Establishment of intestinal microbiota during early life: A longitudinal, explorative study of a large cohort of Danish infants

Anders Bergström¹, Thomas Hjort Skov², Martin Iain Bahl¹, Henrik Munch Roager¹, Line Brinch Christensen²,
Katrine Tschentscher Ejlerskov², Christian Mølgaard², Kim F. Michaelsen², Tine Rask Licht¹*

1Technical University of Denmark, National Food Institute, Denmark; 5
2University of Copenhagen, Faculty of Science, Department of Nutrition, Exercise and Sports, Denmark 6
*Corresponding author: Technical University of Denmark, National Food Institute, Mørkhøj Bygade 26, DK-
2800 Lyngby, trli@food.dtu.dk

Running title

Development of the infant gut microbiota

Keywords

Obesity/Breastfeeding/Enterotypes/Gut Microbiota/SKOT cohort/qPCR

Section

Microbial Ecology

AEM Accepts, published online ahead of print on 28 February 2014
Copyright © 2014, American Society for Microbiology. All Rights Reserved.
Abstract

Fecal samples were obtained from a cohort of 330 healthy Danish infants at 9, 18 and 36 months after birth, enabling characterization of interbacterial relationships by use of quantitative PCR targeting 31 selected bacterial 16S rRNA gene targets representing different phylogenetic levels. Nutritional parameters and measures of growth and body composition were determined and investigated in relation to the observed development in microbiota composition. We found that significant changes in the gut microbiota occurred, particularly from age 9 to 18 months, where cessation of breastfeeding and introduction of a complementary feeding induces replacement of microbiota characterized by lactobacilli, bifidobacteria and Enterobacteriaceae with a microbiota dominated by Clostridium spp. and Bacteroides spp. Classification of samples by a proxy ‘enterotype’ based on the relative levels of Bacteroides spp. and Prevotella spp. showed that ‘enterotype’ establishment occurs between 9 and 36 months. 30% of the individuals shifted ‘enterotype’ between 18 and 36 months. The composition of the microbiota was most pronouncedly influenced by the time of cessation of breastfeeding. From 9 to 18 months, a positive correlation was observed between the increase in Body Mass Index and the increase of the Short Chain Fatty Acid producing clostridia, C. leptum group and E. halii.

Considering previously established positive associations between rapid infant weight gain, early breastfeeding discontinuation and later life obesity, the corresponding microbial findings seen here warrant attention.
Introduction

Establishment of the human intestinal microbiota during infancy is influenced by multiple factors, including delivery mode, sanitary conditions, administration of antibiotics to the infant or mother \( (2;18) \), and level of breast-feeding \( (17) \). Breastfeeding has been shown to significantly increase the relative abundance of bifidobacteria and lactic acid bacteria including lactobacilli and Enterococcus spp. \( (30;69) \). The microbial composition within the first year of life is typically characterized by low species diversity and high instability \( (41;61;75) \). Nevertheless, a number of recent studies suggest that some of the bacteria that become part of the adult microbiota colonize the gut already during the first months of life \( (35;36) \). A more complex, stable and adult-like microbiota is established between 1 and 2 years after birth \( (51;54;73) \), the composition of which is believed to affect the risk of several life-style related disorders including obesity and type-2 diabetes \( (39;59) \). Increased prevalence of childhood obesity is currently a major societal concern due to the high association with adult obesity \( (45;82) \). It has also been proposed that obesity in adults may be related to the capacity of the gut microbiota to harvest energy through breakdown of indigestible polysaccharides \( (80;81) \). Moreover, obesity and insulin resistance are often accompanied by a state of chronic low-grade inflammation \( (31) \). Since species-specific bacterial surface markers are involved in regulation of inflammation, differences in commensal bacterial composition between individuals may differentially predispose to inflammatory-induced diseases \( (78) \). Breastfed infants are leaner than formula fed counterparts \( (25;60) \), and display lower incidences of obesity, diabetes and inflammatory bowel diseases later in life \( (46) \), and a link between diet, infant gut microbiota, obesity development and inflammatory pathways, has thus been suggested \( (71) \). The discovery of the existence of so-called gut enterotypes in adult human subjects has received considerable attention, although final consensus on the number and characteristics of the enterotypes has not yet been achieved \( (4;44) \). In this context, it has recently been proposed by us that adults can be grouped into two distinct groups based on the levels of Bacteroides spp. and particularly Prevotella spp., which are bimodally...
distributed in adult subjects and remain stable over a time-period of at least 6 months, and that these two genera can therefore serve as ‘markers’ for the corresponding enterotypes (68). Each adult individual is estimated to harbor approximately 160 different high-abundant species, many of which are expected to be shared between individuals, but present in very different levels in different individuals (64). Looking at the total pool of species present, earlier studies have suggested that as little as 1% may be shared with another individual (13;61). Therefore, some redundancy of colonic bacterial processing would be expected to exist, and it has been demonstrated that within the intestinal metagenome, the phylogenetic variation is much more pronounced than the variation in functional capacities of the ecosystem (33). While interdependence and recognizable patterns of gut bacterial colonization in the intestinal ecosystem exist (20;24;43;79) and are known to be influenced e.g. by dietary differences (41), only few studies have combined the longitudinal development of representative bacterial taxa with parallel development of growth, body composition and nutritional parameters in early life.

