



Article Colonized Niche, Evolution and Function Signatures of *Bifidobacterium pseudolongum* within Bifidobacterial Genus

Yue Xiao ^{1,2}, Jianxin Zhao ^{1,2}, Hao Zhang ^{1,2,3,4,5}, Qixiao Zhai ^{1,2,6,*} and Wei Chen ^{1,2,3}

- State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi 214122, China; 7170112038@stu.jiangnan.edu.cn (Y.X.); jxzhao@jiangnan.edu.cn (J.Z.); zhanghao@jiangnan.edu.cn (H.Z.); chenwei66@jiangnan.edu.cn (W.C.)
- ² School of Food Science and Technology, Jiangnan University, Wuxi 214122, China
- ³ National Engineering Research Center for Functional Food, Jiangnan University, Wuxi 214122, China
- ⁴ Institute of Food Biotechnology, Jiangnan University, Yangzhou 225004, China
- ⁵ Wuxi Translational Medicine Research Center, Jiangsu Translational Medicine Research Institute Wuxi Branch, Wuxi 214122, China
- ⁶ International Joint Research Laboratory for Probiotics, Jiangnan University, Wuxi 214122, China
- * Correspondence: zhaiqixiao@jiangnan.edu.cn; Tel./Fax: +86-510-85912155

Abstract: Background: Although genomic features of various bifidobacterial species have received much attention in the past decade, information on Bifidobacterium pseudolongum was limited. In this study, we retrieved 887 publicly available genomes of bifidobacterial species, and tried to elucidate phylogenetic and potential functional roles of *B. pseudolongum* within the *Bifidobacterium* genus. Results: The results indicated that B. pseudolongum formed a population structure with multiple monophyletic clades, and had established associations with different types of mammals. The abundance of *B. pseudolongum* was inversely correlated with that of the harmful gut bacterial taxa. We also found that *B. pseudolongum* showed a strictly host-adapted lifestyle with a relatively smaller genome size, and higher intra-species genetic diversity in comparison with the other tested bifidobacterial species. For functional aspects, B. pseudolongum showed paucity of specific metabolic functions, and enrichment of specific enzymes degrading complex plant carbohydrates and host glycans. In addition, B. pseudolongum possessed a unique signature of probiotic effector molecules compared with the other tested bifidobacterial species. The investigation on intra-species evolution of *B. pseudolongum* indicated a clear evolution trajectory in which considerable clade-specific genes, and variation on genomic diversity by clade were observed. Conclusions: These findings provide valuable information for explaining the host adaptability of B. pseudolongum, its evolutionary role, as well as its potential probiotic effects.

Keywords: *Bifidobacterium; Bifidobacterium pseudolongum;* population genomics; niche; evolution; probiotic effector molecules

1. Introduction

Bifidobacteria are important colonizers in the mammalian gut, and abundance of the genus is considered to be positively correlated with host's health outcomes [1–3]. Bifidobacterial composition in the gut has been studied by using isolation, denaturing gradient gel electrophoresis (DGGE) and internally transcribed spacer (ITS) rRNA profiling methods [4–7]. Recently, bifidobacterial composition among a wide range of 291 adult animals was analyzed, indicating their widespread distribution across the mammalian kingdom [7]. In this context, *B. pseudolongum*, together with *B. longum* and *B. adolescentis*, was shown to be the most predominant bifidobacterial species present in the gut of various mammals. Multiple bifidobacterial species were reported to coexist in the gut [7], showing species-level diversity. *B. pseudolongum* were enriched in the gut of some kinds of mammals (e.g., animals from Erinaceidae and Canidae), while were in low abundance for others



Citation: Xiao, Y.; Zhao, J.; Zhang, H.; Zhai, Q.; Chen, W. Colonized Niche, Evolution and Function Signatures of *Bifidobacterium pseudolongum* within Bifidobacterial Genus. *Foods* **2021**, *10*, 2284. https:// doi.org/10.3390/foods10102284

Academic Editors: Leyuan Li, Malgorzata Muc-Wierzgon and Sandra Martin-Pelaez

Received: 11 August 2021 Accepted: 13 September 2021 Published: 27 September 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). (e.g., animals from Cebidae and Cercopithecidae). Notably, Mao et al. previously showed that *B. pseudolongum* was the exclusive bifidobacterial species detected in the gut of Balb/c mice [8].

There are 72 currently recognized (sub)species within the *Bifidobacterium* genus. *B. pseudolongum*, as a species of the *Bifidobacterium* genus, consists of two subspecies, *B. pseudolongum* subsp. *pseudolongum* and *B. pseudolongum* subsp. *globosum* [9]. *B. pseudolongum* subsp. *pseudolongum* was first isolated and described by Mitsuoka in 1969 [10]. *B. globosum* was isolated from bovine rumen by Scardovi in the same year [11], and further classified into a subspecies of *B. pseudolongum*, that is, *pseudolongum* subsp. *globosum* in 1992 [9]. *B. pseudolongum* is a widely distributed gut commensal of the animal kingdom, and can be isolated from the guts of carnivores, herbivores, birds, and reptiles [12]. Nowadays, although comparative genomic analysis among various bifidobacterial species have been conducted [13–17], little attention has been paid to *B. pseudolongum*. Although the phylogenetic structure and metabolic ability of the species have been partially revealed by Lugli [12], the intra-species evolution, the subspecies-specific features, and the phylogenetic and potential functional roles of *B. pseudolongum* within the *Bifidobacterium* genus need further investigation.

Some bifidobacterial species, such as *B. longum*, *B. animalis*, *B. breve* and *B. bifidum*, have been extensively investigated and validated for their uses as probiotics [18–22]. However, very limited studies have focused on probiotic properties of *B. pseudolongum*. B. pseudolongum was evaluated in vitro for its potential probiotic functionality [23,24]. Five *B. pseudolongum* strains isolated from rats were evaluated for their tolerance to low pH and bile, as well as their capacity to adhere to intestinal epithelial cells and mucus [23]. The B. pseudolongum strain 119 has been tested for its growth rate, aerotolerance, antagonistic activity against pathogens, antimicrobial susceptibility profile, and cell wall hydrophobicity [24]. Several in vivo studies have also been conducted to support the potential beneficial effects of B. pseudolongum on the host. The B. pseudolongum strain Patronus has been reported to alleviate oxidative damage in metronidazole-treated rats [25], and in a previous study, oral administration of an isolated B. pseudolongum strain (no strain-specific full name of the strain was provided) was found to protect against hypersensitivity in mice [26]. Similarly, the distribution of genomic features that might contribute to probiotic effects, such as mucin-glycan degrading enzymes, pili, and S-layer protein, has been revealed in some probiotic bifidobacterial species [27,28], but remains overlooked in *B. pseudolongum*.

In this study, we analyzed the bifidobacterial composition and bacterial structure of the gut microbiota of representative mammals to explore niche features of *B. pseudolongum* and identify specific bacterial taxa whose abundances were correlated with that of *B. pseudolongum*. We analyzed 887 bifidobacterial genomes and their niche information, and constructed species-level phylogenetic trees for the selected bifidobacterial species. We also evaluated the phylogenetic role of *B. pseudolongum* within the *Bifidobacterium* genus by comparing the phylogenetic relationship, general genomic features, and genomic diversity among different bifidobacterial species. The profiles of carbohydrate-utilizing enzymes and signatures of clusters of orthologous groups of protein (COG) functions were also compared between *B. pseudolongum* and the other selected bifidobacterial species. In addition, we analyzed the distribution of genes encoding potential probiotic effectors. Finally, we explored the intra-species evolution and the clade-specific genomic features of *B. pseudolongum*.

2. Materials and Methods

2.1. Sample Collection

Fecal samples of C57/B6 male mice (8-week-old, n = 20), Balb/c male mice (8-week-old, n = 20), Wistar female rats (8-week-old, n = 22) and human subjects (n = 60) were collected to determine microbial composition. The mice were bred in the Laboratory Animal Center of the Department of Food Science and Technology, Jiangnan University, Wuxi, China, under specific pathogen-free (SPF) conditions for 3 weeks before sample collection. All of

the study protocols were approved by the Ethics Committee of Jiangnan University, China (JN. No20181130b1200130[261]). All of the applicable institutional and national guidelines (including the ARRIVE guidelines, the UK Animals (Scientific Procedures) Act, 1986 and associated guidelines, EU Directive 2010/63/EU for animal experiments, and the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978)) for the care and use of animals were followed. The human subjects were recruited from the provinces of Qinhai (n = 20), Heilongjiang (n = 20), and Jilin (n = 20), with a median age of the cohort of 48 and a ratio of female/male of 27/32 (information of one sample was missing). The human subjects were self-reported to be healthy and had not consumed antibiotics or probiotic-based products for one month before fecal sample collection.

