Journal Pre-proof

Comparative genomic analyses of *Lactobacillus rhamnosus* isolated from Chinese subjects

Dan Huang, Bo Yang, Yang Chen, Catherine Stanton, R. Paul Ross, Jianxin Zhao, Hao Zhang, Wei Chen

PII: S2212-4292(19)30277-9

DOI: https://doi.org/10.1016/j.fbio.2020.100659

Reference: FBIO 100659

To appear in: *Food Bioscience*

Received Date: 22 March 2019

Revised Date: 5 May 2020

Accepted Date: 23 May 2020

Please cite this article as: Huang D., Yang B., Chen Y., Stanton C., Ross R.P., Zhao J., Zhang H. & Chen W., Comparative genomic analyses of *Lactobacillus rhamnosus* isolated from Chinese subjects, *Food Bioscience* (2020), doi: https://doi.org/10.1016/j.fbio.2020.100659.

Author Contribution Statement

Conceptualization, B.Y. and W.C.; Data curation, D.H. and Y.C.; Formal analysis, D.H. B.Y. and J.Z.; Funding acquisition, W.C.; Investigation, C.S. and P.R.; Methodology, D.H., Y.C. and J.Z.; Project administration, H.Z. and W.C.; Resources, B.Y. and H.Z.; Software, D.H. and H.Z.; Supervision, H.Z. and W.C.; Validation, B.Y.; Visualization, B.Y.; Roles/Writing – original draft, D.H., and B.Y.; Writing – review & editing, C.S., P.R. and W.C.

Comparative genomic analyses of *Lactobacillus rhamnosus* isolated from Chinese subjects

Dan Huang [1,2], Bo Yang [1,2,4], Yang Chen [1,2], Catherine Stanton [4,5,6], R. Paul Ross [4,5,6], Jianxin Zhao [1,2,4], Hao Zhang [1,2,3,8], Wei Chen [1,2,3,7,*]

[1] State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu, China

[2] School of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu, China

[3] National Engineering Research Center for Functional Food, Jiangnan University, Wuxi, Jiangsu, China

[4] International Joint Research Center for Probiotics & Gut Health, Jiangnan University, Wuxi, Jiangsu, China

[5] Teagasc Food Research Centre, Moorepark, Fermoy, Cork, Ireland

[6] APC Microbiome Ireland, University College Cork, Cork, Ireland

[7] Beijing Innovation Center of Food Nutrition and Human Health, Beijing Technology and Business University (BTBU), Beijing, China

[8] Wuxi Translational Medicine Research Center and Jiangsu Translational Medicine Research Institute Wuxi Branch, Wuxi, Jiangsu, China

Running title: *L. rhamnosus* comparative genomics

*Correspondence: Wei Chen

School of Food Science and Technology, Jiangnan University, 1800 Lihu Avenue, Wuxi 214122, China.

E-mail: chenwei66@jiangnan.edu.cn

Tel/Fax: +86 510 8591 2155

Abstract

Lactobacillus rhamnosus has been found in many niches, including human intestine, vagina, mouth and dairy products. To intensively investigate the genomic diversity of this species, draft genomes of 70 *L. rhamnosus* strains isolated from different Chinese subjects were sequenced and further investigated. The pan-genome of *L. rhamnosus* was open. And gene-trait matching (GTM) was done to explore the carbohydrate utilization ability and antibiotic resistance, and to establish a pattern of gene existence/absence and growth/absence. There were no significant correlations between genetic diversity of the strains and the age or region of the donors. The current results extend the understanding of *L. rhamnosus,* which could be used as a reference for subsequent research as well as mining and application of the species.

Keywords: *Lactobacillus rhamnosus*; comparative genomics; carbohydrate utilization; antibiotics resistance

1. Introduction

Lactobacillus is the most diverse genus of lactic acid bacteria (LAB), and is usually found in fermented foods, feeds, oral and gastrointestinal tracts (Barrons and Tassone, 2008). *Lactobacillus* in the gut have an important role, for example, immunity regulation, cholesterol control, and gastrointestinal function improvement, on the intestinal micro-ecological balance and health of hosts (Damodharan et al., 2016; Reuman et al., 1986). In addition, *Lactobacillus* have an important role in food production and processing, such as fermented dairy products (Todorov, 2010).

L. rhamnosus is a Gram-positive bacterium that exists in a variety of niches, such as the human intestine and vagina (Pascual et al., 2008). *L. rhamnosus* GG is one of the best commercialized strain among *L. rhamnosus*. It has a strong resistance to gastrointestinal digestion and potential probiotic characteristics. Previous studies on *L. rhamnosus* focused on its functional benefits including various diarrhea treatments, immunity improvement, and respiratory infections prevention (Barbieri et al., 2017; Evans et al., 2016). Phenotype studies were carried out, such as acid tolerance, bile salt tolerance, carbohydrate utilization and adherence to epithelial cells (Ceapa et al., 2015; 2016; Succi et al., 2005; Tripathi et al., 2013).

To explore the metabolism, evolution and speciation, high-throughput methods (e.g., site sequence typing (MLST) and OmniLog (Biolog) phenotyping (Di Cagno et al., 2010) have been used to analyze genotype and phenotype (Bao et al., 2016). Comparative genomic analysis is a common tool in bioinformatics (Kant et al., 2011) that is able to identify the association between strains and their origins as well as to evaluate the gene distribution of specific species (Kelleher et al., 2017). It is significant for strain characteristics such as combining phenotypic profiles with strain-specific genetic diversity allows the assignment of unknown functions to specific genetic loci and to determine interactions with the hosts (Siezen et al., 2010). It also provides a new approach to effectively assess the diversity of strains. Genetic diversity of several species among the *Lactobacillus* genus had been carried out, such as *L. plantarum* (Duar et al., 2017; Martino et al., 2016), *L. casei* (Broadbent, 2012), *L. reuteri* (Zheng et al., 2015) and *L. salivarius* (Raftis et al., 2011). Current research on the *L. rhamnosus* genome was primarily concerned with the use of carbohydrates. Ceapa et al. (2015; 2016) predicted that *L. rhamnosus* had a number of carbohydrate transport and utilization genes, combined with phenotypic analysis, with which the adhesion-related functional genes were predicted. To investigate the genomic diversity of *L. rhamnosus*, Douillard et al. (2013b) sequenced and compared the genome

and phenotype validation of 100 *L. rhamnosus* strains from diffrent niches.

To further learn about the genetic diversity of *L. rhamnosus* from Chinese subjects, 70 *L. rhamnosus* strains were isolated, draft-genome sequenced and analyzed with comparative genomics approaches as well as gene-trait matching analysis for both carbohydrate utilization and antibiotic resistance based on the genotype-phenotype combination.

2. Methods

2.1 Isolation of strains

One hundred and twenty fecal samples were collected from different regions of China. The fecal samples were mixed with 30% sterile glycerin solution (China National Medicines Corp. Ltd., Beijing, China), stored temporarily at 4°C, and stored at -80°C within 48 h for a maximum of 8 wk. One g of each stool sample was blended with 9 ml sterile physiological saline (China National Medicines Corp. Ltd.) (Ingham, 1999). Serial dilution and plating were done in an anaerobic workstation (AW400TG, Electrotek Scientific Ltd., Shipley, West Yorkshire, UK). For selection of lactobacilli, 100 μl of diluent was plated on *Lactobacillus* selective agar (LBS) (China National Medicines Corp. Ltd.) (Ingham, 1999), and 50 U/ml nystatin (Sangon Biotech Co., Ltd., Shanghai, China). Agar plates were cultured in the anaerobic workstation flushed with 80% N_2, 10% CO_2 and 10% H_2 at 37°C for 72 h. For each sample, colonies on LBS plates were counted. Colonies were selected at random and re-streaked onto LBS agar for purity. The final pure culture was cultured in LBS at 37°C for 24 h and preserved in 30% glycerol (China National Medicines Corp. Ltd.) at -80°C (Bottacini et al., 2018).

DNA was extracted from each strain using the Rapid Bacterial Genomic DNA Isolation Kit (Sangon Biotech Co., Ltd.) according to the manufacturer's instructions. The identity of each putative *Lactobacillus* isolate was confirmed using 16S rRNA sequence analysis. A 1.5-kb 16S rRNA gene fragment was generated using bacterial universal primers (27F: 5'-AGA GTT TGA TCC TGG CTC AG-3'and 1492R: 5'-ACG GCT ACC TTG TTA CGA CTT-3'). Each PCR mixture (25 μl) contained 1.5 mM of $MgCl_2$ (Takara, Dalian, Liaoning, China), 20 mM of Tris-HCl (Takara), 50 mM of KCl (Takara), 200 μM of each deoxynucleoside triphosphate (Takara), 25 pmol of each of the two primers, 1 U of Taq DNA polymerase (Takara), and 50 ng of DNA template. Each PCR (T100^{TM} Thermal Cycle, BioRad, Hercules, California, USA) cycling program consisted of an initial denaturation step of 10 min at 95°C, followed by amplification for 35 cycles as follows: denaturation (30 s at 95°C), annealing (40 s at 58°C), and extension (1 min at 72°C). The PCR was completed with a single elongation step (5 min

at 72°C). PCR fragments were purified using the PCR purification kit (Sangon Biotech Co., Ltd.) according to the manufacturer's instructions and subsequently sequenced by BGI (Shenzhen, Guangdong, China). Strains were assigned to a particular species following comparison of the 16S rRNA sequences using the Genbank database (http://www.ncbi.nlm.nih.gov/BLAST/) to assign to a particular species.

2.2 Sequencing and draft genome assembly

Draft-genome sequencing of all the strains were done using the Illumina HiSeq PE150 platform (Beijing Novogene Bioinformatics Technology Co., Ltd., Beijing, China) and strains were sequenced to a coverage depth no less than 100 ×. A-tailed, ligated to paired-end adaptors and PCR amplified with a 350 bp insert were used for the library construction. The reads were assembled using SOAPdenovo software (https://omictools.com/soapdenovo-tool), the optimal Kmer value was selected to obtain the splicing sequence (Duranti et al., 2016; Tettelin et al., 2008), and local inner gaps were filled by using the software GapCloser (https://sourceforge.net/projects/soap-denovo2/files/GapCloser/) (Luo et al., 2012).

2.3 Genome features prediction

The G+C content and start codon of each genome were predicted with Glimmer 3.02 (http://ccb.jhu.edu/software/glimmer/index.shtml) (Delcher et al., 2007). Transfer RNA (tRNA) was identified using tRNAscan-SE 2.0 (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Eddy, 1997). Open Reading Frame (ORF) prediction using Glimmer3.02 and ORF were annotated using BLASTP analysis against the non-redundant protein databases created by BLASTP based on the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) (Hyatt et al., 2010). Genemark (https://www.genemarks.com/) was used to predict the genetic structure of the spliced sequence and to generate amino acid and nucleotide sequences. The amino acids sequences were subjected to gene annotation using a Swiss-Prot (https://www.uniprot.org/), COG (https://www.ncbi.nlm.nih.gov/COG/), and NR database (https://www.ncbi.nlm.nih.gov/protein) (Lugli et al., 2017).

2.4 Phylogenetic comparison

Cluster analysis was done based on orthologous genes of *L. rhamnosus* using Orthomcl-v2.0.9 software (http://orthomcl.org/common/downloads/software/v2.0/) (Kelleher et al., 2017). All the orthologous genes were extracted, then a phylogenetic tree was constructed using MAFFT alignment (https://mafft.cbrc.jp/alignment/server/) based on the orthologous genes and neighbor-joining (NJ) algorithm was used for evolutionary analysis (Mailund et al., 2006).

2.5 Pan-genome and core-genome analysis

To predict possible variation in the genome, the size of the pan-genome, core genome and unique genes were calculated. Based on the genomic sequence of all the strains, PGAP-1.2.1 (https://sourceforge.net/pro-jects/pgap/files/PGAP-1.2.1) was used for pan-genome calculations (Zhao et al., 2012). The core-genome was measured using the CD-HIT cluster analysis (http://weizhongli-lab.org/cd-hit/). Amino acids had a 50% pairwise identity and 0.7 length difference cut-off threshold (Harris et al., 2017). Then the Venn diagram was drawn to show the relationships among samples. In addition, the average nucleotide identity (ANI) values of each pair of genomes were calculated using ANI Perl (https://github.com/chjp/ANI/blob/master/ANI.pl) (Goris et al., 2007).

