Comparative Genomic Analysis of *Bifidobacterium Catenulatum* Reveal the Genetic Divergence of its Two Subspecies from Infant and Adult Gut

Jiaqi Liu

Inner Mongolia Agricultural University

Weicheng Li

Inner Mongolia Agricultural University

Caiqing Yao

Inner Mongolia Agricultural University

Jie Yu

Inner Mongolia Agricultural University

Heping Zhang (✉ nnd2020106@163.com)

Inner Mongolia Agricultural University

Research Article

Keywords: *Bifidobacterium catenulatum*, comparative genomics, genetic divergence, carbohydrate utilization

Posted Date: October 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-957925/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: *Bifidobacterium catenulatum*, which includes two subspecies that *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum* are usually from infant and adult gut respectively, while the genomic studies of functional difference and genetic divergence in them have been rarely reported. In this study, we analyzed 16 *B. catenulatum* strains through comparative genomics, including two novel sequenced strains.

Results: A phylogenetic tree based on 785 core genes indicated that the two subspecies of *B. catenulatum* were significantly separated and confirmed their colonizing bias in infants and adults. Comparison of general genomic characteristics revealed that the two subspecies had significantly different genomic sizes but similar GC content. Functional annotations found that they peculiarly differ in utilization of carbohydrates and amino acid. Among them, we found that carbohydrate metabolism seems to play an important role in the divergence because of their carbohydrate-active enzymes (CAZyme) present two different clustering patterns. *B. catenulatum* subsp. *kashiwanohense* have functional genes that specifically adapted to the infant gut for glycoside hydrolases 95 (GH95) and carbohydrate-binding modules 51 (CBM51), which specifically participated in the metabolism of Human Milk Oligosaccharides (HMOs), and specific genes *fuc* that related to HMOs were also detected. While *B. catenulatum* subsp. *catenulatum* rich in GH3 and glycosyltransferases 4 (GT4) tended to metabolize plant-derived glycan that may help it metabolize more complex carbohydrates (eg. xylan) in the adult intestine.

Conclusions: Our findings revealed genomic evidence of carbohydrate utilization bias which may be a key leading to the genetic divergence of two subspecies of *B. catenulatum*.

Introduction

Bifidobacterium are gram-positive, anaerobic microorganisms that are present in the gut of humans and other mammals [1]. Some strains of *Bifidobacterium* has attracted wide attention due to its probiotic function in regulating intestinal flora and immune metabolism [2, 3]. As an important member of the *Bifidobacterium* genus, *Bifidobacterium catenulatum* (*B. catenulatum*), which has been confirmed that some strains of it showed good probiotic characteristics, such as treatment of acute liver injury [4], inhibition of pathogenic bacteria *in-vitro*, as well as fermentation potential that presenting good survival in stored yoghurt [5]. These potential probiotic properties suggest that *B. catenulatum* may be a candidate for probiotics in food or medicine.

Bifidobacterium has long been considered as an important intestinal symbiotic bacteria that co-evolved with their hosts. Previous studies have shown that *Bifidobacterium* has obvious species bias in the intestinal environment of adults and infants [6]. For example, *B. bifidum* and *B. breve* are commonly found in the gut of infants, while *B. adolescentis* appear to be more usual in the intestinal tract of adults, though there are no well-defined species boundaries of it [7]. *B. catenulatum*, which has the typical

characteristic of divergence in the human gut [8]. According to the latest taxonomic status of *B. catenulatum* species [9], it contains two subspecies, *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum*. Previous studies have shown that *B. catenulatum* subsp. *kashiwanohense* is a subspecies that lives specifically in the gut of infants, while *B. catenulatum* subsp. *catenulatum* is more suitable for adults [7, 10]. Current research suggests that *B. catenulatum* adapt to different hosts is partially due to the functional preference of different subspecies, such as carbohydrate metabolism [10]. However, the phenomena that lead to different functional preferences in these two subspecies of *B. catenulatum* are mostly based on superficial experiments, and corresponding genomic evidence is lacking. It is necessary to fill the gap in genomic knowledge of genetic divergence and functional differentiation of the two subspecies which is essential for the supplement of *Bifidobacterium* resources and scientific support of their purported health benefits. In-species comparative genomics analysis allows for a deeper understanding of individual characters between genomes [11]. However, largely due to that *Bifidobacterium* genus are strictly anaerobes [12], their harsh culture conditions make a limited number of genomes of *B. catenulatum* sequenced. With the development of second-generation sequencing technology, the genomes of *B. catenulatum* began to be revealed [13]. While there is little literature on genomic analysis of this species, it indicates that overall genomic information of *B. catenulatum* remains largely unexplored.

In the current study, we analyzed a total of 19 genomes of *B. catenulatum* species, including 12 *B. catenulatum* subsp. *catenulatum* and 5 *B. catenulatum* subsp. *kashiwanohense* from the Refseq database, with two novel sequenced strains, IMAUFB085 and IMAUFB087. Employing comparative genomic approaches, we dissected genetic background and functional genomic information in *B. catenulatum*. This work not only allows general insights into the genomic differences between two subspecies of *B. catenulatum*, but also has possibility to reveal the key factors leading to the divergence of them.

Results

Taxonomic status of *B. catenulatum* strains

To confirm that sequence similarity and taxonomic status among the strains in this study, the pairwise ANI and TNI values of all twenty genome assemblies were calculated (Fig. 1). Strains with an ANI value of over 95% are generally considered to be of the same species [14]. ANI and TNI obtained similar clustering results, that distinct subspecies branches were obtained. IMAUFB085 and IMAUFB087 were grouped with most of the *B. catenulatum* subsp. *catenulatum* strains, ANI values of them compared to *B. catenulatum* subsp. *catenulatum* JCM1194^T were 98.41% and 98.42%, and TNI values were 87.45% and 84.48%, respectively. Above results confirmed that IMAUFB085 and IMAUFB087 were classified as *B. catenulatum* subsp. *catenulatum*. These two strains were previously identified as *B. pseudocatenulatum* by 16S rRNA (Additional file 1.), but 16S rRNA could not reliably distinguish the species of *Bifidobacterium* [15], so we revised their taxonomic status to *B. catenulatum* subsp. *catenulatum* here.