2.2. Analysis of Fecal Microbial Community and Composition of Bifidobacterium Species

The microbial DNA of fecal samples were extracted by using the MP FastDNA Spin kit for Feces (MP Biomedicals) according to the manual. The V3-V4 region of 16S rRNA and the 60 kDa chaperonin (groEL) gene were amplified using barcoded fusion primers, and then sequenced with a Miseq TM sequencer separately for determination of microbial community and species-level *Bifidobacterium* composition, as previously described [29]. After sequencing, 16S rRNA sequencing data were analyzed using the QIIME pipeline [30]. For determination of bifidobacterial composition, a local nucleotide database was constructed using Bioedit, and the BLASTn algorithm was used to count the numbers of sequences that belong to each individual species. The PCA analysis based on abundances of key gut bacteria was conducted by the prcomp function and visualized using the ggbiplot package in R software. For correlation analysis, Spearman's correlation with FDR correlation was conducted by using package psych in R. The correlation coefficients of <-0.5 or >0.5, and p < 0.05 were considered as statistically significant.

2.3. Phylogenetic Reconstruction, Hierarchical Bayesian Clustering, Fixation Index (Fst) Caculation, Assembly Re-Annotation, and Pan-Genome Analysis

We retrieved 887 publicly available genomes/strains of the Bifidobacterial species from the NCBI database. The general genomic features (Guanine-Cytosine [GC] content, genome sizes, and numbers of coding sequences [CDs]) and biosample information of these 887 strains were also collected (Table S1). As previously mentioned [31], SNPs of the included genome assemblies were identified via sequence alignment to the corresponding reference genome for each *Bifidobacterium* species via MUMmer [32]. The phylogenetic tree of each *Bifidobacterium* species was constructed by using both the neighbor-joining (NJ) method (via Treebest software) and the maximum-likelihood (ML) method (via FastTree software [33]) based on concatenated sequences of bi-allelic SNPs in the core genome of each corresponding species. Tree-independent hierarchical Bayesian clustering with hierBAPS [34] was used to separately determine the population structure generated from the core genome mapping alignment for each species. For each species with more than two identified populations, SNP sites were used to convert multiple alignments of coregenome to VCF format [35]. Fst values between populations were analyzed by using the R package hierfstat [36]. The genome sequences were re-annotated using Prokka [37], and the annotated results were put into Roary [38] to conduct the pan-genome and gene presence/absence analyses (with a minimum identity of 90%). The core genes of each species were defined by those present in 99% to 100% of the strains of corresponding species. For the phylogenetic reconstruction of the *Bifidobacterium* genus, we selected 10 representative genomes along the tree of each species to cover an intra-species genetic distance for those species with more than 10 publicly available sequenced genomes, and adopted all the genomes for the species with less than 10 publicly available sequenced genomes. After the above selection, a total of 201 strains were adopted, and the homologous genes of the strains were identified by orthomcl [39]. The identified homologous genes were concatenated and then aligned by mafft [40]. The resulted alignment was used to build a NJ tree via the Treebest tool (http://treesoft.sourceforge.net/treebest.shtml, accessed on 25 December 2015). Each above-mentioned tree was visualized and phenotype mapped (e.g., isolated origins and subspecies) by Figtree (http://tree.bio.ed.ac.uk/software/, accessed on 23 January 2018) and iTOL (https://itol.embl.de/, accessed on 22 June 2019).

2.4. Profiles of Glucosyltransferase (GTs) and Glycoside HYDROLASES (GHs), and COG Functions

GH and GT genes were predicted among the genomes of 10 *Bifidobacterium* species by using HMMSCAN (from the HMMER package 3.1b2 (http://hmmer.org/, accessed on 30 March 2019)) to query a hidden Markov models-based CAZyme dbCAN database. For protein sequences beyond 80 aa, 50% coverage and an *E*-value $< 1 \times 10^{-5}$, and for protein sequences below 80 aa, 50% coverage and an *E*-value $< 1 \times 10^{-3}$ were chosen as cutoff thresholds. COG function categories were annotated with protein sequences as inputs by BLASTP against a COG database (ftp://ftp.ncbi.nih.gov/pub/COG/COG2014/data, accessed on 20 May 2019) with a threshold of 45% identity, 50% query coverage, and an *E*-value of 1×10^{-10} . Genes below the thresholds were not included in the analysis. The PCAs based on profiles of bifidobacterial GTs, and GHs, as well as COG functions were respectively conducted by using the prcomp function and visualized with the ggbiplot package in R software. The heatmap plot showing average gene numbers of COGs or carbohydrate enzymes per strain was drawn by pheatmap packages in R. For comparisons of profiles of carbohydrate-utilizing enzymes and COGs of *B. pseudolongum* by clade, PERMANOVA analysis based on bray distance was used with 999 permutations.

2.5. Profiles of Probiotic Effector Factors

For identification of mucin glycan-processing enzymes, the reference protein sequences were collected from the NCBI Refseq database by searching names of the enzymes as keywords (Chitinase (EC 3.2.1.14), neuraminidase/sialidase (EC 3.2.1.18), α galactosidase (EC 3.2.1.22), β -galactosidase (EC 3.2.1.23), α -N-acetylgalactosaminidase (EC 3.2.1.49), α -N-acetylglucosaminidase (EC 3.2.1.50), α -L-fucosidase (EC 3.2.1.51), β -*N*-hexosaminidase (EC 3.2.1.52), and endo- α -*N*-acetylgalactosaminidase (EC 3.2.1.97), as reported by Ravcheev et al. [41]). The separate protein databases for other probiotic effectors were built by retrieving sequences from the Refseq database and/or Uniprot database using the keywords "S-layer protein", "LPXTG", "sortase", "pilus, fimbria, and fimbrial protein", "luxS", "mucus-binding protein", and "serine-rich glycoprotein adhesin". The Antimicrobial Peptide Database (APD) was directly adopted (http://aps.unmc.edu/AP/, accessed on 15 April 2019). The amino acid sequences of each genome from Prokka were taken as a query to search against the above reference protein databases by using BLASTP with an E-value of 1 imes 10⁻³, sequence identity of 45%, and reference coverage of 50% as the cutoff. For identification of the priming-GTF gene (p-gtf) for EPS biosynthesis, the two genes "undecaprenyl-phosphate sugar phosphotransferase" (rfbP, accession number NP_695455) and "galactosyl-transferase" (cpsD, accession number NP_695447) in B. longum NCC2705 were taken as queries separately, as previously reported [42], to search against the proteome of each bifidobacterial strain (amino acid sequences of each bacterial genome from Prokka) with a threshold of query coverage of 50%, E-value of 1×10^{-5} , and sequence identity of 45%.

The BSH genes were identified by searching against the Prokka annotation results using keywords of "choloylglycine hydrolase" or "bile salt hydrolase", as previously reported [43]. The BSH nucleotide sequences with lengths of 900 bp to 1200 bp were included and aligned by mafft [40], and then a NJ tree was constructed by Treebest (http://treesoft.sourceforge.net/treebest.shtml) based on BSH alignments. The PCA analysis based on probiotic effectors was conducted by the prcomp function and visualized using the ggbiplot package in R software.

2.6. Statistical Analysis

For data of normal distribution, a One-Way ANOVA analysis was adopted (Tukey's b or Tamhane's T2). Otherwise, the nonparametric Mann-Whitney U test was used. A *p*-value of less than 0.05 was considered to be statistically significant.

3. Results

3.1. Distribution of B. pseudolongum in the Gut of the Selected Mammals, and Features of B. pseudolongum-Enriched Gut Microbiota

Previous studies have proposed better host adaptability of *B. pseudolongum* among various Bifidobacterium species by demonstrating its prevalence and dominance across the mammalian tree of life [7]. In order to get closer to its niche features, we sequenced and analyzed the composition of gut bifidobacteria of rodents (Balb/c mice, C57/B6 mice, and Wistar rats) and humans by including more individuals for the given types of animals compared with a previous study [7]. Our results suggested that *B. pseudolongum* was the most dominant species among all bifidobacterial species in Balb/c mice. The dominant colonization of this species was also observed for C57 mice. In contrast, the main bifidobacteria in rats were B. pseudocatenulatum and B. animalis, while humans were mainly colonized by B. longum, B. pseudocatenulatum, B. adolescentis, and B. bifidum with varying relative abundances of these species between individuals. We found that population levels of *Bifidobacterium* in rats, humans, and Balb/c mice were comparable, whereas C57 mice showed significantly lower abundances compared with Balb/c mice (p < 0.05, Figure 1B). In general, *B. pseudolongum* was the most dominant bifidobacterial species in the gut of C57 mice and Balb/c mice instead of rats and humans. However, B. pseudolongum showed a higher relative colonized biomass in the gut of Balb/c mice compared with C57 mice. Therefore, we concluded that *B. pseudolongum* prefers to expand in the gut of Balb/c mice compared with C57 mice, rats, and humans.

Next, we explored the bacterial taxa whose abundances were correlated positively or negatively with that of *B. pseudolongum* in Balb/c mice (the most potent niche that is enriched with *B. pseudolongum*). After the correlation test, four taxa stood out, including *unclassified__Clostridiaceae, Dorea, Desulfovibrio,* and *Pseudomonas* ($|\mathbf{R}| > 0.5$ and $p_adj < 0.05$ with a total number of 20,000 tests for FDR correction; Figure 1C). The four kinds of mammals included in this study formed distinct clusters based on the abundances of these four bacterial genera (Figure 1D). In addition, levels of the colonized biomass of these four genera markedly varied between Balb/c mice and each type of the other three kinds of mammals (Figure 1E).

