

Bifidobacterium longum subsp. *iuvenis* subsp. nov., a novel subspecies isolated from the faeces of weaning infants

Monica Modesto¹, Catherine Ngom-Bru², Donatella Scarafile¹, Anne Bruttin², Solenn Pruvost², Shafiqul Alam Sarker³, Tahmeed Ahmed³, Olga Sakwinska², Paola Mattarelli^{1,*} and Stéphane Duboux²

Abstract

The species *Bifidobacterium longum* currently comprises four subspecies: *B. longum* subsp. *longum*, *B. longum* subsp. *infantis*, *B. longum* subsp. *suis* and *B. longum* subsp. *suillum*. Recently, several studies on *B. longum* suggested the presence of a separate clade containing four strains isolated from infants and one from rhesus macaque. These strains shared a phylogenetic similarity to *B. longum* subsp. *suis* DSM 20210^T and *B. longum* subsp. *suillum* JCM1995^T [average nucleotide identity (ANI) of 98.1%] while showed an ANI of 96.5% with both *B. longum* subsp. *infantis* and *B. longum* subsp. *longum*. The current work describes five novel additional *B. longum* strains isolated from Bangladeshi weaning infants and demonstrates their common phylogenetic origin with those of the previously proposed separated clade. Based on polyphasic taxonomic approach comprising loci multilocus sequence analysis and whole genome multilocus sequence typing, all ten examined strains have been confirmed as a distinct lineage within the species *B. longum* with *B. longum* subsp. *suis* and *B. longum* subsp. *suillum* as closest subspecies. Interestingly, these strains are present in weaning infants and primates as opposed to their closest relatives which have been typically isolated from pig and calves. These strains, similarly to *B. longum* subsp. *infantis*, show a common capacity to metabolize the human milk oligosaccharide 3-fucosyllactose. Moreover, they harbour a riboflavin synthesis operon, which differentiate them from their closest subspecies, *B. longum* subsp. *suis* and *B. longum* subsp. *suillum*. Based on the consistent results from genotypical, ecological and phenotypical analyses, a novel subspecies with the name *Bifidobacterium longum* subsp. *iuvenis*, with type strain NCC 5000^T (=LMG 32752^T=CCOS 2034^T), is proposed.

DATA SUMMARY

Supplementary material are available at: <https://doi.org/10.6084/m9.figshare.23537901> [1].

INTRODUCTION

The animal and human gut microbiota harbour a complex microbial community where bifidobacteria are considered one of the key beneficial groups contributing substantially to the health of the host [2]. The genus *Bifidobacterium* belongs to the family *Bifidobacteriaceae* and comprises, up to date, 103 species, of which 12 are typically found in humans [3]. Among them, *Bifidobacterium longum* is one of the most abundant species in the human intestinal microbiota [4], and is one of the

Author affiliations: ¹Department of Agricultural and Food Sciences, University of Bologna, Viale Fanin 42, 40127 Bologna, Italy; ²Nestlé Research, Société des Produits Nestlé SA, 1000 Lausanne 26, Switzerland; ³International Centre for Diarrhoeal Disease Research, Bangladesh (icddr, b), Dhaka, Bangladesh.

***Correspondence:** Paola Mattarelli, paola.mattarelli@unibo.it

Keywords: *Bifidobacterium longum*; subspeciation; MLSA; cgMLST; HMO.

Abbreviations: AMR, anti microbial resistance; AN, accession number; ANI, average nucleotide identity; CDS, coding DNA sequence; cgMLST, core genome multilocus sequence typing; COGs, cluster of orthologous genes; dDDH, digital DNA–DNA hybridization; diFL, diFucosyllactose; 2'FL, 2'-Fucosyllactose; 3'FL, 3'-Fucosyllactose; GGDC, genome-to-genome distance calculator; HCCA, α -4-cyano-hydroxycinnamic acid; HMOs, human milk oligosaccharides; LNT, neutral HMO lacto-N-TETRAOSE; LNT, lacto-N-tetraose; MALDI-TOF/MS, matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry; MLSA, multi-locus sequence analysis; mMRS, modified de Man–Rogosa–Sharpe; MSPs, main spectras profiles; OGRI, Overall Genome Related Index; PAGP, pan-genomes analysis pipeline; UFBoot, ultrafast approximation for phylogenetic bootstrap; UPGMA, unweighted pair group method analysis; wgMLST, whole genome multilocus sequence typing.

The GenBank (EMBL) accession numbers for the 16S rRNA gene sequence and genome of *Bifidobacterium longum* subsp. *iuvenis* NCC 5000^T are OP696622 (GenBank) and Ga0527908 (JGI), respectively.

Eight supplementary figures, 10 supplementary tables and one supplementary file are available with the online version of this article.

006013 © 2023 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

most studied. This species comprises four validly named subspecies, all described by means of both genotypic and phenotypic features [5, 6]: *B. longum* subsp. *infantis* and *B. longum* subsp. *longum*, found in infants and in adults respectively, and *B. longum* subsp. *suis* and *B. longum* subsp. *suillum*, mainly found in pigs. When considering the currently described subspecies of *B. longum*, *B. longum* subsp. *infantis* is the only one with a conserved ability to metabolize human milk oligosaccharides (HMOs) beyond the neutral lacto-*N*-tetraose (LNT) [4], with strains showing a preference for sialylated or fucosylated HMOs [7].

Guidelines for a bacterial species delineation have been well-defined. Average nucleotide identity (ANI) and digital DNA–DNA hybridization (dDDH) have been most widely used, with proposed and generally accepted species boundary values of 95–96 and 70% for ANI and dDDH, respectively. These values, as forms of similarity or distance, were described as the overall genome related index (OGRI) [8].

Conversely, up to now, there are no validly accepted general guidelines for subspecies description, even though for some bacterial groups, a specific cut off has been suggested. Indeed, the ANI threshold value for *Mycobacterium* subspecies delineation has been fixed to 98% [9] whereas an ANI value of 98.4% has been applied for differentiating the novel subspecies *Carnobacterium inhibens* subsp. *gilichinskyi* subsp. nov. [10]. Furthermore, Chun *et al.* [8] suggested the use of the following criteria for novel subspecies description: OGRI between subspecies should be higher than the species level cut-off, strains belonging to different subspecies should be genomically coherent so they should form differentiated clades in the phylogenomic treeing, and finally they should share a sufficient number of phenotypes together with similar host specificity. All these features are the rationale for defining a putative novel subspecies. Moreover, a comprehensive collection of strains is essential to describe a consistent and distinct novel subspecies. According to Pearce *et al.* [11], groups of three or more isolates are needed to ensure that a real subspecies is present. For pathogenic bacteria, the concept of subspecies, other than genotypic clustering, is strictly linked to distinct pathogenic characteristics (mainly determined by different serotypes) and host preference [12]. For non-pathogenic bacteria, robust phenotypic differences have a strong significance for all subspecies description [13].

Over recent years, increasing evidence has emerged showing that some strains included in the grouping of *B. longum* subsp. *suis* and *B. longum* subsp. *suillum* did not cluster with the known clades. Indeed, several studies have highlighted in the *B. longum* core genome tree the presence of a separate clade formed by strains UCD 399, isolated from Rhesus macaque, BSM11-5 isolated from a Kenyan infant [14], CMCC P0001, isolate from a healthy Chinese children, and two other strains closely related to it, respectively BXY01 and JDM 301 (Table S1, available in the online version of this article), thus suggesting the presence of a putative novel subspecies [14–18].

