Emerging process of genetic exchange communities in lactic acid bacteria

Takenaka, Shinkuro

Doctor of Philosophy

Department of Genetics

School of Life Science

The Graduate University for Advanced Studies,

SOKENDAI

Contents

Abstract

In prokaryotes, a major contributor to genomic evolution is gene exchange via horizontal gene transfer (HGT). Bacterial populations with a high HGT frequency are defined as genetic exchange communities (GECs) and often arise in shared ecological niches, characterized by symbiotic interactions and/or phylogenetic closeness. Although some phenotypes are associated with specific ecological niches linked to GECs, little is known about the phenotypic influences on GECs in a taxonomic family with concrete genomic evidence.

I investigated the relationship between bacterial evolution and GECs in ecological niches using phenotypic and genomic data from lactic acid bacteria (LAB). I focused on information on phenotypic features because they reflect the ecological niche of bacteria. LAB produce lactic acid by fermenting carbohydrates and inhabit various ecological niches in food industries, such as fermented foods. They inhabit specific ecological niches, such as fermented milk products, meats, cereals, and vegetables. These are suitable properties of a material for the investigation of GECs in ecological niches. Because they are involved in human activity, genomic and phenotypic data of LAB have been accumulated. The phenotypic and genomic features of LAB can elucidate the relationships between bacterial evolution and GECs in ecological niches.

I selected 178 strains of 24 genera from the *Lactobacillaceae* family to clarify factors contributing to the formation of GECs. In this family, the genus *Lactobacillus* has recently been reclassified into 25 genera, and their phenotypes, including sugar utilization, growth temperature, and oxygen tolerance, have been well documented. Moreover, they

exhibit diverse genomic features. *Lactobacillus apis* has a small genome of 1.70 Mbp, whereas *Lactiplantibacillus plantarum* subsp. *plantarum* has a large genome of 3.45 Mbp. Therefore, the group previously identified as the genus *Lactobacillus* provides a good sandbox to study the influence of ecological niches on HGT in relation to phenotypes, ecologies, and genotypes.

The way that LAB construct GECs in an ecological niche was investigated to analyze their phenotypes, habitats, and ortholog networks. I found that phenotypes to utilize various sugars contribute to forming GECs. The statistical analysis revealed that sugar utilization influences frequent HGT in LAB. To confirm the association between sugar utilization and GECs, the concept of the Average number of Sugar Utilization for the ortholog (ASU) was introduced. Using the ASU, two groups of orthologs were compared, i.e., the orthologs shared dominantly by strains that were able to use a variety of sugars (generalist) and those shared by strains that used only a few sugars (specialist). While the networks of orthologs predominantly shared by the specialist groups for sugar utilization were connected only within the same genera, the networks of the generalist groups were connected across genera. In addition, the genes in the generalist group ortholog encoded not only phenotypes involving sugar utilization but also phenotypes to adapt to various environments, including stress responses, bacteriocin production, antibiotic resistance, survival in the intestinal environment, and heavy metal resistance. The strains in the generalist networks were presumed to use these genes for sharing niches, such as vegetables, dairy products, and brewing-related environments. This feature is consistent with the fact that *Lactobacillaceae* contributes to producing a wide variety of fermented foods. Thus, the results suggested that the phenotype to utilize various sugars, which makes the bacteria become generalists, contributes to forming GECs in the ecological niche of LAB.

Next, I investigated whether the niche construction and GECs affect the genetic diversity in a LAB genome. The bacteria with genetic diversity tended to have potential for gene gain events. Gained genes that encoded phenotypes for adaptation to environments contributed to the formation of GECs in various ecological niches. Through multiplicative events, a higher frequency of gene gain events in generalists may further broaden their niche breadth compared to specialists.

In conclusion, to reveal the formation process of GECs in the ecological niche, I investigated phenotypic and genomic factors in 178 strains of 24 genera in *Lactobacillaceae*. The results suggested that utilizing various sugars substantially influenced the formation of GECs in ecological niches. In addition, genetic diversity might contribute to further increasing potential for gene gain events in LAB. Thus, metabolic capabilities associated with ecological niches contributed to the formation of GECs, which may further promote genetic diversity, balancing it against the pressure to reduce the genomes.

List of figure and table

Figure

Acknowledgments

I would like to express my deepest appreciation to Professor Masanori Arita for his kind support and expert advice as my supervisor. I learned a lot from his attitude and perspective as a researcher.

I would like to express my gratitude to Assistant Professor Takeshi Kawashima. He gave me many time discussions for my study and expert advice, which had been insightful.

I am indebted to Project Associate Professor Nozomu Sakurai and other members of the Biological Networks laboratory. Especially, I would like to thank Dr. Copputa Tada for teaching me the technical skill set for bioinformatics.

I am grateful to my doctoral examination committee members: Professor Yasukazu Nakamura, Professor Ken Kurokawa, Professor Jun Kitano, Professor Niki Hironori and Dr. Masanori Tohno who is a senior researcher in The National Agriculture and Food Research Organization. They provided my study with a wide perspective from various fields.

Special thanks to Prima Meat Packers, Ltd. which paid the enrollment fee and the tuition for the first semester of freshman year in my doctoral course. Shortly after, I left this company because of my health problems and I changed my study theme from the one for this company to bacterial evolution. The company provided me with the opportunity to challenge doctoral courses, which was helpful for my career.

Finally, I would like to express gratitude to my wife for her moral support and warm encouragement. I would not be the person I am today without you.

Publication notes

This thesis is based on the paper (Takenaka et al. 2021).

Chapter 1: General introduction

1.1 Genomic exchange communities (GECs) generated in horizontal gene transfer (HGT) networks promote the evolution of bacterial genomes

Horizontal gene transfer (HGT) is an evolutionary process that allows genetic innovations to spread between distantly related organisms (Andam and Gogarten 2011). HGT is a major contributor to genome evolution and structure in bacteria (Hall et al. 2017). For instance, transfer of gene clusters containing a set of genes involved in the metabolism of carbon sources or resistance to toxins is known (Wiedenbeck and Cohan 2011). In addition, frequent HGT can result in large changes in the genome size (Zimmer and Emlen 2016). Variability in the genome size is frequently observed among closely related strains (Canard and Cole 1989; Harsono et al. 1993; Daniels 1990; Prevost et al. 1992; Tanskanen et al. 1990), and this can be caused by HGT (Bergthorsson and Ochman 1995; Bobay and Ochman 2017). Thus, HGT plays a major role in the evolution of microorganism genomes. When such transfer is described as networks (Puigbò et al. 2010), the HGT bias in preference for transfer partners results in high-density regions in the networks, defined as genetic exchange communities (GECs) (Skippington and Ragan 2011).

1.2 Elucidation of the process of forming GECs in ecological niches provides perspective on the process of bacterial evolution

GECs often occur in shared ecological niches, characterized by symbiotic interactions and phylogenetic closeness (Andam and Gogarten 2011). GECs in ecological niches obscure the definition of a bacterial population, which makes bacterial evolution difficult. Sharing ecological niches causes frequent HGT among multiple bacterial lineages. Indiscriminate exchange of genes via HGT makes the line of descent challenging to follow (Schleifer et al. 2008; Rocha 2018). In addition to HGT mechanisms generating bias to promote gene transfer among closely related organisms, many reports suggest that HGT also occurs among distantly related organisms in ecological niches. For example, different phylum bacteria share genes for surviving in a high-temperature environment (Andam and Gogarten 2011). Distantly related microorganisms can share their features via HGT, which contribute to their adaptation to the environment. This obscures the bacterial population and makes bacterial evolution difficult to understand using population genetics (Rocha 2018). The GECs greatly influence bacterial evolution and spread genetic innovations between distantly related bacterial lineages. Revealing the process of GECs formation will help to elucidate the evolutional process of bacteria.

1.3 How ecological niches form GECs: The approach

Investigating the relationships between environmental factors and GECs among bacteria is an effective approach to reveal the theory of bacterial evolution. Different bacteria existing in different environments vary in physical, chemical, or biological properties. For example, antibiotic resistance bacteria dominate in the hospital.

While investigating ecological niches, finding the niche to which the bacteria belong is difficult. Bacteria have a huge population and fast generation cycle and are less influenced by geographical isolation (Kirchman 2012; Odling-Smee et al. 2003). In addition, the bias generated by culturable bacteria can cause a misunderstanding regarding the ecological niches. Isolation of bacterial strains from an environment does not mean the bacteria are dominant in that environment. For example, although the genus *Streptomyces* and *Bacillus* are often isolated from soil, the 16S rRNA gene clone library analysis indicated these bacteria are not dominant in the soil (Kirchman 2012). Moreover, although the genus *Pseudomonas* and *Vibrio* are frequently detected by seawater cultivation, their 16S rRNA genes rarely exist in seawater. Therefore, bacteria isolated from particular environments are not representative of the microflora in their niche. The ecological niches made by the non-culturable majority may be wrongly annotated because of the culturable minority. These reasons confuse our understanding of the relationships between ecological niches and bacteria.

Meta genome analysis is one of the ways to solve the culturable bias in the investigation of ecological niches. Although metagenome analysis based on 16S rRNA is frequently used, the resolution is not enough for high-precision analysis. For instance, although *Bacillus cereus*, *B. anthracis*, and *B. thuringiensis* are classified as different species, the sequence homology of their 16S rRNA gene is over 97%, which agrees with them being the same species (Kirchman 2012). Meta genome analysis based on 16S rRNA is beneficial to elucidate rough tendency. However, another method is required to investigate ecological niches because even closely related species have variant features and habitats.

The approach focusing on the phenotypic features in bacteria helps assess ecological niches more accurately. Bacterial phenotypes reflect the ecological niche. Genes that encode suitable phenotypes for surviving keep their sequence because of purifying selection in the environment. Genomic data allows high-resolution analysis to reveal the characteristics of bacteria. Furthermore, research on the genomic and phenotypic features in bacteria contributes to discovering the new relationships between ecological niches and bacterial evolution. The detailed analysis of genomic and phenotypic features of bacteria is provided in Section 3.1.

1.4 Lactic acid bacteria (LAB): the target organisms

In this study, I focus on LAB because they have properties suitable for the investigation of ecological niches and evolution in bacteria: variant ecological niches and abundant genomic and phenotypic data. LAB have evolved to adapt to a variety of niches, as explained later. In addition, because LAB strains are used in various fermented foods, their genomic and phenotypic information are available. Various habitats and abundant data in LAB will enhance the investigation of bacterial evolution.

The major conditions regulating the distribution of LAB are nutrients, oxygen, and temperature. LAB strains require carbon sources, amino acids, and vitamins. Moreover, the oxygen condition influences LAB growth. LAB prefer oxygen-free environments because they do not possess catalase to break down the hydrogen peroxide generated in the presence of oxygen. Furthermore, temperature restricts their growth: they can grow in the range of 5–45 °C (Caplice and Fitzgerald 1999). LAB strains are usually distributed in the environments that meet these conditions.

Almost all environments where animals and plants inhabit fulfill the conditions for LAB growth (Yamamoto et al. 2010). In habitats associated with animals, LAB grows in milk, animal intestine, vagina, and feces. LAB strains also inhabit plant-related environments: flower nectar, sap, sedimentary soil of a plant, and damaged fruit. Furthermore, humans have constructed artificial environments for LAB habitat to use them in various foods. Some traditional foods, such as yogurt, cheese, and pickled vegetables, have LAB. In addition, LAB play a major role in liqueur fermenting. These environments are ecological niches for LAB.

Bacteria improve their survivability to become specialists (i.e., microbes adapted to specific habitats) or generalists (i.e., microbes able to adapt to diverse habitats) (Sriswasdi et al. 2017; Douglas 1988). Without exception, LAB also include specialists and generalists.

Some LAB specialize in the niches and adapt to the surrounding environments. There is a tendency for a specialist genome size to be smaller than a generalist's genome size because specialists lack genes not required for survival in the niches (Sriswasdi et al. 2017). For example, *Lactobacillus apis,* which inhabits the intestine of bees, and *Limosilactobacillus vaginalis*, which occupies the animal vagina, have genomes as small as 1.70 Mbp and 1.79 Mbp, respectively (Zheng et al. 2020). A report investigating nine LAB genomes suggested that deletion of genes and simplifying the metabolism are characteristics of evolution. Furthermore, LAB adapt to nutrient-rich environments (Makarova et al. 2006). For instance, LAB require various rich nutrients to grow in synthetic media: amino acid, vitamins, nucleotide acid, and minerals (Yamamoto et al. 2010).

However, some generalists in LAB have diverse habitats. For instance, *Lactiplantibacillus plantarum* subsp. *plantarum* inhabits various environments; they are isolated from dairy products, silage, sauerkraut, pickled vegetables, sourdough, cow dung, the human mouth, intestinal tract and stools, and sewage. In addition, the microbe has a large genome size (3.45 Mbp) (Zheng et al. 2020) because it requires various genetic materials to adapt to diverse environments. The details of the influence on bacterial evolution of specialists and generalists are provided in Section 3.2.

1.5 Contents of this study

Chapter 1 describes the investigation of the process of forming GECs in ecological niches using phenotypic and genomic data of LAB to reveal bacterial evolution. The material and method for investigating the relationships between ecological niches and GECs of LAB are described in Chapter 2. Furthermore, features of genomic and phenotypic factors of LAB are described in Chapter 3. The influence of LAB’s phenotypes on their evolution to contribute to the construction of GECs in ecological niches has also been described. Moreover, the mechanism of LAB evolution in the ecosystem was applied to a model of genetic capitalism. Finally, the relationship between evolution of LAB and their ecology has been described in Chapter 4.

Chapter 2: Material and Methods

2.1 Collection of genome sequences of *Lactobacillaceae* and their features

As discussed in Chapter 1, LAB have properties suitable to investigate the relationships between niches and GECs in bacterial evolution. The group that was previously identified as the genus *Lactobacillus* in the *Lactobacillaceae* family provides an adequate sandbox. The group was selected because of enriched genomic and phenotypic data and presence of various habitats. In addition, the group is suitable for analysis of GECs because members of the group are monophyletic and closely related. These features make members in the group undergo frequent HGT because of the similarity of their genome architecture. Therefore, the data of *Lactobacillaceae* were collected as described below (Supplementary Table 2.1).

2.1.1 Genome sequences and genomic features

The genome sequences and genomic features of 178 strains, previously identified as the genus *Lactobacillus,* were retrieved from the DFAST Archive of Genome Annotation (https://dfast.nig.ac.jp/genomes/) (Tanizawa et al. 2016) database. Except for three strains, I selected type strains in which genomic and phenotypic features correspond to each other. In addition, the genome sequence of *Escherichia coli* ATCC 11775 (accession number: NZ_CP033092) was obtained from NCBI. Six genomic features (genome size, number of coding sequences (CDS), GC content, number of genes encoding rRNAs, number of genes encoding tRNAs, and number of CRISPRs) were used in this study.

2.1.2 Sequences of 16S rRNA gene

The 16S rRNA gene was chosen for this study because this ribosomal gene is traditionally used to investigate the phylogenetic relationship in bacteria. Although the phylogenetic relationship based on the 16S rRNA gene is suspected to not be robust (Sato and Miyazaki 2017), in this investigation, we use the genetic distance as a crude measure for species distance.

The sequences for the 16S rRNA genes were obtained from EZBioCloud (https://www.ezbiocloud.net/resources/16s_download)(Supplementary Table 2.1). In addition, the sequences for the 16S rRNA gene of *Escherichia coli* ATCC 11775 (accession number: NZ_CP033092) were obtained from EZBioCloud.

Because 16S rRNA genes are frequently found as multiple copies in a bacterial genome (Stoddard et al. 2015), they were not extracted from genome data. Because multiple copies make genome assembling in the region difficult, the quality of annotations and sequences for 16S rRNA genes in genome data are not high. Therefore, I selected the EZBioCloud database to obtain 16S rRNA genes.

2.1.3 Phenotypic features

Six phenotypic features of these strains were obtained from the book "Lactic Acid Bacteria: Biodiversity and Taxonomy" (Holzapfel and Wood 2014):

1. Number of sugars the strains can metabolize (sugar utilization value),

2. Growth rate at 15 °C,

3. Growth rate at 45 °C,

4. Microaerobic growth,

5. Facultatively anaerobic growth, and

6. Obligate anaerobic growth.

The sugar utilization value was calculated by counting how many types of sugars the LAB strain can utilize using a Python program. Dummy variables (1 for yes and 0 for no) were used for the other features (Supplementary Table 2.1).

2.1.4 Isolation source

Isolation sources for *Lactobacillaceae* were obtained from the paper by Zheng et al. (2020). Table 2.1 shows the correspondence between old and new species names, genomic features, phenotypic features, and isolation sources. Although genomic and phenotypic features are linked to strains, isolation sources are connected to species. Thus, some LAB have multiple isolation sources.

2.2 Analysis of genomic features

To comprehend the genomic features of *Lactobacillaceae*, I analyzed the genome sequences and 16S rRNA genes. In addition, the result data were subjected to statistical analysis, to detect GECs, and investigation of genetic capitalism in *Lactobacillaceae*.

2.2.1 Ortholog analysis

Orthologs for 178 strains of *Lactobacillaceae* were obtained using SonicParanoid software (Cosentino and Iwasaki 2019) with the default parameters. Given a set of FASTA formatted gene sequences, the software groups similar genes together as orthologs. In the resulting set, singletons were removed as strain-specific genes.

2.2.2 Core- and accessory-genome computation and COG assignment

To understand the characteristics of the LAB genomes, core genomes and accessory genes in *Lactobacillaceae* were determined. Traditionally, the definition of core genome is "the set of genes included in all genomes under investigation" (Satti et al. 2018). However, the definition has problems determining the stable core genome because of its data dependency: when more genomes are used, the number of fully shared genes declines. To avoid this effect, a certain threshold, such as "conserved in *n* percent of the genomes," needs to be used. For the determination of *n*, we need additional information.

For core and accessory-genome analysis, I used clusters of orthologous groups (COG) functional categories to classify the functions of the gene clusters for the 178 genomes of *Lactobacillaceae* (http://www.ncbi.nlm.nih.gov/COG/). Using ortholog analysis data with COG annotation, I determined the core and accessory genomes based on the method described by Satti et al. (2018). The method produces an appropriate *n*-

core, the set of genes conserved in *n* percent of the genomes, based on the COG information for the orthologs. A good parameter *n* needs to provide a robust estimation of the core genome, and the distribution of COG categories should not be susceptible to the small changes in *n*. Therefore, as a necessary condition, slight changes of *n* (e.g., *n*-1 or *n*+1) need to provide a stable distribution of COG categories.

I created 10 *n*-cores, from 100- to 91-cores, and compared the respective COG distribution of the core genome using a handmade Python program. By assessing the robustness of the core genome, a 97-core was selected, indicating that genes shared among >172 of the 178 genomes (97%) were considered the core. The method was performed using Python programs.

2.2.3 Construction of *Lactobacillaceae* phylogenetic tree

Phylogenetic trees for the 178 strains were constructed based on the 16S rRNA gene, and the genes were clustered by ortholog analysis. To generate the phylogenetic tree, MUSCLE, Multiple Sequence Alignment (Edgar 2004), and the neighbor-joining method (Saitou and Nei 1987) were implemented using the program MEGA (Kumar et al. 2018). The 16S rRNA tree was annotated using iTOL (Letunic and Bork 2007).

2.2.4 Detecting HGTs between distantly related organisms

Genes acquired via HGT were predicted by two methods based on the evolutionary distance and codon bias: the DarkHorse v2.0 (Podell and Gaasterland 2007) and COLOMBO v4.0 analysis with SIGI-HMM (Waack et al. 2006). DarkHorse and COLOMBO were run with default parameters. The CDSs were judged as HGT when their lineage probability index was ≥ 0.5 (DarkHorse), or annotation was PUTAL

(COLOMBO). While DarkHorse is based on the taxonomical group name, COLOMBO is based on codon bias. By using two different methods, the detection sensitivity of HGT increases.

2.3 Statistical analysis

To determine the tendency of the evolutionary process in LAB, statistical analyses were performed as described below.

2.3.1 Multiple regression analysis between the size of genome or number of HGT genes and *Lactobacillaceae* family features

Simple and multiple regression analysis was performed using the Python package Statsmodels (https://www.statsmodels.org/stable/). Dummy variables (1 for yes and 0 for no) were used for the following five features: growth at 15 °C, growth at 45 °C, and growth in microaerobic, facultatively anaerobic, and obligate anaerobic conditions. For the strains with missing phenotypic data, average values from all the other strains were assigned. All explanatory variables were normalized using a Z score transformation.

2.3.2 Relationship of COG ratio between ortholog groups

The COG numbers for the chosen ortholog groups were counted, and the ratio of each group was statistically analyzed using a t-test and Benjamini-Hochberg correction for multiple comparisons using the Python package Statsmodels (https://www.statsmodels.org/stable/).

2.4 Detecting GECs formed by sugar utilization in *Lactobacillaceae*

The GECs formed by the influence of the phenotype to utilize various sugars in *Lactobacillaceae* were detected as follows. To determine the GECs, I measured the average number of sugar utilization for the ortholog (ASU). Using this measure, the orthologs that were shared by generalists for sugar utilization were extracted and subjected to network analysis.

2.4.1 Calculation of ASU for the orthologs

To estimate the characteristics for each ortholog, I calculated the average number of metabolizable sugars of strains for each ortholog cluster as the Average number of Sugar Utilization for the ortholog (ASU) (Figure 2.1). Statistically meaningful orthologs were chosen based on their ASU as standard deviation of more/less than 1 from the average of sugar utilization value in the 178 strains. The COG number for the chosen orthologs was counted, and the ratio of each group was statistically analyzed, as described in Section 2.3.

ASU is a measure to confirm GECs generated by the influence of sugar utilization. The judgment of HGT among closely related species in the ortholog networks is complex. The key to this analysis is optimal ortholog selection for generating the ortholog networks. It is difficult to extract GECs in an ecological niche from ortholog networks including phylogenetic genes because the core genome makes ortholog networks become complete graphs. In this analysis using ASU value, two ortholog groups were extracted: the orthologs shared dominantly by strains that could use a variety of sugars (generalist) and those that use only a few sugars (specialist). The networks generated by these two groups were compared. If closely related species share the orthologs, the orthologs are phylogenetic genes or are shared by GECs based on the bias of phylogenetic closeness. If distantly related species share the orthologs, the orthologs are shared by GEC in ecological niches or by gene deletion in the ortholog groups.

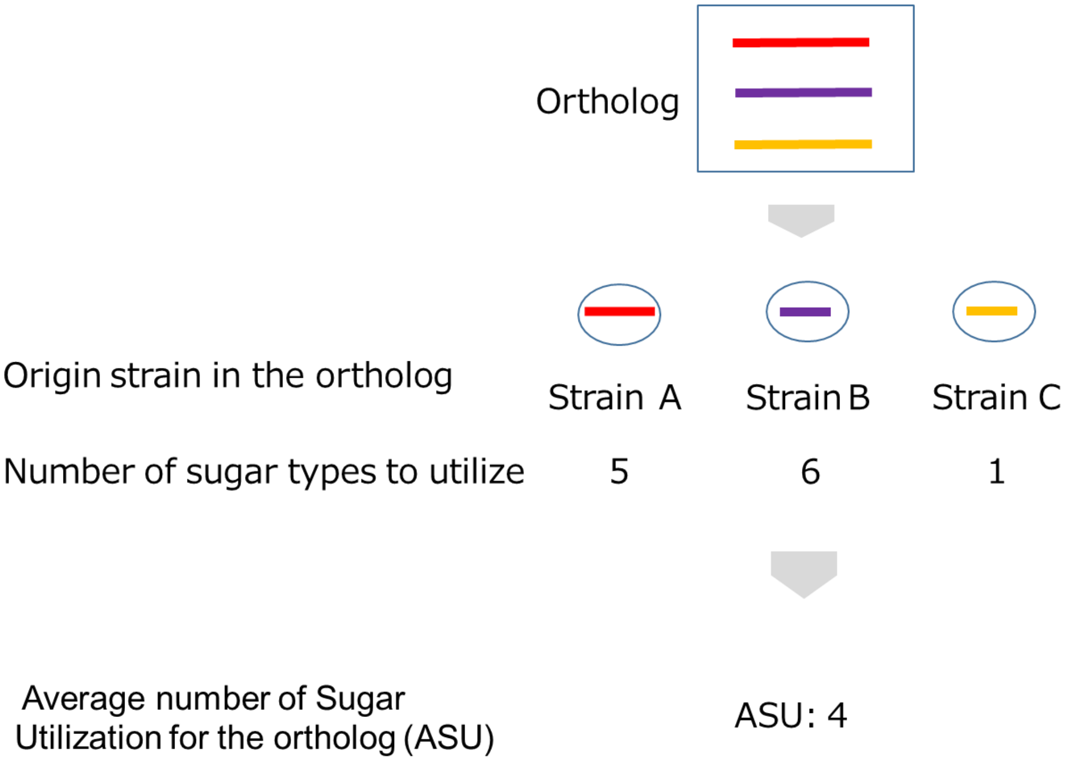

Ortholog
Origin strain in the ortholog
Strain A
Strain B
Strain C
Number of sugar types to utilize
5
6
1
Average number of Sugar Utilization for the ortholog (ASU)
ASU: 4

Figure 2.1: Average number of sugar utilization for the ortholog (ASU).

The average number of metabolizable sugars of strains for each ortholog cluster was calculated. This index was used to select the orthologs dominantly shared by strains that could use a variety of sugars (generalist) and those that used only a few sugars (specialist). The colored bars indicate the genes from each strain.

2.4.2 Construction of networks of shared orthologs

A network graph was constructed for the selected orthologs using the ASU value. Each of the 178 nodes represents a genome of *Lactobacillaceae*, and an edge was created between two genomes when the number of shared orthologs was more than five. Community extraction and visualization were performed with the Python package NetworkX (https://networkx.org/) and CytoScape (version 3.8.2) (Smoot et al. 2011), respectively.

2.5 Analysis of genetic capitalism in *Lactobacillaceae*

Analyses were performed to confirm the tendency of genetic capitalism in *Lactobacillaceae*. The gene gain/loss events were estimated based on phyletic patterns of orthologs and the phylogenetic tree of the 16S rRNA gene. The estimated values of gain/loss events were subjected to statistical analysis to elucidate whether genetic capitalism occurs in LAB.

2.5.1 Multiple sequence alignment of phyletic patterns

Orthologs for 178 strains of *Lactobacillaceae* and *E. coli* were obtained, as described in Section 2.2. Strain-specific genes were included following analysis as orthologs possessed by only one strain. Presence (1) and absence (0) profiles of orthologs (phyletic patterns) were converted to a gap-free multiple sequence alignment (MSA) using a Python program.

2.5.2 Mapping of phylogenetic tree

The estimated value of gene gain/loss events was obtained to apply the MSA of phyletic patterns and phylogenetic tree based on the 16S rRNA gene to GLOOME (Cohen et al. 2010). All parameters were set to the default. The mapped phylogenetic tree of the 16S rRNA gene with the expected value of gain/loss events was obtained from GLOOME analysis.

2.5.3 Analysis of the relationship between gain/loss events and genetic diversity

The normalized expected value of gain/loss events (E_{gl}) for each branch was calculated as follows:

$$E_{gl} = \frac{E_g + E_l}{L_b}$$

E_g indicates the expected value of gain events for each branch. E_l indicates the expected value of loss events for each branch. L_b indicates the branch length for each species.

The expected value of gain/loss events were the values mapped on the branch after the speciation of each species. The branch length is referred to from the tree of the 16S rRNA gene. The value is normalized by the branch length because the expected value depends on branch length. The expected value of gain/loss events indicates how often the bacteria have opportunities to gain and select genes in the genome.

The number of orthologs in the genome (O_n) was used as the index for genetic diversity in the bacteria after speciation. The genetic diversity in the bacteria before speciation (G_d) was calculated as follows:

$$\mathrm{G}_d = O_n - \left(E_g - E_l\right)$$

The normalized net number of the expected value of gain events (N_g) was calculated as follows:

$$N_g = \frac{E_g - E_l}{L_b}$$

Simple regression analysis was performed using the Python package Statsmodels, as in Section 2.5.1, to investigate the genetic capitalism of LAB. There were three combinations of objective and explanatory variables:

1. Normalized expected value of gain/loss events for each branch (E_{gl}) vs. the genetic diversity in bacteria before speciation (G_d),

2. Normalized expected value of gain events (N_g) vs. the genetic diversity in bacteria before speciation (G_d), and

3. Genetic diversity in bacteria after speciation (O_n) vs. the expected value of gain/loss events for each branch (E_{gl}).

The objective and explanatory variables were normalized using the Z score transformation.

Chapter 3: Niche construction and GECs in *Lactobacillaceae*

3.1 Factors affecting HGT in *Lactobacillaceae*

GECs in shared ecological niches influence microbial evolution, providing a selective advantage to microbes and allowing for their expansion into new ecological niches (Soucy et al. 2015; Swithers et al. 2012). However, this complicates the evolution or adaptation within the same GECs (Polz et al. 2013). Ragan and Beiko (2009) suggested that the habitats of donors and recipients are key limitations for HGT. I further investigated the impacts of how environmental range constrains HGT because they may have been previously underestimated.

To better understand the influence of ecological niches on HGT, the relationship of the phenotypes of the microorganism with environmental adaptation should be investigated. Phenotypes such as those for resource utilization enable microbes to survive in various environments and thus help define the range of the habitat of microbes (Chen et al. 2021). Jain et al. (2003) investigated the internal and external environmental factors that regulate HGT in eight bacterial and archaeal genomes. They reported that HGT occurs among organisms with similar characteristics, including host phenotypes, such as carbon utilization and oxygen tolerance. Their analyses provided evidence for the effects of GECs in ecological niches on prokaryote evolution. However, it is unclear if this tendency applies to GECs formed by bacterial groups of the same family in particular ecosystem niches. This is because the HGT among related bacterial groups is affected not only by the bias of the ecological niche they share but also by the bias of their closely related partners with whom they preferentially exchange genes (Andam and Gogarten 2011; Soucy et al. 2015). To clarify this point in more detail, a comparative analysis using a large amount of phenotypic and genomic data for related species is required.

3.1.1 Relationships among the phylogenetic, genomic, and phenotypic features of 178 strains from *Lactobacillaceae*

I first examined the phenotypic and genomic features of each of the 178 strains and mapped them onto a phylogenetic tree (Figure 3.1). Six phenotypes were assessed: two conditions for temperature required for bacterial growth (ability to grow at 15 °C and 45 °C), three levels of oxygen tolerance (microaerobic, facultatively anaerobic, and obligate anaerobic), and sugar utilization value (number of sugars each strain can metabolize). Of the 178 strains, 56.8% grew at 15 °C and 33.3% grew at 45 °C. Furthermore, among these 178 strains, 8.3%, 81.9%, and 9.8% were microaerobic, facultatively anaerobic, and obligate anaerobic, respectively. Sugar utilization values ranged from 0 to 17 (excluding glucose), and the average for all strains was 6.83. For the genomic features, I investigated the total CDS number and estimated the number of CDS gained via HGT for each strain. The total number of CDS for each of the 178 strains ranged from 1191 to 3600. Because the total number of CDS and the genome size were strongly correlated (R = 0.976) (Figure 3.2), they were treated as interchangeable information in this analysis. The number of CDS gained via HGT ranged from 17 to 342 (Supplementary Table 2.1) and indicated a weak correlation between genome size (R = 0.394) and the total number of CDS (R = 0.424) (Figure 3.3).

Variation was observed in the phenotypic features of the groups clustered by the phylogenetic tree (Figure 3.1). In particular, the sugar utilization values varied even within the same genus. For example, in the group for the genus *Lactobacillus*, although *Lactobacillus iners* had sugar-type utilization profile of 0, *Lactobacillus hamster* could utilize 14 kinds of sugar. Additionally, sugar utilization values of the *Ligilactobacillus* genus ranged from 1 to 15, and that of the *Limosilactobacillus* genus ranged from 1 to 16.

The correspondence between the numbers of CDS in a genome and the sugar utilization values was observed (Figure 3.1). The tendency was remarkable in the clusters for the genera *Ligilactobacillus*, *Lacticaseibacillus*, *Limosilactobacillus*, *Apilactobacillus*, *Fructilactobacillus,* and *Secundilactobacillus*. For example, *Lacticaseibacillus manihotivorans*, *Lacticaseibacillus saniviri*, *Lacticaseibacillus casei,* and *Lacticaseibacillus paracasei* ssp. *paracasei* had high numbers of CDS and high sugar utilization values, whereas *Lacticaseibacillus nasuensis*, *Lacticaseibacillus thailandensis*, and *Lacticaseibacillus brantae* had low numbers of CDS and low sugar utilization values.

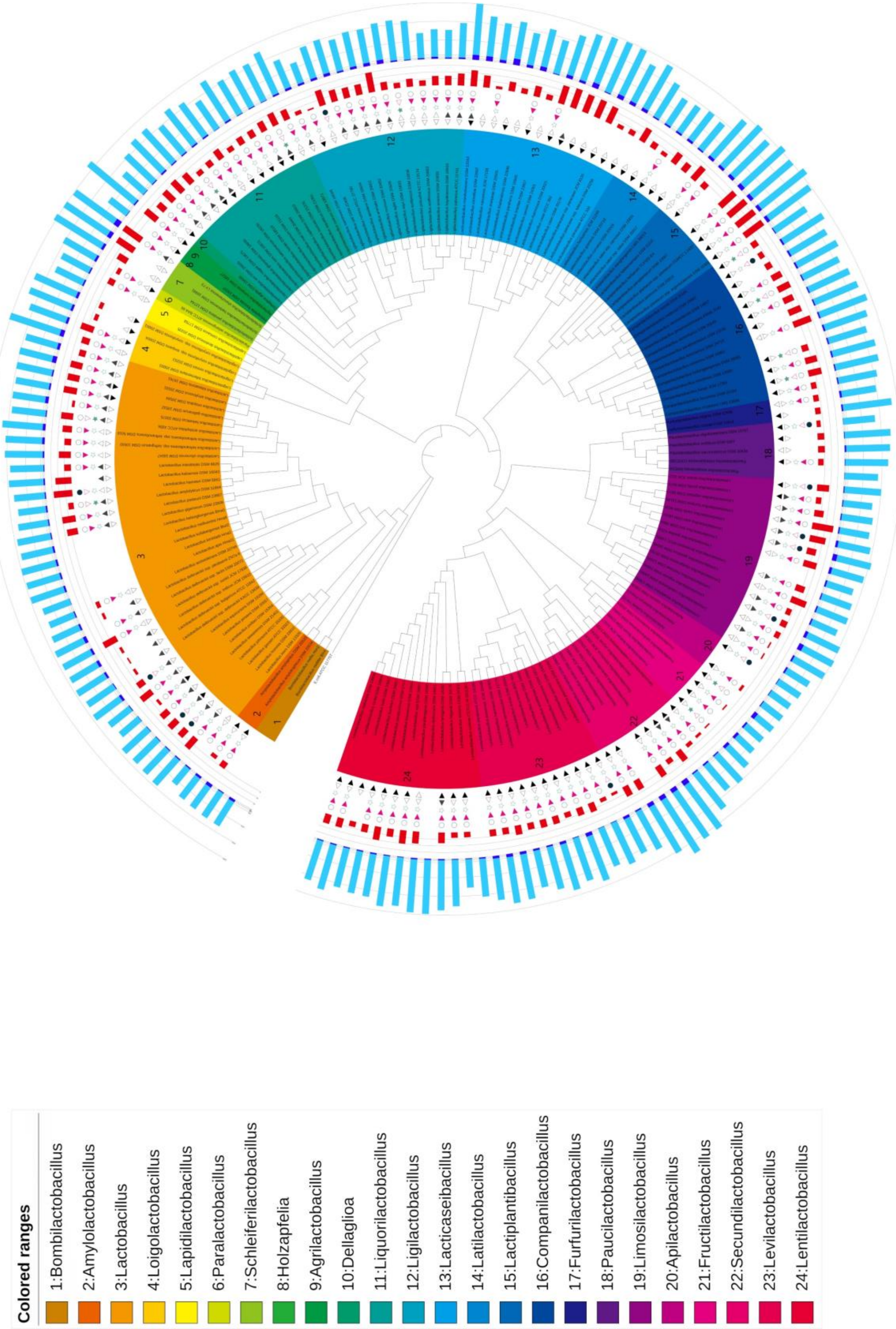

Colored ranges
1:Bombilactobacillus
2:Amylolactobacillus
3:Lactobacillus
4:Loigolactobacillus
5:Lapidilactobacillus
6:Paralactobacillus
7:Schleiferilactobacillus
8:Holzapfelia
9:Agrilactobacillus
10:Dellaglioa
11:Liquorilactobacillus
12:Ligilactobacillus
13:Lacticaseibacillus
14:Latilactobacillus
15:Lactiplantibacillus
16:Companilactobacillus
17:Furfurilactobacillus
18:Paucilactobacillus
19:Limosilactobacillus
20:Apilactobacillus
21:Fructilactobacillus
22:Secundilactobacillus
23:Levilactobacillus
24:Lentilactobacillus

Figure 3.1: Phylogenetic tree based on the 16S rRNA genes of LAB strains with phenotypic and genomic features identified.

