İsmail GÜMÜŞTOP

A Master's Thesis

AGU 2023

NEXT GENERATION SEQUENCING OF A NOVEL *Loigolactobacillus coryniformis* FOL-19 ISOLATED FROM CHEESE AND COMPARATIVE GENOMIC ANALYSIS WITH OTHER *L. coryniformis* STRAINS

A THESIS
SUBMITTED TO THE DEPARTMENT OF BIOENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
İsmail GÜMÜŞTOP

June 2023

NEXT GENERATION SEQUENCING OF A NOVEL *Loigolactobacillus coryniformis* FOL-19 ISOLATED FROM CHEESE AND COMPARATIVE GENOMIC ANALYSIS WITH OTHER *L. coryniformis* STRAINS

A THESIS
SUBMITTED TO THE DEPARTMENT OF BIOENGINEERING
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF
ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

By
İsmail GÜMÜŞTOP
June 2023

SCIENTIFIC ETHICS COMPLIANCE

I hereby declare that all information in this document has been obtained in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

İsmail GÜMÜŞTOP

REGULATORY COMPLIANCE

M.Sc. thesis titled "Next Generation Sequencing Of A Novel *Loigolactobacillus Coryniformis* FOL-19 Isolated From Cheese And Comparative Genomic Analysis With Other *L. coryniformis* Strains" has been prepared in accordance with the Thesis Writing Guidelines of the Abdullah Gül University, Graduate School of Engineering & Science.

Prepared By	Advisor
İsmail GÜMÜŞTOP	Asst. Prof. Fatih ORTAKCI

Head of the Bioengineering Program

Asst. Prof. Altan ERCAN

ACCEPTANCE AND APPROVAL

M.Sc. thesis titled "Next Generation Sequencing Of A Novel *Loigolactobacillus Coryniformis* FOL-19 Isolated From Cheese And Comparative Genomic Analysis With Other *L. coryniformis* Strains" and prepared by İsmail GÜMÜŞTOP has been accepted by the jury in the Bioengineering Graduate Program at Abdullah Gül University, Graduate School of Engineering & Science.

05/06/2023

(Thesis Defense Exam Date)

JURY:

Advisor : Asst. Prof. Fatih ORTAKCI ……

Member : Assoc. Prof. Dr. Funda KARBANCIOĞLU GÜLER ……

Member : Assoc. Prof. Dr. Zafer AYDIN ……...

APPROVAL:

The acceptance of this M.Sc. thesis has been approved by the decision of the Abdullah Gül University, Graduate School of Engineering & Science, Executive Board dated ….. /….. / …….. and numbered ……..……. .

…….. /…….. / ……..

(Date)

Graduate School Dean
Prof. Dr. İrfan ALAN

ABSTRACT

NEXT GENERATION SEQUENCING OF A NOVEL *Loigolactobacillus coryniformis* FOL-19 ISOLATED FROM CHEESE AND COMPARATIVE GENOMIC ANALYSIS WITH OTHER *L. coryniformis* STRAINS

İsmail GÜMÜŞTOP
MSc. in Bioengineering
Advisor: Asst. Prof. Fatih ORTAKCI

June 2023

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 using Illumina NextSeq. 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, and dairy was performed. 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 achieved. Both Type I and II Cas clusters were observed in L. coryniformis. Only one strain (CECT 5711) was predicted to encode a Carnocin CP52 bacteriocin gene cluster. The presence of CRISPR elements and Cas clusters suggests that *L. coryniformis* holds a promising potential for being a reservoir for new CRISPR-based tools. 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.

Keywords: L. coryniformis, Comparative genomics, CRISPR/Cas, Bacteriocin, Fermented foods

ÖZET

PEYNİRDEN İLK DEFA İZOLE EDİLEN *Loigolactobacillus coryniformis* FOL-19'UN YENİ NESİL DİZİLENMESİ VE DİĞER *L. coryniformis* SUŞLARIYLA KARŞILAŞTIRMALI GENOMİK ANALİZLERİ

İsmail GÜMÜŞTOP
Biyomühendislik Anabilim Dalı Yüksek Lisans
Tez Yöneticisi: Dr. Öğr. Üyesi Fatih ORTAKCI
Haziran-2023

Loigolactobacillus coryniformis, çeşitli ekolojik nişlerden izole edilen laktik asit bakterilerinin bir üyesidir. Tulum peynirinden yeni bir *L. coryniformis* suşu olan FOL-19'u izole ettik ve Illumina NextSeq kullanarak FOL-19 için tüm genom dizilimi gerçekleştirdik. Ardından, FOL-19'un kimchi, silaj, fermente et, ahır havası ve süt ürünlerinden izole edilen aynı türe ait mevcut on tam genom dizisine karşı genomik karakterizasyonu gerçekleştirilmiştir. Ortalama genom büyüklüğü 2.93 ±0.1 Mb, GC içeriği %42.96 ±0.002, CDS sayısı 2905 ±165, tRNA sayısı 56 ±10 ve CRISPR element sayısı 6.55 ±1.83 olarak elde edilmiştir. *L. coryniformis*'te hem Tip I hem de II Cas kümeleri gözlenmiştir. Sadece bir suşun (CECT 5711) Carnocin CP52 bakteriyosin gen kümesini kodladığı tahmin edilmiştir. CRISPR elementlerinin ve Cas kümelerinin varlığı, *L. coryniformis*'in yeni CRISPR tabanlı araçlar için bir rezervuar olma konusunda umut verici bir potansiyele sahip olduğunu göstermektedir. Bu bulgular, biyoteknolojik uygulamalar için *L. coryniformis* suşlarının genomik karakterizasyonu için, endüstriyel olarak ilgili özellikleri belirlemek üzere genom güdümlü suş seçimi yoluyla bir adım öne çıkmaktadır.

Anahtar kelimeler: L. coryniformis, Karşılaştırmalı genomik, CRISPR/Cas, Bakteriyosin, Fermente gıdalar

Acknowledgements

I would like to acknowledge Dr. Fatih Ortakci for his excellent support and guidance throughout this process. I also thank my family for their continuous support and motivation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANI	Average Nucleotide Identity
BLAST	Basic Local Alignment Search Tool
CARD	Comprehensive Antibiotic Resistance Database
CAZyme	Carbohydrate active enzyme
CDS	Coding sequence
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
GRAS	Generally Recognized As Safe
HAV	Hepatitis A virus
IgA	Immunoglobulin A
IgG	Immunoglobulin G
LAB	Lactic acid bacteria
MRS	De Man, Rogosa, Sharpe
NCBI	National Center for Biotechnology Information
PCoA	Principal Coordiante Analysis

To my family

Chapter 1

INTRODUCTION

Lactic acid bacteria (LAB) historically known as “milk-souring bacteria” were commonly associated with fermentation of food and feed. 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. LAB were acknowledged as beneficial microorganisms [1]. For example, some LAB species have probiotic effects in a wide range of spectra such as potential prevention of allergies [2], improving feed conversion [3], and potential antidiabetic [4]. Whereas some genera of LAB (i.e. Enterococcus, Streptococcus, Carnobacterium, and Lactococcus) contain pathogenic species [1].

LAB generates energy from substrate-level phosphorylation due to the lack of a functional respiratory system. Two major pathway plays a role in the fermentation of hexoses in LAB. Embden-Meyerhof pathway (glycolysis) is the homofermentative or homolactic pathway and only lactate is synthesized as the end product. However, the heterofermentative or heterolactic pathway (6-phosphogluconate pathway or pentose phosphoketolase pathway) produces lactic acid and ethanol as the end product in addition to CO2 and acetate. The theoretical energy yield of the homofermentative pathway is actualized by synthesis of two moles of ATP per mole of glucose catabolized. However, a single mole of ATP is produced when the intermediate product of acetyl phosphate is reduced to ethanol during the heterofermentative pathway whereas an extra mole of ATP is synthesized if acetate was formed from acetyl phosphate when alternative electron acceptors are available [1]. Pentoses enter heterofermentative pathway to metabolize in either xylulose-5-phosphate or ribulose-5-phosphate [5]. However, CO2 is not produced while the fermentation of pentoses.

Pyruvate is a key intermediate of fermentative pathways that supports the maintenance of redox balance by serving as an electron acceptor in the cell. There are multiple fates of pyruvate that vary according to the presence and/or availability of

substrate and oxygen. One of the most common fates of pyruvate in LAB leads to acetoin and diacetyl. This pathway plays an important role in the dairy industry and happens when there is an extra amount of pyruvate in the cell. Pyruvate is converted to either α-acetolactate by acetolactate synthase or acetylphosphate by oxidizing with pyruvate oxidase. α-acetolactate can be oxidized to yield diacetyl, an important metabolite in the food industry for flavor formation, or decarboxylated acetolactate decarboxylase to yield acetoin [6]. Pyruvate formate-lyase pathway is resorted when the cell is under anaerobic conditions and substrate levels are limited to synthesize acetyl-CoA and formic acid catalyzed by pyruvate formate-lyase enzyme [5, 7]. The acetyl-CoA can either yield to ethanol by playing a critical role as an electron acceptor or acetate as the end product and one mole of ATP by substrate level phosphorylation. The fate of pyruvate leads to lactate in LAB when the abundancy of carbohydrates that are used for fermentation is well under both aerobic and anaerobic conditions. Pyruvate is converted to lactate by receiving an electron from NADH with the help of lactate dehydrogenase [6].

As of 2020, the genus Lactobacillus was segregated into 26 genera by conserved phenotypes and clade-specific signature genes [8]. The genera of Loigolactobacillus (L.) are known for the spoilage potential of fermented foods and drinks. Loigos means havoc, destruction, and ruin in Greek [8]. 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 [8–11]. Currently, eight distinct species of Logiolactobacillus 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* [12]. *Loigolactobacillus bifermentans* is associated with cracks in Dutch-style cheese by fermenting lactic acid into carbon dioxide, ethanol, acetic acid, and hydrogen. *Loigolactobacillus rennini*, originating from the rennet, causes cheese spoilage [10]. *Loigolactobacillus 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 [9].

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 [8]. The Si3 strain of the subsp. coryniformis showed antifungal activity against spoilage yeast on silage [10]. 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 [13]. 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 [14].

Chapter 2

MATERIALS AND METHODS

2.1 Bacterial Isolation

The six-month-aged artisanal Tulum cheese sample, acquired from 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} in the same solution. 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 purposes. 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 extraction and purification of DNA, 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 149R 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 minutes, followed by 30 cycles of denaturation at 94 °C for 30 seconds, annealing at 52 °C for 20 seconds, extension at 72 °C for 90 seconds, and a final

extension at 72 °C for 5 minutes. The amplified DNA fragments were visualized on 1.5% agarose gel and sent for Sanger sequencing. The sequence of DNA fragment was analyzed using Basic Local Alignment Search Tool (BLAST) [15] in NCBI databases to accurately identify the bacterial isolate.

The 16S rDNA gene was sequenced to identify the isolate. Universal primers of 16S rRNA 27F and 1492R were used for nucleotide sequencing. To amplify the PCR reaction mixture, three units of EasyTaq® DNA polymerase, 20 µM of 27F primer, 20 µM of 149R 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 were employed. The PCR reaction mix was amplified under specific conditions, including an initial denaturation of DNA at 94 °C for 5 minutes, followed by 30 cycles of denaturation at 94 °C for 30 seconds, annealing at 52 °C for 20 seconds, extension at 72 °C for 90 seconds, and a final extension at 72 ◦C for 5 minutes. The PCR products were visualized on agarose gel (1.5 %) and sent for Sanger sequencing [16]. The Sanger DNA fragment reads was then analyzed using the Basic Local Alignment Search Tool (BLAST) against previously published DNA sequences in NCBI to identify the isolate.

2.3 Whole-genome Sequencing and Assembly

After the identification of 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 [17]. 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. These steps helped to improve the overall quality of the assembly, 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 [18] with the accession number of 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 [18] with the following accession numbers 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).

Genome quality and completeness of the strains was assessed using BUSCO [19, 20] and CheckM [21]. Prokka software (version 1.14.6) [22] was utilized to perform the annotation of the whole-genome sequences of *L. coryniformis* strains, with the following flag: --kingdom Bacteria. The output of GFF files was then analyzed using Roary (version 3.13.0) [23] 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. To ensure accuracy, the minimum BLASTp identity threshold was set to 95% in Roary. To investigate whether the pangenome of *L. coryniformis* is open, the micropan package [24] utilized 10,000 permutations to fit the Heap's law model. These methods enabled us to gain insights into the genetic makeup of L. coryniformis and the variability of its pangenome.

2.5 Comparative Genomics of *L. coryniformis*

The assessment of genome similarity was carried out by computing Jaccard distance using the prabclus package [25] based on the presence or absence of putative genes. The resulting data was subjected to Principal Coordinate Analysis (PCoA) using the R (version 4.1.1) [26, 27] to evaluate the relationship between genomes. The alignment of the core genome was performed by using FastTree (version 2.1.1) [28]. FastTree employed the Shimodaira-Hasegawa (SH) test with 1000 bootstrap replicates to calculate the reliability of each split in the [29]. The construction of the phylogenetic tree of the core genome alignment was carried out with iTOL [30].

The genomes of *L. coryniformis* were subjected to cluster analysis and phylogenetic tree construction using the TYGS platform with default settings [31], which can be accessed at https://tygs.dsmz.de/. FastME [32] was used as the distance method for constructing the phylogenetic tree. Additionally, a phylogenetic tree of 16S rDNA sequences was constructed using TYGS, which identifies 16S rDNA sequences from closely related type strains using the RNAmmer tool [33] and BLAST+ [15] against each available strain in the TYGS database. Orthologous average nucleotide identity (CDS ANI) was calculated using the GET_HOMOLOGUES [34]. Genomic islands were identified using GIPSy [35] 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% [15, 36].

The CRISPRCasFinder web tool located at https://crisprcas.i2bc.paris-saclay.fr/ [37] was used to identify Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) elements and Cas enzyme clusters in the genomes of *L. coryniformis*. 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) [34], [35] was employed to annotate CAZyme domains. The results of CAZyme annotation were filtered according to suggested 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) [40]. 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 [41]. Plasmid sequences in whole-genome sequences of the *L. coryniformis* strains were identified using the PLSDB web tool [42, 43]. Furthermore, insertion sequences (IS) in the genomes were detected by utilizing the ISfinder web tool, available at https://isfinder.biotoul.fr/ [44].

Chapter 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 3.1. We also identified two plasmids with a length of 0.809 and 1.326 Kb (Table A.5). Annotation of the genome using Prokka yielded 2769 coding sequences, 52 tRNA, and two rRNA. Moreover, 8 CRISPR regions were also detected. Several annotated IS elements and genomic islands were also predicted. Given the increasing global concern over antibiotic resistance, we employed the Comprehensive Antibiotic Resistance Database (CARD) to screen the *L. coryniformis* FOL-19 genome for the presence of antibiotic resistance genes. However, no genes associated with the Generally Recognized As Safe (GRAS) status of this species were found.

3.2 Genetic Diversity of *L. coryniformis*

The genomic features of eleven *L. coryniformis* strains with genome sizes ranging between 2.82 Mb to 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% to 43.51% (average 42.96%). The number of tRNA genes identified in each strain falls

between 32 to 70, while the number of protein-coding sequences (CDS) ranges between 2735 to 3299 (average 2905) (Table 3.1).

