

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

Laura L. Lee,<sup>a,b</sup> Javier A. Izquierdo,<sup>a,b\*</sup> Sara E. Blumer-Schuette,<sup>a,b\*</sup> Jeffrey V. Zurawski,<sup>a,b</sup> Jonathan M. Conway,<sup>a,b</sup> Robert W. Cottingham,<sup>b</sup> Marcel Huntemann,<sup>c</sup> Alex Copeland,<sup>c</sup> I-Min A. Chen,<sup>c</sup> Nikos Kyrpides,<sup>c</sup> Victor Markowitz,<sup>c</sup> Krishnaveni Palaniappan,<sup>c</sup> Natalia Ivanova,<sup>c</sup> Natalia Mikhailova,<sup>c</sup> Galina Ovchinnikova,<sup>c</sup> Evan Andersen,<sup>c</sup> Amrita Pati,<sup>c</sup> Dimitrios Stamatis,<sup>c</sup> T.B.K. Reddy,<sup>c</sup> Nicole Shapiro,<sup>c</sup> Henrik P. Nordberg,<sup>c</sup> Michael N. Cantor,<sup>c</sup> Susan X. Hua,<sup>c</sup> Tanja Woyke,<sup>c</sup> Robert M. Kelly<sup>a,b</sup>

Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, North Carolina, USA<sup>a</sup>; BioEnergy Science Center and Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA<sup>b</sup>; DOE Joint Genome Institute, Walnut Creek, California, USA<sup>c</sup>

\* 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).**

Received 29 March 2015 Accepted 7 April 2015 Published 14 May 2015

**Citation** Lee LL, Izquierdo JA, Blumer-Schuette SE, Zurawski JV, Conway JM, Cottingham RW, Huntemann M, Copeland A, Chen I-MA, Kyrpides N, Markowitz V, Palaniappan K, Ivanova N, Mikhailova N, Ovchinnikova G, Andersen E, Pati A, Stamatis D, Reddy TBK, Shapiro N, Nordberg HP, Cantor MN, Hua SX, Woyke T, Kelly RM. 2015. Complete genome sequences of *Caldicellulosiruptor* sp. strain Rt8.B8, *Caldicellulosiruptor* sp. strain Wai35.B1, and “*Thermoanaerobacter cellulolyticus*.” *Genome Announc* 3(3):e00440-15. doi: 10.1128/genomeA.00440-15.

**Copyright** © 2015 Lee et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/4.0/).

Address correspondence to Robert M. Kelly, [rmkelly@ncsu.edu](mailto:rmkelly@ncsu.edu).

Lignocellulose-degrading microorganisms are of considerable interest for their use in the production of fuels and chemicals from renewable feedstocks. The genus *Caldicellulosiruptor* contains the most thermophilic, plant biomass-degrading bacteria isolated thus far, with optimal growth temperatures between 70 and 78°C (1). To date, complete genome sequences are available for 11 *Caldicellulosiruptor* species (1–7). From this information, it was determined that the *Caldicellulosiruptor* pan genome is still open, indicating that there is additional diversity within the genus, potentially involving yet-to-be discovered biomass-degrading enzymes and pathways. In light of this, three *Caldicellulosiruptor* isolates previously isolated from terrestrial hot springs were genome sequenced. Two were isolated from sites in New Zealand: *Caldicellulosiruptor* sp. strain Rt8.B8 DSM 8990 and strain Wai35.B1 DSM 8977 from Rotorua and Waimangu, respectively (8). The third, from Nozawa Hot Spring, Nagano Prefecture, Japan, was originally classified as “*Thermoanaerobacter cellulolyticus*” strain NA10 DSM 8991 (9, 10), although its genome sequence indicates that it belongs to the genus *Caldicellulosiruptor*.

The draft genomes of *T. 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 *T. 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 National Center for Biotechnology Information (NCBI), UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases to annotate 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 read 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 read 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 [LACN000000000](https://www.ncbi.nlm.nih.gov/nuccore/LACN000000000), [LACO000000000](https://www.ncbi.nlm.nih.gov/nuccore/LACO000000000), and [LACM000000000](https://www.ncbi.nlm.nih.gov/nuccore/LACM000000000) for *T. cellulolyticus* NA10 DSM 8991, *Caldicellu-*

*losiruptor* 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 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 a U.S. Department of Education GAANN fellowship (P200A100004-12).

