

Genome Reports

Draft Genome Sequence of *Weissella koreensis* Strain HJ, a Probiotic Bacterium Isolated from Kimchi

Seung-Min Yang¹, Eiseul Kim¹, So-Yun Lee¹, Soyeong Mun², Hae Choon Chang², and Hae-Yeong Kim^{1*}

¹Institute of Life Sciences and Resources and Department of Food Science and Biotechnology, Kyung Hee University, Yongin 17104, Republic of Korea

²Department of Food and Nutrition, Chosun University, Gwangju 61452, Republic of Korea

Received: January 25, 2023 / Revised: February 14, 2023 / Accepted: February 15, 2023

Here we report the draft genome sequence of *Weissella koreensis* strain HJ and genomic analysis of its key features. The genome consists of 1,427,571 bp with a GC content of 35.5%, and comprises 1,376 coding genes. *In silico* analysis revealed the absence of pathogenic factors within the genome. The genome harbors several genes that play an important role in the survival of the gastrointestinal tract. In addition, a type III polyketide synthase cluster was identified. Pangenome analysis identified 68 unique genes in *W. koreensis* strain HJ. The genome information of this strain provides the basis for understanding its probiotic properties.

Keywords: *Weissella koreensis*, probiotics, whole-genome sequencing, pangenome analysis

Kimchi is a Korean traditional fermented food made by fermenting kimchi cabbage with seasonings including red pepper powder, ginger, garlic, and jeotgal at low temperatures [1]. Kimchi, which has not been sterilized, naturally induces the growth of various microorganisms during the fermentation process, with lactic acid bacteria such as *Lactiplantibacillus*, *Latilactobacillus*, *Leuconostoc*, and *Weissella* being the most involved. The genus *Weissella* is a Gram-positive, non-spore-forming, non-motile short-rod, and heterofermentative lactic acid bacteria belonging to *Leuconostocaceae* [2]. This genus was first reported in 1993 and currently includes 19 species with validly published names [3].

Among the *Weissella* species, *Weissella koreensis* is a dominant lactic acid bacterium in fermented kimchi and has been isolated from diverse habitats, such as fermented fish, fermented sausage, and fermented sour-

dough [1]. This species produces beneficial metabolites that contribute to the taste and flavor of fermented foods and contribute to sourdough fermentation during bread making [4]. Moreover, they have been shown to provide health benefits such as improvement of atopic diseases, antiobesity effects, and immunomodulatory effects [5]. *W. koreensis* was the first reported S-5623^T strain isolated from kimchi [6]. On the 16S rRNA gene sequence, *W. koreensis* is most closely related to *W. kandleri*, but formed distinct a cluster from this species with 97% sequence similarity [7]. Presently, five genomes have been sequenced: strain KACC 15510 (CP002899.1), strain WiKim0080 (CP026847.1), strain SK (CP043431.1), strain CBA3615 (CP046070.1), and strain KCTC 3621^T (AKGG00000000). The purpose of this study was to analyze the genome sequence of *W. koreensis* strain HJ isolated from kimchi and to identify probiotic-related genes present in the genome of this strain.

In a previous study, *W. koreensis* strain HJ was isolated from kimchi [8]. For genomic DNA extraction, *W. koreensis*

***Corresponding author**

Phone: +82-31-201-2660, Fax: +82-31-204-8116
E-mail: hykim@khu.ac.kr

strain HJ was cultivated to MRS broth for 48 h at 30°C and harvested by centrifugation at 16,000 ×g for 10 min. The genomic DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen, Germany) according to the manufacturer's instruction. The genomic DNA was fragmented and into the sequencing libraries using the Illumina TruSeq Nano DNA library preparation kit (Illumina, USA). The genome sequencing was performed using the Illumina Miseq instrument (Illumina) with a 300 bp paired-end protocol. Sequencing reads were assembled using the CLC genomic workbench 10.0.1 (Qiagen) with the de novo assembly of sequences with high quality. The draft genome of the *W. koreensis* strain HJ has been deposited in GenBank with the accession numbers of JAQAZC000000000 and annotated using the prokaryotic genome annotation pipeline of the National Center for Biotechnology Information (NCBI) prokaryotic genome annotation pipeline.