In human beings there is a characteristic ecological occurrence of bifidobacteria at different ages, with *Bifidobacterium breve* and *B. longum* subsp. *infantis* distributed mainly in infants, due to their ability to utilize HMOs as a primary food source [19, 20]. Recently, metagenomic and metabolomic analysis of a longitudinal, community-based cohort study following 267 Bangladeshi infants from birth up to 2 years of age [21] confirmed the high occurrence of *B. longum* subsp. *infantis*, up to 1 year of age [22, 23]. This study furthermore detected a novel clade in *B. longum* (named ‘transitional’ *B. longum*), closest to *B. longum* subsp. *suis* and *B. longum* subsp. *suillum*, which expanded from 10 month onwards during the weaning phase, when solid foods are introduced in combination with continued breastfeeding. The study revealed that this clade possessed a conserved genetic setup to utilize both human milk and diet derived fibres as its energy source. To confirm the presence of these metabolomic traits, five novel *B. longum* strains were obtained from distinct preserved stool samples and were deposited in the Nestlé Culture Collection (NCC) as NCC 5000, NCC 5001, NCC 5002, NCC 5003 and NCC 5004. After *de novo* genome sequencing, comparison to existing genomes of *B. longum* revealed that these novel strains clustered separately from *B. longum* subsp. *infantis* and *B. longum* subsp. *longum*, (ANI values \leq 96.3%). The novel genomes were closely related to the BSM11-5 and CMCC P001 strains while showing similarities to the *B. longum* subsp. *suis* and *B. longum* subsp. *suillum* genomes (ANI \geq 98.1%) [23], which to date have been isolated mainly from livestock animals [6, 24].

In the present work, we aimed at clarifying the phylogenetic position of this group of strains including their classification as a putative novel subspecies of *B. longum*. For that purpose, their genome sequences together with those of selected publicly available *B. longum* strains belonging to the different known subspecies, were subjected to in-depth comparative genome analysis. Furthermore, the taxonomic position and host adaptation of this group of strains were studied using polyphasic approach comprising genotypic and phenotypic analysis.

ECOLOGY AND ISOLATION

Bacterial strains and growth conditions

A total of 24 strains were tested for phenotypic features (highlighted in bold in Tables 1 and S1). Seven strains of the putative novel subspecies were tested together with a selection of strains covering the diversity encompassed within the different *B. longum* subspecies: five strains of *B. longum* subsp. *longum*, six strains of *B. longum* subsp. *infantis*, four strains of *B. longum*

Table 1. Genomic and phylogenetic features of strains belonging to the putative novel subspecies

Strains highlighted in bold were used for phenotypic comparisons.

Strain	Isolated from	Country of origin	Accession no.	Contigs	Genome size (Mb)	G+C content (mol%)	No. of CDS	ORF	tRNA	rRNA
NCC 5000^T	Infant faeces	Bangladesh	Ga0527908 (JGI)	1	2504997	59.76	2136	2066	55	4, 4, 4 (5S, 16S, 23S)
NCC 5001	Infant faeces	Bangladesh	Ga0529016 (JGI)	1	2487584	59.69	2186	2110	61	4, 4, 4 (5S, 16S, 23S)
NCC 5002	Infant faeces	Bangladesh	Ga0529017 (JGI)	1	2518534	59.84	2169	2097	57	4, 4, 4 (5S, 16S, 23S)
NCC 5003	Infant faeces	Bangladesh	Ga0529018 (JGI)	1	2489079	59.79	2111	2040	56	4, 4, 4 (5S, 16S, 23S)
NCC 5004	Infant faeces	Bangladesh	Ga0529019 (JGI)	1	2515591	59.74	2117	2045	57	4, 4, 4 (5S, 16S, 23S)
BSM 11-5	Infant faeces, 6 m old	Kenya	MOAE00000000.1	45	2602875	58.94	2226	2163	51	3, 3, 3 (5S, 16S, 23S)
UCD399	Rhesus macaque	Unknown	2503754041 (JGI GOLD TaxonID)	175	2604157	59.79	2288	2235	47	1, 1, 1 (5S, 16S, 23S)
CMCC P0001	Human gut	China	NZ_KE145390.1	72	2424349	59.62	2066	2006	54	2, 1 (16S, 23S)
JDM301	Human gut	China	CP002010.1	1	2477838	59.81	2083	2016	55	3, 3, 3 (5S, 16S, 23S)
BXY01	Faeces of healthy children	China	CP008885.1	1	2480603	59.82	2091	2024	55	3, 3, 3 (5S, 16S, 23S)

subsp. *suillum*, two strains of *B. longum* subsp. *suis*, two strains closely related to both *B. longum* subsp. *suis* and *B. longum* subsp. *suillum*, named *B. longum* species. All strains were cultivated anaerobically in modified de Man–Rogosa–Sharpe (mMRS; Difco) medium supplemented with cysteine hydrochloride (0.5 g l⁻¹), unless indicated otherwise.

GENOMIC ANALYSIS AND PHYLOGENY

Bifidobacterial genome sequencing and assembly

For the comparison of genetic features, a total of 49 genomes were selected (Tables 1 and S1). Publicly available genomes (completed and draft genomes sequences) were chosen to represent the diversity within the *B. longum* subspecies (up to 10 strains for each subspecies) according to previous papers [15, 16, 25, 26]. Furthermore, the genome sequences of strains NCC 5000–NCC 5004 recently isolated from Vatanen *et al.* [23] belonging to the novel putative subspecies were included in this study (Table 1).

To complement the diversity of publicly available genomes, we sequenced 12 additional *B. longum* strains (Table S1). The DNA extraction method, the sequencing technology and the evaluation of completeness and the contamination of genomes are described in the supplementary file.

The G+C content of all selected *B. longum* genomes ranged from 58.39 to 61.20 mol% and the genome size ranged from 2.08 to 2.20 Mbp. Interestingly, the genome sizes of strains belonging to the putative novel subspecies were significantly higher than those of *B. longum* subsp. *longum*, *B. longum* subsp. *suis*, *B. longum* subsp. *suillum* and *B. longum* species), but were significantly smaller than those of *B. longum* subsp. *infantis* (Fig. S1).

The complete 16S rRNA sequences extracted from genomes were first aligned using ClustalW [27]. The obtained alignment was used to construct a Randomized Axelerated Maximum Likelihood tree with a bootstrap calculation based on 1000 iterations. This analysis was insufficient to resolve the different *B. longum* subspecies (Fig. S2), therefore pairwise ANI and dDDH analysis were performed.

ANI was computed using the orthoani version 1.2 software [28]. The generated matrix of pairwise genome similarities (Table S2) was further used to build a UPGMA phylogenetic tree using the BioNumerics software (version 8.0, bioMérieux) (Fig. 1). Pairwise dDDH values were estimated using the Genome-to-Genome Distance Calculator (GGDC; version 2.1) and formula 2. The GGDC is the most accurate web service for calculating DDH-analogous values, available at www.ggdc.dsmz.de. A matrix of dDDH values has been constructed (Table S3).

Results obtained from both ANI and dDDH revealed the presence of six distinct clusters: putative novel subspecies, *B. longum* subsp. *infantis*, *B. longum* subsp. *longum*, *B. longum* subsp. *suis*, *B. longum* subsp. *suillum*, and a putative new intermediate clade *B. longum* species (Figs 1 and S3). ANI and dDDH values of strains belonging to different *B. longum* lineages were from 95.6 to 98.7% and from 67.7 to 91.3%, respectively, indicating genetic dissimilarity (Tables S2 and S3). Particularly, the putative novel

