



Multi-population cohort meta-analysis of human intestinal microbiota in early life reveals the existence of infant community state types (ICSTs)



Leonardo Mancabelli^a, Chiara Tarracchini^a, Christian Milani^a, Gabriele Andrea Lugli^a, Federico Fontana^a, Francesca Turrone^a, Douwe van Sinderen^b, Marco Ventura^{a,*}

^aLaboratory of Probiogenomics, Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy

^bAPC Microbiome Institute and School of Microbiology, Bioscience Institute, National University of Ireland, Cork, Ireland

ARTICLE INFO

Article history:

Received 25 July 2020

Received in revised form 26 August 2020

Accepted 27 August 2020

Available online 15 September 2020

Keywords:

Metagenomics

16S rRNA profiling

Community state types

ABSTRACT

Appropriate development of the intestinal microbiota during infancy is known to be important for human health. In fact, aberrant alterations of the microbial composition during childhood may cause short- and/or long-term negative health effects. Many factors influence the initial assembly and subsequent progression of the gut microbiota of a neonate, such as feeding type, delivery mode, gestational age, maternal metabolic status and antibiotic exposure.

In the current study, the composition of the infant gut core-microbiota was explored, revealing particular variations of this core-microbiota during the first three years as influenced by delivery mode and feeding type. A multi-population cohort meta-analysis was performed by selecting 15 publicly available datasets pertaining to taxonomic profiles of 1035 fecal samples of healthy infants, as obtained by means of a 16S rRNA gene-based profiling approach. Interestingly, this multi-population cohort meta-analysis revealed great microbial complexity and specific taxonomic shifts in children older than six months, suggesting a major impact by the introduction of solid foods which prompts progression of infant gut microbiota towards that typical of adults. The taxonomic data sets employed in this multi-population cohort meta-analysis possess the statistical robustness to allow the identification of infant community state types (ICSTs). Our analysis therefore reveals the existence of specific taxonomic patterns that correspond to particular nutritional and developmental stages of early life, and that had previously been obscured by the high variability typical of such infant gut microbiota.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The human gastrointestinal tract is colonized by billions of microorganisms, which are commonly referred to as the gut microbiota, and which plays a major role in maintaining host health by influencing and modulating nutritional, metabolic, immunological and physiological functionalities [1]. Consequently, initial gut colonization events and associated development of the gut microbiota in early (human) life is believed to elicit both temporary as well as long-lasting health effects, and to play a key role in defining host health status [2].

The early gut microbiota of healthy infants is characterized by low biodiversity and predominance of genera belonging to the Actinobacteria and Proteobacteria phyla, such as *Escherichia*,

Citrobacter and *Bifidobacterium*, being influenced by several factors such as mode of delivery, diet and gestational age [1,3]. With progressive age, the microbiota becomes more complex and bacteria belonging to Lachnospiraceae, Ruminococcaceae, Bacteroidaceae, Prevotellaceae, and Rikenellaceae families become the dominant taxa [4].

The delivery mode plays an important role in the establishment of the initial gut microbiota composition. In fact, comparison between natural and caesarean section (C-section) delivered infants has revealed several differences in their gut microbial population. In detail, the early gut microbiota of vaginally delivered infants has a composition enriched in vagina-associated microbes such as *Lactobacillus* and *Prevotella* [5,6], while that of C-section delivered infants was shown to have reduced biodiversity, being characterized by the presence of environmental microorganisms that probably originate from maternal skin and hospital environment [7,8].

* Corresponding author at: Department of Life Sciences, University of Parma, Parco Area delle Scienze 11a, 43124 Parma, Italy.

E-mail address: marco.ventura@unipr.it (M. Ventura).

The strong correlation between diet and development of the infant gut microbiota has been highlighted in several studies [9,10]. Compared to formula-fed infants, breastfed infants have been shown to possess higher levels of bifidobacteria and lactobacilli [11], along with a lower abundance of the *Bacteroides* genus and potential pathogens belonging to clostridia, enterobacteria, enterococci and staphylococci [1].

The current study describes a multi-population cohort meta-analysis designed to evaluate the evolution of the infant microbiota composition during the early stages of life and its possible variations due to mode of delivery and diet. In particular, in order to define the composition of the infant core-microbiota from delivery up to the age of 3 years, we analyzed data based on a total of 1035 faecal samples of full-term, healthy infants. For this purpose, we selected 15 public datasets [12–26] obtained through 16S rRNA gene-based microbiota profiling, encompassing infants of an age that ranged between a few days to 3 years. Furthermore, we investigated variations of the infant gut microbiota based on the type of delivery and feeding of the newborn.

In contrast to the large part of the previously published meta-analyses regarding the human gut microbiota [27–32], in our study we applied specific procedures aimed to minimize biases due to cross-study comparisons and maximize statistical accuracy. In detail, we selected metagenomic datasets that rely the same sequencing methodology and that are based on PCR primers with high taxonomic coverage. In this context, the taxonomic resolution of the different regions covered by the primer pairs included in this study was validated to ensure efficient comparison of the results. Furthermore, the raw data were re-analyzed using the same bioinformatics pipeline. Thus, our meta-analysis represents a relevant step-forward respect to currently available scientific literature, thereby allowing an in depth assessment of the microbial taxonomic patterns characterizing the gut microbiota development during infancy, and revealing the existence of distinct microbial assemblies named infant community state types (ICSTs).

2. Materials and methods

2.1. Database selection

In this meta-analysis, we obtained 1035 publicly available data sets from 15 studies involving the determination of the infant gut microbiota. In order to obtain high quality and coverage data, we selected 16S rRNA profiling datasets based on Illumina sequencing technology. The studies employed various primer pairs. In detail, we selected 16S rRNA microbial profiling datasets including only faecal samples from healthy infants with an age ranging between a few days to 3 years. No exclusions were made in the selected cohorts based on the mode of delivery and diet.

2.2. Evaluation of primer pair efficiency

The performance of the 11 primer pairs employed in the selected datasets was evaluated through the web-tool TestPrime 1.0 [33]. The latter executes an *in silico* PCR based on the SILVA database and provides the percentage of amplified sequences for each bacterial group at genus level [33]. The TestPrime was based on RefRN SILVA Database ssu-138 and a maximum of three mismatches was allowed [34].

2.3. 16S rRNA profiling analysis

To avoid biases caused by different bioinformatic analysis pipelines, the sequence read pools of each dataset were filtered and analyzed through the same custom script based on the QIIME

2 software suite [35,36]. Quality control maintained sequences with a length between 140 and 400 bp and average sequence quality score of >20, while sequences with homopolymers of >7 bp and mismatched primers were omitted. 16S rRNA Operational Taxonomic Units (OTUs) were defined at 100% sequence homology using DADA2 [37] and OTUs that were represented by just a single sequence were removed. All reads were classified to the lowest possible taxonomic rank using QIIME 2 [35,36] and a reference dataset from the SILVA database v.132 [38]. In order to evaluate the bacterial biodiversity, alpha-diversity was assessed based on Good's coverage, Observed out, Chao1 and Shannon indexes and represented by box-and-whisker plots. Moreover, the Bray-Curtis dissimilarity index was used to estimate the beta-diversity between different age groups, delivery mode and diet. Dissimilarities were represented through a 3-Dimensional Principal Coordinate Analysis (PCoA).

2.4. Statistical analysis

QIIME 2 and SPSS software (www.ibm.com/software/it/analytics/spss/) were used to compute statistical analyses. PERMANOVA analyses were performed using 1000 permutations to estimate *p*-values for differences among populations in PCoA analyses. Furthermore, differential abundance of bacterial genera and alpha-diversity was tested by ANOVA analysis. Moreover, we also calculated the post hoc analysis LSD (least significant difference) for multiple comparison.

2.5. Infant community state type (ICST) prediction

The hierarchical clustering (HCL) of samples was obtained using bacterial composition at genus level and was calculated through TMeV 4.8.1 software using Pearson correlation as a distance metric based on information at genus level. The data obtained was represented by a cladogram.

3. Results and discussion

3.1. Selection of public datasets

The multi-population cohort meta-analysis performed in this study encompasses data sets corresponding to a total of 1035 faecal samples from healthy, full-term infants ranging from a few hours after birth to 3 years of age (Table 1 and Table S1). These 16S rRNA profiling datasets were selected so as to obtain a high sample number and reach corresponding robust statistical power, in accordance to previous studies [28,29,39]. In detail, we performed an in depth literature search for 16S rRNA profiling datasets based on Illumina sequencing technology, allowing us to retrieve microbiota data from 15 publicly available datasets covering 15 different countries (Table S1). Based on age at sampling, the datasets were subdivided into four macro-groups, i.e. 0–1 M group (<1 month of age), 1–6 M group (between 1 and 6 months of age), 6–12 M group (between 6 and 12 months of age) and 12–36 M group (between 1 and 3 years of age) (Table S1). This subdivision was made on the reported assumption that the infant gut microbiota is variable in composition and constantly evolving during the first three years after birth [40].

3.2. Homogeneity of the samples

16S rRNA gene-based microbiota profiling is the most commonly employed approach to perform gut microbiota profiling. This approach has many advantages, i.e. low cost, high sensitivity and specificity in the identification of bacterial taxonomy and

Table 1
Metadata of samples included in the meta-analysis.

