Computational and Structural Biotechnology Journal 21 (2023) 5111–5124

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Research article

Comparative genomics of *Loigolactobacillus coryniformis* with an emphasis on *L. coryniformis* strain FOL-19 isolated from cheese

Ismail Gumustop [a,1], Fatih Ortakci [b,*,2]

[a] BioEngineering Department, Faculty of Life and Natural Sciences, Abdullah Gul University, Kayseri, Turkey
[b] Food Engineering Department, Faculty of Chemical and Metallurgical Engineering, Istanbul Technical University, Istanbul, Turkey

ARTICLE INFO

Keywords:
Loigolactobacillus coryniformis
Comparative genomics
CRISPR/Cas
Reuterin
Cobalamin

ABSTRACT

Loigolactobacillus coryniformis is a member of lactic acid bacteria isolated from various ecological niches. We isolated a novel *L. coryniformis* strain FOL-19 from artisanal Tulum cheese and performed the whole-genome sequencing for FOL-19. Then, genomic characterization of FOL-19 against ten available whole genome sequences of the same species isolated from kimchi, silage, fermented meat, air of cowshed, dairy, and pheasant chyme was performed to uncover the genetic diversity and biotechnological potential of overall species. The average genome size of 2.93 ± 0.1 Mb, GC content of 42.96% ± 0.002, number of CDS of 2905 ± 165, number of tRNA of 56 ± 10, and number of CRISPR elements of 6.55 ± 1.83 was found. Both Type I and II Cas clusters were observed in *L. coryniformis*. No bacteriocin biosynthesis gene clusters were found. All strains harbored at least one plasmid except KCTC 3167. All strains were predicted to carry multiple IS elements. The most common origin of the IS elements was belong to *Lactiplantibacillus plantarum*. Comparative genomic analysis of *L. coryniformis* revealed hypervariability at the strain level and the presence of CRISPR/Cas suggests that *L. coryniformis* holds a promising potential for being a reservoir for new CRISPR-based tools. All *L. coryniformis* strains except PH-1 were predicted to harbor *pdu* and *cbi-cob-hem* gene clusters encoding industrially relevant traits of reuterin and cobalamin biosynthesis, respectively. These findings put a step forward for the genomic characterization of *L. coryniformis* strains for biotechnological applications via genome-guided strain selection to identify industrially relevant traits.

1. Introduction

Lactic acid bacteria (LAB), historically known as "milk-souring bacteria" were commonly associated with fermentation of food and feed [1]. LAB are comprised of Gram-positive, catalase-negative, non-sporulating, aerotolerant, and non-respiring species, which are either rod-shaped or cocci and able to synthesize lactate as the main product of fermentation [2]. Certain LAB species have probiotic effects in a wide range of spectra, such as potential prevention of allergies [3], improving feed conversion [4], and potential antidiabetic [5].

As of 2020, the genus Lactobacillus was segregated into 26 genera by conserved phenotypes and clade-specific signature genes [6]. The genera of Loigolactobacillus (L.) are known for the spoilage potential of fermented foods and drinks. Loigos means havoc, destruction, and ruin in Greek [6]. Loigolactobacillus are rod-shaped, non-motile, non-spore-forming, homofermentative bacteria that can synthesize L (+) and D (-) lactic acid isomers when fermenting D-mannose and D-mannitol. Loigolactobacillus species can be present in diverse environments such as beer fermentation, cheese, silage, air of dairy barns, rennet, and cabbage [6–9]. Currently, eight distinct species of Loigolactobacillus are reported in the National Center for Biotechnology Information (NCBI) database: *L. backii, L. bifermentans, L. binensis, L. coryniformis, L. iwatensis, L. jiayinensis, L. rennini,* and *L. zhaoyuanensis* [10]. *L. bifermentans* is associated with cracks in Dutch-style cheese by fermenting lactic acid into carbon dioxide, ethanol, acetic acid, and hydrogen. *L. rennini*, originating from the rennet, causes cheese spoilage [8]. *L. backii* (basonym: *Lactobacillus backii*) causes acidification and turbidity while spoiling beer fermentation. It is predicted that *L. backii* is responsible for up to 10% of spoiled beers manufactured between 2010 and 2013 in Germany [7].

* Corresponding author.
E-mail addresses: ismail.gumustop@agu.edu.tr (I. Gumustop), ortakci@itu.edu.tr (F. Ortakci).
[1] ORCID: 0000-0002-1452-1283
[2] ORCID: 0000-0003-1319-0854

https://doi.org/10.1016/j.csbj.2023.10.004
Received 14 August 2023; Received in revised form 4 October 2023; Accepted 4 October 2023
Available online 14 October 2023

L. coryniformis (*coryne* means club, *forma* means shape in Greek) is a coccoid rod-shaped LAB that requires biotin, riboflavin, p-aminobenzoic acid, niacin, and pantothenic acid to grow. Subspecies of *L. coryniformis* are present in diverse environments; for example, subsp. torquens were isolated from yak cheese, tomato pomace silage, and subsp. coryniformis was isolated from table olives, cheese, wheat, and pickled vegetable [6]. The Si3 strain of the subsp. coryniformis showed antifungal activity against spoilage yeast on silage [8]. The *L. coryniformis* K8 CECT 5711 strain is considered probiotic because it increases immunoglobulin G (IgG) levels in elderly COVID-19 patients and IgA levels in elderly people who did not get COVID-19 [11]. In addition, it was reported that consuming CECT 5711 could potentially provide clinical benefits against hepatitis A virus (HAV) infections by increasing total HAV antibody titers [12].

Several LAB strains are capable of producing compounds showing bioactivity in food systems also known as nutraceuticals [13]. Some of these micronutrients such as vitamin B_{12} are utilized as cofactor in several enzymatic reactions [14]. *Limosilactobacillus reuteri* and *Furfurilactobacillus rossiae* were described for the biosynthesis and genetic structure of *de novo* biosynthetic pathway of vitamin B_{12} [15–19]. *L. coryniformis* CRL 1001 was also reported to produce vitamin B_{12} [14] although the extent of this capability in other *L. coryniformis* strains remained to be clarified. *L. coryniformis* is one of the organisms that can produce broad-spectrum antimicrobial compounds [20,21] such as 3-HPA (3-hydroxypropionaldehyde) also referred to as reuterin [22]. Reuterin is biosynthesized as an intermediate during the transformation of glycerol to 1,3-propanediol [23]. The probiotic efficacy of *Limosilactobacillus reuteri* has been attributed to the production of reuterin which is an effective food protection antimicrobial agent [24,25]. The characteristic of reuterin biosynthesis by *L. reuteri* strains is the basis of their frequent utilization as probiotic adjunct cultures in commercial dairy foods applications [25].

Despite the interest in *L. coryniformis*, relatively few studies have been conducted for this species, and a mere ten *L. coryniformis* unique genomes are available in NCBI, originating from dairy, silage, kimchi, meat fermentation, air of cowshed, and pheasant chyme. (https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/3312/). As far as our knowledge goes, we did not come across any study comparing the genomes of *L. coryniformis* which highlights the limited knowledge of genomic diversity and biotechnological potential at the species level. We isolated a novel *L. coryniformis* strain FOL-19 from artisanal Tulum cheese manufactured in the Eastern region of Anatolia. In this study, we evaluated FOL-19 through comparative genomic analysis against ten distinct phenotypes to establish the diversity and *de novo* biosynthesis of vitamin B_{12} and reuterin capabilities.

2. Materials and methods

2.1. Bacterial isolation

The six-month-aged artisanal Tulum cheese sample, acquired from a local store in Eastern Anatolia, was homogenized in a 0.1% peptone (w/v) solution and then subjected to serial dilutions ranging from 10^{-1} to 10^{-7}. Each dilution was then plated on De Man, Rogosa, and Sharpe (MRS) agar (Condalab, Spain), which was prepared according to the manufacturer's instructions. The MRS agar plates were then incubated at 37 °C under anaerobic conditions for three days. After incubation, individual colonies were selected based on their colony morphologies, and selected bacterial colonies were streaked twice on MRS agar for purification and cryopreservation. This approach enabled us to obtain pure bacterial isolates from the cheese sample for further analysis.

2.2. DNA extraction and 16S rDNA sequencing

Cells grown in MRS broth under anaerobic conditions at 37 °C were collected to extract their DNA. During DNA extraction and purification, the standard protocol of PureLink™ Genomic DNA Mini Kit by Invitrogen™ was followed. Identification of bacterial isolate was conducted by 16S rDNA sequencing. The purified DNA samples were amplified with PCR using universal primers of 16S rDNA 27F (AGAGTTTGATCCTGGCTCAG) and 1492R (GGTTACCTTGTTACGACTT). The PCR reaction mixture contained three units of EasyTaq® DNA polymerase, 20 μM of 27F primer, 20 μM of 1492R primer, 3 μL of 10X EasyTaq® Buffer, 2.4 μL 2.5 mM dNTP, 23 μL nuclease-free water, and 50 ng of genomic DNA of the bacterial sample. The PCR reaction mix was amplified following conditions, including an initial denaturation of DNA at 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 52 °C for 20 s, extension at 72 °C for 90 s, and a final extension at 72 °C for 5 min. The amplified DNA fragments were visualized on 1.5% agarose gel and sent for Sanger sequencing [27]. The sequence of DNA fragments was analyzed using Basic Local Alignment Search Tool (BLAST) [26] in NCBI databases to identify the bacterial isolate accurately.

2.3. Whole-genome sequencing and assembly

After identifying the bacterial isolate as a novel *L. coryniformis*, the whole-genome sequencing of the novel *L. coryniformis* strain was performed with the Illumina NextSeq, next-generation sequencing technology. The paired reads of the draft genome were assembled in the PATRIC AutoAssembly pipeline [28]. Quality control measures were taken to ensure the accuracy of the whole-genome assembly, such as trimming reads and filtering out reads with a minimum length of 300 bp and a minimum read contig coverage of 5 by using Trim Galore [29]. These steps helped to improve the assembly's overall quality, ensuring that the genome sequence was accurate and reliable. The whole-genome sequence of the novel *L. coryniformis* strain was named FOL-19 and deposited in NCBI GenBank [30] with the accession number GCA_028439555.1.

2.4. Genome annotation

Whole-genome sequences of a total of ten fellow *L. coryniformis* strains both complete and draft were acquired from NCBI GenBank [30] with the following accession numbers of GCA_000166795.1 (KCTC 3167), GCA_000184285.2 (KCTC 3535), GCA_000283115.1 (CECT 5711), GCA_001742375.1 (CRL 1001), GCA_002706425.1 (DSM 20001), GCA_002706705.1 (DSM 20004), GCA_007954685.1 (CBA3616), GCA_019390135.1 (14I), GCA_019390175.1 (42L), and GCA_023483865.1 (PH-1).

Prokka software (version 1.14.6) [31] was utilized to annotate the whole-genome sequences of *L. coryniformis* strains, with the following flag: –kingdom Bacteria.

2.5. Comparative genomics of L. coryniformis

The assessment of genome similarity was carried out by computing Jaccard distance using the prabclus package [35] based on the presence or absence of putative genes. The resulting data were subjected to Principal Coordinate Analysis (PCoA) using the R (version 4.1.1) [36,37] to evaluate the relationship between genomes. The alignment of the core genome was performed by using FastTree (version 2.1.1) [38]. FastTree employed the Shimodaira-Hasegawa (SH) test with 1000 bootstrap replicates to calculate the reliability of each split according to a previous study [39]. The visualization of the phylogenetic tree of the core genome alignment was carried out with iTOL [40].

