



Contents lists available at SciOpen

Food Science and Human Wellness

journal homepage: <https://www.sciopen.com/journal/2097-0765>Comparative genomic analysis revealed that dietary habits affected the adaptation of *Bifidobacterium bifidum* to the intestinal tract in different geographic populationsMin Li^{a,b,c,1}, Jie Yu^{a,c,d,1}, Weicheng Li^{a,c,d}, Qiong Wu^{a,b,c}, Jiaqi Sun^{a,b,c}, Zhihong Sun^{a,b,e,*}^a Key Laboratory of Dairy Biotechnology and Engineering, Inner Mongolia Agricultural University, Ministry of Education, Inner Mongolia Agricultural University, Hohhot 010018, China^b Key Laboratory of Dairy Products Processing, Ministry of Agriculture and Rural Affairs, Inner Mongolia Agricultural University, Hohhot 010018, China^c Inner Mongolia Key Laboratory of Dairy Biotechnology and Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China^d Collaborative Innovative Center for Lactic Acid Bacteria and Fermented Dairy Products, Ministry of Education, Inner Mongolia Agricultural University, Hohhot 010018, China^e Center for Applied Mathematics Inner Mongolia, Hohhot 010018, China

ARTICLE INFO

Article history:

Received 6 February 2023

Received in revised form 23 February 2023

Accepted 23 March 2023

Keywords:

Bifidobacterium bifidum

Comparative genomics

Genetic diversity

Adaptation

Carbohydrate-active enzymes

ABSTRACT

Recent research on the genome of *Bifidobacterium bifidum* has mainly focused on the isolation sources (intestinal tract niche) recently, but reports on the isolation region are limited. This study analyzed the differences in the genome of *B. bifidum* isolated from different geographical populations by comparative genomic analysis. Results at the genome level indicated that the GC content of American isolates was significantly higher than that of Chinese and Russian isolates. The phylogenetic tree, based on 919 core genes showed that *B. bifidum* might be related to the geographical characteristics of isolation region. Furthermore, functional annotation analysis demonstrated that copy numbers of carbohydrate-active enzymes (CAZs) involved in the degradation of polysaccharide from plant and host sources in *B. bifidum* were high, and 18 CAZs showed significant differences across different geographical populations, indicating that *B. bifidum* had adapted to the human intestinal environment, especially in the groups with diets rich in fiber. Dietary habits were one of the main reasons for the differences of *B. bifidum* across different geographical populations. Additionally, *B. bifidum* exhibited high diversity, evident in glycoside hydrolases, the CRISPR-Cas system, and prophages. This study provides a genetic basis for further research and development of *B. bifidum*.

© 2024 Beijing Academy of Food Sciences. Publishing services by Tsinghua University Press.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Bifidobacterium is a Gram-positive microorganism with high GC characteristics, which is widely distributed in different niches such as the gastrointestinal tract, waste water, and oral cavity of humans and non-human mammals, birds, and social insects^[1–2]. The ecological distribution may be due to the direct transmission of *Bifidobacterium*

cells from the mother or caregiver to the offspring. Recent studies have shown that, *B. animalis*, *B. adolescentis*, *B. dentium*, and *B. catenulatum* have a broader lifestyle, known as the “world bifidobacteria group”^[3]. However, *B. bifidum*, *B. breve*, and *B. longum* are less distributed, especially in the human intestine, and have been proven to be part of the predominant bacteria in the intestinal flora of breastfeeding infants^[4]. Numerous studies have been reported that *B. bifidum* has beneficial effects, such as antibacterial effects on pathogens like *Helicobacter pylori*^[5–6], regulation of the host immune system^[7–8], and alleviation of inflammatory activities associated with certain chronic intestinal dysfunction^[9–10].

Genomic analysis of many bifidobacteria provided molecular evidence of environmental adaptation, including molecules responsible for intestinal colonization and survival in the human intestines. The presence of a large glycobiome in *B. bifidum*, which

¹ These authors have contributed equally to this work.

* Corresponding author at: Key Laboratory of Dairy Biotechnology and Engineering, Inner Mongolia Agricultural University, Ministry of Education, Inner Mongolia Agricultural University, Hohhot 010018, China.

E-mail address: sunzhihong78@imau.edu.cn (Z.H. Sun)

Peer review under responsibility of Tsinghua University Press.

SciOpen

Publishing services by Tsinghua University Press

represents all genes responsible for carbohydrate metabolism, including glycoside hydrolases (GHs), glycosyl transferases (GTs), polysaccharide lyases (PLs), carbohydrate esterases (CEs), and carbohydrate-binding modules (CBMs), with the GH family being especially abundant^[11], was obvious evidence for genome adaptation. Carbohydrates in the human diet (but are not those that are present in human milk), which usually represent polysaccharides found in vegetables, grains and whole grains, are called dietary carbohydrates. While most bifidobacteria can metabolize mono-, di-, and oligosaccharides, there is more variability in the metabolism of polysaccharidic dietary carbohydrates between species such as resistant starch, (arabino) xylans, (arabino) galactans, and arabinans, which are major components of the human diet^[12]. The genomes of *B. longum*, *B. adolescentis*, and *B. pseudocatenulatum* encode a wide range of glycan-active enzymes which are predicted to target plant-based carbohydrates^[13–16]. Current research on the *Bifidobacterium* genomics was revealed a conserved gene set in *B. bifidum*, which encoded various GH families, and these GH families decomposed and utilized the host-derived polysaccharides involved in mucin^[17]. Another study on the genome of *B. bifidum* PRL2010^[18] revealed a nutrient-acquisition strategy that targets host-derived glycans, such as those present in mucin.

Comparative genomics^[19] has become a commonly used analytical method to identify the genetic diversity of strains and analyze the evolutionary structure of specific populations, so as to deeply understand the diversity, metabolic capacity and phylogeny of intestinal symbiotic bacteria (e.g. *Bifidobacterium*)^[20–21], which is very important for the strains' characteristics. For example, according to phylogeny, the relationship between strains and their sources can be determined using the background of known strains' sources^[22]. Through genome comparison and functional annotation, the adaptation mechanism among strains in specific environment was revealed^[23]. It also provides a new method for effectively evaluating the diversity of strains. Lu et al.^[24] also sequenced and compared the genomes of 140 *B. bifidum* strains from different niches to explore their genome diversity.

Previous studies have focused on isolated sources (e.g., intestinal niche) of the *B. bifidum* genome. However, the study of isolated sites is limited. Therefore, in this study, 4 strains of *B. bifidum* isolated from the China population and 106 strains of *B. bifidum* from other regions published by NCBI were analyzed by comparative and functional genomics, and the genetic diversity and potential host adaptability of this species were discussed.

2. Materials and methods

2.1 Bacterial strains culturing, genome sequencing and assembly

Four strains of *B. bifidum* were isolated from adult intestinal tracts from different regions of China (Table S1). These four were cultured at 37 °C for 24 h with reinforced medium for *Clostridia* (RCM) under anaerobic conditions. The species of all strains were preliminarily determined using 16S rRNA gene-based identification (GENEWIZ Co., Ltd., Suzhou, China). All reagents used in this experiment were from Sinopharm Chemical Reagent Co., Ltd., (Shanghai, China) and were of analytical grade.

The draft genomes of these 4 strains were sequenced by Illumina HiSeq 4000 high-throughput sequencing platforms and the clean data (the linker and positive control sequence were removed) was filtered. Then, selecting appropriate k-mer value and using SOAPdenovo v2.04 software^[25], the filtered data are spliced, assembled, and single base corrected. The genome size, scaffold number, N50 and N90, and GC content were used as criteria to evaluate the assembly results. Assembly results with a small number of scaffolds, high N50 and N90 and near GC content to the reference genome were selected. Finally, the sequence with better assembly results was selected for subsequent soap verification. Additionally, another 108 publicly available genomes of *B. bifidum* from NCBI (<https://www.ncbi.nlm.nih.gov/>) were used for comparison (accessed on April 25, 2021) (Table S1).

2.2 General features prediction

The coding sequence (CDS) in the genome was predicted using Glimmer v3.02^[26] (<http://ccb.jhu.edu/software/glimmer/index.shtml>). The transfer RNA (tRNA) in the genome was predicted using tRNAscan-SE v2.0 software^[27] (<http://trna.ucsc.edu/software/>), and the amino acid sequences were annotated using the COG^[28] database by BLASTP^[29], and the specific parameters are (-a 20 -m 8 -b 1 -v 3 -e 1e-10).

2.3 The average nucleotide identity (ANI) values and phylogenetic analyses

Pairwise ANI values among 110 *B. bifidum* genomes were calculated using an in-house script. Based on the core gene sequence, a Neighbor-Joining tree (NJ) was constructed by TreeBeST (V1.9.2) software^[30]. The bootstrap parameter was set to 1 000. Interactive Tree of Life (iTOL) v5.0^[31] was performed to construct a phylogenetic tree. The ANI values of one strain (GCF_900445665.1_55147_E01 isolated by 108 people from different regions downloaded from NCBI) and the model strain were less than 96%, and the CDS number of another strain (GCF_001311705.1_ASM131170v1) was 0. ANI represents a mean of identity/similarity values between homologous genomic regions shared by 2 genomes. It is now generally accepted that ANI values of 95%–96% can be used as a boundary for species delineation^[32]. Therefore, these 2 strains were excluded from the follow-up analysis.

