**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 20210T and *B. longum* subsp. *suillum* JCM1995T (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 5000T (=LMG 32752T=CCOS 2034T), 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...
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 genometrically 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 an 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 ≥96.3%). The novel genomes were closely related to the BSM11-5 and CMCC P0001 strains while showing similarities to the B. longum subsp. suis and B. longum subsp. suillum genomes (ANI ≥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
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 orthoaniu 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

---

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

<table>
<thead>
<tr>
<th>Strain</th>
<th>Isolated from</th>
<th>Country of origin</th>
<th>Accession no.</th>
<th>Contigs</th>
<th>Genome size (Mb)</th>
<th>G+C content (mol%)</th>
<th>No. of CDS</th>
<th>ORF</th>
<th>tRNA</th>
<th>rRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCC 5000⁷</td>
<td>Infant faeces</td>
<td>Bangladesh</td>
<td>Ga0527908 (JGI)</td>
<td>1</td>
<td>2504997</td>
<td>59.76</td>
<td>2136</td>
<td>2066</td>
<td>55</td>
<td>4, 4, 4 (SS, 16S, 23S)</td>
</tr>
<tr>
<td>NCC 5001</td>
<td>Infant faeces</td>
<td>Bangladesh</td>
<td>Ga0529016 (JGI)</td>
<td>1</td>
<td>2487584</td>
<td>59.69</td>
<td>2186</td>
<td>2110</td>
<td>61</td>
<td>4, 4, 4 (SS, 16S, 23S)</td>
</tr>
<tr>
<td>NCC 5002</td>
<td>Infant faeces</td>
<td>Bangladesh</td>
<td>Ga0529017 (JGI)</td>
<td>1</td>
<td>2518534</td>
<td>59.84</td>
<td>2169</td>
<td>2097</td>
<td>57</td>
<td>4, 4, 4 (SS, 16S, 23S)</td>
</tr>
<tr>
<td>NCC 5003</td>
<td>Infant faeces</td>
<td>Bangladesh</td>
<td>Ga0529018 (JGI)</td>
<td>1</td>
<td>2489079</td>
<td>59.79</td>
<td>2111</td>
<td>2040</td>
<td>56</td>
<td>4, 4, 4 (SS, 16S, 23S)</td>
</tr>
<tr>
<td>NCC 5004</td>
<td>Infant faeces</td>
<td>Bangladesh</td>
<td>Ga0529019 (JGI)</td>
<td>1</td>
<td>2515591</td>
<td>59.74</td>
<td>2117</td>
<td>2045</td>
<td>57</td>
<td>4, 4, 4 (SS, 16S, 23S)</td>
</tr>
<tr>
<td>BSM 11–5</td>
<td>Infant faeces, 6 m old</td>
<td>Kenya</td>
<td>MOAE00000000.1</td>
<td>45</td>
<td>2602875</td>
<td>58.94</td>
<td>2226</td>
<td>2163</td>
<td>51</td>
<td>3, 3, 3 (SS, 16S, 23S)</td>
</tr>
<tr>
<td>UCD399</td>
<td>Rhesus macaque</td>
<td>Unknown</td>
<td>2503754041 (JGI GOLD TaxonID)</td>
<td>175</td>
<td>2604157</td>
<td>59.79</td>
<td>2288</td>
<td>2235</td>
<td>47</td>
<td>1, 1, 1 (SS, 16S, 23S)</td>
</tr>
<tr>
<td>CMCC P0001</td>
<td>Human gut</td>
<td>China</td>
<td>NZ_KE145390.1</td>
<td>72</td>
<td>2242349</td>
<td>59.62</td>
<td>2066</td>
<td>2006</td>
<td>54</td>
<td>2, 1 (16S, 23S)</td>
</tr>
<tr>
<td>JDM301</td>
<td>Human gut</td>
<td>China</td>
<td>CP002010.1</td>
<td>1</td>
<td>2477838</td>
<td>59.81</td>
<td>2083</td>
<td>2016</td>
<td>55</td>
<td>3, 3, 3 (SS, 16S, 23S)</td>
</tr>
<tr>
<td>BXY01</td>
<td>Faeces of healthy children</td>
<td>China</td>
<td>CP008885.1</td>
<td>1</td>
<td>2480603</td>
<td>59.82</td>
<td>2091</td>
<td>2024</td>
<td>55</td>
<td>3, 3, 3 (SS, 16S, 23S)</td>
</tr>
</tbody>
</table>
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. *iuvens* 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 20213T (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* and *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 (690 bp), recA (315 bp), rpoB (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.](image-url)
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. suis DSM 20211T, B. longum subsp. suis ATCC 15697T and B. longum subsp. suis DSM 20211T, 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/triptitak/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. suis, B. longum subsp. suis 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 15707T, B. longum subsp. suis DSM 20211T 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. suis. It was shown to be absent from strains of B. longum subsp. longum and B. longum subsp. suis. 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 5000T 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 (PubA and PubB) 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 harbour 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. suis, B. longum subsp. suis 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 metabolization. This operon contained two α-fucosidases (GH95 and GH29), an l-fucose mutarotase (fumB), an l-fucose dehydratase (fumC), an l-fucuronate dehydratase (fumE) and an 1,2-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 (2FL)/3′-fucosyllactose (3FL) and 3FL, respectively [48, 49]. The set of genes implicated in the metabolization of fucosylated HMO was shown to be mostly absent from all selected B. longum subsp. suis, B. longum
subsp. suillum and the novel putative intermediate clade of B. longum species strains with the only exception of 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 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\(^t\).
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 hydrochloride was added. All strains were inoculated at an optical density measured at 600 nm (OD 600) 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.

