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, Ministry of Education, Ministry of Agriculture and Rural Affairs,
Inner Mongolia Agricultural University

Weicheng Li

Inner Mongolia Agricultural University, Ministry of Education, Ministry of Agriculture and Rural Affairs,
Inner Mongolia Agricultural University

Caiqing Yao

Inner Mongolia Agricultural University, Ministry of Education, Ministry of Agriculture and Rural Affairs,
Inner Mongolia Agricultural University

Jie Yu

Inner Mongolia Agricultural University, Ministry of Education, Ministry of Agriculture and Rural Affairs,
Inner Mongolia Agricultural University

Heping Zhang (✉ nnd2020106@163.com)

Inner Mongolia Agricultural University, Ministry of Education, Ministry of Agriculture and Rural Affairs,
Inner Mongolia Agricultural University

Research Article

Keywords: *Bifidobacterium catenulatum*, Genomics, Carbohydrate utilization, Plant-derived glycan, Human milk oligosaccharides

Posted Date: February 3rd, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-957925/v2>

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

[Read Full License](#)

Abstract

Background

The two subspecies of *Bifidobacterium catenulatum*, *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum*, are usually from the infant and adult gut, respectively. However, the genomic analysis of their functional difference and genetic divergence has been rare. Here, 16 *B. catenulatum* strains, including 2 newly sequenced strains, were analysed through comparative genomics. A phylogenetic tree based on 785 core genes indicated that the two subspecies were significantly divergent.

Results

A phylogenetic tree based on 785 core genes indicated that the two subspecies of *B. catenulatum* were significantly separated. The comparison of genomic characteristics revealed that the two subspecies had significantly different genomic sizes ($p < 0.05$) but similar GC contents. The functional comparison revealed the most significant difference in carbohydrate utilisation. Carbohydrate-active enzymes (CAZyme) present two clustering patterns in *B. catenulatum*. The *B. catenulatum* subsp. *kashiwanohense* specially including the glycoside hydrolases 95 (GH95) and carbohydrate-binding modules 51 (CBM51) families involved in the metabolism of human milk oligosaccharides (HMO) common in infants, also, the corresponding fucosylated HMO gene clusters were detected. Meanwhile, *B. catenulatum* subsp. *catenulatum* rich in GH3 may metabolise more plant-derived glycan in the adult intestine.

Conclusions

These findings provide genomic evidence of carbohydrate utilisation bias, which may be a key cause of the genetic divergence of two *B. catenulatum* subspecies.

Introduction

Bifidobacterium is a genus of gram-positive, anaerobic microorganisms found in the human gut [1]. Some strains of *Bifidobacterium* have attracted significant attention due to their probiotic function in regulating microbiota and immune metabolism [2–3]. *Bifidobacterium catenulatum* (*B. catenulatum*) is an important member of the genus; some of its strains demonstrate favourable probiotic characteristics, such as the preclinical treatment of acute liver injury [4], in vitro inhibition of pathogenic bacteria as well as the ability to stay alive in yoghurt for a long period [5]. These potential probiotic properties suggest that *B. catenulatum* may be a candidate for probiotics in food or medicine.

Bifidobacterium has long been considered an important intestinal symbiotic bacterium co-evolving with its hosts. *Bifidobacterium* has significant 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* usually appears in the intestinal tract of adults [7–8]. Meanwhile, *B. catenulatum* has the typical characteristic of divergence in the human gut [9]. According to the latest taxonomy [10], *B. catenulatum* contains two subspecies, *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum*. *B. catenulatum* subsp. *kashiwanohense* lives specifically in the gut of infants, while *B. catenulatum* subsp. *catenulatum* is more suitable for the adult gut [7–8, 11]. Current research suggests that *B. catenulatum*'s adaptation to different hosts is partially due to the functional preference of different subspecies, such as carbohydrate metabolism [11]. However, there is no genomic evidence corresponding to the different functional preferences of the two subspecies. Therefore, it is necessary to fill the gap in the genomic knowledge of the genetic divergence and functional differentiation of the two subspecies; the additional information will be useful for supplementing the existing knowledge on the bacterium and providing scientific support for their purported health benefits.

In-species comparative genomics analysis allows for a deeper understanding of the individual characteristics between genomes [12]. However, because the *Bifidobacterium* genus is strictly anaerobic, it is difficult to culture, thus limiting the number of *B. catenulatum* genomes sequenced [13]. Recently, newly developed sequencing technologies have begun to uncover the *B. catenulatum* genomes. [14]. While there have been genomic analyses of this species, most of the genomic information of *B. catenulatum* remains unexplored.

In the current study, a total of 19 genomes of *B. catenulatum* species were analysed, including 12 *B. catenulatum* subsp. *catenulatum* and 5 *B. catenulatum* subsp. *kashiwanohense* from the Refseq database, and 2 newly sequenced (IMAUFB085 and IMAUFB087) strains. The study dissected the genetic background of and functional genomic information in *B. catenulatum* using comparative genomic approaches. This work not only provides general insights into the genomic differences between two subspecies of *B. catenulatum* but also reveal the key factors leading to their divergence.