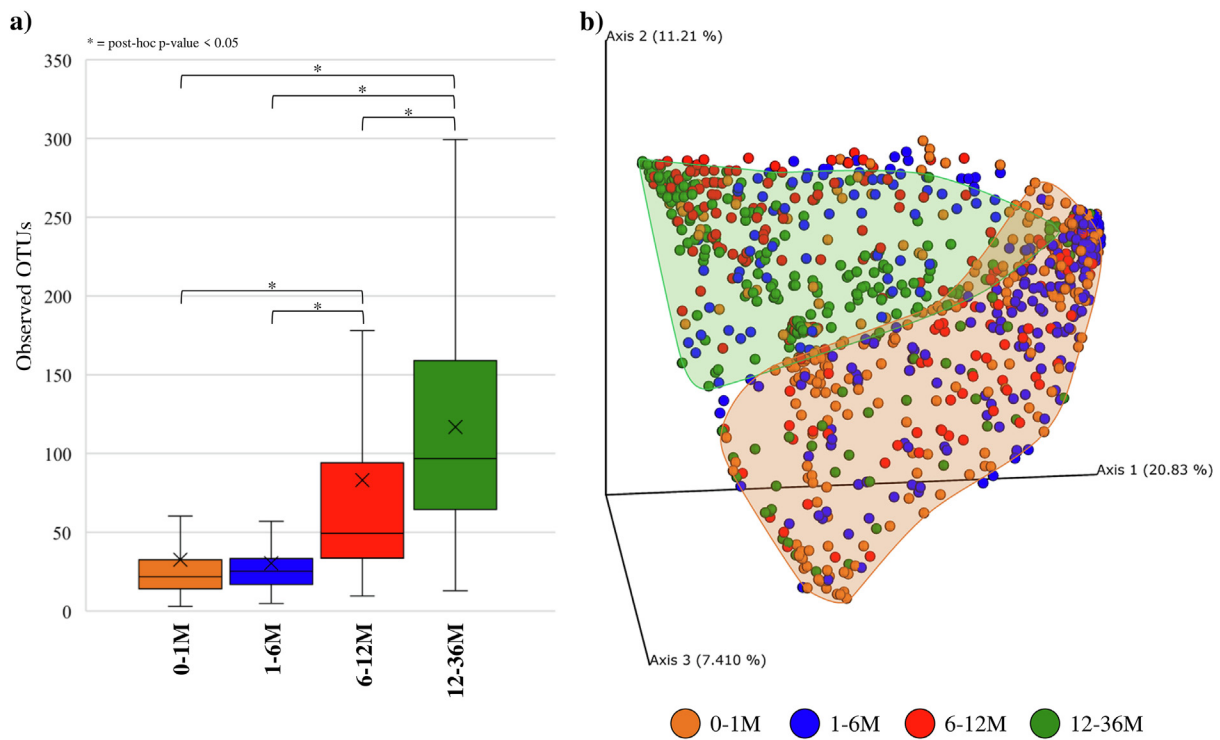
Study (PMID)	Bioproject	Number of samples	Age (days) (mean ± st.dev)	Age (months) (mean ± st.dev)	Nation	Settlement	Socioeconomic status	Technology	Primer region
24896187	PRJEB5482	190	258 ± 233	8 ± 8	Bangladesh	Urban slum	low socioeconomic status	Illumina MiSeq	V1-V2
27160322	PRJNA318053	15	442 ± 262	15 ± 9	Ireland	Urban area	–	Illumina_MiSeq	V4–V5
27306664	PRJEB14529	225	246 ± 207	8 ± 7	USA	Urban area	–	Illumina MiSeq	V4
27717398	PRJNA331150	50	322 ± 104	11 ± 3	USA	Urban area	–	Illumina MiSeq	V4
28079170	PRJEB15633	18	115 ± 34	4 ± 1	Gambia	Urban area	–	Illumina_MiSeq	V4
28095889	PRJNA339264	167	8 ± 38	0 ± 0	Ireland	Urban area	–	Illumina MiSeq	V4–V5
28733284	PRJNA377056	9	383 ± 3	12 ± 0	Italy	Urban area	middle socioeconomic status	Illumina_MiSeq	V3-V4
28789705	PRJNA362530	10	28 ± 6	1 ± 0	Spain	Urban area	–	Illumina_MiSeq	V3
28877893	PRJEB21196	64	118 ± 51	4 ± 2	Germany	Urban area	–	Illumina_MiSeq	V3-V4
29184093	PRJNA403824	47	25 ± 31	1 ± 1	Canada	Urban area	–	Illumina MiSeq	V3
29217369	PRJEB21946	44	70 ± 0	2 ± 0	India	Densely populated Urban area	–	Illumina MiSeq	V4
29795809	PRJEB11419	127	646 ± 266	22 ± 9	USA Thailand United Kingdom Australia Ireland Norway Slovakia Canada	Urban area	–	Illumina MiSeq	V4
29884786	PRJNA450946 PRJNA450998 PRJNA451090 PRJNA451108 PRJNA451156 PRJNA451314 PRJNA451320 PRJNA451359 PRJNA451398 PRJNA451432 PRJNA453125 PRJNA451346	24	186 ± 182	6 ± 6	Norway	Urban area	–	Illumina HiSeq 2500	V4
30131575	PRJNA350676	20	387 ± 0	13 ± 0	Italy	Urban area	–	Illumina MiSeq	V3-V4
30579350	PRJNA491825	25	806 ± 215	27 ± 7	Denmark	Urban area	–	Illumina MiSeq	V3-V4

availability of user-friendly bioinformatic pipelines, all of which have promoted its wide-spread application. Despite these advantages, there is extensive methodological variability, with subsequent output biases that may preclude consistent and meaningful comparisons between datasets obtained from different studies [41]. In particular, one of the main reasons for variable results is that different studies have employed different PCR primer pairs to amplify hypervariable region(s) of the 16S rRNA gene [41]. In order to evaluate the accuracy of the 11 primer pairs employed in the selected datasets (Table S2), we tested the primer pairs through the web-tool TestPrime 1.0 [33,42]. TestPrime allows evaluation of the performance of a given primer pair by running an *in silico* PCR on the SILVA databases. Notably, this analysis showed rather similar amplification performances suggesting homogeneity of the selected datasets (Table S2). In particular, all tested primers showed an *in silico* efficiency of >95% in their ability to amplify the targeted 16S rRNA gene sequences (Table S2). While *in vitro* validation may provide additional evidences, these findings suggest the absence of any major bias in sequencing data generation. In order to confirm this result, we generated artificial communities constituted by genomes of species representative of the bacterial communities typically found in adult and infant feces (Table S3 and Table S4) [43]. These artificial communities were exploited to

obtain PCR amplicons of the primer pairs included in this study and the data sets were processed using the bioinformatics pipeline based on the QIIME 2 software suite, as reported in material and methods section. Remarkably, the results obtained revealed that the profiles generated by the PCR primers pairs included in this study are comparable at the genus level, while, as expected, misclassifications may be observed at species level (Table S3 and Table S4). Thus, highlighting the homogeneity in microbial taxonomic efficiency/accuracy of the PCR primer pairs included in this genus-level meta-analysis. In addition, an identical bioinformatics pipeline based on the QIIME 2 software suite was employed for re-analysis of all the samples in order to allow a reliable and robust cross-meta-analysis of all datasets produced by these different research projects.

3.3. Meta-analysis of full-term infants' microbiota over time

A total of 1035 publicly available datasets from 15 cohorts corresponding to faecal samples of full-term infants were retrieved. These combined cohorts were then divided in 285, 271, 215 and 264 samples belonging to the 0–1 M, 1–6 M, 6–12 M and 12–36 M groups, respectively (Table S1). Quality filtering resulted in a total of 40,014,085 reads with an average of 38,661 ± 48,855



Phylum	Genus	0-1M	1-6M	6-12M	12-36M
Actinobacteria	<i>Bifidobacterium</i>	38.07%	44.16%	20.72%	10.49%
Actinobacteria	<i>Rothia</i>	-	0.27%	-	-
Bacteroidetes	<i>Bacteroides</i>	6.34%	10.47%	16.40%	15.63%
Firmicutes	<i>Staphylococcus</i>	1.34%	-	-	-
Firmicutes	<i>Enterococcus</i>	-	1.02%	-	-
Firmicutes	<i>Lactobacillus</i>	-	1.65%	-	1.61%
Firmicutes	<i>Streptococcus</i>	4.41%	6.70%	3.94%	2.57%
Firmicutes	<i>Clostridium sensu stricto</i> 1	9.51%	2.94%	1.41%	0.92%
Firmicutes	<i>Ruminococcus gnavus</i> group (Lachnospiraceae family)	-	1.91%	1.81%	1.33%
Firmicutes	<i>Ruminococcus torques</i> group (Lachnospiraceae family)	-	-	-	0.67%
Firmicutes	<i>Anaerostipes</i>	-	-	-	0.98%
Firmicutes	<i>Blautia</i>	-	-	1.37%	3.41%
Firmicutes	<i>Lachnoclostridium</i>	-	-	0.72%	0.74%
Firmicutes	Lachnospiraceae UCG-008	-	-	-	0.58%
Firmicutes	U. m. of Lachnospiraceae family	-	-	0.69%	0.60%
Firmicutes	<i>Intestinibacter</i>	-	-	-	0.39%
Firmicutes	<i>Faecalibacterium</i>	-	-	-	5.06%
Firmicutes	<i>Ruminiclostridium</i> 5	-	-	-	0.20%
Firmicutes	<i>Erysipelatoclostridium</i>	-	-	-	0.55%
Firmicutes	<i>Dialister</i>	-	-	-	0.92%
Firmicutes	<i>Veillonella</i>	1.79%	2.47%	1.73%	0.58%
Proteobacteria	<i>Escherichia-Shigella</i>	14.37%	10.12%	6.38%	6.94%
Proteobacteria	<i>Klebsiella</i>	-	3.73%	1.63%	1.02%
Core		56.86%	63.44%	27.10%	13.90%
Accessory		18.98%	21.99%	29.71%	41.30%

Fig. 1. Exploration of the diversity and bacterial composition of healthy, full-term infant faecal samples over time. Panel a shows the box and whisker plot of the alpha-diversity calculated through Observed OTU index of the four age groups. The bottom and top of the box represent the first and third quartiles, and the band inside the box is the median. Moreover, the ends of the whiskers represent the minimum and maximum of all data of the sample. Panel b reports the principal coordinate analysis (PCoA) of the collected infant samples of the four age groups. Panel c displays the bacteria belonging to core- and accessory-microbiota of each age group. The genera belonging to the core are highlighted with a green outline. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

per sample (Table S1). Evaluation of the Good's coverage index resulted in an average value of 0.99 ± 0.006 , thereby indicating that sequencing depth of the assessed samples is adequate for a thorough analysis of the gut microbiota [44] (Table S5).

