Complete genome sequences for the anaerobic, extremely thermophilic plant biomass-degrading bacteria *Caldicellulosiruptor hydrothermalis*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor kronotskyensis*, *Caldicellulosiruptor owensensis*, and *Caldicellulosiruptor lactoaceticus*

Sara E. Blumer-Schuette¹, Inci Ozdemir¹, Dhaval Mistry¹, Susan Lucas², Alla Lapidus³, Jan-Fang Cheng², Lynne A. Goodwin⁵, Samuel Pitluck², Miriam L. Land⁴, Loren J. Hauser⁴, Tanja Woyke², Natalia Mikhailova², Amrita Pati², Nikos C. Kyrpides², Natalia Ivanova², John C. Detter⁵, Karen Walston-Davenport⁵, Shunsheng Han⁵, Michael W.W. Adams⁶, and Robert M. Kelly¹

¹Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695-7905

²DOE Joint Genomes Institute, Walnut Creek, CA 94598

³Fox Chase Cancer Center, Philadelphia, PA 19111-2497

⁴BioEnergy Science Center and Biosciences Division, Oak Ridge National Laboratory, Oak Ridge Tennessee 37831

⁵Los Alamos National Laboratory, Bioscience Division B-6 Genome Science, Joint Genome Institute, Los Alamos, New Mexico 87545

⁶Dept. of Biochemistry and Molecular Biology, University of Georgia, Athens, GA 30602-7229

Submitted to: *Journal of Bacteriology* (December, 2010)

Keywords: *Caldicellulosiruptor*, extreme thermophile, plant biomass, glycoside hydrolases

*Address correspondence to: Robert M. Kelly
Department of Chemical and Biomolecular Engineering
North Carolina State University
EB-1, 911 Partners Way
Raleigh, NC 27695-7905
Phone: (919) 515-6396 / Fax:(919) 515-3465
Email: rmkelly@eos.ncsu.edu
The genus *Caldicellulosiruptor* contains the most thermophilic, plant biomass-degrading bacteria isolated to date. Previously, genome sequences from three cellulolytic members of this genus were reported (*C. saccharolyticus*, *C. bescii*, and *C. obsidiansis*). To further explore the physiological and biochemical basis for polysaccharide degradation within this genus, five additional genomes were sequenced: *C. hydrothermalis*, *C. kristjanssonii*, *C. kronotskyensis*, *C. lactoaceticus*, and *C. owensensis*. Taken together, the seven completed and one draft-phase *Caldicellulosiruptor* genomes suggest that, while central metabolism is highly conserved, significant differences in glycoside hydrolase inventories and numbers of carbohydrate transporters exist, which likely relates to variability observed in plant biomass degradation capacity.

Members of the genus *Caldicellulosiruptor* are asporogenic, plant biomass degrading, hydrogen-generating members of the order *Clostridiales* (13, 17). The genus is globally distributed; species have been isolated from terrestrial geothermal hot springs in Russia (15, 18, 20), Iceland (3, 16), Yellowstone National Park in the United States (9) and New Zealand (17) and, in one case, from solar-heated mud flats in Owens Lake, California (11). With optimal growth temperatures ranging from 70-78°C, the genus *Caldicellulosiruptor* contains the most thermophilic microorganisms capable of biological cellulose hydrolysis known. While 16S ribosomal RNA phylogeny indicates a very close relationship within the eight species studied thus far (94.8% to 99.4% identity), microbiological analysis indicated that the genus is more physiologically divergent than previously thought (2). Three complete genome sequences are currently available for this genus, *C. saccharolyticus* (19), *C. bescii* (12), and *C. obsidiansis* (5), all of which are capable of crystalline cellulose hydrolysis. To further probe the plant biomass
degrading capacity among *Caldicellulosiruptor* species, genome sequences of five additional members of this genus, including some hemicellulolytic but less cellulolytic members, were completed: *C. hydrothermalis*, *C. kristjanssonii*, *C. kronotskyensis*, *C. owensensis* and *C. lactoaceticus* (draft-phase). This also provides additional geographical diversity for examining *Caldicellulosiruptor* physiology.

All members of the *Caldicellulosiruptor* genus have similar sized chromosomes, approximately 2.4 to 2.97 Mb. Their genomes are also A+T rich, ranging from 35-36% G+C. Similar to *C. bescii* (4, 12), *C. kristjanssonii* also possesses an extra-chromosomal element, the 15.9 kb plasmid, pCALKR01. There are no similarities, however, between pATHE01/pATHE02 and pCALKR01, suggesting that there are no ubiquitous *Caldicellulosiruptor* plasmids. The other six genomes do not contain any extra-chromosomal elements, although *C. hydrothermalis* possesses an inverted region that could exist outside of the chromosome.

**Sequencing strategy.** All general aspects of library construction and sequencing can be found at http://www.jgi.doe.gov/. The genomes of *Caldicellulosiruptor* species were sequenced at the U.S. Department of Energy Joint Genome Institute (JGI) using a combination of Illumina (1) and 454 technologies (14) similar to the sequencing strategy of *C. obsidiansis* (5). Sequencing data was assembled with either VELVET (21), or by being converted into a phrap assembly. The Phred/Phrap/Consed software package was used for sequence assembly and quality assessment (6-8) in the finishing process. Illumina data were used to correct base errors and increase consensus quality using Polisher software (Alla Lapidus, unpublished). After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC), possible mis-assemblies were corrected with gapResolution (Cliff Han, unpublished), Dupfinisher (10), or sequencing cloned bridging PCR fragments. Gaps between contigs were
closed by editing in Consed, by PCR and by Bubble PCR primer walks. The creation of additional reactions and shatter libraries were necessary to close gaps and to raise the quality of the finished sequences. Final assemblies are based on a minimum of 30x coverage of the genome.

**Nucleotide Accession Numbers.** Genbank accession numbers for *Caldicellulosiruptor* genomes announced here are as follows: *C. hydrothermalis* (CP002219), *C. kristjanssonii* (CP002326), *C. kronotskyensis* (CP002330), *C. lactoaceticus* (AEKD0000000) and *C. owensensis* (CP002216).

**ACKNOWLEDGMENTS**

This research was supported by a grant (DE-PS02-06ER64304) from the Bioenergy Science Center (BESC), Oak Ridge National Laboratory, a U.S. Department of Energy (DOE) 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 is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.
REFERENCES


