Complete Genome Sequences of *Caldicellulosiruptor* sp. Strain Rt8.B8, *Caldicellulosiruptor* sp. Strain Wai35.B1, and “Thermoanaerobacter cellulolyticus”

Laura L. Lee,a,b Javier A. Izquierdo,a,b Sara E. Blumer-Schuette,a,b Jeffrey V. Zurawski,a,b Jonathan M. Conway,a,b Robert W. Cottingham,a Marcel Huntemann,a Alex Copeland,a I-Min A. Chen,a Nikos Kyrpides,a Victor Markowitz,a Krishnaveeni Palaniappan,b Natalia Ivanova,b Natalia Mikhailova,b Galina Ovchinnikova,b Evan Andersenc, Dimitrios Stamatis,c T.B.K. Reddy,a Nicole Shapiro,a Henrik P. Nordberg,a Michael N. Cantor,a Susan X. Hu,a Tanja Woyke,c Robert M. Kellya,b

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina, USA; BioEnergy Science Center and Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA; DOE Joint Genome Institute, Walnut Creek, California, USA.

* Present address: Javier A. Izquierdo, Department of Biology, Hofstra University, Hempstead, New York, USA; Sara E. Blumer-Schuette, Oakland University, Rochester, Michigan, USA.

The genus *Caldicellulosiruptor* contains extremely thermophilic, cellulolytic bacteria capable of lignocellulose deconstruction. Currently, complete genome sequences for eleven *Caldicellulosiruptor* species are available. Here, we report genome sequences for three additional *Caldicellulosiruptor* species: Rt8.B8 DSM 8990 (New Zealand), Wai35.B1 DSM 8977 (New Zealand), and “Thermoanaerobacter cellulolyticus” strain NA10 DSM 8991 (Japan).

The draft genomes of *Caldicellulosiruptor* contain 12 contigs in 12 scaffolds, totaling 2,514,985 bp, with an input read coverage of 88.9×. The final draft assemblies of all three species were classified into “Thermoanaerobacter cellulolyticus” strain NA10 DSM 8991 (9, 10), although its genome sequence indicates that it belongs to the genus *Caldicellulosiruptor*.

The draft genomes of *C. cellulolyticus* NA10, *Caldicellulosiruptor* sp. strain Rt8.B8, and *Caldicellulosiruptor* sp. strain Wai35.B1 were produced by constructing Pacific Biosciences (PacBio) SMRTBELL libraries, sequencing on the PacBio RS platform (11), and correcting errors on the Illumina platform for each microorganism at the DOE Joint Genome Institute (JGI). This generated 81,181, 150,202, and 166,964 filtered subreads totaling 349.1, 591.3, and 670.4 Mbp for *C. cellulolyticus* NA10, *Caldicellulosiruptor* sp. strain Rt8.B8, and *Caldicellulosiruptor* sp. strain Wai35.B1, respectively. All raw reads were accumulated via Hgap (12) and classified into genes using Prodigal (13) along with GenePRIMP (14). The predicted coding regions were then checked against the genomes within the Integrated Microbial Genomes (IMG) platform (http://img.jgi.doe.gov). Specific genes, such as tRNAs, rRNAs, and other noncoding RNAs, were identified by searching the genome with the tRNAScanSE tool (15), SILVA rRNA gene models (16), and INFERNAL (http://infernal.janelia.org).

The final draft assembly of *T. cellulolyticus* NA10 DSM 8991 contained 12 contigs in 12 scaffolds, totaling 2,514,985 bp, with an input read coverage of 88.9×. The final draft assembly of *Caldicellulosiruptor* sp. strain Rt8.B8 contained 2 contigs in 2 scaffolds, totaling 2,488,483 bp in size, with an input reads coverage of 165.6×. Lastly, the final draft assembly of *Caldicellulosiruptor* sp. strain Wai35.B1 contained 1 contig in 1 scaffold, totaling 2,834,482 bp in size, with an input reads coverage of 175.6×. The G+C content was 35.39%, 36.49%, and 35.78% for *T. cellulolyticus* NA10, *Caldicellulosiruptor* sp. strain Rt8.B8, and *Caldicellulosiruptor* sp. strain Wai35.B1, respectively. We expect that the novel features we are identifying in these genomes will further contribute to our understanding of the metabolic diversity of lignocellulolytic capabilities within the *Caldicellulosiruptor* genus.

**Nucleotide sequence accession numbers.** These whole-genome shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accession numbers LACN00000000, LACO00000000, and LACM00000000 for *T. cellulolyticus* NA10 DSM 8991, *Caldicellulo-
losirruptor sp. strain Rt8.B8 DSM 8990, and Caldicellulosiruptor sp. strain Wai35.B1 DSM 8977, respectively. The versions described in this paper are versions LACN01000000, LACO01000000, and LACM01000000.

ACKNOWLEDGMENTS

The BioEnergy Science Center (BESC) is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The work was conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. L.L.L. acknowledges support from an NIH Biotechnology Traineeship (NIH T32 GM008776-11) and J.M.C. acknowledges support from an NIH Biotechnology Traineeship (NIH T32 GM008776-11) and J.M.C. acknowledges support from a U.S. Department of Education GAANN fellowship (P200A100004-12).

REFERENCES