Establishment of gut microbiota in infants from a large Norwegian cohort

Etablering av tarmflora i spedbarn fra en stor norsk kohort

Philosophiae Doctor (PhD) Thesis

Ekaterina Avershina

Department of Chemistry, Biotechnology and Food Science
Norwegian University of Life Sciences

Ås 2015

Thesis nr.: 2015:76

ISBN-nr.: 978-82-575-1313-9

ISSN-nr.: 1894-6402

Table of Contents

Acknowledgements

The present work was performed at the Department of Chemistry, Biotechnology and Food Sciences, Norwegian University of Life Sciences with financial support from the university. The PACT study was funded by the Norwegian University of Science and Technology and the Norwegian Department of Health and Social affairs. The IMPACT study was funded by GlaxoSmithKline AS, Norway. The ProPACT study was funded by the Norwegian Research Council, Nidarosfondet and Tine BA, including the study milk.

I am deeply thankful to my main supervisor, Prof. Knut Rudi, for all his support, enthusiasm and dedication to the work, as well as for that he is always open for a discussion. I would also like to thank my co-supervisors, prof. Thore Egeland and Dr. Ola Storrø for their valuable advice and help throughout the project. Further I am grateful to Dr. Torbjørn Øien and Prof. Roar Johnsen for letting me use the PACT dataset in my work. I am also grateful to Dr. Anne McCartney from the University of Reading and her PhD student at that time, Beer Thanikan, for hosting me in their lab and giving a hands-on experience with bifidobacteria isolation and culturing. Dr. Monika Sekelja, thank you for sharing your experience in data analysis and I really hope for further collaborations with you.

I am very thankful to all my colleagues and friends here at IKBM for everything that made my work days so much easier and nicer. Anu, Jane, Sabina, Davide, thanks for all the discussions and laughs we shared both at and off work. Raul, thanks for all the music that was cheering me up. Come on! ☺ Paulo, thanks for letting my mind off thesis writing and for diluting my days with brainstorming. Camilla, Melissa, Ida, I learned a lot from you while working together.

Above all, I would like to send my deep thanks to my friends and family. Katya and Tanya, you are always there when I need you and I really appreciate that. Håvard, Tanya, Sergei, Erik, Monika and my Swedish family, thanks for being in my life. Mom, dad, Nadya and Vika, I wouldn't reach what I reached without your guidance and belief in me. Alex, thank you for your unconditional understanding and for taking care of our family. Dima, thank you simply for being my lovely son, for your priceless smile and for bringing joy in our family's life. I love you all very much!

☺ Katya

Summary

Human gut microbiota, our so-called forgotten organ, is crucial for proper functioning of our body and for health maintenance. Although the topic of the gut microbiota establishment is very important, scientists still have limited knowledge on the subject. The reason for this knowledge gap partly stems from the lack of large longitudinal studies. The aim of this thesis was therefore to assess the development of gut microbial community in infants using information from a large cohort. We collectively analyzed 16S rRNA gene amplicons from stool samples of about 350 mother and child pairs followed from pregnancy up to two years of age using various sequencing techniques.

The number of operational taxonomic units (OTUs) which correlated in detection between mothers and children, significantly decreased after three months of age despite the overall increase in the mother-child shared microbiota. Moreover, there was low evidence for direct at-birth transmission of bacteria from the mother to child. Also during the entire period of study, stool samples from children shared as many OTUs with their related mothers as with unrelated mothers and other age-matching children. The distribution of OTUs, as well as microbial diversity estimates, suggested that the development of bacterial community was directed towards neutrality, though dominating bacteria were higher represented than in an entirely neutral assembly.

There were pronounced differences in microbial assemblies between samples from as early as three and ten days of age, indicating rather hyper establishment of gut community at very early days of life. Re-structuring of the gut microbiota in the child population to an adult-like state occurred between one and two years of age, though it was still different from the adult population. The majority of the top prevalent OTUs, though, reached the same prevalence in a population of two-year-olds as in the adult population. Moreover, we propose *Bifidobacterium* OTU as a potential driver of this re-structuring since its prevalence differentially correlated to early and late-colonizing Clostridia OTUs.

Analysis of bifidobacterial community revealed five species (*B. longum*; *B. adolescentis*; *B. bifidum*; *B. breve* and *B. dentium*) that were characteristic for the study cohort with clear age-related correlation patterns between them. Negative correlations between *B. longum* and *B.*

adolescentis were typical for stool samples from two-year-olds and adults, whereas negative correlations between *B. longum* and *B. breve* were common for stool samples from newborns and four-month-olds. Interestingly, both infant and adult correlation profiles were detected at one year of age, also suggesting this age to be a transitional stage in the community re-structuring.

Taken together our results suggest a model for the establishment of gut microbiota in which bacteria are acquired from a common pool of the adult gut associated bacteria in a relatively random manner. In the model, however, the acquisition is controlled by a keystone bacteria.

Sammendrag

Menneskets tarm flora, såkalt 'det glemte organet vårt', er avgjørende for en korrekt funksjon av vår kropp og vedlikehold av vår helse. Selv om temaet om etablering av tarmen bakterieflora er svært viktig, har forskere fortsatt en svært begrenset kunnskap om emnet. Grunnen til den begrensingen stammer delvis fra mangel på store studiekohorter over tid. Målet med denne oppgaven var derfor å vurdere utviklingen av tarmflora hos spedbarn ved bruk av en stor kohort. Vi kollektivt analyserte den mikrobielle sammensetning i avføringsprøver fra ca. 350 mor og barn par som var fulgt fra svangerskapet opp til to år.

Til tross for den generelle økningen i bakterieflora delt mellom mor og barn, så sank antall OTUer, som kollererte med deteksjon hos mor og barn, betraktelig etter tre måneders alder. I tillegg var det lite bevis for direkte ved-fødsel overføring av bakterier fra mor til barn. Dessuten, under hele studietiden, delte avføringsprøver fra barn like mange OTUer med tilhørende mødre som med urelaterte mødre og andre barn med lik alder. Fordeling av OTUer og mikrobiell diversitet estimater foreslo at utviklingen av bakteriefloraen gikk i retning nøytralitet, selv om dominerende bakterier var høyere representert enn ved en helt nøytral sammenstilling.

Det var markante forskjeller i mikrobiell sammensetning allerede mellom prøver fra tre og ti dager, noe som indikerer en aktiv etablering av tarmflora i de første dagene av livet. Omstrukturering av tarmbakteriene fra en barne-populasjon til en voksen-lignende tilstand skjedde når barnet var mellom ett og to år, men bakteriafloraen var allikevel ulik den til en voksen-populasjon. De fleste av de utbredte bakteriene ved denne alderen nådde samme utbredelse som i mødre. Videre foreslår vi at *Bifidobacterium* OTU er en potensiell pådriver for denne re-strukturering siden dens prevalens korrelerte forskjellig med tidlig- og sen-koloniserende Clostridia OTUer.

Analyse av bifidobacteria avdekket fem arter (*B. longum, B. adolescentis, B. bifidum, B. breve* og *B. dentium*) som var karakteristiske for studiekohorten. Det fantes også et klart aldersrelatert korrelasjonsmønster mellom dem. Negative korrelasjoner mellom *B. longum* og *B. adolescentis* var typiske for avføringsprøver fra to-åringer og voksne, mens negative korrelasjoner mellom *B. longum* og *B. breve* var vanlig for avføringsprøver fra nyfødte og fire måneder gamle barn. Både

spedbarn- og voksne korrelasjonsprofiler ble oppdaget ved ett års alder, noe som indikerer at denne alderen er viktig i utviklingen av tarmfloraens struktur.

Våre resultater tyder på en modell for etablering av tarmfloraen hvor bakterier er ervervet fra en felles sammenblanding av bakterier, assosiert med en voksen tarmflora, på en forholdsvis tilfeldig måte. I modellen vår, derimot, styres denne ervervelsen av nøkkelart bakterier.

List of papers

PAPER 1

Avershina E., Storrø O., Øien T., Johnsen R., Pope P. & Rudi K. (2014) Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS microbiology ecology* **87**: 280-290. DOI: 10.1111/1574-6941.12223

PAPER 2

Avershina E., Storrø O., Øien T., Johnsen R., Wilson R., Egeland T. & Rudi K. (2013) Bifidobacterial succession and correlation networks in a large unselected cohort of mothers and their children. *Applied and environmental microbiology* **79**: 497-507. DOI: 10.1128/AEM.02359-12

PAPER 3

Avershina E., Lundgård K., Sekelja M., Dotterud C., Storrø O., Øien T., Johnsen R. & Rudi K. (2015) Transition from infant- to adult-like gut microbiota. (Submitted manuscript)

List of publications not included in the thesis

1. Avershina, E. & Rudi, K. (2015) Dominant short repeated sequences in bacterial genomes. *Genomics* **105**: 175-181. DOI: 10.1016/j.ygeno.2014.12.009
2. Avershina, E. & Rudi, K. (2015) Confusion about the species richness of human gut microbiota. *Benef Microbes*: 1-4. DOI: 10.3920/BM2015.0007
3. Schanche, M., Avershina, E., Dotterud, C., Oien, T., Storro, O., Johnsen, R. & Rudi, K. (2015) High-Resolution Analyses of Overlap in the Microbiota Between Mothers and Their Children. *Curr Microbiol* **71**(2): 283-290 DOI: 10.1007/s00284-015-0843-5
4. Dotterud, C.K., Avershina, E., Sekelja, M., Simpson, M.R., Rudi, K., Storro, O. et al. (2015) Does Maternal Perinatal Probiotic Supplementation Alter the Intestinal Microbiota of Mother and Child? A Randomized Controlled Trial. *J Pediatr Gastroenterol Nutr*. DOI: 10.1097/MPG.0000000000000781
5. Rodriguez, J.M., Murphy, K., Stanton, C., Ross, R.P., Kober, O.I., Juge, N., Avershina E. et al. (2015) The composition of the gut microbiota throughout life, with an emphasis on early life. *Microb Ecol Health Dis* **26**: 26050. DOI: 10.3402/mehd.v26.26050
6. Ravi, A., Avershina, E., Ludvigsen, J., L'Abee-Lund, T.M. & Rudi, K. (2014) Integrons in the intestinal microbiota as reservoirs for transmission of antibiotic resistance genes. *Pathogens* **3**: 238-248. DOI: 10.3390/pathogens3020238
7. Naseribafrouei, A., Hestad, K., Avershina, E., Sekelja, M., Linlokken, A., Wilson, R. & Rudi, K. (2014) Correlation between the human fecal microbiota and depression. *Neurogastroenterol Motil* **26**: 1155-1162. DOI: 10.1111/nmo.12378
8. Nwosu, F.C., Avershina, E., Wilson, R. & Rudi, K. (2014) Gut Microbiota in HIV Infection: Implication for Disease Progression and Management. *Gastroenterol Res Pract* 2014: 803185. DOI: 10.1155/2014/803185
9. Storro, O., Avershina, E. & Rudi, K. (2013) Diversity of intestinal microbiota in infancy and the risk of allergic disease in childhood. *Curr Opin Allergy Clin Immunol* **13**: 257-262. DOI: 10.1097/ACI.0b013e328360968b
10. Thorkildsen, L.T., Nwosu, F.C., Avershina, E., Ricanek, P., Perminow, G., Brackmann, S. et al. (2013) Dominant fecal microbiota in newly diagnosed untreated inflammatory bowel disease patients. *Gastroenterol Res Pract* 2013: 636785. DOI: 10.1155/2013/636785

11. Avershina, E., Frisli, T. & Rudi, K. (2013) De novo semi-alignment of 16S rRNA gene sequences for deep phylogenetic characterization of next generation sequencing data. *Microbes Environ* **28**: 211-216. DOI: 10.1264/jsme2.ME12157
12. Nwosu, F.C., Thorkildsen, L.T., Avershina, E., Ricanek, P., Perminow, G., Brackmann, S. et al. (2013) Age-dependent fecal bacterial correlation to inflammatory bowel disease for newly diagnosed untreated children. *Gastroenterol Res Pract* 2013: 302398. DOI: 10.1155/2013/302398
13. Avershina, E. & Rudi, K. (2013) Is it who you are or what you do that is important in the human gut? *Benef Microbes* **4**: 219-222. DOI: 10.3920/BM2013.0016

Abbreviations

16S rRNA – 16S ribosomal ribonucleic acid

ClpC – caseinolytic protease C

ITS – Internal transcribed spacer

OTU – Operational taxonomic unit

PACT – Prevention of Allergy among Children of Trondheim

PCR – Polymerase chain reaction

qPCR – quantitative PCR

Introduction

Human gut microbiota

Gut microbiota composition

The gut microbiota is a complex community with about 200 different species co-existing at large densities [1, 2]. Most of the gut microbiota is comprised of bacteria, though eukaryotes, archae and viruses are also present [3, 4]. Disruptions in gut microbial assembly have been associated to a large number of diseases [5-11] and psychiatric disorders [12, 13]. Restoration of the microbiota using probiotics (live bacteria), prebiotics (oligosaccharides that promote growth of certain bacteria), synbiotics (combination of pro- and prebiotics) or by fecal transplants from healthy donors, have in many cases shown efficacy in treatment or symptom reduction of these disorders as reviewed in [14]. Some studies also suggest that the timing of gut colonization may have an impact on individuals' liability towards disease development [15]. Therefore understanding of the gut microbiota establishment is of paramount importance.

In the first days of life, the gut microbiota is dominated by *Escherichia*, *Streptococcus* and *Staphylococcus*. Later on, redox potential declines and anaerobic bacteria like *Bacteroides* and *Bifidobacterium* take over [16, 17]. Bifidobacteria are highly specialized in utilization of human milk oligosaccharides [18, 19]. In breast-fed infants, they may substitute up to 90 % of the gut microbial community [20], with *B. longum* and *B. breve* being among the most commonly isolated species [21, 22]. In formula fed infants, bifidobacteria are reportedly less diverse and abundant [23, 24], though addition of prebiotic to the infant formula may increase bifidobacteria diversity [25]. Some reports, however, suggest no difference in bifidobacterial composition between formula fed and breast milk fed infants [26]. When infant diet switches from breast milk to solid foods, bifidobacteria are depleted and gut microbiota becomes enriched in adult-associated Clostridia [16, 27].

The majority of adult gut bacteria belongs to two phyla of Firmicutes and Bacteroides and the relation between them might be associated to the metabolic capacity of the gut [28, 29]. Other phyla comprising the gut microbiota are Actinobacteria, Proteobacteria, Verrucomicrobia, Fusobacteria, Tenericutes and Cyanobacteria [30, 31]. On a species and strain level, gut microbiota is often compared to a fingerprint, since each individual has a unique gut microbial assembly [32,

33]. Gut microbiota is highly impacted by the host immune system [34], lifestyle factors [33, 35, 36] and age [37, 38]. On a functional level, though, gut microbial assemblages seem to be rather similar among individuals [32, 39].

Establishment of gut microbiota

Sterility of the amnion is nowadays debated in a scientific community since indications for non-pathological *in utero* colonization appeared [40]. Numerous studies reported isolation of gut-associated bacteria from meconium of healthy newborns [41], as well as from umbilical cord [42], placenta [43, 44] and amniotic fluid [45]. In a probiotic administration trial, children of mothers who received *L. rhamnosus* GG during their late pregnancy, were colonized by this species regardless delivery mode [46], possibly indicating prenatal translocation of this species. Despite prenatal exposure to gut-associated bacteria, the infant gut microbiome is largely shaped by mode of birth [47]. At passage through the birth canal, the mother's vaginal lactobacilli colonize newborns and can stay at detectable levels in the infant's intestine up until one month after birth [48]. Several studies also reported possible transmission of Bacteroides at birth, since this group of bacteria is often underrepresented in C-section born infants [26, 49] whose stool samples are enriched in skin-derived microbiota [47].

In addition to breast milk's primary role of being a source of nutrition and immune protection for a baby, it also serves as an inoculum of gut bacteria. More than 200 bacterial species have been collectively isolated from human milk, though each mother has from 2 to 18 cultivable species in her milk [50]. Commonly isolated species belong to *Staphylococcus*, *Streptococcus*, *Lactococcus*, *Lactobacillus*, *Enterococcus* and *Bifidobacterium* [50-52]. While aerobic and facultative anaerobic bacteria are likely of the mother's skin and the baby's mouth origin, anaerobes are probably translocated from the gut via a bacterial enteromammary pathway [53]; - a route that is most likely also involved in pre-natal shaping of fetal gut microbiome [54]. Since this kind of translocation assumes bacteria in a blood stream [55], the selection of bacteria to be translocated must be a strictly controlled process. Only few bacteria have been shown to be potentially vertically transmitted so far. When giving *L. rhamnosus* GG, *L. acidophilus* and *B. animalis* subsp. *lactis* to mothers during pregnancy and breast feeding, only *L. rhamnosus GG* was demonstrated to successfully colonize infant gut [56]. Vertical transmission of certain *Staphylococcus* strains from the mother's to the infant's gut through the breast milk was also demonstrated [51].

The mother's role in shaping the child's gut microbiota after breast milk cessation remains vague. So is the question of how adult-associated bacteria are acquired. Two general hypotheses suggest that they either are transmitted perinatally and stay low in abundance until conditions become right or are acquired from the environment later on. In support of a delayed bacterial acquisition, lack of microbial exposure in early life is a known factor that increases the risk of allergy development [57], with even dishwashing (hand vs machine) or pacifier cleaning (mouth vs tap water) routines being associated to it [58, 59]. Interestingly, it has been also noted that if the child has siblings, his/her gut microbial diversity tends to be higher than in case the child is the only child in house [60]. However, no hard evidence for either of the hypotheses has been reported so far.

Taxonomic characterization of gut microbiota

Deep DNA sequencing of 16S rRNA gene [61] is widely used for microbial composition assessment [62]. Sequences with a given similarity level are binned into so-called Operational Taxonomic Units (OTUs) [63]. Although classification of bacteria into species is rather arbitrary [64], commonly used thresholds are 97 % and 95 % to reflect species and genus delineation respectively [65].Two widely used analytical tools for sequencing data analysis are Quantitative Insights Into Microbial Ecology (QIIME) [66] and Mothur [67]. These tools enable complete sequencing data analysis from raw data filtering to diversity estimation.

There are two widely used approaches for binning sequences: supervised clustering, i.e. comparison of sequences against a reference database, and unsupervised clustering of sequences into OTUs based on their pairwise distances [68]. Several large depositories of 16S rRNA sequences are widely used as reference databases (RDP [69]; SILVA [70]; Greengenes [71]). However, reference-based approaches are not applicable for poorly described communities since the information on novel organisms will be discarded [72]. Unsupervised clustering on the other hand, does not rely on reference databases, and can be used to determine previously undescribed organisms [65]. Algorithms of unsupervised clustering can be largely separated into hierarchical and greedy heuristic clustering, though new alternative statistical approaches are also available [73]. Hierarchical clustering (ESPRIT [74], DOTUR [75]) requires calculation of the distance matrix for an input dataset and is thus potentially computationally demanding [76]. Greedy heuristic clustering algorithms (UCLUST [77], CD-HIT [78]) skip distance matrix computation and process one sequence at a time, either assigning it to an already-existing-cluster or marking it

as a new cluster in case its distance to existing clusters is higher than a threshold [63]. Mothur is based on a hierarchical clustering algorithm [67], whereas QIIME implements 12 various OTU picking techniques that cover all groups of approaches (http://qiime.org/scripts/pick_otus.html).

OTU clustering does not necessarily reflect species delineation, and the same sequence can be classified as different OTUs using various clustering techniques [63]. At the same time, sequences belonging to the same species can be classified as several independent OTUs for both biological and technical reasons [79]. This can largely influence data interpretation and lead to erroneous conclusions. To overcome this problem, new algorithms that implement stringent error filtering appear [80]. In addition, sequencing platforms are in constant development to allow longer reads [81] and new sequencing approaches emerge [2].

Children study cohorts

Generally, microbiome studies balance between two extremes of either including large number of individuals [16] or following few individuals through a long period with multiple sampling points [82]. There are several study cohorts though, for example KOALA from Netherlands [83], SKOT from Denmark [84], CHILD from Canada [85], NoMic [86] and PACT [87] from Norway, that have a luxury of both. The CHILD cohort includes more than 3600 children and the research team has collected meconium (i.e. first stool of the newborn) along with stool samples at three months and one year of age [85]. KOALA enrolled a total of 2500 children followed from birth to adulthood [83]. However, the largest longitudinal gut microbiota study from this cohort reported to date involved 606 infants followed from five weeks to seven months of age with one extra sampling point at 13 weeks [88]. The research group of the SKOT cohort reported gut microbiota composition of 330 children that were followed from nine months up until three years of age [60, 89].

Norway is home for two large prospective children cohort studies: The Norwegian Microflora Study (NoMic) and Prevention of Allergy among Children of Trondheim (PACT). The NoMic is a prospective study that started in 2002 and in total recruited 524 families in a course of three years with the aim of studying establishment of gut flora in infancy and its impact on child health [86]. So far, data on microbial composition of 360 NoMic children followed through the first two years of life have been reported (reviewed in [90]).

The PACT study started in 2000 in Trondheim region and it followed women and their children up until two years of age [87]. The aim of the PACT was to reduce risk of allergy, asthma and eczema development in children by decreasing exposure to tobacco smoke and indoor dampness, and by increasing fish oil intake. Pregnant women were included in the study after they gave a written consent to participate and the only inclusion criteria was their ability to read and understand Norwegian language since they had to fill out questionnaires during pregnancy, six weeks, one and two years after birth [91]. The control group of the study was formed prior to the start of the intervention, with the last participant being included in March 2005. Intervention was initiated in June 2002 and the last questionnaire, two years after birth, was submitted in March 2009. Two studies were formed based on the PACT cohort. IMPACT (Immunology and Microbiology in the PACT) was formed from the control group of the PACT and its aim was to investigate correlations between gut microbiota development and establishment of allergic diseases by the age of two years [92]. ProPACT (Probiotics in the PACT) was formed from the intervention cohort of the PACT and its aim was to assess whether consumption of probiotic supplementation during last four weeks of pregnancy and first three months of breast-feeding would reduce the incidence of allergic diseases in children at two years of age [91].

Aim of the thesis

Timely acquisition of 'healthy' intestinal bacteria seem to prevent number of disorders later in life [93-95]. The important issue is then to understand how the normal gut microbial community develops and what drives its establishment. The overall aim of this thesis was therefore **to address the population ecology of children gut microbiota and its establishment at early stages of life** using stool samples as a proxy for gut microbial assessment.

The work was divided into following tasks:

- Characterize gut microbial composition and assess a pace of shifts in microbial community structure during first years of life (PAPER 1)
- Characterize development of bifidobacteria in children population and correlations within this key bacterial group in infancy (PAPER 2)
- Address bacterial recruitment in the child population and the mother's contribution therein (PAPER 1, PAPER 3)

In order to address these questions, we used two nested cohorts from the PACT study. The IMPACT cohort was used in PAPER 1 and PAPER 2 (86 mother-child pairs included), whereas ProPACT samples were utilized in PAPER 3 (287 mother-child pairs included). For both studies, stool samples were collected from mothers once or twice during pregnancy and four to five times from children within the first two years of their lives. In ProPACT, the mothers also donated their stool samples three months post-partum. At end point, allergic sensitization was assessed by pediatrician based on medical history and clinical examinations [91, 92].

Numerous techniques were implemented during the study. The total gut microbiota community was assessed by analysis of V3-V4 region of 16S rRNA gene (PAPER 1; PAPER 3). Caseinolytic protease C (*clpC*) gene was used as a marker for bifidobacterial composition (PAPER 2). Sequencing analysis included Sanger sequencing (PAPER 1; PAPER 2), pyrosequencing (PAPER 1) and Illumina sequencing (PAPER 3) performed in-house and in the Norwegian Sequencing Centre (Oslo, Norway). Results were further confirmed by multilocus sequencing of cultured isolates (PAPER 2) and by qPCR analysis of 16S rRNA (PAPER 1) and of 16S-23S ITS region (PAPER 2).

Results and discussion

Earlier study have suggested rather varying temporal patterns of microbial succession between children [17]. Due to high inter-individual variability [30], deduction of common information on gut microbial community in a population is very problematic. Therefore in our work we utilized a mathematical approach that extracts highly dominant and at the same time highly prevalent sequencing information from bulk sequencing data and leaves the rest of the information aside [96]. This enabled us to reveal highly structured age-specific gut microbial assemblies on a population level (PAPER 1; PAPER 2) that were not apparent when comparing individuals one to another. We then expanded on acquisition of bacterial taxa in the population using stringent analysis of deep 16S rRNA gene sequencing [80] (PAPER 3).

Interactions within gut microbiota

Residents of human gut largely interact with each other [27, 97]. Identification of different cross-talks between bacteria might lead to better understanding of disease development or resilience towards such development in individuals. In our work, we observed clear age-related correlation networks between *Bifidobacterium* species (PAPER 2). Negative correlations between *B. longum* and *B. adolescentis* were typical for stool samples from one- and two-year-olds and adults ($p < 0.001$), whereas negative correlations between *B. longum* and *B. breve* were common for stool samples from newborns and four-month-olds ($p < 0.001$). The second year of life was a turning period in gut microbiota restructuring from infant-like to adult-like configuration (PAPER 2; PAPER 3). We observed a pronounced shift in microbial ecology at that time - OTUs that prevailed among the child population during the entire first year of life, became rarer detected by two years; whereas OTUs that became most prevalent among two-year-olds, had same rate of detection as in mothers (PAPER 3). Moreover, bifidobacterial community of one-year-old children resembled both infants (*B. breve* vs *B. longum*) and adults (*B. adolescentis* vs *B. longum*) with regards to correlation patterns (PAPER 2), also indicating the ongoing restructuring of the community.

B. adolescentis is often regarded as an adult-associated bacteria [20, 98] which is more suited for utilization of plant-derived oligosaccharides [99]. *B. breve* in its turn is commonly isolated from infants [20, 25] and is well equipped for both human milk and plant oligosaccharides degradation [100]. At four months of age, when *B. breve* peaked in abundance, nearly half of children received

both starch-containing food and breast milk (PAPER 2). The switch from *B. breve* to *B. adolescentis* in the adult profile then probably might be explained by cessation of breast feeding rather than by introduction of solid foods into the diet, concordant with other report on driving force of gut community structure [27]. *B. longum* group showed significant correlation patterns both within infant and adult profiles, suggesting its central role in structuring of gut community. In our cohort, *B. longum* group comprised two subspecies of *B. longum* subsp. *longum* and *B. longum* subsp. *infantis*, which excluded each other. *B. longum* subsp. *infantis* was largely detected in stool samples of 4-months-olds, whereas *B. longum* subsp. *longum* was characteristic of all other sampled time points (PAPER 2). Previous genomic analyses of these two species revealed that *B. longum* subsp. *infantis* was more equipped to human milk utilization, whereas *B. longum* subsp. *longum* harbored plant-derived oligosaccharides breakdown machinery [18, 101, 102], which may explain the observed co-exclusion.

The time from one to two years of age was characterized by very active recruitment of Clostridia and Bacteroides. Strikingly, we observed negative correlation between the detection of late-appearing Clostridia and of *Bifidobacterium* OTU ($p < 0.05$) which coincided with positive correlation between detection of this OTU and of early-appearing Clostridia ($p < 0.05$) (PAPER 3). This observation indicates possible gate-keeping role of this bifidobacteria OTU in the developing infant gut. Bifidobacteria are largely breast-milk selected [103] and we observed instant drop in the abundance of this OTU as a response to breast feeding cessation, though decrease in its prevalence was delayed (PAPER 3). Concordant with the notion of gate-keeping role of bifidobacteria, negative correlation between *Bifidobacterium* and gut microbiota diversity was previously reported [26]. Recently, the mechanism for host regulation of the gut colonization through the immune mediation of commensal microbiota, has been demonstrated in mice [104].

Population ecology of gut microbiota

We observed low rate of at-birth transmission of bacteria in a population (PAPER 1) with after-birth acquisition of these bacteria being more likely to account for the increase in the mother-child shared microbiota (PAPER 1; PAPER 3). As such, contrary to a previous study [26], we found higher frequency of *E. coli* detection in children of *E.coli*-positive mothers at ten days as compared to three days of age, suggesting delayed colonization by these species (PAPER 1). Although mode of delivery is an important determinant of the initial gut inoculum [27, 105], low rates of vaginal

flora transmission from mother to child were previously demonstrated [48]. In a study from Azad and colleagues, children who were born by caesarean delivery, had lower levels of *Escherichia/Shigella* group as compared to vaginally delivered infants [26]. The authors reported that all women who underwent caesarean delivery, were administered antibiotics. This intake might have contributed to the altered breast milk composition and thus could have hindered transmission of this breast-milk associated bacteria [106] from mother to child at a later stage. Moreover, children of *B. bifidum*-positive mothers have higher chance of being colonized by this species at six months as compared to one month of age [107], which also points towards acquisition of this bacteria through breast milk rather than at delivery.

We found Bacteroides class to be enriched in mother-child shared pool of bacteria during the whole study period (PAPER 3). In line with this observation, *Bacteroides* is known to be associated with vaginal delivery since it depletes in C-section delivered baby regardless antibiotic intake [105]. At the same time, although lactobacilli are one of the main constituents of breast milk microbiome [50] and nearly all children of the cohort were exclusively breastfed for six weeks after birth, we observed drop in lactobacilli abundance from three to ten days after birth. Since lactobacilli are characteristic of vaginal swabs from pregnant women [108, 109], we believe that the observed reduction might evidence for the acquisition of this bacteria at birth. Therefore we believe that only certain bacteria are 'allowed' to colonize the newborn at a passage through the mother's birth canal and that other factors such as diet and the newborn's environment, take leading role in priming of the infant gut.

Concordant with other reports [27, 105], stool samples from newborns were characterized by low alpha- and high beta-diversity, converging towards more similar communities with higher species richness with age (PAPER 1; PAPER 2). Moreover, in line with previous observations [16, 110], stool samples from children tended to share as many OTUs with unrelated mothers as with their related mothers and other age-matching children (PAPER 1; PAPER 3). From an ecological point of view, one may argue that direct vertical transmission from mother to child would lead to highly dominated communities, which in turn are more vulnerable to perturbations and thus less temporally stable [111]. On the other hand a near-to-neutral community assembly with redundant bacteria occupying the same niches implicates a higher chance of colonization by 'outside' species, thus ensuring higher diversity and community resilience [112, 113]. Interestingly, we detected

lower deviation from neutrality in the group of mothers and their one- and two-year-old children than in newborns and three-month-olds ($p < 0.05$) (PAPER 1; PAPER 3). At the same time, the number of OTUs which were more likely to be detected in children given they were detected in their mothers, gradually decreased despite the overall increase in the mother-child shared fraction of OTUs. Early appearance of OTUs in the child population, as well as their prevalence therein, also positively correlated to OTUs' prevalence in the mother population (PAPER 3). This indicates that OTUs that were widely distributed in the adult population were recruited into the child population first-line.

During the first year of life, infants' stool samples were largely dominated by *Bifidobacterium* and other Actinobacteria (PAPER 1; PAPER 3). Within bifidobacteria, five species (*B. longum*; *B. adolescentis*; *B. bifidum*; *B. breve* and *B. dentium*) were found most prevalent and most abundant in the population across the study period (PAPER 2), corroborating other reports [20, 25]. The second year of life was marked by dynamic OTUs recruitment (PAPER 3) while abundance profiles of dominant gut residents seemed to be stabilized (PAPER 1; PAPER 2). At that period, Clostridia and Bacteroidia were largely recruited in infant gut, though many of the adult-associated OTUs within these classes were still not detected in the child population at the end point of the study (PAPER 1; PAPER 3). Representatives of both classes are commonly isolated from soil [114, 115]. Therefore they can potentially be acquired while children learn about the world around them by actively interacting with it.

It remains unclear how long time is needed for shaping of an adult bacterial profile. In our cohort we observed that at two years of age, though children gut microbiota resembled that of adults, there were still pronounced differences between the two with regards to abundance and richness profiles (PAPER 1; PAPER 3). A study by Yatsunenko et al., which included individuals aged 0 to 80 years, suggested that shaping of gut microbiota towards an adult state takes three years [16], whereas others claim this shaping to continue all way through childhood to adolescence [116].

A study by Koren and colleagues suggested pronounced loss of gut microbial richness during the course of pregnancy coincinding with enrichment of Proteobacteria and Actinobacteria [117]. Although we did not detect such drastic changes in microbial diversity and richness during pregnancy (PAPER 1; PAPER 3), we observed slight increase of Bacilli in last trimester of pregnancy as compared to three months postpartum samples (PAPER 3). A recent study of 40

women sampled weekly during pregnancy also suggested lack of gut microbiota remodeling at that period [118].

Model of gut microbiota establishment

Recently a model of non-random early life gut microbiota establishment was proposed [119]. In this model, the colonization process is separated into subsequent stages of dominance by Bacilli (premature gut only), Proteobacteria, Actinobacteria and Firmicutes/Bacteroides. The occurrence and duration of each stage is largely defined by maturation of the gut, delivery mode and feeding style. We expand on the model and believe that within each stage, colonization of the gut occurs from the common pool of bacteria capable of thriving in the gut (Figure 1).