3.2. Phylogenetic Structures and Niche Distribution

We analyzed the phylogenetic structures and niche distribution modes of *B. pseudolongum* and nine more *Bifidobacterium* species (each species with a number of publicly available genomes >10) (Figures 2, 3 and S1). We observed that *B. pseudolongum* and *B. animalis*, which showed large associations with a variety of animals, appeared to form multi-clade population structures. In particular, *B. pseudolongum* formed four distinct Bayesian Analysis of Population Structure (BAPS) clusters with an average differentiation value (Fst) of 0.70, and *B. animalis* formed three distinct BAPS clusters with an average Fst of 0.85. In contrast, some of the (sub)species that were frequently found to be associated with humans showed radiating structures (a possible indication of recombination, Figure 3). In particular, *B. longum* subsp. *longum* and *B. adolescentis* showed no genetically distant clusters via BAPS analysis, and two clusters of *B. breve* represented limited population differentiation (Fst = 0.17).

Although the numbers of isolates from particular niches were limited, such as *B. pseudolongum* strains isolated from primates and *B. animalis* strains isolated from Artiodactyla, preliminary host-specific lineages were shown for the two bifidobacterial species (Figure S2A,B). In general, *Bifidobacterium* species, a group of strictly anaerobic bacteria, adopted a hostadapted lifestyle (only occasionally found in free-living niches). *B. pseudolongum* with multi-clade population structures was more frequently isolated from various mammalian animals rather than being mainly found in humans according to the current dataset.



Figure 1. Distribution of *B. pseudolongum* in the gut of selected mammals, and features of *B. pseudolongum*-enriched gut microbiota. (**A**) Bifidobacterial composition in the gut of Balb/c mice (n = 15), C57/B6 mice (n = 6), humans (n = 60) and rats (n = 5). Due to good repeatability of sequencing data for C57 mice and rats, only 5 to 6 samples were sequenced. For Balb/c mice, more samples (totally 15) were successfully sequenced in order to conduct the following correlation analysis. For human fecal samples, all the 60 samples were sequenced. (**B**) Relative abundance of the genus *Bifidobacterium*. (**C**) Abundance correlation between relative abundances of *B. pseudolongum* and that of the other genera in the gut of Balb/c mice. Spearman's method was used for correlation analysis. $|\mathbf{R}| > 0.5$ and p < 0.05 (with a total number of 20,000 tests for FDR correction) were considered to be statistically significant. (**D**) PCA plot of the four types of included mammals (Balb/c mice, C57 mice, humans and rats) with the relative abundances of the four representative genera as inputs. (**E**) Relative abundances of the four representative genera. Statistical analysis was conducted between Balb/c group (n = 20) and each of the other three groups (C57 group (n = 20), human group (n = 60) and rat group (n = 22). *, p < 0.05; **, p < 0.01; ***, p < 0.001. All the fecal samples of included humans (60), mice (20 for Balb/c and 20 for C57) and rats (22) were sequenced for 16S rRNA gene data.



Figure 2. Niche distribution. Niche distribution of bifidobacterial species. The percentage of genomes belonging to the given niche was illustrated. Niches were categorized into human-associated habitats (red), free-living environments, such as food matrices or plants (green), animals (orange), and pure bacterial cultures or commercial supplements for which the isolation origins of microbial strains cannot be determined (gray).

3.3. General Genomic Features

To capture the intra-genus evolution of *Bifidobacterium*, we constructed a phylogenetic tree based on core orthologs by selecting representative strains from each species. Six main groups were categorized, including the *B. pseudolongum* group, *B. adolescentis* group, *B. pullorum* group, *B. boum* group, *B. bifidum* and *B. longum* complex, and *B. asteroides* group (Figure 4A). We observed that the *B. asteroides* phylogenetic group was shown to be positioned closest to the root of the tree, therefore indicating that members of this group most closely resemble the evolutionary ancestor of the *Bifidobacterium* genus. In addition, the bootstrap values of the clades were high (Figure S3), suggesting that the root and the phylogenetic structure revealed here were robust. We observed that *B. pseudolongum*

belonged to one of the three groups (*B. pseudolongum* group, *B. adolescentis* group, and *B. bifidum* and *B. longum* complex) that were shown to fit in the deepest branch of the resulting phylogenetic tree, and position farthest to the root, suggesting the members of the *B. pseudolongum* group were divergent from the evolutionary ancestor of the genus.



Figure 3. Phylogenetic structures. Phylogenetic structures of bifidobacterial species. The *Bifidobacterium* species with more than 10 publicly available genomes in the NCBI database were selected, and a maximum likelihood (ML) tree was constructed based on bi-SNPs in the core genome of each species. The corresponding NJ trees were also constructed (Figure S1B), and similar topologies as those ML trees are shown. Reference genomes of each species used for tree construction are shown in bold in Table S1. For *B. longum*, population structures of *B. longum* (upper) and *B. longum* subsp. *longum* (lower) are shown. The scale bars for trees were uniformly 0.1. The node symbols represent population clusters based on Bayesian Analysis of Population Structure (BAPS) hierarchical clustering. Within each species, the average Fst values between BAPS populations are also shown.

We further evaluated the general genomic features and intra-species genomic diversity of these *Bifidobacterium* species (Figure 4B–G). The genome sizes varied significantly among different bifidobacterial species. We found that *B. pseudolongum* harbored the second smallest genome size following *B. animalis* (Figure 4B). There was no apparent difference in ratio of accessory genes between the various species (Figure 4F). However, *B. pseudolongum* was among the species with the highest numbers of pair-wise single nucleotide polymorphisms (SNPs) (Figure 4G).



Figure 4. Features of *B. pseudolongum* in terms of phylogenetic role and general genomic features among various *Bifidobacterium* species. (**A**) Phylogenetic tree (NJ tree) of genus *Bifidobacterium* based on the protein sequences of nine orthologs of various species. For species with more than 10 publicly available sequenced genomes, 10 representative strains distributed across the tree, which covered the overall genetic distance of each species, were selected; for those with less than 10 publicly available sequenced genomes, all the strains were included. (**B**,**C**) Genome size (Mb) (**B**) and GC content (%) (**C**) of different species. The statistical analysis was conducted between *B. pseudolongum* and each of the other species. ***, *p* < 0.001. (**D**) Association between genome size (Mb) and the number of CDSs, Pearson *r* = 0.94, *p* < 0.0001. (**E**) Association between genome size (Mb) and the number of CDSs, Pearson *r* = 0.94, *p* < 0.0001. (**E**) Association between genome size (Mb) and SNP distance (**G**). Columns and dots representing *B. pseudolongum* are highlighted in red. All the available 887 sequenced genomes of *Bifidobacterium* in the NCBI database were used for analysis (**A**). Each of the *Bifidobacterium* species with more than 10 publicly available sequenced genomes in the NCBI database were analyzed, resulting in 786 strains in total (**B**–**G**).

3.4. COG Functions and Profiles of Carbohydrate-Utilizing Enzymes

For COG categories and signatures of the carbohydrate-utilizing enzymes examined, the animal-associated species, *B. pseudolongum* and *B. animalis*, clustered together and showed more similar profiles (Figures 5A,B and 6A,B). *B. pseudolongum* demonstrated significant differences on 19 of 23 COG functions in comparison to other *Bifidobacterium* species (Figure 5A). Most of these significantly different COGs were in the lower levels in

B. pseudolongum (15/19), covering three main metabolic processes—carbohydrate transport and metabolism, amino acid transport and metabolism, and lipid transport and metabolism. In addition, *B. pseudolongum* seemed to possess lower numbers of most glycosyltransferases (GTs) and glycoside hydrolases (GHs) than other species (Figure 6B).



Figure 5. Characterization of *B. pseudolongum* in terms of COG functions among various *Bifidobacterium* species. (**A**) Heatmap of average numbers of COG function categories per strain by species. Significantly different COG functions in the comparison of *B. pseudolongum* strains and the strains of the other studied bifidobacterial species were marked. ***, *p* < 0.001; **, *p* < 0.01. A: RNA processing and modification; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; M: Cell wall/membrane/envelope biogenesis; N: Cell motility; O: Posttranslational modification, protein turnover, chaperones; P:Inprganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; W: Extracellular structures; X: Mobilome: prophages, transposons. (**B**) PCA plot of COG functions by *Bifidobacterium* species with more than 10 publicly available sequenced genomes in the NCBI database were analyzed (including 786 strains).



