

# Complete Genome Sequence of *Geobacillus thermoglucosidans* TNO-09.020, a Thermophilic Sporeformer Associated with a Dairy-Processing Environment

Yu Zhao,<sup>a,b,c</sup> Martien P. Caspers,<sup>a,b</sup> Tjakko Abee,<sup>a,c</sup> Roland J. Siezen,<sup>a,e</sup> and Remco Kort<sup>a,b,d</sup>

TI Food and Nutrition, Wageningen, The Netherlands<sup>a</sup>; Netherlands Bioinformatics Centre, Nijmegen, The Netherlands<sup>b</sup>; TNO, Microbiology and System Biology, Zeist, The Netherlands<sup>c</sup>; Laboratory of Food Microbiology, Wageningen University, Wageningen, The Netherlands<sup>d</sup>; and Molecular Cell Physiology, VU University Amsterdam, Amsterdam, The Netherlands<sup>e</sup>

**Thermophilic spore-forming bacteria are a common cause of contamination in dairy products. We isolated the thermophilic strain *Geobacillus thermoglucosidans* TNO-09.020 from a milk processing plant and report the complete genome of a dairy plant isolate consisting of a single chromosome of 3.75 Mb.**

One of the major problems in the production of dairy concentrates is contamination by heat-resistant spores from the thermophilic bacteria *Anoxybacillus* and *Geobacillus*. In fouling samples of a dairy production plant, we identified the species *Geobacillus thermoglucosidasius*, which has been recently renamed *Geobacillus thermoglucosidans* (4). Our present view on genomic diversity, metabolic diversity, and adaptation of *G. thermoglucosidans* is limited, as only the genomes of the strains *G. thermoglucosidasius* C56-YS93, which was derived from petroleum reservoirs (GenBank accession number CP002835.1), and *Geobacillus* sp. Y4.1MC1 (GenBank accession number CP002293.1), which was isolated from Bath Hot Spring in Yellowstone National Park, have been reported. The isolated strain *G. thermoglucosidans* TNO-09.020 showed significant planktonic growth and strong biofilm-forming capacity at high temperatures (60°C to 70°C) under lab conditions (data not shown). This finding allows studies on the development of biofilm formation with the purpose of controlling the contamination of spores in end products. Therefore, we sequenced the genome of *G. thermoglucosidans* TNO-09.020.

Sequencing was performed by GATC Biotech (Konstanz, Germany) using 454 Life Sciences GS20 pyrosequencing (1 M Titanium; 454-Life Sciences, Roche). This sequence produced 457,270 reads totaling 157 million base pairs (Mb). In addition, Illumina paired-end sequencing was also conducted. Assembly into contigs was performed using a whole-genome shotgun (WGS) assembler and SSPACE (1), which is a stand-alone program for scaffolding preassembled contigs using paired-end read data. A pseudoassembly was created by mapping the scaffolds to the most similar complete genome sequence, *Geobacillus* sp. Y4.1MC1 (GenBank accession number CP002293.1). Specific gap filling was performed by using traditional Sanger sequencing of PCR products. Automatic open reading frame (ORF) calling and annotation were performed using Integrated Services for Genomic Analysis (ISGA). Manual curation of ORFs was conducted using Artemis (2) and ACT (Artemis Comparing Tool) (3) by comparison with the reference *Geobacillus* genomes. Improved manual annotation was performed using Pfam (5), InterProScan (6), and NCBI BLAST.

The genome of *G. thermoglucosidans* TNO-09.020 consists of a single chromosome (3.75Mb; 43.9% GC content) in 7 scaffolds containing 32 contigs; no plasmid is found. The chromosome

contains approximately 4,300 protein-encoding genes, 90 tRNA-encoding genes, and 10 rRNA-encoding operons. As this is the first reported genome of a thermophilic isolate from a dairy processing environment, a more detailed analysis of this genome and a comparative analysis with other thermophilic isolates will provide further insight into the specific properties related to the adaptation of *G. thermoglucosidans* TNO-09.020 to this processing environment.

**Nucleotide sequence accession number.** The complete genome of *G. thermoglucosidans* TNO-09.020 has been deposited in GenBank under accession number AJJN00000000.

## ACKNOWLEDGMENTS

We thank Bernadet Renckens and Jos Boekhorst for help with sequence data analysis. We gratefully acknowledge Roy Moezelaar and Masja Nierop Groot for coordination of the project.

The project was funded by the Top Institute Food and Nutrition, Wageningen, The Netherlands.

## REFERENCES

1. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579.
2. Carver T, et al. 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24:2672–2676.
3. Carver TJ, et al. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* 21:3422–3423.
4. Coorevits A, et al. 2011. Taxonomic revision of the genus *Geobacillus*: emendation of *Geobacillus*, *G. stearothermophilus*, *G. jurassicus*, *G. toebii*, *G. thermodenitrificans* and *G. thermoglucosidans* (nom. corr., formerly ‘*thermoglucosidasius*’); transfer of *Bacillus thermantarcticus* to the genus as *G. thermantarcticus*; proposal of *Caldibacillus debilis* gen. nov., comb. nov.; transfer of *G. tepidamans* to *Anoxybacillus* as *A. tepidamans* and proposal of *Anoxybacillus caldiproteolyticus* sp. nov. *Int. J. Syst. Evol. Microbiol.* doi:10.1099/ijs.0.030346-0.
5. Punta M, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301.
6. Quevillon E, et al. 2005. InterProScan: protein domains identifier. *Nucleic Acids Res.* 33:W116–W120.

Received 5 March 2012 Accepted 16 May 2012

Address correspondence to Remco Kort, remco.kort@tno.nl.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.00318-12