Results

3.1 Taxonomic status of *B. catenulatum* strains

The sequence similarity and taxonomic status among the strains used in this study were confirmed by calculating the pairwise ANI (Fig. 1A) and TNI (Fig. 1B) values of all 20 genome assemblies. Strains with an ANI value of over 95% are generally considered the same species [15]. The ANI and TNI analyses produced similar clustering results, displaying distinct subspecies branches. IMAUFB085 and IMAUFB087 were grouped with most of the *B. catenulatum* subsp. *catenulatum* strains; their ANI values compared to that of *B. catenulatum* subsp. *catenulatum* JCM1194^T were 98.41% and 98.42%, and TNI values were 87.45% and 84.48%, respectively. These results confirmed the classification of IMAUFB085 and IMAUFB087 as *B. catenulatum* subsp. *catenulatum*.

ANI analysis revealed that 3 *B. catenulatum* subsp. *catenulatum* strains, 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%, respectively, less than the threshold value of 95%. Therefore, these strains were subsequently excluded. In addition, cluster analysis distinguished two subspecies. The ANI value was greater than 95% between the 2 subspecies groups, and greater than 98% within the subspecies, indicating that these strains belonged to the same species.

3.2 Comparison of general genomic features between two subspecies

The general information of the strains shows that all *B. catenulatum* subsp. *kashiwanohense* strains are derived from infants, while only two strains of *B. catenulatum* subsp. *catenulatum* are known to be infantile isolates (**Table S1**). The genomic features of 17 *B. catenulatum* strains and two novel strains, IMAUFB085 and IMAUFB087, are summarised (**Table 1**). The genomic characteristics within the *B. catenulatum* species exhibited 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. A comparison of the basic genomic characteristics of the two subspecies (**Fig. S1**) indicated that the genome size of *B. catenulatum* subsp. *kashiwanohense* (2.36 ± 0.05 Mb) was significantly larger than that of *B. catenulatum* subsp. *catenulatum* (2.09 ± 0.07 Mb) ($p=0.0021$), while there were no significant differences in GC content ($p > 0.05$). The substantial genomic differences reflected the speciation boundaries of the two subspecies, while the similarity in GC content represented a close relationship between them [16–17]. In addition, *B. catenulatum* subsp. *kashiwanohense* contained more coding genes (CDSs) than *B. catenulatum* subsp. *catenulatum* ($p=0.0046$) and there were no statistical differences in the number of tRNAs ($p > 0.05$).

The overall genomic differences between the two subspecies were further explored using the BLAST Ring Image Generator (BRIG) to graphically compare *B. catenulatum* strains with *B. catenulatum* subsp. *kashiwanohense* strain JCM15439^T as the reference (**Fig. S2**). Overall, most of the sequences in JCM15439^T were also in all other strains, and the genomes were more than 90% identical. However, two large genomic gaps (GGs) existed separately in the two newly sequenced strains, IMAUFB085 and IMAUFB087, which had less than 70% of the matched degree compared to JCM15439^T. In general, the GG sequences represent hypothetical CDSs, genomic islands or prophages [18]. These data indicate that these two strains have many unknown functions to be explored.

3.3 Phylogenetic divergence of two subspecies of *B. catenulatum*

Classification of species and establishment of intraspecific relationships are frequently based on phylogenetic analysis. This study reconstructed the phylogenetic structure of *B. catenulatum* species while specifying *B. pseudocatenulatum* JCM1200^T as an outgroup. An NJtree based on 785 core genes was constructed and visualised with the bootstrap support of 1000 replications (**Fig. 2A**). This phylogenetic tree confirmed the subspecies divergence of *B. catenulatum*. For example, 16 *B.*

catenulatum strains were clearly divided into two subgroups, *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum* subgroup, indicating the diacritical differences between the two subspecies at the gene level. Interestingly, the annotation of the source of the isolates suggested a significant association between the *B. catenulatum* strains and their isolated source. Infant isolates, including all *B. catenulatum* subsp. *kashiwanohense* strains and 2 *B. catenulatum* subsp. *catenulatum* strains, exhibited intraspecific genetic similarity, while the rest were adult isolates in another cluster, indicating close phylogenetic relationships. These data suggest that the trend of divergence of the *B. catenulatum* strains may be dependent on their hosts. *B. catenulatum* may adapt its functions to infant and adult intestines respectively, thus gradually differentiating into different subspecies.

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 16 *B. catenulatum* strains; a total of 4608 pan genes were searched. The above phenomenon is that the two subspecies of *B. catenulatum* shared 998 core genes (21.66%) (Fig. 2B), indicating a common ancestor of them. The genetic distribution of *B. catenulatum* showed that there were unique core gene sets in the 2 subspecies, with 87 core genes in *B. catenulatum* subsp. *kashiwanohense* and 63 in *B. catenulatum* subsp. *catenulatum* (Table S2). The unique core gene sets can provide an internal basis for the differentiation of *Bifidobacterium* species [19]. Additionally, there were different numbers of strain-specific genes in the *B. catenulatum* subspecies; their numbers ranged from 20 to 578, suggesting the potential genetic diversity among *B. catenulatum* species.