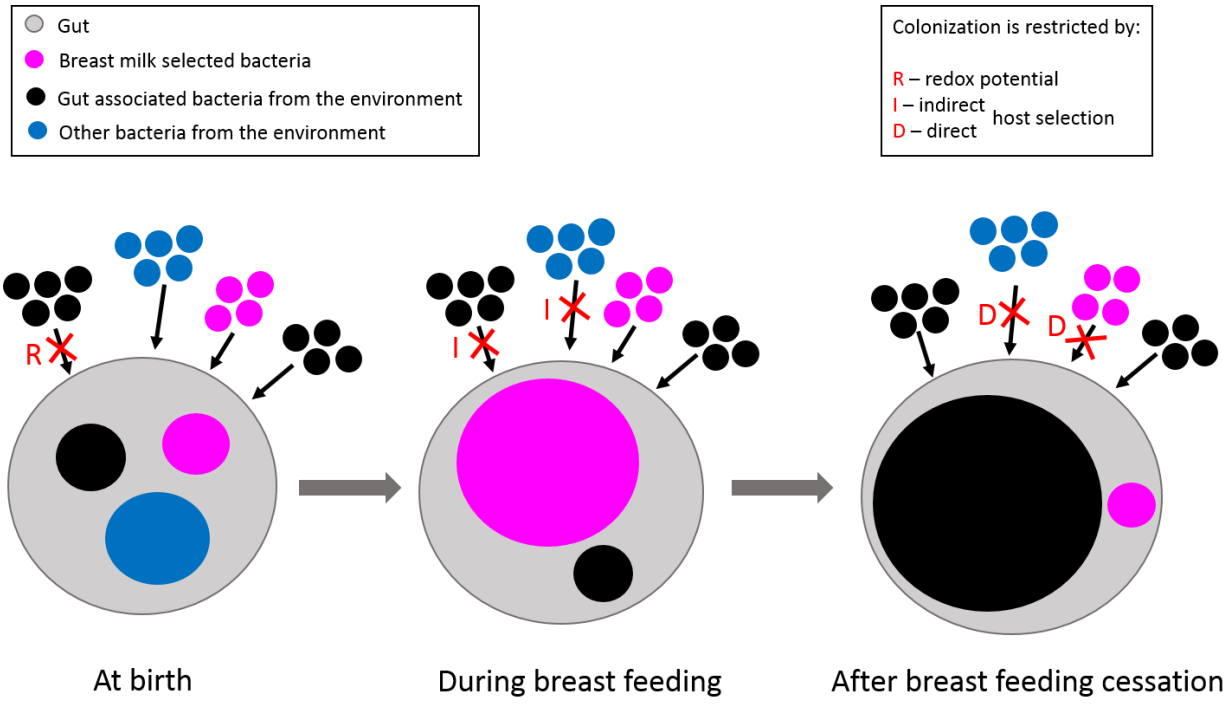


Figure 1. Proposed model of gut colonization. During birth and right after birth the newborn's gut is bombarded by bacteria from the mother's birth canal, skin, mouth, hospital, breast milk, etc. At first, all bacteria capable of thriving in high redox potential, can colonize the gut. With establishment of breast feeding, breast milk promoted bacteria take over and grant access to selected bacteria while keeping others away. When breast feeding stops, the 'keepers' drop in counts and all bacteria that are capable of survival and proliferation in the gut are allowed to colonize it.

We propose that initially many surrounding bacteria that can survive in a positive redox potential environment [120], colonize the gut based on the 'first-come, first-served' principle. Later on, as breast-feeding becomes established and the redox potential declines, colonization switches to a bifidobacteria-controlled process where only those with 'valid ID' can pass guard control and

enter. When breast milk selection stops and bifidobacteria drop, all healthy-gut-associated bacteria are allowed to colonize the gut thus ensuring high diversity and stability of the community towards various perturbations. We propose that gut bacteria are recruited in a relatively random manner with more prevalent bacteria in a population having higher probability of being recruited earlier. It is well known that household members, as well as individuals living in remote tribes and villages, have more similar microbiota to each other than to other people [33, 36, 110, 121]. At the same time, some bacterial groups are shared among inhabitants of separate continents [122]. Hypothetically human-associated gut microbiota can be then represented as a combination of common pools such that people who live in close proximity share more bacteria than those living further apart. Our model corresponds to a combination of two ecological scenarios of community structuring – random sampling and dispersal limitation [123]. However more evidence is needed to draw conclusion on which geographical scale common pools are formed at.

Conclusion

In our study cohort, development of gut microbial community in infancy was directed towards a highly diverse community across the population where more redundant bacteria had higher chance of being acquired faster. Gut colonization by adult-associated Clostridia was at least partially controlled by a potentially breast milk regulated bifidobacteria OTU. We also detected correlation networks between the three most abundant and most prevalent bifidobacteria species in the population. *B. breve* was characteristic of the infant correlation profile, whereas *B. adolescentis* – of the adult profile. *B. longum* group, mostly consisting of *B. longum* subsp. *longum*, played a central role in these correlations throughout the entire study period.

Future prospective

This study was based on a descriptive analysis of taxonomic information. Our results offer interesting insights into re-structuring from infant-like to adult-like microbiota profiles. In the future, it will be important to perform shotgun metagenomic analysis of samples within the cohort in order to shed light on mechanisms of this re-structuring.

Genome sequencing of bifidobacteria isolates from the cohort may yield clues on their correlation networks. It would be also beneficial to identify a gate-keeping bifidobacteria OTU, and attempt its isolation and further characterization by genome sequencing. Next, it would be interesting to co-culture the isolate with various clostridia strains (potential early and late colonizers) in order to divide them into those promoted and inhibited by bifidobacteria. After the separation, comparison of genomes from these two pools of clostridia may give indications to what contributes to the aforementioned separation.

Clostridia and Bacteroidia are actively recruited from one year of age onwards. It is therefore also very interesting to investigate deeper what are the sources of these bacteria. Since toddlers are very curious and taste nearly everything they touch thus potentially contributing to gut colonization, it would be exciting to perform a study with simultaneous sampling of one- and two-year-olds, people who they have frequent contact with (parents, kindergarten teachers) and potential inoculum sources including environmental samples (sand/grass), food products and water.

Although gut bacteria are numerically main residents of the gut, its other dwellers are also very likely to have a beneficial impact on our health and gut diversity [124]. It would therefore be of utmost interest to analyze the eukaryome and virome of these samples in a search for potential correlations between all intestinal ecosystems.

References

1. Qin, J., et al., *A metagenome-wide association study of gut microbiota in type 2 diabetes.* Nature, 2012. **490**(7418): p. 55-60.
2. Faith, J.J., et al., *The long-term stability of the human gut microbiota.* Science, 2013. **341**(6141): p. 1237439.
3. Virgin, H.W. and J.A. Todd, *Metagenomics and personalized medicine.* Cell, 2011. **147**(1): p. 44-56.
4. Breitbart, M., et al., *Viral diversity and dynamics in an infant gut.* Res Microbiol, 2008. **159**(5): p. 367-73.
5. Scaldaferri, F., et al., *Bacterial flora as a cause or treatment of chronic diarrhea.* Gastroenterol Clin North Am, 2012. **41**(3): p. 581-602.
6. Casen, C., et al., *Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD.* Aliment Pharmacol Ther, 2015.
7. Penders, J., et al., *Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study.* Gut, 2007. **56**(5): p. 661-7.
8. Sjogren, Y.M., et al., *Altered early infant gut microbiota in children developing allergy up to 5 years of age.* Clin Exp Allergy, 2009. **39**(4): p. 518-26.
9. Vujkovic-Cvijin, I., et al., *Dysbiosis of the gut microbiota is associated with HIV disease progression and tryptophan catabolism.* Sci Transl Med, 2013. **5**(193): p. 193ra91.
10. Wang, T., et al., *Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers.* ISME J, 2012. **6**(2): p. 320-9.
11. Dapito, D.H., et al., *Promotion of hepatocellular carcinoma by the intestinal microbiota and TLR4.* Cancer Cell, 2012. **21**(4): p. 504-16.
12. Foster, J.A. and K.A. McVey Neufeld, *Gut-brain axis: how the microbiome influences anxiety and depression.* Trends Neurosci, 2013. **36**(5): p. 305-12.
13. Mulle, J.G., W.G. Sharp, and J.F. Cubells, *The gut microbiome: a new frontier in autism research.* Curr Psychiatry Rep, 2013. **15**(2): p. 337.
14. Ho, J.T., G.C. Chan, and J.C. Li, *Systemic effects of gut microbiota and its relationship with disease and modulation.* BMC Immunol, 2015. **16**: p. 21.
15. Cong, X., et al., *Early Life Experience and Gut Microbiome: The Brain-Gut-Microbiota Signaling System.* Adv Neonatal Care, 2015.
16. Yatsunenko, T., et al., *Human gut microbiome viewed across age and geography.* Nature, 2012. **486**(7402): p. 222-7.
17. Palmer, C., et al., *Development of the human infant intestinal microbiota.* PLoS Biol, 2007. **5**(7): p. e177.
18. Sela, D.A., et al., *The genome sequence of Bifidobacterium longum subsp. infantis reveals adaptations for milk utilization within the infant microbiome.* Proc Natl Acad Sci U S A, 2008. **105**(48): p. 18964-9.
19. Kitaoka, M., *Bifidobacterial enzymes involved in the metabolism of human milk oligosaccharides.* Adv Nutr, 2012. **3**(3): p. 422S-9S.
20. Turroni, F., et al., *Diversity of bifidobacteria within the infant gut microbiota.* PLoS One, 2012. **7**(5): p. e36957.
21. Mikami, K., et al., *Influence of maternal bifidobacteria on the establishment of bifidobacteria colonizing the gut in infants.* Pediatr Res, 2009. **65**(6): p. 669-74.

22. Turroni, F., et al., *Microbiomic analysis of the bifidobacterial population in the human distal gut.* ISME J, 2009. **3**(6): p. 745-51.
23. Roger, L.C., et al., *Examination of faecal Bifidobacterium populations in breast- and formula-fed infants during the first 18 months of life.* Microbiology, 2010. **156**(Pt 11): p. 3329-41.
24. Roger, L.C. and A.L. McCartney, *Longitudinal investigation of the faecal microbiota of healthy full-term infants using fluorescence in situ hybridization and denaturing gradient gel electrophoresis.* Microbiology, 2010. **156**(Pt 11): p. 3317-28.
25. Barrett, E., et al., *The neonatal gut harbours distinct bifidobacterial strains.* Arch Dis Child Fetal Neonatal Ed, 2015.
26. Azad, M.B., et al., *Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months.* CMAJ, 2013. **185**(5): p. 385-94.
27. Backhed, F., et al., *Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life.* Cell Host Microbe, 2015. **17**(5): p. 690-703.
28. Tilg, H. and A. Kaser, *Gut microbiome, obesity, and metabolic dysfunction.* J Clin Invest, 2011. **121**(6): p. 2126-32.
29. Rosenbaum, M., R. Knight, and R.L. Leibel, *The gut microbiota in human energy homeostasis and obesity.* Trends Endocrinol Metab, 2015.
30. Lozupone, C.A., et al., *Diversity, stability and resilience of the human gut microbiota.* Nature, 2012. **489**(7415): p. 220-30.
31. Di Rienzi, S.C., et al., *The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria.* Elife, 2013. **2**: p. e01102.
32. Turnbaugh, P.J., et al., *A core gut microbiome in obese and lean twins.* Nature, 2009. **457**(7228): p. 480-4.
33. Schnorr, S.L., et al., *Gut microbiome of the Hadza hunter-gatherers.* Nat Commun, 2014. **5**: p. 3654.
34. Hooper, L.V., D.R. Littman, and A.J. Macpherson, *Interactions between the microbiota and the immune system.* Science, 2012. **336**(6086): p. 1268-73.
35. De Filippo, C., et al., *Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa.* Proc Natl Acad Sci U S A, 2010. **107**(33): p. 14691-6.
36. Zhang, J., et al., *Mongolians core gut microbiota and its correlation with seasonal dietary changes.* Sci Rep, 2014. **4**: p. 5001.
37. Biagi, E., et al., *Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians.* PLoS One, 2010. **5**(5): p. e10667.
38. Claesson, M.J., et al., *Composition, variability, and temporal stability of the intestinal microbiota of the elderly.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4586-91.
39. Human Microbiome Project, C., *Structure, function and diversity of the healthy human microbiome.* Nature, 2012. **486**(7402): p. 207-14.
40. Funkhouser, L.J. and S.R. Bordenstein, *Mom knows best: the universality of maternal microbial transmission.* PLoS Biol, 2013. **11**(8): p. e1001631.
41. Jimenez, E., et al., *Is meconium from healthy newborns actually sterile?* Res Microbiol, 2008. **159**(3): p. 187-93.
42. Jimenez, E., et al., *Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section.* Curr Microbiol, 2005. **51**(4): p. 270-4.

43. Aagaard, K., et al., *The placenta harbors a unique microbiome.* Sci Transl Med, 2014. **6**(237): p. 237ra65.
44. Satokari, R., et al., *Bifidobacterium and Lactobacillus DNA in the human placenta.* Lett Appl Microbiol, 2009. **48**(1): p. 8-12.
45. DiGiulio, D.B., *Diversity of microbes in amniotic fluid.* Semin Fetal Neonatal Med, 2012. **17**(1): p. 2-11.
46. Schultz, M., et al., *Administration of oral probiotic bacteria to pregnant women causes temporary infantile colonization.* J Pediatr Gastroenterol Nutr, 2004. **38**(3): p. 293-7.
47. Dominguez-Bello, M.G., 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): p. 11971-5.
48. Matsumiya, Y., et al., *Molecular epidemiological study of vertical transmission of vaginal Lactobacillus species from mothers to newborn infants in Japanese, by arbitrarily primed polymerase chain reaction.* J Infect Chemother, 2002. **8**(1): p. 43-9.
49. Mueller, N.T., et al., *The infant microbiome development: mom matters.* Trends Mol Med, 2015. **21**(2): p. 109-17.
50. Jeurink, P.V., et al., *Human milk: a source of more life than we imagine.* Benef Microbes, 2013. **4**(1): p. 17-30.
51. Martin, V., et al., *Sharing of bacterial strains between breast milk and infant feces.* J Hum Lact, 2012. **28**(1): p. 36-44.
52. Hunt, K.M., et al., *Characterization of the diversity and temporal stability of bacterial communities in human milk.* PLoS One, 2011. **6**(6): p. e21313.
53. Rodriguez, J.M., *The origin of human milk bacteria: is there a bacterial entero-mammary pathway during late pregnancy and lactation?* Adv Nutr, 2014. **5**(6): p. 779-84.
54. Abrahamsson, T.R., R.Y. Wu, and M.C. Jenmalm, *Gut microbiota and allergy: the importance of the pregnancy period.* Pediatr Res, 2015. **77**(1-2): p. 214-9.
55. Perez, P.F., et al., *Bacterial imprinting of the neonatal immune system: lessons from maternal cells?* Pediatrics, 2007. **119**(3): p. e724-32.
56. Dotterud, C.K., et al., *Does Maternal Perinatal Probiotic Supplementation Alter the Intestinal Microbiota of Mother and Child? A Randomized Controlled Trial.* J Pediatr Gastroenterol Nutr, 2015.
57. West, C.E., *Gut microbiota and allergic disease: new findings.* Curr Opin Clin Nutr Metab Care, 2014. **17**(3): p. 261-6.
58. Hesselmar, B., A. Hicke-Roberts, and G. Wennergren, *Allergy in children in hand versus machine dishwashing.* Pediatrics, 2015. **135**(3): p. e590-7.
59. Hesselmar, B., et al., *Pacifier cleaning practices and risk of allergy development.* Pediatrics, 2013. **131**(6): p. e1829-37.
60. Laursen, M.F., et al., *Having older siblings is associated with gut microbiota development during early childhood.* BMC Microbiol, 2015. **15**(1): p. 154.
61. Woese, C.R., *Bacterial evolution.* Microbiol Rev, 1987. **51**(2): p. 221-71.
62. Weinstock, G.M., *Genomic approaches to studying the human microbiota.* Nature, 2012. **489**(7415): p. 250-6.
63. Sun, Y., et al., *A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis.* Brief Bioinform, 2012. **13**(1): p. 107-21.
64. Achtman, M. and M. Wagner, *Microbial diversity and the genetic nature of microbial species.* Nat Rev Microbiol, 2008. **6**(6): p. 431-40.

65. Hao, X., R. Jiang, and T. Chen, *Clustering 16S rRNA for OTU prediction: a method of unsupervised Bayesian clustering.* Bioinformatics, 2011. **27**(5): p. 611-8.
66. Caporaso, J.G., et al., *QIIME allows analysis of high-throughput community sequencing data.* Nat Methods, 2010. **7**(5): p. 335-6.
67. Schloss, P.D., et al., *Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.* Appl Environ Microbiol, 2009. **75**(23): p. 7537-41.
68. Chen, W., et al., *A comparison of methods for clustering 16S rRNA sequences into OTUs.* PLoS One, 2013. **8**(8): p. e70837.
69. Wang, Q., et al., *Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.* Appl Environ Microbiol, 2007. **73**(16): p. 5261-7.
70. Pruesse, E., et al., *SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB.* Nucleic Acids Res, 2007. **35**(21): p. 7188-96.
71. DeSantis, T.Z., et al., *Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB.* Appl Environ Microbiol, 2006. **72**(7): p. 5069-72.
72. Wang, Y., et al., *MetaCluster 5.0: a two-round binning approach for metagenomic data for low-abundance species in a noisy sample.* Bioinformatics, 2012. **28**(18): p. i356-i362.
73. Flynn, J.M., et al., *Toward accurate molecular identification of species in complex environmental samples: testing the performance of sequence filtering and clustering methods.* Ecol Evol, 2015. **5**(11): p. 2252-66.
74. Sun, Y., et al., *ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences.* Nucleic Acids Res, 2009. **37**(10): p. e76.
75. Schloss, P.D. and J. Handelsman, *Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness.* Appl Environ Microbiol, 2005. **71**(3): p. 1501-6.
76. Cai, Y. and Y. Sun, *ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time.* Nucleic Acids Res, 2011. **39**(14): p. e95.
77. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-1.
78. Li, W., L. Jaroszewski, and A. Godzik, *Clustering of highly homologous sequences to reduce the size of large protein databases.* Bioinformatics, 2001. **17**(3): p. 282-3.
79. Avershina, E. and K. Rudi, *Confusion about the species richness of human gut microbiota.* Benef Microbes, 2015. **in press**.
80. Edgar, R.C., *UPARSE: highly accurate OTU sequences from microbial amplicon reads.* Nat Methods, 2013. **10**(10): p. 996-8.
81. Salipante, S.J., et al., *Performance comparison of Illumina and ion torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling.* Appl Environ Microbiol, 2014. **80**(24): p. 7583-91.
82. Koenig, J.E., et al., *Succession of microbial consortia in the developing infant gut microbiome.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4578-85.
83. Kummeling, I., et al., *Etiology of atopy in infancy: the KOALA Birth Cohort Study.* Pediatr Allergy Immunol, 2005. **16**(8): p. 679-84.
84. Madsen, A.L., et al., *Determinants of blood glucose and insulin in healthy 9-month-old term Danish infants; the SKOT cohort.* Diabet Med, 2010. **27**(12): p. 1350-7.

85. Takaro, T.K., et al., *The Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort study: assessment of environmental exposures.* J Expo Sci Environ Epidemiol, 2015.
86. Eggesbo, M., et al., *Development of gut microbiota in infants not exposed to medical interventions.* APMIS, 2011. **119**(1): p. 17-35.
87. Storrø, O., et al., *A primary health-care intervention on pre- and postnatal risk factor behavior to prevent childhood allergy. The Prevention of Allergy among Children in Trondheim (PACT) study.* BMC Public Health, 2010. **10**: p. 443.
88. Penders, J., et al., *Establishment of the intestinal microbiota and its role for atopic dermatitis in early childhood.* J Allergy Clin Immunol, 2013. **132**(3): p. 601-607 e8.
89. Bergstrom, A., et al., *Establishment of intestinal microbiota during early life: a longitudinal, explorative study of a large cohort of Danish infants.* Appl Environ Microbiol, 2014. **80**(9): p. 2889-900.
90. Eggesbo, M., S. Mandal, and T. Midtvedt, *Factors affecting infant gut microbiota and possible consequences for health.* Microb Ecol Health Dis, 2015. **26**: p. 28062.
91. Dotterud, C.K., et al., *Probiotics in pregnant women to prevent allergic disease: a randomized, double-blind trial.* Br J Dermatol, 2010. **163**(3): p. 616-23.
92. Øien, T., O. Storrø, and R. Johnsen, *Intestinal microbiota and its effect on the immune system--a nested case-cohort study on prevention of atopy among small children in Trondheim: the IMPACT study.* Contemp Clin Trials, 2006. **27**(4): p. 389-95.
93. Conroy, M.E., H.N. Shi, and W.A. Walker, *The long-term health effects of neonatal microbial flora.* Curr Opin Allergy Clin Immunol, 2009. **9**(3): p. 197-201.
94. Neufeld, K.A., et al., *Effects of intestinal microbiota on anxiety-like behavior.* Commun Integr Biol, 2011. **4**(4): p. 492-4.
95. Diaz Heijtz, R., et al., *Normal gut microbiota modulates brain development and behavior.* Proc Natl Acad Sci U S A, 2011. **108**(7): p. 3047-52.
96. Zimonja, M., et al., *Multivariate curve resolution of mixed bacterial DNA sequence spectra: identification and quantification of bacteria in undefined mixture samples.* Journal of Chemometrics, 2008. **22**(5): p. 309-322.
97. Ley, R.E., D.A. Peterson, and J.I. Gordon, *Ecological and evolutionary forces shaping microbial diversity in the human intestine.* Cell, 2006. **124**(4): p. 837-48.
98. Ishikawa, E., et al., *Ethnic diversity of gut microbiota: species characterization of Bacteroides fragilis group and genus Bifidobacterium in healthy Belgian adults, and comparison with data from Japanese subjects.* J Biosci Bioeng, 2013. **116**(2): p. 265-70.
99. Duranti, S., et al., *Exploration of the genomic diversity and core genome of the Bifidobacterium adolescentis phylogenetic group by means of a polyphasic approach.* Appl Environ Microbiol, 2013. **79**(1): p. 336-46.
100. Bottacini, F., et al., *Comparative genomics of the Bifidobacterium breve taxon.* BMC Genomics, 2014. **15**: p. 170.
101. LoCascio, R.G., et al., *Broad conservation of milk utilization genes in Bifidobacterium longum subsp. infantis as revealed by comparative genomic hybridization.* Appl Environ Microbiol, 2010. **76**(22): p. 7373-81.
102. Chaplin, A.V., et al., *Intraspecies Genomic Diversity and Long-Term Persistence of Bifidobacterium longum.* PLoS One, 2015. **10**(8): p. e0135658.
103. Hinde, K. and Z.T. Lewis, *MICROBIOTA. Mother's littlest helpers.* Science, 2015. **348**(6242): p. 1427-8.

104. Guo, X., et al., *Innate Lymphoid Cells Control Early Colonization Resistance against Intestinal Pathogens through ID2-Dependent Regulation of the Microbiota.* Immunity, 2015. **42**(4): p. 731-43.
105. Jakobsson, H.E., et al., *Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section.* Gut, 2014. **63**(4): p. 559-66.
106. Martin, R., et al., *Cultivation-independent assessment of the bacterial diversity of breast milk among healthy women.* Res Microbiol, 2007. **158**(1): p. 31-7.
107. Gronlund, M.M., et al., *Influence of mother's intestinal microbiota on gut colonization in the infant.* Gut Microbes, 2011. **2**(4): p. 227-33.
108. Aagaard, K., et al., *A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy.* PLoS One, 2012. **7**(6): p. e36466.
109. Romero, R., et al., *The composition and stability of the vaginal microbiota of normal pregnant women is different from that of non-pregnant women.* Microbiome, 2014. **2**(1): p. 4.
110. Song, S.J., et al., *Cohabiting family members share microbiota with one another and with their dogs.* Elife, 2013. **2**: p. e00458.
111. Yachi, S. and M. Loreau, *Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis.* Proc Natl Acad Sci U S A, 1999. **96**(4): p. 1463-8.
112. Magurran, A.E., *Measuring biological diversity*. 2004: Blackwell Science Ltd.
113. Tap, J., et al., *Gut microbiota richness promotes its stability upon increased dietary fiber intake in healthy adults.* Environ Microbiol, 2015.
114. Zhalnina, K., et al., *Soil pH determines microbial diversity and composition in the park grass experiment.* Microb Ecol, 2015. **69**(2): p. 395-406.
115. Hartmann, M., et al., *Distinct soil microbial diversity under long-term organic and conventional farming.* ISME J, 2015. **9**(5): p. 1177-94.
116. Agans, R., et al., *Distal gut microbiota of adolescent children is different from that of adults.* FEMS Microbiol Ecol, 2011. **77**(2): p. 404-12.
117. Koren, O., et al., *Host remodeling of the gut microbiome and metabolic changes during pregnancy.* Cell, 2012. **150**(3): p. 470-80.
118. DiGiulio, D.B., et al., *Temporal and spatial variation of the human microbiota during pregnancy.* Proc Natl Acad Sci U S A, 2015.
119. Dogra, S., et al., *Rate of establishing the gut microbiota in infancy has consequences for future health.* Gut microbes, 2015.
120. Schwiertz, A., *Studies on the human gut microbiota. Relevance for maturation, inflammation and nutritional processes.* , in *Agricultural Sciences, Nutritional Sciences and Environmental Management*. 2011, Justus-Liebig-University of Giessen. p. 115.
121. Kwok, L.Y., et al., *Characterization of fecal microbiota across seven Chinese ethnic groups by quantitative polymerase chain reaction.* PLoS One, 2014. **9**(4): p. e93631.
122. Sekelja, M., et al., *Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach.* ISME J, 2011. **5**(3): p. 519-31.
123. Nicholson, J.K., et al., *Host-gut microbiota metabolic interactions.* Science, 2012. **336**(6086): p. 1262-7.
124. Lukes, J., et al., *Are Human Intestinal Eukaryotes Beneficial or Commensals?* PLoS Pathog, 2015. **11**(8): p. e1005039.

PAPER 1

Avershina E., Storrø O., Øien T., Johnsen R., Pope P. & Rudi K. (2014) Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS microbiology ecology* **87**: 280-290. DOI: 10.1111/1574-6941.12223

RESEARCH ARTICLE

Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children

Ekaterina Avershina[1], Ola Storrø[2], Torbjørn Øien[2], Roar Johnsen[2], Phil Pope[1] & Knut Rudi[1]

[1]Department of Chemistry, Biotechnology and Food Science, University of Life Sciences, Ås, Norway; and [2]Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway

Correspondence: Ekaterina Avershina, Department of Chemistry, Biotechnology and Food Science, University of Life Sciences, Ås, Norway. Tel.: +47 64 96 59 00; fax: +47 64 96 59 01; e-mail: ekaterina.avershina@umb.no

Received 26 April 2013; revised 13 September 2013; accepted 13 September 2013.

DOI: 10.1111/1574-6941.12223

Editor: Julian Marchesi

Keywords
16S rRNA gene; infant; gut microbiota.

Abstract

Despite the importance, the diversity of the human infant gut microbiota still remains poorly characterized at the regional scale. Here, we investigated the faecal microbiota diversity in a large *16S rRNA* gene data set from a healthy cohort of 86 mothers and their children from the Trondheim region in Norway. Samples were collected from mothers during early and late pregnancy, as well as from their children at 3 days, 10 days, 4 months, 1 year and 2 years of age. Using a combination of Sanger sequencing of amplicon mixtures (without cloning), real-time quantitative PCR and deep pyrosequencing, we observed a clear age-related colonization pattern in children that was surprisingly evident between 3- and 10-day samples. In contrast, we did not observe any shifts in microbial composition during pregnancy. We found that alpha-diversity was highest at 2 years and lowest at 4 months, whereas beta-diversity estimates indicated highest interindividual variation in newborns. Variation significantly decreased by the age of 10 days and was observed to be convergent over time; however, there were still major differences between 2 years and adults whom exhibited the lowest interindividual diversity. Taken together, the major age-affiliated population shift within gut microbiota suggests that there are important mechanisms for transmission and persistence of gut bacteria that remain unknown.

Introduction

While it is widely accepted that the human gut is one of the most densely populated bacterial communities on the Earth (Whitman *et al.*, 1998), the general mechanisms for host–bacterial interactions are not yet completely described (Avershina & Rudi, 2013). Previously, the scientific community unanimously assumed that humans are born sterile (Ley *et al.*, 2006; Marques *et al.*, 2010), although an evidence now exists for prenatal colonization (Jimenez *et al.*, 2008; Satokari *et al.*, 2009). Regardless of the time required for initial colonization, it is absolute that development of this unique and intricate community takes several years to reach its maturity (Marchesi, 2011). There are many factors that supposedly play a role in the development of gut microbiota; initial inoculation occurs via the mother's birth canal when a child is born vaginally; subsequently, an infant will frequently receive bacteria via breast milk (Martin *et al.*, 2007), and the surrounding environment also exerts a constant influence. Existing reports have addressed various environmental influences towards gut microbiota such as age (Palmer *et al.*, 2007; Claesson *et al.*, 2011), geography and diet (De Filippo *et al.*, 2010; Yatsunenko *et al.*, 2012). There are also recent suggestions of immunological modulations of the microbiota during pregnancy (Koren *et al.*, 2012). However, much less is known about transmission and persistence of gut bacteria in a population during the host's first years of life. We have previously described transmission of some particular gut bacteria from mother to child (Bjerke *et al.*, 2011; de Muinck *et al.*, 2011; Avershina *et al.*, 2013), while we have not yet addressed general patterns of bacterial persistence and diversity in a healthy randomly selected population of children and their mothers.

FEMS MICROBIOLOGY ECOLOGY

The aim of this study was therefore to address longitudinal faecal microbiota shifts in composition and diversity in children and their mothers in a geographically restricted cohort. We analysed stool samples from 86 mother/child pairs, collected two times during the mothers pregnancy (15.0 ± 4.2 and 37.5 ± 1.8 gestation weeks) and five times from infants (ages 3 days and 10 days, 4 months, 1 year and 2 years). We used a polyphasic analytical approach consisting of direct mixed 16S rRNA gene Sanger sequencing (analysis of electropherograms containing information on all amplicon variants) (Zimonja *et al.*, 2008), real-time quantitative PCR (Ginzinger, 2002) and 454-sequencing (Ronaghi, 2001). We present results suggesting highly age-dependent bacterial persistence and diversity patterns within the population. Furthermore, we also present support for mother to child transmission of adult associated gut bacteria – surprisingly not during the birth process, but at a later stage.

Materials and methods

Study material and sample preparation

Faecal samples were collected from the IMPACT cohort study among small children and mothers in Trondheim, which is a nested cohort within the PACT study (Prevention of Allergy among Children in Trondheim) (Storro *et al.*, 2010). Most of the children were delivered vaginally (90%) and at term (90%). There was a high frequency of breast feeding, and 97% of infants were breast-fed during the first 6 weeks of life. By the age of 4 months, 66.7% of infants were exclusively breast-fed, 23.8% were receiving either formula or solid food (fruits, vegetables, wheat, bread, corn, rice) complementary to breast milk, and 9.5% of infants were receiving only formula and/or solid food. More details about the cohort characteristics are given by Storro *et al.* (2011).

Faecal specimens were stored in sterile Cary–Blair transport and holding medium (BD Diagnostics Sparks, MD 21152). Each specimen was frozen at −20 °C within 2 h after defecation and transported to the laboratory for further storage at −80 °C within 1 day (for children) or 4 weeks (for pregnant women). Details about the IMPACT faecal material is given by Oien *et al.* (2006). The data set analysed contained samples from both early (first to second trimester) and late pregnancy (third trimester) from the mothers and 3 days, 10 days, 4 months, 1 year and 2 years from the children.

We purified faecal DNA with paramagnetic beads in accordance with an optimized and automated protocol (Skanseng *et al.* 2006). Briefly, this protocol involved mechanical lysis with glass beads and DNA purification with silica particles. Mechanical lysis was chosen because the compositions of the gut bacteria cell walls are largely unknown.

Direct mixed sequence analysis

The V3–V4 region of *16S rRNA* gene was PCR-amplified using the primers targeting universally conserved gene regions (Nadkarni *et al.*, 2002). Subsequently the V4 region (198 bp) was targeted for sequencing using a mixed Sanger approach. The resulting sequence spectra contained information for the *16S rRNA* genes representative of all the bacteria in a given sample.