2.4 Pan-genomes and core-genome analysis

Functional genes of each strain were predicted and annotated using Prokka v1.11 software^[33] and then, Roary v3.6.1 software^[34] was used to divide the gene families based on the amino acid identity of more than 95%. Both the core and pan-genomes of 110 strains of *B. bifidum* were constructed and visualized by PanGP v1.0.1 software^[35].

2.5 Carbohydrate activity enzymes (CAZy) analysis

HMMER software^[36] (<http://csbl.bmb.uga.edu/dbCAN/>) was used to perform annotation of the genome sequence of carbohydrate-

active enzymes. According to the results, detailed information on carbohydrate-active enzyme family was searched on: <http://www.cazy.org/>.

2.6 CRISPR identification

CRISPRCasTyper^[37] was used to discover CRISPR loci in *B. bifidum* genomes and predict CRISPR repeats and spacers, and the CRISPR subtypes designation was based on the signature of Cas proteins^[38]. Furthermore, the structural composition diagrams of different subtypes of CRISPR were drawn. According to the nucleotide sequence of CRISPR repeat sequence, phylogenetic analysis was performed. The phylogenetic tree was constructed with 1 000 repetitions using the Neighbor-Joining method in TreeBeST v1.9.2 software^[30].

2.7 Prophage identification

Prophage Hunter^[39] (<https://pro-hunter.bgi.com/>) webserver was used to identify and annotate the prophages within all the strains.

2.8 Data visualization and statistics

Box plot, Venn diagram, bar plot, heatmap and principle component analysis (PCA) were all visualized by R v.4.0.4 software^[40], using packages “ggplot2”, “ggpubr”, “plotrix”, “pheatmap”, “FactoMineR”. Statistical significance ($P < 0.05$) among the 3 groups was determined by Adonis, PERMANOVA and Wilcoxon tests.

3. Results

3.1 Genomic features of *B. bifidum* isolates from different geographical populations

Four strains of *B. bifidum* were isolated from different samples in China and sequenced by the Illumina HiSeq 4000 high-throughput sequencing platform. A total of 110 *B. bifidum* genomes were compared, among which 106 were obtained from the NCBI RefSeq database (Table S1). The genome size of all strains ranged 2.14–2.55 Mb, with an average size of (2.22 ± 0.07) Mb; the average GC content was $(62.68 \pm 0.15)\%$, and the highest and lowest GC contents were 63.00% and 62.01%, respectively. The range of CDS was 1 430 to 1 969, with an average of $1\,718 \pm 78$. Each genome contained 51 ± 4 tRNAs on an average. According to different geographical populations, the strains were divided into 26 Chinese isolates, 39 American isolates, 7 Russian isolates, and 38 isolates of unknown origin. By comparing the genomes of isolates, it was found that the GC content of the Russian isolates was significantly lower than that of China and USA ($P < 0.05$), while that of China and USA was no difference. The genomic size of the American strain was significantly lower than that of the Russian strain ($P < 0.05$), suggesting that there might be regional differences in the genomic size and GC content of *B. bifidum* (Fig. 1). The imbalance of the number of strains might affect the reliability of the results, which urged evaluating the possibility of genome differences of *B. bifidum* in different regions based on the same number of strains. The results showed that the difference in genomic size between

American and Russian isolates disappeared, while the GC content of American isolates was significantly higher than that of Chinese and Russian ($P < 0.05$; Fig. S1). Furthermore, the strains with the lowest GC content were isolated from China, while the strains with the largest genome were isolated from Americans. These data indicated that there were significant differences in the genomes of *B. bifidum* in different regions of the population, which was reflected in the GC content.

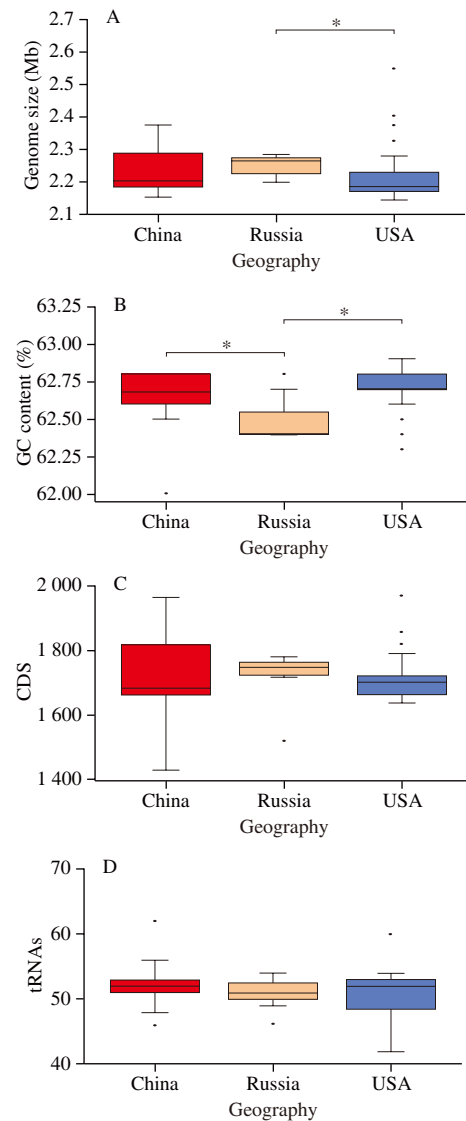


Fig. 1 Genome comparison of *B. bifidum* isolates from different geographical populations. (A) Genome size; (B) GC content; (C) CDS; (D) tRNA. Statistically significant differences among the 3 groups are denoted by * (Wilcoxon test, $P < 0.05$).

3.2 ANI values of *B. bifidum*

The paired ANI values of all strains ranged 98.02%–99.99%, which confirmed that all 110 strains of *B. bifidum* were the same species and there were no potential subspecies. In addition, the higher ANI values indicated that the genomes of *B. bifidum* was highly similar. According to the ANI cluster relationship, more than half of *B. bifidum* strains had regional correlation, among which 61.54% (16/26) Chinese strains, 71.43% (5/7) Russian strains, and 61.54% (24/39) American strains showed geographical clustering (Fig. 2).

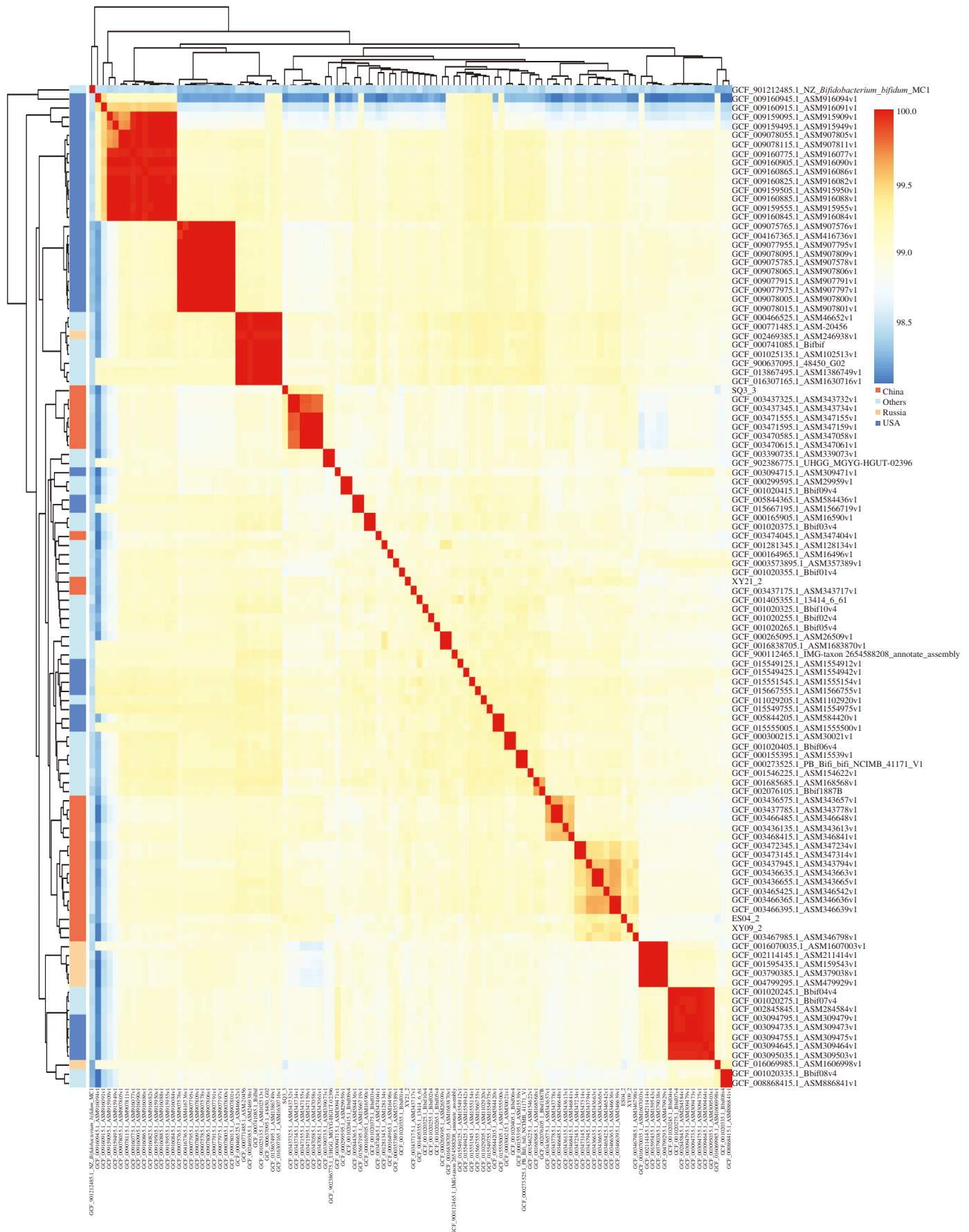


Fig. 2 ANI values of *B. bifidum*. Pairwise average nucleotide identity values across 110 *B. bifidum* genomes. The color coding for the genomes on the x-axis and y-axis was used to differentiate the strains.