Subsequently, the pan-core gene curves for the genomes of the *B. catenulatum* species were established (Fig. S3A). With the augmentation by the new genomes, the number of pan genes increased, indicating the existence of an open pan-genome within the species of *B. catenulatum*. In contrast, the number of core genes was not expected to be significantly reduced by the addition of the new genomes since the exponential trendline reached the number of 1000. Notably, *B. catenulatum* subsp. *catenulatum* has a fairly open pan-core genome (Fig. S3B), while *B. catenulatum* subsp. *kashiwanohense*'s genome tends to be closed (Fig. S3C). These results indicate that *B. catenulatum* subsp. *catenulatum* may have a stronger ability to adapt to the environment, while *B. catenulatum* subsp. *kashiwanohense* exists in a more specific and conserved habitat. This finding is consistent with the information on the sources of the *B. catenulatum* strains (Table S1); *B. catenulatum* subsp. *kashiwanohense* only exists in the intestinal tract of infants and is relatively rare [11], while *B. catenulatum* subsp. *catenulatum* can exist in the intestinal tract of both infants and adults and has a large number. Therefore, more novel strains of *B. catenulatum* subsp. *catenulatum* may be discovered than *B. catenulatum* subsp. *kashiwanohense* in the future.

Comparison of the main functions between two subspecies

The above results have uncovered the genetic differences between the two subspecies at the general genomics level, which are usually associated with functional differentiation [19]. Therefore, it is necessary to conduct further functional genomic comparisons between the two subspecies of *B. catenulatum*. Their functional genomic differences were obtained by annotating all the strains through

the RAST website. The functional annotations of 16 *B. catenulatum* genomes were examined in 23 functional categories (**Table S3**). These results suggest that the function of amino acid derivatives (21.06%) is the most highly represented category within *B. catenulatum* followed by protein metabolism (21.00%), carbohydrate metabolism (15.73%) and cofactors, vitamins, prosthetic groups and pigments (9.76%). These data indicate that a strong ability to utilise substrates by *B. catenulatum*. The comparison of the main functional differences between the two subspecies showed the subspecies differ significantly in their metabolism of carbohydrates ($p=0.01$), amino acids ($p=0.011$) and proteins ($p=0.012$) (Fig. 3A, 3B, 3C). In contrast, there was no statistical difference in the cofactors, vitamins, prosthetic groups and pigments-related categories ($p>0.05$) (Fig. 3D). Because of the most significant difference between the two subspecies was in carbohydrate function, the *B. catenulatum* genes involved in carbohydrate utilisation were analysed.

3.6 Different carbohydrate utilisation patterns in two subspecies of *B. catenulatum*

The carbohydrate utilisation abilities of *B. catenulatum* subspecies at the genomic level were compared by analysing the functional genes of carbohydrate active enzymes of 16 *B. catenulatum* strains. As shown in Fig. 4A, 16 *B. catenulatum* strains were distributed in all six carbohydrate-active enzyme families, indicating that they had rich carbohydrate functions. Notably, the clustering results of CAZyme were roughly consistent with those of the phylogenetic trees in that the two subspecies were distinct. This finding not only suggests that the two subspecies have different metabolic patterns in terms of carbohydrate utilisation, but also indicates that CAZyme-related genes are closely associated with the divergence of *B. catenulatum* subspecies.

Among the identified GH families in *B. catenulatum* species, the most dominant ones were GH3, GH13 and GH43; meanwhile, GT2 and GT4 were the main carbohydrate enzyme families within *B. catenulatum* species. Comparing the main carbohydrate hydrolase families in the subspecies revealed the number of GH3 family members was significantly higher in *B. catenulatum* subsp. *catenulatum* than those in *B. catenulatum* subsp. *kashiwanohense* ($p=0.0038$, Fig. 4B). GH3 is mainly involved in the metabolism of plant-derived glycan common in the adult diet, such as β -glucosidase and xylosidase. [20]. However, there was no statistically significant difference in the function of GH13, GH43, GT2 and GT4 between the two subspecies ($p>0.05$) (Fig. 4C, 4D, 4E, 4F). Therefore, GH3 may be a key factor in the divergence of carbohydrate functional genes between the two subspecies of *B. catenulatum*.