Figure 6. Characterization of *B. pseudolongum* in terms of carbohydrate-utilizing enzymes among various *Bifidobacterium* species. (**A**) PCA plot of GH and GT profiles of different species. (**B**) Heatmap of average numbers of GHs and GTs per strain by *Bifidobacterium* species. (**C**) Selected enriched GHs and GTs in *B. pseudolongum* compared with the other studied bifidobacterial species. GH13_1 (encompassing α -amylases), GH13_28 (encompassing α -amylases), GH13_5 (encompassing α -amylases), GH30 (representing fucosidases), GH73 (including activities of β -*N*-acetylglucosaminidases), GH49 (including activities of dextranases), and GT32 (including activities of mannosyltransferases). The *Bifidobacterium* species with more than 10 publicly available sequenced genomes in the NCBI database were analyzed (including 786 strains).

However, four COGs were in higher levels in *B. pseudolongum* compared with the other tested bifidobacterial species, including replication, recombination and repair (L), defense mechanisms (V), intracellular trafficking, secretion, and vesicular transport (U), and transcription (K) (Figure 5A,C). Similarly, several GHs and GTs were also enriched in *B. pseudolongum*, such as GH13_1 (encompassing α -amylases), GH13_28 (encompassing α -amylases), GH13_5 (encompassing α -amylases), GH30 (representing fucosidases), GH73 (including activities of β -*N*-acetylglucosaminidases), GH49 (including activities of dextranases), and GT32 (including activities of mannosyltransferases) (Figure 6B,C). The enzymes in the GH13 family (encompassing α -amylases) are involved in the breakdown of complex plant carbohydrates [44]. GH13_28 (63/74) and GH13_5 (73/74) distributed in most *B. pseudolongum* strains, but were not frequently found in the genomes of the other tested bifidobacterial species. GH13_1 (1/786), GH73 (including activities of β -*N*-acetylglucosaminidases; 4/786), and GH49 (including activities of dextranases; 8/786) were rare GH families within the studied *Bifidobacterium* species, and the genes encoding these enzymes were more likely to be present in *B. pseudolongum* compared with the other

tested bifidobacterial species (Figure 6C). GH30, which represented fucosidases and was believed to degrade host glycans [16], was one of the core genes of *B. pseudolongum*, but showed large paucity in the other studied bifidobacterial species.

3.5. Probiotic Effector Molecules

To explore potential probiotic effects of *B. pseudolongum*, we mined the distribution of the current elucidated probiotic effector molecules [27,28] in *B. pseudolongum* in the background of hundreds of genomes of bifidobacterial strains.

First, we evaluated the signatures of nine key mucin-glycan foraging enzymes (Figure 7A) [41]. None of the studied *Bifidobacterium* species possessed α -*N*-acetylgalactosaminidase, and *B. bifidum* was the only member that encoded β -*N*-hexosaminidase and α -*N*-acetylglucosaminidase. Five enzymes for mucin-glycan degradation were found in the genomes of *B. pseudolongum*, and the distribution of these enzymes significantly fluctuated among different bifidobacterial species. Compared with other bifidobacteria, *B. pseudolongum* showed lower levels of neuraminidase (sialidase) and β -galactosidase, and higher levels of α -galactosidase, chitinase, and α -L-fucosidase. We found that *B. bifidum*, *B. longum*, and *B. asteroides* possessed the ability to utilize O-linked glycans (executed by Endo- α -*N*-acetylgalactosaminidase, with which bacteria could cleave glycans from mucin proteins). Although the other seven bifidobacterial species including *B. pseudolongum* did not encode this enzyme, they possessed mucin-glycan foraging potentials by cleaving side-chains of mucin-glycans (executed by one of the other seven enzymes other than Endo- α -*N*-acetylgalactosaminidase).

For other probiotic effector molecules, the mucus-binding protein and serine-rich glycoprotein adhesion could not be found in the genomes of all the tested bifidobacterial strains. A previous study has reported that complete mucus-binding domains, called MUB, were found exclusively in lactic acid bacteria [45], which supported the absence of mucus-binding proteins within the *Bifidobacterium* genus. The average gene numbers of bile salt hydrolase (BSH), antimicrobial peptide, and luxS in B. pseudolongum were comparable to the other tested bifidobacterial species (p > 0.05) (Figure 7B). B. asteroides, mainly found in insect guts, did not harbor BSH genes. The levels of pili-related molecules, including sortase, pilus, and LPXTG, significantly varied between *B. pseudolongum* and the other tested bifidobacterial species (p < 0.001). Notably, S-layer proteins were markedly enriched in *B. pseudolongum* strains. Next, we evaluated the presence and absence of exopolysaccharide (EPS) clusters in bifidobacterial genomes by searching for the priming-GTF gene (p-gtf). This gene encoded the enzyme that was in charge of the initial step of the EPS-unit biosynthesis; therefore, it should be present in all EPS clusters. We found that major bifidobacterial species showed the presence of two kinds of priming-GTF genes, while B. pseudolongum only harbored one priming-GTF gene (cpsD). The principal component analysis (PCA) results based on the gene numbers of all these probiotic factors also supported the fact that higher levels of EPS_rfbP and S-layer protein contributed to the separation of *B. pseudolongum* from the other tested bifidobacterial species in the PCA plot (Figure 7C).

Notably, we found that different bifidobacterial species possessed distinct types of BSH genes according to a phylogenetic tree constructed by BSH sequences (Figure 7D). After adding the reference sequences of seven types of BSH genes [43] (see Table S2 for details), we observed that the BSH genes of all the studied *Bifidobacterium* species (except that three BSH sequences from *B. longum* could not be categorized into any currently elucidated BSH types) belonged to the BSH-T4 type, and all the studied species except *B. longum* did not harbor paralogs of BSH. In addition, *B. pseudolongum* clustered with *B. animalis*, indicating they had more similar BSH sequences (Figure 7D).

A

1

0.8

0.6

0.4

0.2 0

1

0.8 0.6 0.4 0.2 0 **Beta-N-hexosaminidase**

α-N-acetylglucosaminidase

a-L-fucosidase

pseudolongum

1.2 1 0.8

0.6 0.4 0.2

0

2.4 2 1.6 1.2 0.8 0.4 0



longum cbh 1



B

1.5

1

0.5

0

1

0.8

0.6 0.4 0.2

0

Endo-a-N-acetylgalactosaminidase

Neuraminidase(sialidase)

a-galactosidase

P <0.001

Figure 7. Distribution of probiotic effector molecules among *B. pseudolongum* and other *Bifidobacterium* species. (A) Average gene numbers of mucin-glycan foraging enzymes per strain by Bifidobacterium species. Each column represents a bifidobacterial species. From left to right: *B. adolescentis, B. animalis, B. asteroides, B. bifidum, B. breve, B. dentium, B. longum, B. moukalabense, B. pseudocatenulatum, and B. pseudolongum.* (B) Average gene numbers of probiotic effectors per strain for each

species. Each column represents a bifidobacterial species. From left to right: *B. adolescentis*, *B. animalis*, *B. asteroides*, *B. bifidum*, *B. breve*, *B. dentium*, *B. longum*, *B. moukalabense*, *B. pseudocatenulatum*, and *B. pseudolongum*. (C) PCA plot of the strains from different *Bifidobacterium* species by taking the numbers of each included probiotic effector as inputs. (D) Phylotypes of BSH sequences by *Bifidobacterium* species. The BSH reference sequences were removed after evaluating the BSH types. The information for all the used BSH sequences are listed in Table S2. The *Bifidobacterium* species with more than 10 publicly available sequenced genomes in the NCBI database were analyzed (including 786 strains). Statistical analyses were performed between *B. pseudolongum* strains and the strains of the other studied bifidobacterial species.

3.6. Sub-Clade-Specific Evolution within B. pseudolongum

We analyzed subspecies-specific genomic features of *B. pseudolongum* to observe its evolutionary trajectory at a higher resolution (Figure 8A–G). B. pseudolongum represented a promiscuously host-adapted lifestyle, and could be found in diverse habitats including Mammalia, Aves, and Reptilia (Figure 8A). Four clades were shown after phylogenetic reconstruction based on core-genome bi-SNPs. Sub-clade A, corresponding with B. pseudo*longum* subsp. *pseudolongum*, showed the greatest numbers of the clade-specific genes, as represented by 43 unique genes that were present in all the strains of clade A and absent from all strains in the other three clades (Table S3 and Figure 8G). These clade-A-specific core genes covered a broad range of bacterial functions, including cell wall biosynthesis (e.g., β-hexosaminidase and LL-diaminopimelate aminotransferase), regulation (e.g., HTHtype transcriptional regulator DegA and transcriptional regulator LytR), signaling (e.g., pheromone autoinducer 2 transporter), stress response (e.g., sensor histidine kinase DesK), nucleotide metabolism (tRNA pseudouridine synthase B and DNA utilization protein GntX) and lipid metabolism (holo-[acyl-carrier-protein] synthase and dephospho-CoA kinase). The COG functions (PERMANOVA $R^2 = 0.31$ and p < 0.001; Figure 8D) and profiles of carbohydrate-utilizing enzymes (GTs and GHs: PERMANOVA $R^2 = 0.30$ and p < 0.001; Figure 8E) were preliminarily separated by clade. For pair-wise comparisons, COGs and carbohydrate-utilizing enzymes of bifidobacterial strains were significantly varied by clade (p < 0.05), except that clade B and clade D could not be distinguished with each other in terms of the profiles of GTs and GHs (p = 0.724).



