Current Microbiology (2020) 77:1871–1882
https://doi.org/10.1007/s00284-020-02000-8

# Complete Genome Sequence of *Lactobacillus plantarum* EM, A Putative Probiotic Strain with the Cholesterol-Lowering Effect and Antimicrobial Activity

**Eiseul Kim[1] · Hae Choon Chang[2] · Hae-Yeong Kim[1]**

Received: 10 January 2020 / Accepted: 15 April 2020 / Published online: 22 April 2020


**Abstract**
*Lactobacillus plantarum* EM is a probiotic strain with antimicrobial activity, cholesterol-lowering effects, and tolerance to acid and bile. To understand the genetic basis of the probiotic characteristics of this strain, genome sequencing and probiotic-related genetic analysis were performed. The genomic characteristics of *L. plantarum* EM were confirmed by comparative genomic analysis with 41 probiotic lactic acid bacteria, including 10 *L. plantarum* strains. *L. plantarum* EM was shown to contain a circular chromosome of 3,184,808 bp and eight plasmids with various lengths from 5,027 to 76,369 bp. The *L. plantarum* EM genome had a total of 3560 protein-coding genes, including probiotic-related genes, such as tolerance to acid and bile, temperature stress, and oxidative stress. Comparative genomic analysis showed that *L. plantarum* EM contained plantaricin and bovicin gene clusters, which are related to antimicrobial activity, and five bile salt hydrolase genes related to serum cholesterol-lowering effects. The genomic analysis confirmed the probiotic properties of *L. plantarum* EM, and our results indicated that this strain has potential application for use as an industrially important probiotic.

## Introduction

The *Lactobacillus plantarum* species constitutes extremely flexible and versatile lactic acid bacteria (LAB), which have been isolated from many different environmental niches, such as animals, plants, and the gastrointestinal and vaginal tract, as well as various food materials, such as vegetables, dairy products, meat products, and fermented foods [1, 2]. *L. plantarum* is applied to a variety of fermented foods, and some strains are used as probiotics that may confer beneficial health effects to humans or animals [3].

Probiotics are living microorganisms that provide beneficial effects to the host and are used to prevent a variety of diseases associated with diarrhea, hyperlipidemia, inflammatory bowel disease, and immune function [4, 5]. In the genus *Lactobacillus*, some strains of the species, such as *L. acidophilus*, *L. gasseri*, *L. rhamnosus*, *L. plantarum*, and *L. fermentum*, act as important probiotics [6]. To function as a probiotic, a bacterial strain should be resistant to bile and the acidity of the gastrointestinal tract to enter the small intestine. Other functional properties for characterizing probiotics are the ability to produce antimicrobial compounds and reduce serum cholesterol levels [6, 7]. Cholesterol-lowering effects are closely related to the bile salt hydrolase (bsh). Bile acid conjugated with taurine or glycine helps to absorb cholesterol in the small intestine. However, when bile acid is removed by bacterial bsh, bile acid is excreted and cholesterol is consumed as a precursor for the synthesis of new bile acid, thereby lowering serum cholesterol [8]. The bsh activity present in microorganisms has been reported in strains, such as *Bifidobacterium*, *Lactobacillus,* and *Streptococcus*, and contributes to the probiotic properties in the gastrointestinal tract of humans and animals [9]. One of the antimicrobial compounds, bacteriocin, is an antimicrobial peptide synthesized in ribosomes and works against closely related species [10]. Plantaricin is a bacteriocin produced by *L. plantarum*, most of which, such as plantaricin A and the two-peptide bacteriocins, plantaricin EF and plantaricin JK, belong to class IIc. Some plantaricins have antimicrobial activity against both gram-negative and gram-positive bacteria, indicating

✉ Hae-Yeong Kim
hykim@khu.ac.kr

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

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

the potential of *L. plantarum* as an antimicrobial agent [11]. A general mechanism for the probiotic effects may be related to the genus or species of bacteria, but specific mechanisms tend to be strain-specific [12]. Thus, genome sequencing is the best way to identify the metabolic pathways, phylogenetic relationships, the health and safety of specific strains, and genetically understand the biological specificity of new strains [13, 14].

The previous researcher isolated *L. plantarum* EM from kimchi, a traditional Korean food, and this strain has been shown to reduce serum cholesterol levels [15]. *L. plantarum* EM has been shown to meet the functional criteria required for probiotics, such as bile and acid tolerance, antimicrobial activity against pathogenic bacteria and fungi, and antibiotic susceptibility [15]. Here, we performed genome sequencing and comparative genomic analysis to uncover the mechanism of the probiotic effect of *L. plantarum* EM.

## Materials and Methods

### Strain Isolation and DNA Extraction

Before use, *L. plantarum* EM was activated in MRS broth (Difco, Becton & Dickinson, Sparks, MD, USA) at 30 °C for 48 h under anaerobic conditions. The genomic DNA of *L. plantarum* EM was extracted with a DNeasy Blood and Tissue kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The total genomic DNA purity and concentration were determined by absorbance using an Ultrospec 2100 Pro-spectrophotometer (Amesham Biosciences, Cambridge, UK) [16].

### Genome Sequencing, Assembly, Annotation, and Analysis

The genome sequencing of *L. plantarum* EM was performed using the PacBio RS II platform (Pacific Biosciences, Menlo Park, CA, USA). A 20 kb library was generated using a SMARTbell Template Preparation Kit 1.0 and sequenced with P4-P2 chemistry on two cells. The raw data were obtained as 91,147 reads with an average read length of 10,774 bp. The filtered subreads were de novo assembled using HGAP version 2.0. Functional annotation of the genome sequence was performed using RAST version 2.0 [17], and probiotic-related genes were identified based on the annotation results. A circular genomic map was constructed using the CGView server [18]. The ResFinder version 3.1 was used to identify the antibiotic resistance genes in plasmids from pEM1 to pEM8 [19].

### Comparative Genomic Analysis of Probiotic *Lactobacillus* Species

To confirm the genetic characteristics of *L. plantarum* EM, a comparative genomic analysis was performed with 41 probiotic *Lactobacillus* strains. The probiotic *Lactobacillus* strains used were those with probiotic functions identified in previous studies, including *L. acidophilus*, *L. brevis*, *L. rhamnosus*, *L. paracasei*, *L. casei*, *L. fermentum*, *L. helveticus*, *L. plantarum*, and *L. reuteri*. The genomes of these strains were retrieved from the National Center for Biotechnology Information (NCBI; ftp://ftp.ncbi.nlm.nih.gov/genomes/) genome database (Table 1). For estimation of the phylogenetic tree, the 16S rRNA gene sequences extracted from the genome sequence of 42 probiotic *Lactobacillus* strains were aligned using ClustalW with default parameters, and phylogenetic analysis was performed using the maximum-likelihood method with 1,000 bootstraps in MEGA (version 6.06). The genes related to bacteriocin synthesis in each *Lactobacillus* species were identified using BAGEL4 [20]. The genes related to probiotic properties and cholesterol-lowering were identified using BLASTp. The bsh genes, the genes related to cholesterol-lowering effects, were extracted from the genome sequence of 21 *L. plantarum* strains, and the alignment and phylogenetic tree were constructed using the ETE3 module with its default parameters for protein sequences [21]. The pan-genome analysis and visualization of the probiotic *Lactobacillus* strains were analyzed using Anvi'o version 6.0 pan-genomic workflow [22, 23]. The number of pan-, core-, accessory-, and unique-genomes were analyzed using computational pipeline Bacterial Pan Genome Analysis (BPGA) version 1.3 with the default parameters [24]. In order to classify the genomes of each strain by functional categories, clusters of orthologous groups (COGs) were assigned to the amino acid sequences using USEARCH version 8.0 against the COG database [25].