The inner band shows species colored by genus. The next five symbols show phenotypic characteristics for each LAB strain: the inward-facing triangle indicates growth at 15 °C; the outward-facing triangle indicates growth at 45 °C; the star indicates micro aerophilic; the red inward-facing symbol indicates facultatively anaerobic; the circle indicates obligate anaerobic. A filled symbol means a strain has the phenotype, and an open symbol indicates it does not. A blank means that there is no relevant information available. The next red band shows the number of sugar types that could be utilized. The outer bands indicate the number of coding sequences (CDS) for each strain: navy blue is the estimated number of CDS acquired by horizontal gene transfer (HGT), and light blue is the number of native CDS. This figure was adapted from Takenaka et al. (2021).

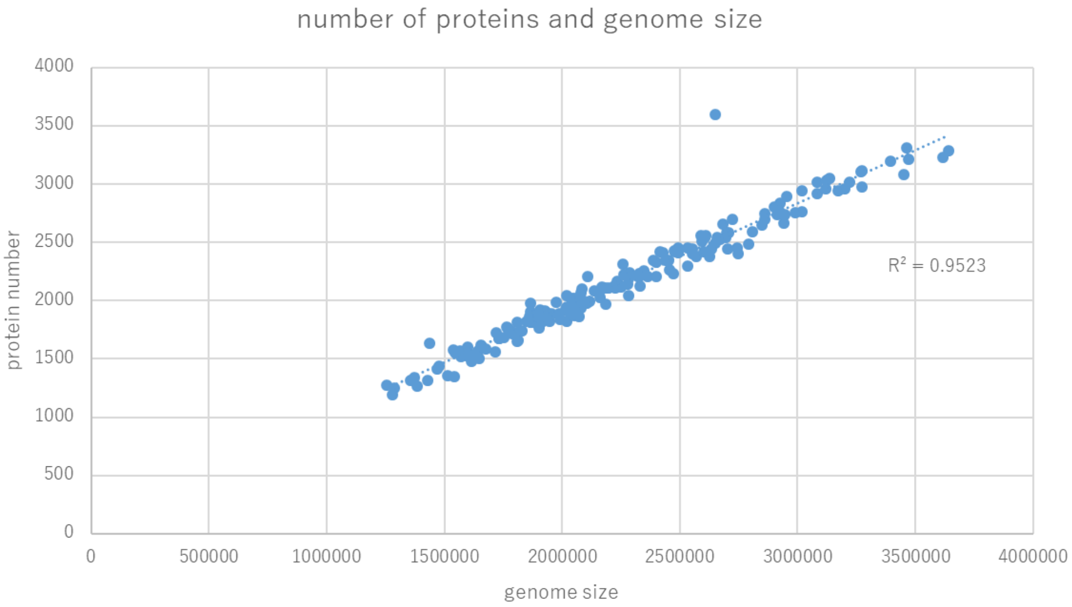


Figure 3.2: Correlation between the number of proteins and genome size

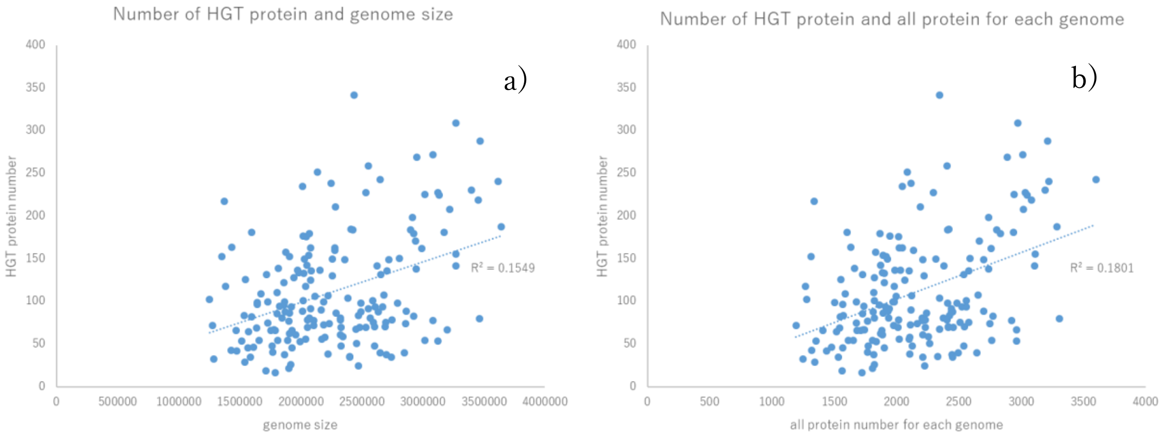


Figure 3.3: Horizontal gene transfer (HGT) protein number with (a) genome size and (b) total number of proteins for each genome.

3.1.2 Influence of phenotypic features on genome size and number of HGT genes

Multiple regression analyses were performed to confirm the relationship between genomic features and sugar utilization, as shown in Figure 3.1. The six phenotypes (sugar utilization value, growth at 15 °C, growth at 45 °C, and growth in microaerobic, facultatively anaerobic, and obligate anaerobic conditions) and four genomic features (G/C content, number of rRNA genes, number of tRNA genes, and number of CRISPRs) were subjected to multiple regression analysis as explanatory variables (Supplementary Table 2.1, Figure 3.4).

The genome sizes of 178 strains in *Lactobacillaceae* were used as the objective variable. The six phenotypic and four genomic features were used as the explanatory variables. The coefficient of determination (R2) obtained was 0.484, and the correlation coefficient (R) was 0.696. For sugar utilization values, growth at 15 °C, growth at 45 °C, G/C content, and number of CRISPRs, P-value was < 0.05. The coefficient for growth at 45 °C was negative, whereas that for G/C content, growth at 15 °C, and the number of CRISPRs was positive. The sugar utilization value had the largest coefficient among these factors (Figure 3.5(a)).

CDS that were transferred from other taxa (HGT gene) were also set as an objective variable, and the 10 factors used to analyze the genome size were used as explanatory variables. As a result, the coefficient of determination (R2) obtained was 0.298, and the correlation coefficient (R) was 0.546. For both the sugar utilization value and the G/C composition, P-value was < 0.05, and they had a positive correlation (Figure 3.5(b)).

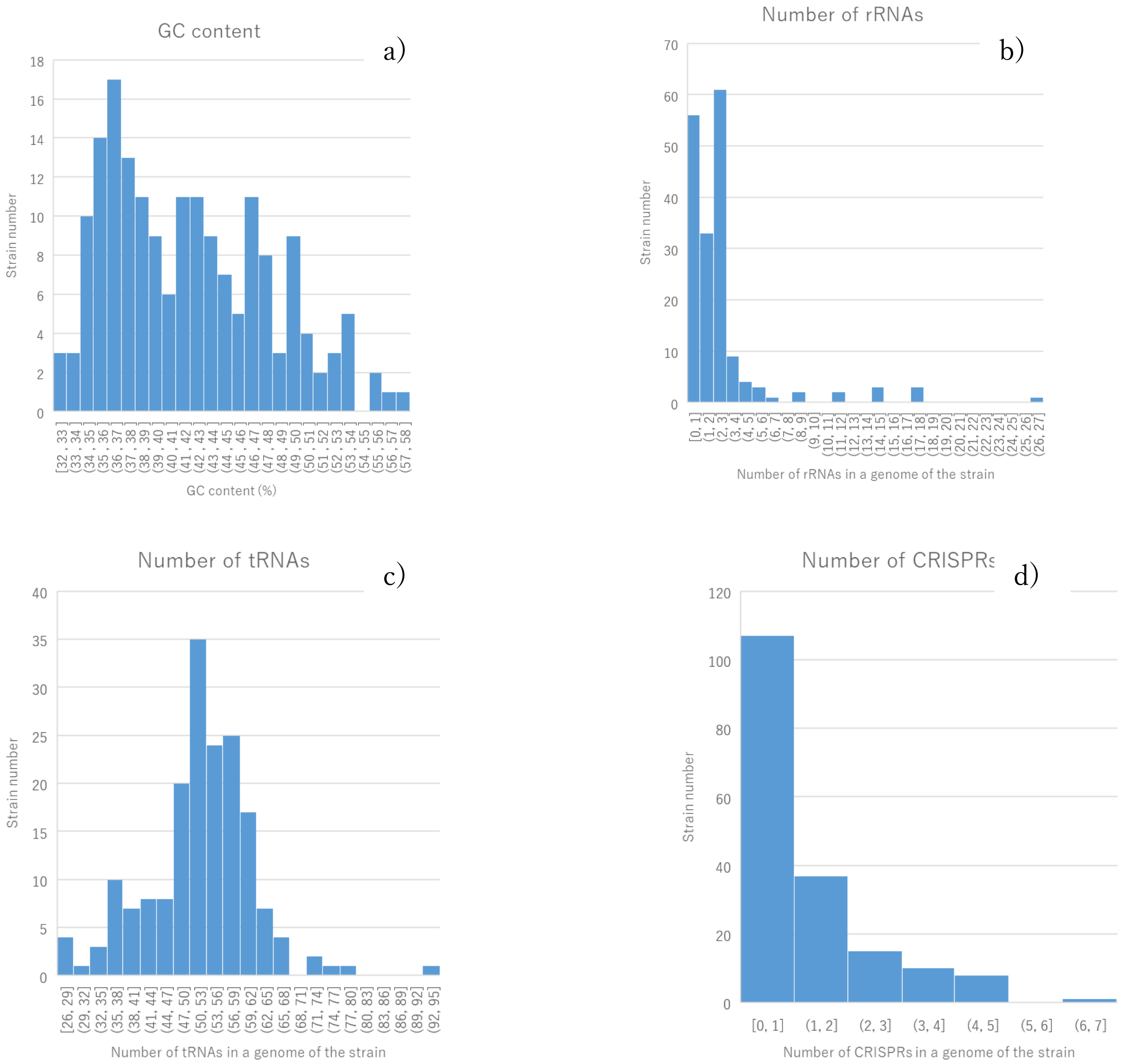


Figure 3.4: a) GC content; b) number of rRNAs; c) number of tRNAs; and number of CRISPRs in genomes of *Lactobacillaceae*.

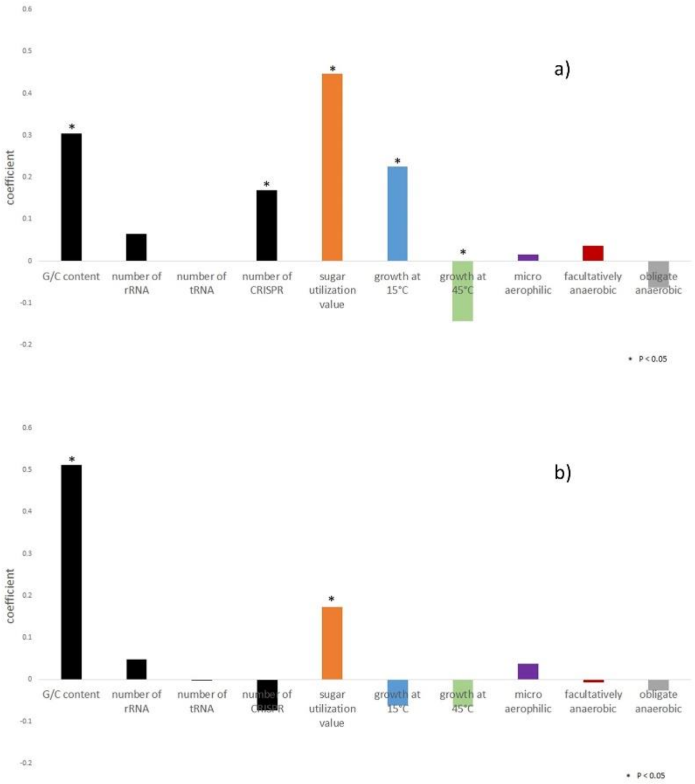


Figure 3.5: Values of coefficients of multiple aggression analysis for a) genome size and b) number of CDS judged to be HGTs.

The genome size or number of CDS judged to be HGTs was set as the objective variable, and the six phenotypic features (sugar utilization value, growth at 15 °C, growth at 45 °C, microaerobic, facultatively anaerobic, and obligate anaerobic) and four genomic features (G/C content, number of rRNAs, number of tRNAs, and number of CRISPRs) were subjected to multiple regression analysis as explanatory variables. * indicates a P-value ≤ 0.05. This figure is adapted from Takenaka et al. (2021).

3.1.3 Influence of sugar utilization phenotype on HGT in *Lactobacillaceae*

Section 3.1 indicates that various sugar utilization and GC content influence HGT frequency in LAB. This result is the first evidence that the phenotype to utilize a variety of sugars influences HGT frequency in LAB strains.

The phenotypes for carbon utilization and oxygen tolerance were previously shown to influence HGT (Jain et al. 2003). However, our results did not support this. Instead, the sugar utilization value, which means the number of sugar types that can be utilized, was found to contribute frequently to HGT. The sugar utilization values in this study differed from the carbon utilization feature defined as heterotroph or autotroph in the previous study. The gaps in optimum conditions for growth in the laboratory and environment may hide possible effects on HGT (Jain et al. 2003). However, as all LAB are heterotrophic organisms, I did not analyze this factor. In addition, no HGT was related to oxygen tolerance, but there was a bias as approximately 80% of the strains in this study were facultatively anaerobic. This may have prevented the detection of a correlation between oxygen tolerance and HGT. The results of Jain et al. may be different because they investigated HGT across domains (empires), whereas I investigated HGT in the same family.

The G/C content in the genome of *Lactobacillaceae* was correlated with the number of HGT (Figure 3.5). HGT occurs among microorganisms with similar genomic G/C contents and it could affect the incorporation of new DNA into microorganisms (Jain et al. 2003). Genomes from the bacterial groups from the phylum Firmicutes, which includes the family *Lactobacillaceae*, have low G/C contents. Foreign genes from outside the phylum Firmicutes may have a higher G/C content, which was correlated with the number of HGT genes and genome size.

In summary, these results suggest that factors influencing HGT are sugar utilization and G/C content in LAB. In particular, sugar utilization may contribute to constructing an ecological niche and forming GECs because resource utilization enables microbes to help define the range of the microbes' habitat (Chen et al. 2021).

3.2 GECs in *Lactobacillaceae*

In Section 3.1, the results indicate that sugar utilization influences HGT among LAB. This suggests that sugar utilization contributes to constructing ecological niches and forming GEC. Ability to utilize a variety of sugars expands the range of the habitat of LAB, increasing the potential of HGT and thereby forming GECs.

Sugar utilization in bacteria has a large role in determining survival in a niche. Bacteria that have genes encoding enzymes that utilize particular carbon sources dominate the environment which contains enrich of the carbon source (Kirchman 2012). For example, bacteria that can use fructose are often found in niches enriched with fructose, such as flowers and fruits (Endo et al. 2009).

Bacteria that can utilize a wide range of sugars may be regarded as generalists because resource utilization helps define the range of habitats for the microbes (Chen et al. 2021). Sriswasdi et al. (2017) reported that generalists maintain the diversity of species and drive bacterial evolution to adapt to a wide range of environments while specializing in particular niches. Considering HGT, the effect of generalists on bacterial evolution is larger.

LAB construct ecological niches in an environment with enriched nutrients; they inhabit fermented dairy products, plants, and meat (Caplice and Fitzgerald 1999). Various lineages of LAB including the *Enterococcaceae*, *Leuconostocaceae*, and *Lactobacillaceae* family construct ecological niches in silage (Cai 1999, Cai et al. 1998). LAB share ecological niches across families.

Bacteria may form GECs in ecological niches across distant lineages in *Lactobacillaceae.* However, besides sharing ecological niches, biases that form GECs include symbiotic interactions and phylogenetic closeness. Phylogenetic closeness greatly influences GECs in *Lactobacillaceae* because the GECs formed include closely related species. The results in Section 3.1 suggested correspondence between sugar utilization, a phenotype that is associated with niches, and HGT frequency. To obtain a deeper perspective of GECs in ecological niches, the relationship between phylogenetic closeness and GECs in the ecological niche should be integrated into this analysis.

In Section 3.2, I investigated how sugar utilization forms the GEC in *Lactobacillaceae*, a group of closely related species. Thereafter, I detected HGT among *Lactobacillaceae* strains by combining ortholog and network analyses because the above-mentioned methods (DarkHorse and COLOMBO software) are suitable only for detecting HGTs between distantly related organisms. To analyze the relationship between sugar utilization and GECs, I introduced the concept of ASU.

3.2.1 COG ratios of orthologs in the core and accessory genome

To understand the characteristics of HGT genes in *Lactobacillaceae*, I focused on "accessory genomes." The variable portion of the genome between individual strains is often called the "accessory genome" and differs from the core genome (Sim et al. 2008). Here, I compared the functions of accessory genomes, except for strain-specific singletons, to those of core genomes.

To classify all genes into core and accessory genomes, I first conducted an ortholog analysis for the CDS present in the 178 strains; as a result, 384,737 putative protein sequences were grouped into 12,884 ortholog clusters. The core and accessory genomes were determined using the COG assignment of each ortholog; 532 and 12,352 ortholog clusters corresponded to the core and accessory genes, respectively. The COG ratios of the core and accessory genomes were quite different (Figure 3.6). Metabolism-related genes were enriched in the accessory genomes.

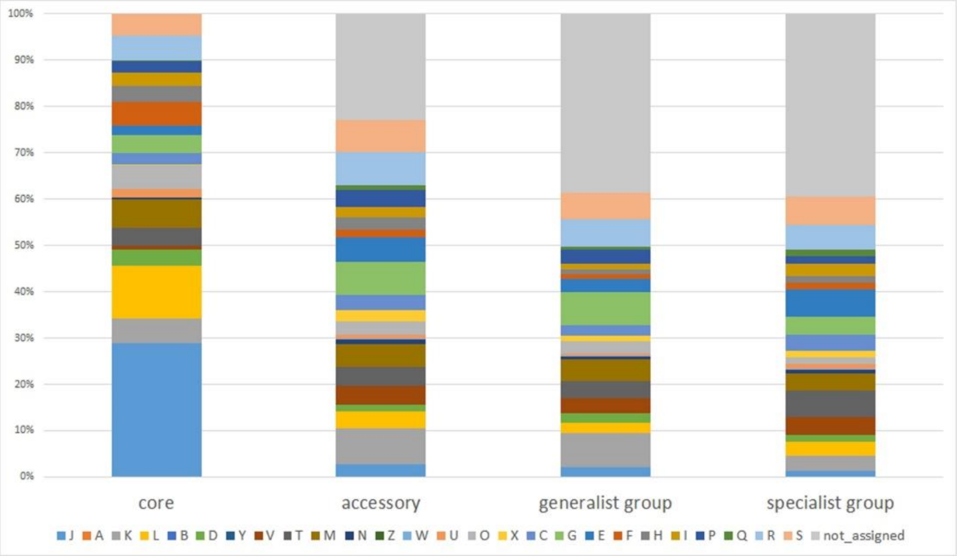


Figure 3.6: Clusters of orthologous groups (COG) ratios for each group of orthologs.

The COG ratios of the core genome, accessory genome, generalist group orthologs, and specialist group orthologs are displayed. Orthologs not assigned COG are indicated in gray. More metabolism-related genes, such as “carbohydrate transport and metabolism” (G), “amino acid transport and metabolism” (E), “transcription” (K), and “defense mechanisms” (V) were enriched in the accessory genome than in the core genome. However, “translation, ribosomal structure, and biogenesis” (J) and “replication, recombination and repair” (L) occurred less in the accessory genome than in the core genome. Figure adapted from Takenaka et al. (2021).

[J] Translation, ribosomal structure, and biogenesis,
[A] RNA processing and modification,
[K] Transcription,
[L] Replication, recombination, and repair,
[B] Chromatin structure and dynamics,
[D] Cell cycle control, cell division, chromosome partitioning,
[Y] Nuclear structure,
[V] Defense mechanisms,
[T] Signal transduction mechanisms,
[M] Cell wall/membrane/envelope biogenesis,
[N] Cell motility,
[Z] Cytoskeleton,
[W] Extracellular structures,
[U] Intracellular trafficking, secretion, and vesicular transport,
[O] Posttranslational modification, protein turnover, chaperones,

[X] Mobilome: prophages, transposons,
[C] Energy production and conversion,
[G] Carbohydrate transport and metabolism,
[E] Amino acid transport and metabolism,
[F] Nucleotide transport and metabolism,
[H] Coenzyme transport and metabolism,
[I] Lipid transport and metabolism,
[P] Inorganic ion transport and metabolism,
[Q] Secondary metabolites biosynthesis, transport, and catabolism,
[R] General function prediction only,
[S] Function unknown.

3.2.2 Ortholog features shared by generalists or specialists for sugar utilization

To confirm that sugar utilization values influence the HGT bias, two groups of orthologs were compared: the orthologs shared dominantly by strains that were able to use a variety of sugars (generalist) and those that use only a few sugars (specialist). Here, the ASU value was used to extract generalist and specialist group orthologs as follows (see also material and methods).

1. The overall average and standard deviation of the sugar utilization values in all 178 strains were calculated.

2. The generalist/specialist orthologs were selected when they had ASU values that were more or /less than the mean ± 1 standard deviation (Figure 3.7).

The ratio of the COG functions between the generalist and specialist group orthologs showed no significant differences (Figure 3.6, Table 3.1), but more strains shared the generalist orthologs. This suggests that the genes are neutrally acquired by HGT regardless of the phenotypic differences.

Among the generalist orthologs, some genes were involved in adaptations to various niches (Table 3.2).

- Stress response: Cell division protein FtsK (Diez et al. 2000), xenobiotic response element (XRE) family transcriptional regulator (Hu et al. 2018), and phenolic acid-responsive transcriptional regulator (PadR) family (Gury et al. 2004).

- Antibiotics: bacteriocin precursor peptides PlnE and PlnF (Anderssen et al. 1998) and multiple antibiotic resistance protein (MarR) family transcriptional regulator (Silva et al. 2018).

- Detoxification: peptide methionine sulfoxide reductase (Walter et al. 2005), mercuric resistance operon regulatory protein (MerR) family transcriptional regulator (Brown et al. 2003), and arsenical resistance operon repressor (ArsR) family transcriptional regulators (Wu and Rosen 1991) for heavy metal resistance.

- Sugar utilization: L-fucose isomerase is involved in the carbohydrate metabolism of bacteria (Seemann and Schulz 1997).

Indeed, phylogenetic trees of these orthologs conflicted with the trees of the host lineages, suggesting HGT events (Figure 3.8).

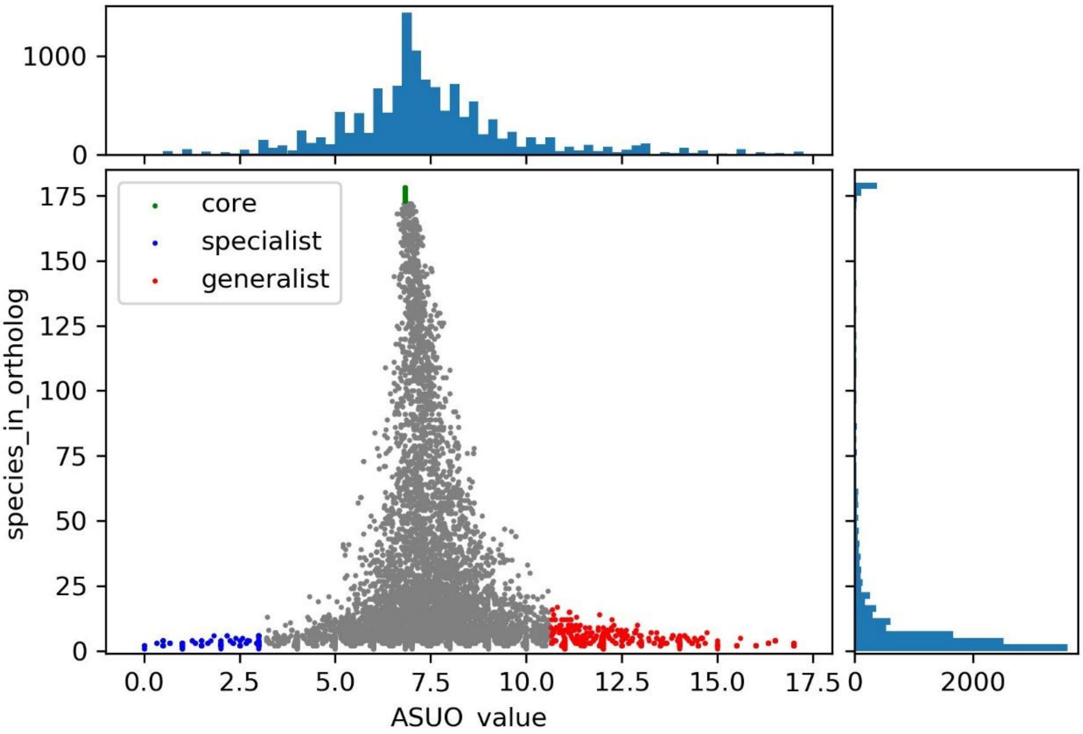


Figure 3.7: ASU value and number of strains for each ortholog.

The vertical axis indicates the number of strains in each ortholog, and the horizontal axis indicates the ASU value for each ortholog. We introduced the concept of ASU (average of sugar utilization for the ortholog) value. For example, two sequences derived from strains A and B were clustered as an ortholog and then their ASU value was calculated as the average sugar utilization value for A and B. We calculated the overall average and standard deviation of the sugar utilization value in 178 strains. Then ortholog clusters were chosen when their ASU values were more/less than the means ± one standard deviation. The orthologs with high ASU values were designated generalist group orthologs (red dots), and the low-value group was designated specialist group orthologs (blue dots). Core genes from the 178 LAB strains are indicated as green dots. The top and side histograms show the number of orthologs on each axis. Figure adapted from Takenaka et al. (2021).

Figure 3.8(a)

XRE family transcriptional regulator

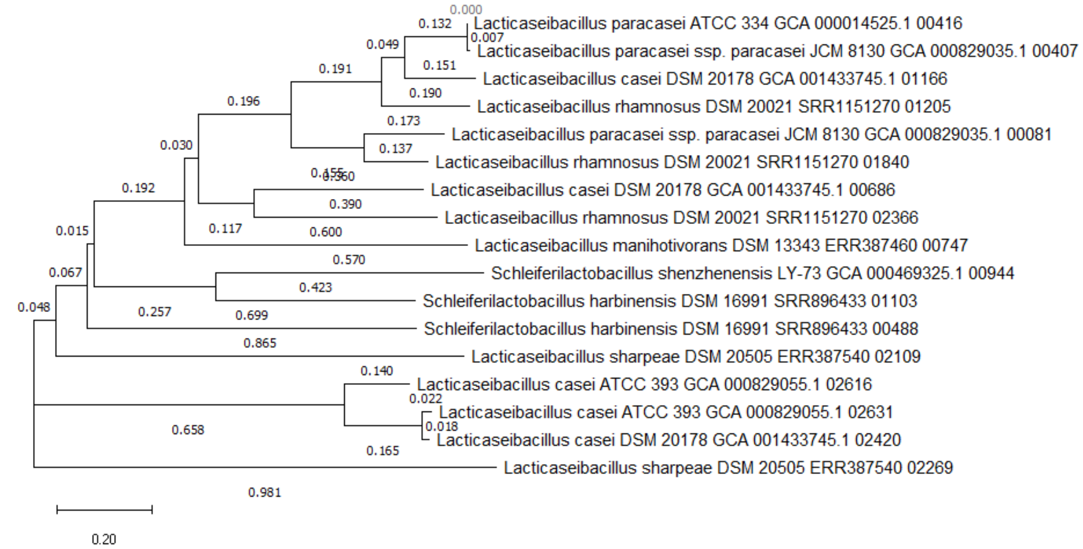

0.000
0.132
Lacticaseibacillus paracasei ATCC 334 GCA 000014525.1 00416
0.007
0.049
Lacticaseibacillus paracasei ssp. paracasei JCM 8130 GCA 000829035.1 00407
0.191
0.151
Lacticaseibacillus casei DSM 20178 GCA 001433745.1 01166
0.196
0.190
Lacticaseibacillus rhamnosus DSM 20021 SRR1151270 01205
0.173
Lacticaseibacillus paracasei ssp. paracasei JCM 8130 GCA 000829035.1 00081
0.030
0.137
Lacticaseibacillus rhamnosus DSM 20021 SRR1151270 01840
0.155
0.360
0.192
Lacticaseibacillus casei DSM 20178 GCA 001433745.1 00686
0.390
Lacticaseibacillus rhamnosus DSM 20021 SRR1151270 02366
0.015
0.117
0.600
Lacticaseibacillus manihotivorans DSM 13343 ERR387460 00747
0.570
0.067
Schleiferilactobacillus shenzhenensis LY-73 GCA 000469325.1 00944
0.423
0.048
Schleiferilactobacillus harbinensis DSM 16991 SRR896433 01103
0.257
0.699
Schleiferilactobacillus harbinensis DSM 16991 SRR896433 00488
0.865
Lacticaseibacillus sharpeae DSM 20505 ERR387540 02109
0.140
Lacticaseibacillus casei ATCC 393 GCA 000829055.1 02616
0.022
Lacticaseibacillus casei ATCC 393 GCA 000829055.1 02631
0.658
0.018
Lacticaseibacillus casei DSM 20178 GCA 001433745.1 02420
0.165
Lacticaseibacillus sharpeae DSM 20505 ERR387540 02269
0.981
0.20

Figure 3.8(b)

Integral membrane protein PlnU

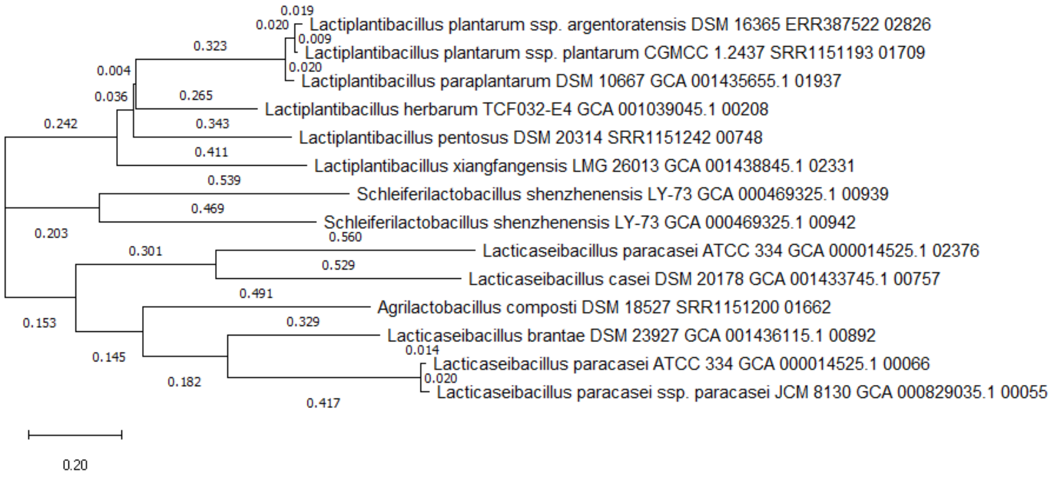

0.019
0.020
Lactiplantibacillus plantarum ssp. argentoratensis DSM 16365 ERR387522 02826
0.323
0.009
Lactiplantibacillus plantarum ssp. plantarum CGMCC 1.2437 SRR1151193 01709
0.004
0.020
Lactiplantibacillus paraplantarum DSM 10667 GCA 001435655.1 01937
0.036
0.265
Lactiplantibacillus herbarum TCF032-E4 GCA 001039045.1 00208
0.242
0.343
Lactiplantibacillus pentosus DSM 20314 SRR1151242 00748
0.411
Lactiplantibacillus xiangfangensis LMG 26013 GCA 001438845.1 02331
0.539
Schleiferilactobacillus shenzhenensis LY-73 GCA 000469325.1 00939
0.469
0.203
Schleiferilactobacillus shenzhenensis LY-73 GCA 000469325.1 00942
0.560
0.301
Lacticaseibacillus paracasei ATCC 334 GCA 000014525.1 02376
0.529
Lacticaseibacillus casei DSM 20178 GCA 001433745.1 00757
0.491
Agrilactobacillus composti DSM 18527 SRR1151200 01662
0.153
0.329
Lacticaseibacillus brantae DSM 23927 GCA 001436115.1 00892
0.014
0.145
Lacticaseibacillus paracasei ATCC 334 GCA 000014525.1 00066
0.182
0.020
Lacticaseibacillus paracasei ssp. paracasei JCM 8130 GCA 000829035.1 00055
0.417
0.20

Figure 3.8(c)

MerR family transcriptional regulator

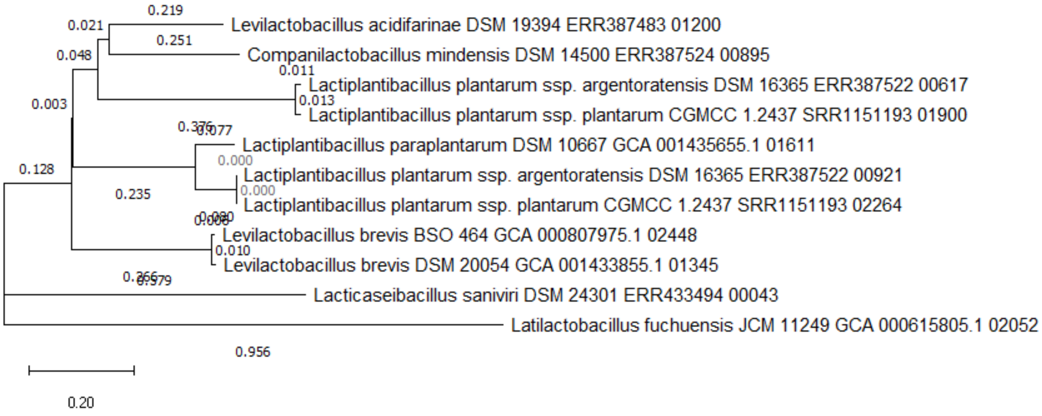

0.219
0.021
Levilactobacillus acidifarinae DSM 19394 ERR387483 01200
0.251
0.048
Companilactobacillus mindensis DSM 14500 ERR387524 00895
0.011
Lactiplantibacillus plantarum ssp. argentoratensis DSM 16365 ERR387522 00617
0.003
0.013
Lactiplantibacillus plantarum ssp. plantarum CGMCC 1.2437 SRR1151193 01900
0.376
0.077
Lactiplantibacillus paraplantarum DSM 10667 GCA 001435655.1 01611
0.000
0.128
Lactiplantibacillus plantarum ssp. argentoratensis DSM 16365 ERR387522 00921
0.235
0.000
Lactiplantibacillus plantarum ssp. plantarum CGMCC 1.2437 SRR1151193 02264
0.000
0.080
Levilactobacillus brevis BSO 464 GCA 000807975.1 02448
0.010
Levilactobacillus brevis DSM 20054 GCA 001433855.1 01345
0.266
0.579
Lacticaseibacillus saniviri DSM 24301 ERR433494 00043
Latilactobacillus fuchuensis JCM 11249 GCA 000615805.1 02052
0.956
0.20

