Contents lists available at ScienceDirect

# Heliyon



journal homepage: www.cell.com/heliyon

Research article

5<sup>2</sup>CelPress

# Genomic and functional diversity of cultivated *Bifidobacterium* from human gut microbiota

Wenxi Li<sup>a,b</sup>, Hewei Liang<sup>b</sup>, Wenxin He<sup>b</sup>, Xiaowei Gao<sup>b</sup>, Zhinan Wu<sup>b,c</sup>, Tongyuan Hu<sup>b</sup>, Xiaoqian Lin<sup>a,b</sup>, Mengmeng Wang<sup>b,c</sup>, Yiyi Zhong<sup>b,d</sup>, Haifeng Zhang<sup>b,d</sup>, Lan Ge<sup>b,d</sup>, Xin Jin<sup>b</sup>, Liang Xiao<sup>b,c,f,g</sup>, Yuanqiang Zou<sup>b,e,f,g,\*</sup>

<sup>a</sup> School of Biology and Biological Engineering, South China University of Technology, Guangzhou, 510006, China

<sup>b</sup> BGI Research, Shenzhen, 518083, China

<sup>c</sup> College of Life Sciences, University of Chinese Academy of Sciences, Beijing, 100049, China

<sup>d</sup> BGI Precision Nutrition (Shenzhen) Technology Co., Ltd, Shenzhen, China

e Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Universitetsparken 13, 2100, Copenhagen,

Denmark

<sup>f</sup> BGI Research, Qingdao, 266555, China

<sup>g</sup> Shenzhen Engineering Laboratory of Detection and Intervention of human intestinal microbiome, BGI-Shenzhen, Shenzhen, China

# ARTICLE INFO

Keywords: Bifidobacterium Gut microbiome Bile salt hydrolase Phages Horizontal gene transfers

# ABSTRACT

The genus Bifidobacterium widely exists in human gut and has been increasingly used as the adjuvant probiotics for the prevention and treatment of diseases. However, the functional differences of Bifidobacterium genomes from different regions of the world remain unclear. We here describe an extensive study on the genomic characteristics and function annotations of 1512 genomes (clustered to 849 non-redundant genomes) of Bifidobacterium cultured from human gut. The distribution of some carbohydrate-active enzymes varied among different Bifidobacterium species and continents. More than 36% of the genomes of B. pseudocatenulatum harbored biosynthetic gene clusters of lanthipeptide-class-iv. 99.76% of the cultivated genomes of Bifidobacterium harbored genes of bile salt hydrolase. Most genomes of B. adolescentis, and all genomes of B. dentium harbored genes involved in gamma-aminobutyric acid synthesis. B. longum subsp. infantis were characterized harboring most genes related to human milk oligosaccharide utilization. Significant differences between the distribution of antibiotic resistance genes among different species and continents revealed the importance to use antibiotics precisely in the clinical treatment. Phages infecting Bifidobacterium and horizontal gene transfers occurring in genomes of Bifidobacterium were dependent on species and region sources, and might help Bifidobacterium adapt to the environment. In addition, the distribution of Bifidobacterium in human gut was found varied from different regions of the world. This study represents a comprehensive view of characteristics and functions of genomes of cultivated Bifidobacterium from human gut, and enables clinical advances in the future.

\* Corresponding author. BGI-Shenzhen, Beishan Industrial Zone, Shenzhen, 518083, China. *E-mail address:* zouyuanqiang@genomics.cn (Y. Zou).

#### https://doi.org/10.1016/j.heliyon.2024.e27270

Received 27 August 2023; Received in revised form 14 February 2024; Accepted 27 February 2024

Available online 2 March 2024

<sup>2405-8440/</sup><sup>©</sup> 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

#### 1. Introduction

The genus *Bifidobacterium*, which belongs to Actinomycetota, is one of the dominant members of microbes in the human gut, particularly during the early life [1]. *Bifidobacterium* makes up 60–70% of infant gut microbes and approximately 10–15% of adult gut microbes [2]. A higher average *Bifidobacterium* level was observed in Japanese adults than individuals from other countries [3]. Based on the development of culture-based methods and metagenomic analyses, the distribution and function of *Bifidobacterium* in the human gut have been studied in a lot [4,5] and their relationships with human health have been investigated. Large-scale culture-based collections of human gut microbiota have been established and have provided a large number of bacterial strains and reference genomes of *Bifidobacterium* which could be used for clinical research [6–9].

*Bifidobacterium* are closely associated with human health. The phylogenomics of *Bifidobacterium* revealed relatively less ancestral gene loss occurrences and more gene acquisition events during the evolution development of this genus [10]. Antibiotic resistance genes are often carried by gene acquisitions [11]. The ability to degrade diet- and host-derived carbohydrate solidified the long-term colonization of *Bifidobacterium* in human intestine [12]. Most strains of *Bifidobacterium* can consume monosaccharides, disaccharides, and oligosaccharides, but there are still many variations in the metabolism of other carbohydrates among different species [13,14]. Through bile salt hydrolase (BSH), conjugated bile acids are transformed into free bile acids, which is important for lowering cholesterol and preventing cardiovascular diseases (CVD) [15]. The genus *Bifidobacterium* is considered one of the BSH-active bacteria that could be utilized as a combination of therapeutic probiotics for CVD in the future [16]. Gamma-aminobutyric acid (GABA) is an important inhibitory neurotransmitter with an anti-depressive function [17]. It can regulate Parkinson's disease [18], Huntington's chorea [19] and Alzheimer's disease [20]. *Bifidobacterium adolescentis* has been confirmed to synthesize GABA and has the potential to modulate the gut-brain axis response [21]. The *Bifidobacterium longum* subspecies *infantis* (*B. longum* subsp. *infantis*), which harbors genes encoding enzymes involved in human milk oligosaccharides (HMO) degradation, can regulate the immune system of infants in their early lives [22], and can be applied to the treatment of acutely malnourished infants [23]. Other species of *Bifidobacterium* have also been found to have antibacterial activities, which may prevent the endogenous *Escherichia coli* infection [24].

Although *Bifidobacterium* has been utilized in the treatment of various diseases as a probiotic for a long time [25,26], the underlying mechanisms of its probiotic effects still need to be explored. Evidence has revealed that people from different nationalities have different enterotypes of the gut microbiome, and the functions of the gut microbiome also varied [27]. The functional differences of cultivated *Bifidobacterium* among different nationalities also need extensive exploration. To gain a deeper understanding of *Bifidobacterium* in the human gut, we collected 1512 *Bifidobacterium* genomes from three large databases of human gut cultivated microbiota and performed an extensive analysis of their functions, infecting phages, and horizontal gene transfer (HGT) to explore the functional differences of *Bifidobacterium* strains among different regions. We then analyzed the distribution of these *Bifidobacterium* species in 12, 461 metagenomes of three large cohorts from different regions.

# 2. Materials and methods

#### 2.1. Collection of cultivated genomes of Bifidobacterium from human gut

We collected 451 *Bifidobacterium* genomes from the expanded version of Cultivated Genome Reference (CGR2) [6], 851 genomes from the Broad Institute-OpenBiome Microbiome Library (BIO-ML) [9], and 210 genomes from the Unified Human Gastrointestinal Genome (UHGG) [7] (Supplementary Table 1). The genomes of CGR2 can be accessed from the CNGB Sequence Archive (CNSA) [28] of the China National GeneBank DataBase (CNGBdb) [29] with accession number CNP0001833. The genomes of BIO-ML were accessed from https://www.ncbi.nlm.nih.gov/bioproject/PRJNA544527. The genomes of UHGG were downloaded from http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify\_genomes/human-gut/. Genome quality was evaluated using CheckM (v1.0.12) [30].

