Scientific Reports

https://doi.org/10.1038/s41598-026-51219-1

Article in Press

# Pangenomics of *Limosilactobacillus fermentum* reveals genomic diversity and bacteriocin activity against *Staphylococcus aureus* and *Escherichia coli*

Received: 14 August 2025

Accepted: 27 April 2026

Published online: 17 May 2026

Cite this article as: Kasapoglu M.Z., Acar I., Gumustop I. *et al.* Pangenomics of *Limosilactobacillus fermentum* reveals genomic diversity and bacteriocin activity against *Staphylococcus aureus* and *Escherichia coli*. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-51219-1

Muhammed Zahid Kasapoglu, Idil Acar, Ismail Gumustop, Ismail Erol, Ibrahim Cagri Kurt & Fatih Ortakci

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

# Pangenomics of *Limosilactobacillus fermentum* reveals genomic diversity and bacteriocin activity against *Staphylococcus aureus* and *Escherichia coli*

Muhammed Zahid Kasapoglu[1*], Idil Acar[1], Ismail Gumustop[2], Ismail Erol[3], Ibrahim Cagri Kurt[2], Fatih Ortakci[4*]

[1]Department of Nanotechnology, Istanbul University-Cerrahpaşa, Istanbul, Türkiye

[2]Department of Molecular Biology and Genetics, Boğaziçi University, Istanbul, Türkiye

[3]School of Pharmacy, Department of Analytical Chemistry, Bahçeşehir University, Istanbul, Türkiye

[4]Department of Food Engineering, Istanbul Technical University, Istanbul, Türkiye

***Corresponding authors:** muhammed.kasapoglu@iuc.edu.tr, ortakci@itu.edu.tr

ARTICLE IN PRESS

## Abstract

*Limosilactobacillus fermentum* is a versatile lactic acid bacterium with significant probiotic and biotechnological potential, yet the genomic determinants underlying its ecological adaptation and therapeutic applications remain underexplored. This study performed a comparative genomic analysis of 52 *L. fermentum* strains, with a focus on the novel strain ATT-06 isolated from traditional Turkish shalgam. We uncovered substantial genetic diversity, evidenced by an open pan-genome (18,647 genes) and a small core genome (718 genes). CRISPR-Cas systems were prevalent (46 strains), with Type-IE and Type-IIA being most common and mutually exclusive from Type-IC. Strain ATT-06, which uniquely produced gamma-aminobutyric acid at 17.74 µg/mL, harbored a Type-IIA CRISPR system and a single prophage. In silico molecular docking and dynamics simulations revealed that the bacteriocin Lactococcin, encoded by strain ATA-LTC-Lf170503, exhibited strong binding affinities (ΔG: -8.1 to -13.0 kcal/mol) against Rho proteins of Staphylococcus aureus and Escherichia coli, outperforming

ARTICLE IN PRESS

Acidocin A. These findings highlight the genomic plasticity and adaptive mechanisms of *L. fermentum*, and position strain ATT-06 as a promising probiotic candidate with potential neuroactive and antimicrobial applications.

**Key words:** *Limosilactobacillus fermentum*, Comparative genomics, CRISPR-Cas systems, Bacteriocins, Gamma-aminobutyric acid

## Introduction

Lactic acid bacteria (LAB) are Gram (+), catalase (-), nonspore forming bacteria that play a crucial role in many food fermentation ecosystems [1,2]. Due to their extensive metabolic diversity, lactobacilli are found and utilized as functional starter cultures in a wide range of fermented foods, such as cheese, plant-based fermented products, fermented meats, wine, beer, and sourdoughs [3]. Additionally, lactobacilli are significant contributors to the fermentation process of fermented shalgam products [4]. Along with their roles as starters and probiotics, lactic acid bacteria (LAB) serve as potential bioprotective cultures due to their ability to produce peptide-based antimicrobials like nisin, enterocin, pediocin, salivaricin, and plantaricin [5,6] which can be found in various fermented dairy and vegetables [7,8].

ARTICLE IN PRESS

Bacteriocins exhibit various attributes, such as being nontoxic or having a low molecular weight; possessing cationic or neutral peptides; or demonstrating hydrophobic or amphiphilic properties. As a result, their spectrum and mechanism of action can differ [9,10]. Additionally, bacteriocins possess diverse molecular structures, methods of biosynthesis, biochemical characteristics, molecular weights, thermal stabilities, and genetic factors. They are widely acknowledged as safe substances with specific bacteriocins such as nisin and pediocin having obtained formal regulatory approval (e.g., GRAS status) for use in foods. Some bacteriocins may inhibit activity against closely related microorganisms, while others can suppress a wide variety of microorganisms, including pathogens like *Listeria monocytogenes*, *Bacillus cereus*,

ARTICLE IN PRESS

*Staphylococcus aureus*, *Escherichia coli*, *Vibrio parahaemolyticus*, Pseudomonas spp., and *Clostridium tyrobutyricum* [10,11].

Diet is a primary modulator of gut microbiota composition and function, directly influencing host health [12]. Fermented foods, in particular, contribute both live microbes and bioactive metabolites (i.e., short-chain fatty acids, gamma-aminobutyric acid, bacteriocins) that can enhance microbial diversity, strengthen intestinal barrier function, and exert systemic immunomodulatory and neuroactive effects [12–16]. Consequently, the isolation and characterization of novel LAB from dietary-fermented sources like shalgam represent a strategic approach to identifying strains with combined technological, probiotic, and therapeutic potential.

Shalgam, a traditional lactic acid-fermented beverage native to the Southern Anatolia region of Turkey, is recognized as a rich source of lactobacilli. These include species such as *Lactiplantibacillus plantarum*, *Lacticaseibacillus paracasei*, *Levilactobacillus brevis*, *Levilactobacillus parabrevis*, *Lactobacillus acidophilus*, *Lactobacillus gasseri*, *Lactobacillus helveticus*, and *Limosilactobacillus reuteri*, with the latter two species being particularly predominant [17,18]. These autochthonous species are promising candidates for next-generation probiotics due to their historical association with food matrices and presumed resilience to gastrointestinal conditions [19]. *Limosilactobacillus fermentum* is another heterofermentative species of lactic acid bacteria (LAB) found in various food fermentation ecosystems, including sourdough fermentation [20], fermenting cocoa pulp-bean mass [21], and other plant-based fermentations [9]. It is also present in environments such as breast milk [22] and is part of human-associated microbiota, including the colon, urogenital tract, and oral cavity [23,24]. While not always the dominant species in shalgam, its metabolic versatility and probiotic potential make it of significant interest. In this context, our group isolated a novel *Limosilactobacillus fermentum* ATT-06 from artisanal shalgam. Its ecological versatility and

ARTICLE IN PRESS

genomic plasticity make it a valuable model for studying adaptation mechanisms and mining for beneficial traits such as antimicrobial production, vitamin synthesis, and neurotransmitter generation [25,26]. Notably, certain strains exhibit psychobiotic potential through the production of gamma-aminobutyric acid (GABA), a key inhibitory neurotransmitter involved in gut-brain communication [27].

In the present study, we employ a comprehensive comparative genomics approach to explore the strain level diversity across *L. fermentum* species that were isolated from various fermented foods and human gut microbiome and to identify the genetic basis of the traits associated with *L. fermentum* ATT-06. Our objectives were to: i) elucidate the genomic diversity, CRISPR-Cas systems, and mobile genetic elements across the species; ii) characterize the carbohydrate-active enzyme profile and GABA biosynthesis capability of strain ATT-06; and iii) mine the *L. fermentum* pangenome for potential bacteriocins and evaluate their therapeutic potential against *Staphylococcus aureus* and *Escherichia coli*.

## Materials and Methods

### Strain Isolation and Cultivation

*L. fermentum* ATT-06 was isolated from fermented shalgam that was obtained from an artisanal local producer in Adana, Turkey. A 10 mL sample of shalgam was mixed with 90 mL of Maximum Recovery Diluent (Merck, GmbH, Darmstadt, Germany) in a Schott bottle and vortexed at high speed for 1 minute (MS-3 Basic, IKA-Werke GmbH, Staufen, Germany). Serial dilutions were prepared, and 100 µL of the diluted sample was spread onto MRS agar (Merck). The plates were incubated anaerobically at 30°C for 48 hours. The isolate, later named ATT-06, was selected and purified through two rounds of colony isolation. Gram staining and catalase tests were performed on the pure isolate. Cryovial stocks of AGA58 were prepared using MRS broth (Merck) supplemented with 25% glycerol and stored at -80°C.

ARTICLE IN PRESS

**Genomic DNA Extraction**

The cryopreserved culture of *L. fermentum* ATT-06 was sub-cultured twice in MRS broth and incubated anaerobically at 32°C for 24 hours. A 1 mL aliquot of the fresh culture was transferred to a sterile 2 mL microcentrifuge tube and centrifuged at 6000 × g for 10 minutes at 4°C. The supernatant was discarded, and genomic DNA was extracted from the cell pellet using the PureLink Genomic DNA Mini Kit (Invitrogen, Thermo-Fisher Scientific, Carlsbad, CA, USA), following the manufacturer's instructions. The quality and concentration of the extracted DNA were assessed using a Nanodrop and 1.5% agarose gel electrophoresis.

**Whole-Genome Sequencing and De Novo Assembly**

Sequencing libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA), and sequencing was performed on the Illumina Novaseq platform, generating paired-end (PE) reads of 2 × 150 bases. Adapter sequences and contaminants were removed using the JGI-RQC Filter pipeline (BBTools v38.22), retaining only reads corresponding to the ATT-06 strain. The genome was assembled using the BACTERIAL AND VIRAL BIOINFORMATICS RESOURCE CENTER genome assembly service (https://www.bv-brc.org/app/Assembly2) with an auto-assembly strategy [28] and deposited to NCBI with the following BioSample accession number of SAMN57472466.

ARTICLE IN PRESS

**Strains and Data Acquisition**

The 52 strains were selected to represent the ecological diversity of *L. fermentum*, encompassing isolates from fermented foods, human-associated environments, and probiotic products. This selection ensured phylogenetic breadth, provided context for the novel strain ATT-06, and included high-quality, publicly available genomes to facilitate a robust comparative analysis. The following strains were evaluated towards comparative genomic analysis 279, 712, 1019, 45756, A51, ACA-DC, 179, AGA52, ATA-LTC-Lf170503, ATCC 23271, CECT 9269,

ARTICLE IN PRESS

CIRM-BIA 1951, CIRM-BIA 2157, CIRM-BIA 2221, CIRM-BIA 697, CRL1446, CYLB55, DSM 20052, EFEL6800, FAM 25337, FI11572, FI11600, FI11615, FOSU-YHD19, FTDC8312, FUA3588, GD181, GD752, GR1103, ICIS24, ike38, JX201, KA16MA, KMB_612, LfU21, LMT2-75, NKN-51, QAULFN21, QH221, R17-5, RI-508, SCB0035, SH272, SN4-1, SP331, TCI757, UBLF-31, USM 8633, WX91, XC416, YC74, YGRT89 and ATT-06.

**Pangenome and Comparative Analysis**

The completeness of over 350 *L. fermentum* genomes genome assemblies available to date in NCBI Genbank was assessed using BUSCO analysis with the bacteria_odb10 lineage dataset [29,30]. Genome assemblies with over 95% completeness were annotated using Prokka (version 1.14.6) [31]. Core- and pan genome analyses were performed by first annotating the genomes with Prokka [32] using the argument --kingdom Bacteria. The resulting Prokka output files were then processed with Roary (version 3.13.0) [33] using the arguments -e -n -v -r. Functional annotation of pan- and core genes was conducted using EggNOG-Mapper [34], with the COG database [35] (available at https://www.ncbi.nlm.nih.gov/COG/). The core and pan COGs were visualized using R and ggplot2 [36]. A SNP-based phylogenetic tree was constructed using TYGS and iTOL [37,38]. The ANI for each pair of genomes was determined through a Perl script that is part of the GET_HOMOLOGUES pipelines, and the resulting matrix was then clustered and visualized with the Pheatmap R package [39]. Carbohydrate-active enzyme (CAZyme)-related genes were identified using the CAZy database (v10) on the dbCAN server and HMMER (version 3.3.2) [40], following the recommended dbCAN protocol. *L. fermentum* strains were categorized based on the number of CAZymes they contained.

**CRISPR and Mobile Genetic Elements**

CRISPR elements, including repeats and spacers, were identified, aligned, and visualized using the CRISPRviz tool [41]. CRISPR types were confirmed using CRISPRCasFinder [18]. CRISPR

ARTICLE IN PRESS

loci types were further classified using CRISPRClassify [42]. Open reading frames (ORFs) within +/- 20 kb of each CRISPR array were predicted using Prodigal (v2.6.3) [43]. Annotation of Cas proteins in ORFs was performed using hmmsearch with curated HMM profiles [43], applying the criteria: HMM E-value <1e-5, HMM alignment coverage >70%, and bitscore >= 40. Partially predicted proteins from Prodigal [44] were excluded. Plasmids in *L. fermentum* genomes were identified using PLSDB (version 2021_06_23_v2) with default settings [45,46]. Prophages were detected using the Phage Search Tool Enhanced Release (PHASTEST) [47,48]. Potential bacteriocins and bacteriocin-expressing regions were identified using the BAGEL4 web tool [38].

