

# Complete Genome Sequence of *Clostridium* sp. Strain BNL1100, a Cellulolytic Mesophile Isolated from Corn Stover

Luen-Luen Li,<sup>a</sup> Safiyh Taghavi,<sup>a,b</sup> Javier A. Izquierdo,<sup>b</sup> and Daniel van der Lelie<sup>a,b</sup>

Biology Department, Brookhaven National Laboratory, Upton, New York, USA,<sup>a</sup> and Center for Agricultural and Environmental Biotechnology, RTI International, Research Triangle Park, North Carolina, USA<sup>b</sup>

**We present the full genome sequence of *Clostridium* sp. strain BNL1100, a Gram-positive, endospore-forming, lignocellulolytic bacterium isolated from a corn stover enrichment culture. The 4,613,747-bp genome of strain BNL1100 contains 4,025 putative protein-coding genes, of which 103 are glycoside hydrolases, the highest detected number in cluster III clostridia.**

The discovery and exploration of novel cellulolytic clostridia are essential for the development of technologies for the direct conversion of lignocellulosic biomass to fuels (6, 7). Here we report whole-genome sequencing of *Clostridium* sp. strain BNL1100, a novel Gram-positive, endospore-forming, lignocellulosic bacterium most closely related to *Clostridium cellulolyticum* and *Clostridium papyrosolvans*. The genome information will be useful for exploring *Clostridium* sp. BNL1100 as a biomass-decomposing microorganism for biofuel production.

*Clostridium* sp. BNL1100 was isolated from a corn stover enrichment culture. Genome sequencing was performed using a combination of Illumina (1) and 454 technologies (8) by the DOE Joint Genome Institute. The initial draft assembly contained 67 contigs in 1 scaffold. All 454 data were assembled with Newbler version 2.3. The Newbler consensus sequences were computationally shredded into 2-kb overlapping shreds. Illumina sequencing data were assembled with Velvet version 1.0.13 (9), and the consensus sequence was computationally shredded into 1.5-kb overlapping shreds. All shreds were integrated with the 454 library read pairs using parallel phrap version SPS-4.24 (High Performance Software, LLC). Consed (2–4) software was used in the finishing process. Illumina data were used to correct potential base errors and increase consensus quality using the software Polisher (A. Lapidus, unpublished data). Possible misassemblies were corrected using gapResolution (C. Han, unpublished data) or Dupfinisher (5) or through subcloning. Gaps between contigs were closed by editing in Consed and by Bubble PCR primer walks. The final assembly is based on 56.8 Mb of 454 draft data which provides an average coverage of 12.3× and 1,980 Mb of Illumina draft data which provides an average coverage of 430.4×.

The complete genome is composed of a circular chromosome of 4,613,747 bp (37.26% GC content), which includes 4,025 coding genes, 61 tRNAs, and 8 rRNA operons. A total of 3,012 coding genes (73.16% of the total) have putative functions assigned on the basis of annotation.

A preliminary analysis of the genome reveals that it contains a total of 103 glycoside hydrolases, the highest detected number in cluster III clostridia. These enzymes belong to 33 glycoside hydrolase families, with the most abundant being GH5, GH9, and GH43 (8, 14, and 13 enzymes, respectively). Despite having a great deal of overlap with the close relatives *C. cellulolyticum* and *C. papyrosolvans*, *Clostridium* sp. BNL1100 is unique in that it is the only

*Clostridium* species to have a xylan  $\alpha$ -1,2-glucuronidase (EC 3.2.1.131) belonging to the GH115 family. The genome sequence also revealed that this organism possesses a cellulosome, with a scaffolding protein with organization similar to that of the scaffolding protein of *C. cellulolyticum* and with 39 glycoside hydrolases containing dockerin domains. In terms of its key metabolic pathways, *Clostridium* sp. BNL1100 possesses glycolytic and fermentative pathways identical to those of *C. cellulolyticum* while possessing a complete pentose phosphate pathway identical to the one found in *C. papyrosolvans*.

**Nucleotide sequence accession number.** The complete genome sequence of *Clostridium* sp. BNL1100 has been deposited in GenBank under accession number CP003259.

## ACKNOWLEDGMENTS

This research was supported by a grant from the BioEnergy Science Center (BESC), Oak Ridge National Laboratory, a U.S. Department of Energy (DOE) Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The *Clostridium* sp. BNL1100 genome sequencing was conducted by the U.S. Department of Energy Joint Genome Institute, which is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231.

## REFERENCES

- Bennett S. 2004. Solexa Ltd. Pharmacogenomics 5:433–438.
- Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8:186–194.
- Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.
- Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res. 8:195–202.
- Han C, Chain P. 2006. Finishing repeat regions automatically with Dupfinisher, p 141–146. In Arabnia HR, Valafar H (ed), Proceedings of the 2006

Received 28 September 2012 Accepted 3 October 2012

Address correspondence to Daniel van der Lelie, vdlelie@rti.org.

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

doi:10.1128/JB.01908-12

- international conference on bioinformatics & computational biology. CSREA Press, Las Vegas, NV.
6. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* **66**:506–577.
  7. Maki M, Leung KT, Qin W. 2009. The prospects of cellulase-producing bacteria for the bioconversion of lignocellulosic biomass. *Int. J. Biol. Sci.* **5**:500–516.
  8. Margulies M, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
  9. Zerbino D, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**:821–829.