The aims of the present study were to describe patterns of microbial establishment during the first three years of life, and to identify putative correlations of such patterns with dietary habits and physiological parameters, particularly focusing on development of bodyweight. We used the previously published quantitative PCR–based GUt Low Density Array (GULDA) (7) to determine relative abundances and interbacterial relations of 31 different bacterial 16S rRNA gene targets, representing different phylogenetic levels, in fecal samples from a longitudinal cohort study of approximately 300 healthy Danish infants sampled at 9, 18 and 36 months after birth. Results were correlated to measures of growth, body composition and nutritional records obtained for the same children, in order to reveal possible associations to the microbial composition and temporal development. Additionally, we performed the first comprehensive analysis of establishment of gut enterotypes (4;44) in early life, using the Prevotella spp./Bacteroides spp. ratio as a proxy for the enterotypes driven by the abundance of these genera.
**Materials and Methods**

**The SKOT cohort**

The present study is based on data and samples collected during an observational cohort study of approximately 300 apparently healthy Danish singleton term infants. The cohort, entitled ‘SKOT’ based on a Danish abbreviation, was followed for a period of 3 years with visits at 9, 18 and 36 months of age and several papers based on the cohort have been published (15;28;34;40;52;53). The study protocol was approved by The Committees on Biomedical Research Ethics for the Capital Region of Denmark (H-KF-2007-0003). Fecal samples and information on birth mode, gender, measures of growth and body composition, food questionnaires and background interviews, were collected at the visits during the study. The numbers of participants completing the 9, 18 and 36 month visits were 311, 290 and 264 respectively.

**DNA extraction and qPCR analysis**

Total community DNA was extracted from a total of 698 fecal samples on the Maxwell® 16 system using the Maxwell® 16 DNA Tissue DNA purification kit (Promega Biotech AB, Sweden). The DNA concentrations were determined fluorometrically (Qubit® dsDNA HS assay; Invitrogen) and adjusted to 1 ng/ul prior to use as template in qPCR. The qPCR analysis was performed using the GULDA platform previously described (7). Briefly, each 384-well PCR-plate accommodated simultaneous analysis of four DNA samples to determine the relative abundance of 31 bacterial selected 16S rRNA gene targets (Table 1) representing different phylogenetic levels. A universal bacterial primer set was included as reference gene. All qPCR reactions were performed in duplicate in transparent 384-well MicroAmp® Optical reaction plates (Applied Biosystems) sealed with MicroAmp® Optical Adhesive Film on an ABI prism 7900HT (Applied Biosystems, Nærum, Denmark). Following
the thermocycling program, the raw fluorescence data recorded by the SDS software were exported to the LinRegPCR program (66:70). This software was used to perform baseline correction and calculate the mean PCR efficiency per amplicon group. This was used to calculate the initial quantities $N_0$ (arbitrary fluorescence units) for each amplicon using the formula $N_0 = \text{threshold} / (\text{Eff}^{\text{Ct}_0}_{\text{mean}})$, where $\text{Eff}_{\text{mean}}$ denotes the mean PCR efficiency per amplicon, threshold is the optimal ‘cutoff’ in the exponential region, and $C_t$ is the cycle number, where each sample exceeds this threshold. The relative abundances of the 31 specific amplicon groups were obtained by normalization to the $N_0$-value obtained for the universal bacterial amplicon group determined in the same array (Figure S1).

A universal limit of detection (LODU) $= 10^{-5} (N_0, \text{specific} / N_0, \text{universal})$ was applied to the normalized $N_0$-values due to qPCR analysis limitations. LODU was set to this value based on previous results using the GULDA setup and roughly correspond to target $C_t$ values above 30 cycles. All normalized $N_0$-values equal to or above LODU were included in the analysis, while samples below LODU were set to $0.5\text{LODU}$. Results were calculated as the arithmetic mean of normalized $N_0$-values of the two technical repeats. Differences between $C_t$ values of technical replicates were typically less than 0.5. For samples, where qPCR was successful for only one of the replicates, this value was used. Water as template was used as negative control. For 40 samples, no PCR amplification was detectable using the universal bacterial primer and thus the final number of fecal DNA samples for gut microbiota analysis was 658 representing 218, 232, and 208 at 9, 18 and 36 months, respectively. For 132 subjects samples from all three time-points were obtained. Note that a small subset of the data (obtained from 6 infants at two time points), have previously been published in order to illustrate the applicability of the PCR-based GULDA platform (7)

Parameters of nutrition, growth and body composition

Information on duration of breast feeding and level of iron supplementation was obtained from background interviews (Table S1). Parameters of infant diet (Table S2) were estimated from parent-completed pre-coded
dietary records over seven consecutive days (27). Dietary intake was calculated with GIES software (version 1.000d, The National Food Institute, DTU food, Soborg, Denmark). Based on anthropometric measures described previously (34;53) age- and gender-specific Z-scores at birth, 9, 18 and 36 months (Table S3) were calculated by the WHO Anthro software (Department of Nutrition, World Health Organization, Geneva, Switzerland). At 3 years, body composition was estimated by both Dual-energy X-ray absorptiometry (DXA) and Bioelectrical impedance (BIA). BIA is a simple method for measuring body composition (Houtkooper et al, 1996) and whole body resistance, reactance and impedance were measured using a single frequency (50 kHz) tetra polar BIA (Quantum III, RJL Systems, Michigan, USA) between right hand and right foot. Whole body DXA scans were performed in a subgroup of the SKOT children (n = 101) with a Lunar Prodigy Advance (GE Healthcare, Madison, WI, USA) using the software enCore, version 12.30 (procedure described in detailed by Jensen et al. 2012). Here DXA Fat Free Mass (FFM) and Fat Mass (FM), resistance index (height^2/resistance) and FFM and FM predicted from BIA were all used as measures for body composition at 3 years.