The alpha- and beta-diversity of each spectrum was assessed by means of a modified Simpson's diversity index c_{mixed} (Eqn. 1) and modified Bray–Curtis dissimilarity index (Eqn. 2), respectively. Calculations were based on the fluorescence intensity fractions of each nucleotide position. The rationale is that these intensity fractions will reflect diversity. In case there is only one bacterium in a sample, there will be only one nucleotide in every position of the sequence spectrum, and therefore, nucleotide fractions in every position will equal 1 : 0 : 0 : 0. In the case of a mixture of a range of different bacteria, though, the fractions will converge towards 0.25 : 0.25 : 0.25 : 0.25. Based on these fractions, one could estimate diversity in a sample, which is independent of operational taxonomic units (OTUs).

$$1/c_{\text{mixed}} = \frac{\sum_{i=1}^{n}(G_i)^2 + \sum_{i=1}^{n}(A_i)^2 + \sum_{i=1}^{n}(T_i)^2 + \sum_{i=1}^{n}(C_i)^2}{n} \quad (1)$$

$$\text{BC}_{ij} = \frac{\sum_{k=1}^{n}|G_{ki} - G_{kj}| + \sum_{k=1}^{n}|A_{ki} - A_{kj}| + \sum_{k=1}^{n}|T_{ki} - T_{kj}| + \sum_{k=1}^{n}|C_{ki} - C_{kj}|}{\sum_{k=1}^{n}|G_{ki} + G_{kj}| + \sum_{k=1}^{n}|A_{ki} + A_{kj}| + \sum_{k=1}^{n}|T_{ki} + T_{kj}| + \sum_{k=1}^{n}|C_{ki} + C_{kj}|} \quad (2)$$

Detailed description of the diversity indices calculations is given in the study by Avershina *et al.* (2013). Beta-diversity was assessed both between samples belonging to the same age group and between samples belonging to the same mother–child pair, but at different time points.

Significant difference between indices at various time points was tested using Friedman's test – a nonparametric version of two-way ANOVA – which takes into account possible correlation between the measurements (MATLAB® documentation, 2010). For those samples, where we did not expect the correlation, Kruskal–Wallis test was used. The null hypothesis was rejected at the level of 5%.

Information on the most dominant bacteria was subsequently resolved using multivariate curve resolution analysis (MCR-ALS). This analysis allows recovery of the common information contained between the samples of interest into so-called components, as well as simultaneous relative quantification of this information in all the samples (Zimonja *et al.*, 2008). Taxonomic level of components' resolution for nondefined bacterial assemblages directly depends on the diversity represented within a data set (Rudi *et al.*, 2012; Sekelja *et al.*, 2012). If a given phylum is represented by one clearly dominant genus, then the signature sequence for this genus will be resolved as a component. While if there were several equally distributed genera within the same family, then the signature sequence for this family would have been recovered. Prior to MCR-ALS, one needs to specify the number of components to be resolved. In case the set number is too high, the 'real' component would be split, and thus, at least two of the resolved components would contain the same information. This can be detected by biological reasoning because these components will then represent the same taxonomic group. To define the initial number of components (initial estimates i), we used both principal component analysis (PCA) and evolving factor analysis (EFA) as recommended (Tauler *et al.*, 1995). The detailed description of use of MCR-ALS for mixed sequence resolution can be found in the study by Avershina *et al.* (2013). Resolved components spectra were manually base-called and classified by Ribosomal Database Project (RDP) hierarchical classifier (Wang *et al.*, 2007).

To address the longitudinal structure of the MCR-ALS score data, that is, relative abundance of resolved components, parallel factor analysis (PARAFAC) method was used. PARAFAC is a multiway generalization of the two-way PCA. However, unlike PCA, the rotation problem is omitted so that pure components can be resolved (Bro, 1997). The core consistency index was used as a criterion for determining the number of components.

Real-time quantitative PCR

We have previously qPCR-amplified the 16S rRNA gene of commonly identified gut bacteria, as well as some pathogenic bacterial species (Storro *et al.*, 2011) for the same study cohort. Among tested species were *Bacteroides fragilis*, *Bifidobacterium longum*, *Bifidobacterium breve*, *Bifidobacterium animalis* subsp. *lactis*, genus *Bifidobacterium*, *Clostridium difficile*, *Clostridium perfringens*, *Lactobacillus rhamnosus*, *Lactobacillus reuteri* and *Helicobacter pylori*. For this work, we binarized these data based on whether the given bacterium was detected in a sample. For every age, unweighted Cohen's kappa indices (Sim & Wright, 2005) were calculated to evaluate whether there was an agreement between detection of a given bacteria in mothers and in children. Interpretation of the index was performed using guidelines provided in the MATLAB® script for Cohen's kappa index calculation (Cardillo, 2007). The relative amount of the detected vs. nondetected populations of bacteria is represented in Supporting Information, Fig. S1. 'Nondetected' populations were defined as populations that did not show amplification after 40 cycles. Some bacteria (*L. rhamnosus* and *C. difficile*) were not detected in any of the mothers, whereas others (e.g. *H. pylori*) were detected only in two mothers (Table S1). Therefore, to ensure sufficient amount of information, only bacterial groups that were detected in more than 11 mothers were included in the analysis. The bacterial groups that satisfied this criterion were *B. longum*, genus *Bifidobacterium*, *B. fragilis* and *Escherichia coli*. We also addressed the persistence patterns of these four bacteria in a population by calculating the fraction of individuals, in which the species was detected at a time point 'x' given it was detected at a time point '$x-1$'.

Pyrosequencing analysis

A subset of seven random mother and child pairs were selected for deep 454-sequencing from the pairs with the most complete temporal series in the main study cohort. DNA isolation, amplicon and PCR conditions were the same as for direct sequencing approach. The only difference was the modification of PCR primers targeting V3–V4 region of *16S rRNA* gene, to be adapted to the GS-FLX Titanium Chemistry (454 Life Sciences). Sequencing was performed according to the manufacturer's recommendations at the Norwegian High-Throughput Sequencing Centre (Oslo, Norway). Pyrosequencing data were analysed using QIIME pipeline (Caporaso *et al.*, 2010). Error correction, chimera removal and operational taxonomic unit (OTUs) clustering were performed using USEARCH quality filtering with QIIME, which incorporates UCHIME (Edgar *et al.*, 2011) and a 97% sequence identity threshold. The RDP classifier (Wang *et al.*, 2007) was used to assign taxonomic identity to the resulting OTUs. For a phylogeny-based diversity assessment, we used weighted UniFrac hierarchical clustering (Lozupone & Knight, 2005) based on 10 rarefactions with 1600 randomly selected sequences per sample for each rarefaction.

To investigate what shapes gut microbiota in both infancy and adulthood, we fitted observed species distributions to commonly used distributions using the Species Diversity and Richness, version 4.1.2 (PISCES Conservation Ltd., UK), software. Hubbell's model of neutrality, often used as a null model of community structure (Magurran, 2004), assumes that when an individual dies in a saturated community, the probability of its replacement by an offspring of rare species is the same as by an offspring of a more abundant species. Jabot & Chave (2011) have developed a generalization of this model introducing a parameter δ. This parameter estimates the non-neutrality of the system based on the deviation of observed species evenness as opposed to the system being best described by neutral model. When δ is positive, dominant species have higher chance of taking the place of the dead individual, whereas negative values indicate that rare species' chances increase. Based on 1000 randomly selected sequences per sample from the chimera- and noise-free pyrosequencing data set, we calculated

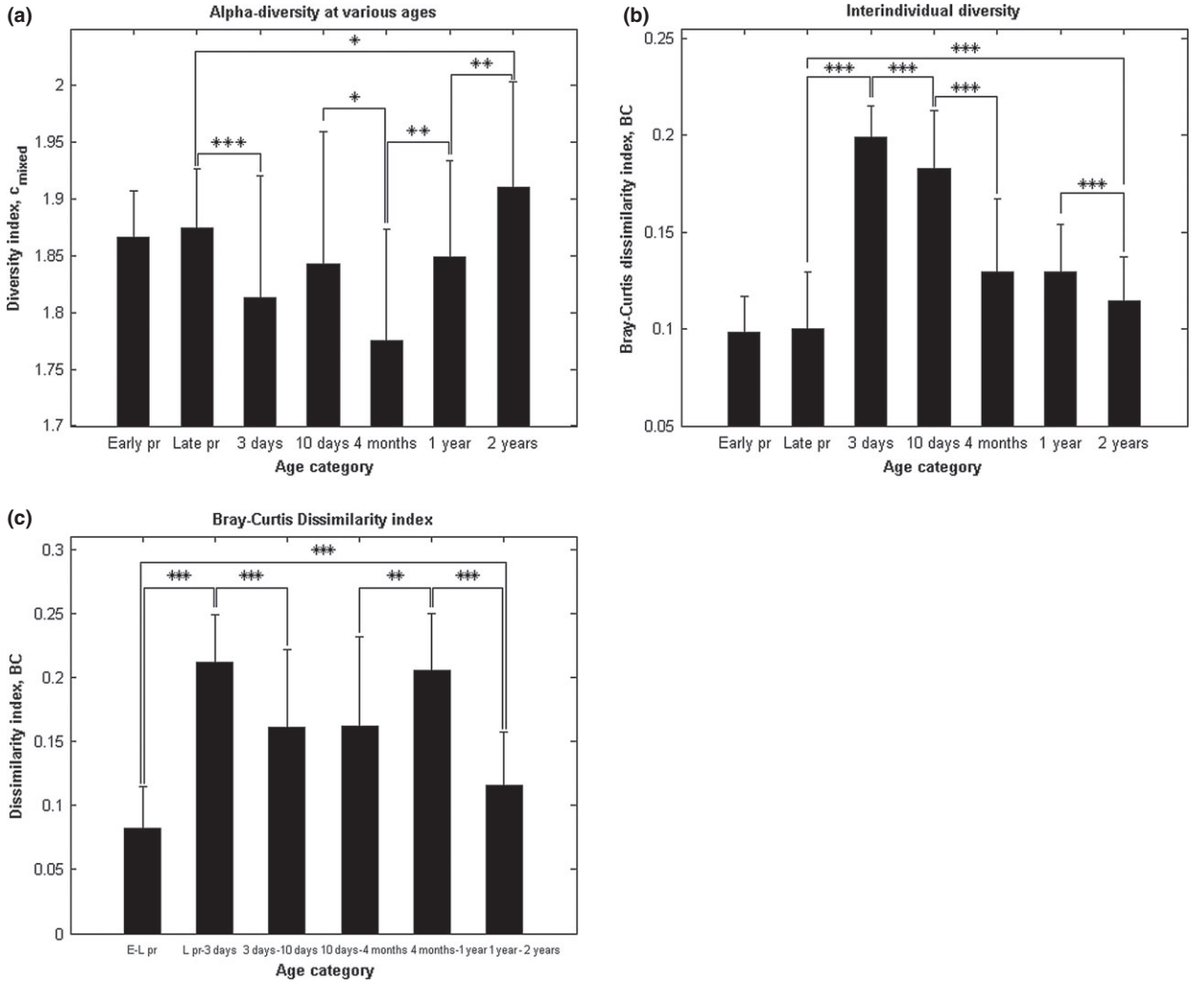


Fig. 1. Nucleotide diversity measurements. The significance in difference between diversity indices at two subsequent time points was calculated with the Friedman's (a and b) and Kruskal–Wallis (c) tests. $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$. Early period (pr) and Late pr: Early (8–20 weeks) and late (30–40 weeks) pregnancy periods, respectively. (a) The modified Simpson's index of nucleotide spectra diversity c_{mixed} at various ages. (b) The modified Bray–Curtis index of nucleotide dissimilarity (BC) between individuals at various ages. Early pr and Late pr: early (8–20 weeks) and late (30–40 weeks) pregnancy periods, respectively. (c) The modified Bray–Curtis index of nucleotide dissimilarity (BC) between the subsequent time points. E–L pr: the period between early (8–20 weeks) and late (30–40 weeks) pregnancy periods; L pr–3 day: comparison between 3-day-old newborns and their mothers during the late pregnancy stage; 3–10 day: between 3 and 10 days of age; 10 day–4 month: between 10 days and 4 months of age; 4 month–1 year: between 4 months and 1 year of age; 1–2 year: between 1 and 2 years of age. The error bars represent standard error of the mean.

non-neutrality parameter δ using Parthy, version 1.0, software (Jabot & Chave, 2011).

Results

Mixed sequence analysis

Nucleotide alpha-diversity (Simpson's diversity index) of mixed spectra ranged from 1.77 ± 0.10 [mean ± standard deviation] at 4-month-old to 1.91 ± 0.09 at 2-year-old infants (Fig. 1a). Generally, diversity of adult' stool samples was higher than that of newborns ($P = 0.0001$) and 4-month-old infants ($P = 2.26 \times 10^{-9}$). At 1 year of age, the diversity increased compared with 4-month-olds ($P = 0.0028$) and then further increased by 2 years of age ($P = 0.0054$).

Newborns exhibited highest beta-diversity between individuals (modified Bray–Curtis index BC = 0.20 ± 0.02 and 0.18 ± 0.03 for 3- and 10-day-old infants, respectively; Fig. 1b). By the age of 4 months, the variation within the population had significantly decreased ($P = 7.51 \times 10^{-13}$) and remained the same up to 1 year. Although the beta-diversity between stool samples from 2-year-olds was significantly lower than that of 1-year-olds ($P = 1.54 \times 10^{-5}$), it was still significantly higher than the beta-diversity between adult stool samples ($P = 4.38 \times 10^{-6}$). In addition to interindividual comparisons, beta-diversity estimations were used to analyse intraindividual variation that developed within an individual from one time point to another (Fig. 1c). The highest variation (highest beta-diversity) was observed between the spectra of mothers at their late pregnancy stage and 3-day-old infants (BC = 0.21 ± 0.04), as well as between 4-month-old and 1-year-old children (BC = 0.20 ± 0.04), whereas the least variation (lowest beta-diversity) was observed between stool samples collected from mothers at two pregnancy trimesters (BC = 0.08 ± 0.03) and also between 1- and 2-year-olds (BC = 0.12 ± 0.02).

Both PCA and EFA suggested six components to be resolved by MCR-ALS. When six components were used, the information on *Bacteroidetes* group was entirely absent. Therefore, MCR-ALS was repeated by gradually increasing the number of components to be resolved until the duplication event. In total, eight components accounting for 70% of the variation in the system were resolved by MCR-ALS and classified by RDP classifier (Table S2).

Taxonomically, stool samples analysed from mothers were rich in *Lachnospiraceae*- and *Faecalibacterium*-affiliated components (Fig. 2). At 3 days, all eight components seemed to be evenly represented, but by the age of 10 days, there was a significant decrease in the level of *Lactobacillales* ($P = 0.0191$). By the age of 4 months, bifidobacteria constituted 57.6% of total gut microbiota, whereas *Lactobacillales*- and *Streptococcus*-affiliated components were diminished ($P = 0.0135$ and $P = 0.0001$, respectively). At 1 and 2 years of age, average composition resembled that of pregnant women, although there were several pronounced differences. For example, the *Bifidobacterium*-affiliated ($P = 0.0042$ and $P = 0.0021$ for 1 and 2 years, respectively) and other *Actinobacteria*-($P = 0.0016$ and $P = 2.3 \times 10^{-5}$ for 1 and 2 years, respectively) components were higher in children than in their mothers, whereas *Faecalibacterium*- ($P = 4.3 \times 10^{-6}$ and $P = 5.9 \times 10^{-7}$ for 1 and 2 years, respectively) and *Bacteroides*-affiliated ($P = 1.4 \times 10^{-5}$ and $P = 5.6 \times 10^{-8}$ for 1 and 2 years, respectively) components were lower.

Due to the fact that the majority of infants were born vaginally, were at term and were breast-fed during the first days of life, we could not investigate the effect of birth mode and diet. However, we could test whether implementation of solid food (wheat, rice, corn) at

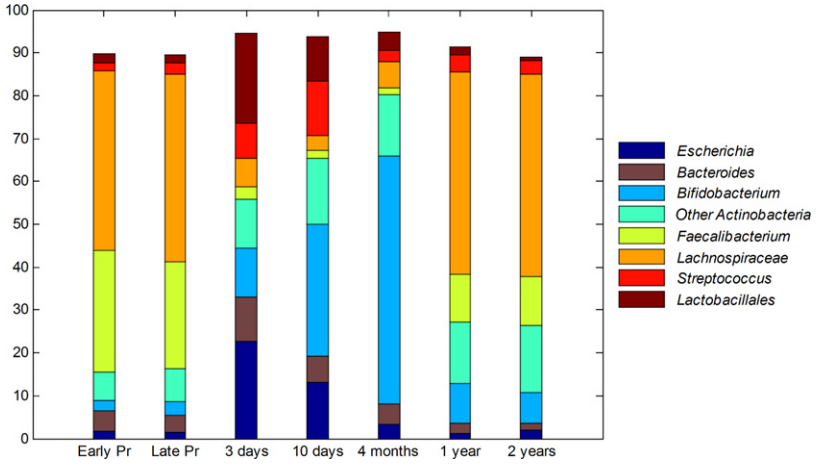


Fig. 2. Bacterial species composition in stool samples of infants (from 3 days to 2 years of age) and their mothers during pregnancy as revealed by MCR-ALS. Early pr and Late pr: early (8–20 weeks) and late (30–40 weeks) pregnancy periods, respectively.

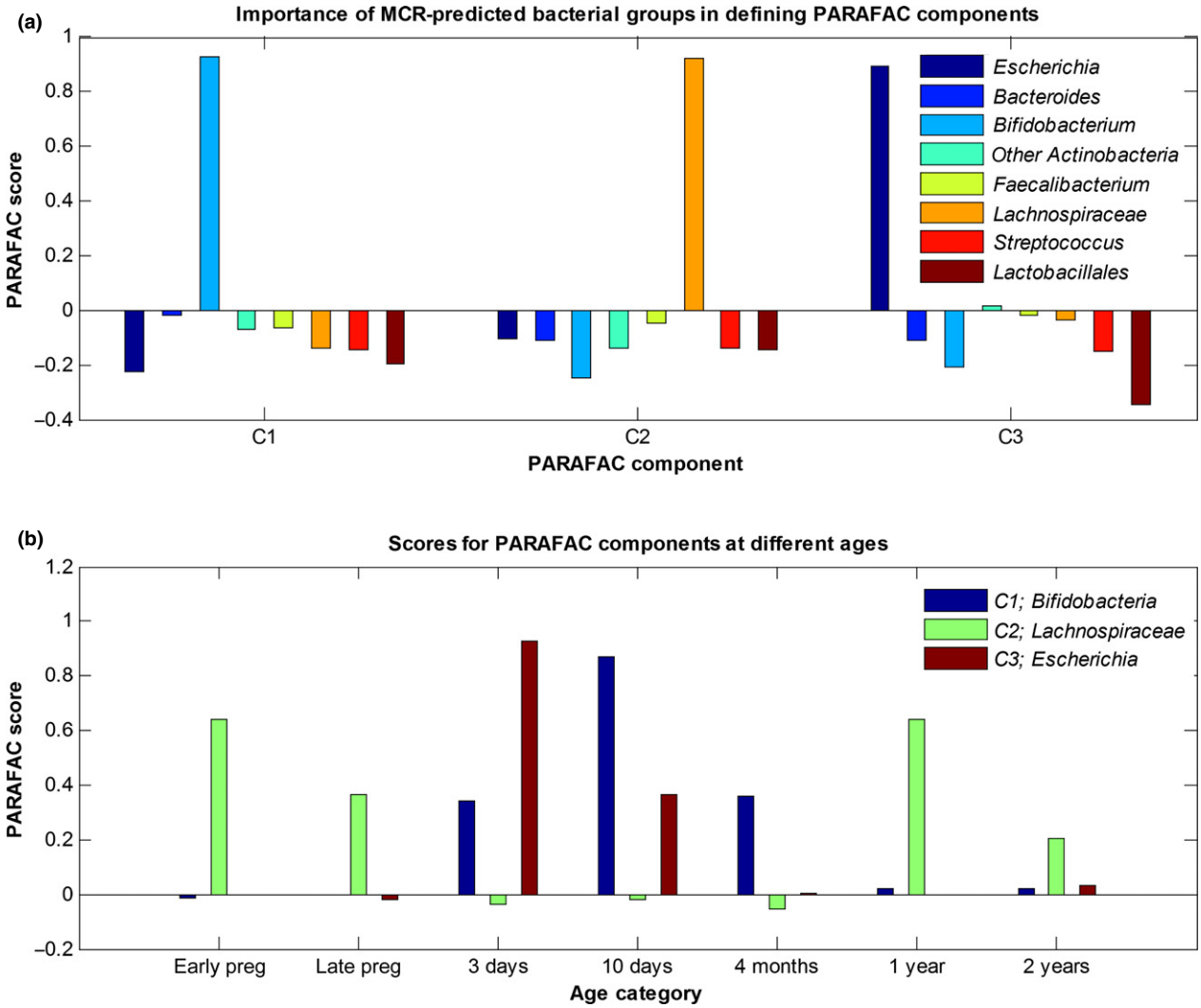


Fig. 3. Summary of PARAFAC on relative abundances of MCR-ALS resolved bacterial groups. C1, C2, C3 – PARAFAC components. Early pr and Late pr: early (8–20 weeks) and late (30–40 weeks) pregnancy periods, respectively. (a) PARAFAC-suggested components C1, C2 and C3 represent *Bifidobacterium*, *Lachnospiraceae* and *Escherichia* components, respectively. (b) At early days of life, C1 and C3 determined the variation in the system, whereas at pregnancy, 1 and 2 years of life, C2 became more important.

4 months would affect faecal microbial composition. These analyses showed no significant difference in relative composition of gut microbiota.

To investigate longitudinal structure in the data (i.e. individual sharing of bacteria for more than one time point), 3-component PARAFAC model was deduced based on a core consistency index of more than 99%. The loadings for the MCR-ALS components dimension indicate that *Escherichia*-, *Bifidobacterium*- and *Lachnospiraceae*-affiliated components influenced the longitudinal structure of the data (Fig. 3a). In particular, the *Escherichia*-affiliated component was associated with 3 and 10 days, *Bifidobacterium*-affiliated component was associated with 3 days, 10 days and 4 months, while *Lachnospiraceae*-affiliated component was associated with early and late pregnancy, in addition to 1 and 2 years (Fig. 3b).

Real-time quantitative PCR analysis of prevalence

Figure 4 illustrates qPCR prevalence data calculated for selected bacterial groups both for the whole study cohort and for a subpopulation of children whose mothers tested positive for the target bacterium (mother–child positive subpopulation). At 10 days, *E. coli* was more frequently detected in those children whose mothers also tested positive for this bacterium ($P = 0.002$). Interestingly, the

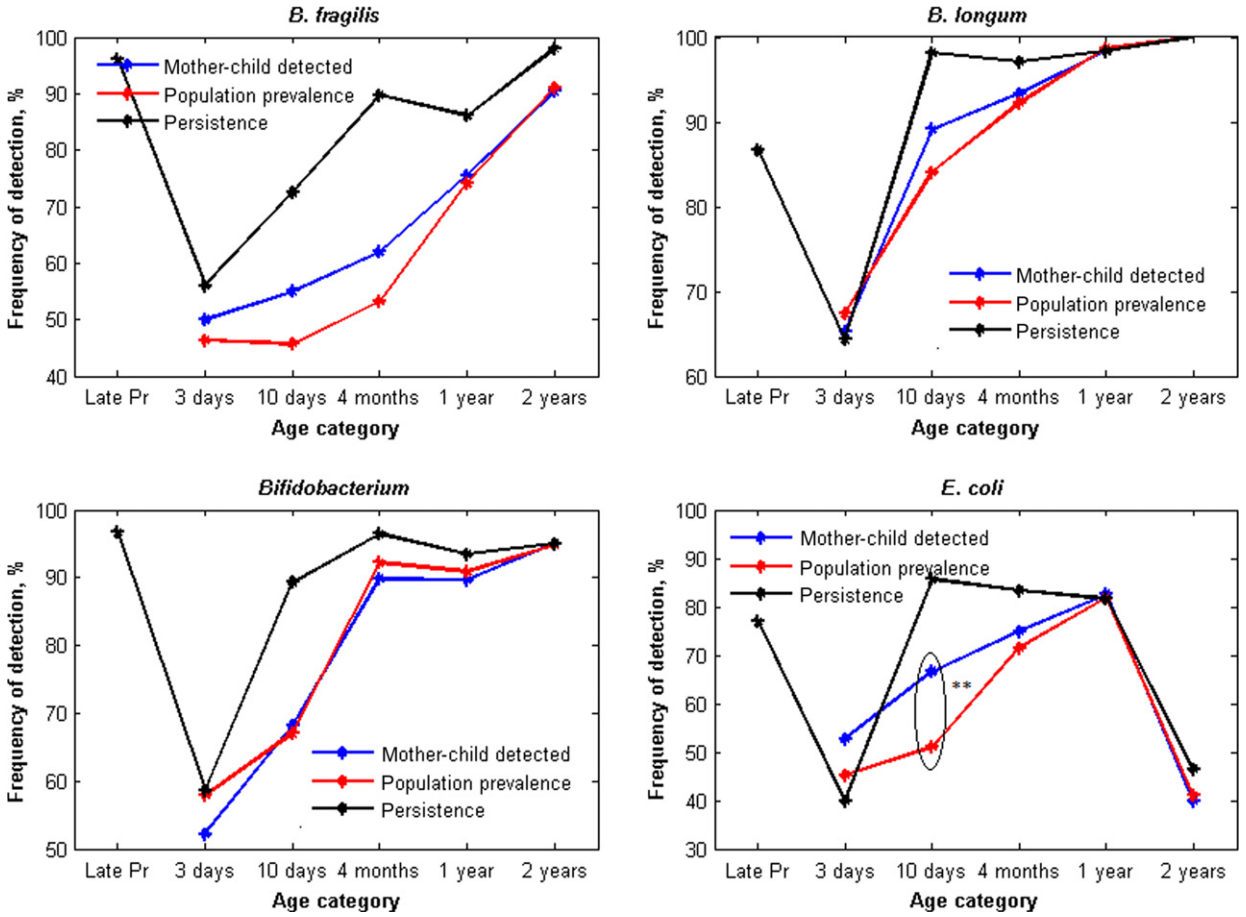


Fig. 4. Prevalence of bacterial species in a population of children at various ages. Blue line indicates prevalence of bacteria in a subpopulation of children in whose mothers it was also detected; red line – in a total population of children of a given age. Black line depicts the percentage of individuals in whom bacteria were detected in both a given and a previous time point compared with a total number of individuals where it was detected in a previous time point. Late pr: late (30–40 weeks) pregnancy period. **one-sided binomial test *P*-value < 0.01.

difference between detection frequencies of this bacterium in mother–child positive subpopulation and total children population was higher in 10 days as compared to 3 days. This may indicate either postnatal or very low at-birth transmission of this bacterial species. *B. longum* was deemed to be one of the most persistent colonizers among the four bacterial groups tested. Already by the age of 10 days, it was detected in nearly all infants who tested positive at 3 days after birth (Fig. 4). Even by the age of 2 years, this species persisted in the majority of infants who previously tested positive. In contrast, *E. coli* detection was observed to be stable during the first year (80–85% of population). However, by 2 years, a detection limit had decreased to 45% of children who previously tested positive.

Cohen's kappa index was used to indicate the magnitude of agreement between the detection of a given bacteria in an individual mother and her child (in the whole cohort). In our data set, the index ranged from −0.05 (poor agreement) to 0.30 (fair agreement) and was observed to decrease with age, indicating that the detection of a given bacterium in 1- to 2-year-old children was less dependent on their mother testing positive (Table 1). In concurrence with qPCR prevalence data (Fig. 4), Cohen's kappa indices indicated slight to fair agreement for both *E. coli* and *B. fragilis*. The ranking is based on the guidelines to the MATLAB® script for the index calculation (Cardillo, 2007). *Bifidobacterium* was observed to be

Table 1. Cohan's kappa index – estimate of an agreement in detection of a given bacteria in mothers and in their infants.

Age	*B. fragilis*	*B. longum*	*Bifidobacterium*	*E. coli*
3 days	0.18	0.07	0.04	0.17
10 days	0.24	0	0.04	0.3
4 months	0.27	−0.03	−0.05	0.02
1 year	0.1	−0.02	−0.05	0.01
2 years	0.1	0	−0.04	−0.07

Calculations are based on detection of a given bacteria by RT-PCR.

negative at 4 months, indicating poor agreement in mother–child detection patterns. High *P*-values (> 0.05) also support low correspondence between detection of a given bacteria in mothers and in children.

Pyrosequencing data analysis

Eight samples, mostly belonging to one mother–child pair, were removed from the analysis due to a low number of recovered sequences (less than 2000 sequences per sample). Therefore, the analysis was performed on a total of 39 samples from 6 children and 5 mothers. After quality filtering, chimera removal and normalization, 370 207 sequences were used for subsequent analysis with a mean of 9492 sequences per sample (ranging from 2146 to 21 317 sequences per sample). Apart from one sample, stool samples from mothers' and 1- and 2-year-old infants clustered separately from stool samples of newborns and 4-month-olds based on weighted UniFrac distances (1600 sequences per sample; bootstrap values are based on 10 rarefactions; Fig. S3A). To examine how similar the faecal microbiota from different age groups was, we used Jaccard distance index calculated for detected OTUs (Fig. S3B). Overall, there was higher variation in microbiota from children when compared to mothers ($P = 0.0011$ and $P = 0.0001$ at 3 days and 2 years of age, respectively), although the microbiota of newly born children were more similar to each other than to their related ($P = 0.0010$, $P = 0.0011$ and $P = 0.0034$ for 3 days, 10 days and 4 months, respectively) and unrelated mothers ($P = 0.0011$, $P = 0.0006$ and $P = 0.0024$ for 3 days, 10 days and 4 months, respectively). By the age of 1 year, their microbiota was as similar to adults as it was to other children from the same age group.

We compared how many OTUs were shared between five children at various time points and their mothers (both related and unrelated). In total, 30 samples were used for these comparisons. From birth to 4 months of age, only one child had more OTUs shared with his own mother than with any other unrelated mother. However, by the age of 2 years, the number of children who shared more OTUs with their mothers than with other unrelated mothers increased to 3 of 5 (Table S3). We also examined which OTUs were underrepresented in children at various ages compared with their mothers (Tables S4–S8). In the immediate period after birth (days 1–3), 1230 OTUs were absent in all infant samples, of which 44% were affiliated to the family of *Lachnospiraceae*. At the age of 1–2 years, 500 OTUs were absent, composed of *c.* 30% that were affiliated to the *Lachnospiraceae*. Overall, *Lachnospiraceae*–affiliated OTUs that had representatives in all children at a given age were first detected at 1 year, although in one child, OTUs affiliated to this clostridial family were detected right after birth. In contrast, within the first days after birth, only OTUs affiliated to the *Bifidobacteriaceae*, *Streptococcaceae* and *Staphylococcaceae* were shared among all infants, and by 4 months, only *Bifidobacteriaceae*-affiliated OTUs were shared. By the age of 1 year, the majority of OTUs were affiliated to the *Clostridiales*, whereas at 2 years, shared *Bacteroidales*-affiliated OTUs also appeared.

Depending on ecological forces that structure communities, species within these communities may follow different distributions that can be described mathematically (Magurran, 2004). We therefore fitted OTU distributions to these common distribution curves (Table S9). The majority of samples fitted well to truncated log normal distribution, two samples, belonging to one child at 3 and 10 days of age, fitted log series distribution. The geometric and broken stick distributions did not fit the data. We also tested whether distributions fitted a neutral model and how much they deviate from it. All these samples showed higher dominance than it would be expected in case of neutrality (Fig. S2), although there was a significant difference in deviation between mothers and 3-day-olds ($P = 0.0091$). Moreover, when combined, in infancy as well as at 4 months, the dominance was significantly higher than in adults and in 1- and 2-year-olds ($P = 0.0001$).

Data consistency

To address whether MCR-ALS and pyrosequencing predictions of faecal microbiota correspond to each other, we selected all OTUs belonging to taxonomical groups predicted by MCR-ALS from a pyrosequencing data set. We then grouped those OTUs in correspondence with MCR-ALS components and calculated their relative amounts based on the total number of OTUs. Pearson's correlation analysis revealed high correlation between MCR-ALS predictions and pyrosequencing results (correlation coefficient = 0.7463, $P = 4.47 \times 10^{-51}$).

Discussion

Interestingly, there was a significant drop in interindividual beta-diversity in a short period of time after birth (3–10 days), as assessed by mixed sequencing. Due to practical reasons, many temporal research studies of faecal microbiota face a trade-off between sampling frequency and number of individuals included in the study. To our knowledge, all temporal faecal microbiota studies to date that have extensive sampling during first weeks of life (Favier *et al.*, 2003; Palmer *et al.*, 2007; Koenig *et al.*, 2011) have few individuals analysed, whereas studies with high sample numbers often have

fewer or more infrequent time points (Yatsunenko *et al.*, 2012). However, our results illustrate that significant differences in average bacterial composition and beta-diversity occur between 3 and 10 days of age. These data therefore suggest that to better understand the development of gut microbiota, gaps between sampling periods should be reduced, particularly for those studies that compare different populations (Yatsunenko *et al.*, 2012).