Figure 8. Clade-specific genomic features and intra-species evolution of *B. pseudolongum*. (A) Phylogenetic tree of *B. pseudolongum* constructed by the NJ method based on the bi-SNPs in the core genome of 74 strains. The tree was rooted

by the clade that was nearest to the outgroup (*B. animalis* subsp. *animalis* ATCC 25527). Niche of each strain was marked on the corresponding clade and the inner strip by the taxonomic unit of "order". Each sub-clade was indicated by color in the outer strip. (**B**,**C**) Comparison of intra-clade genomic diversity by SNP distance (**B**) and accessory genome size (**C**). (**D**) PCA plot of COG functions of *B. pseudolongum* by sub-clade. (**E**) PCA plot of GHs and GTs of *B. pseudolongum* by sub-clade. (**F**) The gene presence and absence conditions of 74 strains of *B. pseudolongum*. (**G**) Numbers of clade-specific core genes that were present in all the strains of a designated clade and absent from any strains of the other three clades. See detailed information in Table S3.

For genomic diversity, although limited strains were included in the clade B, we observed the largest genetic distance within this cluster in terms of the pair-wise SNP numbers (Figure 8B). Clade D, corresponding to the traditional *B. pseudolongum* subsp. *globosum*, showed the largest ratio of accessory gene number/total gene number (Figure 8C). Genomic diversity in terms of the accessory gene sets seemed to be expanded across the phylogenetic tree (Figure 8C). Gene presence and absence analysis indicated the frequent appearance of the strain-specific genes for *B. pseudolongum*, which contributed to accruing accessory gene sets of the species (Figure 8F).

4. Discussion

The studies on evolutionary and functional genomics of various *Bifidobacterium* species have been performed for decades, yet these pieces of research largely focused on the limited number of type strains of various bifidobacterial species and mainly showed an emphasis on functional aspects instead of evolutionary roles [14–17,46]. In particular, *B. pseudolongum* seemed an overlooked species compared with those bifidobacterial species (e.g., *B. longum*) with well-documented health-promoting effects and colonization advantage [22,47]. Using population genomics analysis, we included 887 bifidobacterial genomes to investigate the evolutionary role and functional signature of *B. pseudolongum* within the bifidobacterial genus, and try to reveal the intra-species evolution and function separation. Our findings reveal important implications regarding the molecular mechanism of *Bifidobacterium* diversification.

Our results shed important light on the potential probiotic roles of B. pseudolongum on the host. We demonstrated that *B. pseudolongum* is more likely to colonize the gut of mice compared with rats and humans, and four bacterial taxa, including unclassified__Clostridiaceae, Dorea, Desulfovibrio, and Pseudomonas were observed to be correlated positively or negatively in abundance with *B. pseudolongum*. Notably, *Desulfovibrio* spp. are the most abundant sulfate-reducing bacteria in the mammalian gut [48]. Desulfovibrio spp. utilize hydrogen and other organic substrates as electron donors to reduce sulfate to hydrogen sulfide (H₂S). The increased abundance of *Desulfovibrio* has been linked to development of inflammatory bowel disease [49] and autism spectrum disorder [50], likely caused by the accumulation of H₂S. *Pseudomonas* spp. showed increased levels in the gut of patients with multiple sclerosis [51] and end-stage renal disease [52], represented a higher positive rate in the ileum of Crohn's disease patients compared with the healthy control group [53], and reflected a positive correlation with pro-inflammatory cytokine IL-6 [54]. In addition, one member of the genus, *P. aeruginosa*, is a Gram-negative opportunistic bacterium that causes various infections [55]. Therefore, the negative correlation between levels of *B. pseudolongum* and levels of either *Psuedomonas* spp. (R = -0.618) or *Desulfovibrio* spp. (R = -0.582) raised the question of whether administration by *Bifidobacterium* could inhibit pathogenic Desulfovibrio and Pseudomonas. Indeed, bifidobacterial species, like *B. longum*, had been reported to be able to successfully treat ulcerative colitis with multidrug-resistant pseudomonas aeruginosa infection in a man [56] and prevent gut-derived pseudomonas aeruginosa sepsis in mice [57].

The probiotic effects of some bacterial molecules, such as mucin glycan foraging enzymes, pili, EPS, S-layer protein, bacteriocins, BSH, mucus-binding proteins, and serinerich proteins have been revealed in various *Bifidobacterium* strains [27,28]. These bacterial molecules are known as "probiotic effector molecules" [27]. It was believed that these molecules exerted beneficial effects on host via at least five modes of action: Regulation of the composition and activity of the indigenous microbiota, improvement of epithelial barrier function, regulation of the immune system, regulation of systemic metabolic responses, and functioning via the central nervous system [27]. Here, the analysis on distribution of probiotic effector molecules in the genomes of various bifidobacterial species indicated that B. pseudolongum harbored mucin-glycan foraging enzymes, showed comparable numbers of BSH, antimicrobial peptide, and luxS with the other tested bifidobacterial species, but presented enriched genes encoding S-layer proteins, possible unique EPS organization, and the original BSH subtype. In particular, we observed that the major analyzed bifidobacterial species showed the presence of two kinds of priming-GTF genes, while B. pseudolongum only harbored one priming-GTF gene (cpsD). A previous study based on the analysis of 28 complete genomes of bifidobacterial strains reported that functional-structural organization of bifido-EPS was of large dissimilarity, and structures and lengths of the clusters with one priming-GTF gene were very different from those with two priming-GTF genes [42]. These results could provide a valuable knowledge basis for exploring similarity and uniqueness with respect to probiotic effects on the host between *B. pseudolongum* and each of the other tested bifidobacterial species that are with well-studied health-promoting phenotypes and related mechanistic insights.

This is the first-known example to reveal species-level population structures of various bifidobacterial species and give insight into their genomic diversity based on genome information. Our results indicated that those animal-associated species, including B. pseu*dolongum* (Fst = 0.70) and *B. animalis* (Fst = 0.85), appeared to form multi-clade population structures, whereas some of human-associated bifidobacterial (sub)species showed radiating structures with limited population differentiation. Correspondingly, a multi-clade phylogenetic shape has been reported for another champion colonizer in the gut of various animals, L. reuteri [55]. For L. reuteri, it was considered that distinct monophyletic clades represented host origins but not geographical locations, which was regarded to be consequences of long-term colonization by L. reuteri lineages in the gut of specific vertebrate species and host-driven diversification [58,59]. Similarly, we also observed host-specific lineages for *B. pseudolongum* and *B. animalis*, indicating host-driven genomic diversification. For the intra-genus evolutionary role of *B. pseudolongum*, consistent with previous studies [15,17], six main groups were categorized, including the *B. pseudolongum* group, B. adolescentis group, B. pullorum group, B. boum group, B. bifidum and B. longum complex, and *B. asteroides* group. We observed that the *B. asteroides* phylogenetic group was shown at the position closest to the root of the tree, while *B. pseudolongum* belonged to one of the three groups (the B. pseudolongum group, B. adolescentis group, and B. bifidum and B. longum complex) that were shown to fit in the deepest branch of the resulting phylogenetic tree, suggesting the members of the *B. pseudolongum* group were divergent from the evolutionary ancestor of the genus. A previous study has reported that in a family-based supertree of the Bifidobacteriaceae, bifidobacteria were demonstrated to locate in the deepest branch of the phylogenetic tree, and separated them from other genera within this family. Meanwhile, in the same study, the *B. asteroides* phylogenetic group was shown to position closest to the root in this family-based supertree, therefore indicating that members of this group most closely resemble the evolutionary ancestor of the Bifidobacterium genus. The observed substantial genomic size difference among various bifidobacterial species was previously considered as a reminiscent evidence of an evolutionary route that has involved a lot of gene loss and/or gain events [16]. It was previously reported that host-adapted species tend to harbor a reduced genome size during the lifestyle transition from free-living to host-adapted when the deletion of redundant genetic content occurs [60]. Here, we found that B. pseudolongum harbored the second smallest genome size, which could possibly explain its better host adaptability. In addition, we found that *B. pseudolongum* showed the highest number of pair-wise SNPs, reflecting its higher intra-species genetic diversity [61].

The present work implies considerable functional divergence of *B. pseudolongum* from the other included bifidobacterial species in terms of COG terms and carbohydrateutilizing enzymes. We found that *B. pseudolongum* demonstrated a significant difference on 19 of 23 COG functions in comparison to other *Bifidobacterium* species, and most of these significantly different COGs were in lower levels in *B. pseudolongum* (15/19). The analysis on carbohydrate-utilizing enzymes indicated that *B. pseudolongum* seemed to possess lower numbers of the most of GTs and GHs than other species. It has been reported that bacteria tend to shut down carbohydrate transportation and reduce energy metabolism in response to environmental stresses [62–64], and even discard some redundant genes encoding carbohydrate-utilizing enzymes in order to reduce energy consumption and improve niche fitness [65,66]. In contrast, several GHs and GTs involved in the breakdown of complex plant carbohydrates and host glycans were enriched in *B. pseudolongum*. Very limited studies have focused on carbohydrate metabolic ability of *B. pseudolongum*. Studies have reported that *B. pseudolongum* could use starch as a sole source of carbon and energy, possibly due to its α -glucosidase activity [12,67]. Additionally, *B. pseudolongum* encoded pectin-degrading enzymes and could degrade pectin [68].

