

## Sequencing of Multiple Clostridial Genomes Related to Biomass Conversion and Biofuel Production<sup>∇</sup>

Christopher L. Hemme,<sup>1,2</sup> Housna Mouttaki,<sup>1,2,3</sup> Yong-Jin Lee,<sup>1,2</sup> Gengxin Zhang,<sup>14</sup> Lynne Goodwin,<sup>5</sup> Susan Lucas,<sup>4</sup> Alex Copeland,<sup>4</sup> Alla Lapidus,<sup>4</sup> Tijana Glavina del Rio,<sup>4</sup> Hope Tice,<sup>4</sup> Elizabeth Saunders,<sup>5</sup> Thomas Bretin,<sup>5,6</sup> John C. Detter,<sup>5</sup> Cliff S. Han,<sup>5</sup> Sam Pitluck,<sup>4</sup> Miriam L. Land,<sup>6</sup> Loren J. Hauser,<sup>6</sup> Nikos Kyrpides,<sup>4</sup> Natalia Mikhailova,<sup>4</sup> Zhili He,<sup>1,2</sup> Liyou Wu,<sup>1,2</sup> Joy D. Van Nostrand,<sup>1,2</sup> Bernard Henrissat,<sup>7</sup> Qiang He,<sup>8</sup> Paul A. Lawson,<sup>1</sup> Ralph S. Tanner,<sup>1</sup> Lee R. Lynd,<sup>9</sup> Juergen Wiegel,<sup>10</sup> Matthew W. Fields,<sup>11</sup> Adam P. Arkin,<sup>12</sup> Christopher W. Schadt,<sup>13</sup> Bradley S. Stevenson,<sup>1</sup> Michael J. McInerney,<sup>1</sup> Yunfeng Yang,<sup>13</sup> Hailiang Dong,<sup>14</sup> Defeng Xing,<sup>15</sup> Nanqi Ren,<sup>15</sup> Aijie Wang,<sup>15</sup> Raymond L. Huhnke,<sup>16</sup> Jonathan R. Mielenz,<sup>17</sup> Shi-You Ding,<sup>18</sup> Michael E. Himmel,<sup>18</sup> Safiyh Taghavi,<sup>19</sup> Daniël van der Lelie,<sup>19</sup> Edward M. Rubin,<sup>4</sup> and Jizhong Zhou<sup>1,2\*</sup>

Department of Botany and Microbiology, University of Oklahoma, Norman, Oklahoma<sup>1</sup>; Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma<sup>2</sup>; German Research Center for Environment and Health, Neuherberg, Germany<sup>3</sup>; Joint Genome Institute Production Genomics Facility, Walnut Creek, California<sup>4</sup>; Joint Genome Institute, Los Alamos National Laboratory, Los Alamos, New Mexico<sup>5</sup>; Computational Biology and Bioinformatics Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee<sup>6</sup>; CNRS, Universit s Aix-Marseille, 13628 Aix-en-Provence, France<sup>7</sup>; Department of Civil and Environmental Engineering, University of Tennessee—Knoxville, Knoxville, Tennessee<sup>8</sup>; Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire<sup>9</sup>; Department of Microbiology, University of Georgia, Athens, Georgia<sup>10</sup>; Department of Microbiology, Montana State University, Bozeman, Montana<sup>11</sup>; University of California—Berkeley, Berkeley, California<sup>12</sup>; Molecular Microbial Ecology Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee<sup>13</sup>; Miami University, Oxford, Ohio<sup>14</sup>; State Key Lab of Urban Water Resource and Environment, Harbin 160090, People's Republic of China<sup>15</sup>; Oklahoma State University, Stillwater, Oklahoma<sup>16</sup>; Bioconversion and Science Technology Group, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee<sup>17</sup>; National Renewable Energy Laboratory, Golden, Colorado<sup>18</sup>; and Brookhaven National Laboratory, Upton, New York<sup>19</sup>

