

Improved Draft Genome Sequence of *Clostridium pasteurianum* Strain ATCC 6013 (DSM 525) Using a Hybrid Next-Generation Sequencing Approach

Michael E. Pyne,^a Sagar Utturkar,^b Steven D. Brown,^{b,c} Murray Moo-Young,^a Duane A. Chung,^{a,d} C. Perry Chou^a

Department of Chemical Engineering, University of Waterloo, Waterloo, Ontario, Canada^a; Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, Tennessee, USA^b; Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA^c; Neemo Inc., Hamilton, Ontario, Canada^d

We present an improved draft genome sequence for *Clostridium pasteurianum* strain ATCC 6013 (DSM 525), the type strain of the species and an important solventogenic bacterium with industrial potential. Availability of a near-complete genome sequence will enable strain engineering of this promising bacterium.

Received 16 July 2014 Accepted 21 July 2014 Published 7 August 2014

Citation Pyne ME, Utturkar S, Brown SD, Moo-Young M, Chung DA, Chou CP. 2014. Improved draft genome sequence of *Clostridium pasteurianum* strain ATCC 6013 (DSM 525) using a hybrid next-generation sequencing approach. *Genome Announc.* 2(4):e00790-14. doi:10.1128/genomeA.00790-14.

Copyright © 2014 Pyne 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 Duane A. Chung, duane.chung@uwaterloo.ca, or C. Perry Chou, cpchou@uwaterloo.ca.

Clostridium pasteurianum is a mesophilic, anaerobic Gram-positive bacterium that is apathogenic, can be easily cultivated in chemically defined media, and is more aerotolerant than many clostridia (1, 2). Historically, *C. pasteurianum* has been utilized extensively as a model organism for the study of nitrogen fixation (3) and clostridial ferredoxins (4). More recently, *C. pasteurianum* has received significant biotechnological attention because of its capacity to ferment waste glycerol (5) and glycerol-rich thin stillage (6), which are major by-products of biodiesel and bioethanol production, respectively. In addition to acids and carbon dioxide, glycerol is converted to appreciable quantities of butanol, 1,3-propanediol, and hydrogen gas (7), which have industrial potential as chemicals or biofuels. To allow genetic manipulation of this organism, several recent studies have outlined the need for a genome sequence of *C. pasteurianum* (5, 8). A concurrent effort has reported a draft genome sequence for the type strain (9), while partial or full genome sequences are also available for two *C. pasteurianum* isolates (BC1 [<http://www.ncbi.nlm.nih.gov/GenBank/>] and NRRL B-598 [10]), further demonstrating the appeal of this species. Here we report an improved draft genome assembly for the type strain of *C. pasteurianum*, which was generated using a hybrid next-generation sequencing approach.

The genome of *C. pasteurianum* ATCC 6013 was sequenced using 454, Illumina MiSeq, and single-molecule real-time (SMRT) RS I and RS II sequencing platforms with sequence coverages of 20×, 335×, 80×, and 90×, respectively. *De novo* genome assembly of PacBio data was performed using HGAP.1 protocol from SMRT Analysis software version 2.1. The resulting draft genome sequence of *C. pasteurianum* ATCC 6013 comprises 4,420,124 bp and 12 contigs, an improvement on the 37 contigs reported previously (9). The N50 contig size was improved from 229 kb to 859 kb. Geneious software (Biomatters Ltd., Auckland, New Zealand) identified two supercontigs, with putative contig orderings of ctg10-ctg1C-ctg5-ctg3C-ctg9-ctg12-ctg4-ctg11 and ctg6-ctg7-ctg2-ctg8, respectively (C=complement of contig). Mapping of Illumina and 454 reads against a PacBio assembly

detected only 5 SNPs, which were corrected and defines the high quality of assembly resulting from PacBio reads.

Genome annotation was performed using the Oak Ridge National Laboratory annotation pipeline, based on the Prodigal gene prediction algorithm (11). Ribosomal RNAs were annotated using RNAmmer 1.2 (12), and transfer RNAs were predicted using tRNAscan-SE (13). The G+C content of the genome is 30%. Approximately 4,047 protein-coding, 95 tRNA, and 29 rRNA (14 × 16 S and 15 × 23 S) genes were predicted in the genome. No putative extrachromosomal elements were identified. Gene sequences for enzymes involved in the production of primary metabolites, including ethanol, butanol, and 1,3-propanediol, are present in the genome sequence, many of which are organized in operons similar to *C. acetobutylicum* (14). An acetone formation locus (*adhE-ctfAB-adc*) possessing significant sequence similarity to that of *C. acetobutylicum* (15) is encoded in the genome, yet, as indicated previously (9), acetone production by *C. pasteurianum* has not been reported. It is expected that the draft genome sequence presented herein will guide future strain improvement efforts involving *C. pasteurianum* (16).

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession [JPGY00000000](https://www.ncbi.nlm.nih.gov/nuclink/JPGY00000000). The version described in this report is version [JPGY01000000](https://www.ncbi.nlm.nih.gov/nuclink/JPGY01000000).

