REVISED AEM02037-15

Exploring vertical transmission of bifidobacteria from mother to child

Running title: origin of the gut microbiota

Key words: microbiota, bifidobacteria, metagenomics, genomics

Christian Milani[1], Leonardo Mancabelli[1], Gabriele Andrea Lugli[1], Sabrina Duranti[1], Francesca Turroni[2], Chiara Ferrario[1], Marta Mangifesta[3], Alice Viappiani[3], Pamela Ferretti[4], Valentina Gorfer[5], Adrian Tett[4], Nicola Segata[4], Douwe van Sinderen[2] and Marco Ventura[1]

Laboratory of Probiogenomics, Department of Life Sciences, University of Parma, Italy[1]; Alimentary Pharmabiotic Centre and Department of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland[2]; GenProbio, Parma, Italy[3]; Centre for Integrative Biology, University of Trento, Trento, Italy[4], Azienda Provinciale per i Servizi Sanitari, Trento, Italy[5]

Corresponding author. Mailing address for Marco Ventura Department of Life Sciences, University of Parma, Parco Area delle Scienze 11a, 43124 Parma, Italy. Phone: ++39-521-905666. Fax: ++39-521-905604. E-mail: marco.ventura@unipr.it

Abstract

Passage through the birth canal and consequent exposure to the mother's microbiota is considered to represent the initiating event for microbial colonization of the gastro-intestinal tract of the new-born. However, a precise evaluation of such suspected vertical microbiota transmission is yet to be performed. Here, we evaluated the microbiomes of four sample sets, each consisting of a mother's fecal and milk sample, and corresponding infant fecal sample, by means of amplicon-based profiling supported by shotgun metagenomics data of two key samples. Notably, targeted genome reconstruction from microbiome data revealed vertical transmission of a *Bifidobacterium breve* and *Bifidobacterium longum* subsp. *longum* strain from mother to infant, a notion confirmed by strain isolation and genome sequencing. Furthermore, PCR analyses targeting unique genes from these two strains highlighted their persistence in the infant gut at six months. Thus, this study demonstrates the existence of specific bifidobacterial strains that are common between mother and her child, thus indicative of vertical transmission, and that are maintained in the infant for at least relative short time spans.

INTRODUCTION

Human beings may be considered as superorganisms that consist of both eukaryotic and prokaryotic cells (1). It is generally accepted that not only the nuclear and mitochondrial genomes are transferred to the next generation, but that such transfer may also include certain members and associated genomes of the symbiotic community, whose microbial genes outnumber those of the eukaryotic host by over 100-fold (2). Passage through the birth canal together with breast-feeding represent very important transfer opportunities for symbionts from one generation to the next. Notably, this transmission is facilitated by the maternal holobiont, i.e. the organism together with its associated microbial communities, and the mother is both actively and passively engaged in providing a symbiotic and perhaps long-lasting microbial community to her offspring. However, each individual develops a specific microbial community by adulthood, suggesting that stochastic colonization is more important than direct vertical transmission. In fact, while individual strains may be directly transmitted, the majority of the long-lasting community is probably not. In mammals, it is generally accepted that the fetus develops in an essentially sterile environment within the amnion, and that bacterial colonization of the fetus is made possible as soon as the amnion breaks prior to its delivery through the birth canal (3). Bifidobacteria are amongst the first bacterial colonizers of the human gut, and have been subject to extensive scientific scrutiny (4, 5). It has become clear that certain species of the genus *Bifidobacterium*, e.g. *Bifidobacterium breve* and *Bifidobacterium bifidum*, are genetically adapted to colonize the infant gut [for a review see (6)]. Such bacteria have evolved genetic strategies that allow them to metabolize particular glycans present in human milk (7, 8) (9). However, human milk not only represents a reservoir of glycan compounds that significantly impact on the composition of the infant gut microbiota, but also appears to act as a repository of

64 bacteria for vertical transmission from mother to infant. In fact, the isolation of bifidobacteria
65 from human milk has been reported (9, 10), even though it is currently not known how
66 bifidobacteria reach this human bodily fluid (11). In addition, PCR-based approaches have
67 provided suggestive evidence for the occurrence of direct transmission of bifidobacterial strains
68 and other gut commensals from mother to newborn (12-14).
69 In this study, we investigated this notion of vertical transmission of bifidobacteria from mother
70 to newborn by assessing the gut/milk microbiomes of four mother-child sample sets, revealing
71 the existence of shared bifidobacterial strains.

MATERIALS AND METHODS

Subject recruitment and sample collection. The study is a pilot for a larger infant gut investigation protocol approved by the Ethical Committee of the University of Trento and informed written consent was obtained from the participating mothers. Four mother-infant pairs were enrolled in the study; on the same day, stool samples were collected from both 3-month-old infants and mothers, while a fresh breast milk sample was also collected. At the time of the first sampling, breast-feeding was exclusive for mother-infant pair 2 and 4, whereas 70 % and 30 % supplement of formula milk was given to the infant for pair 1 and 3, respectively. Sampling of stool and milk samples was repeated following weaning six months after the first sample collection. At the time of the second sampling, the infant of pair 4 was still receiving a 20 % dietary supplementation of breast milk from the mother (calculated as % of total calories). All subjects were healthy, which was established based on self-reporting, and had not received any antibiotic or probiotic in the previous month. Stool samples consisted of 6–10 gr of fresh fecal material, which was immediately frozen upon collection at −80°C until processed for DNA extraction. Human milk samples were collected as previously described (9) and were immediately subjected to DNA extraction.

Recovery of bifidobacteria on selective media. Fecal samples were pour-plated onto *Bifidobacterium* selective agar (BSM) for selective outgrowth of bifidobacteria. The BSM selective medium was prepared by the addition to MRS agar (Scharlau Chemie, Barcelona, Spain) of 0.05 % (w/v) L-Cysteine hydrochloride and 50 mg mupirocin (Delchimica, Italy) per litre of MRS as described previously (15). Agar plates were incubated in an anaerobic atmosphere (2.99 % H2, 17.01 % CO2, and 80 % N2) in a chamber (Concept 400; Ruskin) at 37°C for 24 h. Ten colonies were taken as an adequate representation of the major bacterial

strains cultured on a selective medium. DNA was extracted from each isolate through rapid mechanic cell lysis as described previously (15) and subjected to (sub)species identification through ITS amplification and DNA sequencing.

ITS and 16S rRNA gene amplification. Partial ITS sequences were amplified from extracted DNA using primer pair Probio-bif_Uni/Probio-bif_Rev (16). Partial 16S rRNA gene sequences were amplified from extracted DNA using primer pair Probio_Uni and/Probio_Rev, which targets the V3 region of the 16S rRNA gene sequence (17). These primers were designed in order to include, at their 5′ end, one of the two adaptor sequences used in the Ion Torrent-sequencing library preparation protocol linking a unique Tag barcode of 10 bases to identify different samples. The PCR conditions used were 5 min at 95°C, 35 cycles of 30 s at 94°C, 30 s at 55°C and 90 s at 72°C, followed by 10 min at 72°C. Amplification was carried out using a Veriti Thermocycler (Applied Biosystems). The integrity of the PCR amplicons was analyzed by electrophoresis on an Experion workstation (BioRad, UK).

Ion Torrent PGM sequencing of ITS and 16S rRNA Gene-based amplicons. PCR products obtained following amplification of the ITS and 16S rRNA gene sequences were purified by electrophoretic separation on an 2 % agarose gel and the use of a Wizard SV Gen PCR Clean-Up System (Promega), followed by a further purification step involving the Agencourt AMPure XP DNA purification beads (Beckman Coulter Genomics GmbH, Bernried, Germany) in order to remove primer dimers. DNA concentration of the amplified sequence library was estimated through the MultiTape system (Agilent). From the concentration and the average size of each amplicon library, the amount of these DNA fragments per microliter was calculated and libraries for each run were diluted to $3E^9$ DNA molecules prior to clonal amplification. Emulsion PCR was carried out using the Ion OneTouch™ 200 Template Kit v2 DL (Life Technologies)

according to the manufacturer's instructions. Sequencing of the amplicon libraries was carried out on a 316 chip using the Ion Torrent PGM system and employing the Ion Sequencing 200 kit (Life Technologies) according to the supplier's instructions at GenProbio srl (Parma, Italy). After sequencing, the individual sequence reads were filtered by PGM-provided software to remove low quality and polyclonal sequences. Sequences matching the PGM 3′ adaptor were also automatically trimmed. All PGM quality-approved, trimmed and filtered data sets were exported as BAM files.