## REFERENCES

- Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MW, Kelly RM. 2014. Thermophilic lignocellulose deconstruction. *FEMS Microbiol Rev* 38:393–448. <http://dx.doi.org/10.1111/1574-6976.12044>.
- Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin Y, Xu Y, Kataeva I, Poole FL, 2nd, Adams MW, Hamilton-Brehm SD, Elkins JG, Larimer FW, Land ML, Hauser LJ, Cottingham RW, Hettich RL, Kelly RM. 2012. *Caldicellulosiruptor* core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. *J Bacteriol* 194:4015–4028. <http://dx.doi.org/10.1128/JB.00266-12>.
- Blumer-Schuette SE, Ozdemir I, Mistry D, Lucas S, Lapidus A, Cheng JF, Goodwin LA, Pitluck S, Land ML, Hauser LJ, Woyke T, Mikhailova N, Pati A, Kyrpides NC, Ivanova N, Detter JC, Walston-Davenport K, Han S, Adams MW, Kelly RM. 2011. Complete genome sequences for the anaerobic, extremely thermophilic plant biomass-degrading bacteria *Caldicellulosiruptor hydrothermalis*, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor kronotskyensis*, *Caldicellulosiruptor owensensis*, and *Caldicellulosiruptor lactoaceticus*. *J Bacteriol* 193:1483–1484. <http://dx.doi.org/10.1128/JB.01515-10>.
- Elkins JG, Lochner A, Hamilton-Brehm SD, Davenport KW, Podar M, Brown SD, Land ML, Hauser LJ, Klingeman DM, Raman B, Goodwin LA, Tapia R, Meincke LJ, Detter JC, Bruce DC, Han CS, Palumbo AV, Cottingham RW, Keller M, Graham DE. 2010. Complete genome sequence of the cellulolytic thermophile *Caldicellulosiruptor obsidiansis* OB47<sup>T</sup>. *J Bacteriol* 192:6099–6100. <http://dx.doi.org/10.1128/JB.00950-10>.
- Onyenwoke RU, Lee YJ, Dabrowski S, Ahring BK, Wiegel J. 2006. Reclassification of *Thermoanaerobium acetigenum* as *Caldicellulosiruptor acetigenus* comb. nov. and emendation of the genus description. *Int J Syst Evol Microbiol* 56:1391–1395. <http://dx.doi.org/10.1099/ijs.0.63723-0>.
- Van de Werken HJ, Verhaart MR, VanFossen AL, Willquist K, Lewis DL, Nichols JD, Goorissen HP, Mongodin EF, Nelson KE, van Niel EW, Stams AJ, Ward DE, de Vos WM, van der Oost J, Kelly RM, Kengen SW. 2008. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Appl Environ Microbiol* 74:6720–6729. <http://dx.doi.org/10.1128/AEM.00968-08>.
- Yang SJ, Kataeva I, Wiegel J, Yin Y, Dam P, Xu Y, Westpheling J, Adams MW. 2010. Classification of ‘*Anaerocellum thermophilum*’ strain DSM 6725 as *Caldicellulosiruptor bescii* sp. nov. *Int J Syst Evol Microbiol* 60:2011–2015. <http://dx.doi.org/10.1099/ijs.0.017731-0>.
- Rainey FA, Janssen PH, Morgan HW, Stackebrandt E. 1993–1994. A biphasic approach to the determination of the phenotypic and genotypic diversity of some anaerobic, cellulolytic, thermophilic, rod-shaped bacteria. *Antonie Van Leeuwenhoek* 64:341–355. <http://dx.doi.org/10.1007/BF00873092>.
- Taya M, Hinoki H, Kobayashi T. 1985. Tungsten requirement of an extremely thermophilic, cellulolytic anaerobe (strain NA10). *Agric Biol Chem* 49:2513–2515. <http://dx.doi.org/10.1271/abb1961.49.2513>.
- Taya M, Hinoki H, Yagi T, Kobayashi T. 1988. Isolation and characterization of an extremely thermophilic, cellulolytic, anaerobic bacterium. *Appl Microbiol Biotechnol* 29:474–479. <http://dx.doi.org/10.1007/BF00269071>.
- Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S. 2009. Real-time DNA sequencing from single polymerase molecules. *Science* 323:133–138. <http://dx.doi.org/10.1126/science.1162986>.
- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <http://dx.doi.org/10.1038/nmeth.2474>.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
- Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 7:455–457. <http://dx.doi.org/10.1038/nmeth.1457>.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35:7188–7196. <http://dx.doi.org/10.1093/nar/gkm864>.