Having access to complete genome sequences of *L. coryniformis* FOL-19 and type strain DSM 20001, we determined how FOL-19 compares with other *L. coryniformis* strains and performed comparative genomics. Eleven strains, including FOL-19, were chosen for comparative analysis using the 16S rDNA (Figure 3.1), CDS ANI (Figure A.2), and whole-genome sequence (Figure A.3). According to the 16S rDNA sequence-based phylogenetic tree, *L. coryniformis* strains split into two main clades. The first clade consists of strains isolated from plant-based environments such as silage (CRL 1001 and DSM 20001) and kimchi (KCTC 3167 and CBA 3616). Interestingly, isolation sources of strains in the second clade are a variety of non-plant-based niches such as fermented meat (14I and 42L), cheese (FOL-19 and CECT 5711), pheasant chyme (PH-1), air of cow shed (DSM 20004) except kimchi (KCTC 3535). OrthoANI results according to CDS show that the highest identity was achieved at 99.88% between KCTC 3167 and DSM 20001 which were isolated from kimchi and silage, respectively. Likewise, KCTC 3167 showed the second highest similarity with another silage isolate of CRL 1001. Moreover, fermented meat isolates of 14I and 42L revealed the highest similarity between each other.

The whole-genome sequence-based phylogenetic tree revealed that FOL-19 separated from other *L. coryniformis* strains that are 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 which was the first clade was consist of strains (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.

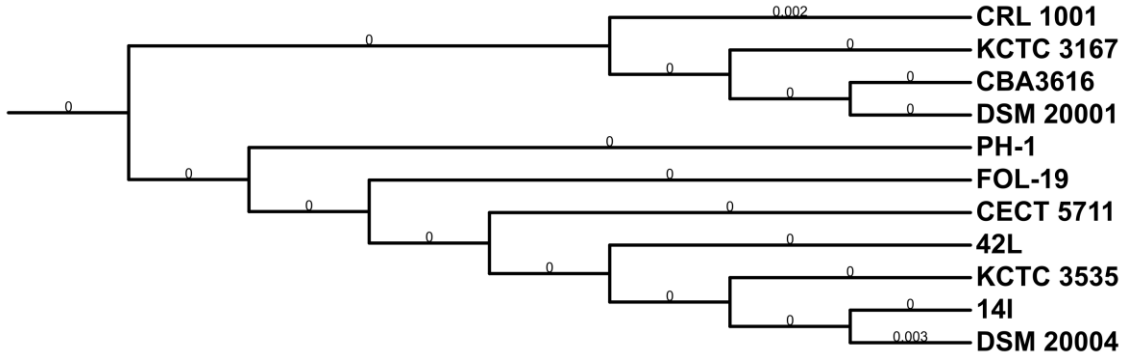

Figure 3.1 16S rDNA sequence-based cladogram of eleven *L. coryniformis* strains

Table 3.1 Whole-genome sequence statistics of eleven *L. coryniformis* strains

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

Assembly Accession
GCA_000166795.1
GCA_000184285.2
GCA_000283115.1
GCA_001742375.1
GCA_002706425.1
GCA_002706705.1
GCA_007954685.1
GCA_019390135.1
GCA_019390175.1
GCA_023483865.1
GCA_028439555.1

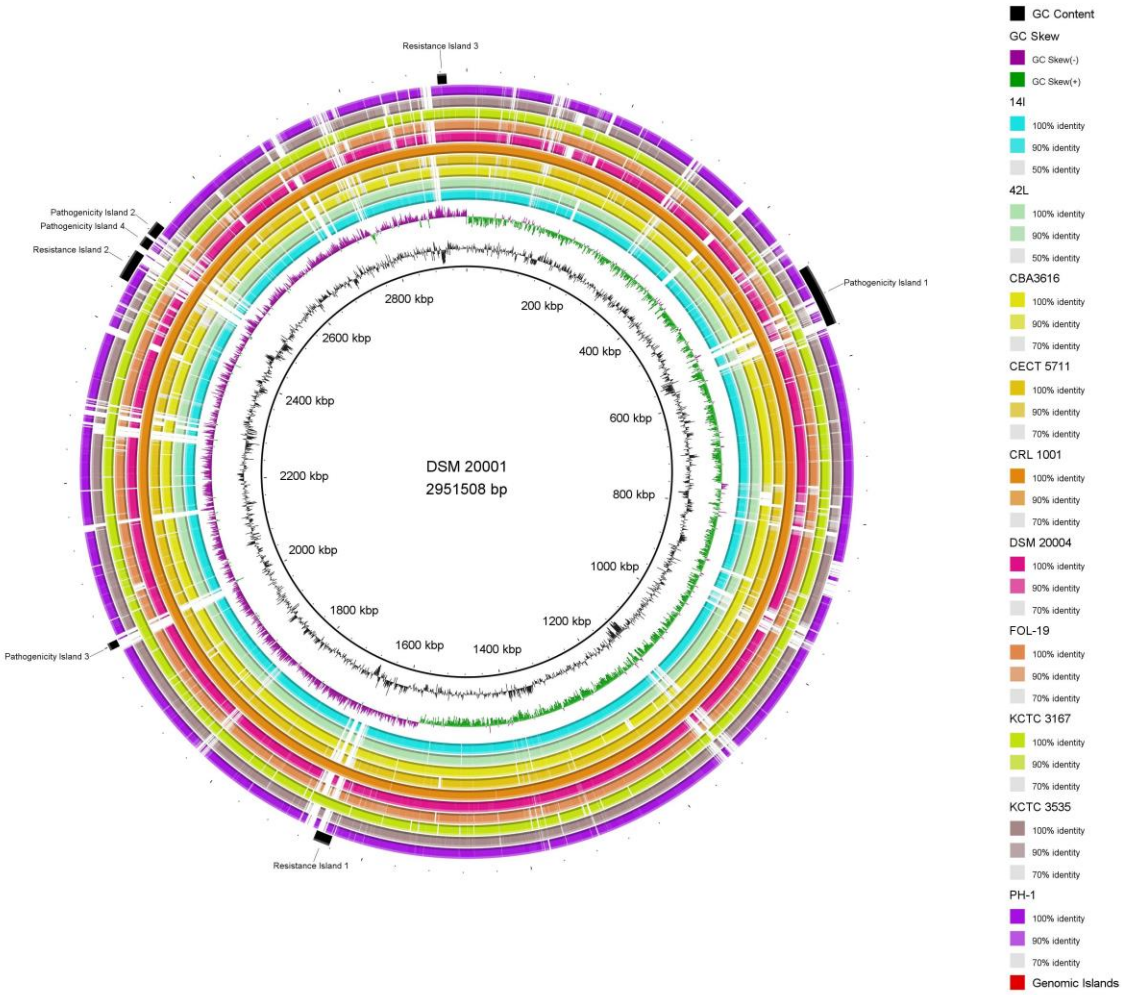

Figure 3.2 Whole-genome-based BLAST comparison of ten *L. coryniformis* strains against reference strain DSM 20001. The innermost rings show GC Content (black) and GC Skew (purple-green). The remaining circles show BLAST comparisons of ten other complete *L. coryniformis* genomes against the reference genome DSM 20001. The outermost rings highlight genomic islands.

BRIG image shows the alignment of ten *L. coryniformis* strains and their GC content and GC skews against the reference genome DSM 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 2.00Mb, and a group of genomic islands of resistance island 2, pathogenicity islands 2 and 4 between 2.44 and 2.55 Mb.

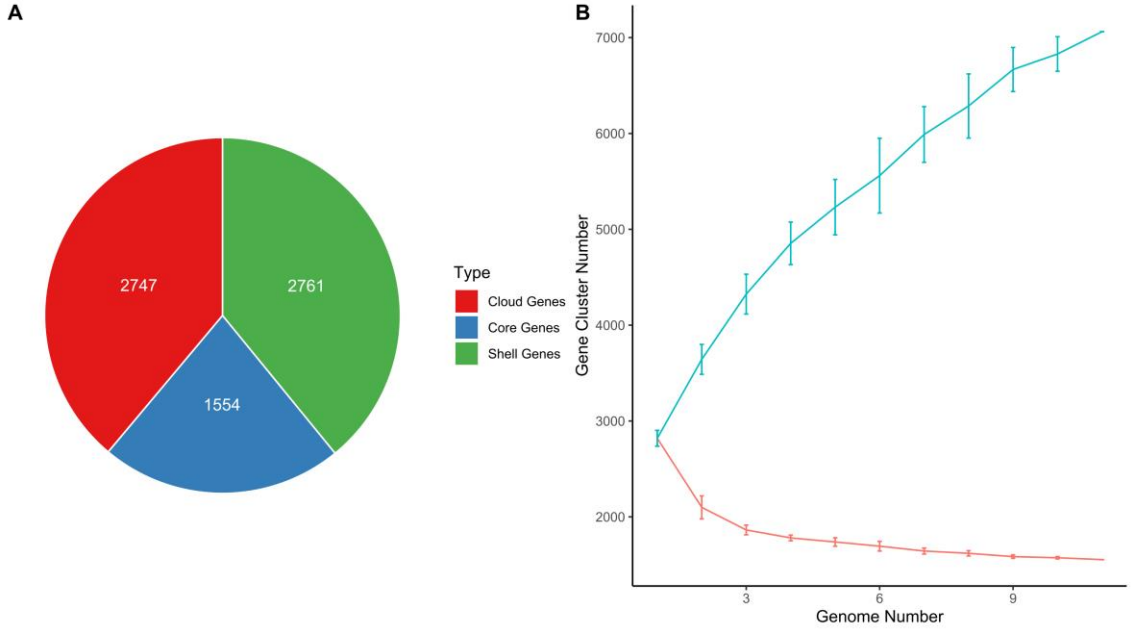

Figure 3.3 (A) Coding sequence distributions in the eleven *L. coryniformis* pan-genome. Cloud genes (red), core genes (blue), and shell genes (green). (B) Estimation of core- (red line) and pan-genomes (blue line) of the eleven *L. coryniformis* strains by including genomes one by one.

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 (Figure 3.3A). Of the 7062 total CDS, 1554 were shared by all eleven strains which is the core genome. The accessory genome, also called the non-core genome, contained 5508 total CDS, perhaps determining fundamental differences in phenotypic traits across different strains, as reported by [45]. We performed randomized subsampling to the strain order for visualizing the trendlines of core- and pangenomes (Figure 3.B). The pangenome size doesn'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 [46].

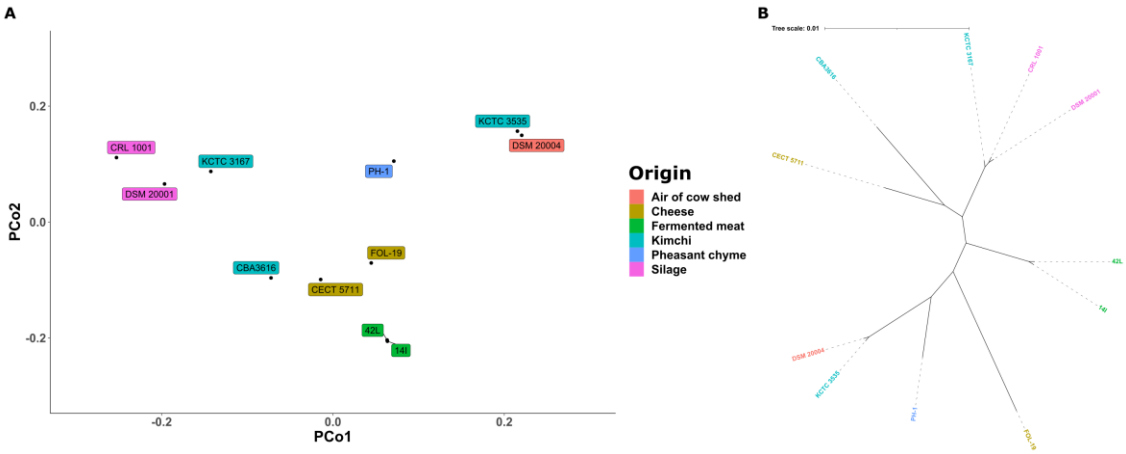

Figure 3.4 (A) PCoA visualization of Jaccard distances based on shared genes across eleven *L. coryniformis* genomes screened. The color of each box indicates a unique isolation source. (B) Neighbor-joining unrooted phylogenetic tree based on core genome alignment. Each font color indicates a unique isolation source.

PCoA across genomes based on Jaccard distance among presence/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 very close 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 each other except kimchi isolates of KCTC 3167 and CBA3616 are separated compared to other strain pairs and KCTC 3535 showed the highest similarity against DSM 20004 which was isolated from air of cow shed (Figure 3.4A). The phylogenetic neighborhood across eleven strains was calculated based on relative hierarchical clustering via core genome alignment (Figure 3.4B). 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.

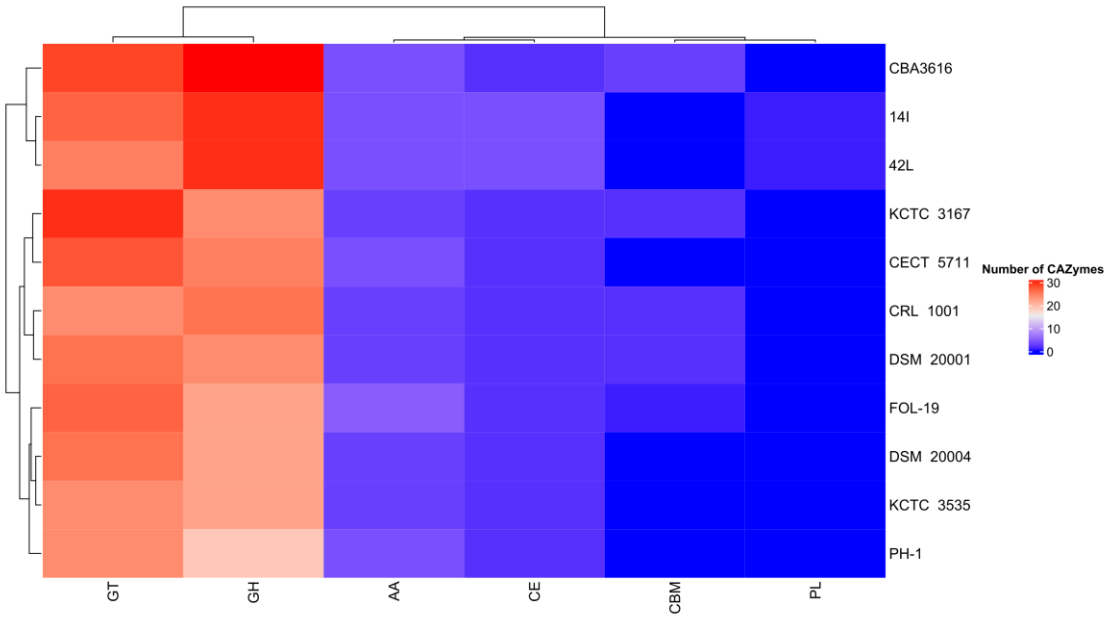

Figure 3.5 Heatmap showing the distribution of CAZyme families across the eleven *L. coryniformis* genomes. Color gradients represent the number of CAZymes. The number of CAZymes is increasing from lighter to darker colors. AA: Auxiliary activities, CE: Carbohydrate esterase family, CMB: Carbohydrate-binding module family, GH: Glycoside hydrolase, GT: Glycosyltransferase, PL: Polysaccharide lyase.