Alpha-diversity analysis calculated through the Observed OTU, Chao1 and Shannon indexes showed statistically supported complexity differences between samples belonging to each of the four age groups (Fig. 1 and Figure S1). In fact, the 0–1 M and 1–6 M groups showed a statistically significantly lower biodiversity when compared to that of the 6–12 M and 12–36 M groups (ANOVA post-hoc p -value < 0.05 with Observed OTU, Chao1 and Shannon indexes) (Fig. 1a). These data support the notion that the infant gut microbiota gradually diversifies until weaning and subsequently develops towards a microbiota similar to that of adults [45–48]. Remarkably, these results also underline the marked impact of weaning in shaping the evolution of the gut microbiota.

Analysis of the beta-diversity represented through Principal Coordinate Analysis (PCoA) based on Bray-Curtis dissimilarity revealed clear compositional differences between the four age-based groups (Fig. 1b). In particular, most of the samples belonging to 0–1 M and 12–36 M age sub-groups cluster separately (PERMANOVA p -value of < 0.001), while samples belonging to the 1–6 M and 6–12 M groups were shown to exhibit high inter-variability. Such findings indicate that the microbiota is constantly changing over time. These results were statistically validated by a pairwise PERMANOVA p -value of < 0.001 .

In order to identify the main taxonomic differences between the gut microbiota of the 0–1 M, 1–6 M, 6–12 M and 12–36 M age-differentiated groups, taxonomic profiling at genus level was employed to reconstruct the core elements of the gut microbiota of each group [49]. This so-called core-microbiota was obtained by selecting bacterial genera that occur with a prevalence $> 70\%$ among collected samples (of a particular age group) in at least 70% of the studies (Tables S6, S7, S8 and S7). Moreover, bacterial genera present in at least 70% of the studies and with a prevalence ranging from 30% to 70% were used to reconstruct the accessory-microbiota (Tables S6, S7, S8 and S9). The cut-off values were chosen based on previous publications [50,51] and adapted to the high variability of the infant microbiota. Interestingly, the achieved results revealed that the core-microbiota (taxa with prevalence $> 70\%$) of 0–1 M and 1–6 M groups was mostly represented by the *Bifidobacterium* genus (average abundance of $38.07\% \pm 34\%$ and $44.16\% \pm 29.10\%$, respectively) (Fig. 1c, Tables S6 and S7). Conversely, the core-microbiota of 6–12 M and 12–36 M groups showed a relative decrease in this bacterial genus (average relative abundance of $20.72\% \pm 22.41\%$ and $10.49\% \pm 15.81\%$, respectively) (p -value < 0.05) and a gradual increase in the core microbial complexity (Fig. 1c, Tables S8 and S9) characterized by the presence of *Bacteroides*, *Feacalibacterium*, *Blautia* and *Ruminococcus* genera (p -value < 0.05) (Fig. 1c). Furthermore, analysis of the data sets corresponding to the infant group younger than six months showed that the taxa with prevalence $> 70\%$ in at least 70% of the studies, i.e. core-microbiota, represents on average of $60.07\% \pm 29.95\%$ of the total identified bacterial community, while in infants aged between 12 and 36 months the core-microbiota comprises on average $27.19\% \pm 26.68\%$ of the total bacterial community. Consequently, during the first 36 months following birth the gut microbiota of a newborn undergoes a progressive increase in accessory bacterial genera with prevalence ranging from 30% to 70%, corresponding to a reduction in the total relative abundance of the core microbiota.

3.4. Birth delivery mode

Several studies have highlighted the impact of delivery mode on the infant microbiota composition [1]. In order to identify possible

differences between subjects born by natural or Caesarean delivery (C-section), samples within each age-differentiated group were further divided according to delivery method (Tables S10, S11, S12, S13, S14, S15, S16 and S17). Evaluation of microbial diversity through alpha-diversity (Fig. 2a) showed significant differences in 1–6 M, 6–12 M and 12–36 M age groups (p -value 0.01). In detail, the gut microbiota of infants born by C-section was shown to elicit a greater microbial complexity compared to vaginally born babies in these different age-based sub-groups (Fig. 2a). This variation can be explained by the contact of C-section newborns with environmental bacteria, resulting from contact of the baby with the mother's skin and hospital environment, which will colonize and affect the composition of the infant gut microbiota [52]. Furthermore, lack of exposure of the newborn to the mother's vaginal/rectal microbiota may cause reduced mother-to-baby bacterial transfer, and allow skin/hospital-associated microbes to colonize in the absence of competition [1].

Analysis of beta-diversity, based on Bray-Curtis dissimilarity and represented through PCoA plots, revealed a spatial division based on delivery mode for all four age sub-groups. Furthermore, these results were confirmed by PERMANOVA analysis (p -value < 0.01) (Fig. 2b).

Scrutiny of the bacterial composition of babies born by C-section indicated a microbiota characterized by the presence of microorganisms that typically colonize hospital environments and the human skin. In detail, C-section delivered infants belonging to the 0–1 M group revealed a 197.98% higher abundance of *Clostridium sensu stricto* 1 genus compared to those born vaginally (p -value < 0.05) (Table S18). Similarly, a 265.78% (p -value < 0.01) higher abundance of the *Escherichia-Shigella* genus was observed in subjects born by C-section within the 6–12 M group (Table S18). In contrast, infants born by natural delivery showed in all age groups a higher abundance of the *Bacteroides* genus compared to C-section delivered infants, being 149.91%, 682.14%, 166.64% and 53.74% more abundant in 0–1 M, 1–6 M, 6–12 M and 12–36 M groups, respectively (p -value 0.01) (Table S18). As highlighted in previous studies, the lower abundance of *Bacteroides* in infants born by C-section (compared to naturally delivered infants), with a corresponding increase in other taxa, may contribute to an increased risk in the development of inflammation and obesity [14,53–55]. Focusing on the *Bifidobacterium* genus, one of the most representative bacteria in all four age groups, significant differences are observed only in the 6–12 M group (Table S18) suggesting rather minor, if any impact of the type of delivery on this bacterial genus.

3.5. Influence of feeding on the infant microbiota

Breastfeeding, as opposed to the use of formula milk, is known to modulate the gut microbial population. In fact, the oligosaccharides present in human milk (i.e. human milk oligosaccharides or HMOs) cannot be digested by the infant itself and are metabolized by certain bacteria present in the infant gut, as these bacteria possess and express the necessary genetic repertoire to allow HMO metabolism, thus facilitating the selection of specific intestinal bacterial species [56]. In order to highlight possible differences between breastfeeding and formula feeding, samples within each age-differentiated group were divided by type of feeding (Tables S19, S20, S21, S22, S23, S24, S25 and S26). Evaluation of alpha-diversity based on the Observed OTU index did indeed reveal significant microbiota compositional differences between breastfed babies compared to infants fed with formula milk, but such differences reached statistical significance only in the 0–1 M and 6–12 M age groups (p -value < 0.01) (Fig. 3a). These results indicate that the complexity of the microbiota, at least at genus level, is not strictly related to the feeding type, but is more likely to be subject

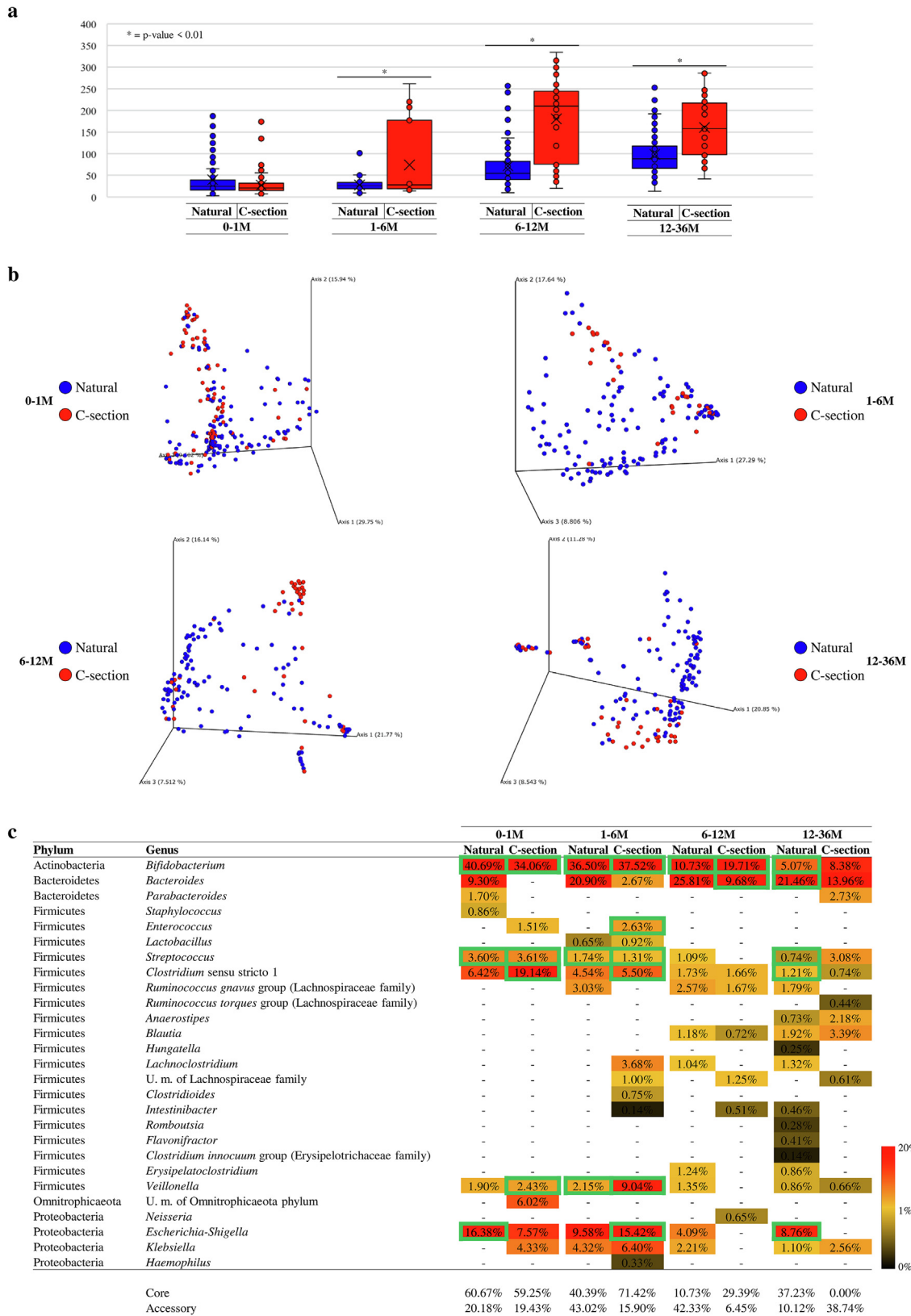


Fig. 2. Evaluation of the diversity and bacterial composition of healthy, full-term infant faecal samples as based on birth delivery mode. Panel a indicates the box and whisker plot of the infant alpha-diversity calculated through Observed OTU index based on birth delivery methods. The bottom and top of the box represent the first and third quartiles, and the band inside the box is the median. Moreover, the ends of the whiskers represent the minimum and maximum of all data of the sample. Panel b reports the principal coordinate analysis (PCoA) of the collected infant samples, respectively, subdivided by age and birth delivery mode. Panel c displays the bacteria belonging to core- and accessory-microbiota of each delivery mode groups. The genera belonging to the core are highlighted with a green outline. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