ITS-based Microbiota Analysis. Fastq files obtained by sequencing of the ITS amplicons were analysed using a custom script, named bif_ITS_analysis.sh script available at (http://probiogenomics.unipr.it/sw/bif_ITS_analysis.zip). This script requires QIIME (18) to be installed (or works in a QIIME virtual machine) and accepts .bam or .fastq input files containing sequencing reads. Input data was processed as previously described (16). The trees were saved in Phylip format and each sequence was named including the name of the OTU and the number of sequences it represents. This script is easily modifiable in order to obtain a profiling based on a different marker sequence, as long as a reliable database is available for such a marker sequence.

Evaluation of *B. longum* subsp. *longum* BLOI2 and *B. breve* BBRI4 persistence. DNA extracted from stool or milk samples was submitted to PCR amplification using BLOI2_0636 fw (5'-GAACTTGAAGGGCTGCTGGA-3') and BLOI2_0636 rev (5'-CTCGGTCTTGAACTGTTCGA-3') or BBR14_0962 fw (5'-GTCTCCTCTACCCGAACCT) and BLOI2_0636 rev (5'-TCCTCGTTGATCCAATCCTC-3') specific primers. Each PCR mixture (25 µl) contained 1.5 mM of $MgCl_2$; 20 mM of Tris-HCl, 50 mM of KCl; 200 µM of each deoxynucleoside triphosphate; 25 pmol of each of the two primers; 1 U of Taq DNA polymerase (Taq PCR Master Mix Kit-QIAGEN, UK) and 50 ng of DNA template. Each PCR

cycling program consisted of an initial denaturation step of 3 min at 94°C followed by amplification for 35 cycles as follows: denaturation (30 sec at 94°C) annealing (30 sec at 56.5°C) and extension (1 min at 72°C). The PCR reaction was completed with a single elongation step (10 min at 72°C). The resulting amplicons were separated on a 0.8 % agarose gel followed by ethidium bromide staining.

16S rRNA gene-based Microbiota Analysis. The fastq files were processed using QIIME (18) as previously described (17). Quality control retained sequences with a length between 140 and 400 bp, mean sequence quality score >25, with truncation of a sequence at the first base if a low quality rolling 10 bp window was found. Sequences with mismatched primers were omitted.

Shotgun metagenomics. Sequencing libraries for samples SS2_infant and SS4_infant were prepared starting from the extracted DNA, as per manufacturer's instructions, using the Illumina Nextera-XT DNA kit (Illumina inc, San Diego, CA, USA), pooled, and sequenced on the Illumina HiSeq-2000 platform (100 bp paired end). We generated a total of 673,922 raw paired-end reads for the SS2_infant , and 4,950,848 raw paired-end reads for the SS4_infant samples.

Analysis of metagenomic datasets. The fastq outputs were filtered for reads with quality < 25 and presence of human DNA, as well as reads < 80 bp. Bases were also removed from the end of the reads until the average quality in a window of 5 bp was > 25. Taxonomic classification of SS2_infant and SS4_infant reads was obtained using RapSearch2 software (19) for sequence homology in the NCBI nr database, followed by data processing using MEGAN5 software (20).

Genome reconstruction of bacterial strains from shotgun metagenomic data. Fastq files of quality- and alien DNA-filtered metagenomics datasets were used for metagenomic assembly with spades assembler (21). Obtained contigs were subjected to ORF prediction using PRODIGAL software (22) and taxonomic classification using RapSearch2 software (19) for

Applied and Environmental Microbiology

homology searches in the NCBI nr database, allowing identification and collection of contigs encoding at least 40 % of ORFs attributed to a species of interest. These contigs were used as a backbone for mapping of the metagenomic reads. Reads that successfully mapped were collected in a fastq file and used as input for assembly with MIRA software (23). Generated contigs encompassing at least 40 % of ORFs attributed to a species of interest and with coverage diverging less than 33 % from the average were then ordered based on a reference genome. Protein-encoding ORFs were predicted using Prodigal (22) and assignment of protein function to predicted coding regions was performed using a custom script based on RapSearch2 software (19), PFAM database (24) and the non-redundant protein database provided by the National Center for Biotechnology Information. Additional bioinformatic analyses included: transfer RNA genes identification using tRNAscan-SE (25) and ribosomal RNA genes detection using RNAmmer (http://www.cbs.dtu.dk/services/RNAmmer/) and IS families finding using ISFinder (http://www-is.biotoul.fr/is.html). Attribution of ORFs to a specific COG family was performed by searching against the EggNog database (26). Prediction of ORFs putatively involved in HGT events was performed by COLOMBO software with sensitivity value set at 0.9 in order to assure maximum sensitivity.

Targeted genome sequencing and bioinformatics analyses. DNA extracted from the bifidobacterial isolates was subjected to whole genome sequencing using MiSeq (Illumina, UK) at GenProbio srl (Parma, Italy) following the supplier's protocol (Illumina, UK). Fastq files obtained from targeted genome sequencing of the isolated strains were used as input for assembly with MIRA software (23). Protein-encoding ORFs were predicted using Prodigal (22) and assignment of protein function to predicted coding regions was performed using a custom script based on RapSearch2 software (19), PFAM database (24) and the non-redundant protein

Applied and Environmental Microbiology

database provided by the National Center for Biotechnology Information. Whole genome alignments between contigs obtained from assembly of metagenomics and isolate datasets were obtained using MAUVE software (27) and the metagenomic genome as reference. Average nucleotide identity (ANI) values based on BLAST and mummer software were calculated using JSpecies (28).

Pan-genome and extraction of shared and unique genes. For all *B. breve* and *B. longum* genomes used in this study a pan-genome calculation was performed using the PGAP pipeline (29); the ORF content of all genomes was organized in functional gene clusters using the GF (Gene Family) method involving comparison of each protein against all other proteins using BLAST analysis (cut-off: E-value 1 x 10^{-4} and 50 % identity over at least 50 % of both protein sequences), followed by clustering into protein families using MCL (graph-theory-based Markov clustering algorithm) (30). Pan-genome profiles were built using all possible BLAST combinations for each genome being sequentially added. Following this, the unique protein families for each of the analyzed bifidobacterial genomes were classified. Protein families shared between all genomes where defined by selecting the families that contained at least one single protein member for each genome. Each set of orthologous proteins constituting core COGs with one member per genome were aligned using MAFFT (31) and phylogenetic trees were constructed using the Neighbor Joining method in Clustal W version 2.1 (32). The supertree was built using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

Data Deposition. The assembled *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 genomes were deposited under the accession numbers listed in Table S2. WGS data for SS2_Infant and SS4_Infant, as well as bifidobacterial ITS-profiling and 16S rRNA gene-profiling data for

Applied and Environmental Microbiology

SS1_T6m, SS2_T6m, SS3_T6m and SS4_T6m sample sets are accessible through SRA study accession number SRP059631.

RESULTS AND DISCUSSION

Characteristics of the subjects included in the study. The work described here represents a pilot study to determine optimal experimental conditions for a much larger metagenomic survey, which is aimed at exploring the gut microbial biodiversity in infants and corresponding mothers. The current study enrolled four mother-infant pairs, which were selected based on the following criteria: subjects had to be in good health and had not been prescribed antibiotics and were not taking any probiotics. Furthermore, since we were interested in evaluating the composition of bifidobacterial populations occurring in the gut of mother-child pairs, we only included mothers that were exclusively or for large part breast-feeding their newborns and babies that had been vaginally delivered. Conditions that were previously shown to be associated with the highest densities of bifidobacteria in infants (4).