Pyrosequencing and mixed sequence analysis both demonstrated individualized clustering of the faecal microbiota during early and late pregnancy in our cohort, with little or no evidence for population-based changes during pregnancy. We were therefore not able to reproduce the results of a major change in the faecal microbiota between early and late pregnancy, as recently reported by Koren *et al.* (2012). Because our sampling times match that of Koren *et al.* with $\pm$ 3 weeks, we believe that sampling time cannot explain the differences in microbiota detected between the two studies. The most likely explanation would therefore be that there are true differences in the gut microbiota composition among pregnant women in the two cohorts.

qPCR analysis suggested a relatively low direct transmission of gut bacteria from mother to child; at 10 days of age, there was better overall agreement between detection of bacteria in mother–child pairs than at 3 days (Table 1). Even early colonizers such as *E. coli* were not likely to be directly transmitted at birth, but rather during first days of life (Fig. 4). The difference in detection of this species in mother–child positive subpopulation and the total population was higher at 10 days than at 3 days. Based on differences between weighted UniFrac (takes into account relative amounts) and Jaccard (takes into account only presence/absence data) distances, it may be suggested that by 1–2 years of age, adult characteristic OTUs already appeared in the gut, although they were still rare. Interestingly, many OTUs affiliated to *Lachnospiraceae* were shared between mothers and 1- to 2-year-old children, suggesting that these species possibly originate from the mother. PARAFAC data based on mixed sequencing also supported sharing of this component between mothers and infants. Even though detection of bifidobacteria seemed to be independent of the mother, frequency of *B. longum* was higher in a mother–child positive subpopulation, which is in line with a recent model suggesting transmittance of *B. longum* subsp. *longum* from mother to child (Makino *et al.*, 2011).

At 3 days of age, there was a relatively high abundance of *Lactobacillales* in stool samples (Fig. 2). Lactobacilli are often isolated from human breast milk (Martin *et al.*, 2003, 2007), and it was noted that the majority of infants (98%) in our cohort were exclusively breast-fed during the first 6 weeks of life. Interestingly, by the age of 10 days, the level of this bacterial group was observed to decline despite no changes in diet with respect to breast milk intake. As such, we hypothesize that lactobacilli detected in this study were possibly acquired via the vaginal microbiota of the mother during the infant's passage through the birth channel.

If we assume that neutral processes (i.e. random replacement of a dead individual in a community by an offspring of other species regardless of relative abundance of this species) are not involved in shaping gut microbiota, one would expect low individual alpha-diversity coinciding with high interindividual beta-diversity. In contrast, we observed steady decreases in beta-diversity over time (lowest among adult women), suggesting that overall microbiota development is ultimately directed towards a more stable community. Furthermore, delta values, which characterize a deviation from neutrality, were significantly lower in adulthood than in infancy.

In contrast to our findings, it has recently been argued that niche selection is also the main force shaping the distal gut community (Jeraldo *et al.*, 2012). This conclusion was based on the fact that microbial OTUs in the gut were more closely related to each other than what would be expected in case of neutrally shaped community. The discrepancy, however, could be explained by the fact that niche selection will always limit the phylotypes allowed in a given environment (Magurran, 2004) and that the distal gut represents a highly selective environment (Marchesi, 2011), whereas among the allowed phylotypes, neutral processes could be important. Probably, because we did not take phylogenetic distances into account, we also discovered the neutral processes as a potential contributor. This explanation is coherent with our recently proposed interface model for bacterial–host interactions, suggesting host selection independent of the actual services provided (Avershina & Rudi, 2013).

In conclusion, our analyses of a large longitudinal cohort of mothers and their children have revealed new knowledge about the ecology of human gut bacteria, suggesting that there are still important mechanisms that remain unknown.

Acknowledgements

Funding for the IMPACT study was obtained from GlaxoSmithKline AS, Norway. The PACT study was funded by the Norwegian Department of Health and Social affairs from 1997–2003. A university scholarship from NTNU funded the research fellows. The mixed sequencing analyses were funded by a research levy on certain agricultural products from the Norwegian Government. P.P. is funded by Norwegian Research Council project 214042. Authors have no conflict of interest to declare.

References

Avershina E & Rudi K (2013) Is it who you are or what you do that is important in the human gut? *Benef Microbes* **4**: 219–222.

Avershina E, Storro O, Oien T, Johnsen R, Wilson R, Egeland T & Rudi K (2013) Succession and correlation-networks of bifidobacteria in a large unselected cohort of mothers and their children. *Appl Environ Microbiol* **79**: 497–507.

Bjerke GA, Wilson R, Storro O, Oyen T, Johnsen R & Rudi K (2011) Mother-to-child transmission of and multiple-strain colonization by *Bacteroides fragilis* in a cohort of mothers and their children. *Appl Environ Microbiol* **77**: 8318–8324.

Bro R (1997) PARAFAC. Tutorial and applications. *Chemometr Intell Lab Syst* **38**: 149–171.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Cardillo G (2007) Cohen's kappa: compute the Cohen's kappa ratio on a 2 × 2 matrix. *MATLAB Central File Exchange*, MathWorks. Available at: http://www.mathworks.com/matlabcentral/fileexchange/15365.

Claesson MJ, Cusack S, O'Sullivan O *et al.* (2011) Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *P Natl Acad Sci USA* **108**(suppl 1): 4586–4591.

De Filippo C, Cavalieri D, Di Paola M *et al.* (2010) Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *P Natl Acad Sci USA* **107**: 14691–14696.

de Muinck EJ, Øien T, Storrø O, Johnsen R, Stenseth NC, Rønningen KS & Rudi K (2011) Diversity, transmission and persistence of *Escherichia coli* in a cohort of mothers and their infants. *Environ Microbiol Rep* **3**: 352–359.

Edgar RC, Haas BJ, Clemente JC, Quince C & Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194–2200.

Favier CF, de Vos WM & Akkermans AD (2003) Development of bacterial and bifidobacterial communities in feces of newborn babies. *Anaerobe* **9**: 219–229.

Ginzinger DG (2002) Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol* **30**: 503–512.

Jabot F & Chave J (2011) Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference. *Am Nat* **178**: E37–E47.

Jeraldo P, Sipos M, Chia N, *et al.* (2012) Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *P Natl Acad Sci USA* **109**: 9692–9698.

Jimenez E, Marin ML, Martin R *et al.* (2008) Is meconium from healthy newborns actually sterile? *Res Microbiol* **159**: 187–193.

Koenig JE, Spor A, Scalfone N *et al.* (2011) Succession of microbial consortia in the developing infant gut microbiome. *P Natl Acad Sci USA* **108**(suppl 1): 4578–4585.

Koren O, Goodrich JK, Cullender TC *et al.* (2012) Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150**: 470–480.

Ley RE, Peterson DA & Gordon JI (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**: 837–848.

Lozupone C & Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.

Magurran AE (2004) *Measuring Biological Diversity*. Blackwell Science Ltd, Oxford, UK.

Makino H, Kushiro A, Ishikawa E *et al.* (2011) Transmission of intestinal *Bifidobacterium longum* subsp. longum strains from mother to infant, determined by multilocus sequencing typing and amplified fragment length polymorphism. *Appl Environ Microbiol* **77**: 6788–6793.

Marchesi JR (2011) Human distal gut microbiome. *Environ Microbiol* **13**: 3088–3102.

Marques TM, Wall R, Ross RP, Fitzgerald GF, Ryan CA & Stanton C (2010) Programming infant gut microbiota: influence of dietary and environmental factors. *Curr Opin Biotechnol* **21**: 149–156.

Martin R, Langa S, Reviriego C *et al.* (2003) Human milk is a source of lactic acid bacteria for the infant gut. *J Pediatr* **143**: 754–758.

Martin R, Heilig HG, Zoetendal EG, Jimenez E, Fernandez L, Smidt H & Rodriguez JM (2007) Cultivation-independent assessment of the bacterial diversity of breast milk among healthy women. *Res Microbiol* **158**: 31–37.

Nadkarni MA, Martin FE, Jacques NA & Hunter N (2002) Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. *Microbiology* **148**: 257–266.

Oien T, Storro O & Johnsen R (2006) Intestinal microbiota and its effect on the immune system–a nested case-cohort study on prevention of atopy among small children in Trondheim: the IMPACT study. *Contemp Clin Trials* **27**: 389–395.

Palmer C, Bik EM, DiGiulio DB, Relman DA & Brown PO (2007) Development of the human infant intestinal microbiota. *PLoS Biol* **5**: e177.

Ronaghi M (2001) Pyrosequencing sheds light on DNA sequencing. *Genome Res* **11**: 3–11.

Rudi K, Moen B, Sekelja M, Frisli T & Lee MR (2012) An eight-year investigation of bovine livestock fecal microbiota. *Vet Microbiol* **160**: 369–377.

Satokari R, Gronroos T, Laitinen K, Salminen S & Isolauri E (2009) *Bifidobacterium* and *Lactobacillus* DNA in the human placenta. *Lett Appl Microbiol* **48**: 8–12.

Sekelja M, Rud I, Knutsen SH, Denstadli V, Westereng B, Naes T & Rudi K (2012) Abrupt temporal fluctuations in the chicken fecal microbiota are explained by its gastrointestinal origin. *Appl Environ Microbiol* **78**: 2941–2948.

Sim J & Wright CC (2005) The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther* **85**: 257–268.

Skanseng B, Kaldhusdal M & Rudi K (2006) Comparison of chicken gut colonisation by the pathogens *Campylobacter jejuni* and *Clostridium perfringens* by real-time quantitative PCR. *Mol Cell Probes* **20**: 269–279.

Storro O, Oien T, Dotterud CK, Jenssen JA & Johnsen R (2010) A primary health-care intervention on pre- and postnatal risk factor behavior to prevent childhood allergy. The Prevention of Allergy among Children in Trondheim (PACT) study. *BMC Public Health* **10**: 443.

Storro O, Oien T, Langsrud O, Rudi K, Dotterud C & Johnsen R (2011) Temporal variations in early gut microbial colonization are associated with allergen-specific immunoglobulin E but not atopic eczema at 2 years of age. *Clin Exp Allergy* **41**: 1545–1554.

Tauler R, Smilde A & Kowalski B (1995) Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution. *J Chemom* **9**: 31–58.

Wang Q, Garrity GM, Tiedje JM & Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–5267.

Whitman WB, Coleman DC & Wiebe WJ (1998) Prokaryotes: the unseen majority. *P Natl Acad Sci USA* **95**: 6578–6583.

Yatsunenko T, Rey FE, Manary MJ *et al.* (2012) Human gut microbiome viewed across age and geography. *Nature* **486**: 222–227.

Zimonja M, Rudi K, Trosvik P & Næs T (2008) Multivariate curve resolution of mixed bacterial DNA sequence spectra: identification and quantification of bacteria in undefined mixture samples. *J Chemom* **22**: 309–322.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. Relative bacterial amounts in a population of mothers and their infants as assessed by RT-PCR.
Fig. S2. Deviation from a fit to a neutral model as assessed by delta-value (Jabot & Chave, 2011).
Fig. S3. Summary of OTU-based analyses of pyrosequencing data.
Table S1. Number of mothers in who a given bacterium was detected during pregnancy by RT-PCR (detected before cycle 40 of the PCR reaction).
Table S2. RDP Database Project hierarchical classification (Wang *et al.*, 2007) of MCR-ALS resolved components.
Table S3. Number of shared OTUs between mothers and children.
Table S4. List of OTUs which were detected in mothers, but were absent in 3-days-old infants.
Table S5. List of OTUs which were detected in mothers, but were absent in 10-days-old infants.
Table S6. List of OTUs which were detected in mothers, but were absent in 4-month-old infants.
Table S7. List of OTUs which were detected in mothers, but were absent in 1-year-old infants.
Table S8. List of OTUs which were detected in mothers, but were absent in 2-year-old infants.
Table S9. Fit of OTU distributions to common distribution curves.

PAPER 2

Avershina E., Storrø O., Øien T., Johnsen R., Wilson R., Egeland T. & Rudi K. (2013) Bifidobacterial succession and correlation networks in a large unselected cohort of mothers and their children. *Applied and environmental microbiology* **79**: 497-507. DOI: 10.1128/AEM.02359-12

AEM
Journals.ASM.org

Bifidobacterial Succession and Correlation Networks in a Large Unselected Cohort of Mothers and Their Children

E. Avershina,[a,c] O. Storrø,[b] T. Øien,[b] R. Johnsen,[b] R. Wilson,[a] T. Egeland,[c] K. Rudi[a,c]

Faculty of Education and Natural Sciences, Hedmark University College, Hamar, Norway[a]; Department of Public Health and General Practice, Norwegian University of Science and Technology, Trondheim, Norway[b]; Department of Chemistry, Biotechnology and Food Science, University of Life Sciences, Ås, Norway[c]

Bifidobacteria are a major microbial component of infant gut microbiota, which is believed to promote health benefits for the host and stimulate maturation of the immune system. Despite their perceived importance, very little is known about the natural development of and possible correlations between bifidobacteria in human populations. To address this knowledge gap, we analyzed stool samples from a randomly selected healthy cohort of 87 infants and their mothers with >90% of vaginal delivery and nearly 100% breast-feeding at 4 months. Fecal material was sampled during pregnancy, at 3 and 10 days, at 4 months, and at 1 and 2 years after birth. Stool samples were predicted to be rich in the species *Bifidobacterium adolescentis*, *B. bifidum*, *B. dentium*, *B. breve*, and *B. longum*. Due to high variation, we did not identify a clear age-related structure at the individual level. Within the population as a whole, however, there were clear age-related successions. Negative correlations between the *B. longum* group and *B. adolescentis* were detected in adults and in 1- and 2-year-old children, whereas negative correlations between *B. longum* and *B. breve* were characteristic for newborns and 4-month-old infants. The highly structured age-related development of and correlation networks between bifidobacterial species during the first 2 years of life mirrors their different or competing nutritional requirements, which in turn may be associated with specific biological functions in the development of healthy gut.

Mother-to-child transmission and temporal development of the human gut microbiota are population-based processes. Understanding these processes is essential to the identification of gut microbiota-associated functionalities. Certain members of the genus *Bifidobacterium* represent very abundant early colonizers of the infant gut (1), making them a prime target for investigation.

The high abundance of *Bifidobacterium* species in infants is considered to promote development and maturation of the immune system to sustain health (2–4). Furthermore, our recent studies suggest that the succession of bifidobacteria is important for the proper immunological development (5, 6).

Eight *Bifidobacterium* species have been associated with the human gastrointestinal tract (GIT): *Bifidobacterium adolescentis*, *B. breve*, *B. longum* subsp. *longum*, *B. longum* subsp. *infantis*, *B. pseudolongum*, *B. bifidum*, *B. pseudocatenulanum*, and *B. dentium*. Some bifidobacterial strains, e.g., *B. pseudolongum*, appear to be exclusively associated with adult gut microbiota, and some, especially *B. longum* subsp. *infantis*, are typically isolated from infants (7). The population-wise, age-related development of bifidobacteria in infants remains to be investigated.

Due to the fact that there is evidence for previously uncharacterized diversity of bifidobacteria in the human gut (8, 9), we believe that to obtain a comprehensive description of bifidobacterial composition, the study should include a large number of individuals, and various techniques should be applied. It is also very important to avoid targeting specific bifidobacteria, since this may lead to the exclusion of as-yet-undiscovered, but potentially important bifidobacterial groups.

The aim of the present study, therefore, was to describe bifidobacterial composition, temporal development, and possible correlations in a large, randomly selected cohort of mothers and their children using a combination of both culture-dependent and -independent techniques. This was done by analyzing the series of stool samples from the IMPACT cohort (10). The samples were collected during early and late pregnancy stages and at 3 days, 10 days, 4 months, and 1 and 2 years after birth. We present results showing a highly structured, age-related succession of bifidobacterial species within the study population, as well as correlations between the abundances of these species during the first 2 years of life.

MATERIALS AND METHODS

The simplified workflow of the data analysis is represented in Fig. 1.

Study cohort. The PACT study (Prevention of Allergy among Children in Trondheim), a controlled nonrandomized prospective intervention study of pregnant women and their children up to 2 years of age, was started in 2000 (11). At the same time, a prospective observational substudy (IMPACT [Immunology and Microbiology in PACT]) of 720 pregnant women and their offspring, recruited from the PACT control cohort, was also initiated (10). All pregnant women in the control cohort were eligible for participation if they were willing and able to answer a questionnaire in Norwegian and supply biological material. We analyzed a randomly selected subset of the IMPACT study comprising stool samples of 87 infants and their mothers. At the point of inclusion participants were normally healthy children not screened for allergic disease hereditary diseases, allergy or atopy in the family, exposure factors, lifestyle, parental health, siblings health, smoking exposure, or any possible confounder. The samples were collected during the first or second trimester (8 to 20 gestation weeks) and the third trimester (30 to 40 gestation weeks) from mothers, and at 3 days, 10 days, 4 months, and 1 year and 2 years after

Received 27 July 2012 Accepted 30 October 2012
Published ahead of print 2 November 2012
Address correspondence to E. Avershina, ekaterina.avershina@umb.no.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.02359-12.

doi:10.1128/AEM.02359-12

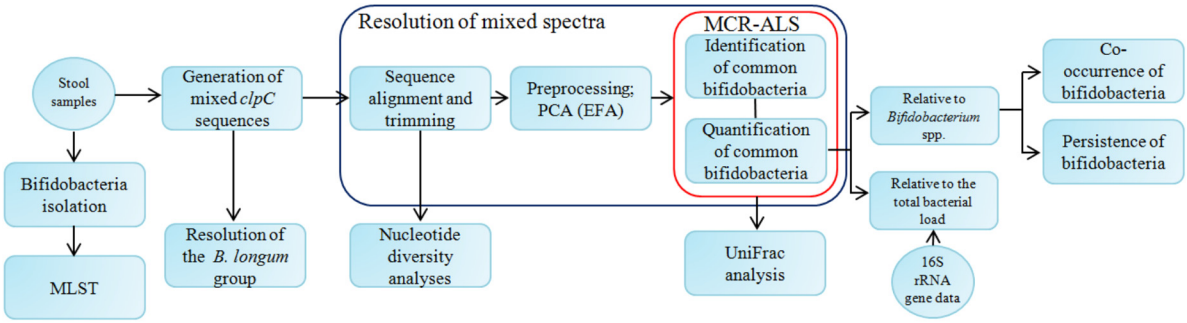


FIG 1 Flow diagram of the study. Boxes and circles represent processes and the data or materials obtained from them, respectively. At first, generated mixed spectra were trimmed to ensure that each spectrum brings an equal amount of information into the system. Then these spectra were analyzed using principal component analysis (PCA), evolving factor analysis (EFA), and multivariate curve resolution analysis with alternating least squares (MCR-ALS) to identify and quantify common bifidobacterial species in the data set. Quantification was performed relative to both the *Bifidobacterium* group and the total bacterial load. The *B. longum* group was resolved separately. Diversity was assessed using both taxon-based (UniFrac analysis based on MCR-ALS predictions) and taxon-independent (based on mixed spectra) techniques. The co-occurrence of various bifidobacteria, as well as their persistence over time, was also evaluated. Also, bifidobacteria were isolated from stool samples, and MLST analysis of several isolates was performed.

birth from children. Most children (~90%) were delivered vaginally. The information on infants' diet (Table 1) was received from a questionnaire filled out by parents at 2 years. More details about the IMPACT study cohort can be found in Storrø et al. (5).

In some cases, the stool sample from one or several collection time points was missing, whereas for some other individuals, more than one sample was present for the same age point (see Fig. S1A in the supplemental material). Moreover, several samples had sequence spectra of poor quality (many abnormally shaped fluorescence peaks resembling dye blobs or a low fluorescence signal) that would compromise the performance of multivariate curve resolution (MCR) analysis (see Fig. S2 in the supplemental material) and therefore had to be removed. In all, there were 330 samples in the final data set. For 52 children, bifidobacteria in the sample from the mother were also detected. The numbers of samples included in the study at each time point, as well as the numbers of mother-child pairs that contained data for all or just a few time points, are summarized in Fig. S1B and C in the supplemental material.

Isolation of DNA from stool samples. Fecal specimens were stored in sterile Cary Blair transport and holding medium (BD Diagnostics, Sparks, MD). Each specimen was frozen at −20°C no later than 2 h after sampling and transported to the laboratory for further storage at −80°C. Stool samples were first diluted 1:1 in solution 1 (50 mM glucose, 25 mM Tris-HCl [pH 8.0], 10 mM EDTA [pH 8.0]). The resulting suspension was then diluted 1:4 in 4 M guanidinium thiocyanate. The sample was then transferred to a sterile FastPrep-tube (Qbiogene, Inc., USA) with 250 mg of acid-washed glass beads (≤106 μm; Sigma-Aldrich, Germany), homogenized in FastPrep Instrument (Qbiogene) for 40 s and then centrifuged at 13,500 rpm for 5 min. Then, 170 μl of supernatant, together with 10 μl of Silica particles (Merck, Germany), was transferred to a 96-well Greiner U-plate (Greiner Bio-One, Germany), which was then placed in a Biomek 2000 Workstation (Beckman Coulter, USA). After the addition of Sarkosyl (1%), the plate was incubated at 65°C for 10 min, followed by 10 min at room temperature. The supernatant was discarded, and a bead pellet was washed twice with 50% ethanol. The beads were finally suspended in 100 μl of buffer C (1 mM EDTA [pH 8.0], 10 mM Tris-HCl [pH 8.0]), followed by incubation at 65°C for 30 min, and the solution with the eluted DNA was collected.

TABLE 1 Number of infants who were breast- and formula-fed and who received solid food at 6 weeks, 4 months, and 1 year of age

Diet (no. of infants)	% infants[a] receiving different diets		
	6 wk	4 mo	1 yr
Breast-feeding (84)	97.6	95.2	29.8
Formula feeding (84)	2.3	4.8	29.8
Combined (85)	10.6	21.2	55.3
Solids (87)	NA	13.8	98.9

[a] Data are based on the questionnaire filled out by parents at 2 years. The total number of infants in the study was 87. NA, not applicable.

Generation of mixed *clpC* sequences. Based on the evaluation of six housekeeping genes (12), we chose a mixed *Bifidobacterium clpC* (caseinolytic protease C) gene sequencing approach to obtain a comprehensive description of the bifidobacterial composition. *Bifidobacterium clpC* gene was amplified using *clpC*-specific primers developed by Ventura et al. (12). Sequencing of the forward strand was performed using BigDye Terminator v1.1 chemistry (Applied Biosystems, USA). The *B. longum* group-specific primer (5′-AGAAGCTGGAAGCCGAT-3′) was designed based on 34 *Bifidobacterium clpC* alleles downloaded from the *Bifidobacterium* multilocus sequence typing (MLST) database (13).

Mixed sequence analysis. The simplified scheme of mixed sequence analysis steps is represented in Fig. 2.

At first, mixed sequence spectra were aligned to small (~20 bp) start and end fragments flanking the variable region of the *clpC* gene and trimmed (Fig. 2A). To correct for retention shift differences and to increase the linearity of the data, all spectra were preprocessed after the alignment by using correlation-optimized warping (14, 15). Preprocessed aligned mixed sequences were then analyzed using a multivariate curve resolution (MCR-ALS) approach (Fig. 2B). This method allows simultaneous qualitative and quantitative identification of the components (in this case, groups of *Bifidobacterium* species), which are common to all of the samples in the data set of interest (16). The specific feature of the MCR-ALS method is that it particularly searches for components that are common for all of the samples in the data set, whereas other information is regarded as noise. Mathematically, the MCR-ALS method presents the initial experimental matrix of spectral data D ($m \times n$) as the combination of the concentration matrix C ($m \times i$) and the matrix of pure components S ($i \times n$), and a residual term E ($m \times n$):

$$D = CS + E \quad (1)$$

To define the initial number of components (initial estimates i), we used both principal component analysis (PCA) and evolving factor analysis (EFA) with nucleotide fluorescence intensities at each position being used as variables. PCA transforms the initial data into a new coordinate system, and the number of components is determined by the amount of variance explained (L. Smith, unpublished data). EFA on the other hand, repeatedly applies PCA to sections of the data set starting from the first two samples and subsequently adding the next one every run of the PCA

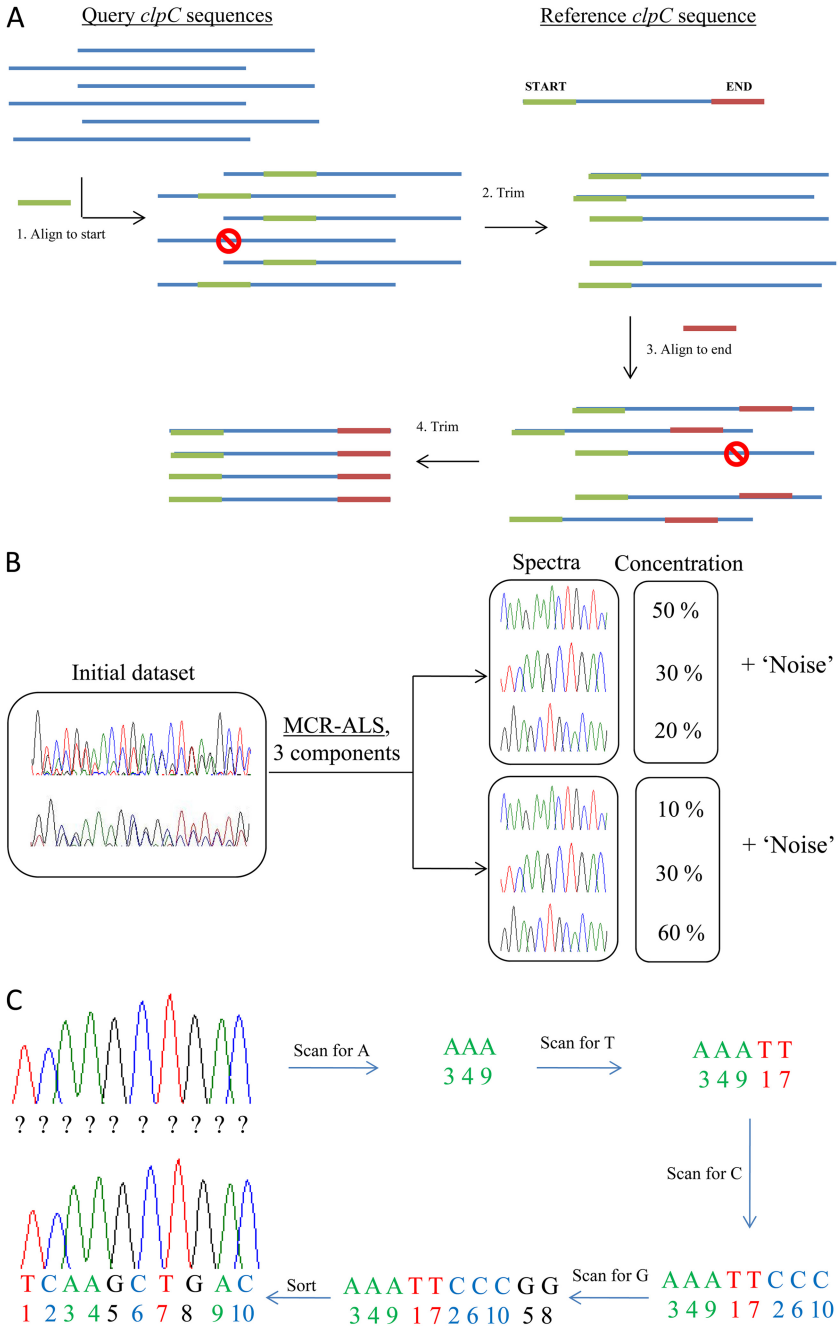


FIG 2 Simplified scheme of mixed sequence analysis. (A) Alignment of mixed sequences. All sequences have the same start and end, as well as a variable region in the middle. (B) MCR-ALS. Pure spectra common for all (the majority of) samples and their relative amounts in each sample are identified. (C) Base calling of pure MCR-ALS-resolved spectra. The nucleotide sequence of each of the pure resolved spectra is identified.

(17). In both cases, the number of suggested components, which is defined based on the drop in eigenvalue, is used as the initial number of components *i* for MCR-ALS.

Resolved spectra were then base called using an in-house-developed program that works in MATLAB, the same programming environment in which MCR-ALS is performed (Fig. 2C [the source code can be made available upon request]). Identification of resolved components was then performed using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) searches against the National Center for Biotechnology Information (NCBI) database.

Due to a high noise ratio in the T-channel, which could compromise results of MCR-ALS analysis, the information on T nucleotides was excluded from all spectra prior to resolution and then reintroduced at the base-calling stage. The validity of T-channel information removal was tested using predefined mixtures of bifidobacterial strains. MCR-ALS resolution of spectra without information on T nucleotides was comparable to that of spectra with information on all four nucleotides available (see Table S1 in the supplemental material).

All of the analyses of sequence spectra were performed using MATLAB R2010a software (The MathWorks, Inc., Natick, MA), Statistical and Bioinformatics toolboxes for MATLAB. For EFA, PCA, and MCR-ALS

analyses, PLS Toolbox v5.8 for MATLAB (Eigenvector Research, Inc., USA) was used.

Diversity measures. Nucleotide α-diversity was measured using a modified Simpson's index, calculated based on the fraction of each nucleotide's fluorescence intensity at every position, normalized for the length of the spectrum:

$$1/c_{\text{mixed}} = \frac{\sum_{i=1}^{n}(G_i)^2 + \sum_{i=1}^{n}(A_i)^2 + \sum_{i=1}^{n}(T_i)^2 + \sum_{i=1}^{n}(C_i)^2}{n} \quad (2)$$

The rationale is that in case of a pure sequence, there is only one nucleotide at each position (indexed by i), while in a very unified mixed sample, each position would contain all four nucleotides, and their intensity fractions would approach 0.25. To prevent misinterpretation based on the variation in length of the sequence (n), the value is then normalized. Furthermore, we consider the inverse relationship $1/c_{\text{mixed}}$ to support the more intuitive interpretation: the higher the c_{mixed} value, the more mixed the sample is (1 corresponds to pure, not mixed, samples, whereas a c_{mixed} value of 4 is the maximum and indicates a uniform mixture). In the case of bifidobacteria, however, the maximum of 4 will never be achieved due to high GC content. Although the information from the T-channel was noisy, we included these data into diversity measures analyses, since the fluorescence intensity at each position was normalized to 1, and abnormally high T-peaks would diminish the fraction of other nucleotides only in few positions.

To calculate individual diversity (i.e., the diversity within one individual), the raw mixed spectrum of each sample was used. The diversity within the population was calculated based on the average spectrum for each age group.

The β-diversity at various ages within individuals, as well as within the whole population (using the average spectra for each age group), was assessed using the modified Bray-Curtis dissimilarity index as follows:

$$BC_{ij} = \frac{\sum_{k=1}^{n}|G_{ki} - G_{kj}| + \sum_{k=1}^{n}|A_{ki} - A_{kj}| + \sum_{k=1}^{n}|T_{ki} - T_{kj}| + \sum_{k=1}^{n}|C_{ki} - C_{kj}|}{\sum_{k=1}^{n}(G_{ki} + G_{kj}) + \sum_{k=1}^{n}(A_{ki} + A_{kj}) + \sum_{k=1}^{n}(T_{ki} + T_{kj}) + \sum_{k=1}^{n}(C_{ki} + C_{kj})} \quad (3)$$

where N_{ki} and N_{kj} denote the fluorescence intensity fractions of a given nucleotide N at position k of the spectra belonging to ages i and j, respectively. In the extreme case when there is no single position with a shared nucleotide between the two sequences, the sum of the differences between the fluorescence intensity fractions divided by the sum of the fractions equals 1. When both sequences are the same, the similarity index equals 0.

For taxon-based assessment of diversity, we used UniFrac analysis (18) of species composition predicted by MCR-ALS. We also calculated Simpson's index of diversity and the Bray-Curtis dissimilarity index of MCR-ALS-predicted bifidobacterial composition.

Temporal development. The percentage and development in time were calculated relative to the *Bifidobacterium* group (MCR-predicted percentages). However, these percentages only reflect within-group composition, not taking into account the abundance of bifidobacteria as a whole compared to the other bacterial groups. In the previous study of 16S rRNA amplicons of IMPACT samples (see Text S1 in the supplemental material), one of MCR-ALS-resolved components was classified as *Bifidobacterium* genus (see Table S2 in the supplemental material). Therefore, using the information on relative abundance of bifidobacteria component in the samples (see Table S3 in the supplemental material), we also recalculated the percentages of *Bifidobacterium* species relative to total bacterial load. The significance of the change in the abundance between two subsequent time points was calculated with the Friedman test, which is a nonparametric version of two-way analysis of variance for repeated measurements (MATLAB documentation, 2010). The null hypothesis was rejected at the level of 5%.

Correlations in the amount of bifidobacteria. Pairwise comparisons between relative amounts of bifidobacteria species at each age group were performed using the Pearson correlation coefficient. To minimize false significant correlations, results were adjusted for multiple testing within each age category using the Bonferroni correction.

To test the correlations between the levels of every bacteria at two subsequent time points, all values were binarized. Values higher than mean abundance of a given bifidobacterial species at a particular point, were marked as "high" values, and smaller values were marked as "low" values. The co-occurrence of high and low values was tested using the Fisher exact test.