The study on intra-species evolution of *B. pseudolongum* here highlights taxonomic inconsistencies and reveals novel clade-associated features. We demonstrated that four clades were shown after phylogenetic reconstruction based on core-genome bi-SNPs, despite the current recognition of only two subspecies—B. pseudolongum subsp. pseudolongum and B. pseudolongum subsp. globosum [12]. In addition, considerable clade-specific genes, functional separation by clade, and expansion of genomic diversity along the phylogenetic tree were observed. In a previous study, 47 bifidobacterial (sub)species have been clustered into three distinct groups based on the profiles of GHs (GHP/A, GHP/B, and GHP/C) [69]. In this context, two subspecies of B. thermacidophilum were clustered into GHP/B, and two subspecies of B. animalis were clustered into GHP/A. However, B. pseudolongum subsp. pseudolongum and B. pseudolongum subsp. globosum were classified into two separate groups—GHP/A and GHP/B, respectively. GHP/A cluster showed an extensive array of putative GH43 family members (involved in the degradation of complex plant glycans), suggesting adaptation of bifidobacterial (sub)species within the group GHP/A to hosts that adopted a vegetarian or omnivorous lifestyle [16,69]. Members of the GHP/B cluster were featured by the paucity of GH43 and GH3 enzymes. Therefore, the separation of GH and GT repertoires according to the clades of *B. pseudolongum* observed here and in the above-mentioned previous study suggested possible intra-species metabolic diversity.

5. Limitations

Our study has several limitations. For the analysis of colonized niches for *B. pseu*dolongum, we only selected fecal samples from rodents (Balb/c mice, C57/B6 mice, and Wistar rats) and humans that were easily available and collected. Although the results demonstrated distribution differences of *B. pseudolongum* and the suggested preference of *B. pseudolongum* in the gut of Balb/c mice, additional insights could be reached when fecal samples from more types of animals (such as ungulates, carnivores, chicken, and pigeons) are included. In addition, for the analysis of sub-clade-specific evolution within B. pseudolongum, the sub-clade-specific genes were analyzed by Roary software. This software is based on protein sequences, and thus might introduce a minute quantity of false-positive and false-negative hits. If further studies are focused on one or several of these specific genes, it is necessary to validate the present/absent status of them via BLASTN and/or PCR experiments. Furthermore, most parts of this study were based on genomic analysis. Some interesting findings here, such as possible different activity of BSH sub-types, and the structural and functional organizations of the EPS clusters of B. pseudolongum and other bifidobacterial species, need further investigations both in vitro and in vivo. Finally, we only gave details of and analyzed the bifidobacterial species with more than 10 publicly available genomes. With increased access to more genomes of other bifidobacterial species that were not analyzed here, more information could be obtained, and we could further contribute to the understanding of genomic diversity and evolution of the genus Bifidobacterium.

6. Conclusions

Taken together, *B. pseudolongum* showed a multi-clade population structure, and established in various mammalian animals a rather strict association with humans. Its abundance was negatively correlated with two harmful gut bacterial taxa (*Psuedomonas* sp. and *Desulfovibrio* sp.). Within the *Bifidobacterium* genus, *B. pseudolongum* represented a relatively smaller genome size and higher intra-species genetic diversity. The paucity of metabolism-related functions and enrichment of specific enzymes degrading complex plant carbohydrates and host glycans might provide a possible explanation for its prevalence and dominance across the mammalian branch of the tree of life. *B. pseudolongum* had a higher level of S-layer proteins, unique BSH subtype, and possible particular EPS cluster organization. Finally, considerable sub-clade specific genes, separation of profiles of carbohydrate-utilizing enzymes and COGs by clade, and variations on genomic diversity across the phylogenetic tree were also observed.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10 .3390/foods10102284/s1. Figure S1. Numbers of publicly available genomes for each bifidobacterial species (A), and maximum likelihood (ML) of them (B). Numbers of publicly available sequenced genomes for each *Bifidobacterium* species. Red, more than 50 genome assemblies; light orange, more than 10; light pink, more than 5; green, more than 1. Species in the right section behind the perpendicular line were without sequenced genomes in the NCBI database. Figure S2. Niche distribution along the phylogenetic tree (NJ tree) for *B. pseudolongum* (A) and *B. animalis* (B). Figure S3. Phylogenetic tree (NJ tree) of genus *Bifidobacterium* with the bootstrap values of the clades indicated. Bootstrap of 100 was set for tree construction. Table S1. The information on genome assemblies used in the study. Table S2. The information of sequences used in phylogenetic reconstruction of BSH genes. Table S3. The clade-specific core genes of *B. pseudolongum*.

Author Contributions: Y.X.: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing. J.Z.: Conceptualization, Project administration, Resources, Supervision, Writing—original draft, Writing—review & editing. H.Z.: Conceptualization, Project administration, Resources, Supervision, Writing—review & editing. W.C.: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing—review & editing. Q.Z.: Conceptualization, Funding acquisition, Project administration, Project administration, Project administration, Supervision, Writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China Program [No.31820103010 and No. 31871773]; Projects of Innovation and Development Pillar Program for Key Industries in Southern Xinjiang of Xinjiang Production and Construction Corps [2018DB002]; National Key Research and Development Project [No. 2018YFC1604206]; National First-Class Discipline Program of Food Science and Technology [JUFSTR20180102]; the BBSRC Newton Fund Joint Centre Award; and Collaborative Innovation Center of Food Safety and Quality Control in Jiangsu Province.

Institutional Review Board Statement: All of the study protocols for animal experiments were approved by the Ethics Committee of Jiangnan University, China (JN. No20181130b1200130[261]). All of the applicable institutional and national guidelines (including the ARRIVE guidelines, the U.K. Animals (Scientific Procedures) Act, 1986 and associated guidelines, EU Directive 2010/63/EU for animal experiments, and the National Institutes of Health guide for the care and use of Laboratory animals (NIH Publications No. 8023, revised 1978)) for the care and use of animals were followed.

Data Availability Statement: The metagenomic sequencing data have been deposited into sequence read archive (SRA) database under accession number PRJNA576558.

Conflicts of Interest: All authors declared no conflict of interest.

Abbreviations

DGGE: Denaturing gradient gel electrophoresis; ITS: Internally transcribed spacer; COG: Clusters of orthologous groups of protein; GHs: Glycoside hydrolases; GTs: Glycosyltransferases; BSH: Bile salt hydrolase; EPS: Exopolysaccharide; GC: Guanine-Cytosine; CD: Coding sequence; SNP: Single nucleotide polymorphisms; NJ: Neighbor-joining; ML: Maximum-likelihood; PCA: Principal component analysis.