### Nucleotide Sequence Accession Numbers

The genome sequence of *Lactobacillus plantarum* EM was deposited in the DDBJ/EMBL/GenBank with accession numbers CP037429.1 (chromosome), CP037430.1 (pEM1), CP037431.1 (pEM2), CP037432.1 (pEM3), CP037433.1 (pEM4), CP037434.1 (pEM5), CP037435.1 (pEM6), CP037436.1 (pEM7), and CP037437.1 (pEM8).

## Results and Discussion

### General Genome Features

The complete genome of *L. plantarum* EM was composed of a circular chromosome and eight plasmids (Fig. 1).

**Table 1** General genome features of 42 *Lactobacillus* strains

| Species | Size (Mb) | GC (%) | Gene | Protein | Accession no |
|---|---|---|---|---|---|
| *L. acidophilus* NCFM | 1.99356 | 34.7 | 1927 | 1832 | CP000033.3 |
| *L. acidophilus* La-14 | 1.99158 | 34.7 | 1978 | 1862 | CP005926.2 |
| *L. acidophilus* LA1 | 1.9912 | 34.7 | 2002 | 1886 | CP017062.1 |
| *L. brevis* KB290 | 2.58788 | 45.6 | 2610 | 2457 | AP012167.1 |
| *L. brevis* 100D8 | 2.47773 | 45.8 | 2509 | 2355 | CP015338.1 |
| *L. brevis* TMW 1.2108 | 2.91798 | 45.3 | 2926 | 2738 | CP019734.1 |
| *L. brevis* TMW 1.2111 | 2.88201 | 45.3 | 2884 | 2506 | CP019743.1 |
| *L. rhamnosus* LOCK900 | 2.88338 | 46.8 | 2901 | 2734 | CP005484.1 |
| *L. rhamnosus* LOCK908 | 2.9909 | 46.8 | 2988 | 2811 | CP005485.1 |
| *L. rhamnosus* DSM 14870 | 3.01315 | 46.7 | 3057 | 2842 | CP006804.1 |
| *L. rhamnosus* Pen | 2.88497 | 46.8 | 2908 | 2719 | CP020464.1 |
| *L. rhamnosus* GG | 3.01011 | 46.7 | 3062 | 2860 | FM179322.1 |
| *L. paracasei* N1115 | 3.06428 | 46.5 | 3263 | 2952 | CP007122.1 |
| *L. paracasei* L9 | 3.07644 | 46.3 | 3169 | 2926 | CP012148.1 |
| *L. paracasei* CAUH35 | 2.97335 | 46.3 | 3153 | 2777 | CP012187.1 |
| *L. casei* LC5 | 3.13287 | 47.9 | 3093 | 2893 | CP017065.1 |
| *L. casei* BL23 | 3.0792 | 46.3 | 3236 | 3029 | FM177140.1 |
| *L. casei* W56 | 3.1321 | 46.3 | 3326 | 3010 | HE970764.1 |
| *L. fermentum* CECT 5716 | 2.10045 | 51.5 | 2177 | 1631 | CP002033.1 |
| *L. fermentum* F-6 | 2.06462 | 51.7 | 2074 | 1874 | CP005958.1 |
| *L. fermentum* SNUV175 | 2.27233 | 51.1 | 2304 | 2108 | CP019030.1 |
| *L. helveticus* H9 | 1.87112 | 37.0 | 1936 | 1546 | CP002427.1 |
| *L. helveticus* R0052 | 2.12921 | 36.8 | 2173 | 1761 | CP003799.1 |
| *L. helveticus* KLDS1.8701 | 2.10663 | 36.9 | 2172 | 1740 | CP009907.1 |
| *L. helveticus* MB2-1 | 2.08406 | 36.9 | 2204 | 1771 | CP011386.1 |
| *L. helveticus* CAUH18 | 2.16058 | 36.8 | 2219 | 1837 | CP012381.1 |
| *L. helveticus* D76 | 2.05832 | 37.0 | 2097 | 1694 | CP016827.1 |
| *L. helveticus* D75 | 2.05307 | 37.0 | 2092 | 1693 | CP020029.1 |
| *L. plantarum* WCFS1 | 3.34862 | 44.5 | 3174 | 3063 | AL935263.2 |
| *L. plantarum* ST-III | 3.30794 | 44.5 | 3194 | 3020 | CP002222.1 |
| *L. plantarum* ZJ316 | 3.29976 | 44.4 | 3209 | 2972 | CP004082.1 |
| *L. plantarum* 16 | 3.36102 | 44.3 | 3281 | 3076 | CP006033.1 |
| *L. plantarum* ATCC14917 | 3.21226 | 44.5 | 3094 | 2922 | ACGZ02 |
| *L. plantarum* CAUH2 | 3.27461 | 44.5 | 3188 | 3010 | CP015126.1 |
| *L. plantarum* LZ206 | 3.26371 | 44.5 | 3182 | 2891 | CP015966.1 |
| *L. plantarum* KLDS1.0391 | 2.91091 | 44.7 | 2918 | 2691 | CP019348.1 |
| *L. plantarum* 10CH | 3.31106 | 44.5 | 3192 | 3013 | CP023728.1 |
| *L. plantarum* EM | 3.61869 | 44.2 | 3706 | 3618 | CP037429.1 |
| *L. reuteri* SKKU-OGDONS-01 | 2.25997 | 38.9 | 2164 | 1944 | CP029615.1 |
| *L. reuteri* ZLR003 | 2.2341 | 38.7 | 2217 | 2024 | CP014786.1 |
| *L. reuteri* WHH1689 | 2.04418 | 39.3 | 2067 | 1704 | CP027805.1 |
| *L. reuteri* I5007 | 2.09328 | 38.9 | 2091 | 1910 | CP006011.1 |

The complete genome of *L. plantarum* EM consisted of 3,618,689 bp with a G+C content of 44.2% (Table 1). The genome had a chromosome of 3,184,808 bp with a G+C content of 44.7%. The plasmids, designated pEM1 to pEM8, had various lengths ranging from 21,426 to 76,369 bp. The genome size and G+C content of the *L. plantarum* EM chromosome were similar to *L. paracasei* N1115 (3,064,279 bp, 46.46%), *L. rhamnosus* DSM 14870 (3,013,149 bp, 46.7%), and *L. casei* LC5 (3,132,867 bp, 47.9%), but not to *L. fermentum* F-6 (2,064,620 bp, 51.7%) or *L. acidophilus* NCFM (1,993,560 bp, 34.7%). Among the species used for analysis, *L. plantarum* had the largest genome and the greatest number of plasmids. This fact is related to the ecological flexibility of *L. plantarum* and