ANI analysis revealed that three *B. catenulatum* subsp. *catenulatum* strains, namely, JGBg468, BCJG468, and MC1, significantly differed from the other *B. catenulatum* subsp. *catenulatum* strains (Their ANI values compared to JCM1194^T were 93.83%, 93.88% and 93.86%). It is assumed that the taxonomic information of these three strains was wrongly recorded, so these three strains were subsequently excluded. Cluster analysis clearly distinguished two subspecies, ANI between the two subspecies groups were greater than 95%, and within the subspecies were greater than 98%, indicating that these strains belonged to the same species.

Comparison of general genomic features between two subspecies

The general information of the strains shows that all of *B. catenulatum* subsp. *kashiwanohense* are derived from infants, while only two strains of *B. catenulatum* subsp. *catenulatum* are known to be infantile isolates (Additional file 2.). The general genomic features of 17 strains of *B. catenulatum* species and two novel strains, IMAUFB085 and IMAUFB087, are summarized in Table 1. The genomic characteristics within *B. catenulatum* species showed different degrees of difference. The genome size and GC content of *B. catenulatum* isolates were 2.16 ± 0.13 Mb and $56.21\% \pm 0.11\%$, respectively. Comparing the basic genomic characteristics between the two subspecies (Additional file 3.), the results showed that the genome size of *B. catenulatum* subsp. *kashiwanohense* (2.36 ± 0.05 Mb) was significantly higher than that of *B. catenulatum* subsp. *catenulatum* (2.09 ± 0.07 Mb) ($p < 0.0001$), while there was no significant difference in GC content ($p > 0.05$). This strong genomic differences reflected the speciation boundaries of the two subspecies, while the consistency of GC content represented the intimate relationship between them [16, 17]. In addition, *B. catenulatum* subsp. *kashiwanohense* contained more CDSs than *B. catenulatum* subsp. *catenulatum* ($p < 0.01$), and there was no statistical difference in the number of tRNA ($p > 0.05$).

Table 1

General genomic features of *B. catenulatum* genomes

Collection strain	Genome size (Mb)	GC content (%)	No of CDSs	No of tRNAs
IMAUFB087	2.01	56.06	1,834	56
IMAUFB085	1.98	55.94	1,781	54
<i>B. catenulatum</i> subsp. <i>catenulatum</i> JCM1194 ^T	2.08	56.20	1,616	56
<i>B. catenulatum</i> subsp. <i>catenulatum</i> DSM16992	2.06	56.10	1,606	56
<i>B. catenulatum</i> subsp. <i>catenulatum</i> LMG11043	2.08	56.11	1,515	56
<i>B. catenulatum</i> subsp. <i>catenulatum</i> DSM16992(2)	2.11	56.41	1,616	56
<i>B. catenulatum</i> subsp. <i>catenulatum</i> 1899B	2.12	56.25	1,656	56
<i>B. catenulatum</i> subsp. <i>catenulatum</i> A2	2.02	56.15	1,584	54
<i>B. catenulatum</i> subsp. <i>catenulatum</i> A1	2.06	56.21	1,659	56
<i>B. catenulatum</i> subsp. <i>catenulatum</i> A3	2.15	56.36	1,707	59
<i>B. catenulatum</i> subsp. <i>catenulatum</i> HGUT-01490	2.08	56.20	1,615	56
<i>B. catenulatum</i> subsp. <i>kashiwanohense</i> PV20-2	2.37	56.12	1,876	58
<i>B. catenulatum</i> subsp. <i>kashiwanohense</i> JCM15439 ^T	2.34	56.30	1,842	54
<i>B. catenulatum</i> subsp. <i>kashiwanohense</i> APCKJ1	2.45	56.20	1,968	54
<i>B. catenulatum</i> subsp. <i>kashiwanohense</i> DSM21854	2.31	56.20	1,758	53
<i>B. catenulatum</i> subsp. <i>kashiwanohense</i> DSM21854(2)	2.32	56.30	1,854	68

Phylogenetic divergence of two subspecies of *B. catenulatum*

The classification of species and the establishment of intraspecific relationships are frequently based on phylogenetic analysis. We reconstructed the phylogenetic structure of *B. catenulatum* species and specified *B. pseudocatenulatum* JCM1200^T as the outgroup. A NJtree on the basis of 785 core genes was constructed and visualized with bootstrap support of 1000 replications (Fig. 2A). This phylogenetic tree well confirmed the phenomenon of subspecies divergence of *B. catenulatum*, 16 *B. catenulatum* strains in our analysis were clearly divided into two subgroups, *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum* subgroup, and showed that there were diacritical differences between the two subspecies at gene level. By annotating the source of the isolates, interestingly, there was a significant association between the *B. catenulatum* strains and their isolated source. Infant isolates,

including all strains of *B. catenulatum* subsp. *kashiwanohense* and two of *B. catenulatum* subsp. *catenulatum*, showed intra-specific genetic similarity, while the rest were adult isolates in another cluster that they also appeared close phylogenetic relationships. This suggests that the trend of divergence of the *B. catenulatum* strains may be dependent on their hosts. We believe that *B. catenulatum* may adapt its functions to infant and adult intestines respectively, thus gradually differentiating into subspecies of *B. catenulatum* from the species level.

