Complete Genome Sequence of the Thermophilic Bacterium *Geobacillus thermoleovorans* CCB_US3_UF5

Muhd Khairul Luqman Muhd Sakaff,[a] Ahmad Yamin Abdul Rahman,[a] Jennifer A. Saito,[a] Shaobin Hou,[b] and Maqsudul Alam[a]

Centre for Chemical Biology, Universiti Sains Malaysia, Penang, Malaysia,[a] and Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, Honolulu, Hawaii, USA[b]

Geobacillus thermoleovorans CCB_US3_UF5 is a thermophilic bacterium isolated from a hot spring in Malaysia. Here, we report the complete genome of *G. thermoleovorans* CCB_US3_UF5, which shows high similarity to the genome of *Geobacillus kaustophilus* HTA 426 in terms of synteny and orthologous genes.

Geobacillus spp. are rod-shaped, Gram-positive, thermophilic bacteria (1, 12, 13). These bacteria have drawn interest for their potential in biotechnology applications as sources of thermostable enzymes (5, 11). *Geobacillus* spp. can be found in various terrestrial and marine environments, ranging from hot geothermal locations to cold regions on Earth (5, 11). Here, we report the full genome sequence of *Geobacillus thermoleovorans* CCB_US3_UF5, isolated from Ulu Slim hot spring, Malaysia (92.4°C, pH 7).

The genomic DNA of *G. thermoleovorans* CCB_US3_UF5 was extracted from the bacterium, which was grown in glucose-yeast extract-tryptone (GYT) medium at 60°C using a modified phenol-chloroform protocol (6). The whole-genome sequencing was performed using Roche 454 and Solexa paired-end sequencing technology. A 3-kb genomic library was constructed, and 100,835 paired-end and 63,267 single-end reads were generated using the GS FLX system, giving ~17-fold coverage of the genome. A total of 5,559,066 reads (3-kb library) were generated to reach a depth of ~120-fold coverage with an Illumina Solexa GA IIx (Illumina, San Diego, CA) and mapped to the scaffolds using the Burrows-Wheeler alignment (BWA) tool (8).

The complete genome of *G. thermoleovorans* CCB_US3_UF5 contains a circular chromosome of 3,596,620 bp, with a mean GC content of 52.3%. There are 4,002 predicted genes, with 3,887 coding sequences (CDS), 27 rRNA genes (9 operons), and 88 tRNA genes for 20 amino acids. Automated annotation was performed using the internal DIYA (Do It Yourself Annotator) annotation pipeline (14), which utilizes several analysis tools, such as Glimmer (4), RNAmmer (7), tRNAscan-SE (9), BLAST (2) against UniRef (15), RPS BLAST against CDD (10), and Asgard (3). Genome analysis was done using CLC Genomics Workbench 4.8 (CLC Bio, Aarhus, Denmark), Pathway Studio 8.1 (Ariadne Genomics, MD), and other bioinformatics software.

The genome of *G. thermoleovorans* CCB_US3_UF5 shows extensive similarity to that of *G. kaustophilus* HTA 426 in terms of synteny and orthologous genes.

Nucleotide sequence accession number. The annotated genome sequence of *G. thermoleovorans* CCB_US3_UF5 has been deposited in GenBank under accession number CP003125.

ACKNOWLEDGMENT

REFERENCES

1. **Abd Rahman RN, Leow TC, Salleh AB, Basri M.** 2007. *Geobacillus zalihae* sp. nov., a thermophilic lipolytic bacterium isolated from palm oil mill effluent in Malaysia. BMC Microbiol. **7**:77.
2. **Altschul SF, et al.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.
3. **Alves JM, Buck GA.** 2007. Automated system for gene annotation and metabolic pathway reconstruction using general sequence databases. Chem. Biodivers. **4**:2593–2602.
4. **Delcher AL, Bratke KA, Powers EC, Salzberg SL.** 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**:673–679.
5. **Feng L, et al.** 2007. Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. Proc. Natl. Acad. Sci. U. S. A. **104**:5602–5607.
6. **Fulton J, Douglas T, Young M.** 2009. Isolation of viruses from high temperature environments. Methods Mol. Biol. **501**:43–54.
7. **Lagesen K, et al.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108.
8. **Li H, Durbin R.** 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754–1760.
9. **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.
10. **Marchler-Bauer A, et al.** 2011. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. **39**:D225–D229.
11. **McMullan G, et al.** 2004. Habitat, applications and genomics of the aerobic, thermophilic genus *Geobacillus*. Biochem. Soc. Trans. **32**:214–217.
12. **Nazina TN, et al.** 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 kautophilus*, *Bacillus thermoglucosidasius* and *Bacillus therdenitrificans* 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.
13. **Ren Y, Strobel G, Sears J, Park M.** 2010. *Geobacillus* sp., a thermophilic soil bacterium producing volatile antibiotics. Microb. Ecol. **60**:130–136.
14. **Stewart AC, Osborne B, Read TD.** 2009. DIYA: a bacterial annotation pipeline for any genomics lab. Bioinformatics **25**:962–963.
15. **Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH.** 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics **23**:1282–1288.