Comparative genome analysis of four *Leuconostoc* strains with a focus on carbohydrate-active enzymes and oligosaccharide utilization pathways

Anshul Sharma [a,1], Neha Sharma [b,1], Deepshikha Gupta [c,d], Hae-Jeung Lee [a,e], Young-Seo Park [b,*]

[a] *Department of Food and Nutrition, Gachon University, Seongnam-si 13120, Gyeonggi-do, South Korea*
[b] *Department of Food Science and Biotechnology, Gachon University, Seongnam-si 13120, Gyeonggi-do, South Korea*
[c] *Department of Plant Sciences, University of Hyderabad, P.O. Central University, Hyderabad 500046, India*
[d] *DBT-National Institute of Animal Biotechnology (NIAB), Hyderabad 500032, Telangana, India*
[e] *Institute for Aging and Clinical Nutrition Research, Gachon University, Seongnam-si 13120, Gyeonggi-do, South Korea*

A R T I C L E   I N F O

A B S T R A C T

Leuconostoc is mostly found in food, plants, and dairy products. Due to their innate genomic features, such as the presence of carbohydrate-active enzymes, bacteriocins, and plasmids, *Leuconostoc* spp. have great biotechnological potential. In this study, four strains were isolated and identified as *Leuconostoc mesenteroides* SG315 (LA), *L. citreum* SG255 (LB), *L. lactis* CCK940 (LC), and *L. lactis* SBC001 (LD). Comparative analysis was performed using their draft genome sequences. Differences among the four strains were analyzed using the average nucleotide identity, dot plot, and multiple alignments of conserved genomic sequences. Functional profiling revealed 2134, 1917, 1751, and 1816 open reading frames; 2023, 1823, 1655, and 1699 protein-coding genes; 60, 57, 83, and 82 RNA-coding genes; and GC content of 37.5 %, 38.8 %, 43.3 %, and 43.2 %, in LA, LB, LC, and LD, respectively. The total number of genes encoding carbohydrate-active enzymes was 76 (LA), 73 (LB), 57 (LC), and 67 (LD). These results indicate that the four strains shared a large number of genes, but their gene content is different. Furthermore, most genes with unknown functions were observed in the prophage regions of the genome. This study also elucidated the oligosaccharide utilization and folate biosynthesis pathways in *Leuconostoc* spp. Taken together, our findings provide useful information on the genomic diversity of CAZymes in the four *Leuconostoc* strains and suggest that these species could be used for potent exploitation.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creative-commons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Bacteria and their genomes have evolved and adapted over time due to mutations and gene rearrangements. This evolutionary pattern can be observed as nucleotide insertions or deletions in specific parts of the genomes. In addition to genes encoding essential functions, many organisms have accessory genes that allow bacteria to adapt to their surroundings [1].

Several *Leuconostoc* species, which are lactic acid bacteria (LAB), are commonly found in plants, silage, fermented foods (e.g., sour cream, quark, kefir, kimchi, and sauerkraut), and dairy products (e.g., fresh and raw milk, cream, butter, and cheese) [2–4]. *Leuconostoc* is a member of the *Leuconostocaceae* family, one of the most commonly used LAB groups [3] and is closely related to Lactobacillaceae [5]. Cienkowski isolated *Leuconostoc* from a slime 46

epidemic in a sugar manufacturing unit in 1878 [6]. *Leuconostoc* species are Gram-positive, facultative anaerobes, non-motile, catalase-negative, asporogenous, psychrotrophic, or psychrotolerant catenation-shaped bacteria with an average GC content of 37.5 % and an optimum growth at 25–30 °C [7]. According to physiological studies, *Leuconostoc* species are closely related to heterofermentative lactobacilli [3]. The genus has been designated as "generally recognized as safe" (GRAS), indicating that it is reasonably safe for use in culinary products [3]. Recently, *Leuconostoc* spp. has recently been identified as one of the most promising probiotic candidates due to its ability to produce bioactive antimicrobial peptides and vitamins, as well as improve human health by modulating immune responses [8–10]. A texturing *Leuconostoc* strain was found to produce hetero-polysaccharides rather than homopolysaccharides in milk via a Wzy-dependent pathway for the first time [11].

Various *Leuconostoc* spp. have produced polysaccharides (fructans and glucans) and oligosaccharides [12–15]. We previously reported oligosaccharide production and immunostimulatory

---

activity in RAW264.7 of *L. lactis* CCK940 [16], the draft genome sequence of *L. citreum* SG255, an oligosaccharide-synthesizing strain [17], and the oligosaccharide production using glucansucrase activity of *L. lactis* CCK940 [18]. Our recent study elucidated the production, optimization, and structural characterization of *L. lactis* SBC001 *gluco*-oligosaccharides, as well as their prebiotic and anti-inflammatory effects [15]. Different *Leuconostoc* species or strains have a wide range of capabilities, implying that their genomes are highly diverse in content and type.

Although numerous studies have been conducted to investigate the diversity and evolution of *Leuconostoc* [19–21], many fundamental studies on genomic diversity, species dynamics, and functional properties are still required to fully understand this genus' potential. Comparative genome analysis of *Leuconostoc* species allows for comparisons and presentation of data at the species or strain level, as well as functional genetic potential at the molecular level [7]. Recently, a comparative genomic analysis of *L. mesenteroides* MTCC 10,508 and *L. carnosum* has been reported [19,20].

Carbohydrates are the primary energy sources for LAB fermentation and acidification. Before being transported inside the bacterial cell by specific transport systems, all sugars must be converted into monomeric forms by specific enzymes [22,23]. Oligosaccharides are the main carbohydrates found in fruits, cereals, milk, and the upper intestines of animals. Although the pathways for mono-and disaccharide metabolism are well understood, there is little information on the metabolism of higher oligosaccharides, which are found in many ecosystems. The bioinformatic analysis of *Leuconostoc* genes involved in carbohydrate metabolism has begun. However, there is still a lack of knowledge about this topic.

Draft genome sequencing and characterization of *L. mesenteroides* SG315 (LA), *L. citreum* SG255 (LB), and *L. lactis* CCK940 (LC), and *L. lactis* SBC001(LD) were performed to advance our understanding of the genome diversity. A comparative genomic analysis was performed to reveal the evolutionary relationships among these species. The number and diversity of carbohydrate-active enzymes (CAZymes) in *Leuconostoc* spp., as well as their association with oligosaccharide utilization pathways, were investigated. This is the first time genomes of four different strains of *Leuconostoc* spp. have been compared.

2. Materials and methods

2.1. Bacterial species and culture conditions

In this study, four *Leuconostoc* strains were used: LA, LB, LC, and LD. In our laboratory, we isolated LA, LB, LC, and LD from *baechu*-kimchi, young radish kimchi, Chinese cabbage kimchi, and chive, respectively. The strains were identified using standard morphological, biochemical, and physiological tests, as well as 16 s rRNA gene sequencing with the National Center for Biotechnology Information's nucleotide BLAST tool. Under microaerophilic conditions, a single colony of each strain was inoculated in modified De Man, Rogosa, and Sharpe (MRS) broth at 37 °C for 24 h. Stock cultures were stored at − 80 °C in modified MRS broth mixed with 20 % glycerol.

2.2. Genome assembly and annotation

For genome sequencing of the selected strains, library construction was performed using the Illumina TruSeq DNA library prep kit and genome sequencing was performed by Illumina Miseq 300 bp paired-end. For quality control, overrepresented sequences were carefully checked, and low quality reads and adaptors were removed. For the sequencing analysis, a de novo assembly of high quality reads was selected for the analysis. Overall, the whole genome sequencing of the four strains was performed by Sanigen Co.,

ltd. company in the Republic of Korea. The following accession numbers have already been given to the genome sequences of all four strains in NCBI GenBank: SG315 NZ_JACGML000000000 (LA), SG255 NZ_JACGMK000000000 (LB), CCK940 NZ_NQLF 00,000,000 (LC), and SBC001 NZ_JACGAK000000000 (LD). The ANI Calculator was used to perform average nucleotide identity (ANI) analysis, as described in a previous study [24]. For pathway analysis, the complete genomes were mapped using the Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.kegg.jp/) and Clusters of Orthologous Groups (COG, https://www.ncbi.nlm.nih.gov/COG) databases with default parameters. PHASTER (http://phast.wishartlab.com/index.html) was used to detect putative prophage insert regions, and MinCED 3 (https://sourceforge.net/projects/minced/) was used to identify clustered regularly interspaced short palindromic repeats (CRISPR).

The genomes were compared using Kbase to generate dot plot matrices [25]. The dot plots are visual representations of conserved regions of genomes discovered through pairwise alignment. MAUVE [26] was used to align the genomes with 940 LC serving as the reference. The dot plot matrices and MAUVE alignment predict the homologous region rearrangements such as inversion or translocation events, as well as insertions or deletions in the genome.

2.3. Phylogenetic analysis

A phylogenetic tree was constructed based on whole genome data from LA, LB, LC, and LD strains. A whole-genome level comparison was performed amongst 119 Leuconostoc strains, covering 65 L. mesenteroides genomes, 28 L. lactis genomes, and 26 L. citreum genomes. The phylogenetic relationship was obtained using PATRIC with the mafft alignment program [27] (https://mafft.cbrc.jp/alignment/software/) and the following parameters: single copy gene requested, 100; max deletions allowed, 10; max duplications allowed, 10. This resulted in the alignment of 97,704 nucleotides. The tree was constructed and viewed graphically using FigTree (https://tree.bio.ed.ac.uk/software/figtree/).

2.4. Genome Island prediction using the IslandViewer web tool

The presence of genomic islands (GIs) was predicted using Island Viewer 4 [28], which is an open-access web tool that uses three prediction algorithms: SIGI-HMM, Island Path-DIMOB, and IslandPick, to calculate the codon usage and dinucleotide bias within a genome and generate a dataset of GIs and non-GIs from phylogenetically related organisms.

2.5. Identification of CAZymes

CAZymes (https://www.cazy.org/) in the *Leuconostoc* species obtained using genome similarity analysis were identified and annotated using dbCAN2 [29]. The dbCAN2 scans the genome with several algorithms, including the hidden Markov model (HMM) profile, which uses HMMdb v7 [30], DIAMOND [31], and Hotpep [32] to improve prediction. Genes annotated by at least two methods were chosen for further analysis.

CAZymes are essential in sugar metabolism, particularly in the biosynthesis, binding, and catabolism of carbohydrates. Glycosyltransferase (GT), glycoside hydrolase (GH), polysaccharide lyase (PL), carbohydrate esterase (CE), and auxiliary activity (AA) are the different classes of CAZymes [33].