References

- Woodmansey, E.J.; McMurdo, M.E.; Macfarlane, G.T.; Macfarlane, S. Comparison of compositions and metabolic activities of fecal microbiotas in young adults and in antibiotic-treated and non-antibiotic-treated elderly subjects. *Appl. Environ. Microbiol.* 2004, 70, 6113–6122. [CrossRef]
- 2. Hopkins, M.; Sharp, R.; Macfarlane, G. Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. *Gut* **2001**, *48*, 198–205. [CrossRef]
- Tojo, R.; Suárez, A.; Clemente, M.G.; de los Reyes-Gavilán, C.G.; Margolles, A.; Gueimonde, M.; Ruas-Madiedo, P. Intestinal microbiota in health and disease: Role of bifidobacteria in gut homeostasis. World J. Gastroenterol. 2014, 20, 15163–15176. [CrossRef]
- 4. Satokari, R.M.; Vaughan, E.E.; Akkermans, A.D.; Saarela, M.; de Vos, W.M. Bifidobacterial diversity in human feces detected by genus-specific PCR and denaturing gradient gel electrophoresis. *Appl. Environ. Microbiol.* **2001**, *67*, 504–513. [CrossRef]
- Milani, C.; Lugli, G.A.; Turroni, F.; Mancabelli, L.; Duranti, S.; Viappiani, A.; Mangifesta, M.; Segata, N.; van Sinderen, D.; Ventura, M. Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. *FEMS Microbiol. Ecol.* 2014, 90, 493–503. [CrossRef]
- Turroni, F.; Foroni, E.; Pizzetti, P.; Giubellini, V.; Ribbera, A.; Merusi, P.; Cagnasso, P.; Bizzarri, B.; de'Angelis, G.L.; Shanahan, F. Exploring the diversity of the bifidobacterial population in the human intestinal tract. *Appl. Environ. Microbiol.* 2009, 75, 1534–1545. [CrossRef]
- 7. Milani, C.; Mangifesta, M.; Mancabelli, L.; Lugli, G.; James, K.; Duranti, S.; Turroni, F.; Ferrario, C.; Ossiprandi, M.; Ventura, M. Unveiling bifidobacterial biogeography across the mammalian branch of the tree of life. *ISME J.* **2017**, *11*, 2834–2847. [CrossRef]
- 8. Mao, B.; Gu, J.; Li, D.; Cui, S.; Zhao, J.; Zhang, H.; Chen, W. Effects of different doses of fructooligosaccharides (FOS) on the composition of mice fecal microbiota, especially the bifidobacterium composition. *Nutrients* **2018**, *10*, 1105. [CrossRef]
- Yaeshima, T.; Fujisawa, T.; Mitsuoka, T. Bifidobacterium globosum, subjective synonym of Bifidobacterium pseudolongum, and description of Bifidobacterium pseudolongum subsp pseudolongum comb. nov. and Bifidobacterium pseudolongum subsp. globosum comb. nov. Syst. Appl. Microbiol. 1992, 15, 380–385.
- Mitsuoka, T. Comparative studies on bifidobacteria isolated from the alimentary tract of man and animals (including descriptions of *Bifidobacterium thermophilum* nov. spec. and *Bifidobacterium pseudolongum* nov. spec). Zent. Bakteriol. Parasitenkd. Infekt. Hygiene. 1. Abt. Med.-Hyg. Bakteriol. Virusforsch. Parasitologie. Orig. 1969, 210, 52–64.
- 11. Scardovi, V.; Trovatelli, L.; Crociani, F.; Sgorbati, B. Bifid bacteria in bovine rumen. Arch. Mikrobiol. 1969, 68, 278–294. [CrossRef]
- Lugli, G.A.; Duranti, S.; Albert, K.; Mancabelli, L.; Napoli, S.; Viappiani, A.; Anzalone, R.; Longhi, G.; Milani, C.; Turroni, F. Unveiling Genomic diversity among members of the species *Bifidobacterium pseudolongum*, a widely distributed gut commensal of the animal kingdom. *Appl. Environ. Microbiol.* 2019, *85*, e03065-18. [CrossRef]
- 13. Xiao, Y.; Zhao, J.; Zhang, H.; Zhai, Q.; Chen, W. Mining genome traits that determine the different gut colonization potential of *Lactobacillus* and *Bifidobacterium* species. *Microb. Genomics* **2021**, *7*, 000581. [CrossRef]
- 14. Bottacini, F.; Medini, D.; Pavesi, A.; Turroni, F.; Foroni, E.; Riley, D.; Giubellini, V.; Tettelin, H.; van Sinderen, D.; Ventura, M. Comparative genomics of the genus *Bifidobacterium*. *Microbiology* **2010**, *156*, 3243–3254. [CrossRef]
- Lugli, G.A.; Milani, C.; Turroni, F.; Duranti, S.; Ferrario, C.; Viappiani, A.; Mancabelli, L.; Mangifesta, M.; Taminiau, B.; Delcenserie, V. Investigation of the evolutionary development of the genus *Bifidobacterium* by comparative genomics. *Appl. Environ. Microbiol.* 2014, 80, 6383–6394. [CrossRef]
- Milani, C.; Turroni, F.; Duranti, S.; Lugli, G.A.; Mancabelli, L.; Ferrario, C.; van Sinderen, D.; Ventura, M. Genomics of the genus *Bifidobacterium* reveals species-specific adaptation to the glycan-rich gut environment. *Appl. Environ. Microbiol.* 2016, *82*, 980–991. [CrossRef]
- Sun, Z.; Zhang, W.; Guo, C.; Yang, X.; Liu, W.; Wu, Y.; Song, Y.; Kwok, L.Y.; Cui, Y.; Menghe, B. Comparative genomic analysis of 45 type strains of the genus *Bifidobacterium*: A snapshot of its genetic diversity and evolution. *PLoS ONE* 2015, 10, e0117912. [CrossRef]
- 18. Quigley, E. Bifidobacterium breve. In *The Microbiota in Gastrointestinal Pathophysiology*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 135–137.
- 19. Quigley, E. Bifidobacterium animalis spp. lactis. In *The Microbiota in Gastrointestinal Pathophysiology;* Elsevier: Amsterdam, The Netherlands, 2017; pp. 127–130.
- 20. Quigley, E. Bifidobacterium longum. In *The Microbiota in Gastrointestinal Pathophysiology*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 139–141.

- 21. Quigley, E. Bifidobacterium bifidum. In *The Microbiota in Gastrointestinal Pathophysiology*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 131–133.
- 22. Xiao, Y.; Yang, C.; Yu, L.; Tian, F.; Wu, Y.; Zhao, J.; Zhang, H.; Yang, R.; Chen, W.; Hill, C.; et al. Human gut-derived *B. longum* subsp. *longum* strains protect against aging in a D-galactose-induced aging mouse model. *Microbiome* **2021**, *9*, 180. [CrossRef]
- 23. Wasilewska, E.; Markiewicz, L.; Bielecka, M. *Bifidobacterium* strains inhabiting the gastrointestinal tract of rat as potential probiotics for animals. *J. Anim. Feed Sci.* 2008, *17*, 398–410. [CrossRef]
- 24. Souza, T.; Silva, A.; Drews, J.; Gomes, D.; Vinderola, C.; Nicoli, J. In vitro evaluation of *Bifidobacterium* strains of human origin for potential use in probiotic functional foods. *Benef. Mirbobes* **2013**, *4*, 179–186. [CrossRef]
- 25. Vasquez, N.; Suau, A.; Magne, F.; Pochart, P.; Pélissier, M.-A. Differential effects of *Bifidobacterium pseudolongum* strain Patronus and metronidazole in the rat gut. *Appl. Environ. Microbiol.* **2009**, *75*, 381–386. [CrossRef]
- Sasajima, N.; Ogasawara, T.; Takemura, N.; Fujiwara, R.; Watanabe, J.; Sonoyama, K. Role of intestinal *Bifidobacterium pseudolongum* in dietary fructo-oligosaccharide inhibition of 2, 4-dinitrofluorobenzene-induced contact hypersensitivity in mice. *Br. J. Nutr.* 2010, 103, 539–548. [CrossRef]
- Lebeer, S.; Bron, P.A.; Marco, M.L.; Van Pijkeren, J.-P.; Motherway, M.O.C.; Hill, C.; Pot, B.; Roos, S.; Klaenhammer, T. Identification of probiotic effector molecules: Present state and future perspectives. *Curr. Opin. Biotechnol.* 2018, 49, 217–223. [CrossRef]
- 28. Ventura, M.; Turroni, F.; Motherway, M.O.C.; MacSharry, J.; van Sinderen, D. Host–microbe interactions that facilitate gut colonization by commensal bifidobacteria. *Trends Microbiol.* **2012**, *20*, 467–476. [CrossRef]
- 29. Wang, L.; Hu, L.; Yan, S.; Jiang, T.; Fang, S.; Wang, G.; Zhao, J.; Zhang, H.; Chen, W. Effects of different oligosaccharides at various dosages on the composition of gut microbiota and short-chain fatty acids in mice with constipation. *Food Funct.* **2017**, *8*, 1966–1978. [CrossRef]
- 30. Caporaso, J.G.; Kuczynski, J.; Stombaugh, J.; Bittinger, K.; Bushman, F.D.; Costello, E.K.; Fierer, N.; Peña, A.G.; Goodrich, J.K.; Gordon, J.I. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **2010**, *7*, 335. [CrossRef]
- 31. Cui, Y.; Yang, X.; Didelot, X.; Guo, C.; Li, D.; Yan, Y.; Zhang, Y.; Yuan, Y.; Yang, H.; Wang, J. Epidemic clones, oceanic gene pools, and eco-LD in the free living marine pathogen *Vibrio parahaemolyticus*. *Mol. Biol. Evol.* **2015**, *32*, 1396–1410. [CrossRef]
- 32. Delcher, A.L.; Salzberg, S.L.; Phillippy, A.M. Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinf.* **2003**, *10*, 1–18. [CrossRef]
- 33. Price, M.; Dehal, P.; Arkin, A. FastTree 2–Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef]
- Corander, J.; Marttinen, P.; Sirén, J.; Tang, J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. BMC Bioinform. 2008, 9, 539. [CrossRef]
- Page, A.; Taylor, B.; Delaney, A.; Soares, J.; Seemann, T.; Keane, J.; Harris, S. SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb. Genom.* 2016, 2, e000056. [CrossRef]
- 36. Goudet, J. Hierfstat, a package for R to compute and test hierarchical F-statistics. Mol. Ecol. Notes 2005, 5, 184–186. [CrossRef]
- 37. Seemann, T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 2014, 30, 2068–2069. [CrossRef] [PubMed]
- 38. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [CrossRef]
- 39. Li, L.; Stoeckert, C.J.; Roos, D.S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003, 13, 2178–2189. [CrossRef] [PubMed]
- 40. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef]
- Ravcheev, D.A.; Thiele, I. Comparative genomic analysis of the human gut microbiome reveals a broad distribution of metabolic pathways for the degradation of host-synthetized mucin glycans and utilization of mucin-derived monosaccharides. *Front. Genet.* 2017, *8*, 111. [CrossRef]
- 42. Hidalgo-Cantabrana, C.; Sánchez, B.; Milani, C.; Ventura, M.; Margolles, A.; Ruas-Madiedo, P. Genomic overview and biological functions of exopolysaccharide biosynthesis in *Bifidobacterium* spp. *Appl. Environ. Microbiol.* **2014**, *80*, 9–18. [CrossRef]
- Song, Z.; Cai, Y.; Lao, X.; Wang, X.; Lin, X.; Cui, Y.; Kalavagunta, P.K.; Liao, J.; Jin, L.; Shang, J. Taxonomic profiling and populational patterns of bacterial bile salt hydrolase (BSH) genes based on worldwide human gut microbiome. *Microbiome* 2019, 7, 9. [CrossRef]
- 44. El, A.K.; Armougom, F.; Gordon, J.; Raoult, D.; Henrissat, B. The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat. Rev. Microbiol.* **2013**, *11*, 497–504.
- 45. Boekhorst, J.; Helmer, Q.; Kleerebezem, M.; Siezen, R. Comparative analysis of proteins with a mucus-binding domain found exclusively in lactic acid bacteria. *Microbiology* (*N. Y.*) **2006**, 152, 273–280. [CrossRef]
- 46. Lugli, G.; Milani, C.; Turroni, F.; Duranti, S.; Mancabelli, L.; Mangifesta, M.; Ferrario, C.; Modesto, M.; Mattarelli, P.; Jiří, K. Comparative genomic and phylogenomic analyses of the *Bifidobacteriaceae* family. *BMC Genom.* **2017**, *18*, 568. [CrossRef]
- 47. Xiao, Y.; Wang, C.; Zhao, J.; Zhang, H.; Chen, W.; Zhai, Q. Quantitative detection of *Bifidobacterium longum* strains in feces using strain-specific primers. *Microorganisms* **2021**, *9*, 1159. [CrossRef]
- 48. Chen, Y.R.; Zhou, L.Z.; Fang, S.T.; Long, H.Y.; Chen, J.Y.; Zhang, G.X. Isolation of Desulfovibrio spp. from human gut microbiota using a next-generation sequencing directed culture method. *Lett. Appl. Microbiol.* **2019**, *68*, 553–561. [CrossRef]

