

Journal Pre-proof

Development of the Microbiota and Associations With Birth Mode, Diet, and Atopic Disorders in a Longitudinal Analysis of Stool Samples, Collected From Infancy Through Early Childhood

Gianluca Galazzo, Niels van Best, Liene Bervoets, Isaac Oteng Dapaah, Paul H. Savelkoul, Mathias W. Hornef, the GI-MDH consortium, Susanne Lau, Eckard Hamelmann, John Penders, GI-MDH consortium, Eileen K. Hutton, Katherine Morrison, Alison C. Holloway, Helen McDonald, Elyanne M. Ratcliffe, Jennifer C. Stearns, Jonathan D. Schertzer, Mike G. Surette, Lehana Thabane, Monique Mommers

PII: S0016-5085(20)30113-X
DOI: <https://doi.org/10.1053/j.gastro.2020.01.024>
Reference: YGAST 63161

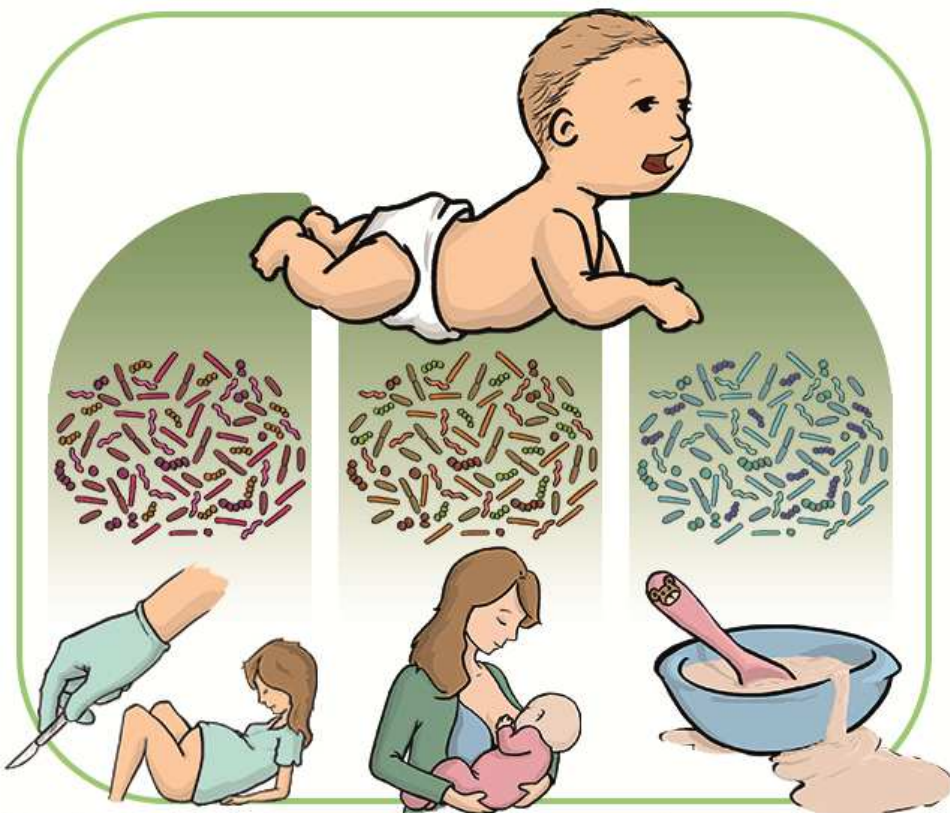
To appear in: *Gastroenterology*
Accepted Date: 9 January 2020

Please cite this article as: Galazzo G, van Best N, Bervoets L, Dapaah IO, Savelkoul PH, Hornef MW, the GI-MDH consortium, Lau S, Hamelmann E, Penders J, GI-MDH consortium, Hutton EK, Morrison K, Holloway AC, McDonald H, Ratcliffe EM, Stearns JC, Schertzer JD, Surette MG, Thabane L, Mommers M, Development of the Microbiota and Associations With Birth Mode, Diet, and Atopic Disorders in a Longitudinal Analysis of Stool Samples, Collected From Infancy Through Early Childhood, *Gastroenterology* (2020), doi: <https://doi.org/10.1053/j.gastro.2020.01.024>.

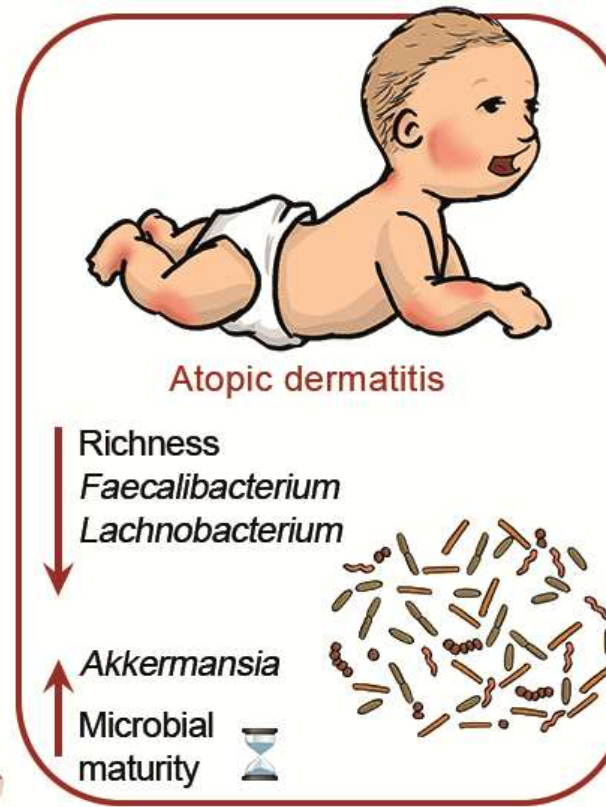
This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 by the AGA Institute

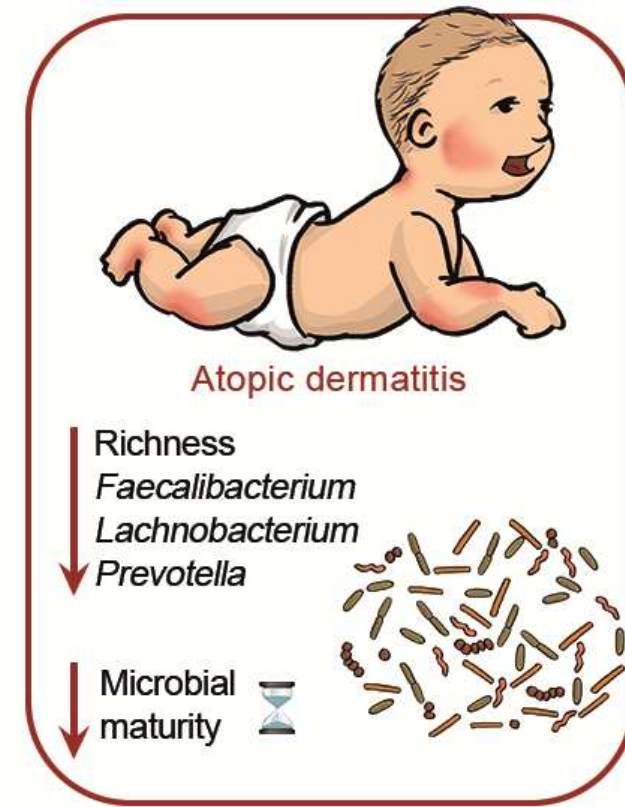




Determinants of microbiome development



First 100 days of life



100-200 days of life

Development of the Microbiota and Associations With Birth Mode, Diet, and Atopic Disorders in a Longitudinal Analysis of Stool Samples, Collected From Infancy Through Early Childhood

Gianluca Galazzo^{1,2*}, Niels van Best^{1,3,4*}, Liene Bervoets^{1,4}, Isaac Oteng Dapaah¹, Paul H. Savelkoul^{1,2,5}, Mathias W. Horne^{f3}, the GI-MDH consortium, Susanne Lau^{6‡}, Eckard Hamelmann^{7‡}, John Penders^{1,2,4‡}

¹School of Nutrition and Translational Research in Metabolism (NUTRIM), Department of Medical Microbiology, Maastricht University Medical Centre, Maastricht, the Netherlands.

²School for Public Health and Primary Care (Caphri), Department of Medical Microbiology, Maastricht University Medical Centre, Maastricht, the Netherlands.

³Institute of Medical Microbiology, RWTH University Hospital Aachen, RWTH University, Aachen, Germany

⁴*in Vivo* Planetary Health: an affiliate of the World Universities Network (WUN), West New York, New Jersey

⁵Department of Medical Microbiology and Infection Control, Amsterdam UMC, location VUmc, Amsterdam, The Netherlands.

⁶Department of Pediatric Pulmonology, Immunology and Intensive Care Medicine, Charité Universitätsmedizin Berlin, Germany.

⁷Children's Center Bethel, Protestant Hospital Bethel, University of Bielefeld, Germany.

*Authors share co-first authorship

‡Authors share co-senior authorship

Prof. Dr. Susanne Lau, M.D., Ph.D. Charité Campus Virchow, Klinik f. Pädiatrie m. S., Pneumologie, Immunologie und Intensivmedizin
Sektionsleitung Päd. Allergologie u. Pneumologie, Augustenburger Platz 1, 13353 Berlin, Germany;

Prof. Dr. Eckard Hamelmann, M.D., Ph.D. Medical Director, Children's Center Bethel
Grenzweg 10, D-33617 Bielefeld, Germany;

Assoc. Prof. Dr. John Penders, Ph.D. Department of Medical Microbiology, Maastricht University Medical Centre+, PO 5800, 6202 AZ, Maastricht, The Netherlands.

Author contributions

S.L. and E.H. designed the study and supervised participant recruitment and diagnosis, sample collection and generation of metadata. J.P., N.v.B. and G.G. drafted the data analyses plan. L.B., P.H.S., M.W.H., S.L. and E.H. provided critical input on the analyses plan.

G.G. and I.O.D. processed the data and performed the analysis. J.P., N.v.B. and G.G. interpreted the data and generated the figures and wrote the paper. All authors critically revised and approved the final manuscript.

Members of the GI-MDH consortium are listed below.

GI-MDH consortium

Collaborators: Eileen K. Hutton¹, Katherine Morrison^{2,7}, Alison C. Holloway¹
Helen McDonald³, Elyanne M. Ratcliffe^{2,6}, Jennifer C. Stearns^{5,6}, Jonathan D. Schertzer^{4,6,7}, Mike
G. Surette^{4,5,6}, Lehana Thabane⁸, Monique Mommers^{9,10}

¹Department of Obstetrics and Gynecology, McMaster University, Hamilton, Ontario, Canada.

²Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada.

³Midwifery Education Program, McMaster University, Hamilton, Ontario, Canada.

⁴Department of Biochemistry & Biomedical Sciences, McMaster University, Hamilton, Ontario, Canada

⁵Department of Medicine, McMaster University, Hamilton, Ontario, Canada.

⁶Farncombe Family Digestive Health Research Institute, McMaster University, Hamilton, Ontario, Canada

⁷Centre for Metabolism, Obesity, and Diabetes Research, McMaster University, Hamilton, Ontario, Canada.

⁸Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada.

⁹Department of Epidemiology, Care and Public Health Research Institute (CAPHRI), Maastricht University,
Maastricht, The Netherlands.

¹⁰*in Vivo* Planetary Health: an affiliate of the World Universities Network (WUN), West New York, New Jersey

Conflict of interest

There is no conflict of interest regarding the contents of the paper.

Funding

This study was financially supported by grants from the Joint Programming Initiative a Healthy Diet for a Healthy Life (HDHL) Joint Action Intestinal Microbiomics (project number 50-52905-98-599) and a ZonMW VIDI grant to JP (grant number 91716427) and grant 2815ERA05E by the German Ministry of Nutrition and Agriculture (Bundesanstalt für Landwirtschaft und Ernährung) paid to Charité and Susanne Lau. The PAPS trial was funded by Symbiopharm Herborn and the German Research Foundation (DFG Ha 2162/4-1).

ABSTRACT**Background & Aims**

Establishment of the gastrointestinal microbiota during infancy affects immune system development and oral tolerance induction. Perturbations in the microbiome during this period can contribute to development of immune-mediated diseases. We monitored microbiota maturation and associations with subsequent development of allergies in infants and children.

Methods: We collected 1453 stool samples, at 5, 13, 21 and 31 weeks post-partum (infants), and once at school-age (6–11 years), from 440 children (49.3% girls, 24.8% born by caesarean section; all children except for 6 were breastfed for varying durations; median 40 weeks; interquartile range, 30–53 weeks). Microbiota were analyzed by amplicon sequencing. Children were followed through 3 years of age for development of atopic dermatitis; data on allergic sensitization and asthma were collected when children were school age.