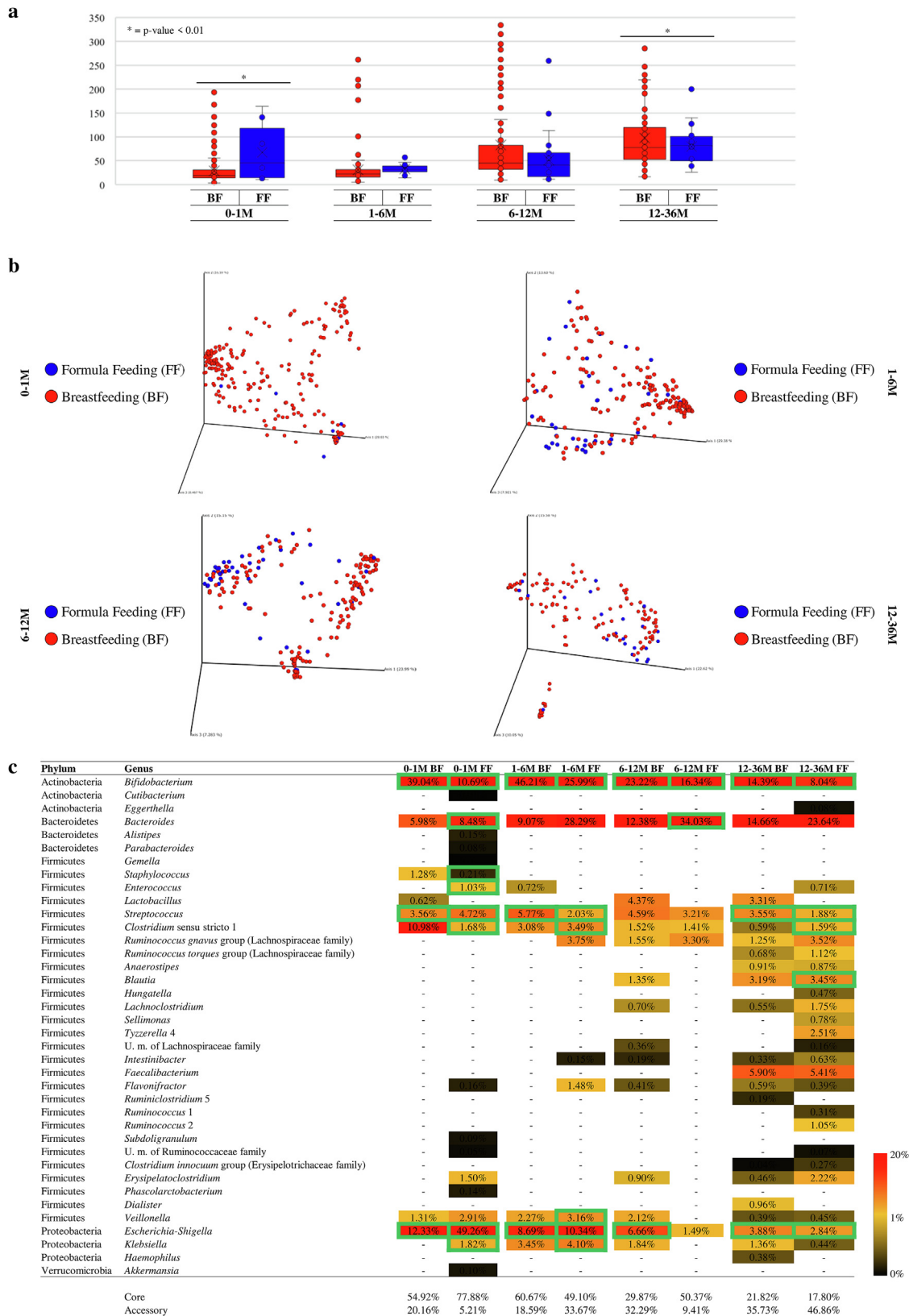


Fig. 3. Examination of the diversity and bacterial composition of healthy, full-term infant faecal samples based on type of feeding. Panel a shows the box and whisker plot of the infant alpha-diversity calculated through Observed OTU index based on type of feeding. The bottom and top of the box represent the first and third quartiles, and the band inside the box is the median. Moreover, the ends of the whiskers represent the minimum and maximum of all the data of the sample. Panel b reports the principal coordinate analysis (PCoA) of the collected infant samples, respectively subdivided by age and type of feeding. Panel c displays the bacteria belonging to core- and accessory-microbiota of each age group. The genera belonging to the core are highlighted with a green outline. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to major changes by the transition to a semi-solid/solid diet that typically occurs at an average age of 6 months [57,58]. Biodiversity analysis revealed significant differences between the breastfed babies compared to infants fed with formula milk in the four distinct age groups (PERMANOVA < 0.05) (Fig. 3b).

In order to identify possible differences in gut microbiota composition between breastfed and formula-fed infants for the four defined age-based macro-groups, the taxonomic changes at genus level were analyzed. This analysis showed that breastfed children possess higher levels of the *Bifidobacterium* genus in their gut microbiota when compared to age-matched children fed with formulated milk, with significant differences in all age groups (p -value < 0.01) (Table S27). Furthermore, the *Escherichia-Shigella* genus showed an average relative abundance of $13.38\% \pm 22.85\%$ in breastfed infants of 0–1 M group, while subjects fed with formula milk revealed an average relative abundance of $35.25\% \pm 37.32\%$ (p -value < 0.01). Additionally, significant differences have been highlighted for the genus *Bacteroides*, which showed the highest levels of relative abundance in the babies fed with formula milk. In detail, comparisons between breastfed and formula-fed infants showed a significant absolute difference (p -value < 0.01) of 19.22%, 21.26% and 16.83% in 1–6 M, 6–12 M and 12–36 M groups, respectively. These results show that the gut microbiota of breastfed infants when compared to that of infants nourished with formula milk possess a higher relative abundance of *Bifidobacterium* [59], confirming the importance of breast milk intake in modulating the gut microbiota during the early stages of life toward presumed health-promoting microbial taxa [60].

3.6. Identification of ICSTs

In 2011 Arumugam et al. [48] proposed that the gut microbiota of adults can be stratified into three distinct clusters driven by discriminative genera. These three clusters, also called enterotypes, include *Bacteroides* (enterotype 1), *Prevotella* (enterotype 2), and *Ruminococcus* (enterotype 3) [48]. Furthermore, a recent study performed an analysis of the enterotypes of children at school-going age by identifying three distinct clusters, dominated by *Bacteroides*, *Prevotella*, and *Bifidobacterium*, respectively [61]. Moreover, a longitudinal study on neonatal gut suggested the possible existence of community state types (CSTs) characterized by the presence of Enterobacteriaceae, *Veillonella*, *Ruminococcus*, *Streptococcus*, *Prevotella*, *Bacteroides*, and *Bifidobacterium* [62]. Consequently, despite the extreme dynamism of the infant gut microbiota that had previously prevented identification of taxonomic patterns and cluster-discriminative genera, we exploited the statistical robustness offered by our extensive meta-analysis to investigate the possible existence of infant community state types (ICSTs) that correspond to the early stages of life. In this context, the distinctly higher microbial gut community complexity of infants aged between 12 and 36 months when compared to that of children less than six months old (see meta-analysis discussed above), prompted us to divide the samples into two macro groups, i.e. a 0–6 M group (<6 months) and a 6–36 M group (>6 months). Interestingly, these two macro groups seem to correspond to the division between pre- and post-weaning infants indicated by the guidelines of the World Health Organization (WHO) [58] and reported in previous studies [57].

Screening for ICSTs present in these two groups was performed by cluster analysis through hierarchical clustering (HCL) (Fig. 4a and 5a) and 3D Bray Curtis PCoA (Fig. 4b and 5b) of the collected samples as based on their 16S rRNA gene-derived microbial profiles (Tables S28 and S29). Each statistically supported cluster had to be represented by at least 10 samples to be considered as a putative ICST. Moreover, the representative taxa of each proposed ICST had to be present at an average abundance and preva-

lence higher than 15% and 90%, respectively (Fig. 4c). Results were verified by PCoA analyses followed by statistical validation through PERMANOVA (p -value < 0.05).

