Complete Genome Sequence of *Lactobacillus harbinensis* Strain NSMJ42, Isolated from Makgeolli, a Traditional Korean Alcoholic Beverage

Ji Young Jung,[a] Sang-Soo Han,[a] Z-Hun Kim,[a] Byung-Gon Ryu,[a] Hyun Mi Jin,[a] Eu Jin Chung[a]

[a]Microbial Research Department, Nakdonggang National Institute of Biological Resources (NNIBR), Sangju-si, Gyeongsangbuk-do, South Korea

ABSTRACT In the present work, we report the complete genome sequence of *Lactobacillus harbinensis* NSMJ42, isolated from makgeolli (a Korean traditional alcoholic beverage) in South Korea. The final genome assembly consists of a 3.29-Mbp chromosome with 3,082 protein-coding sequences and a G+C content of 53.36%.

Makgeolli is a Korean traditional fermented alcoholic beverage with a 6 to 8% alcohol content that is brewed with rice and nuruk. Nuruk, a starchy disk or tablet formed from various cereals as raw material, contains diverse fungal and bacterial strains from the surrounding environment and acts as a starter culture for saccharification and alcoholic fermentation for producing makgeolli (1). Studies on the makgeolli microflora have revealed the presence of amylolytic molds (*Aspergillus*, *Rhizopus*, and *Mucor* spp.), alcohol-producing yeasts (*Saccharomyces* spp.), and lactic acid bacteria (LAB) in makgeolli (1–6). LAB are involved in the production of organic acids, amino acids, vitamins, and aromatic compounds during makgeolli fermentation and also in the prevention of bacterial contamination and spoilage by *Micrococcus*, *Bacillus*, *Aerobacter*, and *Pseudomonas* spp. (1, 7–9). It has been reported that makgeolli has medicinal properties like antioxidant, antihypertensive, antidiabetes, and anticancer activities (1, 10, 11). Moreover, probiotic properties of LAB and yeast isolates in makgeolli have been proven (12, 13). We isolated *Lactobacillus harbinensis* NSMJ42 from makgeolli and sequenced the whole genome to understand its whole metabolic capacity and functional potential.

A traditional makgeolli collected in Gyeongsangbuk Province (South Korea) was diluted in phosphate-buffered saline (PBS) (pH 7.4), and the dilutions were spread over an MRS agar (Difco) plate. The plates were incubated at 30°C for 48 h, and we obtained a single colony of strain NSMJ42. For whole-genome sequencing, genomic DNA was isolated from strain NSMJ42 grown in MRS broth (Difco) at 30°C, using a TruSeq DNA PCR-free kit (Illumina). The whole genome was sequenced at Cosmo Genetech (Seoul, South Korea) by a combination of the PacBio RS II single-molecule, real-time (SMRT) sequencing platform using a 20-kb SMRTbell template library and the Illumina NovaSeq 6000 platform (2 × 101 bp) with an insert size of 550 bp. A total of 70,372 postfilter polymerase reads (783,148,504 bp; mean read length, 11,128 bp) were generated from SMRT sequencing, and 100,364 subreads of clean data (781,769,715 bp; mean subread length, 7,789 bp) were produced with quality filtering (minimum polymerase read quality, 0.75; minimum polymerase read length, 50) and adapter trimming using HGAP.3 within PacBio's SMRT Analysis v2.3.0 (14). To generate long and accurate sequences, preassembly was performed by mapping shorter subreads onto longer subreads (14,557-bp threshold) using HGAP.3 (14). The error-corrected 7,726 long subreads (84,586,769 bp; mean read length, 10,948 bp) were *de novo* assembled to the initial draft genome assembly by HGAP.3 (14). Additionally, 5,037.99 Gbp (1,531.01-fold coverage) with 49,881,092 paired-end reads were generated from the Illumina NovaSeq

6000 system. The raw Illumina reads were used for consensus genome polishing and error correction by mapping onto the initial PacBio draft genome assembly with HGAP.3 (14), and the resulting contig was circularized using NUCmer v3.1 and MUM-merplot v3.5 (15).

The final genome assembly, which had a mean coverage of 162.31-fold and a G+C content of 53.36%, consisted of a 3,290,626-bp circular chromosome. Average nucleotide identity (ANI) analysis was conducted with OrthoANIu (16) to the accurate identification of strain NSMJ42 and resulted in 97.97% similarity to *L. harbinensis* DSM 16991^T (GenBank accession number AUEH00000000). The value is higher than the ANI threshold of 95 to 96% (17), indicating that strain NSMJ42 belongs to the same species, *L. harbinensis*. The NSMJ42 genome was annotated on NCBI PGAP version 4.8 (18), and it contains 3,082 protein-coding genes, 15 rRNA genes, 67 tRNA genes, 4 noncoding RNAs, and 56 pseudogenes. BASys genome annotation (19) showed that specific clusters of orthologous groups (COGs) were assigned to 2,062 coding sequences (CDSs), and genes for carbohydrate transport and metabolism (G) showed the highest prevalence (10.4%), followed by genes for replication, recombination, and repair (L) (6.6%) and transcription (K) (6.1%). The strain NSMJ42 genome contains 160 carbohydrate-active enzyme (CAZyme) genes, as predicted by HMMER searches (E value, <1E−15; coverage, >0.35) in dbCAN (20), including 108 genes encoding glycoside hydrolases (GHs), 18 genes encoding carbohydrate esterases (CEs), 28 genes encoding glycosyltransferases (GTs), 2 genes encoding polysaccharide lyases (PLs), and 4 genes encoding carbohydrate-binding modules (CBMs) involved in the degradation or modification of carbohydrates and their subsequent utilization in fermentative metabolism. In addition, several cell surface proteins (class A and C sortases), LPXTG motif cell wall anchor domain proteins, and D-alanyl-lipoteichoic acid biosynthesis proteins (dltABCD) were detected in the strain NSMJ42 genome, which explains the potential of *L. harbinensis* NSMJ42 to adhere to the intestinal epithelial cells (21, 22). The bacteriocin genome-mining tool BAGEL4 (23) identified one area of interest (AOI) corresponding to class II bacteriocin.