Constructing the pan-core genome of *B. catenulatum*

The gene pool of a population contains all the genetic material and functions of a species. Roary was used to calculate the pan-core genome of the above-mentioned 16 *B. catenulatum* strains, in total 4608 pan genes were searched. The above phenomenon is that the two subspecies of *B. catenulatum* branched from the same ancestor made them share a certain proportion of core genes, with the number of 998 (21.66%) (Fig. 2B). The genetic distribution of *B. catenulatum* showed that there were individual gene set in the two subspecies that 87 core genes in *B. catenulatum* subsp. *kashiwanohense* and 63 in *B. catenulatum* subsp. *catenulatum*, these unique core gene sets can provide a internal basis for differentiation of *Bifidobacterium* species [18]. Additionally, there are different numbers of strain specific genes in the *B. catenulatum* species, their numbers ranged from 20 to 578, which reflected the potential genetic diversity among *B. catenulatum* species.

Subsequently, we established the pan-core genes curves for genomes of the species *B. catenulatum* (Additional file 4.A). With the augment of new genomes, the number of pan genes increases, which indicates the existence of an open pan-genome within the species *B. catenulatum*. Contrary to the pan genome, the core genes are not expected to be significantly reduced in number by the addition of further genomes since the exponential trendline essentially reached the number of 1000. At the same time, we found that *B. catenulatum* subsp. *catenulatum* has a fairly open pan-core genome (Additional file 4.B), while *B. catenulatum* subsp. *kashiwanohense* tends to be closed (Additional file 4.C). These results indicate that *B. catenulatum* subsp. *catenulatum* may have strong ability to be adaptable to the environment, while *B. catenulatum* subsp. *kashiwanohense* exists in a more specific and conservative habitat. This result is consistent with the isolated information of the *B. catenulatum* strains in this study (Additional file 2.), for *B. catenulatum* subsp. *kashiwanohense* only exists in the intestinal tract of infants and it is an infrequently encountered species [10], while *B. catenulatum* subsp. *catenulatum* can exist in both the intestinal tract of infants and adults and has a large number, so we speculated that more novel strains of *B. catenulatum* subsp. *catenulatum* may be discovered in the future than *B. catenulatum* subsp. *kashiwanohense*.

Comparison of the main functions between two subspecies

Above results reveal genetic differences between the two subspecies from the general genomics level, which are usually associated with functional differentiation [18]. Therefore, it is necessary to conduct further functional genomic comparisons between the two subspecies of *B. catenulatum*. In order to acquire the functional genomic difference of them, all strains were annotated online through the RAST website. Examination of the functional annotation of the 16 genomes, functional annotation of *B. catenulatum* genome sequence classified proteins into 23 functional categories (Additional file 5.). Results suggests that the amino acid derivatives (21.06%) functional is the most highly represented functional category within the *B. catenulatum* followed by protein metabolism (21.00%), carbohydrate metabolism (15.73%), and cofactors, vitamins, prosthetic groups, and pigments (9.76%). This indicates that a strong ability exists for substrate utilization by *B. catenulatum*. Comparison of the main functional differences between the two subspecies showed that *B. catenulatum* subsp. *kashiwanohense* had wholly stronger metabolic capacity than *B. catenulatum* subsp. *catenulatum*, and there were significant differences in the metabolism of amino acid ($p<0.01$), carbohydrates ($p<0.01$), and protein ($p<0.05$). While there was no statistical difference in the cofactors, vitamins, prosthetic groups, and pigments-related categories ($p>0.05$) (Fig. 3).

In view of the remarkably significant differences in the metabolic functions of amino acids and carbohydrates in the two subspecies, we first compared the detailed functions of metabolic amino acids in the two subspecies (Additional file 6.A). We found that the main difference lies in a subfunction of amino acids, lysine, urea cycle, polyamines and cysteine, which is also the most abundant amino acid metabolic category in *B. catenulatum*. Further analysis of the details of this function showed that methionine biosynthesis was the main factor affecting the difference of amino acid function between the two subspecies, and methionine biosynthesis of *B. catenulatum* subsp. *kashiwanohense* was significantly stronger than that of *B. catenulatum* subsp. *catenulatum* (Additional file 6.B, C). This suggests that the functional difference in methionine biosynthesis is the key to the difference in the amino acid functional genomes of the two subspecies of *B. catenulatum*.

Different carbohydrate utilization patterns in two subspecies

In order to compare the carbohydrate utilization ability of *B. catenulatum* at the genomic level, the functional genes of carbohydrate active enzymes of 16 *B. catenulatum* strains were analyzed. As shown in fig. 4A, 16 strains of *B. catenulatum* were distributed in all six carbohydrate active enzyme families, which indicated that they had rich carbohydrate function. It is worth noting that the clustering results of CAZyme are roughly consistent with phylogenetic trees that the two subspecies are clearly distinguished. This not only suggests that the two subspecies have different metabolic patterns in terms of carbohydrate utilization, but indicated that CAZy -related genes were closely associated with the divergence of subspecies of *B. catenulatum*.

Among the identified GH families, the most dominant were the GH3, GH13 and GH43 families in *B. catenulatum* species, and GT2 and GT4 were the main family of carbohydrate enzymes within *B. catenulatum* species. By comparing the main carbohydrate hydrolase families between subspecies (Fig. 4B), it was found that the GH3 ($p < 0.0001$) and GT4 ($p < 0.01$) of *B. catenulatum* subsp. *catenulatum* were significantly higher than those of *B. catenulatum* subsp. *kashiwanohense*. Which indicates that *B. catenulatum* subsp. *catenulatum* has stronger ability to utilize β -D-glucosyl, sucrose and cellulose [19]. However, there was not statistically significant difference in the function of GH13, GH43 and GT2 between the two subspecies ($p > 0.05$) (Fig. 4C,D,E).

By analyzing the specific carbohydrase of *B. catenulatum* subsp. *kashiwanohense*, we identified five families closely related to it, including GH18, CBM5, GH95, CBM51 and CBM66 (Fig. 4G). The CBM family is primarily responsible for banding carbohydrates, GH18 family often combines with CBM5 to participate in the function of chitinase, and CBM66 mainly assists in the degradation of fructose [20]. In particular, GH95 is specifically participated in the production of α -L-fucosidase, which is the most abundant substance in HMOs and is closely related to the function of infant-specific species [21]. Additionally, CBM51 is beneficial to bind to GH95 and help it pick up fucose substrates to metabolize HMOs [22]. These two CAZy families, CBM51 and GH95 may be conducive to the colonization of *B. catenulatum* subsp. *kashiwanohense* in infant intestines, especially the utilization of HMOs, which is in contrast to the abundance of plant-derived glycan of *B. catenulatum* subsp. *catenulatum*, and further reflects the bias of the two subspecies in the utilization of carbohydrates.