Figure 3.8(d)

L-fucose isomerase

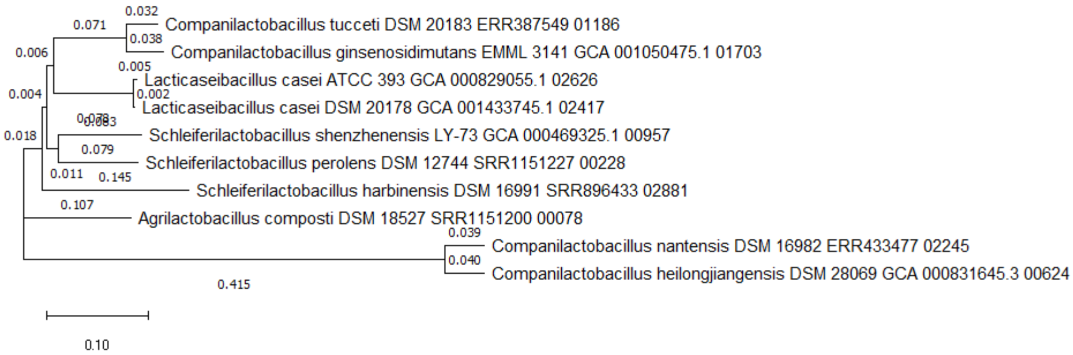

0.032
0.071
Companilactobacillus tucceti DSM 20183 ERR387549 01186
0.006
0.038
Companilactobacillus ginsenosidimutans EMML 3141 GCA 001050475.1 01703
0.005
Lacticaseibacillus casei ATCC 393 GCA 000829055.1 02626
0.004
0.002
Lacticaseibacillus casei DSM 20178 GCA 001433745.1 02417
0.078
0.083
0.018
Schleiferilactobacillus shenzhenensis LY-73 GCA 000469325.1 00957
0.079
Schleiferilactobacillus perolens DSM 12744 SRR1151227 00228
0.011
0.145
Schleiferilactobacillus harbinensis DSM 16991 SRR896433 02881
0.107
Agrilactobacillus composti DSM 18527 SRR1151200 00078
0.039
Companilactobacillus nantensis DSM 16982 ERR433477 02245
0.040
Companilactobacillus heilongjiangensis DSM 28069 GCA 000831645.3 00624
0.415
0.10

Figure 3.8(e)

MarR family transcriptional regulator

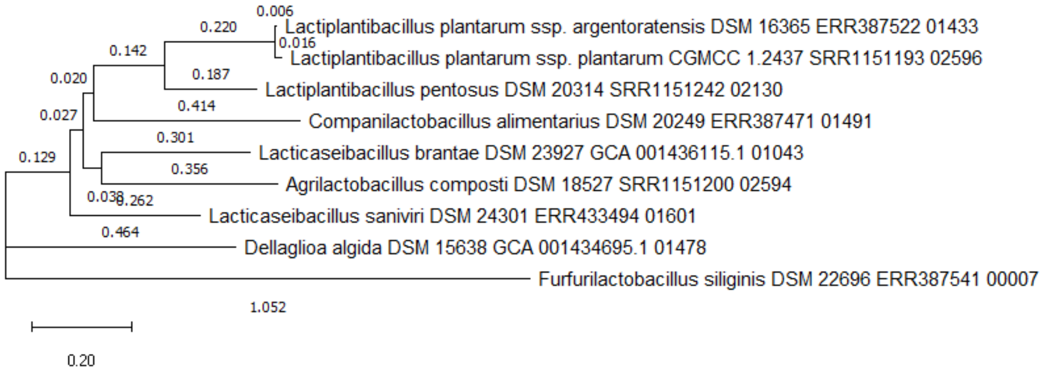

0.006
0.220
Lactiplantibacillus plantarum ssp. argentoratensis DSM 16365 ERR387522 01433
0.016
0.142
Lactiplantibacillus plantarum ssp. plantarum CGMCC 1.2437 SRR1151193 02596
0.020
0.187
Lactiplantibacillus pentosus DSM 20314 SRR1151242 02130
0.414
0.027
Companilactobacillus alimentarius DSM 20249 ERR387471 01491
0.301
0.129
Lacticaseibacillus brantae DSM 23927 GCA 001436115.1 01043
0.356
Agrilactobacillus composti DSM 18527 SRR1151200 02594
0.038
0.262
Lacticaseibacillus saniviri DSM 24301 ERR433494 01601
0.464
Dellaglioa algida DSM 15638 GCA 001434695.1 01478
Furfurilactobacillus siliginis DSM 22696 ERR387541 00007
1.052
0.20

Figure 3.8: Conflicting phylogenetic trees compared to the original lineage for the generalist group orthologs.

a. Xenobiotic response element (XRE) family transcriptional regulator. The clade *Lacticaseibacillus* included genes from *Schleiferilactobacillus*.

b. Integral membrane protein PlnU. The clade *Lacticaseibacillus* included genes from *Agrilactobacillus composti*.

c. Mercuric resistance operon regulatory protein (MerR) family transcriptional regulator. The clade *Companilactobacillus* was mixed with *Levilactobacillus.*

d. L-fucose isomerase. *Companilactobacillus* genes were far split.

e. Multiple antibiotic resistance protein (MarR) family transcriptional regulator. *Lacticaseibacillus* genes were split.

Scale bars are amino acid substitutions per position. Figure adapted from Takenaka et al. (2021).

Table 3.1: T-test and Benjamini-Hochberg method used to compare the functional ratio of COG for each group.

The right side of the table indicates the P-value for the t-test comparing COG ratios between all combinations to choose two from three groups (accessory genome, generalist group orthologs, specialist group orthologs). The left side of the table indicates the Boolean values of the Benjamini-Hochberg correction at a 0.05 false discovery rate (FDR) level. Significant differences indicate TRUE. Table adapted from Takenaka et al. (2021).

	P-value			t-test and Benjamini-Hochberg method		
COG	All accessory vs. generalist	All accessory vs. specialist	Generalist vs. specialist	All accessory vs. generalist	All accessory vs. specialist	Generalist vs. specialist
J	0.326101	0.114384	0.32189	FALSE	FALSE	FALSE
A	0.770197	0.86256	ND	FALSE	FALSE	FALSE
K	0.660644	0.001324	0.005024	FALSE	TRUE	FALSE
L	0.016087	0.454098	0.458151	FALSE	FALSE	FALSE
B	ND	ND	ND	FALSE	FALSE	FALSE
D	0.233915	0.902782	0.498252	FALSE	FALSE	FALSE
Y	ND	ND	ND	FALSE	FALSE	FALSE
V	0.253986	0.908512	0.590247	FALSE	FALSE	FALSE
T	0.546536	0.086224	0.073969	FALSE	FALSE	FALSE
M	0.609181	0.285109	0.484595	FALSE	FALSE	FALSE
N	0.330625	0.666394	0.873454	FALSE	FALSE	FALSE
Z	ND	ND	ND	FALSE	FALSE	FALSE
W	0.795567	0.973348	0.906121	FALSE	FALSE	FALSE
U	0.164648	0.519524	0.133258	FALSE	FALSE	FALSE
O	0.74121	0.073661	0.129009	FALSE	FALSE	FALSE
X	0.003727	0.155424	0.688248	FALSE	FALSE	FALSE
C	0.115125	0.690668	0.197208	FALSE	FALSE	FALSE
G	0.971753	0.014538	0.025503	FALSE	FALSE	FALSE
E	0.000799	0.679508	0.012048	TRUE	FALSE	FALSE
F	0.062515	0.913128	0.279673	FALSE	FALSE	FALSE
H	0.002552	0.136954	0.679383	TRUE	FALSE	FALSE
I	0.018139	0.633887	0.046447	FALSE	FALSE	FALSE
P	0.275201	0.034176	0.140896	FALSE	FALSE	FALSE
Q	0.159491	0.383424	0.094268	FALSE	FALSE	FALSE
R	0.149804	0.147752	0.581598	FALSE	FALSE	FALSE
S	0.145587	0.624075	0.713207	FALSE	FALSE	FALSE
not_assigned	0	0	0.804117	TRUE	TRUE	FALSE

Table 3.2: Annotation of genes in generalist and specialist group orthologs.

The table indicates the genes present in each group of orthologs, and these annotations were based on the genome data from the DFAST Archive of Genome Annotation. Table adapted from Takenaka et al. (2021).

Generalist group ortholog	
16S rRNA methyltransferase	major facilitator superfamily transporter
3',5'-cyclic-nucleotide phosphodiesterase	major head protein Cps
3-dehydroquinate dehydratase	maltodextrose utilization protein malA
4-hydroxyphenylacetate-3-hydroxylase	mannose/fructose/N-acetylgalactosamine-specific PTS system transporter subunit IID
5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase	mannose/fructose/sorbose-specific PTS system IIA component
ABC transporter ATP-binding component	mannose/fructose/sorbose-specific PTS system IID component
ABC transporter ATP-binding protein	mannose-specific adhesin, LPXTG-motif cell wall anchor
ABC transporter permease protein	mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase
ABC transporter substrate-binding protein	MarR family transcriptional regulator
ABC-2 family transporter protein	MATE family efflux transporter
ABC-2 transporter family protein	membrane protein
AbrB family transcriptional regulator	MerR family transcriptional regulator
accessory gene regulator AgrB	methyl-accepting chemotaxis sensory transducer
acetate kinase	Microvirus H protein (pilot protein)
acetylornithine deacetylase/succinyl-diaminopimelate desuccinylase	Microvirus J protein
acetyltransferase	minor capsid protein
acetyltransferase (GNAT) family protein	minor capsid protein from bacteriophage
acyltransferase family protein	Mob
adenylyl transferase	molecular chaperone DnaK
adenylylsulfate kinase	mucus-binding protein
adherence-associated mucus-binding protein, LPXTG-motif cell wall anchor	multidrug ABC transporter ATP-binding and permease protein
alcohol dehydrogenase	muramidase
alpha/beta hydrolase	MutR family transcriptional regulator
alpha-amylase	Na+/xyloside symporter-related transporter
alpha-galactosidase	N-acetyltransferase
alpha-glucosidase	NAD/NADP octopine/nopaline dehydrogenase
alpha-L-fucosidase	NADPH:quinone reductase

amino acid permease

ankyrin repeat family protein

antimicrobial peptide ABC transporter ATP-binding protein

AraC family transcriptional regulator

ArsR family transcriptional regulator

ascorbate-specific PTS system IIC component

aspartate aminotransferase

Assimilatory nitrite reductase [NAD(P)H] small subunit

ATPase component of ABC transporter with duplicated ATPase domains

ATP-dependent nuclease, subunit B

bacteriocin immunity protein

bacteriocin immunity protein PlnL

bacteriocin precursor peptide PlnE

bacteriocin precursor peptide PlnF

bacteriophage replication gene A protein (GPA)

bacteriophage scaffolding protein D

beta-galactosidase

beta-glucosides-specific PTS system IIB component

beta-glucosides-specific PTS system IIC component

beta-lactamase

beta-lactamase family protein

BetT protein

BglG family transcriptional antiterminator/PTS system mannitol/fructose-specific IIA component

branched-chain amino acid ABC transporter ATP-binding protein

butyrate-acetoacetate CoA-transferase, beta subunit

Capsid protein (F protein)

capsular polysaccharide biosynthesis protein

CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase

cell division protein FtsK

cell surface hydrolase

cell surface protein

cell surface protein, CscB family

cellobiose-specific PTS system IIB component

NADPH-dependent FMN reductase family protein

nitroreductase

NUDIX family hydrolase

oligoendopeptidase F

PadR family transcriptional regulator

Pectate lyase precursor

penicillin-binding protein 2B

peptidase family S41

peptidase S41

peptide methionine sulfoxide reductase

peptidoglycan-binding protein

peptidylprolyl isomerase

phage envelope protein

phage holin protein (Holin_LLH)

phage major tail protein

phage portal protein

phage protein

phage protein C

phage related protein

phage single-strand DNA-binding protein

phosphate ABC transporter substrate-binding protein

phosphoglycerate mutase

phosphohydrolase

phosphoketolase

phospholipase/Carboxylesterase

plantaricin A precursor peptide, induction factor

plantaricin biosynthesis protein PlnQ

plantaricin biosynthesis protein PlnR

poly(glycerol-phosphate) alpha-glucosyltransferase

polysaccharide biosynthesis protein

polysaccharide lyase family 8

polysaccharide polymerase

potassium transporter Kup

cellobiose-specific PTS system IIC component

cellulase (glycosyl hydrolase family 5)

chitin-binding protein

chromosome partition protein Smc

chromosome partitioning protein ParA

competence protein TfoX

conjugal transfer protein

Cro/Cl family transcriptional regulator

cupin

cytochrome d ubiquinol oxidase subunit II

cytosolic protein

DeoR family transcriptional regulator

deoxyuridine 5'-triphosphate nucleotidohydrolase

D-galactose-binding periplasmic protein precursor

diacylglyceryl transferase

dipeptide/tripeptide permease

DNA mismatch repair protein MutS

DNA-3-methyladenine glycosylase I

DNA-binding protein

DNA-binding protein with HIRAN domain protein

exopolysaccharide biosynthesis protein

extracellular lipoprotein precursor, Asp-rich

extracellular protein

extracellular zinc metalloproteinase

Fe-S oxidoreductase

Fe-S-cluster oxidoreductase

fibrinogen-binding protein

FIST N domain protein

flagellar biosynthetic protein FlhB

flippase

frv operon regulatory protein

glycerol-3-phosphate dehydrogenase

glycerophosphoryl diester phosphodiesterase family protein

glycoside hydrolase

potassium transporter TrkA

prebacteriocin

preprotein translocase subunit YajC

prophage protein

proton glutamate symport protein

PTS sugar transporter IIA component

PTS sugar transporter subunit IIA

putative chromate transport protein

putative membrane protein

putative nucleotidyltransferase

putative recombinase

putative secreted protein

putative signal transduction protein with a C-terminal HATPase domain protein

putative sporulation-specific glycosylase YdhD

pyruvate kinase

rhomboid family protein

ribitolphosphotransferase

ribonuclease HI

RNA polymerase sigma factor

RNA polymerase sigma factor SigV

RNA-binding protein

RNHCP domain protein

S-adenosyl-L-homocysteine hydrolase, NAD binding domain protein

sensory box protein/response regulator

serine protease

serine transporter

serine/threonine-protein kinase PknD

short-chain dehydrogenase

short-chain dehydrogenase/oxidoreductase

sigma-70, region 4

single-stranded DNA-binding protein

SnoaL-like polyketide cyclase

sodium/sulfate symport protein

sodium:proton antiporter

glycosyl transferase	sortase
glycosyl transferase family 1	spermidine/putrescine ABC transporter permease protein
glycosyl transferase family 2	SpoVT / AbrB like domain protein
GNAT family acetyltransferase	sugar ABC transporter permease protein
GntR family transcriptional regulator	sugar ABC transporter substrate-binding protein
gp1 protein	sugar O-acetyltransferase
group II intron reverse transcriptase/maturase	sugar O-acyltransferase
haloacid dehalogenase	sulfate adenylyltransferase
helix-turn-helix protein	surface antigen
hemagglutinin	tail fiber
Heparinase II/III-like protein	tail protein
holin	tellurite resistance protein TerB
HTH-type transcriptional regulator MhqR	TetR family transcriptional regulator
hydrolase	thioredoxin domain protein
hypothetical protein	thymidylate kinase
integral membrane protein	transcription regulator
integral membrane protein (putative)	transcriptional antiterminator
integral membrane protein PlnU	transcriptional regulator
iron ABC transporter permease protein	transcriptional regulator/sugar kinase NagC
iron ABC transporter substrate-binding protein	transglutaminase-like superfamily protein
iron-sulfur cluster binding protein/lactate utilization protein LutB	transposase
L-fucose isomerase	tryptophan synthase alpha chain
lipoprotein	two-component system response regulator
lipoprotein LipO precursor	two-component system sensor histidine kinase
L-lactate dehydrogenase	type 1 restriction-modification system specificity protein
LPXTG-motif cell wall anchor domain protein	universal stress protein UspA
L-serine dehydratase beta subunit	UTP--glucose-1-phosphate uridylyltransferase
LuxR family transcriptional regulator	WaaG-like sugar transferase
LysR family transcriptional regulator	XRE family transcriptional regulator
major Facilitator Superfamily protein	YhhN-like protein

specialist group ortholog	
2', 3'-cyclic nucleotide 2'-phosphodiesterase	HTH-type transcriptional regulator Hpr

2-dehydropantoate 2-reductase

5'(3')-deoxyribonucleotidase

6-phospho-alpha-glucosidase

ABC transporter ATP-binding protein

ABC transporter permease protein

ABC-2 family transporter protein

acetoacetate decarboxylase

acetyltransferase

acyltransferase

adherence-associated mucus-binding protein, LPXTG-motif cell wall anchor

alkaline phosphatase

alpha/beta hydrolase family protein

alpha-amylase

aluminum resistance protein

amidohydrolase

amino acid ABC transporter ATP-binding protein

aminotransferase

amylopullulanase

antimicrobial peptide ABC transporter ATP-binding protein

arginine/ornithine antiporter

asparagine synthase

ATPase involved in chromosome partitioning

ATP-dependent DNA helicase RecQ

bacterial SH3 domain protein

beta-galactosidase

beta-lactamase class A

branched-chain amino acid permease

catalase

cell division protein

cobalt ABC transporter permease protein

competence protein ComGF

CsbD-like protein

DegV family protein

DeoR family transcriptional regulator

hypothetical protein

L-2,4-diaminobutyrate decarboxylase

L-threonine kinase

LysR family transcriptional regulator

MarR family transcriptional regulator

MATE efflux family protein

MATE family efflux transporter

membrane protein

multidrug ABC transporter ATP-binding and permease protein

Na+/H+ antiporter

N-acetyltransferase

NgoFVII restriction endonuclease

Nuclease-related domain protein

O-acetylhomoserine aminocarboxypropyltransferase

oligopeptide ABC transporter substrate-binding protein

peptidase propeptide and YPEB domain protein

peptidoglycan-binding protein

permease

permease protein

phage Mu protein F like protein

phosphoenolpyruvate carboxykinase

phosphopentomutase

phosphotransferase System HPr-Related protein

Pnp/Udp family phosphorylase

processive diacylglycerol beta-glucosyltransferase

proline dipeptidase

prolyl-tRNA synthetase

putative deoxyribodipyrimidine photolyase

putative helicase

pyridoxamine 5'-phosphate oxidase

rRNA methyltransferase

septum formation initiation protein

serine hydroxymethyltransferase

short-chain dehydrogenase

dihydroorotate dehydrogenase	signal transduction diguanylate cyclase
dipeptidase	small membrane protein
dipeptidase PepV	sugar O-acetyltransferase
D-lactate dehydrogenase	sulfite exporter TauE/SafE family protein
DNA damage-inducible protein DnaD	surface protein Rib
drug/metabolite transporter permease	tagatose-6-phosphate ketose isomerase
esterase	Thiosulfate sulfurtransferase YnjE precursor
exopolysaccharide biosynthesis protein	TM2 domain protein
extracellular zinc metalloproteinase	transcriptional regulator
fumarate hydratase	transcriptional regulator/sugar kinase NagC
fumarate reductase	transposase
fumarate reductase flavoprotein subunit	tricarballylate dehydrogenase
glycerol kinase	type III restriction enzyme, res subunit
glycerol uptake facilitator protein	uracil DNA glycosylase superfamily protein
glycogen phosphorylase	Xaa-Pro aminopeptidase
glycopeptide antibiotics resistance protein	Xylan alpha-(1->2)-glucuronosidase
GntR family transcriptional regulator	YdfK protein
helix-turn-helix domain protein	zinc ABC transporter substrate-binding protein
homoserine O-succinyltransferase	

3.2.3 Network of orthologs shared by strains with high sugar utilization

I constructed networks for the shared orthologs among the 178 strains in the 24 genera to identify the influence of sugar utilization on the GECs for different ecological niches (Figure 3.9). There were 178 nodes representing each genome, which were color-coded according to the 24 genera. An edge was generated between two genomes when they shared more than five orthologs of the generalist or specialist group for sugar utilization. A dense network indicated that the community formed a GEC or had conserved genes inherited from their ancestors. No edges were identified in the investigation among the following genera: *Bombilactobacillus*, *Amylolactobacillus*, *Paralactobacillus*, *Holzapfelia*, *Dellaglioa*, *Furfurilactobacillus*, and *Lentilactobacillus*.

While the networks of orthologs predominantly shared by the specialist groups for sugar utilization were connected only within the same genera, the networks of the generalist groups were connected across genera. The networks of specialists were made by strains from *Lactobacillus*, *Loigolactobacillus*, *Apilactobacillus*, *Fructilactobacillus* and *Secundilactobacillus* independently. The generalist networks connected *Lactobacillus*, *Loigolactobacillus*, *Lapidilactobacillus*, *Schleiferilactobacillus*, *Agrilactobacillus*, *Liquorilactobacillus*, *Lacticaseibacillus*, *Lactilactobacillus*, *Lactiplantibacillus*, *Companilactobacillus*, *Paucilactobacillus*, *Secundilactobacillus*, and *Levilactobacillus*.

In the generalist networks, the edges were connected between distant strains isolated from similar environments. As a result of community extraction, the number of communities was 51, the maximum number of strains in the community was nine, and the minimum value was two (Table 3.3). Communities were often formed among strains of the following three genera, *Schleiferilactobacillus*, *Lacticaseibacillus*, and *Lactiplantibacillus*, or four genera when *Agrilactobacillus* was added. For example, a community was formed among *Schleiferilactobacillus harbinensis*, *Schleiferilactobacillus perolens*, *Lactiplantibacillus paraplantarum*, *Lacticaseibacillus rhamnosus*, *Lacticaseibacillus casei,* and *Agrilactobacillus composti,* with its members isolated from vegetables and brewing-related environments (Supplementary Table 2.1) (Zheng et al. 2020). In addition, some communities including members of the genus *Lactiplantibacillus* and *Liquorilactobacillus* were identified. Members of the community of *Liquorilactobacillus nagelii*, *Lactiplantibacillus paraplantarum,* and *Lactiplantibacillus plantarum* ssp. *plantarum* were isolated from dairy products (Supplementary Table 2.1) (Zheng et al. 2020).

The analysis method aimed to select high ASU value orthologs, and as a result, strains with low sugar utilization values tended not to be included in the generalist networks. For example, for the genus *Lacticaseibacillus*, *L. nasuensis*, *L. thailandensis*, and *L. pantheris* were not included in the generalist network, nor were *L. nasuensis* and *L. thailandensis,* which had small sugar utilization values. Moreover, for the genus *Latilactobacillus*, all strains, except for *L. skei* ssp. *carnosus* and *L. fuchuensis,* had relatively low sugar utilization values and were not included in the network.

Despite this, the generalist network included strains with low sugar utilization values. In these cases, the strains were connected to closely related strains with high values. For example, although *Lacticaseibacillus brantae* had a low sugar utilization value, it shared generalist group orthologs with *Schleiferilactobacillus harbinensis*, *Schleiferilactobacillus shenzhenensis*, and *Lacticaseibacillus saniviri*. *L. brantae* was closely related to *L. saniviri*, which had a high sugar utilization value. In addition, *Lactobacillus paracasei* and *L. paracasei* ssp. *tolerans* were also included in the generalist network, although they had low sugar utilization values as they were closely related to *L. paracasei* ssp. *paracasei,* which had a high sugar utilization value.

The closely related strains in a network of specialists tended to form communities within the same genera. In the genera *Lactobacillus* and *Loigolactobacillus,* there was a tendency for the edges to be connected between the subspecies of each species.

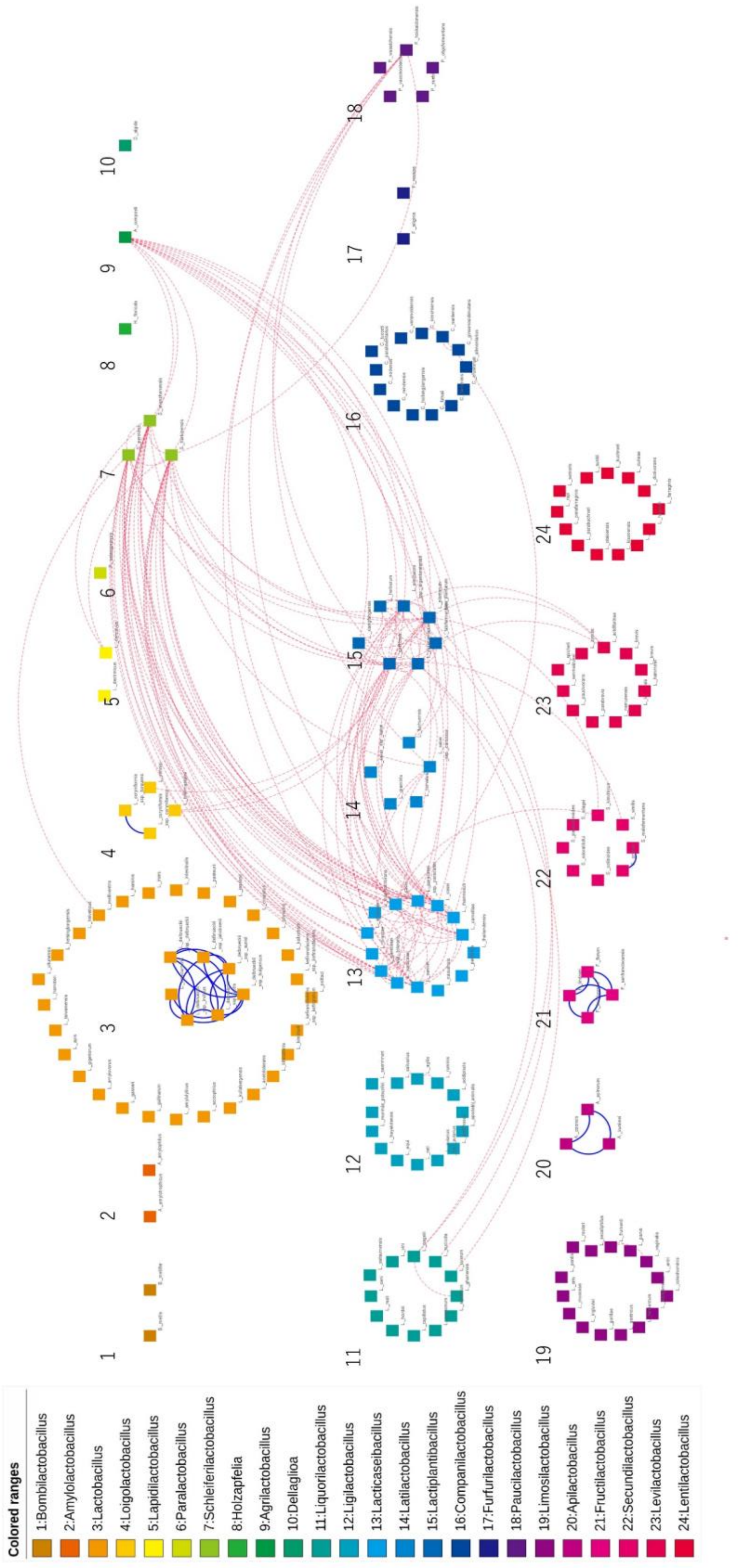

Colored ranges
1:Bombilactobacillus
2:Amylolactobacillus
3:Lactobacillus
4:Loigolactobacillus
5:Lapidilactobacillus
6:Paralactobacillus
7:Schleiferilactobacillus
8:Holzapfelia
9:Agrilactobacillus
10:Dellagliоa
11:Liquorilactobacillus
12:Ligilactobacillus
13:Lacticaseibacillus
14:Latilactobacillus
15:Lactiplantibacillus
16:Companilactobacillus
17:Furfurilactobacillus
18:Paucilactobacillus
19:Limosilactobacillus
20:Apilactobacillus
21:Fructilactobacillus
22:Secundilactobacillus
23:Levilactobacillus
24:Lentilactobacillus
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Figure 3.9: Networks for the generalist and specialist group orthologs.

Each of the 178 nodes represents a LAB genome, colored and numbered by genus. Dotted-red/solid-blue curves indicate edges created between two genomes when the number of sharing generalist/specialist group orthologs is more than five. Figure adapted from Takenaka et al. (2021).

Table 3.3: Community extraction of shared generalist group orthologs networks.

The table indicates the number of strains, genus name, and members in each community for the generalist group ortholog networks. Table adapted from Takenaka et al. (2021).

community	# of strains	genus	member
1	7	*Lactiplantibacillus*	Lactiplantibacillus_fabifermentans_DSM_21115, Lactiplantibacillus_xiangfangensis_LMG_26013, Lactiplantibacillus_paraplantarum_DSM_10667, Lactiplantibacillus_herbarum_TCF032-E4, Lactiplantibacillus_pentosus_DSM_20314, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365
2	3	*Lactiplantibacillus,* *Loigolactobacillus*	Loigolactobacillus _bifermentans_DSM_20003, Lactiplantibacillus_pentosus_DSM_20314, Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365
3	4	*Lactiplantibacillus,* *Levilactobacillus*	Levilactobacillus_acidifarinae_DSM_19394, Lactiplantibacillus_paraplantarum_DSM_10667, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365
4	2	*Limosilactobacillus*	Limosilactobacillus_frumenti_DSM_13145, Limosilactobacillus_vaginalis_DSM_5837
5	2	*Limosilactobacillus*	Limosilactobacillus_frumenti_DSM_13145, Limosilactobacillus_panis_DSM_6035
6	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_perolens_DSM_12744, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lacticaseibacillus_saniviri_DSM_24301
7	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_perolens_DSM_12744, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Agrilactobacillus_composti_DSM_18527
8	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lactiplantibacillus_pentosus_DSM_20314, Lacticaseibacillus_saniviri_DSM_24301
9	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lactiplantibacillus_pentosus_DSM_20314, Agrilactobacillus_composti_DSM_18527

10	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_perolens_DSM_12744, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lacticaseibacillus_saniviri_DSM_24301
11	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_perolens_DSM_12744, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Agrilactobacillus_composti_DSM_18527
12	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lactiplantibacillus_pentosus_DSM_20314, Lacticaseibacillus_saniviri_DSM_24301
13	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lactiplantibacillus_pentosus_DSM_20314, Agrilactobacillus_composti_DSM_18527
14	7	*Schleiferilactobacillus,* *Lactiplantibacillus,* *Latilactobacillus,* *Lacticaseibacillus,* *Paucilactobacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Latilactobacillus_sakei_ssp._carnosus_DSM_15831 Lacticaseibacillus_casei_ATCC_393, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_saniviri_DSM_24301, Lactiplantibacillus_pentosus_DSM_20314, Paucilactobacillus_hokkaidonensis_LOOC260
15	4	*Schleiferilactobacillus,* *Lacticaseibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Lacticaseibacillus_saniviri_DSM_24301, Lacticaseibacillus_brantae_DSM_23927
16	7	*Schleiferilactobacillus,* *Lacticaseibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Schleiferilactobacillus_perolens_DSM_12744 Lacticaseibacillus_saniviri_DSM_24301, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393

17	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Lacticaseibacillus_saniviri_DSM_24301, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lactiplantibacillus_pentosus_DSM_20314
18	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Schleiferilactobacillus_perolens_DSM_12744, Lacticaseibacillus_saniviri_DSM_24301, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437
19	7	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Lacticaseibacillus_saniviri_DSM_24301, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lactiplantibacillus_pentosus_DSM_20314
20	5	*Schleiferilactobacillus,* *Lacticaseibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Schleiferilactobacillus_perolens_DSM_12744, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_sharpeae_DSM_20505
21	5	*Schleiferilactobacillus,* *Lacticaseibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_sharpeae_DSM_20505, Lacticaseibacillus_manihotivorans_DSM_13343
22	9	*Schleiferilactobacillus,* *Lacticaseibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Schleiferilactobacillus_perolens_DSM_12744, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lacticaseibacillus_paracasei_ATCC_334, Lacticaseibacillus_paracasei_ssp._tolerans_DSM_20258
23	9	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Agrilactobacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Schleiferilactobacillus_perolens_DSM_12744, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lacticaseibacillus_paracasei_ATCC_334, Agrilactobacillus_composti_DSM_18527

24	9	*Schleiferilactobacillus,* *Lacticaseibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Schleiferilactobacillus_perolens_DSM_12744, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lacticaseibacillus_paracasei_ATCC_334, Lacticaseibacillus_camelliae_DSM_22697
25	9	*Schleiferilactobacillus,* *Lacticaseibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lacticaseibacillus_paracasei_ATCC_334, Lacticaseibacillus_manihotivorans_DSM_13343, Lacticaseibacillus_camelliae_DSM_22697
26	8	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lactiplantibacillus_pentosus_DSM_20314, Agrilactobacillus_composti_DSM_18527
27	8	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Schleiferilactobacillus_perolens_DSM_12744, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Agrilactobacillus_composti_DSM_18527
28	8	*Schleiferilactobacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Lactiplantibacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lactiplantibacillus_pentosus_DSM_20314, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Agrilactobacillus_composti_DSM_18527,
29	3	*Schleiferilactobacillus,* *Lapidilactobacillus*	Schleiferilactobacillus_harbinensis_DSM_16991, Schleiferilactobacillus_shenzhenensis_LY-73, Lapidilactobacillus_concavus_DSM_17758
30	3	*Lacticaseibacillus,* *Lactiplantibacillus,* *Secundilactobacillus*	Secundilactobacillus_kimchicus_JCM_15530, Lacticaseibacillus_casei_DSM_20178, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437
31	2	*Schleiferilactobacillus,* *Lactobacillus*	Lactobacillus_melliventris_Hma8, Schleiferilactobacillus_perolens_DSM_12744
32	2	*Liquorilactobacillus*	Liquorilactobacillus_nagelii_DSM_13675, Liquorilactobacillus_ghanensis_DSM_18630

33	3	*Lactiplantibacillus,* *Liquorilactobacillus*	Liquorilactobacillus_nagelii_DSM_13675, Lactiplantibacillus_paraplantarum_DSM_10667, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437
34	6	*Lactiplantibacillus,* *Agrilactobacillus*	Lactiplantibacillus_xiangfangensis_LMG_26013, Lactiplantibacillus_paraplantarum_DSM_10667, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lactiplantibacillus_pentosus_DSM_20314, Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365 Agrilactobacillus_composti_DSM_18527
35	8	*Lactiplantibacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Schleiferilactobacillus*	Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Lacticaseibacillus_paracasei_ATCC_334, Schleiferilactobacillus_perolens_DSM_12744 Agrilactobacillus_composti_DSM_18527
36	7	*Lactiplantibacillus,* *Lacticaseibacillus,* *Agrilactobacillus*	Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Lactiplantibacillus_pentosus_DSM_20314 Lacticaseibacillus_rhamnosus_DSM_20021, Agrilactobacillus_composti_DSM_18527, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130
37	7	*Lactiplantibacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Schleiferilactobacillus*	Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lacticaseibacillus_rhamnosus_DSM_20021, Agrilactobacillus_composti_DSM_18527, Schleiferilactobacillus_perolens_DSM_12744
38	7	*Lactiplantibacillus,* *Lacticaseibacillus,* *Agrilactobacillus*	Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_casei_ATCC_393, Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Lactiplantibacillus_paraplantarum_DSM_10667, Lactiplantibacillus_pentosus_DSM_20314, Agrilactobacillus_composti_DSM_18527
39	7	*Lactiplantibacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Schleiferilactobacillus*	Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Agrilactobacillus_composti_DSM_18527, Schleiferilactobacillus_perolens_DSM_12744
40	7	*Lactiplantibacillus,* *Lacticaseibacillus,* *Agrilactobacillus,* *Schleiferilactobacillus*	Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Schleiferilactobacillus_perolens_DSM_12744, Agrilactobacillus_composti_DSM_18527

41	7	*Lactiplantibacillus,* *Lacticaseibacillus,* *Agrilactobacillus*	Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lactiplantibacillus_pentosus_DSM_20314, Lacticaseibacillus_casei_DSM_20178, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_paracasei_ssp._paracasei_JCM_8130, Agrilactobacillus_composti_DSM_18527
42	7	*Lactiplantibacillus,* *Lacticaseibacillus,* *Agrilactobacillus*	Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Lactiplantibacillus_pentosus_DSM_20314, Lactiplantibacillus_paraplantarum_DSM_10667, Lacticaseibacillus_rhamnosus_DSM_20021, Lacticaseibacillus_casei_DSM_20178, Agrilactobacillus_composti_DSM_18527
43	2	*Lactiplantibacillus,* *Liquorilactobacillus*	Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Liquorilactobacillus_sucicola_DSM_21376
44	3	*Lactiplantibacillus,* *Latilactobacillus*	Lactiplantibacillus_plantarum_ssp._argentoratensis_DSM_16365, Lactiplantibacillus_plantarum_ssp._plantarum_CGMCC_1.2437, Latilactobacillus_fuchuensis_JCM_11249
45	2	*Companilactobacillus*	Companilactobacillus_kimchiensis_DSM_24716, Companilactobacillus_nantensis_DSM_16982
46	2	*Companilactobacillus*	Companilactobacillus_ginsenosidimutans_EMML_3141, Companilactobacillus_nantensis_DSM_16982
47	2	*Lactiplantibacillus,* *Secundilactobacillus*	Secundilactobacillus_similis_DSM_23365, Lactiplantibacillus_pentosus_DSM_20314
48	2	*Lactiplantibacillus,* *Liquorilactobacillus*	Liquorilactobacillus_uvarum_DSM_19971, Lactiplantibacillus_paraplantarum_DSM_10667
49	2	*Ligilactobacillus*	Ligilactobacillus_agilis_DSM_20509, Ligilactobacillus_ruminis_ATCC_27780
50	2	*Lactiplantibacillus,* *Loigolactobacillus*	Loigolactobacillus_rennini_DSM_20253, Lactiplantibacillus_pentosus_DSM_20314
51	2	*Lacticaseibacillus,* *Companilactobacillus*	Companilactobacillus_nantensis_DSM_16982, Lacticaseibacillus_casei_DSM_20178

3.2.4 Sugar utilization phenotype contributes to GEC formation in the ecological niche of *Lactobacillaceae*

In Section 3.2, networks of orthologs were analyzed to identify how the phenotypes contributed to the formation of GECs (Figure 3.9). Results in this and Section 3.1 suggested that the ability to utilize a variety of sugars contributed to increased HGT and the formation of GECs in ecological niches among genera. These results will help to improve our understanding of the evolution of related bacteria in ecological niches.

