Draft Genome Sequences for Clostridium thermocellum Wild-Type Strain YS and Derived Cellulose Adhesion-Defective Mutant Strain AD2

Steven D. Brown, Raphael Lamed, Ely Morag, Ilya Borovok, Yuval Shoham, Dawn M. Klingeman, Courtney M. Johnson, Zamin Yang, Miriam L. Land, Sagar M. Utturkar, Martin Keller and Edward A. Bayer

J. Bacteriol. **2012, 194(12):3290. DOI: 10.1128/JB.00473-12.**

JB Journals.ASM.org

GENOME ANNOUNCEMENT

Draft Genome Sequences for *Clostridium thermocellum* Wild-Type Strain YS and Derived Cellulose Adhesion-Defective Mutant Strain AD2

Steven D. Brown,[a,b,c] Raphael Lamed,[d] Ely Morag,[e] Ilya Borovok,[d] Yuval Shoham,[f] Dawn M. Klingeman,[a,b] Courtney M. Johnson,[a,b] Zamin Yang,[a] Miriam L. Land,[a] Sagar M. Utturkar,[b] Martin Keller,[a,b] and Edward A. Bayer[e]

Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA[a]; BioEnergy Science Center, Oak Ridge, Tennessee, USA[b]; Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, Tennessee, USA[c]; Department of Molecular Microbiology and Biotechnology, Tel-Aviv University, Ramat Aviv, Israel[d]; Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel[e]; and Department of Biotechnology and Food Engineering, Technion-Israel Institute of Technology, Haifa, Israel[f]

***Clostridium thermocellum* wild-type strain YS is an anaerobic, thermophilic, cellulolytic bacterium capable of directly converting cellulosic substrates into ethanol. Strain YS and a derived cellulose adhesion-defective mutant strain, AD2, played pivotal roles in describing the original cellulosome concept. We present their draft genome sequences.**

Clostridium thermocellum was characterized over 50 years ago (31), and its ability for efficient degradation and utilization of cellulose for ethanol production was recognized early on (1, 2, 21, 26, 33, 34, 36, 37, 39). It is an obligate anaerobic microorganism that has one of the highest growth rates on cellulose. The bacterium possesses productivity advantages associated with thermophilic growth and is capable of producing its own enzymes for lignocellulosic biomass breakdown (see reviews in references 4, 11, 28, and 29). Recently, important progress has been made in understanding *C. thermocellum* ethanol tolerance (10), a targeted deletion system has been developed (38), and four *C. thermocellum* genome sequences have been determined (16, 19) (GenBank accession no. CP000568.1).

C. thermocellum strain YS was purified from samples derived from hot springs at Yellowstone National Park in the United States, and it was characterized as a potent cellulolytic strain. Strain YS and a derived cellulose adhesion-defective mutant (AD2) played pivotal roles in describing the original cellulosome concept that recognized that *C. thermocellum* cellulases and associated polysaccharide-degrading enzymes are packaged in organized, high-molecular-weight, cellulolytic enzyme complexes (5, 23, 24). Strain YS was used in a number of subsequent studies (e.g., references 6, 8, 13–15, 17, 18, 25, 27, 32, and 35), and the cellulosome concept has served as a model for different clostridia and other related anaerobic bacteria, e.g., *Clostridium cellulovorans*, *Clostridium cellulolyticum*, *Clostridium josui*, *Clostridium papyrosolvens*, *Acetivibrio cellulolyticus*, *Bacteroides cellulosolvens*, and *Ruminococcus flavefaciens* (3, 7, 12, 22).

Draft genome data for strain YS were generated using a combination of 454 (30) and Illumina (9) HiSeq2000 technologies from 3-kb and 500-bp paired-end libraries, respectively. The 454 data consisted of 650,450 reads and generated 207,578,580 bp. After trimming and filtering of Illumina data (CLC Genomics Workbench version 4.9.1), there were 16,806,784,095 and 15,259,925,308 bp of sequence data for strains YS and AD2 from 174,965,956 and 159,631,515 reads, respectively, with an average length of 96 bp. Trimmed Illumina reads were assembled using the CLC Genomics Workbench. The consensus Illumina sequences for strain YS were processed further by generating 1.5-kb overlapping fake reads using the fb_dice.pl script, which is part of the FragBlast module (http://www.clarkfrancis.com/codes/fb_dice.pl). The Newbler application (version 2.6; 454 Life Sciences) was then used to assemble the YS Illumina consensus sequences and the 454 reads into 100 large (≥500-bp) contiguous DNA elements of approximately 3.46 Mb. The average YS contig size was 34,644 bp, the N50 contig size was 126,840 bp, and the largest contig was 330,620 bp. The genome had an overall estimated G+C content of ~39%. Strain AD2 had 132 large contigs with an average size of 26,020 bp, the N50 contig size was 76,137 bp, and the largest contig was 269,895 bp. Draft genome sequences were annotated at Oak Ridge National Laboratory using an automated annotation pipeline, based on the Prodigal gene prediction algorithm (20). Sequence data for DNA contigs, coding and translation models, annotations, and metabolic reconstructions are available online (http://genome.ornl.gov/microbial/guest/YSORG_Nov2011_Hybrid and http://genome.ornl.gov/microbial/guest/AD2_NovRerun).