Analysis of the 0–6 M samples allowed us to identify five main infant community state types (0–6 M-ICST-*Clostr*, -*Strept*, -*Bif*, -*Esch* and -*Kleb*) (Fig. 4), while samples from the 6–36 M macro-group revealed seven main ICSTs (6–36 M-ICST-*Bharg*, -*Odor*, -*Prev*, -*Esch*, -*Bif*, -*Faec* and -*Bact*) (Fig. 4). Interestingly, the high number of detected ICSTs is presumed to reflect the high variability and dynamism of the microbiota at the early stages of life [63]. Particularly, the 0–6 M-ICSTs were shown to be dominated by the *Bifidobacterium* genus (0–6 M-ICST-*Bif*), which is present with an abundance average higher than 15% in 70% of the samples (Fig. 4a). In detail, 174 samples of this bifidobacteria-dominated ICST (0–6 M-ICST-*Bif*) contained high levels of members of the *Bifidobacterium* genus (0–6 M-ICST-*Bif1*, bifidobacterial average abundance of 76.98%), while the 224 remaining samples were characterized by the combination of *Bifidobacterium* genus with other bacterial taxa (0–6 M-ICST-*Bif2*, -*Bif3*, -*Bif4*, -*Bif5*, -*Bif6* and -*Bif7*) (Fig. 4a). Conversely, the 6–36 M groups showed a decrease in community state types characterized by the presence of the genus *Bifidobacterium* (6–36 M-ICST-*Bif* main community state types encompassing 24.84% of the samples) and an increase of ICSTs characterized by the genus *Bacteroides* (6–36 M-ICST-*Bact* main community state types encompassing 37.16% of the samples) (Fig. 5a). The community state type characterized by the dominance of the *Bacteroides* genus (6–36 M-ICST-*Bact1*) and ICSTs dominated by typical adult bacterial genera, such as *Faecalibacterium* (6–36 M-ICST-*Faec*), highlight the shift of the infant's microbiota towards a bacterial community similar to that of adults.

Previous studies have reported that ethnicity in the early stages of life plays a major role in the evolution of the gut microbiota [64–66]. Since the metadata accompanying the 16S rRNA microbial profiling data used in this study do not provide any detail about ethnicities, we decided to investigate possible correlations between the metagenomic data and the geographical origins of the samples. Analysis of the 0–6 M samples revealed that most nations are characterized by high prevalence of 0–6 M-ICST-*Bif* (prevalence in each nation >30%) except for the Spanish samples that seem most represented by 0–6 M-ICST-*Esch* (Table 2). Conversely, samples from the 6–36 M macro-group showed heterogeneity in the distribution of ICSTs. In detail, Danish, German, Norwegian and US samples seem to be characterized by high prevalence of 6–36 M-ICST-*Bact* (>30%), while Italian and English samples showed prevalence >30% of 6–36 M-ICST-*Bif1* and 6–36 M-ICST-*Odor*, respectively (Table 3). These results seem to underline a possible correlation between geographical origin and infant gut microbiota composition, especially in samples older than 6 months probably due to the transition to a solid diet. A specific study regarding geographical origin, race and socioeconomic status categories of infants could provide a much sharper view of their role in the maturation of the gut microbiome.

In addition, we evaluated co-variance between the bacterial genera with a total average abundance of > 0.5%. In detail, covariance analysis based on 0–6 M group samples (Table S30) revealed that the *Bifidobacterium* genus negatively correlates with most other taxa that dominate in non-bifidobacterial ICSTs, such as *Clostridium sensu stricto 1* (0–6 M-ICST-*Clostr*), *Streptococcus* (0–6 M-ICST-*Strept*), *Escherichia-Shigella* (0–6 M-ICST-*Esch*) and *Klebsiella* (0–6 M-ICST-*Kleb*). Similarly, analysis of samples corresponding to 6–36 M infants showed a negative correlation between the *Bifidobacterium* and *Bacteroides* genera, which constitute the most representative community state types, yet revealed a positive correlation between typical adult taxa, such as *Bacteroides*, *Faecalibacterium* and *Alistipes* (Table S31). These results support the notion that during infancy *Bifidobacterium* represents

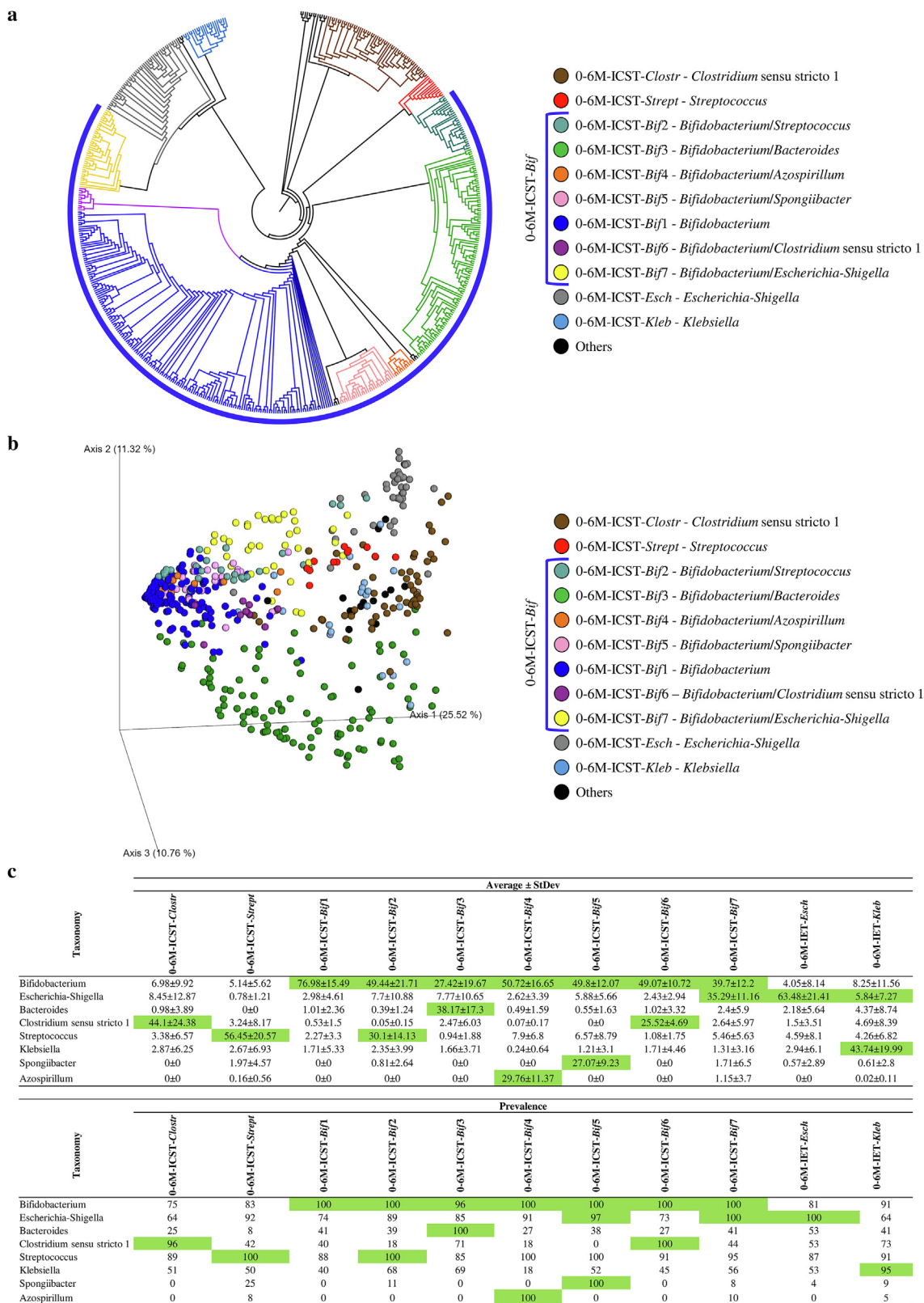
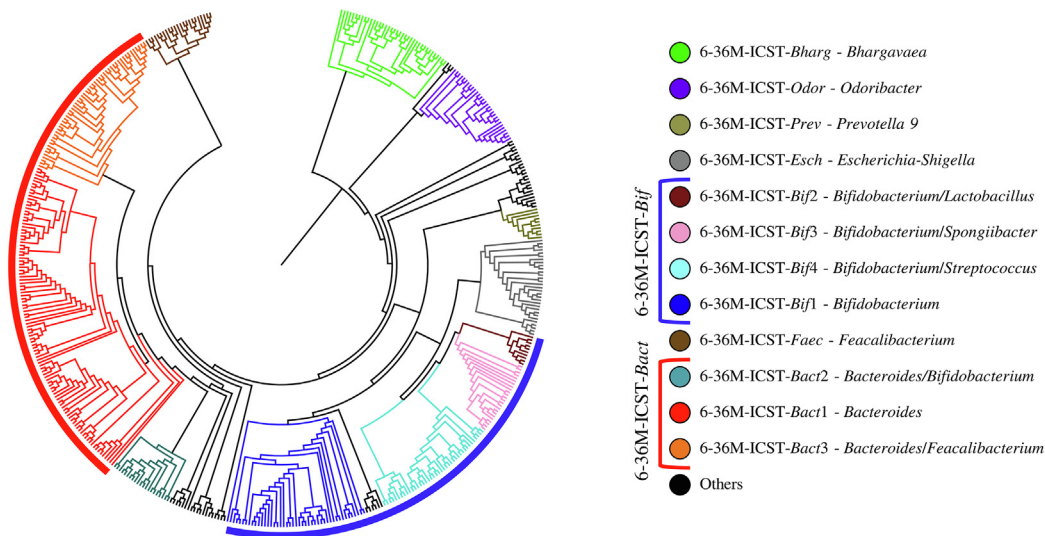
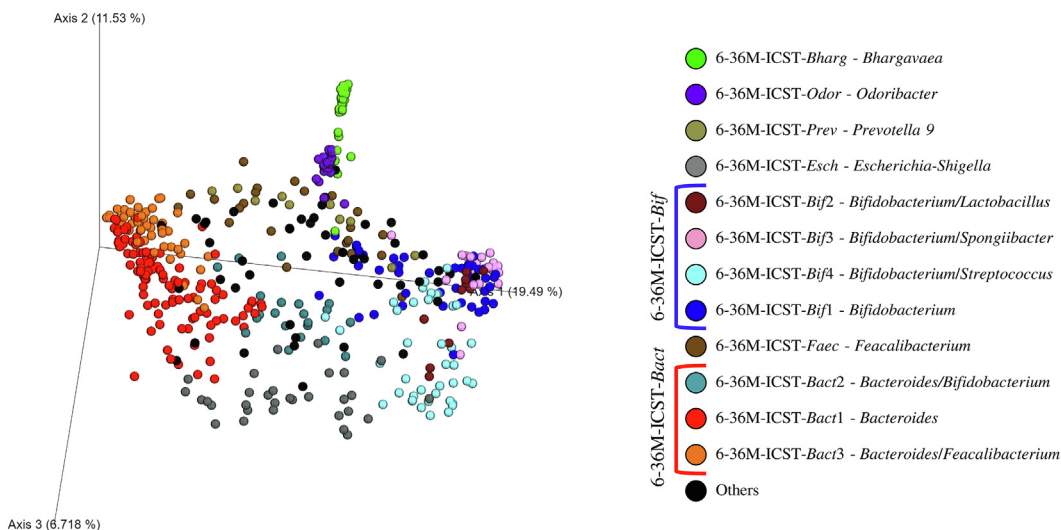