## In silico Docking and MD Simulations

### Protein-Bacteriocin Complex Generation

Two rho proteins extracted from *S. aureus* NCTC 8325 and *E. coli* K-12 were considered as target proteins. The RHO-bacteriocin complexes were obtained using ColabFold [49]. The best ranked complex further analyzed with PRODIGY (PROtein binDIng enerGY prediction) server [50].

ARTICLE IN PRESS

## Experimental Validation

### Gas Formation Test

To confirm the obligatory heterofermentative lifestyle of *L. fermentum* ATT-06, Durham tube method [46] was utilized. The MRS broth was distributed into two 5 mL sterile test tubes, and the sterile Durham tubes were carefully inverted and placed in each test tube, ensuring they were completely submerged. Then, 100 µL of an overnight culture of the bacterial strain was inoculated into each MRS broth tube. The inoculated tubes were incubated at 32°C for 48 hours. After incubation, the Durham tubes were examined for gas bubbles indicating positive gas production.

### Gaba Analysis

ARTICLE IN PRESS

*Limosilactobacillus fermentum* ATT-06 strain was grown in MRS broth supplemented with 2.5 g/L dextrose and 10 g/L monosodium glutamate (MSG) at 32 °C for 48 hours to measure the GABA production ability. The strain, sub-cultured twice, was inoculated into MRS broth (containing 1% MSG and 0.25% dextrose) and incubated at 37 ℃ for 48 hours. The culture medium was centrifuged at 15000 rpm for 10 minutes, and the cells were removed. The cell-free supernatant (CFS) was then freeze-dried using a freeze-dryer. Subsequently, around 20 mg of each CFS, MSG, and GABA were placed into separate cylindrical glass NMR (nuclear magnetic resonance) tubes and dissolved in deuterium oxide. Each prepared sample was analyzed via proton nuclear magnetic resonance ($^1$H-NMR) using an NMR spectrometer (Bruker, Model 400, MA, USA) at 400 MHz for 5 minutes. To confirm GABA production, negative control containing sterile MRS medium was utilized. A commercially available pure GABA standard was used as a positive control, and the obtained spectra and quantitative analyses were verified by comparing them with a known GABA standard [51].

**Results**

Table S1 shows the genome statistics of *Limosilactobacillus fermentum* strains that were isolated from different sources such as Dairy and Fermented Milk Products, Fermented Vegetables and Plant-Based Foods, Grains and Cereals, Traditional Beverages and Substrates, Meat and Probiotic Foods, Miscellaneous Fermented Foods, and Gut or Fecal-Associated Strains. According to phosphoglucomutase-based phylogenetic clustering of all complete genomes available in NCBI GenBank, no clear relationship was seen across strains isolation sources and clade formation (Figure S1). It was interesting to note that lactic acid-fermented shalgam isolate *L. fermentum* ATT-06 was closely associated with those strains that were mostly isolated from gut and feces-related samples.

ARTICLE IN PRESS

Average genome size of 2.04 ± 0.13, GC(%) content of 51.73±0.54, CDS of 2011.75±148.42, and tRNA of 54.92±7.19 were achieved. The *L. fermentum* ATT-06 had a genome size of 1.95 Mbp, GC(%) content of 52.3 and CDS of 1974 which are in the range with the overall genome statistics achieved with the rest of the *L. fermentum* genomes evaluated.




**Figure 1. (A)** SNP (single nucleotide polymorphism) based phylogenetic clustering of 52 *L. fermentum* strains. The phylogenetic tree was built with 1000 bootstraps. Major clades are

ARTICLE IN PRESS

indicated where bootstrap support values greater than 0.7. The scale bar represents nucleotide substitutions per site. (B) PCoA visualization of Jaccard distances based on shared genes across 52 *L. fermentum* genomes screened. The color of each box indicates a unique isolation source.

Phylogenetic analysis utilizing core genome SNPs demonstrated the formation of three major clades (Figure 1A). The first clade from top down was composed of QH221 (i.e., feces), SN4-1 (i.e., yogurt), and SCB0035 (i.e., kefir grains). The second clade members were UBLF-31 (i.e., feces), 712 (i.e., fermented cassava), ike38 (i.e., Shima tofu), TCI757 (i.e., Tempeh), and GD752 (i.e., feces). The remaining strains including ATT-06 and the reference strain EFEL6800 were the members of the third major clade. Strain ATT-06 grouped within a unique subclade alongside strains AGA52, CYLBB55, XC416, and KA16MA. This close association, marked by short branch lengths, potentially suggests a robust evolutionary relationship and confirms that these strains occupy a similar ecological niche, such as those found in fermented plant-based foods. Notably, ATT-06 and AGA52 showed the most similar relationship within this clade, suggesting a significant level of genomic similarity that might arise from enduring selective pressures in comparable environments. In contrast, more distantly related strains like QH221 (i.e., feces isolate) and GR1103 (probiotic supplement food) were positioned in separate, deeper clades, perhaps reflecting taxa that are adapted to different ecological conditions. The placement of ATT-06 in the phylogenetic tree supports its identification within a lineage of lactic acid bacteria that are adapted for plant-based fermentation related ecologies. The PCoA plot of 52 *L. fermentum* genomes constructed based on Jaccard distance shows that there is no clear association between isolation source and PCoA coordination of individual strains (Figure 1B). For example, plant-associated strains were distributed across the negative and positive sides of both PCo1 and PCo2. Similarly, dairy-associated strains were located on both sides of PCo2 while mostly accumulated on the positive side of PCo1. The strain ATT-06 located on the

ARTICLE IN PRESS

negative or positive side of PCo1 and PCo2, respectively, clustered with digestive tract and fecal-associated strains as well as AGA52 that was isolated from a plant-based fermented food [52–54].

**Core and Pan-genome Analysis**

ARTICLE IN PRESS

ARTICLE IN PRESS

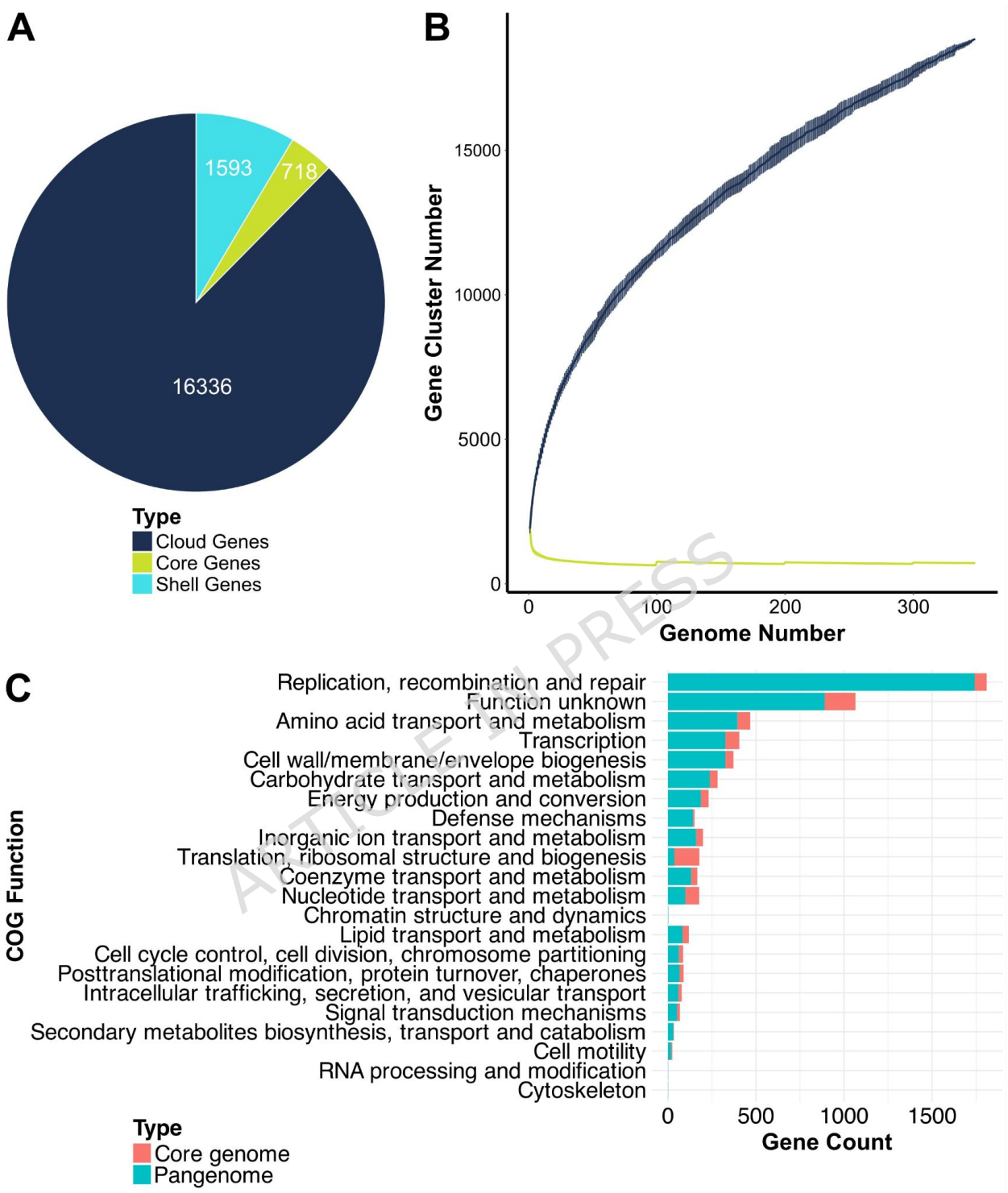



**Figure 2** (A) Distribution of the coding sequences across core and pangenome. (B) Number of core genes (light green curve) vs pan genes (dark blue curve). (C) Functional coding

ARTICLE IN PRESS

sequences in the core genome are shown in red bars, whereas functional coding sequences in the pan-genome are shown in turquoise bars across the genome of 52 *L. fermentum* strains

The genomic conservation of all *L. fermentum* genomes available in NCBI GenBank (i.e., 348 strains) was analyzed with regard to pan- and core genomes. It was found that only 3.85% of the total genes were conserved at 95% blastp identity cutoff. Among the total 18647 coding sequences, 718 of them were shared among 348 genomes that represent the core genome (Figure 2A). The cloud genes comprised 87.6% of the total CDS; however, shell genes represented 8.54% of total coding sequences suggesting the robust phenotypic variations across *B. lactis* isolates. To elaborate further, random subsampling was performed to create trend lines of each isolate's core genome and pan-genome (Figure 2B). Although the size of the core genome was close to a flat line, the pan-genome size never reached a plateau. Because the number of genomes continues to go up and novel genes are still in the process of discovery, the pan-genome stays open among *L. fermentum* strains. The trend of an open pan-genome was statistically validated using the Heap's Law alpha parameter ($\alpha < 0.95$).

ARTICLE IN PRESS

The bar chart produced by eggNOG-mapper illustrates the comparison of the distribution of COG (Clusters of Orthologous Groups) functional categories between the core genome (genes present in all strains) and the pangenome (the overall set of genes found across all strains, comprising both core and accessory genes) (Figure 2C). The core genome primarily consists of essential "housekeeping" functions necessary for fundamental survival, including translation, ribosomal structure and biogenesis, replication, recombination and repair, as well as energy production and conversion. On the other hand, pan-genome is enriched in accessory genes, including: defense mechanisms, secondary metabolites biosynthesis, and carbohydrate transport and metabolism. A remarkable number of coding sequences were assigned to function-unknown in both core and pan-genome categories.

ARTICLE IN PRESS

To quantitatively assess the similarity among the 52 *L. fermentum* genomes including ATT-06, the average nucleotide identity (ANI) was determined (Figure 3). The findings showed that all strains were classified as *L. fermentum*, supported by an ANI value exceeding 96% [55]. According to ANI-based phylogenetic relationships across 52 genomes analyzed, four major clades were identified. The first clade from top down was composed of the single member of RI-508 that was isolated from Cacao bean fermentation. The second major clade was composed of FUA3588, FI11600, FI11572, FI11615, the former was isolated from Mahewu whereas the latter three strains were isolated from pearl millet and its byproducts. The third clade was consisted of five strains (i.e., NKN-51, LMT2-75, YGRT89, 1019, 45756) that were isolated from Yak milk, kimchi, yogurt, Injera fermented dough and Pumiao rice noodle. The remaining forty-two strains including ATT-06 and reference genome EFEL6800 formed the fourth clade that were isolated from diverse ecological niches such as plant based fermented foods, dairy, feces, saliva and fermented meat (i.e., 8633). The closest hit to ATT-06 based on ANI value was achieved with AGA52 that was another plant based fermented food isolate.

