

Complete Genome Sequence of the Acetogenic Bacterium *Moorella thermoacetica* DSM 2955^T

Frank R. Bengelsdorf,^a Anja Poehlein,^b Carola Esser,^a Bettina Schiel-Bengelsdorf,^a Rolf Daniel,^b Peter Dürre^a

Institut für Mikrobiologie und Biotechnologie, Universität Ulm, Ulm, Germany^a; Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Georg-August University, Göttingen, Germany^b

F.R.B. and A.P. contributed equally to this work.

Here, we report the complete genome sequence of *Moorella thermoacetica* DSM 2955^T, an acetogenic bacterium, which uses the Wood–Ljungdahl pathway for reduction of H₂ + CO₂ or CO. The genome consists of a single circular chromosome (2.62 Mb).

Received 21 August 2015 Accepted 25 August 2015 Published 8 October 2015

Citation Bengelsdorf FR, Poehlein A, Esser C, Schiel-Bengelsdorf B, Daniel R, Dürre P. 2015. Complete genome sequence of the acetogenic bacterium *Moorella thermoacetica* DSM 2955^T. *Genome Announc* 3(5):e01157-15. doi:10.1128/genomeA.01157-15.

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

Address correspondence to Peter Dürre, peter.duerre@uni-ulm.de.

We determined the genome sequence of one of the type strains of *Moorella thermoacetica* DSM 2955^T, originally named *Clostridium thermoaceticum*, which was isolated from horse feces in 1942 (1). About 40 years later, Kerby and Zeikus (2) obtained the original isolated strain from Elizabeth McCoy as a “dried spore stock in soil” and showed that this strain uses H₂ + CO₂ as its sole carbon and energy source. Moreover, this strain was adapted to use CO as its sole substrate (2). These were major differences from the previously submitted type strain (DSM 521^T), and therefore *M. thermoacetica* DSM 2955^T was deposited in “Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH” (DSMZ) as the second type strain of the species. However, since isolation of *M. thermoacetica*, the strains DSM 521^T and ATCC 39073 were used as model acetogens for elucidating the biochemistry of the Wood–Ljungdahl pathway (3, 4). In 2008, Pierce et al. (5) published the first complete genome sequence of the non-type strain *M. thermoacetica* ATCC 39073.

Chromosomal DNA was isolated using a MasterPure complete DNA purification kit (Epicentre, Madison, WI, USA) and used to create Illumina shotgun paired-end sequencing libraries. A MiSeq system and MiSeq reagent kit v3 (600 cycles) were used for sequencing as recommended by the manufacturer (Illumina, San Diego, CA, USA). Sequencing resulted in 2,908,542 paired-end reads (2 × 300 bp). Trimmomatic 0.32 (6) was used for quality filtering and adaptor clipping. The *de novo* assembly was performed by employing the SPAdes genome assembler software 3.5.0 (7), which resulted in 43 contigs (>500 bp) with a 231-fold average coverage. For scaffolding, the Move Contigs tool of the Mauve genome alignment software (8) was used together with the genome sequence of *M. thermoacetica* ATCC 39073 (CP000232) as reference. Remaining gaps were closed by PCR-based techniques, primer walking, and Sanger sequencing of the products using BigDye 3.0 chemistry and an ABI3730XL capillary sequencer (Life Technologies GmbH, Darmstadt, Germany). For this purpose, the Gap4 (v4.11) software of the Staden package (9) was employed.

The genome of *M. thermoacetica* DSM 2955^T consists of a cir-

cular chromosome of 2.62 Mb with an overall G+C content of 55.81%. Automatic gene prediction was performed by using Prodigal (10). Genes coding for rRNA and tRNA were identified using RNAmmer (11) and tRNAscan (12), respectively. The Integrated Microbial Genomes–Expert Review (IMG-ER) system (13) was used for automatic annotation, which was subsequently manually curated by using the Swiss-Prot, TrEMBL, and InterPro databases (14). The genome harbored 3 rRNA genes, 52 tRNA genes, and 2,624 protein-encoding genes. The genome of strain DSM 2955^T (2,623,349 bp) is smaller than that of strain ATCC 39073 (2,628,784 bp), but larger than that of the recently published strain DSM 521^T (2,527,564 bp) (15). It harbors additional genes encoding, e.g., hypothetical proteins, methylases, and transposases. Genome comparisons of strain DSM 2955^T with strain ATCC 39073 and strain DSM 2955^T with strain DSM 521^T revealed 120 single nucleotide polymorphisms (SNPs) and 936 SNPs, respectively.

Nucleotide sequence accession number. This complete genome project has been deposited at DDBJ/EMBL/GenBank under the accession number CP012370.

ACKNOWLEDGMENTS

This work was supported by grants from the BMBF Gas-Fermentation project (FKZ 031A468A), the ERA-IB 5 project CO2CHEM (FKZ 031A566A), and the MWK-BW project Nachhaltige und effiziente Biosynthesen (AZ 33-7533-6-195/7/9).

We thank Kathleen Gollnow and Frauke–Dorothee Meyer for technical support.

REFERENCES

- Fontaine FE, Peterson WH, McCoy E, Johnson MJ, Ritter GJ. 1942. A new type of glucose fermentation by *Clostridium thermoaceticum* n. sp. *J Bacteriol* 43:701–715.
- Kerby R, Zeikus JG. 1983. Growth of *Clostridium thermoaceticum* on H₂/CO₂ or CO as energy source. *Curr Microbiol* 8:27–30. <http://dx.doi.org/10.1007/BF01567310>.
- Ragsdale SW. 2004. Life with carbon monoxide. *Crit Rev Biochem Mol Biol* 39:165–195. <http://dx.doi.org/10.1080/10409230490496577>.
- Drake H, Küsel K, Matthies C. 2006. Acetogenic prokaryotes, p 354–420.

- In Dworkin M, Falkow S, Rosenberg E, Schleifer K, Stackebrandt E (ed), *The Prokaryotes*, vol 2. Springer Verlag, New York, NY.
5. Pierce E, Xie G, Barabote RD, Saunders E, Han CS, Detter JC, Richardson P, Brettin TS, Das A, Ljungdahl LG, Ragsdale SW. 2008. The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). *Environ Microbiol*. 10:2550–2573. <http://dx.doi.org/10.1111/j.1462-2920.2008.01679.x>.
 6. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.
 7. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its applications to single cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
 8. Darling AE, Mau B, Perna NT. 2010. ProgressiveMauve: multiple genome alignment with gene gain, loss, and rearrangement. *PLoS One* 5:e11147. <http://dx.doi.org/10.1371/journal.pone.0011147>.
 9. Staden R, Beal KF, Bonfield JK. 2000. The Staden package, 1998. *Methods Mol Biol* 132:115–130.
 10. 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>.
 11. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
 12. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:0955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
 13. Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC. 2014. IMG4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res* 42:560–567. <http://dx.doi.org/10.1093/nar/gkt963>.
 14. Zdobnov EM, Apweiler R. 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847–848. <http://dx.doi.org/10.1093/bioinformatics/17.9.847>.
 15. Poehlein A, Bengelsdorf FR, Esser C, Schiel-Bengelsdorf B, Daniel R, Dürre P. Complete genome sequence of the type strain of the acetogenic bacterium *Moorella thermoacetica* DSM 521^T. *Genome Announc* 3(5): e01159-15. <http://dx.doi.org/10.1128/genomeA.01159-15>.