Heatmap representation of CAZyme families shows that GH and GT family CAZymes are the most abundant ones 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 (Figure 3.5). Three major clades were generated based on the distribution of CAZymes in the genome of *L. coryniformis* strains. From the bottom-up in the heatmap shown in Figure 3.5, the first clade was formed by the members of the non-plant-originated strains 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 20001 except CECT 5711 which was isolated from cheese. The last clade members were isolated from fermented meat except CBA 3616 which originated from kimchi.

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 [40]. 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 except for CECT 5711, which was predicted to carry Carnocin CP52 bacteriocin at 109 aa length and 12.4 kDa in size.

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 rest of the strains had either one or two putative plasmids (Table A.5). 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 A.6). PH-1, CECT 5711, and FOL-19 have the highest numbers of IS elements. All strains harbored IS elements from *Lactobacillus plantarum* which is 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*.

All strains except PH-1 carry at least one Cas cluster. Only CBA316 was predicted to carry both Type I and Type II Cas cluster. The rest of the strains carry either Type I (14I, 42L, and CECT 5711) or Type II (CRL 1001, DSM 20001, FOL-19, KCTC 3167, DSM 20004, and KCTC 3535) Cas clusters in their genome (Table 3.1).

Table 3.2 CRISPR elements and Cas clusters

Strain	Element	CRISPR Id / Cas Type	Start	End	Spacer / Gene	Direction	Evidence Level
14I	Cas cluster	CAS-TypeIC	1701891	1717783	7		
14I	Cas cluster	CAS-TypeIE	2840356	2849746	7		
14I	CRISPR	14I_1	211832	211990	1	ND	1
14I	CRISPR	14I_2	1115855	1116007	1	ND	1
14I	CRISPR	14I_3	1701254	1701551	4	ND	4
14I	CRISPR	14I_4	1709501	1715447	89	ND	4
14I	CRISPR	14I_5	2196094	2196238	1	ND	1

14I	CRISPR	14I_6	2674299	2674530	3	ND	1
14I	CRISPR	14I_7	2823017	2823158	1	ND	1
14I	CRISPR	14I_8	2837487	2840326	46	ND	4
14I	CRISPR	14I_9	2866218	2866365	1	ND	1
42L	Cas cluster	CAS	1848308	1850266	3		
42L	Cas cluster	CAS-TypeIC	2149848	2155978	4		
42L	Cas cluster	CAS-TypeIE	193047	202437	7		
42L	CRISPR	42L_1	175708	175849	1	ND	1
42L	CRISPR	42L_2	190178	193017	46	ND	4
42L	CRISPR	42L_3	218909	219056	1	ND	1
42L	CRISPR	42L_4	379065	379223	1	ND	1
42L	CRISPR	42L_5	1322061	1322213	1	ND	1
42L	CRISPR	42L_6	2158746	2164692	89	ND	4
42L	CRISPR	42L_7	2357721	2357865	1	ND	1
42L	CRISPR	42L_8	2689296	2689527	3	ND	1
CBA3616	Cas cluster	CAS-TypeIC	118445	129647	7		
CBA3616	Cas cluster	CAS-TypeIE	1898208	1916583	8		
CBA3616	Cas cluster	CAS-TypeIIU	724936	729522	3		
CBA3616	CRISPR	CBA3616_1	120780	123399	39	ND	4
CBA3616	CRISPR	CBA3616_2	129986	134275	64	ND	4
CBA3616	CRISPR	CBA3616_3	241833	241984	1	ND	1
CBA3616	CRISPR	CBA3616_4	719581	724883	80	-	4
CBA3616	CRISPR	CBA3616_5	729665	731810	32	-	4
CBA3616	CRISPR	CBA3616_6	1874593	1874737	1	ND	1
CBA3616	CRISPR	CBA3616_7	1907630	1910227	42	ND	4
CECT 5711	Cas cluster	CAS	909481	911439	3		
CECT 5711	Cas cluster	CAS-TypeIC	2518092	2524222	4		
CECT 5711	Cas cluster	CAS-TypeIE	2661194	2665338	5		
CECT 5711	CRISPR	CECT 5711_1	1282702	1282849	1	ND	1
CECT 5711	CRISPR	CECT 5711_2	1438858	1439439	9	ND	4
CECT 5711	CRISPR	CECT 5711_3	1937506	1937664	1	ND	1
CECT 5711	CRISPR	CECT 5711_4	2516815	2517974	17	ND	4
CECT 5711	CRISPR	CECT 5711_5	2665367	2665638	4	ND	4
CRL 1001	Cas cluster	CAS-TypeIIA	2623541	2629268	4		
CRL 1001	CRISPR	CRL 1001_1	76827	76938	1	ND	1
CRL 1001	CRISPR	CRL 1001_2	178526	178676	1	ND	1
CRL 1001	CRISPR	CRL 1001_3	320951	321109	1	ND	1
CRL 1001	CRISPR	CRL 1001_4	2281480	2281624	1	ND	1
CRL 1001	CRISPR	CRL 1001_5	2621564	2623513	29	ND	4
CRL 1001	CRISPR	CRL 1001_6	2764705	2764857	1	ND	1
CRL 1001	CRISPR	CRL 1001_7	2820093	2822042	29	ND	4
DSM 20001	Cas cluster	CAS-TypeIIA	2701667	2707781	4		
DSM 20001	CRISPR	DSM 20001_1	670598	670749	1	ND	1

DSM 20001	CRISPR	DSM 20001_2	724334	724445	1	ND	1
DSM 20001	CRISPR	DSM 20001_3	2185058	2185216	1	ND	1
DSM 20001	CRISPR	DSM 20001_4	2699624	2701639	30	ND	4
DSM 20004	Cas cluster	CAS	778715	780673	3		
DSM 20004	Cas cluster	CAS-TypeIIU	171570	176150	3		
DSM 20004	CRISPR	DSM 20004_1	169018	171493	37	+	4
DSM 20004	CRISPR	DSM 20004_2	176203	179276	46	+	4
DSM 20004	CRISPR	DSM 20004_3	652868	653009	1	ND	1
DSM 20004	CRISPR	DSM 20004_4	778173	778338	2	ND	1
DSM 20004	CRISPR	DSM 20004_5	996379	996522	1	ND	1
DSM 20004	CRISPR	DSM 20004_6	1659706	1659866	1	ND	1
DSM 20004	CRISPR	DSM 20004_7	2028421	2028558	1	ND	1
DSM 20004	CRISPR	DSM 20004_8	2230579	2230737	1	ND	1
FOL-19	Cas cluster	CAS	2715457	2717415	3		
FOL-19	Cas cluster	CAS-TypeIIA	1192332	1198455	4		
FOL-19	CRISPR	FOL-19_1	166369	166508	1	ND	1
FOL-19	CRISPR	FOL-19_2	1198483	1199178	10	ND	4
FOL-19	CRISPR	FOL-19_3	1388395	1388542	1	ND	1
FOL-19	CRISPR	FOL-19_4	1694861	1695004	1	ND	1
FOL-19	CRISPR	FOL-19_5	2024432	2025459	15	-	4
FOL-19	CRISPR	FOL-19_6	2066246	2070305	61	+	4
FOL-19	CRISPR	FOL-19_7	2151798	2151939	1	ND	1
FOL-19	CRISPR	FOL-19_8	2617275	2617433	1	ND	1
KCTC 3167	Cas cluster	CAS-TypeIIA	301940	308054	4		
KCTC 3167	CRISPR	KCTC 3167_1	99000	99158	1	ND	1
KCTC 3167	CRISPR	KCTC 3167_2	299897	301912	30	ND	4
KCTC 3167	CRISPR	KCTC 3167_3	1532887	1534376	22	+	4
KCTC 3167	CRISPR	KCTC 3167_4	1631037	1631188	1	ND	1
KCTC 3535	Cas cluster	CAS	1479538	1481496	3		
KCTC 3535	Cas cluster	CAS-TypeIIU	643542	648122	3		
KCTC 3535	CRISPR	KCTC 3535_1	255008	255164	1	ND	1
KCTC 3535	CRISPR	KCTC 3535_2	379560	379718	1	ND	1
KCTC 3535	CRISPR	KCTC 3535_3	449110	449251	1	ND	1
KCTC 3535	CRISPR	KCTC 3535_4	640416	643489	46	-	4
KCTC 3535	CRISPR	KCTC 3535_5	648199	650674	37	-	4
KCTC 3535	CRISPR	KCTC 3535_6	854596	854733	1	ND	1

KCTC 3535	CRISPR	KCTC 3535_7	1478996	1479161	2	ND	1
KCTC 3535	CRISPR	KCTC 3535_8	1908915	1909058	1	ND	1
PH-1	CRISPR	PH-1_1	1178789	1178945	1	ND	1
PH-1	CRISPR	PH-1_2	1183241	1183394	1	ND	1
PH-1	CRISPR	PH-1_3	1896434	1896576	1	ND	1
PH-1	CRISPR	PH-1_4	2892383	2892541	1	ND	1

Chapter 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* overall is 42.96%, which is typical for the low GC lactobacilli [47]. This finding suggests that *L. coryniformis* has experienced 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 [48]. 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 (Figure 3.1). Phylogenetic analysis showed strain diversity among *L. coryniformis* strains tested due to unclear grouping of strains by isolation source. Two main clades were identified, with *L. coryniformis* FOL-19 laid to the second member clade containing PH-1, CECT 5711, 42L, KCTC 3535, 14I, and DSM 20004 (Fig 3.2.1). Although *L. coryniformis* FOL-19 was isolated from cheese, its clade members were isolated from fermented meat (14I and 42L), air of cowshed (DSM 20004), kimchi (KCTC 3535), and cheese (CECT 5711). We understand that CECT 5711 (cheese isolate) falls in a similar clade with FOL-19 due to sharing similar isolation sources. However, PH-1, pheasant chyme isolate, is the closest match to FOL-19. The 14I and 42L (fermented meat) also share the same clade with FOL-19. We would anticipate that related strains would have similar isolation sources [49]. Since this is mostly not the case for *L. coryniformis* FOL-19, we 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.

Genomic comparison indicated that eleven strains of *L. coryniformis* are a compact group of bacteria. The sequence similarity results were higher than the sequence similarity threshold of 95% (CDS ANI) for species demarcation even though their pangenome was open. Fermented meat isolates 14I and 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 [26]. For example, KCTC 3535 and CBA 3616 do not have any common plasmids and KCTC 3167 has no plasmids even three of which were have the same isolation sources.

Identification of putative carbohydrate metabolism associated genes revealed the sugar metabolism capability *L. coryniformis* in a comparative manner. A bacterial strain’s sugar fermentation capability is a key indicator of strain metabolic function and sets fundamentals for strain selection and cultivation [50]. The presence of phosphoketolase and fructose bisphosphate aldolase genes across all *L. coryniformis* genomes indicated a facultatively heterofermentative carbohydrate metabolism of this species (Table A.2) [1]. 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 [51]. Several types of GTs participate in disaccharide, oligosaccharide, and polysaccharide biosynthesis, which are instrumental in forming glycosidic bonds [52]. 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 possesses a critical role in the carbohydrate glycosidic bond hydrolysis [52]. 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 [52, 53]. The prevalence of lactose intolerance was estimated at around 67% globally [54, 55] due to a lack of β-galactosidase, which hydrolyzes lactose into glucose and galactose. Hydrolysis of lactose also develops the texture of milk products [54]. Among the *L. coryniformis* strains tested, only FOL-19 is the only *L. coryniformis* strain that harbors the *lacZ* gene encoding the β-galactosidase enzyme. 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 [56] 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 [57–59].

The *oppABCDF* operon was found in all *L. coryniformis* strains tested; however, PII-type serine proteinase that is functional against caseins was absent across all *L. coryniformis* strains studied [60]. *pepX* gene encoding x-prolyl dipeptidyl aminopeptidase was complete in FOL-19; however, it was truncated in several LAB strains [60]. *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 do not carry aminopeptidase pepS. Like *pepE* and *pepT* genes, the *pepS* gene was also truncated in silage isolate CRL 1001 (Table A4). *pepV* gene encoding β-ala-dipeptidase, which is known to cut dipeptides by N-terminal D-alanine or β-alanine residue [60, 61], 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 [48, 62].

Horizontal gene transfer (HGT) is the primary factor in bacterial evolution that could bestow fitness and niche adaptation [63].To identify genomic islands, we compared ten *L. coryniformis* genomes against the reference genome DSM 20001 in BRIG, which

resulted in seven genomic islands (Figure 3.2). *L. coryniformis* FOL-19 was predicted to carry several strong metabolic islands which are 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 [49], 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 (Table 3.2). 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 [64]. 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 [65]. Type II in *L. coryniformis* strains was higher than lactobacilli in general [64]. 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 [66].

Among all *L. coryniformis* strains tested, the most common IS element shared across all strains were closely related to *Lactobacillus plantarum*, which is heavily utilized as a probiotic dietary supplement, starter culture in plant fermentations, and bio-protective culture against food-borne pathogens due to its antimicrobial activity by producing bacteriocins [67, 68]. Cheese-originated *L. coryniformis* strains share IS elements with *Lactobacillus 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 [69]. Moreover, they share IS elements with *Lactobacillus plantarum*, an adjunct culture to produce long-shelf-life cheese and to enhance the flavor of fermented milk products [70, 71]. Similarly, *L. coryniformis* strains isolated from fermented meat share IS elements with *Lactobacillus casei*, which can synthesize volatile compounds to enrich the flavor of probiotic food products. Another mutual IS element

belonging to *Lactobacillus sakei* possesses antifungal and antipathogenic activity, thus functional in preserving fresh and fermented food products [72, 73].

None of the strains studied in the present work contained putative bacteriocin genes, except CECT 5711, which encodes Carnocin CP52 bacteriocin. A previous study reported the antagonistic activity of Carnocin CP52 against Listeria monocytogenes, known for food spoilage and poisoning, which seriously threatens human health worldwide. *L. monocytogenes* is a psychotropic organism that could proliferate at colder temperatures where common mesophilic starter cultures could not [74]. The ability to encode Carnocin CP52 implies that CECT 5711 might be a potential bioprotective culture candidate for cheese manufacturing and ripening.

Chapter 5

Conclusions and Future Prospects

5.1 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 phenotypically diverse 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. Carnocin CP52 bacteriocin encoding gene was only found in K8 CECT 5711, a known probiotic strain. Only *L. coryniformis* FOL-19 and DSM 20004 harbor one 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. 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.