Results: Diversity of fecal microbiota, assessed by Shannon index, did not differ significantly among children from 5 through 13 weeks after birth, but thereafter gradually increased to 21 and 31 weeks. Most bacteria within the Bacteroidetes and Proteobacteria phyla were already present at 5 weeks after birth, whereas many bacteria of the Firmicutes phylum were acquired at later times in infancy. At school age, many new Actinobacteria, Firmicutes, and Bacteroidetes bacterial taxa emerged. The largest increase in microbial diversity occurred after 31 weeks. Vaginal, compared with cesarean section delivery, was most strongly associated with an enrichment of Bacteroides species at 5 weeks through 31 weeks. From 13 weeks onwards, diet became most important determinant of microbiota composition—cessation of breastfeeding, rather than solid food introduction, was associated with changes. For example, Bifidobacteria, staphylococci, and streptococci significantly decreased upon cessation of breastfeeding, whereas bacteria within the Lachnospiraceae family (*Pseudobutyrvibrio*, *Lachnobacterium*, *Roseburia*, and *Blautia*) increased. When we adjusted for confounding factors, we found fecal microbiota composition to be associated with development of atopic dermatitis, allergic sensitization, and asthma. Members of the Lachnospiraceae family, as well as the genera *Faecalibacterium* and *Dialister*, were associated with a reduced risk of atopy.

Conclusions: In a longitudinal study of fecal microbiota of children from 5 weeks through 6–11 years, we tracked changes in diversity and composition associated with the development of allergies and asthma.

Keywords: eczema, microbial diversity, microbial age, specific IgE

Colonization of the intestinal tract during the neonatal period is of crucial importance for the maturation of the mucosal immune system and the induction of oral tolerance.¹⁻³ Animal studies have provided compelling evidence to support a causal role of the intestinal microbiota and its metabolites, especially in early life, in the etiology of allergic diseases.³⁻⁶

Numerous epidemiological studies⁷⁻¹² also suggest that the infant intestinal microbiota plays an important role in the manifestation of allergic diseases and asthma, although actual results vary considerably between studies. About half of the studies that examined intestinal microbial diversity in infancy and childhood reported a lower diversity among children with (subsequent) allergies, whereas the remaining studies found no evidence for such an association.¹³ Moreover, despite many specific microbial taxa have been linked to allergies and asthma, it remains unclear which bacterial taxa prevent or promote disease onset.¹⁴

Lack of early stool sampling and different ages of stool sample collection, different microbiological profiling methods, and an inadequate control for potential confounders have been suggested to contribute to the heterogeneity between study results.^{9,13} Additionally, cross-sectional studies are prone to reverse causality, i.e. changes in the microbiota composition as a result of the disease manifestation, and only very few studies have sufficient clinical follow-up to link infant microbiota maturation to the subsequent development of asthma.¹¹

Initial microbial colonization starts upon rupture of the amniotic membranes and subsequent passage through the birth canal when the infant is seeded by maternal microbial strains, a process which is impeded in case of a cesarean section delivery.^{15,16} Subsequently, microbial populations evolve as the diet changes and the host develops. Given the highly dynamic and complex process of microbial assembly, succession and maturation, repeated sampling is important to allow analysis of the overall development of the indigenous infant microbial ecology.¹⁷ Moreover, many of the antenatal and postnatal factors that influence microbial community assembly during infancy, such as birth mode and the presence of older siblings and furry pets in the household, have also been associated with the development of allergic diseases and asthma^{9,13,18,19}, highlighting the importance to account for potential confounding factors.

In the present study, we aimed to monitor microbial assembly, succession and maturation during the first year of life and identify hereditary, perinatal, environmental, lifestyle and dietary factors that drive microbiota development. Through the application of various multivariable longitudinal models, including joint modelling, we next examined the dynamics in microbial diversity, composition and community structure in association with the subsequent risk of developing atopic dermatitis and asthma until to school-age.

Our findings indicate that alterations in microbial diversity and composition precede the onset of allergic manifestations, while emphasizing the importance of possible confounders.

METHODS

Design and clinical outcome measurements

Originally this study was designed as a randomized placebo-controlled clinical trial to examine the impact of a bacterial lysate, containing heat-killed *Escherichia coli* and *Enterococcus faecalis*, on the primary prevention of atopic dermatitis (AD) (registration no. ISRCTN60475069, ISRCTN

registry).²⁰ However, we did not find any evidence that the intervention affected the microbiota composition and therefore pooled both treatment arms in the downstream statistical analyses. Infants were clinically examined by a pediatrician on signs of AD at the ages of 1, 21 and 31 weeks and again at 1, 2 and 3 years of age as described previously.²⁰

The school-age follow-up of the study population (at 6-11 years) took place in 2013 including clinical examination, lung function testing, skin prick tests, and serum analyses of specific IgE to the most common aero-allergens (house dust mite, dog, cat, mold (*Alternaria*, *Cladosporium*), birch and grass pollen. Children were classified as having current asthma in case of a doctor's diagnosis in combination of any indicative symptoms in the last 12 months (wheezing, shortness of breath, nocturnal awakening due to shortness of breath and/or wheezing). Allergic sensitization was assessed by Skin Prick Test and serum sensitization for the above-mentioned allergens.

The study and follow-up were approved by the hospitals local review board Charité Ethics Committee in 2002 and 2012. Parents and participants gave written informed consent.

Microbial profiling of fecal samples

Fecal DNA was isolated by a combination of bead beating and column-based purification as described in detail previously.²¹

Barcoded universal primers adapted from Bartram and colleagues²² were used to amplify the variable 3 region of the 16S rRNA gene. Amplicons were sequenced using the Illumina MiSeq platform using 2x250 paired-end reads. The resulting sequencing data were processed using the short-read library 16S rRNA gene sequencing pipeline (sl1p)²³ (for description see supplementary methods).

This resulted in a total of 93,475,612 reads from 1,468 samples that were clustered into 7,961 OTUs. Removal of OTUs that were observed in only a single sample and discarding OTUs with a fraction of the total number of sequences below 0.001, retained the majority of sequences (92,997,277) while significantly reducing the number of OTUs to 873. Finally, we eliminated 15 samples with a low coverage (<15,000 reads) and normalized the data using *Rhea*.²⁴ In order not to discard informative information, normalization in *Rhea* is performed by dividing OTU counts per sample for their total count (sample depth) and then multiplying the obtained relative abundance for the lowest sample depth (15,540 reads).

Statistical Analysis

All the statistical analyses were performed two-sided using R, version 3.4.3.

GI microbiota richness and maturation

The Chao1 index, as measure for the estimated microbial richness, and the Shannon index, as microbial diversity metric, were computed using the R package *vegan* 2.5.3.²⁵

In order to compare the microbial community structure of samples, we used the unweighted UniFrac which incorporates phylogenetic distances between observed organisms. The R package *GUniFrac* 1.1²⁶ was used to compute the unweighted UniFrac metric.

Ordination of samples based upon the unweighted Unifrac dissimilarity was visualized using Principal Coordinate Analysis (PCoA) using the *capscale* function of the *vegan* package.

The Friedman test was used to test for significant differences in Chao1, Shannon and loadings scores on the principal coordinates between all time-points, followed by Dunn's test for post hoc pairwise comparisons between individual time-points. The p-values were finally FDR adjusted for multiple comparisons.

In order to track the dynamics of individual OTUs within the four main bacterial phyla, we created Sankey plots using Sankeymatic. For readability of the Sankey plots, the OTU table was further filtered to OTUs that were present in $\geq 10\%$ of the samples during one or more time-points.

We next investigated the maturity of the GI microbiota by computing the microbial-by-age z-score (MAZ) of the sample as described previously.²⁷ Shortly, we started feeding a random forest with a training set made of the microbial community composition of healthy children after filtering out all OTUs with a prevalence below 5%. Once the model was trained, we used it to predict the age of all the samples. Finally, the z-score was computed using the following formula:

$$MAZ = \frac{\text{Microbial age} - \text{median of microbial age of healthy children of same chronologic age}}{(\text{standard deviation of microbial age of healthy children of the same chronologic age})}$$

Dirichlet Multinomial Mixture (DMM) clustering, an unsupervised clustering method that uses Laplace approximation to identify groups of communities (enterotypes) with similar composition, was performed as previously described.²⁸ We then analysed the transition of infants through these DMM-clusters with age.²⁹

Analysis of factors shaping the GI microbiota

We examined which hereditary, perinatal, environmental, lifestyle and dietary factors were associated with the establishment of the microbiome during infancy (see supplementary methods for detailed description).

In order to examine which of these factors were associated with the DMM clusters at baseline and/or with the transition of DMM clusters between the age of 5 and 31 weeks, multinomial logistic regression analysis were used. Only factors that were significantly associated with the (transition of) DMM clusters in the univariable analyses, were included in the final multivariable model.

We next used multivariate association with linear models (MaAsLin)³⁰ to examine the association between these factors with and individual microbial taxa and multivariable linear regression models to examine the association with the microbial diversity and maturity.

The effect size and significance of each of the covariates on the microbial community structure was determined using the *envfit* function in *vegan*.²⁵ Ordination was performed using the PCoA based on unweighted UniFrac metric obtained as described above. The significance value was determined based on 999 permutations. All *P* values derived from *envfit* were adjusted for multiple comparisons using FDR adjustment (Benjamini–Hochberg procedure).

In order to understand which of the covariates had the strongest impact on the overall microbial community structure, we performed a Permutational Analysis of the Variance (PERMANOVA)

based on unweighted UniFrac. Only covariates that were statistically significant in the *envfit* analyses were included in the PERMANOVA.

Analysis on association between microbiota and allergic manifestations

To examine how the longitudinal variation of the microbial diversity (Shannon index) and maturity (MAZ) of the GI microbiota affects the time to development of AD, we applied a joint model³¹ using the JM function of the JMpackage³² (for details see supplementary methods).

To examine the impact of microbial diversity and maturity on asthma and allergic sensitization at school-age, a Generalized Linear Model (GLM) was built using *lme4* 1.1.19. The same covariates as included in the JM were incorporated as potential confounding factors. Since both asthma and sensitization were binary outcomes, a binomial distribution was chosen for the GLM.

To identify if specific bacterial genera were differentially abundant in children with and without allergic manifestations, we used the MetaLonDa package.³³ To ensure meaningful p-values we performed 999 permutations. To select only the significant associations, we choose a threshold of 0.05 for the p-values after FDR adjustment.

RESULTS

Study population characteristics

The study, initially designed as a randomized, placebo-controlled trial on the primary prevention of AD by an orally applied lysate of heat-killed *Escherichia coli* and *Enterococcus faecalis*, consisted of healthy newborns (n = 606) with a single or double heredity for atopy.²⁰ During the first 3 years of life, children were deeply phenotyped by physical examination and the collection of detailed questionnaires at 7 clinical visits. At school-age, children were contacted again to determine the establishment of asthma and allergic sensitization.

For the present study, only children with at least three fecal samples collected during the first year and/or feces collected at school-age were included (n = 440). Of these children, 217 (49.3%) were girls, 187 (42.5%) had older siblings, 109 (24.8%) were born by caesarean section and 29 (6.6%) were reportedly treated with antibiotics in the first 31 weeks of life. All except 6 children received breastfeeding, although the duration of breastfeeding varied considerably with a median duration of 40 weeks [IQR: 30-53]. Solid food was introduced at a median age of 25 weeks [IQR: 22 – 27] (Supplementary Table S1).

Development of the microbiota between early infancy and school-age

We first examined the compositional changes in the microbiota during infancy and compared this to the school-age microbiota composition. Samples collected at the age of 5 (n = 306), 13 (n = 287), 21 (n = 268) and 31 (n = 307) weeks post-partum and again at school-age (n = 300) were profiled by amplicon sequencing of the 16S rRNA hypervariable V3 gene region. Upon quality filtering and removal of samples with low sequencing depth (n = 15), 1,453 samples with a median sequencing depth of 62,420 reads/sample (range 15,540 – 168,848) were retained for downstream analysis and clustered into 873 operational taxonomic units (OTUs).

Microbial diversity, assessed by the Shannon index, was not significantly different between ages 5 and 13 weeks but thereafter gradually increased from 13 to 21 and 31 weeks after birth (Fig. 1A). The largest increase in microbial diversity occurred after the age of 31 weeks as indicated by the steep increase in the Shannon index at school-age ($p = 7.99 \times 10^{-28}$). Similar findings were observed for the microbial richness as assessed by the Chao1 (Supplementary Table S2).

Principle coordinate analyses indicated that the microbial community structure as assessed by the unweighted Unifrac metric also gradually shifted during infancy with the most prominent shift between the age of 21 and 31 weeks (Fig 1B, $p = 1.58 \times 10^{-27}$, Supplementary Table S3). The school-age samples, however, clustered separately and showed less inter-individual variation as compared to the infant samples.

Tracking individual OTUs based on their presence or absence revealed different dynamics within the dominant phyla (Supplementary Fig 1). The majority of OTUs within the phyla of Actinobacteria, Bacteroidetes and Proteobacteria found during infancy were already present at 5 weeks after birth, whereas almost half of the OTUs within the phylum of Firmicutes were only acquired at later infant time-points. At school-age, many new Actinobacteria, Firmicutes, and Bacteroidetes OTUs emerged on top of the OTUs already present during infancy. In contrast, only few new OTUs emerged within the phylum Proteobacteria at school-age, while a significant portion of the infant OTUs were lost thereafter.