# 2.2. Phylogenetic, and taxonomic determination, and genomes selection

GTDB-TK (v2.1.0) [31] with the database Release 214 [31] was used to perform taxonomic annotation of each genome and construct a maximum-likelihood phylogenetic tree based on 120 conserved single-copy genes. The pairwise alignment average nucleotide identity (ANI) was calculated using fastANI (v1.32) [32]. The phylogenetic tree was visualized using iTOL (v6.5.6, https://itol.embl.de/). Multi-lineage taxonomy was only considered when calculating the proportions of species in metagenomes.

To avoid the bias of *Bifidobacterium* functions raised by identical genomes, we clustered according to the ANI values using R package hclust (with threshold <99.99%). In each cluster, the genomes with the largest genome sizes were retained. For the genomes of the same cluster but from different sources, the genomes with the largest genome sizes from each continent were retained. Therefore, 849/1512 genomes were retained for functional analysis.

## 2.3. Functional annotation of genomes

The protein-coding sequences (CDS) of each genome were predicted and annotated using Prokka v1.14.6 [33]. Preliminary functional annotation was performed using eggNOG-mapper v2 [34] (eggNOG database version 5.0.2 [35]). The results of COG, EC number, and KO annotation were extracted from the eggNOG-mapper results and counted by functional category.

The genes encoding CAZymes were predicted using dbCAN database [36] (downloaded from http://bcb.unl.edu/dbCAN2/ download/Databases/V11/CAZyDB.08062022.fa). Diamond v0.8.22 [37] (blastp) was used for mapping the CDS to the database, with the identity above 60% and coverage above 50%.

Biosynthetic gene clusters (BGCs) were identified using the antiSMASH 6.0 [38]. The parameters –cb-general –cb-knownclusters –cb-subclusters were used to blast the gene sequences to known classification of stimulated metabolism, and –smcogs were used to analyze the family of secondary metabolic genes. The predicted BGCs were mapped against the MiBIG database [39] to characterize the BGCs with >70% identity as known functions.

We used BLAST (Blastn, v2.9.0) [40] to identify the genes encoding bile salt hydrolases (BSH), mapping to the BSH reference sequence provided by Déjean et al. [41], with identity >95% and coverage >75%. We used EC number 3.5.1.24 to identify the *cbh* gene. And we used EC number 4.1.1.15 and KO number K20265 to identify the *gadB* and *gadC* gene, respectively. The method of identifying HMO-related protein sequences were referred to Sakanaka et al. and the reference sequences were downloaded from each accession number listed by the article [42]. Occurrence of the HMO degrading genes were examined using tblastn v2.5.0+ [40], with the identity >60% and query coverage >50%, e value < 1e-5.

#### 2.4. Annotation of ARGs and VFs

The "main" feature with default parameter of Resistance Gene Identifier (RGI) version 5.2.0 and the Comprehensive Antibiotic Resistance Database (CARD [43], version 3.1.2) was used to annotate ARGs. The VFs annotation of all CDS was performed by BLAST v2.2.26 (-evalue 0.01) against the Virulence Factor Database (VFDB [44], setB, 2021-07) with identity higher than 60% and coverage higher than 50%.

#### 2.5. Phage sequences detection and functional annotation

To study the phages infecting *Bifidobacterium*, the contigs of genomes were scanned using VirSorter2 (v2.2.3) [45] to identify viral-like sequences, and sequences with score >0.7 will be selected. The sequences defined as "not-determined" by CheckV (v0.8.1) [46] were eliminated. The phage proteins were predicted by Prodigal (v2.6.3) [47], and vConTACT2 (v0.11.3) [48] was used to group the viral proteomes into viral clusters (VCs). We then used eggNOG-mapper v2 [34] (eggNOG database version 5.0.2 [35]) for functional annotation of the viral genomes. PFAMs annotation results were extracted from the eggNOG-mapper results.

#### 2.6. Detection of horizontal gene transfer (HGT)

In this study, we focused on horizontal gene transfer occurring between *Bifidobacterium* and other genera. We performed the detection of HGT as described by Mathieu Groussin et al. [49]. HGTs occurring within *Bifidobacterium* species were not considered. Diamond v2.0.11.149 [37] (blastp) was used to detect potentially transferred genes. Protein database file "hgtdb\_20211121.tar. xz" was downloaded from https://arizonastateu-my.sharepoint.com/:f:/g/personal/qzhu44\_asurite\_asu\_edu/ErLl2qExtFhAiS1J0sCpZq gBEebKHtBiJ1IDlaitOVZXg, which included all protein sequences of NCBI RefSeq [50] genomes of bacteria. BLAST hits that were larger than 500 bp, with 100% similarity, 100% coverage, and e-value < 1e-5 were selected. The hit with the highest score for each gene was retained. The COG annotations of transferred genes were extracted from result of eggNOG annotations.

### 2.7. Calculation of the relative abundance of 15 species in metagenomes

We downloaded 3550 gut metagenomes of a Chinese cohort (part of 4D-SZ) [51] from the CNGB Sequence Archive (CNSA) [28] of the China National GeneBank DataBase (CNGBdb) [29] with accession number CNP0000426; 8244 metagenomes from a cohort established in the Netherlands [52], retrieved from the European Genome-Phenome Archive (EGA, https://ega-archive.org/) [53] under accession number EGAS00001005027; and 667 metagenomes from HMP (Human Microbiome Project [54], https://portal.hmpdacc.org/). Fastp (v0.23.1) [55] was used to filter out low-quality reads and bases with partial parameters '-qualified\_quality\_phred 15 -complexity\_threshold 30 -length\_required 30'. Bowtie (v2.4.4) [56] was used to remove host contamination by mapping the reads to the human genome (GRCh38). The default parameter (-sensitive) was used, and '-un-conc-gz' was used to output the sequence file after removing host contamination. Considering that there was a multi-lineage species in the taxonomic annotation, we selected 15 genomes of each species with the longest genome was regarded as a bacterial genome reference in the Kraken2 [57] (v 2.1.1) database, and the combination of Kraken2 and Bracken [58] was used to estimate the abundance of representative genomes.

# 2.8. Statistical analysis

Statistical tests were performed using R v4.1.2. For principal co-ordinates analysis (PCoA), Bray-Curtis dissimilarities were calculated using the vegdist function. The Wilcoxon test was used as the significance test. The packages ggplot2 and pheatmap in R were used for plotting. Venn diagrams were plotted using Venn software (jvenn, http://jvenn.toulouse.inra.fr/app/example.html). Adobe Illustrator CC 2018 was used to adjust the colors and construct figures.