ARTICLE IN PRESS

ARTICLE IN PRESS

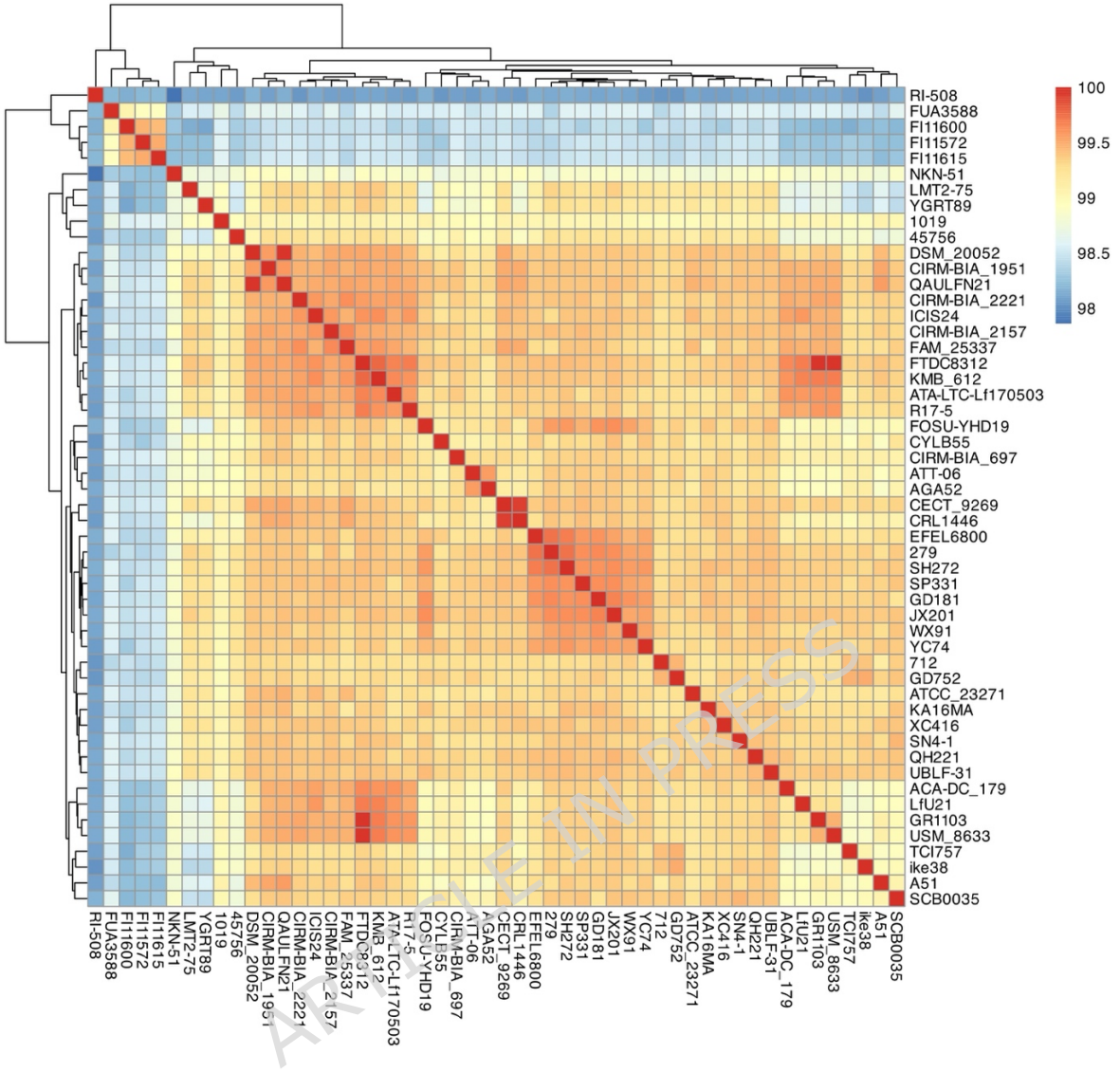



**Figure 3.** Heat map of average nucleotide identity of 52 *L. fermentum* genomes

**Carbohydrate Metabolism**

Three main clades were identified according to CAZy based phylogenetic clustering of 52 *L. fermentum* genomes according to the abundance of AA (Auxiliary Activities), CBM (Carbohydrate-Binding Modules), CE (Carbohydrate Esterases), GH (Glycoside Hydrolases), and GT (GlycosylTransferases) in each strain tested (Figure 4A). The first main clade from the very left of the heatmap was only made of TCI757 that carried the highest number of GH and GT. The second major clade consisted of twenty strains mainly low in GH and GT while

ARTICLE IN PRESS

moderately high in the remaining three CAZy family enzymes. The remaining thirty-one strains including ATT-06 and the reference genome EFEL 6800 formed the third major clade located on the far right hand of the heatmap which were moderately high in GH and GT while mediocre with AA, CBM, and CE. ATT-06 was clustered with three strains (i.e., AGA52, RI-508, and XC416) in the subclade which were isolated from plant based fermented food, Cacao bean fermentation and XC416.

Figure 4B shows the phylogenetic clustering of 52 *L. fermentum* genomes based on the abundance of CAZy subfamilies. Again, three main clades were identified with the first clade from top down being made of four strains namely TCI757, FUA3588, FI11572, and FI11615. Second clade members were YGRT89, A51, and KA16MA which were highest in GH25 content. The remaining forty-five strains including ATT-06 and reference EFEL 6800 formed the third major clade. ATT-06 formed a subclade within third major with ATA-LTC-Lf170503, NKN-51, and AGA52 which were isolated from corn silage, Yak milk, and plant based fermented food respectively. ATT-06 was mediocre in its GT2 and GT4 content, but it had moderately high GH73 content. The differences seen ATT-06 against its subclade members was evident, for example, its GT101, GT32, GH25, GT113 abundance was smaller to that of AGA52. On the other hand, ATT-06 was higher in GH73 and GT2 content compared to NKN-51 and smaller with GT4 against ATA-LTC-Lf170503. ATT-06 had higher number of GH8, GT14, and GT11 compared to its closest subclade member AGA52. Noteworthy, TCI757 and A51 carried the largest number of CBM50 while remaining genomes were poor in CBM50 content. SH272, 712 and EFEL6800 along with GD181, 279, SP331 and JX201 diverged from the rest of the genomes based on the level of their GH13. It's worth emphasizing that these strains were isolated from feces with the exception of EFEL6800. The highest level of CE4 or GH53 was seen in QH221 and CIRM-BIA 697, respectively. The gas production test of ATT-06 in MRS

ARTICLE IN PRESS

broth resulted in gas bubbles in Durham tubes confirming the heterofermentative lifestyle of this novel shalgam isolate [55].

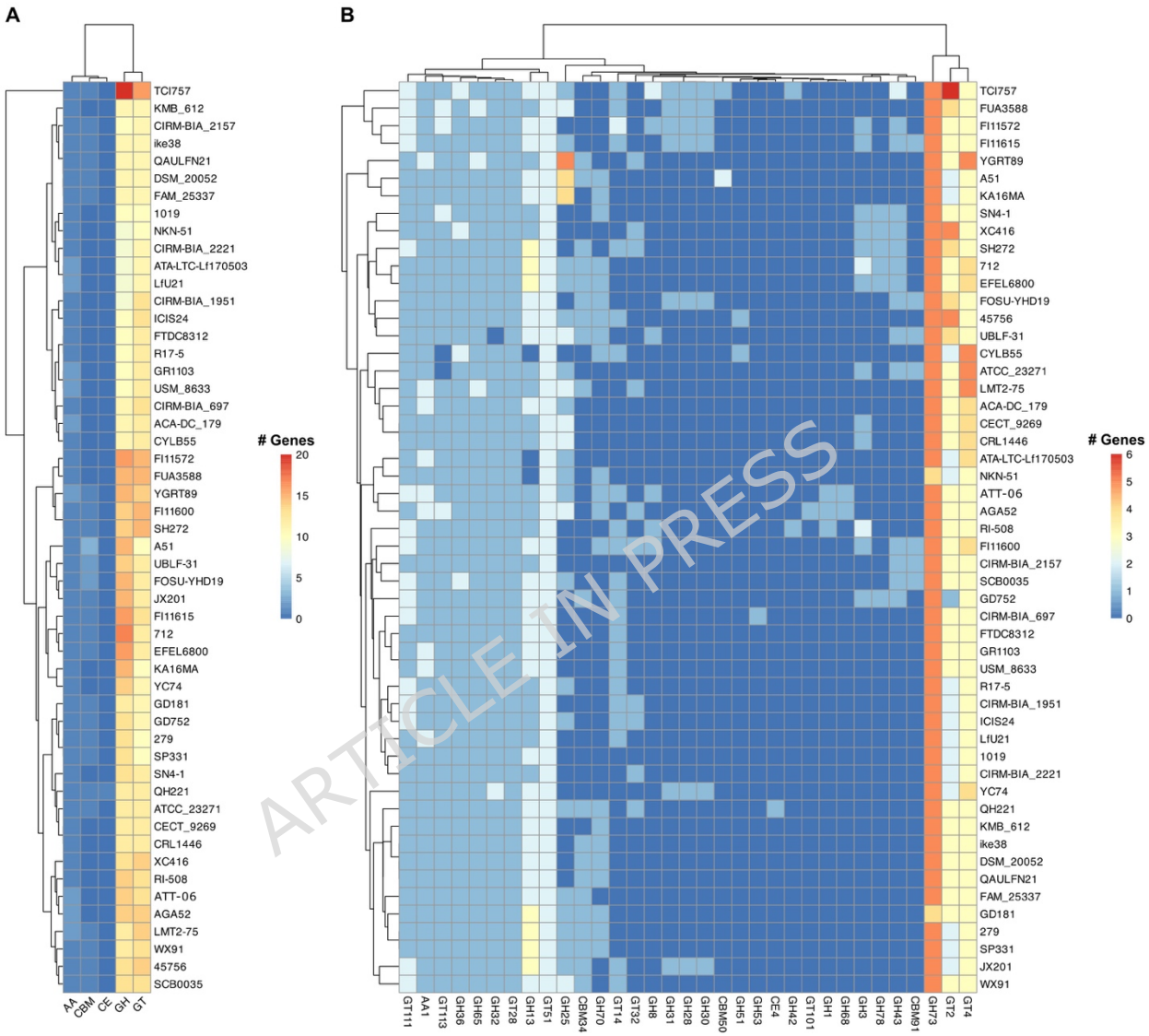



**Figure 4. (A)** Heatmap of CAZymes distribution and clustering across fifty-two *L. fermentum* genomes. The color gradient from blue to red represents the abundance of CAZymes found in each genome. GH: Glycoside hydrolase, GT: Glycosyltransferase, CE: Carbohydrate esterase, AA: Auxiliary activity, CBM: Carbohydrate binding module. **(B)** Heatmap of CAZymes subfamily distribution across fifty-two *L. fermentum* genomes.

ARTICLE IN PRESS

**Mobile Genetic Elements**

Table S2 shows intact prophages incorporated into bacterial genomes, specifying their genomic positions, and GC content. 38 out of 52 genomes were predicted to encode at least one intact prophage (Table S2). It was demonstrated that PHAGE_Lactob_LF1_NC_019486 (60.53%) and PHAGE_Lactob_LfeSau_NC_029068 (31.58%) are the most prevalent prophages found across 38 *L. fermentum* genomes screened. PHAGE_Coryne_Lederberg_NC_048790, PHAGE_Staphy_SPbeta_like_NC_029119 and PHAGE_Paenib_Shelly_NC_041909 were only detected in four, three and two *L. fermentum* strains, respectively. The remaining eight phages were unique to single strains. *L. fermentum* ATT-06 was predicted to encode a single prophage namely PHAGE_Lactob_LfeSau_NC_029068.

Ten *L. fermentum* strains were predicted to carry putative plasmids ranging from 1.3 kb to 75.9 kb. ATA-LTC-Lf170503, CYLB55, LMT2-75, and YGRT89 carried multiple plasmids in their genome (Table 1). While the reference genome EFEL6800 was found to harbor a putative plasmid at a size of 34 kb, Malaysian isolates of GR1103, LMT2-75, USM 8633, and YGRT89 were predicted to carry megaplasmid at a size of 75.9 kb. Noteworthy, LMT2-75 carried three plasmids with a lump sum of over 1 Mb.

ARTICLE IN PRESS

**CRISPR-Cas System**

Forty-six out of fifty-two genomes (88.5%) were predicted to carry at least one CRISPR-Cas system (Table S3). Type-IE is the most common (67.39%) CRISPR system, followed by Type-IIA (63.04%). Type-IIA co-occurs with both IC and IE subtypes, explaining its high prevalence. Noteworthy, Type-IIU is unique to JX201 (2.17%). Type-IIIA is rare (10.87%) and only observed in strains with IC or IE systems. No strain has both Type-IC and Type-IE suggesting mutual exclusivity of both types across *L. fermentum* genomes analyzed. For example, strains with

ARTICLE IN PRESS

TypeIC (e.g., NKN-51, A51) lacked Type-IE. Likewise, strains with Type-IE (e.g., CECT_9269, CRL1446) lacked TypeIC.

ATT-06 harbored a single Cas-Type-IIA similar to another plant based fermented food isolate AGA52. Five genomes (i.e., NKN-51, A51, ICIS24, 1019, SN4-1) carried putative Type I-C among which NKN-51, A51, and ICIS24 encoded Type II-A. It can also be deduced that Type-IIA frequently co-occurs with both Type-IC and Type-IE. Type-IIIA is present in some strains with Type-IC (e.g., NKN-51) and a few with Type-IE (e.g., CECT_9269).

