Bifidobacterium longum subsp. *nexus* nov., a novel subspecies isolated from infant stool

Dena Ennis[1], Sergio Andreu-Sánchez[2], Yarden Levin[1], Omer Goldberger[1], Cyrus A. Mallon[2], Yuvashankar Kavanal Jayaprakash[2,3], Yan Shao[4], Shijie Zhao[6,7], Trishla Sinha[2], Jingyuan Fu[2], Rinse K. Weersma[5], Eric Alm[6,7,8], Orna Amster-Choder[1], Trevor D. Lawley[4], Hermie J.M. Harmsen[3], Douwe van Sinderen[9], Alexandra Zhernakova[2], Moran Yassour[1,10*]

Affiliations:

1. Microbiology & Molecular Genetics Department, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel
2. Department of Genetics, University of Groningen and University Medical Center Groningen, The Netherlands
3. Department of Microbiology, University of Groningen and University Medical Center Groningen, The Netherlands
4. Host–Microbiota Interactions Laboratory, Wellcome Sanger Institute, Hinxton, UK
5. Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, The Netherlands
6. Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
7. Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA, USA
8. The Broad Institute of MIT and Harvard, Cambridge, MA, USA
9. APC Microbiome Institute and School of Microbiology, University College Cork, Cork T12 YT20, Ireland
10. The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel
    *corresponding author: moranta@mail.huji.ac.il

Keywords: Bifidobacterium longum, human gut, infant, microbiome.

Abstract

Bifidobacterium species are well-established members of the human gut microbiome, particularly prominent during infancy, contributing to host health. Within this genus, *Bifidobacterium longum* (*BL.*) is a widespread species found in both infant and adult guts, known for its complexity and functional diversity among its known subspecies: *BL. longum*, *BL. infantis* and *BL. suis*. Here, using genomic and phylogenetic tools we propose a novel subspecies within the *BL.* species, *Bifidobacterium longum* subsp. *nexus* subspecies novel. We analyzed 435 *BL.* genomes using a polyphasic taxonomic approach comprising average nucleotide identity (ANI), digital DNA–DNA hybridization (dDDH), and pangenome analysis. We identified nine *BL.* strains, isolated from human infants and adults stool samples, as members of a distinct lineage within the *BL.* species. The type strain, LL6991, was isolated from the stool of a two-week-old Dutch infant in the Lifelines NEXT birth cohort. Phenotypically, *BL. nexus* exhibits a distinct morphological pattern, predominantly forming rod-shaped cells, often in chains with visible septa, contrasting with the Y-shaped morphology commonly observed for other *BL.* subspecies. Furthermore, *BL. nexus* demonstrates unique metabolic capabilities, including efficient utilization of fructose, and starch, carbohydrates not metabolized well by other tested *BL.* subspecies. This ability may be attributed to specific genes, such as a gene predicted to encode an extracellular amylopullulanase. This characterization expands the known diversity within the *BL.* species and provides insights into *BL. nexus*'s unique adaptations and potential ecological roles within the human gut, especially in infants. Based on the consistent results from genotypic, phylogenetic, and phenotypic analyses, a novel subspecies with the name *Bifidobacterium longum* subsp. *nexus*, with type strain LL6991, is proposed.

Introduction

Bifidobacterium species are well-established members of the human gut microbiome, particularly prominent during infancy, where they contribute to various aspects of host health, including metabolic processes, immune system development, and providing resistance against pathogens[1–6]. Within this genus, *Bifidobacterium longum* (*BL.*) is recognized as a widespread species commonly found in both the infant and adult gut[7–10]. Growing research continues to reveal the complexity and functional diversity within this species in the human gut, where three main subspecies have been reported in several studies: *B. longum* subsp. *longum, B. longum* subsp. *infantis* and *B. longum* subsp. *suis*[11–13]. The original *BL. suis* taxon has recently been divided into three subclusters; *BL. suis*, *B. longum* subsp. *iuvenis*[14], a subspecies originally identified in Bangladesh infants as transitional *BL.*[15], and *B. longum*

subsp. *suillum*, a *BL. suis* subspecies which is urease negative[16]. These subspecies are increasingly understood to possess distinct functional capabilities. For instance, *BL. infantis* is notably adapted for the utilization of human milk oligosaccharides (HMOs)[17], and therefore more prevalent in the infant gut, while *BL. longum* is associated with the metabolism of plant-derived oligosaccharides[18].

The *Bifidobacterium* genus is famously named for the characteristic bifid or Y-shape exhibited by some strains when observed microscopically[19]. Historical accounts trace the identification of these Y-shaped bacteria back to 1899, when French pediatrician Henri Tissier isolated a bacterium with a Y-shaped morphology from the intestinal microbiota of breast-fed infants[20]. He named it "bifidus" recognizing its distinctive shape, which is now a defining feature of the *Bifidobacterium* genus. This distinctive shape has been visualized using various microscopy techniques, such as phase contrast microscopy and scanning electron microscopy (SEM)[21,22]. Although this morphology has long been considered a hallmark of *Bifidobacterium*, studies have shown that their cell shape is pleomorphic, with high variability influenced by environmental and growth conditions[22,23].

Defining bacterial species and subspecies accurately is fundamental to understanding microbial diversity and function. Historically, classification relied heavily on phenotypic characteristics and core gene sequencing, such as the 16S rRNA gene[24,25]. However, these methods often lack sufficient resolution to delineate closely related taxa such as subspecies. For example, 16S rRNA gene can have a slow evolutionary rate so that even full-length sequences may not provide adequate resolution to distinguish between closely related species or subspecies[26]. Modern prokaryotic taxonomy increasingly employs a polyphasic approach that integrates traditional methods with advanced genome-based analyses[25,27]. Advancements in whole-genome sequencing (WGS) have facilitated the adoption of genomic metrics such as average nucleotide identity (ANI) and digital DNA–DNA hybridization (dDDH), offering robust, quantitative measures of overall genomic relatedness[28]. These genome-based approaches have become standard tools in microbial taxonomy, providing higher resolution for species and subspecies delineation compared to traditional methods. Commonly accepted thresholds for species delineation using these methods are 95% for ANI and 70% for dDDH, which have shown a strong correlation with one another. Subspecies are typically distinguished by distinct clustering patterns in these analyses, with dDDH values generally falling below 79–80%. By leveraging WGS data, researchers can gain a more accurate view of microbial diversity and delineate finer taxonomic classifications, including subspecies.

The use of modern taxonomic approaches has significantly refined the classification of complex microbial groups. For example, a recent study combining metagenomic and genomic data revealed that *Segatella copri* is not a monotypic species but consists of four distinct clades, each associated with functional diversity[29]. Further, by integrating targeted isolation with large-scale metagenomic analysis, researchers identified 13 distinct *S. copri*-related species, expanding the known diversity beyond the initial four clades[30]. In a different study of *Escherichia coli*, long-read sequencing enabled the application of ANI and dDDH as robust alternative typing methods, potentially offering higher taxonomic resolution[31]. In another study conducted in Bangladesh, analysis of *BL.* using metagenome-derived marker genes and ANI led to the discovery of a novel *BL.* subspecies known as transitional *BL.*[15]. A recent large-scale phylogenomic analysis of the *BL.* species identified four clades, including a novel subspecies found exclusively in infants from industrialised countries[32].

Given the known diversity and functional specialization within the *BL.* species, and the increasing power of genomic methods to resolve fine-scale taxonomic structures, a comprehensive analysis of a large collection of *BL.* genomes and associated phenotypic data is warranted to fully delineate the species and its subspecies. Such an analysis may reveal new lineages and provide insights into their unique genetic content and potential ecological roles. Here, we applied comprehensive genomic and phenotypic analyses to uncover a previously unrecognized lineage within the *BL.* species, which we describe here as *Bifidobacterium longum* subsp. *nexus* subsp. novel.

Results

Phylogeny of the *Bifidobacterium longum* clade

To explore the taxonomic structure of *Bifidobacterium longum* (*BL.*) species, we compared the genomic sequence of 435 genomes, 284 of which were obtained from public databases and 151 newly generated from isolated strains obtained from stool samples. These new isolates originated from several studies involving British infants[33,34], Irish mothers[35]and infants[36] and both adults and infants in the Netherlands[37–39] (**Supplementary table 1**). Using these genomes, we estimated the pairwise average nucleotide identity (ANI), and observed four distinct clusters of genomes (**Fig. 1A, Methods**). Three of these clusters had genomes that were previously annotated to one of *BL.* subspecies (37 belonged to *BL. infantis*, 122 to *BL. longum* and 13 from the *BL. suis* clade – which included genomes annotated as *BL. suis*, *BL. suillum* and *BL. iuvenis*). Interestingly, the remaining cluster had nine unannotated

genomes, including two public genomes and seven new genomes from internally isolated strains. We hereby propose the name *Bifidobacterium longum* subsp. *nexus* novel for strains belonging to this novel clade, since the strain was isolated from an infant in the Lifelines NEXT cohort. Some publicly available genomes appeared to be misannotated according to their ANI-based clustering, specifically three *BL. longum* genomes clustered within the *BL. suis* group, and one *BL. infantis* genome was found within the *BL. longum* cluster. Previous reports have noted such mis-annotations, including for genome assembly GCA_000092325, which is likely misclassified as *BL. longum* and instead represents *BL. suis*[40,41]. We next compared the ANI across the various clades, and found that *BL. nexus* had an inter-subspecies ANI of 99.3% (**Fig. 1B**). The lowest ANI value observed among all genomes across the clades was 94.7%, supporting the classification of all subspecies within the *BL.* species (**Fig. 1B**).