2.6 Genotype/phenotype association applied to carbohydrate metabolism

All the genomes were annotated using the HMM method in HMMER-3.1 (hmmer.org), and the enzymes involved in carbohydrate metabolism were analyzed using the carbohydrate-active enzymes (CAZY) database (http://www.cazy.org/) (Besemer et al., 2001; Lombard et al., 2014). Seven different carbohydrates including xylooligosaccharides (XOS), D-galactose, sucrose, D-trehalose, L-fucose, D-lactose and D-xylose (Sangon Biotech Co., Ltd.) were further selected for carbohydrate utilization analysis. A 10% (w/v) fresh aqueous solution of those carbohydrates were prepared and filtered through a 0.22 μm sterile membrane filter (Saigon Biotech Co., Ltd.). The utilization assay medium was freshly prepared with the same content as de Man, Rogosa and Sharpe (MRS) medium (China National Medicines Corp. Ltd.) (Ingham, 1999) except glucose, and two drops of 1.6% (w/v) bromcresol purple alcoholic solution (Sangon Biotech Co., Ltd.) were added, the latter as an indicator. After autoclaving and cooling, the sterile carbohydrate was added into the medium at 1% final concentration. To test the utilization capacity of each strain, after being sub-cultured twice in MRS medium, a 1% culture was inoculated into the growth media, each of which was supplemented with a different sugar instead of glucose. The utilization was observed as a color change and measured with a microplate reader at OD_{600nm} (Varioskan Lux, Thermo, Waltham, MA, USA) after anaerobic culture at 37°C for 24 h (Hyatt et al., 2010). The test were done in triplicate.

2.7 Antibiotic resistance gene and tolerance

The antibiotic-resistant genes were analyzed using the comprehensive antibiotic resistance database (CARD) (https://card.mcmaster.ca) (Jia et al., 2017), to obtain information of predicted antibiotic resistance genes encoded by each genome.

According to ISO10932:2010 standard "Milk and Milk Products-Determination of the Minimal Inhibitory Concentration (MIC) of Antibiotics Applicable to Bifidobacteria and Non-Enterococcal Lactic Acid Bacteria (LAB)" (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalo-gue_detail.htm?csnumber=46434)", the MIC of 6 antibiotics (streptomycin, erythromycin, clindamycin, chloramphenicol, tetracycline, and trimethoprim), purchased from Sangon Biotech Co., Ltd., were analyzed for all the strains. Microbial dilutions were incubated in an anaerobic environment at 37°C for 48 h, measured with a microplate reader (Thermo) at OD_{625nm}. And the threshold values for each antibiotic resistance issued by the European Food Safety Authority (EFSA) was used (FEEDAP, 2012).

3. Results

3.1 Isolation of strains

The donors came from 18 different regions in China and their ages varied from 2 d to 102 yr (35 samples of infants < 1 yr; 9 samples from 2-16 yr; 10 samples from 27-54 yr; 10 samples from 60-79 yr; 7 samples from 82-102 yr). Most of donors were located in three cities (Wuhu, Anhui Province; Wuxi, Jiangsu Province; Wusu, Xinjiang Uygur Autonomous Region). A total of 600 isolates were isolated from the LBS agar, and all of them were species identified with 16S rRNA sequencing. Among them, 168 isolates from 70 samples were confirmed as *L. rhamnosus*. Only one isolate from each sample was used for genome sequence to analyze the genetic features of the *L. rhamnosus* species (Table 1).

3.2 General genome features

The draft genome of 70 strains isolated from Chinese subjects were sequenced and compared with *L. rhamnosus* GG. For all strains, the genome size ranged from 2.77 Mb for FXJWS25L4 to 3.10 Mb for FAHWH35L1. The quantity of genomic ORF ranged from 2676 to 3024 with an average of 2840. And only three strains had tRNA genes <40 tRNA genes. Furthermore, the GC content differed from 46.6 to 47.5% for all strains (Table 1).

3.3 Phylogenetic analyses of *L. rhamnosus*

Based on orthologous genes, a phylogenetic tree was constructed to evaluate the evolution of the species. The results showed that the species consisted of 1870 orthologous genes, which were shared among all the sequenced strains. According to the clustering relationship, all the 71 strains assessed were divided into 5 clusters. Focused on the strains isolated from different regions, there was no

significant regional correlation in clustering, although the strains isolated from Xinjiang Uygur Autonomous Region were relatively clustered (Figure 1). In the age distribution, the strains from infants <1 yr were located in the same branch, but there was no significantly clustering relationship when the age was subdivided. Interestingly, two pairs of strains were evolutionarily in the same small branch, which were FHNFQ3L5 and FHNFQ4L1 isolated from a mother and daughter pair, and FQHXN3M6 and FQHXN4M2 isolated from the samples with a father-son relationship. The results indicated that the clustering of strains was independent of age and region of the donors and might be related to the family relationship or daily diet of people from whom the samples were obtained.

3.4 Pan-genome and core genome analysis of *L. rhamnosus*

The 71 genomes of *L. rhamnosus* strains were used to calculate the pan-genome, and the resultant pan-genome was 8395 (Figure 2A). The pan-genome curve showed an upward trend (Figure 2A), while the power exponent in the pan-genome formula was >0.5 (Bosi et al., 2016). These results indicated that the pan-genome would increase if more genomes were included. However, new genes gradually decreased from the initial 190 to the last group at 54.3.

Based on the clusters of orthologous genes (COGS), the core-genome of *L. rhamnosus* was evaluated. The core genome correlation curve decreased with the addition of genomes, and was gradually stable after the 50th genome was included (Figure 2A). Ultimately the core genome of *L. rhamnosus* was 1835. The major role of the core genes were diverse, including energy production and amino acid metabolism (Figure 2C). In addition, the number of specific genes in the strains ranged from 5 to 146 (Figure 2B). In terms of gene content, FBJSY31L2 has the highest number of specific genes, which was highly diverse compared to the other strains.

To analyze the *L. rhamnosus* species and potential subspecies, ANI analysis was carried out. The ANI value of 70 strains as well as *L. rhamnosus* GG was 97.2-99.9%. It was greater than the threshold of 95%, (Richter and Rossello-Mora, 2009) which proved they were the same species without any subspecies. The ANI values of the FHNFQ3L5 and FHNFQ4L1 and the FQHXN3M6 and FQHXN4M2 were > 99%, which indicated that those two pair of strains from one family, might have a closer genetic relationship. However, clustering based on ANI has no significantly correlation with age or region of the donors (Figure 3).

3.5 Genotype/phenotype association applied to carbohydrate metabolism

Carbohydrate-active enzymes were predicted using the CAZY database. The pan-genome of *L.*

rhamnosus contained genes encoding 27 glycosyl hydrolases (GH) families, 14 glycosyl transferases (GT) families, 8 carbohydrate esterases (CE) families, 8 carbohydrate binding modules (CBM) families, two polysaccharide lyases (PL) families and two activities (AA) families (Figure 4A). The heat-map of predicted GH family genes was constructed using HemI (Figure 4A), and the number of GH1, GH13, GH109 and GH25 genes was significantly greater than that of other members of the GH families. On the other hand, the number of genes for GH15, GH23, GH43, GH88, GH115 and GH126 were less than that of other members of the GH families. According to the evolutionary tree, the 71 strains tested were divided into 5 clusters. Some genes of the GH family had clustering differences in distribution, such as the GH115 family only existed with cluster-5, and the GH43 family existed in cluster-4 and -5.

The GH analysis showed the presence of the predicted α-trehalase (GH15, GH65), fucosidase (GH1, GH29, GH30), galactosidase (GH1, GH2, GH4, GH31, GH35, GH36, GH59), mannose glycosidase (GH2, GH31, GH38) and xylanase (GH3, GH31, GH39, GH43). Genotype-phenotype correlations were carried out to identify the characteristics of *L. rhamnosus*. Based on the prediction of metabolic related genes, the 7 types of sugars were selected for the *in vitro* utilization assay. All the strains could use and grow well with D-galactose, sucrose, D-trehalose and L-fucose as the sole carbon source. Whereas the utilization of the other four carbohydrates differed among the strains assessed. D-lactose as the carbon source could support the growth of 43 strains. Only four could use D-xylose, while 20 strains could use XOS (Figure 4B).

Based on the genotype and phenotypic analysis, all the strains utilized D-galactose, sucrose, D-trehalose and L-fucose, indicating that the genotype was consistent with the phenotypic results. However, for some strains, the genotype and phenotype was not consistent. For instance, the gene encoding xylosidase was found in all 71 strains, but only four could utilize D-xylose, which only had a 5.6% genotype-phenotype correlation. The presence of the predicted gene encoding lactase positively correlated with the glycometabolism phenotype for 43 strains (60.6% totally). Particularly, 28 strains having lactase in their genomes could not utilize D-lactose. For XOS, twenty strains could utilize D-xylose with a 28.2% genotype-phenotype correlation. There was no significant correlation between the carbohydrate utilization and the donor's age and region in both genotypes and phenotypes. The genes involved in D-lactose and L-fucose utilization were analyzed in detail for each strain and the results showed that the strains with the utilization ability had the whole related operons (Figure 4C).

3.6 Genotype/phenotype association of antibiotic resistance

The antibiotic resistant genes were analyzed using CARD, and based on the prediction, the heat map was constructed. These results showed that there were no resistance genes for gentamicin, streptomycin and neomycin in all 71 strains, but all of them contained the resistance gene for kanamycin, erythromycin, clindamycin, chloramphenicol, tetracycline, trimethoprim, ampicillin, ciprofloxacin, amoxicillin, vancomycin and rifampicin. The number of resistant genes for erythromycin, tetracycline and ciprofloxacin were significantly higher than for the other antibiotics. However, there was no significant difference in the number of genes of the same antibiotics among the strains. In addition, there was no significant correlation between the resistance gene number and the donor's age or region. The difference in the number of genes only reflected the differences among individual strains (Figure 5).

The MIC value of each strain was compared with the drug resistance threshold reported by EFSA to determine whether the strain was resistant or sensitive (FEEDAP et al., 2012). According to the genotype, six antibiotics were selected for the susceptibility test. Among the 6 antibiotics, no streptomycin resistance gene was predicted in the 71 strains, but the other five antibiotic resistance genes could be found among all the strains. However, all 71 strains showed different resistance to those antibiotics. The MIC of streptomycin for all the strains tested differed from 1 to 32 μg/ml and mostly were 2 to 8 μg/ml (Figure 6A). According to EFSA's guidance, all the strains were sensitive. The MIC of erythromycin for the strains ranged from 0.0625 to 8 μg/ml (Figure 6B), which indicated 69 strains were sensitive. In addition, only 4.23% of the strains were resistant to tetracycline (Figure 6C). And among the 71 strains, the resistance rate for clindamycin reached 9.8% (Figure 6D). Compared with the first four antibiotics, the overall resistance rate for chloramphenicol was 76.1% (Figure 6E). *L. rhamnosus* has a conserved resistance gene for trimethoprim, therefore, there was no accurate threshold. The range of MIC of trimethoprim was 8-64 μg/ml, in which 54 strains had a MIC value of 64 μg/ml (Figure 6F). These results showed that the MIC was not related to the subjects region and age, and was independent of the clustering of the strain.

All 71 strains were sensitive for streptomycin, which indicated that there was a certain correlation between genotype and phenotype. The resistance gene for erythromycin was observed in all 71 strains. However, only one strain was actually resistant to erythromycin, and the genotype-phenotype correlation rate was 2.8%. Three strains were resistant for tetracycline with 4.23% genotype-phenotype correlation. The resistance genes for clindamycin were observed in all 71 strains with only 9.8%

genotype-phenotype correlation. All the strains were resistant to trimethoprim with 100% genotype-phenotypic correlation.

4. Discussion

With sequencing technology improvements, more researches involving the genome of the strain. *L. rhamnosus* contained a 3.0 Mb genome, which was one of the largest for LAB (Lebeer et al., 2008). Several genomic sequences of *L. rhamnosus* have been identified to date, including human intestinal derived *L. rhamnosus* GG (ATCC 53103) (Hidetoshi et al., 2009), cheese-derived *L. rhamnosus* LC705 (Prisciandaro et al., 2011) and soil derived *L. rhamnosus* CASL (Yu et al., 2011). In the current study, 70 strains of *L. rhamnosus* were successfully isolated from different Chinese subjects from different regions, and their draft-genomes were sequenced to determine the genetic diversity and phylogenetic correlations within the species. The genomic characteristics of 70 Chinese *L. rhamnosus* strains were similar to previous literature (Kant et al., 2014). The GC content was significantly higher than that of *L. salivarius* and the average value of the *Lactobacilli* was ~42.4% (Harris et al., 2017).