5.2 Societal Impact and Contribution to Global Sustainability

The biotechnological potential of *Loigolactobacillus coryniformis* FOL-19 strain, which we isolated and identified for the first time from cheese, was revealed in the light of next-generation sequencing. AGU Research Focus Areas Article 2 Health and Medical as this unique organism contains CRISPR region and produces antimicrobial

agent. It is considered that it will contribute to the field of biotechnology and Article 6 Innovation and Entrepreneurship in the medium and long term. As a matter of fact, *L. coryniformis* FOL-19, which is rich in CRISPR, is thought to lead future genome-driven strain selection studies and is thought to contribute to the CRISPR toolkit in the biotechnology industry. In addition, it is evaluated that our microorganism, which we discovered as a result of next-generation sequencing, will contribute to the innovation and entrepreneurship development goal in the context of its potential to produce RiPP (ribosomally synthesized and post-translationally modified peptide product). That is, our organism, which is capable of producing natural antimicrobials in peptide structure, touches innovation and entrepreneurship by laying paving stones in the production of new generation natural antimicrobials, in addition to its contribution to the field of health and medical biotechnology, especially in the search for antibiotic resistance in the world. Therefore, the subject of the master's thesis titled "Next Generation Sequencing Of A Novel *Loigolactobacillus coryniformis* FOL-19 Isolated From Cheese And Comparative Genomic Analysis With Other *L. coryniformis* Strains" is Article 2 and it is evaluated that it will contribute to the AGU Research Focus Areas and the UN Sustainable Development Goals specified in Article 6.

5.3 Future Prospects

Future work should evaluate the in-vitro CRISPR and bacteriocin encoding potential of *L. coryniformis* FOL-19. In particular, the characterization of undefined bacteriocins in terms of size, class, and antagonistic properties should be well studied to further utilize this organism in food protection applications against certain food pathogens. Moreover, the elaboration of the secondary starter culture potential of *L. coryniformis* FOL-19 should be tapped for understanding its contribution to flavor development in ripening cheese.

BIBLIOGRAPHY

[1] von Wright A, Axelsson L. Lactic Acid Bacteria An Introduction. In (Eds: Vinderola G, Ouwehand AC, Salminen S, and Wright A von). Lactic Acid Bacteria: Microbiological and Functional Aspects. CRC Press: Boca Raton (2019).

[2] Żukiewicz-Sobczak W, Wróblewska P, Adamczuk P, Silny W. Probiotic lactic acid bacteria and their potential in the prevention and treatment of allergic diseases. Central European Journal of Immunology. 39(1): 104–108 (2014).

[3] 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. Journal of the Science of Food and Agriculture. 94(2): 341–348 (2014).

[4] 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. Journal of Dairy Science. 102(2): 997–1006 (2019).

[5] Kandler O. Carbohydrate metabolism in lactic acid bacteria. Antonie van Leeuwenhoek. 49(3): 209–224 (1983).

[6] Gänzle MG. Lactic metabolism revisited: metabolism of lactic acid bacteria in food fermentations and food spoilage. Current Opinion in Food Science. 2: 106–117 (2015).

[7] Thomas TD, Ellwood DC, Longyear VMC. Change from Homo- to Heterolactic Fermentation by *Streptococcus lactis* Resulting from Glucose Limitation in Anaerobic Chemostat Cultures. Journal of Bacteriology. 138(1): 109–117 (1979).

[8] 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, (2020).

[9] Geissler Andreas J., Behr Jürgen, Vogel Rudi F. Multiple Genome Sequences of the Important Beer-Spoiling Species Lactobacillus backii. Genome Announcements. 4(4): e00826-16 (2016).

[10] McMahon DJ, Bowen IB, Green I, Domek M, Oberg CJ. Growth and survival characteristics of Paucilactobacillus wasatchensis WDCO4. Journal of Dairy Science. 103(10): 8771–8781 (2020).

[11] 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. 9(2): 274 (2021).

[12] Loigolactobacillus, Available from: https://www.ncbi.nlm.nih.gov/data-hub/taxonomy/2767889/.

[13] 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. 14(1): 228 (2022).

[14] 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. Nutrition & Metabolism. 14(1): 2 (2017).

[15] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, *et al.* BLAST+: architecture and applications. BMC Bioinformatics. 10: 421 (2009).

[16] Sanger F., Nicklen S., Coulson A. R. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences. 74(12): 5463–5467 (1977).
[17] 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 Research. 48(D1): D606–D612 (2020).
[18] Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res. 44(D1): D67–D72 (2016).
[19] Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: Assessing Genomic Data Quality and Beyond. Current Protocols. 1(12) (2021).
[20] Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution. 38(10): 4647–4654 (2021).
[21] Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25(7): 1043–1055 (2015).
[22] Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30(14): 2068–2069 (2014).
[23] Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, *et al.* Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 31(22): 3691–3693 (2015).
[24] Snipen L, Liland KH. micropan: Microbial Pan-Genome Analysis. (2020).
[25] Hennig C, Hausdorf B. prabclus: Functions for Clustering and Testing of Presence-Absence, Abundance and Multilocus Genetic Data. (2020).
[26] Candeliere F, Raimondi S, Spampinato G, Tay MYF, Amaretti A, Schlundt J, *et al.* Comparative Genomics of Leuconostoc carnosum. Frontiers in Microbiology. 11 (2021).
[27] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing: Vienna, Austria (2021).
[28] Price MN, Dehal PS, Arkin AP. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Molecular Biology and Evolution. 26(7): 1641–1650 (2009).
[29] Verce M, De Vuyst L, Weckx S. Comparative genomics of Lactobacillus fermentum suggests a free-living lifestyle of this lactic acid bacterial species. Food Microbiology. 89: 103448 (2020).
[30] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research. 49(W1): W293–W296 (2021).
[31] Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun. 10: 2182 (2019).
[32] Lefort V, Desper R, Gascuel O. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program: Table 1. Mol Biol Evol. 32(10): 2798–2800 (2015).
[33] Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research. 35(9): 3100–3108 (2007).
[34] Contreras-Moreira B, Vinuesa P. GET_HOMOLOGUES, a Versatile Software Package for Scalable and Robust Microbial Pangenome Analysis. Applied and Environmental Microbiology. 79(24): 7696–7701 (2013).

[35] Soares SC, Geyik H, Ramos RTJ, de Sá PHCG, Barbosa EGV, Baumbach J, *et al.* GIPSy: Genomic island prediction software. Journal of Biotechnology. 232: 2–11 (2016).
[36] Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 12(1): 402 (2011).
[37] 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 Research. 46(W1): W246–W251 (2018).
[38] Eddy SR. Accelerated Profile HMM Searches. PLOS Computational Biology. 7(10): e1002195 (2011).
[39] Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, *et al.* dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research. 46(W1): W95–W101 (2018).
[40] 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. 48(D1): D517–D525 (2020).
[41] NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research. 44(D1): D7–D19 (2016).
[42] Galata V, Fehlmann T, Backes C, Keller A. PLSDB: a resource of complete bacterial plasmids. Nucleic Acids Research. 47(D1): D195–D202 (2019).
[43] Schmartz GP, Hartung A, Hirsch P, Kern F, Fehlmann T, Müller R, *et al.* PLSDB: advancing a comprehensive database of bacterial plasmids. Nucleic Acids Research. 50(D1): D273–D278 (2022).
[44] Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Research. 34(suppl_1): D32–D36 (2006).
[45] Daughtry KV, Johanningsmeier SD, Sanozky-Dawes R, Klaenhammer TR, Barrangou R. Phenotypic and genotypic diversity of Lactobacillus buchneri strains isolated from spoiled, fermented cucumber. International Journal of Food Microbiology. 280: 46–56 (2018).
[46] Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. Current Opinion in Microbiology. 11(5): 472–477 (2008).
[47] Brandt K, Barrangou R. Using glycolysis enzyme sequences to inform Lactobacillus phylogeny, https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000187, (2018).
[48] 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 U S A. 103(42): 15611–15616 (2006).
[49] Brandt K, Nethery MA, O'Flaherty S, Barrangou R. Genomic characterization of Lactobacillus fermentum DSM 20052. BMC Genomics. 21(1): 328 (2020).
[50] 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. Frontiers in Microbiology. 11 (2020).
[51] 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).

[52] Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Research. 50(D1): D571–D577 (2022).
[53] Vera C, Guerrero C, Aburto C, Cordova A, Illanes A. Conventional and non-conventional applications of β-galactosidases. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics. 1868(1): 140271 (2020).
[54] Saqib S, Akram A, Halim SA, Tassaduq R. Sources of β-galactosidase and its applications in food industry. 3 Biotech. 7(1): 79 (2017).
[55] Storhaug CL, Fosse SK, Fadnes LT. Country, regional, and global estimates for lactose malabsorption in adults: a systematic review and meta-analysis. The Lancet Gastroenterology & Hepatology. 2(10): 738–746 (2017).
[56] Michlmayr H, Kneifel W. β-Glucosidase activities of lactic acid bacteria: mechanisms, impact on fermented food and human health. FEMS Microbiology Letters. 352(1): 1–10 (2014).
[57] Ellison RT 3rd, Giehl TJ. Killing of gram-negative bacteria by lactoferrin and lysozyme. J Clin Invest. 88(4): 1080–1091 (1991).
[58] 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. Microbiology and Immunology. 25(5): 505–515 (1981).
[59] Clarke TB, Davis KM, Lysenko ES, Zhou AY, Yu Y, Weiser JN. Recognition of peptidoglycan from the microbiota by Nod1 enhances systemic innate immunity. Nature Medicine. 16(2): 228–231 (2010).
[60] 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. Frontiers in Microbiology. 8 (2017).
[61] 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, (1994).
[62] Nethery MA, Henriksen ED, Daughtry KV, Johanningsmeier SD, Barrangou R. Comparative genomics of eight Lactobacillus buchneri strains isolated from food spoilage. BMC Genomics. 20(1): 902 (2019).
[63] Lawrence JG. Gene transfer, speciation, and the evolution of bacterial genomes. Current Opinion in Microbiology. 2(5): 519–523 (1999).
[64] 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. Nature Communications. 6(1): 8322 (2015).
[65] 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. 337(6096): 816–821 (2012).
[66] Barrangou R, Horvath P. CRISPR: New Horizons in Phage Resistance and Strain Identification. Annu Rev Food Sci Technol. 3(1): 143–162 (2012).
[67] Lee K, Lee Y. Effect of Lactobacillus plantarum as a starter on the food quality and microbiota of kimchi. Food Science and Biotechnology. 19(3): 641–646 (2010).
[68] 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. Frontiers in Microbiology. 7 (2016).

[69] Bottari B, Levante A, Neviani E, Gatti M. How the Fewest Become the Greatest. L. casei's Impact on Long Ripened Cheeses. Frontiers in Microbiology. 9 (2018).
[70] 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 Microbiology. 74: 21–33 (2018).
[71] 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. Frontiers in Microbiology. 9 (2019).
[72] 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 Science. 100: 41–51 (2015).
[73] Zagorec M, Champomier-Vergès M-C. Lactobacillus sakei: A Starter for Sausage Fermentation, a Protective Culture for Meat Products. Microorganisms. 5(3) (2017).
[74] Herbin S, Mathieu F, Brulé F, Branlant C, Lefebvre G, Lebrihi A. Characteristics and Genetic Determinants of Bacteriocin Activities Produced by Carnobacterium piscicola CP5 Isolated from Cheese. Current Microbiology. 35(6): 319–326 (1997).

APPENDIX

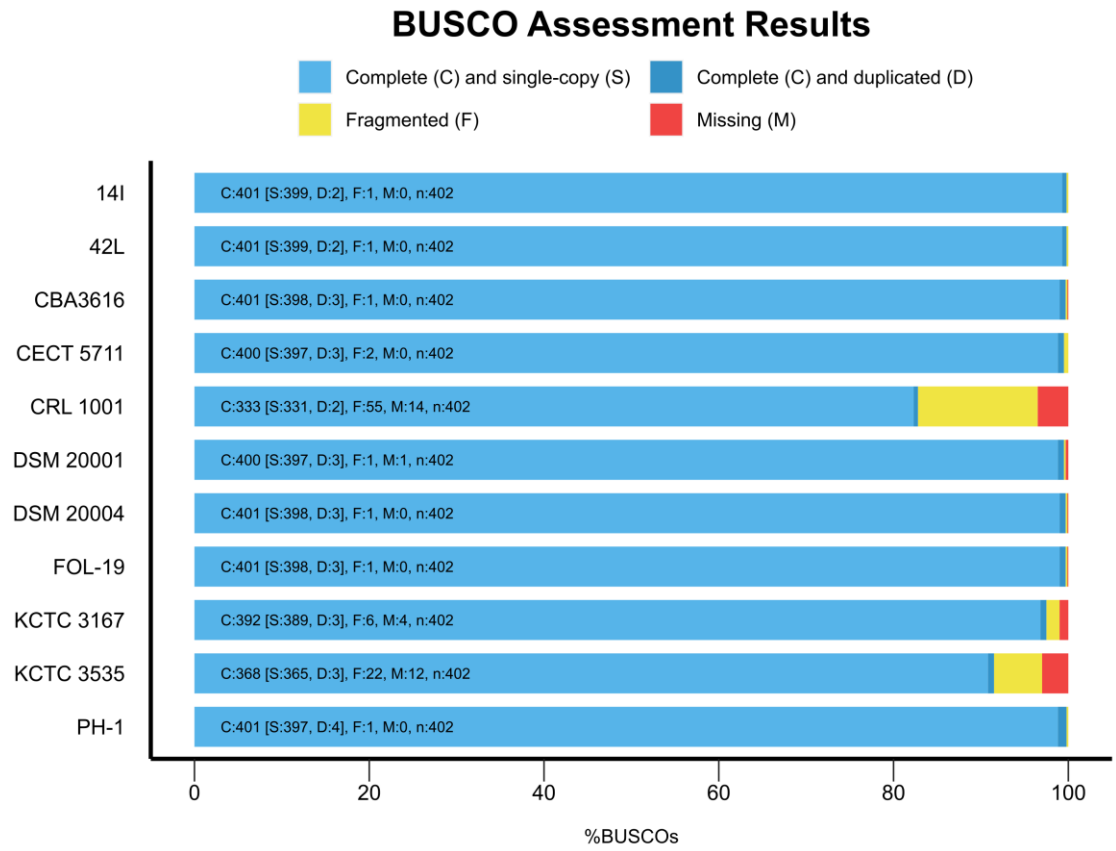

Figure A. 1 BUSCO Assessment Results of eleven *L. coryniformis*.

Table A. 1 CheckM completeness assessment results of eleven *L. coryniformis* strains.