Received 7 September 2010/Accepted 20 September 2010

**Modern methods to develop microbe-based biomass conversion processes require a system-level understanding of the microbes involved. Clostridium species have long been recognized as ideal candidates for processes involving biomass conversion and production of various biofuels and other industrial products. To expand the knowledge base for clostridial species relevant to current biofuel production efforts, we have sequenced the genomes of 20 species spanning multiple genera. The majority of species sequenced fall within the class III cellulosome-encoding *Clostridium* and the class V saccharolytic *Thermoanaerobacteraceae*. Species were chosen based on representation in the experimental literature as model organisms, ability to degrade cellulosic biomass either by free enzymes or by cellulosomes, ability to rapidly ferment hexose and pentose sugars to ethanol, and ability to ferment synthesis gas to ethanol. The sequenced strains significantly increase the number of noncommensal/nonpathogenic clostridial species and provide a key foundation for future studies of biomass conversion, cellulosome composition, and clostridial systems biology.**

Clostridial genomes were sequenced using a combination of Sanger (3× coverage, 8 kb, pMCL200), 454 (20× coverage), and Solexa methods. Standard sequencing protocols are listed on the Joint Genome Institute (JGI) website (<http://www.jgi.doe.gov>

/sequencing/protocols/prots\_production.html). Sanger and 454 reads were assembled as previously described (5). Automatic annotations were conducted for all draft genomes using the JGI-Oak Ridge National Laboratory (ORNL) annotation pipeline, and all draft genomes and annotations were loaded into the JGI Integrated Microbial Resource (IMG) for analysis (11). Due to difficulties in assembling and finishing low-GC, high-repeat genomes, many of the genomes targeted for sequencing could not be finished and are presented as a high-

\* Corresponding author. Mailing address: Stephenson Research and Technology Center, University of Oklahoma, Norman, OK 73019. Phone: (405) 325-6073. Fax: (405) 325-7552. E-mail: jzhou@ou.edu.

<sup>∇</sup> Published ahead of print on 1 October 2010.

TABLE 1. Clostridium genome sequencing projects related to biomass conversion and biofuels production

Organism name	Reference(s)	Genome analysis					Accession no. <sup>c</sup>
		Seq. status <sup>a</sup>	Size (Mb)	No. of contigs	% GC	CDS <sup>b</sup>	
<i>Clostridium sensu stricto</i>							
<i>Clostridium cellulovorans</i> 743B	17	F	5.2	1	31.2	4,256	CP002160
<i>Clostridium carboxidivorans</i> P7	3, 8	PD	5.5	50	29.7	5,462	ACVI00000000
<i>Clostridium ragsdalei</i> P11	3	P					
<i>Ruminococcaceae</i>							
<i>Acetivibrio cellulolyticus</i> CD2	14	D	6.1	193	35.5	5,146	AEDB00000000
<i>Clostridium cellulolyticum</i> H10	15	F	4.1	1	37.0	3,390	CP001348
<i>Clostridium papyrosolvens</i> DSM 2782	10	PD	4.8	31	36.9	4,425	ACXX00000000
<i>Clostridium thermocellum</i> JW20 DSM 4150	2	PD	3.8	26	39.0	3,077	ABVG00000000
<i>Clostridium thermocellum</i> LQR1 DSM 2360	12	PD	3.5	23	39.1	2,914	ACVX00000000
<i>Thermoanaerobacteraceae</i>							
<i>Thermoanaerobacter wiegeli</i> Rt8.B1	1	D	2.7	169	34.2	2,906	Pending
<i>Thermoanaerobacter brockii</i> sp. <i>finnii</i> Ako-1	7, 21	PD	2.2	85	34.3	2,279	ACQZ00000000
<i>Thermoanaerobacter</i> sp. CCSD1	22	PD	2.2	18	34.3	2,146	ACXY00000000
<i>Thermoanaerobacter</i> sp. X513	16	F	2.3	1	34.5	2,330	ACPF00000000
<i>Thermoanaerobacter</i> sp. X561	16	PD	2.4	8	34.5	2,332	ACXP00000000
<i>Thermoanaerobacter pseudethanolicus</i> 39E	13	F	2.4	1	34.0	2,243	CP000924
<i>Thermoanaerobacter</i> sp. X514	16	F	2.5	1	34.0	2,349	CP000923
<i>Thermoanaerobacter italicus</i> Ab9	4	F	2.4	1	34.1	2,271	CP001936
<i>Thermoanaerobacter mathranii</i> subsp. <i>mathranii</i> A3 DSM 11426	6	F	2.3	1	34.3	2,161	CP002032
<i>Thermoanaerobacter ethanolicus</i> JW200	19	D	2.2	376	34.0	2,423	ACXY00000000
<i>Thermoanaerobacter siderophilus</i> L-64	18	P					
<i>Thermoanaerobacterium thermosaccharolyticum</i> DSM 571	7	F	2.7	1	34.1	2,161	CP002171
<i>Thermoanaerobacterium xylanolyticum</i> LX11	7	D	2.5	91	34.8	2,509	Pending
Miscellaneous							
<i>Clostridium saccharolyticum</i> DSM 2544	7	F	4.6	1	45.0	4,160	CP002109
<i>Ethanoligenens harbinense</i> YUAN-3	20	D	3.0	3	55.5	2,787	ADJQ00000000

