Journal of Microbiology (2021) Vol. 59, No. 6, pp. 573–583
DOI 10.1007/s12275-021-0618-6
eISSN 1976-3794
pISSN 1225-8873

Comparative genomics analysis of *Pediococcus acidilactici* species[§]

Zhenzhen Li[1,2,3], Qi Song[3,4], Mingming Wang[3], Junli Ren[3], Songling Liu[3,4], and Shancen Zhao[3*]

[1]*BGI College, Zhengzhou University, Zhengzhou 450052, P. R. China*
[2]*School of Life Sciences, Zhengzhou University, Zhengzhou 450001, P. R. China*
[3]*BGI Institute of Applied Agriculture, BGI-Shenzhen, Shenzhen 518120, P. R. China*
[4]*Key Laboratory of Genomics, Ministry of Agriculture, BGI-Shenzhen, Shenzhen 518120, P. R. China*

(Received Nov 25, 2020 / Revised Feb 5, 2021 / Accepted Feb 23, 2021)

Pediococcus acidilactici **is a reliable bacteriocin producer and a promising probiotic species with wide application in the food and health industry. However, the underlying genetic features of this species have not been analyzed. In this study, we performed a comprehensive comparative genomic analysis of 41** ***P. acidilactici*** **strains from various ecological niches. The bacteriocin production of 41 strains were predicted and three kinds of bacteriocin encoding genes were identified in 11** ***P. acidilactici*** **strains, namely pediocin PA-1, enterolysin A, and colicin-B. Moreover, whole-genome analysis showed a high genetic diversity within the population, mainly related to a large proportion of variable genomes, mobile elements, and hypothetical genes obtained through horizontal gene transfer. In addition, comparative genomics also facilitated the genetic explanation of the adaptation for host environment, which specify the protection mechanism against the invasion of foreign DNA (i.e. CRISPR/Cas locus), as well as carbohydrate fermentation. The 41 strains of** ***P. acidilactici*** **can metabolize a variety of carbon sources, which enhances the adaptability of this species and survival in different environments. This study evaluated the antibacterial ability, genome evolution, and ecological flexibility of** ***P. acidilactici*** **from the perspective of genetics and provides strong supporting evidence for its industrial development and application.**

Keywords: *Pediococcus acidilactici*, comparative genomics, pan-genome, bacteriocin, genetic diversity

Introduction

Pediococcus is a type of lactic acid bacteria (LAB), of which the most prominent characteristics are acid-tolerance, inability to synthesize porphyrins, and possessing of a strictly fermentative (homofermentative) and facultatively anaerobic metabolism, with lactic acid as the major metabolic end-product (Holzapfel *et al.*, 2006). At present, 11 different species of the genus *Pediococcus* have been found, including *P. acidilactici*, *P. argentinicus*, *P. cellicola*, *P. claussenii*, *P. damnosus*, *P. ethanolidurans*, *P. inopinatus*, *P. parvulus*, *P. pentosaceus*, *P. siamensis*, and *P. stilesii*. Among these species, the species *Pediococcus acidilactici* have drawn special attention, owing to its ability to produce antimicrobial agents (Porto *et al.*, 2017), such as pediocin PA-1, pediocin AcH, pediocin SA-1, and pediocin L50 (Biswas *et al.*, 1991; Henderson *et al.*, 1992; Cintas *et al.*, 1995; Anastasiadou *et al.*, 2008). For this reason, many strains of *P. acidilactici* have been used for the preservation and safety assurance of food and beverage during storage by inhibiting pathogenic and spoilage bacteria (Porto *et al.*, 2017), such as *Listeria monocytogenes* (Komora *et al.*, 2020) and *Staphylococcus aureus* (Cintas *et al.*, 1998). Moreover, some strains of this species were found to prevent the colonization of pathogens such as *Shigella* spp., *Salmonella* spp., *Clostridium difficile*, and *Escherichia coli* in the small intestine. Therefore, it was assumed they could also be used as a probiotic supplement (Feng *et al.*, 2016). Indeed, it has been found that some strains of *P. acidilactici* are able to reduce serum cholesterol levels, improve antioxidant status, and promote nutrient digestibility (Shah *et al.*, 2018). It has also been proved that some other strains can reduce serum triglycerides by inhibiting fat absorption and enhance fat degradation and metabolism simultaneously (Ueda *et al.*, 2018). All these properties make *P. acidilactici* a promising probiotic, and not surprisingly more and more attention has been put on this species.

Another property of the species *P. acidilactici* is broad ecological distribution, including fermented food, plants, animal/human gut, wine, and so on. This broad ecological distribution reflects metabolic flexibility that has fueled widespread application of the species in the food and health industry. For example, strains of this species are employed as acid-producing starter cultures in milk fermentation, as adjunct cultures to accelerate or intensify flavor development in bacterial ripened cheeses, and as probiotics to enhance human or animal health (Goldin and Gorbach, 1992; Stiles and Holzapfel, 1997; Mäyra-Mäkinen and Bigret, 2004). Compared with other fastidious probiotic strains, the widespread application of *P. acidilactici* increases its potential to be exploited as probiotic supplement (Abu-Taraboush *et al.*, 1998; Olszewska and Staniewski, 2012).

Despite their probiotic properties and promising application in the food industry, the underlying genetic/genomics features of this species have not been analyzed. In this study, we performed a comprehensive comparative genomics analysis on 41 *P. acidilactici* strains to investigate the genomic biodiversity among the species, to explore their bacteriocin-

*For correspondence. E-mail: zhaoshancen@genomics.cn; Tel.: +86-186-6629-0372
[§]Supplemental material for this article may be found at http://www.springerlink.com/content/120956.
Copyright ⓒ 2021, The Microbiological Society of Korea

producing ability as well as to analyze potential habitat adaptation, which will provide strong evidence for probiotic properties in this species and enable its engineering.

Materials and Methods

Pan-genome calculation

For all genomes used in this study, the pan-genome and core-genome calculation was performed using GET_HOMOLOGUES pipelines (Contreras-Moreira and Vinuesa, 2013), in order to calculate the total gene repertoire and the degree of overlap and diversity of the 41 *P. acidilactici* strains. Sequence comparisons at the protein level were performed using an all-against-all, bi-directional BLAST alignment (Altschul *et al.*, 1990) with 75% match cutoff and value cutoff of 1e-5, followed by clustering into protein families, named Clusters of Orthologous genes (COGs). A presence/absence matrix was built using a custom Perl scripts incorporated in GET_HOMOLOGUES software, which including all identified COGs in the analyzed genomes and allowing for identification of core- and variable-genome. Then, All the COGs in the core- and variable-genome were categorized in predicted functional groups by COG and KEGG databases annotation (Tatusov *et al.*, 2000). Graphical depictions of *P. acidilactici* pan- and core-genome data, were generated using a custom script with several R packages built in the GET_HOMOLOGUES pipelines.

Average nucleotide identity (ANI) values

ANI between any two genomes was calculated using a perl script implemented in the GET_HOMOLOGUES pipelines and the resulting matrix was clustered and visualized using Pheatmap R packages (Kolde and Kolde, 2015).

Phylogenetic analysis

To analyze the phylogenetic relationships among 41 *P. acidilactici* strains based on genomics data, the gene content method was used to construct the phylogenetic tree, and a gene content matrix obtained from pan-genome calculation was parsed indicating the presence (1) or absence (0) of all genes of 41 *P. acidilactici* genomes. The Jaccard distance (one minus the Jaccard coefficient) between pairwise genomes was calculated based on the gene content matrix and the calculation results were uploaded onto the iTOL online website (Letunic and Bork, 2016) to reconstruct the gene content dendrogram.

Functional genome distribution analysis

A Functional Genome Distribution (FGD) analysis (Altermann, 2012) was performed using 41 *P. acidilactici* genome sequences. GenBank files of 41 *P. acidilactici* genomes were subjected to an FGD analysis and the calculated distance matrix was imported into Molecular Evolutionary Genetics Analysis version 4 (MEGA4) (Tamura *et al.*, 2007). The functional distribution was visualized using the unweighted pair group method with arithmetic mean (UPGMA).