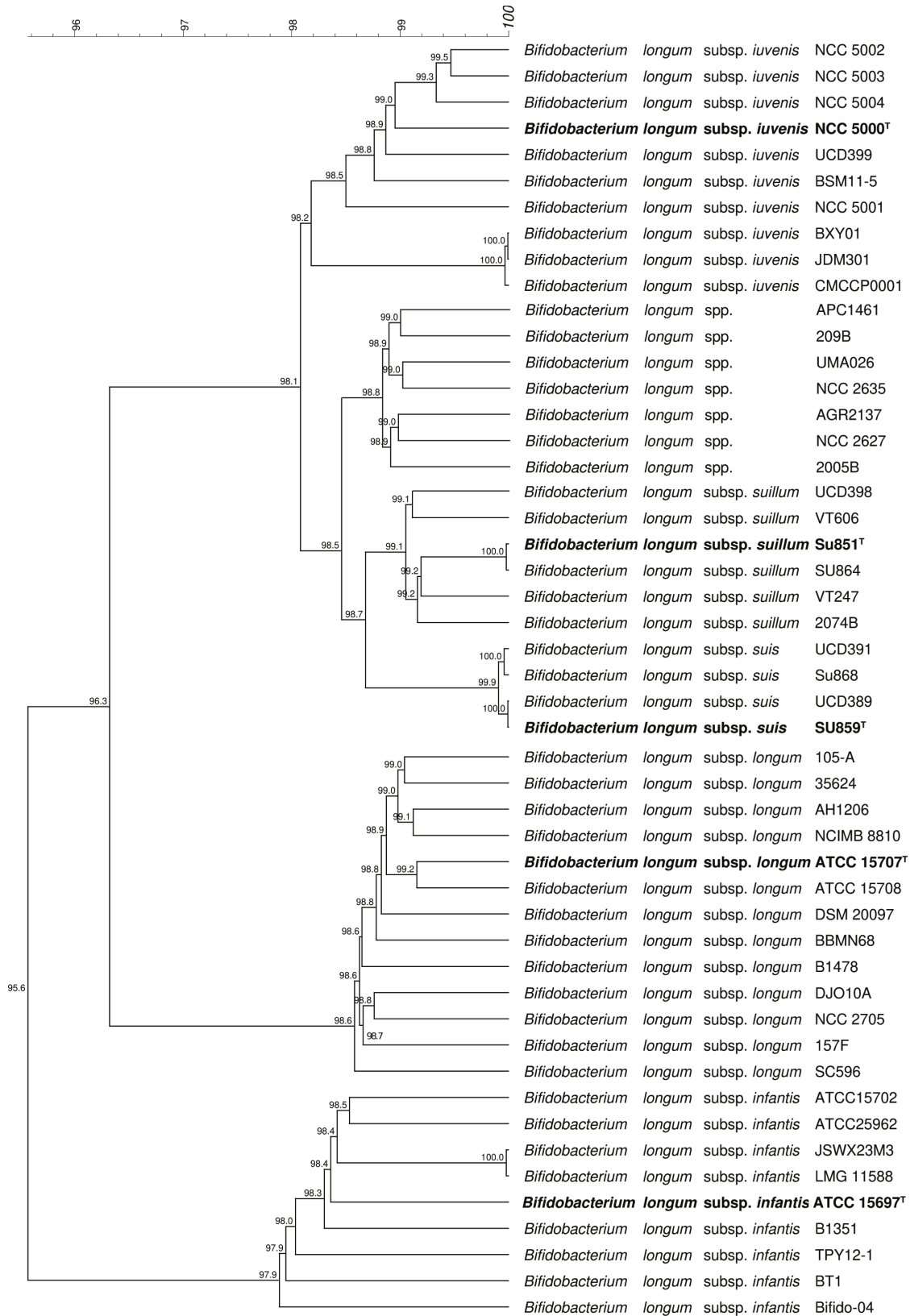


Fig. 1. Whole-genome average nucleotide identity (ANI) based UPGMA tree of selected *B. longum* genomes, including genomes of strains belonging to the newly proposed *B. longum* subsp. *iuvenis* subspecies.

subspecies showed an intersubspecies ANI and dDDH values of 98.1 and 86.6%, respectively, with the closest relatives, *B. longum* subsp. *suis* and *B. longum* subsp. *suillum*.

Core genome-based phylogenetic analysis was also performed using the 49 *B. longum* strains and *Bifidobacterium breve* DSM 20213^T (GCF_000158015) as an outgroup with the PGAP [29]. Determination of the cluster of orthologous genes (COGs) was performed using the Pan Genome Analysis Toolkit [29] (threshold of 0.9 on sequence identity and 0.95 on coverage) and resulted in 479 COGs. Each single COG sequence was first aligned with MAFFT [30], and then concatenated to build a joined alignment which was used for subsequent phylogenetic tree reconstruction using IQ-TREE [31], ModelFinder [32] and UFBoot [33]. The maximum-likelihood tree was reconstructed using previously described parameters [34, 35]. The topology of the obtained tree was proven to be robust as it was not modified using less stringent COG detection thresholds. This core-gene-based phylogenetic tree placed seven out of 10 strains of the putative novel subspecies in a separate clade, while the three clonally related strains CMCCP001, JCM301 and BXY01 clustered separately but in a branch outside of the groups comprising *B. longum* subsp. *suis*, *B. longum* subsp. *suillum*, *B. longum* species and the other seven strains of the putative novel subspecies (Fig. S4).

To gain further insight into the genetic dissimilarities and evolutionary relationships among strains belonging to the putative novel subspecies, a multi-locus sequence analysis (MLSA) of housekeeping-gene sequences was carried out. MLSA can be viewed as an alternative to DDH with promising potential for its efficiency of inter- and intra-species resolution and reproducibility [36]. The phylogenetic location of the selected *B. longum* strains was verified by the analysis of 10 selected genetic markers that have been shown to be discriminative for the classification of the genus *Bifidobacterium* [6, 37]. For this purpose, the ten selected housekeeping-gene sequences were concatenated in the following order, yielding approximately 6314 bp of sequences: *clpC* (646 bp), *dnaG* (892 bp), *dnaJ* (362 bp), *hsp60* (420 bp), *purF* (762 bp), *recA* (690 bp), *rpoB* (315 bp), *rpoC* (1093 bp), *tuf* (948 bp) and *xfp* (186 bp). The resulting concatenated sequences were then aligned and used to reconstruct a maximum-likelihood phylogenetic tree with IQ-TREE 2 (version 2.1.3) using the software's extended model selection followed by tree inference with 1000 non-parametric bootstrap replicates. The MLSA-based tree was visualized using the graphical viewer of iTOL (<https://itol.embl.de/upload.cgi>). MLSA-based results confirmed the presence of a putative novel subspecies clade formed by the strains depicted in Table 1 (Fig. S5).

Furthermore, a NeighborNet analysis [38] was also performed based on the obtained concatenated sequences from the 10 genes with the SplitsTree6 program [39]. Sequence similarities were corrected by using the Log Det method [40]. Both sets of MLSA-based results described the presence of four major clusters: cluster A, *B. longum* subsp. *longum* strains; cluster B, *B. longum* subsp. *infantis*; cluster C, *B. longum* subsp. *suis*, *B. longum* subsp. *suillum* and *B. longum* species; cluster D, putative novel subspecies (Figs 2 and S5).

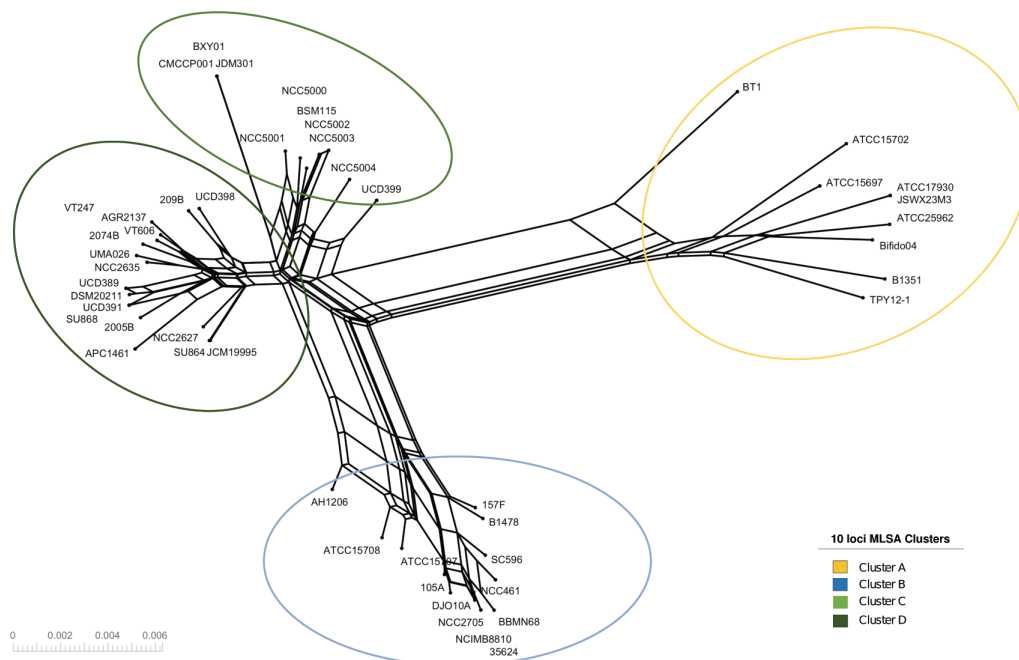


Fig. 2. Concatenated split network tree based on 10 selected housekeeping gene loci. The *clpC*, *dnaG*, *dnaJ*, *hsp60*, *purF*, *recA*, *rpoB*, *rpoC*, *tuf* and *xfp* gene sequences from *B. longum* selected strains were concatenated and phylogenetic tree was generated using the SplitsTree 6 program.

