



# Complete Genome Sequence of *Anaerostipes caccae* Strain L1-92<sup>T</sup>, a Butyrate-Producing Bacterium Isolated from Human Feces

 Kana Morinaga,<sup>a</sup> Hiroyuki Kusada,<sup>a</sup> Miho Watanabe,<sup>a,b</sup>  Hideyuki Tamaki<sup>a,b</sup>

<sup>a</sup>Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan

<sup>b</sup>Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Kana Morinaga and Hiroyuki Kusada contributed equally to this work. Author order was determined by mutual agreement.

**ABSTRACT** *Anaerostipes caccae* strain L1-92<sup>T</sup> is a well-known butyrate-producing bacterium that has been isolated from human feces. In this announcement, we present the complete genome sequence of *A. caccae* strain L1-92<sup>T</sup>, which comprises 3,590,719 bp with a G+C content of 44.30%. The genome harbors 3,369 predicted protein-coding genes.

The genus *Anaerostipes* is one of the most abundant bacterial taxa in the human intestinal microbiome (1, 2). *Anaerostipes* spp. are considered a key gut microbe associated with human health and disease, since they are capable of producing butyrate, known to have beneficial effects on intestinal functions (3). Here, we provide the complete genome sequence of the type species of the genus *Anaerostipes*, *A. caccae* strain L1-92, which is a human feces-derived, Gram-positive, butyrate-producing bacterium (4).

*A. caccae* strain L1-92<sup>T</sup> (=DSM 14662<sup>T</sup>=JCM 13470<sup>T</sup>=NCIMB13811<sup>T</sup>) was obtained from the Japan Collection of Microorganisms (RIKEN BRC, Tsukuba, Japan). The strain was inoculated into anaerobically prepared Gifu anaerobic medium with headspace gas of N<sub>2</sub>/CO<sub>2</sub> (80:20 [vol/vol]) and incubated at 37°C for 2 days. Cells were collected by centrifugation at 8,000 × *g* for 10 min. Genomic DNA was extracted using enzyme-based DNA extraction methods as described previously (5) with some modifications. In brief, lysozyme, achromopeptidase, and proteinase K were used to lyse cells, followed by genomic DNA purification using the phenol-chloroform method. *De novo* sequencing using the HiSeq system (Illumina, San Diego, CA, USA) and PacBio (Menlo Park, CA, USA) sequencing platform and hybrid assembly were conducted at Genewiz, Inc. (South Plainfield, NJ, USA). Illumina paired-end (2 × 150-bp) reads (6,962,694 reads) were generated; then, libraries with different indices were multiplexed and loaded onto an Illumina HiSeq instrument according to the manufacturer's instructions. Default parameters were used for all software. Sequencing was carried out using a 2 × 150-bp paired-end configuration; image analysis and base calling were conducted using the HiSeq control software + OLB + GAPipeline v1.6 (Illumina) on the HiSeq instrument. In addition, long sequence reads (343,145 reads) were generated, and the SMRTbell library was prepared according to the manufacturer's instructions. The library was sequenced and analyzed using the PacBio RS II platform and single-molecule real-time (SMRT) sequencing technology (which yields sequences with ≥99.999% high-quality data) (6). The PacBio reads were assembled using HGAP4 v4.0/Falcon v0.3 of WGS-Assembler v8.2 (7). The *N*<sub>50</sub> value was 5,455 bp. The genome sequence was further corrected using the Illumina HiSeq and PacBio reads with Pilon v1.22 (8) and Quiver (9), respectively. The genome circularization was confirmed by identifying HiSeq reads that map to the beginning and end of the assembly. This yielded a complete genome sequence for *A. caccae* strain L1-92<sup>T</sup>. The resulting sequence was annotated using

**Citation** Morinaga K, Kusada H, Watanabe M, Tamaki H. 2021. Complete genome sequence of *Anaerostipes caccae* strain L1-92<sup>T</sup>, a butyrate-producing bacterium isolated from human feces. *Microbiol Resour Announc* 10:e00056-21. <https://doi.org/10.1128/MRA.00056-21>.

**Editor** David Rasko, University of Maryland School of Medicine

**Copyright** © 2021 Morinaga 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 Hiroyuki Kusada, kusada-hiroyuki@aist.go.jp, or Hideyuki Tamaki, tamaki-hideyuki@aist.go.jp.

**Received** 19 January 2021

**Accepted** 31 March 2021

**Published** 22 April 2021

the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v5.0 (10). The completeness and contamination level of the genome sequence were assessed using CheckM v1.0.7 with the "lineage\_wf" workflow (11).