Assessment of changes in the gut microbial composition

The normalized N0-values obtained for each bacterial taxon were log10-transformed and used as input for multivariate principal component analysis (PCA) using LATENTIX version 2.11 (Latent5 Aps, Frederiksberg, Denmark). Univariate statistical analysis was performed using the GRAPHPAD PRISM software (version 5.03; GraphPad Software Inc., La Jolla, CA). Only individuals for which samples at all three examinations were available, were included. Specific primer results, which never exceeded LODU at neither 9, 18 and 36 months for a given individual, were excluded from further analysis. Consequently, the number of individuals for analysis of each bacterial taxon ranged from N=25 to N=132. Fold-changes (FC) for specific gene targets were calculated as the pairwise (log2) ratio of normalized, but not log10 transformed, abundances at 9, 18 and 36 months, giving three ratios: 18m/9m, 36m/9m and 36m/18m. Mean and corresponding SEM values were calculated and a one-sample t-test was performed to test if the fold changes differed significantly from zero.
Wilcoxon signed rank-sum test was performed as an alternative when data were not normally distributed. Correction for multiple testing, (6) was performed for the 90 comparisons (3x30) presented in Figure 2. Three statistical significance levels were employed: p<0.05, p<0.01, p<0.001. Spearman correlations R and corresponding p values between bacterial fold changes from 9 to 18 months and from 18 to 36 months, respectively, and changes in parameters of growth and body composition and nutritional parameters (Supplementary Tables 1, 2 and 3) were calculated using GraphPad (Table 2). The Mann-Whitney test was used to compare effects of continued or terminated breastfeeding at the 9 months examination on the relative abundances of all bacteria at 9, 18 and 36 months, respectively.

Spearman correlations at 9, 18 and 36 months

For the time-independent analysis, all valid samples for each of the three time points, 9, 18 and 36 months were included (N>200 in each age-group). Pairwise correlations between all measured SKOT parameters, including the microbiota, at all three time points were performed using GraphPad. Spearman correlations were applied for all pairwise correlations. False discovery rates (FDR) were calculated by a classical one-stage method (6).

The ratio between relative abundances of the *Prevotella* spp. (B9) and *Bacteroides* spp.(B3) targets were calculated as a proxy for the corresponding *Prevotella-* or *Bacteroides*-driven gut enterotypes (4). The logged relative abundances of *Bacteroides* spp. and *Prevotella* spp., frequency distributions of *Bacteroides* spp., *Prevotella* spp. and corresponding ratio (P/B) were performed on N=69, 84 and 130 individuals, at 9, 18 and 36 months, respectively. A Kernel density plot was fitted to all histograms using the Kernel add-in package for Microsoft Excel. Characterization of the frequency distributions as uni- or bimodal was tested with a Dip Test, calculated by the R package diptest (R package version 0.75-4, based on Fortran and S-plus from Dario Ringach, NYU.edu).
Samples from the N=79 individuals giving qPCR results for both B3 (Bacteroides spp.) and B9 (Prevotella spp.) at both 18 and 36 months were stratified as either high or low P/B enterotypes and investigated for their putative co-occurrence with specifically high or low levels of physiological parameters (nutrition, growth or body composition). Finally, the fold change from 18 to 36 months using the same stratification was correlated to the longitudinal development of all parameters of nutrition, growth and body composition from 18 to 36 months.

Results and Discussion

Development of the gut microbiota

Although a quite extensive amount of literature on the possible factors involved in microbiota development in early life exist, (for reviews, see (72) and (23)), these studies typically focus on microbial colonization immediately after birth (9;19), during weaning at 4-6 months (17;56) or up to 1 (61) and in a single recent study 2 years of age (5). To our knowledge, no previous studies including as high numbers of participants as in the present study have focused on the development occurring in the microbiota between infancy and 3 years of age.

We observed a clear change in the microbiota during this period, in particular from age 9 to 18 months (Figure 1). The 9 months samples appeared to cluster less closely together than the later samples, and were characterized by more lactic acid bacteria and enterobacteria than seen for samples taken at ages 18 and 36 months. The two other age groups comprised a higher number of different microbes including both Firmicutes and Bacteroidetes. In line with this, the majority of specific changes in abundances of given bacterial taxa occurred between 9 and 18 months (Figure 2; Figure S2). We observed a consistent and significant increase of several species within the Bacteroidetes phylum, which is consistent with reported findings seen after...
introduction of complementary feeding (14;41). Additionally, we observed a significant decrease in the relative abundance of *Bifidobacterium* spp. (FC -2.56; \( p<0.001 \)). Within these, *B. longum* (FC -2.77; \( p<0.001 \)) and *B. breve* (FC -6.47; \( p<0.001 \)) were observed to decrease, while *B. adolescentis* (FC 2.33; \( p<0.01 \)) and *B. catenulatum* (FC 4.00; \( p<0.001 \)) increased during the period from 9 to 36 month of age, indicating that during early childhood, the conditions in the gut and/or the diet changes in ways that favor the latter species of bifidobacteria later in life. For example, breast milk is known to contain bifidogenic human milk oligosaccharides (HMO’s), which are atypical carbohydrates, resistant to enzymatic hydrolysis in the upper gastrointestinal tract (16;26). *B. longum, B. bifidum* and *B. breve* are particularly abundant in breast-fed children (55) and known to be highly proficient in capturing and utilizing HMO’s as their sole carbon source, while *B. adolescentis* is unable to degrade these oligosaccharides (74).

*Lactobacillus* spp. (FC -1.33; \( p<0.05 \)) and *Enterobacteriaceae* (FC=-4.21; \( p<0.001 \)) were found to decrease between 9 and 18 months (Figure 2), while an increase was observed for the butyrate producing taxa *C. leptum* group (FC 2.32;\( p<0.001 \)), *E. hallii* (FC 3.65; \( p<0.001 \)) and *Roseburia* spp. (FC 4.62; \( p<0.001 \)) from age 9 to 36 months. This is in accordance with previous findings seen at cessation of breastfeeding and introduction of formula feeding and/or cow milk (2;8;14;18;46;62;69;76). Conversely, the butyrate producing *C. coccoides* group (FC -3.14; \( p<0.001 \)) was seen to be reduced between 9 and 18 months. A previous report (32), based on a cross-sectional study of 40 children, indicates that *C. coccoides* group increases until 6 months of age, and thereafter remains at a stable level.

