AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of *Lactobacillus helveticus* JCM 1004, an Aminopeptidase-Producing Lactic Acid Bacterium

Kana Morinaga,[a] Hiroyuki Kusada,[a] Miho Watanabe,[a,b] Hideyuki Tamaki[a,b]

[a]Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, Japan
[b]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 their contribution to the data analyses and manuscript preparation.

ABSTRACT We report the complete genome sequence of *Lactobacillus helveticus* JCM 1004, an aminopeptidase-producing lactic acid bacterium. The genome consists of a circular chromosome which comprises 2,261,280 bp, with a G+C content of 37.56%. The genome was predicted to harbor 13 rRNA genes, 64 tRNA genes, and 2,462 protein-coding sequences.

Members of the genus *Lactobacillus* are Gram-positive, fermentative lactic acid bacteria belonging to the phylum *Firmicutes*. Among them, *Lactobacillus helveticus* strains have been widely known as industrial cheese starters (1, 2). The aminopeptidases produced by *L. helveticus* play a significant role in hydrolyzing milk proteins and releasing free amino acids involved in flavors during cheese ripening. *L. helveticus* JCM 1004 has also been reported to produce an aminopeptidase (3), though a gene encoding aminopeptidase has not been identified so far. Here, we performed a complete genome sequence analysis of *L. helveticus* JCM 1004 and identified putative aminopeptidase genes.

L. helveticus JCM 1004 (=LMG 1146 = NCDO 99 = NCIMB 700099) was obtained from the Japan Collection of Microorganisms (RIKEN BRC, Tsukuba, Japan). The strain was cultured on Gifu anaerobic medium (Nissui Pharmaceutical Co., Ltd., Tokyo, Japan) with a headspace gas of N_2/CO_2 (80:20, vol/vol) at 37°C for 48 h under anaerobic conditions. Full-grown cells were harvested by centrifugation and lysed with lysozyme, achromopeptidase, and proteinase K. Genomic DNA was extracted using the phenol-chloroform method as described previously (4). *De novo* sequencing using the HiSeq 2500 system (Illumina, San Diego, CA, USA) and the PacBio sequencing platform (Pacific Biosciences, Menlo Park, CA, USA) was conducted at Genewiz, Inc. (South Plainfield, NJ, USA) as described previously (5). The library was constructed using the VAHTS universal DNA library prep kit for Illumina (Vazyme Biotech, Nanjing, China). Illumina paired-end reads were generated, and the sequencing yielded a total of 9,065,976 reads. Default parameters were used for all bioinformatic software tools. 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. Low-quality sequences and adaptors were removed using Cutadapt v1.9.1 (6). For PacBio sequencing, genomic DNA was sheared using a g-TUBE device (Covaris, Woburn, MA, USA) and then size selected and purified using magnetic beads. A SMRTbell library was constructed using the SMRTbell template prep kit v1.0 (Pacific Biosciences, Menlo Park, CA, USA) and then sequenced and analyzed using the PacBio RS II platform and single-molecule real-time (SMRT) sequencing technology (7). Quality control and adaptor trimming of the PacBio reads were performed using cutadapt v1.9.1; then, the reads were assembled using HGAP4/Falcon v0.3 of WGS-Assembler v8.2 (8). The PacBio sequencing yielded a total of 508,213 reads with an N_{50} value of 4,053 bp.

Citation Morinaga K, Kusada H, Watanabe M, Tamaki H. 2021. Complete genome sequence of *Lactobacillus helveticus* JCM 1004, an aminopeptidase-producing lactic acid bacterium. Microbiol Resour Announc 10:e00641-21. https://doi.org/10.1128/MRA.00641-21.

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Address correspondence to Hiroyuki Kusada, kusada-hiroyuki@aist.go.jp, or Hideyuki Tamaki, tamaki-hideyuki@aist.go.jp.

Received 22 June 2021
Accepted 28 July 2021
Published 19 August 2021

TABLE 1 Summary of putative aminopeptidases found in the *L. helveticus* JCM 1004 genome

Locus tag	Length (aa)[a]	Protein accession no.	Protein name	Source	E value	Identity (%)
LBHL_02250	437	P94870	Aminopeptidase E	*Lactobacillus helveticus*	3.4e−173	53.7
LBHL_02340	438	P94870	Aminopeptidase E	*Lactobacillus helveticus*	0.0	100
LBHL_03880	449	Q10744	Aminopeptidase C	*Lactobacillus helveticus*	0.0	98.4
LBHL_06230	504	M3XFH7	Aminopeptidase Q	*Felis catus*	1.3e−6	23.9
LBHL_14050	437	P94870	Aminopeptidase E	*Lactobacillus helveticus*	0.0	66.3
LBHL_18150	275	P19994	Methionine aminopeptidase 1	*Bacillus subtilis* strain 168	3.9e−69	44.7
LBHL_18880	360	O34924	Putative aminopeptidase YtoP	*Bacillus subtilis* strain 168	2.9e−47	33.1
LBHL_22300	844	Q10730	Aminopeptidase N	*Lactobacillus helveticus*	0.0	99.8

[a] aa, amino acids.