To confirm the taxonomic position of the putative novel subspecies, whole genome multilocus sequence typing (wgMLST) analysis was performed. wgMLST is used to analyse intraspecific diversity, as the distance matrix computed on a large set of loci spread over the genome provides it a high discriminatory capacity [41]. wgMLST was conducted using the chewBBACA (BSR-Based Allele Calling Algorithm) software [42] with default parameters. In chewBBACA, multilocus schemas are composed of loci defined by CDSs and all the called alleles of a given locus are CDSs as defined by Prodigal. Therefore, a very important step to ensure the reproducibility of the allele calls is the use of a Prodigal training file. Due to the absence of a publicly available training file for *Bifidobacterium* species in the software's repository at the time of the analysis, four distinct training files have been created on the genomes of strains *B. longum* subsp. *suillum* JCM 19995^T, *B. longum* subsp. *longum* ATCC 15707^T, *B. longum* subsp. *infantis* ATCC 15697^T and *B. longum* subsp. *suus* DSM 20211^T, using Prodigal (version 2.6.3), according to software's documentation [42]. The resulting allele schemas have been transformed, according to the documentation of the software to account for inferred alleles and missing data calls, and the resulting allele profiles have been used to produce the distance matrices using the program mlst2dist (<https://github.com/tripitakit/mlst2dist>), which implements the 'Hamming distance algorithm' with correction for missing data [43]. The number of loci in each schema and the number of paralog loci detected in the allele-calling step are described in the Table S4. The wgMLST alleles table and the resulting distance matrix obtained using *B. longum* subsp. *infantis* as training file are described in Tables S5 and S6, respectively. A phylogenetic tree was reconstructed from the distance matrix using the neighbour-joining clustering method in MEGA11 software [44] (Fig. 3). The results of wgMLST analysis confirmed the presence of the putative novel subspecies together with other five clearly distinct phylogenetic clades (Fig. 3), which is consistent with previous findings [14–16]: *B. longum* subsp. *infantis*, *B. longum* subsp. *longum*, *B. longum* subsp. *suus*, *B. longum* subsp. *suillum* and a new putative intermediate clade *B. longum* species (Fig. 3). The tree topology was not modified when other genomes such as *B. longum* subsp. *longum* ATCC 15707^T, *B. longum* subsp. *suillum* JCM 19995^T or *B. longum* subsp. *suus* DSM 20211^T were used as training files, confirming a robust taxonomic position for the putative novel subspecies (data not shown). Our obtained data did not consistently support the presence of a new subspecies formed by strains belonging to *B. longum* species, requiring further investigations on strains belonging to this phylogenetic group.

COMPARATIVE METABOLIC ANALYSIS AND PHENOTYPIC EVIDENCE OF HMO UTILIZATION

Urease *in silico* analysis

The presence of a urease-encoding gene was evaluated in all selected *B. longum* genomes. For that purpose, the sequence of the gene encoding the urease subunit alpha (*ureC*; *Blon_0111*) was compared by BLAST in the BioNumerics software (v8.0, bioMérieux) against all selected genome sequences. Analysis revealed the presence of a *ureC* homologue in all strains of *B. longum* subsp. *infantis* and *B. longum* subsp. *suus*. It was shown to be absent from strains of *B. longum* subsp. *longum* and *B. longum* subsp. *suillum*. The presence of an *ureC* homologue was found to be heterogenous in strains belonging to the putative novel subspecies, being present in 6 out of the nine strains evaluated (Table S7).

Unique functional features identified by whole genome-based analysis

Whole genome based alignment was performed in the BioNumerics software (v8.0, bioMérieux) using the NCC 5000^T genome as a reference. Only genes from other genomes with identity and coverage equal or greater than 60 and 70%, respectively, were considered present. In addition, similar genes from a compared genome had to occur in a same synteny to be considered as homologous. This analysis enabled to identify two genomic regions uniquely harboured by all strains of the putative novel subspecies.

A first region harboured six genes uniquely present in all putative novel subspecies and *B. longum* subsp. *infantis* strains. This region encompassed two genes (*PabA* and *PabB*) implicated in the conversion of chorismic acid to 4-amino-4-deoxychorismate, a precursor of *p*-aminobenzoate and folic acid in different micro-organisms. In addition, this region harboured another set of four genes (*RibD*, *RibE*, *RibAB*, *RibH*) implicated in the biosynthesis of riboflavin, as recently demonstrated in *B. longum* subsp. *infantis* [45] (Fig. S6a and b).

The same analysis revealed the presence of a fucosylated HMO utilization gene cluster in all the putative novel subspecies strains, similar to the one previously described in *B. longum* subsp. *infantis* strains [46] and in only one *B. longum* subsp. *longum* strain, respectively SC596 [47]. More specifically, a GH29 enzyme was predicted to be present in all strains of the putative novel subspecies and absent from the closely related *B. longum* subsp. *suus*, *B. longum* subsp. *suillum* and the novel putative intermediate clade of *B. longum* species strains. Interestingly, the gene coding for this glucosyl hydrolase was contained within an operon predicted to be responsible for fucosylated HMO degradation and metabolism. This operon contained two α -fucosidases (GH95 and GH29), an L-fucose mutarotase (*fumB*), an L-fucose dehydratase (*fumC*), an L-fuconate dehydratase (*fumE*) and an L-2-keto-deoxy-fuconate aldolase (*fumF*). It is interesting to note that the predicted GH95 and a GH29 α -fucosidase enzymes found in this cluster were previously demonstrated in *B. kashiwanohense* and *B. longum* subsp. *infantis* to possess specificity to 2'-fucosyllactose (2'FL)/3-fucosyllactose (3FL) and 3FL, respectively [48, 49]. The set of genes implicated in the metabolism of fucose was shown to be mostly absent from all selected *B. longum* subsp. *suus*, *B. longum*