The prediction of the mobilome of *P. acidilaciti* species

Transposable Elements (TEs) were annotated using RepeatMasker software (Tarailo-Graovac and Chen, 2004) and the identification of Prophage in 41 *P. acidilactici* genomes was performed using PHASTER (Arndt *et al.*, 2016).

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) locus, consisting of the CRISPR regions and CRISPR-associated (Cas) proteins were identified using the CRISPR-CasFinder (Couvin *et al.*, 2018) with default settings, and the CRISPR subtypes designation was based on the signature of the Cas proteins (Makarova *et al.*, 2015).

Prediction of horizontally acquired genes was performed by COLOMBO v4.0 implemented with the program SIGIHMM and SigiCRF (Waack *et al.*, 2006).

The prediction of carbohydrate utilization

In silico evaluation of the carbohydrate utilization of *P. acidilactici* was conducted using The Pathway Tools software (Karp *et al.*, 2016), and the role of specific genes associated with carbohydrate unitization was annotated using the RAST Server (Aziz *et al.*, 2008). Carbohydrate-active enzymes were identified based on similarity to the Carbohydrate-Active enzymes (CAZy) database entries (Cantarel *et al.*, 2009).

Bacteriocin identification

All of the strains were assessed for the presence of bacteriocin operons by BAGEL4 (van Heel *et al.*, 2018) and the domains of bacteriocin were determined using BLASTP analysis and distant homology annotation (https://www.ebi.ac.uk/Tools/hmmer/search/jackhmmer).

Results

Genome feature of *P. acidilactici*

41 *P. acidilactici* genomes from the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/2382/) were included in the study. As shown in Table 1, the 41 *P. acidilactici* strains were isolated from different ecological niches, including fermented food, human/animal gut, plant, and wine, which meant a broad representation of genetic, ecological, and geographical diversity in the species. The genome size of all the 41 strains ranged from 1.88 Mb for *P. acidilactici* JKY18 to 2.17 Mb for *P. acidilactici* BCC1, and the average genome size was 1.97 Mb, which was smaller when compared with other bacteria (Burke and Moran, 2011). The GC content of each genome varied slightly, ranging from 42% to 42.4%. The gene number of each genome ranged from 1854 for *P. acidilactici* AS1.2696 to 2135 for *P. acidilactici BCC1* and the number of predicted CDS features in each genome ranged from 776 for *P. acidilactici* JKY18 to 1986, for *P. acidilactici* BCC1, showing obvious plasticity among these genomes.

Pan-genome analysis of *P. acidilactici*

To gain an overall approximation of the total gene pool of *P. acidilactici*, the pan-genome was calculated based on the 41 genomes. A pan-genome with a total of 2733 COGs were

Table 1. General genome features of the 41 *P. acidilactici* strains

Strains	Source	Genome size (Mb)	GC (%)	Gene number	CDS features	Accession number
ZPA017	pig feces	2.13	42.10	2057	1953	SAMN04515708
BCCI	broiler chick cecum	2.17	42.13	2135	1986	SAMN06052263
JQII-5	fermented dairy	2.09	42.20	2025	1841	SAMN07701376
PMC48	perilla leaf kimchi	2.04	42.20	2001	1812	SAMN14362951
CACC 537	canine feces	2.04	42.00	1945	1832	SAMN13871425
SRCM101189	food	2.09	42.10	2059	1926	SAMN07167785
SRCM100424	food	2.09	42.12	2044	1909	SAMN07126147
SRCM100313	food	2.12	42.06	2079	1838	SAMN07125954
SRCM102732	food	2.07	42.14	2038	1883	SAMN08707607
SRCM102731	food	2.06	42.16	2025	1875	SAMN08707606
ATCC 8042	plant Mash	2.01	42.10	1949	1840	SAMN10358339
SRCM103387	food	2.03	42.29	2018	1899	SAMN10743318
SRCM103444	food	1.97	42.10	1899	1785	SAMN10754287
PB22	Korean adult feces	1.96	42.20	1893	1792	SAMN08180273
SRCM103367	food	1.96	42.26	1929	1770	SAMN10743317
DSM 20284	human feces	1.94	42.20	1878	1794	SAMN00116807
DSM 19927	ryegrass silage	2.04	42.10	2000	1895	SAMN02797813
SRCM 103289	food	2.09	42.20	2065	1893	SAMN10737798
AGR20	sheep rumen	1.92	42.40	1904	1735	SAMN02584994
JKY18	wine	1.88	42.30	1930	776	SAMN05804765
FAM 18969	cheese	2.08	42.00	2075	1937	SAMN11653951
S7	wine	2.02	42.10	1958	1860	SAMEA80353168
AS1.2696	food	1.93	42.10	1854	1773	SAMN02797809
MA18/5M	pasture gramineae	1.99	42.10	1915	1799	SAMN02469945
UMNPBX20	*Meleagris gallopavo* ileum	2.03	42.20	1995	1892	SAMN07702371
E24	Meles feces	2.02	42.20	1968	1870	SAMN10822530
F7	Meles feces	2.09	42.20	2061	1937	SAMN10822531
L8-A	bovine	1,95	42.20	1917	1747	SAMN10744163
L2-A	bovine	1.95	42.20	1919	1747	SAMN10744156
E1	honey from beehive	2.01	42.20	1951	1852	SAMN11831832
I32	Meles feces	2.03	42.30	1994	1886	SAMN10822532
SRCM100320	Doenjang	2.08	42.10	2060	1928	SAMN05162513
D3	food	1.96	42.10	2015	1717	SAMN01991044
NBRC 12231	food	1.94	42.20	1892	1734	SAMD00097159
GS1	chicken feces	2.01	42.00	1985	1856	SAMN09531784
L14-B	bovine	1.94	42.30	1915	1746	SAMN10744148
WT	chicken feces	2.02	42.00	2008	1863	SAMN09604260
S1	raw rice wine	1.98	42.00	1961	1788	SAMN04111307
LPBC161	coffee plantation	1.98	42.20	1963	1812	SAMN10690227
M16	Meles feces	1.83	42.20	1922	1735	SAMN10822533
K3	Raw rice wine	1.99	42.10	1941	1548	SAMN04011952

estimated. When plotted on a log-log scale as a function of the number of analyzed genomes, the pan-genome curve displayed an asymptotic trend with a growth rate of an average of 134 COGs in the first 2 iterations and decreasing to an average of 10 COGs in the final addition. As the genome continued to increase, new genes were still discovered, which indicating the existence of an open pan-genome within the *P. acidilactici* species (Fig. 1A).

The core-genome and variable-genome of *P. acidilactici*

The analysis of the core-genome of the 41 *P. acidilactici* strains showed that the number of COGs shared in all strains decreased as more genome sequences were added (Fig. 1B). Eventually, a total of 677 shared COGs were determined as the core-genome that was present in all the considered genomes, which accounted for about 25% of the entire pan-genome. Simultaneously, a total of 2056 variable COGs were identified and 856 of them were classified as unique (Supplementary data Table S1). Among all the studied strains, it was found that *P. acidilactici* M16 and *P. acidilactici* D3 contained the most unique COGs, 124 and 103, respectively (Fig. 1C). Then, a functional characterization was performed by assigning core and variable genes to the COG and KEGG functional class, during this process, 643 and 831 COGs were not included in the analysis due to lack of annotations, respectively. The obtained data regarding the distributions of the COG and KEGG functional classes in each section were