We next examined the bacterial genera that contribute most to the temporal dynamics in microbial diversity and community structure. Towards school-age the prevalence in many of the genera within the phylum of Proteobacteria dramatically decreased, whereas the prevalence of genera within the phylum of Firmicutes, and in particular within the Lachnospiraceae and Ruminococcaceae families strongly increased (Fig 1C). Moreover, with the exception of *Bifidobacterium*, the relative abundance of all of the major bacterial genera changed significantly (Friedman test, all P-values < 0.001, Supplementary Table 4) throughout infancy and childhood (Supplementary Fig. 2A-B). *Escherichia* was the most abundant genus at 5 weeks of age followed by *Bifidobacterium* and *Streptococcus*. *Escherichia* still remained the most abundant genus at 31 weeks of age but was now followed by *Bacteroides* and *Veillonella*. At school-age the most abundant genera were *Blautia*, *Faecalibacterium* and *Ruminococcus*.

The length of breastfeeding represents the main driver of the infant's microbiota composition

To identify covariates associated with the microbiota dynamics during infancy, we continued our analyses focusing on the infant samples. Dirichlet multinomial mixtures (DMM) modelling on OTU-level data formed six clusters (based on lowest Laplace approximation) (Fig. 2A-B).

To illustrate the progression of samples through each DMM cluster with age, we applied a transition model as described previously.²⁹

Clusters 1 and 3 were the most dominant at the age of 5 weeks and thereafter transitions were chaotic, consistent with the previously identified developmental phase of the microbiome during the first 14 months of life.²⁹ Although cluster 1 remained dominant until the age of 31 weeks, cluster 3 gradually disappeared in favor of clusters 4 and 5 (Fig 2C). Multinomial logistic regression analyses indicated that the initial microbiota cluster at 5 weeks of age was mainly determined by birth mode. The chance that a newborn's microbiota belonged to cluster 3 was strongly increased among infants born by cesarean section [Supplementary Table 5]. This cluster was remarkably different with respect to the abundance of several of the driving OTUs. In particular a *Klebsiella* OTU exhibited a high abundance at the expense of an *Escherichia* OTU that dominated many of the other clusters. Additionally, *Citrobacter*, *Leclercia* and *Raoultella* OTUs were characteristic for cluster 3 (Fig 2B).

Analysis of the most common transition trajectories, revealed that for both children starting in cluster 1 as well as for children starting in cluster 3, transition towards clusters 4 and 5 significantly increased when breastfeeding was ceased [Supplementary Table 5].

These results were further supported by the overall bacterial profiles throughout infancy. At the age of 5 weeks, the largest amount of variance was explained by birth mode (Fig. 3A). At the genus level, vaginal as compared to cesarean section delivery was most strongly associated with an enrichment of *Bacteroides* spp., at 5 weeks and until the age of 31 weeks (Supplementary Table 6).

At the age of 13, 21 and 31 weeks, breastfeeding explained by far the greatest variance in bacterial community profiles (Fig 3B-D). Permutational multivariate analyses of variance confirmed that the duration of breastfeeding had a stronger impact than the introduction of solid foods (Fig 3E, Supplementary Table 7). Bifidobacteria, staphylococci and streptococci amongst others significantly decreased upon cessation of breastfeeding, whereas many bacteria within the Lachnospiraceae family (e.g. *Pseudobutyrvibrio*, *Lachnobacterium*, *Roseburia*, *Blautia*) increased (Supplementary Table 6).

A longer duration of breastfeeding was also associated with a lower microbial diversity (Supplementary Table 8) as well as with a lower microbial-by-age z-score (MAZ) (Supplementary Table 9). The MAZ is calculated by training a machine-learning algorithm on the microbiota composition of a dataset with known biological age, thereafter the age of samples is predicted based on its microbiota composition. A lower MAZ is thus indicative for a delayed microbial maturation.

Furthermore, the exposure to older siblings was associated with an increase in several genera within the phylum of Actinobacteria (*Bifidobacterium* and *Corynebacterium* at 5 weeks and *Egghertella* at 21 weeks, Supplementary Table 6) and a higher microbial diversity at 31 weeks of age (Supplementary Table 9). Finally, besides dietary factors, the microbial community structure was most strongly associated with the presence of AD at time of sample collection.

Alterations in microbial composition, diversity and maturity precede manifestations of atopy

To further investigate whether differences in microbiota development precede the onset of atopic disease, we applied several longitudinal analyses while controlling for potential confounding factors by adjusting for other covariates.

We first applied multivariate joint models on the microbial diversity and maturity in association to AD. Joint models have become increasingly popular as a statistical framework to concurrently analyze longitudinal data (e.g. biomarker evolution) and survival data (e.g. time-to-disease onset).³¹ They have, to our knowledge, not been applied in the microbiome research field so far. While accounting for known risk factors for AD, we found that the temporal pattern of microbial diversity was independently and inversely associated with AD (Hazard rate (HR)= 0.21; $p=1.15 \times 10^{-4}$, Figure 4A-B, Supplementary Table 10), indicating that a lower microbial diversity throughout infancy is associated with an increased risk of AD. For the temporal pattern of microbial maturity, expressed as microbial age z-scores, we found a statistically significant positive association with AD (HR=1.14, $p=1.94 \times 10^{-5}$, Figure 4C-D, Supplementary Table 11). Next, we used the recently introduced metagenomics longitudinal differential abundance (MetaLonDa) method³³ to identify time intervals of differentially abundant bacterial genera between infants that did or did not develop AD.

Among children who did not develop AD during follow-up, the relative abundance of *Atopobium* (days 25.6 - 79.4, $\text{FDR}_{\text{adjusted}} p = 7.65 \times 10^{-3}$), *Corynebacterium* (days 126.1 - 151.2, $\text{FDR}_{\text{adjusted}} p = 9.68 \times 10^{-3}$), both members of the phylum Actinobacteria, and *Prevotella* (days 104.6 - 133.3, $\text{FDR}_{\text{adjusted}} p < 0.001$) were temporarily enriched when compared to children who developed AD. Most pronounced were, however, the associations of *Lachnobacterium* and *Faecalibacterium*, which were significantly enriched during the entire period of fecal sampling among children who remained free from AD (Figure 4E-G, Supplementary Table 12).

We next, examined whether the infant microbiota composition was also associated with allergic manifestations at school-age, including allergic sensitization and asthma. Blood samples for the determination of allergic sensitization at school age were available for 292 out of the 440 children included in the present study. Like for AD, we found a higher diversity of the infant microbiota to be associated with decreased risk of allergic sensitization at school-age (Shannon index at 31 weeks $\text{OR}_{\text{adjusted}} = 0.19$, $p = 7.33 \times 10^{-3}$, Supplementary Figure 3A, Supplementary Table 13). A higher microbial maturity very early in life was associated with an increased risk of allergic sensitization (MAZ at 5 weeks $\text{OR}_{\text{adjusted}} = 1.46$, $p = 5.01 \times 10^{-3}$, Supplementary Figure 3B,

Supplementary Table 14), again in line with findings for AD. We could, however, not identify individual bacterial genera with differential abundance over a significant period of time between children who did or did not develop allergic sensitization.

A clear association between microbial diversity and asthma could not be detected. Yet, in line with allergic sensitization and AD, a higher microbial maturity at the age of 5 weeks was also associated with an increased risk for asthma (MAZ at 5 weeks $OR_{adjusted} = 1.43$, $p = 7.78 \times 10^{-3}$, Supplementary Figure 3C-D, Supplementary Tables 15 and 16). Multiple bacterial genera were differentially abundant over time in children who did or did not develop asthma. The genera that were differentially abundant across the entire time-period during which the microbiota composition was monitored included *Lachnobacterium*, *Lachnospira* (both members of the Lachnospiraceae family) and *Dialister* (Veillonellaceae), which were all significantly enriched in healthy as compared to asthmatic children (Supplementary Figure 3E-G, Supplementary Table 17).

DISCUSSION

This study aimed to longitudinally analyze the process of gastrointestinal microbial community assembly, succession and maturation throughout the most critical time-window of immune development and linked microbiota maturation during this time to the development of clinical signs of allergic disease, while carefully controlling for potential confounding factors.

Our results indicate a dynamic microbiota during infancy which is far from completely matured at 31 weeks of age. In early infancy the microbial composition was most strongly affected by birth mode, while from 13 weeks onwards diet became the most important factor. Our data support previous reports, showing that *Bacteroides* are most strongly affected by birth mode.^{29, 34-36}

The difference in microbial community structure and lower abundance of *Bacteroides* in cesarean section as compared to vaginal delivered infants persisted up to the age of 31 weeks of age and withstood mutual adjustment for other determinants, including breastfeeding. This suggests that the impact of cesarean section delivery could not be compensated by breastfeeding. Given the increased risk of future diseases, including allergies and asthma^{37, 38}, among children born by cesarean section, more research is warranted to elucidate the need for and efficacy of restoring the natural microbial colonization process upon cesarean delivery. We furthermore showed that cessation of breastfeeding was more strongly associated with microbial composition and maturity than solid food introduction. In line with previous^{11, 35, 39}, these results suggest that the introduction of solid food does not appear to result in a profound shift in microbial community structure as long as breastfeeding is continued. Only when breastfeeding is ceased, maturation of the microbiota is accelerated with a decrease in degraders of human milk oligosaccharides and an increase in microbial diversity and compositional changes towards bacterial genera specialized in degrading complex dietary carbohydrates. The generally observed lower microbial diversity in infants during breastfeeding⁴⁰ seems at first contradictory with the concept that a “healthy” and resilient microbiome is highly diverse.⁴¹ However, in line with most prospective studies⁴², we did not find a direct association between breastfeeding duration and the risk of AD. The fact that breastfeeding reduces the risk of several other diseases, including metabolic diseases, which on the other hand are also associated with a lower microbial diversity, suggests that the context is of crucial importance when considering microbial diversity. For example, loss of microbiota diversity generally opens up niches for opportunistic invaders⁴¹, while the plethora of bioactive components transferred by breastfeeding protect against colonization by such opportunistic pathogens.⁴³ This further underscores the need for meticulous adjustment for diet as a confounding factor in the association between microbiota and disease outcomes.

Using various multivariable longitudinal analysis, we furthermore demonstrated that the microbial community structure, diversity and maturity as well as the relative abundance of several individual genera were associated with the subsequent development of allergic manifestation. We know from previous animal studies and large longitudinal human cohorts that intestinal microbial dysbiosis in allergic diseases is mainly observed within a critical window in early life.⁴⁴ The comparability between studies is, however, hampered by the highly dynamic microbial communities within this early time window, which likely results in different associations at different sampling time points. The main strengths of the present study are its prospective design, repeated sample collection and the deep clinical phenotyping.

The regular physical examinations of the children throughout the first 3 years of life in combination with the collection of detailed questionnaire data allowed not only deep clinical phenotyping, but also an accurate assessment of the time of disease onset. The follow-up into school-age further facilitated a reliable classification of children who developed allergic asthma as it is well-known that wheezing symptoms at an earlier age are often transient and caused by episodic viral infections.⁴⁵

We observed a lower microbial diversity to be associated with AD development and allergic sensitization, but not with asthma. This is consistent with previous studies that also reported a reduced microbial diversity in association to AD⁴⁶⁻⁴⁹ and sensitization.^{7,50} In contrast, a link between microbial diversity and wheeze or asthma could often not be observed.⁵¹⁻⁵³ Although atopic manifestations are common comorbidities, these results support previous conclusions from COPSAC that extrapolation of risk factors between different atopic disorders may not always be justified.⁵⁴

The increased risk of AD, sensitization and asthma among children with a higher microbial maturity might at first seem in contrast with the above-mentioned results for microbial diversity and with findings of previous studies. Indeed recent results from COPSAC₂₀₁₀ linked a low microbial maturity with later onset of asthma in children born to asthmatic mothers.¹¹ In our study, the microbial maturity was however only significantly increased at 5 weeks of age in children who developed sensitization (as determined by Skin Prick Tests and serum IgE levels to the most common aero-allergens) and asthma. Also, for children with AD, we observed a microbial maturity (MAZ) that was significantly higher at the earliest recorded time-point but gradually decreased and became even lower at the age of 31 weeks when compared to the MAZ of children that did not develop AD. This temporally higher MAZ in very young infants might therefore suggest a dysregulated colonization process, e.g. with some bacterial taxa arriving (too) early, rather than a more mature overall microbial community structure.

Next to differences in microbial diversity and maturity, we were able to identify microbial taxa that were differentially abundant among infants who did or did not develop allergic disease manifestations. *Lachnobacterium* and *Faecalibacterium* were significantly decreased throughout infancy among children who developed AD. Also, *Lachnospira* and *Dialister*, next to *Lachnobacterium*, were significantly decreased among children who developed asthma.