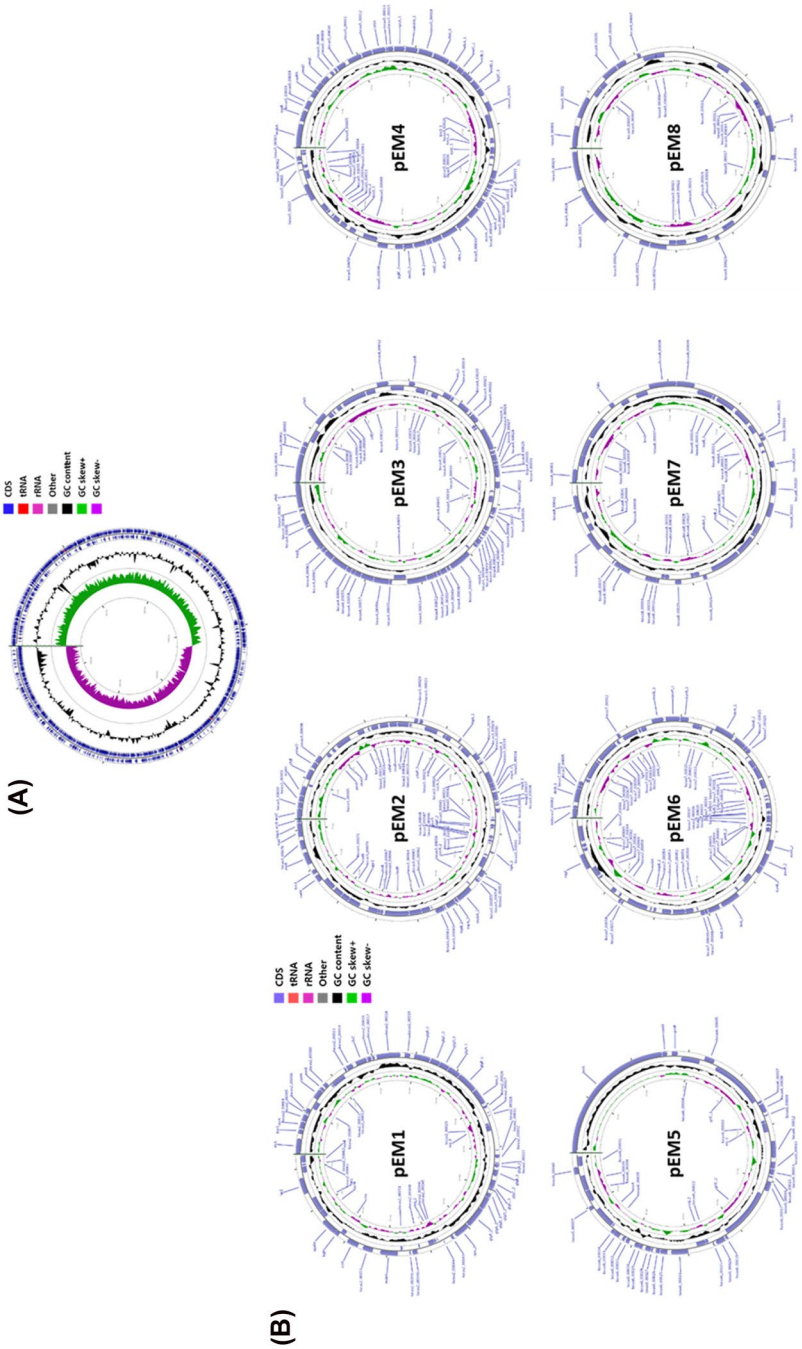


**Fig. 1** Circular genome maps of *L. plantarum* EM **a** Chromosome, **b** Plasmids

the diversity of ecological niches in which *L. plantarum* is encountered [26]. And, in general, *Lactobacillus* reduced the genome size by removing useless functions to adapt to the environment during evolution, whereas *L. plantarum* has a larger genome obtained by horizontal gene transfer via mobile elements, such as plasmids, transposons, prophages, and integrons [27].

The nucleotide sequence blast results revealed that the eight plasmids of *L. plantarum* EM showed similarity to the plasmids or chromosome of the *L. coryniformis*, *L. plantarum*, *L. pentosus*, and *L. curvatus* strains. It was also confirmed that each plasmid had a gene related to a plasmid replication protein. The annotation results showed that the genome had 3,107 coding sequences and 88 RNA genes. Moreover, the protein-coding sequences were functionally divided into 238 SEED subsystem categories. The plasmids of *L. plantarum* EM contained from 29 to 79 different protein-coding genes. The ResFinder database were used to identify the antibiotic resistance genes. The results showed that no antibiotic resistance genes were detected in any of the plasmids. Therefore, in the gastrointestinal tract, antibiotic resistance genes are not expected to be transmitted from *L. plantarum* EM strains to pathogenic microorganisms.

## Probiotic-Related Genes of *L. plantarum* EM

The probiotic properties of *L. plantarum* EM were confirmed in a previous study [15]. This was supported by the genomic analysis data in our study, in which a gene encoding F0F1 ATB syntheses (chr_orf2044 to chr_orf2050), which are related to acid tolerance, and choloylglycine hydrolases (chr_orf56, chr_orf57, chr_orf2236, chr_orf2913, chr_orf3049), which are related to bile salt resistance, were detected (Table 2). Probiotics can experience heat stress in the food industry (e.g., pasteurization and spray-drying) or during storage. Exposure to high temperatures induces the expression of evolutionarily conserved heat shock proteins (HSPs), including chaperones, such as GrpE, DnaK, DnaJ, and GroES/GroEL [28]. The heat shock protein GrpE (chr_orf1738) and the chaperone proteins DnaK (chr_orf1737), DnaJ (chr_orf1736), and GroES/GroEL (chr_orf638 to chr_orf639), which participate in the heat shock response and hyperosmotic response, were detected in the chromosome of *L. plantarum* EM. Cold shock-inducing proteins have been identified in a variety of microorganisms, and these genes are related to the adaptation process required for bacterial survival at low temperatures [29]. In *L. plantarum* EM, the cold shock protein of the CSP family genes was found on the chromosome (chr_orf31, chr_orf886, chr_orf1025). Additionally, catalase katE (chr_orf3077), thiol peroxidase (chr_orf2002), and glutathione peroxidase (chr_orf194), which protect against oxidative stress, were detected.

## Pan-Genomic Analysis of 42 Probiotic *Lactobacillus* Strains

The taxonomic relationship between *L. plantarum* EM and other probiotic *Lactobacillus* species was confirmed by 16S rRNA gene sequence. Phylogenetic tree analysis revealed that *L. plantarum* EM was grouped with *L. plantarum* strains (Fig. 2). The 16S rRNA gene sequence of *L. plantarum* EM was most closely related to ST-III, 10CH, and WCFS1 (100% identity) among the *L. plantarum* strains. Hence, based on the phylogenetic relationship analysis, the EM strain was identified as *L. plantarum*.

To understand the genome of probiotic *Lactobacillus* species and to obtain the unique genes of *L. plantarum* EM, we performed a pan-genome analysis. The pan-genome analysis of 42 *Lactobacillus* strains showed that the remaining strains, except *L. casei* and *L. paracasei* strains, were grouped according to each species (Fig. 3). Based on a comparative genomic analysis of 42 genome sequences of probiotic *Lactobacillus* species, the pan-, accessory-, and core-genome encompassed 15,020, 10,877, and 114 genes, respectively. To investigate the diversity and functionality encoded by the pan-genome, the genes were classified by functional categories using COG analysis. The core-genome was assigned a high percentage of genes for translation, ribosomal structure and biogenesis, and the accessory-genome had the highest percentage of genes for general function prediction and transcription. It contains probiotic-related genes, such as choloylglycine hydrolase, that function in bile resistance. These results suggest that all the strains used for the analysis had probiotic-related genes because they have already been confirmed as probiotic bacteria. Of the 4,029 unique genes identified in the 42 *Lactobacillus* strains, 83 genes were identified as unique genes present only in the *L. plantarum* EM genome. The unique genes identified were those involved in replication, recombination and repair (20.51%), transcription (15.38%), and carbohydrate transport and metabolism (12.82%).