Phylogenetic trees of all the 71 strains (including *L. rhamnosus* GG) based on orthologous genes showed no significant correlation between strain clustering and age or region of the donors (Odamaki et al., 2018). Furthermore, ANI is a classic index to distinguish whether particular strains belong to the same species, usually with a threshold of 95% as the species boundary (Sun et al., 2015). The ANI value of all the strains was >97%, and the similarity of most strains was as high as 99%. The clustering of strains was similar to that of phylogenetic trees established based on orthologous genes, but there were differences, which may be the result of different algorithms and the high similarity among different strains. From the phylogenetic trees constructed using the two methods, the clustering of *L. rhamnosus* strains had no significant correlation with age or region of donors. FHNFQ3L5 and FHNFQ4L1, isolated from two samples with a mother-daughter relationship, living in the same family and sharing the same daily diets, were evolutionarily in the same subgroup. And similar results were observed between FQHXN3M6 and FQHXN4M2, which were isolated from a father-son pair. The reduced differences of strains derived from one family indicated that similar diets and environments would decrease the genetic diversity of strains.

The pan-genome was mainly composed of core genes, accessory genes and specific genes, which represent the whole genome of a species (Borneman, 2012). The trend of the pan-genome in *L.*

rhamnosus was temporarily open, which was significantly higher than that of 13 strains of *L. rhamnosus* reported by Kant (Kant et al., 2014). When the strains increased, the pan-genome showed an upward trend. Core genes are in charge of the basic biological functions and the main phenotypic characteristics of the species (Inglin et al., 2018). The core-gene number in the current work was 1835, less than previous report, indicating when the number of strains increases, the core genes would decline. In the *Lactobacillus* genus, the size of core genes of *L. rhamnosus* was similar to that of *L. paracasei* (Smokvina et al., 2013), while higher than that of *L. salivarius* (Harris et al., 2017). Non-essential genes and specific genes are mainly involved in secondary metabolic pathways (Mols et al., 2010), and the number of specific genes of each strain was independent of the donor's region and age.

Lactobacillus has the great ability to ferment different carbohydrates and subsequently obtain metabolic energy (Ceapa et al., 2015; 2016). To analyze the carbohydrate fermentation capacity of *L. rhamnosus*, the carbohydrate-utilization related genes were predicted using the CAZY database, which contained GH, GT, PL, CE, CBM and AA (Lugli et al., 2017). GH are responsible for the hydrolysis (or modification) of carbohydrate glycoside bonds, which were mainly analyzed *in silico*. All the strains assessed contained α-trehalase (EC 3.2.1.28), xylosidase (EC 3.2.1.37), fucosidase (EC 3.2.1.38), fructan β-(2,1)-fructosidase/1-exohydrolase (EC 3.2.1.153), fructan dfructosidase/6-exohydrolase (EC 3.2.1.154), lactase (EC 3.2.1.108) and α-L-arabinofuranosidase (EC 3.2.1.55), which have important roles in sugar metabolism. The *in vitro* carbohydrate utilization ability of the 71 strains showed that all the strains could use D-galactose, sucrose, D-trehalose and L-fucose, while only 5.6% could use D-xylose, which were consistent with the results in Bergey's Manual of Systematics of Archaea and Bacteria (Hammes and Hertel, 2015). Invertase (EC 3.2.1.26), also known as sucrase or fructofuranosidase, is an important enzyme for the irreversible hydrolysis of sucrose to glucose and fructose (Yu et al., 2017). This enzyme and the corresponding PTS system were present in all the strains indicating that all of them could utilize sucrose. Lactose is hydrolyzed by β-galactosidase (EC 3.2.1.23) to give α-D-glucose and D-galactose, which are utilized by the corresponding PTS pathways. Another pathway for lactose hydration is catalyzed by lactose phosphotransferase (EC 2.7.1.207) and 6-phospho-beta-galactosidase (EC 3.2.1.85) with lactose-6-phosphate, galactose-6-phosphate, following the tagatose-6-phosphate pathway to further catabolism (Kankainen et al., 2009). All of the 71 strains tested contained galactosidase, of which 28 strains were not able to metabolize D-lactose, including the reference strain LGG, and the results were consistent with the previous literature

(Kankainen et al., 2009). All 71 strains had the genes for the lactose PTS pathway (*lac*FEG) and tagatose-6-phosphate pathway (*lac*ABCD), while further analysis indicated that antiterminator (*lac*T) and 6-phospho-β-galactosidase (*lac*G) genes were altered in some of those strains, resulting in the loss of lactose utilization capacity (Douillard et al., 2013a; Kankainen et al., 2009). All the strains in cluster-4 and -5 (Figure 4B) could use lactose and were far away from LGG in their evolutionary relationship, indicating that there was a certain relationship between gene expression and strain clustering. Another typical example is L-fucose, all 71 strains including LGG could use L-fucose. When the L-fucose operon was analyzed it showed that the whole L-fucose catabolic pathway consisting of α-L-fucosidase (*fuc*A, EC 3.2.1.51), *fucU*, *fucI* and *fucS* was present similar to that in LGG (Becerra et al., 2015; Douillard et al., 2013a). Carbohydrate metabolism is not only related to the related hydrolases, but to the corresponding metabolic pathways and transport proteins. For example, the metabolism of xylose is related to the intracellular pentose phosphate pathway (PPP) as well as β-xylosidase. Xylose isomerase could directly convert xylose to xylulose, which could be further phosphorylated by xylulose kinase to form 5-oxokinose-5-phosphate, which eventually enters the glycolysis pathway to produce alcohol by the intermediates 6-phosphate glucose and 3-phosphate glyceraldehyde (Yu et al., 2017). The genetic prediction results showed that the lack of xylose isomerase and xylulose kinase might result in no xylose entering the cell to participate in the PPP, consequently the strain could not metabolize xylose. It was confirmed that only four strains could use D-xylose *in vitro*, which might result from potential horizontal transfer of genes and was consistent with previous results that few strains could use D-xylose (Douillard et al., 2013b). XOS is a functional polymeric sugar composed of 2-7 xylose molecules bound by β-1,4 glycosidic bonds, usually requiring the ABC transport system to transfer to the intracellular space for hydrolysis. During the metabolism, α-L-arabinofuranosidase (EC 3.2.1.55) hydrolyzes the arabinose side chain in the XOS component, and hyxylosidase acts on the end of the XOS to release xylose (Andersen et al., 2013; Arboleya et al., 2018). All of the strains had an α-arabinofuranosidase and a xylosidase, possibly lacking the corresponding transport system. Therefore, 51 strains were unable to use XOS. In addition, the gene involved in glucose metabolism had no correlation with the age and regional origin of the donors, and a similar conclusion was reached for sugar fermentation *in vitro*. Instead, they were consistent with the phylogenetic tree classification.

Antibiotic resistance could be divided into natural resistance and acquired resistance. The innate

resistance gene inherent to the genus or species could be inherited (Campedelli et al., 2019; Karapetkov et al., 2011;). The 71 strains were innately resistant to vancomycin (data not shown) and had a related resistance gene, consistent with the results of Korhonen et al. (2010). D-alanine residues at the pentapeptide end of the cell wall of *L. rhamnosus* were replaced by D-lactic acid, which prevented the binding of vancomycin. Thus, the strains were resistant to vancomycin, and same resistance was observed in *L. casei*, *L. plantarum* and *L. acidophilus* (Hamilton-Miller and Shah, 1998; Korhonen et al., 2010). The resistance gene of trimethoprim is *dfr*, a dihydrofolate reductase encoded by the integron originally found in *Vibrio cholera*, which inhibited the synthesis of tetrahydrofolate and exerted antibacterial activity. A congenital resistance gene to trimethoprim was observed consistent with previous results in *L. casei* and *L. plantarum* (Roberts et al., 1981). Most of the strains were sensitive to macrolide antibiotics (erythromycin and tetracycline), and only two strains (FAHWH5L1, FHeNJZ4L2) were resistant to erythromycin. Among the 71 strains, the ATP-binding cassette (ABC) antibiotic efflux pump for multi-drugs was observed to prevent the resistance for macrolides and tetracyclines (Comunian et al., 2010). A reasonable explanation for FAHWH5L1 and FHeNJZ4L2 erythromycin resistance was that an A-to-G conversion point mutation on the 23S rRNA resulted in the binding of ribosomes to erythromycin or the A to G transition resulted in a lack of binding to the ribosome, making *L. rhamnosus* resistant to macrolide antibiotics (Begovic et al., 2009). All 71 strains of *L. rhamnosus* had the ABC-F ATP binding cassette ribosome protective protein and ABC antibiotic efflux pump involving chloramphenicol (Schwarz et al., 2004). Antibiotics are transported out of cells by efflux, which reduces intracellular antibiotic concentration and improves the drug resistance of the strains. Therefore, 76.1% of the strains were resistant to chloramphenicol. No target sites for the action of streptomycin antibiotics and drug efflux pumps were observed. In the streptomycin resistance test, none of the 71 strains were resistant, i.e., the genotype and phenotype were consistent. The genotype and phenotype were combined to analyze the resistance of the strain, and it was observed that the number of genes had no obvious relationship with the phenotypic results. In addition, genotypes and phenotypes were independent of the source of the strain, reflecting individual differences in the strain.

5. Conclusion

Seventy strains of *L. rhamnosus* were isolated from Chinese infants and adults and their draft-genome sequenced. Combined with LGG, the genome and biological properties of all the strains

were analyzed. The results showed that the pan-genome of *L. rhamnosus* was 8395 and core genome was 1835. All the strains could utilize D-galactose, sucrose, D-trehalose and L-fucose and only a few strains could use XOS and D-xylose, consistent with their genotype. For antibiotics resistance, most genotypic results of *L. rhamnosus* were consistent with phenotypic results. This study increased the genomic information and phenotypic study of *L. rhamnosus*, and provided reference value for future research on *L. rhamnosus*. The results will provide information on the meanings of the genetic diversity observed in this species, especially the carbohydrate utilization capabilities and antibiotic resistance, and can be applied to the further application of characterization of strains for probiotic additives.

Conflict of interest

All authors declared no conflict of interest.

Acknowledgement

This research was supported by the National Natural Science Foundation of China (Nos. 31771953, 31820103010), the National First-Class Discipline Program of Food Science and Technology (JUFSTR20180102), and the Collaborative Innovation Center of Food Safety and Quality Control in Jiangsu Province.

References

Andersen, J.M., Barrangou, R., Abou Hachem, M., Lahtinen, S.J., Goh, Y.J., Svensson, B., & Klaenhammer, T.R. (2013) Transcriptional analysis of oligosaccharide utilization by *Bifidobacterium lactis* Bl-04. *BMC Genomics*, 14, 312.

Arboleya, S., Bottacini F, O'Connell-Motherway, M., Ryan, C., Ross, R.P., van Sinderen, D., & 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(1), 33.

Bao, Q., Song, Y., Xu, H., Yu, J., Zhang, W., Menghe, B., Zhang, H., & Sun, Z. (2016). Multilocus sequence typing of *Lactobacillus casei* isolates from naturally fermented foods in China and Mongolia. *Journal of Dairy Science*, 99(7), 5202-5213.

Barbieri, N., Herrera, M., Salva, S., Villena, J., & Alvarez, S. (2017). *Lactobacillus rhamnosus* CRL1505 nasal administration improves recovery of T-cell mediated immunity against

pneumococcal infection in malnourished mice. *Beneficial Microbes*, 8(3), 393-405.

Barrons, R., & Tassone, D. (2008). Use of *Lactobacillus* probiotics for bacterial genitourinary infections in women: A review. *Clinical Therapeutics*, 30(3), 453-468.

Becerra, J.E., Yebra, M.J., & Monedero, V. (2015) An L-fucose operon in the probiotic *Lactobacillus rhamnosus* GG is involved in adaptation to gastrointestinal conditions. *Applied and Environmental Microbiology*, 81(11), 3880-3888.

Begovic, J., Huys, G., Mayo, B., D'Haene, K., Florez, A.B., Lozo, J., Kojic, M., Strahinic, I., & Topisirovic, L. (2009). Human vaginal *Lactobacillus rhamnosus* harbor mutation in 23S rRNA associated with erythromycin resistance. *Research in Microbiology*, 160(6), 421-426.

Besemer, J., Lomsadze, A., & Borodovsky, M. (2001). GeneMarkS: A self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, 29(12), 2607-2618.

Borneman, A.R. (2012). Comparative analysis of the *Oenococcus oeni* pan genome reveals genetic diversity in industrially-relevant pathways. *BMC Genomics*, 13, 373.

