AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Bacteriology | Announcement

Draft genome sequences of *Butyrivibrio hungatei* DSM 14810 (JK 615^T) and *Butyrivibrio fibrisolvens* DSM 3071 (D1^T)

Keely Berner,[1] Michelle Zoza-Veloz,[1] Matt Nolan,[2] Danielle Graham,[2] Natalia Ivanova,[2] Rekha Seshadri,[2] Stefan Spring,[3] Matthew Escobar[1]

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT Here, we report the draft genome sequences of two *Butyrivibrio*-type strains isolated from rumen fluid. The genome sequence of *Butyrivibrio hungatei* DSM 14810 was 3.3 Mb with 3,093 predicted genes, while the *Butyrivibrio fibrisolvens* DSM 3071 genome sequence was 4.8 Mb with 4,132 predicted genes.

KEYWORDS *Butyrivibrio*, rumen, genomes

Butyrivibrio is commonly found in the rumen of wild and domesticated ruminant animals (1, 2). *Butyrivibrio* spp. are anaerobic, Gram-negative curved rods with a single flagellum, and they ferment rumen oligosaccharides and monosaccharides into butyrate, an essential source of energy for the host (1, 3, 4). *Butyrivibrio fibrisolvens* DSM 3071 (D1^T) was isolated from cow rumen in the 1950s in Maryland, USA (3), while *Butyrivibrio hungatei* DSM 14810 (JK 615^T) was isolated from sheep rumen in the 1990s in Mnichovice, Czech Republic (1). Here, we report draft genome sequences of these two *Butyrivibrio*-type strains as part of the 1000 Microbial Genomes Project (5). These resources will improve our understanding of the complex metabolism and physiology of rumen microbiomes.

Leibniz Institute DSMZ cultured *B. fibrisolvens* DSM 3071 and *B. hungatei* DSM 14810 anaerobically at 37°C and 40°C, respectively, using DSMZ Medium 330 broth (https://mediadive.dsmz.de/medium/330) inoculated from an ampoule preserved by freeze-drying. The centrifuged cell pellet from an approximately 50-mL culture was used to extract genomic DNA using Epicentre's MasterPure Gram-positive DNA Purification Kit. The Joint Genome Institute constructed an Illumina paired-end library with an average insert size of 270 bp. The library was quantified using KAPA Biosystem's sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. Several libraries were multiplexed and prepared for sequencing using a TruSeq paired-end cluster kit (v.3) and Illumina's cBot instrument. Sequencing was performed on an Illumina HiSeq 2000 using TruSeq SBS sequencing kits (v.3) following a 2 × 150 indexed run protocol. A total of 8,786,458 reads totaling 1,318.0 Mbp were generated for *B. fibrisolvens* DSM 3071 and 9,376,984 reads totaling 1,406.5 Mbp were generated for *B. hungatei* DSM 14810. Reads were quality controlled and trimmed through DUK (v.1.0) (6). Filtered reads were assembled using Velvet (v.1.2.07) (velveth: 63 –shortPaired and velvetg: –very clean yes –exportFiltered yes –min contig lgth 500 –scaf- folding no –cov cutoff 10); 1- to 3-kb simulated paired-end reads were created from Velvet contigs using wgsim (v.1.0) (–e 0–1 100–2 100 –r 0 –R 0 –X 0); and these reads were assembled with simulated read pairs with Allpaths-LG (v.r46652) (PrepareAllpathsInputs: PHRED 64 = 0 PLOIDY = 1 FRAG COVERAGE = 125 JUMP COVERAGE = 25 LONG JUMP COV = 50, RunAllpath- sLG TARGETS = standard) (7–9). Gene annotation was performed using Prodigal (v.2.5), with manual curation via GenePRIMP (v.1.0) (10, 11). Predicted coding sequences were translated and used to search National Center for Biotechnology Information nr, UniProt, TIGRFam,

Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine, Baltimore, Maryland, USA

Address correspondence to Rekha Seshadri, rseshadri@lbl.gov, or Matthew Escobar, mescobar@csusm.edu.

Keely Berner and Michelle Zoza-Veloz contributed equally to this article. They are listed alphabetically by surname.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 14 May 2024
Accepted 12 August 2024
Published 28 August 2024

TABLE 1 Genomic features of *Butyrivibrio hungatei* DSM 14810 and *Butyrivibrio fibrisolvens* DSM 3071[a]

Genome feature	*B. hungatei* DSM 14810	*B. fibrisolvens* DSM 3071
Length (bp)	3,394,947	4,837,257
Number of contigs	22	83
Contig N_{50} (bp)	286,711	119,711
Average fold coverage	415×	272×
GC content (%)[b]	39.86	39.66
Predicted genes	3,093	4,132
Predicted protein-coding genes	3,030	4,062
Predicted rRNAs	9	11
Predicted tRNAs	46	44
GH1 domains (β-glucosidase)	1	3
GH3 domains (β-glucosidase)	6	15
GH5 domains (cellulase)	2	5
GH8 domains (cellulase)	1	0
GH9 domains (cellulase)	0	3
GH44 domains (cellulase)	0	0
GH48 domains (cellulase)	0	0
Joint Genome Institute IMG/M taxon ID	2582580726	2585428068
NCBI WGS accession number	GCA_900143205	GCA_900129945
NCBI Bioproject accession number	PRJNA245645	PRJNA245644
NCBI Sequence Read Archive accession number	SRR4096539	SRR4096531
NCBI BioSample number	SAMN02745247	SAMN02745229

[a]NCBI, National Center for Biotechnology Information.
[b]GC, Guanine/Cytosine.