## Genetic Analysis Related to the Cholesterol-Lowering Effect

High cholesterol-removing ability was observed in the *L. plantarum* EM strain [15]. This ability was supported by our genomic analysis results, in which a total of five bsh genes were detected. *L. plantarum* ST-III is a highly cholesterol-resistant strain with four bsh genes on the genome, and the function of these genes was demonstrated in a previous study [30]. As a result of the alignment of the bsh genes of *L. plantarum* ST-III and EM, the bsh1, bsh3, and bsh4 genes of ST-III showed 98–100% identity to chr_orf2191, chr_orf2855, and chr_orf2990 of EM, respectively. Compared to the bsh2 gene of *L. plantarum* ST-III, eleven nucleotide substitutions

**Table 2** Important genes encoding probiotic-related proteins in *L. plantarum* EM

| Locus_tag | Product | EC number | Stress response |
|---|---|---|---|
| chr_orf2044 | ATP synthase epsilon chain acid stress | 3.6.3.14 | Acid tolerance |
| chr_orf2045 | ATP synthase beta chain | 3.6.3.14 | Acid tolerance |
| chr_orf2046 | ATP synthase gamma chain acid stress | 3.6.3.14 | Acid tolerance |
| chr_orf2047 | ATP synthase subunit alpha acid stress | 3.6.3.14 | Acid tolerance |
| chr_orf2048 | ATP synthase subunit delta acid stress | 3.6.3.14 | Acid tolerance |
| chr_orf2049 | ATP synthase subunit b acid stress | – | Acid tolerance |
| chr_orf2051 | ATP synthase subunit a acid stress | 3.6.3.14 | Acid tolerance |
| chr_orf2050 | ATP synthase subunit c acid stress | 3.6.3.14 | Acid tolerance |
| chr_orf56 | Choloylglycine hydrolase | 3.5.1.24 | Bile tolerance |
| chr_orf57 | Choloylglycine hydrolase | 3.5.1.24 | Bile tolerance |
| chr_orf2236 | Choloylglycine hydrolase | 3.5.1.24 | Bile tolerance |
| chr_orf2913 | Choloylglycine hydrolase | 3.5.1.24 | Bile tolerance |
| chr_orf3049 | Choloylglycine hydrolase | 3.5.1.24 | Bile tolerance |
| chr_orf638 | Heat shock protein 60 family co-chaperone GroES | – | Temperature stress |
| chr_orf639 | Heat shock protein 60 family chaperone GroEL | – | Temperature stress |
| chr_orf1738 | Heat shock protein GrpE | – | Temperature stress |
| chr_orf460 | Ribosome-associated heat shock protein | – | Temperature stress |
| chr_orf1736 | Chaperone protein DnaJ | – | Temperature stress |
| chr_orf1737 | Chaperone protein DnaK | – | Temperature stress |
| chr_orf111, orf2905 | Small heat shock protein | – | Temperature stress |
| chr_orf31, orf886, orf1025 | Cold shock protein of CSP family | – | Temperature stress |
| chr_orf3077 | Catalase KatE | 1.11.1.6 | Oxidative stress |
| chr_orf2002 | Thiol peroxidase, Tpx-type | 1.11.1.15 | Oxidative stress |
| chr_orf194 | Glutathione peroxidase | 1.11.1.15 | Oxidative stress |
| chr_orf105 | Halo peroxidase | – | Oxidative stress |
| chr_orf674, orf2248 | Thioredoxin reductase | 1.8.1.9 | Oxidative stress |
| chr_orf323, orf1083, orf1565, orf2833 | Similar to glutathione reductase | – | Oxidative stress |
| chr_orf209, orf1959, orf2284, orf2963 | Thioredoxin | – | Oxidative stress |
| chr_orf667, orf1244, orf1680, orf2213, orf2974 | NADH peroxidase Npx | 1.11.1.1 | Oxidative stress |
| chr_orf759, orf761, orf2281, orf3086, orf3088 | Pyruvate oxidase | 1.2.3.3 | Oxidative stress |

were found in chr_orf56 and chr_orf57 of EM. At the 475 bp position of chr_orf56 and chr_orf57 in *L. plantarum* EM, the TGG for tryptophan was replaced with TAG, causing a premature stop codon. As a result, the bsh gene was divided into two fragments and a total of five bsh genes were present. Comparison of the bsh gene of 23 *L. plantarum* strains showed that they mainly had one bsh gene, similar to bsh1 of *L. plantarum* ST-III. The bsh2, bsh3, and bsh4 genes were present only in strains with three or more bsh genes (Fig. 4). A previous study showed that all bsh genes of *L. plantarum* ST-III were responsible for the hydrolysis activity of many substrates, and the bsh1 gene was highly activate against glycodeoxycholic acid [30].

## Identification of Bacteriocin Gene Clusters

The bacteriocin synthesis gene clusters of *Lactobacillus* species were compared and analyzed. As a result, one or more bacteriocin gene clusters were found in all species except the *L. brevis* and *L. fermentum* strains (Table 3). *L. rhamnosus*, *L. helveticus*, and *L. plantarum* mostly contained carnocin, helveticin, and plantaricin, respectively. *L. casei* and *L. paracasei* mainly contained LSEI bacteriocin derived from *L. casei* ATCC 334, and *L. acidophilus* mainly contained acidocin and helveticin. Bacteriocin gene clusters seemed to be among those that are transferred horizontally, showing similar patterns between closely related genomes [31]. The genome of *L. plantarum* EM consists of encoding genes involved in bovicin (gene start position, 1 bp on the plasmid) and plantaricin (gene start position, 353,842 bp on the chromosome), i.e., plantaricin JK, N, A, and EF (Fig. 5). Plantaricin F was previously identified in probiotic *L. plantarum* with antimicrobial activity against *Micrococcus, Listeria, Staphylococcus,* and *Salmonella* [32]. In an in vitro assay, *L. plantarum* EM showed antimicrobial activity against *Bacillus cereus, Micrococcus*