To further support the classification of *BL. nexus* as a distinct subspecies, we applied gene-level characterization across all genomes described above. Overall, we examined 18,591 genes in the pangenome of the *BL.* species. We identified 789 core genes present in >99% of genomes, and 17,802 accessory genes. Using alignment of the core genes we constructed a phylogenetic tree on a subset of 24 representative genomes, and found four distinct clades with genomes belonging to *BL. nexus* forming their own group (**Fig. 1C**). In addition, using the accessory gene presence/absence profiles of all annotated genomes we constructed a dendrogram revealing the same result (**Supplementary Fig. 1A**).

To further validate the existence of the novel clade, we employed digital DNA-DNA hybridization (dDDH), a widely used *in silico* method for assessing genomic similarity between sequences (**Methods**). The dDDH values between *BL. nexus* and the other subspecies ranged from 62% to 75.9%, with *BL. longum* being its closest subspecies. Inter-subspecies dDDH values ranged from 94% to 99.3% (**Methods, Supplementary Fig. 1B**). These values meet the 79-80% dDDH suggested threshold for delineating subspecies[42]. We next generated a dendrogram based on these values, showing that *BL. nexus* clusters within the *BL.* clade, however being distinct enough to be a novel discrete subspecies (**Fig. 1D)**. Interestingly, *BL. infantis* was shown to exhibit low dDDH values when compared to other subspecies (62-69.4%), supporting the notion that it is more distinct than other *BL.* subspecies, as has previously been suggested[43,44]. The dDDH results showed strong correlation with the estimated ANI values, collectively indicating that *BL. nexus* is most closely related to *BL. longum* and most distantly related to *BL. infantis* (**Fig. 1D**).

Finally, we conducted a phylogenetic analysis based on the complete 16S rRNA gene sequence, using the *Bifidobacterium breve* type strain genome as an outgroup. This

confirmed that all 24 selected genomes share their most recent common ancestor, while *B. breve* forms, as expected, a sister branch (**Supplementary Fig. 1C)**. Although some genomes from the same subspecies clustered closely, the 16S rRNA gene lacked sufficient phylogenetic resolution to clearly distinguish the known *BL.* subspecies from the novel clade (**Supplementary Fig. 1C**).

Genomic features of *Bifidobacterium longum* subsp. *nexus* nov.

To further investigate *BL. nexus*, we analyzed the genomic features of both newly isolated strains and publicly available genomes. The *BL. nexus* strains were isolated from human stool; five from infants in the Netherlands, Ireland, and the United Kingdom, and two from adults in the Netherlands. Publicly available *BL. nexus* genomes were originally isolated from human milk[45] and vaginal microbiomes[46], sourced from Italy and Canada, respectively. The average genome length of *BL. nexus* was 2,400,483 base pairs (bp), comparable to that of *BL. longum* and *BL. suis*, but shorter than *BL. infantis* (t-test, $p= 4*10^{-6}$; **Supplementary Fig. 2A**). The guanine+cytosine (G+C) content across all *BL. nexus* isolates ranged from 59-60 mol%, consistent with other *BL.* subspecies (t-test, $p>0.3$, **Table 1**).

To examine genome structure conservation and variation across *BL.* subspecies, we performed whole-genome synteny analysis using the type strain from each subspecies. We revealed that *BL. nexus* exhibits a high level of genomic similarity to the other *BL.* subspecies (**Fig. 2**). However, we observed a few notable genomic inversions between subspecies, reflecting apparent genome plasticity of the *BL.* genome[43]. Interestingly, *BL. infantis* was shown to contain genomic regions that are absent from genomes of the other subspecies, consistent with its overall larger genome size. In addition, *BL. nexus* and *BL. suis* were determined to lack a genomic region that is present in *BL. longum* and *BL. infantis*.

To investigate *BL. nexus* specific genes, we searched for genes that are unique to the *BL. nexus* genomes. We identified 112 genes that are uniquely present in all *BL. nexus* genomes and were found in less than 1% of other subspecies (**Fig. 3A**). These genes are distributed throughout the entire *BL. nexus* genome (**Fig. 2**). They corresponded to 222 UniRef90 gene clusters and were primarily linked to Gene Ontology (GO) categories involved in carbohydrate metabolic processes (GO:0005975), carbohydrate transport (GO:0008643), and amino acid transport (GO:0006865). Among these unique genes, we identified a gene predicted to encode an extracellular amylopullulanase, previously characterized in *B. breve*[47], suggesting that *BL. nexus* strains possess the ability to degrade starch, a phenotypic trait not observed in other *BL.* subspecies[48–50]. This is consistent with a recent publication that showed enrichment for starch utilisation in an unknown *BL.* subspecies,

based on KEGG (K01200) and COG (COG1523) analysis[32]. The predicted starch utilization trait suggests an adaptation of *BL. nexus* to metabolize starch-like polysaccharides.

We next searched for *BL. nexus*-specific gene-clusters, namely genes that are located in close proximity on the genome. We found two distinct clusters, each consisting of four or more adjacent genes (**Fig. 3B**). Gene annotations indicate that these clusters are/were involved in carbohydrate metabolism. The first cluster appears to be a genetically decayed remnant of a sugar utilization operon. While most genes are likely non-functional, the remaining MFS transporter may still be active and potentially specific for melibiose, which is a component of raffinose and stachyose, though this remains a prediction based on sequence similarity. The second cluster includes genes predicted to encode a carbohydrate-specific ABC-type transport system. The GH35 enzyme encoded within the cluster lacks a signal peptide, suggesting it functions intracellularly, likely acting on a substrate imported by the afore-mentioned-ABC transporter. Notably, some genes within these clusters encode hypothetical proteins, suggesting the presence of novel functionalities that may contribute to the unique physiological traits of *BL. nexus*.

Description of *Bifidobacterium longum* subsp. *nexus* subsp. nov.

Bifidobacterium longum subsp. *nexus* (since the type strain was isolated from an infant in the Lifelines NEXT Dutch cohort).

To investigate the morphological characteristics of *BL. nexus* in comparison to other known subspecies, we cultured three *BL. nexus* strains alongside representative strains of *BL. longum* and *BL. infantis* on rich supplemented brain heart infusion (BHI) media. Cells are Gram-stain-positive. Colonies are circular, with entire margins, smooth, convex, beige and approximately 1 mm in diameter after 3 days cultivation at 37 °C under anaerobic conditions. Growth in supplemented BHI and modified MRS with lactose is good under anaerobic conditions and weak under aerobic conditions.

Subsequently, to explore the microscopic phenotype of *BL. nexus*, we performed fluorescent staining targeting the cell wall and DNA to visualize cell shape during the exponential phase. Using microscopic imaging we revealed that *BL. nexus* cells predominantly exhibit a rod-shaped morphology, frequently forming chains with a visible septum between cells (**Fig. 4A**). This suggests active cell division with incomplete separation of daughter cells. In many cases, we observed accumulation of the cell wall at the cell poles, possibly indicating sites of upcoming division or budding and further cell splitting. In contrast, a significant proportion of *BL. infantis* and *BL. longum* cells displayed Y-shaped morphologies (**Fig. 4A**), a feature

previously documented as characteristic of *Bifidobacterium* species, and particularly visualized in *BL.*[22,23]. Notably, Y-shaped cells were considerably less prevalent in *BL. nexus* strains (<1% of cells), indicating a distinct morphological pattern under similar growth conditions. These observations suggest pleomorphism within the *BL.* species, with morphological variations manifesting at the subspecies level.

The average cell length of *BL. nexus* strains we examined was 1.64 ± 0.41 µm, significantly shorter than that of *BL. infantis* (p = 8.3e-12) and slightly shorter than *BL. longum* (p = 0.017, **Figure 4B**). This difference may be related to the larger genome size of *BL. infantis*[51,52], as well as to the distinct cellular morphologies observed, however this may vary between strains.