Cloning. To ensure the correct identification of *Bifidobacterium* spp. in the study cohort, we selected stool samples for the cloning of *clpC* amplicons using a TOPO TA cloning kit (Invitrogen, USA) according to the manufacturer's instructions. The plasmid inserts from all positively transformed *Escherichia coli* TOP10 colonies were then PCR amplified using primer pairs specific for the vector and then sequenced with the *clpC* primer following ExoI treatment.

Quantitative RT-PCR. The fractions of *B. adolescentis*, *B. longum*, *B. infantis*, *B. breve*, *B. bifidum*, and *B. dentium* in 11 selected stool samples relative to all *Bifidobacterium* spp. and to the total microbiota were quantified by real-time PCR (RT-PCR) with the double-stranded DNA-specific EvaGreen fluorescent dye (Solis BioDyne, Estonia) using 16S-23S ITS primer pairs specific for the named species and designed by Haarman and Knol (19). Each quantitative PCR (qPCR; 20 μl) contained 5× HOT FIREPol EvaGreen qPCR Mix (Solis BioDyne), 200 nM forward primer, 200 nM reverse primer, 1 μl of template DNA, and H_2O. Initial denaturation was performed at 95°C for 15 min, followed by 40 cycles of denaturation at 95°C for 15 s, primer annealing at 60°C for 20 s, and elongation at 72°C for 20 s. PCR efficiency for each reaction was calculated using a linear regression (20).

Isolation of bifidobacteria. Stool samples were diluted 10-, 100-, and 1,000-fold in 0.1% peptone water with the addition of L-cysteine (0.025%) and sodium thioglycolate (0.025%), followed by incubation at 37°C for 4 h to aid the recovery of injured cells (21). Although higher dilution rates are recommended, experiments are normally performed on fresh stool samples; however, in the present study, stool samples were kept at −80°C for up to 8 years. To create anaerobic conditions, Oxyrase for Broth (Oxyrase, United Kingdom) was added to each tube. After the initial incubation, diluted samples were streaked on *Bifidobacterium* agar plates (BD Diagnostics, USA) and further incubated anaerobically at 37°C for 72 h. After incubation, the plates were examined, and randomly selected colonies were inoculated into liquid MRS medium (Merck, Germany) and grown anaerobically at 37°C for 1 day.

Isolates were classified by sequencing of the *clpC* gene. We then selected 13 isolates representing all identified species for MLST analysis performed as described by Deletoile et al. (22).

Pyrosequencing data analysis. 16S rRNA amplicon pyrosequencing data for stool samples from seven randomly selected mother-children pairs were taken from a parallel study of IMPACT samples addressing the total microbial composition (see Text S1 in the supplemental material). Briefly, pyrosequencing data were processed using the QIIME pipeline (23). Sequences were filtered, and chimeras were removed using UCHIME algorithm (24) in addition to the ChimeraSlayer reference-based method. Next, sequences were clustered at the 97% similarity level, and the RDP classifier (25) was used to assign taxonomic identity to the resulting operational taxonomic units (OTUs). We only present results here based on clusters assigned to the *Bifidobacteriaceae* family. To obtain relative estimates of various bifidobacterial species detected by pyrosequencing, OTUs were searched against the NCBI database using BLAST. All of those assigned to a certain bifidobacterial species were grouped together, and their relative abundances were recalculated.

Ethical Committee approval. The Regional Committee for Medical Research Ethics for Central Norway approved the study (reference number 120-2000). The study was approved by the Norwegian Data Inspec-

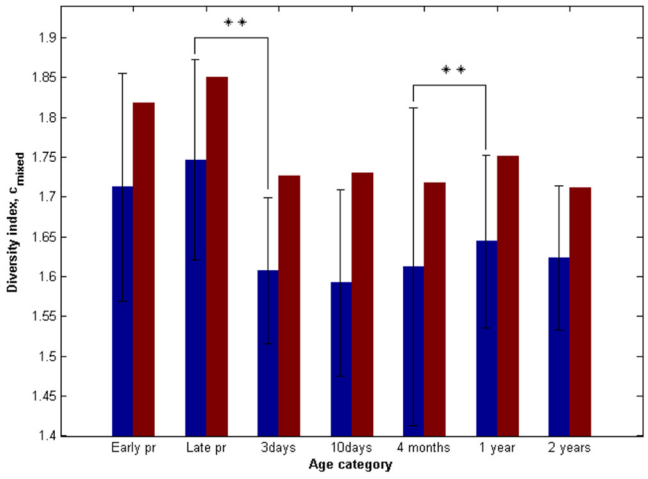


FIG 3 Modified Simpson's index of nucleotide spectra diversity c_{mixed} at various ages. Early pr and Late pr, early (8 to 20 weeks) and late (30 to 40 weeks) pregnancy periods, respectively. Blue bars represent the average nucleotide diversity for individuals with the standard deviation as shown, whereas red bars represent the diversity of the average nucleotide intensity spectra for each age category. The significance in difference between diversity indexes at two subsequent time points was calculated with the Friedman's test. **, $P < 0.01$.

torate, granting license to process personal health data, and one of the parents of each child signed a written informed consent form (reference number 2003/953-3 KBE/-). The Current Controlled Trials registration number is ISRCTN28090297.

RESULTS

Diversity determined directly from the mixed sequence spectra. Nucleotide diversity analysis of the raw mixed spectra (average length of 221 bp) by determination of modified Simpson's α-diversity index showed that the diversities both within individuals and within the population ranged from mean of 1.6 (standard deviation of 0.19) to 1.7 (0.14) and from 1.7 to 1.9, respectively, throughout the whole duration of the study, with 1 corresponding to pure samples and 4 corresponding to the most mixed spectra. The maximum diversity was observed in stool samples from adults (Fig. 3). Immediately after birth, the nucleotide diversity significantly decreased ($P = 0.0016$) and remained unchanged up to the age of 4 months. Interestingly, from 4 months to 1 year, the diversities both within individuals and within the study population increased. However, the diversity did not change from 1 to 2 years of age.

The Bray-Curtis β-diversity index revealed high variability of sequence spectra between subsequent time points within individuals (Fig. 4A). The highest similarity, in this case, was detected between the two pregnancy stages (0.08 [0.03]), whereas the highest dissimilarity was detected between mother-newborn pairs (0.14 [0.05]) and in the period from 4 months to 1 year (0.13 [0.05]). No significant difference between matching mother-newborn pairs and nonmatching mother-newborn pairs was detected. The similarity of the population average, however, was much higher, with the Bray-Curtis index ranging from 0.02 for the period between the two pregnancy stages up to 0.09 when comparing the average mixed spectra from the age of 3 days to that of the mothers in their late pregnancy stage. Clustering of the average mixed sequence spectra, however, suggests a clear age-dependent pattern of the bifidobacterial composition (Fig. 4B).

Resolution of mixed sequence spectra into species components. To define the initial number of components to be resolved by MCR-ALS from stool samples, we used both PCA and EFA. Although iterative methods, such as EFA, are commonly recommended for MCR-ALS analysis (16), the use of PCA also allows identifying outliers, which are best excluded from data sets prior to analysis. Both PCA and EFA analyses proposed six components

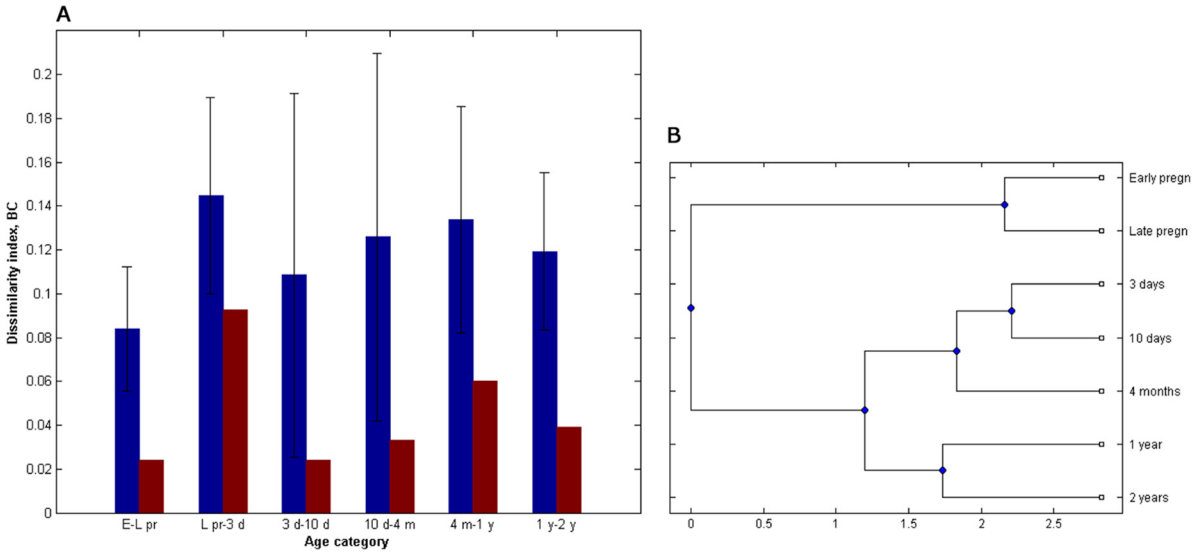


FIG 4 Comparison of mixed sequence spectra. (A) Modified Bray-Curtis index of nucleotide similarity (BC) between the subsequent time points. E-L pr, period between early (8 to 20 weeks) and late (30 to 40 weeks) pregnancy periods; L pr-3 d, comparison between 3-day-old newborns and their mothers during the late pregnancy stage; 3 d-10 d, comparison between 3 and 10 days of age; 10 d-4 m, comparison between 10 days and 4 months of age; 4 m-1 y, comparison between 4 months and 1 year of age; 1 y-2 y, comparison between 1 and 2 years of age. Blue bars represent the average similarity indices of nucleotide intensity spectra within every individual (with the standard deviation as shown), whereas red bars represent the similarity indices between the average nucleotide intensity spectra for each age category. (B) Clustering of the average nucleotide intensity spectra.

TABLE 2 Lowest E-value BLAST search hits of MCR-ALS-resolved spectra sequences against the NCBI nucleotide database[a]

MCR-ALS-resolved spectrum	Sequence length (bp)	Closest BLAST search hit	GenBank accession no.	E value	% identity
Component 1	221	*B. bifidum*	DQ206821.1	1×10^{-102}	98
Component 2	215	*B. dentium*	AY722387.1	3×10^{-109}	98
Component 3	219	*B. adolescentis*	DQ238016.1	3×10^{-107}	99
Component 4	216	*B. breve*	AB437352.1	4×10^{-101}	98
Component 5	221	*B. longum*	AP010889.1	4×10^{-111}	100

[a] Query coverage, 100%.

(see Fig. S3 in the supplemental material). However, when we performed MCR-ALS analysis with six components, one of them was poorly resolved, resulting in a longer nucleotide sequence spectrum with many mixed peaks which were hard to interpret (see Fig. S2 in the supplemental material). Therefore, we repeated the MCR-ALS analysis with five components. All five components were well resolved. The base-called components' spectra were identified by using the resulting sequence as a query in BLAST searches against the NCBI nucleotide database (Table 2). These components are further denoted by their closest matches in the NCBI database, which are, for components 1 to 5, *B. bifidum*, *B. dentium*, *B. adolescentis*, *B. breve*, and *B. longum*, respectively.

The nucleotide intensity of the residual, not extracted, spectra was on average three times lower than that of the resolved components. We then compared the residual spectral information against the NCBI database using BLAST. The analysis revealed the presence of such species as *B. pseudolongum*, *B. longum suis*, *B. magnum*, *B. mongoliense*, *B. scardovii*, and *B. asteroides*. However, in most cases, the hit length was low (~40 bp), indicating that these sequences might rather come from uncharacterized bifidobacteria. There were 33 samples, mostly belonging to 3- to 10-day-old newborns (see Fig. S1C in the supplemental material), which had comparatively high residual MCR component (the score was more than two times higher than the mean score of the residual component). In 25 of these samples, the residual spectra did not show any significant homology to known bifidobacterial species, whereas in the others, *B. adolescentis*, *B. breve*, *B. longum*, *B. cuniculi*, and *B. magnum* were detected. Empirical evaluation of raw spectra suggested that these residuals may reflect technical noise. Despite the noise, the MCR prediction of the most prevalent component in these samples corresponded to the lowest E-value BLAST search hit for raw unresolved sequences.

These results suggest that there were five most commonly identified bifidobacteria species in the study cohort. However, other species were also present and might have been present in high relative amounts in some individuals, but they were not shared among the majority of the samples.

Diversity analyses based on the resolved species components. The species composition based on MCR predictions was used as input in the taxon-based diversity analyses. In addition, we included the sequence information from the residual component representing a taxon. The Simpson's diversity index showed high correlation to that of the modified Simpson's for the direct analyses of the mixed sequence spectra (Pearson correlation coefficient = 0.49, $P = 1.35 \times 10^{-21}$). We also found high correlation between the Bray-Curtis dissimilarity index based on the resolved sequence components, and its modified version used for the analyses of the mixed sequence spectra (Pearson correlation coefficient = 0.47, $P = 1.34 \times 10^{-10}$). Furthermore, the UniFrac clustering supported the direct mixed sequence analyses, with a large diversity at the individual level (see Fig. S4 in the supplemental material).

Temporal changes in *Bifidobacterium* species composition. There was a large variation between individuals at every age, and the confidence levels for the mean varied from 0.6% up to 10 to 13% depending on the species and the age (see Table S4 in the supplemental material). Based on the analysis of total 16S rRNA gene content, bifidobacteria comprised 2% of total bacterial load in adults. In newborns, it constituted nearly one-fourth of the bacterial load, reaching the level of 60% by the age of 4 months (Fig. 5B). Relative to total bifidobacteria, stool samples from pregnant women were predicted to be rich in the *B. adolescentis*, *B. longum*, and *B. bifidum* group, whereas *B. dentium* was present in smaller amounts, and *B. breve* was nearly absent (Fig. 5A). The Friedman test revealed a significant decrease in the relative amount of *B. adolescentis* ($P = 0.0005$) and increase of *B. breve* ($P = 0.0348$) in stool samples of newborns compared to their mothers. The *B. longum* group represented the majority of bifidobacterial load at 3 and 10 days after the birth, whereas *B. breve* was the second most abundant species. Interestingly, *B. breve* separated all 10-day-old infants in two distinct groups. In one group, it accounted for <15%, whereas in the other it accounted for >75% of the bifidobacterial load. At 4 months of age, *B. breve* became the most predominant species in stool samples. However, by the age of 1 year, the percentage of *B. breve* decreased drastically (see Fig. S5A in the supplemental material), and *B. longum* regained its position as the most abundant group. Like adults, in children at 2 years of age, the majority of bifidobacterial load consisted of *B. adolescentis* and *B. longum*, although, unlike adults, *B. longum* was the most prevalent. *B. bifidum* comprised around one fourth of the bifidobacterial load in adults and 1- to 2-year-old infants, whereas in newborns and 4-month-old infants, its levels comprised to 9 to 15%.

Relative to the total microbial load, however, the amount of *B. adolescentis* comprised ca. 1 to 2.5% during the whole duration of the study (Fig. 5B), with the only significant difference in abundance occurring between newborns and mothers (see Fig. S5B in the supplemental material). There also was a significant fluctuation in the relative amount of *B. bifidum* from 10 days to 1 year of age, where it first increased up to 8% at 4 months ($P = 0.0093$) and then decreased to ca. 3% by the age of 1 year ($P = 0.0090$) and further down to 1.5%, still comprising twice as much of the bacterial load compared to adults (see Fig. S5B in the supplemental material). The most pronounced changes, however, were detected with regard to *B. breve* and *B. longum*. The relative amounts of *B. breve* were found to vary significantly between subsequent time points starting from 10 days on (see Fig. S5B in the supplemental material). *B. longum* was also found at significantly higher levels in

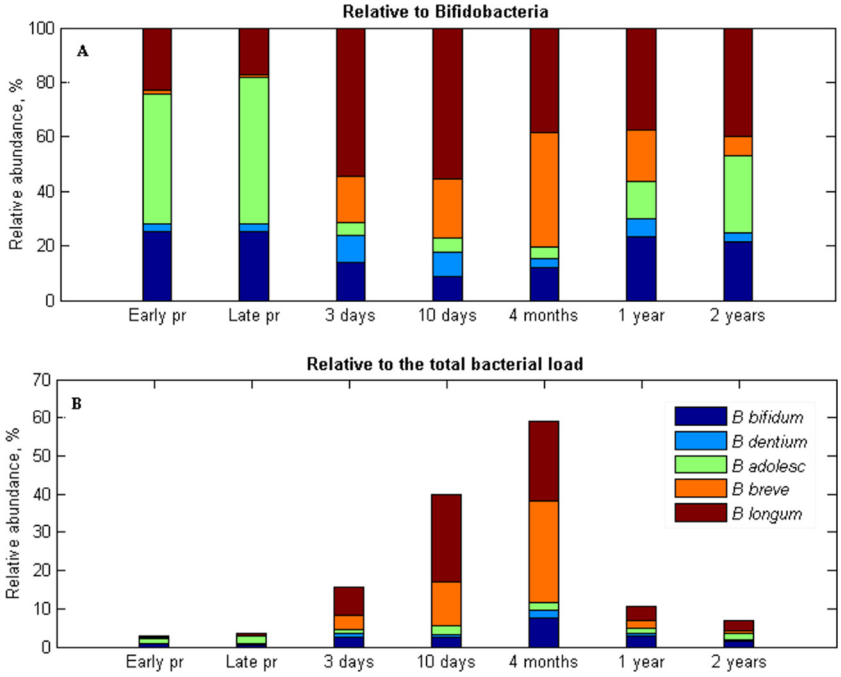


FIG 5 *Bifidobacterium* species composition in stool samples of infants (from 3 days to 2 years of age) and their mothers during pregnancy (pr) based on the results of MCR-ALS analysis relative to the bifidobacteria group (A) and relative to the total bacterial load (B).

10-day-old infants compared to those at 3 days of age ($P = 0.0067$). From 4 months to 1 year of age, however, it decreased significantly ($P = 0.0112$; see Fig. S5B in the supplemental material).

Co-occurrence of bifidobacterial species. Pairwise comparisons of the bifidobacterial co-occurrence revealed that in adulthood, as well as in 2-year-old children, there was a negative correlation between the amount of *B. adolescentis* and *B. longum* ($P = 2.27 \times 10^{-7}$, $P = 4.6 \times 10^{-5}$, and $P = 1.77 \times 10^{-10}$ for early pregnancy, late pregnancy, and 2 years, respectively; Fig. 6A). In newborns and 4-month-old infants, on the other hand, a negative correlation between *B. breve* and *B. longum* ($P = 8.37 \times 10^{-5}$, $P = 7.12 \times 10^{-7}$, and $P = 1.8 \times 10^{-17}$ for 3 days, 10 days, and 4 months, respectively) was detected (Fig. 6B). Interestingly, in 1-year-olds, negative correlations both between *B. adolescentis* and *B. longum* ($P = 0.007$) and between *B. longum* and *B. breve* ($P = 0.048$) was detected.

Correlations between the changes of bifidobacterial species

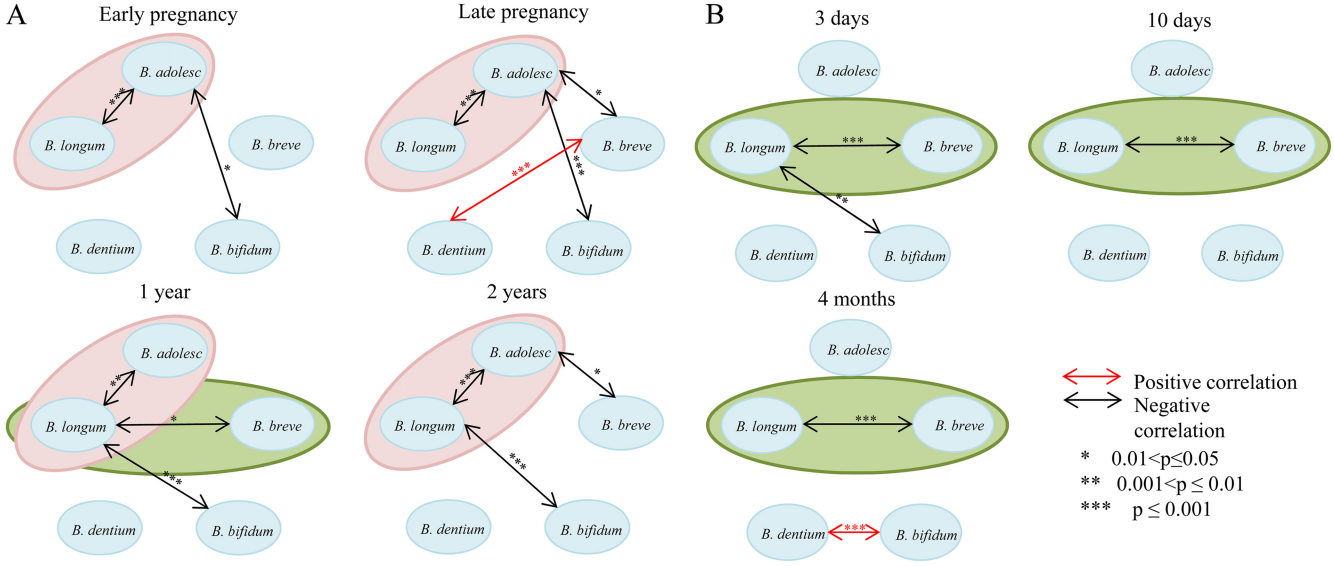


FIG 6 Co-occurrence of five dominant bifidobacterial species. Ellipses represent correlations, detected more than in two various ages. Pink ellipses, correlations detected between *B. adolescentis* and *B. longum* group; green ellipses, correlations detected between *B. longum* group and *B. breve*. (A) Mothers during early (8 to 20 weeks) and late (30 to 40 weeks) pregnancy stages, as well as in infants at 1 and 2 years of age. (B) Infants at 3 and 10 days and 4 months after the birth.

loads from one to another time point were also detected. As such, the change in the relative amount of *B. longum* from mother to newborn was negatively correlated to that of *B. breve*. Fluctuations between these two species loads were also negatively correlated from 3 to 10 days of age and further to 4 months, 1 year of age, and 2 years of age (see Fig. S6 in the supplemental material). The change in the relative amount of *B. longum* from 1 to 2 years was also negatively correlated to that of *B. adolescentis* (see Fig. S6 in the supplemental material).

Mode of delivery and diet effects. Due to high frequency of vaginal delivery, we were unable to test differences in bifidobacterial composition which might have been caused by mode of delivery. For the same reasons, the effect of breast-feeding versus formula feeding could not be investigated. The only factor we could test was solid food (rice, corn, and wheat) consumption at 4 months. Using two-sided permutation test with 10^6 permutations, we did not find any significant difference in relative amounts of bifidobacterial species ($P = 0.11$, $P = 0.74$, $P = 0.14$, $P = 0.84$, and $P = 0.85$ for *B. bifidum*, *B. dentium*, *B. adolescentis*, *B. breve*, and *B. longum*, respectively).

Resolution of *B. longum* group. Most of the nucleotide variation in *clpC* gene amplicons between *B. longum* subsp. *longum* and *B. longum* subsp. *infantis* is located downstream from the part of the *clpC* gene that was used for the MCR-ALS resolution. Therefore, to distinguish between *B. longum* subspecies, we designed a specific sequencing primer that binds closer to a variable site between *B. longum* subsp. *longum* and *B. longum* subsp. *infantis* (see Table S5 in the supplemental material). By resequencing *clpC* gene amplicons with this primer, *B. longum* was detected in 80% of all samples. We then binarized MCR-ALS predictions on the abundance of *B. longum* group in stool samples as present or absent and compared the data to resequencing results obtained with *B. longum* specific primer. In total, there was a high correlation between MCR-ALS predictions and resequencing data (according chi-squared analysis, $\chi^2 = 28.58$ and $P = 8.99 \times 10^{-8}$). Interestingly, all of the obtained sequences were pure. Sequences with a similar nucleotide variation pattern were aligned together, and the consensus sequence of the alignment was queried against the NCBI nucleotide database using BLAST. Four separate clusters were identified. The largest cluster belonged to *B. longum* subsp. *longum*, comprising 251 sequences. The second largest cluster comprised 12 sequences, belonging to *B. longum* subsp. *infantis*, whereas the third cluster, containing 5 sequences, comprised sequences that strongly resembled *B. longum* group but could not be assigned to either *B. longum* subsp. *longum* or *B. longum* subsp. *infantis*. Two sequences belonging to this cluster showed the closest homology to *B. longum* subsp. *suis*. One of the clusters, comprising 8 sequences, could not be assigned to any of the *B. longum* group species. In samples obtained from individuals both during pregnancy and right after birth, only *B. longum* subsp. *longum* was identified. Most of the *B. longum* subsp. *infantis* sequences were characteristic for children aged 4 months; there was only one baby, in whom it was detected earlier, and only two children where it persisted up to 1 year of age. By the age of 2 years, *B. longum* subsp. *longum* was found in 59 of 61 children in whom *B. longum* group was detected (see Fig. S7 in the supplemental material).

Validation of bifidobacterial species composition. To assess the ability of MCR-ALS analysis to resolve five-species mixtures, we verified the method using predefined mixtures of five bifidobacterial species (*B. bifidum* DSM20456, *B. dentium* DSM20436, *B. adolescentis* DSM20083, *B. breve* DSM20213, and *B. longum* subsp. *longum* DSM20219). In general, there was 83 to 97% correlation between actual mixture composition and that predicted by MCR-ALS (see Table S1 in the supplemental material).

To evaluate the MCR resolution for the stool sample, we selected 13 samples for the cloning-based analysis of bifidobacterial diversity. Sequencing of cloned inserts from these samples confirmed the presence of species predicted by MCR (see Table S6 in the supplemental material). All cloned inserts shared 98 to 99% identity to the corresponding reference sequences deposited in GenBank. In one sample, *B. animalis* was detected, and the analysis of the sample's residual spectral information, not resolved by MCR, confirmed the findings. Representative sequences of cloned inserts from all identified bifidobacterial species were deposited in GenBank (accession numbers JQ288967 to JQ288972).

We have previously estimated the relative amounts of *B. breve* and *B. longum* in the same stool samples by RT-PCR (5). When we compared MCR predictions for these species to those from obtained RT-PCR, there was a high correlation between the estimates ($\chi^2 = 149.32$ [$P = 2.44 \times 10^{-34}$] and $\chi^2 = 16.75$ [$P = 4.26 \times 10^{-5}$] for *B. breve* and *B. longum*, respectively). For the comparison, we binarized all of the data as high or low values based on the mean value and performed the chi-squared test. To verify MCR predictions with regard to all detected bifidobacterial species, we then analyzed 11 samples using RT-PCR with 16S-23S-ITS-region-targeting primers, specific for *B. bifidum*, *B. dentium*, *B. longum* subsp. *longum*, *B. longum* subsp. *infantis*, *B. breve*, and *B. adolescentis*. The relative composition revealed by RT-PCR amplification significantly correlated to that predicted by MCR-ALS (Pearson correlation coefficient = 0.76, $P = 8.0 \cdot 10^{-12}$; see Fig. S8 in the supplemental material). However, in four samples, the composition predicted by MCR was more diverse. In particular, *B. dentium* and *B. bifidum* were nearly absent according to RT-PCR. We then performed RT-PCR analysis on DNA isolated from pure bacterial cultures of *B. bifidum* (DSM20456) and *B. dentium* (DSM20436). Both of these species were detected only after 30 cycles, whereas with a universal bacterial primer pair they were detected on the 15th cycle. *In silico* PCR analysis (http://insilico.ehu.es/PCR/) of *B. bifidum* PRL2010, *B. bifidum* S17, and *B. dentium* Bd1 genome sequences with ITS-targeting primer pairs specific for the given species failed to produce any PCR product *in silico*. This may indicate that *B. bifidum* and *B. dentium* primer pairs target regions not universally conserved among all of the species strains.

We have previously analyzed stool samples from seven mother-child pairs using deep 454 sequencing of 16S rRNA amplicons (see Text S1 in the supplemental material). To compare MCR predictions to pyrosequencing findings, we extracted the information on bifidobacterial OTUs, grouped them according to the assigned bifidobacterial species, and calculated their relative abundances. In general, there was a significant correlation between MCR predictions and pyrosequencing data (Pearson correlation coefficient 0.60, $P = 2.11 \times 10^{-15}$; see Fig. S9 in the supplemental material). There was not a single OTU identified as *B. adolescentis*. We therefore reanalyzed the raw unfiltered data, and we found that sequences with high homology to *B. adolescentis* were removed from the final data set during the chimera filtering procedure. However, we were unable to introduce the information about these sequences to the final pyrosequencing data set.

We also isolated *Bifidobacterium* species from nine stool sam-

ples belonging to 4-month-old infants. Sequencing of *clpC* gene PCR products revealed that isolates belong to *B. longum* subsp. *infantis*, *B. longum* subsp. *longum*, *B. adolescentis*, *B. breve*, and *B. animalis* corresponding well with qualitative predictions of MCR analysis on expected bifidobacterial species. We then selected 13 isolates for MLST analysis (22). However, in addition to *clpC* gene, we managed to amplify only four genes (*fusA*, GTP-binding protein chain elongation factor EF-G; *ileS*, isoleucyl-tRNA synthetase; *rplB*, 50S ribosomal subunit protein L2; and *gyrB*, DNA gyrase, subunit B). Apart from the *gyrB* and *fusA* genes, which had very low resolution (see Fig. S13 and S14 in the supplemental material), clustering of sequences with all known profile sequences of a given gene corresponded to the expected *clpC*-gene based species delineations (see Fig. S10 to S12 in the supplemental material). Representative sequences of bifidobacterial isolates were deposited in GenBank (accession numbers JQ288937 to JQ288965).

DISCUSSION

Despite a major interest in bifidobacteria, there is still some inconsistency in the literature with regard to bifidobacterial loads in the human intestine. Estimates of bifidobacterial load in adults range from ca. 4% (26) up to 15% of the total gut microbiota (3). For infants, the range is higher. Some researchers suggest that this genus reaches up to 90% of the total microbiota (27, 28), whereas others claim much lower abundances (ca. 1 to 2%) of these species among infants (29). The inconsistencies in the literature may be due to both methodological differences in identifying the species (28, 30, 31) and the lack of precision among small study populations.

In our study, bifidobacteria comprised ca. 2% of the total bacterial load in adults, whereas the peak of its abundance was observed in 4-month-old infants, when this bacterial group constituted ca. 60% of intestinal microbiota. To our knowledge, our study with 87 mothers and their infants represents the largest comprehensive (including all bifidobacterial species) temporal study of natural development of bifidobacteria within a cohort thus far. Due to the large variation between individuals, and also taking into account the high frequency of missing data in longitudinal human study cohorts, we believe it is essential to analyze large cohorts.

Clustering of the population average mixed sequence spectra indicates that the system seems to be very structured, with significant differences between various ages. Analyses of mixed sequence spectra information at the individual level (UniFrac clustering), however, revealed only minor age-related patterns, apart from the cluster comprising the majority of stool samples from pregnant women and 2-year-old children. As illustrated both by Simpson's diversity index, the Bray-Curtis similarity index, and the information extracted from the residual spectra, there are large individual variations in the composition. This individuality probably brings the "noise" into the system and makes it difficult to deduce structured information when comparing every single individual with one another.

B. bifidum is predicted to possess lacto-*N*-biosidase and galacto-*N*-biosidase activity, enabling it to ferment human milk oligosaccharides (32). However, its relative abundance compared to other *Bifidobacterium* groups was higher in adults and 2-year-old children than in newborns and 4-month-old infants. On the other hand, relative to the total bacterial load, this group of bacteria exhibited peak abundance at 4 months, when most infants were breast-fed, and became nearly absent in 2-year-olds and adults. These observations suggest that *B. bifidum* is less affected by the total reduction of bifidobacteria with age. It has been demonstrated that *B. bifidum* is capable of utilizing host-derived glycans (33, 34), which we believe may partially explain the observed stability of this species. The levels of *B. dentium* remained nearly the same regardless of host age, also pointing to stability of this species. If *B. dentium* is commonly present in the mouth, then it would most likely be transferred constantly to the intestines with food and then excreted. The levels of *B. dentium* detected, however, suggest growth of this species in the gut lumen.

