genomeA Journals.ASM.org

Draft Genome Sequence of the Cellulolytic and Xylanolytic Thermophile *Clostridium clariflavum* Strain 4-2a

Elise A. Rooney,[a] Kenneth T. Rowe,[a] Anna Guseva,[b,c] Marcel Huntemann,[d] James K. Han,[d] Amy Chen,[d] Nikos C. Kyrpides,[d] Konstantinos Mavromatis,[d] Victor M. Markowitz,[d] Krishna Palaniappan,[d] Natalia Ivanova,[d] Amrita Pati,[d] Konstantinos Liolios,[d] Henrik P. Nordberg,[d] Michael N. Cantor,[d] Susan X. Hua,[d] Nicole Shapiro,[d] Tanja Woyke,[d] Lee R. Lynd,[b,c] Javier A. Izquierdo[a]

Department of Biology, Hofstra University, Hempstead, New York, USA[a]; Thayer School of Engineering, Dartmouth College, Hanover, New Hampshire, USA[b]; BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA[c]; DOE Joint Genome Institute, Walnut Creek, California, USA[d]

***Clostridium clariflavum* strain 4-2a, a novel strain isolated from a thermophilic biocompost pile, has demonstrated an extensive capability to utilize both cellulose and hemicellulose under thermophilic anaerobic conditions. Here, we report the draft genome of this strain.**

Received 10 June 2015 **Accepted** 16 June 2015 **Published** 23 July 2015

Citation Rooney EA, Rowe KT, Guseva A, Huntemann M, Han JK, Chen A, Kyrpides NC, Mavromatis K, Markowitz VM, Palaniappan K, Ivanova N, Pati A, Liolios K, Nordberg HP, Cantor MN, Hua SX, Shapiro N, Woyke T, Lynd LR, Izquierdo JA. 2015. Draft genome sequence of the cellulolytic and xylanolytic thermophile *Clostridium clariflavum* strain 4-2a. Genome Announc 3(4):e00797-15. doi:10.1128/genomeA.00797-15.

Address correspondence to Javier A. Izquierdo, javier.izquierdo@hofstra.edu.

The ability of thermophilic clostridia to extensively degrade lignocellulosic materials and produce fermentation products that may serve as biocommodities makes them excellent candidates for consolidating bioprocessing applications (1). The genome sequence of *Clostridium clariflavum* DSM 19732 revealed novel mechanisms among thermophilic clostridia to break down cellulose and hemicellulose (2). *C. clariflavum* strain 4-2a is a novel strain isolated from compost (3), which has demonstrated the additional ability to utilize xylose and considerably degrade unpretreated switchgrass (4).

Genomic DNA of *C. clariflavum* strain 4-2a was obtained through a phenol-chloroform-CTAB extraction procedure, as previously described (2). The draft genome was generated at the DOE Joint Genome Institute (JGI) using Illumina data (5) from a short-insert paired-end library with an average insert size of 270 bp, generating 24,015,970 reads, and an Illumina long-insert paired-end library with an average insert size of 6,484.62 bp, generating 25,583,980 reads and totaling 7,440 Mbp of Illumina data (F. Chen, personal communication).

Protocols for library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov. The initial draft data was assembled with ALLPATHS version 39750 (6) and contained 156 contigs in 18 scaffolds, and the consensus was shredded into 10-kbp overlapping shreds. The Illumina draft data were assembled with Velvet version 1.1.05 (7), and consensus sequences were computationally shredded into 1.5-kbp overlapping shreds. The Illumina draft data were assembled again with Velvet using shreds from the first Velvet assembly to guide the next assembly. The consensus from the second Velvet assembly was shredded into 1.5-kbp overlapping shreds. The shreds from the ALLPATHS and Velvet assemblies and a subset of Illumina CLIP paired-end reads were assembled using parallel Phrap version 4.24 (High Performance Software, LLC). Misassemblies were corrected in Consed (8–10). Gap closure was accomplished using repeat resolution software (W. Gu, personal communication) and sequencing of 104 PCR PacBio consensus sequences (C. Han, personal communication). The total size of the genome is 4.9 Mb and the final assembly is based on 7,440 Mbp of Illumina data, providing 1,518× coverage. Project information is available in the Genomes Online Database (11), and the final annotation is available from the Integrated Microbial Genome (IMG) system (12). Genes were identified using Prodigal (13) and GenePRIMP (14) as part of the JGI's microbial annotation pipeline (15). Additional gene prediction and functional annotation was performed with the IMG-ER platform (16).