3.3 Pan-core genome and phylogenetic analysis of *B. bifidum*

The pan-genomes and core genes were analyzed to study the genetic diversity of *B. bifidum*. The results showed that the pan-genomic asymptotic curve did not reach the plateau, which meant that when more genomes were included, the pan-genomes continued to increase, indicating that the pan-genomes of *B. bifidum* were considered to be in an open state. The curve of the core genes tended to smooth gradually (Fig. 3A). Specifically, there were 6 923 pan-genomes and 919 core genes in the 110 *B. bifidum* genomes. The Venn diagram showed the core genes and accessory genes of 110 *B. bifidum* strains, with a total of 919 genes and the number of accessory genes ranged from 1 712 to 2 060. These findings indicated that *B. bifidum* had a high genetic diversity (Fig. 3B).

By comparing the differences of pan-core genome of *B. bifidum* isolates from different geographical populations, it was found that the pan-genome of American isolates were larger and there were fewer core genes (Fig. 3C). The proportion of accessory genes in China and the USA was higher than in Russia (Fig. 3D). Similarly, the effect of the number of isolates on results was also ruled out. Based on the statistics of the same number of isolates from different regions, it was found that Chinese isolates had the largest pan-genomes and the highest proportion of accessory genes (Fig. S2). In addition, by comparing the unique genes of all isolates, we found that the genomes of *B. bifidum* isolates from 3 geographical populations were hypothetical proteins with unknown functions, although there were overlap (Fig. S3). This reflected that *B. bifidum* has high diversity, and has specific adaptive mechanisms in different regions.

The phylogenetic analysis was performed to further explore this geographical difference, and the phylogenetic tree was constructed based on 919 core genes (Fig. 3E). The results showed that 39 American isolates were distributed among each branch, while almost all Chinese isolates (25/26) clustered in one branch. Except for GCF_002469385.1_ASM246938v1, all the Russian isolates were grouped together. In addition to the American strains, the strains from China and Russia showed prominent geographical aggregation trends. In addition, a recombination event usually affects the DNA sequence, leading to the introduction of multiple mutation sites, thus affecting the phylogenetic structure of the strains. Therefore, Guttman et al.^[41] proposed the r/m value to measure the degree of recombination. r was the number of mutation sites introduced by recombination, and m was the number of mutation sites introduced by spontaneous mutation. The average r/m of 110 isolates of *B. bifidum* was 0.07, indicating a low recombination rate of the *B. bifidum* genome, which excluded the effect of recombination events on the phylogenetic structure of *B. bifidum* (Table S2).

3.4 Function annotation analysis of *B. bifidum*

To explore the functional characteristics of *B. bifidum*, the COG annotation was performed, and the analysis revealed that its function was comprehensive, covering most vital activities, including metabolism (such as amino acid transport metabolism and carbohydrate transport metabolism), information storage and processing, cellular processes and signals, and unknown functions (Fig. S4A). Among them, the core function was metabolism, which accounts for 39.3% of all functions, and carbohydrate metabolism

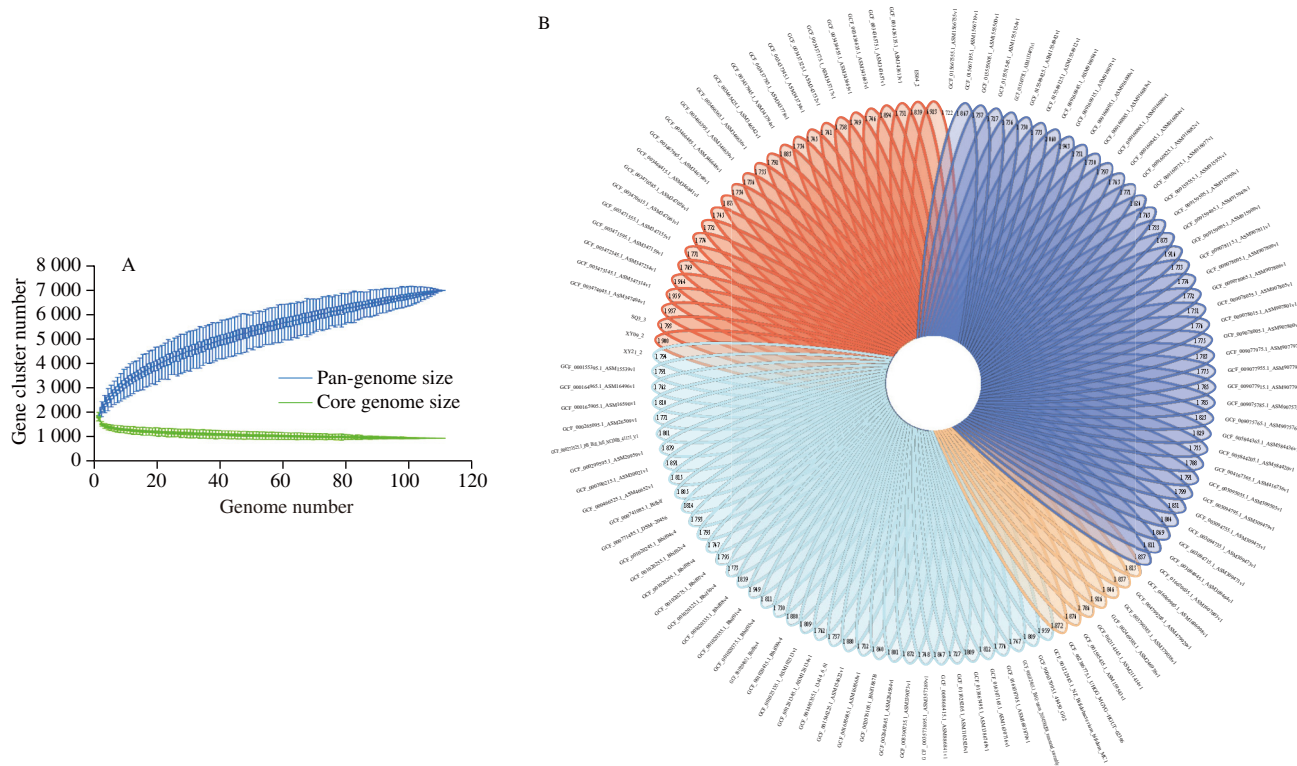


Fig. 3 Pan-core gene, accessory gene and phylogenetic analysis of *B. bifidum*. (A) Pan-genomes (blue) was open, while core genome (green) was conservative. (B) Core gene number and characteristic gene number of 110 *B. bifidum* isolates, and different colors represented the isolates from different geographical populations. (C) Comparison of pan-genomes and core genes of isolates from different geographical populations. (D) The proportion of accessory genes of isolates from different geographical populations in the total number of genes. (E) Phylogenetic tree was constructed based on 919 core genes, and bootstrap was set as 1 000.