# 3. Results

# 3.1. The genome collection of cultivated Bifidobacterium in human gut

In our previous work (CGR2), we obtained 453 *Bifidobacterium* strains from Chinese individuals using 25 anaerobic culture conditions, including modified peptone-yeast extract-glucose (MPYG) medium and blood-brain heart infusion (BHI) medium [8], proving that *Bifidobacterium* can be cultured in a variety of media. Two genomes with completeness of less than 95% in CGR2 were excluded. To extensively investigate the genomic characteristics and functional diversity of cultivated *Bifidobacterium* in the human gut, we further



**Fig. 1.** Genomic diversity of cultivated *Bifidobacterium* in human gut. **a**, Phylogenetic tree of 1512 cultivated genomes of *Bifidobacterium*, based on GTDB annotation. The innermost circle is colored according to the species. The second circle is colored according to the continents from which strains were isolated. The third circle is colored according to the databases from which genomes were obtained. The fourth circle shows the GC content of genomes and the blue line represents a percentage of 62%. The outermost circle represents the genome length, with the red dashed line showing a genome size of 1.9 Mb and the light blue solid line showing a genome size of 2.5 Mb **b**, The number of genomes of *Bifidobacterium* showing genome size and GC content. The cloud arcording to the species. **c**, Scatterplot of genomes of 849 non-redundant cultivated *Bifidobacterium* showing genome size and GC content. The cloud around points represents a simple encirclement of points, using geom\_encircle in ggplot2 package of R. **d**, Principal Co-ordinates Analysis (PCoA) of taxonomic annotation of 849 *Bifidobacterium* genomes, based on ANI values among the genomes (p = 0.001,  $R^2 = 0.99917$ , calculated using Adonis test in R). The clouds around the clusters represents a 95% confidence interval. Panel **b**, **c**, and **d** are colored according to the species shown in panel **a**.

downloaded the *Bifidobacterium* genomes from the Unified Human Gastrointestinal Genome (UHGG) [7] and the Broad Institute-OpenBiome Microbiome Library (BIO-ML) [9], the two largest databases of human gut-cultivated microbes. Three databases were used to generate an integrated catalog of *Bifidobacterium* genomes (1,512) of the human gut (Supplementary Table 1). Genomes



**Fig. 2.** Functional profile of cultivated *Bifidobacterium* (for 849 non-redundant genomes). **a**, The prevalence of some CAZyme families in *B. adolescentis, B. pseudocatenulatum*, and *B. longum* subsp. *longum*, colored according to the continent. The CAZymes with obvious differences among different continents in each species are shown. The prevalence represents the number of genomes harboring corresponding CAZyme relative to the total number of corresponding species in this continent. **b**, The distribution of BGCs in different species. The pie charts show the distribution of continents from which the genomes were obtained. The color of continents is shown in panel **a**. **c**, The distribution of genes related to GABA producing and bile salt hydrolysis harbored by each species. The sizes of bubbles and the number shown in the bubbles represent the prevalence of these genes in each species.

with >95% completeness and <5% contamination, according to CheckM [30], were retained (Supplementary Fig. 1a).

To further examine the taxonomy of the genomes, we annotated them using the Genome Taxonomy Database [31] (GTDB, https://gtdb.ecogenomic.org/). These 1512 genomes were classified into 14 species (including subspecies) (Fig. 1a, Supplementary Table 1). In addition to genomes with unknown regions, 981 genomes were obtained from 13 countries of four continents. Most of the *Bifidobacterium* genomes from Asia were contributed by CGR2. Although the number of genomes of *Bifidobacterium* in UHGG was the lowest, it contributed the greatest diversity of *Bifidobacterium* species among the three databases (Fig. 1b). The average nucleotide identity (ANI) values were calculated to estimate the similarity among the 1512 genomes. To avoid the bias of *Bifidobacterium* functions raised by identical genomes, we removed some redundant genomes (ANI >99.99%) and retained 849 representative genomes for further genomic and functional analysis.

The genome sizes of *Bifidobacterium* ranged from 1.9 Mb to 2.8 Mb and the GC contents ranged from 54.66% to 63.24%. The genomes of the same species tended to have similar genome sizes and GC contents. *B. animalis* subsp. *animalis* and *B. catenulatum* subsp. *catenulatum* had relatively smaller genomes, whereas the genome size of *B. longum* subsp. *infantis* was larger. *B. pseudolongum* subsp. *globosum* and *B. bifidum* had higher GC contents, whereas *B. catenulatum* subsp. *catenulatum*, *B. catenulatum* subsp. *kashiwanohense* and *B. pseudocatenulatum* had lower GC contents (Fig. 1c). The genomes of the same species clustered together according to the ANI (Fig. 1d). *B. longum* subsp. *infantis*, a subspecies of *B. longum*, had the closest relatives to *B. longum* subsp. *longum*, but with an ANI <95%. Most of the ANI between *B. animalis* subsp. *animalis* and other species. Significantly higher ANI values between genomes within the same continent were observed among *B. adolescentis*, *B. longum* subsp. *longum*, *B. pseudocatenulatum*, *B. catenulatum* subsp. *catenulatum* subsp. *catenulatum* subsp. *catenulatum* subsp. *catenulatum* subsp. *catenulatum* subsp. *animalis* subsp. *animalis* subsp. *catenulatum* subsp. *longum*, *B. catenulatum* subsp. *catenulatum* subsp. *cat* 

# 3.2. The functional variations of Bifidobacterium from human gut among species and continents

Bifidobacteria were considered able to metabolize the mono-, di- and oligo-saccharides in the gut, but they were unlikely able to metabolize most of the large, complex plant polysaccharides on their own [59]. We then extensively explored the functional diversity of 849 genomes of cultivated Bifidobacterium in the human gut. Regarding the prediction of carbohydrate-active enzyme (CAZyme) genes, genes belonging to GH13\_9 (α-1,4-glucan branching enzyme), GH13\_11 (amylomaltase/α-glucosidase), GH13\_31 (α-transglucosidase/α-amylase), CBM48 (appended to GH13 modules), GT35 (glycogen or starch phosphorylase), GH32 (hexosyltransferases), GH36 (a-galactosidases), GH77 (amylomaltase/4-a-glucanotransferase), GT2 (mannosyltransferase), GT28 (digalactosyldiacylglycerol synthase), GT35 ( $\alpha$ -1,4-glucan phosphorylase), GT4 (sucrose synthase) and GT51 (peptidoglycan glycosyltransferase) are widely present in the genomes of Bifidobacterium (Supplementary Fig. 2). Some CAZymes are harbored by specific Bifidobacterium species. For example, GH30\_5 (endo-β-1,6-galactanase), GH43\_16 (xylan β-1,4-xylosidase), CBM22 (xylan binding) and CBM6 (cellulose-binding) are mostly harbored by B. longum subsp. longum. Nearly all the genomes of B. bifidum harbor genes belonging to GH110 (a-galactosidase), GH84 (hyaluronidase) and GH89 ( $\alpha$ -N-acetylglucosaminidase), which are barely present in other species. GH101 (Glycopeptide  $\alpha$ -N-acetylgalactosaminidase) was present in most of the genomes of *B. bifidum* and *B. longum* subsp. *longum*, but barely present in genomes of other species. The distribution of CAZymes also varied among the different continents. For B. adolescentis, genes belonging to GH121 ( $\beta$ -L-arabinobiosidase) and GH43 29 ( $\alpha$ -1,2-L-arabinofuranosidase) were present in over 50% of genomes from Asia but not a genome from North America. And most of the genomes from Europe and North America harbor GH1 (α-L-arabinopyranosidase) and CE9 (N-acetylglucosamine-6-phosphate deacetylase) while less than 50% of Asian genomes harbor them. For B. pseudocatenulatum, GT111 ( $\beta$ -1,3-galactofuranosyltransferase) and GH13\_44 ( $\alpha$ -glucosidase) are harbored by most of the genomes from Europe and Asia, while GH43\_22 (α-1,3-L-arabinofuranosidase) is harbored by about 75% of the genomes from Europe and North America. For B. longum subsp. longum, GH95 and GH29 (both α-L-fucosidase) are mostly found in genomes from North America, whereas GH43 4 and GH43 26 (both exo- $\alpha$ -1,5-L-arabinofuranosidase) are mostly present in genomes from Europe and unknown region. GH30 (β-1,6-glucosidase), GH146 (β-L-arabinofuranosidase) and GH136 (lacto-N-biosidase) could be found in less genomes from North America (Fig. 2a). We surmised that the differences in CAZymes harbored by Bifidobacterium might result from regional dietary differences (Supplementary Table 3a).

