

## GENOME ANNOUNCEMENTS

### Complete Genome Sequence of *Clostridium acetobutylicum* DSM 1731, a Solvent-Producing Strain with Multireplicon Genome Architecture<sup>∇</sup>

Guanhui Bao,<sup>1,2</sup> Runjiang Wang,<sup>1,2</sup> Yan Zhu,<sup>1,2</sup> Hongjun Dong,<sup>1,2</sup> Shaoming Mao,<sup>1</sup>  
Yanping Zhang,<sup>1</sup> Zugen Chen,<sup>3</sup> Yin Li,<sup>1\*</sup> and Yanhe Ma<sup>1,4</sup>

*Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*<sup>1</sup>; *Graduate School of Chinese Academy of Sciences, Beijing, China*<sup>2</sup>; *Department of Human Genetics, School of Medicine, University of California, Los Angeles, California 90095*<sup>3</sup>; and *State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China*<sup>4</sup>

Received 16 June 2011/Accepted 29 June 2011

***Clostridium acetobutylicum* is an important microorganism for solvent production. We report the complete genome sequence of *C. acetobutylicum* DSM 1731, a genome with multireplicon architecture. Comparison with the sequenced type strain *C. acetobutylicum* ATCC 824, the genome of strain DSM1731 harbors a 1.7-kb insertion and a novel 11.1-kb plasmid, which might have been acquired during evolution.**

Acetone butanol ethanol (ABE) fermentation by *Clostridium acetobutylicum* is one of the oldest industrial fermentations ever known, and it was historically ranked second to ethanol fermentation by yeast in terms of scale of production. *C. acetobutylicum* is a low-GC-content, Gram-positive, spore-forming obligate anaerobe (2). Recent “-omics” studies, including genomics (4), transcriptomics (12, 13), proteomics (9, 10), and metabolomics (1), have greatly improved our understanding of the physiology of this species. Here we report the complete genome sequence of *Clostridium acetobutylicum* DSM 1731, which was obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany).

The genome was sequenced at the Department of Human Genetics at the University of California, Los Angeles, utilizing Solexa technology according to Illumina’s recommendations. A total of 6,043,147 reads, comprising 217,553,292 bases, were obtained, providing 52× coverage. Assembly was performed by AMOSmp (5) and Velvet (14), using comparative genome assembly and *de novo* assembly. The gaps were filled by sequencing PCR fragments using ABI 3730. The complete genome of *C. acetobutylicum* DSM 1731 consists of a single circular chromosome of 3,942,462 bp (30.9% GC content) and two plasmids, pSMBa (191,996 bp, 30.9% GC content) and pSMBb (11,123 bp, 25.9% GC content).

Putative protein-coding sequences were identified by GeneMark.hmm (8) and Glimmer (3). tRNAscan-SE (7) was used to predict tRNA genes, and the RNAmmer 1.2 server (6) was employed to detect rRNA genes. BLASTP were applied to perform annotation, followed by manual curation. The chro-

mosome of *C. acetobutylicum* DSM 1731 contains 3,765 predicted protein-coding genes, 11 rRNA operons, and 73 tRNA genes, while the two plasmids contain 176 (pSMBa) and 9 (pSMBb) protein-coding genes. Functional classification of all proteins was conducted using NCBI clusters of orthologous groups (COG); 3,072, 126, and 2 COG-classified genes are located on the chromosome, pSMBa, and pSMBb, respectively.

Compared with the genome sequences of *C. acetobutylicum* strains ATCC 824 (11) and EA 2018 (4), the most remarkable feature of strain DSM 1731 is the presence of a 11.1-kb plasmid, pSMBb, which has not been previously reported in this species. In addition, 7 of the 9 predicted genes of pSMBb are hypothetical proteins, suggesting that pSMBb plays an unknown role in clostridial physiology. In addition, 345 single nucleotide variations (SNVs) were identified between the chromosomes of strains DSM 1731 and ATCC 824. The largest insertion is 1,753 bp in the position of 1,278,525 to 1,280,278, located within a hypothetical protein gene. Consistent with strain EA 2018, the megaplasmid in strain DSM 1731 (pSMBa) is 4 bp smaller than that in strain ATCC 824.

**Nucleotide sequence accession numbers.** The complete genome sequence of *Clostridium acetobutylicum* DSM 1731 has been deposited in GenBank under the accession numbers CP002660 to CP002662.

This research was supported by the National Basic Research Programme of China (973 Project, 2011CBA00807) and the CAS/SAFEA International Partnership Program for Creative Research Teams. Yin Li is supported by the Hundreds of Talents Program of the Chinese Academy of Sciences.

\* Corresponding author. Mailing address: Institute of Microbiology, Chinese Academy of Sciences, No.1 West Beichen Road, Chaoyang District, Beijing 100101, China. Phone and fax: 86-10-64807485. E-mail: yli@im.ac.cn.

<sup>∇</sup> Published ahead of print on 8 July 2011.

#### REFERENCES

1. Amador-Noguez, D., et al. 2010. Systems-level metabolic flux profiling elucidates a complete, bifurcated tricarboxylic acid cycle in *Clostridium acetobutylicum*. *J. Bacteriol.* **192**:4452–4461.

2. **Awang, G. M., G. A. Jones, and W. M. Ingledeew.** 1988. The Acetone-Butanol-Ethanol Fermentation. *CRC Crit. Rev. Microbiol.* **15**:S33–S67.
3. **Delcher, A. L., D. Harmon, S. Kasif, O. White, and S. L. Salzberg.** 1999. Improved microbial gene identification with GLIMMER. *Nucleic Acids Res.* **27**:4636–4641.
4. **Hu, S., et al.** 2011. Comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in *Clostridium acetobutylicum* EA 2018. *BMC Genomics* **12**:93.
5. **Kurtz, S., et al.** 2004. Versatile and open software for comparing large genomes. *Genome Biology* **5**:R12.
6. **Lagesen, K., et al.** 2007. RNAmmer: consistent and rapid annotation of rRNA genes. *Nucleic Acids Res.* **35**:3100–3108.
7. **Lowe, T. M., and S. R. Eddy.** 1997. tRNAscan-SE: A program for improved detection of tRNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
8. **Lukashin, A. V., and M. Borodovsky.** 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26**:1107–1115.
9. **Mao, S. M., et al.** 2010. Proteome Reference Map and Comparative Proteomic Analysis between a Wild Type *Clostridium acetobutylicum* DSM 1731 and its Mutant with Enhanced Butanol Tolerance and Butanol Yield. *J. Proteome Res.* **9**:3046–3061.
10. **Mao, S. M., et al.** 2011. Comparative analysis on the membrane proteome of *Clostridium acetobutylicum* wild type strain and its butanol-tolerant mutant. *Mol. Biosystems* **7**:1660–1677.
11. **Nolling, J., et al.** 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183**:4823–4838.
12. **Tomas, C. A., et al.** 2003. DNA array-based transcriptional analysis of asporogenous, nonsolventogenic *Clostridium acetobutylicum* strains SKO1 and M5. *J. Bacteriol.* **185**:4539–4547.
13. **Tomas, C. A., N. E. Welker, and E. T. Papoutsakis.** 2003. Overexpression of groESL in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Appl. Environ. Microbiol.* **69**:4951–4965.
14. **Zerbino, D. R., and E. Birney.** 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**:821–829.