HGT tends to occur among prokaryotes with similar phenotypes, as they live in the same environment (Jain et al. 2003). For example, many bacteria in the order Thermotogales of *Thermotogae,* mainly thermophilic bacteria, and in the class Clostridia included in the phylum Firmicutes, share ecological niches and genes, probably because they share thermophilic features (Andam and Gogarten 2011). These reports suggest that some phenotypes contribute to the sharing of ecological niches and the formation of GECs. My study showed that this tendency can apply to bacterial groups within *Lactobacillaceae* and revealed that the utilization of a variety of sugars highly influenced the construction of GECs across genera to share niches such as vegetables, dairy, and brewing-related environments (Figure 3.9, Supplementary Table 2.1, Table 3.3).

One of the problems in this network analysis is that not only orthologs shared by HGT but also those shared from ancestors constitute the networks. However, I consider that three reasons support the results in Section 3.2 and help overcome this problem. First, phylogenetic trees of generalist group orthologs selected by ASU value contradicted the tree based on the 16S rRNA gene, reflecting phylogenetic relationships (Figure 3.1, Figure 3.8). Conflicting trees suggest HGT events. Secondly, the generalist orthologs group networks were connected among distant strains compared with the networks of the specialist orthologs group (Figure 3.9). If sugar utilization did not contribute to forming GECs in ecological niches, both networks should have constituted closely related strains because of phylogenetic genes and formation of GECs by phylogenetic closeness. Finally, the strains in generalist networks were isolated from similar environments (Supplementary Table 2.1, Table 3.3). These reasons support that sugar utilization contributes to forming GEC in the ecological niche of LAB.

Interestingly, the network of the generalist ortholog group includes strains with low sugar utilization values. The result suggests that GECs in *Lactobacillaceae* are generated by two HGT biases: sharing of ecological niche and phylogenetic closeness. The bias of phylogenetic closeness is caused by the similarity of genomes and the specificity of phages. The phylogenetic proximity influences the HGT events of partners (Andam and Gogarten 2011).

Pan-genome refers to potentially available genes for individuals in closely related groups as HGT events increase by the bias of phylogenetic closeness (Soucy et al. 2015). This concept can apply to the GECs biased by sharing of ecological niche and phylogenetic closeness in LAB. The generalists in LAB have increased potential to gain HGT genes to share various ecological niches. The genes gained by the generalists are transferred to the specialists via HGT by biased phylogenetic closeness. These flows maintain the diversity in closely related groups, which may improve the fitness of individuals in the group.

Generalists may be a kind of “gene installer” for their group; they acquire genes to construct GEC in ecological niches and share the genes between groups to form GEC by phylogenetic closeness. There are both generalists and specialists in closely related groups. Phylogenetic closeness generates HGT bias because of HGT mechanisms (Andam and Gogarten 2011). In Chapter 3, the networks of generalist ortholog groups included not only generalists of sugar utilization but also specialists closely related strains of the generalists. This suggests the possibility that generalists install genes into closely related specialists. Sriswasdi et al. (2017) reported that generalists drive bacterial evolution. The hypothesis of “gene installer” supports the report.

GECs among the strains of *Lactobacillaceae* with high sugar utilization values could help to expand their habitats and promote the exchange of genetic material with various functions. According to my results for the functional classification by COG, there were a variety of gene functions in the generalist group orthologs for sugar utilization, but the function proportions were not significantly different from those of the specialist group orthologs (Figure 3.6). In the generalist group orthologs, there were genes related to sugar metabolism and genes to enable the adaptation of various niches related to stress responses, bacteriocin production, antibiotic resistance, survival in the intestinal environment, and heavy metal resistance. These results are consistent with the idea that most HGT genes are acquired with neutral or nearly neutral effects (Soucy et al. 2015). Some HGT genes in the GECs of different ecological niches may thus help recipients to adapt to new habitats and affects population diversification (Baquero et al. 2021). These results allow us to speculate that the GECs composed of strains in *Lactobacillaceae* with high sugar utilization accelerate their adaptations to new niches.

Overall, my results indicate that the phenotype to utilize a variety of sugars was the key factor for the construction of GECs in the family *Lactobacillaceae*. This feature is consistent with the fact that *Lactobacillaceae* contributes to producing a wide variety of fermented foods by sharing niches such as vegetables, dairy products, and brewing-related environments. The results of this study will help to improve our understanding of these ecologies.

3.3 Genetic capitalism in LAB

In Section 3.3, I investigate LAB evolutions using the concept of genetic capitalism. Genetic capitalism is considered the phenomenon of rich becoming richer. Baquero et al. (2004) explained that genetic capitalism increases interactions with environments such as HGT by acquired genes encoding particular phenotypes in natural selection. For instance, gaining genes that provide hosts with selective advantage increases the population size of the group. Consequently, the group has more potential to acquire genes via HGT.

To understand the genetic capitalism in bacteria, biases that increase and decrease genome size (i.e., biases that are acquisition and deletion of genes) should be considered. Bacteria have a compact architecture of genomes whose proportion of genes is high and the intergenic region is low. The compact genome may be formed by frequent point mutations that cause gene deletion (Douglas 1988). However, gene acquisition via HGT increases the genome size (Zimmer and Emlen 2016). These genes are introduced to genomes neutrally in function (Soucy et al. 2015). Both biases to decrease/increase genome size are competing in the bacterial genome.

The bias that decreases/increases genome size can cause genetic capitalism that widens the disparity of genome size or diversity of genes in bacteria. Strains incapable of inhabiting various environments have less potential to exchange genes in ecological niches, strengthening the relative influence of decreasing bias. The bias deletes more extra genes, which makes strains do not inhabit other environments. This consequently makes the strains specialize in particular niches. Strains capable of inhabiting various environments frequently gain genes to form GEC in ecological niches, where bias that increases genome size exceeds decreasing bias. The genes may encode suitable phenotypes to expand habitats of the strains, increasing genome size or diversity in the genome.

The results in Section 3.2 implied that phenotype of utilizing various sugars in LAB contributes to forming GECs in the ecological niches. GECs in the ecological niche may provide the members with potential to acquire genes for inhabiting various environments, which helps them to relocate to new niches. The concept of genetic capitalism can be applied to the ecological flow in LAB.

However, there are few reports that genetic capitalism influences evolution in LAB. Simplifying genomes plays a major role in the evolution of LAB (Makarova et al. 2006). Moreover, although the results in Section 3.2 showed that the phenotype in LAB contributes to forming GEC in the ecological niches, the results did not prove the tendency of the rich becoming richer. Based on the above discussion, possessing diverse genes that encode phenotypes for surviving various environments in the bacterial genome can contribute to increasing the potential to gain genes via HGT. As a consequence, the bacteria gains diverse genes.

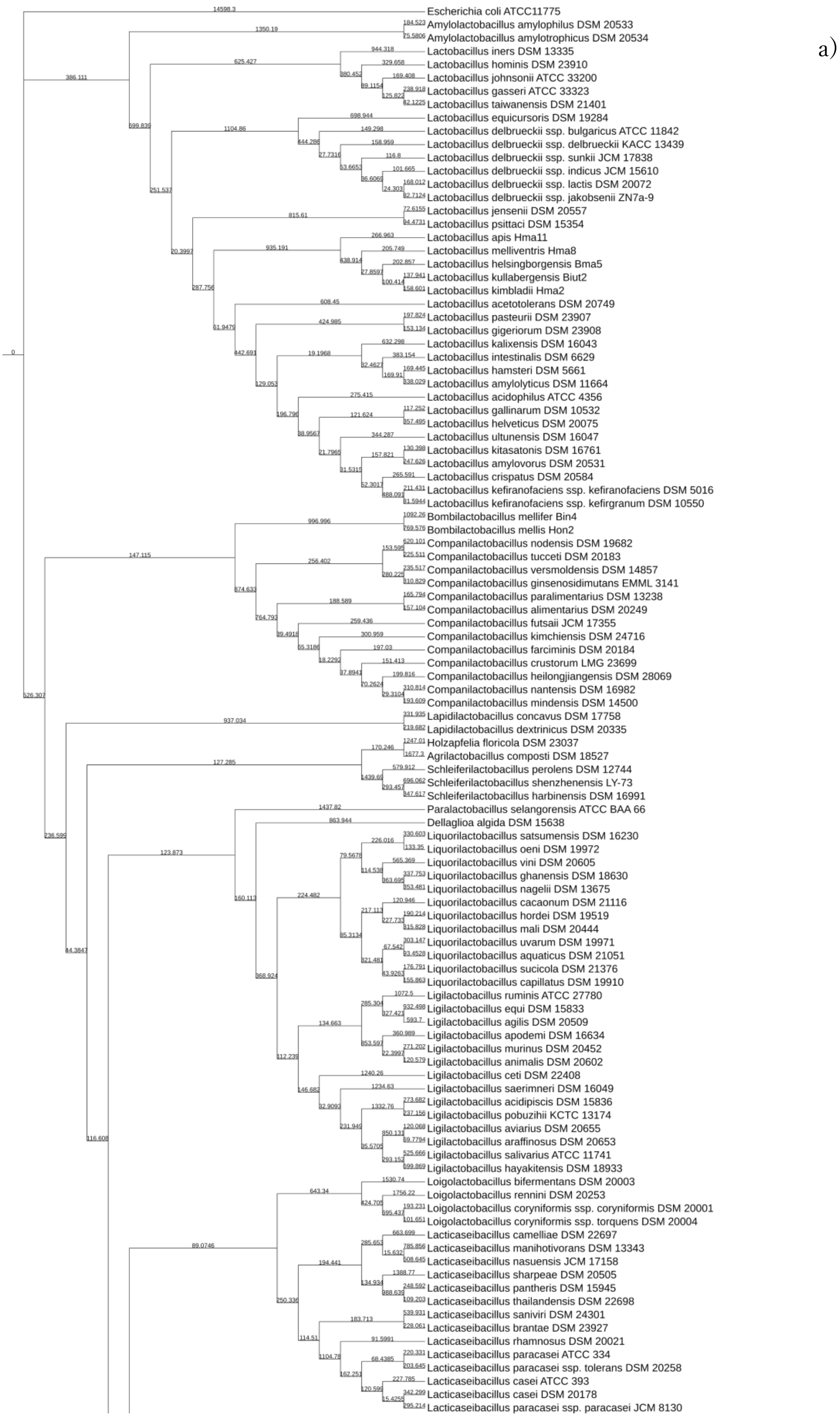
In this Section 3.3, I investigated whether the genetic diversity in the bacterial genome influences the gain and loss of genes. To estimate gene diversity in the bacterial genome before speciation, the expected value of gain/loss events was calculated based on the ortholog in LAB. I hypothesized that if genetic capitalism applies to LAB evolution, the rich that have diversity in the genome before speciation can become richer to obtain more potential to acquire other genes.

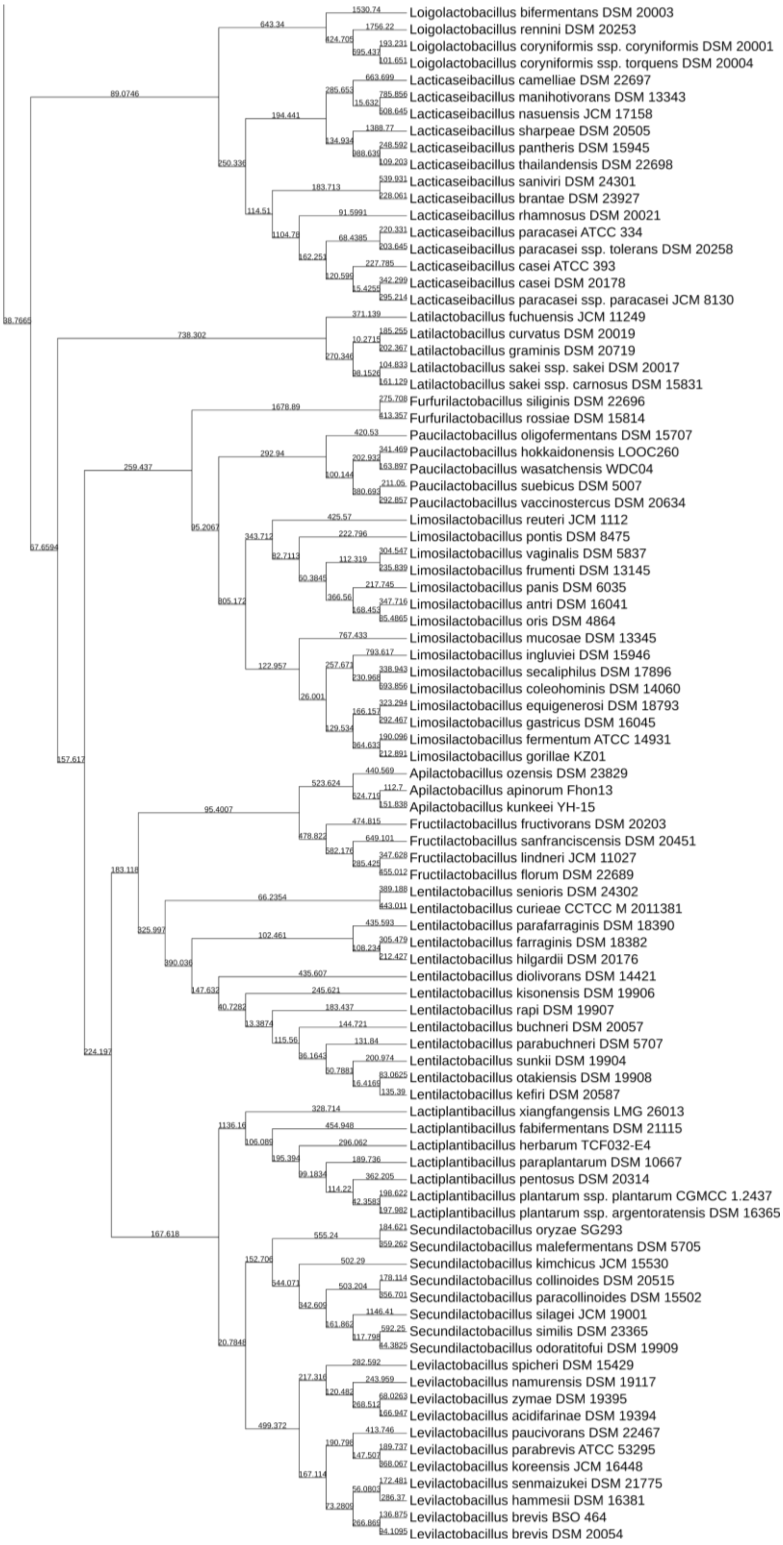
3.3.1 Gene gain and loss in *Lactobacillaceae*

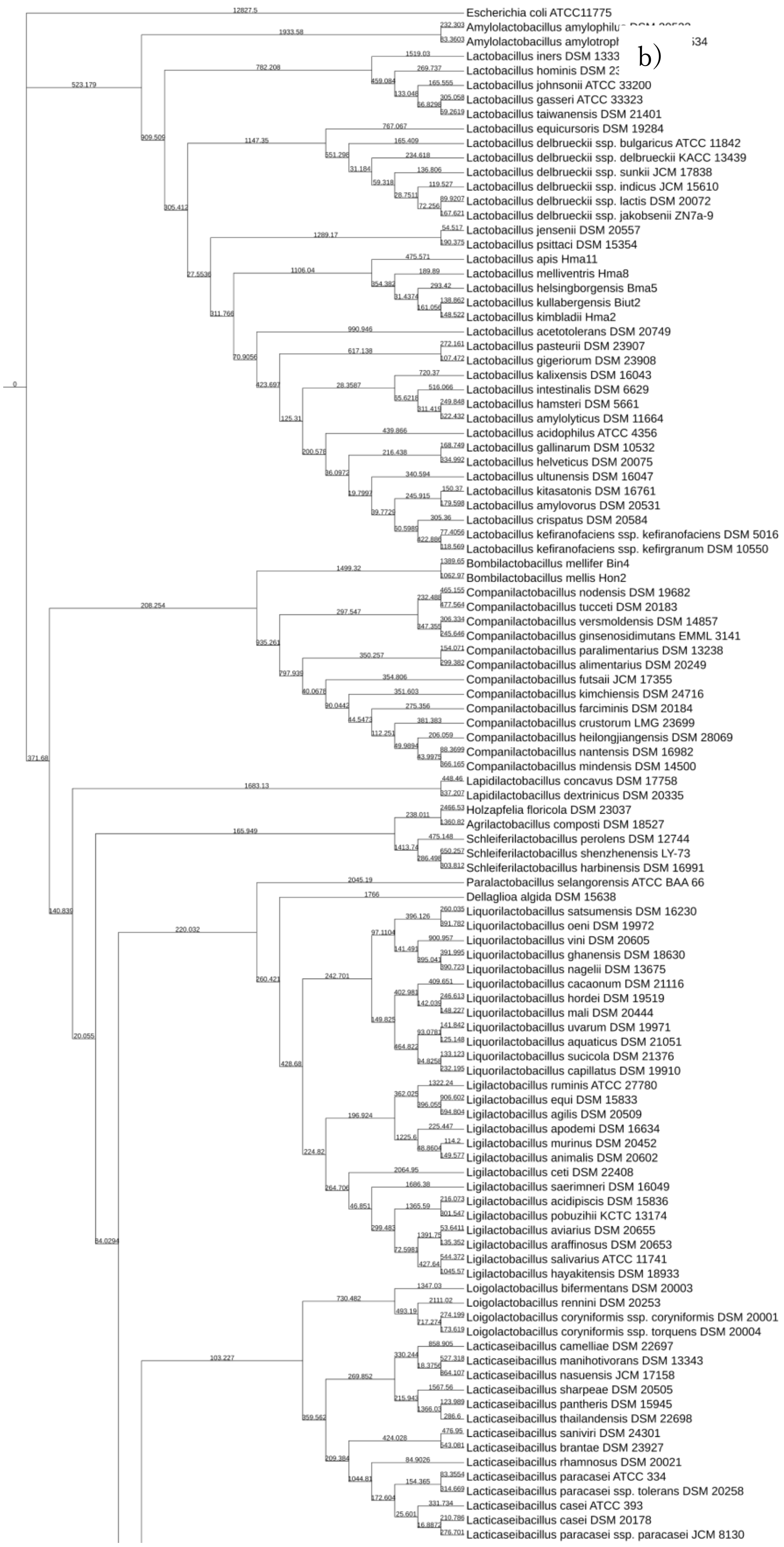
To investigate gain/loss events in *Lactobacillaceae*, the expected value of these events was mapped on a phylogenetic tree based on 16S rRNA (Figure 3.10). The expected number of gain events in the branch of speciation for the 178 strains ranged from 42.12 for *Lactobacillus taiwanensis* to 1756 for *Loigolactobacillus rennini*. The expected number of loss events in the branch of speciation for each of the 178 strains ranged from 42.62 for *Latilactobacillus sakei* ssp. *carnosus* to 2467 for *Holzapfelia floricola* (Figure 3.10, Supplementary Table 3.4).

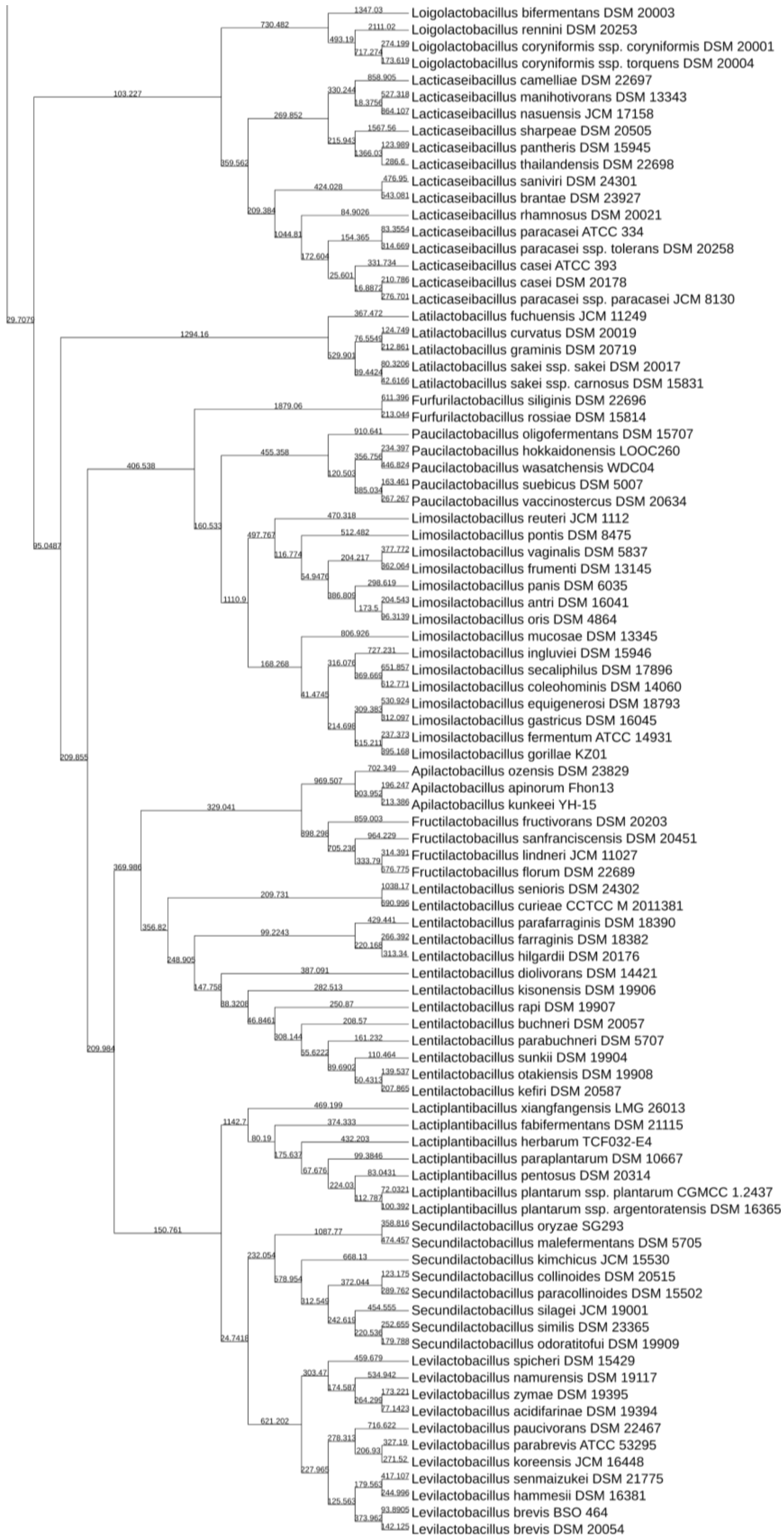
There were a few correlations between expected values of gain/loss events and genomic factors. The coefficients for correlation between the expected value of gain events for each branch and genome size, protein number (number of CDS), and the number of orthologs types in a genome were 0.308, 0.316, and 0.319, respectively. The coefficients for correlation between the expected value of loss events for each branch and genome size, protein number (number of CDS), and the number of orthologs types in a genome were 0.216, 0.239, and 0.247, respectively (Table 3.5).

Even in the same genus, there were various expected values of gain/loss events in the branch of speciation to each species. For instance, the minimum expected value of gain events was 91.6 for *Lacticaseibacillus rhamnosus*, whereas the maximum was 1389 for *Lacticaseibacillus sharpeae*. In addition, the minimum expected value of loss events was 83.36 for *Lacticaseibacillus paracasei*, whereas the maximum was 1568 for *Lacticaseibacillus sharpeae*.


a)
Escherichia coli ATCC11775
Amylolactobacillus amylophilus DSM 20533
Amylolactobacillus amylotrophicus DSM 20534
Lactobacillus iners DSM 13335
Lactobacillus hominis DSM 23910
Lactobacillus johnsonii ATCC 33200
Lactobacillus gasseri ATCC 33323
Lactobacillus taiwanensis DSM 21401
Lactobacillus equicursoris DSM 19284
Lactobacillus delbrueckii ssp. bulgaricus ATCC 11842
Lactobacillus delbrueckii ssp. delbrueckii KACC 13439
Lactobacillus delbrueckii ssp. sunkii JCM 17838
Lactobacillus delbrueckii ssp. indicus JCM 15610
Lactobacillus delbrueckii ssp. lactis DSM 20072
Lactobacillus delbrueckii ssp. jakobsenii ZN7a-9
Lactobacillus jensenii DSM 20557
Lactobacillus psittaci DSM 15354
Lactobacillus apis Hma11
Lactobacillus melliventris Hma8
Lactobacillus helsingborgensis Bma5
Lactobacillus kullabergensis Biut2
Lactobacillus kimbladii Hma2
Lactobacillus acetotolerans DSM 20749
Lactobacillus pasteurii DSM 23907
Lactobacillus gigeriorum DSM 23908
Lactobacillus kalixensis DSM 16043
Lactobacillus intestinalis DSM 6629
Lactobacillus hamsteri DSM 5661
Lactobacillus amylolyticus DSM 11664
Lactobacillus acidophilus ATCC 4356
Lactobacillus gallinarum DSM 10532
Lactobacillus helveticus DSM 20075
Lactobacillus ultunensis DSM 16047
Lactobacillus kitasatonis DSM 16761
Lactobacillus amylovorus DSM 20531
Lactobacillus crispatus DSM 20584
Lactobacillus kefiranofaciens ssp. kefiranofaciens DSM 5016
Lactobacillus kefiranofaciens ssp. kefirgranum DSM 10550
Bombilactobacillus mellifer Bin4
Bombilactobacillus mellis Hon2
Companilactobacillus nodensis DSM 19682
Companilactobacillus tucceti DSM 20183
Companilactobacillus versmoldensis DSM 14857
Companilactobacillus ginsenosidimutans EMML 3141
Companilactobacillus paralimentarius DSM 13238
Companilactobacillus alimentarius DSM 20249
Companilactobacillus futsaii JCM 17355
Companilactobacillus kimchiensis DSM 24716
Companilactobacillus farciminis DSM 20184
Companilactobacillus crustorum LMG 23699
Companilactobacillus heilongjiangensis DSM 28069
Companilactobacillus nantensis DSM 16982
Companilactobacillus mindensis DSM 14500
Lapidilactobacillus concavus DSM 17758
Lapidilactobacillus dextrinicus DSM 20335
Holzapfelia floricola DSM 23037
Agrilactobacillus composti DSM 18527
Schleiferilactobacillus perolens DSM 12744
Schleiferilactobacillus shenzhenensis LY-73
Schleiferilactobacillus harbinensis DSM 16991
Paralactobacillus selangorensis ATCC BAA 66
Dellaglioa algida DSM 15638
Liquorilactobacillus satsumensis DSM 16230
Liquorilactobacillus oeni DSM 19972
Liquorilactobacillus vini DSM 20605
Liquorilactobacillus ghanensis DSM 18630
Liquorilactobacillus nagelii DSM 13675
Liquorilactobacillus cacaonum DSM 21116
Liquorilactobacillus hordei DSM 19519
Liquorilactobacillus mali DSM 20444
Liquorilactobacillus uvarum DSM 19971
Liquorilactobacillus aquaticus DSM 21051
Liquorilactobacillus sucicola DSM 21376
Liquorilactobacillus capillatus DSM 19910
Ligilactobacillus ruminis ATCC 27780
Ligilactobacillus equi DSM 15833
Ligilactobacillus agilis DSM 20509
Ligilactobacillus apodemi DSM 16634
Ligilactobacillus murinus DSM 20452
Ligilactobacillus animalis DSM 20602
Ligilactobacillus ceti DSM 22408
Ligilactobacillus saerimneri DSM 16049
Ligilactobacillus acidipiscis DSM 15836
Ligilactobacillus pobuzihii KCTC 13174
Ligilactobacillus aviarius DSM 20655
Ligilactobacillus araffinosus DSM 20653
Ligilactobacillus salivarius ATCC 11741
Ligilactobacillus hayakitensis DSM 18933
Loigolactobacillus bifermentans DSM 20003
Loigolactobacillus rennini DSM 20253
Loigolactobacillus coryniformis ssp. coryniformis DSM 20001
Loigolactobacillus coryniformis ssp. torquens DSM 20004
Lacticaseibacillus camelliae DSM 22697
Lacticaseibacillus manihotivorans DSM 13343
Lacticaseibacillus nasuensis JCM 17158
Lacticaseibacillus sharpeae DSM 20505
Lacticaseibacillus pantheris DSM 15945
Lacticaseibacillus thailandensis DSM 22698
Lacticaseibacillus saniviri DSM 24301
Lacticaseibacillus brantae DSM 23927
Lacticaseibacillus rhamnosus DSM 20021
Lacticaseibacillus paracasei ATCC 334
Lacticaseibacillus paracasei ssp. tolerans DSM 20258
Lacticaseibacillus casei ATCC 393
Lacticaseibacillus casei DSM 20178
Lacticaseibacillus paracasei ssp. paracasei JCM 8130