Bosi, E., Monk, J.M., Aziz, R.K., Fondia, M., Nizetd, V., & Palsson, B. (2016). Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, 13(26), e3801.

Bottacini, F., Morrissey, R., Esteban-Torres, M., James, K., van Breen, J., Dikareva, E., Egan, M., Lambert, J., van Limpt, K., Knol, J., O'Connell-Motherway, M. & van Sinderen, D. (2018). Comparative genomics and genotype-phenotype associations in *Bifidobacterium breve*. *Scientific Reports*, 8, 10633.

Campedelli, I., Mathur, H., Salvetti, E., Clarke, S., Rea, M.C., Torriani, S., & O'Toole, P.W. (2019). Genus-wide assessment of antibiotic resistance in *Lactobacillus* spp. *Applied and Environmental. Microbiology*, 85(1), e01738-18.

Broadbent, J.R. (2012). Analysis of the *Lactobacillus casei* supragenome and its influence in species evolution and lifestyle adaptation. *BMC Genomics*, 13, 533.

Ceapa, C., Lambert, J., van Limpt, K., Wels, M., Smokvina, T., Knol, J., & Kleerebezem, M. (2015). Correlation of *Lactobacillus rhamnosus* genotypes and carbohydrate utilization signatures determined by phenotype profiling. *Applied and Environmental Microbiology*, 81(16),

5458-5470.

Ceapa, C., Davids, M., Ritari, J., Lambert, J., Wels, M., Douillard, F. P., & Kleerebezem, M. (2016). The variable regions of *Lactobacillus rhamnosus* genomes reveal the dynamic evolution of metabolic and host-adaptation repertoires. *Genome Biology and Evolution*, 8(6), 1889-1905.

Comunian, R., Daga, E., Dupré, I., Paba, A., Devirgiliis, C., Piccioni, V., Perozzi, G., Zonenschain, D., Rebecchi, A., & Morelli, L. (2010). Susceptibility to tetracycline and erythromycin of *Lactobacillus paracasei* strains isolated from traditional Italian fermented foods. *International Journal of Food Microbiology*, 138(1-2), 151-156.

Damodharan, K., Palaniyandi, S.A, Yang, S. H, & Suh, J.W. (2016). Functional probiotic characterization and *in vivo* cholesterol-lowering activity of *Lactobacillus helveticus* isolated from fermented cow milk. *Journal of Microbiology and Biotechnology*, 26(10), 1675-1686.

Delcher, A., Bratke, K., Powers, E., & Salzberg, S. (2007). Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics*, 23(6), 673-679.

Di Cagno, R., Minervini, G., Sgarbi, E., Lazzi, C., Bernini, V., Neviani, E., & Gobbetti, M. (2010). Comparison of phenotypic (biolog system) and genotypic (random amplified polymorphic DNA-polymerase chain reaction, RAPD-PCR, and amplified fragment length polymorphism, AFLP) methods for typing *Lactobacillus plantarum* isolates from raw vegetables and fruits. *International Journal of Food Microbiology*, 143(3), 246-253.

Douillard, F.P., Ribbera, A., Järvinen, H.M., Kant, R., Pietilä, T.E., Randazzo, C., Paulin, L., Laine, P.K., Caggia, C., & von Ossowski, I. (2013a). Comparative genomic and functional analysis of *Lactobacillus casei* and *Lactobacillus rhamnosus* strains marketed as probiotics. *Applied and Environmental Microbiology*, 79(6), 1923-1933.

Douillard, F.P., Ribbera, A., Kant, R., Pietila, T.E., Jarvinen, H.M., Messing, M., Randazzo, C.L., Paulin, L., Laine, P., Ritari, J., Caggia, C., Lahteinen, T., Brouns, S.J., Satokari, R., von Ossowski, I., Reunanen, J., Palva, A., & de Vos, W.M. (2013b). Comparative genomic and functional analysis of 100 *Lactobacillus rhamnosus* strains and their comparison with strain GG. *PLoS Genetics*, 9(8), e1003683.

Duranti, S., Milani, C., Lugli, G.A., Mancabelli, L., Turroni, F., Ferrario, C., Mangifesta, M., Viappiani, A., Sanchez, B., Margolles, A., van Sinderen, D., & Ventura, M. (2016). Evaluation of genetic diversity among strains of the human gut commensal *Bifidobacterium adolescentis*. *Scientific*

Reports, 6, 23971.

Duar, R. M., Lin, X. B., Zheng, J., Martino, M. E., Grenier, T., Pérez-Muñoz, M. E., & Walter, J. (2017). Lifestyles in transition: Evolution and natural history of the genus *Lactobacillus*. *FEMS Microbiology Reviews*, 41(Supp_1), S27-S48.

Evans, M., Salewski, R.P, Christman, M.C, Girard, S.A, & Tompkins, T.A. (2016). Effectiveness of *Lactobacillus helveticus* and *Lactobacillus rhamnosus* for the management of antibiotic-associated diarrhoea in healthy adults: A randomised, double-blind, placebo-controlled trial. *British Journal of Nutrition*, 116(1), 94-103.

EFSA Panel on Additives and Products or Substances used in Animal Feed (FEEDAP). (2012). Guidance on the assessment of bacterial susceptibility to antimicrobials of human and veterinary importance. *EFSA Journal*, 10(6), 2740-2750.

Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Goris J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., & Vandamme, P. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, 57(Pt1), 81-91.

Hamilton-Miller, J.M., & Shah, S. (1998). Vancomycin susceptibility as an aid to the identification of Lactobacilli. Letters in Applied Microbiology, 26(2), 153-154.

Hammes, W.P. & Hertel, C. (2015). *Lactobacillus*. In W.B. Whitman (Eds.), *Bergey's Manual of Systematics of Archaea and Bacteria*. Chichester, UK: John Wiley & Sons, Ltd. p 1-76.

Harris, H.M.B., Bourin, M.J.B., Claesson, M.J., & O'Toole, P.W. (2017). Phylogenomics and comparative genomics of *Lactobacillus salivarius*, a mammalian gut commensal. *Microbial Genomics*, 3(8), e000115.

Hidetoshi, M., Hidehiro, T., Kenshiro, O., Masaru, M., Taylor, T.D., Shizunobu, I., & Masahira, H. (2009). Complete genome sequence of the probiotic *Lactobacillus rhamnosus* ATCC 53103. *Journal of Bacteriology*, 191(24), 7630-7631.

Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., & Hauser, L.J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11, 119.

Ingham, S.C. (1999). Use of modified *Lactobacillus* selective medium and *Bifidobacterium* iodoacetate medium for differential enumeration of *Lactobacillus acidophilus* and *Bifidobacterium* spp. in

519 powdered nutritional products. *Journal of Food Protection*, 62(1), 77-80.

520 Inglin, R.C., Meile, L., & Stevens, M.J.A. (2018). Clustering of pan- and core-genome of *Lactobacillus*
521 provides novel evolutionary insights for differentiation. *BMC Genomics*, 19(1), 284.

522 Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M.,
523 Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R, Williams, L.E., Frye, J.G, Elsayegh,
524 T., Sardar, D., Westman, E.L., Pawlowski, A.C., Johnson, T.A., Brinkman, F.S., Wright, G.D.,
525 & McArthur, A.G. (2017). CARD 2017: Expansion and model-centric curation of the
526 comprehensive antibiotic resistance database. *Nucleic Acids Research*, 45, D566-D573.

527 Kankainen, M., Paulin, L., Tynkkynen, S., von Ossowski, I., Reunanen, J., Partanen, P., Satokari, R.,
528 Vesterlund, S., Hendrickx, A.P., Lebeer, S., de Keersmaecker, S.C., Vanderleyden, J.,
529 Hamalainen, T., Laukkanen, S., Salovuori, N., Ritari, J., Alatalo, E., Korpela, R.,
530 Mattila-Sandholm, T., Lassig, A., Hatakka, K., Kinnunen, K.T., Karjalainen, H., Saxelin, M.,
531 Laakso, K., Surakka, A., Palva, A., Salusjarvi, T., Auvinen, P., & de Vos, W.M. (2009).
532 Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a
533 human-mucus binding protein. *Proceedings of the National Academy of Sciences of the*
534 *United States of America*, 106(40), 17193-17198.

535 Kant, R., Blom, J., Palva, A., Siezen, R.J., & de Vos, W.M. (2011). Comparative genomics of
536 *Lactobacillus*. *Microbial Biotechnology*, 4(3), 323-332.

537 Kant, R., Rintahaka, J., Yu, X., Sigvart-Mattila, P., Paulin, L., Mecklin, J.P., Saarela, M., Palva, A., &
538 von Ossowski, I. (2014). A comparative pan-genome perspective of niche-adaptable
539 cell-surface protein phenotypes in *Lactobacillus rhamnosus*. *PLoS One*, 9(7), e102762.

540 Karapetkov, N., Georgieva, R., Rumyan, N., & Karaivanova, E. (2011). Antibiotic susceptibility of
541 different lactic acid bacteria strains. *Beneficial Microbes*, 2(4), 335-339.

542 Kelleher, P., Bottacini, F., Mahony, J., Kilcawley, K.N., & van Sinderen, D. (2017). Comparative and
543 functional genomics of the *Lactococcus lactis* taxon; insights into evolution and niche
544 adaptation. *BMC Genomics*, 18(1), 267.

545 Korhonen, J.M., Van Hoek, A.H., Saarela, M., Huys, G., Tosi, L., Mayrhofer, S., & Wright, A.V. (2010).
546 Antimicrobial susceptibility of *Lactobacillus rhamnosus*. *Beneficial Microbes*, 1(1), 75-80.

547 Lebeer, S., Vanderleyden, J., & de Keersmaecker, S.C. (2008). Genes and molecules of *Lactobacilli*
548 supporting probiotic action. *Microbiology and Molecular Biology Reviews*, 72(4), 728-764.

549 Lombard, V., Ramulu, H.G., Drula, E., Coutinho, P.M., & Henrissat, B. (2014). The
550 carbohydrate-active enzymes database (CAZY) in 2013. *Nucleic Acids Research*, 42,
551 D490-495.
552 Lowe, T.M., & Eddy, S.R. (1997). tRNAscan-SE: A program for improved detection of transfer RNA
553 genes in genomic sequence. *Nucleic Acids Research*, 25(5), 955-964.
554 Lugli, G.A., Milani, C., Turroni, F., Duranti, S., Mancabelli, L., Mangifesta, M., Ferrario, C., Modesto,
555 M., Mattarelli, P., Jiri, K., van Sinderen, D., & Ventura, M. (2017). Comparative genomic and
556 phylogenomic analyses of the Bifidobacteriaceae family. *BMC Genomics*, 18(1), 568.
557 Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G.,
558 Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng,
559 S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W., & Wang, J. (2012)
560 SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler.
561 *Gigascience*, 1(1), 18.
562 Mailund, T., Brodal, G.S., Fagerberg, R., Pedersen, C.N., & Phillips, D. (2006). Recrafting the
563 neighbor-joining method. *BMC Bioinformatics*, 7, 29.
564 Martino, M.E., Bayjanov, J.R., Caffrey, B.E., Wels, M., Joncour, P., Hughes, S., Gillet, B., Kleerebezem,
565 M., van Hijum, S.A., & Leulier, F. (2016). Nomadic lifestyle of *Lactobacillus plantarum*
566 revealed by comparative genomics of 54 strains isolated from different habitats.
567 *Environmental Microbiology*, 18(12), 4974-4989.
568 Mols, M., de Been, M., Zwietering, M.H., Moezelaar, R., & Abee, T. (2010). Metabolic capacity of
569 *Bacillus cereus* strains ATCC 14579 and ATCC 10987 interlinked with comparative genomics.
570 *Environmental Microbiology*, 9, 2933-2944.
571 Odamaki, T., Bottacini, F., Kato, K., Mitsuyama, E., Yoshida, K., Horigome, A., Xiao, J.Z., & van
572 Sinderen, D. (2018). Genomic diversity and distribution of *Bifidobacterium longum subsp.*
573 *longum* across the human lifespan. *Scientific Reports*, 8(1), 85.
574 Pascual, L.M., Daniele, M.B., Ruiz, F., Giordano, W., Pájaro, C., & Barberis, L. (2008). *Lactobacillus*
575 *rhamnosus* L60, a potential probiotic isolated from the human vagina. *Journal of General and*
576 *Applied Microbiology*, 54(3), 141-148.
577 Prisciandaro, L.D., Geier, M.S., Butler, R.N., Cummins, A.G., & Howarth, G.S. (2011). Evidence
578 supporting the use of probiotics for the prevention and treatment of chemotherapy-induced

intestinal mucositis. *Critical Reviews in Food Science and Nutrition*, 51(3), 239-247.