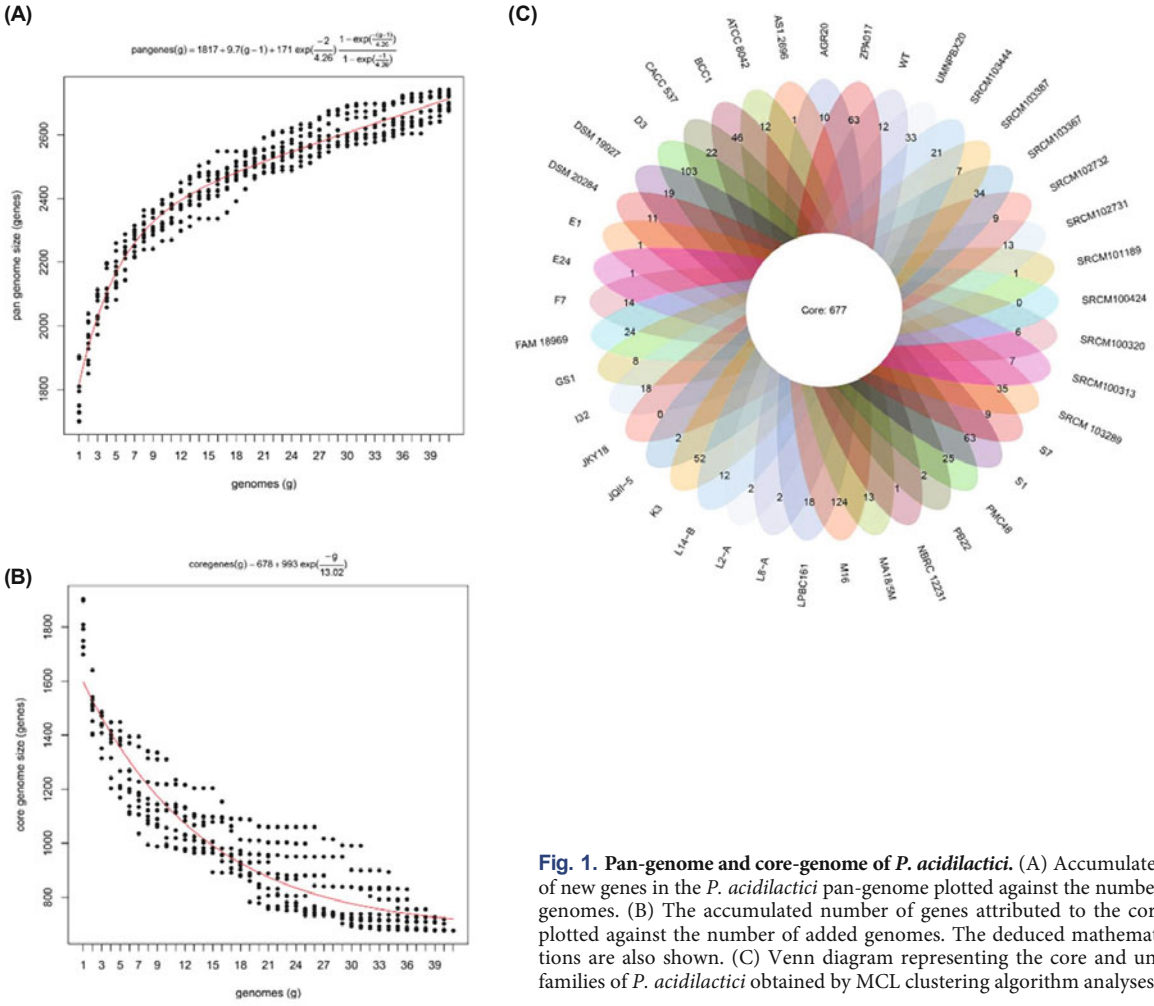

Fig. 1. Pan-genome and core-genome of *P. acidilactici*. (A) Accumulated number of new genes in the *P. acidilactici* pan-genome plotted against the number of added genomes. (B) The accumulated number of genes attributed to the core-genome plotted against the number of added genomes. The deduced mathematical functions are also shown. (C) Venn diagram representing the core and unique gene families of *P. acidilactici* obtained by MCL clustering algorithm analyses.

shown in Supplementary data Fig. S1A and B. Most of the genes in the core-genome were involved in basic physiological functions, including metabolism, genetic information processing and environmental information processing, which further maps the necessity of the core genes for the survival of the bacteria. As expected, the majority of genes of the variable-genome were enriched in carbohydrate metabolism, replication, recombination and repair, defense mechanisms and membrane transport, indicating that the variable genome were closely related to some metabolic pathways or environmental adaptability of the strains, leading to different environmental tolerance of the strains.

ANI and phylogenetic analysis

The average nucleotide identity (ANI) was calculated to quantitatively analyze the similarity between the 41 *P. acidilactici* genomes. The results indicated that all the strains belonged to *P. acidilactici*, as evidenced by the ANI value which was above 96% (Fig. 2A).

To analyze the phylogenetic relationships among the *P. acidilactici* isolates, a genome-relatedness or "pan-genome" tree was generated based on gene content of the 41 genomes. The pan-genome tree (Fig. 2B) shows the presence of seven major groups A, B, C, D, E, F, and G. Group A and Group B consisted of only one strain isolated from fermented food and animal/human gut, respectively. Group E contained mainly strains of animal/human gut, while group G contained mainly strains of fermented food. Apart from this, the remaining three groups contained strains of various origins and were not correlated based on isolation source or sampling region.

There appears to be some correlation between the phylogenetic relatedness and origin of isolation (i.e. fermented food, animal/human gut, etc.) of these strains, but much less than we anticipated (Jiang *et al.*, 2020).

FGD analysis of *P. acidilactici*

The availability of the genome sequences of 41 *P. acidilactici* strains allowed a FGD tree to be constructed (Fig. 2C). In contrast to an evolutionary phylogeny, FGD analyzes the functional relationship between microbes based on their predicted ORFeomes. This type of approach takes genotype adaptations into account which might render organisms more similar to each other than what their respective evolutionary heritage would indicate (Kelly *et al.*, 2016). The FGD tree is largely different from the phylogenetic clusters identified in the pan-genome tree, except for group G, in which five strains