An exploration of biosynthetic gene clusters (BGCs) was performed to predict the secondary metabolites produced by *Bifidobacterium*. Forty-nine non-redundant BGCs were annotated in the 131 genomes (Fig. 2b, Supplementary Fig. 3). Lanthipeptide-class-iv was the most abundant type of BGCs, followed by non-ribosomal peptide synthetase-like (NRPS-like), both of which are mostly produced by *B. pseudocatenulatum*. Lanthipeptide-class-iv was present in the genomes from Asia and Europe but absent in the genomes from North America. NRPS-like BGCs were mostly present in Asian genomes. This suggests that the production of secondary metabolites also varies among continents. Among these BGCs, *B. longum* subsp. *infantis* may have the ability to synthesize the most diverse products, which was consistent with previous research [60]. More than 36% of *B. pseudocatenulatum* were predicted to harbor gene clusters involved in synthesizing class IV lanthipeptide. The BGCs of different classes of lanthipeptide are harbored by different *Bifidobacterium* species. Lanthipeptides, also termed lantibiotics, are resistant to pathogens such as *Staphylococcus aureus*, and have low toxicity to mammals [61]. Owing to the overuse of antibiotics produced by microbes are much safer and have the potential to replace antibiotics as a new type of antibacterial drug in the future [62]. The lanthipeptides are classified into different classes because of their distinct mechanisms of the dehydration reactions [63], and most of current research has focused on class I and class II lanthipeptides [64,65]. Class IV lanthipeptides are also called as venezuelins, with unclear functions as present [66]. And our study showed that *B. pseudocatenulatum* has the potential to produce class IV lanthipeptides, which might contribute to clinical advances (Supplementary Table 3b).

We then identified the genes related to BSH and GABA production of cultivated Bifidobacterium in the human gut (Fig. 2c,

Supplementary Table 4a). A total of 847 of 849 genomes of *Bifidobacterium* in the human gut harbored genes encoding BSH, and 848 genomes harbored the *cbh* gene. Only one genome in our study did not harbor genes encoding any bile salt hydrolases, indicating that most of *Bifidobacterium* strains in this collection have the potential to be used for the treatment and prevention of CVD. Glutamate decarboxylase (GAD) is the sole enzyme for GABA synthesis [67], which is encoded by the genes *gadB* and *gadC*. In our study, 183 of the 186 *B. adolescentis* genomes harbored these two genes. In addition to *B. adolescentis*, all genomes of *B. dentium* also harbored these two genes, which indicated that they might also have the potential to synthesize GABA. In addition, all genomes of *B. angulatum*, *B. animalis* subsp. *animalis*, *B. gallicum* and *B. pseudolongum* subsp. *globosum* harbored only the gene *gadC*.

B. longum subsp. infantis is a well-known major member of HMO utilization [68]. Here, we analyzed the presence of proteins related



**Fig. 3.** Annotation of ARGs in the genomes of *Bifidobacterium* (for 849 non-redundant genomes). **a**, The distribution of annotated ARGs in *Bifidobacterium*. The genomes are annotated by their species and continents and the antibiotics are colored according to the importance of drug usage. **b**, The boxplot of number of ARGs harbored in genomes of different species, colored by their species. Each point represents a genome with annotated ARGs. **c**, The boxplot of number of ARGs harbored in genomes of different continents, colored by corresponding continent. Only the significant p-values are shown in the figure, calculated by paired wilcoxon test in R.

to HMO metabolism using tblastn v2.5.0+ [40]. We investigated the distribution of 26 representative amino acid sequences listed by Sakanaka et al. [42]. Genes encoding enzymes involved in extracellular HMO digestion were mostly harbored by *B. bifidum*. The genes encoding AfcA, AfcB, BbgIII, BbhI, and Siabb2 were conserved in *B.bifidum* strains, and LnbB was also harbored by most of the *B. bifidum* and *B. breve* strains. Strains of *B. bifidum*, *B. breve*, *B. longum* subsp. *infantis*, and *B. longum* subsp. *longum* harbored enzymes involved in uptake of LNB and intracellular LNB digestion. *B. longum* subsp. *infantis* harbored the most genes encoding enzymes related to uptake of intact HMOs and intracellular HMO digestion. Genes encoding fucosyllactose transporter solute-binding proteins (FL transporter SBP), lacto-N-tetraose (LNT) beta-1,3-Galactosidase, beta-N-Acetylglucosaminidase, 1,2-alpha-l-Fucosidase, 1, 3/4-alpha-l-Fucosidase, and 2,3/6-alpha-Sialidase were harbored by all the *B. longum* subsp. *infantis* strains. These results were consistent with the previous research [42]. In addition, the prevalence of HMO-related enzymes might vary among continents. The prevalence of *B. bifidum* harboring LnbP and LacZ6 from Europe were lower than that from other continents. Blon\_2335, Blon\_2336 and Blon\_2202 were harbored by many *B. longum* subsp. *longum* strains of North America while LnbX and LnbY were harbored by the strains of other continents. And the prevalence of *B. pseudocatenulatum* harboring Blon\_0459 or NahA were lower in the strains from North America than that from other continents. (Supplementary Fig. 4, Supplementary Table 4b).



Fig. 4. Taxonomic and functional annotation of phages infecting *Bifidobacterium* (for 849 non-redundant genomes). **a**, Venn diagram and histogram showing distribution of VCs infecting *Bifidobacterium* of different species, colored according to their species. **b**, Venn diagram and histogram showing distribution of PFAMs of phages, colored according to their species. **c**-**e**, Venn diagram showing the distribution of PFAMs encoded by phages infecting *B. longum* subsp. *longum* (**c**), *B. pseudocatenulatum* (**d**), and *B. adolescentis* (**e**) from different continents, colored according to the continents.

#### 3.3. The safety evaluation of cultivated genomes of Bifidobacterium in human gut

In recent years, a comprehensive understanding of antibiotic resistance is urgently needed to prevent the antibiotic-resistant genes spreading [69]. We identified 57 types of antibiotic resistance genes (ARGs) in 847 genomes with resistance to 24 types of antibiotics. 24 of the 57 types of ARGs were multidrug resistant (Supplementary Fig. 5a, Supplementary Table 5a). Many antibiotics are important antimicrobials listed by the World Health Organization (WHO) [70]. Most of the genomes harbored rifamycin-resistance genes. In addition, almost all the genomes of B. adolescentis, B. catenulatum subsp. catenulatum, B. dentium and B. pseudocatenulatum harbored genes resistant to tetracycline, macrolide, aminocoumarin, and monobactam (Fig. 3a). Among the cultivated Bifidobacterium in the human gut, the genomes of B. catenulatum subsp. kashiwanohense and B. catenulatum subsp. catenulatum harbored the most ARGs. Of note, we found a genome of B. adolescentis (GCA\_009094615.1\_ASM909461v1\_genomic) harboring 37 ARGs (Fig. 3b), which might be threatening and unsuitable for application in clinical treatment. For B. longum subsp. longum, the number of North American genomes with genes resistant to penam, cephalosporin, and cephamycin was higher than that of Asian genomes, while genes resistant to lincosamide, streptogramin, tetracycline, and macrolides were mostly harbored by Asian genomes. This indicated that resistance to antibiotics could differ among the continents of the genomes, even from the same species. Significant differences in the number of ARGs in each genome were observed among Europe, Asia, and North America (Fig. 3c, Supplementary Fig. 5b). Evidence has revealed variations in the consumption rates of antibiotic classes among different regions [71]. This suggests that the composition of ARGs in Bifidobacterium might be affected by the use of different antibiotics on different continents. Therefore, it is essential to use antibiotics distinctly and properly according to the resistance of *Bifidobacterium* genomes from different continents if they are used as probiotics.