The fact that these bacterial taxa were not only differentially abundant at a single time-point but throughout infancy strengthens the likelihood of a causal role in the protection against allergic disease. Altogether our results indicate that microbial perturbations in early life are also associated with asthma at school age, although perturbations are not identical to those observed in children that developed AD. In line with our findings, analysis on the microbiota composition at 3 months of age within the Canadian Healthy Infant Longitudinal Development (CHILD) Study revealed *Lachnospira* and *Faecalibacterium* to be significantly decreased among children at risk for allergic wheeze at the age of 1 year.⁵² Moreover, a lower relative abundance of amongst others *Lachnospiraceae incertae sedis*, *Faecalibacterium* and *Dialister* at the age of 1 year in children from COPSAC₂₀₁₀ was associated with an increased risk of asthma at 5 years.¹¹

Fermentation products of these bacteria are a possible explanation for the protective effect of these bacteria. Acetate is one of the fermentation products of *Lachnospira* and to a lesser account *Lachnobacterium*. Animal studies have previously shown that acetate-feeding leads to a marked suppression of allergic airway disease in a mouse-model for human asthma. The

underlying cellular mechanism was related to the effect of acetate on Treg cells, particularly through epigenetic modification of the *Foxp3* protomotor.⁵

Faecalibacterium (prausnitzii) is well-known for its anti-inflammatory effects, amongst others through the production of butyrate⁵⁵ and a microbial anti-inflammatory molecule (MAM) that inhibits the NF- κ B pathway.⁵⁶ Two recent studies have identified another lactate-consuming butyrate-producing genus, *Anaerostipes*, associated with a decreased risk of food allergy⁵⁷ and eczema.⁵⁸ The very low abundance of this genus in our population could potentially explain the lack of association in our study.

The application of several types of longitudinal data-analysis, including the joint modelling of longitudinal and survival data which had previously not been used for microbiota data analyses, enabled us to demonstrate that alterations in microbial diversity, maturity and composition preceded the clinical manifestations of atopic diseases. Although this statistical framework reveals the temporality of associations thereby suggesting causal relationships, causality can never be proven in an observational study. For example, microbial perturbations could be an epiphenomenon of exposure to yet another unknown risk factor for allergy. Also, it cannot be ruled out that early preclinical manifestation of allergies or genetic predisposition for allergy might impact the microbiota composition. Moreover, our findings on fecal collections might not fully reflect alterations in the microbiome on allergy development at the level of (small) intestinal mucosa.

It is therefore of importance not only to replicate findings in similar cohorts, but also to conduct future experimental studies building upon these findings in order to reveal the underlying biological mechanisms and prove causality.

In conclusion, our results demonstrate the importance of birth mode and diet on the early maturation of the infant microbiota and demonstrate that, upon careful adjustment of important confounding factors, alterations in the microbial colonization process of the infant intestinal tract precede the development of AD, sensitization and asthma. In particular members of the Lachnospiraceae family, as well as the genera *Faecalibacterium* and *Dialister* appear to protect against allergies. These findings further support the future development of evidence-based intervention strategies targeting the microbiota to prevent or treat allergic diseases in early life.

REFERENCES

1. Bennek E, Mandic AD, Verdier J, et al. Subcellular antigen localization in commensal *E. coli* is critical for T cell activation and induction of specific tolerance. *Mucosal Immunol* 2019;12:97-107.
2. Cahenzli J, Koller Y, Wyss M, et al. Intestinal microbial diversity during early-life colonization shapes long-term IgE levels. *Cell Host Microbe* 2013;14:559-70.
3. Stefka AT, Feehley T, Tripathi P, et al. Commensal bacteria protect against food allergen sensitization. *Proc Natl Acad Sci U S A* 2014;111:13145-50.
4. Olszak T, An D, Zeissig S, et al. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* 2012;336:489-93.
5. Thorburn AN, McKenzie CL, Shen S, et al. Evidence that asthma is a developmental origin disease influenced by maternal diet and bacterial metabolites. *Nat Commun* 2015;6:7320.

6. Trompette A, Gollwitzer ES, Yadava K, et al. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat Med* 2014;20:159-66.
7. Azad MB, Konya T, Guttman DS, et al. Infant gut microbiota and food sensitization: associations in the first year of life. *Clin Exp Allergy* 2015;45:632-43.
8. Dzidic M, Abrahamsson TR, Artacho A, et al. Aberrant IgA responses to the gut microbiota during infancy precede asthma and allergy development. *J Allergy Clin Immunol* 2017;139:1017-1025 e14.
9. Penders J, Stobberingh EE, van den Brandt PA, et al. The role of the intestinal microbiota in the development of atopic disorders. *Allergy* 2007;62:1223-36.
10. Simonyte Sjodin K, Hammarstrom ML, Ryden P, et al. Temporal and long-term gut microbiota variation in allergic disease: A prospective study from infancy to school age. *Allergy* 2019;74:176-185.
11. Stokholm J, Blaser MJ, Thorsen J, et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat Commun* 2018;9:141.
12. Fujimura KE, Sitarik AR, Havstad S, et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat Med* 2016;22:1187-1191.
13. Zimmermann P, Messina N, Mohn WW, et al. Association between the intestinal microbiota and allergic sensitization, eczema, and asthma: A systematic review. *J Allergy Clin Immunol* 2019;143:467-485.
14. Simonyte Sjodin K, Vidman L, Ryden P, et al. Emerging evidence of the role of gut microbiota in the development of allergic diseases. *Curr Opin Allergy Clin Immunol* 2016;16:390-5.
15. Korpela K, Costea P, Coelho LP, et al. Selective maternal seeding and environment shape the human gut microbiome. *Genome Res* 2018;28:561-568.
16. **Wampach L, Heintz-Buschart A, Fritz JV**, et al. Birth mode is associated with earliest strain-conferred gut microbiome functions and immunostimulatory potential. *Nat Commun* 2018;9:5091.
17. Gilbert JA, Lynch SV. Community ecology as a framework for human microbiome research. *Nat Med* 2019;25:884-889.
18. **Davidson R, Roberts SE, Wotton CJ, Goldracc MJ**. Influence of maternal and perinatal factors on subsequent hospitalisation for asthma in children: evidence from the Oxford record linkage study. *BMC Pulm Med* 2010;10:14.
19. van Nimwegen FA, Penders J, Stobberingh EE, et al. Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy. *J Allergy Clin Immunol* 2011;128:948-55 e1-3.
20. **Lau S, Gerhold K**, Zimmermann K, et al. Oral application of bacterial lysate in infancy decreases the risk of atopic dermatitis in children with 1 atopic parent in a randomized, placebo-controlled trial. *J Allergy Clin Immunol* 2012;129:1040-7.
21. Penders J, Gerhold K, Stobberingh EE, et al. Establishment of the intestinal microbiota and its role for atopic dermatitis in early childhood. *J Allergy Clin Immunol* 2013;132:601-607 e8.
22. Bartram AK, Lynch MD, Stearns JC, et al. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Appl Environ Microbiol* 2011;77:3846-52.

23. Whelan FJ, Surette MG. A comprehensive evaluation of the sl1p pipeline for 16S rRNA gene sequencing analysis. *Microbiome* 2017;5:100.
24. Lagkouvardos I, Fischer S, Kumar N, et al. Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ* 2017;5:e2836.
25. Oksanen J, Guillaume BF, Kindt R, et al. *Vegan: Community Ecology Package*. R Package Version. 2.5-3. CRAN. 2013.
26. Chen J. *GUniFrac: Generalized UniFrac Distances*. R Package Version. 1.1. CRAN. 2018.
27. Subramanian S, Huq S, Yatsunenkov T, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* 2014;510:417-21.
28. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PLoS One* 2012;7:e30126.
29. **Stewart CJ, Ajami NJ**, O'Brien JL, et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 2018;562:583-588.
30. **Morgan XC, Tickle TL, Sokol H**, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol* 2012;13:R79.
31. Cadarso Suarez C, Klein N, Kneib T, et al. Editorial "Joint modeling of longitudinal and time-to-event data and beyond". *Biom J* 2017;59:1101-1103.
32. Rizopoulos D. JM: An R Package for the Joint Modelling of Longitudinal and Time-to-event Data. *Journal of Statistical Software* 2010;35.
33. Metwally AA, Yang J, Ascoli C, et al. MetaLonDA: a flexible R package for identifying time intervals of differentially abundant features in metagenomic longitudinal studies. *Microbiome* 2018;6:32.
34. Azad MB, Konya T, Maughan H, et al. Gut microbiota of healthy Canadian infants: profiles by mode of delivery and infant diet at 4 months. *CMAJ* 2013;185:385-94.
35. Backhed F, Roswall J, Peng Y, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe* 2015;17:690-703.
36. Penders J, Thijs C, Vink C, et al. Factors influencing the composition of the intestinal microbiota in early infancy. *Pediatrics* 2006;118:511-21.
37. Huang L, Chen Q, Zhao Y, et al. Is elective cesarean section associated with a higher risk of asthma? A meta-analysis. *J Asthma* 2015;52:16-25.
38. Thavagnanam S, Fleming J, Bromley A, et al. A meta-analysis of the association between Caesarean section and childhood asthma. *Clin Exp Allergy* 2008;38:629-33.
39. Pannaraj PS, Li F, Cerini C, et al. Association Between Breast Milk Bacterial Communities and Establishment and Development of the Infant Gut Microbiome. *JAMA Pediatr* 2017;171:647-654.
40. Ho NT, Li F, Lee-Sarwar KA, et al. Meta-analysis of effects of exclusive breastfeeding on infant gut microbiota across populations. *Nat Commun* 2018;9:4169.
41. Bello MGD, Knight R, Gilbert JA, et al. Preserving microbial diversity. *Science* 2018;362:33-34.
42. Lin B, Dai R, Lu L, et al. Breastfeeding and Atopic Dermatitis Risk: A systematic Review and Meta-Analysis of Prospective Cohort Studies. *Dermatol* 2019. doi: 10.1159/000503781 [Epub ahead of print].
43. van den Elsen LWJ, Garssen J, Burcelin R, et al. Shaping the Gut Microbiota by Breastfeeding: The Gateway to Allergy Prevention? *Front Pediatr* 2019;7:47.

44. Stiemsma LT, Turvey SE. Asthma and the microbiome: defining the critical window in early life. *Allergy Asthma Clin Immunol* 2017;13:3.
45. Townshend J, Hails S, McKean M. Diagnosis of asthma in children. *BMJ* 2007;335:198-202.
46. Abrahamsson TR, Jakobsson HE, Andersson AF, et al. Low diversity of the gut microbiota in infants with atopic eczema. *J Allergy Clin Immunol* 2012;129:434-40, 440 e1-2.
47. Forno E, Onderdonk AB, McCracken J, et al. Diversity of the gut microbiota and eczema in early life. *Clin Mol Allergy* 2008;6:11.
48. Ismail IH, Oppedisano F, Joseph SJ, et al. Reduced gut microbial diversity in early life is associated with later development of eczema but not atopy in high-risk infants. *Pediatr Allergy Immunol* 2012;23:674-81.
49. Wang M, Karlsson C, Olsson C, et al. Reduced diversity in the early fecal microbiota of infants with atopic eczema. *J Allergy Clin Immunol* 2008;121:129-34.
50. Chen CC, Chen KJ, Kong MS, et al. Alterations in the gut microbiotas of children with food sensitization in early life. *Pediatr Allergy Immunol* 2016;27:254-62.
51. Arrieta MC, Arevalo A, Stiemsma L, et al. Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting. *J Allergy Clin Immunol* 2018;142:424-434 e10.
52. **Arrieta MC, Stiemsma LT, Dimitriu PA, et al.** Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med* 2015;7:307ra152.
53. Bisgaard H, Li N, Bonnelykke K, et al. Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. *J Allergy Clin Immunol* 2011;128:646-52 e1-5.
54. Bisgaard H, Halkjaer LB, Hinge R, et al. Risk analysis of early childhood eczema. *J Allergy Clin Immunol* 2009;123:1355-60 e5.
55. Canani RB, Costanzo MD, Leone L, et al. Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World J Gastroenterol* 2011;17:1519-28.
56. **Quevrain E, Maubert MA, Michon C, et al.** Identification of an anti-inflammatory protein from *Faecalibacterium prausnitzii*, a commensal bacterium deficient in Crohn's disease. *Gut* 2016;65:415-425.
57. **Feehley T, Plunkett CH, Bao R, et al.** Healthy infants harbor intestinal bacteria that protect against food allergy. *Nat Med* 2019;25:448-453.
58. Wopereis H, Sim K, Shaw A, et al. Intestinal microbiota in infants at high risk for allergy: Effects of prebiotics and role in eczema development. *J Allergy Clin Immunol* 2018;141:1334-1342 e5.