3. Results and discussion

3.1. General genomic features

The LA genome is made up of a single circular chromosome of 2.1 Mb in length and has a GC content of 37.5 % (Fig. 1A). Meanwhile,
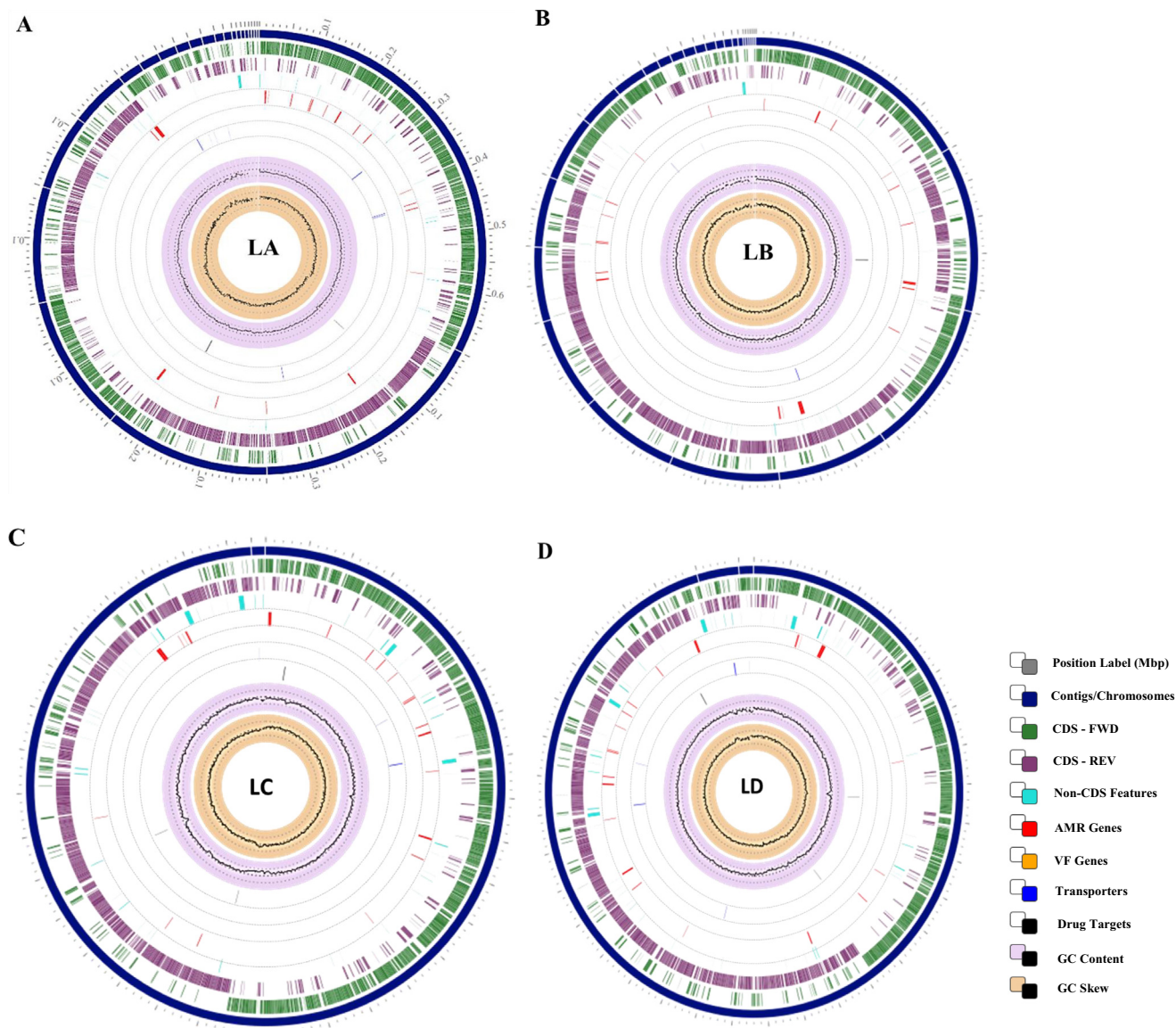
Fig. 1. Circular genomes of (A) LA, (B) LB, (C) LC, and (D) LD. Circles illustrate the following characteristics from the outside to the center: contigs/chromosome (blue), coding sequences on forward and reverse strands." Circles 4–8 represent non-CDS, antimicrobial resistance genes (AMR), virulence factors (VF), transporters, and drug targets, respectively. The next ring represented in pink/black denotes the GC content. The innermost circle shows GC-skew. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the genomes of LB (Fig. 1B), LC (Fig. 1C), and LD (Fig. 1D) each have a single circular chromosome of 1.9 Mb, 1.7 Mb, and 1.8 Mb in length, with a GC content of 38.8 %, 43.3 %, and 43.2 %, respectively. The number of open reading frames (ORFs), protein-coding genes, RNA genes,

and pseudogenes annotated in the four *Leuconostoc* strains is shown in Table 1. Furthermore, in LC and LD, a single CRISPR array was discovered (Table S1). According to Kim et al. [34], the *L. citreum* EFEL2700 isolate from kimchi has a 1.9 Mb long genome sequence,

Table 1
Genome information of the four *Leuconostoc* species.

Attribute	*Leuconostoc mesenteroides* SG315	*Leuconostoc citreum* SG255	*Leuconostoc lactis* CCK940	*Leuconostoc lactis* SBC001
Genome size (bp)	2,099,741	1,869,057	1,741,511	1,835,155
DNA G + C (bp)	786,451	724,875	754,610	791,832
G + C content (%)	37.46	38.78	43.33	43.15
DNA scaffolds	23	28	2	3
N50 (bp)	250,908	152,714	1,726,690	1,758,110
Total genes	2134	1917	1751	1816
Protein coding genes	2023	1823	1655	1699
RNA genes	60	57	83	82
Pseudo genes	51	37	13	35
CRISPR arrays	0	0	1	1

LA: *Leuconostoc mesenteroides* SG315; LB: *L. citreum* SG255; LC: *L. lactis* CCK940; and LD: *L. lactis* SBC001; X= *L. citreum* EFEL2700; Y= *L. lactis* 1.2.28; W: *L. citreum* DmW_111; Z= *L. mesenteroides* SRCM103356; S= *L. mesenteroides* 406; UP: unpublished.

1,853 total genes, and 39.1 % GC content (Table 1). *L. citreum* DmW_111, isolated from wild *Drosophila melanogaster*, had a 1.8 Mb genome length, with a GC content of 38.9 % and a total gene count of 1853 [35]. According to unpublished data, the genomes of *L. lactis* 1.2.28 (GCA_018993775.1), *L. mesenteroides* SRCM 103,356 (GCA_004102585.1) had 1.7 Mb and 2.0 Mb genomes, respectively,

with 43.4 % and 37.7 % GC content and 1724 and 1991 total genes (Table 1). Additionally, *L. mesenteroides* 406 contains a 2.0 Mb chromosome with a GC content of 37.7 % and a total number of genes of 2062 [36], which corroborate our findings.

According to the RAST annotation, LA, LB, LC, and LD harbored 208, 203, 209, and 206 subsystems, respectively, which were
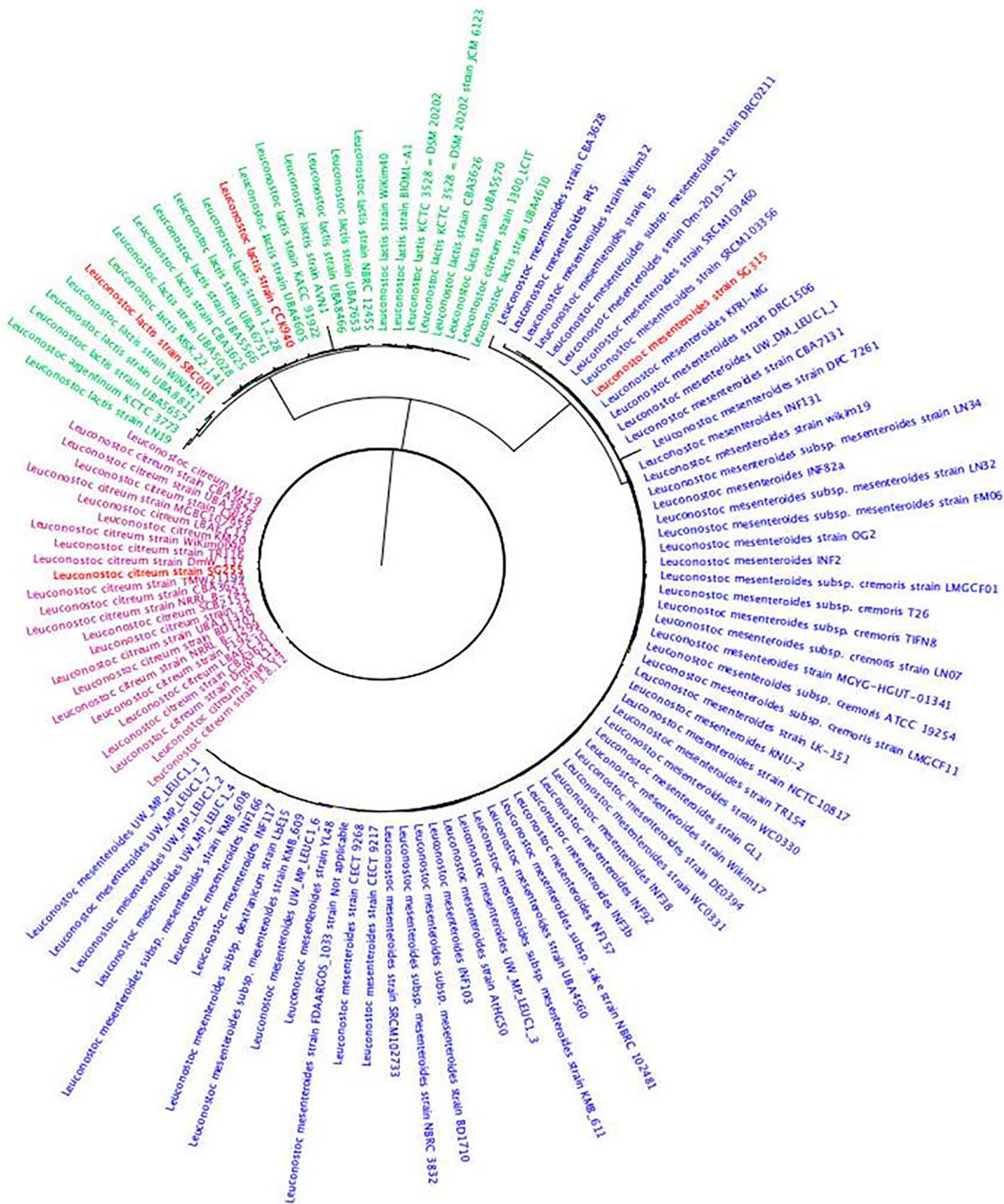

Fig. 2. Phylogenetic comparison of four *Leuconostoc* strains with genomes available in the public domain. Strains used in this study are highlighted.

categorized further into 27 subsystem features. The relevant circular views for LA, LB, LC, and LD are shown in Fig. S1A, B, C, and D, respectively. Orthologous genes are groups of homologous genes that descend from a common ancestor. These genes frequently retain their roles as those of their ancestral genes even after evolution [37]. Approximately 20 COG functions were identified in the three *Leuconostoc* spp., in varying percentages of genes associated with each strain (Table S2). Translation, ribosomal structure, and biogenesis (COG category J) were the most abundant COG function genes (LA, 8.82 %; LB, 9.56 %; LC, 10.51 %; LD, 10.93 %). The genes involved in amino acid transport and metabolism (COG category E) for LA, LB, LC, and LD were found to be 8.63 %, 6.50 %, 7.93 %, and 5.87 %, respectively. Genes involved in carbohydrate transport and metabolism (COG category G) (6.85 %, 6.55 %, 6.34 %, and 7.25 %) and transcription (COG category K) (5.58 %, 5.81 %, 5.34 %, and 7.02 %) were also annotated in LA, LB, LC, and LD, respectively. The functional enrichment of these COGs suggests that they are important for survival and adaptation to harsh environmental conditions, as well as in the utilization of complex carbohydrates for growth. Fig. S2 shows the comparative proteome analysis of the four *Leuconostoc* strains and illustrates that the majority of proteins share a sequence similarity score of less than 80 %, while a few proteins are 95–98 % identical. A phylogenetic tree was constructed with different genome sequences available in the public domain (Fig. 2). The four strains described in this paper are highlighted in red. The findings demonstrated the categorization of strains into species and revealed differences in strain genetic similarity within each species.