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Glucose</th>
<th>2′FL</th>
<th>3FL</th>
<th>diFL</th>
<th>LNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. longum subsp. iuvenis</td>
<td>NCC 5000</td>
<td>+++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. iuvenis</td>
<td>NCC 5001</td>
<td>+++</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. iuvenis</td>
<td>NCC 5002</td>
<td>+++</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. iuvenis</td>
<td>NCC 5003</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td>B. longum subsp. iuvenis</td>
<td>NCC 5004</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>B. longum subsp. iuvenis</td>
<td>BSM11-5</td>
<td>+++</td>
<td>++</td>
<td>+++</td>
<td>–</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. iuvenis</td>
<td>CMCCP0001</td>
<td>+++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. saillum</td>
<td>JCM 19995</td>
<td>+++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. saillum</td>
<td>VT 606</td>
<td>+++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. saillum</td>
<td>SU 864</td>
<td>+++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>B. longum subsp. suis</td>
<td>DSM 20211</td>
<td>+++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. suis</td>
<td>SU 868</td>
<td>++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>B. longum spp.</td>
<td>NCC 2627</td>
<td>+++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>++</td>
</tr>
<tr>
<td>B. longum spp.</td>
<td>NCC 2635</td>
<td>+++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>B. longum subsp. longum</td>
<td>NCC 2705</td>
<td>++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>B. longum subsp. longum</td>
<td>ATCC 15707</td>
<td>+++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>B. longum subsp. longum</td>
<td>ATCC 15708</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>B. longum subsp. longum</td>
<td>NCTM 8810</td>
<td>++</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>B. longum subsp. longum</td>
<td>DSM 20097</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>B. longum subsp. infantis</td>
<td>LMG 11588</td>
<td>++</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>B. longum subsp. infantis</td>
<td>ATCC 15697</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>B. longum subsp. infantis</td>
<td>ATCC 15702</td>
<td>+++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. infantis</td>
<td>ATCC 25962</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>B. longum subsp. infantis</td>
<td>B 1351</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>++</td>
</tr>
</tbody>
</table>

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 hydrochloride was added. All strains were inoculated at an optical density measured at 600 nm (OD 600) 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.
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) were 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 metabolization, 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 search 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 analysis 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 5000T) 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. Aesculín is not hydrolysed.

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

The accession number for the 16S rRNA encoding gene and the genome of NCC 5000T 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**


13. 18.
5.
4.
7.
5 1A in the probiotic potential of among human infant strains.
2021;9:1713.
subsp. mutagenesis through by- pass of restriction-modification systems.
Pangenome analysis of,

the genera,

22.

19.

18.

8.


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