Analysis of the specific CAZymes of *B. catenulatum* subsp. *kashiwanohense* revealed five families that only existed in the subspecies, including GH18, CBM5, GH95, CBM51 and CBM66 (Fig. 4G). The CBM family is primarily responsible for banding carbohydrates. In addition, the GH18 family often combines with CBM5 to participate in the function of chitinase, and CBM66 mainly assists in the degradation of fructose [21]. In particular, the GH95 family is specifically involved in the production of α -L-fucosidase, the most abundant substance in HMO and closely related to the function of infant-specific species [22]. Additionally, the CBM51 family helps GH95 enzymes pick up fucose to metabolise HMO [23]. These

CAZyme families CBM51 and GH95 may be conducive to the colonisation of *B. catenulatum* subsp. *kashiwanohense* in the intestines of infants, especially the utilization of HMO, in contrast to the abundance of plant-derived glycan of *B. catenulatum* subsp. *catenulatum*, further suggesting the bias of the two subspecies in carbohydrate utilisation. In addition, GH29 enzymes often interact with GH95 enzymes to utilise HMO [24], and the study found that GH29 is only in *B. catenulatum* subsp. *kashiwanohense* except for PV20-2.

Identification of HMO gene clusters in *B. catenulatum* genomes

Considering the specific utilisation of fucosylated HMO (FHMO) by GH29 and GH95 enzymes, the FHMO gene cluster in *B. catenulatum* were subsequently examined. Two *Bifidobacterium* strains (*B. longum* subsp. *longum* SC596 and *B. pseudocatenulatum* JCM1200^T) with typically structural FHMO gene clusters were selected as the reference [25] for the search for the homologous FHMO gene cluster in all of the *B. catenulatum* genomes. The homologous alignment showed an integrated FHMO gene cluster in all *B. catenulatum* subsp. *kashiwanohense* genomes but not in *B. catenulatum* subsp. *catenulatum* (Fig. 5), further confirming the unique ability to utilise HMO by *B. catenulatum* subsp. *kashiwanohense*. In the study, two different structures of FHMO gene clusters, named type I and type II, were found in *B. catenulatum* subsp. *kashiwanohense* (Table S4). Type I shared 89.6% homology with *B. longum* subsp. *longum* SC596. The size of type I was about 13.0 kb, including 11 open reading frames (ORF), manifested as GH95, GH29, *fucU*, dihydrodipicolinate synthase family protein (DHP), amidohydrolase family protein, SDR family oxidoreductase, fuconate dehydratase, three ABC transporters and *lacI*. Meanwhile, type II shared 97.8% homology with *B. pseudocatenulatum* JCM1200^T; it was only found in PV20-2 and lacked GH29 and *fucU* genes, consistent with the results of CAZyme.

Notably, the GC content of the FHMO gene clusters in *B. catenulatum* subsp. *kashiwanohense* was significantly lower than the entire subspecies (Fig. S4), suggesting that its FHMO gene clusters might be obtained through horizontal gene transfer (HGT) [26]. The identification of the FHMO gene clusters in *B. catenulatum* subsp. *kashiwanohense* further confirmed its advantage of HMO utilisation, thus providing genomic evidence for its adaptability in the infant intestine.

Discussion

As a typical intestinal symbiotic bacteria, *Bifidobacterium* has experienced a long and extensive evolutionary process in human hosts [1]. For example, *B. catenulatum* has evolved into two subspecies, *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum*. Previous studies have revealed that *B. catenulatum* subsp. *kashiwanohense* and *B. catenulatum* subsp. *catenulatum* have an close phylogenetic relationship [39]. Here, phylogenetic reconstruction has revealed genomic differences between the two subspecies. The genome size and the number of the CDSs of *B. catenulatum* subsp. *catenulatum* were significantly lower than that of *B. catenulatum* subsp. *kashiwanohense*. Also, both subspecies have a unique core gene set, such results represent a marker of genetic divergence [17]. In

addition, there was obvious host differentiation in *Bifidobacterium*. For example, *B. catenulatum* subsp. *kashiwanohense*, *B. longum* subsp. *infantis* and *B. breve* are more common in infants while *B. catenulatum* subsp. *catenulatum*, *B. adolescentis* are more present in adult intestines [7–8, 11, 27]. Thus, the possible association between subspecies divergence and the host was further explored through functional genomic comparisons to explain the divergence of *B. catenulatum* at the genomic level.

Bifidobacterium is a genus of saccharolytic microorganisms whose ability to utilise indigestible carbohydrates is essential for their establishment in the gastrointestinal tract [28]. In this study, functional genomics revealed significant differences in the carbohydrates consumed by the subspecies of *B. catenulatum*. Notably, the CAZyme cluster results are consistent with the phylogenetic tree analysis, suggesting that the functional differences in carbohydrates may be related to the genetic divergence of *B. catenulatum*. This study 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* and is involved in the degradation of plant polysaccharides [29]. The results here indicate that GH3 is also a key factor for the divergence of *B. catenulatum* in carbohydrate function. Studies have shown that the gut environment in adults is more complex than in infants because adults typically consume more difficult-to-digest carbon sources, such as plant-based dietary fibre [7–8]. Kim et al. found that *B. catenulatum* strains can degrade fructooligosaccharides (FOS) in nutritionally restricted environments [30]. Previous studies have shown that a low-fibre diet in adults can cause a significant increase in the abundance of *B. catenulatum* [31]. Here, the results demonstrate that *B. catenulatum* subsp. *catenulatum* has more GH3 that utilises plant-derived glycans; therefore, the subspecies is conducive to the decomposition of difficult-to-use plant-derived glycans in the adult gut.