Contamination	Strain heterogeneity
3.28	0
1.9	0
0.82	0
1.36	0
1.36	0
1.36	0
1.9	0
1.9	0
1.9	20
1.36	0
1.36	0

Bin Id	Marker Lineage	# genomes	# markers	# marker sets	0	1	2	3	4	5+	Completeness
PH-1	o__Lactobacillales (UID355)	490	336	184	2	323	10	1	0	0	99.18
FOL-19	o__Lactobacillales (UID355)	490	336	184	2	329	5	0	0	0	99.18
DSM 20004	o__Lactobacillales (UID355)	490	336	184	2	331	3	0	0	0	99.18
DSM 20001	o__Lactobacillales (UID355)	490	336	184	2	330	4	0	0	0	99.18
CECT 5711	o__Lactobacillales (UID355)	490	336	184	2	330	4	0	0	0	99.18
CBA3616	o__Lactobacillales (UID355)	490	336	184	2	330	4	0	0	0	99.18
42L	o__Lactobacillales (UID355)	490	336	184	2	329	5	0	0	0	99.18
14I	o__Lactobacillales (UID355)	490	336	184	2	329	5	0	0	0	99.18
KCTC 3535	o__Lactobacillales (UID355)	490	336	184	4	327	5	0	0	0	98.1
KCTC 3167	o__Lactobacillales (UID355)	490	336	184	5	327	4	0	0	0	97.83
CRL 1001	o__Lactobacillales (UID355)	490	336	184	11	321	4	0	0	0	95.12

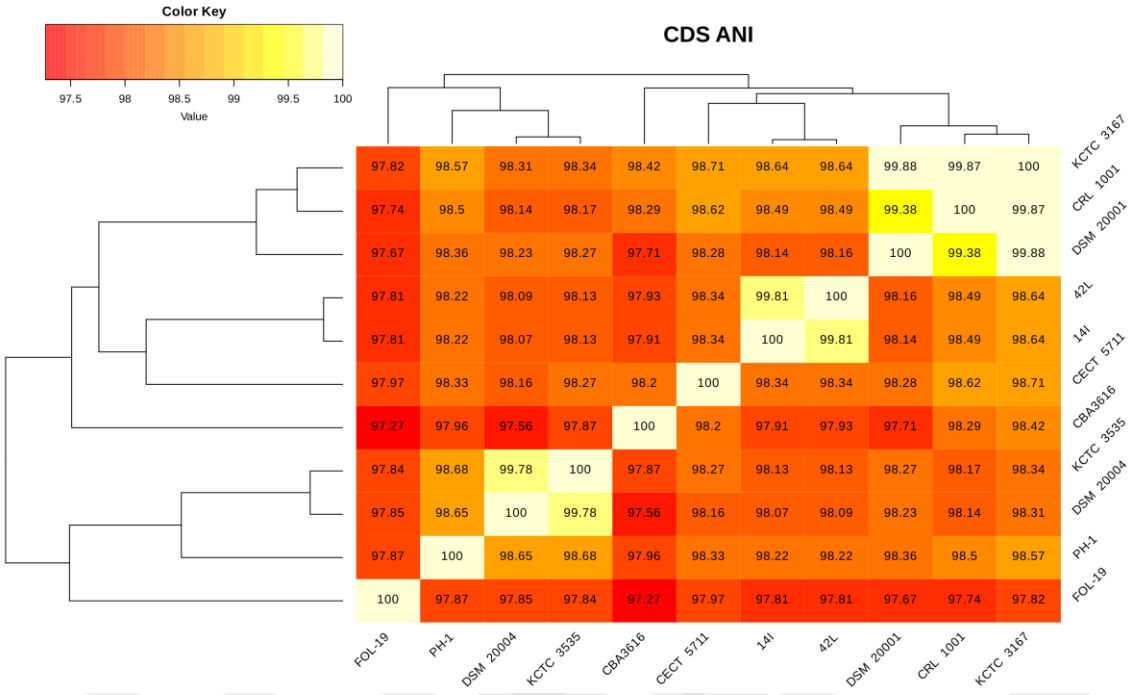

Figure A. 2 CDS average nucleotide identity of eleven *L. coryniformis* strains

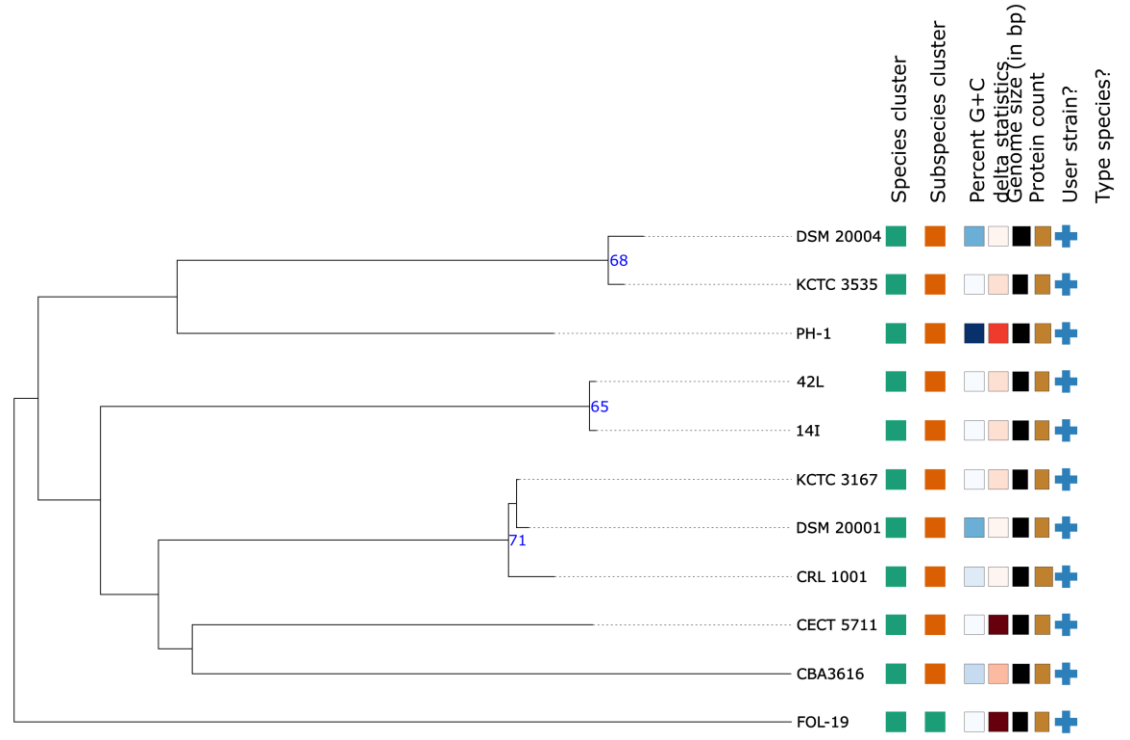

Figure A. 3 Whole-genome sequence based phylogenetic tree of eleven *L. coryniformis* strains.

Table A. 2 Presence (1) and absence (0) of carbohydrate metabolism related genes in eleven *L. coryniformis*.

Gene	Annotation	14I	42L	CBA3616	CECT 5711	CRL 1001	DSM 20001	DSM 20004	FOL-19	KCTC 3167	KCTC 3535	PH-1
bglA	Aryl-phospho-beta-D-glucosidase BglA	1	1	1	1	1	1	1	1	1	1	1
bglF	PTS system beta-glucoside-specific EIIBCA component	1	1	1	0	1	1	1	0	1	1	1
citC	[Citrate [pro-3S]-lyase] ligase	0	0	0	0	1	1	1	0	1	1	0
citD	Citrate lyase acyl carrier protein	0	0	0	0	1	1	1	0	1	1	0
citE	Citrate lyase subunit beta	0	0	0	0	1	1	1	0	1	1	0
citF	Citrate lyase alpha chain	0	0	0	0	1	1	1	0	1	1	0
citG	2-(5''-triphosphoribosyl)-3'-dephosphocoenzyme-A synthase	0	0	1	0	1	1	1	1	1	1	0
crr	PTS system glucose-specific EIIA component	0	0	1	1	1	1	1	1	1	1	1
fba	Fructose-bisphosphate aldolase	1	1	1	1	1	1	1	1	1	1	1
fruA	PTS system fructose-specific EIIABC component	1	1	1	1	1	1	1	1	1	1	1
galE	UDP-glucose 4-epimerase	1	1	1	1	1	1	1	1	1	1	1
galK	Galactokinase	1	1	1	1	1	1	1	1	1	1	1
galT	Galactose-1-phosphate uridylyltransferase	1	1	1	1	1	1	1	1	1	1	1
lacL	Beta-galactosidase large subunit	0	0	0	1	0	0	1	1	0	1	0
lacM	Beta-galactosidase small subunit	0	0	0	1	0	0	1	1	0	1	0
lacS	Lactose permease	1	1	1	1	1	1	1	1	1	1	1
lacZ	Beta-galactosidase	0	0	0	0	0	0	0	1	0	0	0
malL	Oligo-1,6-glucosidase 1	1	1	1	1	1	1	1	1	1	1	1
malP	Maltose phosphorylase	1	1	1	1	1	1	1	1	1	1	1
malR	HTH-type transcriptional regulator MalR	1	1	1	1	1	1	1	1	1	1	1

malX	PTS system maltose-specific EIICB component	0	0	0	1	0	0	1	0	0	1	1
manY	PTS system mannose-specific EIIC component	0	0	0	0	1	1	0	0	1	0	0
manZ	PTS system mannose-specific EIID component	1	1	1	1	1	1	1	1	1	1	1
ptsH	Phosphocarrier protein HPr	1	1	1	1	1	1	1	1	1	1	1
ptsI	Phosphoenolpyruvate-protein phosphotransferase	1	1	1	1	1	1	1	1	1	1	1
scrB	Sucrose-6-phosphate hydrolase	1	1	1	1	1	1	1	0	1	1	1
trePP	Trehalose 6-phosphate phosphorylase	1	1	1	1	1	1	1	1	1	1	1
ulaA	Ascorbate-specific PTS system EIIC component	0	0	0	0	1	1	0	1	1	0	0
xylB	Xylulose kinase	1	1	1	1	1	1	1	1	1	1	1
xpkA	Xylulose-5-phosphate phosphoketolase	1	1	1	1	1	1	1	1	1	1	1

Table A. 3 Presence (1) and absence (0) of genes related proteolytic activity

Gene Name	**Annotation**	**14I**	**42L**	**CBA3616**	**CECT 5711**	**CRL 1001**	**DSM 20001**	**DSM 20004**	**FOL-19**	**KCTC 3167**	**KCTC 3535**	**PH-1**
pepC	Aminopeptidase C	1	1	1	1	1	1	1	1	1	1	1
pepD	Dipeptidase	1	1	1	1	1	1	1	1	1	1	1
pepE	Aminopeptidase E	1	1	1	1	1	1	1	1	1	1	1
pepF1	Oligoendopeptidase F, plasmid	1	1	1	1	1	1	1	1	1	1	1
pepN	Aminopeptidase N	1	1	1	1	1	1	1	1	1	1	1
pepO	Neutral endopeptidase	1	1	1	1	1	1	1	1	1	1	1
pepQ	Xaa-Pro dipeptidase	1	1	1	1	1	1	1	1	1	1	1
pepS	Aminopeptidase PepS	1	1	1	1	1	1	0	0	1	0	1
pepT	Peptidase T	1	1	1	1	1	1	1	1	1	1	1
pepV	Beta-Ala-Xaa dipeptidase	1	1	1	1	1	1	1	1	1	1	1
pepX	Xaa-Pro dipeptidyl-peptidase	1	1	1	1	1	1	1	1	1	1	1

oppA	Oligopeptide-binding protein OppA	1	1	1	1	1	1	1	1	1	1	1
oppB	Oligopeptide transport system permease protein OppB	1	1	1	1	1	1	1	1	1	1	1
oppC	Oligopeptide transport system permease protein OppC	1	1	1	1	1	1	1	1	1	1	1
oppD	Oligopeptide transport ATP-binding protein OppD	1	1	1	1	1	1	1	1	1	1	1
oppF	Oligopeptide transport ATP-binding protein OppF	1	1	1	1	1	1	1	1	1	1	1

Table A. 4 Completeness of peptidase genes in *L.coryniformis*. (+) complete, (-) absent, (#) truncated.

Gene	14I	42L	CBA3616	CECT 5711	CRL 1001	DSM 20001	DSM 20004	FOL-19	KCTC 3167	KCTC 3535	PH-1
pepC	+	+	+	+	+	+	+	+	+	+	+
pepD	+	+	+	+	#	+	+	+	+	+	+
pepE	+	+	+	+	#	+	+	+	+	+	+
pepF1	+	+	+	+	+	+	+	+	+	+	+
pepN	+	+	+	+	+	+	+	+	+	+	+
pepO	+	+	+	+	+	+	+	+	+	+	+
pepQ	+	+	+	+	+	+	+	+	+	+	+
pepS	+	+	+	+	#	+	-	-	+	-	+
pepT	+	+	+	+	#	+	+	+	+	+	+
pepV	+	+	+	+	#	+	+	+	+	+	+
pepX	+	+	+	+	#	+	+	+	+	#	+
oppA	+	+	+	+	#	+	+	+	+	+	+
oppB	+	+	+	+	+	+	+	+	+	+	+
oppC	+	+	+	+	+	+	+	+	+	+	+
oppD	+	+	+	+	+	+	+	+	+	+	+
oppF	+	+	+	+	+	+	+	+	+	+	+

Table A. 5 Predicted putative plasmids in *L. coryniformis*

Strain	Identity	NCBI Accession	Length (bp)	GC (%)
14I	99.01%	NZ_AP012546.1	1326	43.29%
42L	99.24%	NZ_AP012546.1	1326	43.29%
CBA3616	99.73%	LR962096.1	809	36.34%
CBA3616	99.25%	NZ_AP012546.1	1326	43.29%
CECT 5711	99.81%	LR962096.1	809	36.34%
CECT 5711	99.34%	NZ_AP012546.1	1326	43.29%
CRL 1001	99.13%	NZ_AP012546.1	1326	43.29%
DSM 20001	99.16%	NZ_AP012546.1	1326	43.29%
DSM 20004	100.00%	NZ_CP017701.1	15010	40.67%
DSM 20004	100.00%	NZ_CP017700.1	30622	37.77%
DSM 20004	100.00%	NZ_CP017699.1	61336	41.53%
DSM 20004	100.00%	NZ_CP017698.1	75101	40.33%
DSM 20004	99.43%	NZ_AP012546.1	1326	43.29%
FOL-19	99.54%	LR962096.1	809	36.34%
FOL-19	99.05%	NZ_AP012546.1	1326	43.29%
KCTC 3535	99.90%	NZ_CP017701.1	15010	40.67%
KCTC 3535	99.80%	NZ_CP017698.1	75101	40.33%
KCTC 3535	99.74%	NZ_CP017700.1	30622	37.77%
PH-1	99.65%	NZ_CP065812.1	3474	35.81%
PH-1	99.15%	NZ_AP012546.1	1326	43.29%

Table A. 6 Insertion sequencing and their origins in *L. coryniformis*

Strain	Sequences producing significant alignments	IS Family	Origin	Score (bits)	E. value
14I	ISLca4	ISLre2	Lactobacillus casei	3009	0
14I	ISLfr1	ISL3	Lactobacillus fructivorans	2563	0
14I	IS1310	IS256	Enterococcus hirae	2339	0
14I	IS1165	ISL3	Leuconostoc mesenteroides	2101	0
14I	ISLsa1	IS30	Lactobacillus sakei	2008	0