Raftis, E.J., Salvetti, E., Torriani, S., Felis, G.E., & O'Toole, P.W. (2011). Genomic diversity of *Lactobacillus salivarius*. *Applied and Environmental Microbiology*, 77(3), 954-965.

Reuman, P.D., Duckworth, D.H., Smith, K.L., Kagan, R., Bucciarelli, R.L., & Ayoub, E.M. (1986). Lack of effect of *Lactobacillus* on gastrointestinal bacterial colonization in premature infants. *Pediatric Infectious Disease Journal*, 5(6), 663-668.

Richter, M., & Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America*, 106(45), 19126-19131.

Roberts, G.C., Feeney, J., Burgen, A.S., & Daluge, S. (1981). The charge state of trimethoprim bound to *Lactobacillus casei* dihydrofolate reductase. *FEBS Letters*, 131(1), 85-88.

Schwarz, S., Kehrenberg, C., Doublet, B., & Cloeckaert, A. (2004). Molecular basis of bacterial resistance to chloramphenicol and florfenicol. *FEMS Microbiology Reviews*, 28(5), 519-542.

Siezen, R.J., Tzeneva, V.A., Castioni, A., Wels, M., Phan, H.T., Rademaker, J.L., Starrenburg, M.J., Kleerebezem, M., Molenaar, D., & van Hylckama Vlieg, J.E. (2010). Phenotypic and genomic diversity of *Lactobacillus plantarum* strains isolated from various environmental niches. *Environmental Microbiology*, 12(3), 758-773.

Smokvina, T., Wels, M., Polka, J., Chervaux, C., Brisse, S., Boekhorst, J., van Hylckama Vlieg, J.E., & Siezen, R.J. (2013). *Lactobacillus paracasei* comparative genomics: Towards species pan-genome definition and exploitation of diversity. *PLoS One*, 8(7), e68731.

Succi, M., Tremonte, P., Reale, A., Sorrentino, E., Grazia, L., Pacifico, S., & Coppola, R. (2005). Bile salt and acid tolerance of *Lactobacillus rhamnosus* strains isolated from *Parmigiano Reggiano* cheese. *FEMS Microbiology Letters*, 244(1), 129-137.

Sun, Z., Zhang, W., Guo, C., Yang, X., Liu, W., Wu, Y., Song, Y., Kwok, L. Y., Cui, Y., Menghe, B., Yang, R., Hu, L., & Zhang, H. (2015). Comparative genomic analysis of 45 type strains of the genus *Bifidobacterium*: A snapshot of its genetic diversity and evolution. *PLoS One*, 10(2), e0117912.

Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: The bacterial pan-genome. *Current Opinion in Microbiology*, 11, 472-477.

Todorov, S.D. (2010). *Lactobacillus plantarum*: Characterization of the species and application in food

609 production. *Food Reviews International*, 26(3), 205-229.

610 Tripathi, P., Beaussart, A., Alsteens, D., Dupres, V., Claes, I., von Ossowski, I., de Vos, W.M., Palva, A.,
611 Lebeer, S., Vanderleyden, J., & Dufrêne, Y.F. (2013). Adhesion and nanomechanics of pili
612 from the probiotic *Lactobacillus rhamnosus* GG. *ACS Nano*, 7(4), 3685-3697.

613 Yu, B., Su, F., Wang, L., Zhao, B., Qin, J., Ma, C., Xu, P., & Ma, Y. (2011). Genome sequence of
614 *Lactobacillus rhamnosus* strain CASL, an efficient L-lactic acid producer from cheap
615 substrate cassava. *Journal of Bacteriology*, 193(24), 7013-7014.

616 Yu, J., Song, Y., Ren, Y., Qing, Y., Liu, W., & Sun, Z. (2017). Genome-level comparisons provide
617 insight into the phylogeny and metabolic diversity of species within the genus *Lactococcus*.
618 *BMC Microbiology*, 17(1), 213.

619 Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., & Yu, J. (2012). PGAP: Pan-genomes analysis pipeline.
620 *Bioinformatics*, 28(3), 416-418.

621 Zheng, J., Zhao, X., Lin, X.B., & Gänzle, M. (2015). Comparative genomics *Lactobacillus reuteri* from
622 sourdough reveals adaptation of an intestinal symbiont to food fermentations. *Scientific*
623 *Reports*, 5, 18234.

624 **Table 1** *L. rhamnosus* genomes sequenced and analyzed in this study.

Strains	Age	Region	Accession	Size (Mb)	ORF	tRNA	GC%
LGG (Kankainen et al., 2009)	<1 yr	North Carolina	ASM2650v1	3.01	2985	57	46.7
FAHWH26L1	6d	Anhui	PRJNA558200	2.88	2734	51	46.8
FAHWH30L1	7d	Anhui	PRJNA558202	2.90	2902	52	46.7
FAHWH2L1	93yr	Anhui	PRJNA558204	2.94	2856	57	47.3
FAHWH5L1	90yr	Anhui	PRJNA558208	2.91	2874	30	47.3
FAHWHD30L7	86yr	Anhui	PRJNA558583	2.89	2806	56	47.5
FAHWH35L1	70yr	Anhui	PRJNA558584	3.10	3024	53	47.4
FAHWH38L5	77yr	Anhui	PRJNA558585	2.92	2866	54	47.4
FBJCY2L1	9m	Beijing	PRJNA558586	2.93	2818	47	46.6
FBJCY3L1	11m	Beijing	PRJNA558587	2.86	2738	39	46.7
FBJSY7L3	62yr	Beijing	PRJNA558588	3.02	2986	56	47.2
FBJSY31L2	54yr	Beijing	PRJNA558589	2.94	2914	50	46.7
FBJSY60L1	64yr	Beijing	PRJNA558590	2.95	2833	58	47.3
FBJSY66L1	60yr	Beijing	PRJNA558592	2.90	2829	54	47.4
FFJND15L1	6m	Fujian	PRJNA558593	2.92	2795	52	46.7
FFJLY7L1	11yr	Fujian	PRJNA558594	2.94	2876	56	47.3
FGSZY12L6	84yr	Gansu	PRJNA558595	2.94	2865	54	47.3
FHeNJZ4L2	<1yr	Henan	PRJNA558596	2.95	2857	52	46.8
FHeNJZ7L1	6m	Henan	PRJNA558597	2.90	2801	38	46.8
FHeNJZ8L1	7m	Henan	PRJNA558599	2.91	2758	48	46.8
FHNFQ3L5	27yrs	Henan	PRJNA558600	2.95	2914	56	47.4
FHNFQ4L1	8m	Henan	PRJNA558601	2.96	2923	56	47.4
FHNFQ14L7	61yr	Henan	PRJNA558602	3.00	2970	55	47.3
FH28-1	102yr	Hubei	PRJNA558603	2.86	2743	40	46.7
FNMGEL5-1	7m	Inner Mongolia	PRJNA558604	2.90	2700	50	46.8
FNMGHLBE6L3	29yr	Inner Mongolia	PRJNA558605	3.02	2990	58	47.3
FNMGHLBE18L5	70yr	Inner Mongolia	PRJNA558606	2.95	2830	58	47.3
FJSWX1L3	6m	Jiangsu	PRJNA558607	3.02	2980	56	47.3
FJSWX2L6	4m	Jiangsu	PRJNA558608	2.94	2869	54	47.3
FJSWX3-L2	6m	Jiangsu	PRJNA558609	3.00	2926	50	46.6
FJSWX9L1	9m	Jiangsu	PRJNA558610	2.82	2739	54	47.4
FJSWX22-4	10m	Jiangsu	PRJNA558611	2.93	2815	51	46.6
FJSWX24-1	6m	Jiangsu	PRJNA558613	2.92	2789	43	46.7
FJSWX28L2	11m	Jiangsu	PRJNA558614	2.89	2787	54	47.4
FWXBH7L3	79yr	Jiangsu	PRJNA558615	2.94	2830	58	47.3
FJSYC1-5	<1yr	Jiangsu	PRJNA558616	2.81	2681	50	46.8
FJSYC4-1	<1yr	Jiangsu	PRJNA558617	2.90	2765	52	46.7
FJSNJ1-1-M2	1yr	Jiangsu	PRJNA558618	2.86	2732	40	46.7
FJSSZ2L1	<1yr	Jiangsu	PRJNA558620	2.94	2829	53	46.6
FJSZJ2-1	11m	Jiangsu	PRJNA558621	2.90	2762	61	46.7
FJXSRPYH4L2	72yr	Jiangxi	PRJNA558622	2.95	2829	59	47.3
FNXYCHL9M3	82yr	Ningxia	PRJNA558623	3.02	2977	55	47.2

FQHXN3M6	36yr	Qinghai	PRJNA558624	2.94	2827	56	47.3
FQHXN4M2	7yr	Qinghai	PRJNA558626	2.94	2830	56	47.3
FSHMX1-2	7m	Shanghai	PRJNA558627	2.92	2817	51	46.6
FSHMX3-1	3m	Shanghai	PRJNA558628	2.93	2817	48	46.6
FSDLZ7M12	85yr	Shandong	PRJNA558639	2.83	2676	52	46.7
FPAL5	<1yr	Shanxi	PRJNA558640	2.86	2730	51	46.7
FTJDG4G3	3d	Tianjin	PRJNA558641	2.93	2807	51	46.6
FTJDG9L1	4d	Tianjin	PRJNA558642	2.97	2842	44	46.7
FTJDG10L2	2d	Tianjin	PRJNA558643	2.96	2844	52	46.7
FTJDG11L1	2d	Tianjin	PRJNA558644	2.96	2864	60	46.7
FSCYA1L2	8m	Sichuan	PRJNA558646	2.87	2732	49	46.8
FXJWS2M1	43yr	Xinjiang	PRJNA558647	3.01	2970	55	47.2
FXJWS3M4	5m	Xinjiang	PRJNA558648	2.89	2786	54	47.4
FXJWS10M1	6m	Xinjiang	PRJNA558649	2.95	2833	58	47.3
FXJWS12L6	7yr	Xinjiang	PRJNA558650	2.95	2832	56	47.3
FXJWS13L6	16yr	Xinjiang	PRJNA558651	2.95	2832	56	47.3
FXJWS19L2	10yr	Xinjiang	PRJNA558652	2.96	2904	57	47.4
FXJWS25L4	40yr	Xinjiang	PRJNA558654	2.77	2730	47	47.2
FXJWS27L1	10yr	Xinjiang	PRJNA558655	2.95	2830	57	47.3
FXJWS37M4	40yr	Xinjiang	PRJNA558656	2.92	2868	54	47.2
FXJWS38L2	28yr	Xinjiang	PRJNA558657	2.95	2828	56	47.3
FXJWS44-L2	38yr	Xinjiang	PRJNA558658	2.95	2831	58	47.3
FXJSW6-1	31yr	Xinjiang	PRJNA558659	2.94	2828	58	47.3
FXJSW24M2	2yr	Xinjiang	PRJNA558661	2.96	2841	57	47.3
FZJHZ4L6	<1yr	Zhejiang	PRJNA558662	2.94	2824	58	47.3
FZJHZD11L1	10yr	Zhejiang	PRJNA558663	2.95	2830	56	47.3
FZJHZ14L3	8d	Zhejiang	PRJNA558664	2.94	2877	54	47.3
FZJJH6L2	4m	Zhejiang	PRJNA558665	2.86	2791	56	47.4
FZJTZ46L6	67yr	Zhejiang	PRJNA558666	3.04	3017	54	47.2

625

Figure legends

Fig. 1 Phylogenetic analysis of *L. rhamnosus*.

The triangle indicated donor's age with color-coded as follows: yellow for yr 0-1, grey for yr 2-16, orange for yr 27-54, pink for yr 60-79, light-blue for yr 82-102. The star indicated the region of sample collected, and the same area was represented by one color. Less than two samples in a region are represented in white. Red for Anhui, green for Beijing, blue for Henan, pink for Jiangsu, purple for Inner Mongolia, light green for Qinghai, light blue for Fujian, orange for Shanghai, grey for Xinjiang, yellow for Zhejiang, dark red for Tianjin. The strip divided the cluster into 5 large clusters, which are cluster 1 (blue), cluster 2 (green), cluster 3 (yellow), cluster 4 (orange), and cluster 5 (red).

Fig. 2 Pan- and core genes of *L. rhamnosus*.

(A) Pan-genome and core genome. The equation for calculating the pan-genome and core genome is $y=317.5*x^{0.677}+2706$ and $y=573.3*e^{(-0.07*x)}+1830$, respectively, in which x is the genome number of *L. rhamnosus* assayed.