The fact that we observed significant changes still occurring from 18 to 36 months (Figure 1 and 2), suggests that convergence towards adult-like stability, characterized by high levels of Firmicutes and Bacteroidetes and smaller fractions of actinobacteria, proteobacteria and verrucomicrobia (3;49;77) was still occurring during this period. This is in line with a recent large cross-sectional study of humans from different age-groups showing...
that bacterial communities of the gut evolve towards adult-like configuration during the first three years of life (84), but contradicts older reports proposing that full stability is reached already at 12 months (21;41;61).

Correlations between relative abundance of bacterial groups and physiological parameters

In agreement with previous studies from other researchers (74), still breastfeeding at 9 months, was associated positively with high relative abundances of *Lactobacillus* spp., *Bifidobacterium* spp. and *B. longum* at 9 months (Figure 3). Additionally, when compared to infants no longer breastfed at 9 months, infants breastfed at 9 months had lower levels of a number of butyrate producing taxa including *C. leptum* group, *C. coccoides* group, *E. hallii*, and *Roseburia* spp. Breastfeeding at 9 months was also associated with lower levels of *Desulfovibrio* spp. and *A. muciniphila*, as well as of the *Bacteroides* phylum and several taxa herein. For the *C. coccoides* group and some of the *Bacteroides* species, the differences were still present after 18 months, while at 36 months the breastfeeding history did no longer influence the microbiota (Figure 3). Breastfeeding has been shown to significantly reduce risk of overweight/obesity later in childhood as well as in adult life (29), however conflicting reports of the role of breastfeeding on obesity also exist (48). We speculate that the observation that continued breastfeeding at 9 months delays progression of specific bacterial taxa may be of relevance for development and later life health.

We observed only very few correlations between the abundances of specific gut bacteria and the physiological parameters measured (Figure S3). Significant associations with p< 0.001 and false discovery rates (q) below 0.08 were observed at 9 months, where duration of breastfeeding (Breast Milk Days, indicating the number of days with either partial or exclusive breastfeeding as estimated by the mothers ) was shown to correlate positively with *Lactobacillus* spp. and *Bifidobacterium* spp. targets. Less significantly (p<0.01, q<0.32), negative associations to duration of breastfeeding were seen for *C. leptum* group, *E. hallii*, *Roseburia* spp., *Bacteroides/Prevotella* groups, *B. fragilis*, *B. vulgatus*, *Desulfovibrio* spp. and *A. muciniphila*.
Supporting these observations, many of the opposite trends for correlations were observed between these bacterial targets and the intake of infant formula, including a negative correlation (p<0.001, q<0.08) to the abundance of *Lactobacillus* spp. Additionally, the data reflected that duration of breast milk consumption (Breast Milk Days) was negatively associated to the overall energy intake, as previously reported (28). We have previously shown that the breastfed infants had lower BMI both at 9 and 18 months (52). No other significant correlations were found between the gut microbiota and nutritional parameters, or measures of growth and body composition (including DXA and BIA examinations at 36 months), gender or birth mode (vaginally vs. caesarean) at any of the three time points (data not shown). Previous reports on correlations between BMI and gut microbiota composition in young children exist (37;38), however these were observed in cross-sectional cohort studies where sampling was focused on high vs. normal BMI, while the infants included in the present study constituted a younger and leaner population, with only 8% classified as overweight (11).

Exploiting the longitudinal observations from the present study, we found significant (p<0.05) positive correlations between increase of Body Mass Index (ΔBMI) and the increase of the *Firmicutes* phylum, the *C. leptum* group and *E. halii* (belonging to the *C. coccoides* group) between 9 and 18 months (Table 2). Additionally, increases in *M. smithii* were negatively correlated to ΔBMI from 9 to 18 months, while reductions in *Enterobacteriaceae* were associated with higher ΔBMI from 18 to 36 months. Similar results were obtained for ΔBMI-for-age-Z-score (ΔBAZ) and ΔWeight-for-length Z-score (ΔWFL), but not for changes in Z-scores for Weight-for-age (WAZ), Length-for-age (HAZ), Subscapular skinfold for age (SSZ) or Triceps skinfold for age (TSZ) (data not shown). No other significant correlations between changes in body compositional measures and changes in nutritional parameters and/or bacterial targets were observed. Although excessive weight gain during the first six months after birth has been shown to be particularly predictive of later obesity (10;22;58;85), we found no significant characteristics in the microbiota after 9 months, which corresponded to changes in BAZ or WAZ between birth and 9 months (data not shown).
The clostridial targets selected for this study represent colonic butyrate-producing bacteria, assisting in the conversion of polysaccharides to monosaccharides and short-chain fatty acids (SCFA) constituting energy to the host (50;63). The development of abundances of the selected taxa *C. butyricum*, *C. leptum* group, *C. coccoides* group, *E. halli* and *Roseburia* spp. (both of the latter taxa belonging to the *C. coccoides* group), was very different for the different targets (Figure 2). Significant differences in carbohydrate metabolism and butyrate production between different clostridia have been demonstrated to be of relevance to the pathogenesis of obesity (12). It remains to be established whether certain dietary compounds, arguably containing high concentrations of specific complex polysaccharides, is specifically subjected to catabolism by the *C. leptum* group, *E. halli* and *Roseburia* spp. between 9 and 18 months, as suggested by the data (Figure 2).

**Correlation between bacterial groups at 9, 18 and 36 months**

Correlations between the abundances of the bacterial 16S rRNA gene targets were investigated for each of the three age groups (Figure 4). Although there were more differences between the 9 months pattern and the two later groups, differences between 18 and 36 months were also observed.

At 9 months, we found co-abundance (positive Spearman correlations) between the butyrate producing Firmicutes *C. leptum* group, *C. coccoides* group and *C. butyricum*. Similarly, co-abundance was seen for many of the *Bacteroides* species, with the notable exception of *B. eggerthii*. With this specific exception, a high abundance of Firmicutes was also clearly associated with a low abundance of *Bacteroides* species. While the other *Bacteroides* species were increasing in abundance during the study, *B. eggerthii* was reduced (Figure 2). The reverse co-occurrence of *B. eggerthii* and the other *Bacteroides* species was no longer as significant at 36 months (Figure 4), indicating that later in life, the environment in the gut no longer represses this organism more than the other *Bacteroides*. Additionally, the clear association between a high abundance of Firmicutes...
and enterococci, and a low abundance of *Bacteroides* spp. seen at 9 months, has disappeared at 18 and 36 months, indicating that it is particularly in infancy that these groups are mutually exclusive to each other.