To further explore the functional capabilities of *BL. nexus*, we assessed its growth on various carbohydrates and compared the results with the growth capabilities of *BL. infantis* and *BL. longum*. All tested strains of *BL. nexus*, *BL. longum*, and *BL. infantis* exhibited robust growth on lactose (OD600 > 1, **Fig. 4C**). In contrast, growth on fructose and maltodextrin was significantly stronger in *BL. nexus* strains, whereas *BL. longum* and *BL. infantis* displayed only low to moderate growth (OD600 0.2-1, **Fig. 4C**). Notably, while *BL. longum* and *BL. infantis* showed no growth on starch, *BL. nexus* strains demonstrated vigorous growth (**Fig. 4C**). This unique metabolic capability may be attributed to specific genes present in *BL. nexus*, such as the identified amylopullulanase-encoding gene, a homologue of which is known to enable starch utilization in *B. breve*[47].

LL6991, designated as the type strain, was isolated from the stool of a two-week old Dutch infant born vaginally and breastfed. The DNA G+C content of this strain is 60% mol.

The GenBank accession numbers for the 16S rRNA encoding gene and the genome of LL6991 are PV770940 and JBPCCJ000000000, respectively.

Discussion

Here, using comprehensive genomic and phenotypic analyses we showed that *BL. nexus* subsp nov. constitutes a distinct subspecies within the *BL.* species. This classification is robustly supported by multiple complementary genomic approaches, including average nucleotide identity (ANI), digital DNA–DNA hybridization (dDDH), and pangenome analysis. These findings align with modern prokaryotic taxonomy, which employs a polyphasic approach integrating traditional methods with advanced genome-based analyses to resolve finer taxonomic classifications like subspecies[25,27].

One notable finding is the distinct morphological pattern observed in *BL. nexus*. While the *Bifidobacterium* genus is famously named for its characteristic bifid or Y-shape[19,20], microscopic examination revealed that *BL. nexus* cells predominantly exhibited a rod-shaped morphology, frequently forming chains with visible septa, when cultured in rich supplemented BHI media. In contrast, *BL. infantis* and *BL. longum* cells under similar conditions displayed a significant proportion of Y-shaped morphologies (>30% of cells). This suggests pleomorphism within the *BL.* species, with morphological variations evident at the subspecies level. Pleomorphism has previously been observed in *Bifidobacterium* species and is known to vary not only between subspecies but also between individual strains, influenced by environmental and growth conditions[21]. Recent research on *Bifidobacterium breve* indicates that while the Y-shape can be observed, cells are primarily rod-shaped in nutritionally rich media and *in vivo*, with the branched morphology becoming more frequent in sub-optimal environments[53].

We found that *BL. nexus* exhibits unique metabolic capabilities, particularly its unique capability to efficiently utilize fructose, and starch-like carbohydrates (maltodextrin and starch). These carbohydrates could not be metabolized efficiently by *BL. infantis* or *BL. longum* in this study and in most strains in previous studies[48,54]. This distinct metabolic profile is likely attributed to specific genes present in *BL. nexus*, such as the identified amylopullulanase-encoding gene, a homologue of which is known to enable starch utilization in *B. breve*[47]. The ability to utilize fructose is particularly interesting in the context of the infant gut, where *BL. nexus* was originally isolated. In early life, infants lack the GLUT-5 transporter, which is specific for fructose absorption, implying that any consumed fructose is likely metabolized by the gut microbiota, thereby altering its composition[55–57]. Furthermore, fructoselysine, a food process contaminant prevalent in infant formula due to heating processes, seems to lead to an increased fructoselysine degradation activity in the gut microbiota of infants, indicating adaptation towards dietary fructoselysine exposure[58]. We did not identify specific genes directly responsible for fructose utilization in the *BL. nexus* genomes. However, we observed a broader enrichment of genes associated with carbohydrate metabolism. Further research is required to uncover the underlying mechanisms driving fructose utilization in this subspecies.

To deepen our understanding of *Bifidobacterium* diversity, particularly at the subspecies level, it is crucial to carry out systematic studies across varied geographical regions and host populations. In this study, we isolated *BL. nexus* strains from infants in the Netherlands, Ireland, and the United Kingdom, as well as from adults in the Netherlands, highlighting the importance of sampling across different ages and regions. Such targeted isolation, combined

with comprehensive genomic and phylogenetic analysis, lays the foundation for identifying novel taxa.

In conclusion, the characterization o*f BL. nexus* subsp. nov. not only expands our understanding of the diversity within the *BL.* species but also highlights its unique morphological and metabolic adaptations, particularly its ability to utilize various carbohydrates like starch and fructose. These findings provide insights into its potential ecological roles within the human gut microbiome, especially in infants.

Methods

Bacteria isolation and genomic retrieving

In this study we used 435 *BL.* genomes from public data and from strains we isolated. Public genomes were selected based on criteria of over 95% completeness and less than 5% contamination, found by CheckM2 (1.0.2)[59], to minimize analytical noise. A total of 114 genomes were retrieved from NCBI by searching for entries matching *Bifidobacterium longum*. Similarly, 166 genomes were obtained from the JGI IMG/M (Integrated Microbial Genomes & Microbiomes) database. Subspecies annotations were extracted from strain names when available. In addition, four *BL.* genomes from Chilean isolates were included from Díaz et al.[40].

Genomes from strains we isolated were taken from multiple cohorts and geographical locations. 49 genomes were sourced from the Baby Biome Study (BBS) in the United Kingdom[33], with details on isolation, sequencing and assembly as previously described[33,34]. An additional 99 genomes were obtained from the MicrobeMom study in Ireland, with full methods described therein[36].

Two isolates were isolated as part of the Dutch Microbiome Project (DMP)[39], as follows: We collected stool samples from 550 population-based participants from the Lifelines Dutch Microbiome project. Research approvals were obtained from MIT (COUHESprotocol #1612797956) and from the Medical Ethical Committee of the University Medical Center Groningen,the Netherlands; IRB number: 2017.1520. We used two distinct sampling strategies to culture bacterial isolates. In the first round of culturing, we chose selective culturing conditions for specific taxonomic groups, in this caseDe Man–Rogosa–Sharpe (MRS, Millipore, 69964) medium under anaerobic conditions for genus Bifidobacterium (phylum Actinobacterium). We assigned species for each isolate via multiplexed 16S sequencing for each isolate and performed whole-genome sequencing (WGS) on a subset

(on average two isolates per species per donor). In the second round of culturing, we selected dozens of isolates per donor based on diverse morphological characteristics, cultured all in a rich medium, Brain Heart Infusion (Bacto 237300), and performed WGS. DNA extraction was performed using the PureLink Pro Genomic DNA Purification (K182104A) Kit.

The type strain, LL6991, was isolated in April 2025 in Groningen, Netherlands (53.2210° N, 6.5770° E) from a fecal sample from a two-weeks old infant as part of the Lifelines NEXT cohort, a birth cohort designed to study the effects of intrinsic and environmental determinants on health in mothers and children[37]. All participants gave informed consent upon enrolment. A detailed description of the design, inclusion and exclusion criteria has been described previously[38]. The study protocol has been approved by the local Institutional Review Board (METc 2015/600). For isolation, a stool sample that was known to contain the type strain was spread-plated onto Yeast Casitone Fatty acid agar[60] supplemented with 20 mM Lactose and 0.05% (w/v) cystine (YCFA-Lactose) were then incubated in an anaerobic chamber (90% nitrogen, 5% hydrogen, 5 % carbon dioxide) for 72 hours. Each quadrant of the spread plate was subcultured onto fresh YCFA-Lactose agar plates after which individual colonies were picked up randomly and subcultured on fresh YCFA-Lactose agar for colony purification. These purified colonies were identified up to species level using Bruker MALDI TOF (Bruker Daltonics, Germany) where one of the isolates obtained was identified as *Bifidobacterium longum.* Further, the DNA of *B. longum* isolate was extracted using QiAmp DNA Mini kit (Qiagen, Hilden, Germany) and PCR reactions were run using primers that bind to gene loci unique to *BL. nexus* to verify its subspecies identity.

DNA extraction, sequencing and assembly

For the type strain, LL6991, and for an additional *BL. nexus* strain, MB0413, DNA was extracted using MagAttract HMW DNA Kit (#67563, Qiagen). Short read sequencing libraries were prepared using Nextera XT DNA Library Preparation kit (FC-131-1096, Illumina) according to the manufacturer's recommended protocol with half of the volume and the DNA. Samples were sequenced using Illumina single-end 150 bp sequencing on a NextSeq 500 device. Long reads were retrieved using native barcoding sequencing kit (SQK-NBD114-24, ONT) on a MinION device. Basecalling and demultiplexing of long reads was done using Dorado (Oxford Nanopore Technologies). Hybrid assembly was performed using SPAdes (v4.0.0)[61,62].

For the two strains derived from the UK Baby Biome Study infants, 37822_2_275 and 37822_2_213, hybrid genome assemblies were generated. DNA was extracted using the

MasterPure Complete DNA and RNA Purification Kit (Lucigen), then sequenced on the Illumina NovaSeq 6000 (384-plex, 2 x 151bp, Illumina) and on the Oxford Nanopore GridION (24-plex) or PromethION (96-plex) using R9.4.1 and R10.4.1 flow cells with basecalling in 'super-accurate (SUP) model using Dorado. Hybrid genome assemblies were generated using the Dragonflye with Flye for long-read assembly ('--nanohq' with R9.4.1 data and '--nano_hq --read-error 0.03' with R10.4.1 data), Medaka (https://github.com/nanoporetech/medaka) for long-read polishing and Polypolish for short-read polishing.