Author names in bold designate shared co-first authorship

FIGURE LEGENDS

Figure 1. Microbiota maturation throughout infancy and childhood (N= 1,453 stool samples from 440 children). A, Microbial diversity (Shannon index) gradually increased throughout infancy and has markedly risen at school-age ($p = 7.72 \times 10^{-53}$, Friedman, p-values for post-hoc analyses using

Dunn's test are depicted in the table). **B**, Principal Coordinate Analysis (PCoA) based on unweighted Unifrac dissimilarity indicates a gradual shift in microbial community structure along PC1 during infancy and a completely distinct structure at school-age ($p = 6.0 \times 10^{-51}$, Friedman, p -values for post-hoc analyses using Dunn's test are depicted in Supplementary Table 3). **C**, Cladogram depicting the bacterial genera detected in the children's fecal microbiota. Background and branch colors reflect the different phyla. The height of the outer ring reflects the average relative abundance of a genus across all infant time points, whereas the color density of the five inner rings reflects the prevalence of the genus at the individual time points (with opaque color indicating a prevalence of 100 percent and fully transparent indicating a prevalence of 0 percent).

Figure 2. Community typing by Dirichlet Multinomial Mixtures of infant samples revealed six clusters (N= 1,154 stool samples from 312 children). **A**, Heat map showing the relative abundance of the 30 most important/dominant OTUs per DMM cluster. **B**, Principal Component Analysis (PCA) on OTU-level data with samples colored according to DMM cluster. Ellipses indicate the 95% confidence interval. Vertical and horizontal bar charts depict OTUs with the highest loadings on PC1 and PC2 respectively. The OTUs with the highest positive and negative loadings on PC1 and PC2 are plotted in the PCA. **C**, Transition model showing the progression of samples through the six DMM clusters from one sampling time-point to the next time-point.

Figure 3. Microbiota community structure is most strongly influenced by breastfeeding (N= 961 stool samples from 312 children).

A-D, Polar plots visualizing the amount of variance of microbial communities at 5 weeks (**A**), 13 weeks (**B**), 21 weeks (**C**) and 31 weeks (**D**) of age that could be explained by 18 covariates as analyzed using EnvFit. The height of the bars reflects the amount of variance (r^2) explained by each covariate. Covariates are colored to highlight parental health status and ethnicity (purple), perinatal covariates (orange), diet and medication (red) and environmental exposures (green). Asterisks indicate significant covariates (false discovery rate (FDR) $P < 0.05$) at each time point. **E**, Permutational Multivariate Analysis of Variance (PERMANOVA) combining all covariates that were significantly associated with microbial community variation at any given time point in the EnvFit analyses. The size of the dots reflects the R^2 . Only samples without missing data on the included covariates were included in PERMANOVA (5 weeks: N= 238, 13 weeks: N = 233, 21 weeks: N = 231, 31 weeks: N= 259). Asterisks indicate statistical significance with * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure 4. Microbiota composition, diversity and maturity is linked to the subsequent development of atopic dermatitis (N= 961 stool samples from 312 children).

A, Volcano plot depicting the regression coefficients from the joint model on the association between the Shannon index and atopic dermatitis (AD). The dashed line depicts the threshold for statistical significance at $p < 0.05$. Variables depicted below the dashed line were statistically significantly associated with AD in the final model. Positive coefficients (variables to the right of the vertical line) were associated with an increased AD risk. Negative coefficients (variables to

the left of the vertical line) were associated with a decreased AD risk. The Hazard Ratio is given by the exponent of the coefficient (e.g. for Shannon index: $e^{-1.57}$ results in a HR of 0.21).

B, Development of microbial diversity (Shannon index) throughout infancy among children that did (red line) or did not (green line) develop AD as modelled by Loess regression. Grey areas represent the 95% confidence intervals.

C, Volcano plot depicting the regression coefficients from the joint model on the association between the microbial age z-score (MAZ) and the development of atopic dermatitis (AD). The dashed line depicts the threshold for statistical significance at $p < 0.05$. **D**, Development of microbial maturity (MAZ-score) throughout infancy among children that did (red line) or did not (green line) develop AD as modelled by Loess regression. Grey areas represent the 95% confidence intervals.

E-G, Time intervals of differential abundance in *Faecalibacterium* (**E**), *Lachnobacterium* (**F**), and *Prevotella* (**G**) between infants that did or did not develop AD as identified from MetaLonda analyses. Significantly different time-intervals (FDR-adjusted $p < 0.05$) are depicted by gray shading.

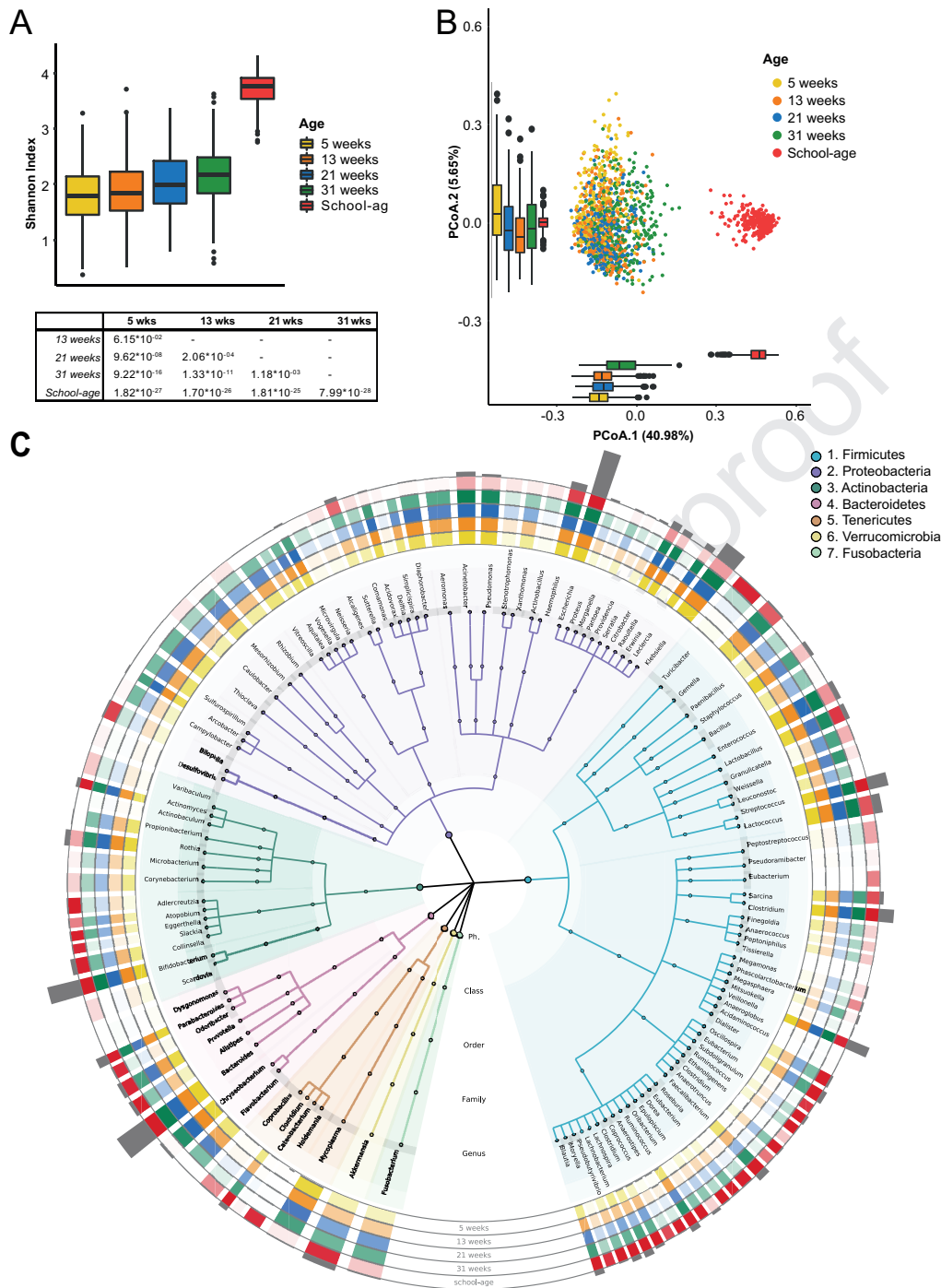


Figure 1. Microbiota maturation throughout infancy and childhood (N= 1,453 stool samples from 440 children).

A. Microbial diversity (Shannon index) gradually increased throughout infancy and has markedly risen at school-age ($p = 7.72 \times 10^{-53}$, Friedman, p -values for post-hoc analyses using Dunn's test are depicted in the table). **B.** Principal Coordinate Analysis (PCoA) based on unweighted Unifrac dissimilarity indicates a gradual shift in microbial community structure along PC1 during infancy and a completely distinct structure at school-age ($p = 6.0 \times 10^{-51}$, Friedman, p -values for post-hoc analyses using Dunn's test are depicted in Supplementary Table 3). **C.** Cladogram depicting the bacterial genera detected in the children's fecal microbiota. Background and branch colors reflect the different phyla. The height of the outer ring reflects the average relative abundance of a genus across all infant time points, whereas the color density of the five inner rings reflects the prevalence of the genus at the individual time points (with opaque color indicating a prevalence of 100 percent and fully transparent indicating a prevalence of 0 percent).

Journal Pre-proof

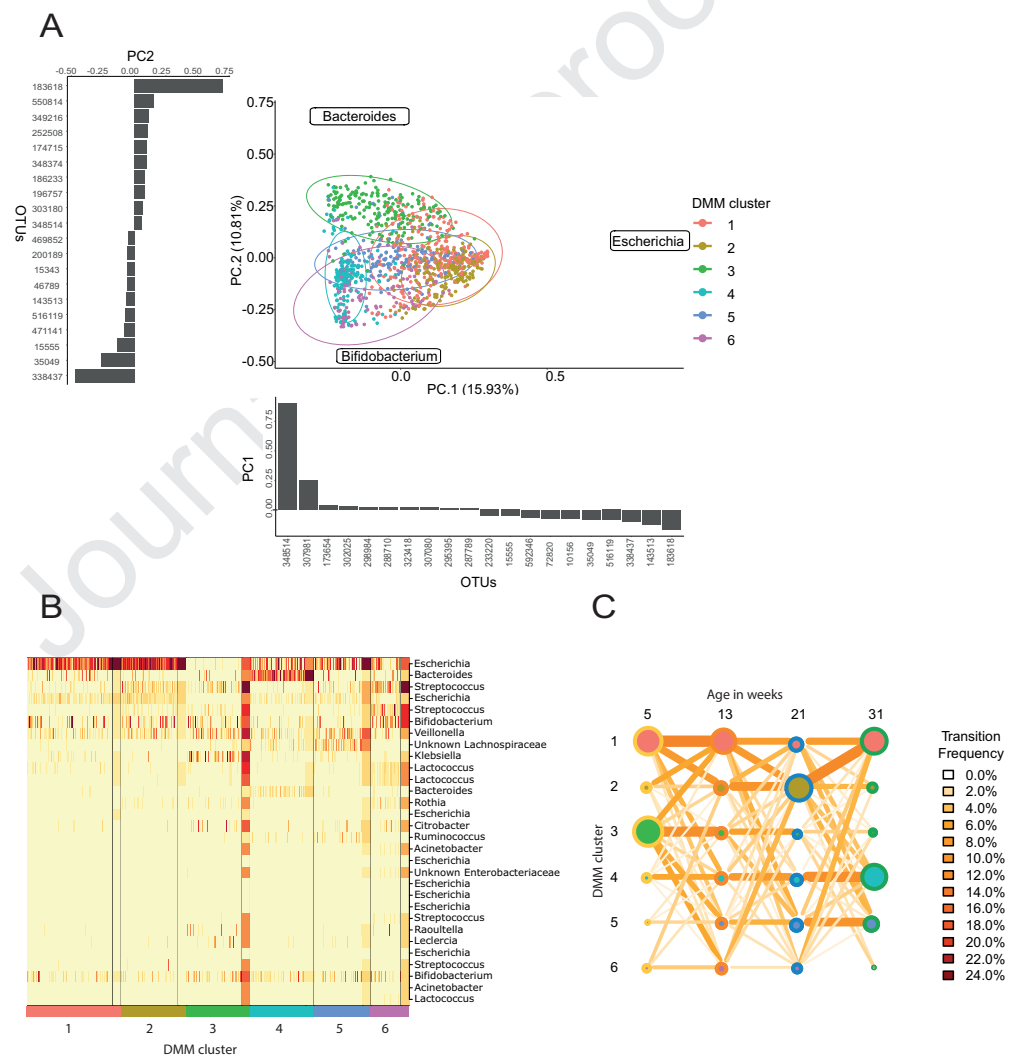


Figure 2. Community typing by Dirichlet Multinomial Mixtures of infant samples revealed six clusters (N= 1,154 stool samples from 312 children). **A**, Heat map showing the relative abundance of the 30 most important/dominant OTUs per DMM cluster.

B, Principal Component Analysis (PCA) on OTU-level data with samples colored according to DMM cluster. Ellipses indicate the 95% confidence interval. Vertical and horizontal bar charts depict OTUs with the highest loadings on PC1 and PC2 respectively. OTUs with the highest positive and negative loadings on PC1 and PC2 are plotted in the PCA. **C**, Transition model showing the progression of samples through the six DMM clusters from one sampling time-point to the next time-point.