Pfam, KEGG, COG, and InterPro (12–18). tRNAScanSE (v.1.3.1) was used to identify tRNA genes, while rRNA genes were found using SILVA (v.123) (19, 20) (Table 1).

Both *B. fibrisolvens* DSM 3071 and *B. hungatei* DSM 14810 metabolize cellobiose (2, 3), and accordingly, their genomes contain many predicted beta glucosidases (GH1 and GH3 domains). However, DSM 3071 can metabolize polysaccharides such as carboxymethyl cellulose, while DSM 14810 cannot (1). This activity is likely attributable to one or more of the eight predicted cellulases in the DSM 3071 genome, compared to three proteins with cellulase domains (GH5/8/9/44/48) in the DSM 14810 genome (Table 1).

ACKNOWLEDGMENTS

The work (proposal DOI: https://doi.org/10.46936/10.25585/60000886) was conducted by the US Department of Energy, Joint Genome Institute (JGI) (https://ror.org/04xm1d337), a DOE Office of Science User Facility that is supported by the Office of Science of the US Department of Energy operated under contract no. DE-AC02-05CH11231. This announcement was largely prepared by undergraduate students, and we gratefully acknowledge JGI for initiating and supporting it as an educational project (the "Adopt-a-Genome" Project).

AUTHOR AFFILIATIONS

[1]Department of Biological Sciences, California State University San Marcos, San Marcos, California, USA
[2]Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California, USA
[3]Department Microorganisms, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

AUTHOR ORCIDs

Natalia Ivanova http://orcid.org/0000-0002-5802-9485
Rekha Seshadri http://orcid.org/0000-0003-3219-2900
Matthew Escobar http://orcid.org/0000-0002-6559-1222

FUNDING

Funder	Grant(s)	Author(s)
United States Department of Energy	DE-AC02-05CH11231	Matt Nolan Danielle Graham Natalia Ivanova Rekha Seshadri

AUTHOR CONTRIBUTIONS

Keely Berner, Formal analysis, Investigation, Writing – original draft | Michelle Zoza-Veloz, Formal analysis, Investigation, Writing – original draft | Matt Nolan, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization | Danielle Graham, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization | Natalia Ivanova, Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Validation | Rekha Seshadri, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review and editing | Stefan Spring, Methodology, Writing – review and editing | Matthew Escobar, Conceptualization, Formal analysis, Investigation, Supervision, Writing – original draft, Writing – review and editing

REFERENCES

1. Kopečný J, Zorec M, Mrázek J, Kobayashi Y, Marinšek-Logar R. 2003. *Butyrivibrio hungatei* sp. nov. and *Pseudobutyrivibrio xylanivorans* sp. nov., butyrate-producing bacteria from the rumen. Int J Syst Evol Microbiol 53:201–209. https://doi.org/10.1099/ijs.0.02345-0
2. Palevich N, Kelly WJ, Leahy SC, Altermann E, Rakonjac J, Attwood GT. 2017. The complete genome sequence of the rumen bacterium *Butyrivibrio hungatei* MB2003. Stand Genomic Sci 12:72. https://doi.org/10.1186/s40793-017-0285-8
3. Bryant MP, Small N. 1956. The anaerobic monotrichous butyric acid-producing curved rod-shaped bacteria of the rumen. J Bacteriol 72:16–21. https://doi.org/10.1128/jb.72.1.16-21.1956
4. Brown DW, Moore WEC. 1960. Distribution of *Butyrivibrio fibrisolvens* in nature. J Dairy Sci 43:1570–1574. https://doi.org/10.3168/jds.S0022-0302(60)90377-5
5. Kyrpides NC, Woyke T, Eisen JA, Garrity G, Lilburn TG, Beck BJ, Whitman WB, Hugenholtz P, Klenk H-P. 2014. Genomic type strains, phase I: the one thousand microbial genomes (KMG-I) project. Stand Genomic Sci 9:1278–1284. https://doi.org/10.4056/sigs.5068949
6. Choi DH, Ahn C, Jang GI, Lapidus A, Han J, Reddy TBK, Huntemann M, Pati A, Ivanova N, Markowitz V, Rohde M, Tindall B, Göker M, Woyke T, Klenk H-P, Kyrpides NC, Cho BC. 2015. High-quality draft genome sequence of *Gracilimonas tropica* CL-CB462^T (DSM 19535^T), isolated from a *Synechococcus* culture. Stand Genomic Sci 10:98. https://doi.org/10.1186/s40793-015-0088-8
7. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107
8. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A 108:1513–1518. https://doi.org/10.1073/pnas.1017351108
9. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
10. 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
11. Hyatt D, Chen G-L, 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
12. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733–D745. https://doi.org/10.1093/nar/gkv1189
13. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh L-SL. 2005. The universal protein resource (UniProt). Nucleic Acids Res 33:D154–D159. https://doi.org/10.1093/nar/gki070
14. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O. 2001. TIGRFAMs: a protein family resource for the functional identification of proteins. Nucleic Acids Res 29:41–43. https://doi.org/10.1093/nar/29.1.41
15. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: the protein families database in 2021. Nucleic Acids Res 49:D412–D419. https://doi.org/10.1093/nar/gkaa913
16. Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27

17. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28:33–36. https://doi.org/10.1093/nar/28.1.33
18. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, et al. 2023. InterPro in 2022. Nucleic Acids Res 51:D418–D427. https://doi.org/10.1093/nar/gkac993
19. 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. https://doi.org/10.1093/nar/25.5.955
20. 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. https://doi.org/10.1093/nar/gkm864