3.2. ANI analysis

The ANI value describes the sequence similarity between the conserved regions of two genomes and measures their genetic relatedness [38]. ANI measurements are considered more informative than 16S rRNA gene identification because the former is based on a larger number of genes [39]. Organisms with ANI values of $\geq$ 95 % are all members of the same species [40]. In this study, the ANI values of he four strains were estimated to investigate their interspecies genetic relatedness. The ANI values ranged from 72.86 % to 98.14 % (Fig. 3). The highest ANI value (98.14 %) was observed between LC and LD which is expected because both strains belong to the same species (Fig. 3). These findings are consistent with previous research, which found that the ANI value between the target *L. lactis* strain EFEL005 and the type strain *L. lactis* KCTC 3528 T was greater than 96 % [36]. The species differences could be attributed to host-specific strain diversity and niche adaption.

3.3. GIs

GIs are genomic regions that reveal evidence of horizontal DNA transfer, particularly in bacteria, and appear to benefit bacterial cells [41,42]. GIs encode various symbiotic or pathogenic functions

that can help an organism adapt to its host or environment [43]. Two steps are involved in GI integration and formation [44]. GIs were predicted in all four *Leuconostoc* strains using the closest complete (or closed) sequence as a reference (*L. mesenteroides* J18 for LA, *L. citreum* KM20 for LB, *L. lactis* WiKim40 for LC, and LD) (Fig. 4A–D). A total of 5 GIs (LA), 8 GIs (LB), 10 (LC), and 10 (LD) were reported during the genome analysis (Table S3). In LA, GIs ranging from 9.13 kb with nine coding sequences (CDS) (GI-2) to 53.95 kb with 65 CDS (GI-4) were predicted (Table S3). In LB, the GI-2, with a length of 50.40 kb and 72 CDS, was the longest, while the GI-1, with a length of 4.89 kb and only 6 CDS, was discovered to be the shortest (Table S3). LC showed GIs that vary in size, with the shortest GI-3 of 5.41 kb with 5 CDS and the largest GI-6 of 79.64 kb with 96 CDS. For LD, the largest GI was 133.48 kb (GI-3) with 131 CDS, and the shortest GI was 6.95 kb (GI-2) with only six CDS (Table S3).

3.4. Prophage regions in the genomes

In this study, we aimed to explore prophages in the genomes of four strains. Prophages integrated into bacterial genomes confer resistance to superinfection [45]. The PHASTER software was used to identify prophage elements in the genomes of LA, LB, LC, and LD. As shown in Fig. 5 (A–D, upper panel), the number of prophage regions were found to be 3, 2, 4, and 2 for LA, LB, LC, and LD, respectively. The presence of prophage in the genomes suggests genomic instability, which is a reservoir for adaptation. As shown in the lower panel of Fig. 5A, in the three-prophage sections of LA, a total of 70 protein-coding genes were identified, 30 of which were phage proteins, and the remaining 40 were hypothetical in nature.

Similarly, the lower panel of Fig. 5B represents 73 proteins found in two prophage regions identified in LB. Of these, 24 were identified as phage proteins, while the remaining 49 were classified as hypothetical proteins. In Fig. 5C, the lower panel shows 17 proteins out of 47 in LC constituted the prophage area, with the remaining 30 classified as hypothetical proteins. Nine phage proteins were identified among the 27 protein-coding genes in LD, with the remaining proteins were hypothetical proteins. The presence of a large number of hypothetical genes in *Leuconostoc* spp. indicates the presence of unrecognized phages [45].

3.5. Dot-plot and MAUVE analysis

Genomic reshuffling has a substantial impact on bacterial evolution [46]. The pairwise alignment dot-plot pattern showed a higher low-level similarity among *Leuconostoc* spp. (Fig. 6). The x- and y-axes show the sequence size scales. There were no continuous lines among the selected species.

The genome sequences of the *Leuconostoc* spp. were aligned using MAUVE [26] to identify the multiple maximal matches and local collinear blocks (LCB). The genome sequence of LC was used as the reference strain. The majority of LCBs were shared by all strains, but the alignments differed, indicating a significant amount of the genetic information in these strains was conserved (Fig. 7). The diagram depicts homologous regions that have been scrambled or inverted as a result of DNA translocation, rearrangement, or recombination. The gaps in the alignment represent missing genome regions Furthermore, chromosomal rearrangements appeared to be common, as the frequency of LCBs with alterations in their relative genomic position was significant, despite their short lengths. A collinear set of homologous (matching) regions between the four *Leuconostoc* genomes is represented by boxes of the same color. After careful observation, the synchronous rearrangement was observed in LC an LD. This result could be due to their close

ANI	LA	LB	LC	LD
LA	X	73.69	72.86	73.25
LB	73.69	X	74.43	74.67
LC	72.86	74.43	X	98.14
LD	73.25	74.67	98.14	X

Fig. 3. Heatmap showing the average nucleotide identity (ANI) among four species of *Leuconostoc*. Green color represents similarity above 97%. As the similarity % decreases, the colour intensity changes from yellow to red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
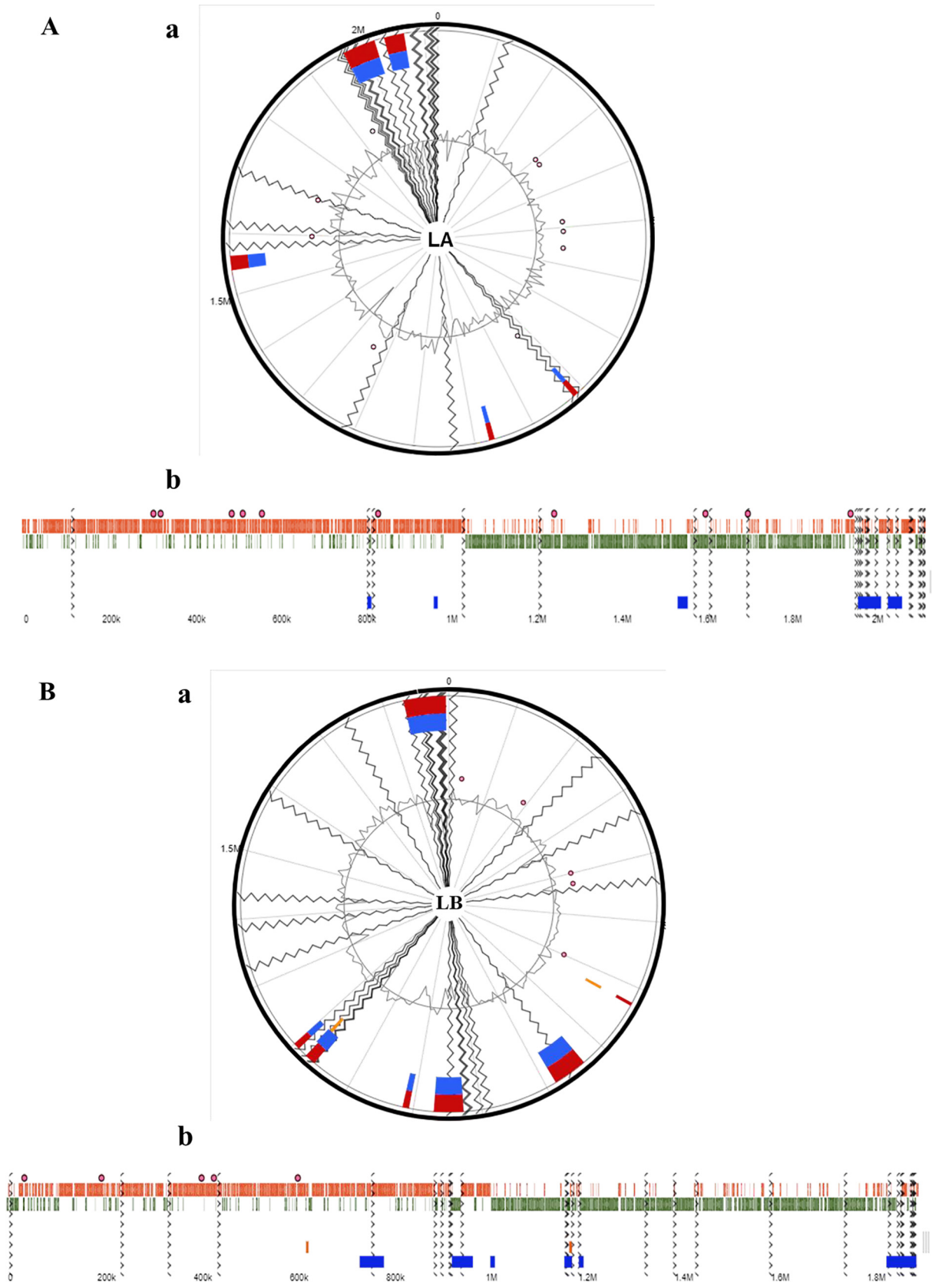
Fig. 4. Genomic islands (GIs) of (A) LA, (B) LB, (C) LC, and (D) LD predicted using SIGI-HMM (orange) and IslandPath-DIMOB (blue). Red shows the integrated genomic island search results. a) Circular and b) horizontal views of the GIs of the four *Leuconostoc* species. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
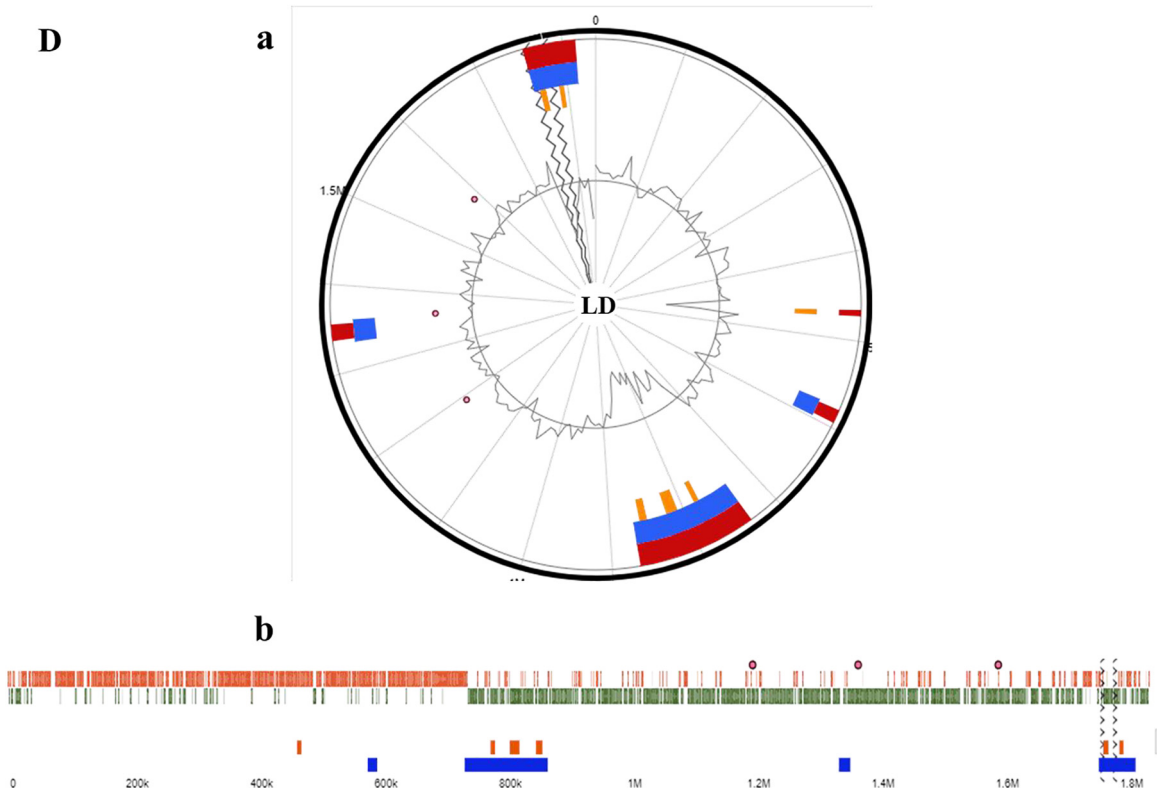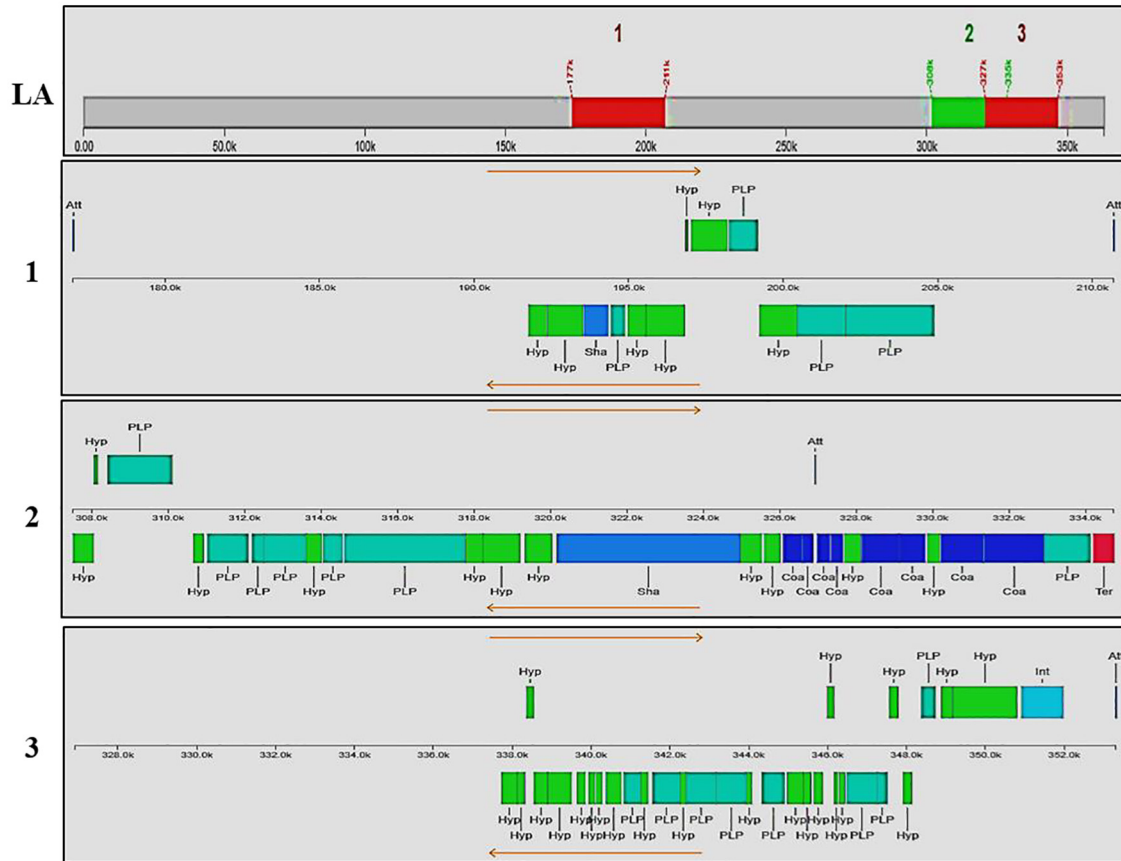
C

