


Article

# Comparative Genomic Analysis of *Lactiplantibacillus plantarum* Isolated from Different Niches

Bingyong Mao <sup>1,2</sup> , Ruimin Yin <sup>1,2</sup>, Xiaoshu Li <sup>1,2</sup>, Shumao Cui <sup>1,2</sup>, Hao Zhang <sup>1,2,3</sup>, Jianxin Zhao <sup>1,2,\*</sup> and Wei Chen <sup>1,2,3</sup>

- <sup>1</sup> State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi 214122, China; maobingyong@jiangnan.edu.cn (B.M.); yinruiminyeah@163.com (R.Y.); sharonli1994@foxmail.com (X.L.); cuishumao@jiangnan.edu.cn (S.C.); zhanghao61@jiangnan.edu.cn (H.Z.); chenwei66@jiangnan.edu.cn (W.C.)
- <sup>2</sup> School of Food Science and Technology, Jiangnan University, Wuxi 214122, China
- <sup>3</sup> National Engineering Research Center for Functional Food, Jiangnan University, Wuxi 214122, China
- \* Correspondence: zhaojianxin@jiangnan.edu.cn; Tel.: +86-510-8591-2155

**Abstract:** *Lactiplantibacillus plantarum* can adapt to a variety of niches and is widely distributed in many sources. We used comparative genomics to explore the differences in the genome and in the physiological characteristics of *L. plantarum* isolated from pickles, fermented sauce, and human feces. The relationships between genotypes and phenotypes were analyzed to address the effects of isolation source on the genetic variation of *L. plantarum*. The comparative genomic results indicate that the numbers of unique genes in the different strains were niche-dependent. *L. plantarum* isolated from fecal sources generally had more strain-specific genes than *L. plantarum* isolated from pickles. The phylogenetic tree and average nucleotide identity (ANI) results indicate that *L. plantarum* in pickles and fermented sauce clustered independently, whereas the fecal *L. plantarum* was distributed more uniformly in the phylogenetic tree. The pan-genome curve indicated that the *L. plantarum* exhibited high genomic diversity. Based on the analysis of the carbohydrate active enzyme and carbohydrate-use abilities, we found that *L. plantarum* strains isolated from different sources exhibited different expression of the Glycoside Hydrolases (GH) and Glycosyl Transferases (GT) families and that the expression patterns of carbohydrate active enzymes were consistent with the evolution relationships of the strains. *L. plantarum* strains exhibited niche-specific characteristics and the results provided better understating on genetics of this species.

**Keywords:** *Lactiplantibacillus plantarum*; comparative genomics; diversity; carbohydrate metabolism



**Citation:** Mao, B.; Yin, R.; Li, X.; Cui, S.; Zhang, H.; Zhao, J.; Chen, W. Comparative Genomic Analysis of *Lactiplantibacillus plantarum* Isolated from Different Niches. *Genes* **2021**, *12*, 241. <https://doi.org/10.3390/genes12020241>

Academic Editor: Radhey S. Gupta  
Received: 23 December 2020  
Accepted: 3 February 2021  
Published: 8 February 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Lactic acid bacteria (LAB) are microorganisms commonly found in the gut of humans and animals [1]. The major LAB in the intestinal tract of humans are streptococci and enterococci, which are usually regarded as pathogens. Other genera are very minor components (<0.01% of the total bacterial population), such as *Ligilactobacillus*, *Lactobacillus*, and *Limosilactobacillus*. However, *Lactobacillus* and their fermented products in the gut can affect multiple physiological aspects, including the composition of the intestinal microbiota, immune responses, tumorigenesis, the serum cholesterol level, and blood pressure [2].

*L. plantarum* is an LAB that can be found in the gastrointestinal tracts of human and animals, meat, fish, fermented vegetables, and dairy products [3–5]. Since the first genome-wide sequencing of *Lactococcus lactis* subsp. *lactis* IL1403 [6], a series of whole-genome sequencing of LAB has been launched worldwide [7–9]. With the rapid development of sequencing techniques, the genomes of multiple *L. plantarum* strains have been sequenced [7]. This advancement has provided a better understanding of the relationships among genotypes and functions and the directions of strain evolution [8,9]. Comparative genomics allows horizontal comparison of the genomes of multiple bacterial strains so inter-strain

differences can be investigated more efficiently and conveniently [10] in order to enable better functional analysis at the molecular level [11].

Several studies have depicted the genetic diversity of *L. plantarum* strains using different phenotypic and genotypic approaches, such as amplified fragment length polymorphism (AFLP) [12], random amplified polymorphic DNA (RAPD) [13], multilocus sequence typing (MLST) [14], and microarray-based comparative genome hybridization (CGH) [15,16]. These studies revealed that various *L. plantarum* strains generally exhibited high conservation for genes involved in protein and lipid synthesis or degradation and high diversity for genes involved in sugar transport and catabolism.

A previous comparative genomics study on 54 *L. plantarum* strains revealed that *L. plantarum* isolated from various niches exhibited high genetic diversity and distinct phylogenetic patterns [10]. However, a genomic characterization of 108 *L. plantarum* strains using pan-genome analysis revealed no direct connections between genomic traits and the niche [17]. Moreover, the associations between genotypes and phenotypes have not been fully considered and investigated [18].

In this study, we studied the genomics of 133 *L. plantarum* strains isolated from various niches. We analyzed their core genes, pan genes, and phylogenetics (on the basis of homologous genes) and explored the functional genomics from a comparative genomics perspective. Furthermore, to establish an efficient method to rapidly select strains with application potential on the basis of their genetic features, we explored the relationships between genotypes and carbohydrate use capacity.

## 2. Materials and Methods

### 2.1. Chemicals and Reagents

Cellobiose, sorbose, xylose, mannose, glucuronic acid, ribose, trehalose, and fucose were purchased from Sinopharm Chemical Reagent Co., Ltd. (Shanghai, China); arabinose and raffinose were purchased from Sigma-Aldrich (Shanghai) Trading Co., Ltd. (Shanghai, China); and fructooligosaccharides were purchased from Bao Lingbao Biotechnology Co., Ltd. (Dezhou, Shandong Province, China).

### 2.2. *L. plantarum* Strains

In this study, 324 samples were collected for isolation of *L. plantarum*, including 126 pickles, 32 fermented sauces, and 166 human feces. The samples were serially diluted, plated on solid MRS medium, and cultured at 37 °C for 24 h. The colonies with the characteristics of lactic acid bacteria were picked up, purified, and identified by 16S rRNA sequencing. If multiple strains of *L. plantarum* are isolated and identified in the same sample, only one strain is selected for the genomic analysis. Thus, a total of 114 strains of *L. plantarum* were obtained in this study, including 32/11/71 strains from pickles/fermented sauce/feces, respectively. The strains were reserved at the Culture Collections of Food Microbiology in Jiangnan University.

Meanwhile, 19 strains of *L. plantarum* with a complete genome were selected from the NCBI database, covering a wide range of niches (Table S1).