This study reveals the *C. thermocellum* YS and AD2 genome sequences for the first time. Access to these genome sequences, which are linked to important prior observations, will facilitate further studies with this genus and species.

Nucleotide sequence accession numbers. The *C. thermocellum* YS and AD2 nucleotide sequences have been deposited in DDBJ/EMBL/GenBank under accession numbers AJGT00000000 and AJGS00000000, respectively, and the versions described in this paper are the first versions. The entire data set has been deposited in the National Center for Biotechnology Information (NCBI) Sequence-Read Archive (SRA) database under accession number SRA049437.

ACKNOWLEDGMENTS

We appreciate the contribution of Ido Lavi for preparation of genomic DNA of strains YS and AD2 used in this study.

This research was supported by the Office of Biological and Environ-

Received 23 March 2012 **Accepted** 29 March 2012

Address correspondence to Edward A. Bayer, ed.bayer@weizmann.ac.il.

doi:10.1128/JB.00473-12

mental Research in the DOE Office of Science through the BioEnergy Science Center, a U.S. DOE Bioenergy Research Center. ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract no. DE-AC05-00OR22725. Grants from the following foundations are gratefully acknowledged: the United States-Israel Binational Science Foundation (BSF), Jerusalem, Israel; the Israel Science Foundation (grant no. 966/09 and 24/11); The Israeli Centers of Research Excellence (I-CORE) program (Center no. 152/11); The Alternative Energy Research Initiative (AERI) Bioenergy Consortium; and the Technion-Niedersachsen Research Cooperation Program. Y.S. holds the Erwin and Rosl Pollak Chair in Biotechnology at the Technion, and E.A.B. is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry.