1530.74
Loigolactobacillus bifermentans DSM 20003
643.34
1756.22
Loigolactobacillus rennini DSM 20253
424.705
193.231
Loigolactobacillus coryniformis ssp. coryniformis DSM 20001
695.437
101.651
Loigolactobacillus coryniformis ssp. torquens DSM 20004
89.0746
663.699
Lacticaseibacillus camelliae DSM 22697
285.653
785.856
Lacticaseibacillus manihotivorans DSM 13343
15.632
608.645
Lacticaseibacillus nasuensis JCM 17158
194.441
1388.77
Lacticaseibacillus sharpeae DSM 20505
134.934
248.592
Lacticaseibacillus pantheris DSM 15945
988.639
109.203
Lacticaseibacillus thailandensis DSM 22698
250.336
539.931
Lacticaseibacillus saniviri DSM 24301
183.713
228.061
Lacticaseibacillus brantae DSM 23927
114.51
91.5991
Lacticaseibacillus rhamnosus DSM 20021
1104.78
220.331
Lacticaseibacillus paracasei ATCC 334
68.4385
203.645
Lacticaseibacillus paracasei ssp. tolerans DSM 20258
162.251
227.785
Lacticaseibacillus casei ATCC 393
120.599
342.299
Lacticaseibacillus casei DSM 20178
15.4255
295.214
Lacticaseibacillus paracasei ssp. paracasei JCM 8130
38.7665
371.139
Latilactobacillus fuchuensis JCM 11249
738.302
185.255
Latilactobacillus curvatus DSM 20019
10.2715
202.367
Latilactobacillus graminis DSM 20719
270.346
104.833
Latilactobacillus sakei ssp. sakei DSM 20017
98.1526
161.129
Latilactobacillus sakei ssp. carnosus DSM 15831
275.708
Furfurilactobacillus siliginis DSM 22696
1678.89
413.357
Furfurilactobacillus rossiae DSM 15814
420.53
Paucilactobacillus oligofermentans DSM 15707
259.437
341.469
Paucilactobacillus hokkaidonensis LOOC260
292.94
202.932
163.897
Paucilactobacillus wasatchensis WDC04
100.144
211.05
Paucilactobacillus suebicus DSM 5007
380.693
292.857
Paucilactobacillus vaccinostercus DSM 20634
425.57
Limosilactobacillus reuteri JCM 1112
95.2067
343.712
222.796
Limosilactobacillus pontis DSM 8475
82.7113
304.547
Limosilactobacillus vaginalis DSM 5837
57.6594
112.319
235.839
Limosilactobacillus frumenti DSM 13145
50.3845
217.745
Limosilactobacillus panis DSM 6035
366.56
347.716
Limosilactobacillus antri DSM 16041
168.453
85.4865
Limosilactobacillus oris DSM 4864
805.172
767.433
Limosilactobacillus mucosae DSM 13345
793.617
Limosilactobacillus ingluviei DSM 15946
257.671
338.943
Limosilactobacillus secaliphilus DSM 17896
122.957
230.968
693.856
Limosilactobacillus coleohominis DSM 14060
26.001
323.294
Limosilactobacillus equigenerosi DSM 18793
166.157
292.467
Limosilactobacillus gastricus DSM 16045
129.534
190.096
Limosilactobacillus fermentum ATCC 14931
364.633
212.891
Limosilactobacillus gorillae KZ01
157.617
440.569
Apilactobacillus ozensis DSM 23829
523.624
112.7
Apilactobacillus apinorum Fhon13
524.719
151.838
Apilactobacillus kunkeei YH-15
95.4007
474.815
Fructilactobacillus fructivorans DSM 20203
478.822
649.101
Fructilactobacillus sanfranciscensis DSM 20451
582.176
347.628
Fructilactobacillus lindneri JCM 11027
285.429
455.012
Fructilactobacillus florum DSM 22689
183.118
389.188
Lentilactobacillus senioris DSM 24302
66.2354
443.011
Lentilactobacillus curieae CCTCC M 2011381
325.997
435.593
Lentilactobacillus parafarraginis DSM 18390
102.461
305.479
Lentilactobacillus farraginis DSM 18382
108.234
390.036
212.427
Lentilactobacillus hilgardii DSM 20176
435.607
Lentilactobacillus diolivorans DSM 14421
147.632
245.621
Lentilactobacillus kisonensis DSM 19906
80.7282
183.437
Lentilactobacillus rapi DSM 19907
13.3874
144.721
Lentilactobacillus buchneri DSM 20057
115.56
131.84
Lentilactobacillus parabuchneri DSM 5707
224.197
36.1643
200.974
Lentilactobacillus sunkii DSM 19904
50.7881
83.0625
Lentilactobacillus otakiensis DSM 19908
16.4169
135.39
Lentilactobacillus kefiri DSM 20587
328.714
Lactiplantibacillus xiangfangensis LMG 26013
1136.16
454.948
Lactiplantibacillus fabifermentans DSM 21115
106.089
296.062
Lactiplantibacillus herbarum TCF032-E4
195.394
189.736
Lactiplantibacillus paraplantarum DSM 10667
99.1834
362.205
Lactiplantibacillus pentosus DSM 20314
114.22
198.622
Lactiplantibacillus plantarum ssp. plantarum CGMCC 1.2437
42.3583
197.982
Lactiplantibacillus plantarum ssp. argentoratensis DSM 16365
184.621
Secundilactobacillus oryzae SG293
167.618
555.24
859.262
Secundilactobacillus malefermentans DSM 5705
152.706
502.29
Secundilactobacillus kimchicus JCM 15530
544.071
178.114
Secundilactobacillus collinoides DSM 20515
503.204
342.609
356.701
Secundilactobacillus paracollinoides DSM 15502
1146.41
Secundilactobacillus silagei JCM 19001
161.862
592.25
Secundilactobacillus similis DSM 23365
117.798
20.7848
84.3825
Secundilactobacillus odoratitofui DSM 19909
282.592
Levilactobacillus spicheri DSM 15429
217.316
243.959
Levilactobacillus namurensis DSM 19117
120.482
68.0263
Levilactobacillus zymae DSM 19395
268.512
166.947
Levilactobacillus acidifarinae DSM 19394
499.372
413.746
Levilactobacillus paucivorans DSM 22467
190.798
189.737
Levilactobacillus parabrevis ATCC 53295
147.507
868.067
Levilactobacillus koreensis JCM 16448
167.114
172.481
Levilactobacillus senmaizukei DSM 21775
56.0803
286.37
Levilactobacillus hammesii DSM 16381
73.2809
136.875
Levilactobacillus brevis BSO 464
266.869
84.1095
Levilactobacillus brevis DSM 20054


b)
12827.5
Escherichia coli ATCC11775
1933.58
232.303
Amylolactobacillus amylophilu
83.3603
Amylolactobacillus amylotropl
534
523.179
782.208
1519.03
Lactobacillus iners DSM 1333
269.737
Lactobacillus hominis DSM 2
459.084
165.555
Lactobacillus johnsonii ATCC 33200
133.048
305.058
Lactobacillus gasseri ATCC 33323
86.8298
59.2619
Lactobacillus taiwanensis DSM 21401
909.509
1147.35
767.067
Lactobacillus equicursoris DSM 19284
165.409
Lactobacillus delbrueckii ssp. bulgaricus ATCC 11842
551.298
234.618
Lactobacillus delbrueckii ssp. delbrueckii KACC 13439
31.184
136.806
Lactobacillus delbrueckii ssp. sunkii JCM 17838
59.318
119.527
Lactobacillus delbrueckii ssp. indicus JCM 15610
28.7511
89.9207
Lactobacillus delbrueckii ssp. lactis DSM 20072
72.256
167.621
Lactobacillus delbrueckii ssp. jakobsenii ZN7a-9
305.412
1289.17
54.517
Lactobacillus jensenii DSM 20557
190.375
Lactobacillus psittaci DSM 15354
1106.04
475.571
Lactobacillus apis Hma11
189.89
Lactobacillus melliventris Hma8
354.382
293.42
Lactobacillus helsingborgensis Bma5
31.4374
138.862
Lactobacillus kullabergensis Biut2
161.056
148.522
Lactobacillus kimbladii Hma2
27.5536
311.766
990.946
Lactobacillus acetotolerans DSM 20749
617.138
272.161
Lactobacillus pasteurii DSM 23907
107.472
Lactobacillus gigeriorum DSM 23908
70.9056
28.3587
720.37
Lactobacillus kalixensis DSM 16043
516.066
Lactobacillus intestinalis DSM 6629
55.6218
249.848
Lactobacillus hamsteri DSM 5661
311.419
522.432
Lactobacillus amylolyticus DSM 11664
0
423.697
125.31
439.866
Lactobacillus acidophilus ATCC 4356
216.438
168.749
Lactobacillus gallinarum DSM 10532
334.992
Lactobacillus helveticus DSM 20075
200.576
36.0972
340.594
Lactobacillus ultunensis DSM 16047
19.7997
245.915
150.37
Lactobacillus kitasatonis DSM 16761
39.7729
179.598
Lactobacillus amylovorus DSM 20531
305.36
Lactobacillus crispatus DSM 20584
50.5989
77.4056
Lactobacillus kefiranofaciens ssp. kefiranofaciens DSM 5016
122.886
118.569
Lactobacillus kefiranofaciens ssp. kefirgranum DSM 10550
1499.32
1389.65
Bombilactobacillus mellifer Bin4
1062.97
Bombilactobacillus mellis Hon2
208.254
297.547
232.488
465.155
Companilactobacillus nodensis DSM 19682
477.564
Companilactobacillus tucceti DSM 20183
935.261
347.358
306.334
Companilactobacillus versmoldensis DSM 14857
245.646
Companilactobacillus ginsenosidimutans EMML 3141
350.257
154.071
Companilactobacillus paralimentarius DSM 13238
797.939
299.382
Companilactobacillus alimentarius DSM 20249
354.806
Companilactobacillus futsaii JCM 17355
80.0678
351.603
Companilactobacillus kimchiensis DSM 24716
90.0442
275.356
Companilactobacillus farciminis DSM 20184
84.5473
381.383
Companilactobacillus crustorum LMG 23699
112.251
206.059
Companilactobacillus heilongjiangensis DSM 28069
49.9894
88.3699
Companilactobacillus nantensis DSM 16982
43.9975
366.165
Companilactobacillus mindensis DSM 14500
371.68
1683.13
448.46
Lapidilactobacillus concavus DSM 17758
337.207
Lapidilactobacillus dextrinicus DSM 20335
238.011
2486.53
Holzapfelia floricola DSM 23037
1360.82
Agrilactobacillus composti DSM 18527
165.949
475.148
Schleiferilactobacillus perolens DSM 12744
1413.74
650.257
Schleiferilactobacillus shenzhenensis LY-73
286.498
803.812
Schleiferilactobacillus harbinensis DSM 16991
2045.19
Paralactobacillus selangorensis ATCC BAA 66
1766
Dellaglioa algida DSM 15638
140.839
220.032
396.126
260.035
Liquorilactobacillus satsumensis DSM 16230
97.1104
391.782
Liquorilactobacillus oeni DSM 19972
900.957
Liquorilactobacillus vini DSM 20605
141.491
391.995
Liquorilactobacillus ghanensis DSM 18630
895.041
890.723
Liquorilactobacillus nagelii DSM 13675
260.421
242.701
409.651
Liquorilactobacillus cacaonum DSM 21116
402.981
246.613
Liquorilactobacillus hordei DSM 19519
142.036
148.227
Liquorilactobacillus mali DSM 20444
149.825
141.842
Liquorilactobacillus uvarum DSM 19971
93.0781
125.148
Liquorilactobacillus aquaticus DSM 21051
20.055
464.822
133.123
Liquorilactobacillus sucicola DSM 21376
94.8258
232.195
Liquorilactobacillus capillatus DSM 19910
428.68
1322.24
Ligilactobacillus ruminis ATCC 27780
362.025
906.602
Ligilactobacillus equi DSM 15833
396.055
694.804
Ligilactobacillus agilis DSM 20509
196.924
225.447
Ligilactobacillus apodemi DSM 16634
1225.6
114.2
Ligilactobacillus murinus DSM 20452
88.8604
149.577
Ligilactobacillus animalis DSM 20602
224.82
2064.95
Ligilactobacillus ceti DSM 22408
1686.38
Ligilactobacillus saerimneri DSM 16049
264.706
1365.59
216.073
Ligilactobacillus acidipiscis DSM 15836
46.851
801.547
Ligilactobacillus pobuzihii KCTC 13174
299.483
53.6411
Ligilactobacillus aviarius DSM 20655
1391.75
135.352
Ligilactobacillus araffinosus DSM 20653
84.0294
72.5981
544.372
Ligilactobacillus salivarius ATCC 11741
427.64
1045.57
Ligilactobacillus hayakitensis DSM 18933
1347.03
Loigolactobacillus bifermentans DSM 20003
730.482
2111.02
Loigolactobacillus rennini DSM 20253
493.19
274.199
Loigolactobacillus coryniformis ssp. coryniformis DSM 20001
717.274
173.619
Loigolactobacillus coryniformis ssp. torquens DSM 20004
858.905
Lacticaseibacillus camelliae DSM 22697
330.244
527.318
Lacticaseibacillus manihotivorans DSM 13343
103.227
18.3756
864.107
Lacticaseibacillus nasuensis JCM 17158
269.852
1567.56
Lacticaseibacillus sharpeae DSM 20505
215.943
123.989
Lacticaseibacillus pantheris DSM 15945
1366.03
286.6
Lacticaseibacillus thailandensis DSM 22698
359.562
424.028
476.95
Lacticaseibacillus saniviri DSM 24301
543.081
Lacticaseibacillus brantae DSM 23927
209.384
84.9026
Lacticaseibacillus rhamnosus DSM 20021
1044.81
154.365
83.3554
Lacticaseibacillus paracasei ATCC 334
314.669
Lacticaseibacillus paracasei ssp. tolerans DSM 20258
172.604
331.734
Lacticaseibacillus casei ATCC 393
25.601
210.786
Lacticaseibacillus casei DSM 20178
16.8872
276.701
Lacticaseibacillus paracasei ssp. paracasei JCM 8130


1347.03
Loigolactobacillus bifermentans DSM 20003
730.482
2111.02
Loigolactobacillus rennini DSM 20253
493.19
274.199
Loigolactobacillus coryniformis ssp. coryniformis DSM 20001
717.274
173.619
Loigolactobacillus coryniformis ssp. torquens DSM 20004
858.905
Lacticaseibacillus camelliae DSM 22697
330.244
527.318
Lacticaseibacillus manihotivorans DSM 13343
103.227
18.3756
864.107
Lacticaseibacillus nasuensis JCM 17158
269.852
1567.56
Lacticaseibacillus sharpeae DSM 20505
215.943
123.989
Lacticaseibacillus pantheris DSM 15945
1366.03
286.6
Lacticaseibacillus thailandensis DSM 22698
359.562
476.95
Lacticaseibacillus saniviri DSM 24301
424.028
543.081
Lacticaseibacillus brantae DSM 23927
209.384
84.9026
Lacticaseibacillus rhamnosus DSM 20021
1044.81
83.3554
Lacticaseibacillus paracasei ATCC 334
154.365
314.669
Lacticaseibacillus paracasei ssp. tolerans DSM 20258
172.604
331.734
Lacticaseibacillus casei ATCC 393
25.601
210.786
Lacticaseibacillus casei DSM 20178
16.8872
276.701
Lacticaseibacillus paracasei ssp. paracasei JCM 8130
29.7079
367.472
Latilactobacillus fuchuensis JCM 11249
1294.16
124.749
Latilactobacillus curvatus DSM 20019
76.5549
212.861
Latilactobacillus graminis DSM 20719
529.901
80.3206
Latilactobacillus sakei ssp. sakei DSM 20017
89.4424
82.6166
Latilactobacillus sakei ssp. carnosus DSM 15831
611.396
Furfurilactobacillus siliginis DSM 22696
1879.06
213.044
Furfurilactobacillus rossiae DSM 15814
910.641
Paucilactobacillus oligofermentans DSM 15707
406.538
455.358
234.397
Paucilactobacillus hokkaidonensis LOOC260
356.756
846.824
Paucilactobacillus wasatchensis WDC04
120.503
163.461
Paucilactobacillus suebicus DSM 5007
385.034
267.267
Paucilactobacillus vaccinostercus DSM 20634
470.318
Limosilactobacillus reuteri JCM 1112
160.533
497.767
512.482
Limosilactobacillus pontis DSM 8475
95.0487
116.774
377.772
Limosilactobacillus vaginalis DSM 5837
204.217
54.9470
362.064
Limosilactobacillus frumenti DSM 13145
298.619
Limosilactobacillus panis DSM 6035
386.809
204.543
Limosilactobacillus antri DSM 16041
1110.9
173.5
96.3139
Limosilactobacillus oris DSM 4864
806.926
Limosilactobacillus mucosae DSM 13345
727.231
Limosilactobacillus ingluviei DSM 15946
316.076
651.857
Limosilactobacillus secaliphilus DSM 17896
168.268
369.669
612.771
Limosilactobacillus coleohominis DSM 14060
41.4745
530.924
Limosilactobacillus equigenerosi DSM 18793
309.383
312.097
Limosilactobacillus gastricus DSM 16045
214.696
237.373
Limosilactobacillus fermentum ATCC 14931
515.211
395.168
Limosilactobacillus gorillae KZ01
209.855
702.349
Apilactobacillus ozensis DSM 23829
969.507
196.247
Apilactobacillus apinorum Fhon13
303.952
213.386
Apilactobacillus kunkeei YH-15
329.041
859.003
Fructilactobacillus fructivorans DSM 20203
898.298
964.229
Fructilactobacillus sanfranciscensis DSM 20451
705.236
314.391
Fructilactobacillus lindneri JCM 11027
333.79
676.775
Fructilactobacillus florum DSM 22689
369.986
1038.17
Lentilactobacillus senioris DSM 24302
209.731
690.996
Lentilactobacillus curieae CCTCC M 2011381
356.82
429.441
Lentilactobacillus parafarraginis DSM 18390
99.2243
266.392
Lentilactobacillus farraginis DSM 18382
248.905
220.168
313.34
Lentilactobacillus hilgardii DSM 20176
387.091
Lentilactobacillus diolivorans DSM 14421
147.756
282.513
Lentilactobacillus kisonensis DSM 19906
88.3208
250.87
Lentilactobacillus rapi DSM 19907
46.8461
208.57
Lentilactobacillus buchneri DSM 20057
308.144
161.232
Lentilactobacillus parabuchneri DSM 5707
55.6222
110.464
Lentilactobacillus sunkii DSM 19904
209.984
89.6902
139.537
Lentilactobacillus otakiensis DSM 19908
60.4313
207.865
Lentilactobacillus kefiri DSM 20587
469.199
Lactiplantibacillus xiangfangensis LMG 26013
1142.7
374.333
Lactiplantibacillus fabifermentans DSM 21115
80.19
432.203
Lactiplantibacillus herbarum TCF032-E4
175.637
99.3846
Lactiplantibacillus paraplantarum DSM 10667
67.676
83.0431
Lactiplantibacillus pentosus DSM 20314
224.03
72.0321
Lactiplantibacillus plantarum ssp. plantarum CGMCC 1.2437
312.787
100.392
Lactiplantibacillus plantarum ssp. argentoratensis DSM 16365
358.816
Secundilactobacillus oryzae SG293
150.761
1087.77
674.457
Secundilactobacillus malefermentans DSM 5705
232.054
668.13
Secundilactobacillus kimchicus JCM 15530
578.954
123.175
Secundilactobacillus collinoides DSM 20515
372.044
312.549
289.762
Secundilactobacillus paracollinoides DSM 15502
454.555
Secundilactobacillus silagei JCM 19001
242.619
252.655
Secundilactobacillus similis DSM 23365
220.536
179.788
Secundilactobacillus odoratitofui DSM 19909
24.7418
459.679
Levilactobacillus spicheri DSM 15429
303.47
534.942
Levilactobacillus namurensis DSM 19117
174.587
173.221
Levilactobacillus zymae DSM 19395
264.299
77.1423
Levilactobacillus acidifarinae DSM 19394
621.202
716.622
Levilactobacillus paucivorans DSM 22467
278.313
327.19
Levilactobacillus parabrevis ATCC 53295
206.93
271.52
Levilactobacillus koreensis JCM 16448
227.965
417.107
Levilactobacillus senmaizukei DSM 21775
179.563
244.996
Levilactobacillus hammesii DSM 16381
125.563
93.8905
Levilactobacillus brevis BSO 464
373.962
142.125
Levilactobacillus brevis DSM 20054

3.10: Phylogenetic tree mapped with gain (a) and loss (b) expected number.

The phylogenetic tree of the 16S rRNA gene mapped with the expected value of gain/loss events was obtained from GLOOME analysis. The numbers attached to each branch indicate the expected number of gain (a) or loss (b) events.

Table 3.4: Pearson correlation values between genomic features.

Each alphabet represents genomic features as follows a) genome size (total_sequence_length), b) number of proteins (Np), c) protein number minus delta of expected value of gain/loss events (Np-(Eg -El)), d) genetic diversity in the bacteria before speciation (Gd), e) rate of gain/loss events (Eg/El), f) expected value of gain events (Eg), g) expected value of loss events (El), h) expected value of gain events per branch length (Eg/Lb), i) expected value of loss events per branch length (El/Lb), j) normalized expected value of gain/loss events for each branch (Egl), k) normalized expected value of gain events (Ng), l) number of orthologs (On).

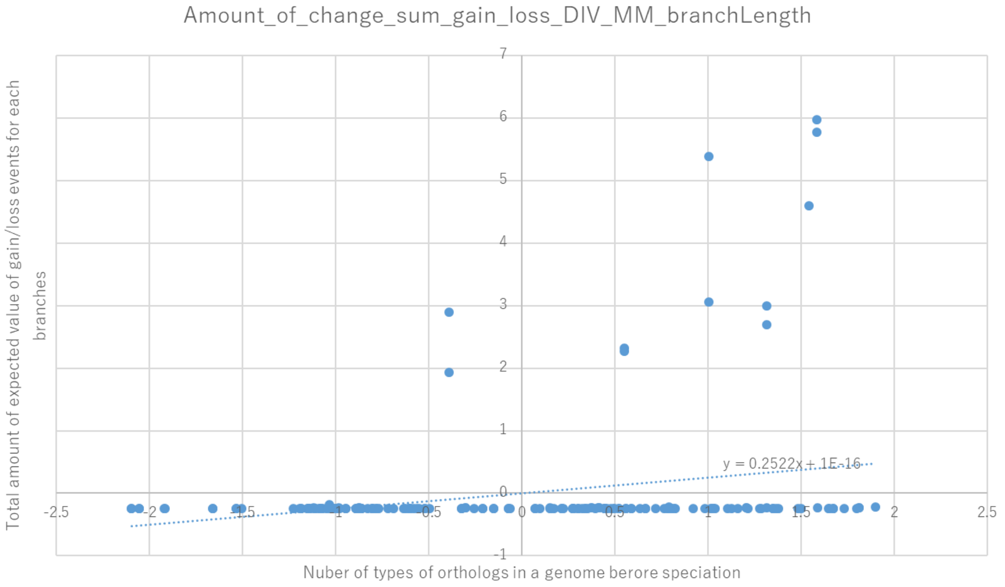
	a	b	c	d	e	f	g	h	i	j	k	l
a	1.00	0.98	0.91	0.82	0.43	0.02	0.25	0.31	0.22	0.28	0.27	0.97
b	0.98	1.00	0.91	0.82	0.46	0.06	0.24	0.32	0.24	0.30	0.26	0.99
c	0.91	0.91	1.00	0.98	0.23	0.21	0.06	0.28	0.23	0.27	0.21	0.90
d	0.82	0.82	0.98	1.00	0.13	0.27	0.19	0.26	0.22	0.25	0.17	0.83
e	0.43	0.46	0.23	0.13	1.00	0.00	0.31	0.42	0.17	0.33	0.54	0.46
f	0.02	0.06	0.21	0.27	0.00	1.00	0.88	0.08	0.10	0.09	0.03	0.08
g	0.25	0.24	0.06	0.19	0.31	0.88	1.00	0.16	0.13	0.15	0.12	0.23
h	0.31	0.32	0.28	0.26	0.42	0.08	0.16	1.00	0.84	0.97	0.71	0.32
i	0.22	0.24	0.23	0.22	0.17	0.10	0.13	0.84	1.00	0.94	0.22	0.25
j	0.28	0.30	0.27	0.25	0.33	0.09	0.15	0.97	0.94	1.00	0.53	0.30
k	0.27	0.26	0.21	0.17	0.54	0.03	0.12	0.71	0.22	0.53	1.00	0.25
l	0.97	0.99	0.90	0.83	0.46	0.08	0.23	0.32	0.25	0.30	0.25	1.00

3.3.2 Influence of ortholog number in a genome on gain and loss events

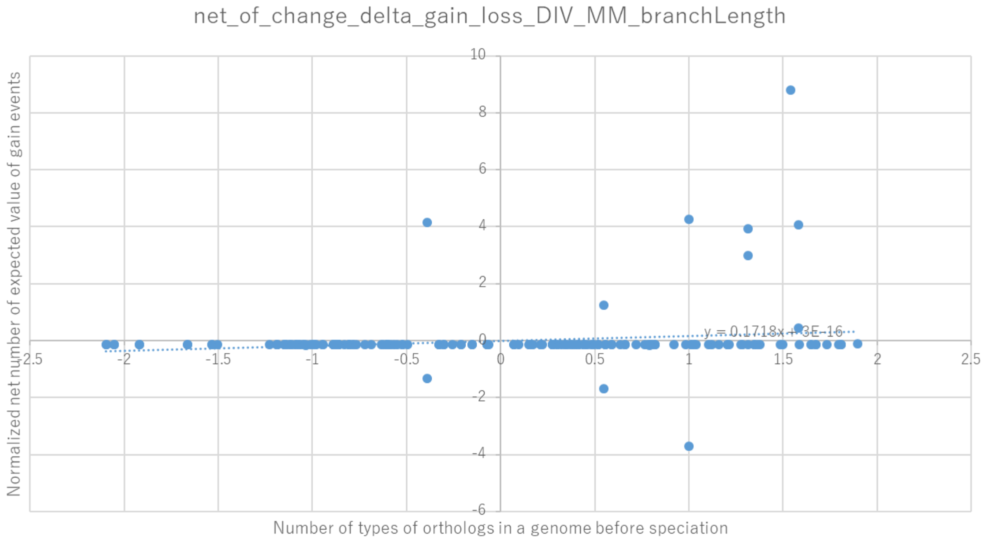
To investigate the influence of genetic diversity in a genome on the number of gain/loss events, simple regression analysis between indexes generated by the number of ortholog types and expected values of gain/loss was performed (Figure 3.11). The number of types classified based on ortholog analysis represents genetic diversity in a genome. As a result, the genetic diversity in a genome before speciation influenced increase in the total expected values of gain/loss in the branch of speciation. The P-value was less than 0.05 (P-value= 0.001). The regression coefficient and the coefficient of determination (R2) obtained were 0.2522 and 0.064, respectively. In addition, the genetic diversity in a genome before speciation influenced increase in the net number of the expected value of gain events. The P-value was less than 0.05 (P-value= 0.022). The regression coefficient and the coefficient of determination (R2) obtained were 0.1718 and 0.030, respectively. Moreover, the total expected values of gain/loss in a branch of speciation influenced the genetic diversity in the current genome. The P-value was less than 0.05 (P-value= 0.000). The regression coefficient and the coefficient of determination (R2) obtained were 0.301 and 0.091, respectively (Table 3.6).

In Section 3.3, the statistical results suggest that the rich (i.e., strains with genetic diversity in a genome) obtain potential for gain/loss events, making the rich richer. This tendency can be interpreted as genetic capitalism in LAB.

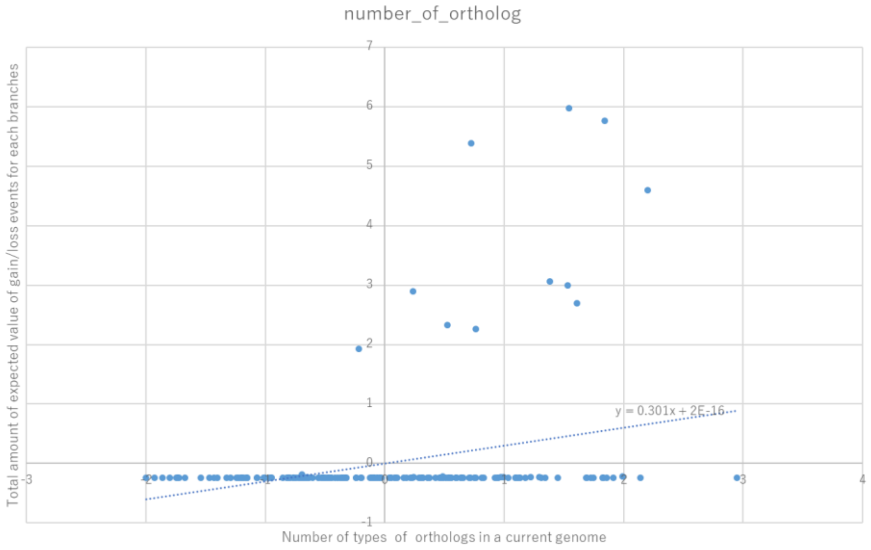
a)


Amount_of_change_sum_gain_loss_DIV_MM_branchLength
Total amount of expected value of gain/loss events for each branches
Nuber of types of orthologs in a genome berore speciation
y = 0.2522x + 1E-16
7
6
5
4
3
2
1
0
-1
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5

b)


net_of_change_delta_gain_loss_DIV_MM_branchLength
Normalized net number of expected value of gain events
Number of types of orthologs in a genome before speciation
y = 0.1718x + 3E-16
10
8
6
4
2
0
-2
-4
-6
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5

c)



3.11: Scatter plot among gain/loss expected number and each parameter.

a) Vertical axis indicates the normalized expected value of gain/loss events for each branch (E_{gl}), and horizontal axis indicates the genetic diversity in the bacteria before speciation (G_d). b) Vertical axis indicates the normalized expected value of gain events (N_g) and horizontal axis indicates the genetic diversity in the bacteria before speciation (G_d), and c) vertical axis indicates the expected value of gain/loss events for each branch (E_{gl}) and horizontal axis indicates the genetic diversity in the bacteria after speciation (O_n).

Table 3.5: Statistics of simple regression analysis for genetic capitalism.

Obj	E_{gl}	N_g	O_n
Exp	G_d	G_d	E_{gl}
R-squared	0.064	0.03	0.091
coef	0.2522	0.1718	0.301
P-value	0.001	0.022	0
std err	0.073	0.074	0.072

Chapter 4: General Discussion

4.1 Niche construction and GECs in LAB

In this study, I elucidated the process of GEC formation in ecological niches, which provided a perspective connected to the process of bacterial evolution. In Chapter 1, I mentioned the influence of GECs in the ecological niche on bacterial evolution. Phenotypic and genomic data in LAB are suitable for this investigation. In Chapter 2, I described the materials and methods used for the investigation of construction of GECs in the ecological niche of LAB. In Chapter 3, I investigated the relationship between phenotypes and HGT in LAB. The results suggested that utilizing various sugars increases potential to acquire multiple genes via HGT. In addition, I indicated the GECs across genera in LAB sharing ecological niches to investigate the ortholog networks. Moreover, genetic diversity in the genome increases the potential of bacteria to undergo gene gain/loss events, which further enriches genetic diversity. These results that phenotypes in LAB contribute to forming GECs suggest that niche construction in LAB forms GECs.

Niche construction is explained as the interaction that organisms change the environment in their habitat and the changing environment affects the evolution of the organisms. To change the environment, organisms have two options: perturbation and relocation. Organisms perturb current habitats or relocate to another habitat, which changes environmental factors that affect organisms (Odling-Smee et al. 2003). In other words, selective pressures from environments to organisms are altered because the organisms affect the environments.

As I mentioned, in this study, the results that phenotype of utilizing various sugars contributes to forming GECs suggest that niche construction in LAB forms GECs. LAB strains can relocate to another habitat using this phenotype, which allows them to form GECs in ecological niches. In addition, the genes obtained via HGT encode phenotypes involving sugar utilization and adaptation to various environments. This suggests that GECs help the members to relocate to new environments. The phenotype to utilize various sugars contributes to sharing ecological niches can paraphrase niche construction in LAB.

4.2 Genetic capitalism in LAB

In Section 3.3, the estimation of gain/loss events and statistical analysis with genomic diversity was performed to investigate the influence of genetic capitalism on LAB evolution. The result suggested that the genetic diversity in a genome before speciation contributes to increasing the potential for gain/loss events, which makes the genome acquire various genetic materials. These results are consistent with the framework of genetic capitalism.

Particular phenotypes cause genetic capitalism, which enriches genetic diversity in the genome. A typical example is antibiotic resistance. Baquero et al. (2003) mentioned that certain genes encode phenotypes that help survival in the local environment, increasing the possibility of gene exchange. As a result, the individuals possessing these genes obtain various genetic materials. In Section 3.3, the result showed that strains containing multiple genes in the genomes had more potential to gain other genes.

In genetic capitalism in LAB, the scenario that the phenotypes allowing adaptation to various environments (phenotype for generalists) increase potential for gene gain/loss events was considered. Genetically rich bacteria have a wide range of habitation because they possess genetic diversity in their genome, including genes to encode phenotype to survive. They acquire genes that help them share ecological niches. Their habitats frequently change because they can inhabit in a wide range of environments. There are many opportunities for gene loss events because purifying selection does not work continuously. Consequently, genetic diversity is increased because of change in genome composition and selective pressure.

The results in Sections 3.1 and 3.2 indicate that the phenotype to utilize various sugars increases HGT events to share ecological niches. Applying the concept of genetic capitalism to formation of GEC in the ecological niche, the generalists that possess phenotypes to utilize various sugars and the specialists can be considered rich and poor bacteria, respectively. The rich obtain opportunities to gain various genes to share ecological niches. Some genes help the rich to relocate to new habitats. As a result, the genetic diversity in the genomes of generalists increases.

Although the statistical analysis in Section 3.3 suggested the tendency of genetic capitalism in LAB, the results do not show a robust model of evolution. The coefficients of determination in simple regression analyses were low. To construct robust models, sophisticated indexes are required. For instance, an index based on genes to enable the adaptation of various niches related to stress responses, bacteriocin production, antibiotic resistance, survival in the intestinal environment, and heavy metal resistance is required.

In conclusion, genetic diversity in a genome before speciation increases the potential for gain/loss events, which further enriches the diversity in the current genome. The results suggest a framework of genetic capitalism underlying construction of ecological niche and GEC in LAB.

4.3 Influence of niche construction in LAB

Niche construction in bacterial evolution changes selective pressure on the bacteria and mutation rate in their genomes because of GEC formation. Niche construction is the process where organisms change their environment using their phenotypes, which affects selective pressure exerted on organisms themselves. The less influence of geographical isolation on bacteria promotes this tendency. In addition, bacteria can distribute in a wide range of environments and construct ecological niches because they have a huge population size. Moreover, bacteria can adapt to the changing environment because their generation cycle is extremely fast (Odling-Smee et al. 2003). Furthermore, bacteria can share the genetic material to form GECs in an ecological niche. These characteristics allow niche construction in bacteria, which has a greater influence on their evolution than on large creatures.

In addition, I investigated whether niche construction causes genetic capitalism in LAB. As a result, the tendency that rich bacteria possessing genetic diversity in the genome have potential to gain genes was observed. Genes encode phenotypes to adapt to environments; this in turn helps construct GECs in various ecological niches. However, poor bacteria that possess less genetic diversity in the genome have few opportunities to gain genes. The non-essential genes to survive in a particular niche are deleted from the genome of poor bacteria because they do not relocate and stay in the same niche. These biases widen the gap between generalists that adapt to a wide range of environments and specialists that adapt to particular niches.

In conclusion, the niche construction in LAB increases the mutation rate to construct GECs in the ecological niche, which may cause genetic capitalism. The niche construction with GEC may play a major role in bacterial evolution.

4.4 Complicated bacterial evolution

The evolution of bacteria is not easy to unravel because they interact with many environmental factors in the ecosystem. The evolution of bacteria, compared to that of eukaryotes, is known to have the following three characteristics: 1. less influence of geographical isolation, 2. Sharing of ecological niche in a suitable environment, 3. exchange of genetic materials between distantly related species. These factors obscure the definition of a population and complicate our understanding of bacterial evolution (VanInsberghe et al. 2020; Rocha 2018; Arevalo et al. 2019). Therefore, for a deep understanding of bacterial evolution, a novel theory is required (Rocha 2018). As described below, these three characteristics have been studied independently. Although the three characteristics interact, few discussions integrate the three characteristics. I build a novel framework of bacterial evolution, including the three characteristics. Through my research, I deduced that introducing concepts such as GECs and niche construction helps to effectively build a novel framework for bacterial evolution.

Bacteria can relocate without the influence of geographic isolation. Environmental factors are fluctuated by relocation in the bacterial evolution (Kirchman 2012). Microorganisms differ from large creatures in relocation behavior. Microorganisms appear in particular habitats and are not influenced by geographic isolation. However, large creatures have habitats according to geographic conditions. For instance, gazelles live in the savannahs of Africa, and pronghorn live in North America, but the reverse is not true. The bacterial relocation is described as "everything is everywhere, but the environment selects" as Becking's hypothesis (Kirchman 2012). As evidence of this hypothesis, he listed the hugeness of the population, the smallness of cell size, and asexual reproduction. These features help bacteria overcome geographic conditions. Although some reports are inconsistent with Becking's hypothesis (Pagaling et al. 2009; Martiny et al. 2006), it is true that bacteria are influenced less than large creatures.

Multiple bacterial lineages coexist in suitable environments; it is easier for bacteria to share the ecological niche because they can relocate without the influence of geographic isolation. Logically, a single cell adapts to a new habitat, and bacteria can grow there. In addition, bacteria are not affected by geographic isolation. Based on these conditions, bacteria can stay together in ecological niches in suitable environments for survival without geographical influence. Multiple bacterial lineages can share the ecological niches unless they do not compete with each other. Evidence consistent with this hypothesis was reported in bacteria and archaea (Martiny et al. 2006). Microflora in soils that have close properties in different latitudes were similar. Moreover, the microflora in water from Antarctica and the Arctic were similar (Fierer and Jackson 2006). This evidence suggests that bacteria share ecological niches until the environments are suitable.

