



Complete Genome Sequence of *Thermoanaerobacterium* sp. Strain RBITD, a Butyrate- and Butanol-Producing Thermophile

Ranjita Biswas,^a Marcel Huntemann,^b Alicia Clum,^b Manoj Pillay,^b Krishnaveni Palaniappan,^b Neha Varghese,^b Natalia Mikhailova,^b Dimitrios Stamatis,^b T. B. K. Reddy,^b Chris Daum,^b Nicole Shapiro,^b Natalia Ivanova,^b Nikos C. Kyrpides,^b Tanja Woyke,^b  Adam M. Guss^c

^aCentre for Rural Development and Technology, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, India

^bDOE Joint Genome Institute, Walnut Creek, California, USA

^cBiosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

ABSTRACT *Thermoanaerobacterium* sp. strain RBITD was isolated from contaminated rich growth medium at 55°C in an anaerobic chamber. It primarily produces butyrate as a fermentation product from plant biomass-derived sugars. The whole-genome sequence of the strain is 3.4 Mbp, with 3,444 genes and 32.48% GC content.

Thermoanaerobacterium sp. strain RBITD was isolated from a contaminated rich growth medium in an anaerobic chamber. It is a thermophilic anaerobic rod-shaped member of the *Firmicutes* that ferments various plant biomass-derived sugars, including glucose, xylose, arabinose, maltose, fructose, cellobiose, galactose, lactose, mannose, maltose, rhamnose, and sucrose, primarily into butyrate, with the additional production of lactate, acetate, H₂, and *n*-butanol, with no detectable ethanol and acetone production. The strain is interesting from an industrial standpoint due to its exceptionally high yield of butyrate from xylose and glucose (approximately 85% and 60% of the theoretical maximum yield, respectively). Butyrate is a 4-carbon organic acid that is primarily petroleum derived, but bio-based processes are in high demand for applications in the food/feed industry; as a biofuel or jet fuel precursor; in the cosmetic, plastic, and textile fiber industries; and as a bioactive compound in the nutraceutical industry (1–3). This strain could help fill the gap between the demand for bio-based butyric acid and the lack of availability of natural microbes to produce butyric acid on a large scale from plant sugars.

The draft genome of *Thermoanaerobacterium* sp. RBITD was generated at the DOE Joint Genome Institute (JGI) using the Pacific Biosciences (PacBio) sequencing technology (4). A PacBio SMRTbell library was constructed and sequenced on the PacBio RS platform, which generated 176,912 filtered subreads totaling 555.0 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found online (<http://www.jgi.doe.gov>). The raw reads were assembled using HGAP version 2.2.0.p1 (5). The final assembly contained 1 contig in 1 scaffold, totaling 3.4 Mbp. The input read coverage was 164.7×.

Genome annotation was performed using the DOE-JGI annotation pipeline (6, 7). Genes were identified using Prodigal (8), followed by a round of manual curation using GenePRIMP (9). The predicted coding sequences (CDSs) were translated and used to search the Integrated Microbial Genomes (IMG) nonredundant database and the UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAscan-SE tool (10) was used to find tRNA genes, whereas rRNA genes were found by searches against models of the rRNA genes built from SILVA (11). Other noncoding RNAs, such as the RNA components of the protein secretion complex and the RNaseP, were identified by searching the genome for the corresponding Rfam profiles using Infernal (12). Addi-

Received 21 November 2017 **Accepted** 29 November 2017 **Published** 11 January 2018

Citation Biswas R, Huntemann M, Clum A, Pillay M, Palaniappan K, Varghese N, Mikhailova N, Stamatis D, Reddy TBK, Daum C, Shapiro N, Ivanova N, Kyrpides NC, Woyke T, Guss AM. 2018. Complete genome sequence of *Thermoanaerobacterium* sp. strain RBITD, a butyrate- and butanol-producing thermophile. *Genome Announc* 6:e01411-17. <https://doi.org/10.1128/genomeA.01411-17>.