<sup>a</sup> Sequencing status: F, finished; PD, permanent high-quality draft; D, draft (ongoing); P, sequence pending.

<sup>b</sup> No. of annotated protein coding sequences.

<sup>c</sup> GenBank accession number.

quality permanent draft. Sequences available at the time of this analysis are listed in Table 1 and are categorized based on the latest Bergey's taxonomy (9).

The biomass-converting clostridium sequencing project was initiated by the Clostridium Sequencing Consortium (see author list) through the Joint Genome Institute. The genome sequencing work was performed under the auspices of the U.S. 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, Los Alamos National Laboratory, under contract no. DE-AC02-06NA25396, and Oak Ridge National Laboratory, under contract no. DE-AC05-00OR22725. This work was supported by NSF-EPSC.R grant 105118300.

#### REFERENCES

- Cook, G. M., F. A. Rainey, B. K. C. Patel, and H. W. Morgan. 1996. Characterization of a new obligately anaerobic thermophile, *Thermoanaerobacter wiegeli* sp. nov. *Int. J. Syst. Bacteriol.* **46**:123–127.
- Freier, D., C. P. Mothershed, and J. Wiegell. 1988. Characterization of *Clostridium thermocellum* JW20. *Appl. Environ. Microbiol.* **54**:204–211.
- Huhnke, R. L., R. S. Lewis, and R. S. Tanner. 27 April 2010. Isolation and characterization of novel clostridial species. U.S. patent 7,704,723.
- Kozianowski, G., F. Canganella, F. A. Rainey, H. Hippe, and G. Antranikian. 1997. Purification and characterization of the most stable pectate-lyases from a newly isolated thermophilic bacterium, *Thermoanaerobacter italicus* sp. nov. *Extremophiles* **1**:171–182.
- Land, M., R. Pukall, B. Abt, M. Göker, M. Rohde, T. G. D. Rio, H. Tice, A. Copeland, J.-F. Cheng, S. Lucas, F. Chen, M. Nolan, D. Bruce, L. Goodwin, S. Pitluck, N. Ivanova, K. Mavromatis, G. Ovchinnikova, A. Pati, A. Chen, K. Palaniappan, L. Hauser, Y.-J. Chang, C. C. Jefferies, E. Saunders, T. Brettn, J. C. Dettler, C. Han, P. Chain, J. Bristow, J. A. Eisen, V. Markowitz, P. Hugenholtz, N. C. Kyrpides, H.-P. Klenk, and A. Lapidus. 2009. Complete genome sequence of *Beutenbergia cavernae* type strain (HKI 0122T). *Stand. Genomic Sci.* doi:10.4056/sjgs.1162.
- Larsen, L., P. Nielsen, and B. K. Ahring. 1997. *Thermoanaerobacter mathranii* sp. nov., an ethanol-producing, extremely thermophilic anaerobic bacterium from a hot spring in Iceland. *Arch. Microbiol.* **168**:114–119.
- Lee, Y.-E., M. K. Jain, C. Lee, S. E. Lowe, and J. G. Zeikus. 1993. Taxonomic distinction of saccharolytic thermophilic anaerobes: description of *Thermoanaerobacterium xylanolyticum* gen. nov., sp. nov., and *Thermoanaerobacterium saccharolyticum* gen. nov., sp. nov.; reclassification of *Thermoanaerobium brockii*, *Clostridium thermosulfurogenes*, and *Clostridium thermohydrosulfuricum* EIO0-69 as *Thermoanaerobacter brockii* comb. nov., *Thermoanaerobacterium thermosulfurogenes* comb. nov., and *Thermoanaerobacter thermohydrosulfuricus* comb. nov., respectively; and transfer of *Clostridium thermohydrosulfuricum* 39E to *Thermoanaerobacter ethanolicus*. *Int. J. Syst. Microbiol.* **43**:41–51.
- Liou, J. S.-C., D. L. Balkwill, G. R. Drake, and R. S. Tanner. 2005. *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *Int. J. Syst. Evol. Microbiol.* **55**:2085–2091.
- Ludwig, W., K. H. Schleifer, and W. B. Whitman. 2008. Revised roadmap to the phylum *Firmicutes*, 2nd ed., vol. 3. Springer, New York, NY.
- Madden, R. H., M. J. Bryder, and N. J. Poole. 1982. Isolation and characterization of an anaerobic, cellulolytic bacterium, *Clostridium papyrosolvens* sp. nov. *Int. J. Syst. Bacteriol.* **32**:87–91.
- Markowitz, V. M., I. M. A. Chen, K. Palaniappan, K. Chu, E. Szeto, Y. Grechkin, A. Ratner, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides. 2010. The integrated microbial genomes system: an