On the other hand, infants, especially those who are breastfed, have many HMOs in their intestines. HMO is a prebiotic unique to breast milk and is especially enriched in human breast milk [32]. The ability of infant-specific *Bifidobacterium* to metabolise HMO has been recognised as a specific marker of its adaptive colonisation and beneficial for strengthening the immune system in infants [33]. For *B. catenulatum* subsp. *kashiwanohense*, which is characterised by infant adaptation [11], its two specific CAZymes, namely GH95 and CBM51, which are notable. GH95 mainly utilises fucosyllactose, a major component of HMO [34]. On the other hand, CBM51 is beneficial to GH95 and helps it pick up FHMO [23]. Thus, this study suggests that GH95 and CBM51 act synergistically in the utilisation of FHMO by *B. catenulatum* subsp. *kashiwanohense*. Particularly, GH29 is often identified with GH95 as the family of metabolic HMO [35]. In *B. catenulatum* subsp. *kashiwanohense*, all strains except PV20-2 contain GH29. Therefore, the study suggests that these three families (GH29, GH95 and CBM51) play an important role in the colonisation of *B. catenulatum* subsp. *kashiwanohense* in the infant intestine.

Based on the findings related to the HMO-related families, this study further confirms the existence of relatively conserved HMO gene clusters in *B. catenulatum* subsp. *kashiwanohense* while not in *B. catenulatum* subsp. *catenulatum*. These HMO gene clusters are highly homologous to those in other typical infantile adapted *Bifidobacterium* that are connected to the GH95 and GH29 families. Only the

PV20-2 strain lacks GH29 and *fucU*, while the genome of PV20-2 shares high homology with the HMO gene cluster of *B. pseudocatenulatum* JCM1200^T, which can grow in purified FHMO [36], the lack of these two genes appears to have little effect on the overall ability to use FHMO. Given that the reference genomes in HMO gene clusters are all from infants, their clusters have been demonstrated to be conducive to their utilisation of HMO [28, 36]. This study suggests that *B. catenulatum* subsp. *kashiwanohense* may have a similar utilisation mechanism of FHMO for adaptive survival in the infant intestine [28, 34–35]. James et al. [11] has confirmed that *B. catenulatum* subsp. *kashiwanohense* APCKJ1 expresses HMO genes related to the consumption of FHMO, its sole carbohydrate source. Given the high similarity of the HMO gene clusters in *B. catenulatum* subsp. *kashiwanohense*, this characteristic may be extrapolated to other strains; however, this hypothesis requires further verification. Notably, *B. catenulatum* subsp. *catenulatum* 1899B and IMAUFB085 belong to infant isolates, but no HMO genes were found in them, further confirming that possession of HMO genes is a genetic trait of *B. catenulatum* subsp. *kashiwanohense*.

The evolution of *Bifidobacterium* involves a large number of gene acquisition events [37]. Garrido et al. [28] propose that the HMO gene clusters have transferred from *B. longum* subsp. *infantis* to *B. longum* subsp. *longum* during evolution. Notably, the HMO gene cluster in *B. catenulatum* subsp. *kashiwanohense* in this study showed a significant decrease in GC content, likely caused by HGT events, that were important in the genomic evolution of species [27]. At present, these types of HMO gene clusters have been found in typical infant-derived strains, such as *B. breve*, *B. longum* and *B. pseudocatenulatum* species, and they have high homology with each other [24, 28, 36]. This study proposes that *B. catenulatum* subsp. *kashiwanohense* acquired HMO gene clusters through HGT from other proximal species (such as *B. longum*), the acquisition of HGT contributed to the specific function of genome divergence and HMO utilisation.

Although the two subspecies of *B. catenulatum* are closely phylogenetically related and share a common ancestor [39], previous studies have confirmed that they showed different tendencies adapted in infants and adult intestines [7, 8, 11]. Taken together, given that the carbohydrate genetic pattern of the two subspecies was consistent with the phylogenetic relationship, 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. Our results are similar to the divergence of *B. longum*, for the *infantis* subspecies of it has specific genes related to the metabolism of HMO and is more suitable for breast-feeding infant intestines, while the *longum* subspecies is present in both infant and adult hosts but has more genes for the utilization of plant-derived sugars and is more suitable for adult diets [28]. The example of this divergence of species in different hosts seems to suggest a potential pattern of genetic divergence of *Bifidobacterium*, in which infant and adult wealthy species have more HMO genes and plant-derived glycan genes respectively in the human gut in order to adapt to their respective hosts.