*Bacteroides* spp. and *Clostridiales* spp. have previously been seen to co-cluster in the so-called enterotype driven by abundance of *Bacteroides* spp. in healthy human adults (4). This is in agreement with our data from age 18 months, and becomes even clearer at 36 months, however at 9 months we found no co-abundance of these groups (Figure 4). It seems plausible that the development of this co-abundance is a logical consequence of adaptation to a Western-type diet after weaning. However, analysis of possible correlations between bacterial targets and the investigated nutritional parameters (Table S2) did not result in statistically significant co-occurrences (data not shown).

We found it noteworthy that at 9 months, where *Bifidobacterium* spp. in general were most abundant (Figure 2), there was a clear co-occurrence of particularly *B. longum* with the other lactic acid producing taxa *Lactobacillus* spp. and *Enterococcus* spp, however no co-occurrence between specific species of *Bifidobacterium* was seen (Figure 4). This pattern was reversed at 18 and 36 months, where the co-occurrence with lactobacilli and enterococci was no longer present, while co-occurrence of the specific *Bifidobacterium* species *B. longum* with *B. bifidum*, *B. adolescentis* and *B. catenulatum*, and further of *B. bifidum* with *B. catenulatum*, and *B. breve* was evident. Interestingly, these co-occurrences existed independently of the fact that the average abundance of certain *Bifidobacterium* species was increased during the experimental period, while others were reduced as discussed above (Figure 2).

‘Enterotype’ development in the infant gut

Lately, the existence of three distinct ‘enterotypes’, driven by the abundance of *Bacteroides* spp., *Prevotella* spp. and *Ruminococcus* spp., respectively, has been given particular attention (4). While evidence is mounting in support of the distinction between the *Bacteroides*- and *Prevotella*- driven groupings, the existence of the
third group, driven by abundance of *Ruminococcus* spp., is not as clearly supported (83). In the present study we used the relative abundance between *Prevotella* spp. and *Bacteroides* spp., measured as the ratio *Prevotella/Bacteroides* (P/B) as a proxy for the enterotypes driven by these two genera as proposed by Arumugam (4). It is important to note that we do not mean to propose that the enterotypes are characterized solely by the abundance of these taxa, but merely that their abundance can be used as a marker for more complex differences characterizing these two types of intestinal bacterial communities. For the first time, this approach allowed addressing the establishment of enterotypes during infancy.

In agreement with the existence of enterotypes, we observed a negative correlation between *Prevotella* spp. and *Bacteroides* spp. at 36 months age, but not at 9 or 18 months (Figure 4). The relative abundances of both *Prevotella* spp. and *Bacteroides* spp. are below detection limits at birth, given the absence of these bacteria in the prenatal environment of the maternal uterus (72). We propose that *Bacteroides* spp. colonize better than *Prevotella* spp., between birth and 9 months (35), as almost all individuals were characterized by a low P/B ratio at 9 months of age (Figure 5A). At 18 months, a smaller subset of individuals established a higher P/B (Figure 5B), a pattern which was even more pronounced after 36 months, where two distinct groups appeared (Figure 5C). We observed a unimodal distribution of *Bacteroides* spp. abundances at all three sampling points (Figure 5D-F), while an increasingly bimodal pattern of *Prevotella* spp. abundances (Figure 5G-I) developed from 9 to 36 months. The ratio of logged P/B values similarly showed an increasingly bimodal pattern with age, with high P/B samples characterized by a logged ratio above a level of -2 (P/B>0.01) and low P/B samples below -2 (P/B<0.01) (Figure 5J-K). When tested statistically for presence of bimodal distribution, there was significance at the 36 months *Prevotella* spp. abundance (P=0.004) (Figure 5I) and an even clearer bimodality for the 36 months P/B ratio (P=6.6E-5) (Figure 5L), but not at the earlier time points. Consequently, development of the two enterotypes starts between 18 and 36 months and is driven by changes in *Prevotella* spp., rather than *Bacteroides* spp. However, stratification of samples into P/B types at either 18 or 36 months
did not reveal any additional correlations between bacterial 16S rRNA gene targets and host physiological phenotypes.

Out of the 79 individuals for whom the P/B-ratio could be calculated for both 18 and 36 months, 70% remained in the same P/B group between 18 and 36 months, while 18% and 11%, respectively, shifted their ‘enterotypes’ from low P/B to high P/B or vice versa (Figure 6). Previous studies have shown that although the adult gut microbiota of any individual is quite resilient to major perturbations, enterotypes may shift in a few individuals when measured over longer periods of time (57;65) and by long-term dietary intervention schemes (83). In the present study, we did not identify any correlation of enterotype to diet or BMI. No data on antibiotic treatment were collected for the cohort, however since antibiotic treatment is known to induce major changes in the composition of the infant gut microbiota (1;62;67), it cannot be excluded that antibiotic usage prior to 18 or 36 months samplings influenced the observed shifts. Nevertheless, our results (Figure 6) strongly support the notion that between 18 and 36 months, the ‘P/B-enterotype’ is still not as stably established as reported in adults (68). In light of recent findings showing that enterotypes driven by Bacteroides spp. and Prevotella spp. affect risk markers for artherosclerosis (42), and that gut microbiome composition correlates with obesity and metabolic markers in adults (47), we propose that the establishment of microbiota during infancy may affect health status in adult life.