ARTICLE IN PRESS

ARTICLE IN PRESS

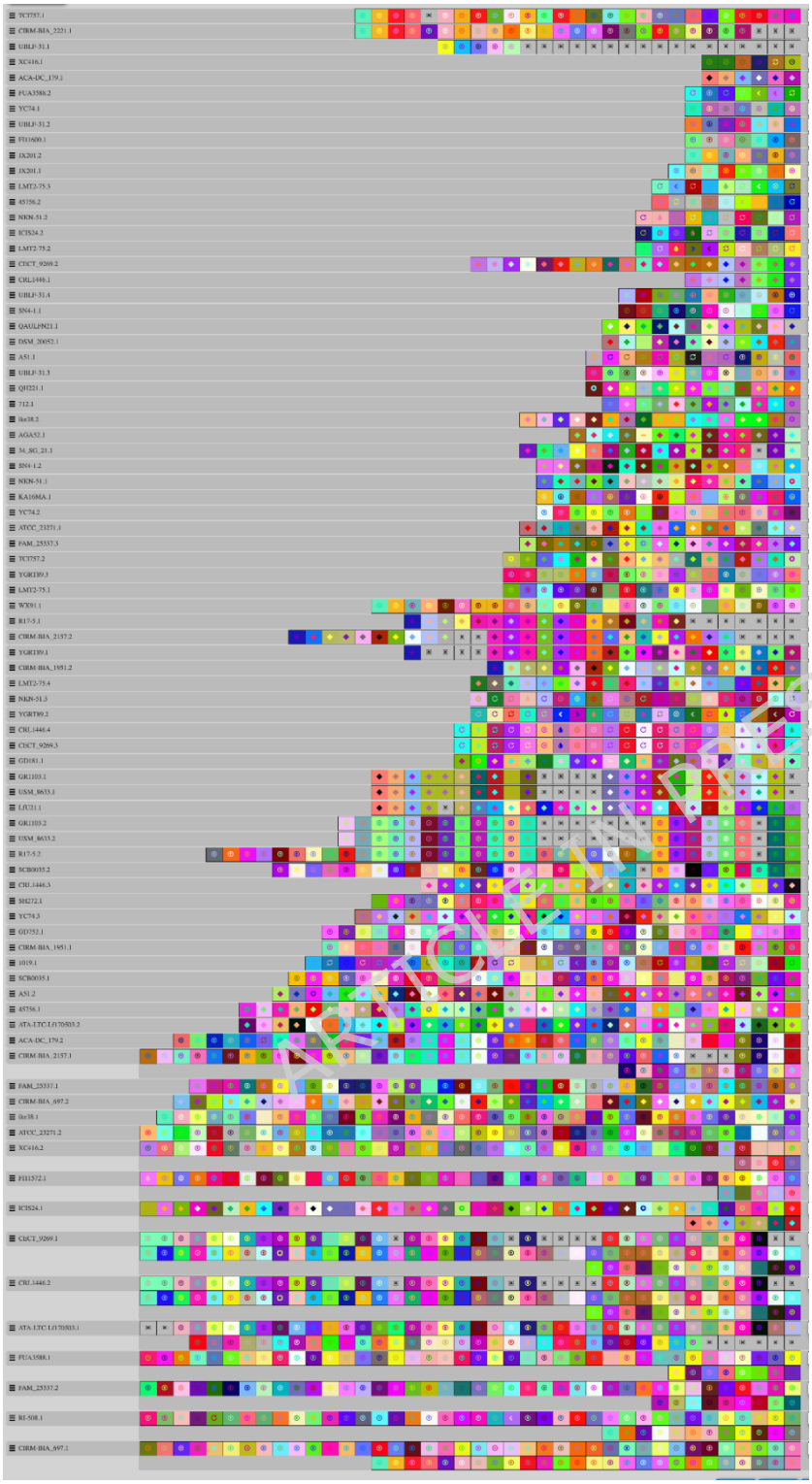


**Figure 5.** Alignment of spacers of each detected CRISPR locus. Each colored square represents a unique spacer in the CRISPR-Cas system. Grey “x” boxes showed missing spacer

ARTICLE IN PRESS

Crisprviz analysis revealed that only four strains had spacer identity across fifty-two isolates analyzed suggesting the presence of evolutionary heterogeneity [56] (Figure 5). CRISPR loci 4 of CRL1446 and CRISPR loci 3 of CECT 9269 had an identical spacer profile. In a parallel manner, CRISPR loci 1 and loci 2 of GR1103 and USM_8633 had an identical spare profile. Noteworthy, CRL1446 and CECT9269 share the same subclade in SNP based phylogenetic tree (Fig 1A). In parallel, GR1103 and USM_8633 are the subclade members.

**Table 1.** Putative plasmids detected in *L. fermentum* genomes analyzed

| Strain | Identity | NCBI Accession | Length (bp) | GC (%) |
|---|---|---|---|---|
| ATA-LTC-Lf170503 | 0.99 | NZ_AP012546.1 | 1326 | 0.43 |
| ATA-LTC-Lf170503 | 1.00 | NZ_CP103294.1 | 15916 | 0.40 |
| CYLB55 | 0.99 | NZ_AP012546.1 | 1326 | 0.43 |
| CYLB55 | 1.00 | NZ_CP120669.1 | 38056 | 0.43 |
| CYLB55 | 1.00 | NZ_CP120670.1 | 8672 | 0.37 |
| EFEL6800 | 1.00 | NZ_CP124738.1 | 34793 | 0.41 |
| GR1103 | 0.99 | CP033372.1 | 759694 | 0.50 |
| ike38 | 1.00 | NZ_AP024321.1 | 29015 | 0.42 |
| LfU21 | 1.00 | NZ_CP103294.1 | 15916 | 0.40 |
| LMT2-75 | 1.00 | CP033372.1 | 759694 | 0.50 |
| LMT2-75 | 1.00 | CP033373.1 | 251760 | 0.50 |

ARTICLE IN PRESS

| | | | | |
|---|---|---|---|---|
| LMT2-75 | 1.00 | NZ_CP034100.1 | 29243 | 0.41 |
| SH272 | 0.99 | NZ_AP012546.1 | 1326 | 0.43 |
| USM 8633 | 0.99 | CP033372.1 | 759694 | 0.50 |
| YGRT89 | 1.00 | CP033372.1 | 759694 | 0.50 |
| YGRT89 | 1.00 | CP033373.1 | 251760 | 0.50 |

**Gaba Biosynthesis**

The $^{1}$H-NMR spectra show the results obtained for the GABA (Figure S2), MSG (Figure S3), and supernatant sample (Figure S4). In the spectrum of the GABA standard, characteristic peaks were observed at chemical shifts of 2.95 ppm, 2.31 ppm, and 1.78 ppm. In the MSG standard, distinct peaks were detected between 3.65 ppm and 2.20-2.05 ppm. When the $^{1}$H-NMR spectrum of the supernatant was examined, the observation of signals belonging to GABA in the 2.95 ppm and 2.31 ppm regions confirmed the conversion of MSG to GABA by ATT-06. This suggests ATT-06 can produce GABA likely through the glutamic acid decarboxylase (GAD) enzyme system.

ARTICLE IN PRESS

**In silico Docking and MD Simulations**

Bacteriocin genes were identified in only one strain, ATA-LTC-Lf170503, encoding Acidocin A and Lactococcin. To evaluate their potential antimicrobial mechanism, we selected Rho transcription termination factor as a docking target due to its essential role in bacterial viability and conservation across Gram-positive and Gram-negative species. Acidocin A was 31 aminoacid long and had a sequence of *VIPGTLCRKQSLPIKQDLKILLFFLPRHSEQ* while Lactococcin was 46 aminoacid long and had a sequence of *MISSHQKTLTDKELALISGGKTHYPTNAWKSLWKGFWESLRYTDGF.*

**Formation of RHO - Bacteriocin Complexes**

ARTICLE IN PRESS

PROtein binDIng enerGY prediction (PRODIGY) server results are shown in Table 2. The table presents the predicted binding energies (ΔG in kcal/mol) for protein-bacteriocin complexes involving two proteins, *E. coli* RHO and *S. aureus* RHO, and two bacteriocins, Acidocin A and Lactococcin. All ΔG values are negative, indicating thermodynamically favorable binding interactions. Lactococcin demonstrates stronger binding affinities compared to Acidocin A for both proteins, with ΔG values of -8.1 kcal/mol (*E. coli* RHO) and -13.0 kcal/mol (*S. aureus* RHO), versus -5.5 kcal/mol and -7.9 kcal/mol for Acidocin A, respectively. Notably, *S. aureus* RHO exhibits significantly stronger binding affinities for both bacteriocins compared to *E. coli* RHO, particularly for Lactococcin, which shows the most stable interaction with a ΔG of -13.0 kcal/mol. These results suggest that Lactococcin may be more effective in targeting both *E. coli* and *S. aureus*, with a pronounced affinity for *S. aureus* RHO. The observed differences in binding affinities may reflect structural or functional variations in the protein binding sites.

ARTICLE IN PRESS

**Table 2. Protein - protein binding energies**

| **Protein-bacteriocin complex** | **PROtein binDIng enerGY prediction results ΔG (kcal/mol)** |
|---|---|
| *E. coli* RHO - Acidocin A | -5.5 |
| *E. coli* RHO - Lactococcin | -8.1 |
| *S. aureus* RHO - Acidocin A | -7.9 |
| *S. aureus* RHO - Lactococcin | -13.0 |

**Interface Analysis**

***E. coli* RHO - Acidocin A**

ARTICLE IN PRESS

The complex that was obtained from ColabFold, subjected the PRODIGY server to check residue-residue pairs that are in the interaction interface. Table S5-S8 shows the residues that reside in the interface.

Analysis of the residue-residue interactions between *E. coli* Rho protein (Chain A) and Acidocin A (Chain B) reveals that PHE24 in Acidocin A exhibits the highest number of contacts, interacting with multiple residues from Rho, including MET415, MET416, PHE412, PHE398, LYS402, SER419, ILE386, and MET390. This suggests that PHE24 may serve as a critical residue in stabilizing the interaction, likely through π-π stacking or hydrophobic interactions. Additionally, ILE20 and LEU21 in Acidocin A form multiple interactions, further supporting the presence of a hydrophobic core that facilitates protein-peptide association. The presence of polar and charged residue interactions, such as ASP17 of Acidocin A interacting with ILE386 and ILE382 of Rho, and GLN16 forming a contact with SER419, suggests potential hydrogen bonding or electrostatic contributions to the binding interface. The interaction between LYS19 of Acidocin A and SER419 of Rho further highlights the role of dipole interactions in stabilizing the complex. Several key residues in Rho, including SER419, ILE386, and PHE398, participate in multiple contacts, indicating their functional significance in mediating Acidocin A binding. Additionally, PRO26 in Acidocin A interacts with PHE398 and MET390 of Rho, which may contribute to local structural rigidity. Collectively, these findings suggest that the interaction between *E. coli* Rho protein and Acidocin A is primarily stabilized by hydrophobic interactions, with hydrogen bonding and electrostatic forces playing supplementary roles. Key residues such as PHE24, ILE20, and LEU21 in Acidocin A appear to serve as central contact points in this interaction, potentially playing a crucial role in its biological function.

ARTICLE IN PRESS

***E. coli* RHO - Lactococcin**

ARTICLE IN PRESS

Analysis of the residue-residue interactions between *E. coli* Rho protein (Chain A) and Lactococcin (Chain B), based on the above table, reveals several key contact points that likely contribute to the stability and function of the complex. Notably, PHE36 in Lactococcin exhibits multiple interactions with residues from Rho, including ILE382, LYS378, GLN378, LYS379, LEU383, and TRP381. This suggests that PHE36 plays a significant role in the binding interface, likely through hydrophobic interactions and potentially π−π stacking with TRP381. Similarly, TRP33 in Lactococcin engages in multiple contacts with LYS417, MET415, LEU383, MET416, and PHE398 from Rho, further supporting the involvement of hydrophobic and aromatic interactions in stabilizing the complex. TRP37 in Lactococcin also forms several interactions, particularly with TRP385, ILE386, LYS379, MET416, and ILE382 from Rho. The multiple Tryptophan residues at the interface (TRP37 and TRP33 in Lactococcin, and TRP381 and TRP385 in Rho) strongly indicate the importance of π−π interactions in the binding. TYR24 in Lactococcin interacts with PRO389, MET390, ALA395, GLY391, and ASP394 from Rho, suggesting a combination of hydrophobic and potentially hydrogen bonding interactions involving the hydroxyl group of tyrosine. TRP29 in Lactococcin also shows multiple contacts with SER419, PHE398, LYS402, and MET390 from Rho, reinforcing the role of hydrophobic and aromatic interactions at the interface.

Other residues in Lactococcin with multiple interactions include ILE17 (interacting with MET385, PRO386, PRO389) and LEU40 (interacting with ILE382, LYS379, GLN378), which contribute to the hydrophobic packing at the interface. The presence of charged residues like ARG41 (interacting with LYS417 and ARG418) and ASP44 (engaging with LYS417) suggests that electrostatic interactions may also play a role in complex formation, although the number of listed electrostatic interactions appears less extensive compared to hydrophobic ones based on

ARTICLE IN PRESS

this table. Furthermore, polar residues like THR43 in Lactococcin form contacts with GLN378 and ILE382 in Rho, indicating possible hydrogen bonding contributions.

Collectively, these findings suggest that the interaction between *E. coli* Rho protein and Lactococcin is primarily driven by a significant network of hydrophobic and π−π interactions, with electrostatic and hydrogen bonding forces likely playing complementary roles in defining the specificity and stability of the complex. The prominent interactions involving PHE36, TRP33, TRP37, TYR24, and TRP29 in Lactococcin, due to their multiple contact points, indicate that these residues are likely critical hotspots for the binding interface and thus for complex stability and function.

