J. Bacteriol.

(Revised 03/26/09: JB256-09)

Genome Announcement

Genome Sequence of the Anaerobic, Thermophilic and Cellulolytic Bacterium *Anaerocellum thermophilum* DSM 6725

Irina A. Kataeva[1,4], Sung-Jae Yang[1,4], Phuongan Dam[1,3,4], Farris L Poole II[1], Yanbin Yin[1,3,4], Fengfeng Zhou[1,3,4], Wen-chi Chou[1,3,4], Ying Xu[1,3,4], Lynne Goodwin[6], David R. Sims[6], John C. Detter[6], Loren J. Hauser[5], Janet Westpheling[2,4] and Michael W. W. Adams[1,4]*

[1]Department of Biochemistry and Molecular Biology, [2]Department of Genetics and [3]Institute of Bioinformatics, University of Georgia, GA 30602; [4]BioEnergy Science Center and [5]Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831; [6]Los Alamos National Laboratory, Bioscience Division B-6 Genome Science, Joint Genome Institute, Los Alamos, NM 87545

* Author to whom correspondence should be addressed:
Michael W. W. Adams
Department of Biochemistry and Molecular Biology
Davison Life Sciences Complex
University of Georgia
Athens, GA 30602-7229
Tel: 706 542-2060 Fax: 706 542-0229
E-mail: adams@bmb.uga.edu

ABSTRACT

Anaerocellum thermophilum **DSM 6725 is a strictly anaerobic bacterium that grows optimally at 75°C. It uses a variety of polysaccharides, including crystalline cellulose and untreated plant biomass, and has potential utility in biomass conversion. Here we report its complete genome sequence of 2.97 Mb, which is contained within one chromosome and two plasmids (of 8.3 and 3.6 kb). The genome encodes a broad set of cellulolytic enzymes, transporters and pathways for sugar utilization and compared to those of other saccharolytic, anaerobic thermophiles is most similar to that of *Caldicellulosiruptor saccharolyticus* DSM 8903.**

Microorganisms that grow at elevated temperatures and are able to utilize a range of carbohydrates have potential utility in the conversion of lignocellulosic biomass to bioenergy. Many hyperthermophilic bacteria and archaea ($T_{opt} \geq 80°C$) from marine environments are able to grow on various α- and β-linked glucans, but none of them are able to efficiently hydrolyze crystalline cellulose and plant biomass (1). The most thermophilic cellulolytic species known at present include the strictly anaerobic bacterium, *Anaerocellum thermophilum*. The type strain (Z-1320) was isolated from thermal springs of Kamchatka (Russia) almost two decades ago (10). It grows optimally at 75°C and utilizes both simple and complex polysaccharides with lactate, acetate, CO_2 and H_2 as end products (10). In particular, *A. thermophilum* DSM 6725 efficiently utilizes the two main components of plant biomass (cellulose and hemicellulose), as well

as untreated grasses with low lignin (napiergrass, bermuda grass) or high lignin (switchgrass) contents and a hardwood (poplar) (12).

The genome of *A. thermophilum* DSM 6725 was sequenced at the Joint Genome Institute (JGI) using an 8 kb library. In addition to Sanger sequencing, 454 pyrosequencing was carried out to a depth of 20x coverage. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov. Draft assemblies were based on 38,121 total reads and all libraries provided 13x coverage. The Phred/Phrap/Consed software package (www.phrap.com <http://www.phrap.com>) was used for sequence assembly and quality assessment (4-6). After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher (7) or transposon bombing of bridging clones (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, custom primer walk or PCR amplification (Roche Applied Science, Indianapolis, IN). A total of 981 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The completed genome sequences contain 41,706 reads, achieving an average of 13.4x coverage in the chromosome (and 56x in pATHE01 and 15x in the pATHE01) with an error rate 0.01 in 100,000.

The chromosome of 2919718 bp has 35.17 % of GC. pATHE01 of 8291 bp and pATHE02 of 3653 bp have 38.53% and 42.92% of GC, respectively. Two native *A. thermophilum* DSM 6725 plasmids pBAL (AX710673) and pBAS (AX710687) had been

sequenced earlier (2). The sequence of pATHE02 perfectly matches that of pBAS2, while that of pATHE01 is similar to pBAL but pBAL contains 19 gaps and 3 mismatches. Similarly, the gene sequences of 16S rRNA (L09180) (9) of a large multi-domain glycoside hydrolase CelA (Z86105) (13) from *A. thermophilum* Z-1320 were previously reported and the corresponding genes in the genome sequence contain 2 inserts and 12 mismatches, and 23 mismatches, respectively.