The taxonomic position of *W. koreensis* strain HJ was confirmed by average nucleotide identity (ANI) and digital DNA-DNA hybridization (dDDH). ANI and dDDH values were calculated using the ANI calculator and Genome-to-Genome Distance Calculator, respectively, based on the default parameters. Whole genome-based taxonomic analysis was performed using type (strain) genome server (TYGS). ResFinder version 4.1 and VirulenceFinder version 2.0 were used to detect antibiotic resistance genes and virulence genes in *W. koreensis* strain HJ. Mobile gene element was detected using MGE version 1.0.3 (database version 1.0.2). The analysis of potential clusters of secondary metabolites was performed using antiSMASH. The probiotic genes in the *W. koreensis* strain HJ genome were identified by Rapid Annotations using Subsystem Technology. For comparative genomics between the genome of *W. koreensis* strain HJ and the genomes of other *W. koreensis* strains (KACC 15510, CP002899.1; WiKim0080, CP026847.1; SK, CP043431.1; CBA3615, CP046070.1; KCTC 3621, AKGG000000000), publicly available genomes of *W. koreensis* were obtained from NCBI. The pangenome analysis for six *W. koreensis* genomes was performed using Bacterial Pan-genome Analysis (BPGA) software version 1.3. The assignment of functional genes was conducted using basic local alignment search tool (BLAST) in the BPGA software against the Kyoto encyclopedia of genes and genomes (KEGG) and cluster of

Table 1. Genome feature of *W. koreensis* strain HJ.

Feature	Values
Genome size (bp)	1,427,571
GC content (%)	35.5
Contig N50	347217
Contig L50	2
Number of scaffolds	15
Number of coding genes	1,376
Number of RNAs	55
Number of subsystems	179

orthologous groups (COG) databases.

A total of 3,577,172 reads with an average length of 295.07 bp were obtained and genome assembly of the reads produced 15 scaffolds (9,113 to 396,269 bp, N₅₀ length 347,217 bp). The total length of draft genome size was 1,427,571 bp with a GC content of 35.5% (Table 1). *W. koreensis* strain HJ has a genome size and GC content similar to *W. koreensis* KACC 15510 (1,441,470 bp, 35.5%), *W. koreensis* WiKim0080 (1,537,013 bp, 35.5%), and *W. koreensis* SK (1,462,088 bp, 35.5%). The annotation results showed that the genome of *W. koreensis* strain HJ had 1,376 coding sequences. A total of 1,376 protein-coding genes were involved in 33 different pathways according to the KEGG database. Arginine deiminase (*arcA*), ornithine carbamoyltransferase (*argF*), and carbamate kinase (*arcC*) were identified as enzymes in the arginine deiminase pathway. In the COG distribution, general function prediction only (R, 13.16%), translation, ribosomal structure and biogenesis (J, 11.26%), and translation (K, 9.11%) were abundant.

The phylogeny relationship among the *W. koreensis* strain HJ and other species was constructed by genome BLAST distance phylogeny approach. The phylogenetic tree showed that *W. koreensis* strain HJ forms a cluster with the *W. koreensis* type strain (KCTC 3621^T) (Fig. 1). The 16S rRNA gene similarity between *W. koreensis* strain HJ and *W. koreensis* strain KCTC 3621^T was 99.9%. Moreover, the 16S rRNA gene similarity showed 93.1 to 98.9% identities between *W. koreensis* strain HJ and other *Weissella* species (*W. bombi* LMG 28290^T, 94.4%; *W. ceti* strain 1119-1A-09^T, 94.7%; *W. cibaria* LMG 17699^T, 95.2%; *W. coleopterorum* HDW19^T, 98.9%; *W. confusa* JCM 1093^T, 95.3%; *W. diestrammenae* ORY33^T, 97.3%; *W. halotolerans* NRIC 1627^T, 93.1%; *W.*