B. breve was nearly absent in adults, whereas in newborns and 4-month-old infants it was, on average, the second most abundant bifidobacterial group both relative to the bifidobacterial and total bacterial loads. Interestingly, with regard to the amount of *B. breve*, all 10-day-old infants were separated into two groups: (i) infants where *B. breve* constituted up to 15% of the bifidobacterial load, and (ii) infants where it accounted for >75% of the bifidobacterial load. *B. breve* was shown to be an efficient inducer of serum IgA (35), a main immunoglobulin found in mucous secretion of intestinal and respiratory tract (36), which averts penetration of pathogenic microorganisms by preventing their adsorption to mucosal epithelium. Therefore, in the future it would be very interesting to investigate whether such a huge gap in *B. breve* abundance has a host-related cause or can be explained by bacterial competition. At 4 months, the difference in *B. breve* abundance was smaller, since all of the infants had quite high numbers of this species. *B. breve* is regarded to be a very important species during the weaning period, since all of the tested strains possess the gene encoding amylopullulanase, the enzyme responsible for starch degradation (37). Because nearly half of infants at this age in the study cohort were given starch containing solid food, we could address the question of whether or not starch consumption would increase *B. breve* fraction. However, we did not detect a significant difference in the relative amount of this species between infants who were given solid food and those who were not. This may indicate that, apart from starch, there are probably other factors that promote *B. breve* at 4 months.

The human gut is a very densely populated site with interactions between numerous microorganisms (38), potentially leading to intricate correlation patterns. We believe that it is crucially important to identify correlation patterns between various bacteria, since this will very likely lead to a better understanding of disease development and health maintenance in individuals. In a recent study by Turroni et al. (28), the researchers noticed underlying interactions between bifidobacteria, although the number of individuals was too low to obtain more precise information. In our work, we could find two distinct patterns of correlation separating all seven ages into two major groups (Fig. 6). Adult-profile correlations comprised stool samples from adults, as well as from 1- and 2-year-old children. Infant-profile correlations comprised stool samples from newborns and 4-month-old infants. Negative correlations between the numbers of the *B. adolescentis* and *B. longum* groups were characteristic of the adult profile. At the same time, in pregnant women, there was also a negative correlation between *B. adolescentis* and *B. breve*, which only reappeared in children at the age of 2 years, also indicating that by the age of 1 year the development of *Bifidobacterium* species is not completely finished. For the infant profile, on the other hand, a negative cor-

relation between the *B. longum* group and *B. breve* was detected. Starting from 3 days of age and up to 4 months, in infants who had a low amount of *B. breve*, *B. longum* accounted for ca. 70% of the load, whereas in those who had high *B. breve* levels, *B. longum* was nearly absent.

Within both adult and infant profiles, *B. longum* comprised one of the central groups for which correlations were detected. This stability of the group regardless of the changes in the gut indicates the high competitiveness of *B. longum* and its vast ability to adjust to various conditions. All sequences obtained from stool samples with the specific *B. longum* group primer were pure. Both during pregnancy and right after the birth, only *B. longum* subsp. *longum* was identified in stool samples (see Fig. S7). This is in accordance with a recent model suggesting the transmittance of *B. longum* subsp. *longum* from mother to child (39). *B. longum* subsp. *infantis* was detected only in 12 samples, mostly belonging to 4-month-old infants, which might indicate horizontal transmittance of these bacteria. In the children who had *B. longum* subsp. *infantis* at 4 months of age, *B. longum* subsp. *longum* was detected earlier and later in life, suggesting a competition between the two subspecies. Genomic analysis of these two subspecies has revealed that *B. longum* subsp. *longum*, although able to ferment HMOs, is more suited for the fermentation of plant-derived oligosaccharides, whereas *B. longum* subsp. *infantis* harbors breakdown machinery for milk-derived oligosaccharides (40, 41).

In conclusion, our analyses of a large unselected cohort revealed structured age-related successions of bifidobacteria and correlations between the species of this genus within the population. We also showed that the *B. longum* group is probably one of the most central bifidobacteria in the human gut with respect to the correlation networks. In the future, it will be very interesting to determine whether the development patterns and correlations revealed here are also valid for populations in other geographical regions. At the same time, it will also be important to address the mechanisms of *Bifidobacterium* transmittance and the biological role of *Bifidobacterium* spp. in the human gut.

ACKNOWLEDGMENTS

This study was supported by Norwegian Department of Health and Social Affairs and Hedmark Sparebank.

We thank Eric de Muinck (University of Oslo, Oslo, Norway), Signe Marie Drømtorp (Nofima, Norway), and Monika Sekelja (Genetic Analysis AS, Norway) for their help with DNA isolation, sequencing, and optimization of the MATLAB script for sequence alignment.

REFERENCES

1. **Tannock GW.** 2010. Analysis of bifidobacterial populations in bowel ecology studies, p 1–15. *In* Mayo B, van Sinderen D (ed), Bifidobacteria: genomics and molecular aspects. Caister Academic Press, Norfolk, England.
2. **Gronlund MM, Gueimonde M, Laitinen K, Kociubinski G, Gronroos T, Salminen S, Isolauri E.** 2007. Maternal breast-milk and intestinal bifidobacteria guide the compositional development of the *Bifidobacterium* microbiota in infants at risk of allergic disease. Clin. Exp. Allergy **37**:1764–1772.
3. **Menard O, Butel MJ, Gaboriau-Routhiau V, Waligora-Dupriet AJ.** 2008. Gnotobiotic mouse immune response induced by *Bifidobacterium* sp. strains isolated from infants. Appl. Environ. Microbiol. **74**:660–666.
4. **Young SL, Simon MA, Baird MA, Tannock GW, Bibiloni R, Spencely K, Lane JM, Fitzharris P, Crane J, Town I, Addo-Yobo E, Murray CS, Woodcock A.** 2004. Bifidobacterial species differentially affect expression of cell surface markers and cytokines of dendritic cells harvested from cord blood. Clin. Diagn. Lab. Immunol. **11**:686–690.
5. **Storrø O, Øien T, Langsrud RK, Dotterud CK, Johnsen R.** 2011. Temporal variations in early gut microbial colonization are associated with allergen-specific immunoglobulin E but not atopic eczema at 2 years of age. Clin. Exp. Allergy **41**:1545–1554.
6. **Vebo HC, Sekelja M, Nestestog R, Storro O, Johnsen R, Oien T, Rudi K.** 2011. Temporal development of the infant gut microbiota in immunoglobulin E-sensitized and nonsensitized children determined by the GA-map infant array. Clin. Vaccine Immunol. **18**:1326–1335.
7. **Ventura M, Turroni F, Bottacini F, Giubellini V, van Sinderen D.** 2010. Bifidobacterial ecology and comparative genomics: perspectives and prospects, p 31–44. *In* Mayo B, van Sinderen D (ed), Bifidobacteria: genomics and molecular aspects. Caister Academic Press, Norfolk, England.
8. **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.
9. **Turroni F, Marchesi JR, Foroni E, Gueimonde M, Shanahan F, Margolles A, van Sinderen D, Ventura M.** 2009. Microbiomic analysis of the bifidobacterial population in the human distal gut. ISME J. **3**:745–751.
10. **Øien T, Storrø O, Johnsen R.** 2006. Intestinal microbiota and its effect on the immune system—a nested case-cohort study on prevention of atopy among small children in Trondheim: the IMPACT study. Contemp. Clin. Trials **27**:389–395.
11. **Storrø O, Øien T, Dotterud CK, Jenssen JA, Johnsen R.** 2010. A primary health-care intervention on pre- and postnatal risk factor behavior to prevent childhood allergy. BMC Public Health **10**:443. doi:10.1186/1471-2458-10-4433.
12. **Ventura M, Canchaya C, Del Casale A, Dellaglio F, Neviani E, Fitzgerald GF, van Sinderen D.** 2006. Analysis of bifidobacterial evolution using a multilocus approach. Int. J. Syst. Evol. Microbiol. **56**:2783–2792.
13. **Brisse, S.** 2009. *Bifidobacterium* MLST database: *Bifidobacterium clpC* alleles. Institut Pasteur, Paris, France. http://www.pasteur.fr/cgi-bin/genopole/PF8/mlstdbnet.pl?page=alleles&format=FASTA&locus=clpC&file=bifido_profiles.xml.
14. **Zimonja M, Rudi K, Trosvik P, Næs T.** 2008. Multivariate curve resolution of mixed bacterial DNA sequence spectra: identification and quantification of bacteria in undefined mixture samples. J. Chemometrics **22**:309–322.
15. **Tomasi G, van den Berg F, Andersson C.** 2004. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. J. Chemometrics **18**:231–241.
16. **Tauler R, Smilde A, Kowalski B.** 1995. Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution. J. Chemometrics **9**:31–58.
17. **Garrido M, Rius FX, Larrechi MS.** 2008. Multivariate curve resolution-alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. Anal. Bioanal. Chem. **390**:2059–2066.
18. **Lozupone C, Hamady M, Knight R.** 2006. UniFrac: an online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics **7**:371. doi:10.1186/1471-2105-7-371.
19. **Haarman M, Knol J.** 2005. Quantitative real-time PCR assays to identify and quantify fecal *Bifidobacterium* species in infants receiving a prebiotic infant formula. Appl. Environ. Microbiol. **71**:2318–2324.
20. **Ruijter JM, Ramakers C, Hoogaars WM, Karlen Y, Bakker O, van den Hoff MJ, Moorman AF.** 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res. **37**:e45. doi:10.1093/nar/gkp045.
21. **Nebra Y, Jofre J, Blanch AR.** 2002. The effect of reducing agents on the recovery of injured *Bifidobacterium* cells. J. Microbiol. Methods **49**:247–254.
22. **Deletoile A, Passet V, Aires J, Chambaud I, Butel MJ, Smokvina T, Brisse S.** 2010. Species delineation and clonal diversity in four *Bifidobacterium* species as revealed by multilocus sequencing. Res. Microbiol. **161**:82–90.
23. **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.

24. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**: 2194–2200.
25. **Wang Q, Garrity GM, Tiedje JM, Cole JR.** 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. **73**:5261–5267.
26. **Lay C, Rigottier-Gois L, Holmstrom K, Rajilic M, Vaughan EE, de Vos WM, Collins MD, Thiel R, Namsolleck P, Blaut M, Dore J.** 2005. Colonic microbiota signatures across five northern European countries. Appl. Environ. Microbiol. **71**:4153–4155.
27. **Harmsen HJ, Wildeboer-Veloo AC, Raangs GC, Wagendorp AA, Klijn N, Bindels JG, Welling GW.** 2000. Analysis of intestinal flora development in breast-fed and formula-fed infants by using molecular identification and detection methods. J. Pediatr. Gastroenterol. Nutr. **30**:61–67.
28. **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. doi:10.1371/journal.pone.0036957.
29. **Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO.** 2007. Development of the human infant intestinal microbiota. PLoS Biol. **5**:e177. doi:10.1371/journal.pbio.0050177.
30. **Sergeant MJ, Constantinidou C, Cogan T, Penn CW, Pallen MJ.** 2012. High-throughput sequencing of 16S rRNA gene amplicons: effects of extraction procedure, primer length, and annealing temperature. PLoS One **7**:e38094. doi:10.1371/journal.pone.0038094.
31. **Sim K, Cox MJ, Wopereis H, Martin R, Knol J, Li MS, Cookson WO, Moffatt MF, Kroll JS.** 2012. Improved detection of bifidobacteria with optimised 16S rRNA-gene based pyrosequencing. PLoS One **7**:e32543. doi:10.1371/journal.pone.0032543.
32. **Wada J, Ando T, Kiyohara M, Ashida H, Kitaoka M, Yamaguchi M, Kumagai H, Katayama T, Yamamoto K.** 2008. *Bifidobacterium bifidum* lacto-*N*-biosidase, a critical enzyme for the degradation of human milk oligosaccharides with a type 1 structure. Appl. Environ. Microbiol. **74**: 3996–4004.
33. **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.
34. **Turroni F, Milani C, van Sinderen D, Ventura M.** 2011. Genetic strategies for mucin metabolism in *Bifidobacterium bifidum* PRL2010: an example of possible human-microbe co-evolution. Gut Microbes **2**:183–189.
35. **Yasui H, Mike NNA, Hayakawa K, Ohwaki M.** 1992. Detection of *Bifidobacterium* strains that induce large quantities of IgA. Microb. Ecol. Health Dis. **5**:155–162.
36. **Wood P.** 2006. Understanding immunology, 2nd ed. Pearson Prentice Hall, Harlow, England.
37. **Ventura M, Canchaya C, Fitzgerald GF, Gupta RS, van Sinderen D.** 2007. Genomics as a means to understand bacterial phylogeny and ecological adaptation: the case of bifidobacteria. Antonie Van Leeuwenhoek **91**:351–372.
38. **Ley RE, Peterson DA, Gordon JI.** 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell **124**:837–848.
39. **Makino H, Kushiro A, Ishikawa E, Muylaert D, Kubota H, Sakai T, Oishi K, Martin R, Ben Amor K, Oozeer R, Knol J, Tanaka R.** 2011. Transmission of intestinal *Bifidobacterium longum* subsp. *longum* strains from mother to infant, determined by multilocus sequencing typing and amplified fragment length polymorphism. Appl. Environ. Microbiol. **77**: 6788–6793.
40. **LoCascio RG, Desai P, Sela DA, Weimer B, Mills DA.** 2010. Broad conservation of milk utilization genes in *Bifidobacterium longum* subsp. *infantis* as revealed by comparative genomic hybridization. Appl. Environ. Microbiol. **76**:7373–7381.
41. **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.
42. **Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O.** 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. **36**:W465–W469.
43. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum-parsimony methods. Mol. Biol. Evol. **28**:2731–2739.

PAPER 3

Avershina E., Lundgård K., Sekelja M., Dotterud C., Storrø O., Øien T., Johnsen R. & Rudi K. (2015) Transition from infant- to adult-like gut microbiota. (Submitted manuscript)

Running title: Gut microbiota transition

Transition from infant- to adult-like gut microbiota

E. Avershina,[1]* K. Lundgård,[1] M. Sekelja,[1,3] C. Dotterud,[2] O. Storrø,[2] T. Øien,[2] R. Johnsen[2] and K. Rudi[1]

[1] Department of Chemistry, Biotechnology and Food Science, University of Life Sciences, 1430 Ås, Norway;
[2]Department of Public Health and General Practice, Norwegian University of Science and Technology, 7491 Trondheim, Norway;
[3]Genetic Analysis AS, 0401 Oslo, Norway;

*Corresponding author information:

email: ekaterina.avershina@nmbu.no;

phone: 004767232447;

fax: 004764965001

Abstract

Transition from an infant to an adult associated gut microbiota with age through establishment of strict anaerobic bacteria remains one of the key unresolved questions in gut microbial ecology. Here we present comprehensive comparative analysis of stool microbiota in a large cohort of mothers and their children using sequencing analysis tool that allows realistic microbial diversity estimates. We provide evidence for the switch from children to adult associated microbial profile between 1 and 2 years of age, suggestively driven by a keystone bifidobacterial Operational Taxonomic Unit (OTU). This OTU was highly prevalent in the population throughout the first year of life, and was negatively associated with detection of a range of adult-like OTUs. Although an adult profile was not fully established, we demonstrate that with regards to the most prevalent pool of OTUs, their prevalence in the child population by 2 years of age already resembled that of the adult population. We also show a decrease in direct mother-child associated OTUs with age despite the increase in the pool of mother-child shared OTUs. Taken together we propose a model of OTUs recruitment from a common adult population pool, which is controlled by gatekeeping OTUs until the age of 1 year.

Introduction

In adulthood, our gut is densely colonized by 150-200 strictly anaerobic bacterial species (Qin et al., 2012; Faith et al., 2013) which have an intricate and not yet fully characterized crosstalk with our bodies (Clemente et al., 2012; Mulle et al., 2013; Antunes et al., 2014). Most likely, the colonization process starts as early as during prenatal development, as numerous gut bacteria were found in placentas (Satokari et al., 2009; Aagaard et al., 2014), umbilical cords (Jimenez et al., 2005), amniotic fluid (DiGiulio et al., 2008; DiGiulio, 2012) and meconium (Jimenez et al., 2008). At birth, a child is exposed to bacteria when passing through the mother's birth canal. Stool samples of newborns resemble the mother's vaginal swabs, whereas stool samples of babies born by Caesarean section resemble skin microbiota (Dominguez-Bello et al., 2010). At the start of feeding, a mother's breast milk becomes an important vector of selected gut bacteria (Martin et al., 2012; Jeurink et al., 2013; Schanche et al., 2015) and a driving factor of microbial community structure (Backhed et al., 2015). As we grow, environmental and lifestyle factors such as family size, rural or urban area, our diet, etc. also play a role in gut microbial structure (Sjogren et al., 2009; Hehemann et al., 2010; Madan et al., 2012; Yatsunenko et al., 2012; Song et al., 2013).

A key yet unresolved question is how strictly anaerobic bacteria are established in the transition from an infant- to an adult-like gut microbiota. There are two general hypotheses for the recruitment of the adult associated strictly anaerobic species: they are either vertically passed on early in life and kept as a very low abundant population until conditions become right for them, or they are acquired later in life though spores or other states that support survival in the environment. In support of a later-life colonization, it has recently been shown that 17 spore-forming *Lachnospiraceae* isolates from a healthy human individual successfully restored butyrate production and immune regulation in mice, protecting them from experimental colitis (Atarashi et al., 2013). We have also previously observed ecological indications for later-life recruitment of bacteria (Avershina et al., 2014).

Until recently, all operational taxonomic units (OTUs) clustering algorithms introduced a high level of noise in the analyzed datasets leading to up to 5 to 10 fold species overestimation and impeding assessment of the proportion of mother-child shared microbiota (Avershina and Rudi, 2015). Therefore, previous investigations of microbial recruitment in children may have yielded biased results. Recently, however, new approaches that allow more accurate biological prediction have been developed (Edgar, 2013; Faith et al., 2013). This opens up the possibility of developing more realistic models of OTUs recruitment with age based directly on the sequencing data.

In this work we aimed at addressing the question of establishment of gut microbiota with age in a large, geographically restricted probiotic intervention cohort (Dotterud et al., 2010; Dotterud et al., 2015) using deep 16S rRNA sequencing with a particular focus on noise reduction approach (Edgar, 2013). To the best of our knowledge, this study is the largest longitudinal study that addresses establishment of infant gut microbiota using improved quality filtering algorithms.

We present results supporting random recruitment of OTUs from a common pool of bacteria which is restricted by a key *Bifidobacterium* OTU. Furthermore, we demonstrate that mother-child associated OTUs gradually decreased with age despite an increase in a shared pool of OTUs. Finally, we show that adult-associated OTUs were not fully established by the age of 2 years and that probiotic intervention did not significantly alter microbial community structure.

Materials and Methods

Study cohort

Probiotics in Prevention of Allergy among Children of Trondheim (ProPACT) was initiated in 2003 as a randomized double-blinded intervention cohort study. In total, 415 pregnant women were recruited for this study. Women received low-fat fermented milk (study group) or heat-

treated sterile skimmed milk (placebo control group) daily starting from the 36th gestation week until 3 months post-partum. The majority of children were born at term and breastfed during at least three months of age; detailed information about the study is given in Dotterud and colleagues (Dotterud et al., 2010).

Stool samples from mothers were collected at the third trimester and 3 months post-partum. Stool samples from children were collected 10 days, 4 months, 1 year and 2 years after birth. Samples were frozen at -20 ºC no later than 2 hours after defecation and then delivered to the laboratory for long-term storage at -80 ºC.

The study was approved by the Regional Committee for Medical Research Ethics for Central Norway (Ref. 097-03) and the Norwegian Data Inspectorate (Ref. 2003/953-3 KBE/). The trial was registered in ClinicalTrials.gov (identifier NCT00159523).

PCR amplification

Stool samples were homogenized by bead-beating and then DNA was isolated using a LGC Mag™ DNA extraction kit (LGC Genomics, UK) following the manufacturer's instructions. The microbial community was characterized by deep sequencing of 16S rRNA gene using primer pair targeting prokaryotic organisms (PRK341F and PRK806R) (Yu et al., 2005). Primers were modified with Illumina adapters and combined in 576 different ways (16 forward and 36 reverse tag variations). PCR amplification was performed in two steps in order to ensure amplification with long and potentially more troublesome-to-anneal Illumina-modified primers. At first, we performed 25 cycles of denaturing at 95 ºC for 30 sec, annealing at 55 ºC for 30 sec and elongation at 72 ºC for 1 min with PRK primers. Then we purified the resulting amplicons with AMPure XP (Beckman-Coulter, USA) following the manufacturers' recommendations, and further performed 10 cycles denaturing at 95 ºC for 30 sec, annealing at 50 ºC for 1 min, and elongation at 72 ºC for 1 min with Illumina-modified PRK primers.

Sequencing library preparation

Due to the high number of samples (n = 1589), DNA sequencing was performed in three runs. All runs were processed identically as described below.

To ensure equal amounts of each sample in the sequencing pools, we quantified PCR amplicons using quantitative real-time PCR (LightCycler 480, Roche, USA) with TaqMan® probe targeting 16S rRNA gene (Yu et al., 2005) and Illumina colony amplification primers. For each plate, standard curves were calculated based on triplicate measurements of 10-fold standard dilutions. Samples were then pooled in equimolar amounts and the final pools were further quantified using PerfeCta® NGS library quantification kit (Quanta Biosciences™, USA). Normalized pools were sequenced on MiSeq platform (Illumina, USA) using V3 sequencing chemistry with 300 base pairs paired-end reads. PhiX DNA (15 % vol.) was added to each sequencing run in order to increase nucleotide heterogeneity during the first cycles.

Sequence data analysis

Resulting data were analyzed using standard workflow of QIIME pipeline (Caporaso et al., 2010). Sequences were paired-end joined (*fastq-join*), quality filtered (average sequence quality score more than 25; not more than 6 ambiguous base pairs, no mismatch in barcode allowed), and assigned to respective samples. Then, all sequences from three separate sequencing batches were combined and clustered with 97 % identity level using closed-reference search (*usearch* (Edgar, 2010) v7 which implements error-minimizing *uparse* algorithm (Edgar, 2013)) against Greengenes database (DeSantis et al., 2006) v 13.8.

To ensure a large number of samples, as well as adequate sequencing depth for the diversity coverage (Supplementary Fig. 1), we chose 2000 sequences per sample as a cutoff for the normalized dataset, though we also used a full dataset to search for low-abundant OTU populations.

The significance of differences between various groups was calculated using the Kruskal-Wallis test. The significance of the change in the abundance between two subsequent time points was calculated with the Friedman test. P-values were corrected for multiple testing using FDR correction. The null hypothesis was rejected at the level of 5 %. All statistical analyses were performed using MATLAB® 2014a software (MathWorks Inc., USA). Deviation of microbial assembly from neutrality was tested using Parthy software (Jabot and Chave, 2011). OTU distributions were fitted using Species Diversity and Richness IV (Pisces Conservation Ltd., UK).

Results

Sequencing dataset characteristics

A complete sequencing dataset comprised 21 158 715 sequences from 1589 samples with a length of 445 basepairs. An average number of sequences per sample was 13 315 (ranging from 9 to 169 024 sequences per sample). The normalized dataset (2000 sequences per sample) comprised 1125 samples belonging to 287 mother-child pairs. Population gut microbial community contained 1315 OTUs belonging to 26 bacterial classes.

Three control groups of gut-associated bacteria, with three replicates in each, were included to assess sequencing reproducibility: low complexity group (10 species), medium complexity group (31 species) and high complexity group (87 species). Ninety-seven percent of species were recovered by sequencing (Supplementary Table 1). Triplicates showed high reproducibility (Pearson's $c = 0.92$ (0.05); 0.78 (0.14); 0.95 (0.01) and $p < 0.01$ for low, medium and high complexity groups respectively). About 6 % of OTUs per control group were assigned to bacterial groups that were not included, collectively representing an average of 5 % of sequencing information per sample.

General characteristics of the microbiota

In line with previous observations (Dotterud et al., 2015), we found no significant difference in gut microbial community structure with respect to maternal probiotic intake (Supplementary Tables 2 and 3). Therefore, we report results on a full dataset regardless probiotic intake.

On average, in adulthood about 130 (50) [mean (standard deviation)] OTUs per individual were detected (Figure 1A). Concordant with a recent study by DiGuilio and colleagues (DiGiulio et al., 2015), no significant differences in the number of OTUs detected within each class during and after pregnancy were observed (Table 1). Immediately after birth, the number of detected OTUs was 6 times less than that of adults, and significantly increased during the entire period of the study ($p < 0.0001$). By the age of 2 years, species richness reached nearly two-thirds of that of adults, though it was still significantly different from it ($p = 1.5*10^{-87}$). All of the Actinobacteria OTUs seem to have appeared by the age of 3 months. At 2 years of age, however, they were still overrepresented as compared to adults (Table 1). On the other hand, we detected a significant increase in the variety of Clostridia and Bacteroidia OTUs from 3 months to 1 year and further to 2 years of age, though both still did not reach adult richness (Figure 1B).

Abundance-wise, 10 of the most abundant classes comprised nearly 100 % of the gut community at all times (Figure 1C). During pregnancy, abundance of Bacilli slightly albeit significantly increased compared to 3 months post-partum (1.5 % and 1.1 % at pregnancy and 3 months post-partum respectively; $p = 0.01$). We did not find indications for stratification of mothers' stool samples into enterotypes (Arumugam et al., 2011) (Supplementary Fig. 2). By the age of 2 years, gut microbiota community resembled that of adults, though with apparent differences between the two (Supplementary Table 4).

We randomly selected 10 individuals from each age group to test for neutrality deviation and rank abundance distributions. Rank abundance plots tended to be shallower with age (Supplementary Fig. 3), indicating the development of a highly dominated community into a

more evenly distributed assemblage (Magurran, 2004). Stool samples from 10 days, 1 year and 2 years fitted both truncated log-normal and log-series distribution, whereas stool samples from 3-month-olds and mothers tended to fit log normal distribution (Supplementary Table 5). At all study periods, we observed a greater chance of abundant OTUs to proliferate in the gut than what would be expected in case of neutral assembly (Supplementary Table 6). We did not detect significant differences in neutrality deviation between subsequent study periods or between 2-year-olds and their mothers. Collectively, though, stool microbial communities at 10 days and 3 months of age deviated from neutrality to a larger extent than at 1 year and 2 years of age and in mothers (deviation from neutrality $\delta = 0.67$ (0.15) and 0.53 (0.19) for 10 days and 3 months and for 1 year, 2 years and mothers respectively; $p = 0.01$).

Sharing of OTUs within the mother population

Both with regards to OTUs detection and OTU abundance profiles, stool samples from mothers were more similar intra- than inter-individually ($p < 1*10^{-10}$; Supplementary Fig. 4).

On average, each mother shared about 50 OTUs with other adults (Figure 2A). Most of these were highly abundant in the population, whereas individual OTUs belonged to low-abundant representatives of the microbial community (Figure 2B). Out of those OTUs that were shared among at least 50 % of the adult population, 34 OTUs persistently colonized over 50 % of the mothers from pregnancy to 3 months after birth (Supplementary Table 7). The majority of these dominant OTUs belonged to the Clostridia class (26 OTUs); other bacterial classes included Bacteroidia (4 OTUs), Actinobacteria (2 OTUs), Bacilli (1 OTU) and Erysipelotrichia (1 OTU).

Sharing of OTUs between mothers and children

The number of OTUs that were shared between mothers and their children increased with age (Figure 3A), and at all times was higher than what would be expected by chance ($p < 1*10^{-35}$; Supplementary Fig. 5), but not significantly different from the number of OTUs shared between

non-related mothers and children (Supplementary Fig. 6), or between children (Supplementary Fig. 7).

As compared to children-associated OTUs, OTUs shared between mothers and children were relatively enriched in Bacteroidia and depleted of Gammaproteobacteria and Erysipelotrichia (Figure 3B). In the first days of life, Actinobacteria were over-represented and Clostridia were under-represented in a mother-child shared group of OTUs as compared to children-associated OTUs, whereas at 2 years of age, the opposite trend was observed.

Early detection of OTUs in the child population positively correlated to prevalence of these OTUs in the mother population during the first year of life (Table 2; $p < 0.005$). As such, 33 out of 34 dominant OTUs appeared in the child population as early as at 10 days of age (Supplementary Table 8). We also observed a positive correlation between prevalence of the appearing OTUs in a population of children and prevalence of these OTUs in their mothers at all times apart from 3 months of age (Spearman $r = 0.4159$; 0.2427; 0.5151 and $p < 0.01$ for 10 days, 1 year and 2 years respectively).

Recruitment of OTUs with age

In order to identify OTUs that are more likely to be recruited from mothers, we selected a set of OTUs that showed significant chi-squared correlation ($p < 0.05$) between the detection of a given OTU in a mother and her child. For each OTU of this set, we then selected those OTUs for which prevalence in mothers of OTU-positive children is significantly higher than that in a general mother population. In total, we identified 29 OTUs that were significantly associated with mothers, with the majority being classified as Bacteroidia (13 OTUs) and Clostridia (8 OTUs) (Table 3). On average, after 3 months of age the average number of OTUs that were significantly associated with mothers decreased slightly albeit significantly ($p < 0.05$), despite the increase in the number of mother/child shared OTUs (Figure 3C).

We then extracted OTUs that were detected only in the mother population in a normalized dataset (351 OTUs), and searched for these OTUs in children samples with a deeper sequencing level (from 2 001 up to 169 024 sequences per sample). Most of these OTUs (70.1 %) were not detected in children at any time point, whereas those that were detected with deeper sequencing occurred at very low levels (2 ± 1 sequence per OTU per sample regardless sequencing depth), and only at one study time point for the absolute majority of cases (Supplementary Table 9). Most of the detected OTUs (67 out of 102) belonged to Clostridia.

Switch in population-prevalent OTUs with age

We observed a significant increase in inter-individual similarity from 1 to 2 years of age with regards to OTUs detection profiles ($p = 5.6*10^{-6}$; Supplementary Fig. 8).

The 10 most prevalent OTUs (hereafter referred to as the top OTUs) in mothers were nearly absent in children during the first months of life, whereas children's top OTUs in that period were rarely found in the mother population (Figure 4A). At 1 year of age, the prevalence of mothers' top OTUs in children increased as compared to earlier time points, but children's top OTUs were less represented in the mother population. The top OTUs of the 2-year-old population, though, reached the same prevalence as in the mother population (Figure 4A). Also, at 2 years of age, there was a shift in prevalent OTUs with 6 OTUs disappearing from a top 10 list. Three of these OTUs (classified as *Bifidobacterium*; *Bifidobacterium longum* and Enterobacteriaceae) dominated the child population throughout the entire first year of life.

We then addressed whether detection of these three OTUs correlated to the detection of Clostridia or Bacteroidia OTUs at 1 and 2 years of age since these two groups significantly increased in variety at this period. Prevalence of 19 Clostridia OTUs and 1 Bacteroidia OTU showed significant correlation to detection of the three top OTUs (Supplementary Table 10), with 17 of Clostridia OTUs correlating to detection of the *Bifidobacterium* OTU (OTU2) at 2 years of age (Figure 4B). These clostridial OTUs clearly separated into two groups: OTUs that

were detected more frequently in a population by 2 years of age had a co-exclusion pattern with OTU2, whereas those that were more frequently observed at 1 years of age co-occurred with this top OTU. The detection of OTU2 exhibited a delayed response to breastfeeding cessation (OTU2 was detected in about 50 % (10 days, 2 years) and 70 % (3 months, 1 year) of the child population), whereas its abundance mirrored the drop in a number of breastfed children with age decreasing from 12.6 % (26.0 %) at 3 months to 3.3 % (8.5 %) at 1 year of age. In mothers, this OTU was detected only in 6.8 % of the population with an average abundance of 0.01 % (0.05 %).

Phylogenetic relatedness between OTU2 and other bifidobacteria

A BLAST search of the NCBI database returned *Bifidobacterium longum* as a closest match to OTU2, though with 95 % identity due to 3 trinucleotide deletions in the V3 region of OTU2. Recently we isolated and genome sequenced 19 bifidobacteria strains from the nested cohort of this study (unpublished). We found the same trinucleotide deletions in 16S rRNA genes of two *B. breve* and one *B. adolescentis*, but not in *B. longum*. However, *B. breve* and *B. adolescentis* isolates exhibited only 95.6 % and 94 % similarity to OTU2 respectively. We were therefore not able to classify the OTU to the species level.

Global distribution of top prevalent OTUs

We searched top prevalent bacteria in the study population against the NCBI database (Supplementary Table 11). Close-matching OTUs were detected in individuals from Bangladesh, China, India, Iran, Russia, Japan and USA; some OTUs also had a close match to fecal microbiota isolated from farm animals. Unfortunately, due to under-representation of healthy cohorts in matching studies, we could not draw any age associations for these OTUs based on BLAST-search results. Nor could we address age associations of our top OTUs in publicly available MG-RAST (Meyer et al., 2008) datasets because of non-consistent age info representation for the samples within these datasets. We then addressed whether these OTUs

were found in a healthy adult population. We randomly selected 20 samples from the Human Microbiome Project (Aagaard et al., 2013) and BLAST searched top OTUs against trimmed 16S rRNA reads from these samples. All OTUs, apart from OTU103 (*Enterococcus*), OTU104 (*Lactobacillus zeae*) and OTU54 (*Granulicatella*) that were prevalent in children at 3 months and 1 year respectively, matched with more than 97 % identity towards HMP reads (Supplementary Table 12).