The genomes of *L. coryniformis* were subjected to cluster analysis and phylogenetic tree construction using the TYGS platform with default settings [41], which can be accessed at https://tygs.dsmz.de/. FastME [42] was used as the distance method for constructing the phylogenetic tree. The output of GFF files was then analyzed using Roary (version 3.13.0) [32] with flags "-e -n -v -r," which enabled us to analyze the pan-

and core genomes of the bacteria and compare the presence or absence of specific genes, including peptidases and aminotransferases. The minimum BLASTp identity threshold was set to 95% in Roary to ensure accuracy. To investigate whether the pangenome of *L. coryniformis* is open, the micropan package [33] utilized 10,000 permutations to fit Heap's law model. These methods enabled us to gain insights into the genetic makeup of *L. coryniformis* and the variability of its pangenome. The truncation of peptidases in *L. coryniformis* genomes were identified by comparing the gene lengths against complete peptidase genes exist in overall lactobacilli [34]. Core genome SNPs were detected with snippy tool [43]. Additionally, a phylogenetic tree of core genome SNPs was constructed using NCBI Genome Workbench [44] and iTOL [40]. Average nucleotide identity (CDS ANI) was calculated using the GET_HOMOLOGUES [45]. Genomic islands were identified using GIPSy [46] by feeding GenBank annotation files from Prokka. The genomes were aligned and visualized using the BLAST Ring Image Generator (BRIG) software with DSM 20001 as the reference genome, using the BLASTn algorithm with a lower identity threshold of 70% and a higher identity threshold of 90% [26,47].

The CRISPRCasFinder web tool located at https://crisprcas.i2bc.paris-saclay.fr/ [48] was used to identify Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) elements and Cas enzyme clusters in the genomes of *L. coryniformis*. Detection of CRISPR spacer and repeat sequences and their aligments were conducted by using CRISPRviz [86]. The dbCAN2 meta server was utilized to annotate Carbohydrate Active Enzymes (CAZyme) encoded by *L. coryniformis* strains. The Database for CAZy annotation (v11) was obtained, and the HMMER (version 3.1b2) [49] was employed to annotate CAZyme domains. The results of CAZyme annotation were filtered according to default thresholds of coverage and e-value scores by dbCAN2. Subsequently, CAZyme families were employed to classify the *L. coryniformis* strains.

Antimicrobial resistance genes were identified in *L. coryniformis* strains by screening with the Comprehensive Antibiotic Resistance Database (CARD) [50]. To detect potential bacteriocin-encoding genes, the BAGEL4 web server was employed. The detected bacteriocin sequences were subsequently validated using the BLASTp utility of the NCBI [51]. Plasmid sequences in whole-genome sequences of the *L. coryniformis* strains were identified using the PLSDB web tool [52,53]. Furthermore, insertion sequences (IS) in the genomes were detected by utilizing the ISfinder web tool, available at https://isfinder.biotoul.fr/ [54].

3. Results

3.1. Complete genome sequence of L. coryniformis FOL-19

The complete genome sequence of *Loigolactobacillus coryniformis* FOL-19 was assembled into a single contig (2.82 Mb) composed of 238 contigs. Genomic features of *L. coryniformis* FOL-19, such as genome size (2.82 Mb) and GC content (42.83%), were summarized in Table 1. We also identified two plasmids with a length of 0.809 and 1.326 Kb (Table S1). Annotation of the genome using Prokka yielded 2769 coding sequences, 52 tRNA, and 2 rRNA. Moreover, 3 unique CRISPR repeat and spacer content were detected among which type II-A was represented. All *L.coryniformis* strains were predicted to carry IS elements with PH-1, CECT 5711, and FOL-19 carrying the highest number of IS elements whereas KCTC 3167, 14I and 42L possessing the least number of those IS elements (Table S2). Comprehensive Antibiotic Resistance Database (CARD) predicted no perfect hit for any of *L. coryniformis* genomes analyzed although strict hits with lower percent identity of matching region were found in each genome (data not shown).

3.2. Genetic diversity of L. coryniformis

The genomic features of eleven *L. coryniformis* strains with genome sizes ranging between 2.82 Mb and 3.14 Mb (average 2.93 Mb) were annotated by identifying putative protein sequences, tRNA, and CRISPR loci. The number of contigs observed for each strain varies from 1 to 1486, while GC content falls within the range of 42.79–43.51% (average 42.96%). The number of tRNA genes identified in each strain falls between 32 and 70, while the number of protein-coding sequences ranges between 2735 and 3299 (average 2905) (Table 1).

Having access to complete genome sequences of *L. coryniformis* FOL-19 and type strain DSM 20001, we determined how FOL-19 compares to other *L. coryniformis* strains and performed comparative genomics. Eleven strains, including FOL-19, were chosen for comparative analysis using the core genome SNPs (Fig. 1A), CDS ANI (Fig. 1B), and whole-genome sequence (Fig. S1). According to the phylogenetic tree based on core genome SNPs, *L. coryniformis* strains split into five main clades. CBA3616 and CECT 5711 formed the first two clades, respectively. Strains isolated from silage (CRL 1001 and DSM 20001) and kimchi (KCTC 3167) shared the third clade. Fermented meat isolates (14I and 42L) formed the fourth clade. The remaining four strains including FOL-19 formed the last clade. ANI results, according to CDS, show that the highest identity was achieved at 99.88% between KCTC 3167 and DSM 20001 isolated from kimchi and silage, respectively. Likewise, KCTC 3167 showed the second-highest similarity with another silage isolate of CRL 1001. FOL-19 showed the highest ANI score similarity (97.97%) with another cheese isolate of CECT 5711. Moreover, fermented meat isolates of 14I and 42L revealed the highest similarity.

The whole-genome sequence-based phylogenetic tree revealed that FOL-19 separated from other *L. coryniformis* strains which clustered into three clades. The first clade members are DSM 20004, KCTC 3535, and PH-1. Members of the second clade (14I and 42L) were isolated from fermented meat. The last clade was separated into two subclades. The first subclade consisted of strains of KCTC 3167, DSM 20001, and CRL 1001 that were isolated from plant-originated niches such as silage and kimchi, and the second subclade was formed by CECT 5711 and CBA3616.

BRIG image shows the alignment of ten *L. coryniformis* strains and their GC content and GC skews against the reference genome DSM

Table 1
Whole-genome sequence statistics of eleven *L. coryniformis* strains.

Assembly Accession	Strain	Origin	Sequencing Technology	Genome Size (Mb)	Contigs	Coverage	GC (%)	CDS	tRNA	CRISPR
GCA_000166795.1	KCTC 3167	Kimchi	454	2.96	55	20.7x	42.80%	2735	32	4
GCA_000184285.2	KCTC 3535	Kimchi	454 GS Titanium	2.82	433	17.0x	42.87%	2917	51	8
GCA_000283115.1	CECT 5711	Cheese	454	2.84	203	25.0x	42.83%	2832	54	5
GCA_001742375.1	CRL 1001	Silage	IonTorrent	2.83	133	40x	42.94%	3298	70	7
GCA_002706425.1	DSM 20001	Silage	PacBio	2.95	1	272.5x	43.12%	2805	64	4
GCA_002706705.1	DSM 20004	Air of cow shed	PacBio	3.14	5	NA	43.11%	3009	64	8
GCA_007954685.1	CBA3616	Kimchi	PacBio RSII	3	1	193.9x	42.96%	2883	64	7
GCA_019390135.1	14I	Fermented meat	Illumina MiSeq	2.9	195	105.0x	42.79%	2793	53	9
GCA_019390175.1	42L	Fermented meat	Illumina MiSeq	2.89	212	71.0x	42.82%	2792	53	8
GCA_023483865.1	PH-1	Pheasant chyme	Illumina MiSeq	3.04	1486	23.0x	43.51%	3120	58	4
GCA_028439555.1	FOL-19	Cheese	Illumina NextSeq	2.82	238	300.0x	42.83%	2769	52	8

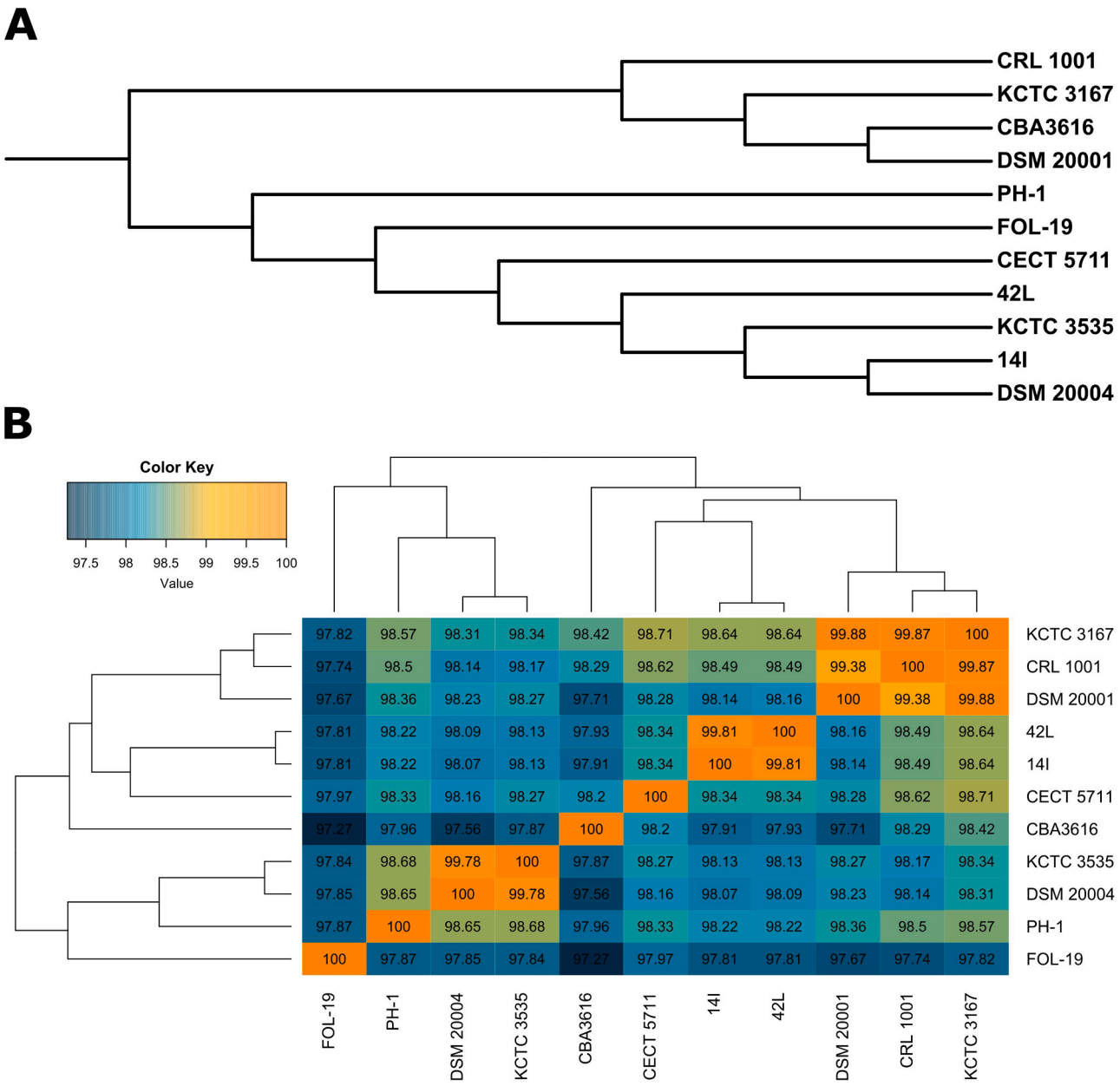

Fig. 1. Core genome SNPs based phylogenetic tree (A) and average nucleotide identity (B) of eleven *L. coryniformis* strains.