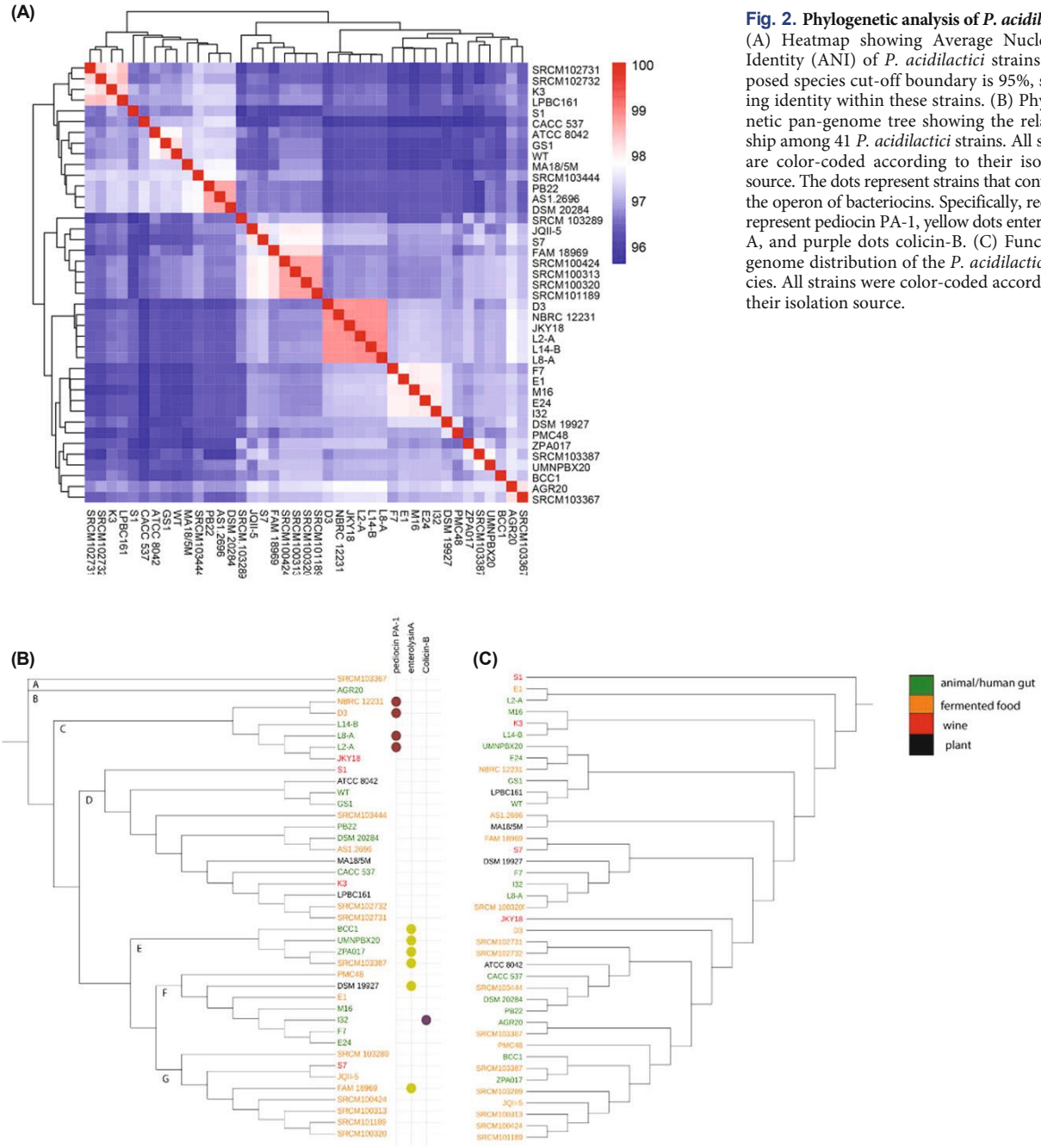

Fig. 2. Phylogenetic analysis of *P. acidilactici*. (A) Heatmap showing Average Nucleotide Identity (ANI) of *P. acidilactici* strains. Proposed species cut-off boundary is 95%, showing identity within these strains. (B) Phylogenetic pan-genome tree showing the relationship among 41 *P. acidilactici* strains. All strains are color-coded according to their isolation source. The dots represent strains that contained the operon of bacteriocins. Specifically, red dots represent pediocin PA-1, yellow dots enterolysin A, and purple dots colicin-B. (C) Functional genome distribution of the *P. acidilactici* species. All strains were color-coded according to their isolation source.

also formed a tight cluster in the FGD tree (*P. acidilactici* S7 and *P. acidilactici* SRCM100320 not included).

In silico prediction of *P. acidilactici* carbohydrate utilization

In order to access possible correlations between the genomic content of the analyzed *P. acidilactici* strains and particular phenotypic abilities, we predicted the carbohydrate metabolic capabilities of each strain using the Pathway Tools. All the 41 strains were predicted to be able to ferment D-ribose, D-mannose, fructose, and cellobiose. In contrast, fermentation capabilities for the other sugars such as D-xylose, lactose, D-galactose, L-arabinose, fucose, D-arabinose, L-rhamnose, D-trehalose, melibiose, and sucrose, were shown to be variable among the tested strains (Fig. 3A).

In order to validate the observed metabolic differences of the *P. acidilactici* species, we predicted the glycosyl hydrolase (GH) enzymes using the CAZy database, which are the key enzymes that metabolize carbohydrates. A total of 32 glycosyl hydrolases were annotated in all 41 strains (Fig. 3B), of which eight glycosyl hydrolases were present in almost all of the strains, including GH1, GH73, GH25, GH2, GH31, GH78, GH29, and GH43_26. Furthermore, some glycoside hydrolases were unique to certain strains, consisting of GH-154 and GH70 for *P. acidilactici* ZPA017, and GH42 and GH5_22 for *P. acidilactici* DSM 19927. Additionally, the last

Table 2. Mobile genetic elements of the *P. acidilactici* genomes

Strain	Transposable elements	Prophage	Plasmid	CRISPR
ZPA017	123	3 (0 intact)	0	1
BCCI	133	4 (0 intact)	2	1
JQII-5	125	3 (1 intact)	0	1
PMC48	126	3 (1 intact)	0	1
CACC 537	96	1 (1 intact)	0	1
SRCM101189	133	2 (2 intact)	2	1
SRCM100424	134	2 (2 intact)	2	1
SRCM100313	129	2 (2 intact)	2	1
SRCM102732	117	3 (2 intact)	1	0
SRCM102731	119	3 (2 intact)	1	0
ATCC 8042	104	1 (1 intact)	0	1
SRCM103387	124	3 (3 intact)	1	1
SRCM103444	113	5 (0 intact)	0	1
PB22	99	1 (1 intact)	0	1
SRCM103367	116	1 (1 intact)	2	1
DSM 20284	95	1 (1 intact)	0	1
DSM 19927	127	1 (0 intact)	0	1
SRCM 103289	125	2 (2 intact)	0	1
AGR20	120	4 (1 intact)	0	1
JKY18	122	7 (2 intact)	0	0
FAM 18969	123	3 (1 intact)	0	1
S7	104	1 (1 intact)	0	1
AS1.2696	92	1 (1 intact)	0	0
MA18/5M	93	1 (1 intact)	0	1
UMNPBX20	108	2 (1 intact)	0	1
E24	123	2 (2 intact)	0	1
F7	128	3 (2 intact)	0	1
L8-A	121	3 (0 intact)	0	0
L2-A	118	2 (1 intact)	0	1
E1	128	2 (2 intact)	0	1
I32	130	2 (2 intact)	0	1
SRCM100320	111	2 (2 intact)	0	1
D3	138	1 (1 intact)	0	1
NBRC 12231	117	1 (1 intact)	0	0
GS1	98	1 (1 intact)	0	1
L14-B	131	2 (0 intact)	0	1
WT	98	1 (1 intact)	0	1
S1	91	2 (1 intact)	0	1
LPBC161	114	6 (0 intact)	0	0
M16	116	1 (1 intact)	0	0
K3	106	1 (1 intact)	0	0

twenty glycosyl hydrolases showed a diverse presence in all of the 41 strains. This analysis essentially confirmed a great abundance of members of the GH family 1, of which the number of members can be as high as 8 in some strains.

Generally, genes involved in the utilization of a given carbohydrate are frequently organized in gene clusters or sugar operons, consisting of genes that encode one or more specific GHs, are associated with the transport system, and are frequently being placed under the control of a transcriptional regulator. For example, the 6-phospho-beta-glucosidase, a member of the GH1 family, which was responsible for the metabolism of cellobiose. All the strains contain 6-phospho-beta-glucosidase, suggesting that they all were able to utilize cellobiose, which is consistent with the predicted results (Fig. 3A). The cellobiose utilization gene cluster consists of the PTS system, the cellobiose-specific IIC component (celB), the 6-phospho-beta-glucosidase (bglA), and the GntR family transcriptional regulator (ydhQ) (Supplementary data Fig. S2).

Beta-galactosidase [EC 3.2.1.23], belonging to the GH2 family, was responsible for the metabolism of D-galactose and lactose. All the 39 strains containing GH2 were able to utilize D-galactose and lactose (Fig. 3A and B). The lactose/galactose operon included LacI (LacI family of the transcriptional regulator), lacZ (beta-galactosidase), lactose, and galactose permease, galM (aldose 1-epimerase), galK (galactokinase), galE (UDP-glucose 4-epimerase), and galT (galactose-1-phosphate uridylyltransferase) (Supplementary data Fig. S2).

Another correlation existed between the presence of the GH13_29 family and D-trehalose utilization, with three exceptions among all the 41 strains, including *P. acidilactici* JQII-5, *P. acidilactici* PMC48, and *P. acidilactici* L8-A (Fig. 3A and B). The trehalose operon consisted of treP (PTS system trehalose-specific EIIBC component), treC (trehalose-6-phosphate hydrolase), and treR (Trehalose operon transcriptional repressor) (Supplementary data Fig. S2).

The pentose sugars arabinose and xylose represented the monosaccharides that are found in many plants derived polysaccharides. The GH3 family, representing beta-glucosidases and xylosidases (Rohman *et al.*, 2019), and the GH43 family consisting of GH43_11 and GH43_26, representing xylosidases (Rohman *et al.*, 2019) and arabinosidases (Yang *et al.*, 2014) were involved in the metabolism of xylose- and arabinose-containing glycans. All the strains, except AGR20, were able to utilize the xylose (Fig. 3A). This xylose locus included xylose isomerase (xylA), xylulokinase (xylB), D-xylose proton-symporter (xylT), xyloside transporter (xynT), and xylose-responsive transcription regulator (xylR) (Supplementary data Fig. S2).