### 2.3. Draft Genome Sequencing of *L. plantarum*

*L. plantarum* strains were inoculated in MRS medium and cultured at 37 °C for 12–16 h, and the bacterial sludge was collected by centrifugation at 8000 × *g* for 5 min. Draft genome sequencing was performed by Novogene Co., Ltd. (Beijing, China) using an Illumina HiSeq 2000 sequencing platform.

### 2.4. Genomic Assembly, Prediction, and Functional Annotation

SOAPdenovo software was used for genome assembly and verification of single bases, GapCloser software was used to fill gaps within the genome [19,20], the open reading frame (ORF) was predicted using Glimmer and Genmark software [21], and protein BLAST

(BLASTp) from Clusters of Orthologous Groups (COG) Proteome Database was used for ORF homology comparison and functional annotation.

### 2.5. Mapping of Homologous Gene Venn Diagram and Phylogenetic Tree Construction

PGAP software was used to calculate the core genome of *L. plantarum* [22]. Orthologous gene prediction was performed on the genome of each strain using OrthoMCL software [23]. The orthologous gene was then aligned using MAFFT v7.3 software [24], and the Venn diagram was generated using Perl script. Phylogenetic trees were constructed on the basis of the 1620 core genes using PHYLIP v3.6 [25].

### 2.6. Average Nucleotide Identity (ANI) Analysis

PYANI was used to calculate the orthologous average nucleotide identity of the genome. PYANI (<https://github.com/widdowquinn/pyani>) is a Python3 module. We clustered and visualized the obtained matrix using R package heatmap software. ANI is the average value based on the comparisons of all orthologous protein-encoding genes of the pairwise genomes and is a classic method for judging whether strains belong to the same species [26]. Strains with an ANI value of over 95% are generally considered to be of the same species [27].

### 2.7. Carbohydrate-Active Enzyme Analysis

Carbohydrate-active enzyme (CAZy) in *L. plantarum* was annotated with HMMER v3.0 [28]. The target protein sequence was aligned with hmmscan-parser.sh from the dbCAN prediction platform and the best hit was reserved [29]. The genome was considered to contain the corresponding CAZys when the CAZys analysis was consistent with the annotation results of the dbCAN prediction platform. The detailed carbohydrate-active enzyme family information was obtained on the CAZyme (<http://www.cazy.org/>) [30]. The database mainly included glycoside hydrolase (GHs), glycosyltransferases (GTs), carbohydrate esterases (CEs), carbohydrate-binding enzymes (CBM), auxiliary active enzymes (AAs), and polysaccharide lyases (PLs).

### 2.8. Determination of the Carbohydrate Utilization Capacity of *L. plantarum*

A series of MRS media were prepared using different carbohydrates as the sole carbon source ( $10 \text{ gL}^{-1}$ ). The pH was adjusted to 7.0, and bromocresol purple was added to the medium as an acid production indicator (0.75%, w/v). A medium with glucose and a carbohydrate-free medium were used as the positive and negative controls, respectively. The sterilized medium was dispensed into sterile 96-well plates (200  $\mu\text{L}$  for each well). The bacterial strains were resuspended in PBS and cultured in the medium (1%) at 37 °C for 12–24 h. The use of carbohydrate by the bacteria was indicated by the color change of the medium (from violet to yellow).

## 3. Results

### 3.1. Genome Characteristics of 133 *L. plantarum* Strains

The detailed information of the 114 *L. plantarum* strains is shown in Table S1. In total, 11/32/71 strains were isolated from fermented sauce/pickles/feces, respectively. While the 19 reference strains were isolated from various sources, including sauerkraut, baby droppings, brewing, saliva, pickles, dairy products, fermented sausage, raw milk, *Drosophila*, silage, fermented bean paste, feces, fermented tamarind pulp, and cheese. The basic genomic characteristics of *L. plantarum* strains are shown in Table 1. We found that the genomic size of the 114 strains ranged from 2.94 to 3.90 Mb, and the number of genes ranged from 2767 to 3650. The genomic size and gene number of the strains isolated from fermented sources exhibited a narrower range than those isolated from feces. The 19 reference strains had similar genomic size and number of genes to the 114 strains, and specific strains had a larger number of genes. The GC content is an important indicator of bacterial classification and is associated with many genetic characteristics [31]. We found



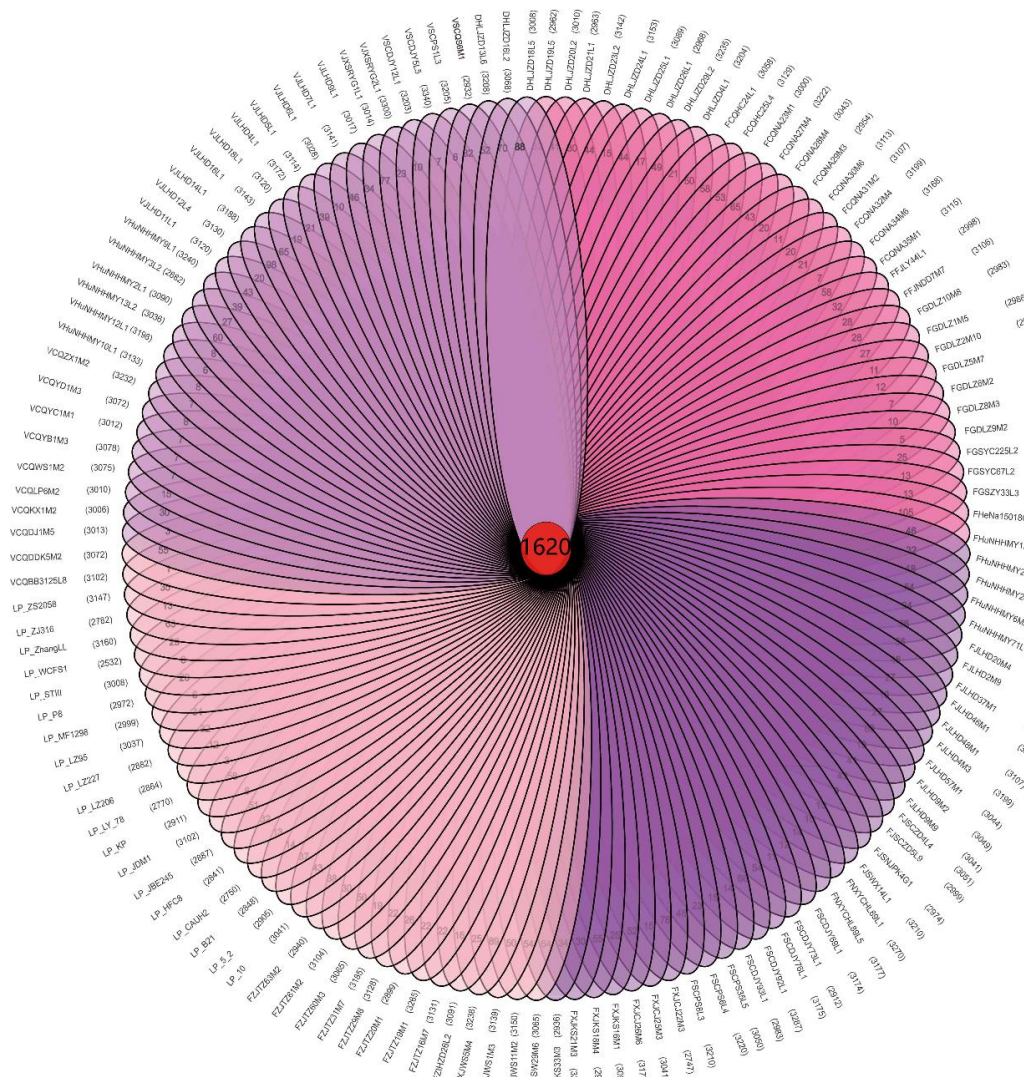
that the GC content of our *L. plantarum* strains ranged from 44.08% to 46.55%, in which seven strains had GC contents over 46% (Table S1). The GC content of the 19 reference strains ranged from 44.08% to 44.90%, which was less than that of the 114 strains.