20001. The genome of the CRL 1001 shows the highest identity with the reference genome. Multiple regions lacking in genomes were labeled as putative resistance island 1 between 1.64 and 1.68 Mb, putative pathogenicity island 3 between 1.96 Mb and 2.00 Mb, and a group of genomic islands of resistance island 2, pathogenicity islands 2 and 4 between 2.44 and 2.55 Mb (Fig. 2). The presence of putative prophages was screened and it was found that all *L. coryniformis* genomes with the exception of CRL 1001, DSM 20001, and KCTC 3167 encoded at least one intact prophage. FOL-19 had one intact prophage in its genome. The prophage regions are represented with a lower GC content as shown in the BRIG image (Fig. 2).

3.3. Pan- and core genome analysis

For characterization of genomic conservation between all isolates, overall coding potential (i.e., pangenome) was determined, and it was observed that about 22% of all genes conserved within 95% BLASTP identity (Fig. 3 A). Of the 7062 total CDS, 1554 were shared by all eleven strains, which is the core genome. The accessory genome contained a total of 5508 CDS which perhaps determines fundamental differences in phenotypic traits across different strains as reported by [55]. We performed randomized subsampling to the strain order to visualize the trendlines of core- and pangenomes (Fig. 3B). The pangenome size didn't reach a plateau at eleven strains; however, the core genome appeared to reach a plateau at eleven strains. Sequencing additional new strains would increase the orthologous gene clusters. Alpha value was calculated as one according to Heap's law; therefore genome of the *L. coryniformis* can be considered open [56]. In addition, a total of 5508 variable COGs were determined, 2747 of which were characterized as unique (Fig. 3 C). Across all *L. coryniformis* strains screened, PH-1 harbored the highest number of unique genes of 661. In the following, CRL 1001 and CBA3616 harbored unique genes of 490 and 376, respectively. FOL-19 had 303 unique genes. On the other hand, 42L and 14I had the lowest number of unique genes across all *L. coryniformis* strains.

PCoA across genomes based on Jaccard distance among presence/

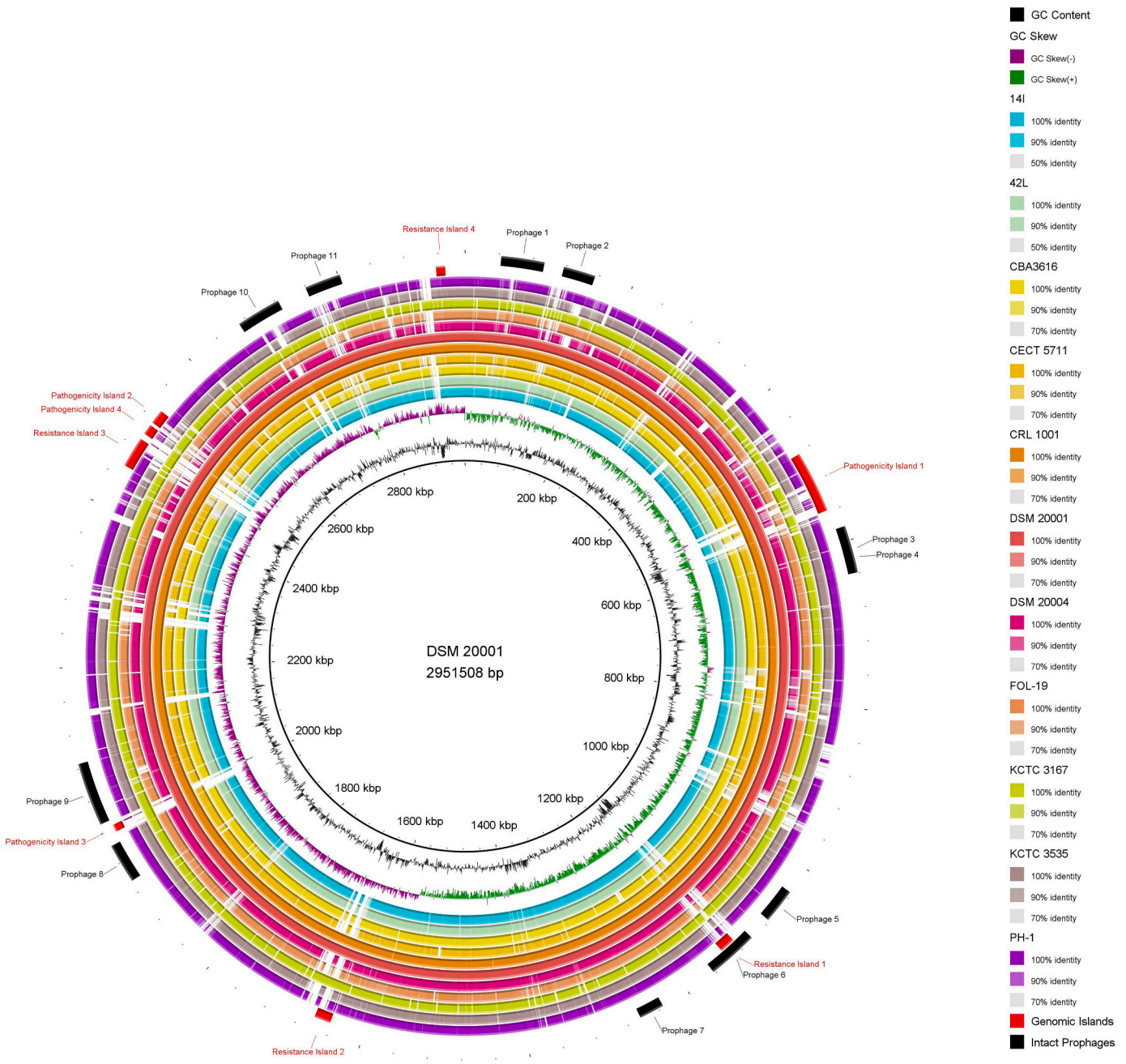

Fig. 2. Whole-genome-based BLAST comparison of ten *L. coryniformis* strains against reference strain DSM 20001. The innermost rings show GC Content (black) and GC Skew (purple-green). The remaining circles show BLAST comparisons of ten other complete *L. coryniformis* genomes against the reference genome DSM 20001. The outermost rings highlight genomic islands.

absence of genes showed that five strains (CRL 1001, DSM 20001, KCTC 3167, CBA3616, CECT 5711) were located at the negative values of PCo1, among which two lay at negative values of PCo2. However, half of the remaining six strains (PH-1, KCTC 3535, DSM 2004) lay at positive values of both PCo1 and PCo2. 14I and 42L strains were closer to each other compared to the remaining strains, which were relatively dispersed. Still, members of each pair of strains from the same isolation source were near to each other, except kimchi isolates of KCTC 3167 and CBA3616. KCTC 3535 showed the highest similarity against DSM 20004 isolated from air of cow shed (Fig. 4 A). The phylogenetic neighborhood across eleven strains was calculated based on relative hierarchical clustering via core genome alignment (Fig. 4B). A parallel trend of similarity with PCoA was observed in the unrooted phylogenetic tree. PH-1 and FOL-19 were separated from other strains and formed their own clades.

3.3.1. *Analysis of carbohydrate active enzymes*

Heatmap representation of CAZyme families shows that GH and GT family CAZymes are the most abundant across eleven *L. coryniformis* genomes. Moreover, FOL-19 harbors the greatest number of AA family CAZymes compared to the rest of the strains. The prevalence of CE family CAZymes was similar across 14I and 42L, which were higher than the rest of the strains (Fig. 5). Three major clades were generated based on the distribution of CAZymes in the genome of *L. coryniformis* strains. From the bottom-up in Fig. 5, the members of the non-plant-originated strains formed the first clade, such as PH-1, DSM 20004, and FOL-19 except KCTC 3535, which was isolated from kimchi. The second clade consisted of plant-associated strains KCTC 3167, CRL 1001, and DSM

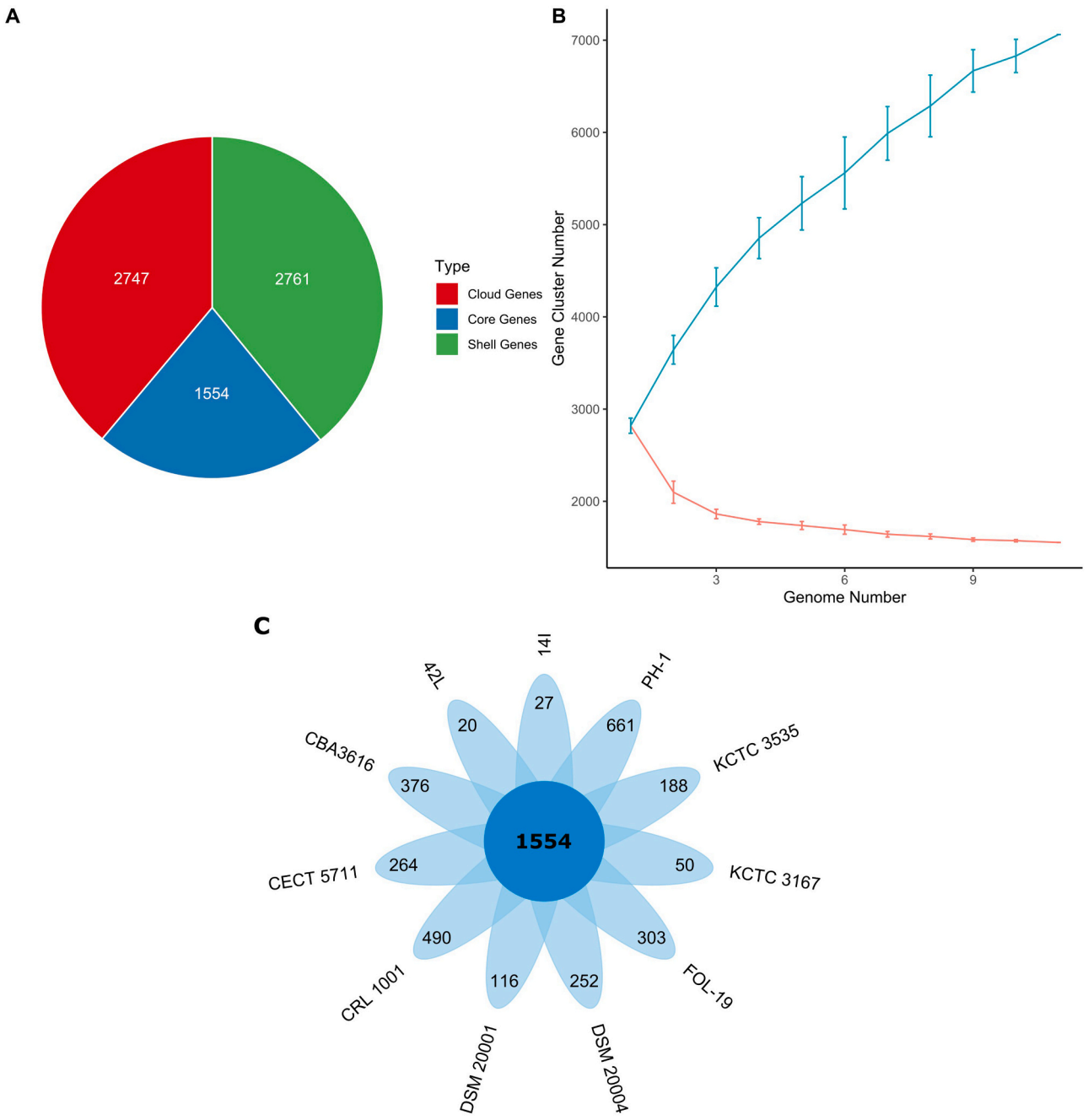

Fig. 3. (A) Coding sequence distributions in the eleven *L. coryniformis* pan-genome. Cloud genes (red), core genes (blue), and shell genes (green). (B) Estimation of core- (red line) and pan-genomes (blue line) of the eleven *L. coryniformis* strains by including genomes one by one. (C) Venn diagram representing the core and unique gene families of *L. coryniformis* obtained.