(B) Venn diagram displaying the unique and core genes.

(C) Functional assignment of the core genome based on the COG database.

Fig. 3 Pairwise average nucleotide identity (ANI) analysis of *L. rhamnosus*.

The darker the color, the greater the similarity.

Fig. 4 Evaluation of carbohydrate utilization in *L. rhamnosus*.

(A) The predicted glycometabolism gene. A gradation of color from blue to white to red represents an increasing number of genes.

(B) The growth performance of the strains on different carbon sources. The red indicated growth, while the blue represented no growth.

(C) Gene cluster for D-lactose and L-fucose utilization.

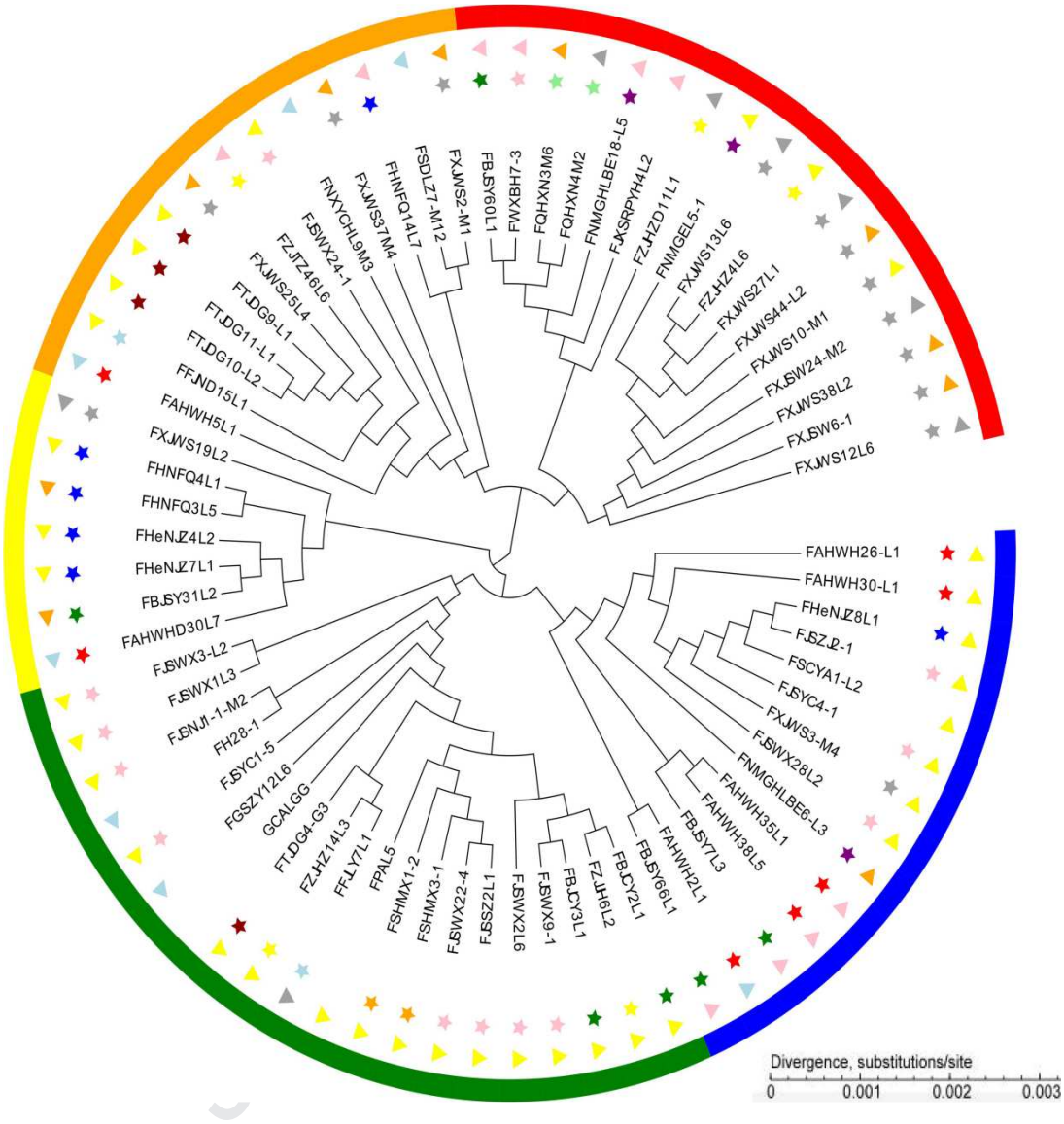
Fig. 5 Antibiotic resistance gene prediction in *L. rhamnosus*.

The darker the color, the more the antibiotic resistance gene.

Fig. 6 Microbiological cut-off threshold values.

Distribution of cut-off threshold values for streptomycin (A), erythromycin (B), tetracycline (C), clindamycin (D), chloramphenicol (E), and trimethoprim (F) in all strains. The red dotted line represented the cut-off threshold value of EFSA.

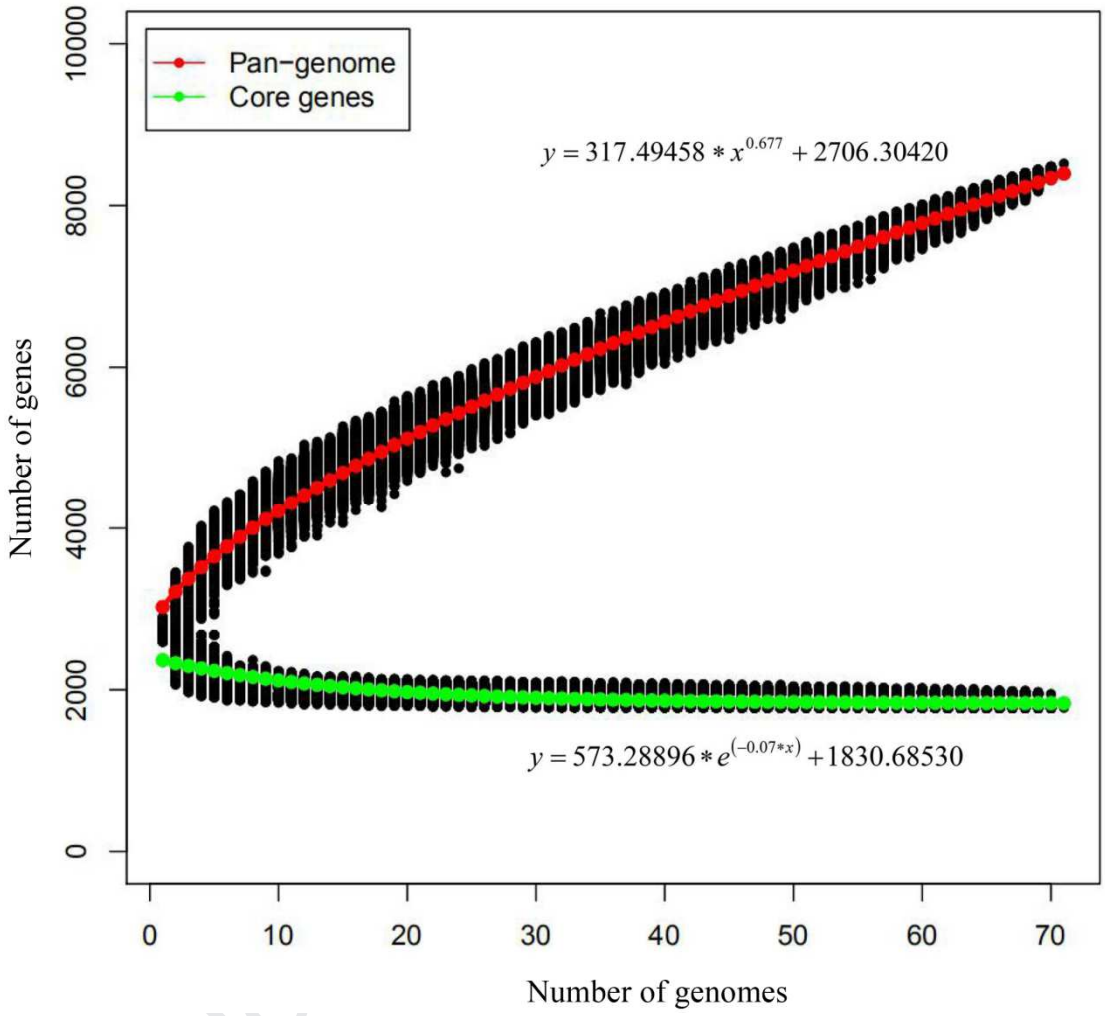
660 **Figure 1**



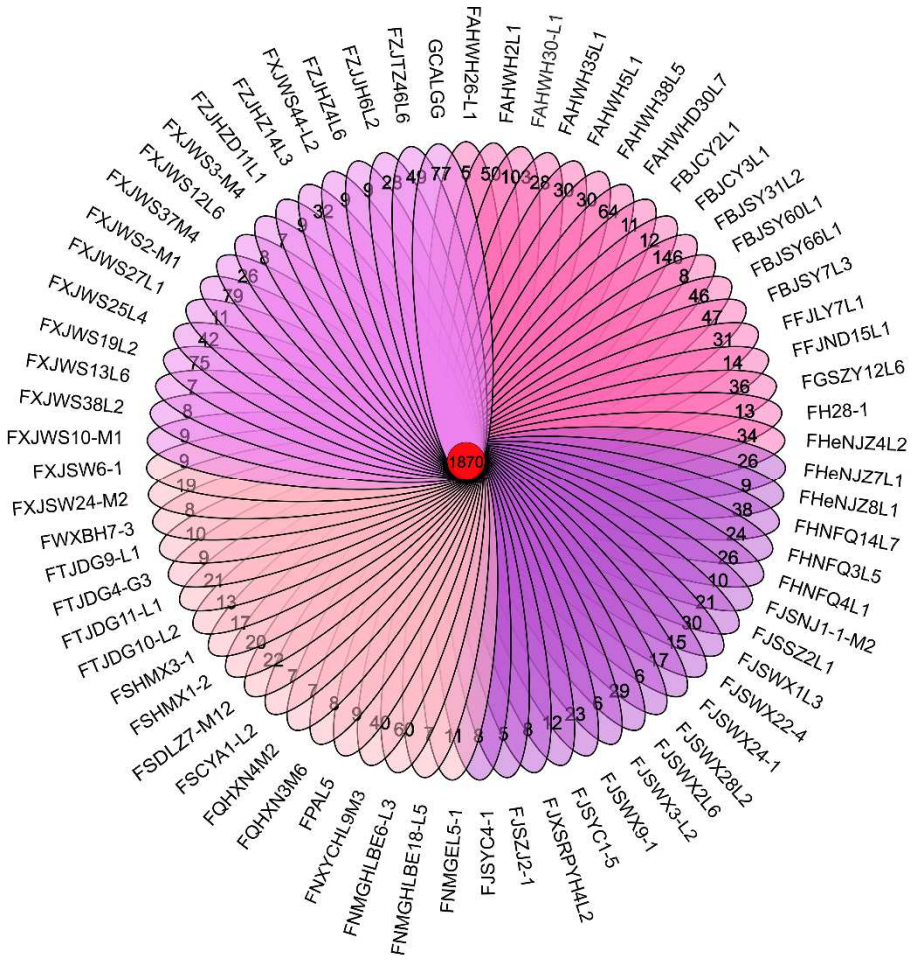
661

Figure 2

(A)