**Table 1.** Basic characteristics of 133 *Lactiplantibacillus plantarum* strains.

Source of Isolation	Number of Strains	Genomic Size (Mb)	Number of Genes	GC Content (%)
Fermented sauce	11	3.14–3.45	3049~3416	45.43–46.33
Pickles	32	3.23–3.50	2931–3464	45.16–46.55
Feces	71	2.94–3.90	2767–3487	45.25–46.29
Others *	19	2.95–3.69	2860–3650	44.08–44.90

\* The source of isolation “others” include sauerkraut, baby droppings, brewing, saliva, pickles, dairy products, fermented sausage, raw milk, *Drosophila*, silage, fermented bean paste, feces, fermented tamarind pulp, and cheese. The detailed information is shown in Table S1.

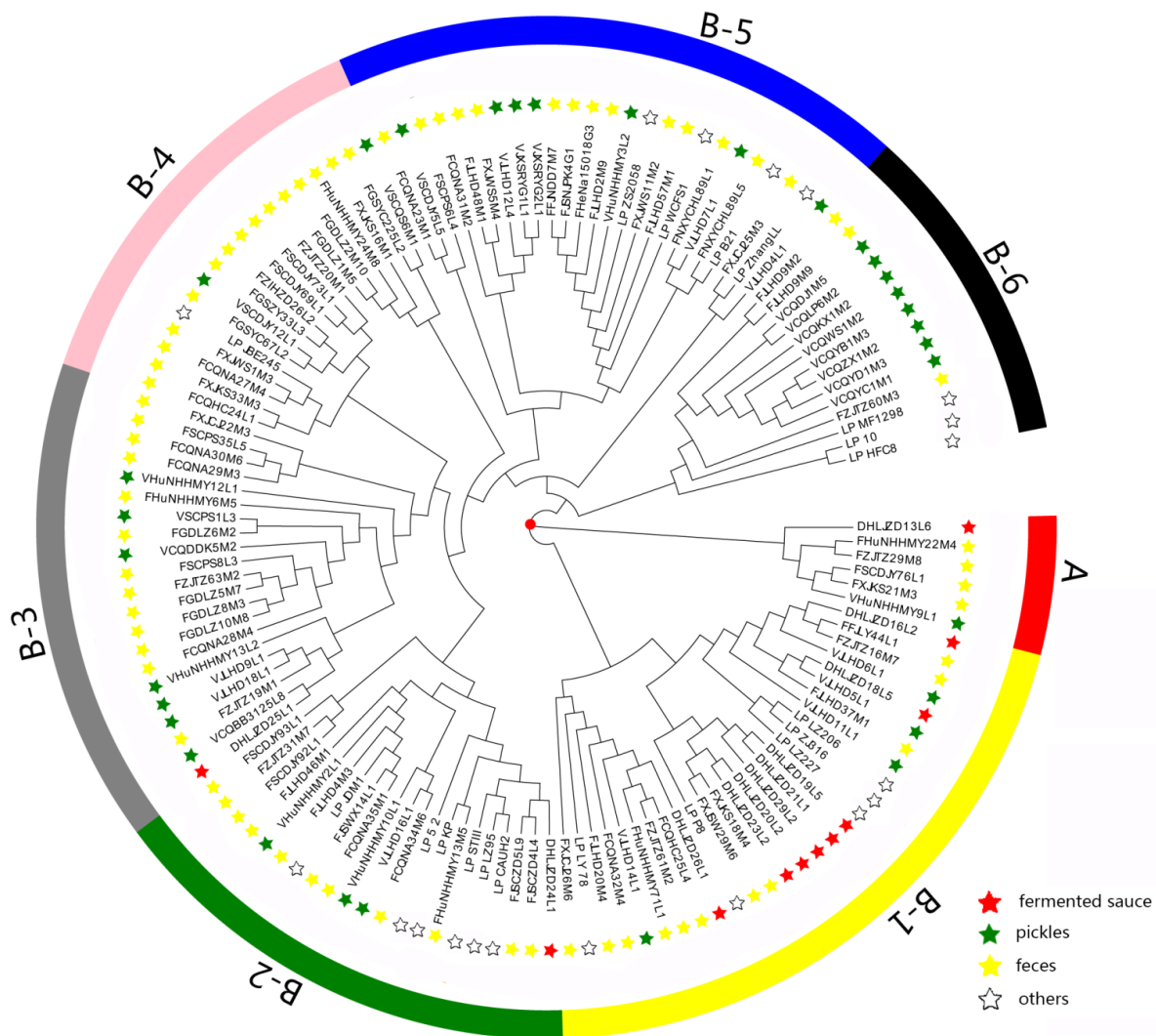
The genes of the bacterial strains can be divided into core genes, non-essential genes, and unique genes [32]. We identified 1620 core genes in the 133 *L. plantarum* strains (Figure 1), which was consistent with the previously reported *L. plantarum* core gene value of 1709 [17]. *L. plantarum* had more core genes than other LAB [24,33], which may explain their high adaptability to diverse niches.



**Figure 1.** The 1620 core genes and unique genes among 133 *L. plantarum* strains.

### 3.2. Analysis of the Phylogenetic Tree with Regard to Homologous Genes

Bacterial adaptation to the environment depends not only on the presence or absence of specific genes, but also on the accumulation of specific allelic variations of conserved genes [34]. We conducted phylogenetic analysis of the 133 *L. plantarum* strains to study the role of genetic variation and to reveal their evolutionary relationships (Figure 2).



**Figure 2.** Phylogenetic tree based on sequence similarities among all the 1620 core genes. “Others” stands for the sources of isolation of the 19 reference strains, including sauerkraut, baby droppings, brewing, saliva, pickles, dairy products, fermented sausage, raw milk, *Drosophila*, silage, fermented bean paste, feces, fermented tamarind pulp, and cheese.

As shown in Figure 2, the 133 strains fell into two clusters (cluster A and B). Cluster A contained six strains, namely, VHuNHMY9L1, FXJKS21M3, FSCDJY76L1, FZJTZ29M8, FHuNHMY22M4, and DHLJZD13L6. We found that the clustering was not affected by various niches in cluster A, possibly due to the small number of strains within the cluster.