Evaluation of the bifidobacterial population in the gut/milk microbiota of mother-newborn pairs. We assessed the composition of fecal/milk microbiota of four corresponding mother-infant sets, named SS1 to SS4, using a previously described bifidobacterial ITS-profiling approach (16). As outlined in Figure 1, bifidobacterial ITS-profiling of fecal samples from the mother and corresponding infant revealed the presence of representatives of the *Bifidobacterium* genus. Notably, members of the species *Bifidobacterium adolescentis, Bifidobacterium angulatum, B. bifidum, B. breve, Bifidobacterium dentium, Bifidobacterium longum, Bifidobacterium pseudolongum* and *Bifidobacterium thermacidophilum* were commonly detected in both mother and infant fecal samples (Fig. 1). As previously suggested, human milk may represent a medium for bacterial transmission from mother to newborn (33). Thus, the bifidobacterial microbiota of human milk samples retrieved from each mother was also evaluated by ITS-profiling analyses. As displayed in Figure 1, we observed the presence of members of the

Applied and Environmental Microbiology

234 *Bifidobacterium* genus in the bacterial communities residing in these milk samples, while
235 comparing these data with those obtained from the mother and infant fecal samples revealed the
236 common presence of *B. adolescentis, B. angulatum, B. breve, B. dentium, B. longum, B.*
237 *pseudolongum* and *B. thermacidophilum.*
238 In order to identify the possible presence of the same or a similar strain in a mother and/or milk
239 sample and a corresponding infant fecal sample, we compared for each ITS dataset the reference
240 sequence of all predicted ITS Operational Taxonomic Units (OTUs), e.g. clusters of identical
241 sequences generated by the ITS-profiling analysis protocol (16). Since the ITS sequence is
242 highly variable, the retrieval of ITS sequences from different samples displaying 100 % identity
243 would be indicative of an identical or a very closely related strain being present in such samples
244 (16). Interestingly, OTUs shared by an entire sample set ranged from 63 in SS3 to four in SS4,
245 for a total of 126 OTUs shown in Figure 2. These OTUs encompass members of the species *B.*
246 *adolescentis, B. angulatum, B. breve*, *B. longum* and *B. pseudolongum*, as well as a putative
247 novel bifidobacterial species (Fig. 2), suggesting a crucial role of these bacterial species in
248 microbiota transmission and colonization of the newborn gut. Additionally, of these 126 shared
249 OTUs, 28 are present at a high relative abundance (among the 100 most represented OTUs) in at
250 least one sample set, thus being relevant in defining the overall bifidobacterial community (Fig.
251 2). Furthermore, we observed that in infant fecal samples of SS2 and SS4 the bifidobacterial
252 population appears to predominantly (97.1 and 99.4 %, respectively) consist of OTUs that belong
253 to the *B. longum* subsp. *longum* and *B. breve* species, respectively (Fig. 1). Interestingly, the
254 OTU corresponding to the most represented *B. longum* subsp. *longum* strain in infant sample
255 SS2, named SS2 Infant_515, was also identified in the corresponding fecal and milk sample of
256 the mother (Fig. 2). The OTU corresponding to the most represented *B. breve* strain in infant

sample SS4, named SS4_Infant_617, was also identified in the mother's milk sample, but not in the mother's fecal sample, possibly due to its low abundance falling below bifidobacterial ITS-profiling limit of detection of 10E4 CFU/gr (16) (Fig. 2). These results indicate that infant fecal samples of SS2 and SS4 may be good candidates for targeted bifidobacterial genome reconstruction (see below). Despite the existence of common bifidobacterial OTUs between the investigated mother-child pairs, there were also various bifidobacterial OTUs that appear to be uniquely present in the data set of the infant, but not in that of the corresponding mother (e.g., SS1 and SS3 datasets). This may be due to the acquisition of such bacteria from the environment (e.g., during birth or from siblings) or due to the presence of these microorganisms below the limit of detection in the fecal samples of the mother.

Notably, comparison between the OTUs identified in the datasets analyzed (SS1, SS2, SS3 and SS4) revealed the presence of identical ITS OTUs in different sample sets. This observation implies that these identical ITS sequences correspond to very related strains that are present in different mother-infant pairs, or that the ITS profiling is unable to accurately distinguish isolates below the species/subspecies level. Thus, these findings need to be confirmed through the use of a more robust method to distinguish intraspecies relationships such as the use of multilocus sequencing or shotgun metagenomics.

Genome reconstruction of bacterial strains from shotgun metagenomic data. In order to validate and precisely map the occurrence of potentially identical bifidobacterial strains in the microbiome of a mother's fecal sample, and that of a corresponding sample from infant feces and/or from breast milk, an in depth shotgun metagenomics analysis was applied to the SS4 and SS2 infant samples. Taxonomic assignment (as based on publicly available genomic data) of shotgun metagenomic reads obtained from the SS4 infant fecal sample showed that 62.92 % of

Applied and Environmental Microbiology

these reads are classified as bifidobacterial DNA, of which 69 % is taxonomically assigned to the *B. breve* species, confirming what had previously been observed by means of 16S rRNA gene profiling (16) (Fig. 3). These data are in good agreement with the ITS-profiling results, which indicate high prevalence of a putative strain in the fecal sample corresponding to *B. breve* OTU SS4_Infant_617, representing 66.8 % of the total collected ITS sequences as well as 67.2 % of the reads constituting OTUs clustering together with the *B. breve* reference ITS sequence, respectively (Fig. S1). Thus, the SS4 infant metagenomic dataset was exploited in order to develop an optimized bioinformatics protocol for the reconstruction of a bacterial genome from shotgun metagenomics data (SI). Such a protocol allows the reconstruction of a final consensus genome sequence that can be entirely attributed to the *B. breve* strain being present at higher abundance in the shotgun metagenomics dataset (see supplementary information for details). The obtained genome corresponding to OTU SS4_Infant_617 was named *B. breve* BBRI4. Notably, the ITS sequence of *B. breve* BBRI4 showed 100% identity with ITS OTU SS4_Infant_617.

The developed pipeline (SI) was also applied to the shotgun sequencing reads achieved from the *Bifidobacterium*-rich SS2 infant fecal sample. Taxonomic classification of this dataset identified that 57.4 % of the obtained microbiome reads was assignable to bifidobacterial DNA, of which 69.8 % was taxonomically annotated as a *B. longum* species, as previously noted at genus level by 16S rRNA gene profiling (16) (Fig. 3). Furthermore, ITS-profiling demonstrated high prevalence of OTU SS2_Infant_515 corresponding to a putative *B. longum* subsp. *longum* strain, named BLOI2 (Fig. S1), which represents 69.9 % of the total bifidobacterial community and 72.1 % of the predicted *B. longum* subsp. *longum* population in this particular sample.

The final assembly of the *B. breve* BBRI4 genome consisted of 14 contigs with a total length of 2,411,812 bp, while chromosome reconstruction based on shotgun metagenomics data of *B.*

longum subsp. *longum* BLOI2 allowed the retrieval of 72 contigs corresponding to a chromosome of 2,417,590 bp. In both cases, the final number of generated contigs was comparable with that achieved when attempting targeted genome sequencing (34-36).