Conclusions

In summary, this study proposes that the *B. catenulatum* species evolved to retain the competitive carbohydrate function genes to adapt to the respective intestinal environment in infants and adults, driving the emergence of two subspecies. This study has provided genomic evidence for the potential host adaptation phenomenon of *B. catenulatum* in infant and adult intestines. However, the number of *B. catenulatum* strains is limited; more strains will need to be sequenced in the future to dissect further the mechanism underlying their genetic divergence.

Methods

Bacterial strains, DNA extraction and publicly available assemblies

The genomes of two strains of *B. catenulatum*, IMAUFB085 and IMAUFB087, were provided by the Lactic Acid Bacteria Collection Center (LABCC). Moreover, IMAUFB085 was isolated from infant faeces and IMAUFB087 from adult faeces in Tibet, China [38].

The two strains were cultured under anaerobic conditions in the Man Rogosa and Sharpe (MRS) broth with L-cysteine hydrochloride at 37°C for 24 h. DNA extraction was performed using the TIANamp Bacteria DNA Kit. Genomic DNA was quantified using a TBS-380 fluorometer. High-quality DNA samples were obtained to construct fragment libraries.

In addition, publicly available assemblies of all *B. catenulatum* strains were obtained from the National Coalition Building Institute (NCBI, <https://www.ncbi.nlm.nih.gov/>) on 4 February 2021, including that of type strains, namely *B. catenulatum* subsp. *catenulatum* (JCM1194^T) and *B. catenulatum* subsp. *kashiwanohense* (JCM15439^T) (**Table S1**). Additionally, the *B. pseudocatenulatum* strain (JCM1200^T) in the *B. adolescentis* group, most closely related to *B. catenulatum* according to the phylogenetic relationship of *Bifidobacterium* genus in a previous study [39], were downloaded to infer phylogenetic relationships across species within it.

Genome sequencing and assembly

Genome sequencing was performed using the Illumina HiSeq platform to generate 150-bp paired-end reads for each sample. Then, the sequences were filtered through the Illumina HiSeq system. The high-quality sequences were assembled using SOAPdenovo2 [40] on a 64-bit Linux system. High-quality data corresponding to a sequencing depth of about 387-fold, was generated for each strain. In addition, local inner gaps were filled, and single-base errors were corrected using 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. The functional gene information of this bacterium was obtained by performing the gene prediction and preliminary annotation of all *B. catenulatum* genomes through the Rapid Annotation

using Subsystems Technology (RAST) server (<https://rast.nmpdr.org/rast.cgi>). In addition, tRNA genes were identified using tRNAscan-SE (<http://trna.ucsc.edu/tRNAscan-SE/>).

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

The genetic relatedness between the two *B. catenulatum* subspecies was evaluated, and the taxonomic status of the strains in this study was confirmed by analysing the ANI and TNI values of all the strains. *B. pseudocatenulatum* JCM1200^T, the type strain most phylogenetically related to *B. catenulatum* [39], was included in the comparison. All pairwise ANI values were calculated according to the method proposed by Goris et al. [41]. TNI values were calculated according to the method proposed by Chen et al. [42]. Finally, the clustering heat map was drawn using TBtools [43].

Construction of pan-core genome and strain-specific genes

The annotated genomes of *B. catenulatum* were obtained using Prokka v1.12 [44] and processed using Roary [45] to identify the pan genes, core genes and specific genes using the default parameters. The intersection groups, representing the unique sets of genes identified only between the intersected genomes, were visualised using the UpSet diagram in TBtools [43].

Phylogenetic analysis

The core gene alignment from Roary was used in TreeBeST [46] with 1,000 bootstrap iterations to build a phylogenetic NJtree through Neighbor-Joining (NJ) [47]. The phylogenetic trees were then visualised and annotated using iTOL (<https://itol.embl.de/>).

BRIG (BLAST Ring Image Generator)

BRIG v0.95 [48] was adapted to compare the genomes of *B. catenulatum* strains based on a JAVA language environment. The image of the circular genomes was also generated through BRIG.

CAZyme (Carbohydrate-active enzymes) online annotation

Given the carbohydrate metabolic activity of *Bifidobacterium*, CAZyme online annotation of *B. catenulatum* strains were accomplished using three annotation tools, HMMER, DIAMOND and Hotpep searches [49]. 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). The identification of CAZymes across the *B. catenulatum* genomes was carried out using the dbCAN2 meta server (<http://bcb.unl.edu/dbCAN2/>). According to the annotation results, the detailed information on the active carbohydrate enzyme family was checked on the CAZyme website (<http://www.aczy.org/>).

Detection of the HMO gene clusters

Taking *B. longum* subsp. *longum* SC596 and *B. pseudocatenulatum* JCM1200^T as the reference, which possess typical HMO gene clusters. In addition, the genome of SC596 was obtained from IMG database

[50]. The corresponding protein-encoding sequences were extracted from the strains and compared using NCBI BLASTp with default parameters. The recognised HMO gene clusters were visualised using the *genoplotR* package.

