Draft Genome Sequence of Butanol-Acetone-Producing *Clostridium beijerinckii* Strain G117

Yi-Rui Wu,^a Yao Li,^a Kun-Lin Yang,^b and Jianzhong He^a

Department of Civil and Environmental Engineering^a and Department of Chemical and Biomolecular Engineering,^b National University of Singapore, Singapore

A recently discovered wild-type strain, *Clostridium beijerinckii* G117, is unique in producing butanol and acetone but negligible amounts of ethanol, unlike previously identified acetone-butanol-ethanol (ABE)-generating microbes. Here we report the draft genome sequence of strain G117 (5,806,675 bp; GC content, 29.7%) and the novel findings obtained from its genome annotations.

The solventogenic *Clostridium* species have attracted renewed attention in the scientific field during the past few years due to the search for sustainable sources of energy (5). Among solventogenic microbes, the Gram-positive anaerobe *Clostridium beijerinckii* NCIMB 8052 is capable of utilizing pentose and hexose sugars to produce acetone, butanol, and ethanol (ABE) without the “glucose repression” effect, which gives it an attractive advantage over another well-known solventogenic organism, *Clostridium acetobutylicum* (10). Recently, a newly discovered isolate, *Clostridium beijerinckii* G117, was distinguished as generating acetone and butanol (AB) but negligible ethanol from fermentation of glucose (1a). The 16S rRNA gene of strain G117 shows 99% identity to that of *C. beijerinckii* NCIMB 8052. The genome sequencing data for *C. beijerinckii* NCIMB 8052 will facilitate our understanding of the “omics” (e.g., genomics, transcriptomics, and metabolomics) of solventogenic microbes (5, 9). To further our understanding of solventogenesis in this organism, we present a draft genome sequence of *C. beijerinckii* G117.

The genome of *C. beijerinckii* G117 was sequenced by a whole-genome shotgun strategy using a high-throughput Illumina HiSeq 2000 at the Beijing Genomics Institute (BGI, Shenzhen, China). A total of 1,766,892 reads with a 296-bp insert size, counting up to 523 Mbp, were obtained, providing 90-fold coverage. Genome sequences were assembled by using the SOAPdenovo program (version 1.05) (6), resulting in 178 contigs with an N_{50} of 95,461 bp and a total length of 5,806,675 bp for the whole genome. These contigs were assembled into 89 scaffolds with a maximum length of 271,884 bp. Putative protein coding sequences were identified by Glimmer (version 3.02) (2) and analyzed by BLASTP. The functions of predicted protein-coding genes were annotated through comparisons with Kyoto Encyclopedia of Genes and Genomes (KEGG) (3), Clusters of Orthologous Groups (COG) (8), and Swiss-Prot (1) databases. tRNA and rRNA were annotated with tRNAscan-SE 1.21 (7) and rRNAmmer 1.2 (4), respectively.

The draft genome includes 5,806,675 bp with a low GC content of 29.7%, and the average nucleotide identity between strain G117 and NCIMB 8052 was determined to be 97.39%. The genome contains 5,262 predicted genes, 5,111 protein-coding sequences (CDSs), 54 tRNA genes, and 20 rRNAs (including 11 5S rRNAs, 3 16S rRNAs, and 6 23S rRNAs). Among the predicted CDSs, 2,687, 2,937, and 2,102 proteins were functionally annotated in the COG, KEGG, and Swiss-Prot databases, respectively. The annotations indicate that a total of 198 CDSs are involved in energy produc-

tion and conversion and 354 CDSs are related to carbohydrate transport and metabolism, as observed from the COG system.

From the genome sequence analysis, 16 alcohol dehydrogenase-related genes are found in the genome of *C. beijerinckii* G117. Among them, one *bdh* gene, encoding butanol dehydrogenase, was found to be responsible for butanol generation (1a), the sequence of which shows 98% and 90% identity with *bdh1* (Cbei2181) and *bdh2* (Cbei1722), respectively, in *C. beijerinckii* NCIMB 8052 (9). Transcriptome analysis indicates that the expression of butanol dehydrogenase gene increases sharply with the production of butyric acid (1a). The other 15 alcohol dehydrogenase-related genes in strain G117 exhibit 95 to 100% similarity to the closest genes in *C. beijerinckii* NCIMB 8052, suggesting the uniqueness of this newly discovered solventogenic *Clostridium* strain.

Nucleotide sequence accession number. The draft sequence of *C. beijerinckii* G117 has been deposited in NCBI GenBank under accession number [AKWA00000000](#).

ACKNOWLEDGMENTS

This research was supported by the Competitive Research Programme from Singapore National Research Foundation under project no. NRF-CRP5-2009-05 and the Singapore Agency for Science, Technology and Research (A*STAR) of the Science and Engineering Research Council under project no. 092 139 0034.

REFERENCES

1. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. 2007. UniProtKB/Swiss-Prot. *Methods Mol. Biol.* 406:89–112.
- 1a. Chua TK, Liang DW, Qi C, Yang KL, He J. Characterization of a butanol-acetone-producing *Clostridium* strain and identification of its solventogenic genes. *Bioresour. Technol.*, in press.
2. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23:673–679.
3. Kanehisa M, et al. 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36:D480–D484.
4. Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of rRNA genes. *Nucleic Acids Res.* 35:3100–3108.

Received 28 June 2012 Accepted 20 July 2012

Address correspondence to Jianzhong He, jianzhong.he@nus.edu.sg.

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

doi:10.1128/JB.01139-12

5. Lee SY, et al. 2008. Fermentative butanol production by clostridia. *Bio-technol. Bioeng.* 101:209–228.
6. Li R, et al. 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265–272.
7. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of tRNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955–964.
8. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28:33–36.
9. Wang Y, Li X, Mao Y, Blaschek HP. 2011. Single-nucleotide resolution analysis of the transcriptome structure of *Clostridium beijerinckii* NCIMB 8052 using RNA-Seq. *BMC Genomics* 12:479.
10. Xiao H, et al. 2012. Metabolic engineering of d-xylose pathway in *Clostridium beijerinckii* to optimize solvent production from xylose mother liquid. *Metab. Eng.* <http://dx.doi.org/10.1016/j.ymben.2012.05.003>.