Comparative genomics of *B. breve* BBRI4 with 16 publicly available *B. breve* genomes led to prediction of a *B. breve* pan-genome encompassing 3889 *B. breve*-specific Cluster of Orthologous Genes (Bb-COGs) of which 1203 are shared by all the *B. breve* analyzed strains, thus constituting the core-genome of this species. These results are consistent with observations from a previous study encompassing 13 *B. breve* genomes (37). In addition, 82 Bb-COGs represented the Truly Unique Genes (TUG) of *B. breve* BBRI4 (Fig. S2). Similarly, 30 genomes, encompassing all publicly available *B. longum* genome sequences, were used in order to identify the *B. longum* pan-genome, which consists of 6077 *B. longum*-specific COGs (Bl-COGs) of which 827 appear to be shared by all 31 genomes analyzed, thus representing the core-genome. The TUG of *B. longum* subsp. *longum* BLOI2 constitutes of 121 Bl-COGs (Fig. S3). Additional information regarding general genome features as well as comparative genomics data of *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 can be found in the supplementary material (SI).

The chromosomes of *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 were screened for genes putatively acquired by Horizontal Gene Transfer (HGT) events employing Colombo software (38) settled at the maximum sensitivity and low specificity. This *in silico* analysis predicted that 76.3 % of the *B. breve* BBRI4 ORFeome and 72,8 % of the *B. longum* subsp. *longum* BLOI2 ORFeome was not acquired by HGT events. Notably, among the BBRI4 genes that had not been predicted to have been acquired by HGT events, we identified 23 ORFs, which are also encompassing the TUG of *B. breve* BBRI4 strain. These ORFs represent optimal marker genes for the *B. breve* BBRI4 strain. Furthermore, of the 1469 *B. longum* subsp. *longum* BLOI2

Applied and Environmental Microbiology

ORFs putatively not involved in HGT events, 34 were observed to be unique to the BLOI2 strain by comparative genomic analyses, thus representing optimal marker genes for identification of this strain.

Isolation and genome sequencing of *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 strains. In order to validate the reconstruction of *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 genomic data from microbiomic datasets we decided to isolate these strains from fecal samples of the SS4 and SS2 infants, respectively, and to subject the isolated strain to targeted genome sequencing. *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 strains were isolated using a selective bifidobacterial medium (15). Selected colonies were identified by PCR amplification, and subsequent sequencing of the amplicons corresponding to the ITS and marker genes specific of *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 (see above). Two positively verified isolates (i.e. one isolate of *B. breve* BBRI4 and one of *B. longum* subsp. *longum* BLOI2) were then subjected to whole genome sequencing employing the Illumina MiSeq platform. Analysis of the assembled genome sequences was performed utilizing the identical pipeline that had been applied for genome reconstruction using the shotgun metagenomics datasets (see above). The genome sequencing of the *B. breve* BBRI4 isolate allowed the identification of 96 contigs with an average coverage of 37 and encoding 1935 ORFs, while the chromosome of *B. longum* subsp. *longum* BLOI2 isolate encompasses 77 contigs with an average coverage of 51.86 and encoding 2047 ORFs. These genomes were used to evaluate the quality of the corresponding assembly obtained from shotgun metagenomics reads. Alignment of the *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 chromosomes achieved by targeted genome sequencing with those obtained from reconstruction by shotgun metagenomics datasets revealed a syntenic structure and very limited differences (see below),

Applied and Environmental Microbiology

most likely attributable to DNA sequencing mistakes (Fig. S4). Moreover, the average nucleotide identity based on mummer alignments (ANIm) between the genomes sequences achieved by targeted genome sequencing or by microbiomic datasets for *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 revealed identities ranging from 99.91 % to 99.92 % (99.96 % and 100.00 % in case of BLAST-based ANI), respectively. These data indicate the reliability of the metagenomic assembly protocol for accurate reconstruction of targeted bacterial genomes. In order to further confirm the reliability of the genomic data reconstructed from microbiomic datasets we compared the predicted marker genes for *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 using both genome sequencing data for each strain. Notably, the predicted marker genes of these isolates show a value of identity at nucleotide level of 100 % between the gene sets retrieved from targeted genome attempts and microbiomic datasets, thus confirming the precise identification of *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2.

Assessing the gut persistence of *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 strain during the lifespan of infants. In order to assess if *B. breve* BBRI4 and *B. longum* subsp. *longum* BLOI2 strains are maintained in the gut of the SS4 and SS2 infants, respectively, at later stages of life, we collected additional fecal samples from these infants six months (SS2_T6m and SS4_T6m) after the initial sample collection, along with fecal and milk samples from the mother. These samples were assayed by bifidobacterial ITS-profiling (16) (Fig. 3). Interestingly, comparison of the results from the 16S rRNA gene-profiling of SS2_T6m and SS4_T6m with taxonomy profiles of SS2 and SS4 (16) highlight a marked decrease of the relative abundance of phylotypes belonging to the genus *Bifidobacterium* especially in the infant fecal samples, with reduction from 56.4 % in SS2 to 2.9 % in SS2_T6m, and a reduction from 50.8 % in SS4 to 0.5 % in SS4_T6m (Fig. 3). This observation is indicative of microbiota evolution towards that

typical of adult human beings as a result of dietary changes (39). Furthermore, bifidobacterial ITS-profiling allowed the identification of nine bifidobacterial species that appear to be shared by the three samples constituting SS2_T6m, and 5 bifidobacterial species shared by the three samples constituting SS4_T6m. Notably, *B. adolescentis*, *B. longum* subsp. *longum* and *B. pseudolongum* that are found in these latter datasets were also shared by all samples constituting SS2 and SS4 (Fig. 1 and Fig. 3). These findings suggest a high persistence of these human-gut adapted bifidobacterial species in the infant and corresponding mother gut, as well as in mother's milk. In addition, a manual search in the T6m sample sets for the presence of the most abundant 15 OTUs in SS2_Infant and SS4_Infant samples showed that one and 15 OTUs are still present in SS2_T6m and in SS4_T6m, respectively (Fig. 3). Interestingly, ITS OTU SS2_Infant_515 corresponding to *B. longum* subsp. *longum* BLOI2 was detected in the SS2_T6m_Mother sample but not in the SS2_T6m_Milk and SS2_T6m_Infant dataset (Fig. 3). This observation was verified by PCR using strain-specific primer pairs based on a predicted marker gene of *B. longum* subsp. *longum* BLOI2 encoding an ATPase domain-containing protein (BLOI2_0636) (Fig. S1). Primers were tested for specificity using the NCBI tool Primer-BLAST and the NCBI nt database containing all the so far sequenced bacterial genomes. Notably, these results allowed the identification of this strain in all sample constituting both SS2 and SS2_T6m (Fig. S5), thus suggesting that the prevalence of the BLOI2 strain was below 10E4 CFU/gr, which is the limit of detection previously estimated by the bifidobacterial ITS-profiling approach in the infant fecal and milk samples of SS2_T6m (16). Notably, ITS OTU SS4_Infant_617, corresponding to *B. breve* BBRI4, is still present in the mother milk and infant fecal sample of SS4_T6m. Validation through PCR using strain-specific primer pairs based on a predicted marker gene of *B. breve* BBRI4, encoding a putative solute-binding component of a dipeptide ABC transporter

Applied and Environmental Microbiology

(BBRI4_0962), confirmed the persistence after six months of this strain in these two samples but its absence (or below detection level) in the mother's gut (Fig. S5). These observations underline the importance of initial colonization of the newborn gut of bacterial strains that may be maintained during the infant lifespan. However, due to the lack of stool-samples collected at later time points we cannot rule out that persistence of these bifidobacterial strains extends beyond 6 months or is just restricted to relative short time spans.

Furthermore, the mother's fecal microbiota was shown to have changed significantly at the six month follow up time point, an observation confirming previous studies that had found high variability in the composition of fecal microbiota during pregnancy and breast-feeding (40).