Statistical analysis

The data were presented as means $\pm$ SEM. The Wilcoxon signed-rank test was used to verify the significance of the difference between the groups, and visualisation was performed using the *ggpubr* packages in R (4.0.3). Lastly, significance was set at a *p*-value of less than 0.05.

Data availability

The assembly and Sequence Read Archive (SRA) data of the two newly isolated sequences in this work were submitted as a Whole Genome project (BioProject No. PRJNA751426) at GenBank under the accessions JAIEWL000000000 (IMAUFB087) and JAIEWM000000000 (IMAUFB085) (available at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA751426>). The phylogenetic trees and alignment files in this study were submitted to the TreeBASE web (Accession No. 28852) (available at <http://purl.org/phylo/treebase/phylows/study/TB2:S28852>).

Abbreviations

LABCC: Lactic acid bacteria collection center; MRS: Man Rogosa and Sharpe; ANI: Average nucleotide identity; TNI: Total nucleotide identity; CDSs: coding sequences; GGs: genome gaps; NJ: Neighbor-Joining; RAST: Rapid Annotation using Subsystems Technology; BRIG: BLAST Ring Image Generator; HMOs: Human Milk Oligosaccharides; FHMO: fucosylatedHMO; *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; HGT: horizontal gene transfer; FOS: fructooligosaccharides.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The assembly and Sequence Read Archive (SRA) data of the two newly isolated sequences in this work were submitted as a Whole Genome project (BioProject No. PRJNA751426) at GenBank under the accessions JAIEWL000000000 (IMAUFB087) and JAIEWM000000000 (IMAUFB085) (available at

<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA751426>). The phylogenetic trees and alignment files in this study were submitted to the TreeBASE web (Accession No. 28852) (available at <http://purl.org/phylo/treebase/phyloids/study/TB2:S28852>).

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;130:1357-67.
3. Luisa M, Michele C, Sonia B, Matteo S, Giacomo B, Susanna V, et al. 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 *Bifidobacteriumcatenulatum*, *Enterococcus caccae* and *Ruminococcus gauvreauii*. *Anaerobe* 2016;42:130-41.
5. Onyibe JE, Oluwole OB, Ogunbanwo ST, Sanni AI. Antibiotic Susceptibility Profile and Survival of *Bifidobacterium adolescentis* and *Bifidobacteriumcatenulatum* of Human and Avian Origin in Stored Yoghurt. *Niger Food J* 2013;31:73-83.
6. Duranti S, Longhi G, Ventura M, van Sinderen D, Turrone F. 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;75:103-18.

8. Turrone F, Milani C, Duranti S, Mancabelli L, Mangifesta M, Viappiani A, et al. Deciphering bifidobacterial-mediated metabolic interactions and their impact on gut microbiota by a multi-omics approach. *The ISME journal* 2016, 10(7) : 1656-68.
9. Ismail IH, Boyle RJ, Licciardi PV, Oppedisano F, Lahtinen S, Robins-Browne RM, et al. Early gut colonization by *Bifidobacterium breve* and *catenulatum* differentially modulates eczema risk in children at high risk of developing allergic disease. *Pediatr Allergy Immunol* 2016;27:838-46.
10. 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.
11. 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;9:1589-96.
12. 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; 19:1173-76.
13. 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.
14. 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.
15. 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: 1281-86.
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;78:5258-69.
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. Holm KO, Bækkel C, Söderberg JJ, Haugen P. Complete Genome Sequences of Seven *Vibrio anguillarum* Strains as Derived from PacBio Sequencing. *Genome Biol Evol* 2018;10:1127-31.
19. Abdelhamid AG, ElDoudgoud NK. Comparative genomics of the gut commensal *Bifidobacterium bifidum* reveals adaptation to carbohydrate utilization. *Biochem Bioph Res Co* 2021;547:155-61.
20. Deboy RT, Mongodin EF, Fouts DE, Tailford LE, Khouri H, Emerson JB. Insights into plant cell wall degradation from the genome sequence of the soil bacterium. *Cellvibrio japonicus* 2008;15:5455-63.
21. Cuskin F, Flint JE, Gloster TM, Morland C, Basle A, Henrissat B, et al. How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc Natl Acad Sci U S A* 2012;109:20889-94.
22. 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;104:5619-31.