The virulence of bacteria is another aspect of safety evaluation. To deeply and comprehensively confirm the potential virulence factors carried by *Bifidobacterium*, we analyzed the virulence factors in the 849 *Bifidobacterium* genomes. We identified 11 types of virulence factors (VF) in 36 genomes of cultivated *Bifidobacterium*, annotated with the Virulence Factor Database [72] (VFDB, http://www.mgc.ac.cn/VFs/) (Supplementary Fig. 6, Supplementary Table 5b). Ten out of the eleven VFs were related to evading the host immune system [73], bacterial adhesion, and survival [74], which may not be that threatening to the host. Hyaluronidase, which was harbored by only one genome (GCA\_009883615.1\_ASM988361v1\_genomic), is a pathogenic bacterial spreading factor, cleaves hyaluronan and may also pave the way for other bacterial toxins [75]. This indicated that most of the *Bifidobacterium* genomes in this study, except those harboring virulence factors, are safe for clinical use.

### 3.4. Identifying the phages infecting Bifidobacterium

The coevolution of bacteria and phages influences the survival of microbes and evolution of microbial communities [76] and the apparently extensive interactions between *Bifidobacterium* and phages were identified [77]. Phages are closely associated with human health and have both beneficial and harmful sides [78]. We found 545 non-singleton phages genomes in 361 *Bifidobacterium*, divided into 38 viral clusters (VCs) (Supplementary Table 6). Among the 38 VCs, 8 could infect all 9 species of this collection, and 34 could infect *B. longum* subsp. *longum*. *B. longum* subsp. *longum* contributed six unique VCs and *B. catenulatum* subsp. *catenulatum* contributed one unique VC (Fig. 4a). This indicated that the VCs might be similar among *Bifidobacterium* in the human gut. A total of 1275 kinds of protein families (PFAMs) were annotated from the phage genomes, and 461 kinds of PFAMs were shared among phages infecting 9



**Fig. 5.** HGTs occurring in *Bifidobacterium* genomes (for 849 non-redundant genomes). **a**, The number of detected HGT genes classified based on the taxa supposed to be the source of the transfer. **b**, The number of genes derived from HGT events according to the COG annotation, colored by the COG types. **c**, The percentage of genes derived from HGT events in each genome, classified and colored according to the species. **d**, The percentage of genes derived from HGT events in each genome of *B. adolescentis, B. bifidum*, and *B. longum* subsp. *longum*, and *B. pseudocatenulatum*, classified and colored according to continents. *P*-values are shown in the figure, calculated by paired wilcoxon test in R.

*Bifidobacterium* species. The phages infecting *B. longum* subsp. *longum* contributed 146 types of unique PFAMs, whereas phages infecting *B. adolescentis* contributed 61 unique PFAMs (Fig. 4b). GrpE, which participates in the response to hyperosmotic stress and heat shock in association with DnaK, was only found in phages infecting *B. longum* subsp. *longum*. PemK\_toxin of the type II toxin-antitoxin system was unique in *B. longum* subsp. *longum*. ParE\_toxin, which is toxic to DNA gyrase, is unique to *B. adolescentis*. This suggests that in these species, phages might have the potential to help the bacteria cope with toxins or compete with other species, promoting the survival of the bacteria.

PFAMs of phages infecting *Bifidobacterium* in the human gut differ among continents. A total of 382 PFAMs were shared among phages infecting *B. longum* subsp. *longum* of Asia, Europe, and North America (Fig. 4c). PFAMs of GH129 ( $\alpha$ -N-acetylgalactosaminidase) and GH10 (endo-1,4- $\beta$ -xylanase) were found only in phages infecting *B. longum* subsp. *longum* in Asia, suggesting that dietary differences across different continents might influence the functions of phages infecting *Bifidobacterium* in the human gut. For phages infecting *B. pseudocatenulatum*, 202 unique kinds of PFAMs were contributed by the genomes from Asia, and 87 types of PFAMs were shared in genomes from the three continents (Fig. 4d). For phages infecting *B. adolescentis*, genomes from North America contributed 125 unique types of PFAMs, while genomes from Asia contributed 103 unique types of PFAMs. A total of 312 PFAMs types were shared by the genomes from Asia, Europe, and North America (Fig. 4e). This suggests that the PFAMs of phages infecting these three species vary among continents.

# 3.5. The occurrence of horizontal gene transfer in Bifidobacterium

Horizontal gene transfer (HGT) [79] is widely recognized as an important mechanism for bacterial adaptation [80]. Here, we searched for *Bifidobacterium* genes that might be transferred from other microbial taxa. In total, 2.45% of the pangenomes of the 849 *Bifidobacterium* strains were potentially transferred (Supplementary Table 7). Most of the genes were mapped to Terrabacteria and Bacteria. In addition to Actinobacteria, HGTs were also frequently found between *Bifidobacterium* and Firmicutes. *Enterococcus* was the genus that was the source of most of the HGT genes detected in *Bifidobacterium*, followed by *Escherichia, Bacteroides* and *Streptomyces* (Fig. 5a). These genera are common in the human gut, reflecting the potentially frequent gene exchange among them in the human gut. Regarding the functions annotated with the Clusters of Orthologous Genes (COG) database, most of the potentially transferred genes were related to information storage and processing (Fig. 5b), even though the genes with metabolic functions accounted for a large proportion of the *Bifidobacterium* genomes.



**Fig. 6.** Distribution of cultivated *Bifidobacterium* in three different cohorts. Boxplot represents the proportion of each species in the abundance of these 15 *Bifidobacterium* species in a sample. Each point represents a sample, colored according to the cohorts. Significant differences between the proportion of the same species of different cohorts are marked on the right of the boxplot, calculated by paired wilcoxon test in R (\*\*\* represents P < 0.001).

*B. longum* subsp. *longum* had the highest proportion of genes acquired by HGT events, with a median of 3.60%, indicating that more gene exchanges with other genera occurred in *B. longum* subsp. *longum* than in other *Bifidobacterium* species (Fig. 5c). In addition, significant differences of the percentage of HGT occurrence in each genome were observed among genomes from different continents in specific species. Among *B. adolescentis*, HGTs that occurred in the genomes from North America and Asia were significantly higher than the genomes from Europe among *B. adolescentis* and *B. bifidum*. And among *B. longum* subsp. *longum*, HGTs that occurred in the genomes from North America were significantly higher than those of Europe and Asia (Fig. 5d). This suggests that the occurrence and frequency of HGT events could vary among different region sources, even for the same species, which might result from the difference in life habits and composition of the gut microbiome among different continents. But this difference could not be observed among *B. pseudocatenulatum*.