CONCLUSIONS

Vertical transmission of gut bacteria from mother to their offspring is considered a pivotal route for microbiota establishment in newborns, although an in-depth evaluation of this process has not been performed. Here, we collected and analyzed four sample sets each encompassing a mother fecal and milk sample, combined with a corresponding infant fecal sample. Bifidobacterial ITS-profiling coupled with shotgun metagenomic analyses allowed the identification of a common bifidobacterial profile in the mother-infant pairs and revealed the presence of identical strains shared between these hosts. Such data suggests the existence of a microbiota transfer process that drives the acquisition and subsequent persistence of specific bacterial strains in the infant gut, which appears to be influenced by the mother's gut and breast milk microbiota. Bifidobacteria represent the dominant members of the gut microbiota of infants (4), thus it is not surprising to identify their occurrence at high numbers in the analyzed infant stool samples. However, it is interesting to note that even though bifidobacteria are generally poorly detectable in adult fecal samples (4, 15), despite the finding that they have been found to persist at high levels in some

Applied and Environmental Microbiology

adult individuals (41), we were able to trace the occurrence of a specific strain being present in the gut of both mother and infant. One may argue that bifidobacteria, even when they decrease in relative abundance following weaning, persist following their initial transfer to the infant gut. The particular route(s) used for such a transfer is (are) still unknown, although it has been speculated that it is facilitated by a human milk route (42, 43), and/or by contamination of fecal and vaginal samples in the partum canal (44). Very recently, the possibility of fetal colonization of the infant gut through the placenta has been put forward (45). Due the small cohort of subjects investigated in this study, we cannot made any statistically robust conclusions about the incidence of vertical transmission of the gut microbiota from mother to child in the human population. Nonetheless, our findings represent a first important molecular evidence of the existence of this route for the generation of the gut microbiota in the earlier stages of life.

Acknowledgments

We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. This work was financially supported by Fondazione Caritro, Trento, Italy. DvS is a member of the Alimentary Pharmabiotic Centre funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (Grant number SFI/12/RC/2273). We thank V. De Sanctis and R. Bertorelli (NGS Facility at the Centre for Integrative Biology and LaBSSAH, University of Trento) for shotgun sequencing.

References

1. **Gilbert SF, Sapp J, Tauber AI.** 2012. A symbiotic view of life: we have never been individuals. Q Rev Biol **87:**325-341.
2. **Ley RE, Peterson DA, Gordon JI.** 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell **124:**837-848.

3. **Funkhouser LJ, Bordenstein SR.** 2013. Mom knows best: the universality of maternal microbial transmission. PLoS Biol **11:**e1001631.
4. **Turroni F, Peano C, Pass DA, Foroni E, Severgnini M, Claesson MJ, Kerr C, Hourihane J, Murray D, Fuligni F, Gueimonde M, Margolles A, De Bellis G, O'Toole PW, van Sinderen D, Marchesi JR, Ventura M.** 2012. Diversity of bifidobacteria within the infant gut microbiota. PLoS One **7:**e36957.
5. **Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE.** 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci U S A **108 Suppl 1:**4578-4585.
6. **Ventura M, Turroni F, Motherway MO, MacSharry J, van Sinderen D.** 2012. Host-microbe interactions that facilitate gut colonization by commensal bifidobacteria. Trends Microbiol **20:**467-476.
7. **Sela DA, Chapman J, Adeuya A, Kim JH, Chen F, Whitehead TR, Lapidus A, Rokhsar DS, Lebrilla CB, German JB, Price NP, Richardson PM, Mills DA.** 2008. The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome. Proc Natl Acad Sci U S A **105:**18964-18969.
8. **Turroni F, Bottacini F, Foroni E, Mulder I, Kim JH, Zomer A, Sanchez B, Bidossi A, Ferrarini A, Giubellini V, Delledonne M, Henrissat B, Coutinho P, Oggioni M, Fitzgerald GF, Mills D, Margolles A, Kelly D, van Sinderen D, Ventura M.** 2010. Genome analysis of Bifidobacterium bifidum PRL2010 reveals metabolic pathways for host-derived glycan foraging. Proc Natl Acad Sci U S A **107:**19514-19519.
9. **Turroni F, Foroni E, Serafini F, Viappiani A, Montanini B, Bottacini F, Ferrarini A, Bacchini PL, Rota C, Delledonne M, Ottonello S, van Sinderen D, Ventura M.** 2011. Ability of Bifidobacterium breve to grow on different types of milk: exploring the metabolism of milk through genome analysis. Appl Environ Microbiol **77:**7408-7417.
10. **Martin R, Jimenez E, Heilig H, Fernandez L, Marin ML, Zoetendal EG, Rodriguez JM.** 2009. Isolation of bifidobacteria from breast milk and assessment of the bifidobacterial population by PCR-denaturing gradient gel electrophoresis and quantitative real-time PCR. Appl Environ Microbiol **75:**965-969.
11. **Wassenaar TM, Panigrahi P.** 2014. Is a foetus developing in a sterile environment? Lett Appl Microbiol **59:**572-579.
12. **Mikami K, Takahashi H, Kimura M, Isozaki M, Izuchi K, Shibata R, Sudo N, Matsumoto H, Koga Y.** 2009. Influence of maternal bifidobacteria on the establishment of bifidobacteria colonizing the gut in infants. Pediatr Res **65:**669-674.
13. **Tannock GW, Fuller R, Smith SL, Hall MA.** 1990. Plasmid profiling of members of the family Enterobacteriaceae, lactobacilli, and bifidobacteria to study the transmission of bacteria from mother to infant. J Clin Microbiol **28:**1225-1228.
14. **Makino H, Kushiro A, Ishikawa E, Kubota H, Gawad A, Sakai T, Oishi K, Martin R, Ben-Amor K, Knol J, Tanaka R.** 2013. Mother-to-infant transmission of intestinal bifidobacterial strains has an impact on the early development of vaginally delivered infant's microbiota. PLoS One **8:**e78331.
15. **Turroni F, Foroni E, Pizzetti P, Giubellini V, Ribbera A, Merusi P, Cagnasso P, Bizzarri B, de'Angelis GL, Shanahan F, van Sinderen D, Ventura M.** 2009. Exploring the diversity of the bifidobacterial population in the human intestinal tract. Appl Environ Microbiol **75:**1534-1545.
16. **Milani C, Lugli GA, Turroni F, Mancabelli L, Duranti S, Viappiani A, Mangifesta M, Segata N, van Sinderen D, Ventura M.** 2014. Evaluation of bifidobacterial community composition in the human gut by means of a targeted amplicon sequencing (ITS) protocol. FEMS Microbiol Ecol **90:**493-503.

17. **Milani C, Hevia A, Foroni E, Duranti S, Turroni F, Lugli GA, Sanchez B, Martin R, Gueimonde M, van Sinderen D, Margolles A, Ventura M.** 2013. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. PLoS One **8:**e68739.
18. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods **7:**335-336.
19. **Zhao Y, Tang H, Ye Y.** 2012. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics **28:**125-126.
20. **Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC.** 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res **21:**1552-1560.
21. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **19:**455-477.
22. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11:**119.
23. **Chevreux.** 1999. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information Computer Science and Biology. Proceedings of the German Conference on Bioinformatics (GCB) **99:**45-56.
24. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M.** 2014. Pfam: the protein families database. Nucleic Acids Res **42:**D222-230.
25. **Lowe TM, Eddy SR.** 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25:**955-964.
26. **Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J, Gabaldon T, Rattei T, Creevey C, Kuhn M, Jensen LJ, von Mering C, Bork P.** 2014. eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Res **42:**D231-239.
27. **Darling AC, Mau B, Blattner FR, Perna NT.** 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14:**1394-1403.
28. **Richter M, Rossello-Mora R.** 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A **106:**19126-19131.
29. **Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J.** 2012. PGAP: pan-genomes analysis pipeline. Bioinformatics **28:**416-418.
30. **Enright AJ, Van Dongen S, Ouzounis CA.** 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res **30:**1575-1584.
31. **Katoh K, Misawa K, Kuma K, Miyata T.** 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **30:**3059-3066.
32. **Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG.** 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:**2947-2948.
33. **Fernandez L, Langa S, Martin V, Maldonado A, Jimenez E, Martin R, Rodriguez JM.** 2013. The human milk microbiota: origin and potential roles in health and disease. Pharmacol Res **69:**1-10.
34. **Milani C, Lugli GA, Duranti S, Turroni F, Bottacini F, Mangifesta M, Sanchez B, Viappiani A, Mancabelli L, Taminiau B, Delcenserie V, Barrangou R, Margolles A, van Sinderen D, Ventura M.** 2014. Genomic encyclopedia of type strains of the genus Bifidobacterium. Appl Environ Microbiol **80:**6290-6302.

