genomeA
Journals.ASM.org

Draft Genome Sequence of the Cellulolytic, Mesophilic, Anaerobic Bacterium *Clostridium termitidis* Strain CT1112 (DSM 5398)

Sadhana Lal,[a] Umesh Ramachandran,[a] Xiangli Zhang,[c] Riffat Munir,[a] Richard Sparling,[b] David B. Levin[a]

Department of Biosystems Engineering, University of Manitoba, Winnipeg, Manitoba, Canada[a]; Department of Microbiology, University of Manitoba, Winnipeg, Manitoba, Canada[b]; Department of Plant Science, University of Manitoba, Winnipeg, Manitoba, Canada[c]

Here, we report the draft genome sequence of *Clostridium termitidis* strain CT1112 (DSM 5398), a mesophilic, cellulolytic bacterium that can utilize a variety of sugars, as well as pure cellulose, as a sole carbon source; it also synthesizes fermentation end products with potential industrial applications.

Received 8 April 2013 **Accepted** 15 April 2013 **Published** 23 May 2013

Citation Lal S, Ramachandran U, Zhang X, Munir R, Sparling R, Levin DB. 2013. Draft genome sequence of the cellulolytic, mesophilic, anaerobic bacterium *Clostridium termitidis* strain CT1112 (DSM 5398). Genome Announc. 1(3):e00281-13. doi:10.1128/genomeA.00281-13.

Copyright © 2013 Lal et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported license.

Address correspondence to David B. Levin, david.levin@ad.umanitoba.ca.

Clostridium termitidis strain CT1112 (DSM 5398) is a Gram-positive, mesophilic, anaerobic, cellulolytic bacterium isolated from the gut of the wood-feeding termite *Nasutitermes lujae* (1). Based on its 16S rRNA, *C. termitidis* belongs to *Clostridium* cluster III (2). It can utilize a wide variety of substrates, such as cellulose, cellobiose, glucose, fructose, and many other sugar monomers, as a sole carbon source, and it produces hydrogen (H_2), carbon dioxide (CO_2), acetate, formate, lactate, and ethanol as major fermentation end products (1, 3).

The genome of *C. termitidis* CT1112 was sequenced by the Genome Québec/McGill University platform using a Roche/454s GS-FLX Titanium sequencer by a whole-genome shotgun strategy, which obtained 303,437 reads. A 454 standard flowgram format (.sff) read file was assembled using Newbler v2.3. The final draft genome assembly has approximately 17-fold coverage and contains 78 contigs (>800 bp in length), with a total size of 6,415,858 bp, an N_{50} contig length of 146,289 nucleotides, and a mean G+C content of 41.18%. The draft genome sequence was automatically annotated using IMG-ER, an online system developed by the U.S. Department of Energy Joint Genome Institute (JGI) (http://www.jgi.doe.gov/). The IMG-ER annotation was processed by a JGI-developed Gene Prediction Improvement Pipeline (GenePRIMP) (4) and was further subjected to manual curation using Artemis (5). The draft genome sequence of *C. termitidis* is estimated to have a total of 5,389 genes, including 5,302 protein-coding genes, 55 tRNAs, and 7 rRNAs.

The *C. termitidis* genome is larger than those of other mesophilic and thermophilic cellulolytic *Clostridium* spp., such as *Clostridium cellulolyticum* H10 (4,068,724 bp), *Clostridium cellulovorans* 743B (5,262,222 bp), *Clostridium phytofermentans* ISDg (4,847,594 bp), *Clostridium thermocellum* ATCC 27405 (3,843,301 bp), and *C. thermocellum* DSM 1313 (3,561,619 bp). The G+C content of *C. termitidis* is higher (41.18%) than those of other cellulolytic *Clostridium* species (31.21% to 39.15%), as is the number of predicted genes (5,389).

C. termitidis protein-coding genes were verified using other cellulolytic *Clostridium* species as reference organisms. Amino acid sequences for each gene product were retrieved from the JGI genome portal (http://genome.jgi-psf.org/) (6) and the NCBI database (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi), and sequence alignments against *C. termitidis* genes were performed. The corresponding gene loci and enzymes for each pathway were identified by percent amino acid sequence identity and were based on a conserved domain of proteins (7). In this manner, the key enzymes involved in *C. termitidis* core metabolism, as well as the major cellulosomal components and glycoside hydrolases, were identified. *C. termitidis* has potential as an industrial microorganism for the production of biofuels and/or other value-added products through direct cellulose fermentation via consolidated bioprocessing.

Nucleotide sequence accession numbers. The genome sequence of *C. termitidis* strain CT1112 (DSM 5398) has been deposited at DDBJ/EMBL/GenBank under the accession no. AORV00000000. The version described in this paper is the first version, accession no. AORV01000000.

ACKNOWLEDGMENTS

This work was supported by funds from Genome Canada, through the Applied Genomics Research in Bioproducts or Crops (ABC) program for the grant titled "Microbial genomics for biofuels and coproducts from biorefining processes," and the Province of Manitoba from the Manitoba Research Innovation Fund (MRIF).

REFERENCES

1. **Hethener P, Brauman A, Garcia JL.** 1992. *Clostridium termitidis* sp. nov., a cellulolytic bacterium from the gut of the woodfeeding termite, *Nasutitermes lujae*. Syst. Appl. Microbiol. 15:52–58.
2. **Collins MD, Lawson PA, Willems A, Cordoba JJ, Fernandez-Garayzabal J, Garcia P, Cai J, Hippe H, Farrow JA.** 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int. J. Syst. Bacteriol. **44**:812–826.
3. **Ramachandrana U, Wrana N, Cicek N, Sparling R, Levin DB.** 2008. Hydrogen production and end product synthesis patterns by *Clostridium termitidis* strain CT1112 in batch fermentation cultures with cellobiose or a-cellulose. Int. J. Hydrogen Energ. **33**:7006–7012.
4. **Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lyki-**

dis A, Kyrpides NC. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat. Methods **7**:455–457.
5. **Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA.** 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics **28**: 464–469.
6. **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.
7. **Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, Fong JH, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Jackson JD, Ke Z, Lanczycki CJ, Lu F, Marchler GH, Mullokandov M, Omelchenko MV, Robertson CL, Song JS, Thanki N, Yamashita RA, Zhang D, Zhang N, Zheng C, Bryant SH.** 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. **39**: D225–D229.