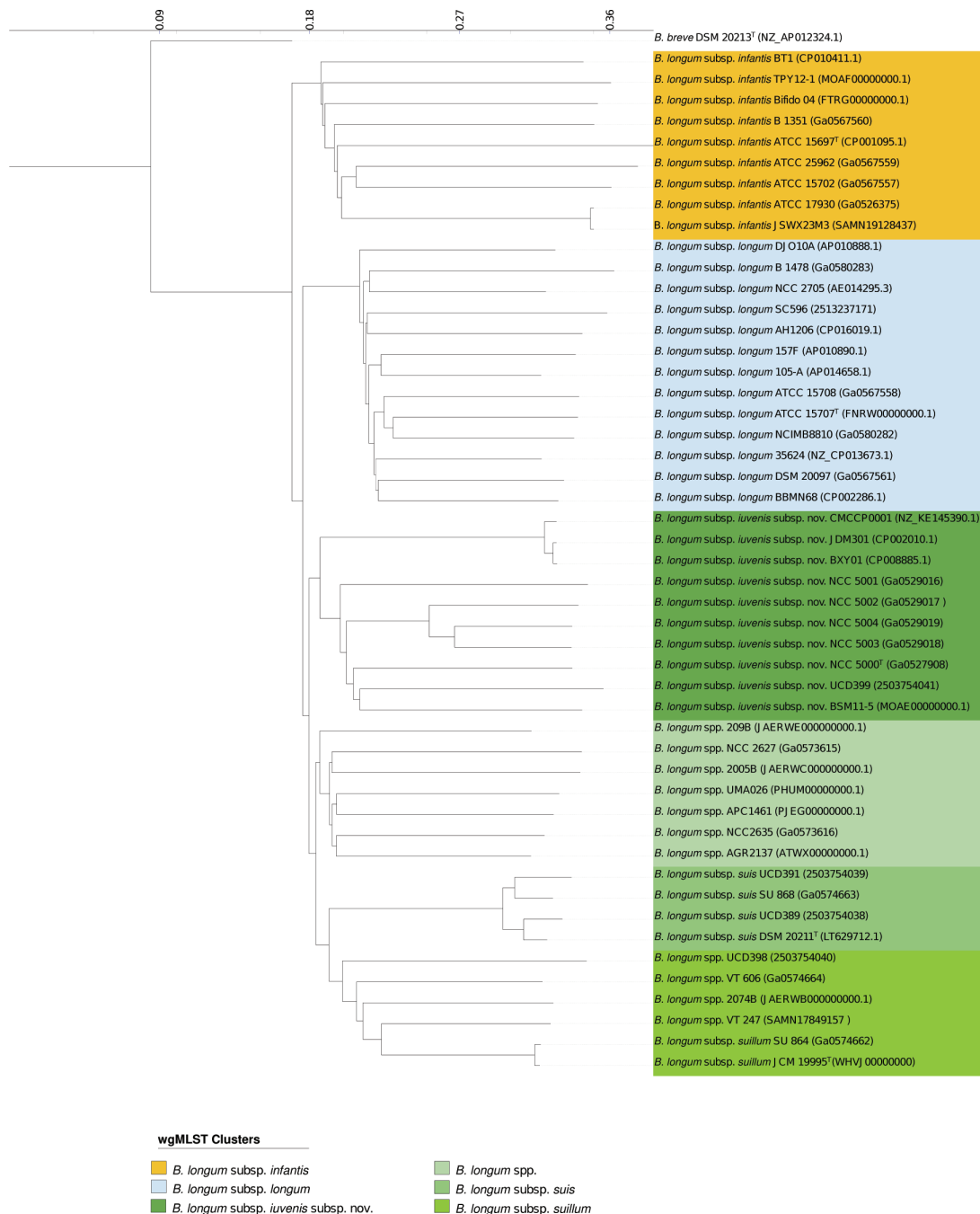


Fig. 3. Whole-genome multi-locus sequence typing (wgMLST) based dendrogram of selected *B. longum* genomes, including genomes of strains belonging to the newly proposed *B. longum* subsp. *iuvenis* subspecies. The phylogenetic tree was reconstructed using the neighbour-joining method from the distance matrix computed from the MLST alleles table, using Hamming distance modified with correction for missing data.

subsp. *suillum* and the novel putative intermediate clade of *B. longum* species strains with the only exception of NCC 2635 and NCC 2627. All strains belonging to the putative novel subspecies, except NCC 5001, harboured an ABC transport system predicted to import fucosylated HMO (*fumS*, *fumT*). However, the NCC 5001 strain harbours an HMO utilization operon highly similar to the one of *B. longum* subsp. *infantis*, containing a GH33 exo-sialidase as well as several ABC transporters for which the specificity has not been described to date [23]. Altogether, these results suggest that all strains of the putative novel subspecies have the genetic capacity to breakdown fucosylated HMOs and to metabolize its derived fucose into pyruvate and L-1,2-propanediol (Fig. 4a, b).

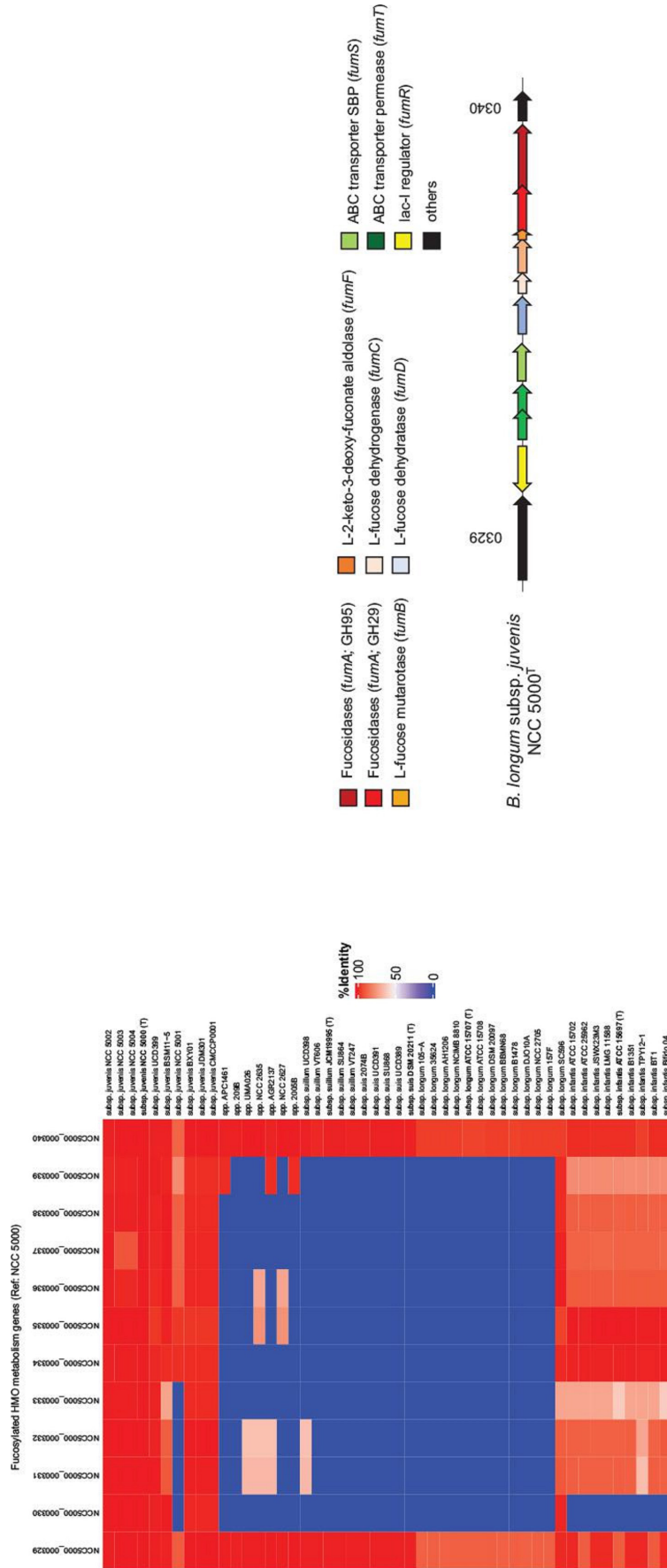


Fig. 4. Conservation (a) and organization (b) of fucosylated HMO metabolism operon found in strains belonging to the newly proposed *B. longum* subsp. *juvenis* subspecies. In panel (a), genes are coloured according to their sequence identity with genes found in *B. longum* subsp. *juvenis* NCC 5000^T. All values <60% identity rounded to 0 and appear as blue. All values greater than this threshold are coloured with a progressive scale ranging from light pink (60% identity) to red (100% identity). Panel (b) represents the gene organization found in *B. longum* subsp. *juvenis* NCC 5000^T.