Discussion

Selective transmission of bacteria from mother to child has been previously demonstrated at birth (Jakobsson et al., 2014) and through breast milk (Martin et al., 2012; Jost et al., 2014). In this work, we observed no difference in the number of shared OTUs between children and related or unrelated mothers. At the same time, OTUs that were highly prevalent in the mother population also tended to appear earlier and at higher prevalence in children. Moreover, from 3 months onward, we observed a declining number of OTUs that had high association to mothers, despite the increase of a mother-child shared pool of bacteria. We propose that these facts indicate recruitment of bacteria from a common pool in a population where most prevalent microbiota is transmitted first-line and less frequently occurring bacteria is recruited later. Interestingly, in line with our previous observations (Avershina et al., 2013), we detected lower deviation from neutrality in the group of mothers and their 1- and 2-year-old children than in newborns and 3-month-olds. We believe that the direction of microbiota assembly towards neutrality with age also supports a relatively random colonization process from common pools. From our data we cannot address how geographically defined these common population pools are. Our study was performed in the Trondheim region of mid-Norway, which is roughly 340.4 km^2. Other studies, however, have shown that household members, as well as individuals living in remote tribes and villages, have more similar microbiota to each other than to other people (Song et al., 2013; Kwok et al., 2014; Schnorr et al., 2014; Zhang et al., 2014).

We observed large differences in the most prevalent OTUs between mothers and children up until 1 year of age. At 2 years of age, though, the majority of the top prevalent OTUs in the child population was replaced by OTUs which had similar prevalence in mothers. Interestingly, detection of one of the top bifidobacteria in a population during infancy negatively correlated to detection of late-appearing and positively correlated to detection of earlier-appearing clostridia in 2-year-old children. Varying mechanistic interaction ability between clostridia and bifidobacteria has been noted previously (Trejo et al., 2006; Vlkova et al., 2008). This may indicate that bifidobacteria, which are promoted by breast-feeding (Hinde and Lewis, 2015), contribute to delayed adult associated stool bacteria colonization by differentially inhibiting growth of those-to-be late colonizers. Interestingly, median duration of breastfeeding in our cohort was 11 months, coinciding with supremacy of OTUs characteristic of the first year of life. In line with this hypothesis, cessation of breastfeeding rather than introduction of solid foods has recently been demonstrated to be a driving force of microbial community structure (Backhed et al., 2015). Though we observed delayed reduction of OTU2 after termination of breastfeeding, there was an apparent drop in its abundance at 1 year of age, when the majority of children were no longer breastfed. OTU2 was detected in less than 10 % of the mother population, and its abundance never exceeded 0.7 %. Therefore, we believe that the switch between children and adult community structure may be at least partially driven by the disappearance of key bifidobacteria. However, further investigations on whether OTU2 is breast-milk selected, and what mechanistic role this OTU has, are needed.

During nearly the entire study period, we observed an increased proportion of OTUs belonging to Bacteroidia class in a subset of OTUs that each child shared with his/her mother. Consistent with these observations, several studies reported depletion of Bacteroides in stool samples of C-section delivered babies (Penders et al., 2006; Jakobsson et al., 2014), suggesting strong association of this bacterial group with mothers and its direct transmission from mother to child.

This is concordant with the inability of Bacteroides to have non-vegetative forms (Wexler, 2007), unlike Clostridia which are capable of spore formation (Meehan and Beiko, 2014).

In conclusion, we propose a model that colonization of the gut can be represented as entering a restricted-access building. We believe that in the early days of life, breast-milk promoted bifidobacteria, in combination with direct host selection, act as a gatekeeper that grants access to the chosen ones and leaves the rest behind. Later on, when breast milk selection stops, the door opens for the entire pool of bacteria that are allowed to colonize the healthy adult gut.

References

Aagaard, K., Ma, J., Antony, K.M., Ganu, R., Petrosino, J., and Versalovic, J. (2014) The placenta harbors a unique microbiome. *Sci Transl Med* **6**: 237ra265.

Aagaard, K., Petrosino, J., Keitel, W., Watson, M., Katancik, J., Garcia, N. et al. (2013) The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB J* **27**: 1012-1022.

Antunes, L.C., McDonald, J.A., Schroeter, K., Carlucci, C., Ferreira, R.B., Wang, M. et al. (2014) Antivirulence activity of the human gut metabolome. *MBio* **5**: e01183-01114.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R. et al. (2011) Enterotypes of the human gut microbiome. *Nature* **473**: 174-180.

Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., Nishikawa, H. et al. (2013) Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**: 232-236.

Avershina, E., and Rudi, K. (2015) Confusion about the species richness of human gut microbiota. *Benef Microbes* **in press**.

Avershina, E., Storro, O., Oien, T., Johnsen, R., Pope, P., and Rudi, K. (2013) Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiol Ecol*.

Avershina, E., Storro, O., Oien, T., Johnsen, R., Pope, P., and Rudi, K. (2014) Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiol Ecol* **87**: 280-290.

Backhed, F., Roswall, J., Peng, Y., Feng, Q., Jia, H., Kovatcheva-Datchary, P. et al. (2015) Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe* **17**: 690-703.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.

Clemente, J.C., Ursell, L.K., Parfrey, L.W., and Knight, R. (2012) The impact of the gut microbiota on human health: an integrative view. *Cell* **148**: 1258-1270.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K. et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**: 5069-5072.

DiGiulio, D.B. (2012) Diversity of microbes in amniotic fluid. *Semin Fetal Neonatal Med* **17**: 2-11.

DiGiulio, D.B., Romero, R., Amogan, H.P., Kusanovic, J.P., Bik, E.M., Gotsch, F. et al. (2008) Microbial prevalence, diversity and abundance in amniotic fluid during preterm labor: a molecular and culture-based investigation. *PLoS One* **3**: e3056.

DiGiulio, D.B., Callahan, B.J., McMurdie, P.J., Costello, E.K., Lyell, D.J., Robaczewska, A. et al. (2015) Temporal and spatial variation of the human microbiota during pregnancy. *Proc Natl Acad Sci U S A*.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and 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.

Dotterud, C.K., Storro, O., Johnsen, R., and Oien, T. (2010) Probiotics in pregnant women to prevent allergic disease: a randomized, double-blind trial. *Br J Dermatol* **163**: 616-623.

Dotterud, C.K., Avershina, E., Sekelja, M., Simpson, M.R., Rudi, K., Storro, O. et al. (2015) Does Maternal Perinatal Probiotic Supplementation Alter the Intestinal Microbiota of Mother and Child? A Randomized Controlled Trial. *J Pediatr Gastroenterol Nutr*.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.

Edgar, R.C. (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* **10**: 996-998.

Faith, J.J., Guruge, J.L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A.L. et al. (2013) The long-term stability of the human gut microbiota. *Science* **341**: 1237439.

Hehemann, J.H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**: 908-912.

Hinde, K., and Lewis, Z.T. (2015) MICROBIOTA. Mother's littlest helpers. *Science* **348**: 1427-1428.

Jabot, F., and Chave, J. (2011) Analyzing tropical forest tree species abundance distributions using a nonneutral model and through approximate Bayesian inference. *Am Nat* **178**: E37-47.

Jakobsson, H.E., Abrahamsson, T.R., Jenmalm, M.C., Harris, K., Quince, C., Jernberg, C. et al. (2014) Decreased gut microbiota diversity, delayed Bacteroidetes colonisation and reduced Th1 responses in infants delivered by caesarean section. *Gut* **63**: 559-566.

Jeurink, P.V., van Bergenhenegouwen, J., Jimenez, E., Knippels, L.M., Fernandez, L., Garssen, J. et al. (2013) Human milk: a source of more life than we imagine. *Benef Microbes* **4**: 17-30.

Jimenez, E., Marin, M.L., Martin, R., Odriozola, J.M., Olivares, M., Xaus, J. et al. (2008) Is meconium from healthy newborns actually sterile? *Res Microbiol* **159**: 187-193.

Jimenez, E., Fernandez, L., Marin, M.L., Martin, R., Odriozola, J.M., Nueno-Palop, C. et al. (2005) Isolation of commensal bacteria from umbilical cord blood of healthy neonates born by cesarean section. *Curr Microbiol* **51**: 270-274.

Jost, T., Lacroix, C., Braegger, C.P., Rochat, F., and Chassard, C. (2014) Vertical mother-neonate transfer of maternal gut bacteria via breastfeeding. *Environ Microbiol* **16**: 2891-2904.

Kwok, L.Y., Zhang, J., Guo, Z., Gesudu, Q., Zheng, Y., Qiao, J. et al. (2014) Characterization of fecal microbiota across seven Chinese ethnic groups by quantitative polymerase chain reaction. *PLoS One* **9**: e93631.

Madan, J.C., Farzan, S.F., Hibberd, P.L., and Karagas, M.R. (2012) Normal neonatal microbiome variation in relation to environmental factors, infection and allergy. *Curr Opin Pediatr* **24**: 753-759.

Magurran, A.E. (2004) *Measuring biological diversity*: Blackwell Science Ltd.

Martin, V., Maldonado-Barragan, A., Moles, L., Rodriguez-Banos, M., Campo, R.D., Fernandez, L. et al. (2012) Sharing of bacterial strains between breast milk and infant feces. *J Hum Lact* **28**: 36-44.

Meehan, C.J., and Beiko, R.G. (2014) A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. *Genome Biol Evol* **6**: 703-713.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M. et al. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9**: 386.

Mulle, J.G., Sharp, W.G., and Cubells, J.F. (2013) The gut microbiome: a new frontier in autism research. *Curr Psychiatry Rep* **15**: 337.

Penders, J., Thijs, C., Vink, C., Stelma, F.F., Snijders, B., Kummeling, I. et al. (2006) Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* **118**: 511-521.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F. et al. (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**: 55-60.

Satokari, R., Gronroos, T., Laitinen, K., Salminen, S., and Isolauri, E. (2009) Bifidobacterium and Lactobacillus DNA in the human placenta. *Lett Appl Microbiol* **48**: 8-12.

Schanche, M., Avershina, E., Dotterud, C., Oien, T., Storro, O., Johnsen, R., and Rudi, K. (2015) High-Resolution Analyses of Overlap in the Microbiota Between Mothers and Their Children. *Curr Microbiol*.

Schnorr, S.L., Candela, M., Rampelli, S., Centanni, M., Consolandi, C., Basaglia, G. et al. (2014) Gut microbiome of the Hadza hunter-gatherers. *Nat Commun* **5**: 3654.

Sjogren, Y.M., Jenmalm, M.C., Bottcher, M.F., Bjorksten, B., and Sverremark-Ekstrom, E. (2009) Altered early infant gut microbiota in children developing allergy up to 5 years of age. *Clin Exp Allergy* **39**: 518-526.

Song, S.J., Lauber, C., Costello, E.K., Lozupone, C.A., Humphrey, G., Berg-Lyons, D. et al. (2013) Cohabiting family members share microbiota with one another and with their dogs. *Elife* **2**: e00458.

Trejo, F.M., Minnaard, J., Perez, P.F., and De Antoni, G.L. (2006) Inhibition of Clostridium difficile growth and adhesion to enterocytes by Bifidobacterium supernatants. *Anaerobe* **12**: 186-193.

Vlkova, E., Rada, V., Smehilova, M., and Killer, J. (2008) Auto-aggregation and co-aggregation ability in bifidobacteria and clostridia. *Folia Microbiol (Praha)* **53**: 263-269.

Wexler, H.M. (2007) Bacteroides: the good, the bad, and the nitty-gritty. *Clin Microbiol Rev* **20**: 593-621.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M. et al. (2012) Human gut microbiome viewed across age and geography. *Nature* **486**: 222-227.

Yu, Y., Lee, C., Kim, J., and Hwang, S. (2005) Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnol Bioeng* **89**: 670-679.

Zhang, J., Guo, Z., Lim, A.A., Zheng, Y., Koh, E.Y., Ho, D. et al. (2014) Mongolians core gut microbiota and its correlation with seasonal dietary changes. *Sci Rep* **4**: 5001.

Tables

Table 1. Number of OTUs detected in each class during the study period. Only classes with significant differences in richness detected at least once throughout the study, are given. * $p < 0.05$ as compared to a previous study time point; ** $p < 0.05$ between mother and child at 2 years of age.

OTU class	OTU richness (average ± standard deviation; values rounded up)					
	Mother		Child			
	Pregnancy	3 months	10 days	3months	1year	2years
Clostridia**	92 ± 28	92 ± 28	6 ± 6	6 ± 9	31 ± 12	55 ± 17*
Bacteroidia**	16 ± 8	19 ± 8	3 ± 4	3 ± 4	6 ± 5*	12 ± 6*
Actinobacteria**	5 ± 2	5 ± 2	5 ± 3	7 ± 3*	6 ± 3	6 ± 3
Erysipelotrichia	5 ± 3	4 ± 2	1 ± 1	1 ± 1	4 ± 2	4 ± 2
Bacilli**	4 ± 3	4 ± 3	6 ± 2	6 ± 3	7 ± 3	6 ± 3
Mollicutes**	4 ± 4	3 ± 4	1 ± 1	1 ± 1	1 ± 1	1 ± 1*
Verrucomicrobiae	2 ± 2	2 ± 2	1 ± 1	1 ± 1	2 ± 2	2 ± 2
Coriobacteriia	3 ± 2	3 ± 2	1 ± 1	2 ± 2*	2 ± 2	3 ± 2
Gammaproteobacteria	2 ± 2	1 ± 2	2 ± 2	3 ± 2	3 ± 2	2 ± 1*
Methanobacteria**	1 ± 1	1 ± 1	1 ± 1	1 ± 1	1 ± 1	1 ± 1

Table 2. Average mothers prevalence of OTUs detected in the child population at a given age. *$p < 0.05$ as compared to a previous point in time.

Number of mothers	OTUs detected for 1st time at			
	10 days	3 months	1 year	2 years
mean	33.2	20.6*	10.5*	8.7
standard dev	29.7	18.8	11.2	7.8

445 **Table 3.** List of OTUs that have a higher probability of being detected in mothers of OTU-
446 positive children.

Age	OTU		Number of mothers of positive children		Avg prevalence in mother population, %
	Number	Taxonomy (class; lowest classified level)	positive mothers	total	
10	151	Bacteroidia; *Bacteroides*	7	9	30.9
	1300	Bacteroidia; *Bacteroides*	38	47	67.1
	1071	Bacteroidia; *Parabacteroides distasonis*	20	27	45.8
90	151	Bacteroidia; *Bacteroides*	5	5	30.9
	1300	Bacteroidia; *Bacteroides*	39	43	67.1
	47	Actinobacteria; *Bifidobacterium*	43	86	36.5
	1071	Bacteroidia; *Parabacteroides distasonis*	21	27	45.8
	199	Bacteroidia; *Bacteroides*	4	4	36.1
	328	Coriobacteriia	3	3	12.4
	1058	Bacteroidia; *Bacteroides*	10	28	17.7
	442	Clostridia	3	3	7.2
	216	Clostridia	5	10	16.5
	476	Coriobacteriia; *Slackia*	2	2	8.8
	963	Alphaproteobacteria; *Sphingomonas*	2	11	2.8
	11	Clostridia; *Veillonella*	1	12	0.4
	631	Erysipelotrichia; *[Eubacterium]*	2	2	1.6
	190	Coriobacteriia; *Collinsella stercoris*	2	2	4.0
360	239	Bacteroidia; *Bacteroides*	4	6	19.3
	159	Clostridia; *Phascolarctobacterium*	4	4	25.3
	37	Bacteroidia; *Bacteroides*	44	49	73.1
	3	Actinobacteria; *Bifidobacterium*	73	101	63.1
	322	Bacteroidia; *Bacteroides eggerthii*	5	6	20.1
	199	Bacteroidia; *Bacteroides*	11	14	36.1
	328	Coriobacteriia	3	5	12.4
	371	Clostridia; *Blautia*	2	2	5.6
	325	Bacteroidia; *Bacteroides*	2	2	18.1
	710	Bacteroidia; *Butyricimonas*	2	4	3.2
	476	Coriobacteriia; *Slackia*	2	3	8.8
	553	Bacteroidia; *Prevotella copri*	2	4	7.6
	362	Clostridia; *Pseudoramibacter_Eubacterium*	2	8	3.2
720	397	Erysipelotrichia; cc_115	2	2	12.4
	159	Clostridia; *Phascolarctobacterium*	10	15	25.3
	322	Bacteroidia; *Bacteroides eggerthii*	6	7	20.1
	639	Bacteroidia; *Porphyromonas*	3	6	11.6
	612	Clostridia	5	12	10.8
	326	Bacteroidia; *Butyricimonas*	2	2	13.3
	747	Clostridia	2	4	7.6
	710	Bacteroidia; *Butyricimonas*	2	2	3.2

447 **Figures**

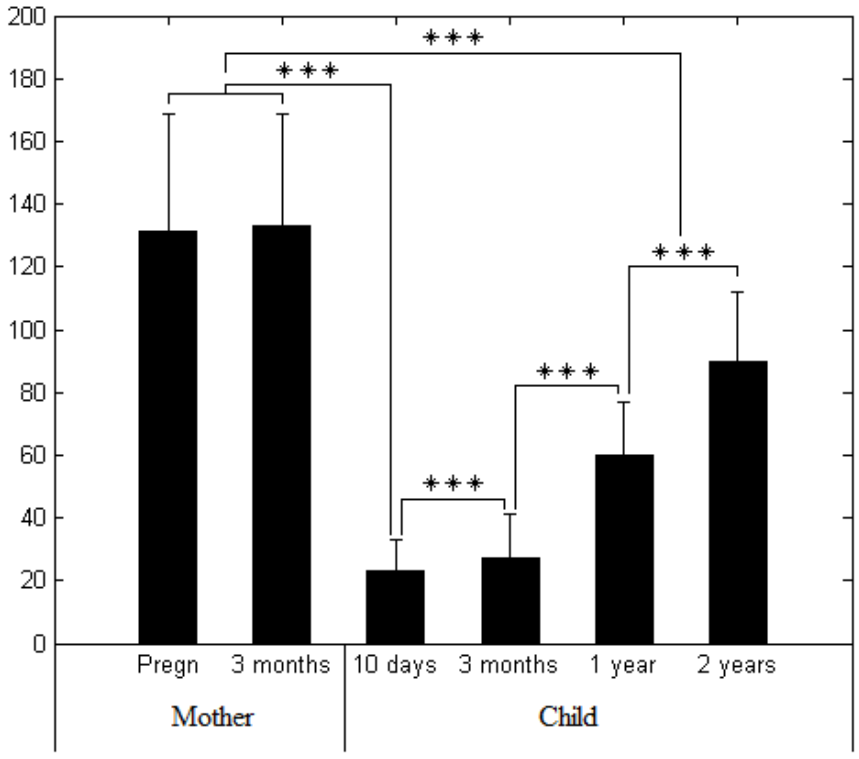
448 **Figure 1.** Gut microbial community structure in a population. **A**. Number of OTUs per
449 individual; **B**. Class distribution of OTUs; **C**. Average abundance of bacterial classes; n –
450 number of individuals

451 **Figure 2**. Sharing of OTUs in the mother population. **A**. Average number of OTUs shared
452 within and between individuals at pregnancy and 3 months postpartum, *** $p < 0.001$; **B**.
453 Average number of OTUs shared between at least a given fraction of the mother population and
454 their relative abundance therein, *from 9 % to 2 mothers.

455 **Figure 3**. Sharing of OTUs between mothers and their children. **A**. Average number of shared
456 and individual OTUs per mother-child pair. **B**. Class distribution of OTUs in children-unique
457 and mother-child shared OTU pools. **C**. Average number of highly mother-associated OTUs
458 per individual.

459 **Figure 4**. Top 10 prevalent OTUs in a population at various ages. **A**. Ratio between prevalence
460 of top 10 OTUs in children and their mothers. **B**. Correlations in detection of top
461 *Bifidobacterium* OTU and Clostridia OTUs at 2 years of age.

462 Figure 1A



463

Figure 1B

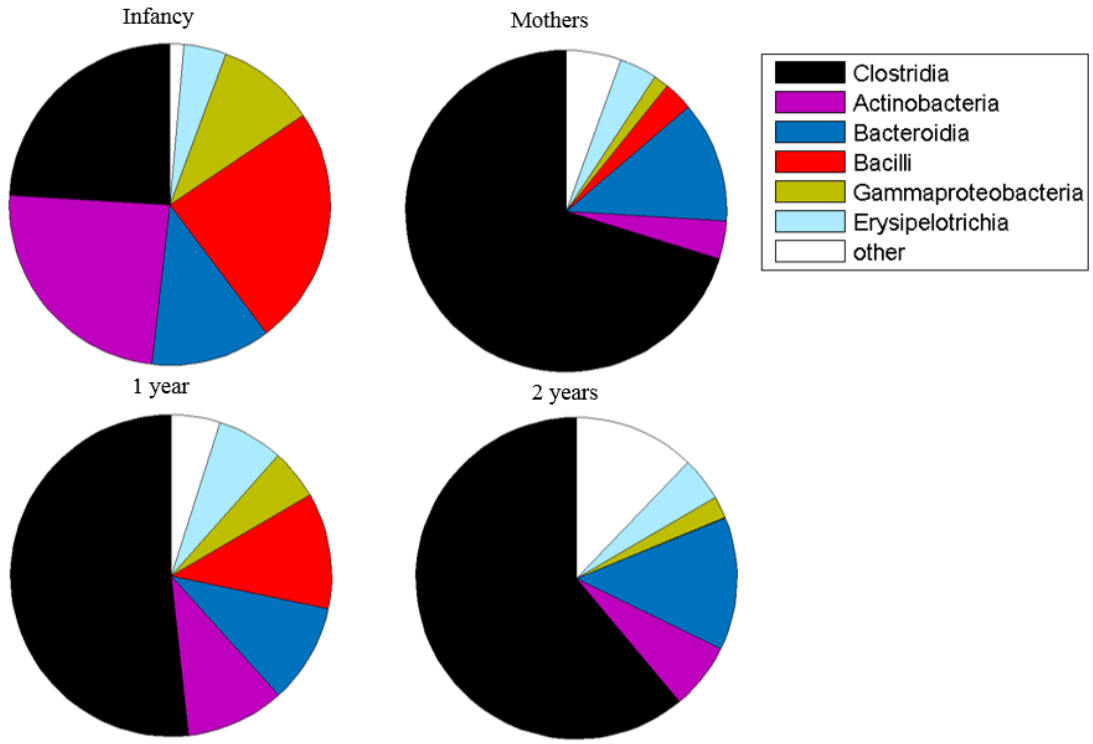
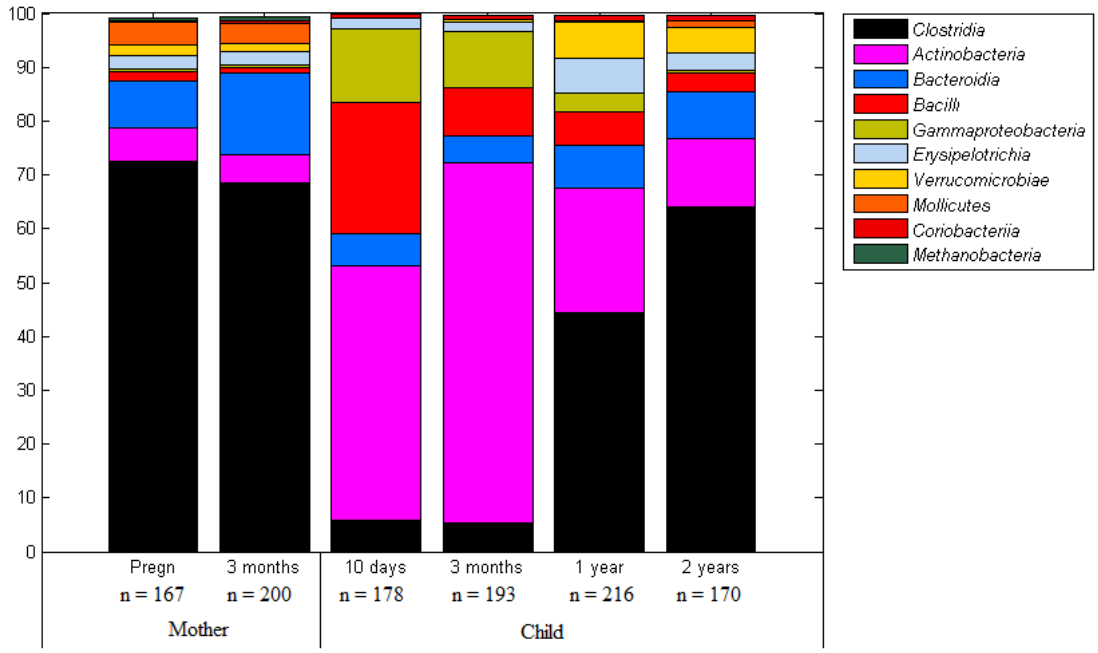
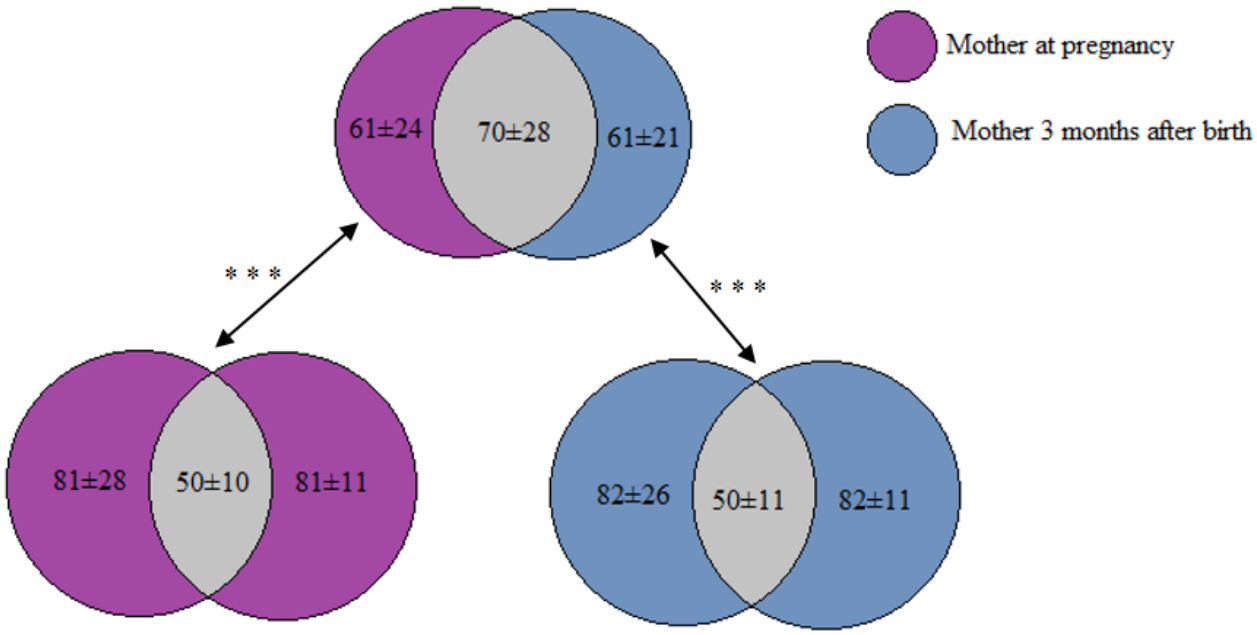