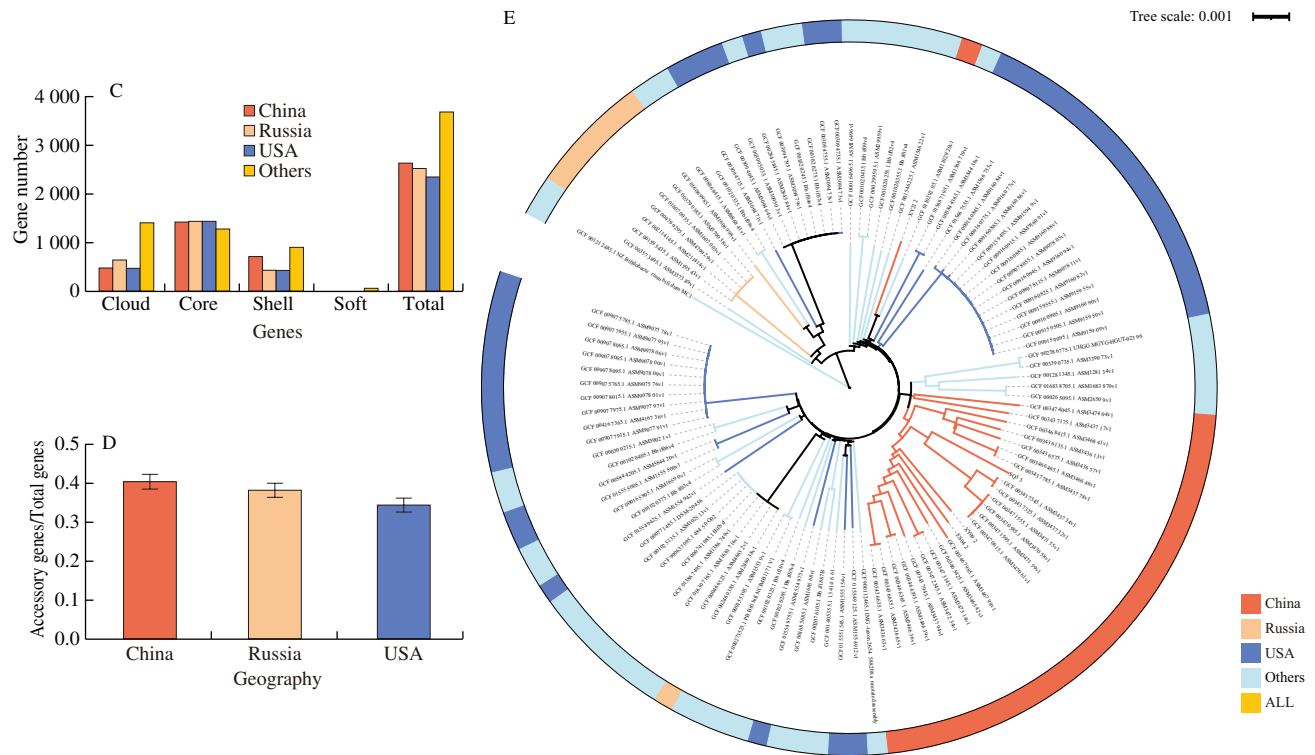


Fig. 3 (Continued)

ranks second, next to amino acids. The carbohydrate and amino acid metabolism genes in *B. bifidum* accounted for 10% and 12% of all genes, respectively. 21.52% of the functions were currently unclear, while 23.29% were for information storage and processing, with cellular processes and signals (Fig. S4B) at the bottom. These results indicated that *B. bifidum* requires an adequate genome to accommodate host-or diet-derived carbohydrate metabolism.

3.5 Prediction of carbohydrate in *B. bifidum*

To clarify the detailed utilization of the carbohydrate in *B. bifidum*, 110 genomes of *B. bifidum* were annotated based on the CAZy database. The results showed that there were many kinds of CAZys in the genome of *B. bifidum*. A total of 77 CAZys were identified, including 45 GHs, 14 CBMs, 11 GTs, and 7 CEs. GH family was particularly rich, accounting for 58.44% of the total CAZys, occupying a dominant position. Among them, the number of GH84, GH33, GH20, GH2, GH13_11, GH112 and GH109 was significantly higher than of other GH family, and they were mainly involved in cellulose metabolism and degradation of host-derived polysaccharides. The CBM32 was the highest member of the CBM family and the most abundant in CAZys of all *B. bifidum* strains, facilitating binding to carbohydrates encountered in the mucin core structure. GT2 and GT51 numbers were the higher members of the GT family, which were mainly involved in the metabolism of starch or sucrose. CE12, CE10, and CE1 were significantly higher than those of other CE family members, which were mainly involved in the degradation process of cellulose and hemicellulose (Fig. 4A). To further characterize the carbohydrate utilization of *B. bifidum* in the genus *Bifidobacterium*, we integrated these genomes

with 569 publicly available genomes of the human gut microbiota bifidobacteria and compared repertoires of CAZyme genes (Table S3). We calculated the gene copy number for each species, and identified 6 CAZyme genes with abundant characteristics in *B. bifidum* (Fig. S5 and Table S4). Four of them contained enzymes involved in degradation of polysaccharides from plant and host, and one enzyme degraded milk oligosaccharides. Therefore, the availability of plant-derived polysaccharides might be characteristic phenotype of *B. bifidum*. By comparing the CAZys genes from different regions, it was found that there were 18 regional differences in CAZys genes (Fig. S6), including 9 GHs, 4 GTs, 3 CEs, and 2 CBMs, specifically GH13, GH42, GH43, GH1, GH112, GH16, GH25, GH43_34, GH5_35, GT2, GT39, GT8, GT81, CE1, CE10, CE3, CBM13, and CBM50. Among them, *B. bifidum* isolates from Russia mainly used plant-derived CAZys, while the isolates from the USA mainly participated in cellulose decomposition. These results showed that the ability of *B. bifidum* to catabolize carbohydrates was linked to fiber-rich diets.

The clustering results with the presence or absence of carbohydrate utilization enzymes showed that the CAZys of *B. bifidum* isolates from the same geographical population were significantly aggregated, but there was no obvious separation trend of *B. bifidum* isolates from different geographic populations, but there were obvious differences in the genetic diversity of the CAZys (Fig. 4B). Furthermore, China population had the highest CAZys gene diversity, followed by the American population and the Russian population had the lowest diversity. No specific CAZys were found in *B. bifidum* isolates from China and Russian, but the specificity of CAZys in *B. bifidum* isolates from the USA primarily encoded a plant-derived degradable dietary fiber protein. These results indicated that *B. bifidum* had adapted to the human intestinal environment, especially those with a diet rich in

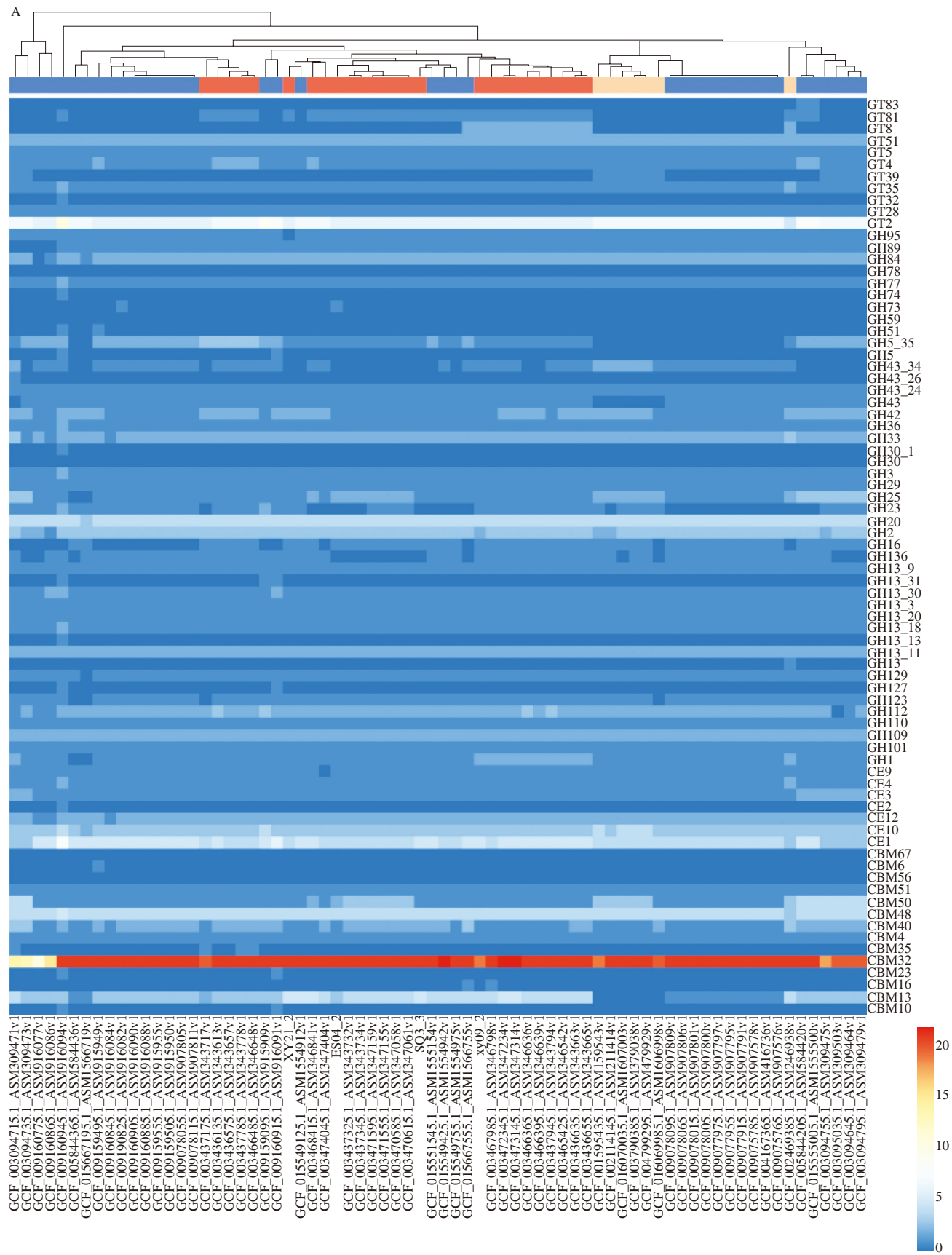


Fig. 4 Carbohydrate analysis of *B. bifidum*. (A) 110 genomes of *B. bifidum* were annotated based on CAZy database. Rows represent CAZys, and columns represent strains. The less the number of CAZys, the closer the color is to blue, and conversely, the closer it is to red. (B) According to the division of people in different regions, clustering and comparative analysis were carried out based on the existence and absence of CAZys. Statistically significant differences among the three groups are denoted by * (PERMANOVA test, $P < 0.05$).