REFERENCES

1. **Ait N, Creuzet N, Cattaneo J.** 1979. Characterization and purification of thermostable beta-glucosidase from *Clostridium thermocellum*. Biochem. Biophys. Res. Commun. **90**:537–546.
2. **Ait N, Creuzet N, Cattaneo J.** 1982. Properties of beta-glucosidase purified from *Clostridium thermocellum*. J. Gen. Microbiol. **128**:569–577.
3. **Bayer EA, Belaich J-P, Shoham Y, Lamed R.** 2004. The cellulosomes: multi-enzyme machines for degradation of plant cell wall polysaccharides. Annu. Rev. Microbiol. **58**:521–554.
4. **Bayer EA, Chanzy H, Lamed R, Shoham Y.** 1998. Cellulose, cellulases and cellulosomes. Curr. Opin. Struct. Biol. **8**:548–557.
5. **Bayer EA, Kenig R, Lamed R.** 1983. Adherence of *Clostridium thermocellum* to cellulose. J. Bacteriol. **156**:818–827.
6. **Bayer EA, Lamed R.** 1986. Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. J. Bacteriol. **167**:828–836.
7. **Bayer EA, Lamed R, White BA, Flint HJ.** 2008. From cellulosomes to cellulosomics. Chem. Rec. **8**:364–377.
8. **Bayer EA, Setter E, Lamed R.** 1985. Organization and distribution of the cellulosome in *Clostridium thermocellum*. J. Bacteriol. **163**:552–559.
9. **Bennett S.** 2004. Solexa Ltd. Pharmacogenomics **5**:433–438.
10. **Brown SD, et al.** 2011. Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*. Proc. Natl. Acad. Sci. U. S. A. **108**:13752–13757.
11. **Demain AL, Newcomb M, Wu JHD.** 2005. Cellulase, clostridia, and ethanol. Microbiol. Mol. Biol. Rev. **69**:124–154.
12. **Doi RH, Kosugi A.** 2004. Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat. Rev. Microbiol. **2**:541–551.
13. **Dror TW, et al.** 2003. Regulation of the cellulosomal *celS* (*cel48A*) gene of *Clostridium thermocellum* is growth rate dependent. J. Bacteriol. **185**: 3042–3048.
14. **Dror TW, Rolider A, Bayer EA, Lamed R, Shoham Y.** 2003. Regulation of expression of scaffoldin-related genes in *Clostridium thermocellum*. J. Bacteriol. **185**:5109–5116.
15. **Dror TW, Rolider A, Bayer EA, Lamed R, Shoham Y.** 2005. Regulation of major cellulosomal endoglucanases of *Clostridium thermocellum* differs from that of a prominent cellulosomal xylanase. J. Bacteriol. **187**:2261–2266.
16. **Feinberg L, et al.** 2011. Complete genome sequence of the cellulolytic thermophile *Clostridium thermocellum* DSM1313. J. Bacteriol. **193**:2906–2907.
17. **Fernandes AC, et al.** 1999. Homologous xylanases from *Clostridium thermocellum*: evidence for bi-functional activity, synergism between xylanase catalytic modules and the presence of xylan-binding domains in enzyme complexes. Biochem. J. **342**:105–110.
18. **Gerwig GJ, et al.** 1989. Novel O-linked carbohydrate chains in the cellulase complex (cellulosome) of *Clostridium thermocellum* 3-O-methyl-*N*-acetylglucosamine as a constituent of a glycoprotein. J. Biol. Chem. **264**: 1027–1035.
19. **Hemme CL, et al.** 2010. Sequencing of multiple Clostridia genomes related to biomass conversion and biofuels production. J. Bacteriol. **192**: 6494–6496.
20. **Hyatt D, et al.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119. doi: 10.1186/1471-2105-11-119.
21. **Johnson EA, Sakajoh M, Halliwell G, Madia A, Demain AL.** 1982. Saccharification of complex cellulosic substrates by the cellulase system from *Clostridium thermocellum*. Appl. Environ. Microbiol. **43**:1125–1132.
22. **Lamed R, Naimark J, Morgenstern E, Bayer EA.** 1987. Specialized cell surface structures in cellulolytic bacteria. J. Bacteriol. **169**:3792–3800.
23. **Lamed R, Setter E, Bayer EA.** 1983. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. J. Bacteriol. **156**:828–836.
24. **Lamed R, Setter E, Kenig R, Bayer EA.** 1983. The cellulosome—a discrete cell surface organelle of *Clostridium thermocellum* which exhibits separate antigenic, cellulose-binding and various cellulolytic activities. Biotechnol. Bioeng. Symp. **13**:163–181.
25. **Lamed R, Tormo J, Chirino AJ, Morag E, Bayer EA.** 1994. Crystallization and preliminary X-ray analysis of the major cellulose-binding domain of the cellulosome from *Clostridium thermocellum*. J. Mol. Biol. **244**: 236–237.
26. **Lamed R, Zeikus JG.** 1980. Ethanol production by thermophilic bacteria: relationship between fermentation product yields of and catabolic enzyme activities in *Clostridium thermocellum* and *Thermoanaerobium brockii*. J. Bacteriol. **144**:569–578.
27. **Lamed RJ, Lobos JH, Su TM.** 1988. Effects of stirring and hydrogen on fermentation products of *Clostridium thermocellum*. Appl. Environ. Microbiol. **54**:1216–1221.
28. **Lynd LR, van Zyl WH, McBride JE, Laser M.** 2005. Consolidated bioprocessing of cellulosic biomass: an update. Curr. Opin. Biotechnol. **16**:577–583.
29. **Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS.** 2002. Microbial cellulose utilization: fundamentals and biotechnology. Microbiol. Mol. Biol. Rev. **66**:506–577.
30. **Margulies M, et al.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437**:376–380.
31. **McBee RH.** 1954. The characteristics of *Clostridium thermocellum*. J. Bacteriol. **67**:505–506.
32. **Morag E, Bayer EA, Hazlewood GP, Gilbert HJ, Lamed R.** 1993. Cellulase S_S (CelS) is synonymous with the major cellobiohydrolase (subunit S8) from the cellulosome of *Clostridium thermocellum*. Appl. Biochem. Biotechnol. **43**:147–151.
33. **Ng TK, Zeikus JG.** 1981. Comparison of extracellular cellulase activities of *Clostridium thermocellum* LQRI and *Trichoderma reesei* QM9414. Appl. Environ. Microbiol. **42**:231–240.
34. **Ng TK, Zeikus JG.** 1981. Purification and characterization of an endoglucanase (1,4-beta-D-glucan glucanohydrolase) from *Clostridium thermocellum*. Biochem. J. **199**:341–350.
35. **Poole DM, et al.** 1992. Identification of the cellulose-binding domain of the cellulosome subunit S1 from *Clostridium thermocellum* YS. FEMS Microbiol. Lett. **99**:181–186.
36. **Saddler JN, Chan MKH.** 1982. Optimization of *Clostridium thermocellum* growth on cellulose and pretreated wood substrates. Eur. J. Appl. Microbiol. Biotechnol. **16**:99–104.
37. **Shinmyo A, Garcia-Martne DV, Demian AL.** 1979. Studies on the extracellular cellulolytic enzyme complex produced by *Clostridium thermocellum*. J. Appl. Biochem. **1**:202–209.
38. **Tripathi SA, et al.** 2010. Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* mutant. Appl. Environ. Microbiol. **76**:6591–6599.
39. **Zertuche L, Zall RR.** 1982. A study of producing ethanol from cellulose using *Clostridium thermocellum*. Biotechnol. Bioeng. **24**:57–68.