Figure 1C

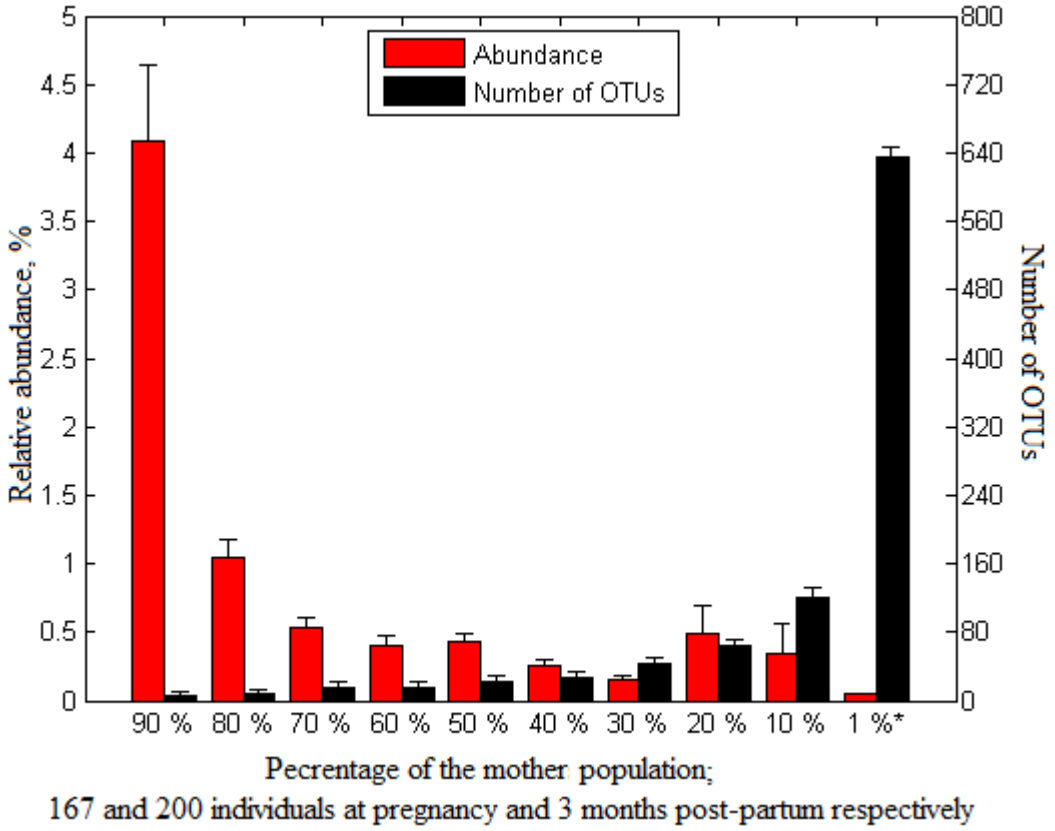


471 Figure 2A



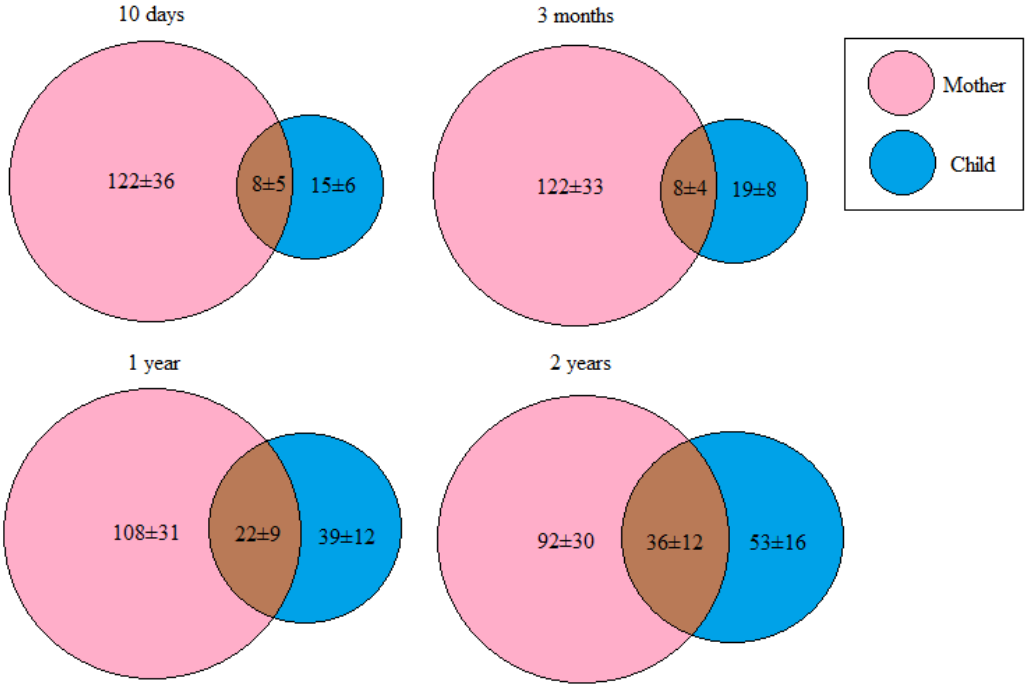
472

473 Figure 2B



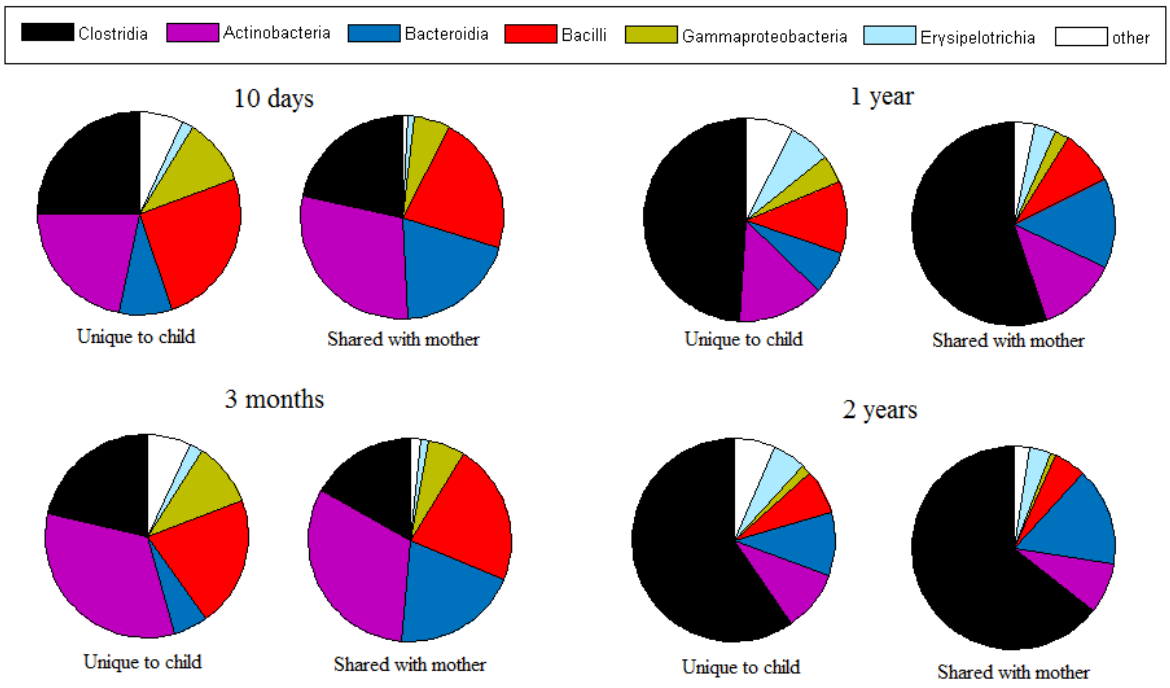
474

475 Figure 3A



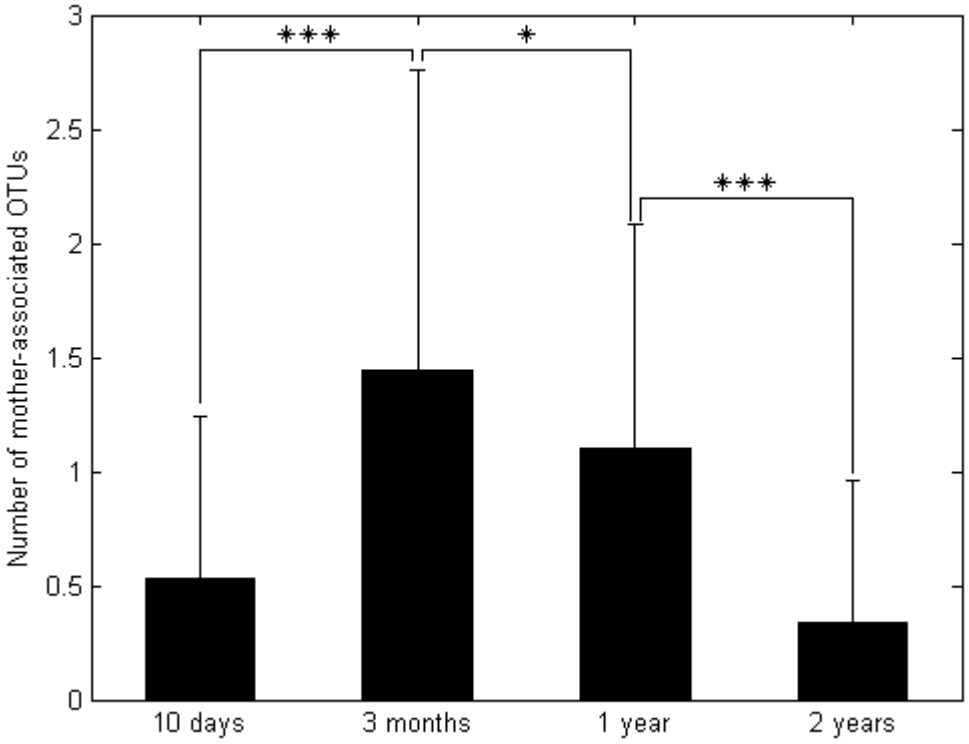
476

477 Figure 3B



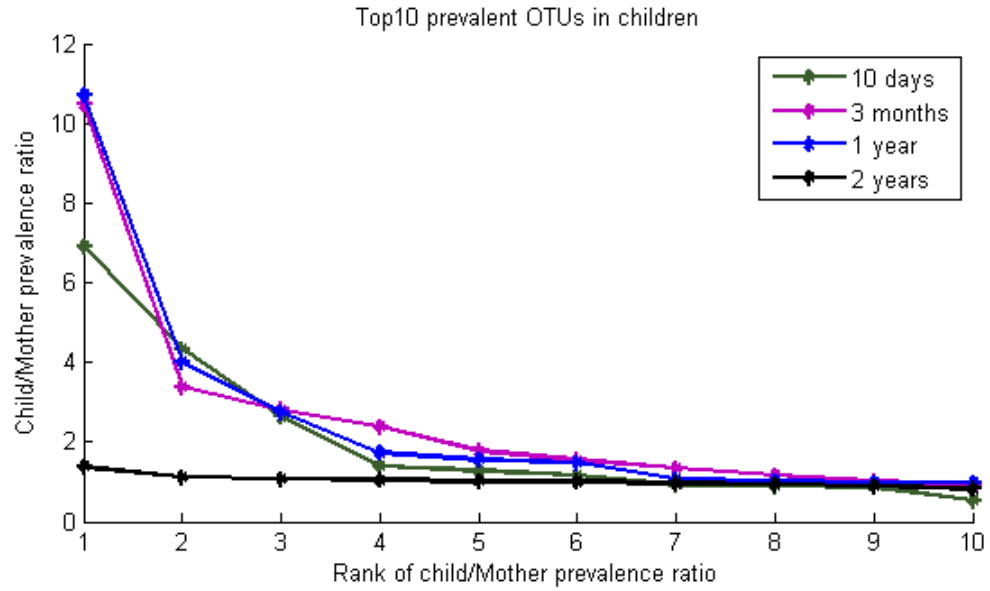
478

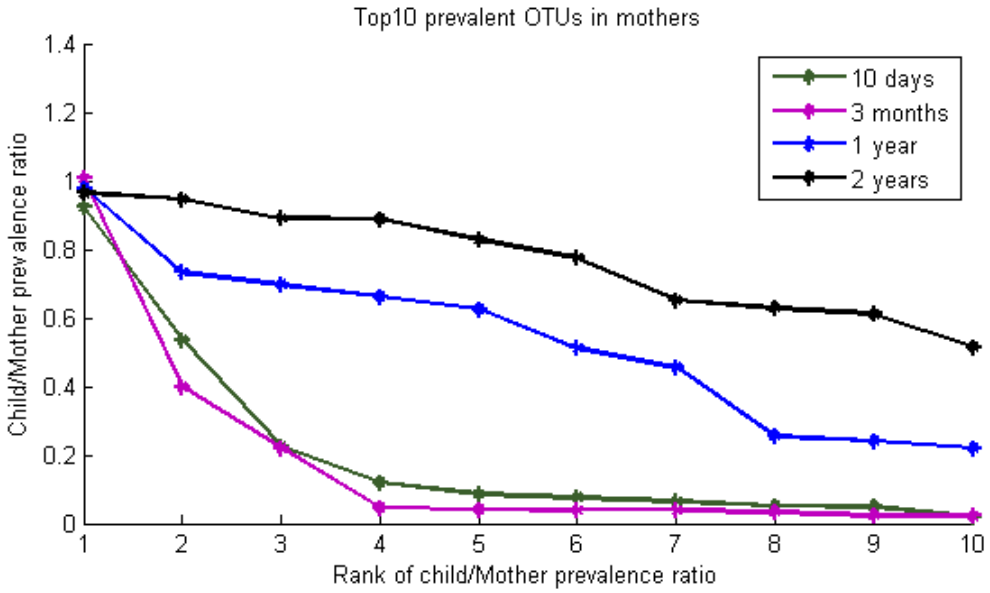
479 Figure 3C



480

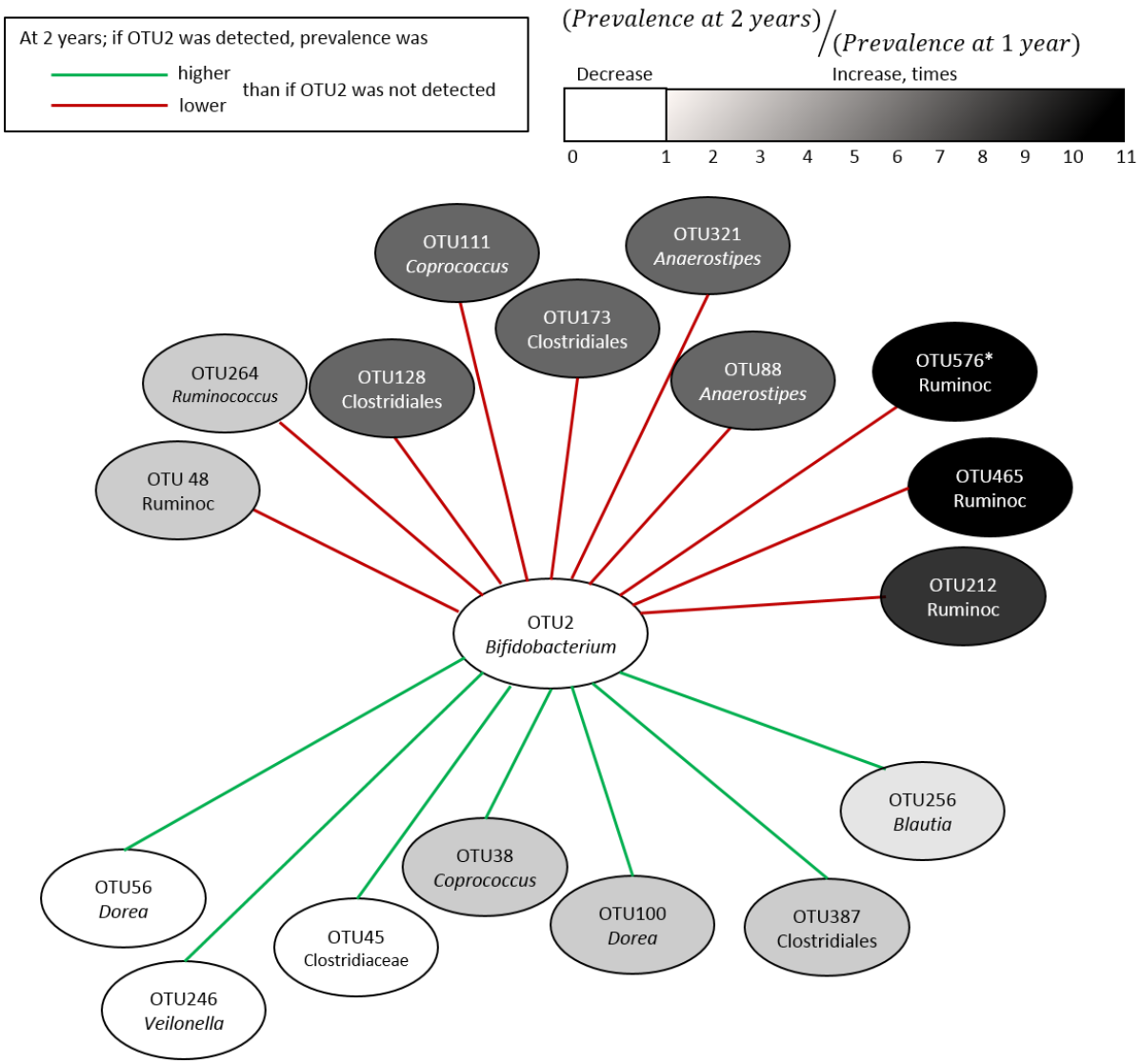
481 Figure 4A





482

483 Figure 4B

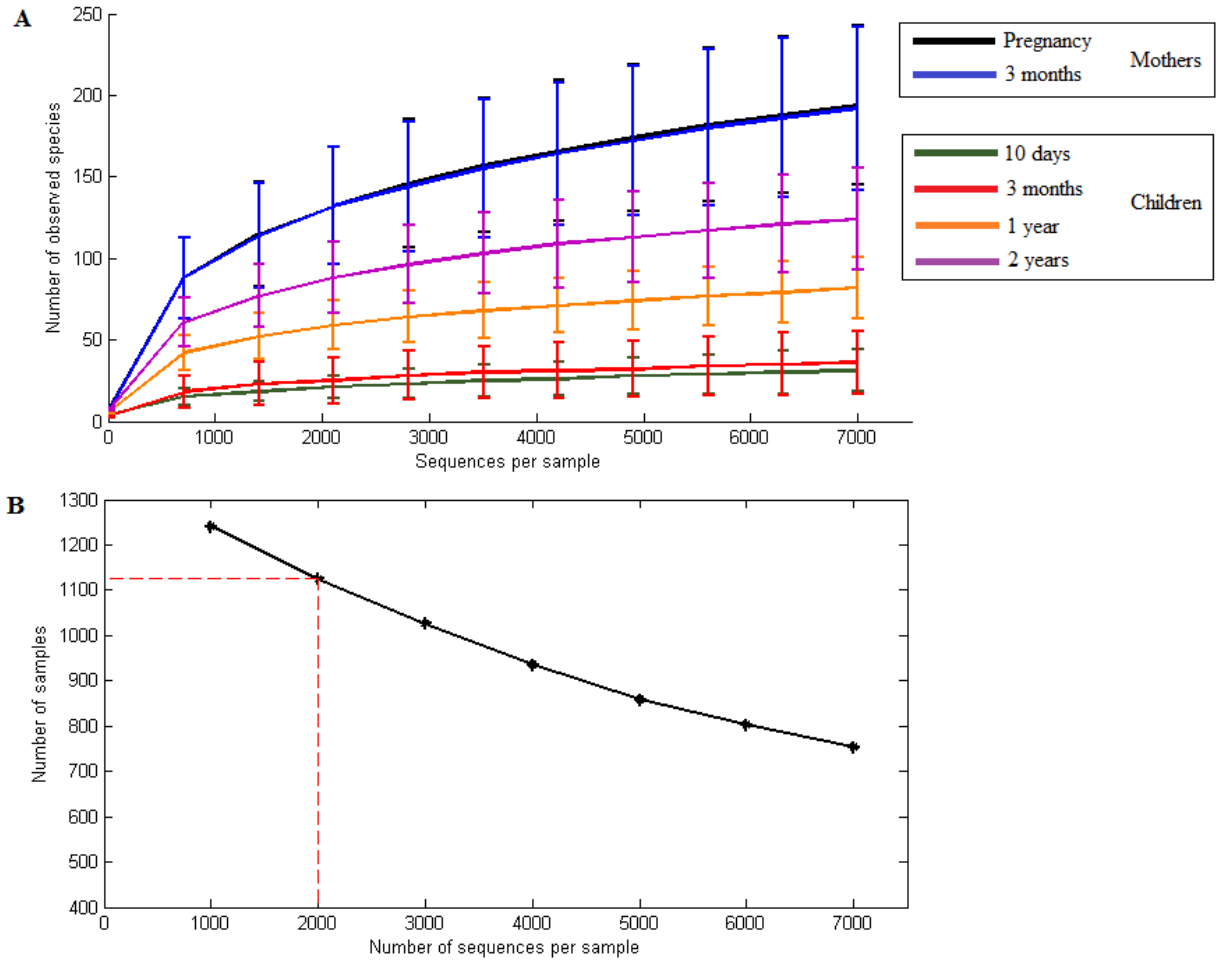


484

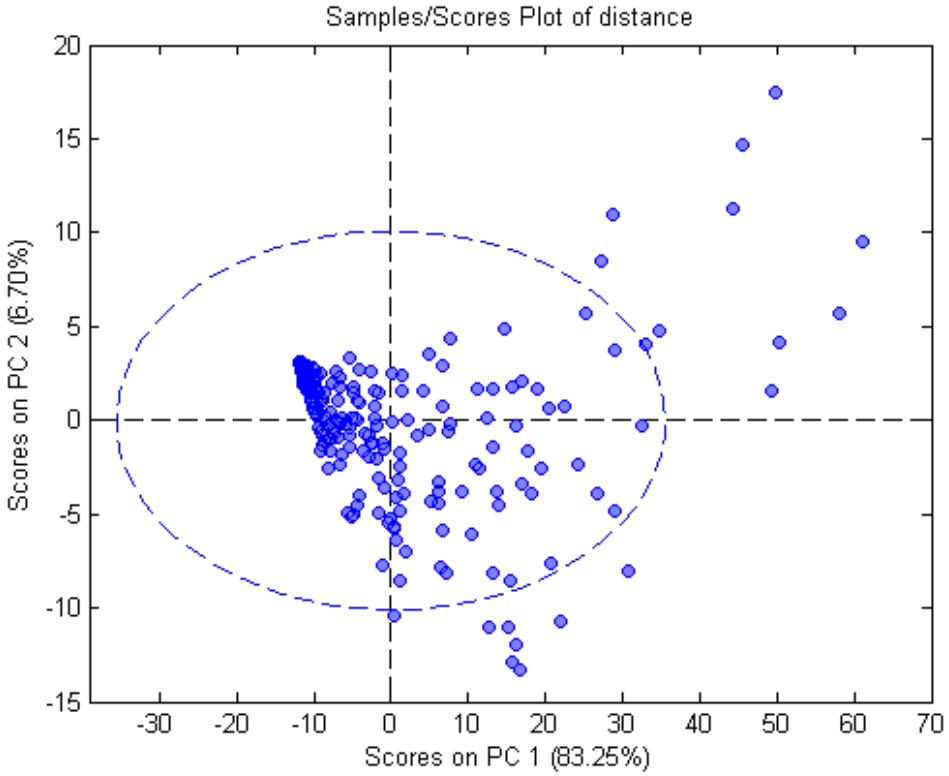
485

486

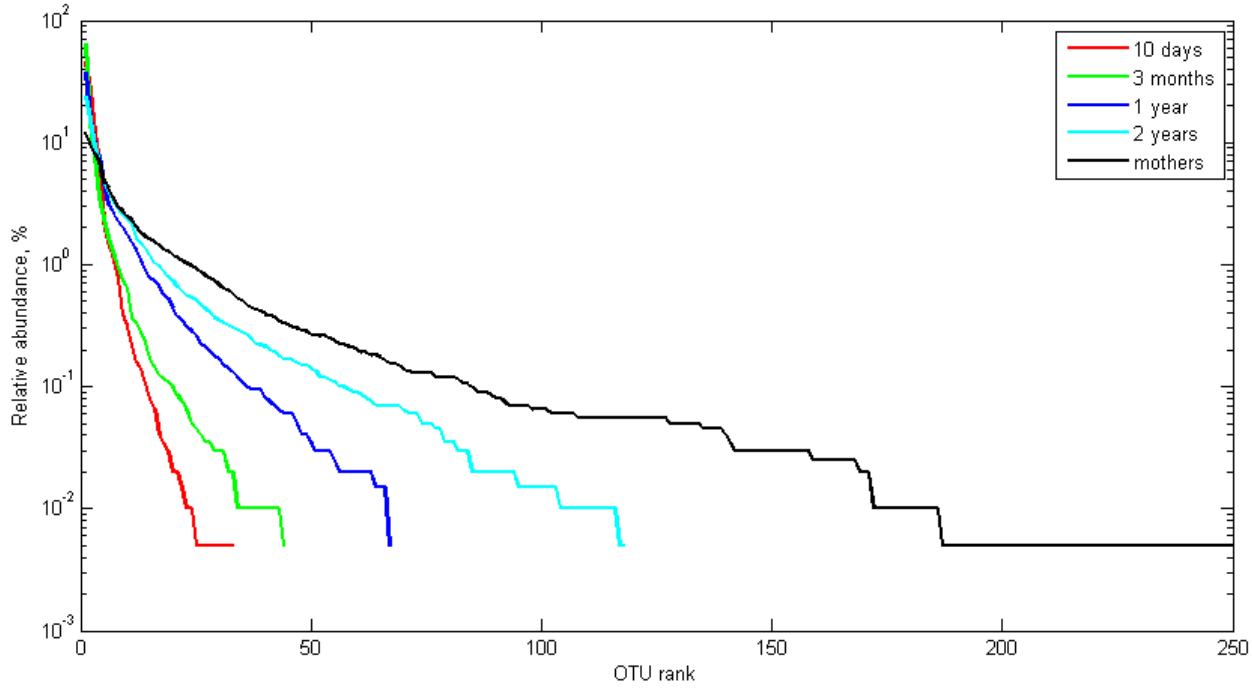
Supplementary Figure 1. General characteristics of the dataset. **A**. Rarefaction curves for the observed species in a dataset with respect to number of sequences per sample included. **B**. Number of samples in a dataset with varying sequences per sample cutoff value. Number of samples in a normalized dataset is highlighted



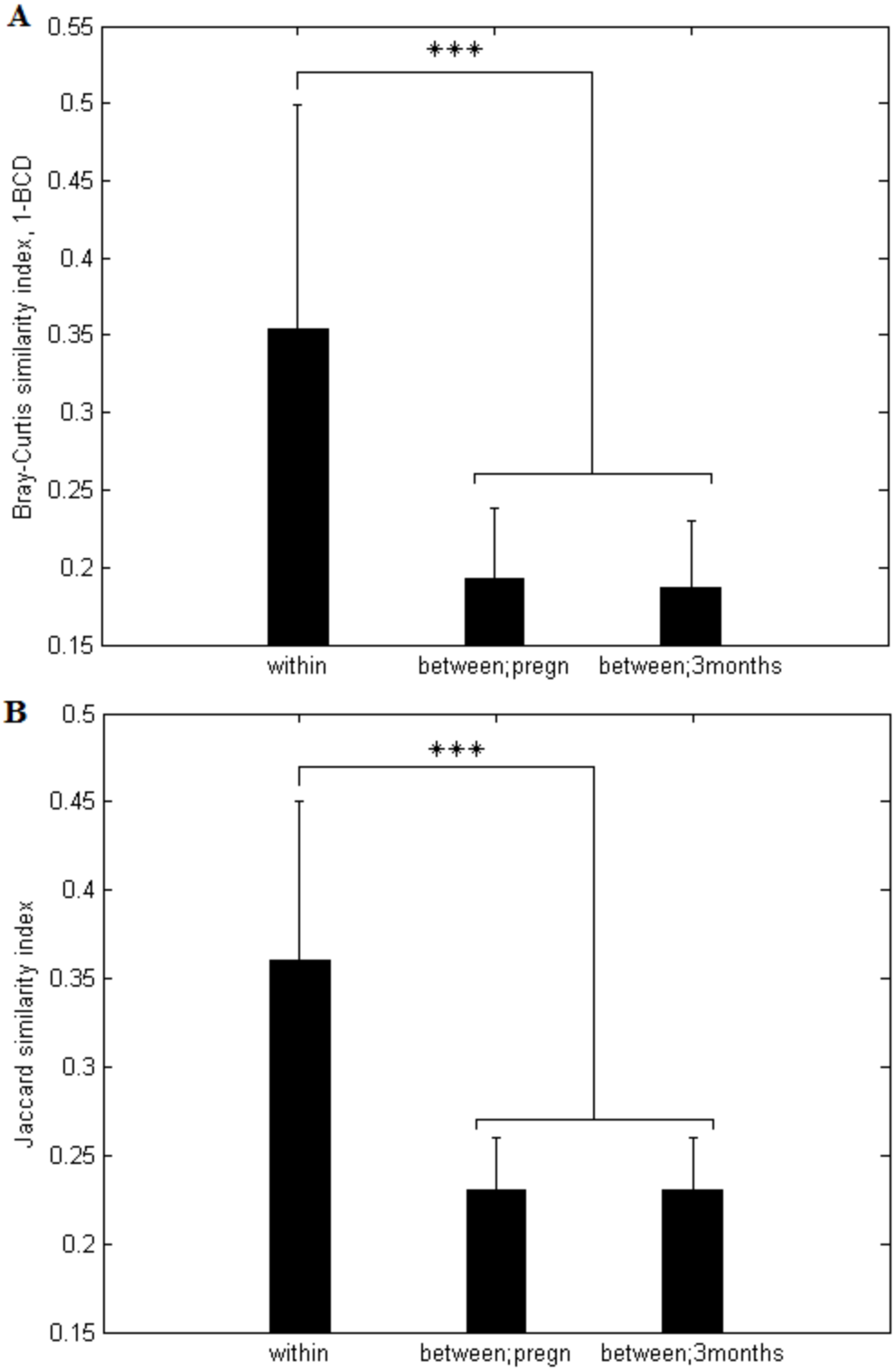
Supplementary Figure 2. PCA on Jensen-Shannon divergence distances between mothers' stool samples. Relative abundance of not-detected OTUs was changed to 0.0000001 in concordance with original publication on enterotype separations (Arumugam et al., 2011; Nature).



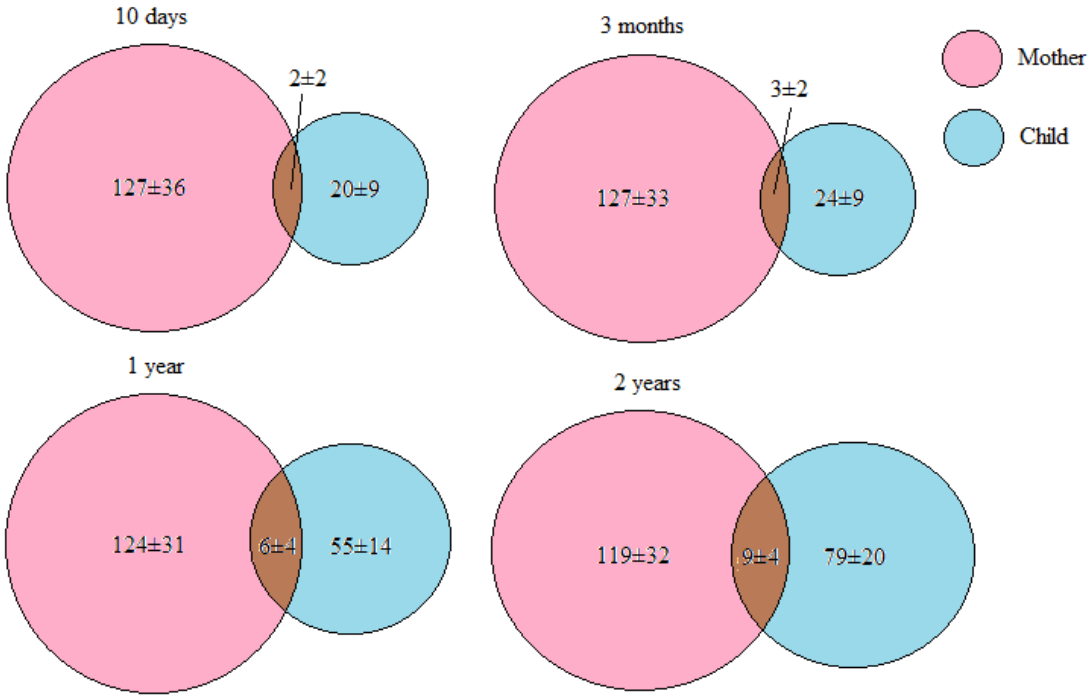
Supplementary Figure 3. Rank abundance distributions.



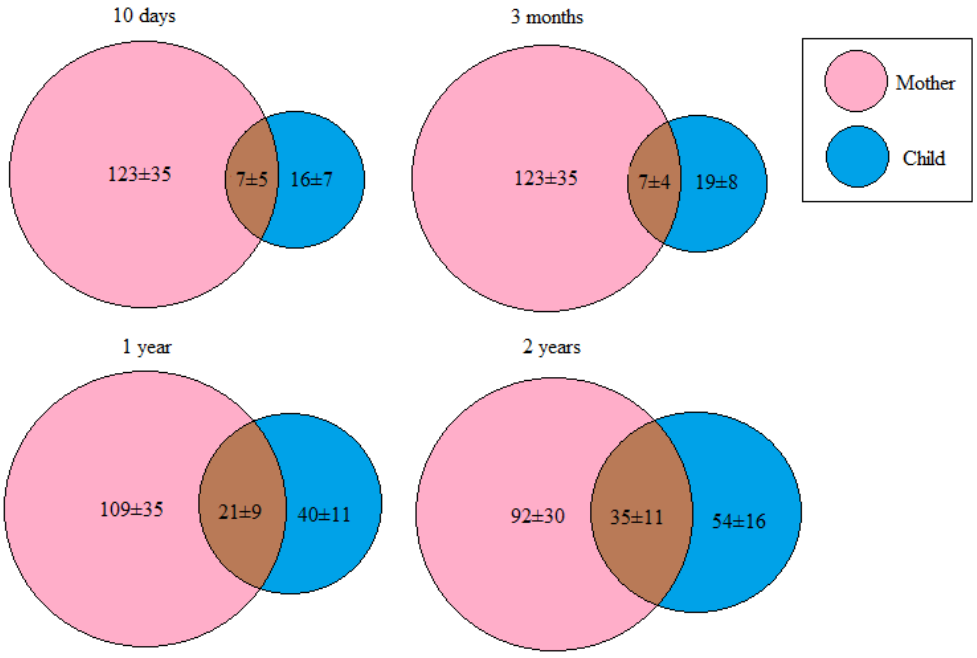
Supplementary Figure 4. Diversity measures between mother's samples. Bray-Curtis similarity index (1-dissimilarity) **B**. Jaccard similarity index; ***p < 0.001



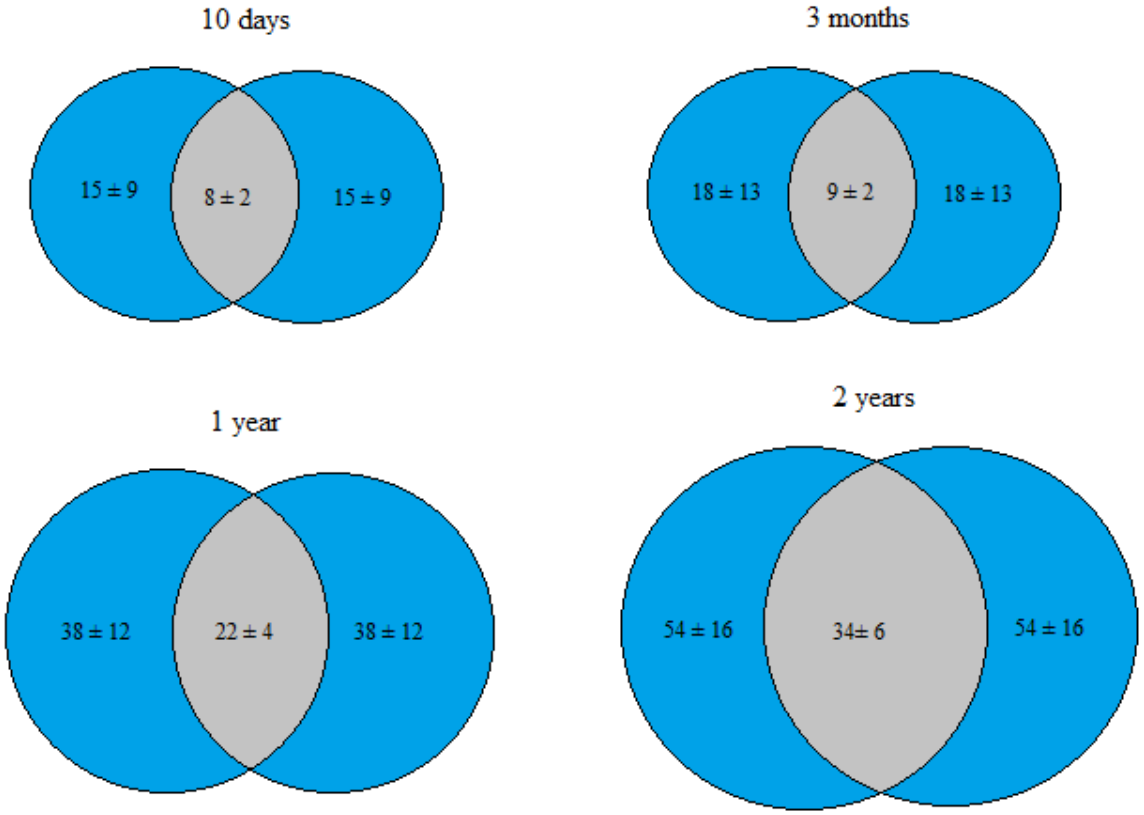
Supplementary Figure 5. Average number of individual/shared OTUs between mothers and their children in a random OTU dataset. OTU profiles from each mother were randomly permuted and then compared to their children.



Supplementary Figure 6. Average number of individual/shared OTUs between random mothers and children. Mother IDs were randomly permuted and then compared to 'their' infants



Supplementary Figure 7. Average number of individual/shared OTUs between children populations at various ages



Supplementary Figure 8. Diversity measures between children samples. **A**. Bray-Curtis similarity index (1–dissimilarity) **B**. Jaccard similarity index; *** $p < 0.001$