Fig. 4. Identification of 0–6 M group infant ICSTs. Panel a shows a circular cladogram of the samples belonging to the 0–6 M group, obtained by means of hierarchical clustering (HCL) analysis. Panel b reports the principal coordinate analysis (PCoA) of the 0–6 M infant samples, subdivided by ICST. Panel c displays the average abundance and prevalence of bacteria that correspond to an identified ICST. Values that represent the ICSTs cut-off are highlighted in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a



b



c

Taxonomy	Average ± StDev											
	6-36M-ICST-Bharg	6-36M-ICST-Odor	6-36M-ICST-Prev	6-36M-ICST-Esch	6-36M-ICST-Bif1	6-36M-ICST-Bif2	6-36M-ICST-Bif3	6-36M-ICST-Bif4	6-36M-ICST-Faec	6-36M-ICST-Bact1	6-36M-ICST-Bact2	6-36M-ICST-Bact3
<i>Bacteroides</i>	0±0	0±0	3.16±4.22	6.62±7.41	0.92±2.49	0.11±0.18	0.12±0.3	0.59±1.83	2.38±4.58	46.16±13.62	23.88±7.58	32.38±8.65
<i>Bifidobacterium</i>	0±0	0±0	6.56±7.56	4.5±6.18	48.01±24.39	28.61±7.12	48.68±14.2	29.27±13.02	9.06±7.89	7.18±7.83	31.92±12.06	2.98±3.68
<i>Escherichia-Shigella</i>	0.62±2.44	0±0.01	0.49±1.36	52.31±17.91	1.98±3.32	5.75±7.12	1.57±3.19	14.84±10.93	1.96±2.39	4.01±6.46	3.41±3.17	0.94±2.26
<i>Faecalibacterium</i>	0±0	0.03±0.17	5.73±3.18	0.87±1.49	0.59±1	0.48±1.11	0.8±1.71	0.4±1.49	21.74±10.34	3.41±4.35	1.51±3.4	16.12±7.65
<i>Streptococcus</i>	0±0	0.42±2.23	3.19±7.05	3.51±8.59	3.81±4.66	6.93±4.48	4.26±5.64	18.29±13.26	2.58±4.37	0.99±1.62	1.77±2.79	0.3±0.67
<i>Bhargavaea</i>	35.6±22.38	0±0	0±0	0±0	0±0	0±0	0±0	0±0	0±0	0±0	0±0	0±0
<i>Lactobacillus</i>	0.15±0.94	0±0	1.62±3.13	0.69±1.85	4.56±7.71	30.57±8.29	4.23±3.75	5.79±5.95	1.62±2.75	0.05±0.12	0.48±1.18	0.04±0.23
<i>Spongibacter</i>	0±0	0±0	0±0	0.71±3.5	0±0	7.8±8.52	24.42±6.09	1.51±4.45	2±3.96	0±0	0±0	0±0
<i>Prevotella 9</i>	0±0	0.36±1.89	36.97±13.5	0.32±0.86	0.84±2.34	1.37±3.47	1.46±3.02	0.59±2.88	2.36±5.45	0.12±0.92	0.01±0.04	1.01±3.13
<i>Odoribacter</i>	0±0.01	15.86±14.53	0.07±0.12	0.03±0.08	0±0.02	0±0	0±0	0±0	0.01±0.03	0.02±0.1	0±0	0.07±0.24

Taxonomy	Prevalence											
	6-36M-ICST-Bharg	6-36M-ICST-Odor	6-36M-ICST-Prev	6-36M-ICST-Esch	6-36M-ICST-Bif1	6-36M-ICST-Bif2	6-36M-ICST-Bif3	6-36M-ICST-Bif4	6-36M-ICST-Faec	6-36M-ICST-Bact1	6-36M-ICST-Bact2	6-36M-ICST-Bact3
<i>Bacteroides</i>	0	3	50	94	47	60	25	65	65	100	100	100
<i>Bifidobacterium</i>	0	0	90	74	100	100	100	100	100	89	100	92
<i>Escherichia-Shigella</i>	50	3	50	100	80	90	88	95	100	94	90	58
<i>Faecalibacterium</i>	0	3	100	65	51	40	42	20	100	63	33	100
<i>Streptococcus</i>	0	17	80	81	91	100	96	98	96	83	86	60
<i>Bhargavaea</i>	100	0	0	0	0	0	0	0	0	0	0	0
<i>Lactobacillus</i>	5	0	50	39	73	100	96	90	48	35	57	13
<i>Spongibacter</i>	0	0	0	6	0	50	100	13	26	0	0	0
<i>Prevotella 9</i>	0	17	100	35	33	40	46	18	39	11	5	19
<i>Odoribacter</i>	3	93	30	19	4	0	0	8	9	10	0	17

Table 2

Evaluation of distribution of 0–6 M-ICSTs in nations included in the study. 0–6 M-ICSTs with a presence greater than 30% are highlighted in green.

	Bangladesh (N=97)	Canada (N=47)	Gambia (N=18)	Germany (N=49)	India (N=44)	Ireland (N=170)	Norway (N=12)	Spain (N=10)	USA (N=109)
0-6M-ICST- <i>Clostr</i>	1%	9%	-	6%	5%	22%	17%	-	6%
0-6M-ICST- <i>Strept</i>	5%	4%	-	-	5%	1%	-	-	1%
0-6M-ICST- <i>Bif1</i>	20%	38%	72%	49%	27%	38%	8%	20%	7%
0-6M-ICST- <i>Bif2</i>	8%	6%	22%	-	25%	1%	-	-	1%
0-6M-ICST- <i>Bif3</i>	-	6%	-	20%	5%	16%	67%	20%	50%
0-6M-ICST- <i>Bif4</i>	11%	-	-	-	-	-	-	-	-
0-6M-ICST- <i>Bif5</i>	30%	-	-	-	-	-	-	-	-
0-6M-ICST- <i>Bif6</i>	-	-	-	4%	-	4%	-	-	2%
0-6M-ICST- <i>Bif7</i>	11%	15%	6%	8%	30%	7%	-	20%	7%
0-6M-ICST- <i>Esch</i>	9%	13%	-	6%	2%	4%	-	30%	17%
0-6M-ICST- <i>Kleb</i>	4%	2%	-	4%	2%	4%	8%	-	6%
Others	-	6%	-	2%	-	4%	-	10%	5%

Table 3

Assessment of distribution of 6–36 M-ICSTs in nations included in the study. 6–36 M-ICSTs with a presence greater than 30% are highlighted in green.

	Bangladesh (N=97)	Canada (N=47)	Gambia (N=18)	Germany (N=49)	India (N=44)	Ireland (N=170)	Norway (N=12)	Spain (N=10)	USA (N=109)
0-6M-ICST- <i>Clostr</i>	1%	9%	-	6%	5%	22%	17%	-	6%
0-6M-ICST- <i>Strept</i>	5%	4%	-	-	5%	1%	-	-	1%
0-6M-ICST- <i>Bif1</i>	20%	38%	72%	49%	27%	38%	8%	20%	7%
0-6M-ICST- <i>Bif2</i>	8%	6%	22%	-	25%	1%	-	-	1%
0-6M-ICST- <i>Bif3</i>	-	6%	-	20%	5%	16%	67%	20%	50%
0-6M-ICST- <i>Bif4</i>	11%	-	-	-	-	-	-	-	-
0-6M-ICST- <i>Bif5</i>	30%	-	-	-	-	-	-	-	-
0-6M-ICST- <i>Bif6</i>	-	-	-	4%	-	4%	-	-	2%
0-6M-ICST- <i>Bif7</i>	11%	15%	6%	8%	30%	7%	-	20%	7%
0-6M-ICST- <i>Esch</i>	9%	13%	-	6%	2%	4%	-	30%	17%
0-6M-ICST- <i>Kleb</i>	4%	2%	-	4%	2%	4%	8%	-	6%
Others	-	6%	-	2%	-	4%	-	10%	5%

the most dominant genus and is replaced during weaning by bacteria typical of an adult microbiota.

3.7. Evaluation of ICSTs corresponding to delivery mode and feeding type

The infant microbiota composition appears to be predominantly influenced by the delivery mode and feeding type (see above). We therefore decided to evaluate if and how ICSTs may be modulated by these two factors. This analysis as based on 0–6 M group samples showed that infants born by vaginal delivery were mainly rep-

resented by the 0–6 M-ICST-*Bif1* (31.16% of the total samples) and 0–6 M-ICST-*Bif3* (33.70% of the total samples) community state types, which are dominated by the *Bifidobacterium* genus (Table 4). Similarly, samples from babies born by C-section stratified within the bifidobacterial community state type 0–6 M-ICST-*Bif1* (33.63% of the total samples), yet were also assigned to the community state type 0–6 M-ICST-*Clostr* (29.20% of the total samples) were dominated by the environmental *Clostridium sensu stricto* 1 (Table 4). Interestingly, samples from infants born by natural delivery and belonging to the 6–36 M group were shown to be frequently assigned to the *Bacteroides* community state types 6–

Fig. 5. Identification of infant ICSTs based on the microbiota of faecal samples associated with the 6–36 M group. Panel a illustrates a circular cladogram of the samples belonging to the 6–36 M group, based on hierarchical clustering (HCL) analysis. Panel b shows the principal coordinate analysis (PCoA) of the 6–36 M samples, subdivided by ICSTs. Panel c reports the average abundance and prevalence of the bacteria that correspond to an identified ICST. The values that represent the ICST cut-off are highlighted in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Evaluation of the ICSTs based on delivery.