Data availability. The genome sequence and raw sequencing reads for strain NSMJ42 were deposited under GenBank accession number CP041364, BioProject accession number PRJNA552757, BioSample accession number SAMN12217290, and SRA accession numbers SRX6406718 and SRX6406719.

REFERENCES

1. Nile SH. 2015. The nutritional, biochemical and health effects of makgeolli—a traditional Korean fermented cereal beverage. J Inst Brew 121:457–463. https://doi.org/10.1002/jib.264.
2. Chai C, Lim GS, Kim YJ, Oh SW. 2015. Microbial community changes in *Makgeolli* during brewing. J Inst Brew 121:304–308. https://doi.org/10.1002/jib.227.
3. Kwon SJ, Ahn TY, Sohn JH. 2012. Analysis of microbial diversity in makgeolli fermentation using PCR-DGGE. J Life Sci 22:232–238. https://doi.org/10.5352/JLS.2012.22.2.232.
4. Lee HL, Kang KW, Seo DH, Jung JH, Jung DH, Kim GW, Park SY, Shin WC, Shim HS, Park CS. 2015. Diversity of lactic acid bacteria (LAB) in makgeolli and their production of γ-aminobutyric acid. Kor J Food Sci Technol 47:204–210. https://doi.org/10.9721/KJFST.2015.47.2.204.
5. Jung MJ, Nam YD, Roh SW, Bae JW. 2012. Unexpected convergence of fungal and bacterial communities during fermentation of traditional Korean alcoholic beverages inoculated with various natural starters. Food Microbiol 30:112–123. https://doi.org/10.1016/j.fm.2011.09.008.
6. Jin J, Kim SY, Jin Q, Eom HJ, Han NS. 2008. Diversity analysis of lactic acid bacteria in Takju, Korean rice wine. J Microbiol Biotechnol 18:1678–1682.
7. Feron G, Bonnarme P, Durand A. 1996. Prospects for the microbial production of food flavours. Trends Food Sci Tech 7:285–293. https://doi.org/10.1016/0924-2244(96)10032-7.
8. Lee CH, Tae WT, Kim GM, Lee HD. 1991. Studies on the pasteurization conditions of Takju. Korean J Food Sci Technol 23:44–51.
9. Park HJ, Lee SM, Song SH, Kim YS. 2013. Characterization of volatile components in Makgeolli, a traditional Korean rice wine, with or without pasteurization, during storage. Molecules 18:5317–5325. https://doi.org/10.3390/molecules18055317.
10. Choi JS, Seo HJ, Lee YR, Kwon SJ, Moon SH, Park SM, Sohn JH. 2014. Characteristics and *in vitro* anti-diabetic properties of the Korean rice wine, makgeolli fermented with *Laminaria japonica*. Prev Nutr Food Sci 19:98–107. https://doi.org/10.3746/pnf.2014.19.2.098.
11. Min JH, Kim YH, Kim JH, Choi SY, Lee JS, Kim HK. 2012. Comparison of microbial diversity of Korean commercial makgeolli showing high β-glucan content and high antihypertensive activity, respectively. Mycobiology 40:138–141. https://doi.org/10.5941/MYCO.2012.40.2.138.
12. Park YU, Kim MD, Jung DH, Seo DH, Jung JH, Park JG, Hong SY, Cho JY, Park SY, Park JW, Shin WC, Park CS. 2015. Probiotic properties of lactic acid bacteria isolated from Korean rice wine Makgeolli. Food Sci Biotechnol 24:1761–1766. https://doi.org/10.1007/s10068-015-0229-2.
13. Lee Y, Seol J, Jeong D, Kim SR. 2016. Application of functional microbial

strains isolated from traditional rice wine in Korea. Microbiol Biotechnol Lett 44:229–235. https://doi.org/10.4014/mbl.1605.05002.

14. 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. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.

15. 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.

16. Yoon SH, Ha SM, Lim J, Kwon SJ, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110:1281–1286. https://doi.org/10.1007/s10482-017-0844-4.

17. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S, Trujillo ME. 2018. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol 68:461–466. https://doi.org/10.1099/ijsem.0.002516.

18. 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.

19. Van Domselaar GH, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS. 2005. BASys: a Web server for automated bacterial genome annotation. Nucleic Acids Res 33:W455–W459. https://doi.org/10.1093/nar/gki593.

20. Yin Y, Mao X, Yang JC, Chen X, Mao F, Xu Y. 2012. dbCAN: a Web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–W451. https://doi.org/10.1093/nar/gks479.

21. Vélez MP, De Keersmaecker SC, Vanderleyden J. 2007. Adherence factors of Lactobacillus in the human gastrointestinal tract. FEMS Microbiol Lett 276:140–148. https://doi.org/10.1111/j.1574-6968.2007.00908.x.

22. Lebeer S, Vanderleyden J, De Keersmaecker SC. 2008. Genes and molecules of lactobacilli supporting probiotic action. Microbiol Mol Biol Rev 72:728–764. https://doi.org/10.1128/MMBR.00017-08.

23. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. 2018. BAGEL4: a user-friendly Web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res 46:W278–W281. https://doi.org/10.1093/nar/gky383.