a

b
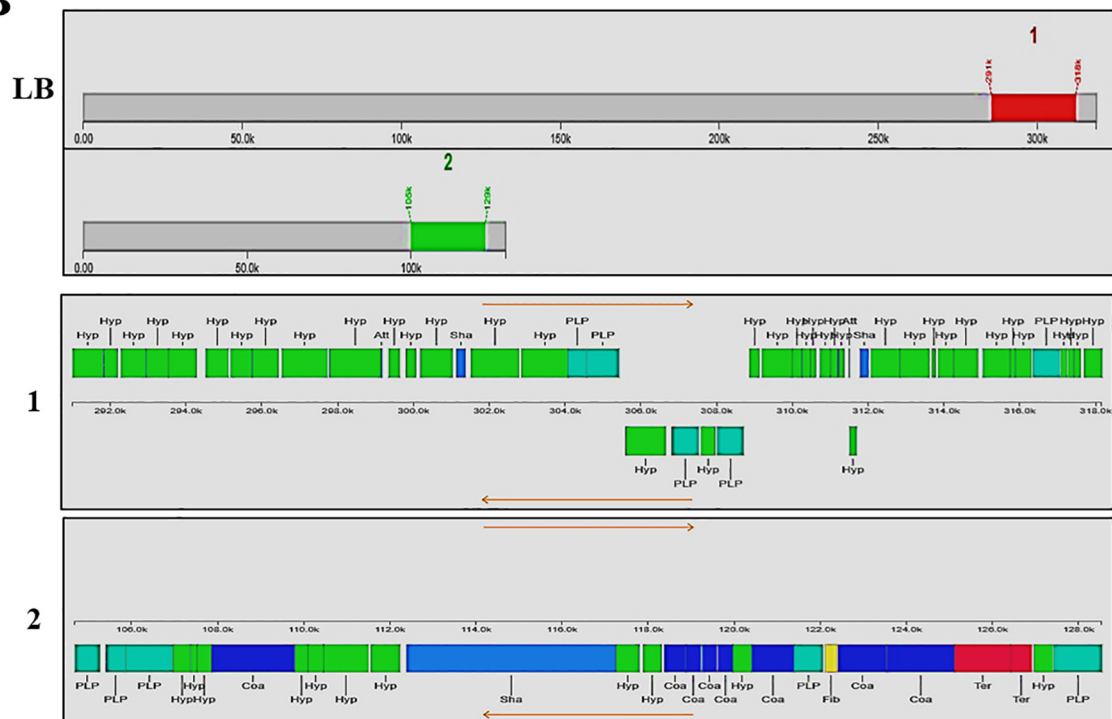
D

a

b
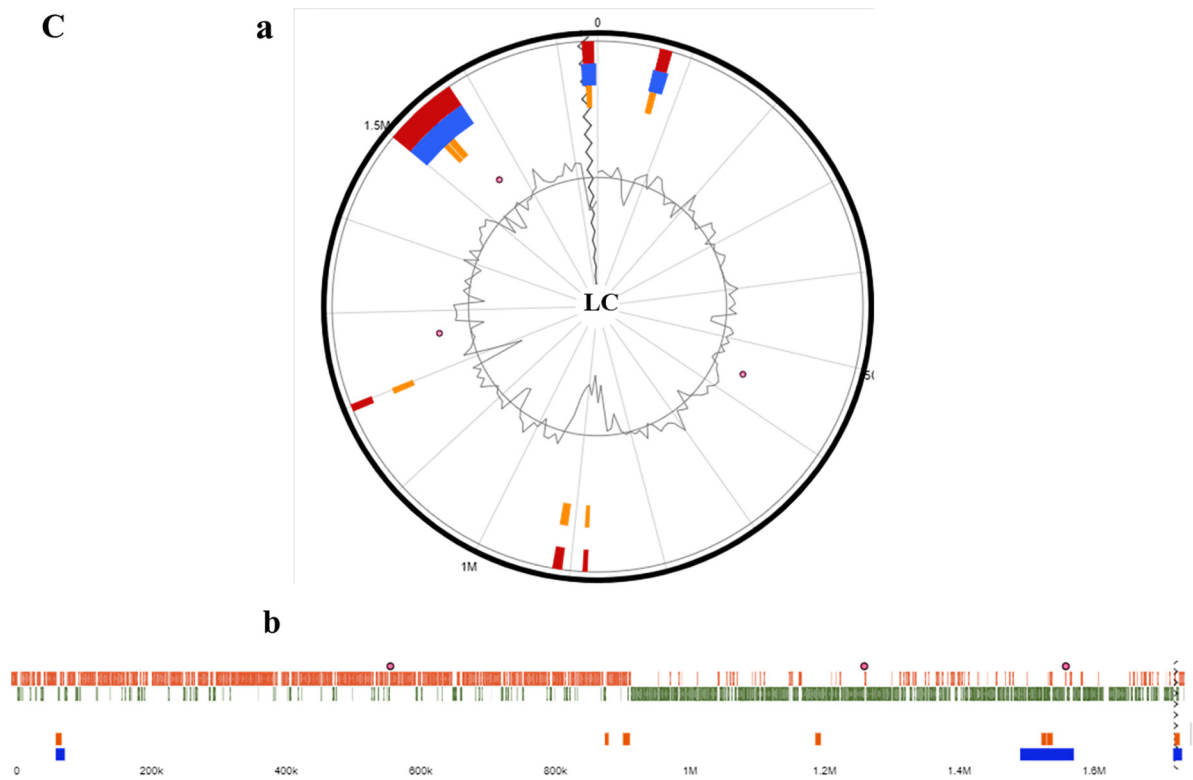
Fig. 4 (*continued*)

