AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of *Bifidobacterium longum* ZJ1, Isolated from a Centenarian in Anhui, China

Zeyu Jin,[a] Wei Li,[a] Wei Wang,[a] Baolin Sun[b]

[a]Chorain Health Technology Co., Ltd., Hefei, Anhui, China
[b]School of Life Sciences, University of Science and Technology of China, Hefei, Anhui, China

ABSTRACT Here, we present the complete genome sequence of a *Bifidobacterium longum* isolate, that of strain ZJ1, and this strain showed a cholesterol degradation ability that is greater than that of five strains we chose for comparison (*Bifidobacterium longum* 536, *B. infantis* 1912, *B. longum* 1941, *B. breve* ATCC 15698, *B. infantis* ATCC 17930). The draft genome of strain ZJ1 consists of 2,414,672 bp, with 2,042 protein-coding genes, 69 noncoding RNA genes, and 60.16% G+C content.

Bifidobacterium spp. have been shown to improve human health conditions and provide protection against infection and potentially health-promoting metabolites (1). The *Bifidobacterium* genus contains probiotics which could improve the clinical symptoms in patients with mild to moderate active ulcerative colitis (2). *Bifidobacterium longum* is one of the most abundant species of the *Bifidobacterium* genus (3).

Bifidobacterium longum ZJ1 was isolated from a fecal sample from a 105-year-old male from Hefei, Anhui, China. The isolate was cultivated on de Man, Rogosa, and Sharpe (MRS) agar (Oxoid, Basingstoke, United Kingdom) supplemented with 0.05% (wt/vol) L-cysteine under aerobic conditions for 48 h at 35°C. Genomic DNA of the cultured isolate was extracted using the Wizard genomic DNA purification kit (Promega, San Luis Obispo, CA, USA), according to the manufacturer's instructions.

The genome sequence of *Bifidobacterium longum* ZJ1 was sequenced using a PacBio RS II platform (20-kb SMRTbell template) and an Illumina HiSeq 4000 platform (TruSeq DNA PCR-free 350-bp library) at the Beijing Genomics Institute (BGI, Shenzhen, China). Genomic DNA was sheared with a g-TUBE (Covaris) and purified using AMPure PB magnetic beads (Beckman Coulter) to construct a 20-kb library. The libraries were further size selected utilizing BluePippin (Sage Scientific, Beverly, MA), with a cutoff size of 10 kb, and sequenced on a PacBio RS II sequencer. Also, the DNA sample was randomly fragmented and ligated with 5′ and 3′ adapters to construct the sequencing library for Illumina HiSeq 4000 platform paired-end sequencing.

A total of 8,396,464 reads were obtained from Illumina and low-quality reads (bases with quality lower than 20, >40%; N content, >10%) were filtered using SOAPnuke v1.5.2 (4) (default settings). In total, 128,072 subreads were obtained from PacBio sequencing, and the subreads with less than 1 kb were removed. The N_{50} and N_{90} values of the subreads are 1,885 bp and 3,551 bp, respectively. The package proovread v2.12 (5) (-t 4 –coverage 60 –mode sr) in the program pbdagcon was used for self-correction. Draft genomic unitigs were assembled using the Celera Assembler v8.3 (6) (doTrim_initialQualityBased = 1, doTrim_finalEvidenceBased = 1, doRemoveSpurReads = 1, doRemoveChimericReads = 1, d properties -U), and then GATK v1.6-13 (7) (-cluster 2 -window 5 -stand_call_conf 50 -stand_emit_conf 10.0 -dcov 200 MQ0> = 4) was used for single-base corrections to improve the accuracy of the genome sequences.

The assembled genome consists of one circular chromosome of 2,414,672 bp, with

Citation Jin Z, Li W, Wang W, Sun B. 2019. Draft genome sequence of *Bifidobacterium longum* ZJ1, isolated from a centenarian in Anhui, China. Microbiol Resour Announc 8:e00878-19. https://doi.org/10.1128/MRA.00878-19.

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

Copyright © 2019 Jin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Baolin Sun, sunb@ustc.edu.cn.