Genomic and phylogeny analysis

Assessment of genome quality and retrieving genome information was performed using CheckM2[59] and Bakta[63–66]. Average nucleotide identity between genomes was estimated using FastANI (1.33)[67]. The selection of the number four for clustering was done using the Silhouette method, where the optimal number of clusters is the one with the highest average silhouette width. Gene annotations were created using Prokka (1.14.6)[68], which uses BioPerl[69], GNU Parallel[70], BLAST+[71] and Prodigal[65], and pangenomes were analyzed with Roary (3.13.1)[72]. 16s rRNA gene sequence was retrieved from genomes using Barrnap (0.9)[73]. Multiple sequence alignment and distance calculation of 16S rRNA gene was performed using Clustal Omega (1.2.4)[74]. 16s rRNA and core gene phylogenetic trees were created using RAxML with raxmlHPC-SSE3[75]. Genome synteny was assessed using minimap2 (2-2.24)[76] with -ax flag asm20, and visualised using the gggenomes (1.0.1) R package[77]. Gene clusters were searched using Integrative Genomics Viewer (IGV)[78] and visualised using gggenomes[77]. Genes annotated as hypothetical proteins by Prokka[68] were further analyzed using BLASTx[71] against non-redundant protein sequences (nr) database and HHpred[79] to infer potential functions. Biosynthetic gene clusters (BGCs) were searched using antiSMASH (8.0.1)[80].

dDDH values were calculated using the genome-to-genome distance web-based calculator (GGDC 3.0[81,82]). Genomes were uploaded to the web server, genomes were aligned using BLAST, distances and dDDH values were calculated using the d2 formula (identities / HSP length) of GGDC, which uses a generalized linear model (GLM) inferred from an empirical reference dataset comprising real DDH values and genome sequence[82].

BL. nexus-specific gene sequences were mapped against the Uniref90 database[83,84] using uniref_annotator.py script[85], which utilizes DIAMOND[86] for alignment. Uniref90 annotations were then assigned a gene Gene Ontology (GO) annotations using UniProt.

Further analysis was done using an in-house R (4.4.0) script utilizing dplyr[87] (1.1.4), tidyr[88] (1.3.1), stringr[89] (1.5.1) and tidyverse[90] (2.0.0). Plots were created using ggplot2[91] (3.5.1) and ggforce[92] (0.4.2), colors were used from RColorBrewer[93] (1.1.3). Heatmaps were created using pheatmap[94] (1.0.12) using the cutree option to divide into clusters. Distance was calculated using the dist function with euclidean distance. Dendrograms were constructed using a neighbour joining algorithm with the nj function from the ape package[95] (5.8.1) and visualized using ggtree[96] (3.12.0). T-tests were pregormed using using R.

Culture Conditions and Microscopy

Culturing and microscopy was performed on a representative of *BL. longum*, *BL. infantis* and three *BL. nexus* isolates; MB0401, MB0413 and LL6991. *BL. infantis* and *BL. longum* were obtained from the BMB cohort[7].

Strains were grown routinely at 37ºC in an anaerobic chamber (COY) that was maintained at > 2.5% H2 by flushing with a gas mix containing 20% H2, 5% CO2, and 75% N2. Strains were grown in supplemented Brain Heart Infusion (sBHI) media, supplemented with 50 ml/L fetal bovine serum (FBS; Sigma, F2442), 10 ml/L trace vitamins (ATCC® MD-VS™), 10 ml/L trace minerals (ATCC® MD-TMS™), 10 ml/L vitamin K1 and hemin (BBL, 212354), 1 g/L D-(+)-Cellobiose (Alfa Aesar, 528507), 1 g/L D-(+)-Maltose (Caisson, 6363537), 1 g/L D-(+)-Fructose (Sigma Aldrich, 1286504), and 0.5 g/L L-Cysteine (Acros Organics, 52904). For imaging, strains were cultured overnight and then diluted 30-fold in fresh medium and grown to exponential phase. Next, they were stained with DAPI (4-,6-diamidino-2-phenylindole, Sigma-Aldrich Cat #D8417) and FM 4-64 Dye (N-(3-Triethylammoniumpropyl)-4-(6-(4-(Diethylamino) Phenyl) Hexatrienyl) Pyridinium Dibromide, Invitrogen Cat# T13320). Cells were placed on an agarose pad with uncoated coverslips and imaged using a Nikon Eclipse Ti-E inverted microscope equipped with Perfect Focus System (PFS) and ORCA Flash 4 camera (Hamamatsu photonics). Images were processed, two-dimensional (2D) deconvolution and bacteria cell length measurement was performed using NIS Elements-AR software.

Additionally, to test growth on various carbohydrates, a 2% w/v sugar solution was mixed with 2x Man-Rogosa-Sharpe (mMRS) medium at a 1:1 ratio (1.5 mL of each) to yield a final 3 mL solution of 1x mMRS medium containing 1% sugar. 2x mMRS medium was made with 20 g/L peptone from casein, tryptic digest; 10 g/L yeast extract; 4 g/L potassium phosphate; 10 g/L sodium acetate; 4 g/L ammonium citrate; 0.4 g/L magnesium sulfate; 0.05 g/L manganese sulfate; and 2 g/L (2 mL/L) Tween 80. 0.5 mg/ml L-cysteine hydrochloride was added before use. The strains were inoculated into this medium at a 1:100 dilution from

cultures grown in liquid supplemented BHI overnight. These cultures were prepared by picking colonies off of CHG plates incubated for 72 hours and transferring them into liquid CHG to incubate overnight. The initial plates were seeded using glycerol stocks of each strain preserved in a -80 degree freezer. Two biological replicates were used for each strain from two isolated colonies on their respective plates.

Funding and conflict of interests

YS and TL acknowledge core funding from the Wellcome Trust (206194 and 220540/Z/20/A). COI: T.D.L. is the co-founder and CSO of Microbiotica. AZ received travel funding from Nestle to attend the Nestlé Nutrition Institute Workshop 101. AZ is supported by the Netherlands Organization for Scientific Research NWO-VICI grant VI.C.232.074, the NWO Gravitation grant ExposomeNL 024.004.017, the NWO KIC grant KICH1.LWV04.21.01.,the ZonMW ME/CFS grant 10091012110015, EU Horizon Europe Program grant INITIALISE (101094099), and EU Horizon Europe Program grant DarkMatter (ID-DarkMatter-NCD, 101136582). MY is the Rosalind, Paul and Robin Berlin Faculty Development Chair in Perinatal Research, and was also funded by the Azrieli Family Foundation.

References

1. Turroni, F. *et al.* The infant gut microbiome as a microbial organ influencing host well-being. *Ital. J. Pediatr.* **46**, 16 (2020).
2. Arboleya, S., Watkins, C., Stanton, C. & Ross, R. P. Gut bifidobacteria populations in human health and aging. *Front. Microbiol.* **7**, 1204 (2016).
3. Derrien, M., Turroni, F., Ventura, M. & van Sinderen, D. Insights into endogenous Bifidobacterium species in the human gut microbiota during adulthood. *Trends Microbiol.* **30**, 940–947 (2022).
4. Yassour, M. *et al.* Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host Microbe* **24**, 146–154.e4 (2018).
5. Taft, D. H. *et al.* Bifidobacterial dominance of the gut in early life and acquisition of antimicrobial resistance. *mSphere* **3**, (2018).
6. Laforest-Lapointe, I. & Arrieta, M.-C. Patterns of early-life gut microbial colonization

during human immune development: An ecological perspective. *Front. Immunol.* **8**, 788 (2017).