All the 41 strains, except SRCM103387 and JKY18, were able to utilize the arabinose (Fig. 3A). 13 strains were only able to utilize L-arabinose, and 4 strains were only able to utilize D-arabinose. Notably, 22 strains were able to metabolize both L-arabinose and D-arabinose. 35 strains containing the araA gene, which encodes L-arabinose isomerase, were able to utilize L-arabinose. The L-arabinose operon was composed of the L-arabinose isomerase (araA), ribulokinase (araB), L-ribulose-5-phosphate 4-epimerase (araD), and arabinose-proton symporter (araE) (Supplementary data Fig. S2).

Prediction of pediocin and other bacteriocins

In total, three kinds of bacteriocin were identified in the 41 strains, namely pediocin PA-1, enterolysin A, and colicin-B, and the results showed that not all the 41 strains could synthesize bacteriocin (Fig. 2B).

Pediocin PA-1 was predicted in four strains, *P. acidilactici* D3, *P. acidilactici* L2-A, *P. acidilactici* L8-A, and *P. acidilactici* NBRC 12231. The gene clusters for pediocin PA-1 consisted of four contiguous genes (Supplementary data Fig. S3), namely pedD, pedC, pedB, and pedA, which encode pediocin PA-1 transport/processing ATP-binding protein, leucocin-A immunity protein, putative immunity protein, and a highly conserved precursor of pediocin PA-1, respectively.

Enterolysin A was also found in six strains, namely *P. aci-*

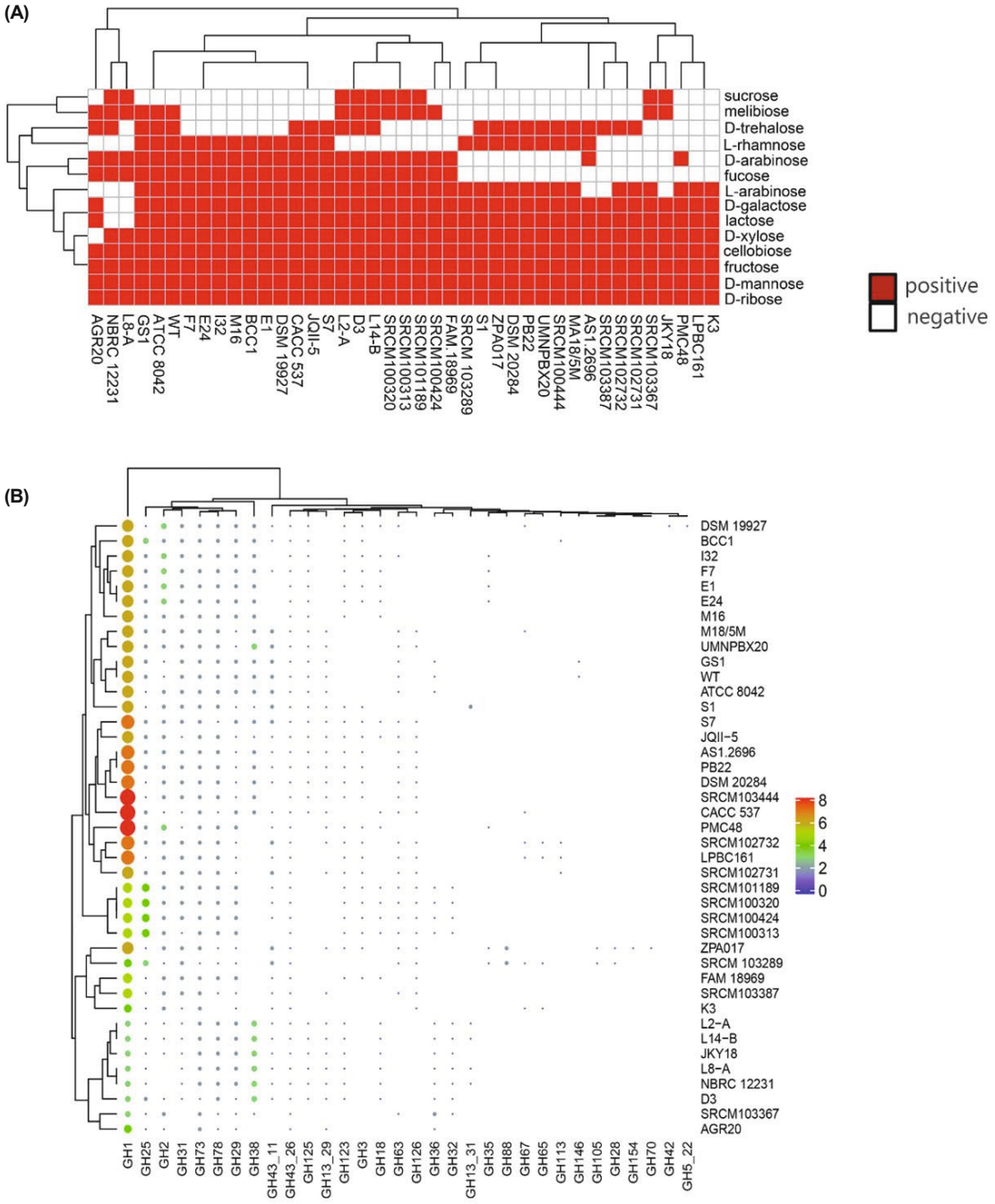

Fig. 3. Analysis of carbohydrate metabolism of *P. acidilactici*. (A) Heatmap of carbohydrate utilization by *P. acidilactici* strains. The horizontal axis represents the genomes tested in this study which were clustered according to the phylogenetic tree. The vertical axis represents the 14 carbohydrates used in this test. The color scales represent the degree of *P. acidilactici* utilizing carbohydrates with white squares (as in blank) representing the inability and red squares representing the ability of strains to metabolize the corresponding sugar. (B) Heatmap showing the distribution of GH families in the 41 *P. acidilactici* strains. The vertical axis represents the genomes that harbor GH families which was clustered according to the phylogenetic tree. The horizontal axis represents the GH families predicted in *P. acidilactici*. The color scales represent the number of GH families within each genome with white dots (as in blank) representing the absence of matches and red squares representing the highest number of GH families.

dilactici BCC1, *P. acidilactici* DSM 19927, *P. acidilactici* UMNPBX20, *P. acidilactici* FAM 18969, *P. acidilactici* SRCM-103387, and *P. acidilactici* ZPA017. And it was annotated as peptidase M23 in this study (Supplementary data Fig. S3). According to the distant homology annotation, these genes were all related to prophage genomes. In addition, by comparison with the phage database, gene clusters related to enterolysin A were homologous to one hypothetical protein predicted in the *Lactobacillus* prophage Lj928 (NC_005354).

Moreover, colicin-B was identified in *P. acidilactici* I32. The gene clusters for colicin-B consisted of cba (colicin-B activity protein), cbi (colicin-B immunity protein), and lysis

protein (Supplementary data Fig. S3). Colicin-B was usually associated with recombinases (e.g., xerD, tyrosine recombinase) and transposases (Cameron *et al.*, 2019). In this study, by comparison with the NCBI nr database, it was found that the contigs containing colicin-B had the highest identity with plasmids either in *Klebsiella pneumoniae* or in *Salmonella enterica*. Thus, the contig sequences containing colicin-B described here are likely plasmid-borne and recombined into the *P. acidilactici* I32 genome.

The predicted mobilome of *P. acidilaciti* species

All 41 *P. acidilaciti* genome sequences were investigated for the presence of mobile genetic elements such as transposable elements, prophage, plasmid, and CRISPR locus.

The analysis of transposable elements, using the RepeatMasker software, revealed that these elements accounted for less than 1% of each *P. acidilaciti* genome. Among them, the *P. acidilaciti* D3 genome encompassed the largest number (i.e. 138) of such mobile elements (Table 2). By classifying the transposon elements in 41 *P. acidilactici* genomes, it was found that LTR elements accounted for the highest proportion, followed by DNA transposon, LINEs and SINEs (Supplementary data Fig. S4).