The genome sequence of *A. caccae* strain L1-92<sup>T</sup> is comprised of a circular chromosome of 3,590,719 bp with a 44.30% G+C content. The numbers of predicted coding sequences and rRNA and tRNA genes in the genome were 3,369, 12, and 57, respectively. CheckM estimated the genome completeness as 99.33% and the contamination rate as 4.03%. The complete genome sequence of *A. caccae* strain L1-92<sup>T</sup> provides essential data for future taxonomic, comparative genomics, and metabolic analysis and for gaining deep insights into how this strain affects human health and disease.

**Data availability.** The complete genome sequence and annotations of *A. caccae* strain L1-92<sup>T</sup> have been deposited at DDBJ/EMBL/GenBank under accession number [AP023027](https://doi.org/10.1093/nar/49/11/6023). The genome sequence has also been submitted to the SRA under BioSample accession number [SAMD00215733](https://www.ncbi.nlm.nih.gov/biosample/SAMD00215733) and BioProject number [PRJDB9542](https://www.ncbi.nlm.nih.gov/bioproject/PRJDB9542). The raw sequence data for strain L1-92<sup>T</sup> were deposited under DRA accession numbers [DRR259155](https://www.ncbi.nlm.nih.gov/dra/RRR259155) (Illumina) and [DRR259156](https://www.ncbi.nlm.nih.gov/dra/RRR259156) (PacBio).

## ACKNOWLEDGMENTS

This study was funded by a Grant-in-Aid for Early-Career Scientists (20K15441) to K.M. from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and by AMED PRIME grant number JP18gm6010019 and JST ERATO grant number JPMJER1502 to H.T. from the Japan Agency for Medical Research and Development (AMED) and the Japan Science and Technology Agency (JST).

## REFERENCES

- Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, Fernandes GR, Tap J, Bruls T, Batto J-M, Bertalan M, Borruel N, Casellas F, Fernandez L, Gautier L, Hansen T, Hattori M, Hayashi T, Kleerebezem M, Kurokawa K, Leclerc M, Levenez F, Manichanh C, Nielsen HB, Nielsen T, Pons N, Poulain J, Qin J, Sicheritz-Ponten T, Tims S, Torrents D, Ugarte E, Zoetendal EG, Wang J, Guarner F, Pedersen O, de Vos WM, Brunak S, Doré J, Weissenbach J, Ehrlich SD, Bork P, MetaHIT Consortium. 2011. Enterotypes of the human gut microbiome. *Nature* 473:174–180. <https://doi.org/10.1038/nature09944>.
- Walker AW, Ince J, Duncan SH, Webster LM, Holtrop G, Ze X, Brown D, Stares MD, Scott P, Bergerat A, Louis P, McIntosh F, Johnstone AM, Lobley GE, Parkhill J, Flint HJ. 2011. Dominant and diet-responsive groups of bacteria within the human colonic microbiota. *ISME J* 5:220–230. <https://doi.org/10.1038/ismej.2010.118>.
- Riviere A, Selak M, Lantin D, Leroy F, De Vuyst L. 2016. Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. *Front Microbiol* 7:979. <https://doi.org/10.3389/fmicb.2016.00979>.
- Schwartz A, Hold GL, Duncan SH, Gruhl B, Collins MD, Lawson PA, Flint HJ, Blaut M. 2002. *Anaerostipes caccae* gen. nov., sp. nov., a new saccharolytic, acetate-utilising, butyrate-producing bacterium from human faeces. *Syst Appl Microbiol* 25:46–51. <https://doi.org/10.1078/0723-2020-00096>.
- Moore ERB, Arnscheidt A, Krüger A, Strömpl C, Mau M. 1999. Simplified protocols for the preparation of genomic DNA from bacterial cultures, p 1–15. In Akkermans ADL, van Elsas JD, de Bruijn FJ (ed), *Molecular microbial ecology manual*. Kluwer Academic Press, Dordrecht, Netherlands.
- McCarthy A. 2010. Third generation DNA sequencing: Pacific Biosciences' single molecule real time technology. *Chem Biol* 17:675–676. <https://doi.org/10.1016/j.chembiol.2010.07.004>.
- Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KHJ, Remington KA, Anson EL, Bolanos RA, Chou H-H, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC. 2000. A whole-genome assembly of *Drosophila*. *Science* 287:2196–2204. <https://doi.org/10.1126/science.287.5461.2196>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Chin C-S, 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. <https://doi.org/10.1038/nmeth.2474>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.