Table 2. Growth profile of selected strains representing the diversity within the *B. longum* species on fucosylated HMOs

Species	Strain	Glucose	2'FL	3FL	diFL	LNT
<i>B. longum</i> subsp. <i>iuvenis</i>	NCC 5000 ^T	+++	++	++	+	++
<i>B. longum</i> subsp. <i>iuvenis</i>	NCC 5001	+++	–	+	–	++
<i>B. longum</i> subsp. <i>iuvenis</i>	NCC 5002	+++	+	++	+	++
<i>B. longum</i> subsp. <i>iuvenis</i>	NCC 5003	+++	++	+++	++	+++
<i>B. longum</i> subsp. <i>iuvenis</i>	NCC 5004	++	++	++	++	+++
<i>B. longum</i> subsp. <i>iuvenis</i>	BSM11-5	+++	++	+++	–	++
<i>B. longum</i> subsp. <i>iuvenis</i>	CMCCP0001	+++	+	+	+	++
<i>B. longum</i> subsp. <i>suillum</i>	JCM 19995 ^T	+++	–	–	–	++
<i>B. longum</i> subsp. <i>suillum</i>	VT 606	+++	–	–	–	++
<i>B. longum</i> subsp. <i>suillum</i>	SU 864	+++	–	–	–	+
<i>B. longum</i> subsp. <i>suis</i>	DSM 20211 ^T	+++	–	–	–	++
<i>B. longum</i> subsp. <i>suis</i>	SU 868	++	–	–	–	+
<i>B. longum</i> spp.	NCC 2627	+++	–	–	–	++
<i>B. longum</i> spp.	NCC 2635	+++	–	–	–	+
<i>B. longum</i> subsp. <i>longum</i>	NCC 2705	++	–	–	–	+
<i>B. longum</i> subsp. <i>longum</i>	ATCC 15707 ^T	+++	–	–	–	+
<i>B. longum</i> subsp. <i>longum</i>	ATCC 15708	+	–	–	–	++
<i>B. longum</i> subsp. <i>longum</i>	NCIMB 8810	++	–	–	–	++
<i>B. longum</i> subsp. <i>longum</i>	DSM 20097	+	–	–	–	++
<i>B. longum</i> subsp. <i>infantis</i>	LMG 11588	++	++	++	+	+
<i>B. longum</i> subsp. <i>infantis</i>	ATCC 15697 ^T	++	+	+	+	+++
<i>B. longum</i> subsp. <i>infantis</i>	ATCC 15702	+++	+	+	+	++
<i>B. longum</i> subsp. <i>infantis</i>	ATCC 25962	+	+	++	+	++
<i>B. longum</i> subsp. <i>infantis</i>	B 1351	++	+	+	+	++

2'FL, 2'-fucosyllactose; 3FL, 3-fucosyllactose; diFL, difucosyllactose; LNT, neutral HMO lacto-*N*-tetraose; –, No growth (OD <0.2); +, low growth (OD 0.2–0.5); ++, medium growth (OD 0.5–0.8); +++, high growth (OD >0.8).

Fucosylated HMO consumption

To determine the ability of the strains belonging to the putative novel subspecies to grow on fucosylated HMOs, all strains selected for phenotypical study (Tables 1 and S1) were grown on different fucosylated HMOs as sole carbon source, including 2'FL, 3FL, difucosyllactose (diFL) and the neutral HMO LNT (Glycom A/S). All carbohydrates were added at a concentration of 0.5% in an MRS-based medium containing no carbohydrates [50] to which 0.05% cysteine hydrochlorine was added. All strains were inoculated at an optical density measured at 600 nm (OD₆₀₀) of 0.1. Optical density was evaluated again after 48 h of incubation at 37 °C.

Similarly to *B. longum* subsp. *infantis* strains, all strains belonging to the putative novel subspecies had the capacity to grow on 3FL, an HMO which concentration rises with lactation time and becomes predominant during weaning [51]. Most strains belonging to this newly described subspecies grew to higher cell densities on 3FL than on 2'FL or diFL (Table 2). Furthermore, we could confirm that all *B. longum* strains had the capacity to grow on LNT, a feature commonly found within the *B. longum* species [52] (Table 2). Until now, within the *B. longum* species, the capacity to metabolize fucosylated HMOs was only described in *B. longum* subsp. *infantis* strains. Therefore the discovery of a similarly conserved ability in the strains belonging the putative novel subspecies, represents an important finding.

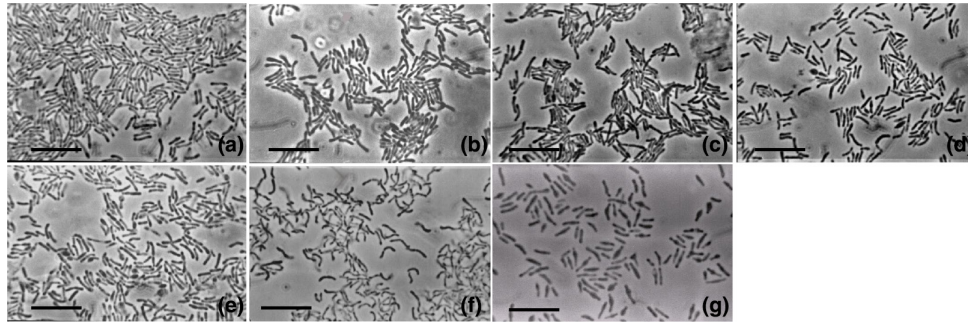


Fig. 5. Cell morphology of *B. longum* subsp. *iuvenis* NCC 5000^T (a), *B. longum* subsp. *iuvenis* NCC 5001 (b), *B. longum* subsp. *iuvenis* NCC 5002 (c), *B. longum* subsp. *iuvenis* NCC 5003 (d), *B. longum* subsp. *iuvenis* NCC 5004 (e), *B. longum* subsp. *iuvenis* CMCCP0001 (f); *B. longum* subsp. *iuvenis* BSM 11-5 (g). Cells of the strains were grown on mMRS agar plates for 48 h at 37 °C under anaerobic conditions. Phase-contrast photomicrograph. Bar, 10 mm.

Other carbohydrate consumption

To determine specific carbohydrate consumption patterns, all available strains belonging to the putative novel described subspecies, as well as strains representing the diversity found within the *B. longum* species, were subjected to an API50CH carbohydrate profiling by using the API 50CHL system (bioMérieux) where 0.05% of cysteine hydrochloride was added to the growth medium supplied.

All strains of the putative novel subspecies acidified the medium on the following carbohydrates: L-arabinose, ribose, D-xylose, galactose, D-glucose, D-mannose, maltose, lactose, melibiose, sucrose, raffinose and turanose (Table S8). Out of those 12 substrates, fucose was metabolized by all strains of putative novel subspecies. Most tested strains of *B. longum* subsp. *infantis* also shared this behaviour, while amongst all other subspecies, only *B. longum* species NCC 2627 displayed this capacity. It is worth noting that both strains isolated from cat (NCC 2627 and NCC 2635, respectively), were identified to possess genes involved in the fucose import and metabolism, while for an unknown reason, to date, only NCC 2627 displayed the phenotype.

Phenotypic characterization

Morphological, cultural and biochemical testing according to standard techniques was performed at 37 °C unless otherwise stated. Gram staining, oxidase and catalase activity were determined by using cells grown on mMRS agar at 37 °C for 2 days, according to Modesto *et al.* [53]. Motility was tested in modified mMRS soft agar (0.15%). Temperature and pH range of growth were tested according to Modesto *et al.* [54]. The ability of the strains to grow under anaerobic and microaerophilic conditions was tested using the GasPack EZ Anaerobic Pouch System (BD) and the Campygen (Oxoid) system, respectively.

The morphology of the cells belonging to the novel putative subspecies corresponds to pleomorphic bastoncellar rods, which occasionally show swollen coccid forms (Fig. 5).

All strains were Gram-stain-positive, catalase- and oxidase-negative, non-motile, and non-spore-forming. They also showed different temperatures and pH ranges of growth, as reported in Table S8. They are predicted to be more sensitive to oxygen than some recently described species such as *B. asteroides* [55]. Using the set of genes predicted to provide *B. asteroides* PRL011 an enhanced resistance to oxygen as reference, we could identify that amongst all described system, strains of *B. longum* subsp. *iuvenis* harboured genes belonging to Complex II (81–72% identity) and Complex IV (64–84% identity). On the other hand, no genes belonging to Complex I, Complex III nor a superoxide dismutase could be identified.

All the genomes belonging to the proposed novel subspecies *B. longum* subsp. *iuvenis* were evaluated for the presence of virulence and toxin producing genes, as well as genes encoding antibiotic resistances. The genomes were screened *in silico* both at nucleotide and protein levels. Virulence and toxin producing genes were searched by sequence similarity (BLAST) using the Virulence Factor Database full dataset downloaded on June 2022 (www.mgc.ac.cn/VFs/download.htm). Genes coding for potential antibiotic resistance were searched using two different software and their related databases: AMRFinder [56] and ResFinder-PointFinder [57]. No virulence nor toxin related genes were found in any of the genomes belonging to the proposed novel *B. longum* subsp. *iuvenis*. No antibiotic resistance gene was found to be common to all strains, However all genomes were predicted to have at least one gene conferring antibiotic resistance to at least one of the following: tetracyclin, trimethoprim, erythromycin and/or streptomycin (Table S9).