Concluding remarks

We have studied the establishment of intestinal microbiota in a large cohort of Danish infants, and analysed the microbial data in relation to a vast amount of dietary and physiological measures. We demonstrate significant differences in microbiota composition between infants either breastfed or no longer breastfed at 9 months, but additionally show that the effects of breastfeeding on the microbiota are no longer prevalent at age 36 months. Positive correlations between increases in BMI, C. leptum group and E. halii were observed.
from 9 to 18 months, indicating that these butyrate producing groups may contribute importantly to host energy harvest. Additionally, we show for the first time that human enterotypes, expressed as a bimodal distribution of the \textit{Prevotella} spp./\textit{Bacteroides} spp. ratio, starts establishing between 18 and 36 months of age. In this period, where we observe an ongoing development of the microbiota towards an adult-like composition, enterotypes are still more susceptible to shifting than previously seen for adults.

Considering the increasing evidence supporting a key role of gut microbiota composition in human health, the presented data constitutes an important new body of knowledge on microbiota development during infancy, which is likely to constitute a window where the microbiota can be more significantly influenced by intervention. In this context, the current development in next generation sequencing is expected to contribute importantly to our understanding of human microbiome establishment during the coming years.

\section*{Acknowledgements}

The authors would like to thank The Danish Council for Independent Research for financial support (Grant no. 10-093725/FTP). The SKOT study was supported by The Danish Directorate for Food, Fisheries and Agriculture (Grant no. 3304-FSE-06-0503). We thank Bodil Madsen and Vivian Julia Anker for excellent technical assistance with qPCR and DNA purification, respectively. Additionally, we thank Inge Tetens, Ellen Trolle, Ulla Gondolf, and Majken Ege for collecting the nutritional data, and Louise Beltoft Borup Andersen for providing the processed nutritional data, while Line Rieck Schmidt, Sara Dyhrberg and Louise Nissen-Schmidt are acknowledged for additional experimental assistance.

\section*{References}


Figure 1: Principal component analysis (PCA) of the GULDA microbiota.
Upper plot: Scores (Individuals); Lower plot: Loadings (Bacterial 16S rRNA gene targets). This figure shows the two primary principal components, PC1 and PC2, which explain 18.02% and 11.29% of data variation, respectively. Bacterial targets primarily associated with the lower left quadrant and thus relatively highly abundant in the 9 months samples were *Enterococcus spp.* (F10), *Enterobacteriaceae* (P1) and *E. coli* (P2) and to a lesser extent *B. breve* (A6) and *Lactobacillus spp.* (F2). Bacterial targets appearing in higher abundances in the 36 months samples were the Bacteroidetes (B1), the *Bacteroides-Prevotella* group (B2), *Bacteroides spp.* (B3), *B. fragilis* group (B4), *B. vulgatus* (B5), *B. thetaiotaomicron* (B6), *Alistipes spp.* (B10), *A. muciniphila* (V1) and *Desulfovibrio spp.* (P3). Complete explanations for all labels are given in Table 1.

Figure 2: Progressive development of the gut microbial composition.
Log$_2$ (Fold changes) of microbial 16S rRNA gene targets occurring from 9 to 18 months (white), 18 to 36 months (grey), and cumulative from 9 to 36 months (black). Statistical significance of one-sided t-tests; *p<0.05, **p<0.01, ***p<0.001. Data were corrected for multiple testing, using a maximal false discovery rate of 5%.

Figure 3: Effect of breastfeeding on infant gut microbiota.
P values of Mann-Whitney statistical tests addressing differences between the relative bacterial abundances at 9, 18 and 36 months, dependent on whether or not the infants were still breastfed at the 9 months examination. Green colour indicates an increase in children breastfed at 9 months and green indicates a corresponding decrease.
Figure 4: Spearman pairwise correlation map of measured bacterial 16S rRNA gene targets at 9, 18, and 36 months. Because each time point was analyzed separately, the included number of individuals was >200 at each of the three samplings. Color gradient denotes Spearman R values. Dots indicate significant correlations, corrected for false discovery rates (q).

Figure 5: ‘Enterotype’ defined as P/B ratio development from 9 to 36 months. Relative abundances of log (Bacteroides spp.) and log (Prevotella spp.) show a distinct development from 9 to 36 months (A-C), moving from low relative abundances of both groups (blue circle) to a Bacteroides prevalent microbiota (red circle) at 9 months. From 9 to 36 months, an increasing subgroup of Prevotella prevalent samples appear (green circle), indicating segregation of specific individuals from a Bacteroides-driven into a Prevotella-driven enterotype. This progressive development is also evident from the corresponding histograms of frequency distributions of log (Bacteroides spp.) (D-F) and log (Prevotella spp.) (G-I) abundance, and particularly from the distributions of the logged P/B-ratio (J-L). The dotted curve in Panels D-L shows a Kernel density plot, which is a modification of the histogram patterns, supporting the underlying statistical distributions found. Panels A, D, G, J (9 months) represent 69 individuals, Panels B, E, H, K (18 months) represent 84 individuals and Panels C, F, I, L (36 months) represent 130 individuals, as only individuals with relative abundances of Bacteroides spp. and Prevotella spp. exceeding the detection limit were included.