Copyright © 2018 Biswas et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Adam M. Guss, gussam@ornl.gov.

tional gene prediction analysis and manual functional annotation were performed within the Integrated Microbial Genomes (IMG) platform (13) developed by JGI (14).

The genome sequence length is 3,402,993 bp, with 32.48% GC content. The total number of predicted genes is 3,444, of which 3,348 are protein-coding genes, and 2,576 genes had a functional prediction. A total of 96 RNA genes were determined, including 5 rRNA operons. The whole-genome sequence of this strain will offer insight into its metabolic network, serve as a new source for thermophilic proteins, and provide necessary information to enable metabolic engineering for the production of renewable fuels and chemicals from plant biomass feedstocks.

Accession number(s). The complete genome sequence of *Thermoanaerobacterium* sp. strain RBITD has been deposited in GenBank under the accession number [LT906662](https://doi.org/10.1093/nar/gkm864).

ACKNOWLEDGMENTS

This work was supported by the BioEnergy Science Center, U.S. DOE Bioenergy Research Center, supported by the Office of Biological and Environmental Research in the DOE Office of Science, Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. DOE under contract DE-AC05-00OR22725. The sequencing and data analysis work were conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH11231. R.B. gratefully acknowledges the award of the Ramalingaswami Fellowship 2014 and research grants by the Department of Biotechnology, Government of India.

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Van Immerseel F, Boyen F, Gantois I, Timbermont L, Bohez L, Pasmans F, Haesebrouck F, Ducatelle R. 2005. Supplementation of coated butyric acid in the feed reduces colonization and shedding of *Salmonella* in poultry. *Poult Sci* 84:1851–1856. <https://doi.org/10.1093/ps/84.12.1851>.
2. Zhang C, Yang H, Yang F, Ma Y. 2009. Current progress on butyric acid production by fermentation. *Curr Microbiol* 59:656–663. <https://doi.org/10.1007/s00284-009-9491-y>.
3. Dwidar M, Park JY, Mitchell RJ, Sang BI. 2012. The future of butyric acid in industry. *ScientificWorldJournal* 2012:471417. <https://doi.org/10.1100/2012/471417>.
4. 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, Veceli 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. <https://doi.org/10.1126/science.1162986>.
5. 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. Non-hybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10:563–569. <https://doi.org/10.1038/nmeth.2474>.
6. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen IM, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand Genomic Sci* 10:86. <https://doi.org/10.1186/s40793-015-0077-y>.
7. Chen IM, Markowitz VM, Palaniappan K, Szeto E, Chu K, Huang J, Ratner A, Pillay M, Hadjithomas M, Huntemann M, Mikhailova N, Ovchinnikova G, Ivanova NN, Kyrpides NC. 2016. Supporting community annotation and user collaboration in the integrated microbial genomes (IMG) system. *BMC Genomics* 17:307. <https://doi.org/10.1186/s12864-016-2629-y>.
8. 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. <https://doi.org/10.1186/1471-2105-11-119>.
9. 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. <https://doi.org/10.1038/nmeth.1457>.
10. 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.
11. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, 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. <https://doi.org/10.1093/nar/gkm864>.
12. Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29:2933–2935. <https://doi.org/10.1093/bioinformatics/btt509>.
13. Chen IA, Markowitz VM, Chu K, Palaniappan K, Szeto E, Pillay M, Ratner A, Huang JH, Andersen E, Huntemann M, Varghese N, Hadjithomas M, Tennessen K, Nielsen T, Ivanova NN, Kyrpides NC. 2017. IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res* 45:D507–D516. <https://doi.org/10.1093/nar/gkw929>.
14. Markowitz VM, Mavromatis K, Ivanova NN, Chen IMA, Chu K, Kyrpides NC. 2009. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25:2271–2278. <https://doi.org/10.1093/bioinformatics/btp393>.