- expanding comparative analysis resource. *Nucleic Acids Res.* **38**:D382–D390.
12. **Ng, T. K., and J. G. Zeikus.** 1981. Comparison of extracellular cellulase activities of *Clostridium thermocellum* LQRI and *Trichoderma reesei* QM9414. *Appl. Environ. Microbiol.* **42**:231–240.
  13. **Onyenwoke, R. U., V. V. Kevbrin, A. M. Lysenko, and J. Wiegel.** 2007. *Thermoanaerobacter pseudethanolicus* sp. nov., a thermophilic heterotrophic anaerobe from Yellowstone National Park. *Int. J. Syst. Evol. Microbiol.* **57**:2191–2193.
  14. **Patel, G. B., A. W. Khan, B. J. Agnew, and J. R. Colvin.** 1980. Isolation and characterization of an anaerobic, cellulolytic microorganism, *Acetivibrio cellulolyticus* gen. nov., sp. nov. *Int. J. Syst. Bacteriol.* **30**:179–185.
  15. **Petitdemange, E., F. Caillet, J. Giallo, and C. Gaudin.** 1984. *Clostridium cellulolyticum* sp. nov., a cellulolytic, mesophilic species from decayed grass. *Int. J. Syst. Bacteriol.* **34**:155–159.
  16. **Roh, Y., S. V. Liu, G. Li, H. Huang, T. J. Phelps, and J. Zhou.** 2002. Isolation and characterization of metal-reducing *Thermoanaerobacter* strains from deep subsurface environments of the Piceance Basin, Colorado. *Appl. Environ. Microbiol.* **68**:6013–6020.
  17. **Sleat, R., R. A. Mah, and R. Robinson.** 1984. Isolation and characterization of an anaerobic, cellulolytic bacterium, *Clostridium cellulovorans* sp. nov. *Appl. Environ. Microbiol.* **48**:88–93.
  18. **Slobodkin, A. I., T. P. Tourova, B. B. Kuznetsov, N. A. Kostrikina, N. A. Chernyh, and E. A. Bonch-Osmolovskaya.** 1999. *Thermoanaerobacter siderophilus* sp. nov., a novel dissimilatory Fe(III)-reducing, anaerobic, thermophilic bacterium. *Int. J. Syst. Bacteriol.* **49**:1471–1478.
  19. **Wiegel, J., and L. G. Ljungdahl.** 1981. *Thermoanaerobacter ethanolicus* gen. nov., spec. nov., a new, extreme thermophilic, anaerobic bacterium. *Arch. Microbiol.* **128**:343–348.
  20. **Xing, D., N. Ren, Q. Li, M. Lin, A. Wang, and L. Zhao.** 2006. *Ethanoligenens harbinense* gen. nov., sp. nov., isolated from molasses wastewater. *Int. J. Syst. Evol. Microbiol.* **56**:755–760.
  21. **Zeikus, J. G., P. W. Hegge, and M. A. Anderson.** 1979. *Thermoanaerobium brockii* gen. nov. and sp. nov., a new chemoorganotrophic, caldoactive, anaerobic bacterium. *Arch. Microbiol.* **122**:41–48.
  22. **Zhang, G., H. Dong, Z. Xu, D. Zhao, and C. Zhang.** 2005. Microbial diversity in ultra-high-pressure rocks and fluids from the Chinese Continental Scientific Drilling Project in China. *Appl. Environ. Microbiol.* **71**:3213–3227.