14I	ISLpl1	IS30	Lactobacillus plantarum	1608	0
14I	ISPp1	IS30	Pediococcus pentosaceus	1505	0
14I	IS153	IS3	Lactobacillus sanfranciscensis	1142	0
14I	ISLasa2	IS3	Lactobacillus salivarius	724	0
42L	ISLca4	ISLre2	Lactobacillus casei	3009	0
42L	ISLfr1	ISL3	Lactobacillus fructivorans	2561	0
42L	IS1310	IS256	Enterococcus hirae	2339	0
42L	IS1165	ISL3	Leuconostoc mesenteroides	2115	0
42L	ISLsa1	IS30	Lactobacillus sakei	2008	0
42L	ISLpl1	IS30	Lactobacillus plantarum	1608	0
42L	ISPp1	IS30	Pediococcus pentosaceus	1505	0
42L	IS153	IS3	Lactobacillus sanfranciscensis	1142	0
42L	ISLasa2	IS3	Lactobacillus salivarius	724	0
CBA3616	ISLrh3	IS5	Lactobacillus rhamnosus	3037	0
CBA3616	ISLca4	ISLre2	Lactobacillus casei	2920	0
CBA3616	IS1310	IS256	Enterococcus hirae	2204	0
CBA3616	ISLpl1	IS30	Lactobacillus plantarum	2012	0
CBA3616	ISPp1	IS30	Pediococcus pentosaceus	1893	0
CBA3616	ISLpl3	IS5	Lactobacillus plantarum	1651	0
CBA3616	IS1165	ISL3	Leuconostoc mesenteroides	1626	0
CBA3616	IS1216E	IS6	Enterococcus faecium	1572	0
CBA3616	IS1216V	IS6	Enterococcus sp.	1540	0
CBA3616	IS1216	IS6	Enterococcus hirae	1487	0
CBA3616	ISS1W	IS6	Lactococcus lactis	1241	0
CBA3616	IS153	IS3	Lactobacillus sanfranciscensis	1158	0
CBA3616	ISLmo19	IS6	Listeria monocytogenes	1132	0
CBA3616	ISLsa1	IS30	Lactobacillus sakei	1090	0
CBA3616	ISLasa2	IS3	Lactobacillus salivarius	724	0
CBA3616	ISLmo13	IS6	Listeria monocytogenes	686	0
CECT 5711	ISLrh3	IS5	Lactobacillus rhamnosus	3037	0

CECT 5711	ISLfr1	ISL3	Lactobacillus fructivorans	2859	0
CECT 5711	ISLca4	ISLre2	Lactobacillus casei	2847	0
CECT 5711	ISL1	IS3	Lactobacillus casei	2397	0
CECT 5711	IS1163	IS3	Lactobacillus sake	2315	0
CECT 5711	IS1310	IS256	Enterococcus hirae	2171	0
CECT 5711	ISLpl1	IS30	Lactobacillus plantarum	2028	0
CECT 5711	ISLsa1	IS30	Lactobacillus sakei	2024	0
CECT 5711	ISPp1	IS30	Pediococcus pentosaceus	1909	0
CECT 5711	ISLpl3	IS5	Lactobacillus plantarum	1643	0
CECT 5711	IS1216E	IS6	Enterococcus faecium	1580	0
CECT 5711	IS1216V	IS6	Enterococcus sp.	1548	0
CECT 5711	IS1216	IS6	Enterococcus hirae	1495	0
CECT 5711	IS1165	ISL3	Leuconostoc mesenteroides	1348	0
CECT 5711	ISS1W	IS6	Lactococcus lactis	1249	0
CECT 5711	IS153	IS3	Lactobacillus sanfranciscensis	1142	0
CECT 5711	ISLmo19	IS6	Listeria monocytogenes	1116	0
CECT 5711	ISLasa2	IS3	Lactobacillus salivarius	724	0
CECT 5711	ISLmo13	IS6	Listeria monocytogenes	702	0
CRL 1001	ISLrh3	IS5	Lactobacillus rhamnosus	3021	0
CRL 1001	ISLpl1	IS30	Lactobacillus plantarum	1996	0
CRL 1001	ISPp1	IS30	Pediococcus pentosaceus	1885	0
CRL 1001	ISLpl3	IS5	Lactobacillus plantarum	1638	0
CRL 1001	IS153	IS3	Lactobacillus sanfranciscensis	1084	0
CRL 1001	ISLasa2	IS3	Lactobacillus salivarius	700	0
DSM 20001	ISLrh3	IS5	Lactobacillus rhamnosus	3037	0
DSM 20001	ISLpl1	IS30	Lactobacillus plantarum	2012	0
DSM 20001	ISPp1	IS30	Pediococcus pentosaceus	1893	0

DSM 20001	ISLpl3	IS5	Lactobacillus plantarum	1651	0
DSM 20001	IS153	IS3	Lactobacillus sanfranciscensis	1100	0
DSM 20001	ISLasa2	IS3	Lactobacillus salivarius	731	0
DSM 20004	IS1165	ISL3	Leuconostoc mesenteroides	2999	0
DSM 20004	ISLca4	ISLre2	Lactobacillus casei	2993	0
DSM 20004	ISLfr1	ISL3	Lactobacillus fructivorans	2859	0
DSM 20004	IS1163	IS3	Lactobacillus sake	2260	0
DSM 20004	IS1310	IS256	Enterococcus hirae	2189	0
DSM 20004	ISLsa1	IS30	Lactobacillus sakei	2026	0
DSM 20004	ISLpl1	IS30	Lactobacillus plantarum	2012	0
DSM 20004	IS1070	IS30	Leuconostoc lactis	1909	0
DSM 20004	ISPp1	IS30	Pediococcus pentosaceus	1901	0
DSM 20004	ISLpl3	IS5	Lactobacillus plantarum	1643	0
DSM 20004	IS153	IS3	Lactobacillus sanfranciscensis	1166	0
FOL-19	IS1165	ISL3	Leuconostoc mesenteroides	2688	0
FOL-19	ISLfr1	ISL3	Lactobacillus fructivorans	1774	0
FOL-19	ISLca3	IS5	Lactobacillus casei	1606	0
FOL-19	IS1216E	IS6	Enterococcus faecium	1564	0
FOL-19	IS1216V	IS6	Enterococcus sp.	1532	0
FOL-19	IS1216	IS6	Enterococcus hirae	1479	0
FOL-19	ISLpl3	IS5	Lactobacillus plantarum	1360	0
FOL-19	ISS1W	IS6	Lactococcus lactis	1249	0
FOL-19	IS153	IS3	Lactobacillus sanfranciscensis	1150	0
FOL-19	ISLmo19	IS6	Listeria monocytogenes	1116	0
FOL-19	ISLpl1	IS30	Lactobacillus plantarum	848	0
FOL-19	ISPp1	IS30	Pediococcus pentosaceus	801	0
FOL-19	ISLmo13	IS6	Listeria monocytogenes	670	0
KCTC 3167	ISLrh3	IS5	Lactobacillus rhamnosus	3037	0

KCTC 3167	ISLfr1	ISL3	Lactobacillus fructivorans	2843	0
KCTC 3167	IS1310	IS256	Enterococcus hirae	2228	0
KCTC 3167	ISLasa2	IS3	Lactobacillus salivarius	731	0
KCTC 3535	ISLca4	ISLre2	Lactobacillus casei	2993	0
KCTC 3535	ISLfr1	ISL3	Lactobacillus fructivorans	2839	0
KCTC 3535	IS1163	IS3	Lactobacillus sake	2244	0
KCTC 3535	IS1310	IS256	Enterococcus hirae	2222	0
KCTC 3535	ISLsa1	IS30	Lactobacillus sakei	2026	0
KCTC 3535	IS1165	ISL3	Leuconostoc mesenteroides	1945	0
KCTC 3535	IS1070	IS30	Leuconostoc lactis	1909	0
KCTC 3535	ISLpl3	IS5	Lactobacillus plantarum	1610	0
KCTC 3535	IS153	IS3	Lactobacillus sanfranciscensis	864	0
KCTC 3535	ISLpl1	IS30	Lactobacillus plantarum	700	0
KCTC 3535	ISPp1	IS30	Pediococcus pentosaceus	644	0
PH-1	IS1165	ISL3	Leuconostoc mesenteroides	2864	0
PH-1	ISLfr1	ISL3	Lactobacillus fructivorans	2759	0
PH-1	ISWci2	IS3	Weissella cibaria	2684	0
PH-1	IS1310	IS256	Enterococcus hirae	2680	0
PH-1	ISLsa1	IS30	Lactobacillus sakei	2026	0
PH-1	IS1070	IS30	Leuconostoc lactis	1917	0
PH-1	ISLpl3	IS5	Lactobacillus plantarum	1643	0
PH-1	ISLpl1	IS30	Lactobacillus plantarum	1181	0
PH-1	ISPp1	IS30	Pediococcus pentosaceus	1110	0
PH-1	IS153	IS3	Lactobacillus sanfranciscensis	882	0
CECT 5711	ISLgar4	IS6	Lactococcus garvieae	551	########
CBA3616	ISLgar4	IS6	Lactococcus garvieae	535	########
FOL-19	ISLgar4	IS6	Lactococcus garvieae	535	########
CRL 1001	IS1310	IS256	Enterococcus hirae	527	########
DSM 20001	IS1310	IS256	Enterococcus hirae	527	########
PH-1	IS640	IS21	Shigella sonnei	525	########

PH-1	IS21	IS21	Pseudomonas aeruginosa	525	########
CRL 1001	IS1165	ISL3	Leuconostoc mesenteroides	508	########
DSM 20001	IS1165	ISL3	Leuconostoc mesenteroides	507	########
KCTC 3167	ISLca2	IS5	Lactobacillus casei	484	########
DSM 20001	ISLca2	IS5	Lactobacillus casei	484	########
KCTC 3167	ISLpl1	IS30	Lactobacillus plantarum	478	########
PH-1	IS911	IS3	Shigella dysenteriae	464	########
KCTC 3167	ISPp1	IS30	Pediococcus pentosaceus	462	########
CRL 1001	ISLca2	IS5	Lactobacillus casei	460	########
PH-1	IS2	IS3	Escherichia coli	436	########
CBA3616	ISLmo4	IS6	Listeria monocytogenes	396	########
CECT 5711	ISLmo4	IS6	Listeria monocytogenes	396	########
FOL-19	ISLmo4	IS6	Listeria monocytogenes	396	########
PH-1	IS91	IS91	Escherichia coli	396	########
42L	ISLpl3	IS5	Lactobacillus plantarum	394	########
PH-1	ISKmi2	IS3	Klebsiella michiganensis	381	########
14I	ISLpl3	IS5	Lactobacillus plantarum	375	########
CECT 5711	ISEnfa1	IS6	Enterococcus faecium	375	########
CBA3616	ISEnfa1	IS6	Enterococcus faecium	359	2.00E-95
FOL-19	ISEnfa1	IS6	Enterococcus faecium	343	1.00E-90
PH-1	ISEc52	IS3	Escherichia coli	323	1.00E-84
KCTC 3167	IS153	IS3	Lactobacillus sanfranciscensis	303	8.00E-79
PH-1	ISLad1	IS3	Leclercia adecarboxylata	297	6.00E-77
CECT 5711	ISLmo14	IS6	Listeria monocytogenes	272	3.00E-69
PH-1	ISEc62	IS21	Escherichia coli	266	2.00E-67
PH-1	ISEc37	IS91	Escherichia coli	262	3.00E-66
PH-1	ISKpn74	IS5	Klebsiella pneumoniae	260	1.00E-65
PH-1	IS102	IS5	Escherichia coli	260	1.00E-65
CBA3616	ISLmo14	IS6	Listeria monocytogenes	256	2.00E-64
PH-1	ISBrsa1	IS3	Brenneria salicis	252	3.00E-63

FOL-19	ISLmo14	IS6	Listeria monocytogenes	248	4.00E-62
CECT 5711	IS946V	IS6	Lactococcus lactis	238	4.00E-59
PH-1	ISEc84	IS91	Escherichia coli	238	5.00E-59
CECT 5711	IS1297	IS6	Leuconostoc mesenteroides	236	2.00E-58
PH-1	ISEc27	IS3	Escherichia coli	236	2.00E-58
CBA3616	IS946V	IS6	Lactococcus lactis	230	1.00E-56
CBA3616	IS1297	IS6	Leuconostoc mesenteroides	228	4.00E-56
CECT 5711	ISS1E	IS6	Lactococcus lactis	228	4.00E-56
CECT 5711	ISS1D	IS6	Lactococcus lactis	228	4.00E-56
FOL-19	IS946V	IS6	Lactococcus lactis	222	3.00E-54
CBA3616	ISS1E	IS6	Lactococcus lactis	220	1.00E-53
CBA3616	ISS1D	IS6	Lactococcus lactis	220	1.00E-53
CECT 5711	ISS1CH	IS6	Lactococcus lactis	220	1.00E-53
FOL-19	IS1297	IS6	Leuconostoc mesenteroides	220	1.00E-53
PH-1	ISPrre1	IS3	Providencia rettgeri	220	1.00E-53
PH-1	ISEc36	IS3	Escherichia coli	220	1.00E-53
CECT 5711	ISS1N	IS6	Lactococcus lactis	216	2.00E-52
CECT 5711	ISS1M	IS6	Lactococcus lactis	214	6.00E-52
FOL-19	ISS1E	IS6	Lactococcus lactis	212	2.00E-51
FOL-19	ISS1D	IS6	Lactococcus lactis	212	2.00E-51
CBA3616	ISS1CH	IS6	Lactococcus lactis	212	3.00E-51
CBA3616	ISS1N	IS6	Lactococcus lactis	208	4.00E-50
CBA3616	ISS1M	IS6	Lactococcus lactis	206	2.00E-49
PH-1	IS1G	IS1	Escherichia coli	206	2.00E-49
PH-1	IS1A	IS1	Escherichia coli	206	2.00E-49
FOL-19	ISS1CH	IS6	Lactococcus lactis	204	6.00E-49
FOL-19	ISS1N	IS6	Lactococcus lactis	200	9.00E-48
FOL-19	ISS1M	IS6	Lactococcus lactis	200	9.00E-48
PH-1	IS1SD	IS1	Shigella dysenteriae	198	4.00E-47
PH-1	IS1S	IS1	Shigella sonnei	198	4.00E-47
PH-1	IS903B	IS5	Escherichia coli	196	2.00E-46
PH-1	IS903	IS5	Escherichia coli	188	4.00E-44
CBA3616	ISTeha2	IS6	Tetragenococcus halophilus	186	1.00E-43
CECT 5711	ISTeha2	IS6	Tetragenococcus halophilus	186	1.00E-43
PH-1	ISSen9	IS1	Salmonella enterica	182	2.00E-42
PH-1	IS1R	IS1	Escherichia coli	182	2.00E-42
PH-1	IS1D	IS1	Escherichia coli	182	2.00E-42