Sharing ecological niches induces HGT among multiple bacterial lineages. Bacteria exchange their genes indiscriminately via HGT, which makes the line of descent difficult (Schleifer et al. 2008; Rocha 2018). Mainly, there are three mechanisms of HGT: conjugation, transformation, and transduction (Soucy et al. 2015). Although these mechanisms generate bias to promote gene transfer among closely related organisms, many reports show that HGT also occurs among distantly related organisms, allowing sharing of ecological niches. For example, different phylum microorganisms share genes for surviving in a high-temperature environment (Andam and Gogarten 2011). Distantly related microorganisms can share their features via HGT, which in turn contributes to their environmental adaptation. This obscures the definition of a bacterial population and makes bacterial evolution difficult to understand using population genetics (Rocha 2018).

4.5 Hypothetical framework: Niche Construction and GECs model

Based on the above discussion, I construct the framework to comprehend bacterial evolution deeply. The relationship between the three characteristics in bacterial evolution should be described correctly. In bacterial evolution, relocation without the influence of geographic isolation allows sharing of ecological niches. Sharing of ecological niches allows frequent HGT among distantly related microorganisms. Thus, in bacterial evolution, the simple flow is suggested as follows: relocation without the influence of geographic isolation makes bacteria share ecological niches, which causes frequent HGT. To understand bacterial evolution better, I need to bring two concepts into this simple flow: niche construction and GECs.

As mentioned in Section 4.1, the results that phenotype of utilizing various sugars contributes to forming GECs suggests that niche construction in LAB forms GECs. LAB can relocate to another habitat using a sugar utilization phenotype, which affects their evolution to form GECs in ecological niches.

Niche construction forms GECs, which have a huge influence on bacterial evolution. Notably, in bacterial evolution, niche construction changes the selective pressure on them, and niche construction influences the mutation rate of their genome to form GECs. A further modified flow of bacterial evolution is as follows: relocation without geographic isolation causes sharing of ecological niches and forming GECs, generating high-density regions in the HGT network. This flow can be paraphrased as niche construction changing mutation rate by frequent HGT (GECs). This phenomenon is not observed in large creatures. Therefore, niche construction influences bacteria more than large creatures.

In conclusion, I suggested that the evolutional model of bacteria integrates the three characteristics of bacterial evolution using two concepts. The model is named the "Niche Construction and GECs model (NCG model)" (Figure 4.1). This model can be described as a simplified flow: relocation without the influence of geographic isolation allows bacteria to share ecological niches and form GECs. This flow can be paraphrased as niche construction causing GECs. This model indicates that niche construction not only changes the selective pressure on bacteria but also influences their mutation rate by forming GECs. In evolutionary theory, this interpretation indicates a large difference between prokaryotic and eukaryotic organisms.

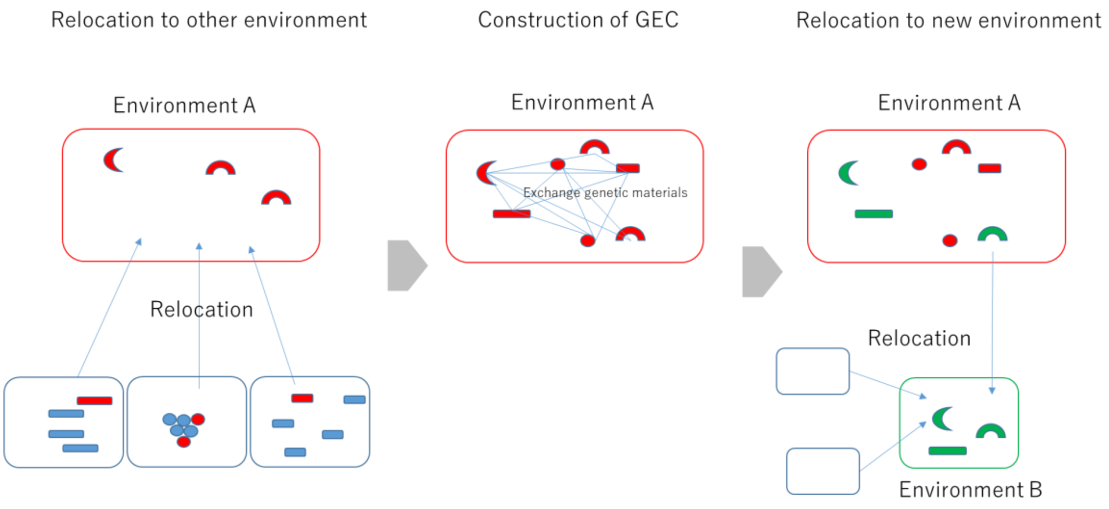


Figure 4.1: Niche construction and the GECs model (NCG model).

Each frame indicates the environment. Each symbol indicates bacterial strains. Different shapes of symbols refer to different lineages. Red symbols indicate that strains possess genetic material to survive in environment A. Emigration without geographic isolation causes sharing of ecological niches and forming GEC, generating high-density regions in the HGT network. This means that niche construction that contributes to forming GECs changes the selective pressure and mutation rate of bacterial genomes.

4.6 Validity of the NCG model

There are a variety of approaches to reveal bacterial evolution. Bacterial evolution is difficult to understand because of their huge population sizes and diverse genomes. The classification of bacteria started by investigating traditional morphology in the 19th century (Schleifer et al. 2009). The development of methods based on ribosomal RNA provided information on phylogenetic relationships in bacterial evolution. Thereafter, analysis based on ribosomal RNA became the gold standard for deducing the phylogenetic relationships of prokaryotes. The development of ultra-high-throughput next-generation sequencing technologies dramatically improved the availability of whole genome sequences for many bacterial strains (Tettelin et al. 2008). The first pan-genome approach compared whole genomes of numerical strains of *Streptococcus agalactiae* to describe their evolution of virulence mechanisms (Tettelin et al. 2002). The results demonstrated shared genetic features and the diversity of genomes in the population. Furthermore, the method of metagenomic analysis also improved because of development of next-generation sequencing. The framework established in the last decade (Caporaso et al., 2010; Qin et al., 2010) describes microflora composition in various niches (Liu et al. 2021). Although these approaches provide us with huge insights, integrated frameworks are required because these approaches are complicated to understand the theory of bacterial evolution.

The NCG model utilized here successfully suggests a simple and integrated framework of the theory of bacterial evolution: less influence of geographical isolation allows formation of GECs in ecological niches, which causes genetic capitalism. This simple scenario helps us to better understand bacterial evolution.

However, to make the model robust, some investigations are required. First, the GECs in ecological niche and genetic capitalism among distantly related lineages should be investigated. In this study, I used *Lactobacillaceae*, a closely related group, because the phenotypic and genomic data of its members have been obtained and provide us with the appropriate sandbox. Furthermore, the research targets should be expanded and phenotypic and genomic data of other bacterial groups should be included. Secondary investigation of more widely phenotypic features is required. Although the basic phenotypic information used for classifying taxon was analyzed in this study, the phenotypic features required to adapt to various environments, such as surviving in the animal intestine and antibiotic resistance, were not. Finally, sophisticated indexes for classifying bacteria as rich or poor in genetic capitalism are required. For instance, genes encoding a phenotype for environmental adaptation can be an effective index of richness.

4.7 Conclusion

I investigated phenotypic and genomic factors in 178 strains of 24 genera in *Lactobacillaceae* to reveal the process of GECs formation in the ecological niche. The results suggested that the capability of utilizing various sugars contributes to the formation of GECs in ecological niches. Moreover, genetic diversity may further increase potential for gene gain events in LAB. Based on the results, I suggested a hypothesis model of the process of forming GECs in ecological niches: the NCG model. The results in this study provide the first evidence that phenotypes associated with ecological niches contribute to forming GECs in the LAB family. Moreover, the results may help to improve our understanding of role of niche construction in bacterial evolution.

References

Andam CP, Gogarten JP. Biased gene transfer in microbial evolution. Nat Rev Microbiol 2011;9;543-55

Anderssen EL, Diep DB, Nes IF et al. Antagonistic activity of *Lactobacillus plantarum* C11: two new two-peptide bacteriocins, plantaricins EF and JK, and the induction factor plantaricin A. Appl Environ Microbiol 1998;64;2269-72

Arevalo P, VanInsberghe D, Elsherbini J et al. A Reverse Ecology Approach Based on a Biological Definition of Microbial Populations. Cell 2019;178;820-834.e14

Baquero F, Coque T, Canton R. Antibiotics, Complexity, and Evolution(2003).Volume 69, ASM News 2003

Baquero F, Coque TM, Galán JC et al. The Origin of Niches and Species in the Bacterial World. Front Microbiol 2021;12;657986

Baquero F. From pieces to patterns: evolutionary engineering in bacterial pathogens. Nat Rev Microbiol 2004;2;510-8

Bergthorsson U, Ochman H. Heterogeneity of genome sizes among natural isolates of Escherichia coli. J Bacteriol 1995;177;5784-9

Bobay LM, Ochman H. The Evolution of Bacterial Genome Architecture. Front Genet 2017, DOI: 10.3389/fgene.2017.00072

Brown NL, Stoyanov JV, Kidd SP, Hobman JL. The MerR family of transcriptional regulators. FEMS Microbiol Rev 2003;27;145-63

Cai Y, Benno Y, Ogawa M et al. Influence of lactobacillus spp. from An inoculant and of *Weissella* and *Leuconostoc* spp. from forage crops on silage fermentation. Appl Environ Microbiol 1998;64;2982-7

Cai Y. Identification and characterization of *Enterococcus* species isolated from forage crops and their influence on silage fermentation. J Dairy Sci 1999;82;2466-71

Canard B, Cole ST. Genome organization of the anaerobic pathogen *Clostridium perfringens*. Proc Natl Acad Sci USA 1989;86;6676-80

Caplice E, Fitzgerald GF. Food fermentations: role of microorganisms in food production and preservation. Int J Food Microbiol 1999;50;131-49

Caporaso JG, Kuczynski J, Stombaugh J et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 2010;7;335-6

Chen YJ, Leung PM, Wood JL et al. Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. ISME J 2021, DOI:10.1038/s41396-021-00988-w.

Cohen O, Ashkenazy H, Belinky F et al. GLOOME: gain loss mapping engine. Bioinformatics 2010;26;2914-5

Cosentino S, Iwasaki W. SonicParanoid: fast, accurate and easy orthology inference. Bioinformatics 2019;35;149-151

Daniels DL. The complete AvrII restriction map of the *Escherichia coli* genome and comparisons of several laboratory strains. Nucleic Acids Res 1990;18;2649-51

Diez A, Gustavsson N, Nyström T et al. The universal stress protein A of *Escherichia coli* is required for resistance to DNA damaging agents and is regulated by a RecA/FtsK-dependent regulatory pathway. Mol Microbiol 2000;36;1494-503

Douglas J. Futuyma and Gabriel Moreno. The Evolution of Ecological Specialization. Annu. Rev. Ecol. Syst. 1988

Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 2004, doi:10.1186/1471-2105-5-113.

Endo A, Futagawa-Endo Y, Dicks LM et al. Isolation and characterization of fructophilic lactic acid bacteria from fructose-rich niches. Syst Appl Microbiol 2009;32;593-600

Fierer N, Jackson RB. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A 2006;103;626-31

Gury J, Barthelmebs L, Tran NP et al. Cloning, deletion, and characterization of PadR, the transcriptional repressor of the phenolic acid decarboxylase-encoding padA gene of *Lactobacillus plantarum*. Appl Environ Microbiol 2004;70;2146-53

Hall JPJ, Brockhurst MA, Harrison E. Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria. Phil. Trans. R. Soc. 2017; B3722016042420160424, DOI: https://doi.org/10.1098/rstb.2016.0424

Harsono KD, Kaspar CW, Luchansky JB et al. Comparison and genomic sizing of *Escherichia coli* O157:H7 isolates by pulsed-field gel electrophoresis. Appl Environ Microbiol 1993;59;3141-4

Holzapfel WH, Wood BJB. Lactic acid bacteria: biodiversity and taxonomy. Chichester:Wiley Blackwell, 2014.

Hu Y, Hu Q, Wei R et al. The XRE Family Transcriptional Regulator SrtR in Streptococcus suis Is Involved in Oxidant Tolerance and Virulence. Front Cell Infect Microbiol 2018, DOI:10.3389/fcimb.2018.00452

Jain R, Rivera MC, Moore JE et al. Horizontal gene transfer accelerates genome innovation and evolution. Mol Biol Evol 2003;20;1598-602

Kirchman D. Processes in Microbial Ecology. Oxford: Oxford University press, 2012.

Kumar S, Stecher G, Li M et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 2018;35;1547-1549

Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 2007;23;127-8

Liu YX, Qin Y, Chen T et al. A practical guide to amplicon and metagenomic analysis of microbiome data. Protein Cell 2021;12;315-330

Makarova K, Slesarev A, Wolf Y et al. Comparative genomics of the lactic acid bacteria. Proc Natl Acad Sci U S A 2006;103;15611-6

Martiny JB, Bohannan BJ, Brown JH et al. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol 2006;4;102-12

Odling-Smee FJ, Laland KN, Feldman MW. Niche Construction: The Neglected Process in Evolution. Princeton: Princeton University Press, 2003.

Pagaling E, Wang H, Venables M et al. Microbial biogeography of six salt lakes in Inner Mongolia, China, and a salt lake in Argentina. Appl Environ Microbiol 2009;75;5750-60

Podell S, Gaasterland T. DarkHorse: a method for genome-wide prediction of horizontal gene transfer. Genome Biol 2007, DOI:10.1186/gb-2007-8-2-r16.

Polz MF, Alm EJ, Hanage WP et al. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. Trends Genet 2013;29;170-5

Prevost G, Jaulhac B, Piemont Y et al. DNA fingerprinting by pulsed-field gel electrophoresis is more effective than ribotyping in distinguishing among methicillin-resistant *Staphylococcus aureus* isolates. J Clin Microbiol 1992;30;967-73

Puigbò P, Wolf YI, Koonin EV et al. The tree and net components of prokaryote evolution. Genome Biol Evol 2010;2;745-56

Qin J, Li R, Raes J et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010 Mar 4;464(7285):59-65.

Ragan MA, Beiko RG. Lateral genetic transfer: open issues. Philos Trans R Soc Lond B Biol Sci 2009;364;2241-51

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4;406-25

Sato M, Miyazaki K. Phylogenetic Network Analysis Revealed the Occurrence of Horizontal Gene Transfer of 16S rRNA in the Genus *Enterobacter*. Front. Microbiol. 2017;8;2225

Satti M, Tanizawa Y, Endo A et al. Comparative analysis of probiotic bacteria based on a new definition of core genome. J Bioinform Comput Biol 2018, DOI: 10.1142/S0219720018400127.

Schleifer KH. Classification of Bacteria and Archaea: past, present and future. Syst Appl Microbiol 2009;32;533-42

Seemann JE, Schulz GE. Structure and mechanism of L-fucose isomerase from Escherichia coli. J Mol Biol 1997;273;256-68

Silva CMG, Silva DNDS, Costa SBD et al. Inactivation of MarR gene homologs increases susceptibility to antimicrobials in Bacteroides fragilis. Braz J Microbiol ;49;200-206

Sim SH, Yu Y, Lin CH et al. The core and accessory genomes of *Burkholderia pseudomallei*: implications for human melioidosis. PLoS Pathog 2008, DOI:10.1371/journal.ppat.1000178.

Skippington E, Ragan MA. Lateral genetic transfer and the construction of genetic exchange communities. FEMS Microbiol Rev 2011;35;707-35

Smoot ME, Ono K, Ruscheinski J et al. Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 2011;27;431-2

Soucy SM, Huang J, Gogarten JP et al. Horizontal gene transfer: building the web of life. Nat Rev Genet 2015;16;472-82

Sriswasdi S, Yang CC, Iwasaki W et al. Generalist species drive microbial dispersion and evolution. Nat Commun 2017;8;1162

Stoddard SF, Smith BJ, Hein R et al. rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Research 2015; 43; D1; D593–D598

Swithers KS, Soucy SM, Gogarten JP et al. The role of reticulate evolution in creating innovation and complexity. Int J Evol Biol 2012, DOI:10.1155/2012/418964.

Takenaka S, Kawashima T, Arita M. A sugar utilization phenotype contributes to the formation of genetic exchange communities in lactic acid bacteria, FEMS Microbiology Letters, Volume 368, Issue 17, September 2021, fnab117

Tanizawa Y, Fujisawa T, Kaminuma E et al. DFAST and DAGA: web-based integrated genome annotation tools and resources. Biosci Microbiota Food Health 2016;35;173-184

Tanskanen EI, Tulloch DL, Hillier AJ et al. Pulsed-Field Gel Electrophoresis of SmaI Digests of Lactococcal Genomic DNA, a Novel Method of Strain Identification. Appl Environ Microbiol 1990;56;3105-11

Tettelin H, Masignani V, Cieslewicz MJ et al. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. Proc Natl Acad Sci U S A 2002;99;12391-6

Tettelin H, Riley D, Cattuto C et al. Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol 2008;11;472-7

VanInsberghe D, Arevalo P, Chien D et al. How can microbial population genomics inform community ecology? Philos Trans R Soc Lond B Biol Sci 2020;375;20190253

Waack S, Keller O, Asper R et al. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics 2006, DOI:10.1186/1471-2105-7-142.

Walter J, Chagnaud P, Tannock GW et al. A high-molecular-mass surface protein (Lsp) and methionine sulfoxide reductase B (MsrB) contribute to the ecological performance of *Lactobacillus reuteri* in the murine gut. Appl Environ Microbiol 2005;71;979-86

Wiedenbeck J, Cohan FM. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. FEMS Microbiology Reviews 2011;35;5;957–976, DOI: https://doi.org/10.1111/j.1574-6976.2011.00292.x

Wu J, Rosen BP. The ArsR protein is a trans-acting regulatory protein. Mol Microbiol 1991;5;1331-6

Yamamoto K, Nakayama J (Eds.). Nyusankin to bifizusukin no saiensu (Science of lactic acid bacteria and bifidobacteria) [published in Japanese]. Kyoto: Kyoto University Press, 2010

Zheng J, Wittouck S, Salvetti E et al. A taxonomic note on the genus Lactobacillus: Description of 23 novel genera, emended description of the genus Lactobacillus Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. Int J Syst Evol Microbiol 2020;70;2782-2858

Zimmer C, Emlen D. Evolution : Making Sense of Life. New York:W.H. Freeman, 2016.

Supplementary Table 2.1: Features of the 178 LAB strains. The accession numbers of the genome sequences, the old and new species names, strain names, type status, seven genomic features, six phenotypic characteristics, and the strains' isolation source are presented. The genomic features are genome size (bp), number of CDS, G/C content (%), number of rRNA, number of tRNA, number of CRISPRs, number of CDS judged to be HGTs. One of phenotypes is sugar utilization value which indicates the number of sugar types that can be utilized. The other five phenotypes, growth at 15 °C, growth at 45 °C, and growth in microaerobic, facultatively anaerobic, and obligate anaerobic conditions were expressed as a dummy variable: If a strain has the feature, 1 was given as the dummy variable and 0 if not. The isolation source indicates the environment in which the species was isolated.

accession number (id)	original name	strain	new name	type status	genome size (total sequence length)	number of CDS	G/C content	number of rRNA	number of tRNA	number of CRISPRS	number of CDS judged HGT	sugar utilization value (number of sugar types to be able to utilize)	growth at 15	growth at 45	micro aerophilic	facultatively anaerobic	anaerobic	isolation source	16S_rRNA_accession
ERR203996	Lactobacillus fermentum	ATCC 14931	Limosilactobacillus fermentum	type strain	1782450	1742	52.8	1	49	1	67	7	0	1	0	1	0	fermented cereals fermenting plant materials dairy products manure sewage the faeces and vagina of humans	AJ575812
ERR387459	Lactobacillus capillatus	DSM 19910	Liquorilactobacillus capillatus	type strain	2224347	2107	37.6	0	41	1	39	6	1	0	0	1	0	fermented brine used for stinky tofu production	AZEF01000036
ERR387460	Lactobacillus manihotivorans	DSM 13343	Lacticaseibacillus manihotivorans	type strain	3081436	3012	47.7	2	50	0	272	12	1	1	0	1	0	sour cassava starch fermentation tomato pomace silage	BBAH01000233
ERR387461	Lactobacillus hayakitensis	DSM 18933	Ligilactobacillus hayakitensis	type strain	1636658	1543	34	3	68	0	69	7	0	1	0	1	0	the faeces of a thoroughbred as predominant species in the intestinal microbiota	BAML01000063
ERR387463	Lactobacillus kefiri	DSM 20587	Lentilactobacillus kefiri	type strain	2322665	2208	41.7	1	57	3	61	4	1	0				kefir as part of the core microbiota	AJ621553
ERR387469	Lactobacillus camelliae	DSM 22697	Lacticaseibacillus camelliae	type strain	2553708	2403	55.4	0	51	0	259	9	0	0				fermented tea (Camellia sinensis) leaves fermented tomato pomace	AYZJ01000044
ERR387471	Lactobacillus alimentarius	DSM 20249	Companilactobacillus alimentarius	type strain	2331920	2232	35.4	3	51	0	70	10	1	0	1	0	0	marinated fish products fermented sausages ready-to-eat meats type I sourdough other plant fermentations	AZDQ01000025
ERR387476	Lactobacillus fabifermentans	DSM 21115	Lactiplantibacillus fabifermentans	type strain	3271316	3111	45	2	60	0	156	5	1	0	0	1	0	cocoa bean heap fermentation fermented grapes fermented cereals	AYGX01000583
ERR387480	Lactobacillus diolivorans	DSM 14421	Lentilactobacillus diolivorans	type strain	3202031	2962	40	0	43	2	67							maize silage vegetable (cucumber) fermentations fermented dairy products	AZEY01000081
ERR387482	Lactobacillus hammesii	DSM 16381	Levilactobacillus hammesii	type strain	2807716	2591	49.4	2	52	3	151	7	1	0	0	1	0	wheat and rye sourdoughs ryegrass silages a municipal biogas plant	AJ632219
ERR387483	Lactobacillus acidifarinae	DSM 19394	Levilactobacillus acidifarinae	type strain	2913834	2738	51.6	1	57	7	199	6	1	0	0	1	0	type I wheat sourdough fermented rice bran	AZDV01000008
ERR387486	Lactobacillus amylotrophicus	DSM 20534	Amylolactobacillus amylotrophicus	type strain	1600645	1602	42.6	2	51	0	181	5	1	0	0	1	0	corn silage	AM236149
ERR387489	Lactobacillus floricola	DSM 23037	Holzapfelia floricola	type strain	1287117	1247	34.5	3	44	3	33	0	1	0	0	1	0	flowers	AYZL01000003
ERR387493	Lactobacillus aquaticus	DSM 21051	Liquorilactobacillus aquaticus	type strain	2399635	2210	37.4	1	48	1	35	10	1	1	1	0	0	eutrophic freshwater pond	AYZD01000026
ERR387495	Lactobacillus futsaii	JCM 17355	Companilactobacillus futsaii	type strain	2490561	2449	35.6	2	51	1	88	9	1	0	0	1	0	traditional fermented mustard products fu-tsai and suan-tsai it has been used experimentally for fermentation of shrimp waste	AZDO01000040

ERR387498	Lactobacillus agilis	DSM 20509	Ligilactobacillus agilis	type strain	2047633	2015	41.7	0	63	3	176	15	0	1	0	1	0	municipal sewage the pigeon crops the gut and cecum of birds human gut and vagina porcine intestinal mucin Nigerian ogi cheese fermented food products such as masau fruits	AYYP01000002
ERR387499	Lactobacillus ingluviei	DSM 15946	Limosilactobacillus ingluviei	type strain	2138634	2086	49.9	0	66	5	252	4	0	0	0	1	0	the crop of a pigeon birds cattle carnivore faeces Korean rice wine (makgeolii)	AZFK01000041
ERR387501	Lactobacillus vaccinostercus	DSM 20634	Paucilactobacillus vaccinostercus	type strain	2553579	2440	43.5	0	52	0	92	5	0	0				cow dung fermented tea leaves fermented cereals	AYYY01000028
ERR387504	Lactobacillus murinus	DSM 20452	Ligilactobacillus murinus	type strain	2159096	2030	40	2	55	0	137	8	0	1	0	1	0	the intestinal tract of mice and rats sourdough	BCVJ01000104
ERR387505	Lactobacillus nagelii	DSM 13675	Liquorilactobacillus nagelii	type strain	2493596	2409	36.7	0	51	1	98	11	0	1	0	1	0	partially fermented wine spontaneous cocoa bean fermentations water kefirs fermented cassava food silage fermentation of fruit residues	AZEV01000015
ERR387506	Lactobacillus pantheris	DSM 15945	Lacticaseibacillus pantheris	type strain	2531803	2293	52.9	2	52	0	228	8	1	0	0	1	0	the faeces of a jaguar in Beijing Zoo fermented vegetables	BCVS01000179
ERR387507	Lactobacillus hamsteri	DSM 5661	Lactobacillus hamsteri	type strain	1790730	1712	35.1	3	58	2	66	14	0	0	0	0	1	the intestine of a hamster	BALY01000063
ERR387508	Lactobacillus gallinarum	DSM 10532	Lactobacillus gallinarum	type strain	1925768	1912	36.5	3	58	0	94	10	1	1	0	1	0	chicken intestine	BALB01000057
ERR387510	Lactobacillus intestinalis	DSM 6629	Lactobacillus intestinalis	type strain	1993045	1838	35.3	3	52	5	53	5	0	1	0	1	0	the intestines of rats, mice and pigs	AZGN01000031
ERR387512	Lactobacillus kitasatonis	DSM 16761	Lactobacillus kitasatonis	type strain	1906076	1917	37.5	2	59	0	153	5	0	1	0	1	0	the intestine of animals including chicken swine	BALU01000027
ERR387520	Lactobacillus psittaci	DSM 15354	Lactobacillus psittaci	type strain	1542511	1344	35.7	3	52	1	29	2	1	1	0	1	0	a hyacinth macaw	AUEI01000022
ERR387522	Lactobacillus plantarum subsp. argentoratensis	DSM 16365	Lactiplantibacillus plantarum ssp. argentoratensis	type strain	3172036	2939	45	0	46	3	181	17	1	0	0	1	0	starchy food fermenting food of plant origin timothy orchardgrass and elephant grass silage fermented Uttapam batter fermented idli batter	CP032751
ERR387524	Lactobacillus mindensis	DSM 14500	Companilactobacillus mindensis	type strain	2326589	2205	38.2	2	53	2	81	5	1	0	1	0	0	type I sourdough	AZEZ01000067
ERR387525	Lactobacillus hordei	DSM 19519	Liquorilactobacillus hordei	type strain	2287468	2239	34.8	0	57	2	87	8	0	0	0	1	0	malted barley water kefirs Turkish tradi tional fermented gilaburu fruit juice	EU074850
ERR387527	Lactobacillus acetotolerans	DSM 20749	Lactobacillus acetotolerans	type strain	1571585	1518	36.2	3	55	3	65	3	0	0	0	1	0	mash fermenta tions for production of grain liquor and vinegar in China and Japan plant fermentations silage intestine of swine ducks	BBBU01000079

																		cattle	
ERR387528	Lactobacillus graminis	DSM 20719	Latilactobacillus graminis	type strain	1829440	1739	40.3	2	51	1	95	5	1	0				grass silage meat products sourdough gut of snail Cornum aspersum grapes	AYZB01000012
ERR387529	Lactobacillus frumenti	DSM 13145	Limosilactobacillus frumenti	type strain	1730467	1676	42.6	2	59	0	75	16	0	1	0	1	0	an industrial rye bran fermentation must wine intestine of poultry and swine	AZER01000001
ERR387530	Lactobacillus aviarius subsp. aviarius	DSM 20655	Ligilactobacillus aviarius	type strain	1674521	1585	40.1	0	43	0	109	7	0	0	0	1	0	the intestine and faeces of birds	AYZA01000007
ERR387532	Lactobacillus selangorensis	ATCC BAA 66	Paralactobacillus selangorensis	type strain	2081509	2064	46.4	1	50	2	125	3	1	0	0	1	0	a Malaysian food ingredient called chili bo	AF049745
ERR387540	Lactobacillus sharpeae	DSM 20505	Lacticaseibacillus sharpeae	type strain	2438466	2344	53.4	3	50	0	342	7	1	0				municipal sewage spoiled meat	M58831
ERR387541	Lactobacillus siliginis	DSM 22696	Furfurilactobacillus siliginis	type strain	2041760	1980	44.1	1	53	0	118	4	0	0	0	1	0	wheat sourdough	AB681446
ERR387542	Lactobacillus similis	DSM 23365	Secundilactobacillus similis	type strain	3452668	3084	47	0	49	3	219	8	1	0	0	1	0	fermented cane molasses at alcohol plants in Thailand rice wine (makgeolii)	AB282889
ERR387543	Lactobacillus spicheri	DSM 15429	Levilactobacillus spicheri	type strain	2742678	2451	55.9	0	46	4	35	3	1	0	0	1	0	wheat and rice sourdoughs fermented vegetables a municipal biogas plant	AJ534844
ERR387546	Lactobacillus taiwanensis	DSM 21401	Lactobacillus taiwanensis	type strain	1865395	1816	33.9	2	52	0	38	7	0	1	0	1	0	the mouse gastrointestinal tract silage cattle feed	AYZG01000031
ERR387549	Lactobacillus tucceti	DSM 20183	Companilactobacillus tucceti	type strain	2170671	2102	34.1	3	56	1	56	5	1	0	1	0	0	sausage	AZDG01000033
ERR387550	Lactobacillus uvarum	DSM 19971	Liquorilactobacillus uvarum	type strain	2671380	2525	36.9	1	53	1	94	8	0	0	0	1	0	Bobal grape musts	AZEG01000088
ERR387553	Lactobacillus animalis	DSM 20602	Ligilactobacillus animalis	type strain	1870553	1812	41.1	1	51	2	88	8	0	1	0	1	0	dental plaques intestines of animals	AEOF01000010
ERR433462	Lactobacillus apodemi	DSM 16634	Ligilactobacillus apodemi	type strain	2082063	2019	38.6	3	52	3	163	9	0	1	0	1	0	the faeces of a wild mouse	BAMM01000051
ERR433476	Lactobacillus namurensis	DSM 19117	Levilactobacillus namurensis	type strain	2470988	2227	52	1	58	5	25	7	1	0	0	1	0	wheat sourdough vegetable fermentations	AZDT01000040
ERR433477	Lactobacillus nantensis	DSM 16982	Companilactobacillus nantensis	type strain	2923132	2774	36.2	2	55	1	83	14	0	0	0	1	0	type I sourdough	AZFV01000069
ERR433478	Lactobacillus odoratitofui	DSM 19909	Secundilactobacillus odoratitofui	type strain	2747284	2403	44.2	1	61	4	79	8	1	0	0	1	0	fermented brine used for stinky tofu production in Taipei County, Taiwan	AZEE01000005
ERR433479	Lactobacillus ozensis	DSM 23829	Apilactobacillus ozensis	type strain	1476372	1439	31.9	3	56	2	42	0	1	0	0	0	1	chrysanthemum flower	AYYQ01000014
ERR433491	Lactobacillus rennini	DSM 20253	Loigolactobacillus rennini	type strain	2261248	2219	40.7	0	54	5	130	10	1	0	0	1	0	rennet and are associated with cheese spoilage	AYYI01000077
ERR433493	Lactobacillus sakei subsp. carnosus	DSM 15831	Latilactobacillus sakei ssp. carnosus	type strain	1975630	1984	41	0	49	2	137	9	1	0				fermented meat products vacuum-packaged meat sauerkraut other fermented plant material	AZFG01000015
ERR433494	Lactobacillus saniviri	DSM 24301	Lacticaseibacillus saniviri	type strain	2429351	2409	47.7	1	56	1	184	14	1	0	0	1	0	the faeces of a healthy man fermented rice fermented fish	JQCE01000025
ERR433495	Lactobacillus satsumensis	DSM 16230	Liquorilactobacillus satsumensis	type strain	2634920	2441	39.9	1	48	0	88	5	1	1	0	1	0	mashes of shochu a traditional Japanese distilled spirit made from fermented riceother starchy materials	AZFQ01000022

ERR433499	Lactobacillus ruminis	ATCC 27780	Ligilactobacillus ruminis	type strain	2025861	1903	43.4	1	58	2	133	12	0	0	0	0	1	rumen of cow and from sewage horses and pigs and bovine uterus the gut of humans	BCVU01000117
ERR438946	Lactobacillus aviarius subsp. araffinosus	DSM 20653	Ligilactobacillus araffinosus	type strain	1470053	1410	38.1	2	46	0	66	4	0	0	0	1	0	the intestine and faeces of birds	AYYZ01000003
ERR485115	Lactobacillus sucicola	DSM 21376	Liquorilactobacillus sucicola	type strain	2456798	2265	38.5	0	55	2	51	8	1	1	0	1	0	the sap of an oak (Quercus sp)	AB458681
GCA_000010005.1	Lactobacillus reuteri	JCM 1112	Limosilactobacillus reuteri	type strain	2039414	2020	38.9	18	65	0	56	9	1	0	0	0	1	the intestinal microbiota of rodents, birds, swine, and in other intestinal ecosystems cereal fermentations particularly type II sour doughs Food isolates are of intestinal origin	AP007281
GCA_000014425.1	Lactobacillus gasseri	ATCC 33323	Lactobacillus gasseri	type strain	1894360	1808	35.3	18	78	0	55	7	0	1	0	0	1	human female lower genital tract the human mouth intes tinal tract the intestine of animals wounds, urine, blood, carious dentine and pus of patients suffering from septic infections	CP000413
GCA_000014525.1	Lactobacillus paracasei	ATCC 334	Lacticaseibacillus paracasei		2924325	2835	46.6	15	60	1	180	4	1	0				a variety of courses including the human oral cavity fermented cereals vegetables meats dairy products invertebrate hosts	KC429784
GCA_000056065.1	Lactobacillus delbrueckii subsp. bulgaricus	ATCC 11842	Lactobacillus delbrueckii ssp. bulgaricus	type strain	1864998	1900	49.7	27	95	1	122	1	0	1				yoghurt cheese intestinal microbiota of suckling piglets	CR954253
GCA_000159395.1	Lactobacillus salivarius	ATCC 11741	Ligilactobacillus salivarius	type strain	2017251	1929	32.6	3	37	1	89	9	0	1	0	1	0	the mouth and intestinal tract of humans cats hamsters chickens dairy products swine	CP024067
GCA_000160855.1	Lactobacillus helveticus	DSM 20075	Lactobacillus helveticus	type strain	2020582	1944	36.8	3	37	1	177	2	0	1	0	1	0	chicken sour milk cheese starter cultures and cheese particularly Emmental and Gruye?re cheeses tomato pomace silage	ACLM01000202
GCA_000160875.1	Lactobacillus iners	DSM 13335	Lactobacillus iners	type strain	1277649	1191	32.5	3	45	0	72	0	0	0	0	1	0	the human female lower genital tract human skin	ACLN01000018
GCA_000192165.1	Lactobacillus delbrueckii subsp. lactis	DSM 20072	Lactobacillus delbrueckii ssp. lactis	type strain	2071079	1864	49.8	3	72	1	180	0	0	1	0	1	0	milk cheese compressed yeasts grain mash	AEXU01000148
GCA_000255495.2	Lactobacillus vini	DSM 20605	Liquorilactobacillus vini	type strain	2195706	2106	37.6	3	44	3	58	10	0	1	0	1	0	fermenting Spanish grape must bioethanol industrial processes in different distilleries of Brazil	AYYX01000149
GCA_000387565.1	Lactobacillus delbrueckii subsp. jakobsenii	ZN7a-9	Lactobacillus delbrueckii ssp. jakobsenii	type strain	1730812	1677	50.2	3	45	2	100	2	0	1				dolo wort used in the production of the fermented African beverge dolo in Burkina Faso	ALPY01000052