The 41 *P. acidilaciti* genomes were also examined for the presence of prophages and plasmids. Notably, our analysis identified 90 prophage-like elements, of which 50 prophage-like elements that appeared to be complete (Table 2), while the following five strains only had incomplete prophage-like elements, *P. acidilaciti* SRCM103444, *P. acidilaciti* DSM 19927, *P. acidilaciti* L8-A, *P. acidilaciti* L14-B, and *P. acidilaciti* LPBC161. In addition, the results also showed that only eight complete *P. acidilaciti* genomes consisted of integrated plasmids (Table 2).

Interestingly, all 41 *P. acidilaciti* genomes were predicted to possess the CRISPR locus, which can make bacteria resistant to infection by foreign DNA such as prophages and plasmids (Horvath and Barrangou, 2010). However, in the prediction process of CRISPRCasFinder, CRISPR varied in evidence level, and only those whose evidence level was above one were considered in this study (Couvin *et al.*, 2018). In total, 32 strains contained complete CRISPR/CAS systems and all of those systems belonged to Type II-A, as they contained Cas1, Cas2, Cas9, and Csn2 (Supplementary data Fig. S5C), while the remaining nine strains were unable to silence foreign DNA, due to the absence of Cas protein (Table 2).

The Direct Repeats (DR) sequence in the CRISPR locus has a palindrome sequence, which is highly conserved, can be transcribed, and form a stem-loop structure. To confirm which CRISPR repeat family the DR sequences in *P. acidilaciti* belong to, multiple sequence alignments were performed for DR sequences from *P. acidilaciti*, the results showed that all the DR sequences from *P. acidilaciti* belonged to the Lsal1 family. In addition, the bioinformatics method was used to genetically analyze the DR sequence of the CRISPR region, the results indicated that all the DR sequences within the 32 CRISPR locus showed high similarity (Supplementary data Fig. S5D). By predicting the secondary structure of the DR RNA sequences, it was found that all the DR sequences were divided into two groups, non-circular and circular, respectively (Supplementary data Fig. S5A and B). The secondary structure of circular DR sequences was a large loop without a stem. On the other hand, RNA secondary structures of non-circular DR sequences contained two rings at both ends with a stem in the middle, of which the MFEs were -0.78 kcal/mol and was larger than the MFE of circular DR sequences (Supplementary data Table S2). Noticeably, on the stem, RNA secondary structures of non-circular DR sequences also had a small loop and G:U base pairs, which is typical of conserved RNA secondary structures and is crucial to the stem-loop in the repeats for the functionality of CRISPRs.

Evolutionary gene gain and loss analysis

In order to identify genes that may have been acquired by horizontal genes transfer (HGT), the genomes of the 41 *P. acidilaciti* strains were analyzed with the software suite COLOMBO v4.0 (Waack *et al.*, 2006). Each *P. acidilaciti* genome consisted of a different proportion of alien genes, ranging from 20% in *P. acidilaciti* M16 to 1.5% in *P. acidilaciti* S1. Interestingly, as can be seen from Fig. 1C, *P. acidilaciti* M16 had the most unique genes, this phenomenon indicated that HGT event was an important driving force for the evolution of *P. acidilaciti* species, which could make *P. acidilactici* species rapidly adapt to the ecological environment. In addition, these genes were annotated using the COG database and the results indicated that 18% of the genes that may have been acquired by HGT were predicted to be involved in carbohydrate metabolism and 11% were predicted to be involved in transcriptional regulation. In addition, 6% of the genes were involved in replication, recombination, and repair (Supplementary data Fig. S6).

Discussion

Pediococcus acidilactici, as a reliable bacteriocin producer, has gained much interest and can prevent the growth of pathogenic and spoilage bacteria during food fermentation and preservation (Parada *et al.*, 2007). In this study, 41 strains of *P. acidilactici*, previously isolated from different ecological niches, were analyzed using a comparative genomics approach to reveal genomic diversity and evolutionary relationship in this species.

It has been reported that various strains of *P. acidilactici* can produce bacteriocins. In this study, three kinds of bacteriocins were found in these strains, namely pediocin PA-1, enterolysin A, and colicin-B. *P. acidilactici* can produce pediocin PA-1, which has been reported countless times (Henderson *et al.*, 1992; Chikindas *et al.*, 1993; Nieto-Lozano *et al.*, 2010; Ueda *et al.*, 2018). Pediocin PA-1 is a 44 amino acid peptide with no posttranslational modifications (Henderson *et al.*, 1992; Lozano *et al.*, 1992). Its N-terminal region is particularly well conserved and contains a conserved "pediocin box" motif (YGNGVXCXK) (Nissen-Meyer and Nes, 1997) and its C-terminal disulfide bridge is a major determinant of the antimicrobial spectrum, which shows particularly strong activity against Listeria monocytogenes (Dabour *et al.*, 2009), a foodborne pathogen of special concern among food industries. It is worth mentioning that enterolysin A and colicin-B have not been reported in *P. acidilactici*

species before. The genes encoding enterolysin A were homologous with predicted prophage associated genes and the genes encoding colicin-B had the highest identity with plasmids in other strains. Therefore, it is possible that *P. acidilactici* seized the ability to produce enterolysin A and colicin-B from other species by the HGT event. This provides a new perspective for the study of evolutionary relationships among species.

Pan-genome analysis showed that *P. acidilactici* species had an open genome, which lead to its functional diversity. Generally speaking, the more widely distributed species usually have open pan-genomes that allow them to continuously acquire genetic material from the environment and adapt to harsh surrounding environments, such as *E. coli* (Fu and Qin, 2012), *B. cereus* and *S. pneumoniae* (Tettelin *et al.*, 2008; Bazinet, 2017). Meanwhile, comparative genomics analysis also identified variable genes in the *P. acidilactici* pan-genome, representing approximately 75% of the total gene content within the *P. acidilactici* species, which indicates a comparatively high level of diversity of this species (Medini *et al.*, 2005; Arboleya *et al.*, 2018). In addition, the number of predicted CDS features in each genome ranged from 776 to 1986 and this clear variability of gene content was evident in comparison across strains of the same species. It once again implied that there is an obvious genomic plasticity among *P. acidilactici,* living in different habits and possessing diverse properties (Marri *et al.*, 2006).

Functional annotation comparison analysis revealed that the variable genes of this species mostly encoded functions involved in carbohydrate transport and metabolism, defense mechanisms (i.e. CRISPER/Cas system), and replication, recombination and repair, which indicating these genes play an important role in adapting to different niches of *P. acidilactici*. In addition, some variable genes were functionally unknown. Such genes could be new genes with no homology in the distant homology databases, or pseudogenes with several primary amino acid sequence disruptions. Indeed, it has been argued that pseudogenes are pervasively distributed in prokaryotes and that a great portion of them derives from "failed" HGT events (Liu *et al.*, 2004). Genes acquired through HGT events may provide no advantage (due to non-efficient expression), or even be detrimental to the host (i.e. ORFs from infectious genetic elements such as a virus), and represent a minor force in the evolution of genomes of the *P. acidilaciti* species.

In the evolutionary analysis of *P. acidilactici*, Pan-genome tree is clustered based on the gene content of each genome, which shows the evolutionary relationship of *P. acidilactici*. While FGD tree does not attempt to represent the evolutionary path a genome has taken, since different genes will have been acquired by different routes. Instead, FGD tree investigates the level of collinearity between microbial genomes based on amino-acid sequences of the predicted complete ORFeomes. In our study, FGD tree was largely different from the phylogenetic clusters identified in the pan-genome tree, which indicating that a large number of gene rearrangement events occurred in the genome of *P. acidilactici*.

The predicted mobilome of *P. acidilactici* highlights a certain proportion of genetic elements, such as transposon elements, prophage-like elements and plasmids. Although these genetic elements only occupy a small part of the bacterial genome, their occasional transposition to other sites on the genome, sometimes coupled with major DNA rearrangements such as DNA inversion, deletion formation, partial duplication, and replicon fusion, represents an important contribution to genetic plasticity and bacterial evolution (Arber, 1991). Such foreign DNA also appears to explain the presence of the CRISPR-Cas system in variable regions of the *P. acidilactici*. CRISPR in combination with Cas constitutes CRISPR-Cas systems, which provides adaptive immunity against invasive elements in bacteria (Barrangou *et al.*, 2007). We observed that 32 strains of all the 41 strains contained type II-A CRISPR/Cas. The DR sequences in the CRISPR locus showed high similarity and they all belonged to the Lsal1 family, which suggests that there is a close relationship between *Pediococcus* and *Lactobacillus*, corresponding to the view that horizontal transfer dominates the evolution of the CRISPR/Cas locus (Horvath *et al.*, 2009; Makarova *et al.*, 2015).