23. 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;283:12604-13.
24. Liu R, Yang B, Stanton C, Ross RP, Zhao J, Zhang H, et al. Comparative genomics and gene-trait matching analysis of *Bifidobacterium breve* from Chinese children. *Food Biosci* 2020;36:100631.
25. Sela DA, Garrido D, Lerno L, Wu S, Tan K, Eom H, et al. *Bifidobacterium longum* subsp *infantis* ATCC 15697 alpha-Fucosidases Are Active on Fucosylated Human Milk Oligosaccharides. *App Env Microbiol* 2011;78(3):795-803.
26. Humbert JF, Barbe V, Latifi A, Gugger M, Calteau A, Coursin T, et al. A Tribute to Disorder in the Genome of the Bloom-Forming Freshwater Cyanobacterium *Microcystis aeruginosa*. *PLoS ONE* 2013;8:e70747.
27. Garrido D, Ruiz-Moyano S, Kirmiz N, Davis JC, Totten SM, Lemay D, et al. A novel gene cluster allows preferential utilization of fucosylated milk oligosaccharides in *Bifidobacterium longum* subsp *longum* SC596. *Sci Rep* 2016;6:35045.
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;5:467-76.
29. 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;80:6383-94.
30. 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.
31. Kang JW, 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.
32. Li J, Hou Q, Zhang J, Xu H, Sun Z, Menghe B, et al. Carbohydrate Staple Food Modulates Gut Microbiota of Mongolians in China. *Front Microbiol* 2017;8:484.
33. Henrick BM, Rodriguez L, Lakshmikanth T, Pou C, Henckel E, Arzoomand A, et al. *Bifidobacteria*-mediated immune system imprinting early in life. *Cell* 2021;184:3884-98.
34. Becerra JE, Yebra MJ, Monedero V. An L-Fucose Operon in the Probiotic *Lactobacillus rhamnosus* GG Is Involved in Adaptation to Gastrointestinal Conditions. *Appl Environ Microbiol* 2015;81:3880-88.
35. Liu R, Yang B, Stanton C, Ross RP, Zhao J, Zhang H, et al. Comparative genomics and gene-trait matching analysis of *Bifidobacterium breve* from Chinese children. *Food Biosci* 2020;36:100631.
36. Shani G, Hoeflinger JL, Heiss BE, Masarweh CF, Larke JA, Jensen NM, et al. Fucosylated human milk oligosaccharide foraging within the species *Bifidobacterium pseudocatenulatum* is driven by glycosyl hydrolase content and specificity. *Appl Environ Microbiol* 2021;AEM0170721.
37. 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

38. 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;9:1979-90.
39. 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.
40. 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.
41. 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.
42. Chen J, Yang X, Chen J, Cen Z, Guo C, Jin T, et al. 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.
43. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Mol Plant* 2020;13: 1194-1202.
44. Torsten S. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068-69.
45. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 2015;31:3691-93.
46. 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;19:327-35.
47. Zhang W, Sun Z. Random local neighbor joining: A new method for reconstructing phylogenetic trees. *Mol Phylogenet Evol* 2008;47:117-28.
48. Alikhan N-F, Petty NK, Zakour NLB, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 2011;12:402.
49. 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.
50. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, et al. The integrated microbial genomes (IMG) system. *Nucleic Acids Res* 2006;34:D344–D348.

Tables

Table1 General genomic features of *B. catenulatum* genomes

Collectionstrain	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

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 2

Phylogenetic tree of two subspecies of *B. catenulatum*

Phylogenetic NJtree of *B. catenulatum* species and taken *B. pseudocatenulatum* JCM1200^T as the outgroup. Bootstrap was set as 1000. All the *B. catenulatum* strains were annotated to isolate location and source. The scale bars represent 0.01 substitutions per site (A). UpSet diagram showing shared and unique core genes distribution among *B. catenulatum* strains. The horizontal bars represent the total number of genes identified of individual strains. The vertical bars or intersections represent the number of genes that were regulated by one or more strains. The orange dots represent unique genes and the yellow dots represent core genes. The green items represent information about *B. catenulatum* subsp. *catenulatum*, and the blue items represent information about *B. catenulatum* subsp. *kashiwanohense*. Groups with fewer than 10 genes were filtered (B).

Figure 3

Comparison of the main functions between *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 4

Prediction of CAZymes in 16 *B. catenulatum* strains.

The Heatmap of CAZymes in *B. catenulatum*. The isolated source of strains was annotated (A). The significance analysis of the key CAZyme families between two subspecies of *B. catenulatum* including GH3 (B), GH13 (C), GH43 (D), GT2 (E), and GT4 (F). Specific CAZymes in *B. catenulatum* subsp. *kashiwanohense* (G).

Figure 5

HMO gene clusters in *B. catenulatum* subsp. *kashiwanohense* and two reference clusters in *Bifidobacterium*.

Arrows represent genes, and numbers on top of each gene indicate the locus tag number in the respective genome. Numbers inside the arrows indicate percent identity between corresponding genes and homologs relative to reference. numbers outside on the left indicate percent identity of full clusters relative to reference. SBP: Solute Binding Protein; cABC: carbohydrate ABC transporter; sABC: sugar ABC transporter; SDR: SDR family oxidoreductase; DHP: dihydrodipicolinate synthase family protein; *fucU*: L-fucose mutarotase; *fucd*: fuconate dehydratase.

Supplementary Files

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

- [FigureS1.pdf](#)
- [FigureS2.pdf](#)
- [FigureS3.pdf](#)
- [FigureS4.pdf](#)
- [TableS1.docx](#)
- [TableS2.docx](#)
- [TableS3.docx](#)
- [TableS4.docx](#)