HMOs-related genes in *B. catenulatum* genomes

To further explore the deletion of HMOs-related genes and the overall genomic differences between the two subspecies, the BLAST Ring Image Generator (BRIG) was used to graphically compare *B. catenulatum* strains using type strain *B. catenulatum* subsp. *kashiwanohense* JCM15439^T as reference (Fig. 5). In order to compare the presence and deletion of HMOs-related genes in the two subspecies, we annotated four *fuc* genes (*fuc1*, *fuc2*, *fuc3* and *fuc4*) that encode α -L-fucosidase in reference genome JCM15439^T based on the close relationship between L-fucose and HMOs mentioned above. The results showed that all *fuc* genes were specifically found in *B. catenulatum* subsp. *kashiwanohense* but not in *B. catenulatum* subsp. *catenulatum*. *fuc2*, *fuc3* and *fuc4* are included in all *B. catenulatum* subsp. *kashiwanohense* genomes except *fuc1*, which is absent in PV20-2. Above phenomenon further confirms the unique functional advantage of *B. catenulatum* subsp. *kashiwanohense* subspecies in metabolic HMOs.

Overall, the figure shows that the majority of sequences present in JCM15439^T are also present in all other strains, the genomes were more than 90% identical. However, it shows that two large genome gaps (GGs) exist separately in the two newly sequenced strains, IMAUFB085 and IMAUFB087, which have less than 70% of the matched-degree compared to JCM15439^T. In general, sequences located in the GGs

represent hypothetical CDSs, genomic islands, or prophages [23], which indicates that there are still many unknown functions of these two strains to be explored.

Discussion

Bifidobacterium, as a typical intestinal symbiotic bacteria, has experienced a long and extensive evolutionary process in the human hosts [1]. *B. catenulatum* is one such species that has evolved into two subspecies, including *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum*. Previous studies have revealed that *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum* have a highly intimate relationship in the phylogenetic tree [24]. Through phylogenetic reconstruction in this study, we obtained clear clustering results for two subspecies of *B. catenulatum*. On the one hand, we found common characteristics of the two subspecies within the species of *B. catenulatum* by general comparative genomic analysis. For example, GC content is the genetic characteristic of the species. Our results indicated that *B. catenulatum* species appeared high GC content, from 55.94–56.41%, which accords with the characteristics of *Bifidobacterium* [16]. On the other hand, we have revealed genomic differences between the two subspecies. For example, the genome size and number of CDSs of *B. catenulatum* subsp. *catenulatum* was significantly lower than that of *B. catenulatum* subsp. *kashiwanohense*, and there are separate core gene sets in both of them. Such genomic differences represent a marker of genetic divergence [17]. Given that *B. catenulatum* subsp. *kashiwanohense* are all separated from infants, and *B. catenulatum* subsp. *catenulatum* mainly colonized in the adult gut, this suggests that they may have retained different functions in order to adapt to intestinal environments of infant and adult respectively. In order to find potential genomic evidence to confirm this hypothesis, we further explored the possible association between subspecies divergence and host through functional genomic comparisons.

Functional genomics revealed significant differences in amino acids and carbohydrates between the two subspecies of *B. catenulatum*. Amino acid metabolism in *Bifidobacterium* appears to have been less studied than carbohydrate metabolism. We found that the function of methionine biosynthesis may be the main factor leading to the difference of amino acid metabolism between the two subspecies, because the number of methionine synthesis functional genes of *B. catenulatum* subsp. *kashiwanohense* were significantly higher than that of *B. catenulatum* subsp. *catenulatum*. Methionine is an essential amino acid for infants and can be used as a biochemical indicator for metabolomics studies [25]. Previous studies have confirmed that the increased metabolic capacity of intestinal microbes is related to their methionine synthesis [26]. Based on this conclusion, we speculated that the metabolic capacity of *B. catenulatum* subsp. *kashiwanohense* was stronger than that of *B. catenulatum* subsp. *catenulatum* from the results of functional genome. For carbohydrates related functions, given the *Bifidobacterium* mainly obtained competitive advantage through the intake of sugar in the animal intestinal [27], so we conducted detailed analysis of the carbohydrate-active enzyme of *B. catenulatum*, it is interesting to note that we found the function of subspecies differentiation and related characteristics of baby and adult host relevance that clustering of carbohydrate function between the two subspecies yielded similar

relationships to phylogenetic trees which based on core genes, and functional genomic evidence of adaptation of the two subspecies to the infant and adult intestines, respectively.

