

Genome Sequence of *Clostridium acetobutylicum* GXAS18-1, a Novel Biobutanol Production Strain

Xinchun Mo,^{a,b} Jianxin Pei,^a Yuan Guo,^a Lihua Lin,^a Lixin Peng,^b Chan Kou,^b Danmin Fan,^b Hao Pang^a

National Engineering Research Center for Non-Food Biorefinery, Guangxi Academy of Science, Nanning, Guangxi, China^a; Department of Life Science, Lijiang Teacher's College, Lijiang, Yunnan, China^b

***Clostridium acetobutylicum* is an organism involved in the production of acetone and butanol by traditional acetone-butanol-ethanol fermentation (ABE). We report the draft genome sequence of *C. acetobutylicum* strain GXAS18-1, which can produce ABE directly from cassava flour.**

Received 14 January 2015 Accepted 16 January 2015 Published 5 March 2015

Citation Mo X, Pei J, Guo Y, Lin L, Peng L, Kou C, Fan D, Pang H. 2015. Genome sequence of *Clostridium acetobutylicum* GXAS18-1, a novel biobutanol production strain. *Genome Announc* 3(2):e00033-15. doi:10.1128/genomeA.00033-15.

Copyright © 2015 Mo et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Hao Pang, haopang@gxas.cn.

Clostridia, well-known commercial biobutanol production organisms, have the capability to convert a wide variety of renewable biomass and agricultural waste materials to butanol and ethanol. *Clostridium acetobutylicum* is one group of *Clostridium* strains that is generally used to produce acetone and butanol through traditional acetone-butanol-ethanol fermentation (ABE) (1). Previous studies have focused on the isolation of these organisms and their adoption for various feedstock and fermentation processing functions (2, 3). However, the industrial application of *Clostridium acetobutylicum* strains for high-performance solvent yields and their extreme tolerance for butanol still require more investigation. *C. acetobutylicum* GXAS18-1 is a special butanol fermentation strain that can produce ABE directly from cassava flour by the addition of ammonium acetate. Here, we briefly describe the draft genome sequence of *C. acetobutylicum* GXAS18-1.

The genomic DNA from *C. acetobutylicum* GXAS18-1 (deposited in the China Center for Type Culture Collection [CCTCC] [accession number CCTCC M2011280]) was sequenced using an Illumina MiSeq platform (Illumina, San Diego, CA) with a paired-end library. The SOAPaligner 2.20 and SOAPdenovo 1.3 software (BGI bio, Shenzhen, China) were used to perform quality trimming and *de novo* assembly of the reads, respectively. The resulting contigs were reordered using SOAPdenovo 1.3 and *C. acetobutylicum* ATCC 824 as a reference genome (GenBank accession number NC_003030.1).

Potential coding sequences (CDSs) were predicted using RAST (4), and the assignment of protein functions was performed by searching against the GenBank and Pfam (5) databases using BLASTp (6), while single-nucleotide polymorphisms (SNPs) were identified using MUMmer 3.06 (7). Ribosomal RNAs, tRNAs, and other RNAs were identified using BLASTn, tRNAscan-SE version 1.21 (8), and RNAmmer 1.2 (9), respectively. The orthologous genes between *C. acetobutylicum* GXAS18-1 and *C. acetobutylicum* ATCC 824 were identified using SEED functional comparison (10). The draft genome of *C. acetobutylicum* GXAS18-1 consists of 49 contigs ranging in size from 590 to 924,670 bases, resulting in a total genome size of 3,796,049 nucleotides, with average genome

coverage of 100-fold and a G+C content of 30.79%. Whole-genome SNP scan revealed that 50,702 SNP loci, which were identified in the *C. acetobutylicum* GXAS18-1 genome. Among them, 43,248 SNPs were found in the gene region accounting for the entire SNP set of 85.30%, whereas 7,454 SNPs are in the intergenic region accounting for 14.70% of the entire SNP set. The chromosome of *C. acetobutylicum* GXAS18-1 contains 3,600 CDSs and 3,648 genes, like other *Clostridium*, 3 noncontiguous rRNAs (5S, 16S, and 23S rRNA), 41 tRNAs, and 48 other RNAs. The *C. acetobutylicum* GXAS18-1 chromosome exhibits a high level of synteny to *C. acetobutylicum* ATCC 824, with the exception of 104 unique genes in the former chromosome and 64 genes in the latter, respectively. Furthermore, 2,306 functional genes were identified by SEED functional comparison involved in the metabolic subsystem, in which most of them were involved in carbohydrate (335 genes), amino acid and derivative (277 genes), and cofactor (244 genes) metabolic pathways, followed by RNA metabolism (153 genes) and the cell wall and capsule (150 genes). Others were involved in life cycle metabolism. A significant difference between *C. acetobutylicum* ATCC 824 and GXAS18-1 is that a clustered regularly interspaced short palindromic repeats (CRISPR) system is present in the GXAS18-1 genome but not in the *C. acetobutylicum* ATCC 824 genome.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession numbers [JRWL01000001](https://www.ncbi.nlm.nih.gov/nuccore/JRWL01000001) to [JRWL01000049](https://www.ncbi.nlm.nih.gov/nuccore/JRWL01000049) (BioProject number PRJNA263987).

ACKNOWLEDGMENTS

This research was funded by the National Science Foundation of China (21366007) and the Guangxi Natural Science Foundation Program (2014GXNSFAA118078 and 2014GXNSFBA118129).

REFERENCES

1. Beesch SC. 1953. Acetone-butanol fermentation of starches. *Appl Microbiol* 1:85–95.
2. Schiel-Bengelsdorf B, Montoya J, Linder S, Dürre P. 2013. Butanol

- fermentation. *Environ Technol* 34:1691–1710. <http://dx.doi.org/10.1080/09593330.2013.827746>.
3. Patakova P, Linhova M, Rychtera M, Paulova L, Melzoch K. 2013. Novel and neglected issues of acetone-butanol-ethanol (ABE) fermentation by clostridia: clostridium metabolic diversity, tools for process mapping and continuous fermentation systems. *Biotechnol Adv* 31:58–67. <http://dx.doi.org/10.1016/j.biotechadv.2012.01.010>.
 4. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
 5. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman ARD, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301. <http://dx.doi.org/10.1093/nar/gkr1065>.
 6. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <http://dx.doi.org/10.1093/nar/25.17.3389>.
 7. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12. <http://dx.doi.org/10.1186/gb-2004-5-2-r12>.
 8. 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. <http://dx.doi.org/10.1093/nar/25.5.0955>.
 9. Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
 10. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res* 42:D206–D214. <http://dx.doi.org/10.1093/nar/gkt1226>.