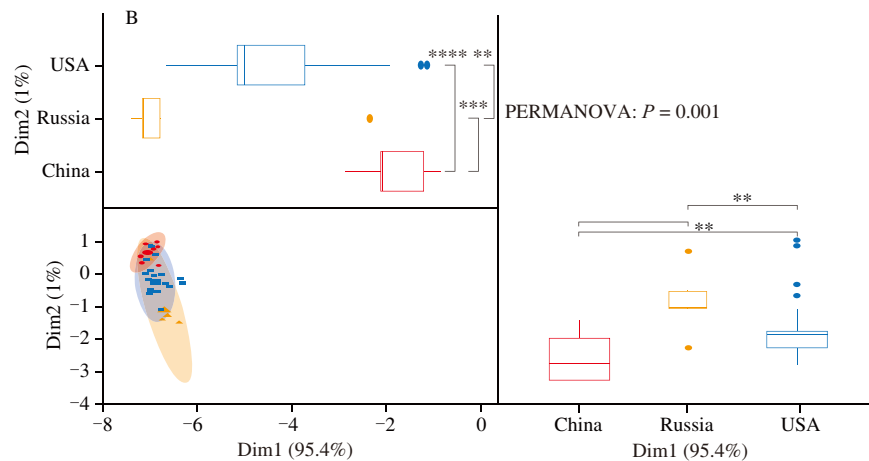


Fig. 4 (Continued)

dietary fiber, and dietary habits were one of the main reasons for the difference in *B. bifidum* in different geographical populations.

3.6 Identification of CRISPR-Cas Systems and Prophages in *B. bifidum*

The CRISPR-Cas analysis was performed to investigate the acquired immunity of *B. bifidum*. 39% (43/110) of the genomes encoded the CRISPR-Cas system based on the presence of Cas marker proteins. There were 3 subtypes of CRISPR-Cas, Type IC (10 strains), Type IIC (10 strains) and Type IIA (23 strains), and the Type II systems (30%) were more common than the Type I systems (9%). Based on the phylogenetic analysis of repeated sequences, the CRISPR-Cas system of *B. bifidum* was classified in more detail (Fig. 5A). Six branches of phylogenetic tree were found. Subtypes IIC, IIA, and IC all had 2 different branches. Two subtypes of IIC and one subtype of IIA were located in one large branch, while 2 branches of IC and 1 subtype of IIA were located in the other large branch, which indicated that the repetitive sequences of the same type of CRISPR/Cas systems had high homology in the phylogenetic tree, and the repetitive sequences of different types of systems were variable. Six strains were selected as representatives of each CRISPR-Cas subtype (Fig. 5B). Cas3 and Cas9 proteins were characteristic genes of type I and type II, respectively, and were used to distinguish different subtypes. They were key proteins for targeted interference with the CRISPR-Cas system. In addition, every CRISPR-Cas system contained Cas1 and Cas2 proteins, which were responsible for inserting new spacer sequences as important adaptive proteins. The CRISPR-Cas systems were not geographically related, and the CRISPR-Cas systems of the three Chinese isolates were all of type IC. There was only one Russian isolate, which CRISPR-Cas system was type IC. Interestingly, 90% of the 30 American isolates were type II (Table S5).

The prophage of *B. bifidum* was predicted. In general, 464 prophage regions were identified in 89.09% (89/110) of 110 strains of *B. bifidum* (Table S6). Only 12 strains of *B. bifidum* did not carry any prophage fragments, which indicated that prophage existed widely in *B. bifidum*. The number of prophages in *B. bifidum* strains from different geographical populations ranged from 0 to 15 (Fig. 5C). Among them, 51 isolates were from China, 6 from Russian

and 71 from the USA. Among the 464 predicted prophage regions, most of them were labeled with uncertain prophage fragments, and 183 were active prophage fragments, of which 38.8% were distributed in the American isolates, 27.87% in China isolates and the rest were distributed in Russian and other countries. No matter all the prophages or active phages, the number of prophages in American isolates was higher than that of China and Russia, while that China isolates was lower than that of Russian isolates, which was contrary to the result of active prophages. However, there was no statistical significance in the number of prophages of *B. bifidum* isolates in different geographical populations ($P > 0.05$; Fig. S7).

4. Discussion

Previous studies on *B. bifidum* mainly focused on its isolation sources, but the reports on its isolation places were limited, which needed to be more to fully understand the genetic diversity and functional characteristics of *B. bifidum*. This study showed that *B. bifidum* was a geographically specific species that had adapted to the human intestinal environment. The availability of host-derived or dietary-derived polysaccharides contributes to their niche expansion, and their adaptive mechanisms in the human intestinal environment may vary among dietary habit. This discovery will provide new ideas for the study of adaptation strategies of *B. bifidum*. In addition, the new knowledge of the phylogeny and geographical characteristics of *B. bifidum*, the CRISPR-Cas system, and the prophages gained in this study expanded our knowledge of this species.

The previously reported genome size of *B. bifidum* was between 2.14 and 2.28 Mb, and the GC content was about 62%^[1,42]. In this study, the genome size of 110 strains of *B. bifidum* was 2.14–2.55 Mb, which was not only consistent with the previous research results but also consistent with the characteristics of *Bifidobacterium*. By comparing the genomic characteristics of the isolates from different geographical populations, it was found that the GC content of American isolates was significantly higher than that of Chinese and Russian isolates ($P < 0.05$), which indicated that the genomes of *B. bifidum* from different geographical populations had differences. Previous studies have shown that *B. bifidum* has a closed pan-genomic structure^[43] and revealed its specific genetic strategy, which enabled its members to successfully establish and persist in the colonization of the human intestinal tract. Our research found that the pan-genomes

were in an open state, which was different from previous results. If more genomes were included, more gene families would be found, and the number of pan-genomes would increase. Open pan-genomes mean more frequent exchanges within species than closed pan-genomes. These results supported the hypothesis that the relative size and content of the pan-genome were potential indicators of genetic plasticity and the environmental adaptability of species. Core genes exist in most bacterial strains and determine the basic functions and phenotypic characteristics of almost all individuals in this bacterial species^[44]. Accessory genes existed only in the genomes of a few strains or genes specific to a particular strain. These genes were usually closely related to the specific living environment, for example, determining the adaptability of bacteria^[45]. On the premise that the number of isolates was the same, the core genes of isolates from 3 regions were the same, and the pan-genomes of Chinese isolates was the largest, while those of American isolates were the smallest. This result indicated that Chinese isolates had the largest accessory genes, which showed strong adaptability of *B. bifidum* in the intestines of the Chinese population.

We found that the ANI value of *B. bifidum* is 98.02%–99.99%, and its span is short, indicating high genome similarity. By constructing

the phylogenetic tree of 110 strains of *B. bifidum*, it was found that the phylogenetic relationships of these strains had some correlation with the isolation places, but it was far lower than expected. A possible reason was the limited number of strains. Lu et al.^[24] studied *B. bifidum* isolates and found a significant geographical correlation consistent with the current results. This may explain the geographical aggregation of *B. bifidum*.

COG functional annotation showed that the genome of *B. bifidum* was rich in protein-encoding genes, which mainly involved the transport and metabolism of carbohydrates, amino acid transport and metabolism, as well as transcription and ribosomal structure. This finding was closely related to the hypothesis that *B. bifidum* needed enough of a genomic library to adapt to the carbohydrate metabolism of the host or dietary sources. Previous studies have further confirmed this explanation, indicating that approximately 14% of genes in the pan-genome of *Bifidobacterium* encode carbohydrates to utilize the required enzymes^[14,46].