Bifidobacterium are saccharolytic microorganisms and their ability to utilize indigestible carbohydrates is an important property to assist in their establishment in the gastrointestinal tract [28]. In our study, functional genomics analysis confirmed that carbohydrate metabolism was one of the main functions of *Bifidobacterium*. Analysis of CAZyme indicated that GH3, GH13 and GH43 families that were involved in the hydrolysis of starch, cellulose, and their related hydrolysed derivatives were highly conserved in *B. catenulatum*, which is consistent with the GHs characteristics of the whole *Bifidobacterium* genus [18]. Notably, the CAZy cluster is basically consistent with the phylogenetic tree, showing obvious differentiation of the two subspecies. This result suggests that functional differences in carbohydrates may be related to genetic divergence of *B. catenulatum*. We found that the GH3 content of *B. catenulatum* subsp. *catenulatum* was significantly higher than that of *B. catenulatum* subsp. *kashiwanohense*. Previous studies have shown that GH3 is a key family in the evolution of *Bifidobacterium* that was involved in the degradation of plant polysaccharides, such as xylan [15]. Our results indicate that GH3 is also a key factor for the divergence of the two subspecies of *B. catenulatum* in carbohydrate function. In addition, the high GT4 content of *B. catenulatum* subsp. *catenulatum* also indicated its strong ability to decompose plant-derived polysaccharides, which is beneficial to the colonization of the strains in the intestinal tract of adults [19]. Studies have shown that the gut environment of adults is far more complex than that of infants because adults typically consume more difficult-to-digest, long-chain carbon sources, such as plant-based dietary fiber [7]. Kim et al. found that *B. catenulatum* strains can degrade fructooligosaccharides (FOS) in nutritionally restricted environments [29]. Previous studies have shown that a low-fiber diet in adults can cause a significant increase in the abundance of *B. catenulatum* [30]. We demonstrate that *B. catenulatum* subsp. *catenulatum* has more carbohydrates that utilize plant-derived glycans, which is conducive to the decomposition of difficult-to-use plant-derived glycans in the adult gut to exert a probiotic effect.

Infants, especially those who are breastfed, have a lot of HMOs in their intestines. HMOs is a prebiotic unique to breast milk that is found specifically in human breast milk and cannot be used by adults because they lack the necessary GHs [31]. *Bifidobacterium* metabolism of HMOs is beneficial to reducing the degradation of gastrointestinal mucin and pathogenic bacteria, thereby strengthening the immune system in infants [32]. For *B. catenulatum* subsp. *kashiwanohense*, which is characterized by infant adaptation [10], has two specific CAZymes, including GH95 and CBM51. It is reported that GH95 mainly utilizes fucosyllactose, a major component of HMOs. CBM51 is beneficial to bind to GH95 and helps it pick up fucose substrates to metabolize HMOs [22]. The ability of infantile specific *Bifidobacterium* to metabolize HMOs has been recognized as a specific marker of its adaptive colonization and plays an important evolutionary role in the infant gut [32]. In a word, *B. catenulatum* subsp. *kashiwanohense* may be adapted to the colonization of infant intestinal tract due to its HMOs-related families. Liu et al. [33] found that the presence of GH95 in *B. breve* could enable the strain to coexist in the intestinal tract of infants with HMOs as the only carbon source. James et al. [10] found *B. catenulatum* subsp. *kashiwanohense* APCKJ1 expressed specific genes that are involved in uptaking and hydrolysing

fucosyllactose, which are related to its ability to consume 2'-fucosyllactose (2'-FL) or 3'-fucosyllactose (3'-FL) as its sole carbohydrate source. Our research confirmed that *fuc* genes encode L-fuconate dehydratase specificity exists in *B. catenulatum* subsp. *kashiwanohense* strains while not in *B. catenulatum* subsp. *catenulatum*. These results further indicate the specificity of *B. catenulatum* subsp. *kashiwanohense* in metabolic HMOs, and its colonization in the intestinal tract of infants may be beneficial to the growth and development of infants. *B. catenulatum* subsp. *catenulatum* 1899B and IMAUFB085 all belong to known infant isolates but no HMOs genes were found in them, thus we speculated that they may be acquired directly through vertical mother-child transmission [27].

Taken together, our results support the host adaptations in *B. catenulatum* species from the perspective of comparative genomics. We speculated that the *B. catenulatum* species evolved to retain the competitive carbohydrate function genes to adapt to the intestinal environment of infants and adults, respectively, driving the emergence of two subspecies. However, the strains of the *B. catenulatum* are limited, and more strains will need to be sequenced in the future to complement their genetic and functional genomics information.

Conclusion

In summary, the genetic divergence and functional differences between two subspecies of *B. catenulatum* were revealed by comparative genomics in this study. Phylogenetic analysis confirmed that *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum* subspecies were generated from *B. catenulatum* and tended to disperse into infant and adult hosts, respectively. To explore whether this genetic differentiation is related to functional preference, further functional genomics identified predominant genes for colonization of two subspecies in their preferred hosts, that the differences in carbohydrate utilization of metabolic HMOs and plant-derived glycan may be a key for the genetic divergence. This study provided genomic evidence for potential host adaptation phenomenon of *B. catenulatum* in infant and adult intestine. However, the strains of the *B. catenulatum* are limited, and more strains will need to be sequenced in the future to complement their genetic and functional genomics information.

Methods

Bacterial strains, DNA extraction and publicly available assemblies

In this study we sequenced the genomes of two strains of *B. catenulatum* named IMAUFB085 and IMAUFB087 respectively, which were provided by the Lactic acid bacteria collection center (LABCC). Moreover, IMAUFB085 was isolated from infant feces and IMAUFB087 from adult feces in Tibet, China [34].

The two strains were cultured under anaerobic conditions in Man Rogosa and Sharpe (MRS) broth added L-cysteine hydrochloride (Oxoid Ltd., Basingstoke, United Kingdom) at 37°C for 24h. DNA extraction was performed using the TIANamp Bacteria DNA Kit (Tiangen, Beijing, China) in accordance with the instructions of manufacturer. Genomic DNA was quantified using a TBS-380 fluorometer (Turner BioSystems Inc., Sunnyvale, CA). High quality DNA samples were obtained to construct fragment libraries.

In addition, we obtained publicly available assemblies of all *B. catenulatum* strains, including *B. catenulatum* subsp. *catenulatum* (12 strains) and *B. catenulatum* subsp. *kashiwanohense* (5 strains) from National Coalition Building Institute (NCBI, <https://www.ncbi.nlm.nih.gov/>) on 4 February 2021, including their type strains, namely *B. catenulatum* subsp. *catenulatum* JCM1194^T and *B. catenulatum* subsp. *kashiwanohense* JCM15439^T respectively. Additionally, in order to infer phylogenetic relationships across species within it, we have downloaded the type strain of *B. pseudocatenulatum* species in the *B. adolescentis* group, *B. pseudocatenulatum* JCM1200^T, which is most closely related to *B. catenulatum* according to the phylogenetic relationship of *Bifidobacterium* genus in previous study [24].