***S. aureus* RHO - Acidocin A**

The interaction analysis between *S. aureus* Rho (Chain A) and Acidocin A (Chain B) highlights key residue-residue contacts that may play a role in complex stability and function. Several charged interactions are observed, particularly between LYS15, LYS17, and LYS19 in Acidocin A and multiple residues in Rho, including ASP209, ALA210, GLU267, PRO208, ASP132, LYS211, and GLU45, suggesting a significant contribution from electrostatic forces. Similarly, ASP17 in Acidocin A interacts with LYS44 and LUE48 in Rho, while ASP132 in Rho also engages with GLN16, reinforcing the importance of charge-based interactions. Hydrophobic interactions are also prevalent, particularly with LEU18, ILE20, LEU21, and LEU22 in Acidocin A, which form multiple contacts with residues in Rho, including GLU64, PHE129, HIS133, LYS44, ASP132, THR109, LYS107, VAL108, MET63, TYR61, and MET55, suggesting that these residues play a role in stabilizing the complex through van der Waals forces and hydrophobic packing. Additionally, PHE23 and PHE24 in Acidocin A interact with several residues in Rho, including TYR61, TYR62, ALA49, GLU52, GLU56, and MET51, indicating the presence of π-π stacking and aromatic interactions, which may further enhance the stability of

ARTICLE IN PRESS

the interface. The interaction of PRO26 in Acidocin A with GLU56 in Rho suggests additional stabilization via hydrophobic and possible hydrogen bonding contributions. These observations suggest that the *S. aureus* Rho–Acidocin A complex is primarily stabilized by a combination of electrostatic interactions, hydrophobic contacts, and aromatic stacking, with key residues in Acidocin A—LYS15, ASP17, LEU18, ILE20, LEU21, PHE23, and PHE24—likely playing critical roles in complex formation and function.

### ***S. aureus*** **RHO - Lactococcin**

The interaction analysis between *S. aureus* Rho (Chain A) and Lactococcin (Chain B) reveals a network of residue-residue contacts that may contribute to the stability and function of the complex. Several electrostatic interactions are observed, particularly involving GLU52, GLU56, GLU64, GLU232, GLU235, and ASP132 in Rho, which form contacts with multiple residues in Lactococcin, including LYS7, LYS30, ARG41, and ASP44, suggesting the presence of charge-based stabilization. Additionally, GLU52, GLU56, and TYR62 in Rho interact with LEU9, LEU14, and LEU16 in Lactococcin, indicating a role for van der Waals forces in stabilizing these contacts. Hydrophobic interactions are also prominent, particularly with ILE2, ILE17, and ILE20 in Lactococcin, which interact with LYS34, LEU35, THR36, and MET63 in Rho, suggesting an important contribution from hydrophobic packing. Aromatic interactions are evident in contacts between PHE23, PHE36, TRP29, and TRP37 in Lactococcin and residues such as TYR61, TYR62, ALA49, and PHE129 in Rho, likely stabilizing the complex through π-π stacking and hydrophobic interactions. Furthermore, ASP44, THR43, and GLY45 in Lactococcin form contacts with GLU45, LEU41, and LYS40 in Rho, suggesting additional stabilization via hydrogen bonding and electrostatic contributions. The extensive contacts involving PHE46 in Lactococcin with multiple residues in Rho, including ALA30, LEU33, GLU45, TYR38, and LEU41, indicate its potential role as a critical residue in complex stabilization. Overall, the *S. aureus* Rho–Lactococcin complex appears to be stabilized by a combination of electrostatic

ARTICLE IN PRESS

interactions, hydrophobic contacts, and aromatic stacking, with key Lactococcin residues—LYS7, ILE17, PHE23, PHE36, TRP29, and PHE46—playing central roles in mediating the interaction.

THR26 and ASN27 from Lactococcin (Chain B) also play notable roles in the interaction with *S. aureus* Rho (Chain A), contributing to the stabilization of the complex through both hydrogen bonding and hydrophobic interactions. THR26 in Lactococcin forms contacts with multiple residues in Rho, including TYR83, ILE66, PHE213, ASN82, VAL81, VAL237, LYS107, ARG258, GLU235, VAL236, and GLU64, suggesting that it is a key interaction site involved in a mix of hydrogen bonding (via the hydroxyl group of THR) and van der Waals interactions. The presence of charged residues (LYS107, ARG258, GLU235, and GLU64) further indicates the potential for electrostatic contributions to the stabilization of this contact.

Similarly, ASN27 in Lactococcin interacts with THR80, VAL81, VAL237, HIS238, GLU235, TYR83, VAL236, ASN82, and GLU228 in Rho, highlighting its involvement in hydrogen bonding and polar interactions. The presence of HIS238 and GLU235 suggests possible polar interactions or charge-dipole interactions, while contacts with VAL81, VAL237, and VAL236 indicate some degree of hydrophobic stabilization. These interactions suggest that THR26 and ASN27 act as important mediators of the interaction, bridging electrostatic, hydrogen bonding, and hydrophobic contacts between the two proteins.

ARTICLE IN PRESS

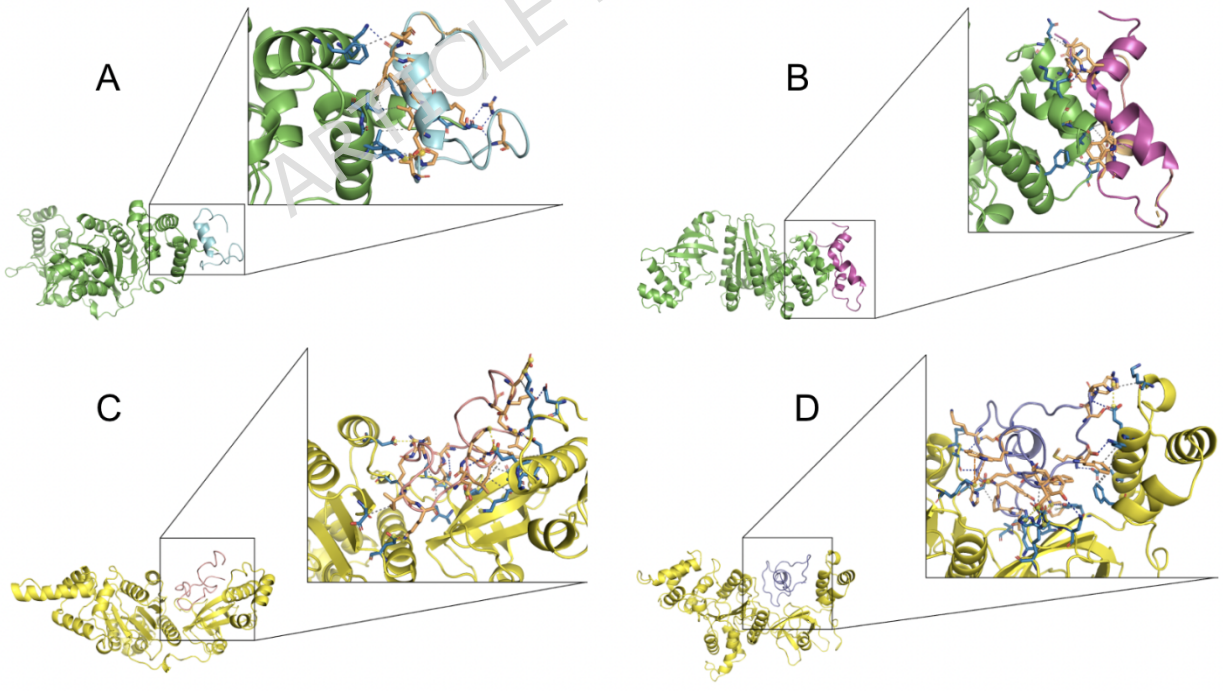
**Cluster Analyses**

To comprehensively analyze the conformational landscape sampled, the trajectories from three independent molecular dynamics (MD) simulations were first concatenated into a single, unified trajectory for all studied systems. This combined trajectory was then subjected to cluster analysis to identify representative structural ensembles. The Gromos clustering algorithm was employed for this purpose, utilizing a root-mean-square deviation (RMSD) based cut-off of 3.0 Å

ARTICLE IN PRESS

to group structurally similar conformations. For the EcoliRHO-Acidocin A complex, 25 clusters were obtained, where the first cluster represents 28.73% of the concatenated trajectory (2nd cluster: 27.44%, 3rd cluster: 18.19%, first three clusters cover ~75% of all frames). For the EcoliRHO-Lactococcin complex, 18 clusters were obtained, where the first cluster represents 31.29% of the concatenated trajectory (2nd cluster: 27.27%, 3rd cluster: 21.35%, first three clusters cover ~80% of all frames). For the SaureusRHO-Acidocin A complex, 23 clusters were obtained, where the first cluster represents 25.97% of the concatenated trajectory (2nd cluster: 22.88%, 3rd cluster : 20.25%, first three clusters cover ~70% of all frames). For the SaureusRHO-Lactococcin complex, 16 clusters were obtained, where the first cluster represents 24.14% of the concatenated trajectory (2nd cluster: 23.48%, 3rd cluster: 20.22%, first three clusters cover ~68% of all frames). Figure 6A shows the first cluster representative for the EcoliRHO-Acidocin A complex.

ARTICLE IN PRESS



ARTICLE IN PRESS

**Figure 6**. (A) EcoliRHO-Acidocin A interactions in the most representative structure, EcoliRHO shown in green, Acidocin A shown in cyan. (B) EcoliRHO-Lactococcin interactions in the most representative structure, EcoliRHO shown in green, Lactococcin shown in purple. (C) SaureusRHO-Acidocin A interactions in the most representative structure, SaureusRHO shown in yellow, Acidocin A shown in salmon. (D) SaureusRHO-Lactococcin interactions in the most representative structure, SaureusRHO shown in yellow, Lactococcin shown in slate.

In EcoliRHO-Acidocin A complex, 8 hydrophobic interactions were observed between ILE382-VAL1, ILE382-ILE20, ILE386-LEU21, ILE386-PHE24, ILE386-ILE20, PHE398-LEU25, MET415-PHE23, ARG418-ILE20, where first residues belong to EcoliRHO. 6 hydrogen bonds were identified between LYS379-PRO3, LYS402-PHE24, SER419-LYS19, SER419-ARG8 (3 bonds). LYS385-ASP17 salt bridge was also observed, together with ARG418-PHE23 pi-cation interaction. In summary, the structural analysis of the EcoliRHO-Acidocin A complex revealed a significant network of stabilizing interactions at the protein-peptide interface. Specifically, eight hydrophobic contacts, notably involving residues ILE382, ILE386, PHE398, and MET415 of EcoliRHO with various residues of Acidocin A, suggest a crucial role for nonpolar interactions in complex formation. Furthermore, the identification of six hydrogen bonds, particularly those involving LYS379, LYS402, and SER419 of EcoliRHO, along with a key LYS385-ASP17 salt bridge and an ARG418-PHE23 pi-cation interaction, highlights the contribution of polar and electrostatic forces to the stability and specificity of this interaction. These findings collectively underscore the multifaceted nature of the binding mechanism between EcoliRHO and Acidocin A, providing valuable insights into potential avenues for further investigation and the rational design of molecules targeting this interaction.

In EcoliRHO-Lactococcin complex, 9 hydrophobic interactions were observed between GLN374-ILE2, GLN378-PHE36, LYS379-PHE36, LYS385-ASP11, LYS385-LEU14, ILE386-ILE17, PRO389-ALA15, PHE389-ILE17, and MET416-TRP33 (Figure 6B). Two hydrogen bonds

ARTICLE IN PRESS

were observed between GLU369-SER3, and GLU414-TRP37. Comparative analysis with the EcoliRHO-Lactococcin complex reveals both shared and distinct features in its interaction profile. Here, nine hydrophobic interactions, involving residues such as GLN374, GLN378, LYS379, LYS385, ILE386, PRO389, PHE389, and MET416 of EcoliRHO with various Lactococcin residues, again emphasize the importance of nonpolar contacts in mediating binding. In contrast to the Acidocin A complex, only two hydrogen bonds were identified, formed between GLU369 and SER3, and GLU414 and TRP37 of the respective binding partners. These differences in the number and nature of polar interactions suggest that while hydrophobic forces play a dominant role in both complexes, the specific contributions of hydrogen bonding and electrostatic interactions may vary, potentially influencing the affinity and specificity of each peptide for EcoliRHO. Further studies exploring the structural dynamics and energetic contributions of these interactions will be crucial for a comprehensive understanding of EcoliRHO's diverse binding capabilities.

In SaureusRHO-Acidocin A complex, 12 hydrophobic interactions were observed between LYS40-GLU30, LYS44-LEU18, GLU45-LEU18, LEU48-LEU21, GLU52-LEU21, TYR62-LEU22, TYR62-VAL1, GLU64-LEU22, ILE66-LEU6, VAL81-LEU6, ASP209-ILE14, and GLU235-LYS9 (Figure 6C). Four hydrogen bonds were detected; ASN37-GLN31, GLU45-ASP17, GLU52-VAL1, and THR80-THR5. Seven salt bridges were identified between LYS40-GLU30, LYS44-ASP17, GLU45-LYS19, GLU52-ARG27, GLU228-ARG8, GLU235-LYS9, and GLU267-LYS15. The structural analysis of the SaureusRHO-Acidocin A complex reveals a dense network of stabilizing interactions at the protein-peptide interface. A significant number of hydrophobic interactions, totaling twelve and involving residues such as LYS40, LYS44, GLU45, LEU48, GLU52, TYR62, ILE66, VAL81, ASP209, and GLU235 of SaureusRHO with various Acidocin A residues, highlights the substantial contribution of nonpolar forces to the complex stability. Furthermore, the presence of four hydrogen bonds, formed by residues ASN37, GLU45,

ARTICLE IN PRESS

GLU52, and THR80 of SaureusRHO, alongside a remarkable seven salt bridges involving residues LYS40, LYS44, GLU45, GLU52, GLU228, GLU235, and GLU267 of SaureusRHO, underscores the critical role of electrostatic interactions in mediating high-affinity binding. These findings collectively suggest a robust and specific interaction mechanism between SaureusRHO and Acidocin A, where both hydrophobic packing and electrostatic attractions contribute significantly to the overall stability and potentially the functional relevance of this complex.