Notably, some potentially transferred genes are related to virulence and resistance. For example, *mprF* gene acquired from *Enterococcus* species was found harbored by *B. longum* subsp. *longum*. This gene is involved in lysylphosphatidylglycerol (L-PG) production and contributes to bacterial virulence [81], which can help microbes escape from the host immune system. In addition, genes encoding proteins related to the (helix\_turn\_helix) arsenic (As) resistance operon repressor and (helix\_turn\_helix) mercury (Hg) resistance were found in the genomes of *B. pseudocatenulatum* and *B. longum* subsp. *longum* respectively, both obtained from North America. These genes are potentially acquired from Firmicutes and Lactobacillaceae. Evidence has shown that both mercury contamination [82] and arsenic contamination [83] exist in North America. This implies that HGT might help *Bifidobacterium* from North America cope with Hg and As contamination and adapt to the environment.

#### 3.6. Distribution of cultivated genomes of Bifidobacterium from different populations

To explore the distribution of the cultivated *Bifidobacterium* in the human gut from different regions, we mapped 15 representative genomes of each species (including multi-lineage species *B. catenulatum* subsp. *kashiwanohense\_A*) collected in this work to 3550 metagenomes from China [51], 8244 metagenomes from the Netherlands [52] and 667 metagenomes from HMP (Human Microbiome Project [54], https://portal.hmpdacc.org/). After normalization, we obtained the proportion of each species in three population cohorts. *B. adolescentis* and *B. longum* subsp. *longum* occupied the most proportion of *Bifidobacterium* in the Dutch individuals' gut, accounting for 21.91% and 17.92% respectively (Fig. 6). *B. longum* subsp. *longum* and *B. pseudocatenulatum* occupied the most proportion of *Bifidobacterium* in the gut of whom participating in Chinese cohort (10.16% and 18.62% respectively) or HMP cohort (12.52% and 12.46% respectively). Significant differences existed in the proportion of each species in three cohorts, indicating that the distribution of *Bifidobacterium* in gut of people in different area might vary a lot. Given that different *Bifidobacterium* species could metabolize different kinds of substances, this phenomenon might be caused by the dietary differences among different continents.

#### 4. Discussion and conclusion

Bifidobacteria, which are common microorganisms in the human gut, are one of the most used probiotics [84]. The application and efficacy of probiotics in the prevention and treatment of intestinal diseases has received great attention. However, even though the same probiotic species are used, different disease types may contribute to the inconsistency of the results [85,86]. Therefore, a deeper understanding of the distribution and function of *Bifidobacterium* is needed to ensure that these probiotic bacteria can be used properly.

Here, we collected 1512 reference genomes of *Bifidobacterium* from three large-scale culturable gut microbiota databases and selected 849 non-redundant genomes for further analysis of bacterial functions, infecting phages, and horizontal gene transfer. CAZymes varied among the genomes of different species and continents. Most genomes of *B. adolescentis* harboring genes encoding  $\alpha$ -L-fucosidase (GH95) were from Europe, and most genomes of *B. longum* subsp. *longum* harboring GH95 were from North America. GH95, which participates in  $\alpha$ -L-fucoside degradation and HMO degradation [87], was also found in most genomes of *B. bifidum* and *B. breve*. Notably, all genomes of *B. longum* subsp. *infantis*, which is known as an HMO utilizer, also harbored genes encoding GH95. We identified BGCs in *Bifidobacterium* and found that many *B. pseudocatenulatum* strains had the potential to synthesize class IV lanthipeptide, which might be a candidate for antibacterial drugs.

Since *Bifidobacterium* has been widely considered a probiotic, we conducted an in-depth study on probiotic functions such as bile salt hydrolysis, GABA synthesis, and HMO degradation. Almost all genomes of *Bifidobacterium* in this study harbored genes encoding BSH, indicating that bile salt hydrolysis might be a conserved function in *Bifidobacterium*. Recently, *B. adolescentis* was confirmed to synthesize GABA in mouse [21]. However, whether it can be used for the prevention and treatment of neurological diseases in humans requires verification *in vivo* and *in vitro*. In addition to *B. adolescentis*, we found another species (*B. dentium*) also harbored *gadB* and *gadC*, which might provide other choices for probiotic application in the clinical treatment of neurological diseases. Both as a subspecies of *B. longum*, *B. longum* subsp. *infantis* harbors unique and more complete sets of genes for HMO utilization that are not all present in other genomes of *B. longum* subsp. *longum*, indicating that *B. longum* subsp. *infantis* is more suitable for application in the prevention and treatment of related diseases such as malnutrition [23].

When using *Bifidobacterium* strains as probiotics, it is necessary to decide which antibiotics could be used according to the species and region source of that strain. In addition, we found a strain of *B. adolescentis* harboring 37 ARGs and a strain of *B. longum* subsp. *longum* harboring the virulent factor hyaluronidase, which might cause more ARGs transfer and harmful effects if used as probiotics.

The phages of 38 VCs were found to infect *Bifidobacterium*, and their functions varied among different species and continents. This revealed that phages might play an important role in helping probiotic *Bifidobacterium* compete with other species and adapt to the environment. In our study, genes related to information storage and processing were transferred more frequently between *Bifidobacterium* and other genera. We also found that *Bifidobacterium* from North America could cope with local Hg and As contamination and

adapt to the environment through HGT. The average proportion of representative genomes of *Bifidobacterium* in human gut varied among different cohorts, which might be raised by the dietary differences between different countries.

However, there are limitations of our work. Our collection which included three large databases of human gut microbiome covered most of the representatives of *Bifidobacterium* genomes in the human gut, but some genomes of *Bifidobacterium* might not be included in our analysis. Even though metagenome-assembled genomes (MAGs) can hardly be used in the experiment, they may represent novel species which can be difficult to isolate. It is worthwhile that we collect more isolated *Bifidobacterium* genomes and MAGs on a large scale to explore the diversity of *Bifidobacterium* from different species and countries. In addition, the genetic diversity and different niches adaption of *Bifidobacterium* isolated from animals, environments, and fermented food remains to be explored. In conclusion, we envisage that our work will serve as a useful summary of the characteristics and functions of genomes of cultivated *Bifidobacterium* in the human gut and a reference for the clinical application of *Bifidobacterium* in the future.

# Data availability statement

The data that support the findings of this study can be accessed in the Zenodo database, by using the https://doi.org/10.5281/zenodo.7874214.

# **Ethics declarations**

This study was reviewed and approved by the Institutional Review Board on Bioethics and Biosafety of BGI, with the approval number: BGI-IRB 20106-T2. Informed consent was not required for this study because the microbial genomic data in this study were all public.

# CRediT authorship contribution statement

Wenxi Li: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Conceptualization. Hewei Liang: Methodology, Investigation, Formal analysis. Wenxin He: Visualization, Formal analysis. Xiaowei Gao: Writing – review & editing. Zhinan Wu: Methodology, Formal analysis. Tongyuan Hu: Methodology, Formal analysis. Xiaoqian Lin: Methodology, Formal analysis. Mengmeng Wang: Visualization. Yiyi Zhong: Resources. Haifeng Zhang: Resources. Lan Ge: Resources. Xin Jin: Supervision. Liang Xiao: Project administration. Yuanqiang Zou: Writing – review & editing, Funding acquisition, Conceptualization.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

# Acknowledgements

This study was supported by a grant from the National Natural Science Foundation of China (No. 32100009). This study was supported by Henan Supercomputer Center.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e27270.