The comparative genomics approach used in this study also facilitated the explanation of certain differences previously observed in the carbon sources utilization of *P. acidilactici* members. The carbohydrate utilization ability of bacteria is an important indicator of strain functionality and lays a foundation for further cultivation and selection of the strain (Drissi *et al.*, 2014). For this reason, the prediction of carbohydrate utilization was performed. The results showed that all 41 strains could ferment fructose, D-cellobiose, D-mannose, and D-ribose. Additionally, each strain showed a great abundance of GH families and contained different types and numbers of GH families, of which the GH1 family were included in all 41 strains.

In this study, most strains could utilize plant-derived polysaccharides, such as cellobiose, D-xylose, and arabinose, which implied that *P. acidilactici* can be used as an intestinal probiotic supplement to digest dietary polysaccharides that are not degraded by the host (Flint *et al.*, 2008). In addition, 39 strains containing beta-galactosidase were able to utilize lactose and galactose, except for *P. acidilactici* L8-A and *P. acidilactici* NBRC 12231, which is contradictory to the view that lactose-positive strains within the naturally occurring *Pediococcus* genus are absent (Semjonovs and Zikmanis, 2008). Hence, those strains might be acceptable starters for dairy fermentation. In the current study, five putative operons related to the utilization of lactose/galactose, trehalose, D-cellobiose, L-arabinose, and D-xylose were analyzed, in which three glycosyl hydrolases and two isomerases were identified to be the key enzymes in the corresponding operons (O'Donnell *et al.*, 2011). Moreover, the degradation of D-arabinose and L-fucose have very close correlations in our study, due to the sharing of the same enzymes in the process of degradation. Such a phenomenon has been reported before in *Escherichia coli* K-12 and *Escherichia coli* B (LeBlanc and Mortlock, 1971; Elsinghorst and Mortlock, 1988). This suggests that bacteria often recruit the enzymes of the L-fucose pathway by a regulatory mutation to metabolize the uncommon pentose D-arabinose (Elsinghorst and Mortlock, 1994).

Overall, in this study, the genomes of 41 *P. acidilactici* were analyzed through a comparative genomics approach. The results showed that *P. acidilactici* is a non-niche-specific species. *P. acidilactici* species showed a high level of genetic diver-

sity, which was mainly due to an open pan-genome, horizontal transfer events and mobile elements. Furthermore, in the process of adaptation to the ecological environment, the carbohydrate metabolic capacity and immune protection mechanism (CRISPR/Cas locus) of *P. acidilactici* species gradually had certain differences. The comprehensive genetic analyses conducted in this study allow for a deeper understanding of the evolutionary relationship and genetic diversity of *P. acidilactici*, facilitating its biotechnological potential for the dairy industry and therapeutics for microbiota-related diseases.

Acknowledgements

We thank all members of the Guangdong Provincial Key Laboratory of core colletion of corp genetic resources research and application (NO.2011A091000047). This work was supported by the projects subsidized by special funds for science technology innovation and industrial development of Shenzhen Dapeng New District (Grand No. KJYF-202001-10).

Conflict of Interest

We have no conflicts of interest to report.

References

Abu-Taraboush, H., Al-Dagal, M., and Al-Royli, M. 1998. Growth, viability, and proteolytic activity of Bifidobacteria in whole camel milk. *J. Dairy Sci.* **81**, 354–361.

Altermann, E. 2012. Tracing lifestyle adaptation in prokaryotic genomes. *Front. Microbiol.* **3**, 48.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Anastasiadou, S., Papagianni, M., Filiousis, G., Ambrosiadis, I., and Koidis, P. 2008. Pediocin SA-1, an antimicrobial peptide from *Pediococcus acidilactici* NRRL B5627: Production conditions, purification and characterization. *Bioresour. Technol.* **99**, 5384–5390.

Arber, W. 1991. Elements in microbial evolution. *J. Mol. Evol.* **33**, 4–12.

Arboleya, S., Bottacini, F., O'Connell-Motherway, M., Ryan, C.A., Ross, R.P., Van Sinderen, D., and Stanton, C. 2018. Gene-trait matching across the *Bifidobacterium longum* pan-genome reveals considerable diversity in carbohydrate catabolism among human infant strains. *BMC Genomics* **19**, 33.

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. 2016. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., *et al.* 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**, 75.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712.

Bazinet, A.L. 2017. Pan-genome and phylogeny of *Bacillus cereus sensu lato*. *BMC Evol. Biol.* **17**, 176.

Biswas, S., Ray, P., Johnson, M., and Ray, B. 1991. Influence of growth conditions on the production of a bacteriocin, pediocin AcH, by *Pediococcus acidilactici* H. *Appl. Environ. Microbiol.* **57**, 1265–1267.

Burke, G.R. and Moran, N.A. 2011. Massive genomic decay in *Serratia symbiotica*, a recently evolved symbiont of aphids. *Genome Biol. Evol.* **3**, 195–208.

Cameron, A., Zaheer, R., Adator, E.H., Barbieri, R., Reuter, T., and McAllister, T.A. 2019. Bacteriocin occurrence and activity in *Escherichia coli* isolated from bovines and wastewater. *Toxins* **11**, 475.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., and Parkhill, J. 2005. ACT: the Artemis comparison tool. *Bioinformatics* **21**, 3422–3423.

Chikindas, M.L., García-Garcerá, M.J., Driessen, A., Ledeboer, A.M., Nissen-Meyer, J., Nes, I.F., Abee, T., Konings, W.N., and Venema, G. 1993. Pediocin PA-1, a bacteriocin from *Pediococcus acidilactici* PAC1. 0, forms hydrophilic pores in the cytoplasmic membrane of target cells. *Appl. Environ. Microbiol.* **59**, 3577–3584.

Cintas, L., Casaus, P., Fernández, M., and Hernández, P. 1998. Comparative antimicrobial activity of enterocin L50, pediocin PA-1, nisin A and lactocin S against spoilage and foodborne pathogenic bacteria. *Food Microbiol.* **15**, 289–298.

Cintas, L.M., Rodriguez, J.M., Fernandez, M.F., Sletten, K., Nes, I.F., Hernandez, P.E., and Holo, H. 1995. Isolation and characterization of pediocin L50, a new bacteriocin from *Pediococcus acidilactici* with a broad inhibitory spectrum. *Appl. Environ. Microbiol.* **61**, 2643–2648.

Contreras-Moreira, B. and Vinuesa, P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl. Environ. Microbiol.* **79**, 7696–7701.

Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B., Rocha, E.P., Vergnaud, G., Gautheret, D., and Pourcel, C. 2018. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res.* **46**, W246–W251.

Dabour, N., Zihler, A., Kheadr, E., Lacroix, C., and Fliss, I. 2009. *In vivo* study on the effectiveness of pediocin PA-1 and *Pediococcus acidilactici* UL5 at inhibiting *Listeria monocytogenes*. *Int. J. Food Microbiol.* **133**, 225–233.

Drissi, F., Merhej, V., Angelakis, E., El Kaoutari, A., Carrière, F., Henrissat, B., and Raoult, D. 2014. Comparative genomics analysis of *Lactobacillus* species associated with weight gain or weight protection. *Nutr. Diabetes* **4**, e109.

Elsinghorst, E.A. and Mortlock, R.P. 1988. D-Arabinose metabolism in *Escherichia coli* B: induction and cotransductional mapping of the L-fucose-D-arabinose pathway enzymes. *J. Bacteriol.* **170**, 5423–5432.

Elsinghorst, E.A. and Mortlock, R.P. 1994. Molecular cloning of the *Escherichia coli* B L-fucose-D-arabinose gene cluster. *J. Bacteriol.* **176**, 7223–7232.