Figure 6: Changes in P/B ratio occurring between age 18 and 36 months. The 79 individuals giving qPCR results for both 18 (blue) and 36 (green) months, sorted after increasing logged P/B ratios at 36 months. Samples above the dotted line belong to the high P/B group (Prevotella-driven ‘enterotype’), while samples below this line belong to the low P/B group (Bacteroides-driven ‘enterotype’).
48/79 individuals remained in the low P/B group, while 8/79 individuals remained in the high P/B group from age 18 to age 36 months. 14/79 and 9/79 individuals shifted from low to high P/B, or from high to low P/B, respectively.
<table>
<thead>
<tr>
<th>Amplicon ID</th>
<th>Phylum</th>
<th>Class/Family/Genus</th>
<th>Species/Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>Universal (all phyla)</td>
<td>Universal</td>
<td>Universal</td>
</tr>
<tr>
<td>F1b</td>
<td>Firmicutes</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>F1</td>
<td>Firmicutes</td>
<td><em>Lactobacillus</em></td>
<td><em>spp.</em></td>
</tr>
<tr>
<td>F3</td>
<td>Firmicutes</td>
<td><em>Lactobacillus</em></td>
<td><em>plantarum</em></td>
</tr>
<tr>
<td>F4</td>
<td>Firmicutes</td>
<td><em>Lactobacillus</em></td>
<td><em>acidophilus</em></td>
</tr>
<tr>
<td>F5</td>
<td>Firmicutes</td>
<td><em>Clostridium</em></td>
<td><em>butyricum</em></td>
</tr>
<tr>
<td>F6</td>
<td>Firmicutes</td>
<td><em>clostridia</em></td>
<td>Cluster IV (<em>C. leptum group</em>)</td>
</tr>
<tr>
<td>F7</td>
<td>Firmicutes</td>
<td><em>clostridia</em></td>
<td>Cluster XIVa (<em>C. coccoides-E. rectale group</em>)</td>
</tr>
<tr>
<td>F8</td>
<td>Firmicutes</td>
<td><em>Eubacterium</em></td>
<td><em>hallii</em></td>
</tr>
<tr>
<td>F9</td>
<td>Firmicutes</td>
<td><em>Roseburia</em></td>
<td><em>spp.</em></td>
</tr>
<tr>
<td>F10</td>
<td>Firmicutes</td>
<td><em>Enterococcus</em></td>
<td><em>spp.</em></td>
</tr>
<tr>
<td>B1</td>
<td>Bacteroidetes</td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>B2</td>
<td>Bacteroidetes</td>
<td><em>Bacteroides-Prevotella</em></td>
<td><em>spp.</em></td>
</tr>
<tr>
<td>B3</td>
<td>Bacteroidetes</td>
<td><em>Bacteroides</em></td>
<td><em>spp.</em></td>
</tr>
<tr>
<td>B4</td>
<td>Bacteroidetes</td>
<td><em>Bacteroides</em></td>
<td>fragilis group</td>
</tr>
<tr>
<td>B5</td>
<td>Bacteroidetes</td>
<td><em>Bacteroides</em></td>
<td><em>vulgatus</em></td>
</tr>
<tr>
<td>B6</td>
<td>Bacteroidetes</td>
<td><em>Bacteroides</em></td>
<td><em>thetaiotaomicron</em></td>
</tr>
<tr>
<td>B7</td>
<td>Bacteroidetes</td>
<td><em>Bacteroides</em></td>
<td><em>eggerthii</em></td>
</tr>
<tr>
<td>B8</td>
<td>Bacteroidetes</td>
<td><em>Bacteroides</em></td>
<td><em>distasonis</em></td>
</tr>
<tr>
<td>B9</td>
<td>Bacteroidetes</td>
<td><em>Prevotella</em></td>
<td><em>spp.</em></td>
</tr>
<tr>
<td>B10</td>
<td>Bacteroidetes</td>
<td><em>Alistipes</em></td>
<td><em>spp.</em></td>
</tr>
<tr>
<td>A1b</td>
<td>Actinobacteria</td>
<td><em>Bifidobacterium</em></td>
<td><em>spp.</em></td>
</tr>
<tr>
<td>A2</td>
<td>Actinobacteria</td>
<td><em>Bifidobacterium</em></td>
<td><em>bifidum</em></td>
</tr>
<tr>
<td>A3</td>
<td>Actinobacteria</td>
<td><em>Bifidobacterium</em></td>
<td><em>adolescentis</em></td>
</tr>
<tr>
<td>A4</td>
<td>Actinobacteria</td>
<td><em>Bifidobacterium</em></td>
<td><em>catenulatum/pseudocatenulatum</em></td>
</tr>
<tr>
<td>A5</td>
<td>Actinobacteria</td>
<td><em>Bifidobacterium</em></td>
<td><em>longum</em></td>
</tr>
<tr>
<td>A6</td>
<td>Actinobacteria</td>
<td><em>Bifidobacterium</em></td>
<td><em>breve</em></td>
</tr>
<tr>
<td>P1</td>
<td>Proteobacteria</td>
<td><em>Enterobacteriaceae</em></td>
<td>Group</td>
</tr>
<tr>
<td>P2</td>
<td>Proteobacteria</td>
<td><em>Escherichia</em></td>
<td><em>coli</em></td>
</tr>
<tr>
<td>P3</td>
<td>Proteobacteria</td>
<td><em>Desulfovibrio</em></td>
<td><em>spp.</em></td>
</tr>
<tr>
<td>V1</td>
<td>Verrucomicrobia</td>
<td><em>Akkermansia</em></td>
<td><em>muciniphila</em></td>
</tr>
<tr>
<td>E1</td>
<td>Euryarchaeota</td>
<td><em>Methanobrevibacter</em></td>
<td><em>smithii</em></td>
</tr>
</tbody>
</table>

**Table 1:** 16S rRNA gene targets included on the Gut Low Density Array, ‘GULDA’ (7).
<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>p</th>
<th>R</th>
<th>N</th>
<th>p</th>
<th>R</th>
<th>N</th>
<th>p</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Firmicutes (F1b)</strong></td>
<td>132</td>
<td>0.02*</td>
<td>0.20</td>
<td>NS</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>C. leptum group (F6)</strong></td>
<td>132</td>
<td>0.02*</td>
<td>0.21</td>
<td>NS</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>E. hallii (F8)</strong></td>
<td>131</td>
<td>0.03*</td>
<td>0.19</td>
<td>NS</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Enterobacteriaceae (P1)</strong></td>
<td>57</td>
<td>0.16</td>
<td>-0.19</td>
<td>0.03*</td>
<td>-0.28</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>M. smithii (E1)</strong></td>
<td>25</td>
<td>0.04*</td>
<td>-0.42</td>
<td>NS</td>
<td>NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Spearman correlation analysis of the relative differences occurring in BMI from 9 to 18 months, 18 to 36 months, and 9 to 36 months (ΔBMI) with corresponding bacterial fold changes in the same period. Since fold change and Δ-calculations require valid measurements for each included individual at all three time points, the number of individuals (N) was different for each bacterial target. R designates the Spearman correlation coefficient. Only bacterial taxa with p-values (p) below 0.05 (*) are included in the table. NS: Not Significant (p>0.05).
PCA Scores and Loadings [Model 1]