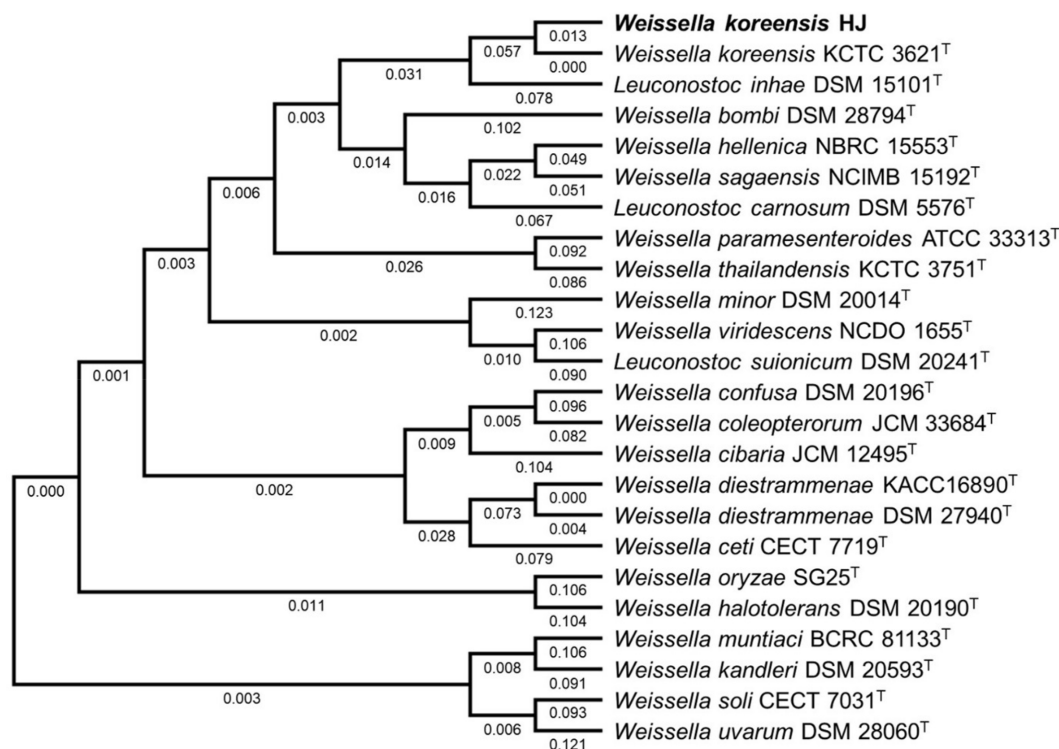


Fig. 1. Whole-genome based phylogenetic tree for classification of *W. koreensis* strain HJ. The genome BLAST distance phylogeny tree was constructed by the Type Strain Genome Server (TYGS; accessed 17 January 2023). Numbers in the figure indicate branch lengths.

hellenica NCFB 2973^T, 94.6%; *W. kandleri* DSM 20593^T, 96.4%; *W. minor* NRIC 1625^T, 95.3%; *W. muntiaci* 8 H-2^T, 95.9%; *W. oryzae* SG 25^T, 95.7%; *W. paramesenteroides* NRIC 1542^T, 94.5%; *W. sagaensis* X0750^T, 94.5%; *W. soli* JCM 12536^T, 96.0%; *W. thailandensis* FS61-1^T, 93.1%; *W. uvarum* B18NM42^T, 95.5%; *W. viridescens* NRIC 1536^T, 95.6%). As a result of calculating the ANI value for *W. koreensis* strain HJ and 19 *Weissella* species, *W. koreensis* strain HJ is most closely related to *W. koreensis* KCTC 3621^T (99.12%). The proposed ANI cut-off for the species boundary is $\geq 95\%$, indicating that *W. koreensis* strain HJ and KCTC 3621^T are the same species. On the other hand, this strain shared 68.88 to 79.54% sequence similarities with genomes of other *Weissella* species. According to the dDDH calculation, *W. koreensis* strain HJ shared 92% sequence identity with the *W. koreensis* type strain (KCTC 3621^T), while other *Weissella* species shared sequence identity ranging from 20.0 to 31.7% below the species threshold (70.0%).

The genome of *W. koreensis* strain HJ contained probiotic-related genes, such as the F0F1 ATP synthases

(accession no. MCZ9310942.1 to MCZ9310949.1) genes related to acid tolerance and *ppaC* (MCZ9310418.1) gene related to bile tolerance. Probiotic strains can be subjected to heat stress in the food industry, hence they must be able to tolerate heat shock when exposed to high temperatures [9]. The heat shock protein *htpX* (MCZ9310452.1), *hrcA* (MCZ9311479.1), *hslO* (MCZ9311177.1), *dnaK* (MCZ9311481.1), *dnaJ* (MCZ9311482.1), *ctsR* (MCZ9311184.1), *grpE* (MCZ9311480.1), *groES* (MCZ9311630.1), and *groEL* (MCZ9311631.1), which participates in the heat shock response, were identified in this genome. In addition to heat shock protein, cold shock protein (MCZ9311541.1) related to the survival of bacteria at low temperature was identified in the genome of *W. koreensis* strain HJ. Additionally, the oxidative stress-resistant genes *fur* (MCZ9310783.1), *nmr* (MCZ9311702.1), and *nrdH* (MCZ10552.1) were identified. Adhesion to intestinal epithelium is an essential feature to exert their effectiveness of probiotics and plays an important role in the interaction between bacteria and the host [9]. *W. koreensis*

strain HJ encodes cell-surface proteins, such as elongation factor Tu (MCZ9311332.1), sortase (MCZ9310935.1), and cell surface protein (MCZ9310653.1). Moreover, genes (*epsC* and *epsD*) involved in exopolysaccharides synthesis related to adhesion were contained within the genome [10]. In addition, *W. koreensis* strain HJ encoded a type III polyketide synthase (T3PKS) cluster that produces secondary metabolites with diverse biological activities, including antimicrobials.