Fig. 5. Prophage regions detected in (A) LA, (B) LB (C) LC and (D) LD using the PHASTER server. The total number of prophage regions discovered in each strain is displayed in the upper panel, and the number of genes discovered in each prophage region is shown in the lower panel.
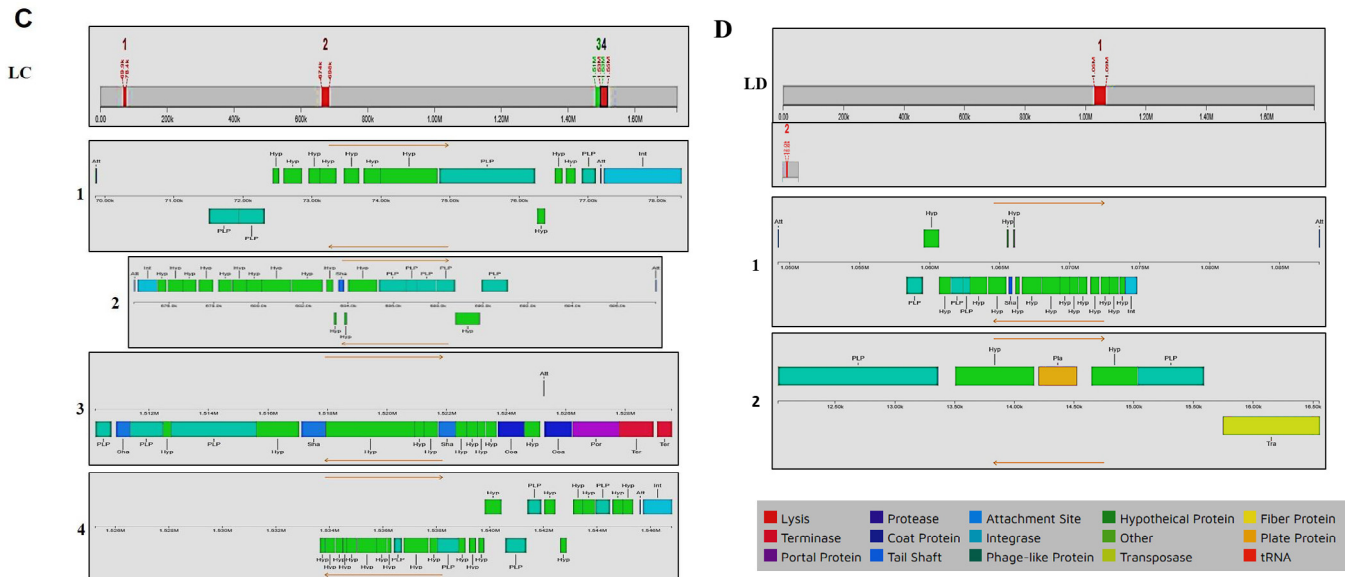
Fig. 5 (*continued*)

taxonomic relationship. Notably, the degree o f homology varied between blocks.

4. CAZymes identified in the four *Leuconostoc* spp.

The total number of genes encoding CAZymes in LA, LB, LC, and LD was 76, 73, 57, and 67, respectively (Table S4). Only those genes that were predicted by at least two of the three algorithms were considered for the downstream analysis. A closer look at the CAZyme families revealed clear differences in the number and distribution of these genes among these four *Leuconostoc* strains. CAZymes from LA, LB, LC, and LD contained 32, 25, 18, and 22 genes encoding glycoside hydrolases (GHs), and 19, 20, 15, and 17 genes encoding glycosyltransferases (GTs), respectively (Fig. 8). Except for LD, which had only one gene for carbohydrate esterases (CEs), all strains had two CE genes With the exception of LA, each strain possessed a single gene encoding polysaccharide lyases (PL) (Table 2) (Fig. 8). These findings are consistent with those of a recent study on 182 *Leuconostoc* spp. genomes, which found that the most prevalent identified gene clusters were the GH > GT > CE > CMB > and PL families [47]. Notably, all selected strains encoded more genes for GHs and GTs, which can participate in the metabolism and transport of functional and active substances. The genes were mainly distributed among the 16 GH
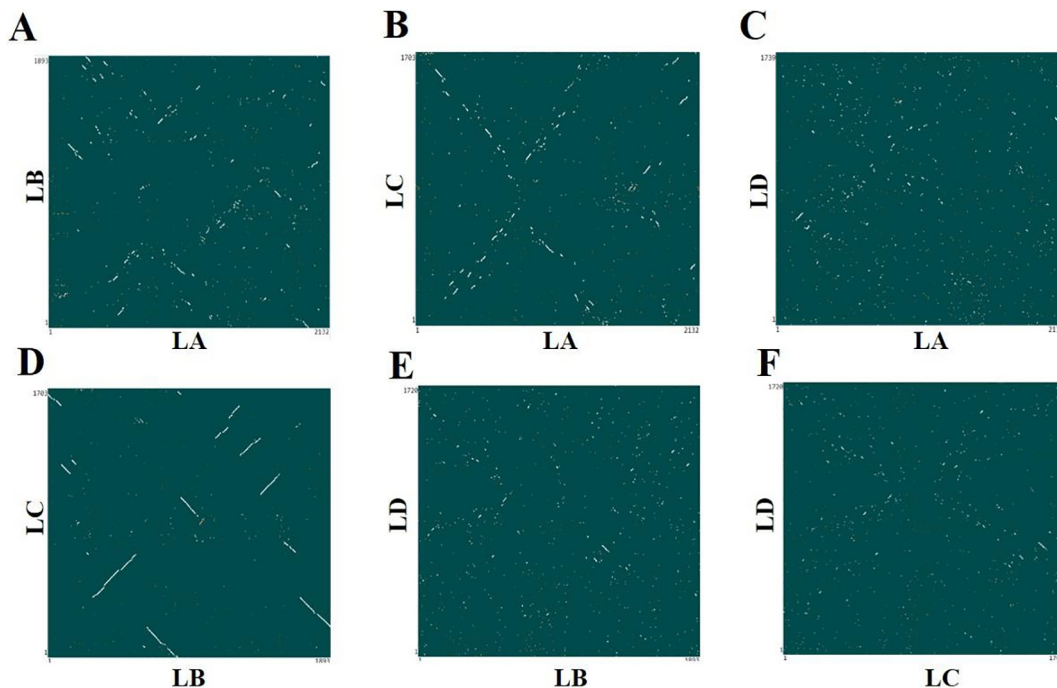

Fig. 6. Dot plot matrices between (A) LA vs LB, (B) LA vs LC, (C) LA vs LD, (D) LB vs LC, (E) LB vs LD, and (F) LC vs LD. Dot plot matrices were generated using Kbase with default parameters.

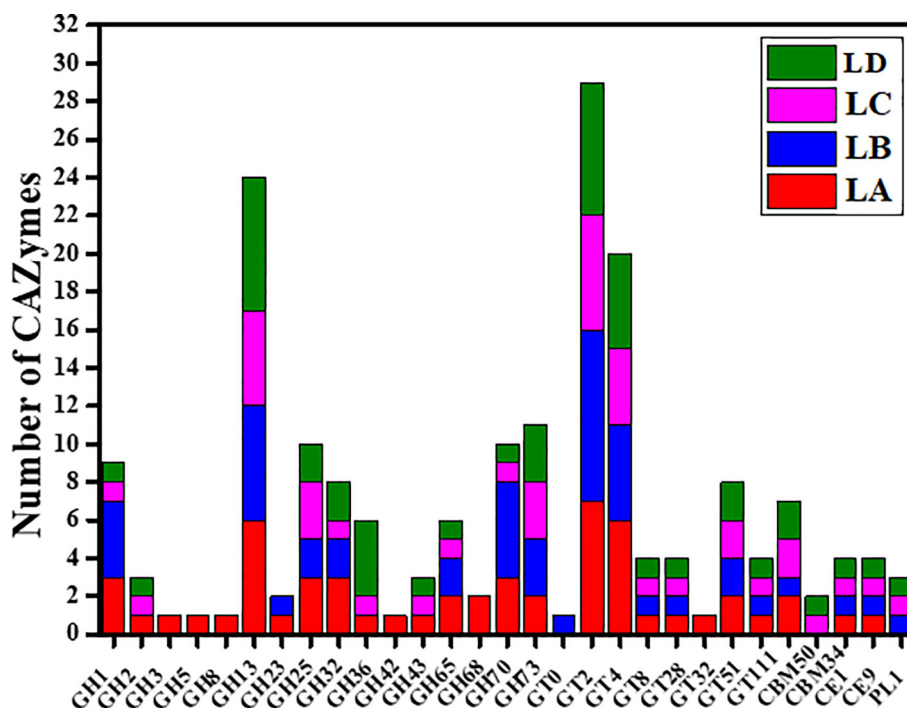**Fig. 7.** MAUVE alignment of (A) LC with (B) LD, (C) LB, and (D) LA.

Fig. 8. Number of CAZymes in the four *Leuconostoc* species. Color codes represent the different bacterial species: red, LA; blue, LB; pink, LC; green, LD; GH, glycoside hydrolase; GT, glycosyltransferase; CBM, carbohydrates-binding module; CE, carbohydrate esterase; PL, polysaccharide lyase. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and 7 GT, and 8 GH and 7 GT families, respectively, in LA and LB. In LC and LD, genes were distributed among 10 GH and 6 GT, and 10 GH and 7 GT gene families, respectively. Notably, in GHs and GTs, GH13 and GT2 were the most dominant in all strains. The enzymes GHs and PLs cleave glycosidic bonds between carbohydrates or carbohydrates and non-carbohydrate moieties. GHs, in particular,

hydrolyze glycosidic bonds, whereas PLs use an elimination mechanism to cleave complex carbohydrates [48–50].

The HMMER, DIAMOND, and Hotpep algorithms predicted 32, 25, 18, and 22 genes and 19, 20, 15, and 17 genes encoding GHs and GTs in LA, LB, LC, and LD, respectively (Table S4). Some GHs in this study are modular enzymes with catalytic and non-

catalytic domains such as carbohydrate-binding modules (CBM) [51]. Furthermore, all strains encoded GH1, GH13 (GH 13–18 in CCK940), GH25, GH32, GH65, GH70, and GH74, whereas GT2, GT4, GT8, GT28, GT51, and GT111 were found in all strains. Major oligosaccharide-degrading enzyme families have been identified as GH32, GH13, and GH2 [52].

In general, GH13 belongs to the α-amylase group, has 4–7 conserved sequence regions and catalytic machinery, and adopts the (α/β) eight-barrel fold of the catalytic domain [53]. A previous study found the GH13 family, which includes amylases, beta-xylosidases, beta-glucosidases, and beta-galactosidase, in the genome of *Leuconostoc* sp. MTCC 10,508 [19]. Meanwhile, the GH70 family is classified as transglycosylase, which catalyzes the inter- or intramolecular substitution of glycoside molecules at the anomeric location, resulting in new glycoside molecules [54]. Peptidoglycan hydrolase (GH73 family) cleaves glycosidic (β-1,4) linkages between *N*-acetylglucosaminyl and *N*-acetylmuramyl residues in bacterial peptidoglycan [43].

GTs participate in glycosidic bond formation by transferring the sugar moiety from the activated sugar donor to the acceptor molecule [49]. GT2 and GT4 (polyspecific families) account for roughly half of the total number of GTs required for cellulose and sucrose synthesis, respectively, among the GT families listed in the CAZy database [44]. Furthermore, the GT2 and GT4 families are responsible for the majority of glycosyltransferases in *Leuconostoc* spp. [47]. In this study, the GT2 family was prominent among the pre-

dicted GTs in the *Leuconostoc* spp. genomes, with seven, nine, six, and seven genesin the LA, LB, LC, and LD, respectively. GT2 functions include galactosyltransferase, chitin synthase, cellulose synthase, glucosyltransferase, rhamnosyltransferase, and mannosyltransferase [55].