The genome size of *A. thermophilum* DSM 6725 is similar to that of the celluloytic *Caldicellulosiruptor saccharolyticus* DSM 8903 (T_{opt} 70^{o}C, 2.97 Mb, 35% GC, CP000679), which was also isolated from a continental hot spring (11), but smaller than that of the cellulolytic bacterium *Clostridium thermocellum* ATCC 27405 (T_{opt} 60^{o}C, 3.8 Mb, 39% GC, NC_009012, US DOE Joint Genome Institute), and larger than that of the xylanolytic bacterium *Thermotoga maritima* MSB8 (T_{opt} 80^{o}C, 1.86 Mb, 46% GC, NC_000853) (8).

The chromosome of *A. thermophilum* DSM 6725 is predicted to contain 2662 coding sequences, three rRNA operons, and 47 tRNA genes. pATHE01 and pATHE01 are predicted to contain 8 and 4 orfs, respectively. The genes in the genome of *A. thermophilum* DSM 6725 are predicted to be organized into 573 multi-gene transcripts and 626 single-gene transcripts (3), and a total of102 transcripts contain genes that are predicted to be involved in the degradation of complex polysaccharides. While most of the genes in *A. thermophilum* DSM 6725 have their best Blast hits (e-value of <1e-20) in the genome of *C. saccharolyticus* DSN 8903, a total of 550 genes do not. Of these, 18

are predicted to have functions relating to biomass degradation, suggesting that these genes may contribute to any nutritional differences between the two organisms. The genomes of *A. thermophilum* DSM 6725 and *C. saccharolyticus* DSM 8903 contain 25 and 68 putative transposase genes, respectively. This might account for the apparent genome plasticity within the two genomes of these closely-related bacteria that were isolated from similar geothermal freshwater environments.

Nucleotide sequence accession number. The genome sequence and annotation of the *Anaerocellum thermophilum* DSM 6725 chromosome and the two plasmids pATHE01 and pATHE02 were deposited in GenBank under accession numbers CP001393, CP001394 and CP001395, respectively.

ACKNOWLEDGEMENTS

This research was supported by a grant (DE-PS02-06ER64304) from the Bioenergy Science Center (BESC), Oak Ridge National Laboratory, a US Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. Work at the Joint Genome Institute is performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

REFERENCES

1. **Blumer-Schuette, S. E., I. Kataeva, J. Westpheling, M. W. Adams, and R. M. Kelly.** 2008. Extremely thermophilic microorganisms for biomass conversion: status and prospects. Curr. Opin. Biotechnol. **19:**210-217.
2. **Clausen, A., M. J. Mikkelsen, I. Schroder, and B. K. Ahring.** 2004. Cloning, sequencing, and sequence analysis of two novel plasmids from the thermophilic anaerobic bacterium *Anaerocellum thermophilum*. Plasmid **52:**131-138.
3. **Dam, P., V. Olman, K. Harris, Z. Su, and Y. Xu.** 2007. Operon prediction using both genome-specific and general genomic information. Nucleic Acids Res. **35:**288-298.
4. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. **8:**186-194.
5. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. **8:**175-185.
6. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195-202.
7. **Han, C. S., and P. Chain.** 2006. Finishing repeat regions automatically with Dupfinisher, p. 141-146. *In* H. R. Arabnia and H. Valafar (ed.), Proceedings of the 2006 International Conference on Bioinformatics and Computational Biology. CSREA Press, Las Vegas, NV.

8. **Nelson, K. E., R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser.** 1999. Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. Nature **399:**323-329.

9. **Rainey, F.A., N. L. Ward, H. W. Morgan, R. Toalster, and E. Stackebrandt.** 1993. Phylogenetic analysis of anaerobic thermophilic bacteria: aid for their reclassification. J Bacteriol. **175:** 4772-4779.

10. **Svetlichnyi, V. A., T. P. Svetlichnaya, N. A. Chernykh, and G. A. Zavarzin.** 1990. *Anaerocellum thermophilum* gen. nov., sp. nov., an extremely thermophilic cellulolytic eubacterium Isolated from hot-springs in the valley of Geysers. Microbiology **59:**598-604.

11. **van de Werken, H. J., M. R. Verhaart, A. L. VanFossen, K. Willquist, D. L. Lewis, J. D. Nichols, H. P. Goorissen, E. F. Mongodin, K. E. Nelson, E. W. van Niel, A. J. Stams, D. E. Ward, W. M. de Vos, J. van der Oost, R. M. Kelly, and S. W. Kengen.** 2008. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Appl. Environ. Microbiol. **74:**6720-6729.

12. **Yang, S.-J., I. Kataeva, S. D. Hamilton-Brehm, N. L. Engle, T. J. Tschaplinski, C. Doeppke, M. Davis, J. Westpheling, and M. W. W. Adams.**

2009. Efficient degradation of lignocellulosic plant biomass without pretreatment by the thermophilic anaerobe, *Anaerocellum thermophilum* DSM 6725. Appl. Environ. Microbiol. **in revision**

13. **Zverlov, V., S. Mahr, K. Riedel, and K. Bronnenmeier.** 1998. Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile '*Anaerocellum thermophilum*' with separate glycosyl hydrolase family 9 and 48 catalytic domains. Microbiology **144:**457-465.