20001, except CECT 5711, isolated from cheese. The last clade members were isolated from fermented meat except CBA 3616, which was isolated from kimchi.

3.4. Mobile genetic elements

Antibiotic resistance genes were assessed to test the safety of *L. coryniformis* FOL-19, and no antibiotic resistance genes were identified using the CARD database [50]. Similarly, antibiotic resistance genes did not exist across other *L. coryniformis* strains. Bacteriocin screening via the BAGEL4 web tool did not yield any bacteriocin gene cluster in any strains analyzed.

Plasmid screening results using the PLSDB web tool revealed that all strains harbored at least one plasmid except KCTC 3167. DSM 20004 has the highest number of putative plasmids, and KCTC 3535 harbored three plasmids. The remaining strains had one or two putative plasmids (Table S1). The average length of the predicted putative plasmids is 16 Kb, and GC content varies between 36% and 43% (average 41%). All eleven *L. coryniformis* strains screened in the present study were predicted to carry IS elements (Table S2). PH-1, CECT 5711, and FOL-19 have the highest numbers of IS elements. All strains harbored IS elements from *Lactiplantibacillus plantarum,* the most predicted origin of the IS elements. The second most predicted origin of the IS elements was *Lactococcus lactis,* which was mainly harbored by CBA3616, CECT 5711, and FOL-19. Interestingly, PH-1 has more than 95% of the IS elements originating from *Escherichia coli*.

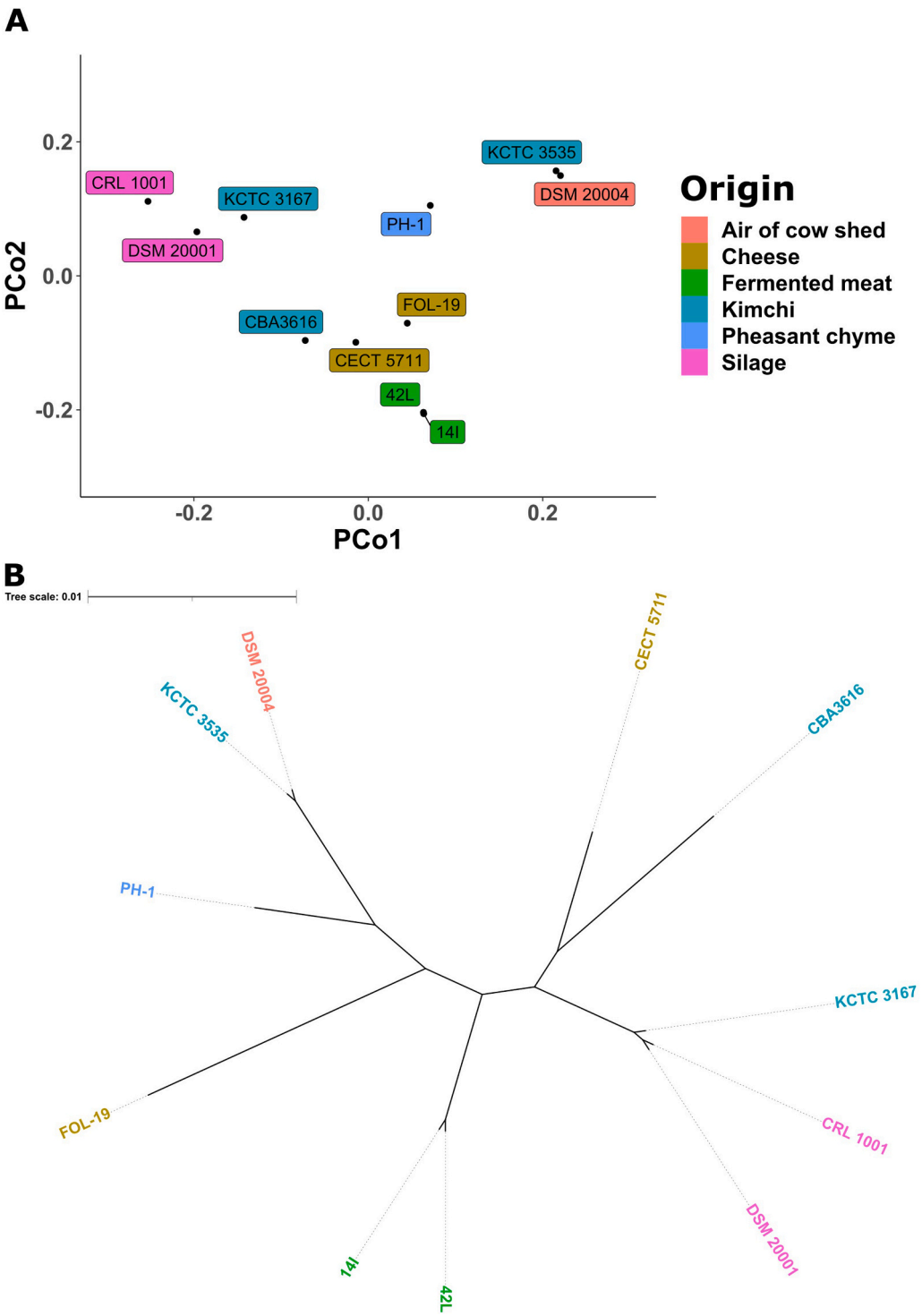

Fig. 4. (A) PCoA visualization of Jaccard distances based on shared genes across eleven *L. coryniformis* genomes screened. The color of each box indicates a unique isolation source. (B) Neighbor-joining unrooted phylogenetic tree based on core genome alignment. Each font color indicates a unique isolation source.

To better understand CRISPR-Cas systems in *L. coryniformis*, we defined and located spacers (Fig. 6A) and repeats (Fig. 6B) and successfully assigned them to canonical types and subtypes [48,57]. All *L. coryniformis* genomes analyzed in the present work was predicted to encode CRISPR-Cas systems with the exception of PH-1. Four different CRISPR-Cas systems were detected which belong to type I-C, I-E, II-A, and II-U subtypes (Table S3). When the subtypes were categorized, a type I-E and I-C was represented by 14I, 42L, CBA3616, and CECT 5711. On the other hand, a type II-A was represented by CRL 1001, DSM 20001, FOL-19, and KCTC 3167. Type II-U was only represented by CBA3616, DSM 20004, and KCTC 3535. Interestingly, CBA 3616 was the only *L. coryniformis* strain represented by type I-C, I-E, and II-U. The alignment of spacers shows four distinct groups. Among ten strains, multiple CRISPR loci were detected in all except for DSM 20001 which had a single CRISPR locus. Noteworthy, FOL-19 carried three different spacers and repeats similar to another cheese isolate of CECT 5711 which also had three unique spacer and repeat content. The highest spacer and repeat content diversity was achieved with CBA3616 isolated

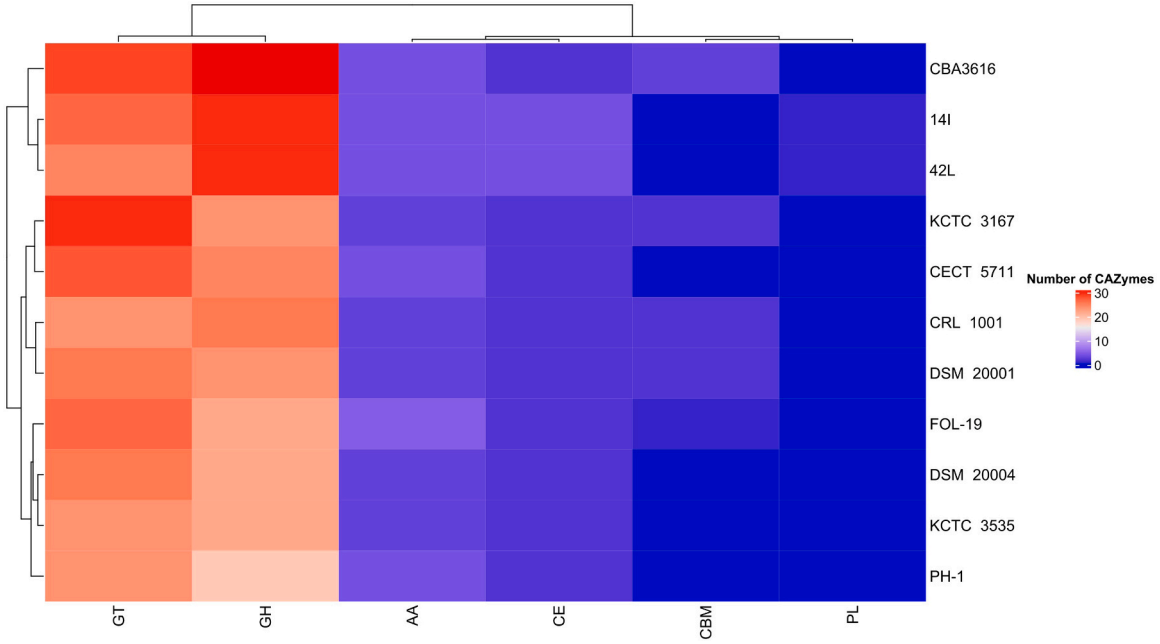

Fig. 5. Heatmap showing the distribution of CAZyme families across the eleven *L. coryniformis* genomes. Color gradients represent the number of CAZymes. The number of CAZymes is increasing from lighter to darker colors. AA: Auxiliary activities, CE: Carbohydrate esterase family, CMB: Carbohydrate-binding module family, GH: Glycoside hydrolase, GT: Glycosyltransferase, PL: Polysaccharide lyase.

from kimchi.