Cluster B contained 127 strains and could be subdivided into B-1, B-2, B-3, B-4, B-5, and B-6 clusters. We found that some strains isolated from pickles and fermented sauce were clearly distinguished in the phylogenetic tree. Eight strains isolated from fermented sauce accumulated in the B-1 cluster, five of which aggregated in the same branch under the B-1 cluster. More strains isolated from pickles accumulated in the B-6 cluster than in the B-1 cluster. Interestingly, fecal *L. plantarum* strains exhibited no specific distribution patterns.

The genomic evolution of *L. plantarum* within the same niche was also affected by other factors. For the strains isolated from pickles, eight strains clustered in cluster B-6 and



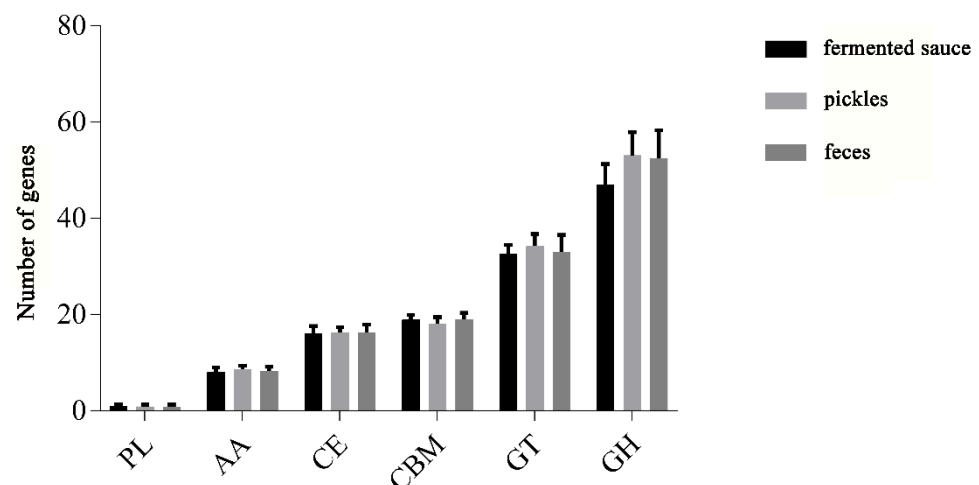


cantly differed from the other 127 strains. The maximum ANI value across these two groups was 95.23%, whereas the ANI values among the six strains were between 99.23% and 99.29%.

The ANI threshold for differentiating strains within the same species was updated to 96% [36] in 2018 (strains with an ANI value greater than 96% are considered the same species). Due to the small difference between our inter-cluster ANI (95.23%) and the accepted threshold, we speculated that these six strains were of a separate species or subspecies in the *L. plantarum* group.

#### 3.4. Analysis of Carbohydrate-Active Enzymes

In viewing the annotation results of carbohydrate-active enzymes (CAZy), we found that the 114 *L. plantarum* strains contained all six major families of glycoside hydrolase (GH), glycosyl transferase (GT), carbohydrate esterase (CE), carbohydrate-binding enzyme (CBM), auxiliary active enzyme (AA), and polysaccharide lyase (PL). We found that the most abundant CAZy genes in the *L. plantarum* strains belonged to the GH family, followed by the GT, CBM, CE, AA, and PL families (Figure 4). We observed no significant differences in the number of genes in the families of PL, AA, CE, and CBM for *L. plantarum* strains isolated from different niches.