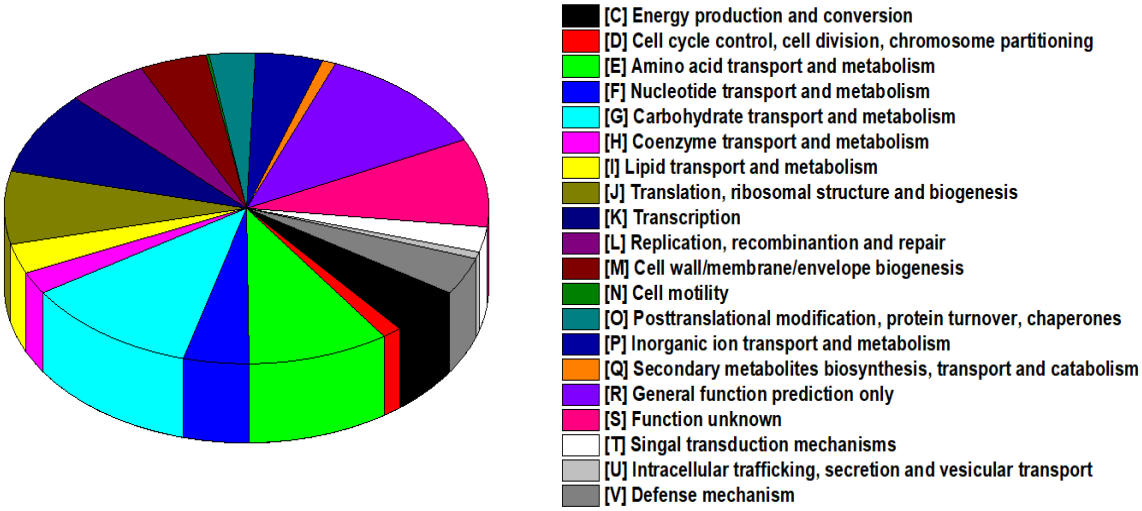

Pan−genome
Core genes
$y = 317.49458 * x^{0.677} + 2706.30420$
$y = 573.28896 * e^{(-0.07*x)} + 1830.68530$
Number of genes
Number of genomes
0
2000
4000
6000
8000
10000
0
10
20
30
40
50
60
70