5. Oligosaccharide utilization pathway in *Leuconostoc* spp.

Based on the predicted KEGG pathways and BLASTp analysis of genes related to carbohydrate metabolism, the metabolic pathways of the selected strains encoded GH families involved in oligosaccharide utilization routes (Fig. 9). The four strains contained The GH1, GH2, GH13_30, GH13_31, GH13_18, and GH65 families, which were primarily associated with the metabolism of gentiobiose-6-P and cellobiose-6-P (GH1), glucooligosaccharide (lactose) (GH2), maltotriose (GH13_30), isomaltose panose (GH13_31), glucose + fructose (GH13_18), and maltose (GH65). Furthermore, GH42 (glucooligosaccharide), GH36 (melibiose, raffinose, and stachyose), GH43 (xylobiose and xylooligosaccharides), and GH42 (lactitol) were found in LA, LC, and LD but not in LB. Notably, GH30, which may be associated with the metabolism of gentiobiose, was not found in any of the strains studied (Fig. 9). Except for strain LB, all strains contained GH36 (α-galactosidase (EC 3.2.1.22) and β-fructofuranosidase (EC 3.2.1.26), which are mainly used for the utilization of raffinose and other sugars

Table 2
Genes related to carbohydrate-active enzymes in the four *Leuconostoc* species.

	GH	GT	CE	CBM	PL
SG315	**32** GH32(14-326), GH13_18(36-375), GH73(288-428), GH13_29(25-371), GH25(58-218), GH73(66-204), GH32 (49-359), GH13_31(28-377), GH42(13-379), GH5_44(48-352), GH43_11(10-321), GH8(30-365), GH25(32-220), GH23, GH32(51-353), GH68(175-614), GH68(179-618), GH25(47-227), GH3(35-252), GH65(317-693), GH36(14-719), GH70(332-1127), GH70 (2011-2804), GH65(317-694), GH13_31(29-380), GH13_31(30-381), GH1(9-482), GH13_31(29-382), GH1(1-474), GH1(3-473), GH2(33-918), GH70(292-1108), GH70(370-1167)	**19** GT4 (197-346), GT4 (160-302), GT51(104-284), GT32 (19-93), GT28 (189-349), GT2_Glycos_transf_2(108-278), GT2_Glycos_transf_2(6-172), GT2_Glycos_transf_2 (6-131), GT2_Glycos_transf_2 (15-147), GT111 (4-216), GT2_Glycos_transf_2 (9-168), GT4 (335-485), GT4 (314-463), GT4 (303-450), GT2_Glycos_transf_2 (58-235), GT8 (3-252), GT51 (70-250), GT4 (177-325), GT2_Glycos_transf_2(13-177)	**2** CE9 (16-381), CE1 (10-253)	**2** CBM50, CBM50	**NI**
SG255	**25** GH73(695-846), GH65(317-693), GH13_31(29-380), GH13_31(30-381), GH73(63-201), GH1(8-490), GH70 (349-1145), GH1(2-480), GH1(3-478), GH25(61-221), GH65(308-722), GH13_31(29-380), GH70(2-465), GH70 (312-626), GH70(851-1647), GH32(14-326), GH25(32-220), GH23, GH1(3-452), GH13_18(1-191), GH32(22-328), GH73(295-434), GH13_18(37-373), GH70(453-1311), GH13_31(29-382)	**20** GT28(188-352), GT51(102-282), GT2_Glycos_transf_2(5-142), GT2_Glycos_transf_2(8-176), GT2_Glycos_transf_2(5-132), GT51(71-249), GT2_Glycos_transf_2(9-173), GT2_Glycos_transf_2(30-166), GT4(214-348), GT2_Glycos_transf_2(14-182), GT2_Glycos_transf_2(12-145), GT111(7-222), GT0, GT2_Glyco_trans_2_3(54-275), GT4(159-301), GT4(197-345), GT8(9-260), GT4(176-325), GT4(303-445), GT2_Glycos_transf_2(57-232)	**2** CE9(5-374), CE1(10-257)	**1** CBM50	**1** PL1_6 (145-354)
CCK940	**18** GH32(14-326), GH13_18(37-373), GH2(37-916), GH36 (14-718), GH13_29(25-371), GH25(63-223), GH25(34-221), GH73(297-437), GH73(65-203), GH43_11(3-304), GH65(317-693), GH13_31(32-365), GH13_31(28-362), GH73(465-601), GH1(6-485), GH13_31(30-381), GH25(536-722), GH70 (345-1141)	**15** GT2_Glycos_transf_2(8-119), GT111(7-223), GT2_Glycos_transf_2(58-234), GT2_Glycos_transf_2(57-232), GT4(302-440), GT4(175-324), GT8(11-261), GT4(195-343), GT4(160-299), GT28(188-352) GT51(104-284), GT2_Glycos_transf_2(8-175) GT2_Glycos_transf_2(5-174) GT2_Glycos_transf_2(9-135), GT51(71-250)	**2** CE9(10-356), CE1(10-257)	**1** CBM34(6-133) + GH13_20 (183-487)	**1** PL1_6 (146-354)
SBC001	**22** CBM34(6-133) + GH13_20(183-487), GH13_31(28-362), GH13_31(32-365), GH65(317-693), GH73(65-203), GH73(297-437), GH70(343-1139), GH43_11(3-304), GH25(63-223), GH13_29(25-371), GH36(14-718), GH2 (37-916), GH13_18(37-373), GH32(14-326), GH25(536-722), GH36(14-719), GH13_31(30-381), GH1(6-485), GH13_31(30-378), GH32(37-339), GH36(12-700), GH36 (12-711)	**17** GT4(39-188), GT51(104-284), GT28(188-352), GT4(160-299), GT4(195-343), GT8(11-261), GT2_Glyco_tranf_2_3 (55-290), GT4, GT2_Glycos_transf_2(7-173), GT2_Glycos_transf_2(17-138), GT111(7-223), GT2_Glycos_transf_2(8-119), GT51(71-250), GT4(302-440), GT2_Glycos_transf_2(9-135), GT2_Glycos_transf_2 (5-174), GT2_Glycos_transf_2(8-175)	**2** CE1(10-257)CE9 (10-356)	**1** CBM50	**1** PL1_6 (146-354)

GH, glycoside hydrolases; GT, glycosyltransferases; CE, carbohydrate esterases; CBM, carbohydrate-binding modules; PL, polysaccharide lyases; NI, not identified; SBC001, *Leuconostoc lactis* SBC001; CCK940, *Leuconostoc lactis* CCK940, SG255: *Leuconostoc citrem* SG255, SG315: *Leuconostoc mesenteroides* SG315.

commonly found in plants. A recent study found that enzymes involved in raffinose metabolism were not present in the genomes of 56 LAB, including *L. mesenteroides* and *L. citreum* [56]. D-raffinose is a trisaccharide composed of galactose, glucose, and fructose, that is hydrolyzed by galactosidase (EC 3.2.1.22) to D-galactose and sucrose. This enzyme is absent in the human digestive tract [56]. According to Gemma Buron-Moles et al. [56] *Leuconostoc* spp. do not metabolize glycogen. In this study, comparable results were observed for all *Leuconostoc* spp. (data not shown). Similarly, with the exception of LB, all strains contained the GH42 and GH43 family (Fig. 9), which includes β-galactosidase, α-L-arabinosidase, and xylanases. Remarkably, both GH42 and GH2 enzymes are annotated as β-galactosidase. GH2 β-galactosidases hydrolyze glucooligosaccharides (GOS), including oligosaccharides from lactose [57]. In *L. lactis*, this enzyme is encoded by two overlapping genes: *lacL* (large subunit, 1,878 bp) and *lacM* (small subunit, 963 bp) [58]. In another comparative study using dairy-isolated species, *L. mesenteroides, L. cremoris*, and *L. pseudomesenteroides*, but not *L. lactis*, harbored *lacL* and *lacM* [59]. Notably, GH42 β-galactosidases (*lacA*) have been reported to have a low activity against lactose [60]. These findings suggest that *Leuconostoc* spp. can utilize specific oligosaccharides. GH13_30 is also found in all *Leuconostoc* strains and is classified as α-glucosidase (EC 3.2.1.20), a carbohydrate-hydrolyzing enzyme that hydrolyzes maltotriose by breaking α-1,4-glycosidic bonds [61]. GH13_31 and GH65, annotated as 1,6-α-glucosidase (EC 3.2.1.10, also known as isomaltose and alpha-methyl glucosidases) and maltose phosphorylase (EC 2.4.1.8), respectively, are present in all *Leuconostoc* spp. Maltose phosphorylase catalyzes the conversion of maltose to glucose and glucose-1-phosphate through phosphorolysis [62]. Previous research found genes encoding maltose phosphorylase

in *L. mesenteroides*, *L. lactis*, and *L. pseudomesenteroides*, but not *L. cremoris* [59]. These results indicate that the maltooligosaccharide utilization locus was present in all selected *Leuconostoc* spp. Furthermore, GH1, also known as 6-phospho-β-glucosidase (EC 3.2.1.86), was found in all selected strains of this study. This enzyme catalyzes the hydrolytic cleavage of cellobiose 6-phosphate (β-1,4-linked) to yield glucose 6-phosphate and glucose [63]. A recent study described the important role of 6-phospho-β-glucosidases in *L. pseudomesenteroides* DSM 20,193 in plant-based fermentation [64]. GH42 and β-galactosidase (EC 3.2.1.23) may be involved in the production of galactose and glucitol. GH43 was identified as β-xylosidase (EC 3.2.1.37). Except for LB, all strains contained β-1,3-xylosidase (EC 3.2.1.-), which may use xylooligosaccharides (XOS) [65]. XOSs are a group of oligosaccharides that include xylotetrose, -triose, and -biose, which comprise xylose residues linked by β-1,4 linkages and are found naturally in vegetables fruits, bamboo shoots, milk, and honey [66]. With the exception of strain LB, all strains consumed galactose *via* the Leloir pathway, indicating the lack of lactose phosphotransferase system activity [67]. The xylosidase enzyme of *L. lactis* SHO-47 has been proposed to be cytoplasmically localized [68].