Scores PC#1 (18.025%)
Scores PC#2 (11.295%)

Firmicutes
Bacteroidetes
Actinobacteria
Other bacteria

Loadings PC#1 (18.025%)
Loadings PC#2 (11.295%)

9 MONTHS
18 MONTHS
36 MONTHS
on September 23, 2017 by guest

http://aem.asm.org/
<table>
<thead>
<tr>
<th><strong>Firmicutes</strong></th>
<th>9 months</th>
<th>18 months</th>
<th>36 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firmicutes (F1b)</td>
<td>0.566</td>
<td>0.182</td>
<td>0.4265</td>
</tr>
<tr>
<td>Lactobacillus spp. (F2)</td>
<td>0.0001***</td>
<td>0.1736</td>
<td>0.4136</td>
</tr>
<tr>
<td>L. acidophilus (F4)</td>
<td>0.6249</td>
<td>0.172</td>
<td>0.9128</td>
</tr>
<tr>
<td>C. butyricum (F5)</td>
<td>0.8274</td>
<td>0.0561</td>
<td>0.6691</td>
</tr>
<tr>
<td>C. leptum group (F6)</td>
<td>0.0267*</td>
<td>0.4015</td>
<td>0.5951</td>
</tr>
<tr>
<td>C. coccoides group (F7)</td>
<td>0.0021**</td>
<td>0.0097**</td>
<td>0.8153</td>
</tr>
<tr>
<td>E. hali (F8)</td>
<td>0.0193*</td>
<td>0.4321</td>
<td>0.6329</td>
</tr>
<tr>
<td>Roseburia spp. (F9)</td>
<td>0.0316*</td>
<td>0.4482</td>
<td>0.6901</td>
</tr>
<tr>
<td>Enterococcus spp. (F10)</td>
<td>0.0731</td>
<td>0.0379*</td>
<td>0.5529</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Bacteroidetes</strong></th>
<th>9 months</th>
<th>18 months</th>
<th>36 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteroidetes (B1)</td>
<td>0.016*</td>
<td>0.243</td>
<td>0.2087</td>
</tr>
<tr>
<td>Bacteroides/Prevotella groups (B2)</td>
<td>0.0177*</td>
<td>0.0128*</td>
<td>0.1326</td>
</tr>
<tr>
<td>Bacteroides spp. (B3)</td>
<td>0.126</td>
<td>0.3384</td>
<td>0.0709</td>
</tr>
<tr>
<td>B. fragilis group (B4)</td>
<td>0.0004***</td>
<td>0.0984</td>
<td>0.8973</td>
</tr>
<tr>
<td>B. vulgatus (B5)</td>
<td>0.0345*</td>
<td>0.0115*</td>
<td>0.1967</td>
</tr>
<tr>
<td>B. thetataomicron (B6)</td>
<td>0.0016**</td>
<td>0.017*</td>
<td>0.3561</td>
</tr>
<tr>
<td>B. eggerthii (B7)</td>
<td>0.1241</td>
<td>0.9103</td>
<td>0.8546</td>
</tr>
<tr>
<td>B. distasonis (B8)</td>
<td>0.2906</td>
<td>0.1402</td>
<td>0.8114</td>
</tr>
<tr>
<td>Prevotella spp. (B9)</td>
<td>0.3372</td>
<td>0.1112</td>
<td>0.076</td>
</tr>
<tr>
<td>Alistipes spp. (B10)</td>
<td>0.1745</td>
<td>0.2494</td>
<td>0.3157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Bifidobacteria</strong></th>
<th>9 months</th>
<th>18 months</th>
<th>36 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bifidobacterium spp. (A1b)</td>
<td>0.0002***</td>
<td>0.51</td>
<td>0.5306</td>
</tr>
<tr>
<td>B. bifidum (A2)</td>
<td>0.6315</td>
<td>0.4082</td>
<td>0.8208</td>
</tr>
<tr>
<td>B. adolescentis (A3)</td>
<td>0.8294</td>
<td>0.9744</td>
<td>0.946</td>
</tr>
<tr>
<td>B. catenulatum (A4)</td>
<td>0.4334</td>
<td>0.4076</td>
<td>0.7047</td>
</tr>
<tr>
<td>B. longum (A5)</td>
<td>0.0477*</td>
<td>0.8546</td>
<td>0.2187</td>
</tr>
<tr>
<td>B. breve (A6)</td>
<td>0.7623</td>
<td>0.467</td>
<td>0.6985</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Other bacteria</strong></th>
<th>9 months</th>
<th>18 months</th>
<th>36 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterobacteriaceae (P1)</td>
<td>0.6179</td>
<td>0.6436</td>
<td>0.0523</td>
</tr>
<tr>
<td>E. coli (P2)</td>
<td>0.6053</td>
<td>0.555</td>
<td>0.1425</td>
</tr>
<tr>
<td>Desulfovibrio spp. (P3)</td>
<td>0.0449*</td>
<td>0.1072</td>
<td>0.9721</td>
</tr>
<tr>
<td>A. muciniphila (V1)</td>
<td>0.0451*</td>
<td>0.0863</td>
<td>0.811</td>
</tr>
<tr>
<td>M. smithii (E1)</td>
<td>0.0753</td>
<td>0.3251</td>
<td>0.6047</td>
</tr>
</tbody>
</table>

- **Increased in breastfed at 9 months**
- **Decreased in breastfed at 9 months**

*p<0.05; **p<0.01; ***p<0.001