In SaureusRHO-Lactococcin complex, 7 hydrophobic interactions were observed between PHE14-PHE46, LYS40-GLN6, LYS44-PHE46, LYS107-THR26, ASP132-TRP37, GLU232-LYS21, GLU235-PRO25 (Figure 6D). Twelve hydrogen bonds were detected between LYS44-PHE46, GLU45-HIS5, GLU45-SER4, GLU52-MET1, TYR62-LYS30, GLU64-TYR24, GLU64-TRP29, VAL81-THR26, LYS107-PRO25, ASN134-TYR42, GLU232-LYS12, and ALA233-HIS23. Seven salt bridges were identified between GLU45-HIS5, GLU64-LYS30, ASP132-LYS30, HIS133-ASP44, LYS211-GLU38, and GLU232-LYS12. Between TYR62-TRP29 a pi-stacking interaction was observed, together between LYS211-TRP33 pi-cation interaction. The structural analysis of the SaureusRHO-Lactococcin complex reveals a complex interplay of diverse intermolecular forces. Hydrophobic interactions, while present with seven contacts involving residues such as PHE14, LYS40, LYS44, LYS107, ASP132, GLU232, and GLU235 of SaureusRHO, appear to be complemented by a significant contribution from polar interactions. A striking twelve hydrogen bonds were identified, engaging a wide array of residues including LYS44, GLU45, GLU52, TYR62, GLU64, VAL81, LYS107, ASN134, GLU232, and ALA233 of SaureusRHO. Furthermore, seven salt bridges, involving residues such as GLU45, GLU64, ASP132, HIS133, LYS211, and GLU232 of SaureusRHO, suggest a crucial role for electrostatic attractions in stabilizing the complex. The observation of a pi-stacking interaction between TYR62 and TRP29, along with a pi-cation interaction between LYS211 and TRP33, further underscores the multifaceted nature of the binding interface. These findings indicate that the

ARTICLE IN PRESS

recognition of Lactococcin by SaureusRHO is driven by a synergistic combination of hydrophobic, hydrogen bonding, salt bridge, and aromatic interactions, highlighting a potentially intricate mechanism for achieving binding specificity and stability.

A comparative analysis of the four RHO-peptide complexes reveals both conserved principles and distinct specificities in their interaction mechanisms. Across all systems, hydrophobic interactions play a significant role in mediating binding, although the number and specific residues involved vary, suggesting a fundamental contribution of nonpolar contacts to complex formation. However, the balance and nature of polar interactions exhibit more pronounced differences. The EcoliRHO-Acidocin A and EcoliRHO-Lactococcin complexes show a more modest number of hydrogen bonds and salt bridges compared to their SaureusRHO counterparts. Notably, SaureusRHO-Acidocin A complex stands out with the highest number of hydrophobic and salt bridge interactions, while the SaureusRHO-Lactococcin complex is characterized by a remarkable abundance of hydrogen bonds alongside aromatic interactions. These variations underscore the evolutionary adaptation of RHO proteins from different bacterial species to recognize and interact with diverse antimicrobial peptides. The differential reliance on specific interaction types – whether predominantly hydrophobic and electrostatic, or more heavily influenced by hydrogen bonding and aromatic stacking – likely reflects the unique structural features of each RHO protein and the specific sequence and conformation of their respective peptide ligands.

ARTICLE IN PRESS

**Protein RMSD**

Based on the provided root mean square deviation (RMSD) values across three independent simulation runs, all four systems—*E. coli*-Acidocin, *E. coli*-Lactococcin, *S. aureus*-Acidocin, and *S. aureus*-Lactococcin—demonstrate relatively low average RMSD values, ranging from 2.84 Å to 3.18 Å (Table 3, Figure S5-S8).

ARTICLE IN PRESS

**Table 3.** Protein backbone RMSD values for the RHO proteins

| | Run 1 (Å) | Run 2 (Å) | Run 3 (Å) | Average (Å) |
|---|---|---|---|---|
| Ecoli-Acidocin | $3.19 \pm 0.53$ | $2.21 \pm 0.43$ | $3.11 \pm 0.56$ | $2.84 \pm 0.55$ |
| Ecoli-Lactococcin | $2.53 \pm 0.48$ | $3.06 \pm 0.49$ | $3.01 \pm 0.65$ | $2.87 \pm 0.29$ |
| Saureus-Acidocin | $3.02 \pm 0.29$ | $3.47 \pm 0.42$ | $2.80 \pm 0.30$ | $3.10 \pm 0.34$ |
| Saureus-Lactococcin | $3.01 \pm 0.26$ | $3.22 \pm 0.35$ | $3.32 \pm 0.46$ | $3.18 \pm 0.15$ |

These values suggest that the overall structures of the systems remained relatively stable and close to their initial conformations throughout the simulations. A direct comparison of the average RMSD values indicates only minor differences between the systems, with *E. coli*-Acidocin showing the lowest average deviation (2.84 Å) and *S. aureus*-Lactococcin showing the highest (3.18 Å). While the averages are similar, examination of individual run data reveals some variability in the RMSD across the three independent simulations for each system, highlighting the importance of performing multiple runs to assess reproducibility. Despite this run-to-run variation, the consistently low average RMSD values across all systems suggest that the simulated complexes maintain a reasonable degree of structural integrity under the tested conditions. One-way ANOVA analysis performed on the independent simulation runs revealed no statistically significant differences in RMSD values among the four RHO-bacteriocin complexes ($p=0.597$). All systems exhibited comparable structural stability, with average RMSD values fluctuating within a narrow range of 2.84 Å to 3.18 Å. Consequently, these results indicate that neither Acidocin A nor Lactococcin induces a distinguishable structural destabilization in *E. coli* or *S. aureus* Rho proteins compared to one another.

ARTICLE IN PRESS

ARTICLE IN PRESS

**Bacteriocin RMSD**

Based on the provided root mean square deviation (RMSD) values across three independent simulation runs, the structural stability of the four systems varies. The Ecoli-Acidocin system exhibits the highest average RMSD of 10.02 Å, with considerable variability observed between individual runs (ranging from 8.22 Å to 12.52 Å) (Table 4).

**Table 4.** Bacteriocin backbone RMSD values

| | Run 1 (Å) | Run 2 (Å) | Run 3 (Å) | Average (Å) |
|---|---|---|---|---|
| *E. coli*-Acidocin | 8.22 ± 1.67 | 12.52 ± 1.38 | 9.34 ± 1.50 | 10.02 ± 2.23 |
| *E. coli*-Lactococcin | 4.90 ± 0.69 | 8.28 ± 1.09 | 5.04 ± 1.06 | 6.07 ± 1.91 |
| *S. aureus*-Acidocin | 6.02 ± 1.21 | 6.99 ± 0.81 | 6.47 ± 0.89 | 6.49 ± 0.48 |
| *S. aureus*-Lactococcin | 6.00 ± 0.92 | 6.62 ± 0.65 | 5.75 ± 0.87 | 6.13 ± 0.45 |

ARTICLE IN PRESS

This suggests that the Ecoli-Acidocin complex undergoes larger conformational changes or explores a wider range of structural space during the simulations compared to the other systems. In contrast, the Ecoli-Lactococcin system shows a notably lower average RMSD of 6.07 Å, with less variation among runs (4.90 Å to 8.28 Å), indicating relatively greater structural stability and closer adherence to its initial conformation. The Saureus-Acidocin and Saureus-Lactococcin systems display comparable average RMSD values of 6.49 Å and 6.13 Å, respectively. Their individual runs also show a tighter clustering of values compared to Ecoli-

ARTICLE IN PRESS

Acidocin, suggesting a moderate level of structural deviation and better reproducibility between independent simulations for these two systems. Overall, while all systems show some degree of structural fluctuation, the Ecoli-Acidocin complex appears to be the least stable structurally under these simulation conditions, while Ecoli-Lactococcin demonstrates the highest relative stability as reflected by the lowest average RMSD. One-way ANOVA analysis reveals a statistically significant difference in structural stability among the complexes ($p=0.0337$), primarily driven by the *E. coli* RHO-Acidocin A system which exhibits a markedly higher average RMSD (10.02±2.23 Å) compared to the other groups (~6 Å). Post-hoc tests confirm that Acidocin A significantly destabilizes *E. coli* RHO compared to Lactococcin ($p=0.049$), resulting in an approximately 4 Å increase in structural deviation. In contrast, no significant differences were observed between the *S. aureus* RHO complexes ($p=0.99$), indicating that the choice of bacteriocin does not affect the stability of the *S. aureus* protein. These findings suggest that Acidocin A exerts a selective and potent destabilizing effect specific to *E. coli* RHO that is not present in the other tested systems.

**MM/GBSA Analysis**

To elucidate the binding energetics and stability of the protein-bacteriocin complex, the Molecular Mechanics/Generalized Born Surface Area (MMGBSA) approach was employed. This method is widely utilized to estimate the binding free energies of biomolecular complexes by combining molecular mechanics energy calculations with continuum solvation models, typically applied to snapshots derived from molecular dynamics (MD) simulations. Table S4 shows the average MM/GBSA values for all studied systems in 3 replicates.

The binding free energies (ΔGbind) for the interactions between RHO proteins (from *E. coli* and *S. aureus*) and the bacteriocins (AcidocinA and Lactococcin) were estimated using the Molecular Mechanics/Generalized Born Surface Area (MMGBSA) approach. All four protein-

ARTICLE IN PRESS

bacteriocin complexes exhibited favorable binding energies, with average ΔGbind values ranging from -80.94 kcal/mol to -161.05 kcal/mol. Notably, SaureusRHO-Lactococcin complex displayed the strongest predicted binding affinity (ΔGbind = -161.05±11.83 kcal/mol), whereas the EcoliRHO-AcidocinA complex showed the weakest (ΔGbind = -80.94 kcal/mol). Comparative analysis of the average energies indicated that Lactococcin consistently bound more strongly than AcidocinA to both EcoliRHO (−129.00±33.94 vs. −80.94±4.50 kcal/mol) and SaureusRHO (-161.05±11.83 vs. −137.70±10.88 kcal/mol). These computational findings suggest differential binding specificities and potential variations in inhibitory potency among the bacteriocins, with Lactococcin indicated as a more potent binder overall and SaureusRHO appearing as a more susceptible molecular target for both bacteriocins investigated. Analysis of binding free energies (MMGBSA) reveals a statistically significant difference (p=0.005), identifying the *E. coli* RHO-Acidocin A complex as having the weakest affinity (−80.94 kcal/mol) compared to the significantly stronger binding seen in *S. aureus* complexes (−138 to −161 kcal/mol). Post-hoc tests confirm that Acidocin A binds significantly less effectively to *E. coli* RHO than to *S. aureus* RHO (p=0.026) or than Lactococcin does (p<0.01). This thermodynamic Instability aligns with the structural instability observed in RMSD analysis, suggesting Acidocin A is less effective against the *E. coli* RHO protein compared to others.

ARTICLE IN PRESS

**Discussion**

A comprehensive comparative evaluation of genomic and functional analysis of 52 *Limosilactobacillus fermentum* strains isolated from diverse ecological niches were explored, with a focus on the novel isolate ATT-06, isolated from lactic acid-fermented beverage, shalgam. We aimed to uncover the genomic diversity, ecological adaptability, GABA production

ARTICLE IN PRESS

capacity and test encoded bacteriocins' in silico efficacy against common Gram-negative and Gram-positive human pathogens. The genome sizes achieved were between 1.91 and 2.17 Mbp which is in alignment with the range of lactic acid bacteria. The G+C% of *L. fermentum* is also in alignment with low G+C% LAB. Whole genome analysis determined either single or multiple plasmid sequences in nine *L. fermentum* strains which is consistent with a single plasmid found in the reference genome EFEL 6800. It was proposed that lactic acid bacteria are highly specialized for their particular ecological environment, a notion that is bolstered by the presence of plasmids that can be rapidly acquired and exchanged during periods of swift environmental changes [57].

Comparative genome analysis of all *L. fermentum* genomes available in NCBI GenBank (i.e., 348 strains) showed an open pan-genome across *L. fermentum* species with a lump sum of 18647 coding sequences and relatively small core genome (i.e., 718). This together with the prophages found in *L. fermentum* possibly highlights the genomic plasticity proposing the adaptation capacity of the species across various ecologies allowing it to survive in fermented foods and host-associated niches [58]. Even though ATT-06 was isolated from shalgam which is a plant-based beverage, its clade members in both SNP-based phylogenetic tree and principle coordinate analysis were those isolated from fecal samples. This perhaps suggests potential horizontal gene transfer events or functional convergence which enable adaptation against gut ecosystems. Its close evolutionary relationship with AGA52, another isolate derived from plants-based environment, further supports that environmental selection pressures might form shared genomic traits among strains from related or even different origins [59].