ACKNOWLEDGMENTS

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chairs (CRC) program. 454 pyrosequencing was performed using 454 FLX technology by Genome Canada at the McGill University Innovation Centre (Montréal, QC).

Illumina sequencing was performed using a MiSeq sequencer at Oak Ridge National Laboratory (Oak Ridge, TN). SMRT sequencing using the RS I and RS II analyzer was carried out by the DNA Sequencing Core at the University of Michigan (Ann Arbor, MI) and the Genomic Resource Center at the Institute for Genome Sciences (University of Maryland School of Medicine; Baltimore, MD), respectively.

This study leveraged resources supported by the BioEnergy Science Center (BESC). BESC is a U.S. Department of Energy (DOE), Bioenergy Research Center supported by the Office of Biological and Environmental Research (BER) in the DOE's Office of Science. S.U. is supported by the BER Genomic Science Program as part of the Plant Microbe Interfaces Scientific Focus Area. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the DOE under contract DE-AC05-00OR22725.

REFERENCES

- Minton NP, Morris JG. 1983. Regeneration of protoplasts of *Clostridium pasteurianum* ATCC 6013. *J. Bacteriol.* 155:432–434.
- Wiegel J, Tanner R, Rainey FA. 2006. An introduction to the family Clostridiaceae, p 654–678. In Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (ed), *The Prokaryotes: bacteria, firmicutes, cyanobacteria*, vol. 4, 3rd ed. Springer Verlag, New York. http://dx.doi.org/10.1007/0-387-30744-3_20.
- Carnahan JE, Mortenson LE, Mower HF, Castle JE. 1960. Nitrogen fixation in cell-free extracts of *Clostridium pasteurianum*. *Biochim. Biophys. Acta* 44: 520–535. [http://dx.doi.org/10.1016/0006-3002\(60\)91606-1](http://dx.doi.org/10.1016/0006-3002(60)91606-1).
- Mortenson LE, Valentine RC, Carnahan JE. 1962. An electron transport factor from *Clostridium pasteurianum*. *Biochem. Biophys. Res. Commun.* 7:448–452. [http://dx.doi.org/10.1016/0006-291X\(62\)90333-9](http://dx.doi.org/10.1016/0006-291X(62)90333-9).
- Taconi KA, Venkataramanan KP, Johnson DT. 2009. Growth and solvent production by *Clostridium pasteurianum* ATCC (R) 6013 (TM) utilizing biodiesel-derived crude glycerol as the sole carbon source. *Environ. Prog. Sustain. Energy* 28:100–110. <http://dx.doi.org/10.1002/ep.10350>.
- Ahn JH, Sang BI, Um Y. 2011. Butanol production from thin stillage using *Clostridium pasteurianum*. *Bioresour. Technol.* 102:4934–4937. <http://dx.doi.org/10.1016/j.biortech.2011.01.046>.
- Biebl H. 2001. Fermentation of glycerol by *Clostridium pasteurianum*—Batch and continuous culture studies. *J. Ind. Microbiol. Biotechnol.* 27: 18–26. <http://dx.doi.org/10.1038/sj.jim.7000155>.
- Pyne ME, Moo-Young M, Chung DA, Chou CP, Zhang H, Chen R, Chen Y, Zheng Y. 2013. Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. *Biotechnol. Biofuels* 6:1–20.
- Rappert S, Song L, Sabra W, Wang W, Zeng A-P. 2013. Draft genome sequence of type strain *Clostridium pasteurianum* DSM 525 (ATCC 6013), a promising producer of chemicals and fuels. *Genome Announc.* 1(1): e00232-12. <http://dx.doi.org/10.1128/genomeA.00232-12>.
- Kolek J, Sedlar K, Provaznik I, Patakova P. 2014. Draft genome sequence of *Clostridium pasteurianum* NRRL B-598, a potential butanol or hydrogen producer. *Genome Announc.* 2(2):e00192-14. [doi:10.1128/genomeA.00192-14](http://dx.doi.org/10.1128/genomeA.00192-14).
- Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, 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>.
- 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>.
- Nölling J, Breton G, Omelchenko MV, Makarova KS, Zeng Q, Gibson R, Lee HM, Dubois J, Qiu D, Hitti J, Wolf YI, Tatusov RL, Sabathe F, Doucette-Stamm L, Soucaille P, Daly MJ, Bennett GN, Koonin EV, Smith DR, Finishing GTCSCP. 2001. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* 183:4823–4838. <http://dx.doi.org/10.1128/JB.183.16.4823-4838.2001>.
- Petersen DJ, Cary JW, Vanderleyden J, Bennett GN. 1993. Sequence and arrangement of genes encoding enzymes of the acetone-production pathway of *Clostridium acetobutylicum* ATCC 824. *Gene* 123:93–97. [http://dx.doi.org/10.1016/0378-1119\(93\)90545-E](http://dx.doi.org/10.1016/0378-1119(93)90545-E).
- Pyne ME, Bruder M, Moo-Young M, Chung DA, Chou CP. 2014. Technical guide for genetic advancement of underdeveloped and intractable *Clostridium*. *Biotechnol. Adv.* 32:623–641. <http://dx.doi.org/10.1016/j.biotechadv.2014.04.003>.