Genome Sequencing and assembly

Genome sequencing was performed using the Illumina HiSeq platform (Illumina, Inc., San Diego, CA, United States), generating 150 bp paired-end reads per sample. Then the sequences were filtered through the Illumina HiSeq system. The high quality sequences were assembled using SOAPdenovo2 [35] on a 64-bit Linux system. High quality data was generated for each strain, corresponding to a sequencing depth of about 387-fold. Local inner gaps were filled and single base errors were corrected using the software GapCloser (<http://sourceforge.net/projects/soapdenovo2/files/GapCloser/>).

Genome annotation

In this study, all the general genomic information of *B. catenulatum* genomes was generated using self-made Perl scripts. To obtain the functional gene information of this bacterium, gene prediction and preliminary annotation of all *B. catenulatum* genomes were performed with the Rapid Annotation using Subsystems Technology (RAST) server (<https://rast.nmpdr.org/rast.cgi>) through RASTtk annotation scheme. Identification of tRNA genes using tRNAscan-SE (<http://trna.ucsc.edu/tRNAscan-SE/>).

ANI (Average nucleotide identity) and TNI (Total nucleotide identity)

To evaluate the genetic relatedness between the two subspecies in *B. catenulatum* and to confirm the taxonomic status of the strains in this study, we analyzed ANI and TNI values of all strains. *B.*

pseudocatenulatum JCM1200^T, as type strain of *B. pseudocatenulatum* which were most phylogenically related to *B. catenulatum* [24], were included in the comparison. All pairwise ANI values were calculated according to the method proposed by Goris et al. [36]. TNI values were calculated according to the method proposed by Chen et al. [37]. The clustering heat map was drawn by TBtools software (38Chen, Chen, Zhang, Thomas, Frank, He, & Xia., 2020).

Construction of pan-core genome and strain-specific genes

The annotated genomes of *B. catenulatum* were obtained with Prokka v1.12 [39] were then processed with Roary [40] to identify the pan genes, core genes and specific genes, using default parameters. Intersection groups, representing unique sets of gene identified only between intersected genomes, were visualized by UpSet diagram in TBtools software.

Phylogenetic analysis

The core gene alignment resulted from Roary was then used in TreeBeST [41] with 1,000 bootstrap iterations to build a phylogenetic NJtree by way of Neighbor-Joining (NJ) [42]. The phylogenetic trees were then visualized and annotated in iTOL (<https://itol.embl.de/>).

BRIG (BLAST Ring Image Generator)

Adopting BRIG v0.95 [43] software to compare the genomes of *B. catenulatum* based on JAVA language environment, it uses CG view [44] for image rendering and BLAST [45] for genome comparisons. BRIG can generate images that show multiple prokaryote genome comparisons, without an arbitrary limit on the number of genomes compared [43].

CAZyme (Carbohydrate-active enzymes) online annotation

Given the carbohydrate metabolic activity of *Bifidobacterium*, CAZyme online annotation of *B. catenulatum* strains were accomplished through three annotation tools, including HMMER, DIAMOND, and Hotpep searches [46]. Carbohydrate active enzymes were annotated in the genome sequence to predict potential families of glycosyltransferases (GTs), glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases (PLs), auxiliary activity (AA), and carbohydrate-binding modules (CBMs). Identification of CAZymes across *B. catenulatum* genomes was carried out using the dbCAN2 meta server (<http://bcb.unl.edu/dbCAN2/>). According to the annotation results, we checked the detailed information of active carbohydrate enzyme family on the CAZy official website (<http://www.aczy.org/>).

Statistical Analysis

Data are presented as means $\pm$ SEM. T-test method was used to verify the significance of difference between groups, visualization with ggpubr packages in R (4.0.3). Significance was set at p -value < 0.05 .

Abbreviations

LABCC: Lactic acid bacteria collection center; MRS: Man Rogosa and Sharpe; ANI: Average nucleotide identity; TNI: Total nucleotide identity; CDSs: coding sequences; NJ: Neighbor-Joining; RAST: Rapid Annotation using Subsystems Technology; BRIG: BLAST Ring Image Generator; HMOs: Human Milk Oligosaccharides; *B. catenulatum*: *Bifidobacterium catenulatum*; CAZyme: Carbohydrate-active enzymes; GTs: glycosyltransferases; GHs: glycoside hydrolases; CEs: carbohydrate esterases; PLs: polysaccharide lyases; AA: auxiliary activity; CBMs: carbohydrate-binding modules; GGs: genome gaps; FOS: fructooligosaccharides; 2'-FL: 2'-fucosyllactose; 3'-FL: 3'-fucosyllactose.

Declarations

Data availability

The assembly of two new isolates sequenced in this work were submitted as a Whole Genome project (BioProject No. PRJNA751426) at GenBank under the accessions JAIEWL000000000 (IMAUFB087) and JAIEWM000000000 (IMAUFB085), which are available at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA751426>. All of the phylogenetic trees and alignment files in this study were submitted to the TreeBASE web (<https://treebase.org/treebase-web>), with the accession no.28852. (<http://purl.org/phylo/treebase/phyloids/study/TB2:S28852>).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The assembly of two isolates sequenced in this work were submitted as a Whole Genome Shotgun project (BioProject No. PRJNA751426) at GenBank under the accessions JAIEWL000000000 (IMAUFB087) and JAIEWM000000000 (IMAUFB085) (see Additional file 2.).

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by China Agriculture Research System of MOF and MARA, and Science and Technology Major Projects of Inner Mongolia Autonomous Region (2021ZD0014).