Feng, J., Wang, L., Zhou, L., Yang, X., and Zhao, X. 2016. Using *in vitro* immunomodulatory properties of lactic acid bacteria for selection of probiotics against *Salmonella* infection in broiler chicks. *PLoS ONE* **11**, e0147630.

Flint, H.J., Bayer, E.A., Rincon, M.T., Lamed, R., and White, B.A. 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat. Rev. Microbiol.* **6**, 121–131.

Fu, J. and Qin, Q. 2012. Analysis of pan-genomic characteristics of 30 strains of *E. coli*. *Genetic* **34**, 765–772.

Fuller, R. 1992. Probiotics: The scientific basis. Chapman & Hall,

London, United Kingdom.
Goldin, B.R. and Gorbach, S.L. 1992. Probiotics for humans. *In* Fuller, R. (ed.) Probiotics. Springer, Dordrecht, Germany.
Henderson, J.T., Chopko, A.L., and Van Wassenaar, P.D. 1992. Purification and primary structure of pediocin PA-1 produced by *Pediococcus acidilactici* PAC-1.0. *Arch. Biochem. Biophys.* **295**, 5–12.
Holzapfel, W.H., Franz, C.M.A.P., Ludwig, W., Back, W., and Dicks, L.M.T. 2006. The genera *Pediococcus* and *Tetragenococcus*. *In* Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., and Stackebrandt, E. (eds.), The Prokaryotes, pp. 229–266. Springer, New York, USA.
Horvath, P. and Barrangou, R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170.
Horvath, P., Coûté-Monvoisin, A.C., Romero, D.A., Boyaval, P., Fremaux, C., and Barrangou, R. 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food Microbiol.* **131**, 62–70.
Jiang, J., Yang, B., Ross, R.P., Stanton, C., and Chen, W. 2020. Comparative genomics of *Pediococcus pentosaceus* isolated from different niches reveals genetic diversity in carbohydrate metabolism and immune system. *Front. Microbiol.* **11**, 253.
Karp, P.D., Latendresse, M., Paley, S.M., Krummenacker, M., Ong, Q.D., Billington, R., Kothari, A., Weaver, D., Lee, T., Subhraveti, P., *et al.* 2016. Pathway tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief. Bioinform.* **17**, 877–890.
Kelly, W.J., Cookson, A.L., Altermann, E., Lambie, S.C., Perry, R., Teh, K.H., Otter, D.E., Shapiro, N., Woyke, T., and Leahy, S.C. 2016. Genomic analysis of three *Bifidobacterium* species isolated from the calf gastrointestinal tract. *Sci. Rep.* **6**, 30768.
Kolde, R. and Kolde, M.R. 2015. Package 'pheatmap'. *R package* **1**, 790.
Komora, N., Maciel, C., Pinto, C.A., Ferreira, V., Brandão, T.R., Saraiva, J.M., Castro, S.M., and Teixeira, P. 2020. Non-thermal approach to *Listeria monocytogenes* inactivation in milk: The combined effect of high pressure, pediocin PA-1 and bacteriophage P100. *Food Microbiol.* **86**, 103315.
LeBlanc, D.J. and Mortlock, R.P. 1971. Metabolism of D-arabinose: a new pathway in *Escherichia coli*. *J. Bacteriol.* **106**, 90–96.
Letunic, I. and Bork, P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245.
Liu, Y., Harrison, P.M., Kunin, V., and Gerstein, M. 2004. Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol.* **5**, R64.
Lozano, J.C.N., Meyer, J.N., Sletten, K., Peláz, C., and Nes, I.F. 1992. Purification and amino acid sequence of a bacteriocin produced by *Pediococcus acidilactici*. *Microbiology* **138**, 1985–1990.
Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., *et al.* 2015. An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736.
Marri, P.R., Hao, W., and Golding, G.B. 2006. Gene gain and gene loss in *Streptococcus*: is it driven by habitat? *Mol. Biol. Evol.* **23**, 2379–2391.
Mäyra-Mäkinen, A. and Bigret, M. 2004. Industrial use and production of lactic acid bacteria. *In* Salminen, S. and von Wright, A. (eds.), Lactic acid bacteria, Chapter 5. CRC Press, New York, USA.
Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. 2005. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589–594.
Nieto-Lozano, J.C., Reguera-Useros, J.I., Peláez-Martínez, M.d.C., Sacristán-Pérez-Minayo, G., Gutiérrez-Fernández, Á.J., and de la Torre, A.H. 2010. The effect of the pediocin PA-1 produced by *Pediococcus acidilactici* against *Listeria monocytogenes* and *Clostridium perfringens* in Spanish dry-fermented sausages and frankfurters. *Food Control* **21**, 679–685.
Nissen-Meyer, J. and Nes, I.F. 1997. Ribosomally synthesized antimicrobial peptides: their function, structure, biogenesis, and mechanism of action. *Arch. Microbiol.* **167**, 67–77.
O'Donnell, M.M., Forde, B.M., Neville, B., Ross, P.R., and O'Toole, P.W. 2011. Carbohydrate catabolic flexibility in the mammalian intestinal commensal *Lactobacillus ruminis* revealed by fermentation studies aligned to genome annotations. *Microb. Cell Fact.* **10**, S12.
Olszewska, M. and Staniewski, B. 2012. Cell viability of *Bifidobacterium lactis* strain in long-term storage butter assessed with the plate count and fluorescence techniques. *Czech J. Food Sci.* **30**, 421–428.
Parada, J.L., Caron, C.R., Medeiros, A.B.P., and Soccol, C.R. 2007. Bacteriocins from lactic acid bacteria: purification, properties and use as biopreservatives. *Braz. Arch. Biol. Technol.* **50**, 512–542.
Porto, M.C.W., Kuniyoshi, T.M., Azevedo, P.O.S., Vitolo, M., and Oliveira, R.P.S. 2017. *Pediococcus* spp.: an important genus of lactic acid bacteria and pediocin producers. *Biotechnol. Adv.* **35**, 361–374.
Rohman, A., Dijkstra, B.W., and Puspaningsih, N.N.T. 2019. β-Xylosidases: structural diversity, catalytic mechanism, and inhibition by monosaccharides. *Int. J. Mol. Sci.* **20**, 5524.
Semjonovs, P. and Zikmanis, P. 2008. Evaluation of novel lactose–positive and exopolysaccharide-producing strain of *Pediococcus pentosaceus* for fermented foods. *Eur. Food Res. Technol.* **227**, 851–856.
Shah, A.A., Yuan, X., Khan, R.U., and Shao, T. 2018. Effect of lactic acid bacteria-treated King grass silage on the performance traits and serum metabolites in New Zealand white rabbits (*Oryctolagus cuniculus*). *J. Anim. Physiol. Anim. Nutr.* **102**, e902–e908.
Stiles, M.E. and Holzapfel, W.H. 1997. Lactic acid bacteria of foods and their current taxonomy. *Int. J. Food Microbiol.* **36**, 1–29.
Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.
Tarailo-Graovac, M. and Chen, N. 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **25**, 4.10.1–4.110.14.
Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36.
Tettelin, H., Riley, D., Cattuto, C., and Medini, D. 2008. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**, 472–477.
Ueda, T., Tategaki, A., Hamada, K., Kishida, H., Nakagawa, K., Hosoe, K., Morikawa, H., and Nakagawa, K. 2018. Effects of *Pediococcus acidilactici* R037 on serum triglyceride levels in mice and rats after oral administration. *J. Nutr. Sci. Vitaminol.* **64**, 41–47.
van Heel, A.J., de Jong, A., Song, C., Viel, J.H., Kok, J., and Kuipers, O.P.J.N.a.r. 2018. BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Res.* **46**, W278–W281.
Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W.F., Surovcik, K., Meinicke, P., and Merkl, R. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics* **7**, 142.
Yang, X., Shi, P., Huang, H., Luo, H., Wang, Y., Zhang, W., and Yao, B. 2014. Two xylose-tolerant GH43 bifunctional β-xylosidase/α-arabinosidases and one GH11 xylanase from *Humicola insolens* and their synergy in the degradation of xylan. *Food Chem.* **148**, 381–387.