7. Ennis, D., Shmorak, S., Jantscher-Krenn, E. & Yassour, M. Longitudinal quantification of Bifidobacterium longum subsp. infantis reveals late colonization in the infant gut independent of maternal milk HMO composition. *Nat. Commun.* **15**, 894 (2024).
8. Xiao, Y. *et al.* Human gut-derived B. longum subsp. longum strains protect against aging in a D-galactose-induced aging mouse model. *Microbiome* **9**, 180 (2021).
9. Zhang, C. *et al.* Colonization and probiotic function of Bifidobacterium longum. *J. Funct. Foods* **53**, 157–165 (2019).
10. Rivière, A., Selak, M., Lantin, D., Leroy, F. & De Vuyst, L. Bifidobacteria and butyrate-producing colon bacteria: Importance and strategies for their stimulation in the human gut. *Front. Microbiol.* **7**, 979 (2016).
11. Parte, A. C., Sardà Carbasse, J., Meier-Kolthoff, J. P., Reimer, L. C. & Göker, M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **70**, 5607–5612 (2020).
12. Odamaki, T. *et al.* Genomic diversity and distribution of Bifidobacterium longum subsp. longum across the human lifespan. *Sci. Rep.* **8**, 85 (2018).
13. Mattarelli, P., Bonaparte, C., Pot, B. & Biavati, B. Proposal to reclassify the three biotypes of Bifidobacterium longum as three subspecies: Bifidobacterium longum subsp. longum subsp. nov., Bifidobacterium longum subsp. infantis comb. nov. and Bifidobacterium longum subsp. suis comb. nov. *Int. J. Syst. Evol. Microbiol.* **58**, 767–772 (2008).
14. Modesto, M. *et al.* Bifidobacterium longum subsp. iuvenis subsp. nov., a novel subspecies isolated from the faeces of weaning infants. *Int. J. Syst. Evol. Microbiol.* **73**, (2023).
15. Vatanen, T. *et al.* A distinct clade of Bifidobacterium longum in the gut of Bangladeshi children thrives during weaning. *Cell* **185**, 4280–4297.e12 (2022).
16. Yanokura, E. *et al.* Subspeciation of Bifidobacterium longum by multilocus approaches

and amplified fragment length polymorphism: Description of B. longum subsp. suillum subsp. nov., isolated from the faeces of piglets. *Syst. Appl. Microbiol.* **38**, 305–314 (2015).

17. Sela, D. A. *et al.* Bifidobacterium longum subsp. infantis ATCC 15697 α-fucosidases are active on fucosylated human milk oligosaccharides. *Appl. Environ. Microbiol.* **78**, 795–803 (2012).
18. Kelly, S. M., Munoz-Munoz, J. & van Sinderen, D. Plant glycan metabolism by bifidobacteria. *Front. Microbiol.* **12**, 609418 (2021).
19. Young, K. D. The selective value of bacterial shape. *Microbiol. Mol. Biol. Rev.* **70**, 660–703 (2006).
20. Tissier G Carre, H. & Naud, C. *Recherches Sur La Flore Intestinale Des Nourrissons (e'tat Normal et Pathologique)*. (Paris, 1900).
21. Bauer, H., Sigarlakie, E. & Faure, J. C. Scanning and transmission electron microscopy of three strains of Bifidobacterium. *Can. J. Microbiol.* **21**, 1305–1316 (1975).
22. Husain, I., Poupard, J. A. & Norris, R. F. Influence of nutrition on the morphology of a strain of Bifidobacterium bifidum. *J. Bacteriol.* **111**, 841–844 (1972).
23. Dhanashree, Rajashekharan, S., Krishnaswamy, B. & Kammara, R. Bifid Shape Is Intrinsic to Bifidobacterium adolescentis. *Front. Microbiol.* **8**, 478 (2017).
24. Schleifer, K. H. Classification of Bacteria and Archaea: past, present and future. *Syst. Appl. Microbiol.* **32**, 533–542 (2009).
25. Prakash, O. *et al.* Polyphasic approach of bacterial classification - An overview of recent advances. *Indian J. Microbiol.* **47**, 98–108 (2007).
26. Kiepas, A. B., Hoskisson, P. A. & Pritchard, L. 16S rRNA phylogeny and clustering is not a reliable proxy for genome-based taxonomy in Streptomyces. *Microb. Genom.* **10**, (2024).
27. Vandamme, P. *et al.* Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microbiol. Rev.* **60**, 407–438 (1996).
28. Riesco, R. & Trujillo, M. E. Update on the proposed minimal standards for the use of

genome data for the taxonomy of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **74**, (2024).

29. Tett, A. *et al.* The Prevotella copri complex comprises four distinct clades underrepresented in Westernized populations. *Cell Host Microbe* **26**, 666–679.e7 (2019).
30. Blanco-Míguez, A. *et al.* Extension of the Segatella copri complex to 13 species with distinct large extrachromosomal elements and associations with host conditions. *Cell Host Microbe* **31**, 1804–1819.e9 (2023).
31. Versmessen, N. *et al.* Average Nucleotide Identity and digital DNA-DNA Hybridization analysis following PromethION nanopore-based whole genome sequencing allows for accurate prokaryotic typing. *Diagnostics (Basel)* **14**, 1800 (2024).
32. Shao, Y. *et al.* Genomic atlas of early life bifidobacteria for rational development of next-generation probiotics for infant health. *Social Science Research Network* (2025) doi:10.2139/ssrn.5095902.
33. Shao, Y. *et al.* Stunted microbiota and opportunistic pathogen colonization in caesarean-section birth. *Nature* **574**, 117–121 (2019).
34. Shao, Y. *et al.* Primary succession of Bifidobacteria drives pathogen resistance in neonatal microbiota assembly. *Nat. Microbiol.* **9**, 2570–2582 (2024).
35. Díaz, R. & Garrido, D. Screening competition and cross-feeding interactions during utilization of human milk oligosaccharides by gut microbes. *Microbiome Res Rep* **3**, 12 (2024).
36. Feehily, C. *et al.* Detailed mapping of Bifidobacterium strain transmission from mother to infant via a dual culture-based and metagenomic approach. *Nat. Commun.* **14**, 3015 (2023).
37. Sinha, T. *et al.* Pregnancy and early life gut microbiome: Influencing factors and health implications. *Research Square* (2024) doi:10.21203/rs.3.rs-5334252/v1.
38. Warmink-Perdijk, W. D. B. *et al.* Lifelines NEXT: a prospective birth cohort adding the next generation to the three-generation Lifelines cohort study. *Eur. J. Epidemiol.* **35**, 157–168 (2020).

39. Gacesa, R. *et al.* Environmental factors shaping the gut microbiome in a Dutch population. *Nature* **604**, 732–739 (2022).
40. Díaz, R., Torres-Miranda, A., Orellana, G. & Garrido, D. Comparative Genomic Analysis of Novel Bifidobacterium longum subsp. longum Strains Reveals Functional Divergence in the Human Gut Microbiota. *Microorganisms* **9**, (2021).
41. Andreu-Sánchez, S. *et al.* Global genetic diversity of human gut microbiome species is related to geographic location and host health. *Cell* **0**, (2025).
42. Meier-Kolthoff, J. P. *et al.* Complete genome sequence of DSM 30083(T), the type strain (U5/41(T)) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. *Stand. Genomic Sci.* **9**, 2 (2014).
43. Tarracchini, C. *et al.* Phylogenomic disentangling of the Bifidobacterium longum subsp. infantis taxon. *Microb. Genom.* **7**, (2021).
44. Sela, D. A. *et al.* The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18964–18969 (2008).
45. Duranti, S. *et al.* Maternal inheritance of bifidobacterial communities and bifidophages in infants through vertical transmission. *Microbiome* **5**, 66 (2017).
46. Freitas, A. C. & Hill, J. E. Quantification, isolation and characterization of Bifidobacterium from the vaginal microbiomes of reproductive aged women. *Anaerobe* **47**, 145–156 (2017).
47. O'Connell Motherway, M. *et al.* Characterization of ApuB, an extracellular type II amylopullulanase from Bifidobacterium breve UCC2003. *Appl. Environ. Microbiol.* **74**, 6271–6279 (2008).
48. Ryan, S. M., Fitzgerald, G. F. & van Sinderen, D. Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains. *Appl. Environ. Microbiol.* **72**, 5289–5296 (2006).
49. Millar, M. E. *et al.* Novel amylase genes enable utilisation of resistant starch by bifidobacteria relevant to early-life microbiome development. *Microbiology* (2024).

50. Rhim, S. L., Park, M. S. & Ji, G. E. Expression and secretion of Bifidobacterium adolescentis amylase by Bifidobacterium longum. *Biotechnol. Lett.* **28**, 163–168 (2006).
51. Sabath, N., Ferrada, E., Barve, A. & Wagner, A. Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol. Evol.* **5**, 966–977 (2013).
52. Mueller, R. L. Genome biology and the evolution of cell-size diversity. *Cold Spring Harb. Perspect. Biol.* **7**, a019125 (2015).
53. Shiver, A. L. *et al.* Genome-scale resources in the infant gut symbiont Bifidobacterium breve reveal genetic determinants of colonization and host-microbe interactions. *Cell* **188**, 2003–2021.e19 (2025).
54. Liu, S. *et al.* Starch and starch hydrolysates are favorable carbon sources for bifidobacteria in the human gut. *BMC Microbiol.* **15**, 54 (2015).
55. Ferraris, R. P., Choe, J.-Y. & Patel, C. R. Intestinal absorption of fructose. *Annu. Rev. Nutr.* **38**, 41–67 (2018).
56. Mokhtari, P. *et al.* Associations between dietary sugar and fiber with infant gut microbiome colonization at 6 mo of age. *J. Nutr.* **154**, 152–162 (2024).
57. Ramne, S. *et al.* Gut microbiota composition in relation to intake of added sugar, sugar-sweetened beverages and artificially sweetened beverages in the Malmö Offspring Study. *Eur. J. Nutr.* **60**, 2087–2097 (2021).
58. van Dongen, K. C. W., Ioannou, A., Wesseling, S., Beekmann, K. & Belzer, C. Differences in gut microbial fructoselysine degradation activity between breast-fed and formula-fed infants. *FEMS Microbiol. Ecol.* **99**, fiac145 (2022).
59. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **20**, 1203–1212 (2023).
60. Duncan, S. H., Louis, P. & Flint, H. J. Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl. Environ. Microbiol.* **70**, 5810–5817 (2004).

61. Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes DE Novo Assembler. *Curr. Protoc. Bioinformatics* **70**, e102 (2020).

62. Antipov, D., Korobeynikov, A., McLean, J. S. & Pevzner, P. A. hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* **32**, 1009–1015 (2016).

63. Schwengers, O. *et al.* Bakta: rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. *Microb. Genom.* **7**, (2021).

64. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).

65. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

66. Larralde, M. Pyrodigal: Python bindings and interface to Prodigal, an efficient method for gene prediction in prokaryotes. *J. Open Source Softw.* **7**, 4296 (2022).

67. Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T. & Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **9**, 5114 (2018).

68. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

69. Stajich, J. E. *et al.* The Bioperl toolkit: Perl modules for the life sciences. *Genome Res.* **12**, 1611–1618 (2002).

70. Tange, O. GNU Parallel - The Command-Line Power Tool, ;login. *The USENIX Magazine, Feb* 42–47 (2011).

71. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

72. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).

73. Seemann, T. *Barrnap: :microscope: Bacterial Ribosomal RNA Predictor*. (Github).

74. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018).
75. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
76. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
77. Hackl, T., Ankenbrand, M., van Adrichem, B., Wilkins, D. & Haslinger, K. gggenomes: effective and versatile visualizations for comparative genomics. *arXiv* Preprint at https://doi.org/10.48550/arXiv.2411.13556 (2024).
78. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
79. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–8 (2005).
80. Blin, K. *et al.* antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. *Nucleic Acids Res.* (2025) doi:10.1093/nar/gkaf334.
81. Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L. & Göker, M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* **50**, D801–D807 (2022).
82. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P. & Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* **14**, 60 (2013).
83. Suzek, B. E. *et al.* UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932 (2015).
84. UniProt Consortium. UniProt: The universal protein knowledgebase in 2025. *Nucleic Acids Res.* **53**, D609–D617 (2025).
85. Beghini, F. *et al.* Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* **10**, (2021).
86. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale

using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).

87. Wickham, H., François, R., Henry, L., Müller, K. & Vaughan, D. dplyr: A Grammar of Data Manipulation. Preprint at (2023).
88. Wickham, H., Vaughan, D. & Girlich, M. tidyr: Tidy Messy Data. Preprint at (2023).
89. Wickham, H. stringr: Simple, Consistent Wrappers for Common String Operations. Preprint at https://CRAN.R-project.org/package=stringr (2023).
90. Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* vol. 4 1686 Preprint at https://doi.org/10.21105/joss.01686 (2019).
91. Wickham, H. Data Analysis. in *ggplot2: Elegant Graphics for Data Analysis* (ed. Wickham, H.) 189–201 (Springer International Publishing, Cham, 2016).
92. Pedersen, T. L. ggforce: Accelerating 'ggplot2'. Preprint at (2022).
93. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. Preprint at https://CRAN.R-project.org/package=RColorBrewer (2022).
94. Kolde, R. pheatmap: Pretty Heatmaps. Preprint at https://CRAN.R-project.org/package=pheatmap (2019).
95. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
96. Xu, S. *et al. Ggtree* : A serialized data object for visualization of a phylogenetic tree and annotation data. *Imeta* **1**, (2022).

Figure captions

Figure 1: Analysis of *Bifidobacterium longum* genomes reveals four distinct clades. (A) Average nucleotide identity (ANI) distances between all 435 B. longum genomes analyzed in this study. Genome annotation data is shown for genomes with available public annotations. *BL. nexus* genomes are highlighted in orange. **(B)** Median pairwise average nucleotide identity (ANI) within and between the four *B. longum* subspecies. **(C)** Phylogenetic tree of 24 selected genomes across subspecies constructed from the core genome alignment. **(D)** Dendrogram of selected genomes generated from digital DNA-DNA hybridization (dDDH) distances among 24 selected *BL.* genomes.

Figure 2: Genomic synteny analysis between type strains of *Bifidobacterium longum* subspecies. Each line represents a genome, when multiple lines are present each one is a single contig. Green shading between genomes denote conserved collinear regions identified via sequence alignment. On the *BL. nexus* genome, *BL. nexus*-specific genes are highlighted in orange.

Figure 3: Unique genomic features of *BL. nexus*. (A) Presence/absence heatmap of accessory genes in the *B. longum* pangenome. Genomes are annotated by subspecies where available (from NCBI or JGI IMG/M). **(B)** Genomic regions containing clusters of ≥4 *BL. nexus*-specific genes are shown in brown; non-*BL. nexus*-specific genes are shown with transparent shading. Gene annotations appear above/below each feature.

Figure 4. Morphology and phenotypic characteristics of *Bifidobacterium longum* strains. (A) Representative microscopic images of *BL. infantis*, *BL. longum*, and three *BL. nexus* strains during the logarithmic growth phase. Cells are stained to visualize the cell membrane and DNA. Scale bar, 2.5 μm. **(B)** Distribution of bacterial cell lengths (μm) for each strain. Boxes indicate the interquartile range (25th–75th percentiles), with the median marked; whiskers extend to 1.5 times the interquartile range. Colors denote subspecies. **(C)** Growth rates of *BL. infantis*, *BL. longum*, and three *BL. nexus* strains on various carbohydrates, measured by OD600. OD600 values are categorized as no growth (0-0.2), low (0.2–0.6), medium (0.6–1), and high (>1).

Table 1. Metadata for *BL. nexus* strains and genomes. This table summarizes key metadata for the nine *BL. nexus* genomes analyzed in this study, including isolation source of the strains, geographic origin, and relevant genomic features.

Supplementary Figure 1: (A) Dendrogram of *B. longum* genomes with subspecies annotations (172 genomes), constructed using presence/absence patterns of accessory genes in the *B. longum* pangenome. **(B)** Heatmap showing digital DNA–DNA hybridization (dDDH) values among selected genomes. **(C)** Phylogenetic tree of selected genomes based on the 16S rRNA gene nucleotide sequences, with *B. breve* used as the outgroup. Subspecies clustering is not as clear in this case due to lack of resolution in the 16s rRNA gene sequence.

Supplementary Figure 2: Histogram showing the genome lengths (in base pairs) of all analyzed genomes.

Supplementary Table 1: List of all genomes used in the study, including their sources and subspecies annotations when publicly available (from NCBI or JGI IMG/M). Cohorts included

are: Baby Biome Study (BBS)[33,34], Diaz et al.[40], Dutch microbiome project (DMP)[39], Vatanen et al. 2022[15], MicrobeMom[36], Lifelines NEXT (LLNEXT)[37,38].