The whole cell protein profile was obtained using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF/MS) analysis, which was performed with an adapted method based on the supplier recommendations [58]. Fresh colonies of all strains were obtained anaerobically on mMRS agar supplemented with 0.05% of cysteine hydrochloride.

Bacterial material was resuspended in HPLC-grade water and pure ethanol was added, before being centrifuged (13000 r.p.m. for 2 min) [59]. The bacterial pellet was air-dried and a further extraction was performed in 50 µl (50% of 70% formic acid; 50% acetonitrile), followed by another centrifugation step. One microlitre of the obtained soluble protein extract was spotted and air-dried on the target plate before addition of 1 µl α -4-cyano-hydrocinnamic acid (HCCA). A minimum of 30 spectra obtained with the AutoXecute mode of the flexControl software (Bruker Daltonics) were selected with the flexAnalysis Software and then used to create main spectra profiles (MSPs) within the MBT Explorer Module (Bruker Daltonics GmbH). Those MSP were then used for species-level identification in the MBT Explorer Module (MBT Compass Library version 11) as well as to reconstruct a phylogenetic UPGMA tree in BioNumerics (Fig. S7).

A logarithmic score between 0 and 3 has been assigned to each spectrum describing the extent of peak matching. A score of 0 to 1.699 defines a situation with no reliable identification; a score between 1.700 and 1.999 defines probable genus identification; a score between 2.000 and 2.299 defines a secure genus identification and probable species identification; a score between 2.300 and 3.000 define highly probable species identification (Table S10). Results from this analysis are shown in Fig. S7. The analysis demonstrated species delineation, while no discrimination at subspecies level could be obtained. This finding confirms the usefulness of MALDI-TOF to be limited to species identification [59]. The protein profiles obtained are provided as supplementary material (Fig. S8).

Using a polyphasic approach including phenotypic and chemotaxonomic characterizations as well as multi-level phylogenetic analysis (i.e. MLSA, wgMLST), we could demonstrate that strains belonging to the previously described 'transitional' *B. longum* clade [23] were genetically and phenotypically discernible from the currently recognized subspecies of *B. longum*; thus, according to minimal standard guidelines [13], they represent a novel taxon for which the name *Bifidobacterium longum* subsp. *iuvenis* subsp. nov. (NCC 5000^T) is proposed.

DESCRIPTION OF *BIFIDOBACTERIUM LONGUM* SUBSP. *IUVENIS* SUBSP. NOV.

Bifidobacterium longum subsp. *iuvenis* (iu.'ve.nis. L. gen. sn. m. iuvenis, of child, because the type strain was isolated from a 2 year old Bangladeshi child).

Cells are Gram-stain-positive, non-motile, non-spore-forming rods, 0.4–0.5×1.5–2.5 µm, that usually occur singly. Colonies on mMRS agar 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 on mMRS agar is good under anaerobic and microaerophilic conditions and weak under aerobic condition. Cells can grow at 25–48°C. The optimum temperature is 37°C. Cells can grow at pH 3.5–8.5. The optimum pH is 5.5. Negative for catalase and oxidase production. Using the API 50 CHL system, acids are produced from L-arabinose, ribose, xylose, galactose, glucose, mannose, maltose, lactose, melibiose, sucrose, raffinose and turanose. Acids are produced weakly from fructose, methyl α -D-mannoside, methyl α -D-glucoside, salicin and L-fucose but not from other carbohydrates in the API50CH system.

Cells are able to utilize fucosylated HMOs, especially the fucosylated 3FL and the neutral LNT. Cells are able to produce riboflavin. Aesculin is not hydrolysed.

The type strain, NCC 5000^T (=LMG 32752^T=CCOS 2034^T), was isolated from the faeces of a Bangladeshi child. The DNA G+C content of the type strain is 59.76mol%.

The accession number for the 16S rRNA encoding gene and the genome of NCC 5000^T are OP696622 (Genbank) and Ga0527908 (JGI), respectively.

Funding information

This work received no specific grant from any funding agency.

Acknowledgements

We are thankful to Aliénor Genevaz for her support regarding the MALDI-TOF analysis. We thank Prof. Dr. Christophe Lacroix from the Eidgenössische Technische Hochschule (ETH) Zürich for providing the BSM 11-5 strain.

Conflicts of interest

C.Ngom-Bru, A. Bruttin, O. Sakwinska and S. Duboux are employed by Société des Produits Nestlé SA.

References

1. Modesto M, Ngom-Bru C, Scarafile D, Bruttin A, Pruvost S, et al. *Bifidobacterium longum* subsp. *iuvenis* subsp. nov., a novel subspecies isolated from the faeces of weaning infants. *Figshare*. 2023. DOI: 10.6084/m9.figshare.23537901.
2. Hidalgo-Cantabrana C, Delgado S, Ruiz L, Ruas-Madiedo P, Sánchez B, et al. Bifidobacteria and their health-promoting effects. *Microbiol Spectr* 2017;5.
3. Mattarelli P, Biavati B. The bifidobacteria and related organisms: biology, taxonomy, applications. In: Mattarelli P, Biavati B and