35. **Duranti S, Turroni F, Lugli GA, Milani C, Viappiani A, Mangifesta M, Gioiosa L, Palanza P, van Sinderen D, Ventura M.** 2014. Genomic characterization and transcriptional studies of the starch-utilizing strain Bifidobacterium adolescentis 22L. Appl Environ Microbiol **80:**6080-6090.
36. **Duranti S, Milani C, Lugli GA, Turroni F, Mancabelli L, Sanchez B, Ferrario C, Viappiani A, Mangifesta M, Mancino W, Gueimonde M, Margolles A, van Sinderen D, Ventura M.** 2014. Insights from genomes of representatives of the human gut commensal Bifidobacterium bifidum. Environ Microbiol doi:10.1111/1462-2920.12743.
37. **Bottacini F, O'Connell Motherway M, Kuczynski J, O'Connell KJ, Serafini F, Duranti S, Milani C, Turroni F, Lugli GA, Zomer A, Zhurina D, Riedel C, Ventura M, van Sinderen D.** 2014. Comparative genomics of the Bifidobacterium breve taxon. BMC Genomics **15:**170.
38. **Waack S, Keller O, Asper R, Brodag T, Damm C, Fricke WF, Surovcik K, Meinicke P, Merkl R.** 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics **7:**142.
39. **Voreades N, Kozil A, Weir TL.** 2014. Diet and the development of the human intestinal microbiome. Front Microbiol **5:**494.
40. **Koren O, Goodrich JK, Cullender TC, Spor A, Laitinen K, Backhed HK, Gonzalez A, Werner JJ, Angenent LT, Knight R, Backhed F, Isolauri E, Salminen S, Ley RE.** 2012. Host remodeling of the gut microbiome and metabolic changes during pregnancy. Cell **150:**470-480.
41. **Dinoto A, Marques TM, Sakamoto K, Fukiya S, Watanabe J, Ito S, Yokota A.** 2006. Population dynamics of Bifidobacterium species in human feces during raffinose administration monitored by fluorescence in situ hybridization-flow cytometry. Appl Environ Microbiol **72:**7739-7747.
42. **Turroni F, Strati F, Foroni E, Serafini F, Duranti S, van Sinderen D, Ventura M.** 2012. Analysis of predicted carbohydrate transport systems encoded by Bifidobacterium bifidum PRL2010. Appl Environ Microbiol **78:**5002-5012.
43. **Khodayar-Pardo P, Mira-Pascual L, Collado MC, Martinez-Costa C.** 2014. Impact of lactation stage, gestational age and mode of delivery on breast milk microbiota. J Perinatol **34:**599-605.
44. **Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R.** 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc Natl Acad Sci U S A **107:**11971-11975.
45. **Aagaard K, Ma J, Antony KM, Ganu R, Petrosino J, Versalovic J.** 2014. The placenta harbors a unique microbiome. Sci Transl Med **6:**237ra265.

570 **Figure legends**

571 **Figure 1:** Bifidobacterial ITS-profiling of four sample sets each encompassing the fecal and
572 milk samples from a mother and a fecal sample from the corresponding infant. Panel a represents
573 a bar plot of the identified bifidobacterial population in the 12 analyzed samples. Panel b shows a
574 heat map of bifidobacterial species that appeared to be present in both infant fecal sample and
575 mother fecal sample (I-M), and/or present in the infant fecal sample and mother milk sample (I-
576 K).

577

578 **Figure 2:** Distribution of computed ITS OTUs in the four analyzed sample sets. Panel a displays
579 the 126 ITS OTUs identified as being present in an entire sample set (e.g. present in the mother
580 fecal and milk samples as well as by the infant fecal sample). The red color indicates presence of
581 a specific OTU in the three samples of a given set; the light blue color is used to highlight those
582 OTUs being among the 100 most represented in at least two members of a sample set. OTUs are
583 also clustered by taxonomic assignment and colored accordingly in the first column of the heat
584 map. Panel b shows a heat map illustrating the presence of the five most abundant OTUs of each
585 analyzed sample, and their presence or absence in the two other samples of the corresponding
586 set, accompanied by their taxonomic assignment. The most represented OTUs in the infant
587 samples of SS2 and SS4 are highlighted in blue and green, respectively.

588

589 **Figure 3:** Evaluation of gut microbiota persistence in SS2 and SS4. Panel a shows a bar plot
590 representing results from 16S rRNA gene profiling of SS2, SS2_T6m, SS4 and SS4_T6m at
591 genus level. Only taxa with relative abundance >5 % in at least one sample are indicated.
592 "Unclassified members" is abbreviated as "u.m.". Panel b depicts the bifidobacterial taxonomic

593 profile in SS2, SS2_T6m, SS4 and SS4_T6m obtained through bifidobacterial ITS-profiling.
594 Panel c represents a heat map showing persistence of bifidobacterial OTUs (putative
595 bifidobacterial strains) in SS2 and SS4 after 6 months. The most represented OTUs in infant
596 sample of SS2 and SS4 are highlighted in blue and green, respectively.
597

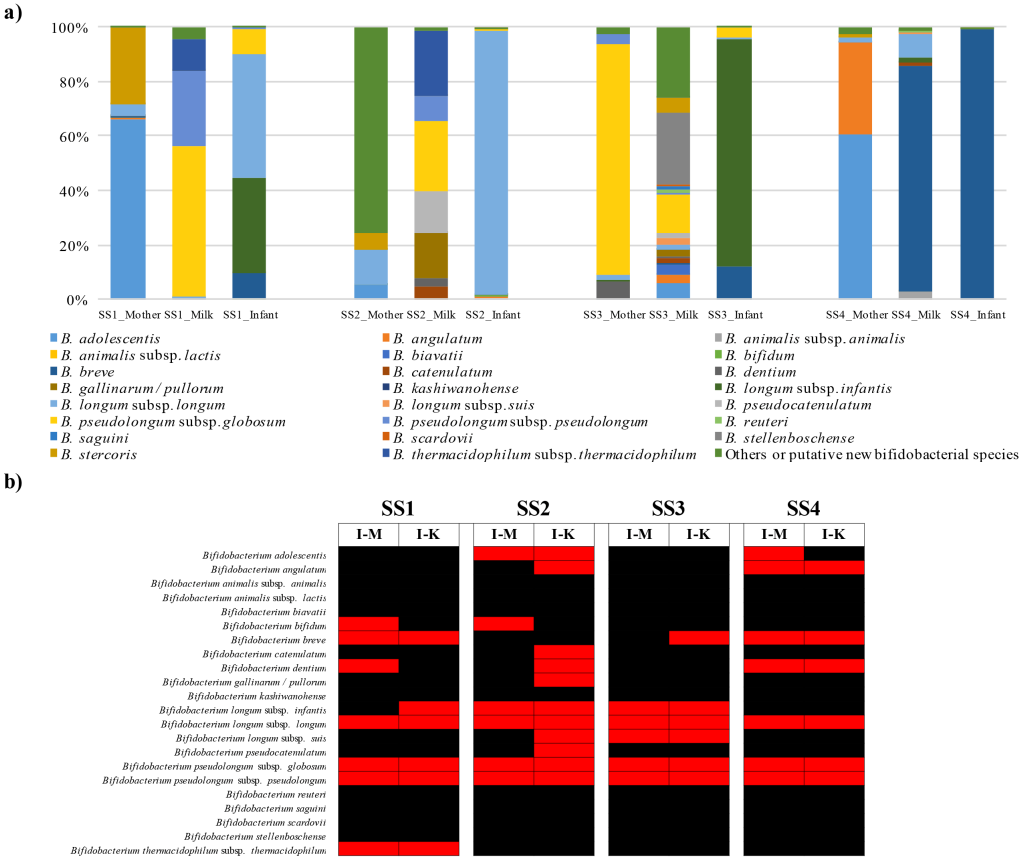
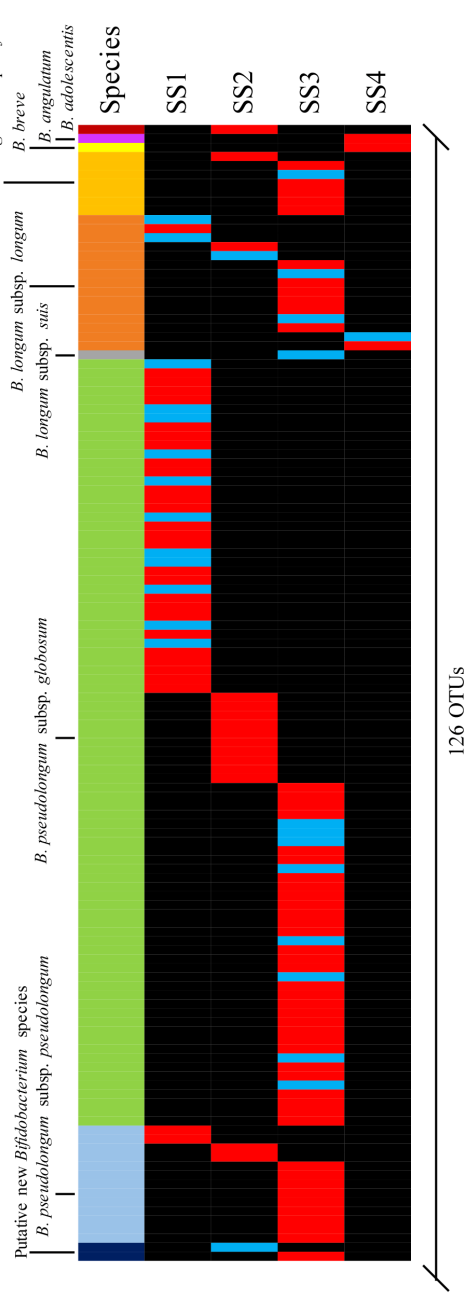