Figure 1

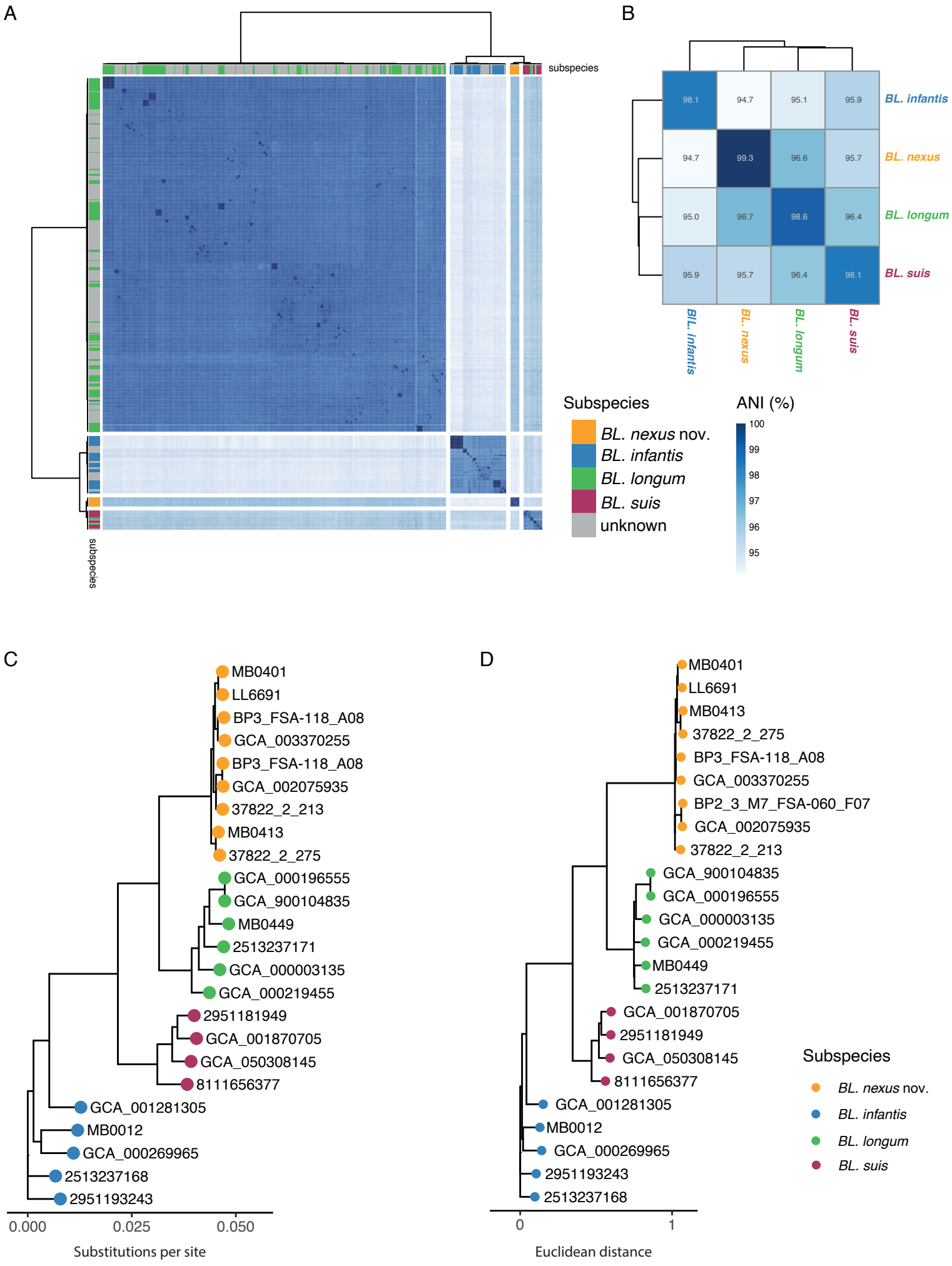

A
subspecies
subspecies
B
98.1
94.7
95.1
95.9
94.7
99.3
96.6
95.7
95.0
96.7
98.6
96.4
95.9
95.7
96.4
98.1
BL. infantis
BL. nexus
BL. longum
BL. suis
Subspecies
BL. nexus nov.
BL. infantis
BL. longum
BL. suis
unknown
ANI (%)
100
99
98
97
96
95
C
MB0401
LL6691
BP3_FSA-118_A08
GCA_003370255
BP3_FSA-118_A08
GCA_002075935
37822_2_213
MB0413
37822_2_275
GCA_000196555
GCA_900104835
MB0449
2513237171
GCA_000003135
GCA_000219455
2951181949
GCA_001870705
GCA_050308145
8111656377
GCA_001281305
MB0012
GCA_000269965
2513237168
2951193243
0.000
0.025
0.050
Substitutions per site
D
MB0401
LL6691
MB0413
37822_2_275
BP3_FSA-118_A08
GCA_003370255
BP2_3_M7_FSA-060_F07
GCA_002075935
37822_2_213
GCA_900104835
GCA_000196555
GCA_000003135
GCA_000219455
MB0449
2513237171
GCA_001870705
2951181949
GCA_050308145
8111656377
GCA_001281305
MB0012
GCA_000269965
2951193243
2513237168
0
1
Euclidean distance
Subspecies
BL. nexus nov.
BL. infantis
BL. longum
BL. suis

Figure 2

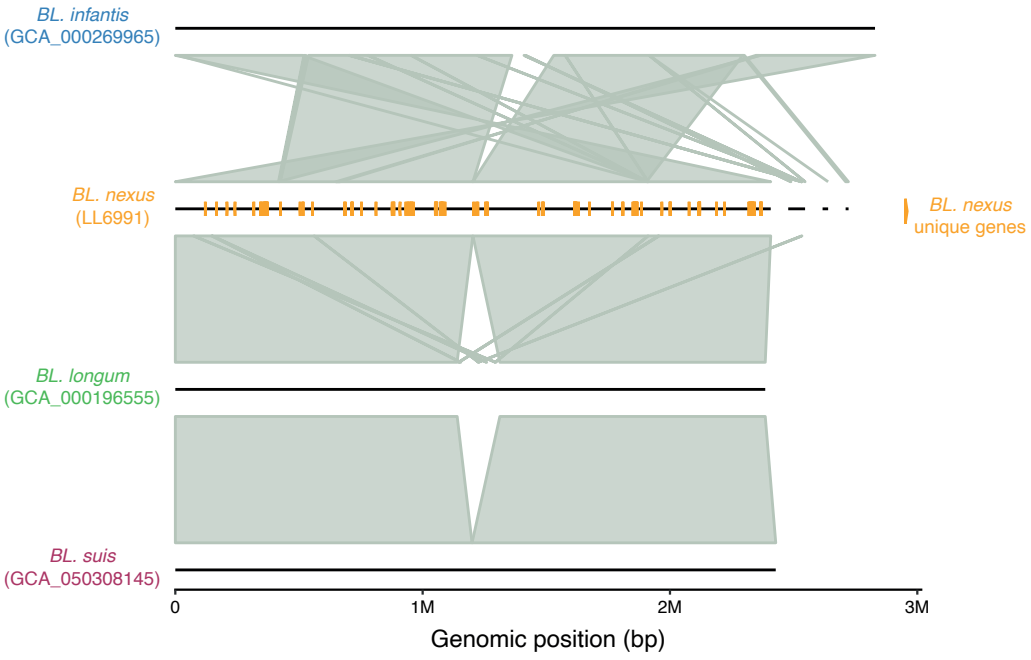
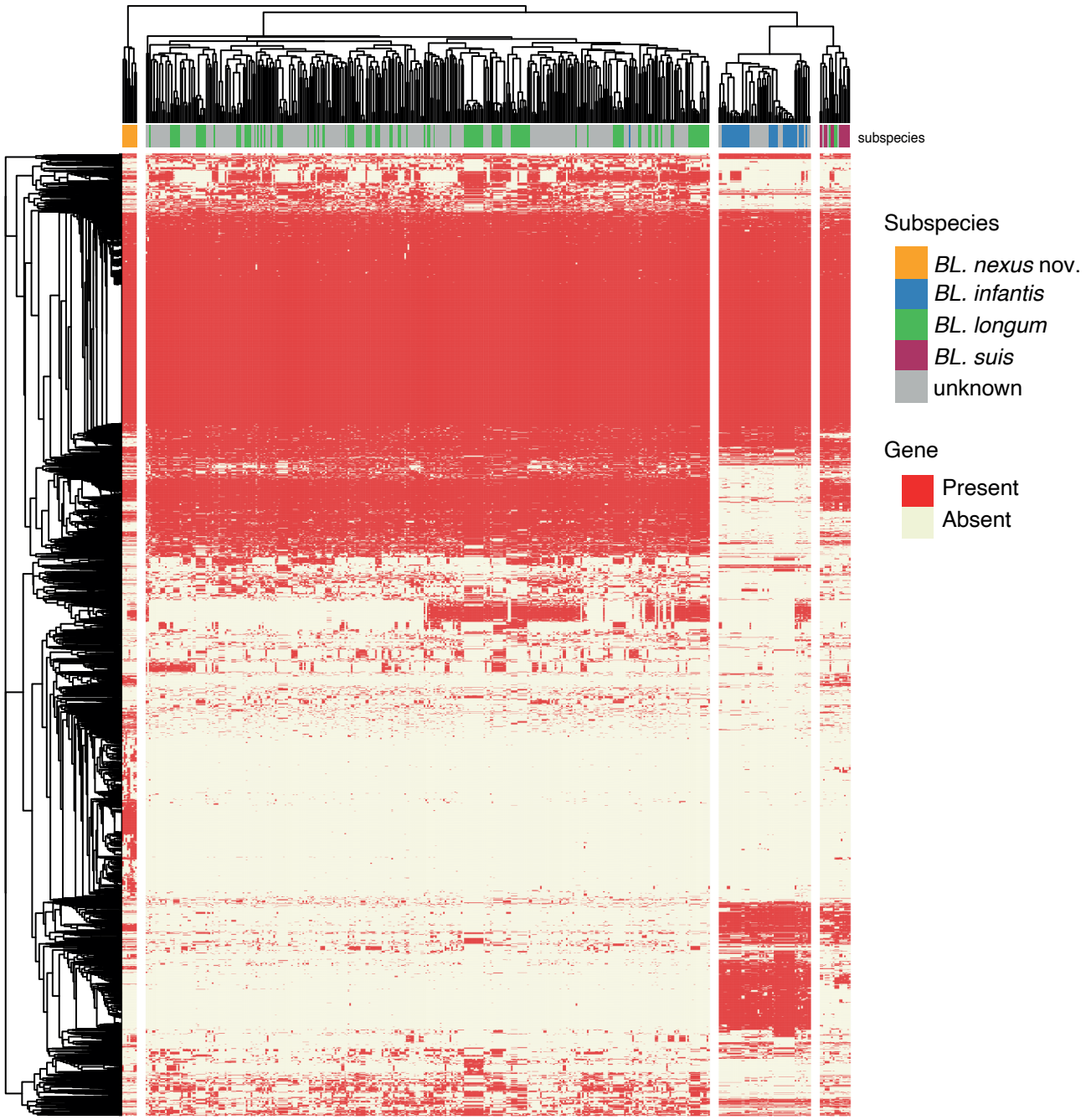