Among dietary factors, dietary carbohydrates were the most frequently reported dietary components positively correlated with *Bifidobacterium* species^[47-48]. We found that the genomes of *B. bifidum* from different geographical populations contained a variety

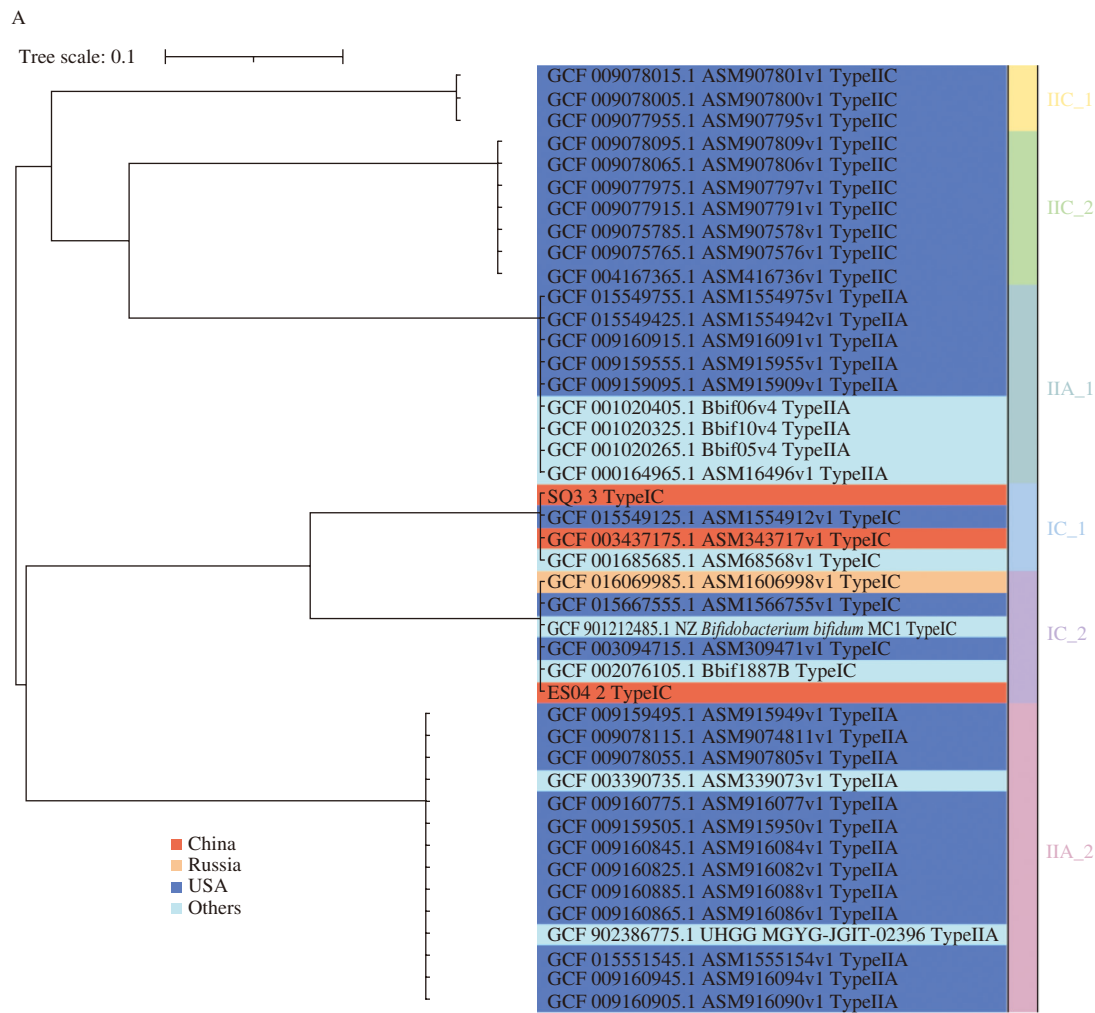


Fig. 5 Identification of CRISPR-Cas Systems and Prophages in *B. bifidum*. (A) The CRISPR-Cas system in *B. bifidum* was classified in more detail based on phylogenetic analysis of repeat sequences. Bootstrap value was set to 1 000. (B) Composition of different CRISPR-Cas subtypes. (C) The number of prophages was predicted in all strains.

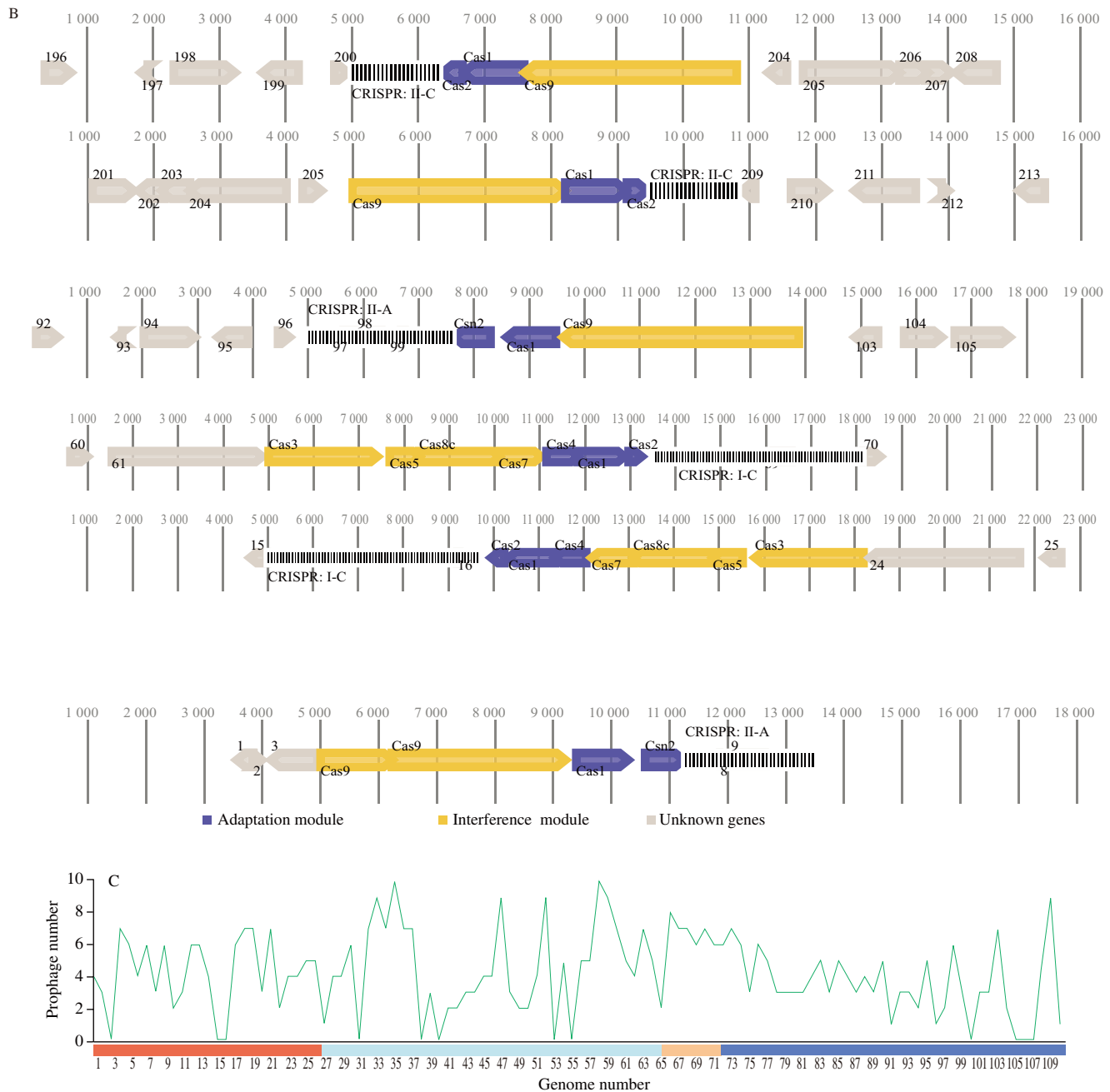


Fig. 5 (Continued)

of CAZys, of which the GHs family was dominant. Compared with other common bifidobacteria in human intestinal tract, we identified 6 CAZys that were characteristically abundant in *B. bifidum*. Four CAZys, CE12, GH84, GH13_11 and GH33 were involved in the degradation of host-derived polysaccharides and dietary fibers^[49]. There were 18 CAZys with geographical differences. The content of GH43 and GH13 in Russian isolates was high, and GH43 was related to the degradation of long-chain carbohydrates^[50-51]. GH13^[46] is a common substance in bacteria. It can decompose a variety of polysaccharides, including plant-derived complex carbohydrates, such as starch and related substrates (such as amylose and amylopectin and (cyclo) maltodextrin), trehalose, stachyose, raffinose, and maltose. GH5_35 and GT2 accounted for a higher proportion in the American

isolates and were related to cellulose degradation. GT2 represented the highest proportion of isolates in America and was the key glycosyltransferase involved in starch or sucrose metabolism. The specific CAZys in American isolates also mainly encoded for proteins, which can degrade the dietary fiber from plant sources. These data indicated that dietary habits affect host adaptation of *B. bifidum*. We hypothesized that dietary habits would affect the host adaptability of *B. bifidum*. One study^[52] has pointed out that “food shapes genes, especially carbohydrates.” Take amylase as an example. During the evolution of mammals, mammals with starch-containing diets had more copies of amylase genes carrying the instructions to construct amylase than those without starch (at least in the species studied). Both existing and emerging evidence supports the capacity of certain

Bifidobacterium species/strains to act as primary degradants of plant-based carbohydrates in complex diets, thereby facilitating metabolic cross-feeding with secondary degradants and resulting in higher yields of specific host health-related metabolites such as short-chain fatty acids^[53-54]. Chinese diet has always had a variety of flavors, with plant-based food as the main food, grain as the staple food, vegetables as the supplementary food, and a small amount of meat^[55]. Russian food is greasy and full of flavor, and bread is the staple food^[56]. Different from China and Russian diet, western diet is characterized by high animal meat, fat, sugar, highly processed food and low fiber^[57]. In addition, it was found that the diversity of CAZys isolated from China was the highest, which was consistent with the results of previous studies on accessory genes. Unfortunately, due to the small proportion of Russian and China isolates in this study and the lack of specific CAZys, the verification of carbohydrate metabolism needs to be further confirmed from future experimental evidence.