Initial findings from the genome of *C. clariflavum* strain 4-2a include the presence of genes coding for xylose isomerase and xylulose kinase in a genomic island not present in the type strain, DSM 19732. The inventory of glycoside hydrolases is almost identical to the one reported for the type strain (2), except for an additional beta-glucosidase/xylosidase from glycoside hydrolase family 3 identified in the same genomic island. Further ongoing analysis is expected to provide insight into this strain's lignocellulolytic capabilities and inform the development of genetic tools.

Nucleotide sequence accession numbers. The genome sequence of *C. clariflavum* 4-2a is deposited in GenBank under the accession numbers ASAA01000001 to ASAA01000016 and the annotated genome in IMG through accession number 2524614797.

ACKNOWLEDGMENTS

The BioEnergy Science Center (BESC) is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231.

REFERENCES

1. **Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS.** 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol Mol

Biol Rev **66:**506–577. http://dx.doi.org/10.1128/MMBR.66.3.506-577.2002.

2. **Izquierdo JA, Goodwin L, Davenport KW, Teshima H, Bruce D, Detter C, Tapia R, Han S, Land M, Hauser L, Jeffries CD, Han J, Pitluck S, Nolan M, Chen A, Huntemann M, Mavromatis K, Mikhailova N, Liolios K, Woyke T, Lynd LR.** 2012. Complete genome of *Clostridium clariflavum* DSM 19732. Stand Genomic Sci **6:**104–115. http://dx.doi.org/10.4056/sigs.2535732.
3. **Sizova MV, Izquierdo JA, Panikov NS, Lynd LR.** 2011. Cellulose- and xylan-degrading thermophilic anaerobic bacteria from biocompost. Appl Environ Microbiol **77:**2282–2291. http://dx.doi.org/10.1128/AEM.01219-10.
4. **Izquierdo JA, Pattathil S, Guseva A, Hahn MG, Lynd LR.** 2014. Comparative analysis of the ability of *Clostridium clariflavum* strains and *Clostridium thermocellum* to utilize hemicellulose and unpretreated plant material. Biotechnol Biofuels **7:**136. http://dx.doi.org/10.1186/s13068-014-0136-4.
5. **Bennett S.** 2004. Solexa Ltd. Pharmacogenomics **5:**433–438. http://dx.doi.org/10.1517/14622416.5.4.433.
6. **Butler J, MacCallum I, Kleber M, Shlyakhter IA, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB.** 2008. ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. Genome Res **18:**810–820. http://dx.doi.org/10.1101/gr.7337908.
7. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res **18:**821–829. http://dx.doi.org/10.1101/gr.074492.107.
8. **Ewing B, Green P.** 1998. Base-calling of automated sequencer traces using Phred, II: error probabilities. Genome Res **8:**186–194.
9. **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. http://dx.doi.org/10.1101/gr.8.3.175.
10. **Gordon D, Abajian C, Green P.** 1998. *Consed*: a graphical tool for sequence finishing. Genome Res **8:**195–202. http://dx.doi.org/10.1101/gr.8.3.195.
11. **Reddy TB, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA, Kyrpides NC.** 2015. The Genomes Online Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. Nucleic Acids Res **43:**D1099–D1106. http://dx.doi.org/10.1093/nar/gku950.
12. **Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC.** 2012. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res **40:**D115–D122. http://dx.doi.org/10.1093/nar/gkr1044.
13. **Hyatt D, Chen GL, 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.
14. **Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC.** 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods **7:**455–457. http://dx.doi.org/10.1038/nmeth.1457.
15. **Mavromatis K, Ivanova NN, Chen IM, Szeto E, Markowitz VM, Kyrpides NC.** 2009. The DOE-JGI standard operating procedure for the annotations of microbial genomes. Stand Genomic Sci **1:**63–67. http://dx.doi.org/10.4056/sigs.632.
16. **Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC.** 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics **25:**2271–2278. http://dx.doi.org/10.1093/bioinformatics/btp393.