PH-1	IS1B	IS1	Escherichia coli	182	2.00E-42
CBA3616	ISP1	ISL3	Lactobacillus plantarum	180	9.00E-42
FOL-19	ISTeha2	IS6	Tetragenococcus halophilus	178	3.00E-41
PH-1	ISPan1	IS5	Pantoea ananatis	176	1.00E-40
PH-1	IS1X3	IS1	Escherichia fergusonii	174	6.00E-40
CECT 5711	ISS1Z	IS6	Lactococcus lactis	168	3.00E-38
CECT 5711	ISS1X	IS6	Lactococcus lactis	168	3.00E-38
CECT 5711	ISS1S	IS6	Lactococcus lactis	168	3.00E-38
CBA3616	ISS1Z	IS6	Lactococcus lactis	167	1.00E-37
CBA3616	ISS1X	IS6	Lactococcus lactis	167	1.00E-37
CBA3616	ISS1S	IS6	Lactococcus lactis	167	1.00E-37
CECT 5711	ISS1RS	IS6	Lactococcus lactis	161	8.00E-36
CECT 5711	ISS1B	IS6	Lactococcus lactis	161	8.00E-36
CBA3616	ISS1RS	IS6	Lactococcus lactis	159	3.00E-35
FOL-19	ISS1Z	IS6	Lactococcus lactis	159	3.00E-35
FOL-19	ISS1X	IS6	Lactococcus lactis	159	3.00E-35
FOL-19	ISS1S	IS6	Lactococcus lactis	159	3.00E-35
PH-1	IS1X4	IS1	Escherichia hermannii	159	3.00E-35
PH-1	IS1X2	IS1	Escherichia vulneris	159	3.00E-35
PH-1	IS1X1	IS1	Shigella flexneri	159	3.00E-35
PH-1	IS1F	IS1	Escherichia coli	159	3.00E-35
CBA3616	ISS1B	IS6	Lactococcus lactis	153	2.00E-33
CECT 5711	ISS1T	IS6	Lactococcus lactis	153	2.00E-33
CBA3616	ISS1T	IS6	Lactococcus lactis	151	8.00E-33
CECT 5711	ISS1A	IS6	Lactococcus lactis	151	8.00E-33
FOL-19	ISS1RS	IS6	Lactococcus lactis	151	8.00E-33
FOL-19	ISS1B	IS6	Lactococcus lactis	145	5.00E-31
CBA3616	ISS1A	IS6	Lactococcus lactis	143	2.00E-30
FOL-19	ISS1T	IS6	Lactococcus lactis	143	2.00E-30
CBA3616	ISLsa2	IS3	Lactobacillus sakei	135	5.00E-28
FOL-19	ISS1A	IS6	Lactococcus lactis	135	5.00E-28
CRL 1001	ISLpl4	IS982	Lactobacillus plantarum	133	2.00E-27
DSM 20001	ISLpl4	IS982	Lactobacillus plantarum	133	2.00E-27
KCTC 3167	ISLpl4	IS982	Lactobacillus plantarum	133	2.00E-27
14I	ISLsa2	IS3	Lactobacillus sakei	127	1.00E-25

14I	ISLpl4	IS982	Lactobacillus plantarum	127	1.00E-25
42L	ISLsa2	IS3	Lactobacillus sakei	127	1.00E-25
42L	ISLpl4	IS982	Lactobacillus plantarum	127	1.00E-25
CBA3616	ISLpl4	IS982	Lactobacillus plantarum	127	1.00E-25
CECT 5711	ISLsa2	IS3	Lactobacillus sakei	127	1.00E-25
CECT 5711	ISLpl4	IS982	Lactobacillus plantarum	127	1.00E-25
CECT 5711	IS1201	IS256	Lactobacillus helveticus	127	1.00E-25
CRL 1001	ISLsa2	IS3	Lactobacillus sakei	127	1.00E-25
DSM 20001	ISLsa2	IS3	Lactobacillus sakei	127	1.00E-25
DSM 20004	ISLho3	IS4	Lactobacillus hokkaidonensis	127	1.00E-25
DSM 20004	ISLsa2	IS3	Lactobacillus sakei	127	1.00E-25
DSM 20004	ISLpl4	IS982	Lactobacillus plantarum	127	1.00E-25
KCTC 3167	ISLsa2	IS3	Lactobacillus sakei	127	1.00E-25
KCTC 3535	ISLsa2	IS3	Lactobacillus sakei	127	1.00E-25
CRL 1001	ISLho3	IS4	Lactobacillus hokkaidonensis	125	4.00E-25
CBA3616	ISWci2	IS3	Weissella cibaria	125	5.00E-25
14I	ISLhe30	IS30	Lactobacillus helveticus	123	2.00E-24
42L	ISLhe30	IS30	Lactobacillus helveticus	123	2.00E-24
CBA3616	ISLhe30	IS30	Lactobacillus helveticus	123	2.00E-24
CRL 1001	ISLhe30	IS30	Lactobacillus helveticus	123	2.00E-24
DSM 20001	ISLhe30	IS30	Lactobacillus helveticus	123	2.00E-24
DSM 20004	ISLhe30	IS30	Lactobacillus helveticus	123	2.00E-24
KCTC 3535	ISLhe30	IS30	Lactobacillus helveticus	123	2.00E-24
PH-1	ISLpl4	IS982	Lactobacillus plantarum	123	2.00E-24
CBA3616	ISLho3	IS4	Lactobacillus hokkaidonensis	119	3.00E-23
CECT 5711	ISLho3	IS4	Lactobacillus hokkaidonensis	119	3.00E-23
DSM 20001	ISLho3	IS4	Lactobacillus hokkaidonensis	119	3.00E-23
KCTC 3535	ISLho3	IS4	Lactobacillus hokkaidonensis	119	3.00E-23

PH-1	ISLho3	IS4	Lactobacillus hokkaidonensis	119	3.00E-23
CECT 5711	ISLhe30	IS30	Lactobacillus helveticus	115	4.00E-22
FOL-19	ISLho3	IS4	Lactobacillus hokkaidonensis	115	4.00E-22
FOL-19	ISLsa2	IS3	Lactobacillus sakei	113	2.00E-21
PH-1	ISKpn14	IS1	Klebsiella pneumoniae	113	2.00E-21
14I	ISLho3	IS4	Lactobacillus hokkaidonensis	111	7.00E-21
42L	ISLho3	IS4	Lactobacillus hokkaidonensis	111	7.00E-21
KCTC 3167	IS1165	ISL3	Leuconostoc mesenteroides	109	3.00E-20
PH-1	ISLhe30	IS30	Lactobacillus helveticus	105	4.00E-19
PH-1	ISPrs1	IS3	Providencia sp.	103	2.00E-18
FOL-19	ISLhe30	IS30	Lactobacillus helveticus	95.6	4.00E-16
FOL-19	ISLpl4	IS982	Lactobacillus plantarum	93.7	2.00E-15
KCTC 3535	ISLpl4	IS982	Lactobacillus plantarum	93.7	2.00E-15
CECT 5711	ISLmo3	IS6	Listeria monocytogenes	87.7	1.00E-13
PH-1	ISSpu4	IS3	Shewanella putrefaciens	83.8	2.00E-12
FOL-19	ISLmo3	IS6	Listeria monocytogenes	79.8	2.00E-11
CBA3616	ISP2	IS1182	Lactobacillus plantarum	79.8	3.00E-11
CBA3616	ISLmo3	IS6	Listeria monocytogenes	79.8	3.00E-11
PH-1	ISCfr25	IS3	Citrobacter freundii	79.8	3.00E-11
PH-1	ISEc31	IS3	Escherichia coli	79.8	3.00E-11
PH-1	ISSlo2	IS3	Shewanella loihica	79.8	3.00E-11
CECT 5711	ISXne2	IS6	Xenorhabdus nematophila	75.8	4.00E-10
PH-1	ISXne2	IS6	Xenorhabdus nematophila	75.8	4.00E-10
PH-1	ISCfr6	IS3	Citrobacter freundii	71.9	6E-09
PH-1	ISEc48	IS3	Escherichia coli	69.9	2E-08
CECT 5711	ISLrh2	IS5	Lactobacillus rhamnosus	67.9	9E-08
CRL 1001	ISLrh2	IS5	Lactobacillus rhamnosus	67.9	9E-08
DSM 20001	ISLrh2	IS5	Lactobacillus rhamnosus	67.9	9E-08
FOL-19	IS1310	IS256	Enterococcus hirae	67.9	9E-08
KCTC 3167	ISLrh2	IS5	Lactobacillus rhamnosus	67.9	9E-08

CBA3616	ISLrh2	IS5	Lactobacillus rhamnosus	67.9	1E-07
PH-1	ISNpu10	ISAzo13	Nostoc punctiforme	67.9	1E-07
PH-1	ISLsa2	IS3	Lactobacillus sakei	65.9	4E-07
CECT 5711	ISP1	ISL3	Lactobacillus plantarum	63.9	1E-06
CECT 5711	IS431R	IS6	Staphylococcus aureus	63.9	1E-06
CECT 5711	IS431mec	IS6	Staphylococcus aureus	63.9	1E-06
CECT 5711	IS431L	IS6	Staphylococcus aureus	63.9	1E-06
CECT 5711	IS257R1	IS6	Staphylococcus aureus	63.9	1E-06
CECT 5711	IS257-1	IS6	Staphylococcus aureus	63.9	1E-06
CECT 5711	IS257R2	IS6	Staphylococcus aureus	63.9	1E-06
FOL-19	ISP1	ISL3	Lactobacillus plantarum	63.9	1E-06
KCTC 3535	ISP1	ISL3	Lactobacillus plantarum	63.9	1E-06
DSM 20004	ISP1	ISL3	Lactobacillus plantarum	63.9	2E-06
PH-1	ISP1	ISL3	Lactobacillus plantarum	63.9	2E-06
PH-1	ISAlg	IS3	Vibrio cholerae	61.9	6E-06
PH-1	ISVch4	IS3	Vibrio cholerae	61.9	6E-06
CBA3616	IS431R	IS6	Staphylococcus aureus	60	0.00002
CBA3616	IS431mec	IS6	Staphylococcus aureus	60	0.00002
CBA3616	IS431L	IS6	Staphylococcus aureus	60	0.00002
CBA3616	IS257R1	IS6	Staphylococcus aureus	60	0.00002
CBA3616	IS257-1	IS6	Staphylococcus aureus	60	0.00002
CBA3616	IS257R2	IS6	Staphylococcus aureus	60	0.00002
CECT 5711	ISSau6	IS6	Staphylococcus aureus	60	0.00002
CECT 5711	IS257-2	IS6	Staphylococcus aureus	60	0.00002
FOL-19	IS431R	IS6	Staphylococcus aureus	60	0.00002
FOL-19	IS431mec	IS6	Staphylococcus aureus	60	0.00002
FOL-19	IS431L	IS6	Staphylococcus aureus	60	0.00002
FOL-19	IS257R1	IS6	Staphylococcus aureus	60	0.00002

FOL-19	IS257-1	IS6	Staphylococcus aureus	60	0.00002
FOL-19	IS257R2	IS6	Staphylococcus aureus	60	0.00002
PH-1	ISNsp4	ISAzo13	Nostoc sp.	60	0.00002
PH-1	ISShfr6	IS3	Shewanella frigidimarina	60	0.00002
PH-1	ISSba5	IS3	Shewanella baltica	60	0.00002
KCTC 3167	ISPrre11	IS66	Providencia rettgeri	58	0.00008
KCTC 3167	ISBame1	IS256	Bacillus megaterium	58	0.00008
14I	ISPrre11	IS66	Providencia rettgeri	58	0.00009
14I	ISBame1	IS256	Bacillus megaterium	58	0.00009
42L	ISPrre11	IS66	Providencia rettgeri	58	0.00009
42L	ISBame1	IS256	Bacillus megaterium	58	0.00009
CBA3616	ISPrre11	IS66	Providencia rettgeri	58	0.00009
CBA3616	ISBame1	IS256	Bacillus megaterium	58	0.00009
CBA3616	ISSau6	IS6	Staphylococcus aureus	58	0.00009
CECT 5711	ISPrre11	IS66	Providencia rettgeri	58	0.00009
CECT 5711	ISBame1	IS256	Bacillus megaterium	58	0.00009
CRL 1001	ISPrre11	IS66	Providencia rettgeri	58	0.00009
CRL 1001	IS1221F	IS3	Mycoplasma hyorhinis	58	0.00009
DSM 20001	ISPrre11	IS66	Providencia rettgeri	58	0.00009
DSM 20001	IS1221F	IS3	Mycoplasma hyorhinis	58	0.00009
FOL-19	ISLiv2	IS256	Listeria ivanovii	58	0.00009
FOL-19	ISPrre11	IS66	Providencia rettgeri	58	0.00009
FOL-19	ISSau6	IS6	Staphylococcus aureus	58	0.00009
FOL-19	IS1661	IS3	Yersinia pestis	58	0.00009
FOL-19	IS1221F	IS3	Mycoplasma hyorhinis	58	0.00009
KCTC 3535	ISPrre11	IS66	Providencia rettgeri	58	0.00009
KCTC 3535	ISBame1	IS256	Bacillus megaterium	58	0.00009
PH-1	ISPrre11	IS66	Providencia rettgeri	58	0.00009
PH-1	ISRisp1	ISAzo13	Rivularia sp.	58	0.00009
PH-1	ISPsy24	IS3	Pseudomonas syringae	58	0.00009
PH-1	ISS1Z	IS6	Lactococcus lactis	58	0.00009
DSM 20004	ISPrre11	IS66	Providencia rettgeri	58	0.0001
DSM 20004	ISBame1	IS256	Bacillus megaterium	58	0.0001

42L	ISLmo7	IS21	Listeria monocytogenes	56	0.0003
CRL 1001	ISLmo8	IS3	Listeria monocytogenes	56	0.0003
CRL 1001	ISLmo7	IS21	Listeria monocytogenes	56	0.0003
FOL-19	ISAba15	IS5	Acinetobacter baumannii	56	0.0003
KCTC 3167	ISLmo8	IS3	Listeria monocytogenes	56	0.0003
KCTC 3167	ISLmo7	IS21	Listeria monocytogenes	56	0.0003
KCTC 3535	ISAba15	IS5	Acinetobacter baumannii	56	0.0003
14I	ISLmo7	IS21	Listeria monocytogenes	56	0.0004
CBA3616	ISLmo7	IS21	Listeria monocytogenes	56	0.0004
DSM 20001	ISLmo8	IS3	Listeria monocytogenes	56	0.0004
DSM 20001	ISLmo7	IS21	Listeria monocytogenes	56	0.0004
DSM 20004	ISAba15	IS5	Acinetobacter baumannii	56	0.0004
PH-1	ISPa126	IS3	Pseudomonas aeruginosa	56	0.0004
14I	ISSlu1	IS30	Streptococcus lutetiensis	54	0.001
14I	ISBth166	IS110	Bacillus thuringiensis	54	0.001
42L	ISSlu1	IS30	Streptococcus lutetiensis	54	0.001
42L	ISBth166	IS110	Bacillus thuringiensis	54	0.001
CBA3616	ISSlu1	IS30	Streptococcus lutetiensis	54	0.001
CECT 5711	ISSlu1	IS30	Streptococcus lutetiensis	54	0.001
CECT 5711	ISBth20	IS6	Bacillus thuringiensis	54	0.001
CECT 5711	ISHar2	IS3	Herminiimonas arsenicoxydans	54	0.001
CRL 1001	ISSlu1	IS30	Streptococcus lutetiensis	54	0.001
CRL 1001	ISBth166	IS110	Bacillus thuringiensis	54	0.001
DSM 20001	ISSlu1	IS30	Streptococcus lutetiensis	54	0.001
DSM 20001	ISBth166	IS110	Bacillus thuringiensis	54	0.001
DSM 20004	ISP2	IS1182	Lactobacillus plantarum	54	0.001