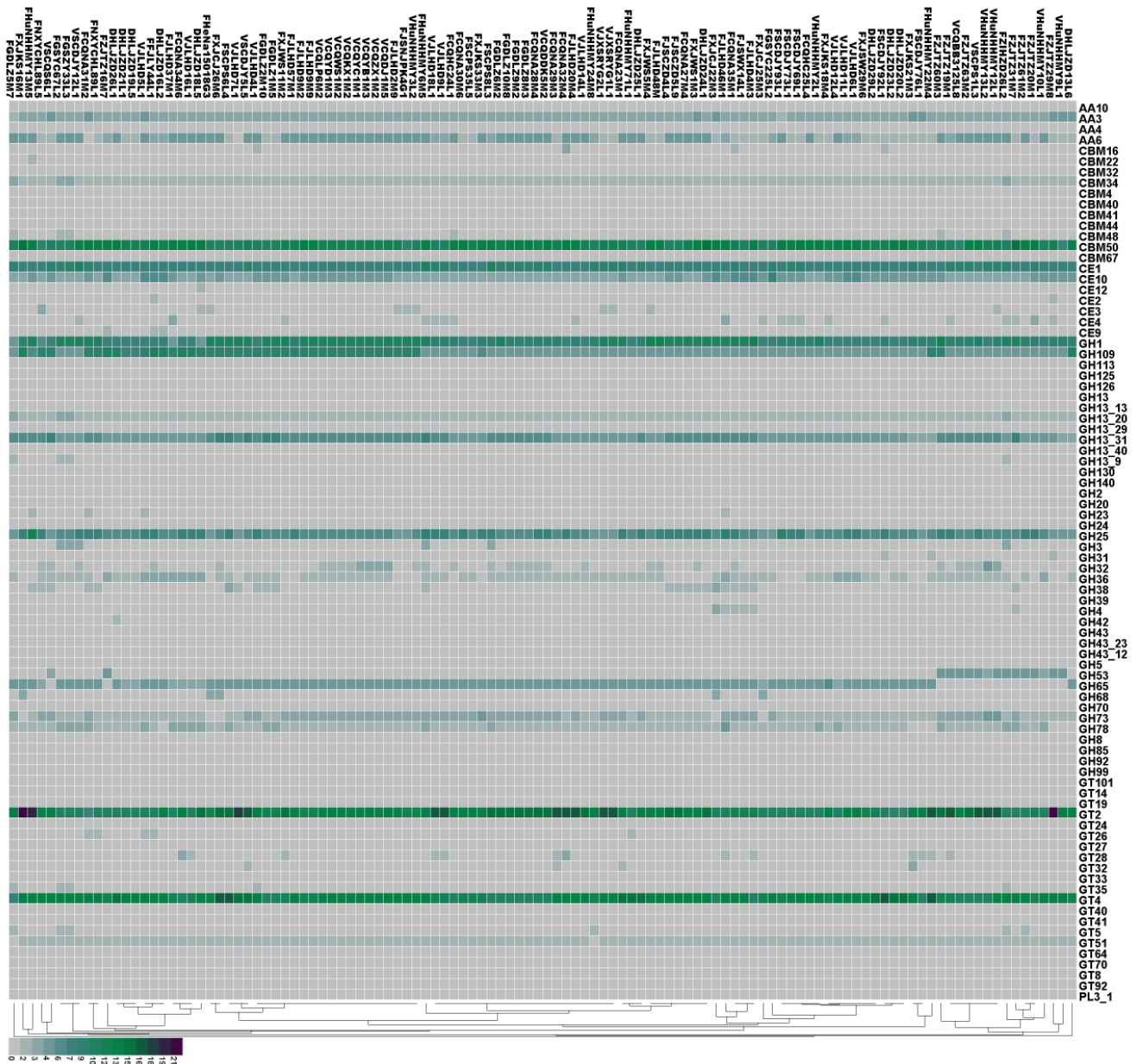


**Figure 4.** Statistics for carbohydrate-active enzyme (CAZy) family of *L. plantarum* isolated from different niches.

To further clarify the differences in CAZy for *L. plantarum* strains isolated from different niches, we next characterized the genes of CAZy using heat maps (Figure 5). Our results indicate that all the strains contained the genes encoding for the AA3 family proteins. The AA3 enzymes belong to the glucose–methanol–choline (GMC) oxidoreductase family and can be divided into four subfamilies: AA3\_1 (mostly cellobiose dehydrogenases), AA3\_2 (including both aryl alcohol oxidase and glucose 1-oxidase), AA3\_3 (alcohol oxidase), and AA3\_4 (pyranose 2-oxidase). CBM50 was present in all strains, and was found attached to various enzymes from families GH18, GH19, GH23, GH24, GH25, and GH73, i.e., enzymes cleaving either chitin or peptidoglycan. The GH family were the most abundant enzymes in the CAZy database, which is required for the degradation of glycosidic linkages between carbohydrates [37]. We found that the GH family members encoded in all strains were GH1, GH109, GH13\_31, and GH25. The GT family mainly encodes for glycosyltransferases. We identified GT2 (required for cellulose synthesis) and GT4 (required for sucrose synthesis) in all strains.

Our results reveal that *L. plantarum* strains isolated from different niches encoded different CAZy enzymes. Specifically, nine strains isolated from feces in Zhejiang did not contain the GH65 family (required for rhamnose use). Moreover, fecal *L. plantarum* strains had fewer GH36 genes than those isolated from pickles and fermented sauce. The increased

number of GH36 genes may help *L. plantarum* strains isolated from pickles to use raffinose, which is a carbohydrate commonly found in plants [38]. The GH43 and GH68 families were found only in the strains isolated from feces, and the GT27 family was found only in the strains isolated from pickles. The GH42 family, which includes  $\beta$ -galactosidase and  $\alpha$ -L-arabinosidase, was only found in the strain DHLJZD26L1.



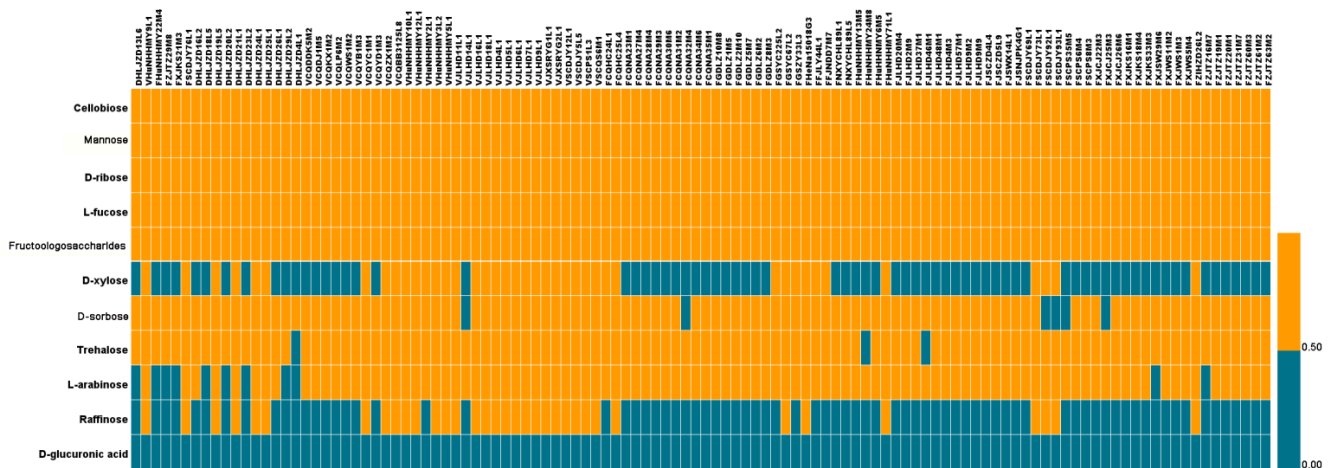
**Figure 5.** Prediction of carbohydrate-active enzymes (CAZy) in 114 *L. plantarum* strains by protein blast. The isolation sources can be differentiated according to the first letter of the strain names. “F” stands for feces, “V” stands for pickles, and “D” stands for fermented sauce.

### 3.5. Analysis of Carbohydrate Use Ability

We further explored the relationships between carbohydrate use-related genes and carbohydrate use capabilities. We analyzed the ability of different *L. plantarum* strains in using 11 carbohydrates (see the Materials and Methods section), including cellobiose, sorbose, xylose, mannose, glucuronic acid, ribose, trehalose, fucose, arabinose, raffinose, and fructooligosaccharides (FOS) (Figure 6). Our results indicate that all *L. plantarum*



strains were able to use cellobiose, mannose, ribose, and fucose, which was consistent with the results in Bergey's Manual of Systematic Bacteriology. Consistent with a previous report [8], we found that all *L. plantarum* strains used FOS. Finally, none of the 114 strains were able to use glucuronic acid, and only 31.25% of the strains could use raffinose.



**Figure 6.** Use of 11 carbohydrates by 114 *L. plantarum* strains.

These findings indicate that the *L. plantarum* strains isolated from different niches showed various abilities to use specific carbohydrates. Although 25% of the *L. plantarum* strains isolated from pickles did not use xylose, 66.67% and 76.39% of *L. plantarum* strains isolated from fermented sauce and feces, respectively, were unable to use xylose. Among the 114 strains, 6 strains, including 5 fecal strains, were unable to use sorbose. We found that three strains did not use trehalose, and 11 strains did not use arabinose, of which five were isolated from fermented sauce in Heilongjiang and six from feces. All strains isolated from pickles were able to use arabinose.

Strains isolated from feces used xylose and raffinose less efficiently than those isolated from pickles and fermented sauce. Thus, our results indicate that the ability of *L. plantarum* to use carbohydrates was strain-specific and, to some extent, niche-specific (dependent on the available carbon sources of that niche).

#### 4. Discussion

In this study, comparative genomic analyses were performed for the 133 *L. plantarum* strains isolated from different sources. The phylogenetic tree and ANI analyses indicate that the niche affected the survival and selection, and thus the genomes, of the various *L. plantarum* strains. Indeed, environmental selection pressure promotes genomic modifying processes such as horizontal gene transfer or gene deletion, which improve the adaptability of the strain to a specific niche [39–41]. The strains isolated from feces were found in each cluster of the phylogenetic tree (Figure 2), which indicates that the fecal strains showed greater similarity to those isolated from food sources. Siezen et al. studied the sugar-metabolizing capabilities of 185 *L. plantarum* strains isolated from different niches and found that the sugar-metabolizing capability of *L. plantarum* differed according to the food niche, whereas the sugar-metabolizing capability of strains isolated from human feces was not uniform [18]. Consistent with our results from this study, these findings indicated that human fecal *L. plantarum* strains originated from various food sources. We found that a higher proportion of strains isolated from feces diverged later in the phylogenetic tree than those isolated from pickles and fermented sauce, indicating that *L. plantarum* found in the human gut is inoculated via diet.