	n° samples	0–6 M-ICST- <i>Clostr</i>	0–6 M-ICST- <i>Strept</i>	0–6 M-ICST- <i>Bif1</i>	0–6 M-ICST- <i>Bif2</i>	0–6 M-ICST- <i>Bif3</i>	0–6 M-ICST- <i>Bif4</i>	0–6 M-ICST- <i>Bif5</i>	0–6 M-ICST- <i>Bif6</i>	0–6 M-ICST- <i>Bif7</i>	0–6 M-ICST- <i>Esch</i>	0–6 M-ICST- <i>Kleb</i>
Natural	276	6.88%	1.09%	31.16%	1.81%	33.70%	0.00%	0.00%	2.54%	5.07%	9.06%	3.26%
C-section	113	29.20%	1.77%	33.63%	0.00%	8.85%	0.00%	0.00%	3.54%	3.54%	8.85%	7.08%
Breastfeeding (BF)	393	11.70%	1.78%	32.32%	3.56%	17.30%	2.80%	7.38%	2.54%	4.58%	7.89%	4.58%
Formula Feeding (FF)	45	2.22%	2.22%	11.11%	0.00%	53.33%	0.00%	0.00%	0.00%	8.89%	15.56%	4.44%

Table 5
Valuation of the ICSTs based on feeding.

	n° samples	6–36 M-ICST- <i>Bharg</i>	6–36 M-ICST- <i>Odor</i>	6–36 M-ICST- <i>Prev</i>	6–36 M-ICST- <i>Esch</i>	6–36 M-ICST- <i>Bif1</i>	6–36 M-ICST- <i>Bif2</i>	6–36 M-ICST- <i>Bif3</i>	6–36 M-ICST- <i>Bif4</i>	6–36 M-ICST- <i>Faec</i>	6–36 M-ICST- <i>Bact1</i>	6–36 M-ICST- <i>Bact2</i>	6–36 M-ICST- <i>Bact3</i>
Natural	229	9.91%	7.76%	1.29%	6.03%	1.72%	0.00%	0.00%	2.16%	5.17%	37.07%	6.90%	12.07%
C-section	89	15.56%	11.11%	1.11%	11.11%	3.33%	3.33%	0.00%	23.33%	0.00%	15.56%	5.56%	4.44%
Breastfeeding (BF)	264	10.07%	5.97%	1.49%	2.99%	8.96%	3.36%	7.84%	11.57%	6.72%	19.03%	5.97%	7.09%
Formula Feeding (FF)	65	7.58%	1.52%	1.52%	3.03%	4.55%	1.52%	4.55%	3.03%	3.03%	48.48%	6.06%	12.12%

36 M-ICST-*Bact1* (37.07% of the total samples) and 6–36 M-ICST-*Bact3* (12.07% of the total samples), while C-section infants were not associated with a dominant ICST except for 6–36 M-ICST-*Bif4* (23.33% of the total samples), which is characterized by *Bifidobacterium* and *Streptococcus* genera (Table 5).

Focusing on the correlation between infant community state types and the mode of feeding, the 0–6 M breast-fed samples were shown to be frequently assigned to the community state types 0–6 M-ICST-*Bif1* (32.32% of the total samples) or 0–6 M-ICST-*Bif3* (17.30% of the total samples) being characterized by dominance of the *Bifidobacterium* and *Bacteroides* genera, respectively, while samples from formula-fed infants were most commonly assigned to the community state type 0–6 M-ICST-*Bif3* (53.33% of the total samples) (Table 4). Instead, the 6–36 M breast-fed samples revealed a high heterogeneity in the distribution of ICSTs, while samples from formula-fed babies showed a prevalence of the *Bacteroides* community state type 6–36 M-ICST-*Bact1* (48.48% of the total samples) (Table 5). These results confirm the impact of the feeding type, i.e. breast- and formula-feeding, on the infant microbiota, suggesting a close correlation between formula-fed babies and the *Bacteroides* ICST.

4. Conclusions

The human gastrointestinal tract is colonized by millions of microorganisms collectively defined as microbiota. In this context, several studies have shown that appropriate development of the intestinal microbiota during the early stages of life may positively influence human health.

In order to investigate changes in the intestinal microbiota during the early stages of life in human, we performed a multi-population cohort meta-analysis of a total of 1035 faecal samples of healthy, full-term infants ranging from a few hours following birth to 3 years of age. The multi-population cohort meta-analysis revealed a higher microbial complexity in samples from infants aged between 12 and 36 months compared to those from children less than one month old. This increase in biodiversity

appears to have caused a reduction in the abundance of the *Bifidobacterium* genus in favor of *Bacteroides*, *Feacalibacterium*, *Blautia* and *Ruminococcus* genera. The multi-population cohort meta-analysis allowed us to observe higher complexity in subjects born by C-section compared to those born by natural delivery. In addition, babies born by C-section were associated with a bacterial colonization characterized by genera commonly present in the environment and/or bacterial genera that may be pathogenic. The multi-population cohort meta-analysis also revealed a positive correlation between breastfeeding and dominance of bifidobacteria in the corresponding gut microbiota.

The statistical robustness offered by our extensive meta-analysis allowed us to identify infant community state types (ICSTs) highlighting that healthy children less than six months old are characterized by the predominance of the ICST with high relative abundance of the *Bifidobacterium* genus in their gut microbiota, while infants aged between 12 and 36 months are characterized by typically adult bacterial genera, such as *Bacteroides* and *Feacalibacterium*. Moreover, analysis of ICSTs highlighted a strong correlation between formula-fed infants and the development and maintenance of a microbiota with a predominance of the *Bacteroides* genus. Certainly, a deep meta-analysis of datasets based on whole-metagenome shotgun (WMS) or built on internal transcribed spacer (ITS)-based profiling, will offer a more detailed species level investigation and will no doubt allow the confirmation of these infant community state types described here and the discovery of further ICSTs or sub-ICSTs. Furthermore, our results highlight the need for additional studies accompanied by accurate metadata that will be crucial in order to precisely determine the factors that influence the evolution of the infant gut microbiota. Moreover, due to the absence of longitudinal studies involving metagenomic data across the 3 years, it was not possible to evaluate if these identified ICSTs remain stable over time. In this context, a specific longitudinal study spanning a wider age range may be important to clarify the evolution and stability of the microbiota in relation to the identified community state types and highlight possible correlations between the ICSTs of infants and adults.

CRedit authorship contribution statement

Leonardo Mancabelli: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Visualization, Writing - original draft. **Chiara Tarracchini:** Methodology, Investigation, Data curation, Visualization. **Christian Milani:** Methodology, Supervision, Writing - review & editing. **Gabriele Andrea Lugli:** Conceptualization, Software. **Federico Fontana:** Data curation. **Francesca Turroni:** Conceptualization, Supervision. **Douwe Sinderen:** Conceptualization, Supervision, Writing - review & editing. **Marco Ventura:** Conceptualization, Supervision, Project administration, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Part of this research is conducted using the High Performance Computing (HPC) facility of the University of Parma. This work was funded by the EU Joint Programming Initiative – A Healthy Diet for a Healthy Life (JPI HDHL, <http://www.healthydietforhealthylife.eu/>) to MV and DvS (Grant no. 15/JPI/HDHL/3280), and the MIUR to MV. We thank GenProbio srl for financial support of the Laboratory of Probiogenomics. DvS is a member of APC Microbiome Ireland funded by Science Foundation Ireland (SFI), through the Irish Government's National Development Plan (Grant no's. SFI/12/RC/2273-P1 and SFI/12/RC/2273-P2).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2020.08.028>.