Journal Pre-proof

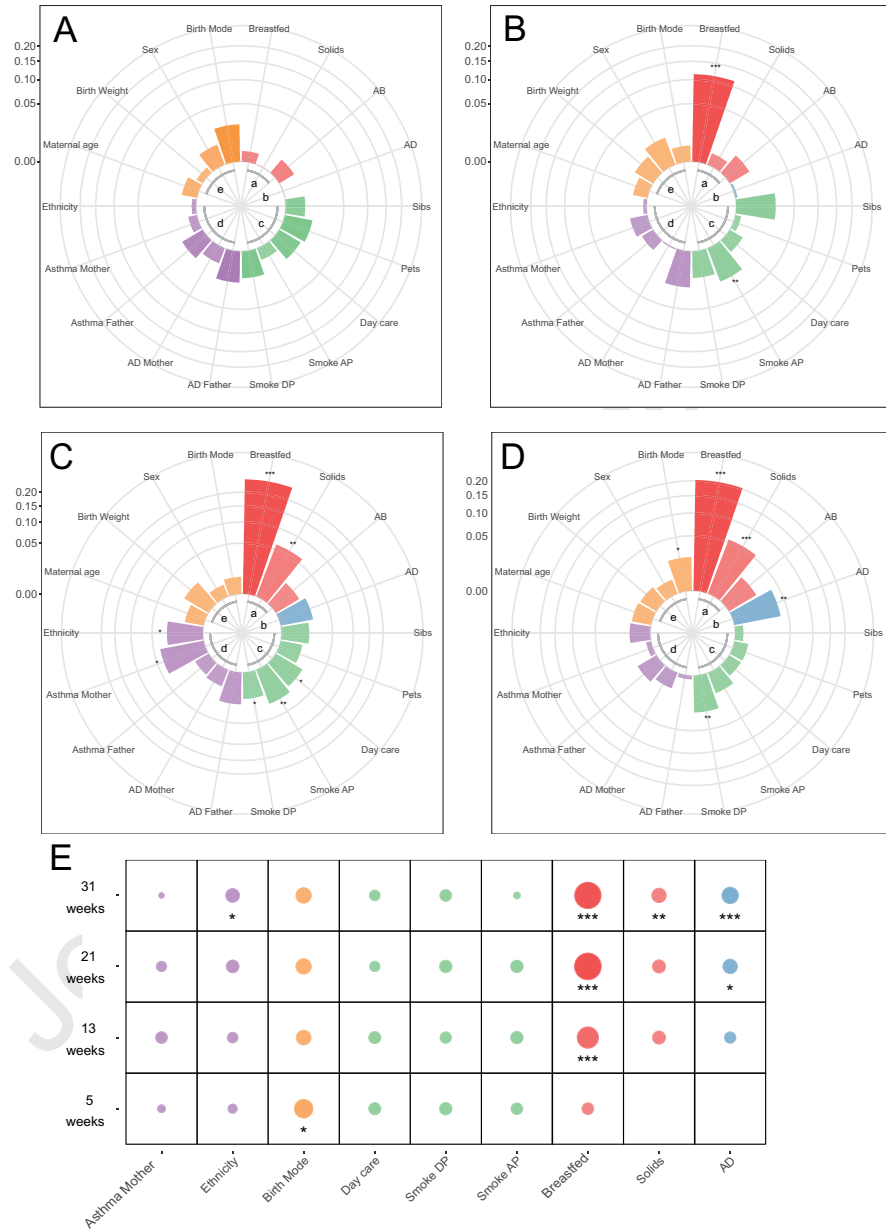


Figure 3. Microbiota community structure is most strongly influenced by breastfeeding (N= 961 stool samples from 312 children). **A-D**, Polar plots visualizing the amount of variance of microbial communities at 5 weeks (**A**), 13 weeks (**B**), 21 weeks (**C**) and 31 weeks (**D**) of age that could be explained by 18 covariates as analyzed using EnvFit. The height of the bars reflects the amount of variance (r^2) explained by each covariate. Covariates are colored to highlight parental health status and ethnicity (purple), perinatal covariates (orange), diet and medication (red) and environmental exposures (green). Asterisks indicate significant covariates (false discovery rate (FDR) $p < 0.05$) at each time point. **E**, Permutational Multivariate Analysis of Variance (PERMANOVA) combining all covariates that were significantly associated with microbial community variation at any given time point in the EnvFit analyses. The size of the dots reflects the R^2 . Only samples without missing data on the included covariates were included in PERMANOVA (5 weeks: N = 238, 13 weeks: N = 233, 21 weeks: N = 231, 31 weeks: N = 259). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

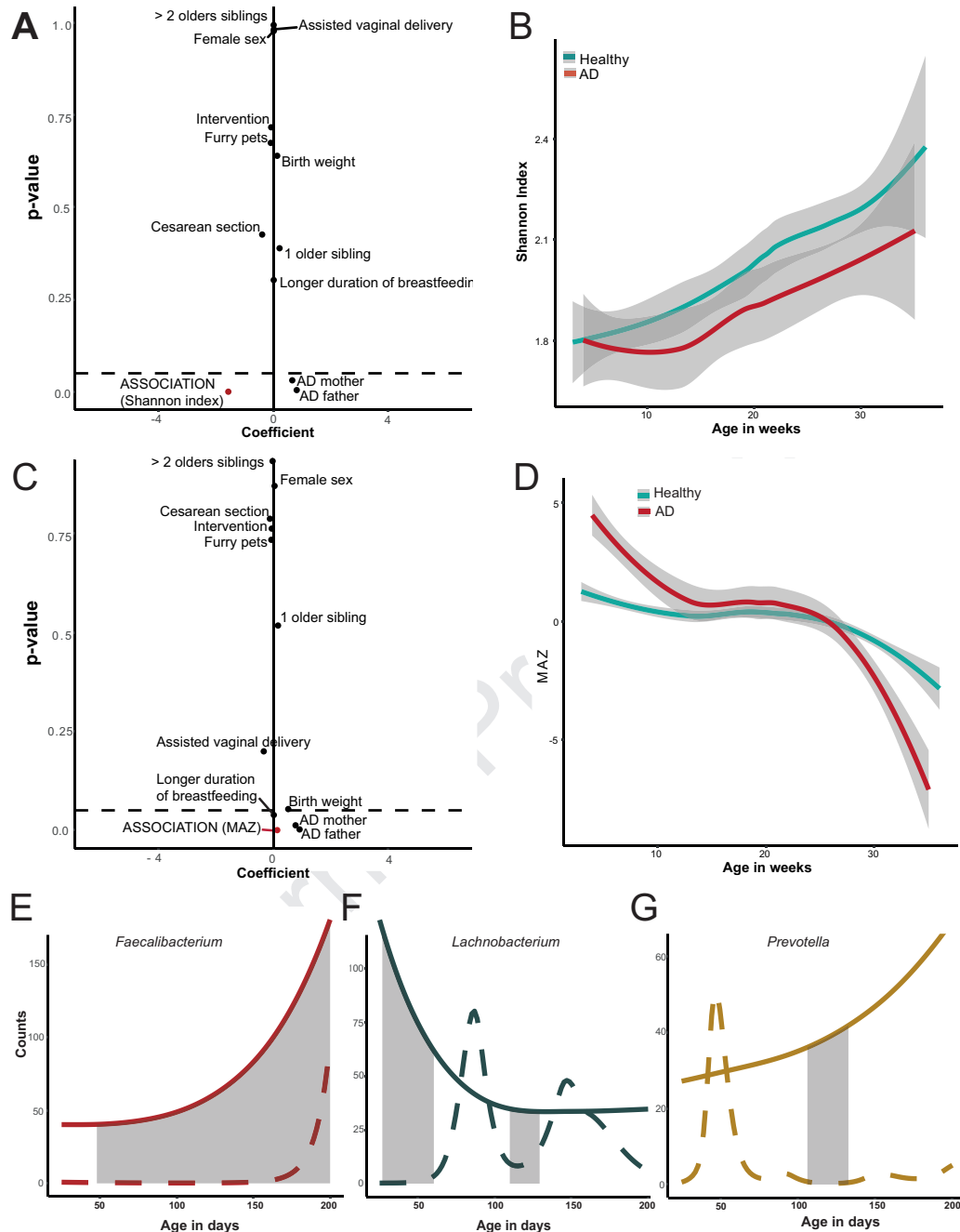


Figure 4. Microbiota composition, diversity and maturity is linked to the subsequent development of atopic dermatitis (N= 961 samples from 312 children). **A**, Volcano plot depicting the regression coefficients from the joint model on the association between the Shannon index and atopic dermatitis (AD). The dashed line depicts the threshold for statistical significance at $p < 0.05$. Variables depicted below the dashed line were statistically significantly associated with AD in the final model. Positive coefficients (variables to the right of the vertical line) were associated with an increased AD risk. Negative coefficients (variables to the left of the vertical line) were associated with a decreased AD risk. The Hazard Ratio is given by the exponent of the coefficient (e.g. for Shannon index: $e^{1.57}$ results in a HR of 0.21). **B**, Development of microbial diversity (Shannon index) throughout infancy among children that did (red line) or did not (green line) develop AD as modelled by Loess regression. Grey areas represent the 95% confidence intervals. **C**, Volcano plot depicting the regression coefficients from the joint model on the association between the microbial age z-score (MAZ) and the development of atopic dermatitis (AD). The dashed line depicts the threshold for statistical significance at $p < 0.05$. **D**, Development of microbial maturity (MAZ-score) throughout infancy among children that did (red line) or did not (green line) develop AD as modelled by Loess regression. Grey areas represent the 95% confidence intervals. **E-G**, Time intervals of differential abundance in *Faecalibacterium* (**E**), *Lachnobacterium* (**F**), and *Prevotella* (**G**) between infants that did or did not develop AD as identified from MetaLonda analyses. Significantly different time-intervals (FDR-adjusted $p < 0.05$) are depicted by gray shading.

Journal Pre-proof

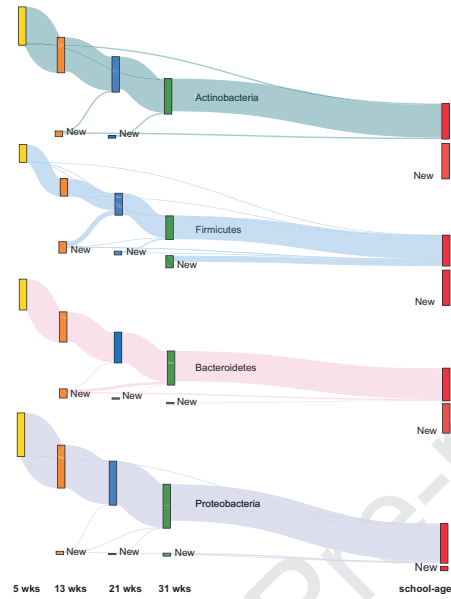
Supplementary Figure 1. Tracking the flow of OTUs within the four main phyla throughout infancy and childhood (N= 1,453 stool samples from 440 children).

OTUs that were shared by at least 10 percent of the population during one or more time points were tracked using Sankey plots in the four major phyla. The rectangle height indicates the relative number of OTUs and the rectangle color reflects the children's age. The lines represent the transfer of OTUs between time points. At school-age many new OTUs within the Actinobacteria, Firmicutes and Bacteroidetes were gained, whereas a major loss of OTUs in Proteobacteria was observed.