Figure 1

a)



b)

	SS1			
	SS1_Mother	SS1_Milk	SS1_Infant	
SS1_Infant_249		SS1_Milk_114	SS1_Infant_249	G
SS1_Infant_419			SS1_Infant_419	G
SS1_Infant_674	SS1_Mother_3131	SS1_Milk_296	SS1_Infant_674	H
SS1_Infant_832			SS1_Infant_832	C
SS1_Infant_1559	SS1_Mother_2064	SS1_Milk_654	SS1_Infant_1559	H
SS1_Milk_209		SS1_Milk_209	SS1_Infant_1113	L
SS1_Milk_495	SS1_Mother_8601	SS1_Milk_495	SS1_Infant_1520	L
SS1_Milk_651	SS1_Mother_4254	SS1_Milk_651	SS1_Infant_1558	L
SS1_Milk_812		SS1_Milk_812		L
SS1_Milk_1153	SS1_Mother_3930	SS1_Milk_1153	SS1_Infant_1304	L
SS1_Mother_324	SS1_Mother_324			P
SS1_Mother_633	SS1_Mother_633			P
SS1_Mother_1733	SS1_Mother_1733	SS1_Milk_1141		A
SS1_Mother_2423	SS1_Mother_2423	SS1_Milk_306		A
SS1_Mother_7363	SS1_Mother_7363			A

	SS2			
	SS2_Mother	SS2_Milk	SS2_Infant	
SS2_Infant_502	SS2_Mother_1116		SS2_Infant_502	H
SS2_Infant_515	SS2_Mother_2967	SS2_Milk_3273	SS2_Infant_515	H
SS2_Infant_744	SS2_Mother_1648		SS2_Infant_744	H
SS2_Infant_1201	SS2_Mother_2607		SS2_Infant_1201	H
SS2_Infant_1218	SS2_Mother_2644		SS2_Infant_1218	H
SS2_Milk_1107		SS2_Milk_1107		N
SS2_Milk_1184		SS2_Milk_1184		I
SS2_Milk_1987		SS2_Milk_1987		N
SS2_Milk_2462		SS2_Milk_2462		F
SS2_Milk_3241		SS2_Milk_3241		I
SS2_Mother_401	SS2_Mother_401		SS2_Infant_166	O
SS2_Mother_915	SS2_Mother_915		SS2_Infant_398	A
SS2_Mother_995	SS2_Mother_995	SS2_Milk_1082	SS2_Infant_438	O
SS2_Mother_1993	SS2_Mother_1993		SS2_Infant_271	O
SS2_Mother_2468	SS2_Mother_2468	SS2_Milk_2844	SS2_Infant_1144	A

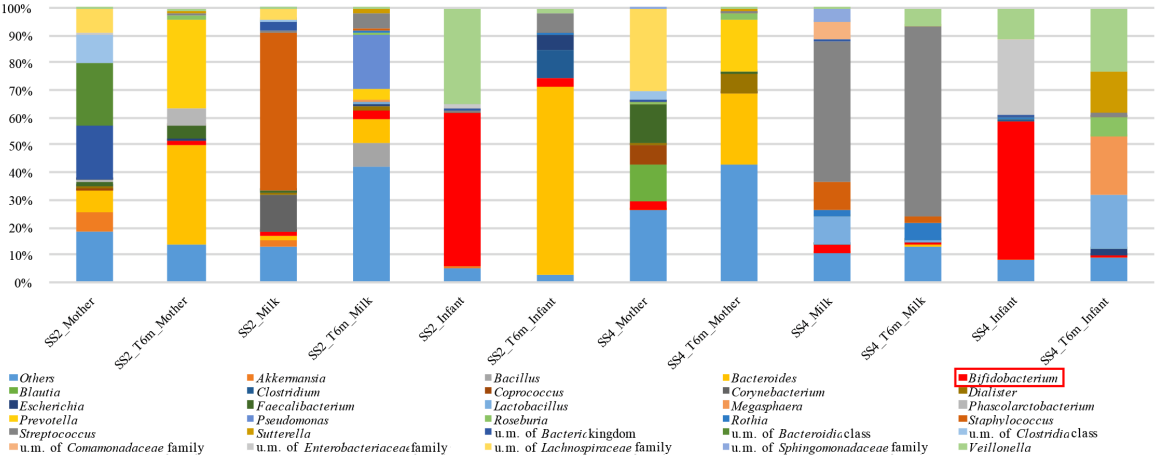
	SS3			
	SS3_Mother	SS3_Milk	S4_Infant	
SS3_Infant_110	SS3_Mother_1380	SS3_Milk_1188	SS3_Infant_110	G
SS3_Infant_192	SS3_Mother_1425		SS3_Infant_192	G
SS3_Infant_703	SS3_Mother_2327	SS3_Milk_2029	SS3_Infant_703	G
SS3_Infant_809		SS3_Milk_1519	SS3_Infant_809	C
SS3_Infant_1080		SS3_Milk_2008	SS3_Infant_1080	C
SS3_Milk_71	SS3_Mother_91	SS3_Milk_71		L
SS3_Milk_138	SS3_Mother_178	SS3_Milk_138		A
SS3_Milk_187		SS3_Milk_187		M
SS3_Milk_281		SS3_Milk_281		D
SS3_Milk_1073		SS3_Milk_1073		A
SS3_Mother_534	SS3_Mother_534	SS3_Milk_443	SS3_Infant_919	L
SS3_Mother_570	SS3_Mother_570	SS3_Milk_474		E
SS3_Mother_1457	SS3_Mother_1457	SS3_Milk_1251	SS3_Infant_265	L
SS3_Mother_1716	SS3_Mother_1716	SS3_Milk_1215	SS3_Infant_816	L
SS3_Mother_1731	SS3_Mother_1731	SS3_Milk_1492	SS3_Infant_129	L