BL. infantis
(GCA_000269965)
BL. nexus
(LL6991)
BL. nexus
unique genes
BL. longum
(GCA_000196555)
BL. suis
(GCA_050308145)
0
1M
2M
3M
Genomic position (bp)

Figure 3

A


subspecies
Subspecies
BL. nexus nov.
BL. infantis
BL. longum
BL. suis
unknown
Gene
Present
Absent

B

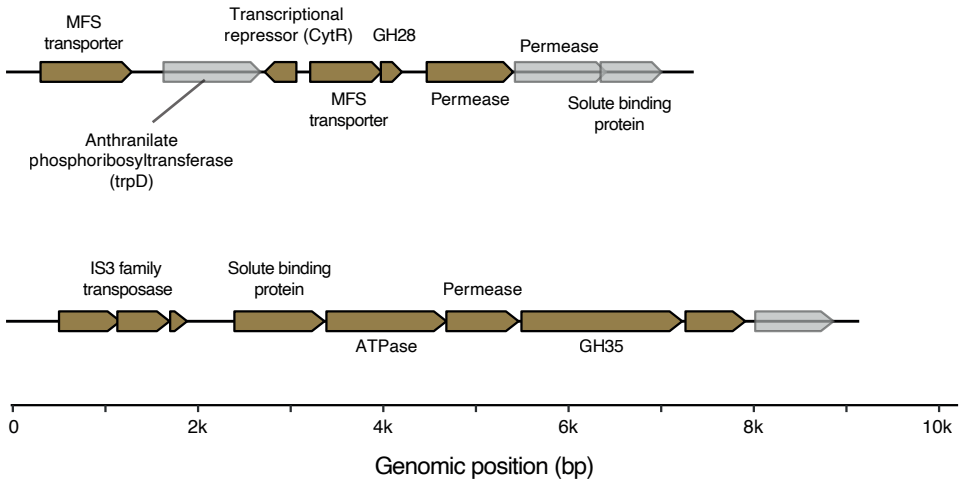
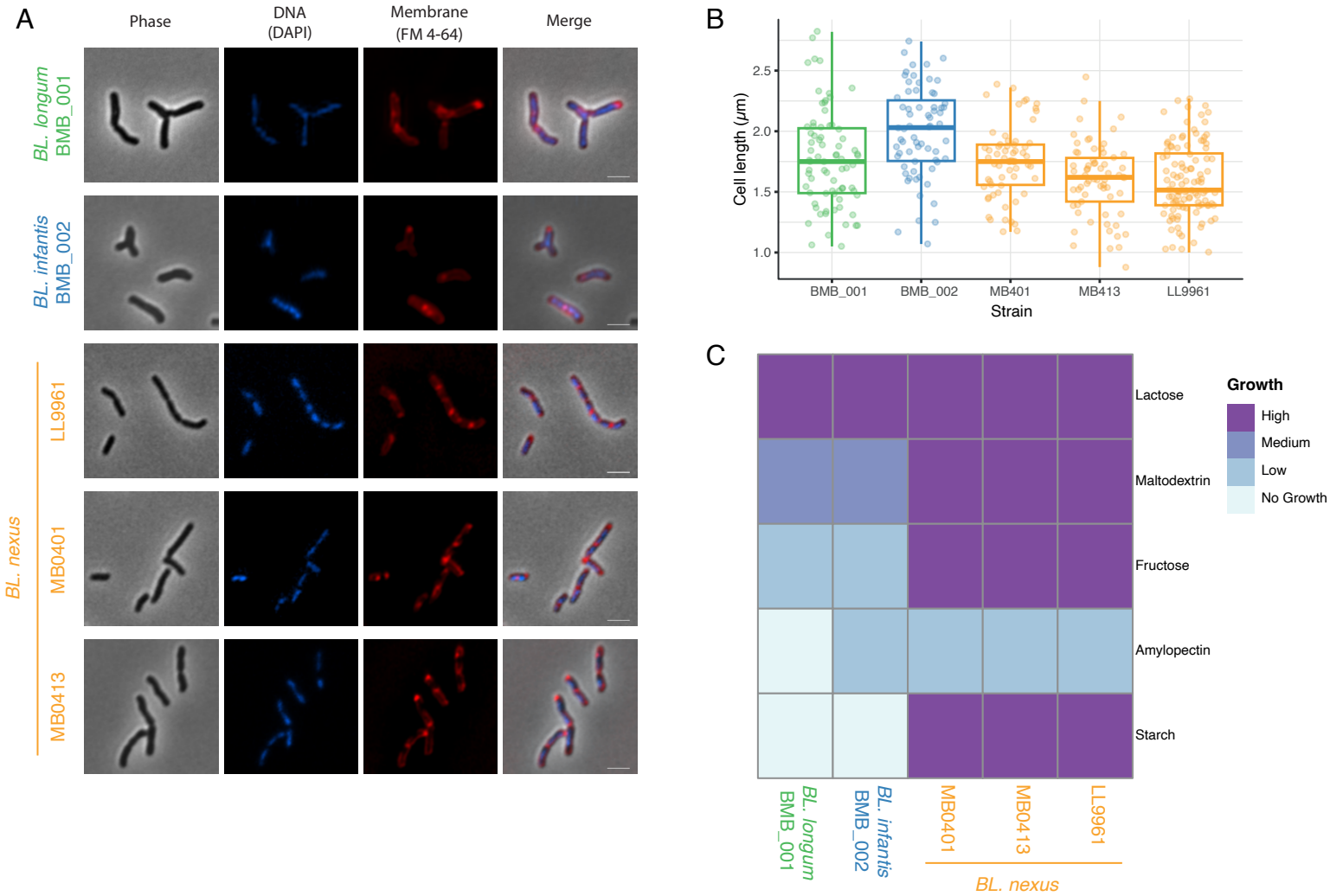

MFS transporter
Transcriptional repressor (CytR)
GH28
Permease
MFS transporter
Permease
Solute binding protein
Anthranilate phosphoribosyltransferase (trpD)
IS3 family transposase
Solute binding protein
Permease
ATPase
GH35
0
2k
4k
6k
8k
10k
Genomic position (bp)

Figure 4


A
Phase
DNA (DAPI)
Membrane (FM 4-64)
Merge
BL. longum BMB_001
BL. infantis BMB_002
BL. nexus
LL9961
MB0401
MB0413
B
Cell length (μm)
2.5
2.0
1.5
1.0
BMB_001
BMB_002
MB401
MB413
LL9961
Strain
C
Lactose
Maltodextrin
Fructose
Amylopectin
Starch
Growth
High
Medium
Low
No Growth
BL. longum BMB_001
BL. infantis BMB_002
MB0401
MB0413
LL9961
BL. nexus

Name	Cohort	Geographical	Isolation source	Genome Size (bp)	G+C Content (nmol%)	Total Contigs
37822_2_213	BBS	United Kingdom	Infant stool	2359727	0.6	42
37822_2_275	BBS	United Kingdom	Infant stool	2479076	0.6	61
BP.3_FSA.118_A08	DMP	Netherlands	Adult stool	2325002	0.59	37
BP2_3_M7_FSA.060_F07	DMP	Netherlands	Adult stool	2414568	0.6	39
GCA_002075935	NCBI	Italy	Human milk	2453605	0.6	55
GCA_003370255	NCBI	Canada: Toronto	Vaginal microbiome	2275165	0.6	116
MB0401	MicrobeMom	Ireland	Infant stool	2491550	0.6	33
MB0413	MicrobeMom	Ireland	Infant stool	2471158	0.6	20
LL6991	Lifelines NEXT	Netherlands	Infant stool	2648729	0.6	215

Completeness (%)	Contamination (%)	Coding Density	Contig N50	Total Coding Sequences	Average Gene Length (bp)	Total rRNA	Total tRNA
99.99	0.07	0.872	278613	2066	332.6	3	57
100	0.06	0.87	200066	2180	330.55	3	56
99.99	0.06	0.867	370448	1965	342.48	5	55
99.99	0.04	0.873	140849	2077	338.94	2	56
99.99	0.07	0.872	185292	2176	328.49	5	55
99.91	0.09	0.855	33459	1947	333.57	4	56
100	0.06	0.871	185207	2149	337	6	55
100	0.02	0.869	2467408	2118	338.48	9	55
100	0.66	0.866	2407308	2411	317.91	9	58