

GENOME ANNOUNCEMENTS

Complete Genome Sequence of the Cellulolytic Thermophile *Caldicellulosiruptor obsidiansis* OB47^{TV}

James G. Elkins,^{1,2*} Adriane Lochner,^{1,2} Scott D. Hamilton-Brehm,^{1,2} Karen Walston Davenport,³
Mircea Podar,^{1,2} Steven D. Brown,^{1,2} Miriam L. Land,² Loren J. Hauser,^{1,2} Dawn M. Klingeman,^{1,2}
Babu Raman,^{1,2} Lynne A. Goodwin,³ Roxanne Tapia,³ Linda J. Meincke,³ J. Chris Detter,⁴
David C. Bruce,³ Cliff S. Han,³ Anthony V. Palumbo,^{1,2} Robert W. Cottingham,^{1,2}
Martin Keller,^{1,2} and David E. Graham²

BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831¹; Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831²; Los Alamos National Laboratory, Los Alamos, New Mexico 87545³; and DOE Joint Genome Institute, Walnut Creek, California 94598⁴

Received 12 August 2010/Accepted 10 September 2010

***Caldicellulosiruptor obsidiansis* OB47^T (ATCC BAA-2073, JCM 16842) is an extremely thermophilic, anaerobic bacterium capable of hydrolyzing plant-derived polymers through the expression of multidomain/multifunctional hydrolases. The complete genome sequence reveals a diverse set of carbohydrate-active enzymes and provides further insight into lignocellulosic biomass hydrolysis at high temperatures.**

Members of the genus *Caldicellulosiruptor* within the order *Clostridiales* can solubilize cellulose at extremely thermophilic growth temperatures (65 to 80°C). *Caldicellulosiruptor obsidiansis* OB47^T was isolated from Obsidian Pool, Yellowstone National Park, in enrichment cultures containing dilute acid-pretreated switchgrass as the primary carbon and energy source for cultivation (5). High-temperature saccharification can promote higher hydrolysis rates while reducing cooling costs following biomass pretreatment and suppressing contamination in reactors (9). Given the organism's rapid growth on cellulosic substrates and ability to use a wide range of plant-derived sugars, a complete genome sequence was determined using a sequencing-by-synthesis approach.

The genome of *C. obsidiansis* OB47^T was sequenced by the U.S. Department of Energy (DOE) Joint Genome Institute (JGI) using a combination of Illumina (1) and 454 technologies (8). All of the general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov/>. Illumina sequencing data were assembled with VELVET (10), and the consensus sequences were shredded into 1.5-kbp overlapped fake reads and assembled together with the 454 data. The initial Newbler assembly contained 64 contigs in two scaffolds. The initial 454 assembly was converted into a Phrap assembly by making fake reads from the consensus and collecting the read pairs in the 454 paired-end library. The Phred/Phrap/Consed software package was used for sequence assembly and quality assessment (2–4) in the following finishing process. Illumina data were used to correct potential

base errors and increase consensus quality using the Polisher software developed at the JGI (Alla Lapidus, unpublished data). After the shotgun stage, reads were assembled with parallel Phrap (High Performance Software, LLC). Possible misassemblies were corrected with gapResolution (Cliff Han, unpublished data), Dupfinisher (6), or sequencing of cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR, and by Bubble PCR primer walks. A total of 773 additional reactions and seven shatter libraries were necessary to close gaps and to raise the quality of the finished sequence. The genome was annotated at Oak Ridge National Laboratory using the automated annotation pipeline, which is driven by the gene prediction algorithm Prodigal (7). Annotation quality was verified by the JGI.

Although many well-characterized bacteria and fungi can use cellulose, *C. obsidiansis* was selected and isolated specifically for its ability to deconstruct potential bioenergy feedstocks (e.g., pretreated switchgrass or *Populus* sp.). Through high-throughput sequencing of novel strains relevant to different aspects of renewable energy production, genome-enabled technologies can be used to discover important cellular properties (such as the secretion of hydrolytic enzymes). Making the genome sequence of *C. obsidiansis* OB47^T available will allow comprehensive comparisons with other members of the genus and enable further investigation into the mechanisms employed by microorganisms to solubilize lignocellulosic materials at elevated temperatures.

Nucleotide sequence accession number. The final annotated genome of *C. obsidiansis* OB47^T has been deposited in GenBank under accession number CP002164.

* Corresponding author. Mailing address: BioEnergy Science Center, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831. Phone: (865) 241-1386. Fax: (865) 576-8646. E-mail: elkinsjg@ornl.gov.

[†] Published ahead of print on 17 September 2010.

We thank Tatiana A. Vishnivetskaya and Marilyn K. Kerley for sequencing and analysis of the 16S rRNA genes from *C. obsidiansis*.

The BioEnergy Science Center is a U.S. DOE Bioenergy Research Center supported by the Office of Biological and Environmental Re-

search in the DOE Office of Science. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. DOE under contract DE-AC05-00OR22725. Work at the JGI is performed under the auspices of the U.S. DOE Office of Science Biological and Environmental Research Program and by the University of California Lawrence Berkeley National Laboratory under contract DE-AC02-05CH11231, by the Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344, and by the Los Alamos National Laboratory under contract DE-AC02-06NA25396.

REFERENCES

1. Bennett, S. 2004. Solexa Ltd. *Pharmacogenomics* **5**:433–438.
2. Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**:186–194.
3. 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.
4. Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* **8**:195–202.
5. Hamilton-Brehm, S. D., J. J. Mosher, T. Vishnivetskaya, M. Podar, S. Carroll, S. Allman, T. J. Phelps, M. Keller, and J. G. Elkins. 2010. *Caldicellulosiruptor obsidiansis* sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park. *Appl. Environ. Microbiol.* **76**:1014–1020.
6. Han, C. S., and P. Chain. 2006. Finishing repetitive regions automatically with Dupfinisher. *In* Proceedings of the 2006 International Conference on Bioinformatics and Computational Biology, Las Vegas, NV. <http://ww1.ucmss.com/books/LFS/CSREA2006/BIC3878.pdf>.
7. Hyatt, D., G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**:119.
8. Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. T. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. G. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.
9. Wiegel, J. 1980. Formation of ethanol by bacteria—a pledge for the use of extreme thermophilic anaerobic-bacteria in industrial ethanol fermentation processes. *Experientia* **36**:1434–1446.
10. Zerbino, D. R., and E. Birney. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.* **18**:821–829.