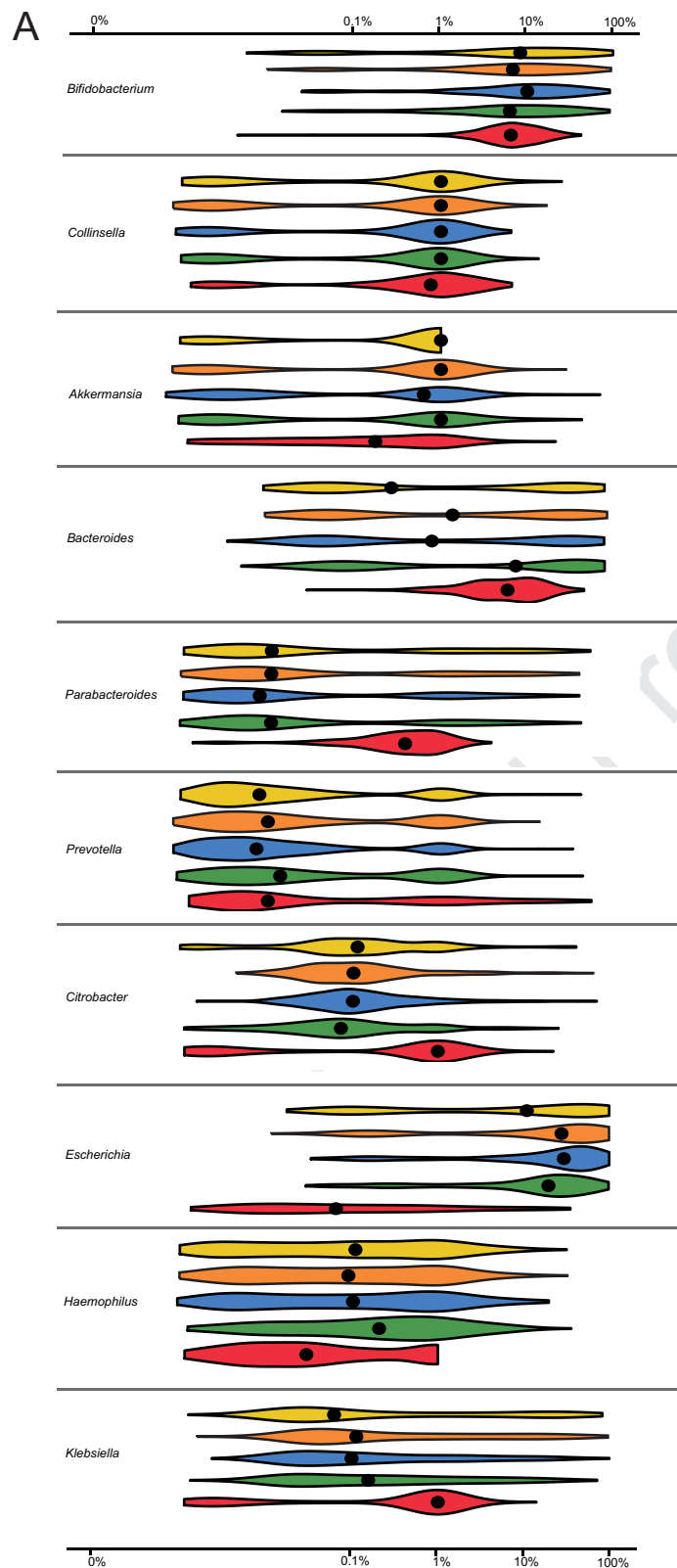
Supplementary Figure 2. Relative abundance of the main bacterial genera at different time-points during infancy and at school-age (N= 1,453 stool samples from 440 children). Comparison were made among the 20 most abundant genera and stratified according to age. The black dots represent the median values and the violins are colored according to age. **A**, relative abundance of the main genera within the phyla of Proteobacteria, Actinobacteria, Bacteroidetes and Verrucomicrobia. **B**, relative abundances of the main genera within the Firmicutes phylum. To test for significant changes in relative abundances with age, the Friedman test was used followed by the Dunn's test for post-hoc analyses. FDR-adjusted p-values are presented in Supplementary Table 4.

Supplementary Figure 3 (N= 961 stool samples from 312 children).

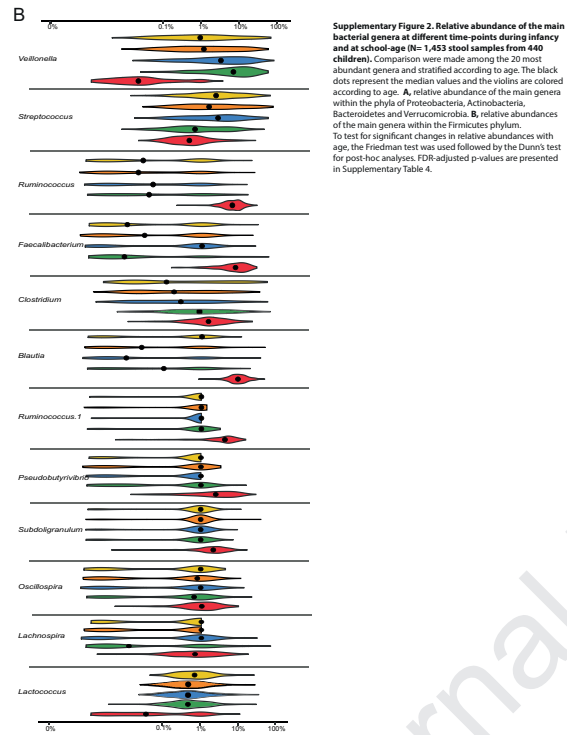
A-B, Volcano plots depicting the regression coefficients from the logistic regression analyses on the association between microbial diversity (Shannon index) and maturity (microbial age z-scores, MAZ) at the ages of 5, 13, 21 and 31 weeks and the development of allergic sensitization at school-age. **C-D**, Volcano plots depicting the regression coefficients from the logistic regression analyses on the association between microbial diversity (Shannon index) and maturity (microbial age z-scores, MAZ) at the ages of 5, 13, 21 and 31 weeks and the development of asthma at school-age. The dashed lines depict the threshold for statistical significance at $p < 0.05$. **E-G**, Time intervals of differential abundance in *Lachnobacterium* (**E**), *Lachnospira* (**F**) and *Dialister* (**G**) between infants that did or did not develop asthma as identified from MetaLonda analyses. Significantly different time-intervals (fdr-adjusted $p < 0.05$) are depicted by gray shading.

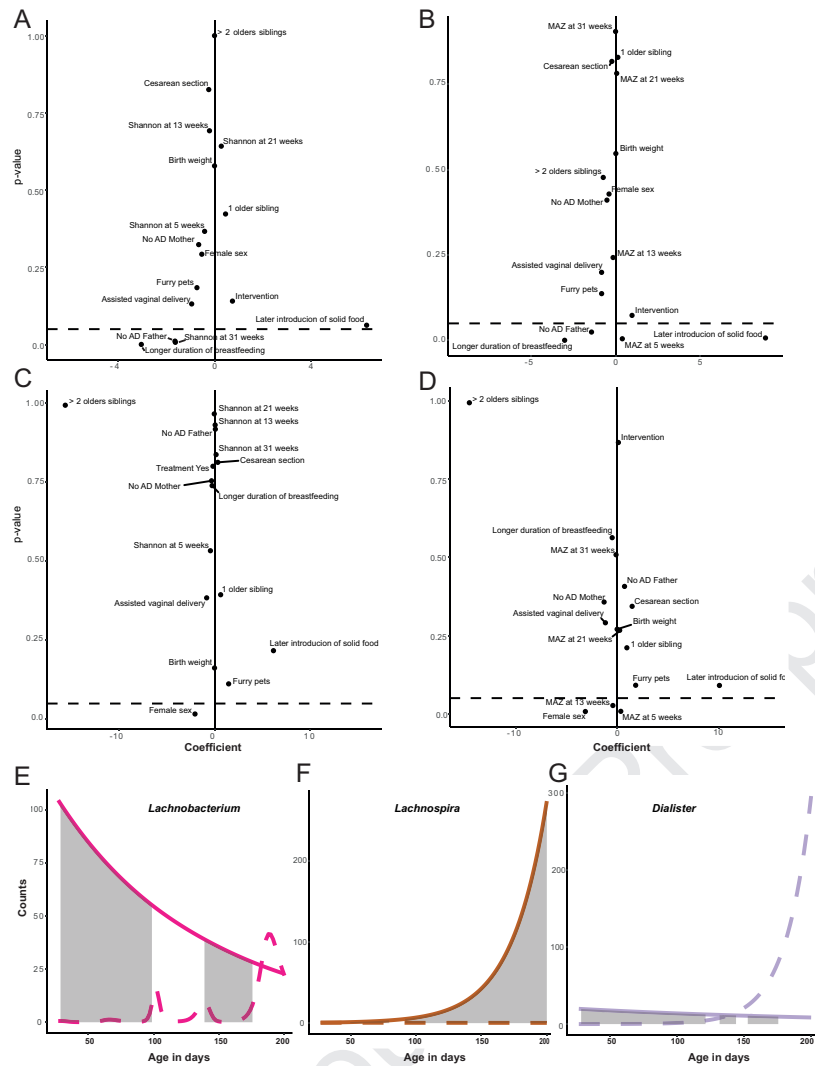


Supplementary Figure 1. Tracking the flow of OTUs within the four main phyla throughout infancy and childhood (N= 1,453 stool samples from 440 children). OTUs that were shared by at least 10 percent of the population during one or more time points were tracked using Sankey plots in the four major phyla. The rectangle height indicates the relative number of OTUs and the rectangle color reflects the children's age. The lines represent the transfer of OTUs between time points. At school-age many new OTUs within the Actinobacteria, Firmicutes and Bacteroidetes were gained, whereas a major loss of OTUs in Proteobacteria was observed.



Supplementary Figure 2. Relative abundance of the main bacterial genera at different time-points during infancy and at school-age (N= 1,453 stool samples from 440 children). Comparison were made among the 20 most abundant genera and stratified according to age. The black dots represent the median values and the violins are colored according to age. **A**, relative abundance of the main genera within the phyla of Proteobacteria, Actinobacteria, Bacteroidetes and Verrucomicrobia. **B**, relative abundances of the main genera within the Firmicutes phylum. To test for significant changes in relative abundances with age, the Friedman test was used followed by the Dunn's test for post-hoc analyses. FDR-adjusted p-values are presented in Supplementary Table 4.





Supplementary Figure 3 (N= 961 stool samples from 312 children). **B**, Volcano plots depicting the regression coefficients from the logistic regression analyses on the association between microbial diversity (Shannon index) and maturity (microbial age z-scores, MAZ) at the ages of 5, 13, 21 and 31 weeks and the development of allergic sensitization at school-age. **C-D**, Volcano plots depicting the regression coefficients from the logistic regression analyses on the association between microbial diversity (Shannon index) and maturity (microbial age z-scores, MAZ) at the ages of 5, 13, 21 and 31 weeks and the development of asthma at school-age. The dashed lines depict the threshold for statistical significance at $p < 0.05$. **E-G**, Time intervals of differential abundance in *Lachnobacterium* (**E**), *Lachnospira* (**F**) and *Dialister* (**G**) between infants that did or did not develop asthma as identified from MetaLonda analyses. Significantly different time-intervals (fdr-adjusted $p < 0.05$) are depicted by gray shading.

Table	
Supplementary Table S1	Demographics of the study population
Supplementary Table S2	Microbial richness as assessed by Chao1 index across time-points
Supplementary Table S3	Testing for statistical significant association between age of sample collection and principal coordinates of PCo1 as depicted in Figure 1
Supplementary Table S4	Testing for statistical significant association between age of sample collection and relative abundance of the main bacterial genera as depicted in Supplementary Figures 2A-B
Supplementary Table S5	Multinomial logistic regression analyses on the association between demographic, lifestyle and medical factors in association with DMM clusters and trajectories
Supplementary Table S6	Multivariate Analysis by Linear Models (MaAsLin) on demographic, lifestyle and medical factors in association with genus level taxa
Supplementary Table S7	Permutational analyses of variance on demographic, dietary, lifestyle and medical factors associated with microbial community structure
Supplementary Table S8	Multivariable linear regression analyses on demographic, dietary, lifestyle and medical factors associated with microbial maturity (Microbial Age z-scores)
Supplementary Table S9	Multivariable linear regression analyses on demographic, dietary, lifestyle and medical factors associated with microbial diversity (Shannon)
Supplementary Table S10	Joint modelling on the association between longitudinal microbial diversity (Shannon index) and the time to development of atopic dermatitis (AD)
Supplementary Table S11	Joint modelling on the association between longitudinal microbial maturity (MAZ score) and the time to development of atopic dermatitis (AD)
Supplementary Table S12	METALONDA analyses on microbial genera in association to atopic dermatitis
Supplementary Table S13	Logistic regression model of microbial diversity (Shannon index) in association to allergic sensitization at school-age
Supplementary Table S14	Logistic regression model of microbial maturity (MAZ score) in association to allergic sensitization at school-age
Supplementary Table S15	Logistic regression model of microbial diversity (Shannon index) in association to asthma at school-age
Supplementary Table S16	Logistic regression model of microbial maturity (MAZ score) in association to asthma at school-age
Supplementary Table S17	METALONDA analyses on microbial genera in association to asthma

Supplementary Table S1: Demographics of the study population

Variable	n (%)*
Sex	
Boys	223 (50.7)
Girls	217 (49.3)
Birth mode	
Natural vaginal delivery	304 (69.1)
Assisted vaginal delivery	26 (5.9)
Cesarean section	109 (24.8)
Number of older siblings	
No older siblings	253 (57.5)
1 older sibling	150 (34.1)
≥ 2 older siblings	37 (8.4)
Eczema mother	
Yes	161 (36.2)
No	279 (63.4)
Eczema father	
Yes	92 (20.9)
No	346 (78.6)
Asthma mother	
Yes	142 (32.3)
No	298 (67.6)
Asthma father	
Yes	102 (23.2)
No	337 (76.6)
Maternal smoking in pregnancy	
No	348 (79.1)
Yes	92 (20.9)
Maternal smoking after pregnancy	
No	341 (77.5)
Yes	99 (22.5)
Households with furry pets	
No	138 (31.4)
Yes	277 (63.0)
Ethnicity	
Caucasian	429 (97.5)
Other	11 (2.5)
Antibiotics before age 31 weeks	
No	411 (93.4)
Yes	29 (6.6%)