- 49. Rowan, F.; Docherty, N.G.; Murphy, M.; Murphy, B.; Coffey, J.C.; O'Connell, P.R. Desulfovibrio bacterial species are increased in ulcerative colitis. *Dis. Colon Rectum* 2010, 53, 1530–1536. [CrossRef]
- 50. Liu, F.; Li, J.; Wu, F.; Zheng, H.; Peng, Q.; Zhou, H. Altered composition and function of intestinal microbiota in autism spectrum disorders: A systematic review. *Transl. Psychiatr.* **2019**, *9*, 43. [CrossRef]
- 51. Chen, J.; Chia, N.; Kalari, K.R.; Yao, J.Z.; Novotna, M.; Soldan, M.M.P.; Luckey, D.H.; Marietta, E.V.; Jeraldo, P.R.; Chen, X. Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. *Sci. Rep.* **2016**, *6*, 28484. [CrossRef]
- 52. Wang, F.; Jiang, H.; Shi, K.; Ren, Y.; Zhang, P.; Cheng, S. Gut bacterial translocation is associated with microinflammation in end-stage renal disease patients. *Nephrology* **2012**, *17*, 733–738. [CrossRef] [PubMed]
- Wagner, J.; Short, K.; Catto-Smith, A.G.; Cameron, D.J.; Bishop, R.F.; Kirkwood, C.D. Identification and characterisation of Pseudomonas 16S ribosomal DNA from ileal biopsies of children with Crohn's disease. *PLoS ONE* 2008, *3*, e3578. [CrossRef] [PubMed]
- 54. Biagi, E.; Nylund, L.; Candela, M.; Ostan, R.; Bucci, L.; Pini, E.; Nikkïla, J.; Monti, D.; Satokari, R.; Franceschi, C. Through ageing, and beyond: Gut microbiota and inflammatory status in seniors and centenarians. *PLoS ONE* **2010**, *5*, e10667. [CrossRef]
- 55. Markou, P.; Apidianakis, Y. Pathogenesis of intestinal *Pseudomonas aeruginosa* infection in patients with cancer. *Front. Cell. Infect. Microbiol.* **2014**, *3*, 115. [CrossRef]
- Nagasaki, A.; Takahashi, H.; Iinuma, M.; Uchiyama, T.; Watanabe, S.; Koide, T.; Tokoro, C.; Inamori, M.; Abe, Y.; Nakajima, A. Ulcerative colitis with multidrug-resistant *Pseudomonas aeruginosa* infection successfully treated with bifidobacterium. *Digestion* 2010, *81*, 204–205. [CrossRef] [PubMed]
- 57. Matsumoto, T.; Ishikawa, H.; Tateda, K.; Yaeshima, T.; Ishibashi, N.; Yamaguchi, K. Oral administration of *Bifidobacterium longum* prevents gut-derived Pseudomonas aeruginosa sepsis in mice. *J. Appl. Microbiol.* **2008**, *104*, 672–680. [CrossRef]
- 58. Phaik, L.O.; Benson, A.; Peterson, D.; Patil, P.; Moriyama, E.; Roos, S.; Walter, J. Diversification of the gut symbiont *Lactobacillus reuteri* as a result of host-driven evolution. *ISME J.* **2010**, *4*, 377–387.
- 59. Frese, S.A.; Benson, A.K.; Tannock, G.W.; Loach, D.M.; Kim, J.; Zhang, M.; Oh, P.L.; Heng, N.C.; Patil, P.B.; Juge, N. The evolution of host specialization in the vertebrate gut symbiont *Lactobacillus reuteri*. *PLoS Genet*. **2011**, 7, e1001314. [CrossRef]
- 60. Duar, R.M.; Lin, X.B.; Zheng, J.; Martino, M.E.; Grenier, T.; Pérez-Muñoz, M.E.; Leulier, F.; Gänzle, M.; Walter, J. Lifestyles in transition: Evolution and natural history of the genus *Lactobacillus*. *Fems Microbiol. Rev.* **2017**, *41*, S27–S48. [CrossRef]
- 61. Yang, C.; Pei, X.; Wu, Y.; Yan, L.; Yan, Y.; Song, Y.; Coyle, N.M.; Martinez-Urtaza, J.; Quince, C.; Hu, Q. Recent mixing of *Vibrio parahaemolyticus* populations. *ISME J.* 2019, *13*, 2578–2588. [CrossRef]
- 62. Wang, A.; Crowley, D.E. Global gene expression responses to cadmium toxicity in *Escherichia coli*. *J. Bacteriol*. **2005**, *187*, 3259–3266. [CrossRef]
- 63. Pagès, D.; Sanchez, L.; Conrod, S.; Gidrol, X.; Fekete, A.; Schmitt-Kopplin, P.; Heulin, T.; Achouak, W. Exploration of intraclonal adaptation mechanisms of *Pseudomonas brassicacearum* facing cadmium toxicity. *Environ. Microbiol.* 2007, *9*, 2820–2835. [CrossRef]
- 64. Zhai, Q.; Xiao, Y.; Zhao, J.; Tian, F.; Zhang, H.; Narbad, A.; Chen, W. Identification of key proteins and pathways in cadmium tolerance of *Lactobacillus plantarum* strains by proteomic analysis. *Sci. Rep.* **2017**, *7*, 1182. [CrossRef] [PubMed]
- Ellegaard, K.M.; Engel, P. Beyond 16S rRNA community profiling: Intra-species diversity in the gut microbiota. *Front. Microbiol* 2016, 7, 1475. [CrossRef] [PubMed]
- 66. Pande, S.; Merker, H.; Bohl, K.; Reichelt, M.; Schuster, S.; Kaleta, C.; Kost, C. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J.* **2014**, *8*, 953–962. [CrossRef] [PubMed]
- Degnan, B.A.; Macfarlane, G.T. Synthesis and activity of α-glucosidase produced by *Bifidobacterium pseudolongum*. *Curr. Microbiol*. 1994, 29, 43–47. [CrossRef]
- 68. Slovakova, L.; Dušková, D.; Marounek, M. Fermentation of pectin and glucose, and activity of pectin-degrading enzymes in the rabbit caecal bacterium *Bifidobacterium pseudolongum*. *Lett. Appl. Microbiol.* **2002**, *35*, 126–130. [CrossRef]
- 69. Milani, C.; Lugli, G.A.; Duranti, S.; Turroni, F.; Mancabelli, L.; Ferrario, C.; Mangifesta, M.; Hevia, A.; Viappiani, A.; Scholz, M. Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut. *Sci. Rep.* **2015**, *5*, 15782. [CrossRef]