Authors' contributions

HZ designed the study. JL and WL performed comparative genomics analyses and wrote the manuscript. CY and JY participated in the culture and sequencing of two new strains in this study. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Turrone F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, *et al.* *Diversity of bifidobacteria within the infant gut microbiota. PLoS ONE.* 2018a;7:e36957.
2. Assad SE, Rolny IS, Minnaard J, Pérez PF. *Bifidobacteria from human origin: interaction with phagocytic cells. J Appl Microbiol.* 2021;130:1357-1367.
3. Luisa M, Michele C, Sonia B, Matteo S, Giacomo B, Susanna V, Rosanna I. *The Role of Bifidobacteria in Predictive and Preventive Medicine: A Focus on Eczema and Hypercholesterolemia. Microorganisms.* 2021;9:836.
4. Firman J, Liu L, Zhang L, Argoty GA, Wang M, Tomasula P. *The effect of quercetin on genetic expression of the commensal gut microbes Bifidobacterium catenulatum, Enterococcus caccae and Ruminococcus gnavus. Anaerobe.* 2016;42:130-141.
5. Onyibe JE, Oluwole OB, Ogunbanwo ST, Sanni AI. *Antibiotic Susceptibility Profile and Survival of Bifidobacterium adolescentis and Bifidobacterium catenulatum of Human and Avian Origin in Stored Yoghurt. Niger Food J.* 2013;31:73-83.
6. Sabrina D, Giulia L, Marco V, Douwe vS, Francesca T. *Exploring the Ecology of Bifidobacteria and Their Genetic Adaptation to the Mammalian Gut. Microorganisms.* 2020;9:8.
7. Turrone F, Milani C, Duranti S, Ferrario C, Lugli GA, Mancabelli L, *et al.* *Bifidobacteria and the infant gut: an example of co-evolution and natural selection. Cell Mol Life Sci.* 2018b;75:103-118.

8. Ismail IH, Boyle RJ, Licciardi PV, Oppedisano F, Lahtinen S, Robins-Browne RM, Tang MLK. *Early gut colonization by Bifidobacterium breve and B. catenulatum differentially modulates eczema risk in children at high risk of developing allergic disease. Pediatr Allergy Immunol.* 2016;27:838-846.
9. Nouioui I, Carro L, Garcia-Lopez M, Meier-Kolthoff JP, Woyke T, Kyrpides NC, et al. *Genome-Based Taxonomic Classification of the Phylum Actinobacteria. Front Microbiol.* 2018;9:2007.
10. James K, Bottacini F, Contreras JIS, Vigoureux M, Egan M, Motherway MOc, et al. *Metabolism of the predominant human milk oligosaccharide fucosyllactose by an infant gut commensal. Sci Rep.* 2019;9:1589-1596.
11. He Q, Hou Q, Wang Y, Li J, Li W, Kwok L-Y, et al. *Comparative genomic analysis of Enterococcus faecalis: insights into their environmental adaptations. BMC Genomics.* 2018; 19:1173-1176.
12. Yang SH, Xie XQ, Ma JM, He XX, Li Y, Du MZ, et al. *Selective Isolation of Bifidobacterium From Human Faeces Using Pangenomics, Metagenomics, and Enzymology. Front Microbiol.* 2021;12:649698.
13. Morita H, Toh H, Nakano A, Oshima K, Takagi M, Suda W, et al. *Complete Genome Sequence of Bifidobacterium kashiwanohense JCM 15439T, Isolated from Feces from a Healthy Japanese Infant. Genome Announc.* 2015;3:e00255-15.
14. Seok-Hwan Y, Sung-Min H, Jeongmin L, Soonjae K, Jongsik C. *A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie van Leeuwenhoek.* 2017;110: 1281-1286.
15. Lugli GA, Milani C, Turrone F, Duranti S, Ferrario C, Viappiani A, et al. *Investigation of the Evolutionary Development of the Genus Bifidobacterium by Comparative Genomics. Appl Environ Microbiol.* 2014;80:6383-6394.
16. Pablo A-M., Matilde F, Mary OC-M, Joan OCK, Nicolas S, F FG, et al. *A conserved two-component signal transduction system controls the response to phosphate starvation in Bifidobacterium breve UCC2003. Appl Environ Microbiol.* 2012;78:5258-5269.
17. Marais GAB, Batut B, Daubin, V. *Genome Evolution: Mutation Is the Main Driver of Genome Size in Prokaryotes. Current biology: CB.* 2020;30(19):R1083-R1085.
18. Abdelhamid AG, ElDoudoug NK. *Comparative genomics of the gut commensal Bifidobacterium bifidum reveals adaptation to carbohydrate utilization. Biochem Bioph Res Co.* 2021;547:155-161.
19. Mao B, Yin R, Li X, Cui S, Zhang H, Zhao J, Chen W. *Comparative Genomic Analysis of Lactiplantibacillus plantarum Isolated from Different Niches. Genes.* 2021;12:241.
20. Cuskin F, Flint JE, Gloster TM, Morland C, Basle A, Henrissat B, Coutinho PM, Strazzulli A, Solovyova AS, Davies GJ, et al. *How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. Proc Natl Acad Sci U S A.* 2012;109:20889-20894.
21. Wan L, Zhu Y, Zhang W, Mu W. *Alpha-L-Fucosidases and their applications for the production of fucosylated human milk oligosaccharides. Appl Microbiol Biotechnol.* 2020;104:5619-5631.
22. Gregg KJ, Finn R, Abbott DW, Boraston AB. *Divergent modes of glycan recognition by a new family of carbohydrate-binding modules. J Biol Chem.* 2008;283:12604-12613.