6. Folic acid biosynthesis

In recent years, there has been increased interest in folate, a water-soluble vitamin, in relation to nutrition [69]. Plants, vegetables, fruits, and meats are the best sources of folate. While many studies have shown that most LAB strains are deficient in folate, certain strains can generate natural folate [70]. These microorganisms can be used as folic acid substitutes.
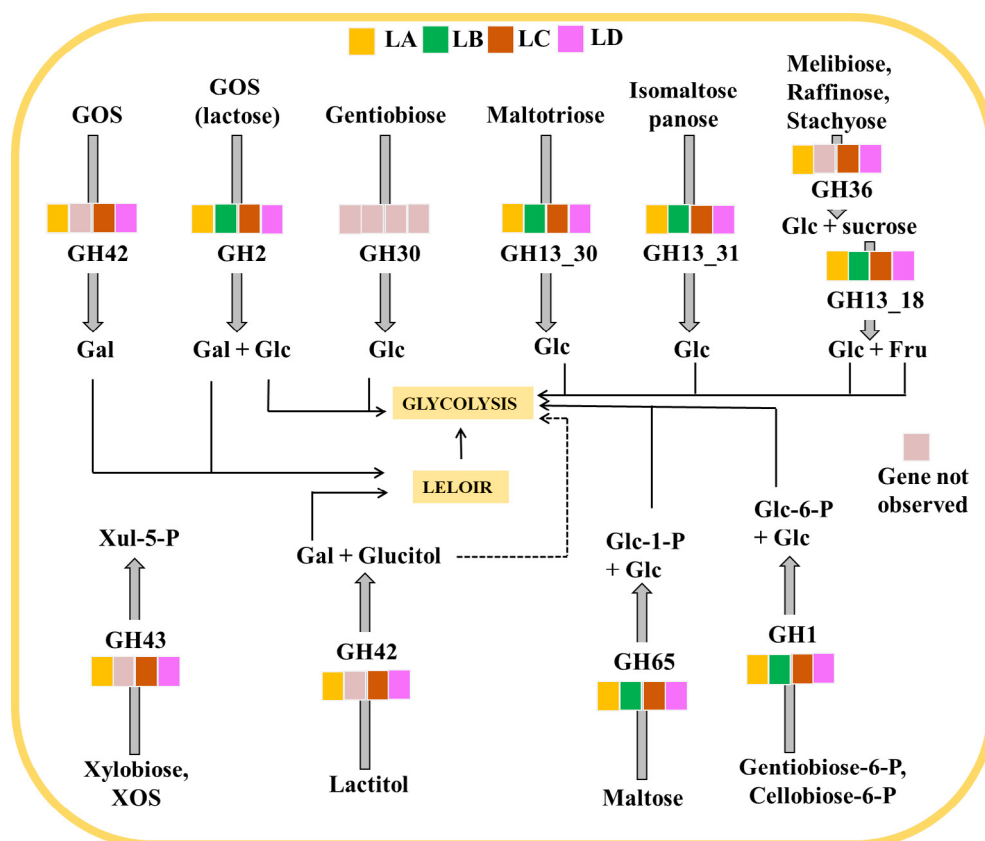

Fig. 9. Predicted oligosaccharide utilization pathways of various carbohydrates in *Leuconostoc* spp. GH, glycoside hydrolases; GOS, glucooligosaccharide; XOS, xylooligosaccharides; Gal, galactose; Glc, glucose; Fru, fructose; Glc-1-P, glucose-1-phosphate; Glc-6-P, glucose-6-phosphate; Xul-5-P, xylulose-5-phosphate.
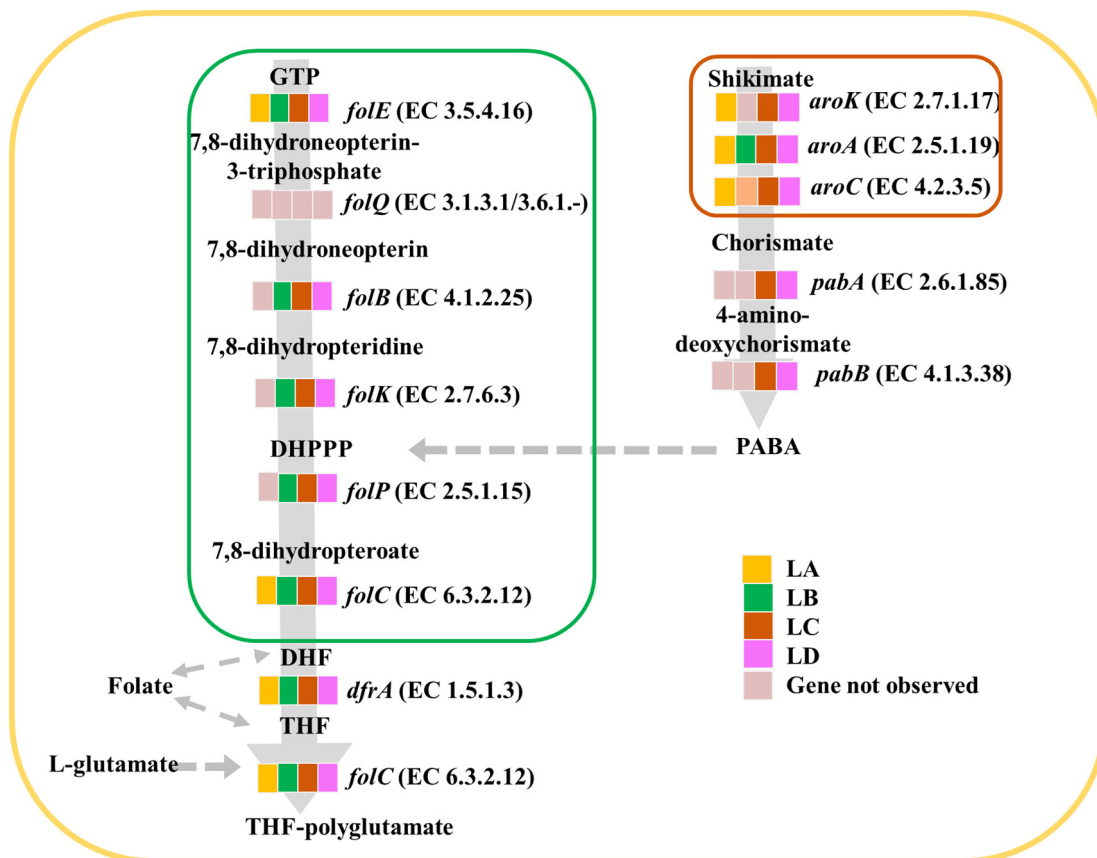
Fig. 10. Overview of the folate biosynthesis pathway in the four *Leuconostoc* species. Different color codes are used to represent the presence of genes in different *Leuconostoc* species. Grey boxes represent the absence of a gene. GTP: guanosine triphosphate, PABA: *para*-aminobenzoic acid, DHPPP: hydroxymethyl-7,8-dihydropterin pyrophosphate, DHF: dihydrofolate, THF: tetrahydrofolate.

In this study, only LC and LD were observed to synthesize folic acid *via* two metabolic pathways commonly found in LAB (Fig. 10). The first pathway converts guanosine triphosphate (GTP) to 6-hy droxymethyl-7,8-dihydropterin pyrophosphate (DHPPP), while the second converts chorismite to *para*-aminobenzoic acid (PABA) [71]. The study also observed the presence of enzymes related to the GTP pathway in LC and LD genomes: *folA* (DHF reductase, EC 1.5.1.3), *folB* (EC 4.1.2.25), *folC* (EC 6.3.2.12 dihydrofolate synthase, and folylpolyglutamate synthase), *folE* (GTP cyclohydrolase I, EC 3.5.4.16), *folK* (2-amino-4-hydroxy-6-hydroxymethyl-dihydropteri dine diphosphokinase, EC 2.7.6.3), and *folP* (dihydropteroate syn-thase, EC 2.5.1.15). Furthermore, as illustrated in Fig. 10, *folC* cat-alyzes the synthesis of DHF and polyglutamylation in *Leuconostoc* spp. [72]. Some LAB lack *folQ* (EC 3.1.3.1/3.6.1.-) which encodes the DHNTPase enzyme responsible for folate biosynthesis; how-ever, they have alternative enzymes that are similar to DHNTPase [73]. The genomes of LC and LD did not contain *folQ.* In the Shiki-mate pathway (alternate route), chorismate is converted into 4-amino-4-deoxychorismate by *pabA* (aminodeoxychorismate syn-thase, EC 2.6.1.85). Following that, 4-amino-4-deoxychorismate is converted to pyruvate, which is cleaved by *pabB* (4-amino-4-deoxychorismate lyase, EC 4.1.3.38), yielding *para*-aminobenzoic acid (PABA), which is then used in folate biosynthesis. However, in LA, this pathway was found to be incomplete, and the xylulokinase-producing gene (*aroK* EC 2.7.1.17) was absent in the LB strain.

7. Conclusion

This study analyzed the draft genome sequences and functional attributes of four *Leuconostoc* strains. Based on ANI, the two differ-

ent strains of *L. lactis* have the highest similarity among the four species, and low strain similarity could be due to gene arrange-ments. Although all of the strains shared a common gene pool, there were differences in their gene contents. Furthermore, many genes with unknown functions have been observed in the proph-age region, which can be investigated further in future studies. Functional profiling revealed the presence of various CAZymes among the four strains, correlating with previous research. The dif-ferent enzymes involved in oligosaccharide metabolism are also highlighted in this study Meanwhile, only two strains of *L. lactis* showed the complete folic acid synthesis pathway In conclusion, these findings will aid in the understanding of genetic similarity among *Leuconostoc* strains and can help in the development of new strategies for industrial applications through the production of genetically-modified strains.

CRediT authorship contribution statement

Anshul Sharma: Methodology, Writing – original draft, Valida-tion, Data curation. **Neha Sharma:** Methodology, Writing – original draft, Validation, Data curation. **Deepshikha Gupta:** Methodology, Software, Visualization, Formal analysis, Investigation, Writing – review & editing. **Hae-Jeung Lee:** Conceptualization, Data curation. **Young-Seo Park:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing finan-cial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.08.032.

References

[1] Schmidt H, Hensel M. Pathogenicity islands in bacterial pathogenesis. Clin Microbiol Rev 2004;17:14–56.

[2] Shin S-Y, Han NS. *Leuconostoc* spp. as starters and their beneficial roles in fermented foods. In: Liong M–T, editor. Beneficial microorganisms in food and nutraceuticals. Switzerland: Springer; 2015. p. 111–32.

[3] Ogier J-C, Casalta E, Farrokh C, Saïhi A. Safety assessment of dairy microorganisms: the *Leuconostoc* genus. Int J Food Microbiol 2008;126:286–90.

[4] Rezac S, Kok CR, Heermann M, Hutkins R. Fermented foods as a dietary source of live organisms. Front Microbiol 2018;9:1785.

[5] Zheng J, Wittouck S, Salvetti E, Franz CM, Harris H, et al. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. Int J Syst Evol Microbiol 2020;70:2782–858.

[6] Hucker G, Pederson C (1930) Studies on the coccaceae. XVI. The genus *Leuconostoc*. New York State Agr Exp Sta Tech Bull No. 167. Geneva.

[7] Schleifer KH. Leuconostocaceae fam. nov. Bergey's Manual of Systematics of Archaea and Bacteria. Wiley & Sons Inc.; 2015.

[8] Sybesma W, Starrenburg M, Tijsseling L, Hoefnagel MH, Hugenholtz J. Effects of cultivation conditions on folate production by lactic acid bacteria. Appl Environ Microbiol 2003;69:4542–8.

[9] Ahn S-B, Lee S-M, Shon M-Y, Kim S-Y, Shin M-S, et al. Immune-enhancing effects of *Leuconostoc* strains isolated from kimchi. J Biomed Res 2012;3:353–6.

[10] Li L, Shin SY, Lee K, Han N. Production of natural antimicrobial compound d-phenyllactic acid using *Leuconostoc mesenteroides* ATCC 8293 whole cells involving highly active d-lactate dehydrogenase. Lett Appl Microbiol 2014;59:404–11.

[11] Poulsen VK, Koza A, Al-Nakeeb K, Oeregaard G. Screening for texturing *Leuconostoc* and genomics behind polysaccharide production. FEMS Microbiol Lett 2020;367:fnaa179.

[12] Du R, Qiao X, Wang Y, Zhao B, Han Y, et al. Determination of glucansucrase encoding gene in *Leuconostoc mesenteroides*. Int J Biol Macromol 2019;137:761–6.

[13] Münkel F, Bechtner J, Eckel V, Fischer A, Herbi F, et al. Detailed structural characterization of glucans produced by glucansucrases from *Leuconostoc citreum* TMW 2.1194. J Agric Food Chem 2019;67:6856–66.

