

# Complete Genome Sequence of *Geobacillus thermoglucosidarius* NCIMB 11955, the Progenitor of a Bioethanol Production Strain

Lili Sheng, Ying Zhang, Nigel P. Minton

Clostridia Research Group, BBSRC/EPSRC Synthetic Biology Research Centre (SBRC), School of Life Sciences, Centre for Biomolecular Sciences, University of Nottingham, University Park, Nottingham, United Kingdom

**The industrially important thermophile *Geobacillus thermoglucosidarius* has the potential to produce chemicals and fuels from biomass-derived sugar feedstocks. Here, we present the genome sequence of strain NCIMB 11955, the progenitor of an ethanologenic industrial strain, revealing 11 single-nucleotide polymorphisms and 2 indels compared to strain DSM 2542 and two novel plasmids.**

Received 5 August 2016 Accepted 10 August 2016 Published 29 September 2016

**Citation** Sheng L, Zhang Y, Minton NP. 2016. Complete genome sequence of *Geobacillus thermoglucosidarius* NCIMB 11955, the progenitor of a bioethanol production strain. *Genome Announc* 4(5):e01065-16. doi:10.1128/genomeA.01065-16.

**Copyright** © 2016 Sheng et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Nigel P. Minton, [nigel.minton@nottingham.ac.uk](mailto:nigel.minton@nottingham.ac.uk).

*Geobacillus thermoglucosidarius* is a Gram-positive, thermophilic, spore-forming bacterium that performs a typical mixed-acid fermentation (1). It has biotechnological potential for chemical and fuel generation from biomass-derived feedstocks (2–4), as illustrated by the creation of TM242, which produces bioethanol at yields close to theoretical maxima (5). The genome sequence of strain DSM2542, originally isolated in Japan (6, 7), has already been published (8). However, as phenotypes vary widely within the same strain due to genotypic changes (9, 10), we have sequenced the direct progenitor of TM242, NCIMB 11955.

Genomic DNA was prepared using phenol-chloroform extraction. Paired-end libraries were sequenced using an Illumina MiSeq bench-top sequencer (Deepseq, University of Nottingham, Nottingham, UK), producing 1.66 Gb of data consisting of a total of 801,181,830 nucleotides (3,203,788 reads with an average length of 250.1 bp). Using CLC Genomics Workbench version 7.0.4 (CLCbio), 3,203,678 reads were produced through trimming with a quality score limit of 0.05 and maximum ambiguous nucleotide of 2. The genome sequence was then derived by mapping to DSM2542 (CP012712) (using 80% as the cutoff for single-nucleotide variant calling) as a reference and *de novo* assembly.

The entire reference was covered by 92% of the reads (average coverage of  $189.38 \pm 41.51$ ) with four gaps (699, 253, 66, 321 bp). While the first three were closed through identification of *de novo* assembled contigs, the 321-bp gap was closed through PCR amplification and Sanger sequencing of the missing region. A total of 11 single-nucleotide polymorphisms (8 nonsynonymous) and 2 indels were identified compared to DSM 2542. Affected genes included those encoding an MFS transporter (AOT13\_RS03255, V429A), *rpoE* (AOT13\_RS03625, A106V), a transposase (AOT13\_RS05685, T145M), *YggS* (AOT13\_RS08685, stop: 83R), *CodY* (AOT13\_RS09075, A143V), *CheA* (AOT13\_RS09210, V322D), *RimP* (AOT13\_RS09295, K137 fs), and adenine phosphoribosyltransferase (AOT13\_RS17360, P65T).

BLAST analysis of two unmapped contigs designated pNCI001 (83,925 bp, 43.5% GC) and pNCI002 (47,893 bp, 39.0% GC),

revealed homology to a number of *Geobacillus* plasmids. Plasmid pNCI001 shared homology with pGS18 (AM886060: 55% coverage, 98% identity) and pLW1071 (CP000558: 64% coverage, 98% identity), whereas plasmid pNCI002 was homologous to pGG01 (DQ146476: 23% coverage, 98% identity), pGEOH01 (CP002836: 66% coverage, 99% identity), and pGY4MC101 (CP002294: 65% coverage, 99% identity).

Genome annotation was carried out by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) service, resulting in a total of 3,844 coding sequences (3,708 on the chromosome, 81 on pNCI001, and 55 on pNCI002). The chromosome shared identical features to that of DSM 2542 (8). The plasmids carry mostly hypothetical proteins. Open reading frames responsible for replication of the plasmids were identified as RepB (BCV53\_19210, BCV53\_19380) for pNCI001 and RepA (BCV53\_19650), both of which belong to the Rep3 superfamily, suggesting a rolling-circle mode of replication (11, 12). The mapping coverage (294× and 585×, respectively) suggests that the copy number of pNCI002 is twice that of pNCI001. The latter encodes a number of proteins of predicted metabolic relevance, including acyl-CoA dehydrogenase (BCV53\_19405), flavin reductase (BCV53\_19410), and acetaldehyde dehydrogenase (BCV53\_19435). However, no proline- and hydroxyproline-related metabolic genes were predicted, which is in contrast to pGEOH01 present in *G. thermoglucosidarius* C56-YS93 (13).