23. Holm KO, Bækkelund C, Söderberg JJ, Haugen P. *Complete Genome Sequences of Seven Vibrio anguillarum Strains as Derived from PacBio Sequencing*. *Genome Biol Evol*. 2018;10:1127-1131.
24. Sun Z, Zhang W, Guo C, Yang X, Liu W, Wu Y. *Comparative genomic analysis of 45 type strains of the genus Bifidobacterium: a snapshot of its genetic diversity and evolution*. *PLoS ONE*. 2017;10:e0117912.
25. He X, Sotelo-Orozco J, Rudolph C, Lonnerdal B, Slupsky CM. *The Role of Protein and Free Amino Acids on Intake, Metabolism, and Gut Microbiome: A Comparison Between Breast-Fed and Formula-Fed Rhesus Monkey Infants*. *Front Pediatr*. 2020;7:563.
26. Kushkevych I, Cejnar J, Trembl J, Dordevic D, Kollar P, Vitezova M. *Recent Advances in Metabolic Pathways of Sulfate Reduction in Intestinal Bacteria*. *Cells*. 2020;9:698.
27. Milani C, Lugli GA, Duranti S, Turrone F, Bottacini F, Mangifesta M. *Genomic encyclopedia of type strains of the genus Bifidobacterium*. *Appl Environ Microbiol*. 2014;80:6290-6302.
28. Milani C, Lugli GA, Duranti S, Turrone F, Mancabelli L, Ferrario C. *Bifidobacteria exhibit social behavior through carbohydrate resource sharing in the gut*. *Sci Letter*. 2015;5:467-476.
29. Kim H, Jeong Y, Kang S, You HJ, Ji GE. *Co-Culture with Bifidobacterium catenulatum Improves the Growth, Gut Colonization, and Butyrate Production of Faecalibacterium prausnitzii: In Vitro and In Vivo Studies*. *Microorganisms*. 2020;8:788.
30. Kang, J. W., Tang, X., & Zivkovic, A. *A Prebiotic Supplement Increases Bifidobacteria Abundance and Microbial Acetate Synthesis Gene in Participants*. *FASEB Journal*. 2021;35.doi: 10.1096/fasebj.2021.35.S1.00442.
31. Li J, Hou Q, Zhang J, Xu H, Sun Z, Menghe B, Zhang H. *Carbohydrate Staple Food Modulates Gut Microbiota of Mongolians in China*. *Front Microbiol*. 2017;8:484.
32. Henrick BM, Rodriguez L, Lakshmikanth T, Pou C, Henckel E, Arzoomand A, et al. *Bifidobacteria-mediated immune system imprinting early in life*. *Cell*. 2021. doi:10.1016/J.CELL.2021.05.030.
33. Liu R, Yang B, Stanton C, Ross RP, Zhao J, Zhang H, Chen W. *Comparative genomics and gene-trait matching analysis of Bifidobacterium breve from Chinese children*. *Food Biosci*. 2020;36:100631.
34. Zhang JC, Guo Z, Xue ZS, Sun ZH, Zhang MH, Wang LF. *A phylo-functional core of gut microbiota in healthy young Chinese cohorts across lifestyles, geography and ethnicities*. *The ISME J*. 2015;9:1979-1990.
35. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J. Erratum: *SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler*. *BioMed Central*. 2015;4:30.
36. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. *DNA–DNA hybridization values and their relationship to whole-genome sequence similarities*. *Int J Syst Evol Microbiol*. 2007;57:81-91.
37. Chen J, Yang X, Chen J, Cen Z, Guo C, Jin T, Cui Y. *SISP: a Fast Species Identification System for Prokaryotes Based on Total Nucleotide Identity of Whole Genome Sequences*. *Infectious Diseases and Translational Medicine*. 2015;1:30-55.

38. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. *TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data*. *Mol Plant*. 2020;13: 1194-1202.
39. Torsten S. *Prokka: rapid prokaryotic genome annotation*. *Bioinformatics*. 2014;30:2068-2069.
40. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG. *Roary: rapid large-scale prokaryote pan genome analysis*. *Bioinformatics*. 2015;31:3691-3693.
41. Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. *EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates*. *Cold Spring Harb Perspect Biol*. 2009;19:327-335.
42. Zhang W, Sun Z. *Random local neighbor joining: A new method for reconstructing phylogenetic trees*. *Mol Phylogenet Evol*. 2008;47:117-128.
43. Alikhan N-F, Petty NK, Zakour NLB, Beatson SA. *BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons*. *BMC Genomics*. 2011;12:402.
44. Stothard P, Wishart DS. *Circular genome visualization and exploration using CGView*. *Bioinformatics*. 2005;21:3.
45. Mount DW. *Using the Basic Local Alignment Search Tool (BLAST)*. *Cold Spring Harb Perspect Biol*. 2007;7: *pdb.top17*.
46. Han Z, Tanner Y, Le H, Sarah E, Peizhi W, Zhenglu Y, *et al*. *dbCAN2: a meta server for automated carbohydrate-active enzyme annotation*. *Nucleic Acids Res*. 2018;46:W95-W101.

Figures

Figure 1

Heatmap of ANI (A) and TNI (B) based on the sequences of 20 genomes. The location and isolated resource of primary *B. catenulatum* isolates were annotated.

Figure 3

Comparison of the main functions of *B. catenulatum* subsp. *catenulatum* and *B. catenulatum* subsp. *kashiwanohense* Amino acid derivatives (A); Protein metabolism (B); Carbohydrate metabolism (C); Cofactors, vitamins, prosthetic groups, and pigments (D).

Figure 5

A display of the genome circle map of 16 *B. catenulatum* genomes. The figure was generated using the BRIG tool based on comparison of 16 *B. catenulatum* genomes according to percentage identity (100%, 90%, or 70%). The numbers on the rings from inside to outside, 1: JCM15439^T, 2: DSM21854(2)^T, 3: DSM21854^T, 4: APCKJ1, 5: PV20-2, 6: HGUT-01490, 7: LMG11043^T, 8: A1, 9: A3, 10: A2, 11: 1899B, 12: DSM16992^T, 13: DSM16992(2)^T, 14: JCM1164^T, 15: IMAUFB085, 16: IMAUFB087.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.pdf](#)
- [Additionalfile2.docx](#)
- [Additionalfile3.pdf](#)

- [Additionalfile4.pdf](#)
- [Additionalfile5.docx](#)
- [Additionalfile6.pdf](#)
- [SupplementaryInformation.docx](#)