Received 24 July 2019
Accepted 27 July 2019
Published 15 August 2019

Downloaded from http://mra.asm.org/ on August 15, 2019 at China University of Science and Technology

a G+C content of 60.16%. There are 2,042 protein-coding genes, as predicted by Glimmer3 v3.02 (8), within the genome of *Bifidobacterium longum* ZJ1. In total, 12 rRNA genes, 56 tRNA genes, and 1 small RNA (sRNA) gene were recognized by tRNAscan-SE v1.3.1 (9) [–Spec_tag(BAOG) –o *. tRNA –f * .tRNA.structure], RNAmmer v1.2 (10) (–s Species –m Type –gff *. rRNA.gff –f *.rRNA.fq), and the Rfam database v9.1 (11) (–p blastn –W 7 –e 1 –v 10000 –b 10000 –m 8 –i subfile –o *.blast.m8).

Two incomplete prophages were found using the PHAge Search Tool (PHAST) v2013.03.20 (12) (default settings); one prophage has 12,784 bp, and the other prophage has 30,265 bp. CRISPRFinder v0.4 (13) (default settings) was used to identify 7 CRISPR sequences. One has 30 spacers, one has 7 spacers, and the others have only 1 spacer.

The synteny of *B. longum* ZJ1 and 4 other *B. longum* strains (*B. longum* 105-A, *B. longum* AH1206, *B. longum* BORI, and *B. longum* KACC91563) was determined using MUMmer v3.22 (14) (-b 200 -c 65 –extend -l 20), and BLAST core/pan genes of these 5 strains were clustered using CD-HIT v4.6.6 (15) (-c 0.5 -n 3 -p 1 -g 1 -d 0 -s 0.7 -aL 0.7 -aS 0.7) with a threshold of 50% pairwise identity and 0.7 length difference cutoff in amino acids. Finally, 1,356 core genes were identified.

Data availability. This whole-genome project has been deposited at DDBJ/EMBL/GenBank under accession number CP040235. The BioProject accession number is PRJNA539831. Raw sequence reads are deposited at DDBJ/EMBL/GenBank under accession numbers SRR9051180 and SRR9051181.

REFERENCES

1. Arboleya S, Watkins C, Stanton C, Ross RP. 2016. Gut bifidobacteria populations in human health and aging. Front Microbiol 7:1204. https://doi.org/10.3389/fmicb.2016.01204.
2. Plaza-Díaz J, Ruiz-Ojeda FJ, Vilchez-Padial LM, Gil A. 2017. Evidence of the anti-inflammatory effects of probiotics and synbiotics in intestinal chronic diseases. Nutrients 9:555. https://doi.org/10.3390/nu9060555.
3. Arboleya S, Stanton C, Ryan CA, Dempsey E, Ross PR. 2016. Bosom buddies: the symbiotic relationship between infants and Bifidobacterium longum ssp. longum and ssp. infantis genetic and probiotic features. Annu Rev Food Sci Technol 7:1–21. https://doi.org/10.1146/annurev-food-041715-033151.
4. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang X, Wang J, Yang H, Fang L, Chen Q. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience 7:1–6. https://doi.org/10.1093/gigascience/gix120.
5. Hackl T, Hedrich R, Schultz J, Forster F. 2014. proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics 30:3004–3011. https://doi.org/10.1093/bioinformatics/btu392.
6. Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, 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.
7. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11.10.1–33. https://doi.org/10.1002/0471250953.bi1110s43.
8. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679. https://doi.org/10.1093/bioinformatics/btm009.
9. 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.
10. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.
11. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, Wilkinson AC, Finn RD, Griffiths-Jones S, Eddy SR, Bateman A. 2009. Rfam: updates to the RNA families database. Nucleic Acids Res 37:D136–D140. https://doi.org/10.1093/nar/gkn766.
12. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. Nucleic Acids Res 39:W347–W352. https://doi.org/10.1093/nar/gkr485.
13. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35:W52–W57. https://doi.org/10.1093/nar/gkm360.
14. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.
15. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

Downloaded from http://mra.asm.org/ on August 15, 2019 at China University of Science and Technology