To analyze *in silico* safety assessment, ResFinder and VirulenceFinder databases were used to identify the antibiotic resistance and virulence genes, the results showed that antibiotic resistance and virulence genes were not detected in the genome of *W. koreensis* strain HJ. Also, in this genome, mobile gene elements related to horizontal gene transfer that promote the transfer of various genes were absent. Therefore, the antibiotic resistance and virulence genes are not transmitted from *W. koreensis* strain HJ to pathogenic microorganisms in the gastrointestinal tract.

The pan-genome of *W. koreensis* strains was analyzed using BPGA software. The five *W. koreensis* yield a pan-genome of 1,610 genes (50.0%), core-genome of 1,132 genes (35.16%), accessory-genome of 278 genes (8.63%), and unique-genome of 200 genes (6.21%). In the core-genome, 979 genes were assigned to 20 COG categories, such as general function prediction only (R, 13.37%) and translation, ribosomal structure and biogenesis (J, 12.0%). In the unique-genome, 63 genes were assigned in 15 COG categories. There were a total of 14 genes enriched in transcription (K, 19.15%), containing transcriptional regulators that can act on specific genes to regulate their expression and confer benefits when present in the gut [9]. Also, 11 genes were assigned in cell wall/membrane/envelope biogenesis (M, 15.05%). Among the 200 unique genes, *W. koreensis* strain HJ had 68 unique genes, including 30 genes with hypothetical proteins of unknown function and 38 genes with known functions, such as low temperature requirement protein A (MCZ9310806.1), *Rrf2* family transcriptional regulator (MCZ9311114.1), *ImmA/IrrE* family metallo-endopeptidase (MCZ9310437.1), and *IspD/TarI* family cytidyltransferase (MCZ9310875.1).

Acknowledgments

This work was funded by the Strategic Initiative for Microbiomes in Agriculture and Food, Ministry of Agriculture, Food and Rural Affairs, Republic of Korea (Grant number 918005-4).

Conflict of Interest

The authors have no financial conflicts of interest to declare.

References

- Jeong SE, Chun BH, Kim KH, Park D, Roh SW, Lee SH, *et al.* 2018. Genomic and metatranscriptomic analyses of *Weissella koreensis* reveal its metabolic and fermentative features during kimchi fermentation. *Food Microbiol.* **76**: 1-10.
- Apostolakos I, Paramithiotis S, Mataragas M. 2022. Functional and safety characterization of *Weissella paramesenteroides* strains isolated from dairy products through whole-genome sequencing and comparative genomics. *Dairy* **3**: 799-813.
- Kim E, Yang S-M, Kim H-Y. 2023. *Weissella* and the two Janus faces of the genus. *Appl. Microbiol. Biotechnol.* **107**: 1119-1127.
- Kang BK, Cho MS, Ahn TY, Lee ES, Park DS. 2015. The influence of red pepper powder on the density of *Weissella koreensis* during kimchi fermentation. *Sci. Rep.* **5**: 15445.
- Anandharaj M, Sivasankari B, Santhanakaruppu R, Manimaran M, Rani RP, Sivakumar S. 2015. Determining the probiotic potential of cholesterol-reducing *Lactobacillus* and *Weissella* strains isolated from gherkins (fermented cucumber) and south Indian fermented koozh. *Res. Microbiol.* **166**: 428-439.
- Lee J-S, Lee KC, Ahn J-S, Mheen T-I, Pyun Y-R, Park Y-H. 2002. *Weissella koreensis* sp. nov., isolated from kimchi. *Int. J. Syst. Evol. Microbiol.* **52**: 1257-1261.
- Lee JH, Bae JW, Chun J. 2012. Draft genome sequence of *Weissella koreensis* KCTC 3621T. *J. Bacteriol.* **194**: 5711-5712.
- Mun SY, Moon SH, Chang HC. 2020. Characterization of high-ornithine-producing *Weissella koreensis* DB1 isolated from kimchi and its application in rice bran fermentation as a starter culture. *Foods* **9**: 1545.
- Kim E, Yang SM, Kim D, Kim HY. 2022. Complete genome sequencing and comparative genomics of three potential probiotic strains, *Lactocaseibacillus casei* FBL6, *Lactocaseibacillus chiayiensis* FBL7, and *Lactocaseibacillus zeae* FBL8. *Front. Microbiol.* **12**: 794315.
- Kandasamy S, Yoo J, Yun J, Lee KH, Kang HB, Kim JE, *et al.* 2022. Probiogenomic in-silico analysis and safety assessment of *Lactiplantibacillus plantarum* DJF10 strain isolated from Korean raw milk. *Int. J. Mol. Sci.* **23**: 14494.