	SS4			
	SS4_Mother	SS4_Milk	SS4_Infant	
SS4_Infant_177		SS4_Milk_382	SS4_Infant_177	C
SS4_Infant_548		SS4_Milk_11	SS4_Infant_548	C
SS4_Infant_617		SS4_Milk_563	SS4_Infant_617	C
SS4_Infant_754		SS4_Milk_453	SS4_Infant_754	C
SS4_Infant_805		SS4_Milk_552	SS4_Infant_805	C
SS4_Milk_11		SS4_Milk_11	SS4_Infant_548	C
SS4_Milk_281	SS4_Mother_2277	SS4_Milk_281		H
SS4_Milk_453		SS4_Milk_453	SS4_Infant_754	C
SS4_Milk_552		SS4_Milk_552	SS4_Infant_805	C
SS4_Milk_563		SS4_Milk_563	SS4_Infant_617	C
SS4_Mother_260	SS4_Mother_260		SS4_Infant_846	A
SS4_Mother_1570	SS4_Mother_1570	SS4_Milk_604	SS4_Infant_830	B
SS4_Mother_2277	SS4_Mother_2277	SS4_Milk_281		H
SS4_Mother_2972	SS4_Mother_2972			A
SS4_Mother_3694	SS4_Mother_3694	SS4_Milk_796		A

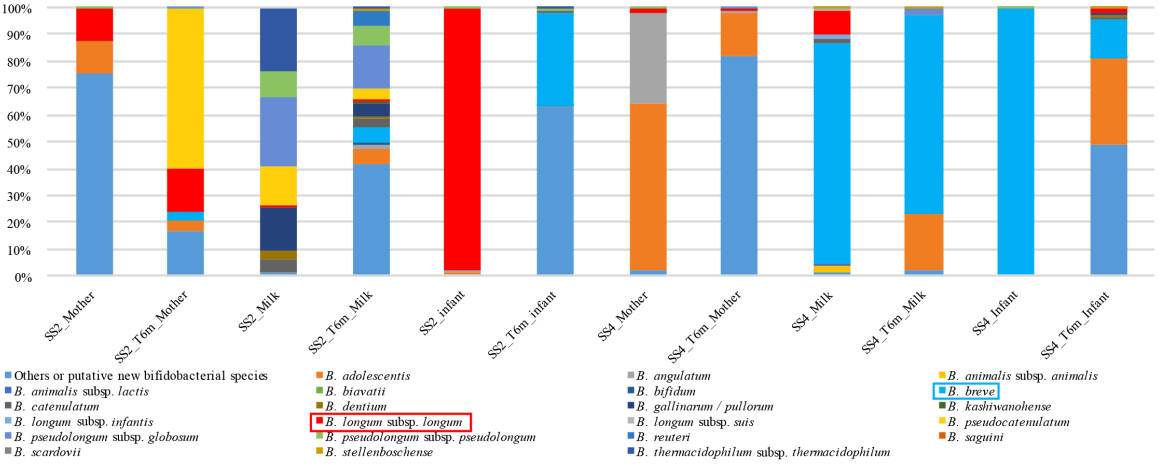
A *B. adolescentis*
B *B. angulatum*
C *B. breve*
D *B. callitrichos*
E *B. dentium*
F *B. gallinarum* / *B. pullorum*
G *B. longum* subsp. *infantis*
H *B. longum* subsp. *longum*
I *B. pseudocatenulatum*
L *B. pseudolongum* subsp. *globosum*
M *B. stellenboschense*
N *B. thermacidophilum* subsp. *thermacidophilum*
O Putative new *Bifidobacterium* species
P Unknown *Bifidobacteriaceae* member

Figure 2

a)



b)



c)

	SS2			SS2_T6m			
	M2	K2	I2	M2_T6m	K2_T6m	I2_T6m	
SS2_Infant_96		SS2_Milk_241	SS2_Infant_96		SS2_T6m_Milk_210	SS2_T6m_Infant_256	A
SS2_Infant_316	SS2_Mother_731		SS2_Infant_316				D
SS2_Infant_408	SS2_Mother_718		SS2_Infant_408				D
SS2_Infant_502	SS2_Mother_1116		SS2_Infant_502				D
SS2_Infant_515	SS2_Mother_2967	SS2_Milk_3273	SS2_Infant_515	SS2_T6m_Mother_2457			D
SS2_Infant_590	SS2_Mother_1318		SS2_Infant_590				D
SS2_Infant_595		SS2_Milk_1436	SS2_Infant_595		SS2_T6m_Milk_198		B
SS2_Infant_656	SS2_Mother_1458		SS2_Infant_656				D
SS2_Infant_744	SS2_Mother_1648		SS2_Infant_744				D
SS2_Infant_879	SS2_Mother_1938		SS2_Infant_879				D
SS2_Infant_1096	SS2_Mother_2375		SS2_Infant_1096				D
SS2_Infant_1201	SS2_Mother_2607		SS2_Infant_1201				D
SS2_Infant_1218	SS2_Mother_2644		SS2_Infant_1218	SS2_T6m_Mother_2218			D
SS2_Infant_1344	SS2_Mother_924		SS2_Infant_1344				D
SS2_Infant_1506	SS2_Mother_3280		SS2_Infant_1506				D

	SS4			SS4_T6m			
	M4	K4	I4	M4_T6m	K4_T6m	I4_T6m	
SS4_Infant_177		SS4_Milk_382	SS4_Infant_177		SS4_T6m_Milk_121	SS4_T6m_Infant_346	C
SS4_Infant_533		SS4_Milk_1065	SS4_Infant_533		SS4_T6m_Milk_341	SS4_T6m_Infant_1881	C
SS4_Infant_548		SS4_Milk_11	SS4_Infant_548		SS4_T6m_Milk_3	SS4_T6m_Infant_1013	C
SS4_Infant_550		SS4_Milk_16	SS4_Infant_550		SS4_T6m_Milk_5	SS4_T6m_Infant_1017	C
SS4_Infant_617		SS4_Milk_563	SS4_Infant_617		SS4_T6m_Milk_183	SS4_T6m_Infant_2452	C
SS4_Infant_660		SS4_Milk_246	SS4_Infant_660			SS4_T6m_Infant_1205	C
SS4_Infant_754		SS4_Milk_453	SS4_Infant_754		SS4_T6m_Milk_146	SS4_T6m_Infant_1381	C
SS4_Infant_805		SS4_Milk_552	SS4_Infant_805		SS4_T6m_Milk_175	SS4_T6m_Infant_1066	C
SS4_Infant_980		SS4_Milk_710	SS4_Infant_980		SS4_T6m_Milk_235	SS4_T6m_Infant_1818	C
SS4_Infant_1101		SS4_Milk_53	SS4_Infant_1101		SS4_T6m_Milk_17	SS4_T6m_Infant_2034	C
SS4_Infant_1242		SS4_Milk_346	SS4_Infant_1242		SS4_T6m_Milk_110	SS4_T6m_Infant_2270	C
SS4_Infant_1244		SS4_Milk_350	SS4_Infant_1244		SS4_T6m_Milk_112	SS4_T6m_Infant_1894	C
SS4_Infant_1343		SS4_Milk_556	SS4_Infant_1343		SS4_T6m_Milk_177	SS4_T6m_Infant_2447	C
SS4_Infant_1900		SS4_Milk_598	SS4_Infant_1900		SS4_T6m_Milk_196	SS4_T6m_Infant_1226	C
SS4_Infant_2061		SS4_Milk_923	SS4_Infant_2061		SS4_T6m_Milk_23	SS4_T6m_Infant_3794	C

A *B. adolescentis* B *B. angulatum* C *B. breve* D *B. longum* subsp. *longum*

Figure 3