#### References

- [1] C.F. Favier, et al., Molecular monitoring of succession of bacterial communities in human neonates, Appl. Environ. Microbiol. 68 (1) (2002) 219–226.
- [2] S. Arboleya, et al., Gut bifidobacteria populations in human health and aging, Front. Microbiol. 7 (2016) 1204.
- [3] S. Nishijima, et al., The gut microbiome of healthy Japanese and its microbial and functional uniqueness, DNA Res. 23 (2) (2016) 125–133.
- [4] V.K. Gupta, et al., A predictive index for health status using species-level gut microbiome profiling, Nat. Commun. 11 (1) (2020) 4635.
- [5] L. Drago, et al., Cultivable and pyrosequenced fecal microflora in centenarians and young subjects, J. Clin. Gastroenterol. 46 (Suppl) (2012) S81-S84.
- [6] X. Lin, et al., The genomic landscape of reference genomes of cultivated human gut bacteria, Nat. Commun. 14 (1) (2023) 1663.
- [7] A. Almeida, et al., A unified catalog of 204,938 reference genomes from the human gut microbiome, Nat. Biotechnol. 39 (1) (2021) 105–114.
- [8] Y. Zou, et al., 1,520 reference genomes from cultivated human gut bacteria enable functional microbiome analyses, Nat. Biotechnol. 37 (2) (2019) 179–185.
- [9] M. Poyet, et al., A library of human gut bacterial isolates paired with longitudinal multiomics data enables mechanistic microbiome research, Nat. Med. 25 (9) (2019) 1442–1452.
- [10] C. Milani, 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.
- [11] W. Mancino, et al., Mobilome and resistome reconstruction from genomes belonging to members of the Bifidobacterium genus, Microorganisms 7 (12) (2019).

[12] F. Bottacini, D. van Sinderen, M. Ventura, Omics of bifidobacteria: research and insights into their health-promoting activities, Biochem. J. 474 (24) (2017) 4137–4152.

[13] M. Derrien, et al., Insights into endogenous Bifidobacterium species in the human gut microbiota during adulthood, Trends Microbiol. 30 (10) (2022) 940–947.

- [14] S.M. Ryan, G.F. Fitzgerald, D. van Sinderen, Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains, Appl. Environ. Microbiol. 72 (8) (2006) 5289–5296.
- [15] J.M. Ridlon, et al., Consequences of bile salt biotransformations by intestinal bacteria, Gut Microb. 7 (1) (2016) 22-39.
- [16] M.L. Jones, et al., Cholesterol lowering with bile salt hydrolase-active probiotic bacteria, mechanism of action, clinical evidence, and future direction for heart health applications, Expet Opin. Biol. Ther. 13 (5) (2013) 631–642.
- [17] S.B. Sarasa, et al., A brief review on the non-protein amino acid, gamma-amino Butyric acid (GABA): its production and role in microbes, Curr. Microbiol. 77 (4) (2020) 534–544.
- [18] M.D. Muñoz, N. de la Fuente, A. Sánchez-Capelo, TGF-β/Smad3 signalling modulates GABA Neurotransmission: implications in Parkinson's disease, Int. J. Mol. Sci. 21 (2) (2020).
- [19] B.V. Manyam, et al., Isoniazid-induced elevation of CSF GABA levels and effects on chorea in Huntington's disease, Ann. Neurol. 10 (1) (1981) 35–37.
- [20] G.A. Czapski, J.B. Strosznajder, Glutamate and GABA in Microglia-Neuron Cross-Talk in Alzheimer's disease, Int. J. Mol. Sci. 22 (21) (2021).
- [21] S. Duranti, et al., Bifidobacterium adolescentis as a key member of the human gut microbiota in the production of GABA, Sci. Rep. 10 (1) (2020) 14112.
- [22] B.M. Henrick, et al., Bifidobacteria-mediated immune system imprinting early in life, Cell 184 (15) (2021) 3884–3898.e11.
   [23] M. J. Barratt, et al., Bifidobacteria-mediated immune system imprinting early in Bangladeshi infants with savere acute malnutrition. Sci. Transl. Med 1
- [23] M.J. Barratt, et al., Bifidobacterium infantis treatment promotes weight gain in Bangladeshi infants with severe acute malnutrition, Sci. Transl. Med. 14 (640) (2022) eabk1107.
- [24] T. Asahara, et al., Antibacterial effect of fermented milk containing Bifidobacterium breve, Bifidobacterium bifidum and Lactobacillus acidophilus against indigenous Escherichia coli infection in mice 13 (1) (2001) 16–24.
- [25] X. Ma, et al., Bifidobacterium infantis strain YLGB-1496 possesses excellent antioxidant and skin barrier-enhancing efficacy in vitro, Exp. Dermatol. 31 (7) (2022) 1089–1094.
- [26] K. Kajander, et al., Clinical trial: multispecies probiotic supplementation alleviates the symptoms of irritable bowel syndrome and stabilizes intestinal microbiota, Aliment. Pharmacol. Ther. 27 (1) (2008) 48–57.
- [27] M. Arumugam, et al., Enterotypes of the human gut microbiome, Nature 473 (7346) (2011) 174–180.
- [28] X. Guo, et al., CNSA: a Data Repository for Archiving Omics Data, Database, oxford, 2020.
- [29] F.Z. Chen, et al., CNGBdb: China national GeneBank DataBase, Yi Chuan 42 (8) (2020) 799-809.
- [30] D.H. Parks, et al., CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, Genome Res. 25 (7) (2015) 1043–1055.
- [31] D.H. Parks, et al., GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genomebased taxonomy, Nucleic Acids Res. 50 (D1) (2022) D785–d794.
- [32] C. Jain, et al., High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries, Nat. Commun. 9 (1) (2018) 5114.
- [33] T. Seemann, Prokka: rapid prokaryotic genome annotation, Bioinformatics 30 (14) (2014) 2068–2069.
- [34] C.P. Cantalapiedra, et al., eggNOG-mapper v2: functional annotation, orthology Assignments, and Domain prediction at the metagenomic scale, Mol. Biol. Evol. 38 (12) (2021) 5825–5829.
- [35] J. Huerta-Cepas, et al., eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, Nucleic Acids Res. 47 (D1) (2019) D309–d314.
- [36] J. Zheng, et al., dbCAN-seq update: CAZyme gene clusters and substrates in microbiomes, Nucleic Acids Res. 51 (D1) (2022) D557–D563.
- [37] B. Buchfink, K. Reuter, H.-G. Drost, Sensitive protein alignments at tree-of-life scale using DIAMOND, Nat. Methods 18 (4) (2021) 366-368.
- [38] K. Blin, et al., antiSMASH 6.0: improving cluster detection and comparison capabilities, Nucleic Acids Res. 49 (W1) (2021) W29-w35.
- [39] S.A. Kautsar, et al., MIBiG 2.0: a repository for biosynthetic gene clusters of known function, Nucleic Acids Res. 48 (D1) (2020) D454–d458.
- [40] C. Camacho, et al., BLAST+: Architecture and Applications, vol. 10, 2009, pp. 1–9.
  [41] G. Déjean, et al., Identifying a novel bile salt hydrolase from the Keystone gut Bacterium Christensenella minuta, Microorganisms 9 (6) (2021).
- [41] G. Depan, et al., iterating a hover one san hydroiase from the existing bacterian constension and the existing a hover one san hydroiase from the existing bacterian constension and the existing and the existin
- [43] B.P. Alcock, et al., Card 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database, Nucleic Acids Res. 48 (D1) (2020) D517–d525.
- [44] L. Chen, et al., Vfdb 2016: hierarchical and refined dataset for big data analysis-10 years on, Nucleic Acids Res. 44 (D1) (2016) D694–D697.
- [45] J. Guo, et al., VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses, Microbiome 9 (1) (2021) 37.
- [46] S. Nayfach, et al., CheckV assesses the quality and completeness of metagenome-assembled viral genomes, Nat. Biotechnol. 39 (5) (2021) 578-585.
- [47] D. Hyatt, et al., Prodigal: prokaryotic gene recognition and translation initiation site identification, BMC Bioinf. 11 (2010) 119.
- [48] H. Bin Jang, et al., Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks, Nat. Biotechnol. 37 (6) (2019) 632–639.
- [49] M. Groussin, et al., Elevated rates of horizontal gene transfer in the industrialized human microbiome, Cell 184 (8) (2021) 2053–2067.e18.
- [50] N.A. O'Leary, et al., Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation, Nucleic Acids Res. 44 (D1) (2016) D733–D745.
- [51] Z. Jie, et al., A transomic cohort as a reference point for promoting a healthy human gut microbiome, Medicine in Microecology 8 (2021).
- [52] R. Gacesa, et al., Environmental factors shaping the gut microbiome in a Dutch population, Nature 604 (7907) (2022) 732-739.
- [53] M.A. Freeberg, et al., The European genome-phenome archive in 2021, Nucleic Acids Res. 50 (D1) (2022) D980-d987.
- [54] I.H.i.R.N. Consortium, The integrative human microbiome Project, Nature 569 (7758) (2019) 641–648.
- [55] S. Chen, et al., fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (17) (2018) i884–i890.
- [56] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (2012) 357–359.
- [57] D.E. Wood, J. Lu, B. Langmead, Improved metagenomic analysis with Kraken 2, Genome Biol. 20 (1) (2019) 257.
- [58] J. Lu, et al., Bracken: Estimating Species Abundance in Metagenomics Data, vol. 3, 2017, p. e104.
- [59] S.M. Kelly, J. Munoz-Munoz, D. van Sinderen, Plant glycan metabolism by bifdobacteria, Front. Microbiol. 12 (2021) 609418.
- [60] D. Yu, et al., Bifidobacterium longum subsp. infantis as widespread bacteriocin gene clusters carrier stands out among the Bifidobacterium, Appl. Environ. Microbiol. 89 (9) (2023) e0097923.
- [61] S. Yoganathan, J.C. Vederas, Fracturing rings to understand lantibiotics, Chem. Biol. 15 (10) (2008) 999–1001.
- [62] J. Dischinger, S. Basi Chipalu, G. Bierbaum, Lantibiotics: promising candidates for future applications in health care, Int J Med Microbiol 304 (1) (2014) 51-62.
- [63] L.M. Repka, et al., Mechanistic understanding of lanthipeptide biosynthetic enzymes, Chem. Rev. 117 (8) (2017) 5457–5520.
- [64] J. Barbosa, T. Caetano, S. Mendo, Class I and class II lanthipeptides produced by Bacillus spp, J. Nat. Prod. 78 (11) (2015) 2850–2866.
- [65] J.H. Lee, X. Li, D.J. O'Sullivan, Transcription analysis of a lantibiotic gene cluster from Bifidobacterium longum DJO10A, Appl. Environ. Microbiol. 77 (17) (2011) 5879–5887.
- [66] Q. Zhang, et al., Expanded natural product diversity revealed by analysis of lanthipeptide-like gene clusters in actinobacteria, Appl. Environ. Microbiol. 81 (13) (2015) 4339–4350.
- [67] H. Ueno, Enzymatic and structural aspects on glutamate decarboxylase, J. Mol. Catal. B Enzym. 10 (1) (2000) 67–79.
- [68] D.A. Sela, et al., The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome, Proc. Natl. Acad. Sci. U. S. A. 105 (48) (2008) 18964–18969.
- [69] S. Bhardwaj, et al., Antibiotics and antibiotic resistance- flipsides of the same coin, Curr. Pharmaceut. Des. 28 (28) (2022) 2312–2329.
- [70] W.H. Organization, Critically Important Antimicrobials for Human Medicine, World Health Organization, 2018. 6th revision.
- [71] A.J. Browne, et al., Global antibiotic consumption and usage in humans, 2000-18: a spatial modelling study, Lancet Planet. Health 5 (12) (2021) e893-e904.
- [72] B. Liu, et al., Vfdb 2022: a general classification scheme for bacterial virulence factors, Nucleic Acids Res. 50 (D1) (2022) D912–d917.