The use of carbohydrates by *L. plantarum* was not only closely related to the CAZy expressions but also to the expressions of transporters (phosphoglucose transferase (PTS) or ABC transport system) and regulatory proteins. Transporters, regulatory proteins, and

related CAZys together constitute an operon that enables sugar metabolism. Wang et al. reported that *L. plantarum* ST-III, WCFS1, and JDM1 contained a large number of phosphoglucose transferase (PTS) systems and displayed broad adaptability to various substrates [42]. In *L. plantarum*, hexose is transported to the cells, phosphorylated by the PTS and permease systems, and fed into the glycolytic pathway and converted to pyruvate [43].

The abilities of the 114 *L. plantarum* strains to use 11 sugars were determined in this study. Our results indicate that all of the strains could use cellobiose, mannose, ribose, fucose, and FOS (Figure 6). According to the CAZy analysis, all of the strains contained 5–12 GH1 families (including  $\beta$ -glucosidase). The reaction catalyzed by  $\beta$ -glucosidase is the most important step in cellobiose degradation. In addition, some strains contained the PTS system, including EIIA, EIIB, EIIC, and HPr. Thus, our findings indicated that all of the *L. plantarum* strains were able to use cellobiose. Mannose-6-phosphoisomerase (EC 5.3.1.8) and the PTS system IIA and IIB (for transporting mannose) are mainly involved in the metabolism of mannose and were found in all the strains. We identified genes involved in ribose metabolism in all strains, such as ABC transporter, purine nucleoside phosphorylase (EC 2.4.2.1), pyrimidine nucleoside phosphorylase (EC 2.4.2.2), and ribose 5-phosphate isomerase A (EC 5.3.1.6). FOS, generally considered as prebiotics, can be used by the intestinal bacteria to produce SCFAs, which have beneficial properties for the host [44]. FOS is made up of a sucrose molecule linked (to its fructose residue) to an extra 1–3 fructose molecules via  $\beta$ -1,2 glycosidic linkages.  $\beta$ -Fructofuranosidase belongs to the GH32 family and is important for the hydrolysis of FOS. In this study, GH32 family was only found in part of the 114 strains, although all the strains metabolized FOS.

Most of the 114 *L. plantarum* strains used trehalose, L-arabinose, and sorbitol, and fewer than half of the strains used xylose and raffinose. None of the strains used glucuronic acid. Trehalose consists of two glucose molecules linked via 1,1 glycosidic bonds, which can be hydrolyzed by  $\alpha$ -glucosidase (GH13). In this study, we found that 111 strains used trehalose. L-arabinose is a naturally occurring pentose. In lactobacilli, L-arabinose is transported into the cells via transporters, converted to ribulose (by L-arabinose isomerase—EC 5.3.1.4), and then phosphorylated (by ribulose kinase—EC 2.7.1.16), converted to xylulose phosphate (by L-ribulose-5-phosphate 4-epimerase—EC 5.1.3.4), and finally metabolized. In this study, we observed differences in the use of L-arabinose among the 114 strains. We found that 103 strains used L-arabinose. Genes related to arabinose metabolism were also identified, including L-arabinose isomerase, L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4), and L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4). The gene *araA* (encoding L-arabinose isomerase) was not found in the other 11 strains that did not use L-arabinose. In lactobacilli, the xylose is transported into the cell by the cotransporter Xy1E/Xy1T and the binding protein system Xy1FGH. Intracellular xylose is converted to xylulose by xylose isomerase (encoded by *xylA*) under the regulation of Xy1R. Xylulose is phosphorylated (to form xylulose-5 phosphate) by xylulose kinase (encoded by *xylB*), which then enters the pentose phosphate pathway [45]. Our results indicate that only 41 of the 114 strains used xylose. We also identified XylH and the xylulose kinase XylB in these strains. However, *xylH* was not found in the other 73 strains that did not use xylose. Raffinose is a trisaccharide composed of galactose, glucose, and fructose molecules linked by  $\alpha$ -1,6 and  $\alpha$ -1,3 glycosidic linkages [46]. The genes for  $\alpha$ -galactosidase (EC 3.2.1.22) and  $\beta$ -fructofuranosidase (EC 3.2.1.26), which hydrolyze raffinose, were found in all 114 strains. However, not all strains used raffinose. The gene *rafP* (encoding for raffinose permease) was only found in the strains capable of using raffinose. In addition, we found that none of the strains use glucuronic acid, probably due to the lack of the GT1 and GT101 families.

## 5. Conclusions

In summary, the comparative genomic results show that the number of strain-specific genes has a certain relationship with the niche in which they are located. The strains from kimchi and fermented sauce were distinctly distinguished, while the strains from feces were mixed in the branches of the phylogenetic tree. The core functions of *L. plantarum* were

concentrated on carbohydrate metabolism and amino acid metabolism, which provide molecular support for strains to metabolize multiple sugars. Moreover, we found that the ability of *L. plantarum* strains to use carbohydrates correlated with the carbohydrate metabolism-related genes, including genes encoding glycosidases and transporters. We have previously established that protein BLAST (BLASTp) alignment of the glycosidases and transporters against the protein database of gut bacteria provides effective and reliable predictions on the oligosaccharide-using abilities of gut bacteria [47–49]. Further literature mining of the carbohydrate metabolism-related genes would improve the prediction accuracy of this method. Here, we showed that the carbohydrate-using phenotypes of *L. plantarum* strains could be predicted on the basis of certain genetic traits. Finally, we found that a small number of genes did not correspond to the carbohydrate-metabolizing phenotype, probably because the genes were lost in specific niches.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/2073-4425/12/2/241/s1>: Table S1: Genomic information and sample information of 133 *L. plantarum* strains.