**Accession number(s).** The complete genome sequence has been deposited in GenBank under the accession numbers CP016622 (chromosome), CP016623 (pNCI001), and CP016624 (pNCI002).

## FUNDING INFORMATION

This work, including the efforts of Ying Zhang and Nigel P. Minton, was funded by Biotechnology and Biological Sciences Research Council (BBSRC) (BB/K020358/1).

This work was supported by the U.K. Biotechnology and Biological Sciences Research Council (BBSRC), grant no. BB/K020358/1, and is part of

the joint BBSRC and DBT (Department of Biotechnology, India) Sustainable Bioenergy and Biofuels (SuBB) programme grant RICEFUEL.

## REFERENCES

1. Förster AH, Gescher J. 2014. Metabolic engineering of *Escherichia coli* for production of mixed-Acid fermentation end products. *Front Bioeng Biotechnol* 2:16. <http://dx.doi.org/10.3389/fbioe.2014.00016>.
2. Taylor MP, Eley KL, Martin S, Tuffin MI, Burton SG, Cowan DA. 2009. Thermophilic ethanogenesis: future prospects for second-generation bioethanol production. *Trends Biotechnol* 27:398–405. <http://dx.doi.org/10.1016/j.tibtech.2009.03.006>.
3. Lin L, Xu J. 2013. Dissecting and engineering metabolic and regulatory networks of thermophilic bacteria for biofuel production. *Biotechnol Adv* 31:273–287. <http://dx.doi.org/10.1016/j.biotechadv.2013.03.003>.
4. Xiao Z, Wang X, Huang Y, Huo F, Zhu X, Xi L, Lu JR. 2012. Thermophilic fermentation of acetoin and 2,3-butanediol by a novel *Geobacillus* strain. *Biotechnol Biofuels* 5:88. <http://dx.doi.org/10.1186/1754-6834-5-88>.
5. Cripps RE, Eley K, Leak DJ, Rudd B, Taylor M, Todd M, Boakes S, Martin S, Atkinson T. 2009. Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production. *Metab Eng* 11: 398–408. <http://dx.doi.org/10.1016/j.ymben.2009.08.005>.
6. Suzuki Y, Kishigami T, Inoue K, Mizoguchi Y, Eto N, Takagi M, Abe S. 1983. *Bacillus thermoglucosidasius* sp. nov., a new species of obligately thermophilic bacilli. *Syst Appl Microbiol* 4:487–495. [http://dx.doi.org/10.1016/S0723-2020\(83\)80006-X](http://dx.doi.org/10.1016/S0723-2020(83)80006-X).
7. Nazina TN, Tourova TP, Poltarau 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. <http://dx.doi.org/10.1099/00207713-51-2-433>.
8. Chen J, Zhang Z, Zhang C, Yu B. 2015. Genome sequence of *Geobacillus thermoglucosidasius* DSM2542, a platform hosts for biotechnological applications with industrial potential. *J Biotechnol* 216:98–99. <http://dx.doi.org/10.1016/j.jbiotec.2015.10.002>.
9. Zhou J, Wu K, Rao C. 22 April 2016. Evolutionary engineering of *Geobacillus thermoglucosidasius* for improved ethanol production. *Biotechnol Bioeng* [Epub ahead of print.] <http://dx.doi.org/10.1002/bit.25983>.
10. Brown CT, Fishwick LK, Chokshi BM, Cuff MA, Jackson JM, Oglesby T, Rioux AT, Rodriguez E, Stupp GS, Trupp AH, Woollcombe-Clarke JS, Wright TN, Zaragoza WJ, Drew JC, Triplett EW, Nicholson WL. 2011. Whole-genome sequencing and phenotypic analysis of *Bacillus subtilis* mutants following evolution under conditions of relaxed selection for sporulation. *Appl Environ Microbiol* 77:6867–6877. <http://dx.doi.org/10.1128/AEM.05272-11>.
11. Abeles AL. 1986. P1 plasmid replication. *J Biol Chem* 261:3548–3555.
12. Boe L, Gros MF, Te Riele H, Ehrlich SD, Gruss A. 1989. Replication origins of single-stranded-DNA plasmid pUB110. *J Bacteriol* 171: 3366–3372.
13. Brumm P, Land ML, Hauser LJ, Jeffries CD, Chang Y, Mead DA. 2015. Complete genome sequences of *Geobacillus* sp. Y412MC52, a xylan-degrading strain isolated from obsidian hot spring in Yellowstone National Park. *Stand Genomic Sci* 10:81. <http://dx.doi.org/10.1186/s40793-015-0075-0>.