielle Genomforschung, Georg-August-Universität, Göttingen, Germany[a]; Abteilung für Mikrobiologie und Biotechnologie, Biozentrum Klein Flottbek, Universität Hamburg, Hamburg, Germany[b]

***Geobacillus* sp. strain GHH01 was isolated during a screening for producers of extracellular thermostable lipases. The completely sequenced and annotated 3.6-Mb genome encodes 3,478 proteins. The strain is genetically equipped to utilize a broad range of different substrates and might develop natural competence.**

Received 8 February 2013 **Accepted** 13 March 2013 **Published** 25 April 2013

Citation Wiegand S, Rabausch U, Chow J, Daniel R, Streit WR, Liesegang H. 2013. Complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase-secreting bacterium. Genome Announc. 1(2):e00092-13. doi:10.1128/genomeA.00092-13.

Address correspondence to Heiko Liesegang, hlieseg@gwdg.de.

The genus *Geobacillus* contains thermophilic strains, which produce a variety of thermostable hydrolytic extracellular enzymes, such as proteases, amylases, and lipases. These features are interesting for future production platforms used in industrial applications (1).

Here, we present the complete genome sequence of *Geobacillus* sp. strain GHH01, a thermophilic lipase producer. The strain was isolated from an enrichment culture originally sampled at Botanischer Garten, University of Hamburg, Germany, and was cultivated at 60°C with 1.5% native olive oil as the sole carbon source. Recombinant expression in *Escherichia coli* revealed that the *Geobacillus* sp. GHH01 lipase (locus tag GHH_c20570) is highly active but only moderately thermostable.

The genome sequence of *Geobacillus* sp. GHH01 was determined by a combined approach of 454 GS-FLX Titanium XL paired-end sequencing (454 Life Sciences, Branford, CT) and Genome Analyzer II single-read sequencing (TruSeq Chemistry, Illumina, San Diego, CA), resulting in average coverages of 10.91-fold and 33.03-fold, respectively. The assembly employing the MIRA v3.4.1.1 software (2) yielded 84 contigs >3 kbp. Gap closure and quality improvement were performed by PCR-based techniques and subsequent Sanger sequencing (ABI 3730xl, Life Technologies, Carlsbad, CA). Initial gene prediction was performed with IMG/ER (3), followed by manual curation based on comparisons to the Swiss-Prot, TrEMBL (4), and InterPro (5) databases. For the identification of rRNA and tRNA genes, RNAmmer v1.2 and tRNAscan-SE v1.4 (6, 7) were used, respectively.

The complete genome consists of a 3,582,992-bp chromosome with a G+C content of 52.3%. In total, 3,597 genes were identified, including 10 rRNA gene clusters and 88 tRNA genes. The annotation resulted in 2,724 protein-encoding genes with assigned functions.

16S rRNA gene phylogenetic analysis confirmed the affiliation of *Geobacillus* sp. GHH01 to the genus *Geobacillus*, whereas an assignment to a described species was not possible. We determined average nucleotide identities (8) of approximately 96% between the *Geobacillus* sp. GHH01 genome and the genomes of *Geobacillus kaustophilus* HTA426 (9) and *Geobacillus thermoleovorans* CCB_US3_UF5 (10). The recently mentioned (10) high synteny between *G. thermoleovorans* CCB_US3_UF5 and *G. kaustophilus* HTA426 (97.94%) calls into question their assignment to distinct species. Hence, a sequence similarity-based assignment of *Geobacillus* sp. GHH01 to a distinct species could not be employed.

Geobacillus sp. GHH01 is predicted to secrete 139 enzymes by the Sec-dependent pathway (11, 12), including the identified lipase, diverse peptidases, proteinases, an amylopullanase (GHH_c32620), an alpha-amylase (GHH_c32630), and an alkaline phosphatase (GHH_c27900). Several substrate-binding proteins of ABC transporters indicate the potential for utilization of a broad range of substrates. The ability to take up extracellular DNA is a crucial mechanism for strain development. Eighteen out of 25 main competence-related structural genes identified for *Bacillus subtilis* (13) were detected, featuring a possible mechanism of DNA uptake.

Genome comparisons revealed seven distinct GHH01-specific genomic islands (14). Furthermore, 123 putative transposases, five clustered regularly interspaced short palindromic repeat (CRISPR) regions, and nine CRISPR-associated genes of subtype III-B (15) could be detected.

Nucleotide sequence accession number. The genome sequence of *Geobacillus* sp. GHH01 has been deposited in GenBank under accession no. CP004008. The strain is available upon request at the Bacillus Genetic Stock Center (BGSC).

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung) within the network grants GenoMik-Plus and GenoMik-Transfer.

We thank Anja Poehlein for support during sequencing, annotation, and submission and Stefanie Offschanka for technical assistance.

REFERENCES

1. **McMullan G, Christie JM, Rahman TJ, Banat IM, Ternan NG, Marchant R.** 2004. Habitat, applications and genomics of the aerobic, thermophilic genus *Geobacillus*. Biochem. Soc. Trans. **32**:214–217.
2. **Chevreux B, Wetter T, Suhai S.** 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer Science and Biology: Proceedings of the German Conference on Bioinformatics. CiteSeer, Pennsylvania State University, University Park, PA.
3. **Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25**:2271–2278.
4. **UniProt Consortium.** 2010. The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res. **37**:D169–D174.
5. **Zdobnov EM, Apweiler R.** 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics **17**: 847–848.
6. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. **25**: 955–964.
7. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
8. **Richter M, Rosselló-Móra R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. **106**: 19126–19131.
9. **Takami H, Horikoshi K.** 2000. Analysis of the genome of an alkaliphilic *Bacillus* strain from an industrial point of view. Extremophiles **4**:99–108.
10. **Muhd Sakaff MKL, Abdul Rahman AY, Saito JA, Hou S, Alam M.** 2012. Complete genome sequence of the thermophilic bacterium *Geobacillus thermoleovorans* CCB_US3_UF5. J. Bacteriol. **194**:1239.
11. **Petersen TN, Brunak S, Von Heijne G, Nielsen H.** 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods **8**:785–786.
12. **Sonnhammer EL, Von Heijne G, Krogh A.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. **6**:175–182.
13. **Kovács AT, Smits WK, Mirończuk AM, Kuipers OP.** 2009. Ubiquitous late competence genes in *Bacillus* species indicate the presence of functional DNA uptake machineries. Environ. Microbiol. **11**:1911–1922.
14. **Langille MG, Brinkman FS.** 2009. IslandViewer: an integrated interface for computational identification and visualization of genomic islands. Bioinformatics **25**:664–665.
15. **Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, Van der Oost J, Koonin EV.** 2011. Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. **9**:467–477.