3.5. Genome identification of reuterin and vitamin B_{12} clusters

The comparative genome analysis performed across eleven *L. coryniformis* genomes identified *pdu* and *cbi-cob-hem* clusters and displayed a high degree of homology with previously reported *pdu* and *cbi-cob-hem* cluster exist in the genome of *Limosilactobacillus reuteri* JCM 1112 [21]. *In silico* genomic analysis also detected the existence of 36 open reading frames related to coenzyme B_{12} and reuterin biosynthesis. Among the predicted *pdu* and *cbi-cob-hem* genes, the orientation was heterogeneously distributed. Across *pdu* genes, *pduC*, *pduD*, and *pduE* were found. These genes form the *gup* operon. Moreover, *pduX* gene encoding the L-threonine kinase was located in the genome of *L. coryniformis* genomes. *cysG* gene and seventeen *cbi* genes were identified via *in silico* analysis at the downstream of the *pdu* operon. Five *cob* genes (*cobA*, *cobD*, *cobU*, *cobS*, and *cobC*) were identified adjacent to *cbi* and *hem* gene clusters (i.e., downstream of the *cbi* but upstream of *cob* genes). Four hem genes (*hemA*, *hemC*, *hemB*, and *hemL*) were identified. Interestingly, *hemD* was not detected in any *L. coryniformis* genomes (Fig. 7).

Fig. 7 shows the *pdu* and *cbi-cob-hem* gene clusters of all *L. coryniformis* strains except for PH-1 which did not encode these gene clusters. The *pdu* and *cbi-cob-hem* clusters are responsible for 3-HPA and cobalamin biosynthesis in lactobacilli species [14,15]. *pdu* and *cbi--cob-hem* gene clusters were distributed across different loci of *L. coryniformis* genomes. In the genomes of several *L. coryniformis* strains, those gene clusters were located on the negative strand.

FOL-19 was predicted to possess gene sets of *cbi*, *cob*, and *hem* which are required for cobalamin biosynthesis. These gene sets are positioned adjacent to *pdu* operon (Fig. 8). 1,3-propanediol dehydrogenase encoding gene was also detected in FOL-19. *cbiOQNM* gene set is responsible for the ABC transport of cobalt and glutamine across the cytoplasm, which were further condensed into cobalamin. *cobCSU*, *hemLBCA*, *cbiP*, *cbiLK*, *cobA*, *cbiJHGFTEDCBA*, and *cobD* gene sets play key role for the transformation of glutamate to cobalamin. Glycerol is transported to cytoplasm through *pduF* encoding for propanediol diffusion facilitator protein followed by converted to reuterin via *pduEDC* gene cluster. Glycerol can also feed into the glycolysis shunt from the lower half of the pathway. *L. coryniformis* FOL-19 was predicted to encode fructose bisphoshate aldolase implying a homofermentative lifestyle while utilizing hexose sugars such as glucose through glycolysis. It also carried phosphoketolase perhaps indicating that the pentose phosphate pathway is available for fermenting five carbon sugars (Fig. 8).

4. Discussion

In this study, we performed a comparative genomic evaluation of *Loigolactobacillus coryniformis* species, focusing on the novel strain FOL-19 isolated from cheese in the present work. The GC content of FOL-19 is 42.83%, and the average GC content of *L. coryniformis* is 42.96%, typical for low-GC lactobacilli [58]. This finding suggests that *L. coryniformis* has experienced less genomic drift since lactobacilli are generally considered low-GC organisms. It has been reported that lactobacilli are highly adapted to their microenvironment by undergoing genome decay or gene loss [59]. The high portion (39%) of unknown/hypothetical genes indicates that there is still more to discover about *L. coryniformis* FOL-19.

After evaluating the genome of *L. coryniformis* FOL-19, we conducted a phylogeny of *L. coryniformis* using eleven genomes (Fig. 1). Phylogenetic analysis showed unclear grouping of strains by isolation source. For example, five main clades were identified, with *L. coryniformis* FOL-19 laid to the fifth member clade containing PH-1, KCTC 3535, and DSM 20004 (Fig. 1A). Although *L. coryniformis* FOL-19 was isolated from cheese, its clade members were isolated from pheasant chyme (PH-1), kimchi (KCTC 3535), and air of cowshed (DSM 20004). PH-1, pheasant chyme isolate, is the closest match to FOL-19. It would be anticipated that related strains would have similar isolation sources [60]. Since this is mostly not the case for *L. coryniformis* FOL-19, one could speculate that *L. coryniformis* might contaminate the cheese milk through dairy production environments. Instead of being a permanent member, it might be a transient member "allochthonous" of the cheese microbiome.

The sequence similarity results were higher than the sequence similarity threshold of 95% (i.e., CDS ANI) for species demarcation even though their pangenome was open. Fermented meat isolates of 14I and

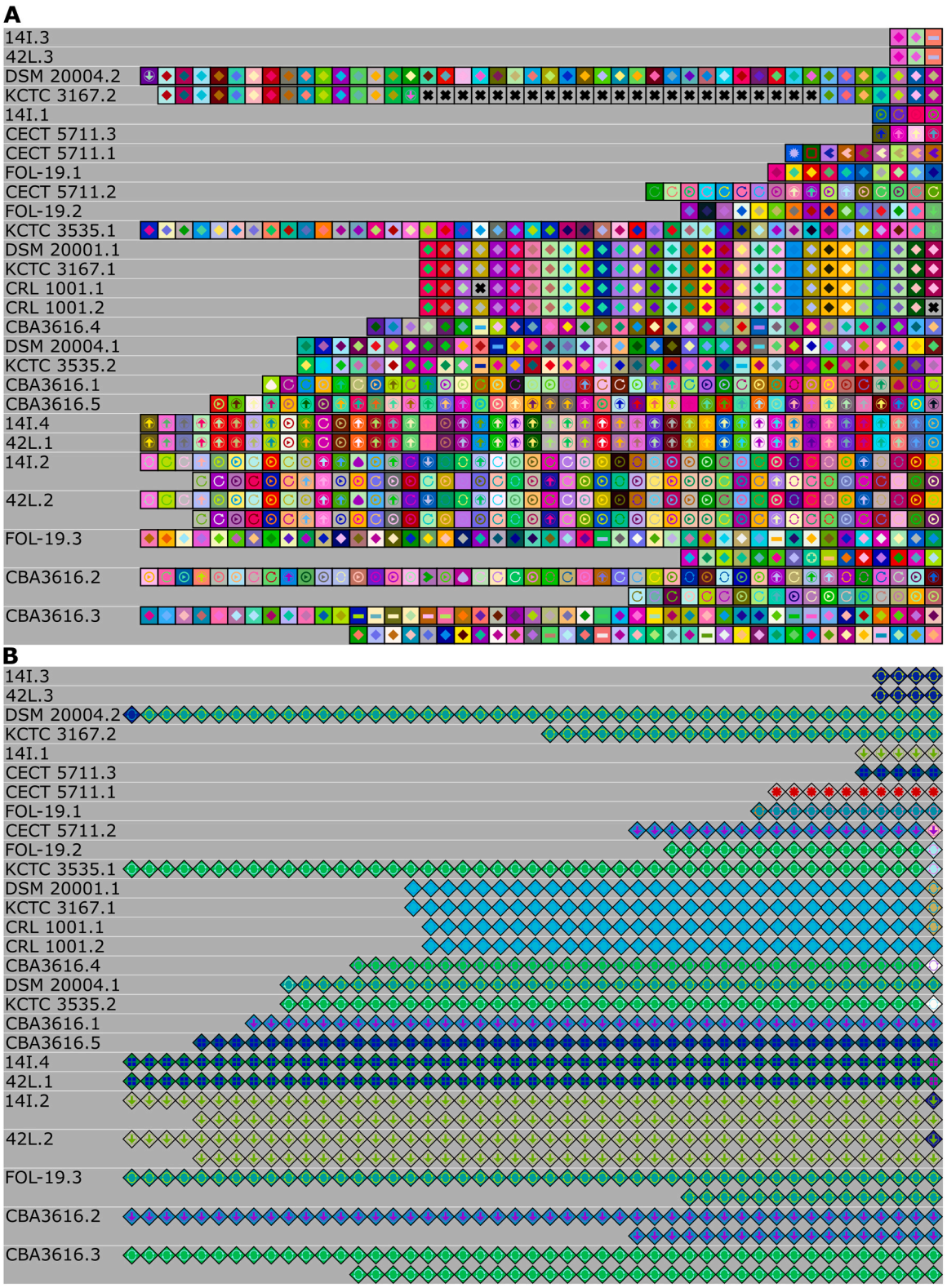

Fig. 6. Alignment of spacers (A) and repeats (B) of each detected CRISPR locus. Each colored diamond represents a unique repeat, and each colored square represents a unique spacer in the CRISPR-Cas system. Grey "x" boxes showed missing spacer.

42L share the same clade and PCoA cluster. Similarly, isolated from silage, CRL 1001 and DSM 20001 share the same clade and PCoA cluster. The differences in PCoA locations regarding phylogenetic distances could be attributed to accessory genes with the contribution of plasmid encoding genes [36]. For example, KCTC 3535 and CBA 3616 do not have any common plasmids, and KCTC 3167 has no plasmids, although they share the same isolation sources.

Identification of putative carbohydrate metabolism-associated genes revealed the sugar metabolism capability of *L. coryniformis* in a comparative manner. A bacterial strain's sugar fermentation capacity is

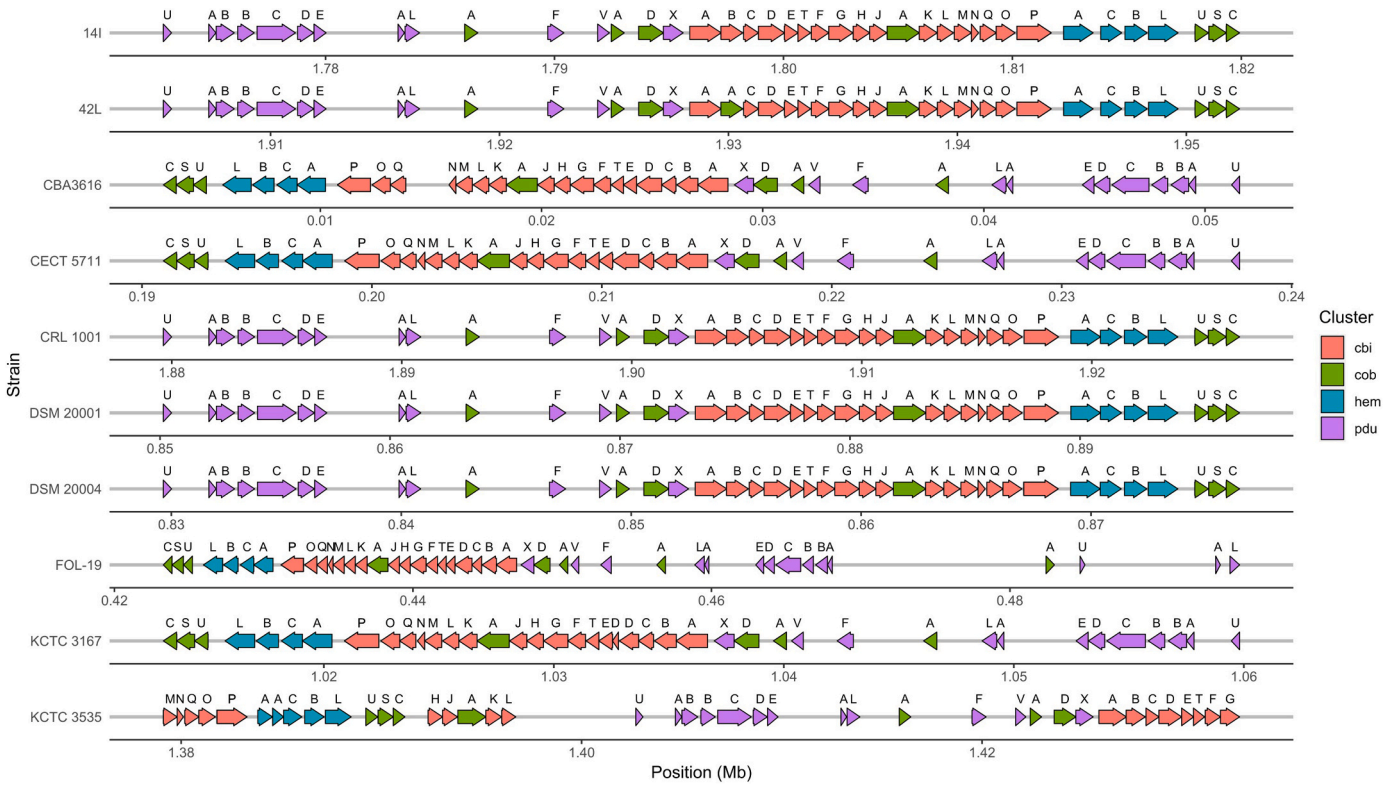

Fig. 7. Schematic representation of *pdu* and *cbi-cob-hem* clusters of ten *L. coryniformis* genomes.