- BW WH (eds). *The Bifidobacteria and Related Organisms: Biology, Taxonomy, Applications*. Elsevier GmbH; 2018. pp. 9–60.
4. Tarracchini C, Milani C, Lugli GA, Mancabelli L, Fontana F, et al. Phylogenomic disentangling of the *Bifidobacterium longum* subsp. *infantis* taxon. *Microb Genom* 2021;7.
 5. 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* 2008;58:767–772.
 6. Yanokura E, Oki K, Makino H, Modesto M, Pot B, 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* 2015;38:305–314.
 7. Duboux S, Ngom-Bru C, De Bruyn F, Bogicevic B. Phylogenetic, functional and safety features of 1950s *B. infantis* strains. *Microorganisms* 2022;10:203.
 8. Chun J, Oren A, Ventosa A, Christensen H, Arahall DR, et al. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 2018;68:461–466.
 9. Tortoli E, Meehan CJ, Grottola A, Fregni Serpini G, Fabio A, et al. Genome-based taxonomic revision detects a number of synonymous taxa in the genus *Mycobacterium*. *Infect Genet Evol* 2019;75:103983.
 10. Nicholson WL, Zhalnina K, de Oliveira RR, Triplett EW. Proposal to rename *Carnobacterium inhibens* as *Carnobacterium inhibens* subsp. *inhibens* subsp. nov. and description of *Carnobacterium inhibens* subsp. *gilichinskyi* subsp. nov., a psychrotolerant bacterium isolated from Siberian permafrost. *Int J Syst Evol Microbiol* 2015;65:556–561.
 11. Pearce ME, Langridge GC, Lauer AC, Grant K, Maiden MCJ, et al. An evaluation of the species and subspecies of the genus *Salmonella* with whole genome sequence data: proposal of type strains and epithets for novel *S. enterica* subspecies VII, VIII, IX, X and XI. *Genomics* 2021;113:3152–3162.
 12. Tran QT, Han XY. Subspecies identification and significance of 257 clinical strains of *Mycobacterium avium*. *J Clin Microbiol* 2014;52:1201–1206.
 13. Mattarelli P, Holzapfel W, Franz CMAP, Endo A, Felis GE, et al. Recommended minimal standards for description of new taxa of the genera *Bifidobacterium*, *Lactobacillus* and related genera. *Int J Syst Evol Microbiol* 2014;64:1434–1451.
 14. O'Callaghan A, Bottacini F, O'Connell Motherway M, van Sinderen D. Pangenome analysis of *Bifidobacterium longum* and site-directed mutagenesis through by-pass of restriction-modification systems. *BMC Genomics* 2015;16:832.
 15. Arboleya S, Bottacini F, O'Connell-Motherway M, Ryan CA, Ross RP, et al. Gene-trait matching across the *Bifidobacterium longum* pangenome reveals considerable diversity in carbohydrate catabolism among human infant strains. *BMC Genomics* 2018;19:33.
 16. Li M, Zhou X, Stanton C, Ross RP, Zhao J, et al. Comparative genomics analyses reveal the differences between *B. longum* subsp. *infantis* and *B. longum* subsp. *longum* in carbohydrate utilisation, CRISPR-Cas systems and bacteriocin operons. *Microorganisms* 2021;9:1713.
 17. da Silva JGV, Vieira AT, Sousa TJ, Viana MVC, Parise D, et al. Comparative genomics and in silico gene evaluation involved in the probiotic potential of *Bifidobacterium longum* 5^{1A} Gene 2021;795:145781.
 18. 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* 2021;9:1906.
 19. Garrido D, Dallas DC, Mills DA. Consumption of human milk glycoconjugates by infant-associated bifidobacteria: mechanisms and implications. *Microbiology* 2013;159:649–664.
 20. Zivkovic AM, German JB, Lebrilla CB, Mills DA. Human milk glyco-biome and its impact on the infant gastrointestinal microbiota. *Proc Natl Acad Sci* 2011;108:4653–4658.
 21. Vidal K, Sultana S, Patron AP, Binia A, Rahman M, et al. Microbiota and health study: a prospective cohort of respiratory and diarrheal infections and associated risk factors in Bangladeshi infants under two years. *medRxiv* 2019:1–31. DOI: 10.1101/19000505.
 22. Underwood MA, German JB, Lebrilla CB, Mills DA. *Bifidobacterium longum* subspecies *infantis*: champion colonizer of the infant gut. *Pediatr Res* 2015;77:229–235.
 23. Vatanen T, Ang QY, Siegwald L, Sarker SA, Le Roy CI, et al. A distinct clade of *Bifidobacterium longum* in the gut of Bangladeshi children thrives during weaning. *Cell* 2022;185:4280–4297.
 24. Satti M, Modesto M, Endo A, Kawashima T, Mattarelli P, et al. Host-diet effect on the metabolism of *Bifidobacterium*. *Genes* 2021;12:609.
 25. Alessandri G, van Sinderen D, Ventura M. The genus *Bifidobacterium*: from genomics to functionality of an important component of the mammalian gut microbiota running title: bifidobacterial adaptation to and interaction with the host. *Comput Struct Biotechnol J* 2021;19:1472–1487.
 26. Albert K, Rani A, Sela DA. Comparative pangenomics of the mammalian gut commensal *Bifidobacterium longum*. *Microorganisms* 2019;8:7.
 27. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680.
 28. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek* 2017;110:1281–1286.
 29. Zhao Y, Wu J, Yang J, Sun S, Xiao J, et al. PGAP: pan-genomes analysis pipeline. *Bioinformatics* 2012;28:416–418.
 30. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;30:3059–3066.
 31. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;32:268–274.
 32. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
 33. Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol* 2013;30:1188–1195.
 34. Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol* 2008;25:1307–1320.
 35. Soubrier J, Steel M, Lee MSY, Der Sarkissian C, Guindon S, et al. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol Biol Evol* 2012;29:3345–3358.
 36. Rong X, Huang Y. Taxonomic evaluation of the *Streptomyces hygroscopicus* clade using multilocus sequence analysis and DNA-DNA hybridization, validating the MLSA scheme for systematics of the whole genus. *Syst Appl Microbiol* 2012;35:7–18.
 37. Bottacini F, Ventura M, van Sinderen D, O'Connell Motherway M. Diversity, ecology and intestinal function of bifidobacteria. *Microb Cell Fact* 2014;13.
 38. Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 2004;21:255–265.
 39. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 2006;23:254–267.
 40. Steel M. Recovering a tree from the leaf colourations it generates under a Markov model. *Appl Math Lett* 1994;7:19–23.
 41. Uelze L, Grütze J, Borowiak M, Hammerl JA, Juraschek K, et al. Typing methods based on whole genome sequencing data. *One Health Outlook* 2020;2:3.

42. Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, et al. chewBBACA: a complete suite for gene-by-gene schema creation and strain identification. *Microb Genom* 2018;4:e000166.
43. Galpern P, Manseau M, Hettinga P, Smith K, Wilson P. Allelematch: an R package for identifying unique multilocus genotypes where genotyping error and missing data may be present. *Mol Ecol Resour* 2012;12:771–778.
44. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Mol Biol Evol* 2021;38:3022–3027.
45. Solopova A, Bottacini F, Venturi Degli Esposti E, Amaretti A, Raimondi S, et al. Riboflavin biosynthesis and overproduction by a derivative of the human gut commensal *Bifidobacterium longum* subsp. *infantis* ATCC 15697. *Front Microbiol* 2020;11:573335.
46. LoCascio RG, Desai P, Sela DA, Weimer B, Mills DA. Broad conservation of milk utilization genes in *Bifidobacterium longum* subsp. *infantis* as revealed by comparative genomic hybridization. *Appl Environ Microbiol* 2010;76:7373–7381.
47. Garrido D, Ruiz-Moyano S, Kirmiz N, Davis JC, Totten SM, et al. A novel gene cluster allows preferential utilization of fucosylated milk oligosaccharides in *Bifidobacterium longum* subsp. *longum* SC596. *Sci Rep* 2016;6:35045.
48. Sela DA, Garrido D, Lerno L, Wu S, Tan K, et al. *Bifidobacterium longum* subsp. *infantis* ATCC 15697 α -fucosidases are active on fucosylated human milk oligosaccharides. *Appl Environ Microbiol* 2012;78:795–803.
49. James K, Bottacini F, Contreras JIS, Vigoureux M, Egan M, et al. Metabolism of the predominant human milk oligosaccharide fucosyllactose by an infant gut commensal. *Sci Rep* 2019;9:15427.
50. Duboux S, Golliard M, Muller JA, Bergonzelli G, Bolten CJ, et al. Carbohydrate-controlled serine protease inhibitor (serpin) production in *Bifidobacterium longum* subsp. *longum*. *Sci Rep* 2021;11:7236.
51. Plows JF, Berger PK, Jones RB, Alderete TL, Yonemitsu C, et al. Longitudinal changes in human milk oligosaccharides (HMOs) over the course of 24 months of lactation. *J Nutr* 2021;151:876–882.
52. Fushinobu S, Abou Hachem M. Structure and evolution of the bifidobacterial carbohydrate metabolism proteins and enzymes. *Biochem Soc Trans* 2021;49:563–578.
53. Modesto M, Michelini S, Stefanini I, Sandri C, Spiezio C, et al. *Bifidobacterium lemurum* sp. nov., from faeces of the ring-tailed lemur (*Lemur catta*). *Int J Syst Evol Microbiol* 2015;65:1726–1734.
54. Modesto M, Michelini S, Sansosti MC, De Filippo C, Cavalieri D, et al. *Bifidobacterium callitrichidarum* sp. nov. from the faeces of the emperor tamarin (*Saguinus imperator*). *Int J Syst Evol Microbiol* 2018;68:141–148.
55. Bottacini F, Milani C, Turrone F, Sánchez B, Foroni E, et al. *Bifidobacterium asteroides* PRL2011 genome analysis reveals clues for colonization of the insect gut. *PLoS One* 2012;7:e44229.
56. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother* 2019;63:e00483–19.
57. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J Antimicrob Chemother* 2020;75:3491–3500.
58. Bastin B, Bird P, Benzinger MJ, Crowley E, Agin J, et al. Confirmation and identification of *Salmonella* spp., *Cronobacter* spp., and other Gram-negative organisms by the Bruker MALDI biotyper method: collaborative study method extension to include *Campylobacter* species, revised first action 2017.09. *J AOAC Int* 2019;102:1595–1616.
59. Sánchez-Juanes F, Teixeira-Martín V, González-Buitrago JM, Velázquez E, Flores-Félix JD. Identification of species and subspecies of lactic acid bacteria present in Spanish cheeses type “Torta” by MALDI-TOF MS and *pheS* gene analyses. *Microorganisms* 2020;8:301.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org