GCA_000423245.1	Lactobacillus ceti	DSM 22408	Ligilactobacillus ceti	type strain	1385752	1269	33.7	5	37	0	118	1	1	0	0	1	0	the lungs of a beaked whale	JQBZ01000004
GCA_000423265.1	Lactobacillus saerimneri	DSM 16049	Ligilactobacillus saerimneri	type strain	1720753	1726	42.5	4	35	0	132	3	0	1	0	1	0	pig faeces the intestines of pigs the human gut and vagina the cecum of chicken	AY255802
GCA_000428925.1	Lactobacillus rossiae	DSM 15814	Furfurilactobacillus rossiae	type strain	2862294	2700	43.3	5	58	1	89	4	1	0				wheat sourdough related cereal fermentations beer fruit fecal samples of children and swine it was used experimentally as starter culture for cactus pear fermentation [243]	AKZK01000036
GCA_000469325.1	Lactobacillus shenzhenensis	LY-73	Schleiferilactobacillus shenzhenensis	type strain	3271684	2975	56.4	2	43	5	309	16	0	1	0	1	0	fermented dairy beverage	JX523627
GCA_000615805.1	Lactobacillus fuchuensis	JCM 11249	Latilactobacillus fuchuensis	type strain	2107444	2205	41.8	3	34	1	78	11	1	0				vacuum-packaged refrigerated beef common carp intestine other seafood products	BAMJ01000063
GCA_000740055.1	Lactobacillus oryzae	SG293	Secundilactobacillus oryzae	type strain	1860394	1859	42.8	6	40	1	96	3	1	1	0	1	0	fermented rice grains in Tochigi, Japan	BBAZ01000072
GCA_000785105.1	Lactobacillus curieae	CCTCC M 2011381	Lentilactobacillus curieae	type strain	2185962	2112	39.6	6	56	1	73							stinky tofu brine cocoa bean fermentations cheese curd powder	CP018906
GCA_000786395.1	Lactobacillus acidophilus	ATCC 4356	Lactobacillus acidophilus	type strain	1956698	1884	34.6	4	55	1	61	9	0	1	1	0	0	intestinal tract of humans and animals human mouth human vagina sourdough wine	CBLQ010000054
GCA_000807975.1	Lactobacillus brevis	BSO 464	Levilactobacillus brevis		2723202	2700	45.4	18	48	1	149	6	1	0	0	1	0	milk cheese sauerkraut and rrelated vegetable fermentations sourdough silage cow manure faeces the mouth and intestinal tract of humans and rats	GCA_000807975.1_00077
GCA_000829035.1	Lactobacillus paracasei subsp. paracasei	JCM 8130	Lacticaseibacillus paracasei ssp. paracasei	type strain	3017804	2945	46.6	15	62	0	226	13	1	0				dairy products sewage silage humans and clinical sources	ACGY01000162
GCA_000829055.1	Lactobacillus casei	ATCC 393	Lacticaseibacillus casei	type strain	2952961	2890	47.9	15	59	0	269	14	1	0				chinese traditional pickle infant faeces corn liquor oat silage commercial dietary supplements sputum nasopharynx	AP012544
GCA_000829395.1	Lactobacillus hokkaidonensis	LOOC260	Paucilactobacillus hokkaidonensis	type strain	2400586	2328	38.2	12	56	1	36	4	1	0	0	1	0	grass silage	AP014680
GCA_000831645.3	Lactobacillus heilongjiangensis	DSM 28069	Companilactobacillus heilongjiangensis	type strain	2790548	2485	37.5	12	55	1	98							fermented vegetables type I sourdough	CP012559
GCA_000876205.1	Lactobacillus wasatchensis	WDC04	Paucilactobacillus wasatchensis	type strain	1904253	1807	39.8	3	51	4	22							spoiled cheddar cheese silage	AWTT01000084

GCA_000967245.1	Lactobacillus mellis	Hon2	Bombilactobacillus mellis	type strain	1810599	1650	36.2	3	53	0	55							the honey stomach of the honeybee Apis mellifera	KQ033880
GCA_000970735.1	Lactobacillus apis	Hma11	Lactobacillus apis		1717379	1564	36.6	3	50	1	19							stomach contents of honeybees	KF386017
GCA_000970755.1	Lactobacillus kimbladii	Hma2	Lactobacillus kimbladii	type strain	2186983	1972	35.8	3	50	2	100							the honey stomach of the honeybee A. mellifera	JX099549
GCA_000970775.1	Lactobacillus melliventris	Hma8	Lactobacillus melliventris	type strain	2116151	1994	35.8	3	51	1	106							the homey stomach of honeybees	JX099551
GCA_000970795.1	Lactobacillus mellifer	Bin4	Bombilactobacillus mellifer	type strain	1815047	1661	39.3	3	50	0	139							the honey stomach of the honeybee Apis mellifera	JX099543
GCA_000970855.1	Lactobacillus helsingborgensis	Bma5	Lactobacillus helsingborgensis	type strain	2020254	1823	36.3	3	51	2	101							the honey stomach of the honeybee A. mellifera mellifera alfalfa silage	JX099553
GCA_001039045.1	Lactobacillus herbarum	TCF032-E4	Lactiplantibacillus herbarum	type strain	2899876	2805	43.5	4	36	0	184							fermented radish	LFEE01000051
GCA_001050475.1	Lactobacillus ginsenosidimutans	EMML 3141	Companilactobacillus ginsenosidimutans	type strain	2590556	2558	36.7	9	55	0	101	8	1	0	0	1	0	kimchi	CP012034
GCA_001189855.1	Lactobacillus delbrueckii subsp. indicus	JCM 15610	Lactobacillus delbrueckii ssp. indicus	type strain	1877412	1832	49.5	7	64	1	158	2	0	1				a fermented dairy product dahi from India	LGAS01000062
GCA_001190005.1	Lactobacillus delbrueckii subsp. sunkii	JCM 17838	Lactobacillus delbrueckii ssp. sunkii	type strain	1945263	1823	50.1	9	74	2	128	11	0	0	0	1	0	a traditionally fermented Japanese red turnip	LGHR01000024
GCA_001263315.1	Lactobacillus delbrueckii subsp. delbrueckii	KACC 13439	Lactobacillus delbrueckii ssp. delbrueckii	type strain	1766190	1769	50	1	50	0	48	2	0	1				vegetable source sour grain mash fermented grains	CP018615
GCA_001281265.1	Lactobacillus kunkeei	YH-15	Apilactobacillus kunkeei	type strain	1515712	1353	36.4	3	62	0	54	2	1	0	0	1	0	a sluggish grape wine fermentation honey bees and flowers	JXDB01000004
GCA_001293735.1	Lactobacillus gorillae	KZ01	Limosilactobacillus gorillae	type strain	1641621	1568	48.1	3	53	1	97							the faeces of a captive gorillas wild western lowland gorillas	AB904716
GCA_001311115.1	Lactobacillus lindneri	JCM 11027	Fructilactobacillus lindneri	type strain	1436854	1632	34.1	3	55	2	164	1	1	0	1	0	0	spoiled beer wine	BBAF01000027
GCA_001313225.1	Lactobacillus silagei	JCM 19001	Secundilactobacillus silagei	type strain	2650200	3600	44.8	3	61	4	243							silage	AB786910
GCA_001433745.1	Lactobacillus zeae	DSM 20178	Lacticaseibacillus casei	type strain	3121340	2961	47.7	5	53	3	54	14	1	0				chinese traditional pickle infant faeces corn liquor oat silage commercial dietary supplements sputum nasopharynx	D86516
GCA_001433765.1	Lactobacillus coryniformis subsp. coryniformis	DSM 20001	Loigolactobacillus coryniformis ssp. coryniformis	type strain	2705076	2579	42.9	3	38	1	136	3	1	0				silage cow dung dairy barn air and sewage table olives wheat pickled vegetable cheese and ting a fermented sorghum porridge	GL544638
GCA_001433855.1	Lactobacillus brevis	DSM 20054	Levilactobacillus brevis	type strain	2474438	2423	46	4	42	0	84	6	1	0	0	1	0	milk cheese sauerkraut and rrelated vegetable fermentations sourdough silage	KI271266

																		cow manure faeces the mouth and intestinal tract of humans and rats	
GCA_001433985.1	Lactobacillus amylovorus	DSM 20531	Lactobacillus amylovorus	type strain	2017377	2045	37.8	4	36	0	235	9	0	1	0	1	0	swine intestinal sourdough cattle waste-corn fermentation	AZCM01000082
GCA_001434005.1	Lactobacillus crispatus	DSM 20584	Lactobacillus crispatus	type strain	2057071	2017	36.6	3	43	1	79	9	1	1	0	1	0	human faeces vagina and buccal cavities crops and caeca of chicken patients with purulent pleurisy leucorrhea and urinary tract infections type II sourdoughs	AZCW01000112
GCA_001434145.1	Lactobacillus otakiensis	DSM 19908	Lentilactobacillus otakiensis	type strain	2346188	2255	42.4	3	28	2	59	7	1	0	0	1	0	sunki, a fermented turnip product kefir	BASH01000017
GCA_001434365.1	Lactobacillus gastricus	DSM 16045	Limosilactobacillus gastricus	type strain	1848461	1819	41.6	3	37	1	81	12	0	0	0	1	0	biopsy of a human stomach human　milk	AZFN01000048
GCA_001434465.1	Lactobacillus oris	DSM 4864	Limosilactobacillus oris	type strain	2031774	1925	50	2	51	2	150	10	0	0	0	1	0	the human saliva other human body sites including the vagina and　mother milk foods such as corn dough and bran	AZGE01000048
GCA_001434475.1	Lactobacillus suebicus	DSM 5007	Paucilactobacillus suebicus	type strain	2651315	2495	39	3	56	2	40	5	1	0				fermented cherry mashes cider silage	AM113785
GCA_001434695.1	Lactobacillus algidus	DSM 15638	Dellaglioa algida	type strain	1590323	1531	36	3	33	0	35	7	1	0	0	1	0	refrigerated beef and pork meat	AZDI01000021
GCA_001434775.1	Lactobacillus farciminis	DSM 20184	Companilactobacillus farciminis	type strain	2480845	2417	36.4	3	40	2	70	10	1	0	1	0	0	meat products sourdough fermentend fish cold-smoked salmon soy sauce mash dairy products table　olives fermented vegetables corn silage	AEOT01000034
GCA_001434815.1	Lactobacillus equicursoris	DSM 19284	Lactobacillus equicursoris	type strain	2052598	1873	47.7	3	28	2	143	6	0	1	0	0	1	a thoroughbred racehorse	BBBW01000097
GCA_001435555.1	Lactobacillus nodensis	DSM 19682	Companilactobacillus nodensis	type strain	2683197	2654	37.6	3	55	2	108	4	1	0	0	0	1	fermented rice bran paste it has been used experimentally as adjunct culture in cheese	BAMN01000046
GCA_001435655.1	Lactobacillus paraplantarum	DSM 10667	Lactiplantibacillus paraplantarum	type strain	3395753	3192	43.7	0	26	0	231	15	1	0	0	1	0	beer human faeces grape marmalade dairy products jangajji, Korean fermented food fermented vegetables fermented fruits fermented dates rice bran pickles silage cocoa beans fermented sourdough fermented slurry faecal microbita of healthy dogs traditional fura processing wine sow milk	AJ306297

GCA_001435735.1	Lactobacillus equi	DSM 15833	Ligilactobacillus equi	type strain	2284210	2188	39	2	50	4	211	6	0	1	0	1	0	faeces of horses	BAMI01000114
GCA_001435755.1	Lactobacillus acidipiscis	DSM 15836	Ligilactobacillus acidipiscis	type strain	2326083	2230	39.1	2	32	1	79	4	0	0	1	0	0	fermented fish (pla-ra and pla-chom) in Thai land but also found in dairy products soy sauce mash table olives sake starter tropical grasses forage crops bee pollen Chinese DaQu a saccharification starter for production of vinegar liquor from cereals	AB326356
GCA_001435875.1	Lactobacillus farraginis	DSM 18382	Lentilactobacillus farraginis	type strain	2859511	2749	42	3	40	2	74	8	1	1	0	1	0	a compost of distilled shochu residue	BAKI01000097
GCA_001435895.1	Lactobacillus parafarraginis	DSM 18390	Lentilactobacillus parafarraginis	type strain	3081674	2921	45.2	3	52	5	78	9	0	0	0	1	0	compost of distilled shochu residue silage fermented vegetables kefir grains	AZFZ01000113
GCA_001435975.1	Lactobacillus collinoides	DSM 20515	Secundilactobacillus collinoides	type strain	3616190	3224	46.1	3	50	3	241	9	1	0	0	1	0	compost apple cider table olives dairy products fermented durian fruit wines	BBEQ01000098
GCA_001436115.1	Lactobacillus brantae	DSM 23927	Lacticaseibacillus brantae	type strain	1929842	1900	47.5	3	27	2	46	4	0	0				the faeces of wild Canada goose (Branta canadensis) experimental sourdoughs	AYZQ01000010
GCA_001436555.1	Lactobacillus senioris	DSM 24302	Lentilactobacillus senioris	type strain	1566789	1568	39.1	3	44	0	46	4	1	0	0	1	0	the faeces of a 100-year-old female	LC519995
GCA_001436675.1	Lactobacillus senmaizukei	DSM 21775	Levilactobacillus senmaizukei	type strain	2222963	2122	48.6	2	61	1	107	5	1	0	0	1	0	senmaizuke a fermented turnip product	AB682140
GCA_001437055.1	Lactobacillus secaliphilus	DSM 17896	Limosilactobacillus secaliphilus	type strain	1646143	1503	47.7	2	39	0	99	1	0	0	0	1	0	type II sourdough	LC480808
GCA_001437125.1	Lactobacillus paucivorans	DSM 22467	Levilactobacillus paucivorans	type strain	2362603	2210	49.1	3	45	2	149	2	1	0	0	0	1	storage tank of a brewery	JQCA01000055
GCA_001438805.1	Lactobacillus kimchiensis	DSM 24716	Companilactobacillus kimchiensis	type strain	2698724	2579	35.5	2	38	2	76	10	1	0				kimchi	JQCF01000055
GCA_001438825.1	Lactobacillus crustorum	LMG 23699	Companilactobacillus crustorum	type strain	2235695	2165	35	5	48	1	74	4	1	0	0	1	0	sourdough dairy products forages	JQCK01000058
GCA_001438845.1	Lactobacillus xiangfangensis	LMG 26013	Lactiplantibacillus xiangfangensis	type strain	2989578	2757	45.1	4	50	0	162	12	0	0	0	1	0	pickle sourdough	JQCL01000078
SRR1151124	Lactobacillus bifermentans	DSM 20003	Loigolactobacillus bifermentans	type strain	3134903	3049	44.3	3	60	3	225	6	0	0	0	1	0	spoiled Edam Gouda cheeses fermented masau fruits Himalayan fermented milk products	M58809
SRR1151125	Lactobacillus curvatus	DSM 20019	Latilactobacillus curvatus	type strain	1807340	1814	42	0	37	2	111	3	1	0	0	1	0	cow dung fermented and vacuum-packaged refrigerated meat fermented and vacuum-packaged refrigerated fish dairy products such as milk and cheese fermented plant products like	BBBQ01000060

																		sauerkraut sourdough (including prepacked finished dough and pressed yeast) radish pickles kimchi other plant derived materials like honey the environmental fermentation process of corn or grass silage	
SRR1151129	Lactobacillus paracasei subsp. tolerans	DSM 20258	Lacticaseibacillus paracasei ssp. tolerans	type strain	2413718	2419	46.4	1	46	0	185	3	1	0				dairy products tomato pomace silage	D16550
SRR1151132	Lactobacillus concavus	DSM 17758	Lapidilactobacillus concavus	type strain	1903092	1765	43.3	1	49	1	77	5	0	0	0	1	0	the walls of a distilled-spirit-fermenting cellar in China	AZFX01000066
SRR1151133	Lactobacillus coryniformis subsp. torquens	DSM 20004	Loigolactobacillus coryniformis ssp. torquens	type strain	2657964	2541	43	2	38	0	132	3	1	0				cheese yaks' milk cheese silage tomato pomace silage	AEOS01000123
SRR1151134	Lactobacillus paracollinoides	DSM 15502	Secundilactobacillus paracollinoides	type strain	3470681	3214	46.9	2	67	2	288	4	1	0	0	1	0	beer cider fermented olives	AZFD01000165
SRR1151138	Lactobacillus equigenerosi	DSM 18793	Limosilactobacillus equigenerosi	type strain	1599169	1545	42.7	1	60	0	82	4	0	1	0	1	0	the intestinal tract of a thoroughbred horse	AZGC01000017
SRR1151139	Lactobacillus johnsonii	ATCC 33200	Lactobacillus johnsonii	type strain	1770443	1767	34.4	3	52	0	41	8	1	1	0	0	1	humans (gut, vagina) the faeces of birds rodents calves and pigs type II sourdoughs	ACGR01000047
SRR1151144	Lactobacillus versmoldensis	DSM 14857	Companilactobacillus versmoldensis	type strain	2386851	2344	38.3	3	55	1	104	5	1	0				poultry salami	BACR01000055
SRR1151148	Lactobacillus antri	DSM 16041	Limosilactobacillus antri	type strain	2249658	2113	51.1	2	61	2	239	7	0	1	0	0	1	biopsy of a healthy human gastric mucosa the intestine of other vertebrate animals	ACLL01000037
SRR1151152	Lactobacillus coleohominis	DSM 14060	Limosilactobacillus coleohominis	type strain	1865893	1979	41.1	3	62	0	100	1	0	1	0	1	0	the human vagina in human intestinal microbiota swine	AZEW01000320
SRR1151155	Lactobacillus gigeriorum	DSM 23908	Lactobacillus gigeriorum	type strain	1906781	1870	37	2	56	0	63	8	0	1	0	1	0	a crop of a chicken	CAKC01000053
SRR1151158	Lactobacillus hominis	DSM 23910	Lactobacillus hominis	type strain	1930068	1882	35.2	2	56	1	66	10	0	1	0	1	0	the human intestine	CAKE01000027
SRR1151162	Lactobacillus jensenii	DSM 20557	Lactobacillus jensenii	type strain	1615929	1478	34.3	3	39	1	47	10	0	1	0	1	0	the human female lower genital tract.	AYYU01000057
SRR1151163	Lactobacillus kalixensis	DSM 16043	Lactobacillus kalixensis	type strain	2073352	1935	36.1	3	62	1	81	12	0	1				a biopsy of the healthy human gastric mucosa	AZFM01000074
SRR1151164	Lactobacillus mucosae	DSM 13345	Limosilactobacillus mucosae	type strain	2280266	2044	46.4	0	76	2	163	5	0	1	0	0	1	the intestine of a pig the intestine of other vertebrates including humans type II sourdough related cereal fermentations	AF126738
SRR1151168	Lactobacillus parabuchneri	DSM 5707	Lentilactobacillus parabuchneri	type strain	2568303	2377	43.4	1	58	4	81	10	1	0	0	1	0	dairy products saliva silage spoiled beer some strains were shown to persist over month in whiskey mashes in Scottish	BCVT01000078

																		distilleries	
SRR1151169	Lactobacillus pasteurii	DSM 23907	Lactobacillus pasteurii	type strain	1753652	1684	38.5	1	54	1	66	10	0	1	0	1	0	the human intestine	CAKD01000001
SRR1151174	Lactobacillus ultunensis	DSM 16047	Lactobacillus ultunensis	type strain	2169096	2115	36	3	60	0	90	9	0	0	0	1	0	a biopsy of a healthy human gastric mucosa	ACGU01000081
SRR1151175	Lactobacillus vaginalis	DSM 5837	Limosilactobacillus vaginalis	type strain	1781526	1733	40.5	0	58	0	67	7	0	1	0	1	0	microbiota of the human vagina	AF243177
SRR1151187	Lactobacillus oligofermentans	DSM 15707	Paucilactobacillus oligofermentans	type strain	1789353	1722	35.5	2	52	1	17	3	1	0	0	0	1	marinated poultry meat at the end of its shelf life fermented olives	AZFE01000013
SRR1151190	Lactobacillus fructivorans	DSM 20203	Fructilactobacillus fructivorans	type strain	1372674	1336	38.9	1	58	0	218	0	1	0	1	0	0	the intestinal microbiota of fruit flies spoiled sake mashes spoiled mayonnaise salad dressings sour dough dessert wines aperitifs	AEQY01000004
SRR1151193	Lactobacillus plantarum subsp. plantarum	CGMCC 1.2437	Lactiplantibacillus plantarum ssp. plantarum	type strain	3220167	3019	44.5	4	63	0	208	17	1	0	0	1	0	dairy products and dairy environments silage sauerkraut pickled vegetables sourdough cow dung the human mouth intestinal tract and stools sewage	ACGZ01000098
SRR1151196	Lactobacillus buchneri	DSM 20057	Lentilactobacillus buchneri	type strain	2451635	2345	44.4	3	61	1	68	6	1	0	0	1	0	pressed yeast milk cheese fermenting plant material the human mouth used commercially as silage inoculant	AB205055
SRR1151197	Lactobacillus cacaonum	DSM 21116	Liquorilactobacillus cacaonum	type strain	1917961	1823	33.9	1	50	1	26	4	1	0	0	1	0	cocoa fermentation	AYZE01000006
SRR1151200	Lactobacillus composti	DSM 18527	Agrilactobacillus composti	type strain	3463695	3306	44	2	51	3	80	11	1	0	0	1	0	compost from shochu mash solids pulque, a Mexican alcoholic beverage	AZGA01000039
SRR1151201	Lactobacillus dextrinicus	DSM 20335	Lapidilactobacillus dextrinicus	type strain	1807580	1725	38	3	49	2	86							silage fermenting vegetables beer sliced vacuum-packed cooked sausage	AYYK01000012
SRR1151205	Lactobacillus florum	DSM 22689	Fructilactobacillus florum	type strain	1354760	1313	41.1	3	47	1	153	0	1	0	0	1	0	peony bietou flowers grapes wine	AYZI01000021
SRR1151207	Lactobacillus ghanensis	DSM 18630	Liquorilactobacillus ghanensis	type strain	2602751	2416	37.1	2	54	1	71	9	0	1	0	1	0	cocoa fermentations	AZGB01000014
SRR1151211	Lactobacillus kefiranofaciens subsp. kefiranofaciens	DSM 5016	Lactobacillus kefiranofaciens ssp. kefiranofaciens	type strain	2258515	2316	37.3	3	58	2	150	7	0	0	0	1	0	kefir grains fermented dairy products	BAMG01000091
SRR1151212	Lactobacillus kefiranofaciens subsp. kefirgranum	DSM 10550	Lactobacillus kefiranofaciens ssp. kefirgranum	type strain	2084861	2099	37.5	2	58	4	136	5	0	0	0	1	0	kefir grains	AZEM01000027
SRR1151214	Lactobacillus kimchicus	JCM 15530	Secundilactobacillus kimchicus	type strain	2593829	2511	46.6	2	60	4	76	9	1	1	0	1	0	kimchi	EU678893

SRR1151216	Lactobacillus kisonensis	DSM 19906	Lentilactobacillus kisonensis	type strain	3017560	2765	41.8	0	58	4	55	6	1	0	0	1	0	pickle brine	BBAU01000086
SRR1151217	Lactobacillus koreensis	JCM 16448	Levilactobacillus koreensis	type strain	2940897	2666	49.2	1	55	2	171	5	1	0	0	1	0	cabbage kimchi sourdough	FJ904277
SRR1151218	Lactobacillus mali	DSM 20444	Liquorilactobacillus mali	type strain	2611318	2559	36.1	2	40	1	94	6	1	0	0	1	0	wine must fermenting cider fermented　molasses water kefirs cocoa bean fermentations table　olives	BACP01000083
SRR1151220	Lactobacillus nasuensis	JCM 17158	Lacticaseibacillus nasuensis	type strain	2278732	2137	57	6	64	0	160	2	0	0	0	1	0	Sudan grass [Sorghum sudanense (Piper)　Stapf] silage	AZDJ01000014
SRR1151225	Lactobacillus parabrevis	ATCC 53295	Levilactobacillus parabrevis	type strain	2625389	2379	49	0	51	1	142	5	1	0	0	1	0	farmhouse red Cheshire cheese wheat sourdough fermented vegetables a municipal biogas plant	JQCI01000059
SRR1151227	Lactobacillus perolens	DSM 12744	Schleiferilactobacillus perolens	type strain	3269427	3106	49.2	1	57	2	142	8	0	0				spoiled soft drinks brewery environments	Y19167
SRR1151229	Lactobacillus pobuzihii	KCTC 13174	Ligilactobacillus pobuzihii	type strain	2332525	2124	37.7	0	59	0	48	6	0	0	0	1	0	pobuzihi fermented cummincordia fermented　fish traditional vinegar	AZCL01000058
SRR1151230	Lactobacillus rapi	DSM 19907	Lentilactobacillus rapi	type strain	2848015	2645	42.9	0	57	3	40	10	1	0	0	1	0	sunki other vegetable fermentations	AZEI01000033
SRR1151235	Lactobacillus sunkii	DSM 19904	Lentilactobacillus sunkii	type strain	2693190	2545	42.1	1	58	0	71	7	1	0	0	1	0	sunki, a fermented turnip product kefir	AZEA01000056
SRR1151237	Lactobacillus thailandensis	DSM 22698	Lacticaseibacillus thailandensis	type strain	2064913	1893	53.5	1	52	1	154	4	0	0				fermented fish (pla-ra) in Thailand	AYZK01000017
SRR1151242	Lactobacillus pentosus	DSM 20314	Lactiplantibacillus pentosus	type strain	3642579	3285	46.3	4	68	5	188	17	1	0	0	1	0	diverse sources including corn silage fermenting　olives sewage fermented mulberry leaf powders fermented　teas glutinous rice dough corn noodles chili sauce mustard　pickles stinky tofu dairy products mustard pickle fermented　idli batter tempoyak human vagina human stools sourdoughs	AZCU01000047
SRR1151250	Lactobacillus panis	DSM 6035	Limosilactobacillus panis	type strain	1986287	1887	48.1	1	59	2	134	12	0	1	0	0	1	type II sourdough fermenting plant material the intestine of birds	X94230
SRR1151251	Lactobacillus paralimentarius	DSM 13238	Companilactobacillus paralimentarius	type strain	2533817	2454	35.1	2	53	3	70	8	1	0	0	1	0	sourdough other cereal fermentations poultry meat	BAMH01000179
SRR1151252	Lactobacillus pontis	DSM 8475	Limosilactobacillus pontis	type strain	1656883	1614	53.5	1	65	0	55	4	1	1	0	1	0	type I and type II sourdough the intestinal microbiota of swine silage dairy products mezcal fermentation	AJ422032

																		wet wheat distillers' grain	
SRR1151254	Lactobacillus sanfranciscensis	DSM 20451	Fructilactobacillus sanfranciscensis	type strain	1253219	1278	34.7	4	57	1	103	1	1	0	0	1	0	traditional sourdoughs agave mash	X76327
SRR1151256	Lactobacillus zymae	DSM 19395	Levilactobacillus zymae	type strain	2700869	2444	53.6	3	63	5	38	4	1	0	0	1	0	type I wheat sourdough forages fermented onions	AZDW01000036
SRR1151257	Lactobacillus amylolyticus	DSM 11664	Lactobacillus amylolyticus	type strain	1539298	1574	38.3	3	55	0	84	4	0	1	1	0	0	malt mash and unhopped wort in breweries sourdough tofu whey	ADNY01000006
SRR1151258	Lactobacillus amylophilus	DSM 20533	Amylolactobacillus amylophilus	type strain	1546306	1550	43.7	0	52	1	126	3	1	0	0	1	0	swine waste-corn fermentation corn-starch processing industrial wastes kocho (Ensete ventricosum) bread	BBBR01000026
SRR1151260	Lactobacillus hilgardii	DSM 20176	Lentilactobacillus hilgardii	type strain	2605214	2538	39.6	0	59	1	48	4	1	0	0	1	0	spoiled wine kefir grains mezcal fermentations silage	ACGP01000200
SRR1151262	Lactobacillus malefermentans	DSM 5705	Secundilactobacillus malefermentans	type strain	2054106	2013	41	3	61	4	70	2	1	0	0	1	0	beer	BACN01000105
SRR1151264	Lactobacillus oeni	DSM 19972	Liquorilactobacillus oeni	type strain	2105430	1976	37.3	1	52	2	72	4	1	1	1	0	0	Bobal wine	AZEH01000040
SRR1151267	Lactobacillus sakei subsp. sakei	DSM 20017	Latilactobacillus sakei ssp. sakei	type strain	1907928	1891	41.1	0	51	0	87	7	1	0				sake starter fermented meat products vacuum packaged meat sauerkraut other fermented plant material human faeces	BALW01000030
SRR1151270	Lactobacillus rhamnosus	DSM 20021	Lacticaseibacillus rhamnosus	type strain	2945929	2738	46.7	3	56	0	138	15	1	1				a broad range of habitats including dairy products fermented meat fish vegetables and cereals sewage humans (oral, vaginal and intestinal) invertebrate hosts and clinical sources	BALT01000058
SRR1745849	Lactobacillus kullabergensis	Biut2	Lactobacillus kullabergensis	type strain	2080753	1939	35.5	3	53	1	92							the honey stomach of the honeybee A. mellifera mellifera	JX099550
SRR1752129	Lactobacillus apinorum	Fhon13	Apilactobacillus apinorum	type strain	1428890	1317	34.6	1	59	2	43							honey stomach of the honeybee	JX099541
SRR896433	Lactobacillus harbinensis	DSM 16991	Schleiferilactobacillus harbinensis	type strain	3123257	3031	53.1	1	62	2	228	15	0	0				fermented vegetables 'Suan Cai' the brewery environment fermented cereals tomato pomace spoiled soft drinks	AZFW01000057

Supplementary Table 3.4: Gain/loss expected number and other stats for each strains.

id	protein_id	expected value of gain events (E_g)	expected value of loss events (E_l)	branch Length (L_b)	number of ortholog (O_n)	curated name	strain	rate of gain/loss events (E_g/E_l)	protein number minus delta of expected value of gain/loss events (N_p-(E_g - E_l))	expect value of gain events par branch length (E_g/L_b)	expect value of loss events par branch length (E_l/L_b)	genetic diversity in the bacteria before speciation (Gd)	normalized amount of expected value of gain/loss events for each branches (Egl)	normalized net number of expected value of gain events (Ng)
SRR1151200	SRR1151200.protein.faa	1677	1361	0.1285	2710	Agrilactobacillus composti	DSM 18527	1.232182	2990	13050.58	10591.44	2394	23642.02	2459.144
SRR1151258	SRR1151258.protein.faa	184.5	232.3	0.01941	1457	Amylolactobacillus amylophilus	DSM 20533	0.794232	1597.8	9505.41	11968.06	1504.8	21473.47	-2462.65
ERR387486	ERR387486.protein.faa	75.58	83.36	0.006439	1497	Amylolactobacillus amylotrophicus	DSM 20534	0.90667	1609.78	11737.85	12946.11	1504.78	24683.96	-1208.26
SRR1752129	SRR1752129.protein.faa	112.7	196.2	0.0223	1247	Apilactobacillus apinorum	Fhon13	0.574414	1400.5	5053.812	8798.206	1330.5	13852.02	-3744.39
GCA_001281265.1	GCA_001281265.1.protein.faa	151.8	213.4	0.02954	1269	Apilactobacillus kunkeei	YH-15	0.71134	1414.6	5138.795	7224.103	1330.6	12362.9	-2085.31
ERR433479	ERR433479.protein.faa	440.6	702.3	0.06238	1348	Apilactobacillus ozensis	DSM 23829	0.627367	1700.7	7063.161	11258.42	1609.7	18321.58	-4195.25
GCA_000970795.1	GCA_000970795.1.protein.faa	1092	1390	0.128	1519	Bombilactobacillus mellifer	Bin4	0.785612	1959	8531.25	10859.38	1817	19390.63	-2328.13
GCA_000967245.1	GCA_000967245.1.protein.faa	769.6	1063	0.09406	1523	Bombilactobacillus mellis	Hon2	0.723989	1943.4	8182.011	11301.3	1816.4	19483.31	-3119.29
ERR387471	ERR387471.protein.faa	157.1	299.4	0.01503	1921	Companilactobacillus alimentarius	DSM 20249	0.524716	2374.3	10452.43	19920.16	2063.3	30372.59	-9467.73
GCA_001438825.1	GCA_001438825.1.protein.faa	151.4	381.4	0.01333	1869	Companilactobacillus crustorum	LMG 23699	0.396959	2395	11357.84	28612.15	2099	39969.99	-17254.3
GCA_001434775.1	GCA_001434775.1.protein.faa	197	275.4	0.01667	2095	Companilactobacillus farciminis	DSM 20184	0.715323	2495.4	11817.64	16520.7	2173.4	28338.33	-4703.06
ERR387495	ERR387495.protein.faa	259.4	354.8	0.02661	2129	Companilactobacillus futsaii	JCM 17355	0.731116	2544.4	9748.215	13333.33	2224.4	23081.55	-3585.12
GCA_001050475.1	GCA_001050475.1.protein.faa	310.8	245.6	0.01674	2215	Companilactobacillus ginsenosidimutans	EMML 3141	1.265472	2492.8	18566.31	14671.45	2149.8	33237.75	3894.863
GCA_000831645.3	GCA_000831645.3.protein.faa	199.8	206.1	0.006609	2113	Companilactobacillus heilongjiangensis	DSM 28069	0.969432	2491.3	30231.5	31184.75	2119.3	61416.25	-953.246
GCA_001438805.1	GCA_001438805.1.protein.faa	301	351.6	0.02768	2149	Companilactobacillus kimchiensis	DSM 24716	0.856086	2629.6	10874.28	12702.31	2199.6	23576.59	-1828.03
ERR387524	ERR387524.protein.faa	193.6	366.2	0.005058	1932	Companilactobacillus mindensis	DSM 14500	0.528673	2377.6	38276	72400.16	2104.6	110676.2	-34124.2
ERR433477	ERR433477.protein.faa	310.8	88.37	0.003524	2327	Companilactobacillus nantensis	DSM 16982	3.517031	2551.57	88195.23	25076.62	2104.57	113271.9	63118.62