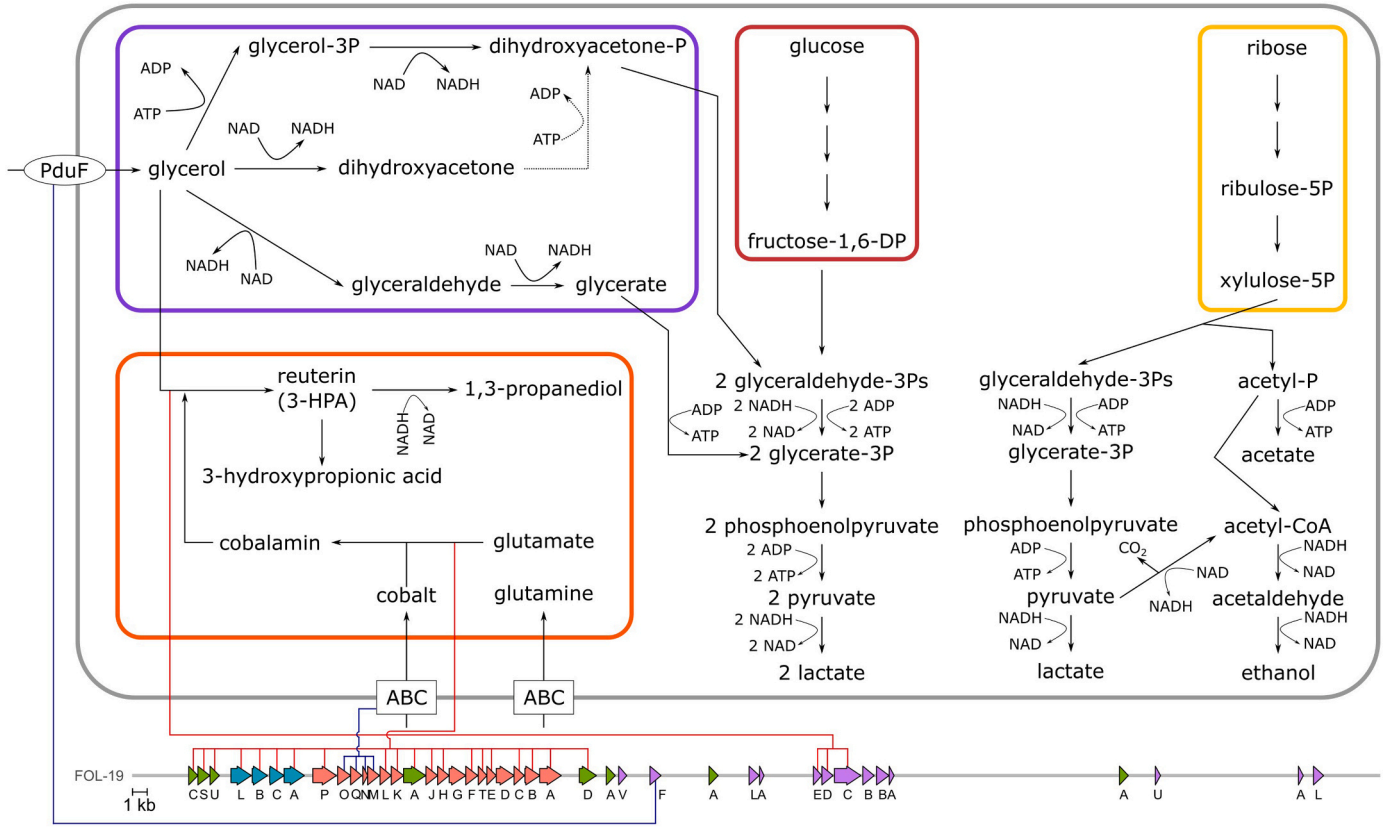

Fig. 8. Putative glucose, ribose and glycerol metabolism in *L. coryniformis* FOL-19. The bottom piece of the figure depicts the structure of *pdu-cbi-cob-hem* gene cluster in *L. coryniformis* FOL-19. Genes are shown with arrows depicting the transcription direction with the following colors: purple, *pdu* genes; green, *cob* genes; blue, *hem* genes; red, *cbi* genes. Lines connect corresponding genes across the pathway and the cluster (red, enzymes; blue, transporters).

a key indicator of strain metabolic function and sets the fundamentals for strain selection and cultivation [61]. The presence of phosphoketolase and fructose bisphosphate aldolase genes across all *L. coryniformis* genomes indicated this species' facultatively heterofermentative carbohydrate metabolism (Table S4) [62]. CAZymes participate in biosynthesis (glycosyltransferases, GTs), degradation (glycoside hydrolases, GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and enzymes for auxiliary activities (AAs), and recognition (carbohydrate-binding module (CBM)) of various complex sugars functional in carbohydrate metabolism [63]. Several types of GTs participate in disaccharide, oligosaccharide, and polysaccharide biosynthesis, which are instrumental in forming glycosidic bonds [64]. CAZyme identified 6 GT families in the *L. coryniformis* FOL-19 genome, and enzymes belonging to GT2 and GT4 families represent 70.37% of all GTs responsible for cellulose synthase, chitin synthase, sucrose synthase, galactosyltransferase, and glucosyltransferase biosynthesis. GH is the main enzyme family that functions in the metabolism of carbohydrates and plays a critical role in the carbohydrate glycosidic bond hydrolysis [64]. The FOL-19 genome is predicted to carry genes functional in beta-glucosidase (GH1, GH3), beta-galactosidase (GH2), and hexosyltransferase (GH13_31) biosynthesis. These enzymes are functional in carbohydrate metabolism; for instance, the utilization of lactose, sucrose, and oligosaccharides is essential for the proliferation of organisms in various microenvironments, including dairy-associated niches [64,65]. The prevalence of lactose intolerance was estimated at around 67% globally [66,67] due to a lack of β-galactosidase, which hydrolyzes lactose into glucose and galactose. Hydrolysis of lactose also develops the texture of milk products [66]. Among the *L. coryniformis* strains tested, FOL-19 is the only *L. coryniformis* strain that harbors the *lacZ* gene encoding the β-galactosidase enzyme (Table S4). This could make FOL-19 a potential adjunct culture candidate in the dairy industry for reducing the amount of lactose in final dairy products. FOL-19 genome was predicted to carry lysozyme (GH73) encoding genes generally linked to catalysis of beta-1,4 bond hydrolysis across N-acetylglucosamine and N-acetylmuramic acid of the bacterial cell wall. Lysozyme could also show an antimicrobial spectrum [68] by disrupting the bacterial cellular integrity and causing death. Moreover, hydrolysis products of the bacterial cell wall could enhance immunoglobulin A secretion, activation of macrophages, and bacterial pathogen clearance [69–71].

The *oppABCDF* operon was found in all *L. coryniformis* strains tested (Table S5); however, PII-type serine proteinase that is functional against caseins was absent across all *L. coryniformis* strains studied [34]. *pepX* gene encoding x-prolyl dipeptidyl aminopeptidase was complete in FOL-19; however, it was truncated in several LAB strains [34]. *pepE* and *pepT* genes encoding aminopeptidase E and peptidase T enzymes did exist in all *L. coryniformis* strains, including FOL-19 though both genes were truncated in CRL 1001. Interestingly, all strains except FOL-19, DSM 20004, and KCTC 3535, were predicted to carry aminopeptidase *pepS*. Like *pepE* and *pepT* genes, the *pepS* gene was also truncated in silage isolate CRL 1001 (Table S6). *pepV* gene encoding β-ala-dipeptidase, which is known to cut dipeptides by N-terminal D-alanine or β-alanine residue [34,72], was carried by all eleven *L. coryniformis* strains. However, the *pepV* gene was truncated in CRL 1001. It has been reported that LAB is heavily adapted to their corresponding ecological niches and have smaller genomes than other bacteria due to genome reduction, which results in the maintenance of the required number of genes necessary for niche-specific survivability [59,73].

Horizontal gene transfer (HGT) is the primary factor in bacterial evolution that could bestow fitness and niche adaptation [74]. To identify genomic islands, we compared ten *L. coryniformis* genomes against the reference genome DSM 20001 in BRIG, which resulted in seven genomic islands (Fig. 2). *L. coryniformis* FOL-19 was predicted to carry several strong metabolic islands absent in other *L. coryniformis* genomes. These genomic islands were detected based on their presence and absence in other genomes and an apparent reduction in their GC content. Moreover, two CRISPR regions were identified adjacent to genomic islands in FOL-19, revealing that HGT events might have contributed to the acquisition of CRISPR.

CRISPR/Cas systems are invaluable tools for genome editing [60], and we screened CRISPR systems in eleven *L. coryniformis* genomes. We found on a species level that a hundred percent of strains encoded at least four predicted CRISPR elements and at least one Cas cluster except PH-1 (Table S3). This is ~58% higher than the lactobacilli in general and ~117% higher than bacteria as a whole, which implies that *L. coryniformis* holds a promising potential for being a reservoir for new CRISPR-based tools [75]. Type II was the most common Cas system in *L. coryniformis* (53%), whereas Type I was present in 36% of strains tested. Type II is the most popular Cas-based genome editing tool in the CRISPR toolbox [76]. Type II in *L. coryniformis* strains was higher than lactobacilli in general [75]. A putative CRISPR/Cas locus was also identified in FOL-19, which implies immunity against phage infections and a crucial biotechnological trait of starter or adjunct LAB used in the fermented food industry. It is also instrumental in plasmid interference to prevent the uptake of unwanted plasmids carrying antibiotic-resistance genes [77].

Fermented meat isolates of 14I and 42L were predicted to have the same spacer length and identity which were also found in the same clade (Fig. 4B). Interestingly, silage isolate of DSM 20001 had the same spacer length and identity with kimchi isolate of KCTC 3167. These two strains shared the same clade in the core genome SNPs based phylogenetic tree. While KCTC 3535 was the closest genome to DSM 20004 according to PCoA analysis based on the shared genes, neither spacers nor repeats were identical between these strains. While FOL-19 and CECT 5711 were both isolated from cheese, these two strains had no spacer or repeat identity. According to these findings, it can be speculated that *L. coryniformis* strains are notably diverse with regard to CRISPR-Cas system and genomic rearrangements. The pronounced spacer diversity across strains including those isolated from the same sources suggests that each individual strain was exposed to different ecological circumstances and evolutionary history [60].