The genome sequence was further corrected using the Illumina HiSeq and PacBio reads with Pilon v1.22 (9) and Quiver (10), respectively. The genome circularization was confirmed by PCR and Sanger sequencing across the ends of the chromosome. The resulting circular genome was evaluated for its completeness and contamination level using CheckM v1.0.7 (11). Protein-coding genes and prophage sequences were predicted using DFAST (12) and PHASTER (13), respectively. The amino acid sequences of putative aminopeptidases were further analyzed using the UniProt Web BLAST server (14).

The complete *L. helveticus* JCM 1004 genome comprises a circular 2,261,280-bp chromosome with an average coverage of 650×, a G+C content of 37.56%, 2,462 protein-coding sequences, 1 intact prophage, 13 rRNA genes, and 64 tRNA loci. Based on the CheckM analysis, the *L. helveticus* JCM 1004 genome was 99.03% complete with no contamination. The *L. helveticus* JCM 1004 genome encodes eight different aminopeptidase coding genes (Table 1). The complete genome sequence information of strain JCM 1004 will be useful for better understanding the physiological characteristics of the cheese starter and to further verify the enzymatic functions and production mechanisms of its aminopeptidases.

Data availability. The complete genome sequence and annotations of *L. helveticus* JCM 1004 have been deposited at DDBJ/EMBL/GenBank under accession number AP023028. The genome sequence also has been submitted to the SRA under BioSample accession number SAMD00215734 and BioProject accession number PRJDB9543. The raw sequence data for strain JCM 1004 were deposited under DRA accession numbers DRR287345 (Illumina) and DRR287346 (PacBio).

ACKNOWLEDGMENTS

K.M. was supported by a Grant-in-Aid for Early-Career Scientists (20K15441) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and a Grant-in-Aid for JSPS Fellows (21J01569) from the Japanese Society for the Promotion of Science (JSPS). H.T. was supported by Japan Agency for Medical Research and Development (AMED) PRIME grant number JP18gm6010019 and Japan Science and Technology Agency (JST) ERATO grant number JPMJER1502.

REFERENCES

1. Jensen MP, Ardö Y. 2010. Variation in aminopeptidase and aminotransferase activities of six cheese related *Lactobacillus helveticus* strains. Int Dairy J 20:149–155. https://doi.org/10.1016/j.idairyj.2009.09.007.
2. Moser A, Schafroth K, Meile L, Egger L, Badertscher R, Irmler S. 2018. Population dynamics of *Lactobacillus helveticus* in Swiss Gruyere-type cheese manufactured with natural whey cultures. Front Microbiol 9:637. https://doi.org/10.3389/fmicb.2018.00637.
3. Pan D, Tanokura M. 2004. Purification and characterization of an aminopeptidase from *Lactobacillus helveticus* JCM 1004. Food Chem 88:511–516. https://doi.org/10.1016/j.foodchem.2004.01.082.
4. Moore ERB, Arnscheidt A, Krüger A, Strömpl C, Mau M. 1999. Simplified protocols for the preparation of genomic DNA from bacterial cultures, p 1.6.1.1–1.6.1.15. *In* Akkermans AD, van Elsas JD, de Bruijn FJ, Kluwer D (ed), Molecular microbial ecology manual. Springer, Dordrecht, Netherlands. https://doi.org/10.1007/978-1-4020-2177-0_1.
5. Morinaga K, Kusada H, Watanabe M, Tamaki H. 2021. Complete genome sequence of *Anaerostipes caccae* strain L1-92^T, a butyrate-producing bacterium isolated from human feces. Microbiol Resour Announc 10:e00056-21. https://doi.org/10.1128/MRA.00056-21.
6. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
7. 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.
8. 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.

9. 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.
10. 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.
11. 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.
12. Tanizawa Y, Fujisawa T, Nakamura Y. 2018. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 34:1037–1039. https://doi.org/10.1093/bioinformatics/btx713.
13. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, Wishart DS. 2016. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res 44:W16–W21. https://doi.org/10.1093/nar/gkw387.
14. UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res 49:D480–D489. https://doi.org/10.1093/nar/gkaa1100.