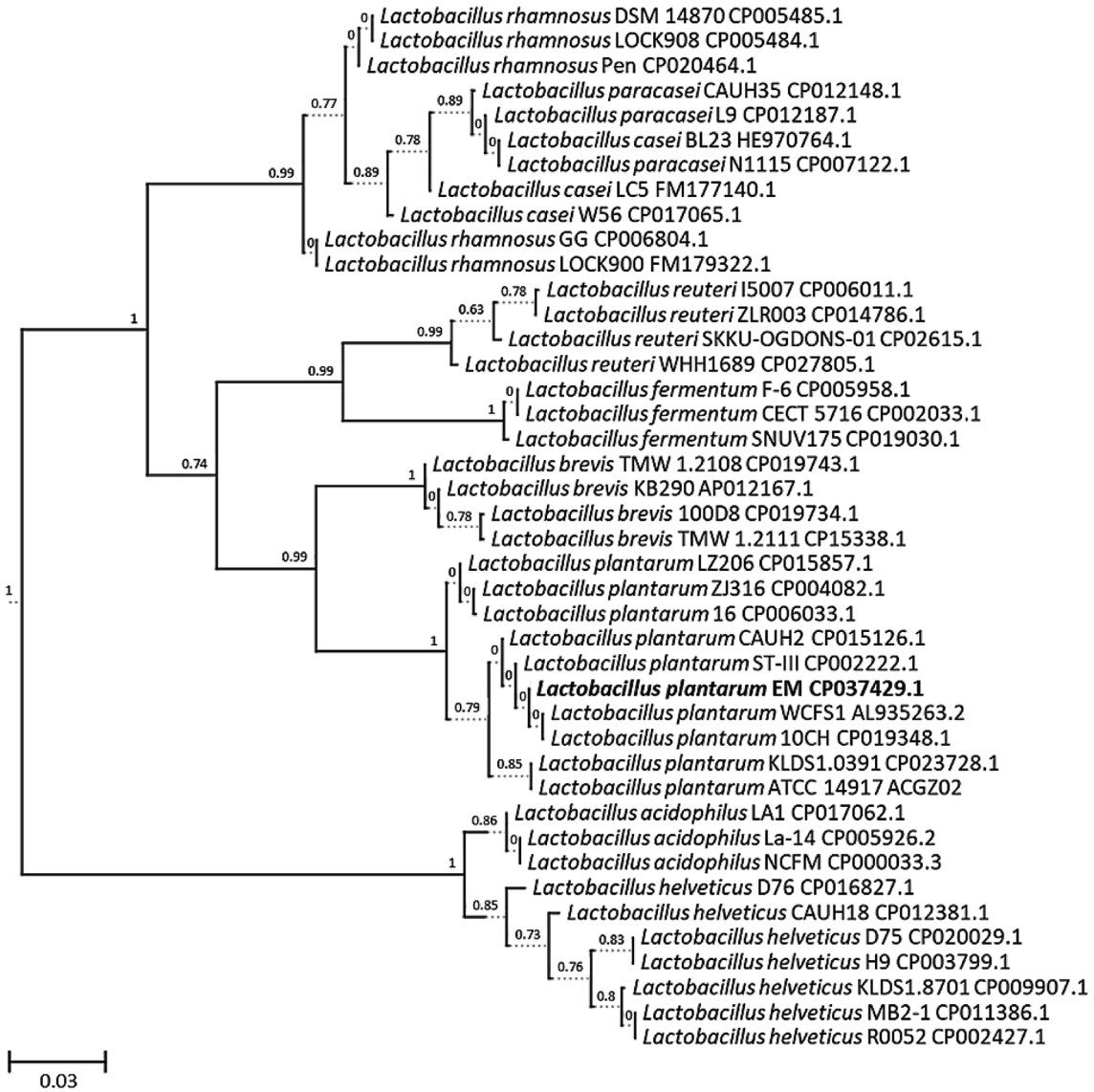


**Fig. 2** Phylogenetic analysis was based on 16S rRNA gene sequences for 42 probiotic *Lactobacillus* strains

*leteus, Staphylococcus aureus, Escherichia coli, Salmonella* Typhi, *Vibrio parahaemolyticus*, and *Pseudomonas aeruginosa* [15]. These results were assumed to be related to the bacteriocin gene cluster present on the genome of *L. plantarum* EM.

In this study, we performed genome sequencing and analysis of *L. plantarum* EM, which has already confirmed probiotic properties. The genome sequence of *L. plantarum* EM provided genetic information on probiotic-related functions, such as cholesterol-lowering, antimicrobial activity, and tolerance to bile and acid. The *bsh* gene, bacteriocin gene cluster, and F0F1 ATB syntheses were identified through genomic analysis of *L. plantarum* EM. This strain may be used in foods or industries as a probiotic for human health.

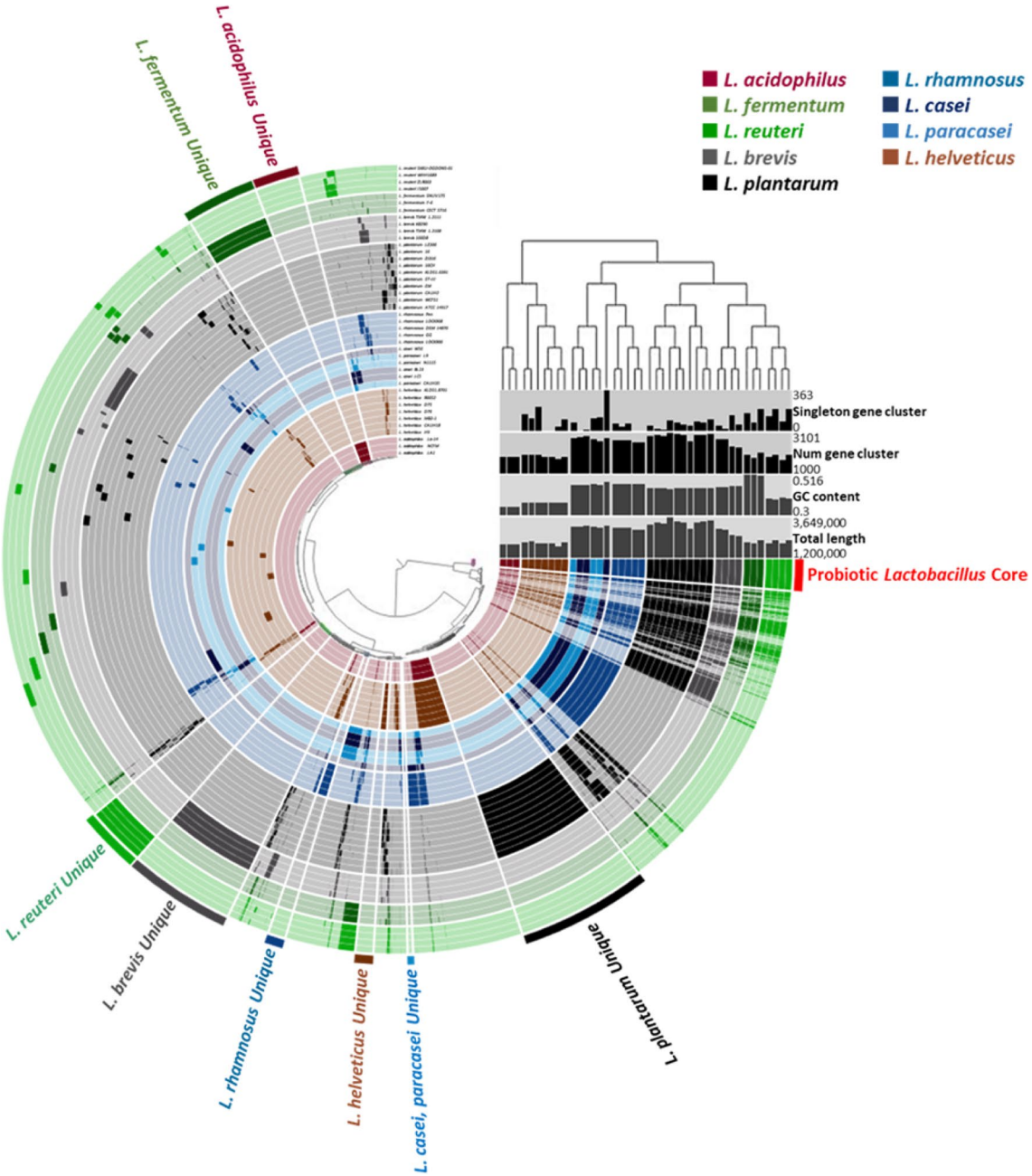


**Fig. 3** Pan-genome distribution across 42 *Lactobacillus* species. The center figure shows the hierarchical clustering of pan-genome based on their presence/absence