Among all *L. coryniformis* strains tested, the most common IS element shared across all strains was closely related to *Lactiplantibacillus plantarum*, which is heavily utilized as a probiotic dietary supplement, starter culture in plant based fermentations, and bio-protective culture against food-borne pathogens due to its antimicrobial activity by producing bacteriocins [78,79]. Cheese-originated *L. coryniformis* strains share IS elements with *Lacticaseibacillus casei*, which plays a significant role in cheese ripening and can survive the acidic and ketone-rich environment of ripened cheese such as Parmigiano Reggiano and Grana Padano [80]. Moreover, they share IS elements with *Lactiplantibacillus plantarum*, an adjunct culture to produce long-shelf-life cheese and to enhance the flavor of fermented milk products [81,82]. Similarly, *L. coryniformis* strains isolated from fermented meat share IS elements with *Lacticaseibacillus casei*, which can synthesize volatile compounds to enrich the flavor of probiotic food products. Another mutual IS element belonging to *Latilactobacillus sakei* possesses antifungal and antipathogenic activity, thus functional in preserving fresh and fermented food products [83,84].

Although it was reported that *L. coryniformis* is a potential cobalamin and 3-HPA producer, we did not come across any study describing the extent of those metabolisms across different *L. coryniformis* isolates. We discovered that reuterin and cobalamin biosynthesis are highly prevalent among most if not all *L. coryniformis* strains. Adjacency between *pdu* operon and gene sets of *cbi-cob* can be reflected as dependency on cobalamin for glycerol dehydratase activity [74,85]. As mentioned above, FOL-19 possesses *cbi-cob-hem* gene cluster and *pdu* operon for biosynthesis of cobalamin and reuterin, respectively. We propose that *pdu-cbi-cob-hem* gene cluster provides a competitive advantage to FOL-19 over other organism found in the same ecological niche and contribute to the probiotic potential of FOL-19. However, it should be confirmed whether putative gene products of *pdu-cbi-cob-hem* gene

cluster is consistent with FOL-19's capacity to produce both cobalamin and reuterin under in vitro conditions.

5. Conclusions

Overall, the present study puts forth a basis for genomic analysis of *L. coryniformis* strains, focusing on FOL-19 isolated from artisanal Tulum cheese manufactured in the Eastern Anatolia region. Whole-genome sequence analysis of eleven different strains revealed that these strains are highly variable and enriched in the CRISPR/Cas system, IS elements, genomic islands, and plasmids. *L. coryniformis* strains FOL-19, KCTC 3167, and CRL 1001 were predicted to carry a single Type II-A CRISPR/Cas system. Only *L. coryniformis* FOL-19 and DSM 20004 harbor a single plasmid. *L. coryniformis* FOL-19 was predicted to be the only strain that harbors the *lacZ* gene encoding β-galactosidase, which plays a crucial role in improving dairy products' digestibility by hydrolyzing lactose sugar. All *L. coryniformis* strains except PH-1 were predicted to possess *pdu-cbi-cob-hem* gene cluster encoding for cobalamin and reuterin biosynthesis strengthening the probiotic potential of this species. These observations pave the way for new means for functional evaluations of *Loigolactobacillus coryniformis* strains, closely related species, and further discoveries of the biotechnologically relevant phenotypes.

Ethics approval and consent to participate

Not applicable.

Funding

This work has been supported by Istanbul Technical University Scientific Research Projects Unit with grant number MAB-2023-44444.

CRediT authorship contribution statement

FO: Conceptualization, Supervision, Investigation, Data curation, Formal analysis, Visualization, Writing – original draft. **IG:** Investigation, Data curation, Formal analysis, Visualization, Writing – original draft. All authors contributed to the article and approved the submitted version.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgment

Not applicable.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.10.004.

References

[1] von Wright A., Axelsson L. Lactic Acid Bacteria: An Introduction. Lactic Acid Bacteria Microbiological And Functional Aspects. 5th ed., CRC Press; 2019.
[2] Mokoena MP. Lactic acid bacteria and their bacteriocins: classification, biosynthesis and applications against uropathogens: a mini-review. Molecules 2017:22. https://doi.org/10.3390/molecules22081255.
[3] Żukiewicz-Sobczak W, Wróblewska P, Adamczuk P, Silny W. Probiotic lactic acid bacteria and their potential in the prevention and treatment of allergic diseases. Cent Eur J Immunol 2014;39:104–8. https://doi.org/10.5114/ceji.2014.42134.
[4] Mookiah S, Sieo CC, Ramasamy K, Abdullah N, Ho YW. Effects of dietary prebiotics, probiotic and synbiotics on performance, caecal bacterial populations and caecal fermentation concentrations of broiler chickens. J Sci Food Agric 2014; 94:341–8. https://doi.org/10.1002/jsfa.6365.
[5] Niibo M, Shirouchi B, Umegatani M, Morita Y, Ogawa A, Sakai F, et al. Probiotic Lactobacillus gasseri SBT2055 improves insulin secretion in a diabetic rat model. J Dairy Sci 2019;102:997–1006. https://doi.org/10.3168/jds.2018-15203.
[6] Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, et al. A taxonomic note on the genus Lactobacillus: Description of 23 novel genera, emended description of the genus Lactobacillus Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. Int J Syst Evolut Microbiol 2020;70: 2782–858. https://doi.org/10.1099/ijsem.0.004107.
[7] Geissler Andreas J, Behr Jürgen, Vogel Rudi F. Multiple genome sequences of the important beer-spoiling species Lactobacillus backii. e00826-16 Genome Announc 2016;4. https://doi.org/10.1128/genomeA.00826-16.
[8] McMahon DJ, Bowen IB, Green I, Domek M, Oberg CJ. Growth and survival characteristics of Paucilactobacillus wasatchensis WDCO4. J Dairy Sci 2020;103: 8771–81. https://doi.org/10.3168/jds.2020-18597.
[9] Li X, Chen F, Wang X, Sun L, Guo L, Xiong Y, et al. Impacts of low temperature and ensiling period on the bacterial community of oat silage by SMRT. Microorganisms 2021;9:274. https://doi.org/10.3390/microorganisms9020274.
[10] Loigolactobacillus. NCBI 2022. ⟨https://www.ncbi.nlm.nih.gov/data-hub/taxonomy/2767889/⟩ (accessed March 29, 2022).
[11] Fernández-Ferreiro A, Formigo-Couceiro FJ, Veiga-Gutierrez R, Maldonado-Lobón JA, Hermida-Cao AM, Rodriguez C, et al. Effects of loigolactobacillus coryniformis K8 CECT 5711 on the immune response of elderly subjects to COVID-19 vaccination: a randomized controlled trial. Nutrients 2022;14:228. https://doi.org/10.3390/nu14010228.
[12] Redondo N, Nova E, Gheorghe A, Díaz LE, Hernández A, Marcos A. Evaluation of Lactobacillus coryniformis CECT5711 strain as a coadjuvant in a vaccination process: a randomised clinical trial in healthy adults. Nutr Metab 2017;14:2. https://doi.org/10.1186/s12986-016-0154-2.
[13] Wildman R.E., Wildman R., Wallace T.C. Handbook of nutraceuticals and functional foods. CRC press; 2016.
[14] Torres AC, Vannini V, Bonacina J, Font G, Saavedra L, Taranto MP. Cobalamin production by Lactobacillus coryniformis: biochemical identification of the synthetized corrinoid and genomic analysis of the biosynthetic cluster. BMC Microbiol 2016;16:240. https://doi.org/10.1186/s12866-016-0854-9.
[15] Morita H, Toh H, Fukuda S, Horikawa H, Oshima K, Suzuki T, et al. Comparative genome analysis of lactobacillus reuteri and lactobacillus fermentum reveal a genomic island for reuterin and cobalamin production. DNA Res. 2008;15:151–61. https://doi.org/10.1093/dnares/dsn009.
[16] Santos F, Vera JL, van der Heijden R, Valdez G, de Vos WM, Sesma F, et al. The complete coenzyme B12 biosynthesis gene cluster of Lactobacillus reuteri CRL1098. Microbiology 2008;154:81–93. https://doi.org/10.1099/mic.0.2007/011569-0.
[17] Taranto María P, Vera José L. Lactobacillus reuteri CRL1098 produces cobalamin. J Bacteriol 2003;185:5643–7. https://doi.org/10.1128/jb.185.18.5643-5647.2003.
[18] Sriramulu Dinesh Diraviam Liang Mingzhi, Hernandez-Romero Diana Raux-Deery Evelyne, Lünsdorf Heinrich Parsons Joshua B, et al. Lactobacillus reuteri DSM 20016 produces cobalamin-dependent diol dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation. J Bacteriol 2008;190:4559–67. https://doi.org/10.1128/jb.01535-07.
[19] De Angelis M, Bottacini F, Fosso B, Kelleher P, Calasso M, Di Cagno R, et al. Lactobacillus rossiae, a vitamin B12 producer, represents a metabolically versatile species within the genus lactobacillus. PLOS ONE 2014;9:e107232. https://doi.org/10.1371/journal.pone.0107232.
[20] Martín R, Olivares M, Marín ML, Xaus J, Fernández L, Rodríguez JM. Characterization of a reuterin-producing Lactobacillus coryniformis strain isolated from a goat's milk cheese. Int J Food Microbiol 2005;104:267–77. https://doi.org/10.1016/j.ijfoodmicro.2005.03.007.
[21] TANAKA O, KOMATSU T, OSHIBE A, CAI Y, MIYAZAKI S, NAKANISHI K. Production of 3-hydroxypropionaldehyde in silage inoculated with lactobacillus coryniformis plus glycerol. Biosci, Biotechnol, Biochem 2009;73:1494–9. https://doi.org/10.1271/bbb.80863.
[22] Talarico TL, Casas IA, Chung TC, Dobrogosz WJ. Production and isolation of reuterin, a growth inhibitor produced by Lactobacillus reuteri. Antimicrob Agents Chemother 1988;32:1854–8. https://doi.org/10.1128/aac.32.12.1854.
[23] Lüthi-Peng Q, Dileme F, Puhan Z. Effect of glucose on glycerol bioconversion by Lactobacillus reuteri. Appl Microbiol Biotechnol 2002;59:289–96. https://doi.org/10.1007/s00253-002-1002-z.
[24] Kitamoto HK, Hasebe A, Ohmomo S, Suto EG, Muraki M, Iimura Y. Prevention of aerobic spoilage of maize silage by a genetically modified killer yeast, kluyveromyces lactis, defective in the ability to grow on lactic acid. Appl Environ Microbiol 1999;65:4697–700. https://doi.org/10.1128/AEM.65.10.4697-4700.1999.
[25] Cai Y, Ohmomo S, Ogawa M, Kumai S. Effect of NaCl-tolerant lactic acid bacteria and NaCl on the fermentation characteristics and aerobic stability of silage. J Appl Microbiol 1997:83.
[26] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST +: architecture and applications. BMC Bioinforma 2009;10:421. https://doi.org/10.1186/1471-2105-10-421.
[27] Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci 1977;74:5463–7. https://doi.org/10.1073/pnas.74.12.5463.