665 **(B)**



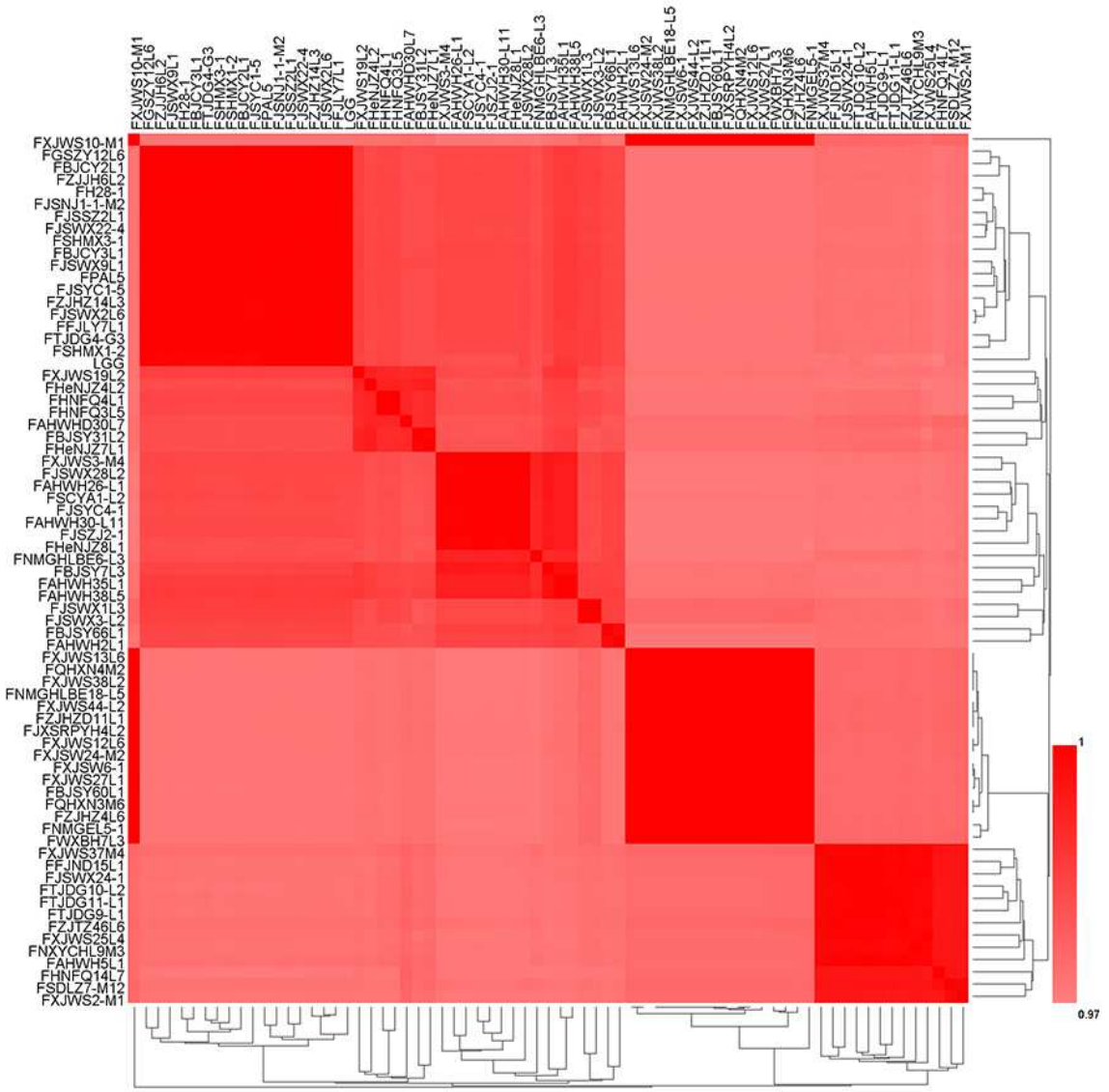
666

667 **(C)**



668
669

Figure 3

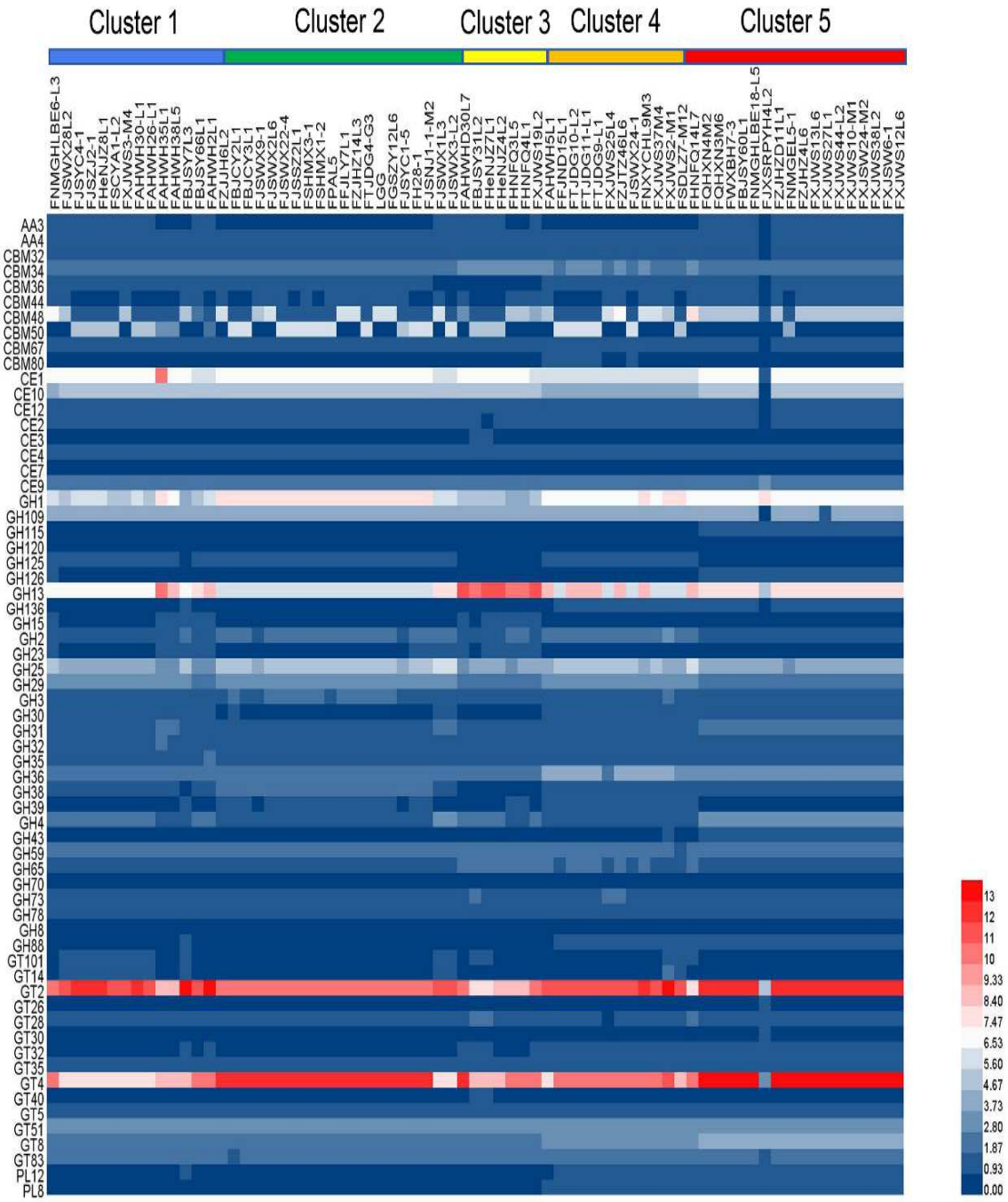


671

672

673 **Figure 4**

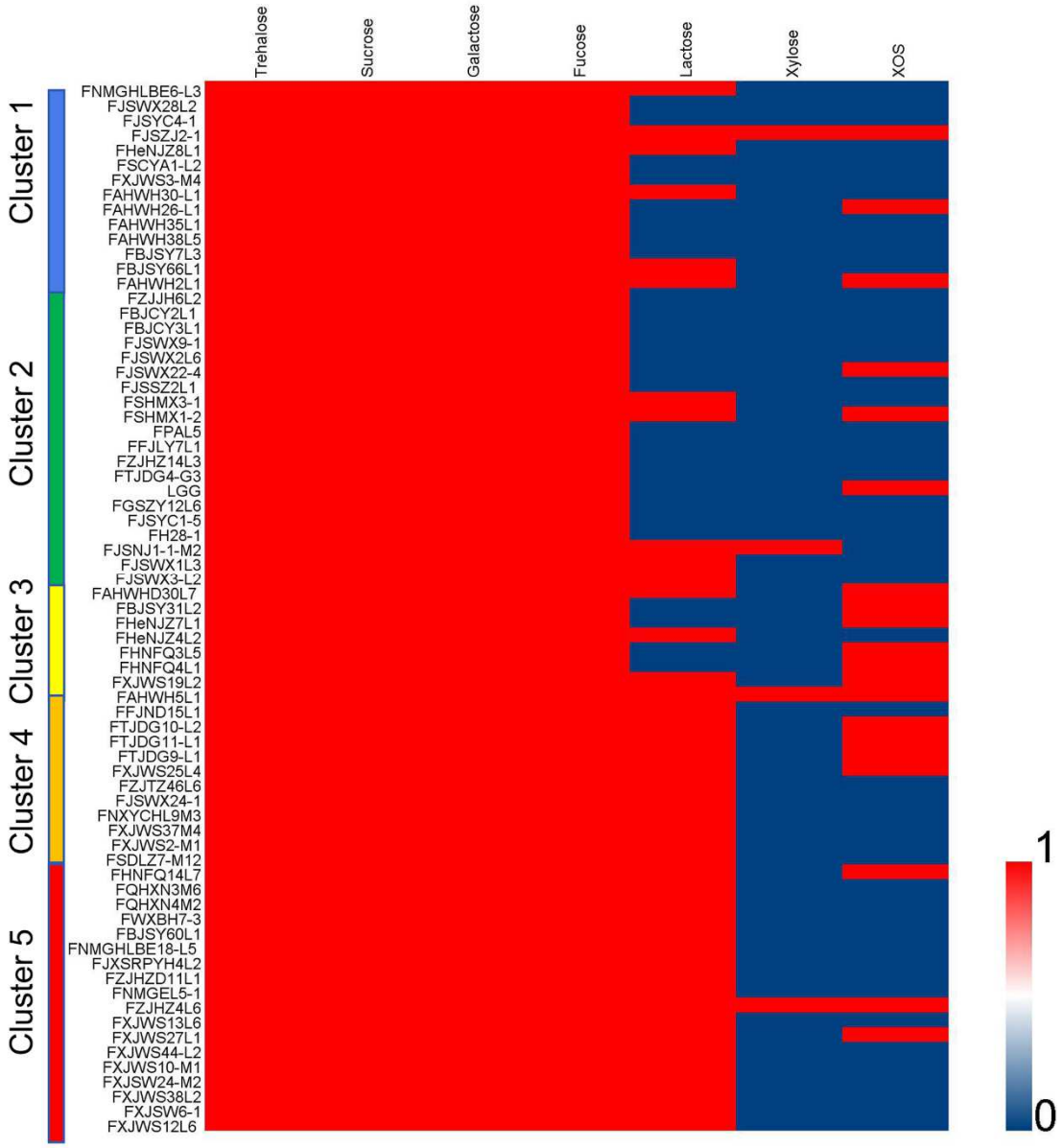
674 **(A)**



675

676

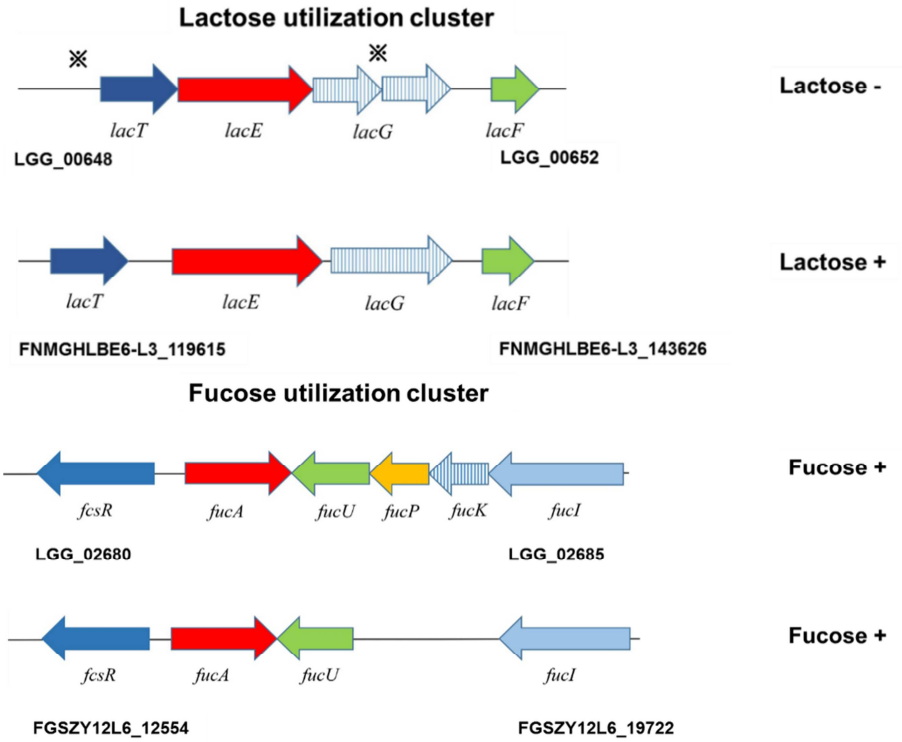
677 **(B)**


Trehalose
Sucrose
Galactose
Fucose
Lactose
Xylose
XOS
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
FNMGHLBE6-L3
FJSWX28L2
FJSYC4-1
FJSZJ2-1
FHeNJZ8L1
FSCYA1-L2
FXJWS3-M4
FAHWH30-L1
FAHWH26-L1
FAHWH35L1
FAHWH38L5
FBJSY7L3
FBJSY66L1
FAHWH2L1
FZJJH6L2
FBJCY2L1
FBJCY3L1
FJSWX9-1
FJSWX2L6
FJSWX22-4
FJSSZ2L1
FSHMX3-1
FSHMX1-2
FPAL5
FFJLY7L1
FZJHZ14L3
FTJDG4-G3
LGG
FGSZY12L6
FJSYC1-5
FH28-1
FJSNJ1-1-M2
FJSWX1L3
FJSWX3-L2
FAHWHD30L7
FBJSY31L2
FHeNJZ7L1
FHeNJZ4L2
FHNFQ3L5
FHNFQ4L1
FXJWS19L2
FAHWH5L1
FFJND15L1
FTJDG10-L2
FTJDG11-L1
FTJDG9-L1
FXJWS25L4
FZJTZ46L6
FJSWX24-1
FNXYCHL9M3
FXJWS37M4
FXJWS2-M1
FSDLZ7-M12
FHNFQ14L7
FQHXN3M6
FQHXN4M2
FWXBH7-3
FBJSY60L1
FNMGHLBE18-L5
FJXSRPYH4L2
FZJHZD11L1
FNMGEL5-1
FZJHZ4L6
FXJWS13L6
FXJWS27L1
FXJWS44-L2
FXJWS10-M1
FXJSW24-M2
FXJWS38L2
FXJSW6-1
FXJWS12L6
1
0

678

679

680 **(C)**

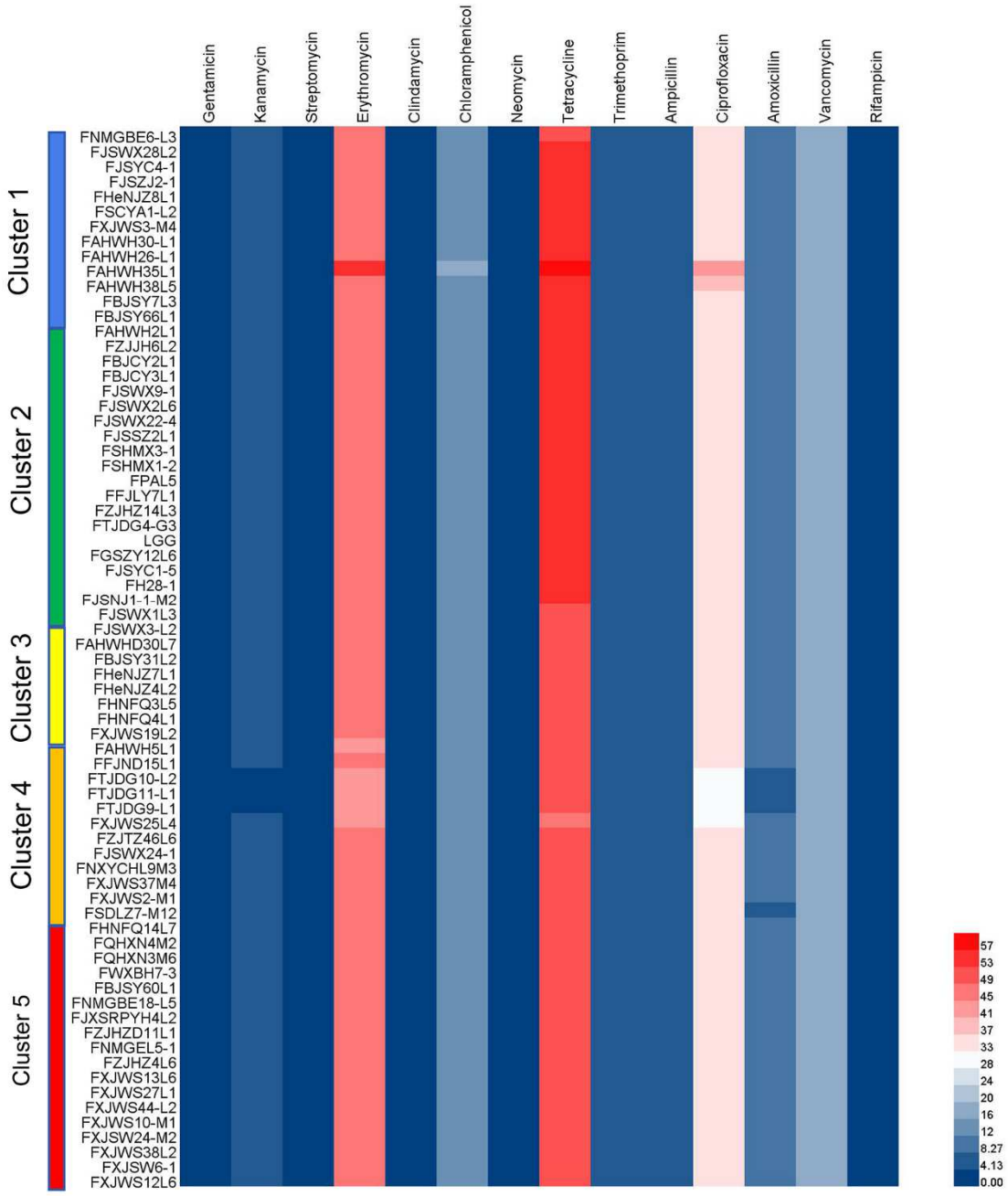


681

682

683

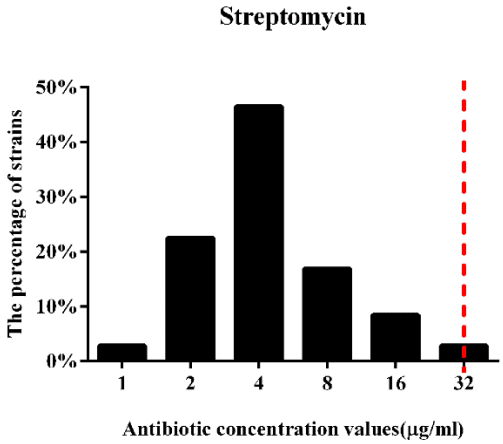
684 **Figure 5**



685

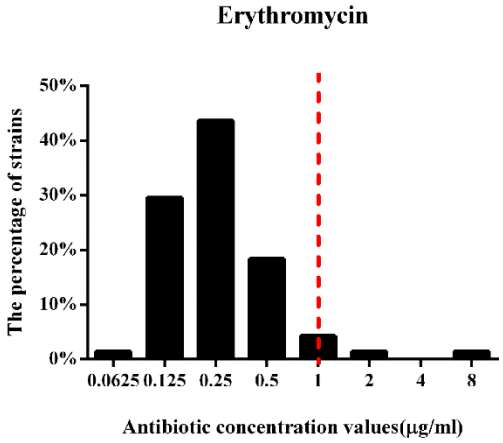
686 **Figure 6**

687 **(A)**



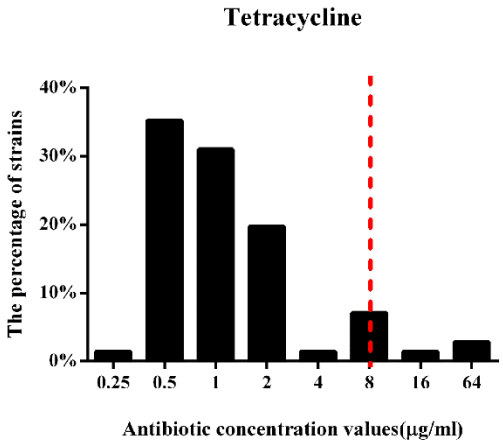
688

689 **(B)**



690

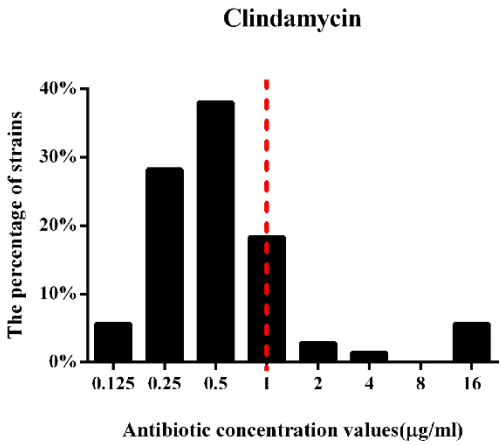
691 **(C)**



692

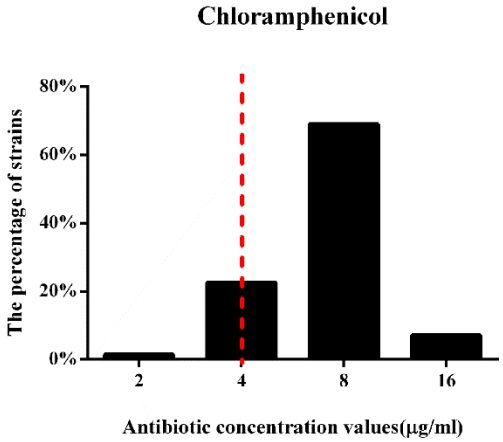
693

694 **(D)**



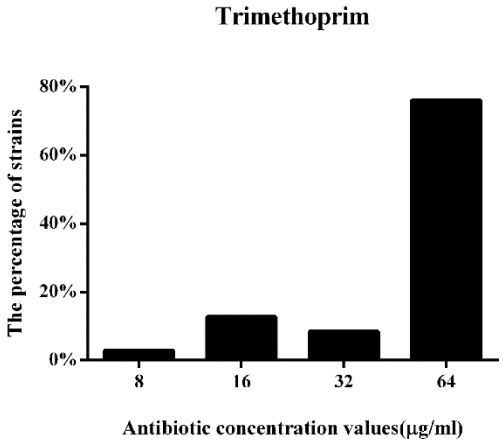
695

696 **(E)**



697

698 **(F)**



699

Conflict of interest

All authors declared no conflict of interest.