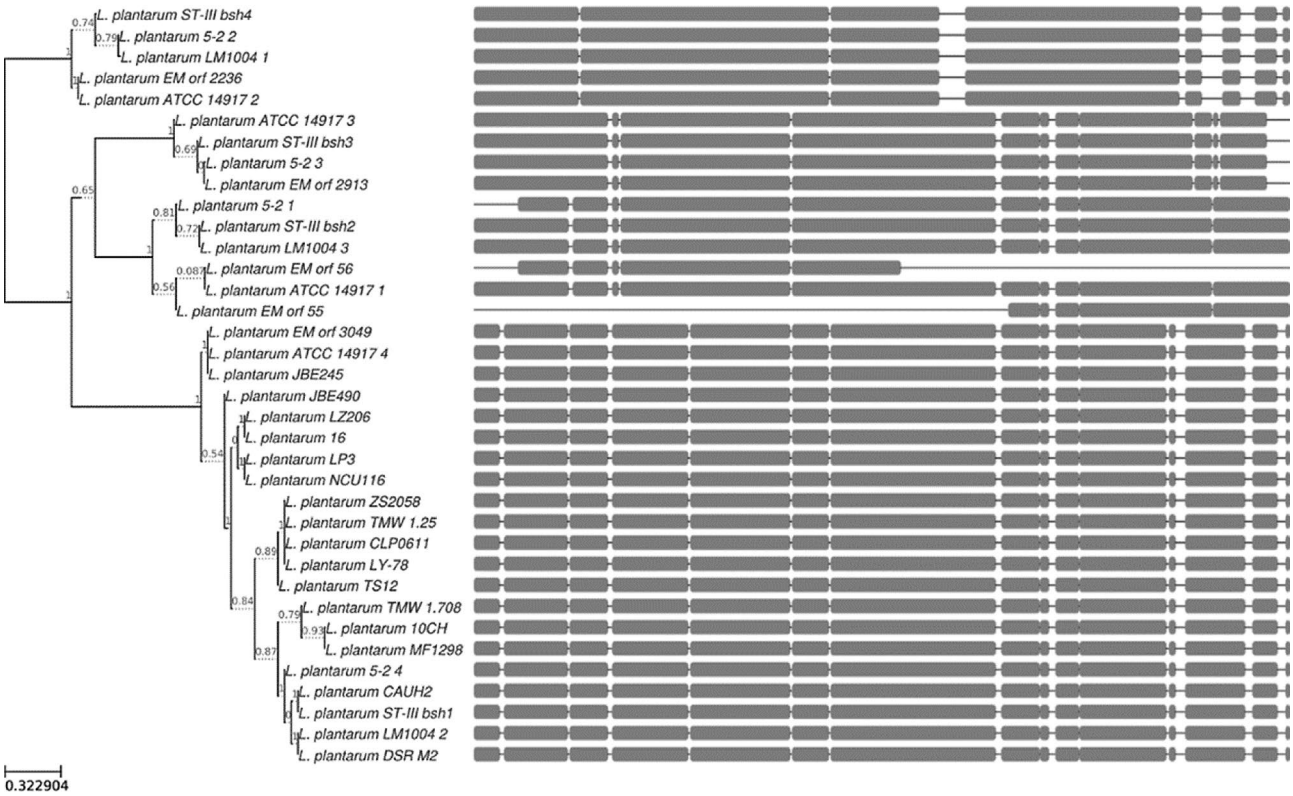


**Fig. 4** Phylogenetic analysis was based on the bile salt hydrolase genes for 21 *L. plantarum* strains

**Table 3** Putative bacteriocin gene cluster identified in *Lactobacillus* species

| Strain | Size (bp) | Class (bacteriocin gene name) |
|---|---|---|
| *L. acidophilus* NCFM | 25,766 | Acidocin J1132 (Enterocin X, Acidocin J1132) |
| | 20,483 | Enterolysin A (Enterolysin A) |
| | 20,933 | Helveticin J (Helveticin J) |
| *L. acidophilus* La-14 | 25,769 | Acidocin J1132 (Enterocin X, Acidocin J1132) |
| | 20,483 | Enterolysin A (Enterolysin A) |
| | 20,933 | Helveticin J (Helveticin J) |
| *L. acidophilus* LA1 | 25,766 | Acidocin J1132 (Enterocin X, Acidocin J1132) |
| | 20,483 | Enterolysin A (Enterolysin A) |
| | 20,933 | Helveticin J (Helveticin J) |
| *L. rhamnosus* LOCK900 | 25,853 | Carnocin CP52 (Enterocin X, Carnocin CP52) |
| *L. rhamnosus* LOCK908 | 25,850 | Carnocin CP52 (Enterocin X, Carnocin CP52, LSEI 2386) |
| *L. rhamnosus* DSM 14870 | 25,850 | Carnocin CP52 (Enterocin X, Carnocin CP52) |
| *L. rhamnosus* Pen | 25,853 | Carnocin CP52 (Enterocin X, Carnocin CP52) |
| *L. rhamnosus* GG | 25,853 | Carnocin CP52 (Enterocin X, Carnocin CP52, LSEI 2386) |
| *L. paracasei* N1115 | 20,114 | LSEI 2163 (LSEI 2163) |
| | 33,617 | LSEI 2386 (LSEI 2386, Enterocin X, Carnocin CP52) |
| *L. paracasei* L9 | 20,231 | Thermophilin A (Thermophilin A) |
| | 30,992 | LSEI 2386 (LSEI 2386, Enterocin X, Carnocin CP52) |
| *L. paracasei* CAUH35 | 28,577 | LSEI 2386 (LSEI 2386, Enterocin X, Carnocin CP52) |
| *L. casei* LC5 | 32,672 | Carno bacteriocin A (Carno A, Enterocin X, Acidocin LF) |
| *L. casei* BL23 | 20,114 | LSEI 2163 (LSEI 2163) |
| | 29,141 | LSEI 2386 (LSEI 2386, Enterocin X, Carnocin CP52) |
| | 20,180 | Enterolysin A (Enterolysin A) |
| *L. casei* W56 | 20,114 | LSEI 2163 (LSEI 2163) |
| | 29,138 | LSEI 2386 (LSEI 2386, Enterocin X, Carnocin CP52) |
| | 20,180 | Enterolysin A (Enterolysin A) |
| *L. helveticus* H9 | 20,285 | Enterolysin A (Enterolysin A) |
| | 20,936 | Helveticin J (Helveticin J) |
| *L. helveticus* R0052 | 20,762 | Helveticin J (Helveticin J) |
| | 20,483 | Enterolysin A (Enterolysin A) |
| | 20,696 | Helveticin (Helveticin J) |
| | 20,882 | Helveticin J (Helveticin J) |
| | 20,000 | LAPs |
| | 20,000 | LAPs |
| *L. helveticus* KLDS1.8701 | 20,516 | Helveticin J (Helveticin J) |
| | 20,429 | Enterolysin A (Enterolysin A) |
| | 20,876 | Helveticin J (Helveticin J) |
| *L. helveticus* MB2-1 | 20,966 | Helveticin J (Helveticin J) |
| | 20,285 | Enterolysin A (Enterolysin A) |
| *L. helveticus* CAUH18 | 20,798 | Helveticin (Helveticin J) |
| | 20,147 | Enterolysin A (Enterolysin A) |
| | 20,936 | Helveticin J (Helveticin J) |
| *L. helveticus* D76 | 20,762 | Helveticin J (Helveticin J) |
| | 20,468 | Enterolysin A (Enterolysin A) |
| | 20,954 | Helveticin J (Helveticin J) |
| *L. helveticus* D75 | 20,762 | Helveticin J (Helveticin J) |
| | 20,468 | Enterolysin A (Enterolysin A) |
| | 20,954 | Helveticin J (Helveticin J) |
| *L. plantarum* WCFS1 | 29,495 | Plantaricin J (Plantaricin JK, N, A, EF) |
| *L. plantarum* ST-III | 29,495 | Plantaricin J (Plantaricin JK, N, A, EF) |
| *L. plantarum* ZJ316 | 24,131 | Plantaricin J (Plantaricin JK, Plantaricin NC8-alpha, beta) |