GCA_001435555.1	GCA_001435555.1.protein.faa	620.1	465.2	0.04427	2293	Companilactobacillus nodensis	DSM 19682	1.332975	2499.1	14007.23	10508.24	2138.1	24515.47	3498.984
SRR1151251	SRR1151251.protein.faa	165.8	154.1	0.01079	2075	Companilactobacillus paralimentarius	DSM 13238	1.075925	2442.3	15366.08	14281.74	2063.3	29647.82	1084.337
ERR387549	ERR387549.protein.faa	225.5	477.6	0.02927	1886	Companilactobacillus tucceti	DSM 20183	0.472152	2354.1	7704.134	16317.05	2138.1	24021.18	-8612.91
SRR1151144	SRR1151144.protein.faa	235.5	306.3	0.01775	2079	Companilactobacillus versmoldensis	DSM 14857	0.768854	2414.8	13267.61	17256.34	2149.8	30523.94	-3988.73
GCA_001434695.1	GCA_001434695.1.protein.faa	863.9	1766	0.1156	1434	Dellaglioa algida	DSM 15638	0.489185	2433.1	7473.183	15276.82	2336.1	22750	-7803.63
SRR1151205	SRR1151205.protein.faa	455	676.8	0.06512	1243	Fructilactobacillus florum	DSM 22689	0.672281	1534.8	6987.101	10393.12	1464.8	17380.22	-3406.02
SRR1151190	SRR1151190.protein.faa	474.8	859	0.0726	1252	Fructilactobacillus fructivorans	DSM 20203	0.552736	1720.2	6539.945	11831.96	1636.2	18371.9	-5292.01
GCA_001311115.1	GCA_001311115.1.protein.faa	347.6	314.4	0.03489	1498	Fructilactobacillus lindneri	JCM 11027	1.105598	1598.8	9962.74	9011.178	1464.8	18973.92	951.5621
SRR1151254	SRR1151254.protein.faa	649.1	964.2	0.09556	1198	Fructilactobacillus sanfranciscensis	DSM 20451	0.673201	1593.1	6792.591	10090	1513.1	16882.59	-3297.4
GCA_000428925.1	GCA_000428925.1.protein.faa	413.4	213	0.01938	2315	Furfurilactobacillus rossiae	DSM 15814	1.940845	2499.6	21331.27	10990.71	2114.6	32321.98	10340.56
ERR387541	ERR387541.protein.faa	275.7	611.4	0.02818	1779	Furfurilactobacillus siliginis	DSM 22696	0.450932	2315.7	9783.534	21696.24	2114.7	31479.77	-11912.7
ERR387489	ERR387489.protein.faa	1247	2467	0.1572	1174	Holzapfelia floricola	DSM 23037	0.505472	2467	7932.57	15693.38	2394	23625.95	-7760.81
GCA_001436115.1	GCA_001436115.1.protein.faa	228.1	543.1	0.0294	1768	Lacticaseibacillus brantae	DSM 23927	0.419996	2215	7758.503	18472.79	2083	26231.29	-10714.3
ERR387469	ERR387469.protein.faa	663.7	858.9	0.06414	2103	Lacticaseibacillus camelliae	DSM 22697	0.772733	2598.2	10347.68	13391.02	2298.2	23738.7	-3043.34
GCA_000829055.1	GCA_000829055.1.protein.faa	227.8	331.7	0.003763	2364	Lacticaseibacillus casei	ATCC 393	0.686765	2993.9	60536.81	88147.75	2467.9	148684.6	-27610.9
GCA_001433745.1	GCA_001433745.1.protein.faa	342.3	210.8	5.35E-06	2598	Lacticaseibacillus casei	DSM 20178	1.623814	2829.5	63957399	39387145	2466.5	1.03E+08	24570254
ERR387460	ERR387460.protein.faa	785.9	527.3	0.04547	2554	Lacticaseibacillus manihotivorans	DSM 13343	1.490423	2753.4	17283.92	11596.66	2295.4	28880.58	5687.266
SRR1151220	SRR1151220.protein.faa	508.6	864.1	0.05458	1940	Lacticaseibacillus nasuensis	JCM 17158	0.588589	2492.5	9318.432	15831.81	2295.5	25150.24	-6513.37
ERR387506	ERR387506.protein.faa	248.6	124	0.004593	2009	Lacticaseibacillus pantheris	DSM 15945	2.004839	2168.4	54125.84	26997.61	1884.4	81123.45	27128.24
GCA_000014525.1	GCA_000014525.1.protein.faa	220.3	83.36	5.35E-06	2424	Lacticaseibacillus paracasei	ATCC 334	2.642754	2698.06	41162182	15575486	2287.06	56737668	25586697
GCA_000829035.1	GCA_000829035.1.protein.faa	295.2	276.7	5.35E-06	2485	Lacticaseibacillus paracasei ssp. paracasei	JCM 8130	1.066859	2926.5	55156951	51700299	2466.5	1.07E+08	3456652
SRR1151129	SRR1151129.protein.faa	203.6	314.7	5.35E-06	2176	Lacticaseibacillus paracasei ssp. tolerans	DSM 20258	0.646965	2530.1	38041854	58800448	2287.1	96842302	-2.1E+07
SRR1151270	SRR1151270.protein.faa	91.6	84.9	0.000995	2390	Lacticaseibacillus rhamnosus	DSM 20021	1.078916	2731.3	92051.05	85318.06	2383.3	177369.1	6732.992
ERR433494	ERR433494.protein.faa	539.9	477	0.04861	2146	Lacticaseibacillus saniviri	DSM 24301	1.131866	2346.1	11106.77	9812.796	2083.1	20919.56	1293.972
ERR387540	ERR387540.protein.faa	1389	1568	0.1321	2083	Lacticaseibacillus sharpeae	DSM 20505	0.885842	2523	10514.76	11869.8	2262	22384.56	-1355.03
SRR1151237	SRR1151237.protein.faa	109.2	286.6	0.003988	1707	Lacticaseibacillus thailandensis	DSM 22698	0.381019	2070.4	27382.15	71865.6	1884.4	99247.74	-44483.5

ERR387476	ERR387476.protein.faa	454.9	374.3	0.02255	2593	Lactiplantibacillus fabifermentans	DSM 21115	1.215335	3030.4	20172.95	16598.67	2512.4	36771.62	3574.279
GCA_001039045.1	GCA_001039045.1.protein.faa	296.1	432.2	0.0182	2396	Lactiplantibacillus herbarum	TCF032-E4	0.685099	2941.1	16269.23	23747.25	2532.1	40016.48	-7478.02
GCA_001435655.1	GCA_001435655.1.protein.faa	189.7	99.38	0.00078	2654	Lactiplantibacillus paraplantarum	DSM 10667	1.908835	3101.68	243361.1	127492	2563.68	370853.1	115869.1
SRR1151242	SRR1151242.protein.faa	362.2	83.04	5.35E-06	2733	Lactiplantibacillus pentosus	DSM 20314	4.361753	3005.84	67675635	15515695	2453.84	83191330	52159940
ERR387522	ERR387522.protein.faa	198	100.4	5.35E-06	2481	Lactiplantibacillus plantarum ssp. argentoratensis	DSM 16365	1.972112	2841.4	36995516	18759342	2383.4	55754858	18236173
SRR1151193	SRR1151193.protein.faa	198.6	72.03	5.35E-06	2510	Lactiplantibacillus plantarum ssp. plantarum	CGMCC 1.2437	2.757185	2892.43	37107623	13458520	2383.43	50566143	23649103
GCA_001438845.1	GCA_001438845.1.protein.faa	328.7	469.2	0.02909	2346	Lactiplantibacillus xiangfangensis	LMG 26013	0.700554	2897.5	11299.42	16129.25	2486.5	27428.67	-4829.84
ERR387527	ERR387527.protein.faa	608.4	990.9	0.08151	1402	Lactobacillus acetotolerans	DSM 20749	0.613987	1900.5	7464.115	12156.79	1784.5	19620.91	-4692.68
GCA_000786395.1	GCA_000786395.1.protein.faa	275.4	439.9	0.0358	1639	Lactobacillus acidophilus	ATCC 4356	0.626051	2048.5	7692.737	12287.71	1803.5	19980.45	-4594.97
SRR1151257	SRR1151257.protein.faa	338	522.4	0.04404	1439	Lactobacillus amylolyticus	DSM 11664	0.647014	1758.4	7674.841	11861.94	1623.4	19536.78	-4187.1
GCA_001433985.1	GCA_001433985.1.protein.faa	247.6	179.6	0.01244	1780	Lactobacillus amylovorus	DSM 20531	1.378619	1977	19903.54	14437.3	1712	34340.84	5466.238
GCA_000970735.1	GCA_000970735.1.protein.faa	267	475.6	0.04087	1414	Lactobacillus apis	Hma11	0.561396	1772.6	6532.909	11636.9	1622.6	18169.81	-5103.99
GCA_001434005.1	GCA_001434005.1.protein.faa	265.6	305.4	0.02398	1752	Lactobacillus crispatus	DSM 20584	0.869679	2056.8	11075.9	12735.61	1791.8	23811.51	-1659.72
GCA_000056065.1	GCA_000056065.1.protein.faa	149.3	165.4	0.01131	1659	Lactobacillus delbrueckii ssp. bulgaricus	ATCC 11842	0.90266	1916.1	13200.71	14624.23	1675.1	27824.93	-1423.52
GCA_001263315.1	GCA_001263315.1.protein.faa	159	234.6	0.01456	1596	Lactobacillus delbrueckii ssp. delbrueckii	KACC 13439	0.677749	1844.6	10920.33	16112.64	1671.6	27032.97	-5192.31
GCA_001189855.1	GCA_001189855.1.protein.faa	101.7	119.5	0.000871	1656	Lactobacillus delbrueckii ssp. indicus	JCM 15610	0.851046	1849.8	116802.6	137245.9	1673.8	254048.5	-20443.3
GCA_000387565.1	GCA_000387565.1.protein.faa	82.71	167.6	0.007749	1541	Lactobacillus delbrueckii ssp. jakobsenii	ZN7a-9	0.493496	1761.89	10673.64	21628.6	1625.89	32302.23	-10955
GCA_000192165.1	GCA_000192165.1.protein.faa	168	89.92	0.005134	1704	Lactobacillus delbrueckii ssp. lactis	DSM 20072	1.868327	1785.92	32723.02	17514.61	1625.92	50237.63	15208.41
GCA_001190005.1	GCA_001190005.1.protein.faa	116.8	136.8	0.004599	1646	Lactobacillus delbrueckii ssp. sunkii	JCM 17838	0.853801	1843	25396.83	29745.6	1666	55142.42	-4348.77
GCA_001434815.	GCA_001434815.1.protein.faa	698.9	767.1	0.08097	1714	Lactobacillus equicursoris	DSM 19284	0.911094	1941.2	8631.592	9473.879	1782.2	18105.47	-842.287

1														
ERR387508	ERR387508.protein.faa	117.3	168.7	0.007222	1660	Lactobacillus gallinarum	DSM 10532	0.695317	1963.4	16242.04	23359.18	1711.4	39601.22	-7117.14
GCA_000014425.1	GCA_000014425.1.protein.faa	238.9	305.1	0.00848	1592	Lactobacillus gasseri	ATCC 33323	0.783022	1874.2	28172.17	35978.77	1658.2	64150.94	-7806.6
SRR1151155	SRR1151155.protein.faa	153.1	107.5	0.0063	1657	Lactobacillus gigeriorum	DSM 23908	1.424186	1824.4	24301.59	17063.49	1611.4	41365.08	7238.095
ERR387507	ERR387507.protein.faa	169.4	249.8	0.02081	1543	Lactobacillus hamsteri	DSM 5661	0.678143	1792.4	8140.317	12003.84	1623.4	20144.16	-3863.53
GCA_000970855.1	GCA_000970855.1.protein.faa	202.9	293.4	0.0267	1613	Lactobacillus helsingborgensis	Bma5	0.691547	1913.5	7599.251	10988.76	1703.5	18588.01	-3389.51
GCA_000160855.1	GCA_000160855.1.protein.faa	357.5	335	0.01858	1734	Lactobacillus helveticus	DSM 20075	1.067164	1921.5	19241.12	18030.14	1711.5	37271.26	1210.98
SRR1151158	SRR1151158.protein.faa	329.7	269.7	0.02137	1703	Lactobacillus hominis	DSM 23910	1.222469	1822	15428.17	12620.5	1643	28048.67	2807.674
GCA_000160875.1	GCA_000160875.1.protein.faa	944.3	1519	0.1226	1147	Lactobacillus iners	DSM 13335	0.621659	1765.7	7702.284	12389.89	1721.7	20092.17	-4687.6
ERR387510	ERR387510.protein.faa	383.2	516.1	0.04595	1632	Lactobacillus intestinalis	DSM 6629	0.742492	1970.9	8339.499	11231.77	1764.9	19571.27	-2892.27
SRR1151162	SRR1151162.protein.faa	72.62	54.52	0.003547	1362	Lactobacillus jensenii	DSM 20557	1.331988	1459.9	20473.64	15370.74	1343.9	35844.38	5102.904
SRR1151139	SRR1151139.protein.faa	169.4	165.6	0.007801	1603	Lactobacillus johnsonii	ATCC 33200	1.022947	1763.2	21715.16	21228.05	1599.2	42943.21	487.117
SRR1151163	SRR1151163.protein.faa	632.3	720.4	0.07594	1710	Lactobacillus kalixensis	DSM 16043	0.877707	2023.1	8326.31	9486.437	1798.1	17812.75	-1160.13
SRR1151211	SRR1151211.protein.faa	211.4	77.41	5.35E-06	1991	Lactobacillus kefiranofaciens ssp. kefiranofaciens	DSM 5016	2.730913	2182.01	39499253	14463752	1857.01	53963004	25035501
SRR1151212	SRR1151212.protein.faa	81.59	118.6	5.35E-06	1820	Lactobacillus kefiranofaciens ssp. kefirgranum	DSM 10550	0.687943	2136.01	15244768	22159940	1857.01	37404709	-6915172
GCA_000970755.1	GCA_000970755.1.protein.faa	158.6	148.5	0.01833	1653	Lactobacillus kimbladii	Hma2	1.068013	1961.9	8652.482	8101.473	1642.9	16753.96	551.0093
ERR387512	ERR387512.protein.faa	130.4	150.4	0.004751	1692	Lactobacillus kitasatonis	DSM 16761	0.867021	1937	27446.85	31656.49	1712	59103.35	-4209.64
SRR1745849	SRR1745849.protein.faa	137.9	138.9	0.01614	1642	Lactobacillus kullabergensis	Biut2	0.992801	1940	8543.99	8605.948	1643	17149.94	-61.9579
GCA_000970775.1	GCA_000970775.1.protein.faa	205.7	189.9	0.02423	1723	Lactobacillus melliventris	Hma8	1.083202	1978.2	8489.476	7837.392	1707.2	16326.87	652.0842
SRR1151169	SRR1151169.protein.faa	197.8	272.2	0.01519	1537	Lactobacillus pasteurii	DSM 23907	0.726672	1758.4	13021.72	17919.68	1611.4	30941.41	-4897.96
ERR387520	ERR387520.protein.faa	94.47	190.4	0.009344	1248	Lactobacillus psittaci	DSM 15354	0.496166	1439.93	10110.23	20376.71	1343.93	30486.94	-10266.5
ERR387546	ERR387546.protein.faa	42.12	59.26	0.000107	1641	Lactobacillus taiwanensis	DSM 21401	0.710766	1833.14	394382	554868.9	1658.14	949250.9	-160487
SRR1151174	SRR1151174.protein.faa	344.3	340.6	0.03281	1812	Lactobacillus ultunensis	DSM 16047	1.010863	2111.3	10493.75	10380.98	1808.3	20874.73	112.7705
SRR1151132	SRR1151132.protein.faa	331.9	448.5	0.03593	1613	Lapidilactobacillus concavus	DSM 17758	0.740022	1881.6	9237.406	12482.61	1729.6	21720.01	-3245.2
SRR1151201	SRR1151201.protein.faa	219.7	337.2	0.02456	1612	Lapidilactobacillus dextrinicus	DSM 20335	0.651542	1842.5	8945.44	13729.64	1729.5	22675.08	-4784.2
SRR1151125	SRR1151125.protein.faa	185.3	124.7	0.005087	1693	Latilactobacillus curvatus	DSM 20019	1.485966	1753.4	36426.18	24513.47	1632.4	60939.65	11912.72
GCA_000615805.1	GCA_000615805.1.protein.faa	371.1	367.5	0.03192	1962	Latilactobacillus fuchuensis	JCM 11249	1.009796	2201.4	11625.94	11513.16	1958.4	23139.1	112.782

ERR387528	ERR387528.protein.faa	202.4	212.9	0.003495	1622	Latilactobacillus graminis	DSM 20719	0.950681	1749.5	57911.3	60915.59	1632.5	118826.9	-3004.29
ERR433493	ERR433493.protein.faa	161.1	42.62	0.001767	1826	Latilactobacillus sakei ssp. carnosus	DSM 15831	3.779916	1865.52	91171.48	24119.98	1707.52	115291.5	67051.5
SRR1151267	SRR1151267.protein.faa	104.8	80.32	0.002523	1732	Latilactobacillus sakei ssp. sakei	DSM 20017	1.304781	1866.52	41537.85	31835.12	1707.52	73372.97	9702.735
SRR1151196	SRR1151196.protein.faa	144.7	208.6	0.01118	2062	Lentilactobacillus buchneri	DSM 20057	0.693672	2408.9	12942.75	18658.32	2125.9	31601.07	-5715.56
GCA_000785105.1	GCA_000785105.1.protein.faa	443	691	0.05631	1867	Lentilactobacillus curieae	CCTCC M 2011381	0.6411	2360	7867.164	12271.35	2115	20138.52	-4404.19
ERR387480	ERR387480.protein.faa	435.6	387.1	0.02254	2448	Lentilactobacillus diolivorans	DSM 14421	1.125291	2913.5	19325.64	17173.91	2399.5	36499.56	2151.73
GCA_001435875.1	GCA_001435875.1.protein.faa	305.5	266.4	0.02181	2330	Lentilactobacillus farraginis	DSM 18382	1.146772	2709.9	14007.34	12214.58	2290.9	26221.92	1792.756
SRR1151260	SRR1151260.protein.faa	212.4	313.3	0.01696	2190	Lentilactobacillus hilgardii	DSM 20176	0.677944	2638.9	12523.58	18472.88	2290.9	30996.46	-5949.29
ERR387463	ERR387463.protein.faa	135.4	207.9	0.01008	1971	Lentilactobacillus kefiri	DSM 20587	0.651275	2280.5	13432.54	20625	2043.5	34057.54	-7192.46
SRR1151216	SRR1151216.protein.faa	245.6	282.5	0.009466	2315	Lentilactobacillus kisonensis	DSM 19906	0.869381	2801.9	25945.49	29843.65	2351.9	55789.14	-3898.16
GCA_001434145.1	GCA_001434145.1.protein.faa	83.06	139.5	0.007113	1987	Lentilactobacillus otakiensis	DSM 19908	0.595412	2311.44	11677.21	19611.98	2043.44	31289.19	-7934.77
SRR1151168	SRR1151168.protein.faa	131.8	161.2	0.004001	2077	Lentilactobacillus parabuchneri	DSM 5707	0.817618	2406.4	32941.76	40289.93	2106.4	73231.69	-7348.16
GCA_001435895.1	GCA_001435895.1.protein.faa	435.6	429.4	0.03884	2409	Lentilactobacillus parafarraginis	DSM 18390	1.014439	2914.8	11215.24	11055.61	2402.8	22270.85	159.6292
SRR1151230	SRR1151230.protein.faa	183.4	250.9	0.009215	2251	Lentilactobacillus rapi	DSM 19907	0.730969	2712.5	19902.33	27227.35	2318.5	47129.68	-7325.01
GCA_001436555.1	GCA_001436555.1.protein.faa	389.2	1038	0.05697	1466	Lentilactobacillus senioris	DSM 24302	0.374952	2216.8	6831.666	18220.12	2114.8	25051.78	-11388.5
SRR1151235	SRR1151235.protein.faa	201	110.5	0.007022	2168	Lentilactobacillus sunkii	DSM 19904	1.819005	2454.5	28624.32	15736.26	2077.5	44360.58	12888.07
ERR387483	ERR387483.protein.faa	166.9	77.14	0.003509	2321	Levilactobacillus acidifarinae	DSM 19394	2.163599	2648.24	47563.41	21983.47	2231.24	69546.88	25579.94
GCA_001433855.1	GCA_001433855.1.protein.faa	94.11	142.1	5.35E-06	2099	Levilactobacillus brevis	DSM 20054	0.66228	2470.99	17584081	26550822	2146.99	44134903	-8966741
GCA_000807975.1	GCA_000807975.1.protein.faa	136.9	93.89	5.35E-06	2190	Levilactobacillus brevis	BSO 464	1.458089	2656.99	25579223	17542975	2146.99	43122197	8036248
ERR387482	ERR387482.protein.faa	286.4	245	0.01593	2172	Levilactobacillus hammesii	DSM 16381	1.16898	2549.6	17978.66	15379.79	2130.6	33358.44	2598.87
SRR1151217	SRR1151217.protein.faa	368.1	271.5	0.02554	2256	Levilactobacillus koreensis	JCM 16448	1.355801	2569.4	14412.69	10630.38	2159.4	25043.07	3782.302
ERR433476	ERR433476.protein.faa	244	534.9	0.03559	1936	Levilactobacillus namurensis	DSM 19117	0.45616	2517.9	6855.858	15029.5	2226.9	21885.36	-8173.64
SRR1151225	SRR1151225.protein.faa	189.7	327.2	0.0219	2022	Levilactobacillus parabrevis	ATCC 53295	0.579768	2516.5	8662.1	14940.64	2159.5	23602.74	-6278.54
GCA_001437125.1	GCA_001437125.1.protein.faa	413.7	716.6	0.05527	1916	Levilactobacillus paucivorans	DSM 22467	0.57731	2512.9	7485.073	12965.44	2218.9	20450.52	-5480.37
GCA_001436675.1	GCA_001436675.1.protein.faa	172.5	417.1	0.01417	1886	Levilactobacillus senmaizukei	DSM 21775	0.41357	2366.6	12173.61	29435.43	2130.6	41609.03	-17261.8
ERR387543	ERR387543.protein.faa	282.6	459.7	0.03569	2104	Levilactobacillus spicheri	DSM 15429	0.614749	2628.1	7918.184	12880.36	2281.1	20798.54	-4962.17
SRR1151256	SRR1151256.protein.faa	68.03	173.2	0.005075	2126	Levilactobacillus zymae	DSM 19395	0.392783	2549.17	13404.93	34128.08	2231.17	47533	-20723.2

GCA_001435755.1	GCA_001435755.1.protein.faa	273.7	216.1	0.01622	1989	Ligilactobacillus acidipiscis	DSM 15836	1.266543	2172.4	16874.23	13323.06	1931.4	30197.29	3551.171
ERR387498	ERR387498.protein.faa	593.7	694.8	0.06249	1855	Ligilactobacillus agilis	DSM 20509	0.854491	2116.1	9500.72	11118.58	1956.1	20619.3	-1617.86
ERR387553	ERR387553.protein.faa	120.6	149.6	0.004226	1674	Ligilactobacillus animalis	DSM 20602	0.80615	1841	28537.62	35399.91	1703	63937.53	-6862.28
ERR433462	ERR433462.protein.faa	361	225.4	0.02354	1865	Ligilactobacillus apodemi	DSM 16634	1.601597	1883.4	15335.6	9575.191	1729.4	24910.79	5760.408
ERR438946	ERR438946.protein.faa	69.78	135.4	0.00485	1320	Ligilactobacillus araffinosus	DSM 20653	0.515362	1475.62	14387.63	27917.53	1385.62	42305.15	-13529.9
ERR387530	ERR387530.protein.faa	120.1	53.64	0.003736	1452	Ligilactobacillus aviarius	DSM 20655	2.239001	1518.54	32146.68	14357.6	1385.54	46504.28	17789.08
GCA_000423245.1	GCA_000423245.1.protein.faa	1240	2065	0.1534	1221	Ligilactobacillus ceti	DSM 22408	0.600484	2094	8083.442	13461.54	2046	21544.98	-5378.1
GCA_001435735.1	GCA_001435735.1.protein.faa	932.5	906.6	0.07263	1982	Ligilactobacillus equi	DSM 15833	1.028568	2162.1	12839.05	12482.45	1956.1	25321.49	356.602
ERR387461	ERR387461.protein.faa	699.9	1046	0.08396	1447	Ligilactobacillus hayakitensis	DSM 18933	0.66912	1889.1	8336.112	12458.31	1793.1	20794.43	-4122.2
ERR387504	ERR387504.protein.faa	271.2	114.2	0.008657	1860	Ligilactobacillus murinus	DSM 20452	2.374781	1873	31327.25	13191.64	1703	44518.89	18135.61
SRR1151229	SRR1151229.protein.faa	237.2	301.5	0.01826	1867	Ligilactobacillus pobuzihii	KCTC 13174	0.786733	2188.3	12990.14	16511.5	1931.3	29501.64	-3521.36
ERR433499	ERR433499.protein.faa	1072	1322	0.1143	1775	Ligilactobacillus ruminis	ATCC 27780	0.810893	2153	9378.828	11566.05	2025	20944.88	-2187.23
GCA_000423265.1	GCA_000423265.1.protein.faa	1235	1686	0.1355	1580	Ligilactobacillus saerimneri	DSM 16049	0.732503	2177	9114.391	12442.8	2031	21557.2	-3328.41
GCA_000159395.1	GCA_000159395.1.protein.faa	525.7	544.4	0.05131	1774	Ligilactobacillus salivarius	ATCC 11741	0.96565	1947.7	10245.57	10610.02	1792.7	20855.58	-364.451
SRR1151148	SRR1151148.protein.faa	347.7	204.5	0.0254	1879	Limosilactobacillus antri	DSM 16041	1.700244	1969.8	13688.98	8051.181	1735.8	21740.16	5637.795
SRR1151152	SRR1151152.protein.faa	693.9	612.8	0.06123	1767	Limosilactobacillus coleohominis	DSM 14060	1.132343	1897.9	11332.68	10008.17	1685.9	21340.85	1324.514
SRR1151138	SRR1151138.protein.faa	323.3	530.9	0.03231	1447	Limosilactobacillus equigenerosi	DSM 18793	0.608966	1752.6	10006.19	16431.45	1654.6	26437.64	-6425.26
ERR203996	ERR203996.protein.faa	190.1	237.4	0.02524	1600	Limosilactobacillus fermentum	ATCC 14931	0.800758	1789.3	7531.696	9405.705	1647.3	16937.4	-1874.01
ERR387529	ERR387529.protein.faa	235.8	362.1	0.03637	1543	Limosilactobacillus frumenti	DSM 13145	0.651201	1802.3	6483.365	9956.008	1669.3	16439.37	-3472.64
GCA_001434365.1	GCA_001434365.1.protein.faa	292.5	312.1	0.02816	1635	Limosilactobacillus gastricus	DSM 16045	0.9372	1838.6	10387.07	11083.1	1654.6	21470.17	-696.023
GCA_001293735.1	GCA_001293735.1.protein.faa	212.9	395.2	0.03529	1465	Limosilactobacillus gorillae	KZ01	0.538715	1750.3	6032.871	11198.64	1647.3	17231.51	-5165.77
ERR387499	ERR387499.protein.faa	793.6	727.2	0.07039	1891	Limosilactobacillus ingluviei	DSM 15946	1.091309	2019.6	11274.33	10331.01	1824.6	21605.34	943.3158
SRR1151164	SRR1151164.protein.faa	767.4	806.9	0.07659	1859	Limosilactobacillus mucosae	DSM 13345	0.951047	2083.5	10019.58	10535.32	1898.5	20554.9	-515.733
GCA_001434465.1	GCA_001434465.1.protein.faa	85.49	96.31	0.004733	1725	Limosilactobacillus oris	DSM 4864	0.887654	1935.82	18062.54	20348.62	1735.82	38411.16	-2286.08
SRR1151250	SRR1151250.protein.faa	217.7	298.6	0.02646	1660	Limosilactobacillus panis	DSM 6035	0.729069	1967.9	8227.513	11284.96	1740.9	19512.47	-3057.45
SRR1151252	SRR1151252.protein.faa	222.8	512.5	0.04001	1466	Limosilactobacillus pontis	DSM 8475	0.434732	1903.7	5568.608	12809.3	1755.7	18377.91	-7240.69
GCA_000010005.1	GCA_000010005.1.protein.faa	425.6	470.3	0.0466	1745	Limosilactobacillus reuteri	JCM 1112	0.904954	2064.7	9133.047	10092.27	1789.7	19225.32	-959.227

GCA_001437055.1	GCA_001437055.1.protein.faa	338.9	651.9	0.05191	1373	Limosilactobacillus secaliphilus	DSM 17896	0.519865	1816	6528.607	12558.27	1686	19086.88	-6029.67
SRR1151175	SRR1151175.protein.faa	304.5	377.8	0.0417	1596	Limosilactobacillus vaginalis	DSM 5837	0.805982	1806.3	7302.158	9059.952	1669.3	16362.11	-1757.79
ERR387493	ERR387493.protein.faa	93.45	125.1	0.001389	1993	Liquorilactobacillus aquaticus	DSM 21051	0.747002	2241.65	67278.62	90064.79	2024.65	157343.4	-22786.2
SRR1151197	SRR1151197.protein.faa	120.9	409.7	0.01697	1719	Liquorilactobacillus cacaonum	DSM 21116	0.295094	2111.8	7124.337	24142.6	2007.8	31266.94	-17018.3
ERR387459	ERR387459.protein.faa	155.9	232.2	0.006388	1923	Liquorilactobacillus capillatus	DSM 19910	0.671404	2183.3	24405.13	36349.41	1999.3	60754.54	-11944.3
SRR1151207	SRR1151207.protein.faa	337.8	392	0.03296	2128	Liquorilactobacillus ghanensis	DSM 18630	0.861735	2470.2	10248.79	11893.2	2182.2	22141.99	-1644.42
ERR387525	ERR387525.protein.faa	190.2	246.6	0.00453	2037	Liquorilactobacillus hordei	DSM 19519	0.77129	2295.4	41986.75	54437.09	2093.4	96423.84	-12450.3
SRR1151218	SRR1151218.protein.faa	315.8	148.2	0.004053	2261	Liquorilactobacillus mali	DSM 20444	2.130904	2391.4	77917.59	36565.51	2093.4	114483.1	41352.08
ERR387505	ERR387505.protein.faa	353.5	390.7	0.03627	2145	Liquorilactobacillus nagelii	DSM 13675	0.904786	2446.2	9746.347	10771.99	2182.2	20518.33	-1025.64
SRR1151264	SRR1151264.protein.faa	133.4	391.8	0.01734	1812	Liquorilactobacillus oeni	DSM 19972	0.34048	2234.4	7693.195	22595.16	2070.4	30288.35	-14902
ERR433495	ERR433495.protein.faa	330.6	260	0.02578	2141	Liquorilactobacillus satsumensis	DSM 16230	1.271538	2370.4	12823.89	10085.34	2070.4	22909.23	2738.557
ERR485115	ERR485115.protein.faa	176.8	133.1	0.006495	2043	Liquorilactobacillus sucicola	DSM 21376	1.328325	2221.3	27220.94	20492.69	1999.3	47713.63	6728.253
ERR387550	ERR387550.protein.faa	303.1	141.8	0.007192	2186	Liquorilactobacillus uvarum	DSM 19971	2.137518	2363.7	42144.05	19716.35	2024.7	61860.4	22427.7
GCA_000255495.2	GCA_000255495.2.protein.faa	565.4	901	0.07325	1878	Liquorilactobacillus vini	DSM 20605	0.627525	2441.6	7718.771	12300.34	2213.6	20019.11	-4581.57
SRR1151124	SRR1151124.protein.faa	1531	1347	0.1157	2624	Loigolactobacillus bifermentans	DSM 20003	1.1366	2865	13232.5	11642.18	2440	24874.68	1590.32
GCA_001433765.1	GCA_001433765.1.protein.faa	193.2	274.2	0.003227	2269	Loigolactobacillus coryniformis ssp. coryniformis	DSM 20001	0.704595	2660	59869.85	84970.56	2350	144840.4	-25100.7
SRR1151133	SRR1151133.protein.faa	101.7	173.6	0.001064	2278	Loigolactobacillus coryniformis ssp. torquens	DSM 20004	0.585829	2612.9	95582.71	163157.9	2349.9	258740.6	-67575.2
ERR433491	ERR433491.protein.faa	1756	2111	0.1809	2017	Loigolactobacillus rennini	DSM 20253	0.831833	2574	9707.02	11669.43	2372	21376.45	-1962.41
ERR387532	ERR387532.protein.faa	1438	2045	0.1509	1829	Paralactobacillus selangorensis	ATCC BAA 66	0.703178	2671	9529.49	13552.02	2436	23081.51	-4022.53
GCA_000829395.1	GCA_000829395.1.protein.faa	341.5	234.4	0.02745	2020	Paucilactobacillus hokkaidonensis	LOOC260	1.456911	2220.9	12440.8	8539.162	1912.9	20979.96	3901.639
SRR1151187	SRR1151187.protein.faa	420.5	910.6	0.06001	1597	Paucilactobacillus oligofermentans	DSM 15707	0.461783	2212.1	7007.165	15174.14	2087.1	22181.3	-8166.97
GCA_001434475.1	GCA_001434475.1.protein.faa	211.1	163.5	0.005677	2110	Paucilactobacillus suebicus	DSM 5007	1.291131	2447.4	37185.13	28800.42	2062.4	65985.56	8384.71
ERR387501	ERR387501.protein.faa	292.9	267.3	0.01151	2088	Paucilactobacillus vaccinostercus	DSM 20634	1.095773	2414.4	25447.44	23223.28	2062.4	48670.72	2224.153
GCA_000876205.1	GCA_000876205.1.protein.faa	163.9	446.8	0.02433	1630	Paucilactobacillus wasatchensis	WDC04	0.366831	2089.9	6736.539	18364.16	1912.9	25100.7	-11627.6
SRR896433	SRR896433.protein.faa	347.6	303.8	0.02393	2538	Schleiferilactobacillus harbinensis	DSM 16991	1.144174	2987.2	14525.7	12695.36	2494.2	27221.06	1830.338
SRR1151227	SRR1151227.protein.faa	579.9	475.1	0.04373	2592	Schleiferilactobacillus perolens	DSM 12744	1.220585	3001.2	13260.92	10864.4	2487.2	24125.31	2396.524
GCA_000469325.1	GCA_000469325.1.protein.faa	696.1	650.3	0.05847	2540	Schleiferilactobacillus shenzhenensis	LY-73	1.070429	2929.2	11905.25	11121.94	2494.2	23027.19	783.3077
GCA_001435975.1	GCA_001435975.1.protein.faa	178.1	123.2	0.002381	2591	Secundilactobacillus collinoides	DSM 20515	1.445617	3169.1	74800.5	51742.97	2536.1	126543.5	23057.54

SRR1151214	SRR1151214.protein.faa	502.3	668.1	0.04856	2209	Secundilactobacillus kimchicus	JCM 15530	0.751834	2676.8	10343.9	13758.24	2374.8	24102.14	-3414.33
SRR1151262	SRR1151262.protein.faa	359.3	474.5	0.04662	1762	Secundilactobacillus malefermentans	DSM 5705	0.757218	2128.2	7706.993	10178.04	1877.2	17885.03	-2471.04
ERR433478	ERR433478.protein.faa	44.38	179.8	0.000693	2086	Secundilactobacillus odoratitofui	DSM 19909	0.24683	2538.42	64040.4	259451.7	2221.42	323492.1	-195411
GCA_000740055.1	GCA_000740055.1.protein.faa	184.6	358.8	0.02252	1703	Secundilactobacillus oryzae	SG293	0.514493	2033.2	8197.158	15932.5	1877.2	24129.66	-7735.35
SRR1151134	SRR1151134.protein.faa	356.7	289.8	0.006206	2603	Secundilactobacillus paracollinoides	DSM 15502	1.230849	3147.1	57476.64	46696.75	2536.1	104173.4	10779.89
GCA_001313225.1	GCA_001313225.1.protein.faa	1146	454.6	0.02671	3016	Secundilactobacillus silagei	JCM 19001	2.520897	2908.6	42905.28	17019.84	2324.6	59925.12	25885.44
ERR387542	ERR387542.protein.faa	592.3	252.7	0.01219	2561	Secundilactobacillus similis	DSM 23365	2.343886	2744.4	48589.01	20730.11	2221.4	69319.11	27858.9