- [73] J. Cuccui, et al., Characterization of the Burkholderia pseudomallei K96243 capsular polysaccharide I coding region, Infect. Immun. 80 (3) (2012) 1209–1221.
- [74] C. Rouquette, et al., Identification of a ClpC ATPase required for stress tolerance and in vivo survival of Listeria monocytogenes, Mol. Microbiol. 21 (5) (1996) 977–987.
- [75] G. Kayaoglu, D. Ørstavik, Virulence factors of Enterococcus faecalis: relationship to endodontic disease, Crit. Rev. Oral Biol. Med. 15 (5) (2004) 308–320.
- [76] B. Koskella, M.A. Brockhurst, Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities, FEMS Microbiol. Rev. 38 (5) (2014) 916–931.
- [77] M. Ventura, et al., Comparative analyses of prophage-like elements present in bifidobacterial genomes, Appl. Environ. Microbiol. 75 (21) (2009) 6929-6936.
- [78] A. Wahida, K. Ritter, H.P. Horz, The Janus-Face of Bacteriophages across human Body Habitats, PLoS Pathog. 12 (6) (2016) e1005634.
- [79] H. Ochman, J.G. Lawrence, E.A. Groisman, Lateral gene transfer and the nature of bacterial innovation, Nature 405 (6784) (2000) 299–304.
- [80] S.M. Soucy, J. Huang, J.P. Gogarten, Horizontal gene transfer: building the web of life, Nat. Rev. Genet. 16 (8) (2015) 472-482.
- [81] P. Staubitz, et al., MprF-mediated biosynthesis of lysylphosphatidylglycerol, an important determinant in staphylococcal defensin resistance, FEMS Microbiol. Lett. 231 (1) (2004) 67–71.
- [82] C.A. Eagles-Smith, et al., Mercury in western North America: a synthesis of environmental contamination, fluxes, bioaccumulation, and risk to fish and wildlife, Sci. Total Environ. 568 (2016) 1213–1226.
- [83] T.R. McClintock, et al., Arsenic exposure in Latin America: biomarkers, risk assessments and related health effects, Sci. Total Environ. 429 (2012) 76-91.
- [84] A. Rivière, et al., Bifidobacteria and Butyrate-producing Colon bacteria: importance and Strategies for their stimulation in the human gut, Front. Microbiol. 7 (2016) 979.
- [85] I. Bjarnason, G. Sission, B. Hayee, A randomised, double-blind, placebo-controlled trial of a multi-strain probiotic in patients with asymptomatic ulcerative colitis and Crohn's disease, Inflammopharmacology 27 (3) (2019) 465–473.
- [86] M.L. Ritchie, T.N. Romanuk, A meta-analysis of probiotic efficacy for gastrointestinal diseases, PLoS One 7 (4) (2012) e34938.
- [87] H. Ashida, et al., Two distinct alpha-L-fucosidases from Bifidobacterium bifidum are essential for the utilization of fucosylated milk oligosaccharides and glycoconjugates, Glycobiology 19 (9) (2009) 1010–1017.