**Table 3** (continued)

| Strain | Size (bp) | Class (bacteriocin gene name) |
|---|---|---|
| *L. plantarum* 16 | 20,342 | Plantaricin F (Plantaricin F, Plantaricin E) |
| *L. plantarum* ATCC 14917 | 29,498 | Plantaricin J (Plantaricin JK, N, A, EF) |
| *L. plantarum* CAUH2 | 29,495 | Plantaricin J (Plantaricin JK, N, A, EF) |
| *L. plantarum* LZ206 | 24,130 | Plantaricin J (Plantaricin JK, Plantaricin NC8-alpha, beta) |
| *L. plantarum* KLDS1.0391 | 20,342 | Plantaricin F (Plantaricin F) |
| *L. plantarum* 10CH | 29,495 | Plantaricin J (Plantaricin JK, N, A, EF) |
| *L. plantarum* EM | 29,495 | Plantaricin J (Plantaricin JK, N, A, EF) |
| | 20,105 | Bovicin 255 (Bovicin 255) |
| *L. reuteri* I5007 | 20,474 | Enterolysin A (Enterolysin A) |
| *L. reuteri* ZLR003 | 20,459 | Enterolysin A (Enterolysin A) |
| *L. reuteri* SKKU-OGDONS-01 | 20,459 | Enterolysin A (Enterolysin A) |

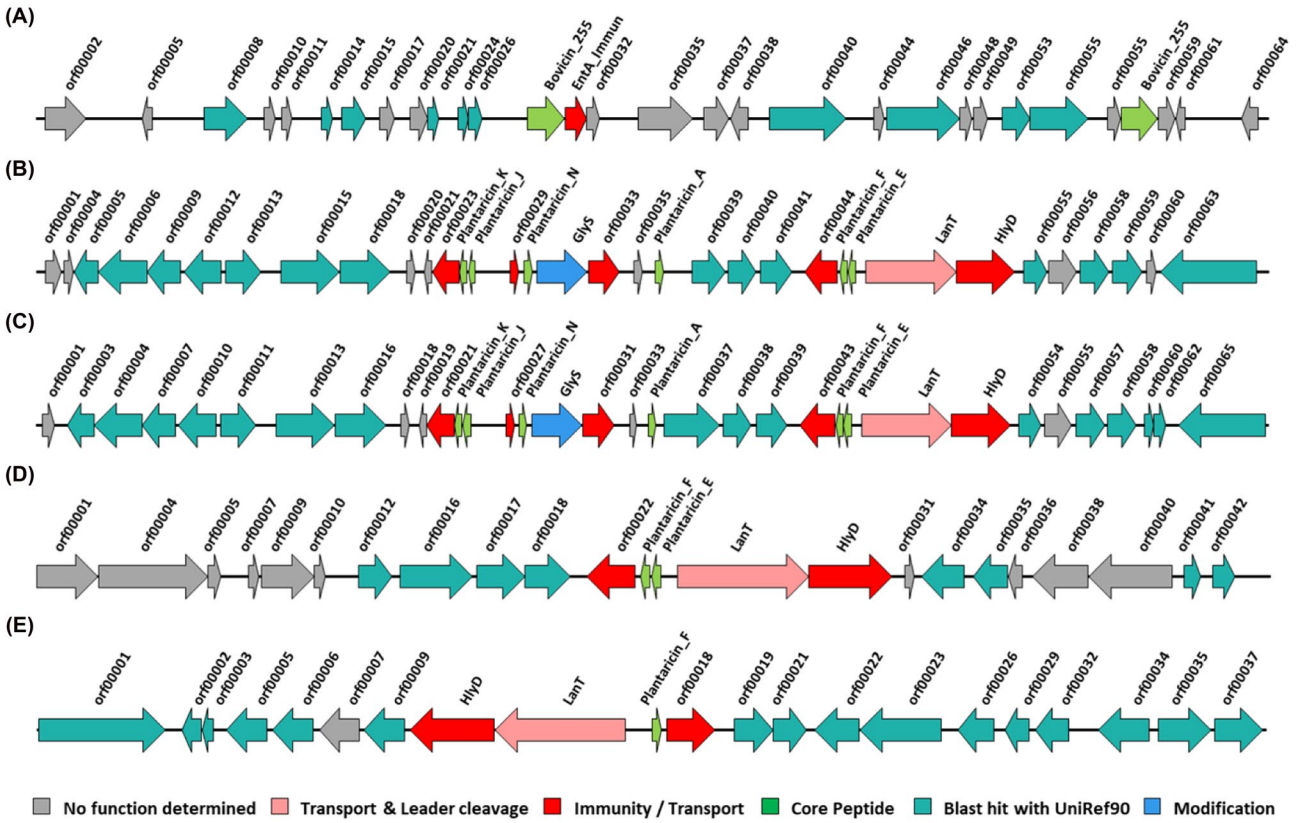


**Fig. 5** Genetic organization of putative bacteriocin synthesis genes: **a** Bovicin gene cluster of *L. plantarum* EM on plasmid, **b** Plantaricin gene cluster of *L. plantarum* EM on chromosome, **c** Plantaricin gene cluster of *L. plantarum* WCFS1, **d** Plantaricin gene cluster of *L. plantarum* 16, **e** Plantaricin gene cluster of *L. plantarum* KLDS1.0391

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

**Author Contributions** EK performed the experiments, analyzed data and wrote the manuscript. HCC reviewed and edited the manuscript. HYK designed this study, supervised all experiments and reviewed the manuscript. All authors read and approved the final manuscript.