**Author Contributions:** Conceptualization, B.M. and J.Z.; methodology, B.M. and R.Y.; software, R.Y., X.L., and S.C.; investigation, R.Y. and X.L.; resources, H.Z., W.C., and J.Z.; writing—original draft preparation, R.Y. and B.M.; writing—review and editing, B.M. and S.C.; supervision, J.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by National Natural Science Foundation of China (grant no. 31972086, 31801530), National First-Class Discipline Program of Food Science and Technology (JUFSTR20180102), and Collaborative Innovation Center of Food Safety and Quality Control in Jiangsu Province.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** This study generated sequencing data for 114 *L. plantarum* isolates, and all sequence data have been deposited in the National Center for Biotechnology Information under project no. PRJNA658852; the accession numbers of all isolates are listed in Table S1.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Grangette, C.; Nutten, S.; Palumbo, E.; Morath, S.; Hermann, C.; Dewulf, J.; Pot, B.; Hartung, T.; Hols, P.; Mercenier, A. Enhanced anti-inflammatory capacity of a *Lactobacillus plantarum* mutant synthesizing modified teichoic acids. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10321–10326. [CrossRef]
2. Broekaert, I.J.; Walker, W.A. Probiotics and chronic disease. *J. Clin. Gastroenterol.* **2006**, *40*, 270–274. [CrossRef]
3. Mayo, B.; Gonzalez, B.; Arca, P.; Suarez, J.E. Cloning and expression of the plasmid-encoded  $\beta$ -d-galactosidase gene from a *Lactobacillus-plantarum* strain of dairy origin. *FEMS Microbiol. Lett.* **1994**, *122*, 145–151. [CrossRef] [PubMed]
4. Kongnum, K.; Hongpattarakere, T. Effect of *Lactobacillus plantarum* isolated from digestive tract of wild shrimp on growth and survival of white shrimp (*Litopenaeus vannamei*) challenged with *Vibrio harveyi*. *Fish Shellfish. Immun.* **2012**, *32*, 170–177. [CrossRef]
5. de Vries, M.C.; Vaughan, E.E.; Kleerebezem, M.; de Vos, W.M. *Lactobacillus plantarum*- survival, functional and potential probiotic properties in the human intestinal tract. *Int. Dairy J.* **2006**, *16*, 1018–1028. [CrossRef]
6. Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malmgren, K.; Weissenbach, J.; Ehrlich, S.D.; Sorokin, A. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp *lactis* IL1403. *Genome Res.* **2001**, *11*, 731–753. [CrossRef] [PubMed]
7. Siezen, R.J.; Francke, C.; Renckens, B.; Boekhorst, J.; Wels, M.; Kleerebezem, M.; van Hijum, S.A.F.T. Complete Resequencing and Reannotation of the *Lactobacillus plantarum* WCFS1 Genome. *J. Bacteriol.* **2012**, *194*, 195–196. [CrossRef]
8. Liu, C.J.; Wang, R.; Gong, F.M.; Liu, X.F.; Zheng, H.J.; Luo, Y.Y.; Li, X.R. Complete genome sequences and comparative genome analysis of *Lactobacillus plantarum* strain 5-2 isolated from fermented soybean. *Genomics* **2015**, *106*, 404–411. [CrossRef] [PubMed]
9. Axelsson, L.; Rud, I.; Naterstad, K.; Blom, H.; Renckens, B.; Boekhorst, J.; Kleerebezem, M.; van Hijum, S.; Siezen, R.J. Genome Sequence of the Naturally Plasmid-Free *Lactobacillus plantarum* Strain NC8 (CCUG 61730). *J. Bacteriol.* **2012**, *194*, 2391–2392. [CrossRef] [PubMed]
10. Martino, M.E.; Bayjanov, J.R.; Caffrey, B.E.; Wels, M.; Joncour, P.; Hughes, S.; Gillet, B.; Kleerebezem, M.; van Hijum, S.A.F.T.; Leulier, F. Nomadic lifestyle of *Lactobacillus plantarum* revealed by comparative genomics of 54 strains isolated from different habitats. *Environ. Microbiol.* **2016**, *18*, 4974–4989. [CrossRef] [PubMed]

11. Kwak, W.; Kim, K.; Lee, C.; Lee, C.; Kang, J.; Cho, K.; Yoon, S.H.; Kang, D.K.; Kim, H.; Heo, J.; et al. Comparative Analysis of the Complete Genome of *Lactobacillus plantarum* GB-LP2 and Potential Candidate Genes for Host Immune System Enhancement. *J. Microbiol. Biotechnol.* **2016**, *26*, 684–692. [[CrossRef](#)]
12. Torriani, S.; Clementi, F.; Vancanneyt, M.; Hoste, B.; Dellaglio, F.; Kersters, K. Differentiation of *Lactobacillus plantarum*, *L. pentosus* and *L. paraplantarum* species by RAPD-PCR and AFLP. *Syst. Appl. Microbiol.* **2001**, *24*, 554–560. [[CrossRef](#)] [[PubMed](#)]
13. Johansson, M.L.; Quednau, M.; Molin, G.; Ahrne, S. Randomly amplified polymorphic DNA (rapd) for rapid typing of *Lactobacillus plantarum* strains. *Lett. Appl. Microbiol.* **1995**, *21*, 155–159. [[CrossRef](#)] [[PubMed](#)]
14. de las Rivas, B.; Marcobal, A.; Munoz, R. Development of a multilocus sequence typing method for analysis of *Lactobacillus plantarum* strains. *Microbiol. Sgm.* **2006**, *152*, 85–93. [[CrossRef](#)]
15. Molenaar, D.; Bringel, F.; Schuren, F.H.; de Vos, W.M.; Siezen, R.J.; Kleerebezem, M. Exploring *Lactobacillus plantarum* genome diversity by using microarrays. *J. Bacteriol.* **2005**, *187*, 6119–6127. [[CrossRef](#)] [[PubMed](#)]
16. Siezen, R.J.; Vlieg, J.E.T.V. Genomic diversity and versatility of *Lactobacillus plantarum*, a natural metabolic engineer. *Microb. Cell Factories* **2011**, *10*, S3. [[CrossRef](#)]
17. Choi, S.; Jin, G.D.; Park, J.; You, I.; Kim, E.B. Pan-Genomics of *Lactobacillus plantarum* Revealed Group-Specific Genomic Profiles without Habitat Association. *J. Microbiol. Biotechnol.* **2018**, *28*, 1352–1359. [[CrossRef](#)]
18. Siezen, R.J.; Tzeneva, V.A.; Castioni, A.; Wels, M.; Phan, H.T.K.; Rademaker, J.L.W.; Starrenburg, M.J.C.; Kleerebezem, M.; Molenaar, D.; Vlieg, J.E.T.V. Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches. *Environ. Microbiol.* **2010**, *12*, 758–773. [[CrossRef](#)] [[PubMed](#)]
19. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.Z.; Chen, Y.X.; Pan, Q.; Liu, Y.J.; et al. Erratum: SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2015**, *4*, 30. [[CrossRef](#)] [[PubMed](#)]
20. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.Z.; Chen, Y.X.; Pan, Q.; Liu, Y.J.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *Gigascience* **2012**, *1*, 18. [[CrossRef](#)] [[PubMed](#)]
21. Borodovsky, M.; Mills, R.; Besemer, J.; Lomsadze, A. Prokaryotic Gene Prediction Using GeneMark and GeneMark.hmm. *Curr. Protoc. Bioinform.* **2003**, *1*, 4.5.1–4.5.16. [[CrossRef](#)] [[PubMed](#)]
22. Zhao, Y.; Wu, J.; Yang, J.; Sun, S.; Xiao, J.; Yu, J. PGAP: Pan-genomes analysis pipeline. *Bioinformatics* **2012**, *28*, 416–418. [[CrossRef](#)]
23. Skarp-de Haan, C.P.A.; Culebro, A.; Schott, T.; Revez, J.; Schweda, E.K.H.; Hanninen, M.L.; Rossi, M. Comparative genomics of unintrogressed *Campylobacter coli* clades 2 and 3. *BMC Genom.* **2014**, *15*, 129. [[CrossRef](#)]
24. Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066. [[CrossRef](#)]
25. Wickstead, B.; Gull, K.; Richards, T.A. Patterns of kinesin evolution reveal a complex ancestral eukaryote with a multifunctional cytoskeleton. *BMC Evol. Biol.* **2010**, *10*, 110. [[CrossRef](#)]
26. Richter, M.; Rossello-Mora, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19126–19131. [[CrossRef](#)]
27. Yoon, S.H.; Ha, S.M.; Lim, J.; Kwon, S.; Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Leeuw Int. J. G.* **2017**, *110*, 1281–1286. [[CrossRef](#)]
28. Mistry, J.; Finn, R.D.; Eddy, S.R.; Bateman, A.; Punta, M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* **2013**, *41*, e121. [[CrossRef](#)] [[PubMed](#)]
29. Yin, Y.; Mao, X.; Yang, J.; Chen, X.; Mao, F.; Xu, Y. dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2012**, *40*, 445–451. [[CrossRef](#)] [[PubMed](#)]
30. Lombard, V.; Ramulu, H.G.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [[CrossRef](#)]
31. Rodin, S.N.; Parkhomchuk, D.V. Position-associated GC asymmetry of gene duplicates. *J. Mol. Evol.* **2004**, *59*, 372–384. [[CrossRef](#)]
32. Jorgensen, R. Altered gene expression in plants due to trans interactions between homologous genes. *Trends Biotechnol.* **1990**, *8*, 340–344. [[CrossRef](#)]
33. Kant, R.; Rintahaka, J.; Yu, X.; Sigvart-Mattila, P.; Paulin, L.; Mecklin, J.P.; Saarela, M.; Palva, A.; von Ossowski, I. A Comparative Pan-Genome Perspective of Niche-Adaptable Cell-Surface Protein Phenotypes in *Lactobacillus rhamnosus*. *PLoS ONE* **2014**, *9*, e102762. [[CrossRef](#)]
34. Zhang, S.; Song, W.; Yu, M.; Lin, X. Comparative genomics analysis of five *Psychrobacter* strains isolated from world-wide habitats reveal high intra-genus variations. *Extremophiles* **2017**, *21*, 581–589. [[CrossRef](#)] [[PubMed](#)]
35. Zhao, N.; Zhang, C.; Yang, Q.; Guo, Z.; Yang, B.; Lu, W.; Li, D.Y.; Tian, F.W.; Liu, X.M.; Zhang, H.; et al. Selection of Taste Markers Related to Lactic Acid Bacteria Microflora Metabolism for Chinese Traditional Paocai: A Gas Chromatography-Mass Spectrometry -Based Metabolomics Approach. *J. Agric. Food Chem.* **2016**, *64*, 2415–2422. [[CrossRef](#)]
36. Ciufu, S.; Kannan, S.; Sharma, S.; Badretdin, A.; Clark, K.; Turner, S.; Brover, S.; Schoch, C.L.; Kimchi, A.; DiCuccio, M. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 2386–2392. [[CrossRef](#)]
37. Duan, C.J.; Xian, L.; Zhao, G.C.; Feng, Y.; Pang, H.; Bai, X.L.; Tang, J.L.; Ma, Q.S.; Feng, J.X. Isolation and partial characterization of novel genes encoding acidic cellulases from metagenomes of buffalo rumens. *J. Appl. Microbiol.* **2009**, *107*, 245–256. [[CrossRef](#)] [[PubMed](#)]