[28] Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, et al. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res 2020;48:D606–12. https://doi.org/10.1093/nar/gkz943.
[29] Krueger F. Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for *Msp*I-digested RRBS-type (Reduced Representation Bisufite-Seq) libraries. 2012.
[30] Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2016;44:D67–72. https://doi.org/10.1093/nar/gkv1276.
[31] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics 2014; 30:2068–9. https://doi.org/10.1093/bioinformatics/btu153.
[32] Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 2015;31:3691–3. https://doi.org/10.1093/bioinformatics/btv421.
[33] Snipen L., Liland K.H. micropan: Microbial Pan-Genome Analysis. 2020.
[34] Frantzen CA, Kot W, Pedersen TB, Ardö YM, Broadbent JR, Neve H, et al. Genomic characterization of dairy associated leuconostoc species and diversity of leuconostocs in undefined mixed mesophilic starter cultures. Front Microbiol 2017: 8.
[35] Hennig C., Hausdorf B. prabclus: Functions for Clustering and Testing of Presence-Absence, Abundance and Multilocus Genetic Data. 2020.
[36] Candeliere F, Raimondi S, Spampinato G, Tay MYF, Amaretti A, Schlundt J, et al. Comparative genomics of leuconostoc carnosum. Front. Microbiol. 2021:11.
[37] Core Team R. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
[38] Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 2009;26:1641–50. https://doi.org/10.1093/molbev/msp077.
[39] Verce M, De Vuyst L, Weckx S. Comparative genomics of Lactobacillus fermentum suggests a free-living lifestyle of this lactic acid bacterial species. Food Microbiol 2020;89:103448. https://doi.org/10.1016/j.fm.2020.103448.
[40] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res 2021;49:W293–6. https://doi.org/10.1093/nar/gkab301.
[41] Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 2019;10:2182. https://doi.org/10.1038/s41467-019-10210-3.
[42] Lefort V, Desper R, Gascuel O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program: table 1. Mol Biol Evol 2015;32: 2798–800. https://doi.org/10.1093/molbev/msv150.
[43] Seemann T. Snippy: rapid haploid variant calling and core SNP phylogeny. GitHub Available at: Github Com/Tseemann/Snippy 2015.
[44] Kuznetsov A, Bollin CJ. NCBI genome workbench: desktop software for comparative genomics, visualization, and genbank data submission. In: Katoh K, editor. Multiple Sequence Alignment: Methods and Protocols, New York, NY. US: Springer; 2021. p. 261–95. https://doi.org/10.1007/978-1-0716-1036-7_16.
[45] Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl Environ Microbiol 2013;79:7696–701. https://doi.org/10.1128/AEM.02411-13.
[46] Soares SC, Geyik H, Ramos RTJ, de Sá PHCG, Barbosa EGV, Baumbach J, et al. GIPSy: genomic island prediction software. J Biotechnol 2016;232:2–11. https://doi.org/10.1016/j.jbiotec.2015.09.008.
[47] Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genom 2011;12:402. https://doi.org/10.1186/1471-2164-12-402.
[48] Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res 2018;46:W246–51. https://doi.org/10.1093/nar/gky425.
[49] Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res 2018;46:W200–4. https://doi.org/10.1093/nar/gky448.
[50] Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res 2020;48:D517–25. https://doi.org/10.1093/nar/gkz935.
[51] NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2016;44:D7–19. https://doi.org/10.1093/nar/gkv1290.
[52] Galata V, Fehlmann T, Backes C, Keller A. PLSDB: a resource of complete bacterial plasmids. Nucleic Acids Res 2019;47:D195–202. https://doi.org/10.1093/nar/gky1050.
[53] Schmartz GP, Hartung A, Hirsch P, Kern F, Fehlmann T, Müller R, et al. PLSDB: advancing a comprehensive database of bacterial plasmids. Nucleic Acids Res 2022;50:D273–8. https://doi.org/10.1093/nar/gkab1111.
[54] Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 2006;34:D32–6. https://doi.org/10.1093/nar/gkj014.
[55] Daughtry KV, Johanningsmeier SD, Sanozky-Dawes R, Klaenhammer TR, Barrangou R. Phenotypic and genotypic diversity of Lactobacillus buchneri strains isolated from spoiled, fermented cucumber. Int J Food Microbiol 2018;280:46–56. https://doi.org/10.1016/j.ijfoodmicro.2018.04.044.
[56] Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 2008;11:472–7. https://doi.org/10.1016/j.mib.2008.09.006.
[57] Nethery MA, Barrangou R. CRISPR Visualizer: rapid identification and visualization of CRISPR loci via an automated high-throughput processing pipeline. RNA Biol 2019;16:577–84. https://doi.org/10.1080/15476286.2018.1493332.
[58] Brandt K, Barrangou R. Using glycolysis enzyme sequences to inform Lactobacillus phylogeny. Microb Genom 2018:4.
[59] Makarova K, Slesarev A, Wolf Y, Sorokin A, Mirkin B, Koonin E, et al. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci USA 2006;103:15611–6. https://doi.org/10.1073/pnas.0607117103.
[60] Brandt K, Nethery MA, O'Flaherty S, Barrangou R. Genomic characterization of Lactobacillus fermentum DSM 20052. BMC Genom 2020;21:328. https://doi.org/10.1186/s12864-020-6740-8.
[61] Jiang J, Yang B, Ross RP, Stanton C, Zhao J, Zhang H, et al. Comparative genomics of pediococcus pentosaceus isolated from different niches reveals genetic diversity in carbohydrate metabolism and immune system. Front Microbiol 2020:11.
[62] von Wright A., Axelsson L. Lactic Acid Bacteria An Introduction. In: Vinderola G., Ouwehand A.C., Salminen S., Wright A. von, editors. Lactic Acid Bacteria: Microbiological and Functional Aspects. 5th ed., Boca Raton: CRC Press; 2019. https://doi.org/10.1201/9780429057465.
[63] Oliveira FS, da Silva Rodrigues R, de Carvalho AF, Nero LA. Genomic analyses of pediococcus pentosaceus ST65ACC, a bacteriocinogenic strain isolated from artisanal raw-milk cheese. Probiotics Antimicro Prot 2022. https://doi.org/10.1007/s12602-021-09894-1.
[64] Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res 2022;50:D571–7. https://doi.org/10.1093/nar/gkab1045.
[65] Vera C, Guerrero C, Aburto C, Cordova A, Illanes A. Conventional and non-conventional applications of β-galactosidases. Biochim Et Biophys Acta (BBA) - Proteins Proteom 2020;1868:140271. https://doi.org/10.1016/j.bbapap.2019.140271.
[66] Saqib S, Akram A, Halim SA, Tassaduq R. Sources of β-galactosidase and its applications in food industry. 3 Biotech 2017;7:79. https://doi.org/10.1007/s13205-017-0645-5.
[67] Storhaug CL, Fosse SK, Fadnes LT. Country, regional, and global estimates for lactose malabsorption in adults: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2017;2:738–46. https://doi.org/10.1016/S2468-1253(17)30154-1.
[68] Michlmayr H, Kneifel W. β-Glucosidase activities of lactic acid bacteria: mechanisms, impact on fermented food and human health. FEMS Microbiol Lett 2014;352:1–10. https://doi.org/10.1111/1574-6968.12348.
[69] Ellison 3rd RT, Giehl TJ. Killing of gram-negative bacteria by lactoferrin and lysozyme. J Clin Invest 1991;88:1080–91. https://doi.org/10.1172/JCI115407.
[70] Kawano M, Namba Y, Hanaoka M. Regulatory factors of lymphocyte-lymphocyte interaction: I. Con a-induced mitogenic factor acts on the late G1 stage of T-cell proliferation. Microbiol Immunol 1981;25:505–15. https://doi.org/10.1111/j.1348-0421.1981.tb00052.x.
[71] Clarke TB, Davis KM, Lysenko ES, Zhou AY, Yu Y, Weiser JN. Recognition of peptidoglycan from the microbiota by Nod1 enhances systemic innate immunity. Nat Med 2010;16:228–31. https://doi.org/10.1038/nm.2087.
[72] Vongerichten KF, Klein JR, Matern H, Plapp R. Cloning and nucleotide sequence analysis of pepV, a carnosinase gene from Lactobacillus delbrueckiisubsp. lactisDSM 7290, and partial characterization of the enzyme. Microbiology 1994; 140:2591–600.
[73] Nethery MA, Henriksen ED, Daughtry KV, Johanningsmeier SD, Barrangou R. Comparative genomics of eight Lactobacillus buchneri strains isolated from food spoilage. BMC Genom 2019;20:902. https://doi.org/10.1186/s12864-019-6274-0.
[74] Lawrence JG. Gene transfer, speciation, and the evolution of bacterial genomes. Curr Opin Microbiol 1999;2:519–23. https://doi.org/10.1016/S1369-5274(99)00010-7.
[75] Sun Z, Harris HMB, McCann A, Guo C, Argimón S, Zhang W, et al. Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera. Nat Commun 2015;6:8322. https://doi.org/10.1038/ncomms9322.
[76] Martin Jinek Chylinski Krzysztof, Fonfara Ines Hauer Michael, Doudna Jennifer A. Charpentier emmanuelle. a programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. Science 2012;337:816–21. https://doi.org/10.1126/science.1225829.
[77] Barrangou R, Horvath P. CRISPR: new horizons in phage resistance and strain identification. Annu Rev Food Sci Technol 2012;3:143–62. https://doi.org/10.1146/annurev-food-022811-101134.
[78] Lee K, Lee Y. Effect of Lactobacillus plantarum as a starter on the food quality and microbiota of kimchi. Food Sci Biotechnol 2010;19:641–6. https://doi.org/10.1007/s10068-010-0090-2.
[79] Arena MP, Silvain A, Normanno G, Grieco F, Drider D, Spano G, et al. Use of lactobacillus plantarum strains as a bio-control strategy against food-borne pathogenic microorganisms. Front Microbiol 2016:7.
[80] Bottari B, Levante A, Neviani E, Gatti M. How the fewest become the Greatest. L. casei's impact on long ripened cheeses. Front Microbiol 2018:9.
[81] Papadopoulou OS, Argyri AA, Varzakis EE, Tassou CC, Chorianopoulos NG. Greek functional Feta cheese: Enhancing quality and safety using a Lactobacillus plantarum strain with probiotic potential. Food Microbiol 2018;74:21–33. https://doi.org/10.1016/j.fm.2018.02.005.
[82] Dan T, Chen H, Li T, Tian J, Ren W, Zhang H, et al. Influence of Lactobacillus plantarum P-8 on fermented milk flavor and storage stability. Front Microbiol 2019:9.
[83] Sidira M, Kandylis P, Kanellaki M, Kourkoutas Y. Effect of immobilized Lactobacillus casei on the evolution of flavor compounds in probiotic dry-

fermented sausages during ripening. Meat Sci 2015;100:41–51. https://doi.org/10.1016/j.meatsci.2014.09.011.

[84] Zagorec M, Champomier-Vergès M-C. Lactobacillus sakei: a starter for sausage fermentation, a protective culture for meat products. Microorganisms 2017:5. https://doi.org/10.3390/microorganisms5030056.

[85] Daniel R, Bobik TA, Gottschalk G. Biochemistry of coenzyme B12-dependent glycerol and diol dehydratases and organization of the encoding genes. FEMS Microbiol Rev 1998;22:553–66. https://doi.org/10.1111/j.1574-6976.1998.tb00387.x.

[86] Nethery MA, Barrangou R. CRISPR visualizer: rapid identification and visualization of CRISPR loci via an automated high-throughput processing pipeline. RNA Biol 2019;16:577–84. https://doi.org/10.1080/15476286.2018.1493332.