CAZymes analysis showed that ATT-06 encodes a moderate number of GTs and GHs, in particular GH73 which might enable the effective utilization of polysaccharides in shalgam. This supports the ecological fitness of ATT-06 in shalgam microenvironment. Moreover, sub-clade

ARTICLE IN PRESS

members of ATT-06 were AGA52 and RI-508 that were isolated from plant-based ecologies which is an indication of emphasis against degradation of plant-derived substrates. Noteworthy, slight differences in abundances of the CAZy subfamily was seen for example higher levels of GT11, GT14 and GH8 found in ATT-06 as opposed to AGA52 perhaps imply metabolic divergence or niche-specific adaptation [59].

Mobile genetic elements of prophages and plasmids were explored to uncover the potential horizontal gene transfer and genomic plasticity across 52 *L. fermentum* genomes. We found that ATT-06 encodes a single prophage (PHAGE_Lactob_LfeSau_NC_029068) as opposed to other strains possessing large plasmids and megaplasmids. We speculate that lack of a plasmid and existence of a single prophage in ATT-06 might imply a relatively stable genome architecture on one hand but on the other hand it might be an indication of an absence of phage hits or existence of robust phage resistance mechanisms. This was supported by identification of a Type-IIA CRISPR-Cas system in ATT-06 which is one of the two most abundant CRISPR types among 52 analyzed genomes. The lack of Type-IC and Type-IE in ATT-06 along with other isolates underscore unique evolutionary paths in CRISPR immunity [60]. The co-existence of Type-IIA with both Type-IC and Type-IE in other strains may suggest a complex relationship between phage pressure and the diversification of CRISPR systems [61].

The high prevalence of prophage types of PHAGE_Lactob_LF1_NC_019486 (60.53%) and PHAGE_Lactob_LfeSau_NC_029068 (31.58%) likely reflects host-phage co-evolution in lactobacilli species. Strains that were found to carry multiple putative prophages might show enhanced genomic plasticity or niche adaptability. Prophages such as PHAGE_Staphy_SPbeta_like suggest cross-genus interactions or horizontal gene transfer. Strains with TypeIE (e.g., CECT_9269, CRL1446) often carry Lactob phages, perhaps suggesting CRISPR-Cas systems may target these invaders [62].

ARTICLE IN PRESS

As mentioned above, only ATA-LTC-Lf170503 among 52 *L. fermentum* genomes screened was predicted to carry bacteriocins namely Acidocin A and Lactococcin.

Recent *in silico* studies have highlighted the therapeutic potential of bacteriocins against resistant pathogens. Hasnat et al. identified Rho and HupB as conserved targets in clinical mastitis pathogens, demonstrating that peptides BP8 and BP32 form stable complexes that inhibit Rho by disrupting critical RNA and ADP binding sites [63]. Similarly, Halder and Mandal characterized a 13.1 kDa bacteriocin from *L. fermentum* LMEM22 (MIC 41–199 µg/ml), utilizing molecular docking to validate that its J32 peptide binds strongly to *P. aeruginosa* Exotoxin A (-70.35 kcal/mol) and *S. aureus* Sortase A (-49.90 kcal/mol) [64]. Khan et al. identified Enterocin B as a leading candidate against MRSA, exhibiting superior binding affinity (-7.5 kcal/mol) and stability with the resistance determinant PBP2a compared to other bacteriocins [65]. Collectively, these findings underscore the transformative role of computational modeling in accelerating the discovery of bacteriocin-based therapeutics as viable alternatives to conventional antibiotics. Building upon the established potential of Rho as a therapeutic target, the current study extends this research by conducting a comparative evaluation of Lactococcin and Acidocin A against the Rho proteins of human pathogens. The docking scores revealed that Lactococcin showed higher binding affinities compared to Acidocin A, for *S. aureus* and *E. coli*, implying potential antimicrobial capacity [66,67]. These findings could guide future strategies in peptide engineering focused on improving the specificity and efficacy of bacteriocins [68,69]. Stable bacteriocin–Rho interactions observed in silico suggest a non-canonical antimicrobial mechanism targeting bacterial transcription termination. Rho is an essential ATP-dependent helicase that mediates transcription termination via RNA binding and translocation, and its inhibition causes lethal transcriptional dysregulation [70]. Structurally exposed and conserved hydrophobic and charged regions involved in RNA recognition render Rho a plausible interaction partner for peptide-based antimicrobials. Docking and molecular dynamics analyses

ARTICLE IN PRESS

indicate that bacteriocin binding is stabilized predominantly by hydrophobic contacts, π–π stacking, and complementary electrostatic interactions, which are characteristic of class II bacteriocins [71,72]. Lactococcin exhibited more persistent interactions and lower inter-run RMSD variability than Acidocin, suggesting stronger anchoring to conserved Rho surface patches. Importantly, such local interaction stability may impair Rho dynamics without inducing major global conformational changes. Although Rho is cytoplasmic, bacteriocin-mediated membrane perturbation may enable intracellular access, allowing interference with transcriptional regulation that complements classical membrane-associated antimicrobial effects [70].

We also demonstrated the GABA biosynthesis capability of ATT-06 using $^{1}$H-NMR which implies the conversion of monosodium glutamate to GABA likely through the glutamate decarboxylase pathway [73]. The confirmed conversion of monosodium glutamate to GABA by ATT-06 demonstrates the activity of the glutamate decarboxylase pathway, a well-characterized acid resistance and metabolic system in lactic acid bacteria [74]. The GAD enzyme catalyzes the irreversible decarboxylation of L-glutamate, consuming a cytoplasmic proton in the process which contributes to intracellular pH homeostasis, a key survival mechanism in acidic environments [75] like lactic acid-fermented shalgam. Beyond its role in bacterial physiology, microbial GABA production which serves as a primary inhibitory neurotransmitter within the human central nervous system, the in situ microbial synthesis of GABA by a probiotic strain may hold significant therapeutic potential, especially in the context of stress-related or mood disorders [76,77]. The dual function of the GAD pathway—enhancing survival under acidic conditions while producing a neuroactive metabolite—supports the classification of ATT-06 as a potential psychobiotic candidate, although in vivo validation is required to substantiate functional relevance in the host context [78].

ARTICLE IN PRESS

Overall, ATT-06 stood out with its capability to biosynthesize GABA, carry CRISPR-based defence against phages, and inhabited a phylogenetic clade associated with plant- and gut-related strains, placing it as a robust candidate for probiotic development. However, even though in silico docking experiments put forth promising preliminary information regarding bacteriocin-Rho interactions, these insights are prone to further experimental validations via antimicrobial assays. To confirm psychobiotic potential of ATT-06, the efficacy of GABA biosynthesis should be explored under GI conditions and host model systems. While the current study establishes the genomic and in vitro functional profile of *L. fermentum* ATT-06, future translational and clinical studies will be required to determine an optimal route of administration. Given its origin from a fermented beverage and intended probiotic application, oral delivery is the most feasible route; subsequent research should focus on formulation, dosage, and survival in the gastrointestinal tract to validate its efficacy in vivo.

We acknowledge that while our genomic survey of 52 strains provides meaningful insights into *L. fermentum* diversity, it represents a curated subset of the species' pangenome. Furthermore, the promising probiotic and antimicrobial traits identified here—including GABA production and predicted bacteriocin activity—are derived from preclinical, in silico, and in vitro analyses. Consequently, their clinical relevance remains to be established through future translational studies involving formulation, in vivo models, and human trials. Finally, the molecular mechanisms proposed, such as GABA synthesis via the glutamate decarboxylase pathway and bacteriocin inhibition via Rho protein binding, require direct experimental validation through targeted gene editing, expression studies, and purified antimicrobial assays. Including this discussion contextualizes our findings and provides a clear roadmap for subsequent research.

**Conclusion**

This study establishes *Limosilactobacillus fermentum* ATT-06 as a novel strain of biotechnological interest through a comprehensive genomic and functional analysis. The

ARTICLE IN PRESS

primary novel contribution lies in its unique combination of functional traits: its ability to produce gamma-aminobutyric acid in vitro, its possession of a Type-IIA CRISPR-Cas system and specific prophage profile, and its genetic linkage to a potent bacteriocin (Lactococcin) with high predicted binding affinity against pathogenic *Staphylococcus aureus* and *Escherichia coli*. These attributes, coupled with its genomic adaptation to both plant-based and gut-associated environments, position ATT-06 as a multifaceted candidate for probiotic development and natural antimicrobial applications. Our findings are not without limitations. The antimicrobial potential of the identified bacteriocins is based on in silico predictions and requires experimental validation through antimicrobial assays. The genetic basis for GABA production and other adaptive traits, while strongly indicated, needs confirmation via targeted molecular studies. Furthermore, as a genomic and in vitro investigation, this study does not address translational parameters essential for probiotic application. Future research should therefore focus on: i) experimental validation of Lactococcin's antimicrobial activity and its mechanism of action; ii) molecular characterization of the GABA biosynthesis pathway in ATT-06; iii) formulation and dose optimization studies to ensure survival and functionality in the gastrointestinal tract; and iv) controlled pre-clinical and clinical trials to evaluate its safety, efficacy, and mental health benefits in vivo. Addressing these directions will be crucial to translate the promising genomic potential of *L. fermentum* ATT-06 into tangible health and biotechnological applications.

ARTICLE IN PRESS

**Data Availability**

Genomes studied in the present work are available with the following accession numbers from NCBI GenBank: GCA_002794275.1 (279), GCA_025186015.1 (712), GCA_025185835.1 (1019), GCA_018732245.1 (45756), GCA_030845215.1 (A51), GCA_019890875.1 (ACA-DC 179), GCA_021582915.1 (AGA52), GCA_037966785.1 (ATA-LTC-Lf170503), GCA_018398375.1 (ATCC 23271), GCA_900290185.1 (CECT 9269), GCA_025191825.1 (CIRM-BIA 1951), GCA_025191625.1 (CIRM-BIA 2157), GCA_025122475.1 (CIRM-BIA 2221),

ARTICLE IN PRESS

GCA_025191705.1 (CIRM-BIA 697), GCA_002899655.1 (CRL1446), GCA_029537335.1 (CYLB55), GCA_013394085.1 (DSM 20052), GCA_029961225.1 (EFEL6800), GCA_045527705.1 (FAM 25337), GCA_036431425.1 (FI11572), GCA_036431385.1 (FI11600), GCA_036431445.1 (FI11615), GCA_040790835.1(FOSU-YHD19), GCA_000417005.1 (FTDC8312), GCA_004683835.1 (FUA3588), GCA_037990855.1 (GD181), GCA_037990595.1 (GD752), GCA_017068315.1 (GR1103), GCA_040347825.1 (ICIS24), GCA_016861955.1 (ike38), GCA_037987985.1 (JX201), GCA_041380155.1 (KA16MA), GCA_003346325.1 (KMB_612), GCA_002869825.2 (LfU21), GCA_003855655.1 (LMT2-75), GCA_016918895.1 (NKN-51), GCA_022509425.1 (QAULFN21), GCA_037987775.1 (QH221), GCA_029872335.1 (R17-5), GCA_001982185.1 (RI-508), GCA_022819245.1 (SCB0035), GCA_037987505.1 (SH272), GCA_041413225.1 (SN4-1), GCA_037987235.1 (SP331), GCA_040267735.1 (TCI757) GCA_015160755.1 (UBLF-31), GCA_009676625.1 (USM 8633), GCA_037987015.1 (WX91), GCA_037986975.1 (XC416), GCA_037986705.1 (YC74), GCA_042434925.1 (YGRT89).

**Funding**

The study was supported by Scientific Research Commission of Istanbul Technical University with Project Number: MGA-2024-45355

**Acknowledgement**

The authors acknowledge the use of Generative AI for polishing the grammar of the manuscript.

**Declarations**

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare no competing interests.

ARTICLE IN PRESS

ARTICLE IN PRESS

**References**

1. Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* **44**, W16–W21 (2016).

2. Bourdichon, F. *et al.* Food fermentations: Microorganisms with technological beneficial use. *International Journal of Food Microbiology* **154**, 87–97 (2012).

3. Camargo, A. P. *et al.* IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Research* **51**, D733–D743 (2022).

4. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* **38**, 5825–5829 (2021).

5. Anwar, F. *et al.* Antiviral effects of probiotic metabolites on COVID-19. *Journal of Biomolecular Structure and Dynamics* **39**, 4175–4184 (2021).

6. Manna, S., Chowdhury, T., Chakraborty, R. & Mandal, S. M. Probiotics-Derived Peptides and Their Immunomodulatory Molecules Can Play a Preventive Role Against Viral Diseases Including COVID-19. *Probiotics and Antimicrobial Proteins* **13**, 611–623 (2021).

7. Behera, S. S., El Sheikha, A. F., Hammami, R. & Kumar, A. Traditionally fermented pickles: How the microbial diversity associated with their nutritional and health benefits? *Journal of Functional Foods* **70**, 103971 (2020).

8. Jiang, J. *et al.* Characterization of a novel bacteriocin produced by Lactobacillus sakei LSJ618 isolated from traditional Chinese fermented radish. *Food Control* **23**, 338–344 (2012).

9. Parada, J. L., Caron, C. R., Medeiros, A. B. P. & Soccol, C. R. Bacteriocins from lactic acid bacteria: purification, properties and use as biopreservatives. *Brazilian Archives of Biology and Technology* **50**, (2007).

ARTICLE IN PRESS

ARTICLE IN PRESS

10. Hwanhlem, N., Ivanova, T., Biscola, V., Choiset, Y. & Haertlé, T. Bacteriocin producing Enterococcus faecalis isolated from chicken gastrointestinal tract originating from Phitsanulok, Thailand: Isolation, screening, safety evaluation and probiotic properties. *Food Control* **78**, 187–195 (2017).