DSM 20004	ISSlu1	IS30	Streptococcus lutetiensis	54	0.001
DSM 20004	ISBth166	IS110	Bacillus thuringiensis	54	0.001
DSM 20004	ISLh1	IS982	Lactobacillus helveticus	54	0.001
FOL-19	ISLpl2	IS3	Lactobacillus plantarum	54	0.001
FOL-19	ISBth166	IS110	Bacillus thuringiensis	54	0.001
KCTC 3167	ISSlu1	IS30	Streptococcus lutetiensis	54	0.001
KCTC 3167	ISBth166	IS110	Bacillus thuringiensis	54	0.001
KCTC 3535	ISP2	IS1182	Lactobacillus plantarum	54	0.001
KCTC 3535	ISSlu1	IS30	Streptococcus lutetiensis	54	0.001
KCTC 3535	ISBth166	IS110	Bacillus thuringiensis	54	0.001
KCTC 3535	ISLh1	IS982	Lactobacillus helveticus	54	0.001
PH-1	ISPmo5	IS3	Pseudomonas monteilii	54	0.001
PH-1	ISBth166	IS110	Bacillus thuringiensis	54	0.001
PH-1	ISVsa11	IS3	Aliivibrio salmonicida	54	0.001
PH-1	ISLh1	IS982	Lactobacillus helveticus	54	0.001
14I	ISEfa9	IS3	Enterococcus faecium	52	0.005
14I	ISTet2	IS6	Thermoanaerobacter ethanolicus	52	0.005
42L	ISEfa9	IS3	Enterococcus faecium	52	0.005
42L	ISTet2	IS6	Thermoanaerobacter ethanolicus	52	0.005
CECT 5711	ISEfa9	IS3	Enterococcus faecium	52	0.005
CECT 5711	ISTet2	IS6	Thermoanaerobacter ethanolicus	52	0.005
CRL 1001	ISLmo5	IS3	Listeria monocytogenes	52	0.005
CRL 1001	ISTet2	IS6	Thermoanaerobacter ethanolicus	52	0.005
FOL-19	ISDet4	IS256	Dehalococcoides ethenogenes	52	0.005
FOL-19	ISEf1	IS256	Enterococcus faecalis	52	0.005
FOL-19	IS257-2	IS6	Staphylococcus aureus	52	0.005
FOL-19	IS1223	IS3	Lactobacillus johnsonii	52	0.005

KCTC 3167	ISEfa9	IS3	Enterococcus faecium	52	0.005
KCTC 3167	ISLmo5	IS3	Listeria monocytogenes	52	0.005
KCTC 3535	ISEfa9	IS3	Enterococcus faecium	52	0.005
CBA3616	ISEfa9	IS3	Enterococcus faecium	52	0.006
CBA3616	ISTet2	IS6	Thermoanaerobacter ethanolicus	52	0.006
CBA3616	IS660	IS1182	Bacillus halodurans	52	0.006
CBA3616	IS257-2	IS6	Staphylococcus aureus	52	0.006
DSM 20001	ISLmo5	IS3	Listeria monocytogenes	52	0.006
DSM 20001	ISTet2	IS6	Thermoanaerobacter ethanolicus	52	0.006
DSM 20004	ISEfa9	IS3	Enterococcus faecium	52	0.006
PH-1	ISEfa9	IS3	Enterococcus faecium	52	0.006
PH-1	IS621	IS110	Escherichia coli	52	0.006
PH-1	ISMac7	IS200/IS605	Methanosarcina acetivorans	52	0.006
KCTC 3167	ISEch7	IS30	Dickeya dadantii	50.1	0.02
KCTC 3167	MICBce1	IS4	Bacillus cereus	50.1	0.02
KCTC 3167	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.02
CECT 5711	ISAbe16	IS3	Acinetobacter bereziniae	50.1	0.021
CECT 5711	ISEch7	IS30	Dickeya dadantii	50.1	0.021
CECT 5711	ISStin2	IS30	Streptococcus iniae	50.1	0.021
CECT 5711	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.021
CECT 5711	ISBce15	IS3	Bacillus cereus	50.1	0.021
CECT 5711	ISRde1	IS3	Roseobacter denitrificans	50.1	0.021
CRL 1001	ISEch7	IS30	Dickeya dadantii	50.1	0.021
CRL 1001	ISStin2	IS30	Streptococcus iniae	50.1	0.021
CRL 1001	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.021
CRL 1001	ISSpn5	IS1380	Streptococcus pneumoniae	50.1	0.021
CRL 1001	IS1221I	IS3	Mycoplasma hyopneumoniae	50.1	0.021
CRL 1001	IS1221H	IS3	Mycoplasma hyopneumoniae	50.1	0.021

CRL 1001	IS1221E	IS3	Mycoplasma hyopneumoniae	50.1	0.021
CRL 1001	IS1221C	IS3	Mycoplasma hyorhinis	50.1	0.021
CRL 1001	IS1221B	IS3	Mycoplasma hyorhinis	50.1	0.021
CRL 1001	IS1221A	IS3	Mycoplasma hyorhinis	50.1	0.021
FOL-19	ISEch7	IS30	Dickeya dadantii	50.1	0.021
FOL-19	ISStin2	IS30	Streptococcus iniae	50.1	0.021
FOL-19	ISPeth4	IS256	Pelotomaculum thermopropionicum	50.1	0.021
FOL-19	ISSpn5	IS1380	Streptococcus pneumoniae	50.1	0.021
FOL-19	IS1221I	IS3	Mycoplasma hyopneumoniae	50.1	0.021
FOL-19	IS1221H	IS3	Mycoplasma hyopneumoniae	50.1	0.021
FOL-19	IS1221E	IS3	Mycoplasma hyopneumoniae	50.1	0.021
FOL-19	IS1221C	IS3	Mycoplasma hyorhinis	50.1	0.021
FOL-19	IS1221B	IS3	Mycoplasma hyorhinis	50.1	0.021
FOL-19	IS1221A	IS3	Mycoplasma hyorhinis	50.1	0.021
KCTC 3535	ISEch7	IS30	Dickeya dadantii	50.1	0.021
KCTC 3535	ISStin2	IS30	Streptococcus iniae	50.1	0.021
KCTC 3535	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.021
KCTC 3535	ISSpn5	IS1380	Streptococcus pneumoniae	50.1	0.021
14I	ISStin2	IS30	Streptococcus iniae	50.1	0.022
14I	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.022
42L	ISStin2	IS30	Streptococcus iniae	50.1	0.022
42L	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.022
CBA3616	ISLhe63	IS1182	Lactobacillus helveticus	50.1	0.022
CBA3616	ISStin2	IS30	Streptococcus iniae	50.1	0.022
CBA3616	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.022
CBA3616	ISBco1	IS1380	Bacillus coagulans	50.1	0.022
DSM 20001	ISEch7	IS30	Dickeya dadantii	50.1	0.022
DSM 20001	ISStin2	IS30	Streptococcus iniae	50.1	0.022
DSM 20001	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.022

DSM 20001	ISSpn5	IS1380	Streptococcus pneumoniae	50.1	0.022
DSM 20001	IS1221I	IS3	Mycoplasma hyopneumoniae	50.1	0.022
DSM 20001	IS1221H	IS3	Mycoplasma hyopneumoniae	50.1	0.022
DSM 20001	IS1221E	IS3	Mycoplasma hyopneumoniae	50.1	0.022
DSM 20001	IS1221C	IS3	Mycoplasma hyorhinis	50.1	0.022
DSM 20001	IS1221B	IS3	Mycoplasma hyorhinis	50.1	0.022
DSM 20001	IS1221A	IS3	Mycoplasma hyorhinis	50.1	0.022
DSM 20004	ISEch7	IS30	Dickeya dadantii	50.1	0.023
DSM 20004	ISStin2	IS30	Streptococcus iniae	50.1	0.023
DSM 20004	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.023
DSM 20004	ISSpn5	IS1380	Streptococcus pneumoniae	50.1	0.023
PH-1	ISEch7	IS30	Dickeya dadantii	50.1	0.023
PH-1	ISStin2	IS30	Streptococcus iniae	50.1	0.023
PH-1	ISAve3	IS3	Aeromonas veronii	50.1	0.023
PH-1	ISAs17	IS3	Aeromonas salmonicida	50.1	0.023
PH-1	ISSne1	IS256	Sporosarcina newyorkensis	50.1	0.023
PH-1	ISBame1	IS256	Bacillus megaterium	50.1	0.023
PH-1	ISCap1	ISAzo13	Candidatus Accumulibacter	50.1	0.023
PH-1	ISBco1	IS1380	Bacillus coagulans	50.1	0.023
PH-1	ISSpi4	IS3	Shewanella piezotolerans	50.1	0.023
PH-1	ISSpn5	IS1380	Streptococcus pneumoniae	50.1	0.023
PH-1	IS946V	IS6	Lactococcus lactis	50.1	0.023
PH-1	IS1520	IS3	Lactobacillus sakei	50.1	0.023
KCTC 3167	ISArsp14	ISNCY	Arthrobacter sp.	48.1	0.081
KCTC 3167	ISLpl2	IS3	Lactobacillus plantarum	48.1	0.081
KCTC 3167	ISCfr4	IS30	Citrobacter freundii	48.1	0.081
KCTC 3167	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.081
CRL 1001	ISArsp14	ISNCY	Arthrobacter sp.	48.1	0.083
CRL 1001	ISLpl2	IS3	Lactobacillus plantarum	48.1	0.083
CRL 1001	ISCfr4	IS30	Citrobacter freundii	48.1	0.083

CRL 1001	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.083
CRL 1001	ISMmy3	IS3	Mycoplasma mycoides	48.1	0.083
CRL 1001	ISLh1	IS982	Lactobacillus helveticus	48.1	0.083
CRL 1001	IS1201	IS256	Lactobacillus helveticus	48.1	0.083
FOL-19	ISArsp14	ISNCY	Arthrobacter sp.	48.1	0.083
FOL-19	ISLmo8	IS3	Listeria monocytogenes	48.1	0.083
FOL-19	ISCfr4	IS30	Citrobacter freundii	48.1	0.083
FOL-19	ISNaoc1	IS200/IS605	Natronococcus occultus	48.1	0.083
FOL-19	ISLgar5	IS256	Lactococus garvieae	48.1	0.083
FOL-19	ISEfm2	IS256	Enterococcus faecium	48.1	0.083
FOL-19	ISEfa13	IS256	Enterococcus faecium	48.1	0.083
FOL-19	ISLh1	IS982	Lactobacillus helveticus	48.1	0.083
KCTC 3535	ISLmo8	IS3	Listeria monocytogenes	48.1	0.083
KCTC 3535	ISCfr4	IS30	Citrobacter freundii	48.1	0.083
KCTC 3535	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.083
KCTC 3535	IS1201	IS256	Lactobacillus helveticus	48.1	0.083
CECT 5711	ISArsp14	ISNCY	Arthrobacter sp.	48.1	0.084
CECT 5711	ISLmo8	IS3	Listeria monocytogenes	48.1	0.084
CECT 5711	ISCfr4	IS30	Citrobacter freundii	48.1	0.084
CECT 5711	ISBce18	IS3	Bacillus cereus	48.1	0.084
CECT 5711	ISCARN89	IS21	Metagenomic data	48.1	0.084
CECT 5711	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.084
CECT 5711	ISRme13	IS3	Ralstonia metallidurans	48.1	0.084
CECT 5711	ISAzvi9	IS3	Azotobacter vinelandii	48.1	0.084
CECT 5711	ISLh1	IS982	Lactobacillus helveticus	48.1	0.084
CECT 5711	IS257-3	IS6	Staphylococcus aureus	48.1	0.084
CECT 5711	IS1397	IS3	Escherichia coli	48.1	0.084
14I	ISLmo20	IS256	Listeria monocytogenes	48.1	0.085

14I	ISLpl2	IS3	Lactobacillus plantarum	48.1	0.085
14I	ISLmo8	IS3	Listeria monocytogenes	48.1	0.085
14I	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.085
42L	ISLmo20	IS256	Listeria monocytogenes	48.1	0.085
42L	ISLpl2	IS3	Lactobacillus plantarum	48.1	0.085
42L	ISLmo8	IS3	Listeria monocytogenes	48.1	0.085
42L	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.085
42L	IS1201	IS256	Lactobacillus helveticus	48.1	0.085
DSM 20001	ISArsp14	ISNCY	Arthrobacter sp.	48.1	0.087
DSM 20001	ISLpl2	IS3	Lactobacillus plantarum	48.1	0.087
DSM 20001	ISCfr4	IS30	Citrobacter freundii	48.1	0.087
DSM 20001	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.087
DSM 20001	ISMmy3	IS3	Mycoplasma mycoides	48.1	0.087
DSM 20001	ISLh1	IS982	Lactobacillus helveticus	48.1	0.087
DSM 20001	IS1201	IS256	Lactobacillus helveticus	48.1	0.087
CBA3616	ISLmo8	IS3	Listeria monocytogenes	48.1	0.088
CBA3616	ISBf4	IS1182	Bacteroides fragilis	48.1	0.088
CBA3616	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.088
CBA3616	ISEnfa364	IS30	Enterococcus faecalis	48.1	0.088
CBA3616	ISLh1	IS982	Lactobacillus helveticus	48.1	0.088
CBA3616	IS1201	IS256	Lactobacillus helveticus	48.1	0.088
PH-1	ISPa127	IS3	Pseudomonas aeruginosa	48.1	0.09
PH-1	ISPa107	IS3	Pseudomonas aeruginosa	48.1	0.09
PH-1	ISLpl2	IS3	Lactobacillus plantarum	48.1	0.09
PH-1	ISLmo8	IS3	Listeria monocytogenes	48.1	0.09
PH-1	ISCfr4	IS30	Citrobacter freundii	48.1	0.09
PH-1	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.09

PH-1	IS222	IS3	Pseudomonas aeruginosa	48.1	0.09
PH-1	IS1201	IS256	Lactobacillus helveticus	48.1	0.09
DSM 20004	ISLmo8	IS3	Listeria monocytogenes	48.1	0.093
DSM 20004	ISCfr4	IS30	Citrobacter freundii	48.1	0.093
DSM 20004	ISLhe2	ISL3	Lactobacillus helveticus	48.1	0.093
DSM 20004	IS1201	IS256	Lactobacillus helveticus	48.1	0.093

CURRICULUM VITAE

2016 – 2021 B.Sc., Molecular Biology and Genetics, Abdullah Gul University, Kayseri, TURKEY

2021 – Present M.Sc., Bioengineering, Abdullah Gul University, Kayseri, TURKEY

SELECTED PUBLICATIONS AND PRESENTATIONS

J1) Gumustop I, Ortakci F. Analyzing the genetic diversity and biotechnological potential of *Leuconostoc pseudomesenteroides* by comparative genomics. Frontiers in Microbiology. 13 (2023).

J2) Gumustop I, Ortakci F. Comparative Genomics of *Lentilactobacillus parabuchneri* isolated from dairy, KEM complex, Makgeolli, and Saliva Microbiomes. BMC Genomics. 23(1): 803 (2022).

J3) Gumustop I, Ortakci F. Comparative genomics of *Leuconostoc lactis* strains isolated from human gastrointestinal system and fermented foods microbiomes. BMC Genomic Data. 23(1): 61 (2022).

J4) Gumustop I, Ortakci F. Evaluating the microbial growth kinetics and artificial gastric digestion survival of a novel *Pichia kudriavzevii* FOL-04. Biotech Studies. 31(1): 28–35 (2022).