References

- Milani C, Duranti S, Bottacini F, Casey E, Turroni F, Mahony J, et al. The first microbial colonizers of the human gut: composition, activities, and health implications of the infant gut microbiota. *Microbiol Mol Biol Rev* 2017;81(4).
- Ventura M, Milani C, Lugli GA, van Sinderen D. Health benefits conferred by the human gut microbiota during infancy. *Microb Biotechnol* 2019;12(2):243–8.
- Fouhy F, Ross RP, Fitzgerald GF, Stanton C, Cotter PD. Composition of the early intestinal microbiota: knowledge, knowledge gaps and the use of high-throughput sequencing to address these gaps. *Gut Microbes* 2012;3(3):203–20.
- Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, et al. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* 2019;7(1).
- Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, et al. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* 2010;107(26):11971–5.
- Biasucci G, Rubini M, Riboni S, Morelli L, Bessi E, Retetangos C. Mode of delivery affects the bacterial community in the newborn gut. *Early Hum Dev* 2010;86(Suppl 1):13–5.
- Backhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* 2015;17(5):690–703.
- Salminen S, Gibson GR, McCartney AL, Isolauri E. Influence of mode of delivery on gut microbiota composition in seven year old children. *Gut* 2004;53(9):1388–9.
- Yatsunenkeno T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. *Nature* 2012;486(7402):222–7.
- O'Sullivan A, Farver M, Smilowitz JT. The influence of early infant-feeding practices on the intestinal microbiome and body composition in infants. *Nutr Metab Insights* 2015;8(Suppl 1):1–9.
- Stewart CJ, Ajami NJ, O'Brien JL, Hutchinson DS, Smith DP, Wong MC, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 2018;562(7728):583–8.
- Subramanian S, Huq S, Yatsunenkeno T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* 2014;510(7505):417–21.
- Hill CJ, Brown JR, Lynch DB, Jeffery IB, Ryan CA, Ross RP, et al. Effect of room temperature transport vials on DNA quality and phylogenetic composition of faecal microbiota of elderly adults and infants. *Microbiome* 2016;4(1):19.
- Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* 2016;8(343):343–82.
- Davis MY, Zhang H, Brannan LE, Carman RJ, Boone JH. Rapid change of fecal microbiome and disappearance of *Clostridium difficile* in a colonized infant after transition from breast milk to cow milk. *Microbiome* 2016;4(1):53.
- Davis JC, Lewis ZT, Krishnan S, Bernstein RM, Moore SE, Prentice AM, et al. Growth and morbidity of gambiai infants are influenced by maternal milk oligosaccharides and infant gut microbiota. *Sci Rep* 2017;7:40466.
- Hill CJ, Lynch DB, Murphy K, Ulaszewska M, Jeffery IB, O'Shea CA, et al. Evolution of gut microbiota composition from birth to 24 weeks in the INFANTMET Cohort. *Microbiome* 2017;5(1):4.
- Berni Canani R, De Filippis F, Nocerino R, Laiola M, Paparo L, Calignano A, et al. Specific signatures of the gut microbiota and increased levels of butyrate in children treated with fermented cow's milk containing heat-killed *Lactobacillus paracasei* CBA L74. *Appl Environ Microbiol* 2017;83(19).
- Nogacka A, Salazar N, Suarez M, Milani C, Arboleya S, Solis G, et al. Impact of intrapartum antimicrobial prophylaxis upon the intestinal microbiota and the prevalence of antibiotic resistance genes in vaginally delivered full-term neonates. *Microbiome* 2017;5(1):93.
- Bazanella M, Maier TV, Clavel T, Lagkouvardos I, Lucio M, Maldonado-Gomez MX, et al. Randomized controlled trial on the impact of early-life intervention with bifidobacteria on the healthy infant fecal microbiota and metabolome. *Am J Clin Nutr* 2017;106(5):1274–86.
- Stearns JC, Simioni J, Gunn E, McDonald H, Holloway AC, Thabane L, et al. Intrapartum antibiotics for GBS prophylaxis alter colonization patterns in the early infant gut microbiome of low risk infants. *Sci Rep* 2017;7(1):16527.
- Parker EPK, Praharaj I, Zekavati A, Lazarus RP, Giri S, Operario DJ, et al. Influence of the intestinal microbiota on the immunogenicity of oral rotavirus vaccine given to infants in south India. *Vaccine* 2018;36(2):264–72.
- McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, et al. American gut: an open platform for citizen science microbiome research. *mSystems* 2018;3(3).
- de Muinck EJ, Trosvik P. Individuality and convergence of the infant gut microbiota during the first year of life. *Nat Commun* 2018;9(1):2233.
- Berni Canani R, De Filippis F, Nocerino R, Paparo L, Di Scala C, Cosenza L, et al. Gut microbiota composition and butyrate production in children affected by non-IgE-mediated cow's milk allergy. *Sci Rep* 2018;8(1):12500.
- Mortensen MS, Hebbelstrup Jensen B, Williams J, Breen AD, O'Brien Andersen L, Roser D, et al. Stability and resilience of the intestinal microbiota in children in daycare - a 12 month cohort study. *BMC Microbiol* 2018;18(1):223.
- Lozupone CA, Stombaugh J, Gonzalez A, Ackermann G, Wendel D, Vazquez-Baeza Y, et al. Meta-analyses of studies of the human microbiota. *Genome Res* 2013;23(10):1704–14.
- Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun* 2017;8(1):1784.
- Bisanz JE, Upadhyay V, Turnbaugh JA, Ly K, Turnbaugh PJ. Meta-analysis reveals reproducible gut microbiome alterations in response to a high-fat diet. *Cell Host Microbe* 2019;26(2):265–272 e264.
- Drewes JL, White JR, Dejea CM, Fathi P, Iyadorai T, Vadivelu J, et al. Erratum: author correction: high-resolution bacterial 16S rRNA gene profile meta-analysis and biofilm status reveal common colorectal cancer consortia. *NPJ Biofilms Microbiomes* 2019;5:2.
- Pammi M, Cope J, Tarr PI, Warner BB, Morrow AL, Mai V, et al. Intestinal dysbiosis in preterm infants preceding necrotizing enterocolitis: a systematic review and meta-analysis. *Microbiome* 2017;5(1):31.
- Bhute S, Pande P, Shetty SA, Shelar R, Mane S, Kumbhare SV, et al. Molecular characterization and meta-analysis of gut microbial communities illustrate enrichment of prevotella and megasphaera in indian subjects. *Front Microbiol* 2016;7:660.
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 2013;41(1):e1.
- Lefever S, Pattyn F, Hellemans J, Vandesompele J. Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. *Clin Chem* 2013;59(10):1470–80.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7(5):335–6.
- Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 2018;6(1):90.
- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;13(7):581–3.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO: The SILVA ribosomal RNA gene database project: improved data

- processing and web-based tools. *Nucleic Acids Res* 2013; 41(Database issue): D590–596.
- [39] Greathouse KL, White JR, Padgett RN, Perrotta BG, Jenkins GD, Chia N, et al. Gut microbiome meta-analysis reveals dysbiosis is independent of body mass index in predicting risk of obesity-associated CRC. *BMJ Open Gastroenterol* 2019;6(1):e000247.
- [40] Derrien M, Alvarez AS, de Vos WM. The Gut Microbiota in the First Decade of Life. *Trends Microbiol* 2019;27(12):997–1010.
- [41] Milani C, Hevia A, Foroni E, Duranti S, Turroni F, Lugli GA, et al. Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS ONE* 2013;8(7):e68739.
- [42] Mancabelli L, Milani C, Lugli GA, Fontana F, Turroni F, van Sinderen D, et al. The impact of primer design on amplicon-based metagenomic profiling accuracy: detailed insights into bifidobacterial community structure. *Microorganisms* 2020;8(1).
- [43] Milani C, Alessandri G, Mangifesta M, Mancabelli L, Lugli GA, Fontana F, et al. Untangling species-level composition of complex bacterial communities through a novel metagenomic approach. *mSystems* 2020;5(4).
- [44] Alessandri G, Milani C, Mancabelli L, Mangifesta M, Lugli GA, Viappiani A, et al. The impact of human-facilitated selection on the gut microbiota of domesticated mammals. *FEMS Microbiol Ecol* 2019;95(9).
- [45] Fallani M, Amarri S, Uusijarvi A, Adam R, Khanna S, Aguilera M, et al. Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. *Microbiology* 2011;157(Pt 5):1385–92.
- [46] Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, et al. Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* 2011;108(Suppl 1):4578–85.
- [47] Mancabelli L, Milani C, Lugli GA, Turroni F, Ferrario C, van Sinderen D, et al. Meta-analysis of the human gut microbiome from urbanized and pre-agricultural populations. *Environ Microbiol* 2017;19(4):1379–90.
- [48] Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. *Nature* 2011;473(7346):174–80.
- [49] Salonen A, Salojarvi J, Lahti L, de Vos WM. The adult intestinal core microbiota is determined by analysis depth and health status. *Clin Microbiol Infect* 2012;18(Suppl 4):16–20.
- [50] Alessandri G, Milani C, Mancabelli L, Mangifesta M, Lugli GA, Viappiani A, et al. Metagenomic dissection of the canine gut microbiota: insights into taxonomic, metabolic and nutritional features. *Environ Microbiol* 2019;21(4):1331–43.
- [51] Alessandri G, Milani C, Mancabelli L, Longhi G, Anzalone R, Lugli GA, et al. Deciphering the bifidobacterial populations within the canine and feline gut microbiota. *Appl Environ Microbiol* 2020;86(7).
- [52] Shin H, Pei Z, Martinez 2nd KA, Rivera-Vinas JI, Mendez K, Cavallin H, et al. The first microbial environment of infants born by C-section: the operating room microbes. *Microbiome* 2015;3:59.
- [53] Thavagnanam S, Fleming J, Bromley A, Shields MD, Cardwell CR. A meta-analysis of the association between Caesarean section and childhood asthma. *Clin Exp Allergy* 2008;38(4):629–33.
- [54] Huh SY, Rifas-Shiman SL, Zera CA, Edwards JW, Oken E, Weiss ST, et al. Delivery by caesarean section and risk of obesity in preschool age children: a prospective cohort study. *Arch Dis Child* 2012;97(7):610–6.
- [55] Lozupone CA, Stombaugh JI, Gordon JL, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. *Nature* 2012;489(7415):220–30.
- [56] Lugli GA, Duranti S, Milani C, Mancabelli L, Turroni F, Alessandri G, et al. Investigating bifidobacteria and human milk oligosaccharide composition of lactating mothers. *FEMS Microbiol Ecol* 2020.
- [57] D'Auria E, Bergamini M, Staiano A, Banderali G, Pendezza E, Penagini F, et al. Italian Society of P: Baby-led weaning: what a systematic review of the literature adds on. *Ital J Pediatr* 2018;44(1):49.
- [58] World Health Organization U: *Global strategy for infant and young child feeding*. 2003.
- [59] Tanaka M, Sanefuji M, Morokuma S, Yoden M, Momoda R, Sonomoto K, et al. The association between gut microbiota development and maturation of intestinal bile acid metabolism in the first 3 y of healthy Japanese infants. *Gut Microbes* 2020;11(2):205–16.
- [60] Le Doare K, Holder B, Bassett A, Pannaraj PS. Mother's milk: a purposeful contribution to the development of the infant microbiota and immunity. *Front Immunol* 2018;9:361.
- [61] Zhong H, Penders J, Shi Z, Ren H, Cai K, Fang C, et al. Impact of early events and lifestyle on the gut microbiota and metabolic phenotypes in young school-age children. *Microbiome* 2019;7(1):2.
- [62] Grier A, McDavid A, Wang B, Qiu X, Java J, Bandyopadhyay S, et al. Neonatal gut and respiratory microbiota: coordinated development through time and space. *Microbiome* 2018;6(1):193.
- [63] Dogra S, Sakwinska O, Soh SE, Ngom-Bru C, Bruck WM, Berger B, et al. Dynamics of infant gut microbiota are influenced by delivery mode and gestational duration and are associated with subsequent adiposity. *mBio* 2015;6(1).
- [64] Stearns JC, Zulyniak MA, de Souza RJ, Campbell NC, Fontes M, Shaikh M, et al. Ethnic and diet-related differences in the healthy infant microbiome. *Genome Med* 2017;9(1):32.
- [65] Brooks AW, Priya S, Blekhan R, Bordenstein SR. Gut microbiota diversity across ethnicities in the United States. *PLoS Biol* 2018;16(12):e2006842.
- [66] Sordillo JE, Zhou Y, McGeachie MJ, Ziniti J, Lange N, Laranjo N, et al. Factors influencing the infant gut microbiome at age 3–6 months: Findings from the ethnically diverse Vitamin D Antenatal Asthma Reduction Trial (VDAART). *J Allergy Clin Immunol* 2017;139(2):482–491 e414.