11. Alegría, Á., Delgado, S., Roces, C., López, B. & Mayo, B. Bacteriocins produced by wild Lactococcus lactis strains isolated from traditional, starter-free cheeses made of raw milk. *International Journal of Food Microbiology* **143**, 61–66 (2010).

12. Aziz, T., Hussain, N., Hameed, Z. & Lin, L. Elucidating the role of diet in maintaining gut health to reduce the risk of obesity, cardiovascular and other age-related inflammatory diseases: recent challenges and future recommendations. *Gut Microbes* **16**, 2297864 (2024).

13. Ahmad, W. *et al.* Lacticaseibacillusparacasei BNCC345679 revolutionizes DSS-induced colitis and modulates gut microbiota. *Frontiers in Microbiology* **Volume 15-2024**, (2024).

14. Wang, Y. *et al.* Probiotic Lactobacillus casei Zhang reduces pro-inflammatory cytokine production and hepatic inflammation in a rat model of acute liver failure. *European Journal of Nutrition* **55**, 821–831 (2016).

15. Aziz, T. *et al.* Exploring the multifaceted probiotic potential of Lactiplantibacillus plantarum NMGL2, investigating its antimicrobial resistance profiles and bacteriocin production. *Journal of Microbiological Methods* **236**, 107178 (2025).

16. Alkay, Z., Tuncil, Y. E., Durak, M. Z., Yilmaz, M. T. & Dertli, E. Enhancing γ-Aminobutyric Acid (GABA) in Boza, a Fermented Turkish Beverage: Role of Monosodium Glutamate and Lactobacillus Strains. *Journal of Food Processing and Preservation* **2025**, 7139222 (2025).

17. Ciciani, M. *et al.* Automated identification of sequence-tailored Cas9 proteins using massive metagenomic data. *Nature Communications* **13**, 6474 (2022).

18. Couvin, D. *et al.* CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Research* **46**, W246–W251 (2018).

ARTICLE IN PRESS

19. Akman, P. K., Ozulku, G., Tornuk, F. & Yetim, H. Potential probiotic lactic acid bacteria isolated from fermented gilaburu and shalgam beverages. *LWT* **149**, 111705 (2021).

20. De Vuyst, L. *et al.* Microbial ecology of sourdough fermentations: Diverse or uniform? *Food Microbiology* **37**, 11–29 (2014).

21. De Vuyst, L. & Weckx, S. The cocoa bean fermentation process: from ecosystem analysis to starter culture development. *Journal of Applied Microbiology* **121**, 5–17 (2016).

22. Dhakal, R., Bajpai, V. K. & Baek, K.-H. Production of gaba (γ - aminobutyric acid) by microorganisms: a review. *Brazilian Journal of Microbiology* **43**, (2012).

23. Dia, M., Wehner, T. C. & Arellano, C. RGxE: An R Program for Genotype x Environment Interaction Analysis. *AJPS* **08**, 1672–1698 (2017).

24. Ekinci, F. Y. *et al.* Characterization of chemical, biological, and antiproliferative properties of fermented black carrot juice, shalgam. *European Food Research and Technology* **242**, 1355–1368 (2016).

25. Heredia-Castro, P. Y. *et al.* Novel bacteriocins produced by Lactobacillus fermentum strains with bacteriostatic effects in milk against selected indicator microorganisms. *Journal of Dairy Science* **104**, 4033–4043 (2021).

26. Zhao, Y. *et al.* Lactobacillus fermentum and its potential immunomodulatory properties. *Journal of Functional Foods* **56**, 21–32 (2019).

27. Yetiman, A., Horzum, M., Bahar, D. & Akbulut, M. Assessment of Genomic and Metabolic Characteristics of Cholesterol-Reducing and GABA Producer Limosilactobacillus fermentum AGA52 Isolated from Lactic Acid Fermented Shalgam Based on “In Silico” and “In Vitro” Approaches. *Probiotics and Antimicrobial Proteins* **16**, 334–351 (2024).

28. Endo, A., Mizuno, H. & Okada, S. Monitoring the bacterial community during fermentation of sunki, an unsalted, fermented vegetable traditional to the Kiso area of Japan. *Letters in Applied Microbiology* **47**, 221–226 (2008).

ARTICLE IN PRESS

ARTICLE IN PRESS

29. Galata, V., Fehlmann, T., Backes, C. & Keller, A. PLSDB: a resource of complete bacterial plasmids. *Nucleic Acids Research* **47**, D195–D202 (2019).

30. Grissa, I., Vergnaud, G. & Pourcel, C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research* **35**, W52–W57 (2007).

31. Grover Sunita, Sharma Vineet K., Mallapa Rashmi H., & Batish Virender K. Draft Genome Sequence of Lactobacillus fermentum Lf1, an Indian Isolate of Human Gut Origin. *Genome Announcements* **1**, 10.1128/genomea.00883-13 (2013).

32. HAYWARD, A. C. Detection of gas production from glucose by heterofermentative lactic acid bacteria. *Microbiology (Reading, England)* vol. 16 9–15 (1957).

33. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

34. Jiménez Esther *et al.* Complete Genome Sequence of Lactobacillus fermentum CECT 5716, a Probiotic Strain Isolated from Human Milk. *Journal of Bacteriology* **192**, 4800–4800 (2010).

35. van Heel, A. J. *et al.* BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research* **46**, W278–W281 (2018).

36. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

37. Wishart, D. S. *et al.* PHASTEST: faster than PHASTER, better than PHAST. *Nucleic Acids Research* **51**, W443–W450 (2023).

38. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* **49**, W293–W296 (2021).

39. Kolde, R. *Pheatmap: Pretty Heatmaps*. https://CRAN.R-project.org/package=pheatmap (2019).

ARTICLE IN PRESS

ARTICLE IN PRESS

40. Manni, M., Berkeley, M. R., Seppey, M. & Zdobnov, E. M. BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols* **1**, (2021).

41. Martín, R. *et al.* Human milk is a source of lactic acid bacteria for the infant gut. *The Journal of Pediatrics* **143**, 754–758 (2003).

42. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nature Methods* **19**, 679–682 (2022).

43. Morita, H. *et al.* Comparative Genome Analysis of Lactobacillus reuteri and Lactobacillus fermentum Reveal a Genomic Island for Reuterin and Cobalamin Production. *DNA Research* **15**, 151–161 (2008).

44. Nethery, M. A. *et al.* CRISPRclassify: Repeat-Based Classification of CRISPR Loci. *The CRISPR Journal* **4**, 558–574 (2021).

45. Schmartz, G. P. *et al.* PLSDB: advancing a comprehensive database of bacterial plasmids. *Nucleic Acids Research* **50**, D273–D278 (2022).

46. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

47. Tatusov, R. L. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28**, 33–36 (2000).

48. Leroy, F. & De Vuyst, L. Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends in Food Science & Technology* **15**, 67–78 (2004).

49. Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **32**, 3676–3678 (2016).

50. Yetiman, A. E. & Ortakci, F. Genomic, probiotic, and metabolic potentials of Liquorilactobacillus nagelii AGA58, a novel bacteriocinogenic motile strain isolated from lactic acid-fermented shalgam. *Journal of Bioscience and Bioengineering* **135**, 34–43 (2023).

ARTICLE IN PRESS

ARTICLE IN PRESS

51. Wu, X. *et al.* Determination of air pollution-related biomarkers of exposure in urine of travellers between Germany and China using liquid chromatographic and liquid chromatographic-mass spectrometric methods: a pilot study. *Biomarkers* **22**, 525–536 (2017).

52. Zhang, H. *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research* **46**, W95–W101 (2018).

53. Zheng, J. *et al.* dbCAN-seq update: CAZyme gene clusters and substrates in microbiomes. *Nucleic Acids Research* **51**, D557–D563 (2023).

54. Li, Z. *et al.* Comparative genomics analysis of Pediococcus acidilactici species. *Journal of Microbiology* **59**, 573–583 (2021).

55. Ortakci, F., Broadbent, J. R., Oberg, C. J. & McMahon, D. J. Growth and gas formation by Lactobacillus wasatchensis, a novel obligatory heterofermentative nonstarter lactic acid bacterium, in Cheddar-style cheese made using a Streptococcus thermophilus starter. *Journal of Dairy Science* **98**, 7473–7482 (2015).

56. Gumustop, I. & Ortakci, F. Comparative genomics of Leuconostoc lactis strains isolated from human gastrointestinal system and fermented foods microbiomes. *BMC Genomic Data* **23**, 61 (2022).

57. Nethery, M. A., Henriksen, E. D., Daughtry, K. V., Johanningsmeier, S. D. & Barrangou, R. Comparative genomics of eight Lactobacillus buchneri strains isolated from food spoilage. *BMC Genomics* **20**, 902 (2019).

58. Medini, D., Donati, C., Tettelin, H., Masignani, V. & Rappuoli, R. The microbial pan-genome. *Current Opinion in Genetics & Development* **15**, 589–594 (2005).

59. Zhao, Y. *et al.* Environment-Related Genes Analysis of Limosilactobacillus fermentum Isolated from Food and Human Gut: Genetic Diversity and Adaption Evolution. *Foods* **11**, 3135 (2022).

ARTICLE IN PRESS

ARTICLE IN PRESS

60. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology* **13**, 722–736 (2015).

61. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology* **15**, 169–182 (2017).

62. Pei Zhangming *et al.* Comprehensive Scanning of Prophages in Lactobacillus: Distribution, Diversity, Antibiotic Resistance Genes, and Linkages with CRISPR-Cas Systems. *mSystems* **6**, 10.1128/msystems.01211-20 (2021).

63. Hasnat, S. *et al.* Natural Bacteriocins as Potential Drug Candidates Targeting Core Proteins in Mastitis Pathogens of Dairy Cattle. *bioRxiv* 2024.11.05.622021 (2024) doi:10.1101/2024.11.05.622021.

64. Halder, D. & Mandal, S. The in vitro studies combined with molecular docking and MM-GBSA binding free energy calculations reveal broad-spectrum antibacterial activity of bacteriocin from curd-derived Limosilactobacillus fermentum with probiotic attributes. *In Silico Research in Biomedicine* **1**, 100009 (2025).

65. A. Khan, Ishika, & J. G. Sharma. Computational Insights into Bacteriocin Activity Against MRSA: Exploration of Bacteriocin Dynamics in Countering Methicillin-Resistant Staphylococcus Aureus (MRSA) and Overcoming Beta-Lactamase Enzyme Defense Mechanisms in the Antibiotic Resistance Warfare. in *2024 International Conference on Big Data Analytics in Bioinformatics (DABCon)* 01–06 (2024). doi:10.1109/DABCon63472.2024.10919392.

66. De Vuyst, L. & Vandamme, E. J. *Bacteriocins of Lactic Acid Bacteria: Microbiology, Genetics and Applications*. (Springer US, 2012).

67. Erol, I., Kotil, S. E., Ortakci, F. & Durdagi, S. Exploring the binding capacity of lactic acid bacteria derived bacteriocins against RBD of SARS-CoV-2 Omicron variant by molecular simulations. *Journal of Biomolecular Structure and Dynamics* **41**, 10774–10784 (2023).

ARTICLE IN PRESS

68. Cotter, P. D., Hill, C. & Ross, R. P. Bacteriocins: developing innate immunity for food. *Nature Reviews Microbiology* **3**, 777–788 (2005).

69. Drider Djamel, Fimland Gunnar, Héchard Yann, McMullen Lynn M., & Prévost Hervé. The Continuing Story of Class IIa Bacteriocins. *Microbiology and Molecular Biology Reviews* **70**, 564–582 (2006).

70. Cardinale, C. J. *et al.* Termination Factor Rho and Its Cofactors NusA and NusG Silence Foreign DNA in E. coli. *Science* **320**, 935–938 (2008).

71. Cotter, P. D., Ross, R. P. & Hill, C. Bacteriocins — a viable alternative to antibiotics? *Nature Reviews Microbiology* **11**, 95–105 (2013).

72. Alvarez-Sieiro, P., Montalbán-López, M., Mu, D. & Kuipers, O. P. Bacteriocins of lactic acid bacteria: extending the family. *Applied Microbiology and Biotechnology* **100**, 2939–2951 (2016).

73. Komatsuzaki, N., Shima, J., Kawamoto, S., Momose, H. & Kimura, T. Production of γ-aminobutyric acid (GABA) by Lactobacillus paracasei isolated from traditional fermented foods. *Food Microbiology* **22**, 497–504 (2005).

74. De Biase, D. & Pennacchietti, E. Glutamate decarboxylase-dependent acid resistance in orally acquired bacteria: function, distribution and biomedical implications of the gadBC operon. *Molecular Microbiology* **86**, 770–786 (2012).

75. Feehily, C. & Karatzas, K. A. G. Role of glutamate metabolism in bacterial responses towards acid and other stresses. *Journal of Applied Microbiology* **114**, 11–24 (2013).

76. Braga, J. D., Thongngam, M. & Kumrungsee, T. Gamma-aminobutyric acid as a potential postbiotic mediator in the gut–brain axis. *npj Science of Food* **8**, 16 (2024).

77. Strandwitz, P. Neurotransmitter modulation by the gut microbiota. *Brain Research* **1693**, 128–133 (2018).

78. Dinan, T. G., Stanton, C. & Cryan, J. F. Psychobiotics: A Novel Class of Psychotropic. *Biological Psychiatry* **74**, 720–726 (2013).

ARTICLE IN PRESS