CRISPR loci exist in the genomes of a large number of prokaryotic organisms and provide acquired immunity against exogenous genetic elements^[58]. However, there was little research on CRISPR in *B. bifidum*. In this study, the CRISPR-Cas system of 110 strains of *B. bifidum* was analyzed, and it was found that 39% (43/110) of the genomes encoded the complete CRISPR-Cas system. This indicated that *B. bifidum* has advantages in constructing a mobile gene editing tool and screening strains with specific genotypes^[59]. Incomplete CRISPR-Cas loci in the remaining strains may be due to genome recombination, loss of ability to obtain other CRISPR loci or incomplete genome assembly in these strains. It was found that type IC, type IIA, and type IIC systems were widespread in *B. bifidum* strains, and type II was more common than type I. This was consistent with previous studies^[58]. Notably, the CRISPR-Cas system was not geographically relevant, with 90% of US isolates classified as Type II. The reasons remained to be further explored.

5. Conclusion

Our study confirmed that dietary habits lead to the differences of *B. bifidum* in different geographical populations, and then the availability of complex polysaccharides from host or dietary sources is specific. To be sure, the validation of carbohydrate metabolism needs further exploration through future experimental evidence. In addition, the genetic diversity of *B. bifidum* is mainly embodied in GHs family, CRISPR-Cas system and prophages. These results broaden the understanding of this species and provide a theoretical basis for further study on the genetic characteristics and functions of the strain.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgment

We are profoundly thankful to Research Fund for the National Key R&D Program of China (2022YFD21007002), the National Natural Science Foundation of China (32325040), Inner Mongolia Science & Technology planning project (2022YFSJ0017), and the earmarked fund for CARS36.

Data availability

Raw sequence data are available in the NCBI sequence read archive (project PRJNA895329).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://doi.org/10.26599/FSHW.2022.9250243>.

References

- [1] M. Ventura, C. Canchaya, A. Tauch, et al., Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum, *Microbiol. Mol. Biol. Rev.* 71(3) (2007) 495-548. <https://doi.org/10.1128/MMBR.00005-07>.
- [2] F. Turrone, D. van Sinderen, M. Ventura, Genomics and ecological overview of the genus *Bifidobacterium*, *Int. J. Food Microbiol.* 149(1) (2011) 37-44. <https://doi.org/10.1016/j.ijfoodmicro.2010.12.010>.
- [3] R. Lamendella, J.W.S. Domingo, C. Kelty, et al., Bifidobacteria in feces and environmental waters, *Appl. Environ. Microbiol.* 74(3) (2008) 575-584. <https://doi.org/10.1128/AEM.01221-07>.
- [4] F. Turrone, E. Foroni, P. Pizzetti, et al., Exploring the diversity of the bifidobacterial population in the human intestinal tract, *Appl. Environ. Microbiol.* 75(6) (2009) 1534-1545. <https://doi.org/10.1128/AEM.02216-08>.
- [5] E. Chenoll, B. Casinos, E. Bataller, et al., Novel probiotic *Bifidobacterium bifidum* CECT 7366 strain active against the pathogenic bacterium *Helicobacter pylori*, *Appl. Environ. Microbiol.* 77(4) (2011) 1335-1343. <https://doi.org/10.1128/AEM.01820-10>.
- [6] Y. Shirasawa, H. Shibahara-Sone, T. Iino, et al., *Bifidobacterium bifidum* BF-1 suppresses *Helicobacter pylori*-induced genes in human epithelial cells, *J. Dairy Sci.* 93(10) (2010) 4526-4534. <https://doi.org/10.3168/jds.2010-3274>.
- [7] D. Philippe, E. Heupel, S. Blum-Sperisen, et al., Treatment with *Bifidobacterium bifidum* 17 partially protects mice from Th1-driven inflammation in a chemically induced model of colitis, *Int. J. Food Microbiol.* 149(1) (2011) 45-49. <https://doi.org/10.1016/j.ijfoodmicro.2010.12.020>.
- [8] Y.R. Fu, Z.J. Yi, J.L. Pei, et al., Effects of *Bifidobacterium bifidum* on adaptive immune senescence in aging mice, *Microbiol. Immunol.* 54(10) (2010) 578-583. <https://doi.org/10.1111/j.1348-0421.2010.00255.x>.
- [9] S. Guglielmetti, D. Mora, M. Gschwender, et al., Randomised clinical trial: *Bifidobacterium bifidum* MIMBb75 significantly alleviates irritable bowel syndrome and improves quality of life-a double-blind, placebo-controlled study, *Aliment. Pharm. Ther.* 33(10) (2011) 1123-1132. <https://doi.org/10.1111/j.1365-2036.2011.04633.x>.
- [10] F. Mouni, E. Aissi, J. Hernandez, et al., Effect of *Bifidobacterium bifidum* DSM 20082 cytoplasmic fraction on human immune cells, *Immunol. Invest.* 38(1) (2009) 104-115. <https://doi.org/10.1080/08820130802608303>.
- [11] F. Turrone, C. Milani, S. Duranti, et al., Glycan utilization and cross-feeding activities by bifidobacteria, *Trends Microbiol.* 26(4) (2018) 339-350. <https://doi.org/10.1016/j.tim.2017.10.001>.
- [12] S.M. Kelly, J. Munoz-Munoz, D. van Sinderen, Plant glycan metabolism by bifidobacteria, *Front. Microbiol.* 12 (2021) 609418. <https://doi.org/10.3389/fmicb.2021.609418>.
- [13] H. Chung The, C. Nguyen Ngoc Minh, C. Tran Thi Hong, et al., Exploring the genomic diversity and antimicrobial susceptibility of *Bifidobacterium pseudocatenulatum* in a Vietnamese population, *Microbiol. Spectrum.* 9(2) (2021) e00526-00521. <https://doi.org/10.1128/Spectrum.00526-21>.
- [14] C. Milani, F. Turrone, S. Duranti, et al., Genomics of the genus *Bifidobacterium* reveals species-specific adaptation to the glycan-rich gut environment, *Appl. Environ. Microbiol.* 82(4) (2016) 980-991. <https://doi.org/10.1128/AEM.03500-15>.
- [15] S. Duranti, C. Milani, G.A. Lugli, et al., Evaluation of genetic diversity among strains of the human gut commensal *Bifidobacterium adolescentis*, *Sci. Rep.* 6(1) (2016) 23971.
- [16] Y. Watanabe, Y. Saito, T. Hara, et al., Xylan utilisation promotes adaptation of *Bifidobacterium pseudocatenulatum* to the human gastrointestinal tract, *ISME Commun.* 1(1) (2021) 62.