38. Zdunczyk, Z.; Jankowski, J.; Juskiewicz, J.; Lecewicz, A.; Slominski, B. Application of soybean meal, soy protein concentrate and isolate differing in  $\alpha$ -galactosides content to low- and high-fibre diets in growing turkeys. *J. Anim. Physiol. Anim. Nutr.* **2010**, *94*, 561–570. [[CrossRef](#)] [[PubMed](#)]
39. Douillard, F.P.; Ribbera, A.; Kant, R.; Pietila, T.E.; Jarvinen, H.M.; Messing, M.; Randazzo, C.L.; Paulin, L.; Laine, P.; Ritari, J.; et al. Comparative Genomic and Functional Analysis of 100 *Lactobacillus rhamnosus* Strains and Their Comparison with Strain GG. *PLoS Genet.* **2013**, *9*, e1003683. [[CrossRef](#)] [[PubMed](#)]
40. Lee, J.Y.; Han, G.G.; Kim, E.B.; Choi, Y.J. Comparative genomics of *Lactobacillus salivarius* strains focusing on their host adaptation. *Microbiol. Res.* **2017**, *205*, 48–58. [[CrossRef](#)]
41. Yu, J.; Zhao, J.; Song, Y.; Zhang, J.; Yu, Z.; Zhang, H.; Sun, Z.H. Comparative Genomics of the Herbivore Gut Symbiont *Lactobacillus reuteri* Reveals Genetic Diversity and Lifestyle Adaptation. *Front. Microbiol.* **2018**, *9*, 1151. [[CrossRef](#)] [[PubMed](#)]
42. Wang, Y.; Chen, C.; Ai, L.; Zhou, F.; Zhou, Z.; Wang, L.; Zhang, H.; Chen, W.; Guo, B.H. Complete Genome Sequence of the Probiotic *Lactobacillus plantarum* ST-III. *J. Bacteriol.* **2011**, *193*, 313–314. [[CrossRef](#)]
43. Miller, K.A.; Phillips, R.S.; Mrazek, J.; Hoover, T.R. *Salmonella* Utilizes D-Glucosaminic Acid via a Mannose Family Phosphotransferase System Permease and Associated Enzymes. *J. Bacteriol.* **2013**, *195*, 4057–4066. [[CrossRef](#)] [[PubMed](#)]
44. Gibson, G.R.; Hutkins, R.; Sanders, M.E.; Prescott, S.L.; Reimer, R.A.; Salminen, S.J.; Scott, K.; Stanton, C.; Swanson, K.S.; Cani, P.D.; et al. The International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nat. Rev. Gastroenterol. Hepatol.* **2017**, *14*, 491–502. [[CrossRef](#)]
45. Jo, S.; Yoon, J.; Lee, S.M.; Um, Y.; Han, S.O.; Woo, H.M. Modular pathway engineering of *Corynebacterium glutamicum* to improve xylose utilization and succinate production. *J. Biotechnol.* **2017**, *258*, 69–78. [[CrossRef](#)]
46. Jeffrey, G.A.; Huang, D.B. The hydrogen bonding in the crystal structure of raffinose pentahydrate. *Carbohydr. Res.* **1990**, *206*, 173–182. [[CrossRef](#)]
47. Mao, B.; Li, D.; Zhao, J.; Liu, X.; Gu, Z.; Chen, Y.Q.; Zhang, H.; Chen, W. In Vitro Fermentation of Lactulose by Human Gut Bacteria. *J. Agric. Food Chem.* **2014**, *62*, 10970–10977. [[CrossRef](#)]
48. Mao, B.; Li, D.; Zhao, J.; Liu, X.; Gu, Z.; Chen, Y.Q.; Zhang, H.; Chen, W. In vitro fermentation of fructooligosaccharides with human gut bacteria. *Food Funct.* **2015**, *6*, 947–954. [[CrossRef](#)]
49. Mao, B.; Tang, H.; Gu, J.; Li, D.; Cui, S.; Zhao, J.; Zhang, H.; Chen, W. In vitro fermentation of raffinose by the human gut bacteria. *Food Funct.* **2018**, *9*, 5824–5831. [[CrossRef](#)] [[PubMed](#)]