[14] Zikmanis P, Brants K, Kolesovs S, Semjonovs P. Extracellular polysaccharides produced by bacteria of the *Leuconostoc* genus. World J Microbiol Biotechnol 2020;36:1–18.

[15] Kim M, Jang J-K, Park Y-S. Production optimization, structural analysis, and prebiotic-and anti-inflammatory effects of gluco-oligosaccharides produced by *Leuconostoc lactis* SBC001. Microorganisms 2021;9:200.

[16] Lee S, Park G-G, Jang J-K, Park Y-S. Optimization of oligosaccharide production from *Leuconostoc lactis* using a response surface methodology and the immunostimulating effects of these oligosaccharides on macrophage cells. Molecules 2018;23:2118.

[17] Kim S, Lee S, Park Y-S. Draft genome sequence of oligosaccharide-synthesizing *Leuconostoc citreum* SG255 isolated from radish kimchi in Korea. Korean J Microbiol 2021;57:69–71.

[18] Lee S, Park Y. Oligosaccharide production by *Leuconostoc lactis* CCK940 which has glucansucrase activity. Food Eng Prog 2017;21:383–90.

[19] Kaushal G, Singh SP. Comparative genome analysis provides shreds of molecular evidence for reclassification of *Leuconostoc mesenteroides* MTCC 10508 as a strain of *Leu. suionicum*. Genomics 2020;112:4023–31.

[20] Candeliere F, Raimondi S, Spampinato G, Tay MYF, Amaretti A, et al. Comparative genomics of *Leuconostoc carnosum*. Front Microbiol 2020;11:605127.

[21] Kumar S, Bansal K, Sethi SK. Comparative genomics analysis of genus Leuconostoc resolves its taxonomy and elucidates its biotechnological importance. Food Microbiol 2022;106:104039.

[22] Honda H, Yajima N, Saito T. Characterization of lactose utilization and β-galactosidase in *Lactobacillus brevis* KB290, the hetero-fermentative lactic acid bacterium. Curr Microbiol 2012;65:679–85.

[23] Jeckelmann JM, Erni B. Transporters of glucose and other carbohydrates in bacteria. Pflugers Arch - Eur J Physiol 2020;472:1129–53. https://doi.org/10.1007/s00424-020-02379-0.

[24] Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 2017;110:1281–6.

[25] Allen B, Drake M, Harris N, Sullivan T (2017) Using KBase to assemble and annotate prokaryotic genomes. Curr Protoc Microbiol 46: 1E. 13.1-1E. 13.18.

[26] Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE 2010;5:e11147.

[27] Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res 2020;48:D606–12.

[28] Bertelli C, Laird MR, Williams KP, Group SFURC, Lau BY, et al. (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 45(W1): W30-W35.

[29] Zhang H, Yohe T, Huang L, Entwistle S, Wu P, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 2018;46(W1):W95–W101.

[30] Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, et al. HMMER web server: 2018 update. Nucleic Acids Res 2018;46(W1):W200–4.

[31] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods 2015;12:59–60.

[32] Busk PK, Pilgaard B, Lezyk MJ, Meyer AS, Lange L. Homology to peptide pattern for annotation of carbohydrate-active enzymes and prediction of function. BMC Bioinform 2017;18:1–9.

[33] Tarrah A, Pakroo S, Junior WJFL, Guerra AF, Corich V, et al. Complete genome sequence and Carbohydrates-Active EnZymes (CAZymes) analysis of *Lactobacillus paracasei* DTA72, a potential probiotic strain with strong capability to use inulin. Curr Microbiol 2020;77:2867–75.

[34] Kim S-A, Jang Y-J, Heo JE, Li L, Moon JS, Han NS. Complete genome sequence of *Leuconostoc citreum* EFEL2700, a host strain for transformation of pCB vectors. J Biotechnol 2018;287:52–8.

[35] Wright SM, Carroll C, Walters A, Newell PD, Chaston JM. Genome sequence of *Leuconostoc citreum* DmW_111, isolated from wild Drosophila. Genome Announc 2017;5:e00507–17.

[36] Morita H, Toh H, Oshima K, Nakano A, Hano C, Yoshida S, et al. Draft genome sequence of *Leuconostoc mesenteroides* 406 isolated from the traditional fermented mare milk Airag in Tuv Aimag. Mongolia Genome Announc 2016;4(2). e00166–00116.

[37] Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, et al. The COG database: an updated version includes eukaryotes. BMC Bioinform 2003;4:1–14.

[38] Qin QL, Xie BB, Yu Y, Shu YL, Rong JC, et al. Comparative genomics of the marine bacterial genus G laciecola reveals the high degree of genomic diversity and genomic characteristic for cold adaptation. Environ Microbiol 2014;16:1642–53.

[39] Rodriguez-R LM, Castro JC, Kyrpides NC, Cole JR, Tiedje JM, et al. How much do rRNA gene surveys underestimate extant bacterial diversity? Appl Environ Microbiol 2018;84:e00014–8.

[40] Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci 2009;106:19126–31.

[41] Croucher NJ, Coupland PG, Stevenson AE, Callendrello A, Bentley SD, et al. Diversification of bacterial genome content through distinct mechanisms over different timescales. Nature Commun 2014;5:1–12.

[42] Bellanger X, Payot S, Leblond-Bourget N, Guédon G. Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol Rev 2014;38:720–60.

[43] Zaloba P, Bailey-Elkin BA, Derksen M, Mark BL. Structural and biochemical insights into the peptidoglycan hydrolase domain of FlgJ from *Salmonella typhimurium*. PLoS ONE 2016;11:e0149204.

[44] Mao B, Yin R, Li X, Cui S, Zhang H, et al. Comparative genomic analysis of *Lactiplantibacillus plantarum* isolated from different niches. Genes 2021;12:241.

[45] Canchaya C, Proux C, Fournous G, Bruttin A, Brüssow H. Prophage genomics. Microbiol Mol Biol Rev 2003;67:238–76.

[46] Sun S, Ke R, Hughes D, Nilsson M, Andersson DI. Genome-wide detection of spontaneous chromosomal rearrangements in bacteria. PLoS ONE 2012;7:e42639.

[47] Kumar S, Bansal K, Sethi SK, Bindhani BK. Phylo-taxonogenomics of 182 strains of genus *Leuconostoc* elucidates its robust taxonomy and biotechnological importance. BioRxiv 2021. https://doi.org/10.1101/2021.07.20.453025.

[48] Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 1991;280:309–16.

[49] Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 2009;1:D233–8.

[50] Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, et al. A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J 2010;432:437–44.

[51] Cuskin F, Flint JE, Gloster TM, Morland C, Baslé A, et al. How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. Proc Natl Acad Sci 2012;109:20889–94.

[52] Wang L, Zhang G, Xu H, Xin H, Zhang Y. Metagenomic analyses of microbial and carbohydrate-active enzymes in the rumen of holstein cows fed different forage-to-concentrate ratios. Front Microbiol 2019;10:649.

[53] Martinovičová M, Janeček Š. In silico analysis of the α-amylase family GH57: eventual subfamilies reflecting enzyme specificities. 3. Biotech 2018;8:1–11.

[54] Bissaro B, Monsan P, Fauré R, O'Donohue MJ. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. Biochem J 2015;467:17–35.

[55] Breton C, Šnajdrová L, Jeanneau C, Koča J, Imberty A. Structures and mechanisms of glycosyltransferases. Glycobiology 2006;16:29R–37R.

[56] Buron-Moles G, Chailyan A, Dolejs I, Forster J, Mikš MH. Uncovering carbohydrate metabolism through a genotype-phenotype association study of 56 lactic acid bacteria genomes. Appl Microbiol Biotechnol 2019;103:3135–52.

[57] Gänzle MG. Enzymatic synthesis of galacto-oligosaccharides and other lactose derivatives (hetero-oligosaccharides) from lactose. Int Dairy J 2012;22:116–22.

[58] David S, Stevens H, Van Riel M, Simons G, De Vos W. Leuconostoc lactis beta-galactosidase is encoded by two overlapping genes. J Bacteriol 1992;174:4475–81.

[59] Frantzen CA, Kot W, Pedersen TB, Ardö YM, Broadbent JR, et al. Genomic characterization of dairy associated Leuconostoc species and diversity of Leuconostocs in undefined mixed mesophilic starter cultures. Front Microbiol 2017;8:132.

[60] Schwab C, Sørensen KI, Gänzle MG. Heterologous expression of glycoside hydrolase family 2 and 42 β-galactosidases of lactic acid bacteria in Lactococcus lactis. Syst Appl Microbiol 2010;33:300–7.

[61] Wangpaiboon K, Laohawuttichai P, Kim S-Y, Mori T, Nakapong S, et al. A GH13 α-glucosidase from Weissella cibaria uncommonly acts on short-chain maltooligosaccharides. Acta Crystallogr D 2021;77:1064–76.

[62] Stolz P, Hammes WP, Vogel RF. Maltose-phosphorylase and hexokinase activity in lactobacilli from traditionally prepared sourdoughs. Advances Food Sci 1996;18:1–6.

[63] Yu W-L, Jiang Y-L, Pikis A, Cheng W, Bai X-H, et al. Structural insights into the substrate specificity of a 6-phospho-β-glucosidase BglA-2 from Streptococcus pneumoniae TIGR4. J Biol Chem 2013;288:14949–58.

[64] Acin-Albiac M, Filannino P, Arora K, Da Ros A, Gobbetti M, et al. Role of lactic acid bacteria phospho-β-glucosidases during the fermentation of cereal by-products. Foods 2021;10:97.

[65] Falck P, Linares-Pastén JA, Adlercreutz P, Karlsson EN. Characterization of a family 43 β-xylosidase from the xylooligosaccharide utilizing putative probiotic Weissella sp. strain 92. Glycobiology 2016;26:193–202.

[66] Vazquez M, Alonso J, Domınguez H, Parajo J. Xylooligosaccharides: manufacture and applications. Trends Food Sci Technol 2000;11:387–93.

[67] Huang D, Prevost H, Divies C. Interrelationship of sugar metabolism (glucose, galactose, lactose) by Leuconostoc mesenteroides subsp mesenteroides. Le Lait 1994;74:207–15.

[68] Ohara H, Owaki M, Sonomoto K. Xylooligosaccharide fermentation with Leuconostoc lactis. J Biosci Bioeng 2006;101:415–20.

[69] Rossi M, Amaretti A, Raimondi S. Folate production by probiotic bacteria. Nutrients 2011;3:118–34.

[70] Chen L, Gu Q, Li P, Chen S, Li Y. Genomic analysis of Lactobacillus reuteri WHH 1689 reveals its probiotic properties and stress resistance. Food Sci Nutr 2019;7:844–57.

[71] Magnúsdóttir S, Ravcheev D, de Crécy-Lagard V, Thiele I. Systematic genome assessment of B-vitamin biosynthesis suggests co-operation among gut microbes. Front Genetics 2015;6:148.

[72] de Crécy-Lagard V, El Yacoubi B, de la Garza RD, Noiriel A, Hanson AD. Comparative genomics of bacterial and plant folate synthesis and salvage: predictions and validations. BMC Genomics 2007;8:1–15.

[73] He S, Li X, Liu C. Research progress of the folate synthesized by lactic acid bacteria. Microbiol China 2015;42:1994–2001.