**Data Availability** The data used to support the findings of this study are included within the article.

## Compliance with Ethical Standards

**Conflict of interest** The authors declare no conflict of interest.

**Ethical Approval** Not applicable.

## References


1. Siezen RJ, Tzeneva VA, Castioni A, Wels M, Phan HT, Rademaker JL, Starrenburg MJ, Kleerebezem M, Molenaar D, van Hylckama Vlieg JE (2010) Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches. Environ Microbiol 12(3):758–773
2. Martino ME, Bayjanov JR, Caffrey BE, Wels M, Joncour P, Hughes S, Gillet B, Kleerebezem M, van Hijum SA, Leulier F (2016) Nomadic lifestyle of *Lactobacillus plantarum* revealed by comparative genomics of 54 strains isolated from different habitats. Environ Microbiol 18(12):4974–4989
3. Kleerebezem M, Boekhorst J, van Kranenburg R, Molenaar D, Kuipers OP, Leer R, Tarchini R, Peters SA, Sandbrink HM, Fiers MW, Stiekema W, Lankhorst RM, Bron PA, Hoffer SM, Groot MN, Kerkhoven R, de Vries M, Ursing B, de Vos WM, Siezen RJ (2003) Complete genome sequence of *Lactobacillus plantarum* WCFS1. Proc Natl Acad Sci USA 100(4):1990–1995
4. Kang J, Chung WH, Lim TJ, Whon TW, Lim S, Nam YD (2017) Complete genome sequence of *Lactobacillus casei* LC5, a potential probiotics for atopic dermatitis. Front Immunol 8:413
5. Reid G, Jass J, Sebulsky MT, Mccormick JK (2003) Potential uses of probiotics in clinical practice. Clin Microbiol Rev 16:658–672
6. Oh NS, Joung JY, Lee JY, Kim Y (2018) Probiotic and anti-inflammatory potential of *Lactobacillus rhamnosus* 4B15 and *Lactobacillus gasseri* 4M13 isolated from infant feces. PLoS ONE 13:e0192021
7. Bao Y, Zhang Y, Zhang Y, Liu Y, Wang S, Dong X, Wang Y, Zhang H (2010) Screening of potential probiotic properties of *Lactobacillus fermentum* isolated from traditional dairy products. Food Control 21:695–701
8. Kumar M, Nagpal R, Kumar R, Hemalatha R, Verma V, Kumar A, Chakraborty C, Singh B, Marotta F, Jain S, Yadav H (2012) Cholesterol-lowering probiotics as potential biotherapeutics for metabolic diseases. Exp Diabetes Res 2012:1–14
9. Patel AK, Singhania RR, Pandey A, Chincholkar SB (2010) Probiotic bile salt hydrolase: current developments and perspectives. Appl Biochem Biotechnol 162(1):166–180
10. Leroy F, Vuyst LD (2004) Lactic acid bacteria as a functional starter culture for the food fermentation industry. Trends Food Sci Technol 15:67–78
11. Li P, Gu Q, Zhou Q (2016) Complete genome sequence of *Lactobacillus plantarum* LZ206, a potential probiotic strain with antimicrobial activity against food-borne pathogenic microorganisms. J Biotechnol 238:52–55
12. Hill C, Guarner F, Reid G, Gibson GR, Merenstein DJ, Pot B, Morelli L, Canani RB, Flint HJ, Salminen S, Calder PC, Sanders ME (2014) The international scientific association for probiotics and prebiotics consensus statement on the scope and appropriate use of the term probiotic. Nat Rev Gastroenterol Hepatol 11:506–514
13. Abriouel H, Pérez Montoro B, Casimiro-Soriguer CS, Pérez Pulido AJ, Knapp CW, Caballero Gómez N, Castillo-Gutiérrez S, Estudillo-Martínez MD, Gálvez A, Benomar N (2017) Insight into potential probiotic markers predicted in *Lactobacillus pentosus* MP-10 genome sequence. Front Microbiol 8:891
14. Maiden MC, Jansen van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND (2013) MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol 11:728–736
15. Choi EA, Chang HC (2015) Cholesterol-lowing effects of a putative probiotic strain *Lactobacillus plantarum* EM isolated from kimchi. LWT - Food Sci Technol 62:210–217
16. Kim E, Cho Y, Lee Y, Han SK, Kim CG, Choo DW, Kim YR, Kim HY (2017) A proteomic approach for rapid identification of *Weissella* species isolated from Korean fermented foods on MALDI-TOF MS supplemented with an in-house database. Int J Food Microbiol 243:9–15
17. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R (2014) The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res 42:D206–D214
18. Grant JR, Stothard P (2008) The CGView server: a comparative genomics tool for circular genomes. Nucleic Acids Res 36:W181–W184
19. Kleinheinz KA, Joensen KG, Larsen MV (2014) Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage 4:e27943
20. Patel AK, Singhania RR, Pandey A, Chincholkar SB (2018) BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res 46:W278–W281
21. Huerta-Cepas J, Serra F, Bork P (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol 33(6):1635–1638
22. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319
23. Delmont TO, Eren AM (2018) Linking pangenomes and metagenomes: the *Prochlorococcus* metapangenome. PeerJ 6:e4320
24. Chaudhari NM, Gupta VK, Dutta C (2016) BPGA: an ultra-fast pan-genome analysis pipeline. Sci Rep 6:24373
25. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26(19):2460–2461
26. Molenaar D, Bringel F, Schuren FH, de Vos WM, Siezen RJ, Kleerebezem M (2005) Exploring *Lactobacillus plantarum* genome diversity by using microarrays. J Bacteriol 187(17):6119–6127
27. Evanovich E, de Souza Mendonça Mattos PJ, Guerreiro JF (2019) Comparative Genomic Analysis of *Lactobacillus plantarum*: an overview. Int J Genomics 2019:4973214
28. Haghi M, Koseoglu AE, Karaboz I, Un C (2016) Detection of heat shock protein (DnaK, DnaJ and GrpE) horizontal gene transfers among Acanthamoeba polyphaga, Acanthamoeba polyphaga mimivirus (APMV), amoeba-infecting bacteria and sputnik virophage. Int J Adv Biotechnol Res 7:1618–1622
29. Kim WS, Khunajakr N, Ren J, Dunn NW (1998) Conservation of the major cold shock protein in lactic acid bacteria. Curr Microbiol 37:333–336
30. Ren J, Sun K, Wu Z, Yao J, Guo B (2011) All 4 bile salt hydrolase proteins are responsible for the hydrolysis activity in *Lactobacillus plantarum* ST-III. J Food Sci 76:M622–M628
31. Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, Pavlov A, Pavlova N, Karamychev V, Polouchine N, Shakhova V, Grigoriev I, Lou Y, Rohksar D, Lucas S, Huang K, Goodstein DM, Hawkins T, Plengvidhya V, Welker D, Hughes J, Goh Y, Benson A, Baldwin K, Lee JH, Díaz-Muñiz I, Dosti B, Smeianov V, Wechter W, Barabote R, Lorca G, Altermann E, Barrangou R, Ganesan B, Xie Y, Rawsthorne H, Tamir D, Parker C, Breidt F, Broadbent J, Hutkins R, O'Sullivan D, Steele J, Unlu G, Saier M, Klaenhammer T, Richardson P, Kozyavkin S, Weimer B, Mills D (2006) Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A 103(42):15611–15616
32. da Silva SS, Vitolo M, González JMD, Oliveira RPS (2014) Overview of *Lactobacillus plantarum* as a promising bacteriocin producer among lactic acid bacteria. Food Res Int 64:527–536