- [17] A.G. Abdelhamid, N.K. El-Dougoud, Comparative genomics of the gut commensal *Bifidobacterium bifidum* reveals adaptation to carbohydrate utilization, *Biochem. Biophys. Res. Commun.* 547 (2021) 155-161. <https://doi.org/10.1016/j.bbrc.2021.02.046>.
- [18] F. Turrioni, F. Bottacini, E. Foroni, et al., Genome analysis of *Bifidobacterium bifidum* PRL2010 reveals metabolic pathways for host-derived glycan foraging, *Proc. Natl. Acad. Sci.* 107(45) (2010) 19514-19519. <https://doi.org/10.1073/pnas.1011100107>.
- [19] R. Kant, J. Blom, A. Palva, et al., Comparative genomics of *Lactobacillus*, *Microbial Biotechnol.* 4(3) (2011) 323-332. <https://doi.org/10.1111/j.1751-7915.2010.00215.x>.
- [20] Z.H. Sun, W.Y. Zhang, C.Y. Guo, et al., Comparative genomic analysis of 45 type strains of the genus *Bifidobacterium*: a snapshot of its genetic diversity and evolution, *PLoS One* 10(2) (2015) e0117912. <https://doi.org/10.1371/journal.pone.0117912>.
- [21] M.J. Kwak, S.K. Kwon, J.K. Yoon, et al., Evolutionary architecture of the infant-adapted group of *Bifidobacterium* species associated with the probiotic function, *Syst. Appl. Microbiol.* 39(7) (2016) 429-439. <https://doi.org/10.1016/j.syapm.2016.07.004>.
- [22] J.G.V. da Silva, A.T. Vieira, T.J. Sousa, et al., Comparative genomics and *in silico* gene evaluation involved in the probiotic potential of *Bifidobacterium longum* 51A, *Gene* 795 (2021) 145781. <https://doi.org/10.1016/j.gene.2021.145781>.
- [23] T. Vatanen, D.R. Plichta, J. Somani, et al., Genomic variation and strain-specific functional adaptation in the human gut microbiome during early life, *Nat. Microbiol.* 4(3) (2019) 470-479. <https://doi.org/10.1038/s41564-018-0321-5>.
- [24] W.W. Lu, Z.M. Pei, M.N. Zang, et al., Comparative genomic analysis of *Bifidobacterium bifidum* strains isolated from different niches, *Genes* 12(10) (2021) 1504. <https://doi.org/10.3390/genes12101504>.
- [25] R.B. Luo, B.H. Liu, Y.L. Xie, et al., SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler, *Gigascience* 1(1) (2012) 18. <https://doi.org/10.1186/2047-217X-1-18>.
- [26] A.L. Delcher, K.A. Bratke, E.C. Powers, et al., Identifying bacterial genes and endosymbiont DNA with Glimmer, *Bioinformatics* 23(6) (2007) 673-679. <https://doi.org/10.1093/bioinformatics/btm009>.
- [27] T.M. Lowe, S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence, *Nucleic Acids Res.* 25(5) (1997) 955-964. <https://doi.org/10.1093/nar/25.5.955>.
- [28] R.L. Tatusov, N.D. Fedorova, J.D. Jackson, et al., The COG database: an updated version includes eukaryotes, *BMC Bioinform.* 4(1) (2003) 1-14. <https://doi.org/10.1186/1471-2105-4-41>.
- [29] S.F. Altschul, W. Gish, W. Miller, et al., Basic local alignment search tool, *J. Mol. Biol.* 215(3) (1990) 403-410.
- [30] A.J. Vilella, J. Severin, A. Ureta-Vidal, et al., EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates, *Genome Res.* 19(2) (2009) 327-335. <https://doi.org/10.1101/gr.073585.107>.
- [31] I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Res.* 49 (2021) W293-W296. <https://doi.org/10.1093/nar/gkab301>.
- [32] M. Richter, R. Rosselló-Móra, Shifting the genomic gold standard for the prokaryotic species definition, *Proc. Natl. Acad. Sci.* 106(45) (2009) 19126-19131.
- [33] T. Seemann, Prokka: rapid prokaryotic genome annotation, *Bioinformatics* 30(14) (2014) 2068-2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- [34] A.J. Page, C.A. Cummins, M. Hunt, et al., Roary: rapid large-scale prokaryote pan genome analysis, *Bioinformatics* 31(22) (2015) 3691-3693. <https://doi.org/10.1093/bioinformatics/btv421>.
- [35] Y.B. Zhao, X.M. Jia, J.H. Yang, et al., PanGP: a tool for quickly analyzing bacterial pan-genome profile, *Bioinformatics* 30(9) (2014) 1297-1299. <https://doi.org/10.1093/bioinformatics/btu017>.
- [36] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, *Nucleic Acids Res.* 39(suppl_2) (2011) W29-W37. <https://doi.org/10.1093/nar/gkr367>.
- [37] J. Russel, R. Pinilla-Redondo, D. Mayo-Muñoz, et al., CRISPRCasTyper: automated identification, annotation, and classification of CRISPR-Cas Loci, *CRISPR J.* 3(6) (2020) 462-469. <https://doi.org/10.1089/crispr.2020.0059>.
- [38] K.S. Makarova, Y.I. Wolf, O.S. Alkhnbashi, et al., An updated evolutionary classification of CRISPR-Cas systems, *Nat. Rev. Microbiol.* 13(11) (2015) 722-736. <https://doi.org/10.1038/nrmicro3569>.
- [39] W. Song, H.X. Sun, C. Zhang, et al., Prophage Hunter: an integrative hunting tool for active prophages, *Nucleic Acids Res.* 47 (2019) W74-W80. <https://doi.org/10.1093/nar/gkz380>.
- [40] D.J. Olive, Software for data analysis: programming with R, *Technometrics* 52(2) (2010) 261. https://doi.org/10.1007/978-0-387-75936-4_14.
- [41] D.S. Guttman, D.E. Dykhuizen, Clonal divergence in *Escherichia coli* as a result of recombination, not mutation, *Science* 266(5189) (1994) 1380-1383. <https://doi.org/10.1126/science.7973728>.
- [42] F. Turrioni, M. Ventura, L.F. Buttó, et al., Molecular dialogue between the human gut microbiota and the host: a *Lactobacillus* and *Bifidobacterium* perspective, *Cell. Mol. Life Sci.* 71(2) (2014) 183-203. <https://doi.org/10.1007/s00018-013-1318-0>.
- [43] S. Duranti, C. Milani, G.A. Lugli, et al., Insights from genomes of representatives of the human gut commensal *Bifidobacterium bifidum*, *Environ. Microbiol.* 17(7) (2015) 2515-2531. <https://doi.org/10.1111/1462-2920.12743>.
- [44] H. Tettelin, V. Masignani, M.J. Cieslewicz, et al., Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome", *Proc. Natl. Acad. Sci.* 102(39) (2005) 13950-13955. <https://doi.org/10.1073/pnas.0506758102>.
- [45] A. Mira, A.B. Martín-Cuadrado, G. D'auria, et al., The bacterial pan-genome: a new paradigm in microbiology, *Int. Microbiol.* 13(2) (2010) 45-57. <https://doi.org/10.2436/20.1501.01.110>.
- [46] C. Milani, G.A. Lugli, S. Duranti, et al., Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut, *Sci. Rep.* 5(1) (2015) 1-14. <https://doi.org/10.1038/srep15782>.
- [47] L.A. Bolte, A.V. Vila, F. Imhann, et al., Long-term dietary patterns are associated with pro-inflammatory and anti-inflammatory features of the gut microbiome, *Gut* 70(7) (2021) 1287-1298.
- [48] F. Asnicar, S.E. Berry, A.M. Valdes, et al., Microbiome connections with host metabolism and habitual diet from 1,098 deeply phenotyped individuals, *Nat. Med.* 27(2) (2021) 321-332.
- [49] Y.L. Wang, Y. Nie, D.H. Yu, et al., Genome-wide study of saprotrophy-related genes in the basal fungus *Conidiobolus heterosporus*, *Appl. Microbiol. Biotechnol.* 104(14) (2020) 6261-6272. <https://doi.org/10.1007/s00253-020-10698-6>.
- [50] K. Makarova, A. Slesarev, Y. Wolf, et al., Comparative genomics of the lactic acid bacteria, *Proc. Natl. Acad. Sci.* 103(42) (2006) 15611-15616. <https://doi.org/10.1073/pnas.0607117103>.
- [51] S. Arbolea, F. Bottacini, M. O'connell-Motherway, et al., Gene-trait matching across the *Bifidobacterium longum* pan-genome reveals considerable diversity in carbohydrate catabolism among human infant strains, *BMC Genomics.* 19(1) (2018) 1-16. <https://doi.org/10.1186/s12864-017-4388-9>.
- [52] P. Pajic, P. Pavlidis, K. Dean, et al., Independent amylase gene copy number bursts correlate with dietary preferences in mammals, *Elife* 8 (2019) e44628. <https://doi.org/10.7554/eLife.44628>.
- [53] F.M. Cerqueira, A.L. Photenhauer, R.M. Pollet, et al., Starch digestion by gut bacteria: crowdsourcing for carbs, *Trends Microbiol.* 28(2) (2020) 95-108. <https://doi.org/10.1016/j.tim.2019.09.004>.
- [54] F. Backhed, R.E. Ley, J.L. Sonnenburg, et al., Host-bacterial mutualism in the human intestine, *Science* 307(5717) (2005) 1915-1920.
- [55] S.W. Mintz, C.M. Du Bois, The anthropology of food and eating, *Ann. Rev. Anthropol.* 31(1) (2002) 99-119. <https://doi.org/10.1146/annurev.anthro.32.032702.131011>.
- [56] C. Dufresne, E. Farnworth, Tea, Kombucha, and health: a review, *Food Res. Int.* 33(6) (2000) 409-421. <https://doi.org/10.4315/0362-028x-63.7.976>.
- [57] P. Rozin, The meaning of food in our lives: a cross-cultural perspective on eating and well-being, *J. Nutr. Educ. Behav.* 37 (2005) S107-S112. [https://doi.org/10.1016/s1499-4046\(06\)60209-1](https://doi.org/10.1016/s1499-4046(06)60209-1).
- [58] R. Barrangou, C. Fremaux, H. Deveau, et al., CRISPR provides acquired resistance against viruses in prokaryotes, *Science* 315(5819) (2007) 1709-1712. <https://doi.org/10.1126/science.1138140>.
- [59] C. Hidalgo-Cantabrana, S. O'flaherty, R. Barrangou, CRISPR-based engineering of next-generation lactic acid bacteria, *Curr. Opin. Microbiol.* 37 (2017) 79-87. <https://doi.org/10.1016/j.mib.2017.05.015>.