**Numbers may not add up to 440 (100%) due to missing data*

Supplementary Table S1: Demographics of the study population continued

Variable	Median (IQR)
Maternal age at delivery (in years)	33 (29-36)
Birth Weight (in grams)	3497 (3202-3800)
Breast feeding in weeks	40 (30-53)
Age at introduction of solids (in weeks)	25 (22-27)
Gestational age (in weeks)	40 (38-40)
Start at daycare (in months)	13 (11-22)

Supplementary Table S2: Microbial richness as assessed by Chao1 index accross time-points

Age	Median Chao1	Interquartile range
5 weeks	147.23	41.6 - 188.64
13 weeks	170.8	65.13 - 214.3
21 weeks	182.71	66.11 - 220.88
31 weeks	189.83	74.13 - 239.52
School-age	379.83	216.52 - 432.17

Test	P-value
Friedman test	7.724E-53

Variable 1	Variable 2	Test	P-value
Chao1 at 5 weeks	Chao1 at 13 weeks	Dunn's test	1.37E-02
Chao1 at 5 weeks	Chao1 at 21 weeks	Dunn's test	1.52E-05
Chao1 at 5 weeks	Chao1 at 31 weeks	Dunn's test	1.65E-07
Chao1 at 5 weeks	Chao1 at school-age	Dunn's test	7.55E-21
Chao1 at 13 weeks	Chao1 at 21 weeks	Dunn's test	4.09E-03
Chao1 at 13 weeks	Chao1 at 31 weeks	Dunn's test	4.09E-03
Chao1 at 13 weeks	Chao1 at school-age	Dunn's test	7.68E-21
Chao1 at 21 weeks	Chao1 at 31 weeks	Dunn's test	9.09E-01
Chao1 at 21 weeks	Chao1 at school-age	Dunn's test	7.55E-21
Chao1 at 31 weeks	Chao1 at school-age	Dunn's test	8.56E-21

Related samples statistics based only on children with samples collected at every single time-point (N = 606 samples of 121 children)

Supplementary Table S3: Testing for statistical significant association between age of sample collection and principal coordinates of PCo1 (Figure 1)

Test	P-value
Friedman	6.002E-51

Post-hoc comparisons

Variable 1	Variable 2	Test	P-value
PCo1 at 5 weeks	PCo1 at 13 weeks	Dunn's test	2.32E-04
PCo1 at 5 weeks	PCo1 at 21 weeks	Dunn's test	2.12E-03
PCo1 at 5 weeks	PCo1 at 31 weeks	Dunn's test	5.65E-34
PCo1 at 5 weeks	PCo1 at school-age	Dunn's test	1.21E-27
PCo1 at 13 weeks	PCo1 at 21 weeks	Dunn's test	3.27E-01
PCo1 at 13 weeks	PCo1 at 31 weeks	Dunn's test	1.21E-27
PCo1 at 13 weeks	PCo1 at school-age	Dunn's test	8.53E-27
PCo1 at 21 weeks	PCo1 at 31 weeks	Dunn's test	1.58E-27
PCo1 at 21 weeks	PCo1 at school-age	Dunn's test	1.03E-25
PCo1 at 31 weeks	PCo1 at school-age	Dunn's test	3.99E-28

Related samples statistics based only on children with samples collected at every single time-point (N = 606 samples of 121 children)

Supplementary Table S4: Testing for statistical significant association between age of sample collection and relative abundance of the main bacterial genera as depicted in Supplementary Figures 2A-B

	<i>Akkermansia</i>	<i>Prevotella</i>	<i>Oscillospira</i>	<i>Subdoligranulum</i>	<i>Haemophilus</i>	<i>Lachnospira</i>	<i>Citrobacter</i>	<i>Pseudobutyrvibrio</i>
Friedman-test p-value	7.69E-37	4.60E-05	3.76E-45	5.91E-75	1.48E-11	8.79E-57	4.37E-39	4.96E-64

Post-hoc analyses Dunn's test Bonferroni adjusted p-value

Time-point 1	Time-point 2	<i>Akkermansia</i>	<i>Prevotella</i>	<i>Oscillospira</i>	<i>Subdoligranulum</i>	<i>Haemophilus</i>	<i>Lachnospira</i>	<i>Citrobacter</i>	<i>Pseudobutyrvibrio</i>
5 weeks	13 weeks	NS	NS	NS	NS	NS	NS	NS	NS
5 weeks	21 weeks	<0.001	NS	NS	NS	NS	NS	0.05	NS
5 weeks	31 weeks	0.002	NS	<0.001	NS	<0.001	<0.001	NS	0.028
5 weeks	school-age	<0.001	NS	<0.001	<0.001	0.012	<0.001	<0.001	<0.001
13 weeks	21 weeks	0.012	NS	NS	NS	NS	NS	NS	NS
13 weeks	31 weeks	0.039	NS	NS	NS	0.012	<0.001	NS	NS
13 weeks	school-age	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
21 weeks	31 weeks	NS	NS	NS	NS	0.042	<0.001	NS	NS
21 weeks	school-age	<0.001	0.037	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
31 weeks	school-age	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Related samples statistics based only on children with samples collected at every single time-point (N = 606 samples of 121 children)

	<i>Ruminococcus</i>	<i>Lactococcus</i>	<i>Parabacteroides</i>	<i>Ruminococcus</i>	<i>Faecalibacterium</i>	<i>Blautia</i>	<i>Clostridium</i>	<i>Klebsiella</i>
Friedman-test p-value	6.30E-55	7.75E-42	1.54E-11	3.75E-78	8.95E-59	2.16E-59	2.55E-10	3.42E-39

Post-hoc analyses Dunn's test Bonferroni adjusted p-value

Time-point 1	Time-point 2	<i>Ruminococcus</i>	<i>Lactococcus</i>	<i>Parabacteroides</i>	<i>Ruminococcus</i>	<i>Faecalibacterium</i>	<i>Blautia</i>	<i>Clostridium</i>	<i>Klebsiella</i>
5 weeks	13 weeks	NS	NS	NS	NS	NS	NS	NS	NS
5 weeks	21 weeks	0.005	NS	NS	0.005	NS	0.012	NS	NS
5 weeks	31 weeks	<0.001	NS	NS	<0.001	0.002	<0.001	<0.001	NS
5 weeks	school-age	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
13 weeks	21 weeks	NS	NS	NS	NS	NS	NS	NS	NS
13 weeks	31 weeks	0.023	NS	NS	0.023	<0.001	0.003	0.005	NS
13 weeks	school-age	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
21 weeks	31 weeks	NS	NS	NS	NS	<0.001	NS	0.007	NS
21 weeks	school-age	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
31 weeks	school-age	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Related samples statistics based only on children with samples collected at every single time-point (N = 606 samples of 121 children)

Supplementary Table S4: Testing for statistical significant association between age of sample collection and relative abundance of the main bacterial genera as depicted in Supplementary Figures 2A-B

	<i>Streptococcus</i>	<i>Veillonella</i>	<i>Bacteroides</i>	<i>Bifidobacterium</i>	<i>Escherichia</i>
Friedman-test p-value	1.61E-14	6.00E-53	0.000067	3.00E-01	2.56E-40

Post-hoc analyses Dunn's test Bonferroni adjusted p-value

Time-point 1	Time-point 2	<i>Streptococcus</i>	<i>Veillonella</i>	<i>Bacteroides</i>	<i>Bifidobacterium</i>	<i>Escherichia</i>
5 weeks	13 weeks	NS	NS	NS	-	NS
5 weeks	21 weeks	NS	NS	NS	-	NS
5 weeks	31 weeks	<0.001	<0.001	<0.001	-	NS
5 weeks	school-age	<0.001	<0.001	NS	-	<0.001
13 weeks	21 weeks	NS	NS	NS	-	NS
13 weeks	31 weeks	<0.001	<0.001	0.011	-	0.003
13 weeks	school-age	<0.001	<0.001	NS	-	<0.001
21 weeks	31 weeks	<0.001	0.007	<0.001	-	0.02
21 weeks	school-age	<0.001	<0.001	NS	-	<0.001
31 weeks	school-age	NS	<0.001	NS	-	<0.001

Related samples statistics based only on children with samples collected at every single time-point (N = 606 samples of 121 children)

SUPPLEMENTARY TABLE S5: FINAL MULTIVARIABLE MULTINOMIAL LOGISTIC REGRESSION MODEL ON COVARIATES IN ASSOCIATION TO DMM CLUSTERS AT 5 WEEKS OF AGE.				
	Variable	Exp(B)	95% Confidence Interval for Exp (B)	P-value
DMM Cluster 1 (n = 75) - Reference category	-	-	-	-
DMM Cluster 2 (n = 29)	Breastfed (in weeks)	0.998	0.983-1.014	8.27E-01
	C-section delivery	1.610	0.520 - 4.974	4.10E-01
	Eczema mother	1.574	0.476 - 5.204	4.57E-01
	Asthma father	0.685	0.188 - 2.497	5.67E-01
	1 older sibling	2.097	0.815 - 5.398	1.25E-01
	≥ 2 older siblings	2.931	0.645 - 13.315	1.64E-01
DMM Cluster 3 (n = 76)	Breastfed (in weeks)	0.997	0.985 - 1.010	6.81E-01
	C-section delivery	3.346	1.488 - 7.528	3.00E-03
	Eczema mother	2.238	0.895 - 5.596	8.50E-02
	Asthma father	0.700	0.274 - 1.787	4.56E-01
	1 older sibling	1.019	0.487 - 2.133	9.60E-01
	≥ 2 older siblings	0.765	0.180 - 3.254	7.17E-01
DMM Cluster 4 (n = 20)	Breastfed (in weeks)	0.965	0.938 - 0.992	1.20E-02
	C-section delivery	0.242	0.028 - 2.109	1.99E-01
	Eczema mother	2.564	0.713 - 9,220	1.49E-01
	Asthma father	3.338	1.006 - 11.075	4.90E-02
	1 older sibling	2.527	0.802 - 7.956	1.13E-01
	≥ 2 older siblings	2.815	0.443 - 17.903	2.73E-01
DMM Cluster 5 (n = 11)	Breastfed (in weeks)	0.990	0.964 - 1.017	4.57E-01
	C-section delivery	6.347	1.467 - 27.456	1.30E-02
	Eczema mother	0,808	0.105 - 6.212	8.37E-01
	Asthma father	0.931	0.140 - 6.175	9.41E-01
	1 older sibling	1.635	0.323 - 8.277	5.52E-01
	≥ 2 older siblings	16.669	2.535 - 109.587	3.00E-03
DMM Cluster 6 (n = 14)	Breastfed (in weeks)	1.013	0.993 - 1.034	1.98E-01
	C-section delivery	1.371	0.310 - 6.065	6.78E-01
	Eczema mother	2.233	0.514 - 9.703	2.84E-01
	Asthma father	3.455	0.937 - 12.748	6.30E-02
	1 older sibling	0.343	0.066 - 1.783	2.03E-01
	≥ 2 older siblings	0.502	0.045 - 5.555	5.74E-01

SUPPLEMENTARY TABLE S5		FINAL MULTIVARIABLE MULTINOMIAL LOGISTIC REGRESSION MODEL ON COVARIATES IN ASSOCIATION TO MOST COMMON DMM CLUSTER TRAJECTORIES		
	Variable	Exp(B)	95% Confidence Interval for Exp (B)	P-value
DMM Cluster trajectory 1 -> 1 (n =26) - Reference category	-	-	-	-
DMM Cluster trajectory 1 -> 4 (n = 22)	Breastfed (in weeks)	0.959	0.928 - 0.992	1.40E-02
	Age introduction solids (weeks)	0.943	0.774-1.148	5.57E-01
	Age at start day care (months)	0.958	0.876 - 1.048	3.53E-01
DMM Cluster trajectory 1 -> 5 (n = 14)	Breastfed (in weeks)	0.948	0.909 - 0.989	1.20E-02
	Age introduction solids (weeks)	0.818	0.660 - 1.014	6.70E-02
	Age at start day care (months)	1.004	0.909 - 1.108	0./944
DMM Cluster trajectory 3 -> 1 (n = 16)	Breastfed (in weeks)	0.980	0.950 - 1.010	1.92E-01
	Age introduction solids (weeks)	1.074	0.876 - 1.315	4.93E-01
	Age at start day care (months)	0.981	0.890 - 1.081	6.95E-01
DMM Cluster trajectory 3 -> 3 (n = 11)	Breastfed (in weeks)	0.997	0.958 - 1.037	8.69E-01
	Age introduction solids (weeks)	0.836	0.636 - 1.100	2.01E-01
	Age at start day care (months)	1.130	1.016 - 1.257	2.50E-02
DMM Cluster trajectory 3 -> 4 (n = 13)	Breastfed (in weeks)	0.939	0.896 - 0.985	1.00E-02
	Age introduction solids (weeks)	0.776	0.618 - 0.974	2.80E-02
	Age at start day care (months)	0.970	0.865 - 1.087	5.97E-01
DMM Cluster trajectory 3 -> 5 (n = 25)	Breastfed (in weeks)	0.949	0.916 - 0.983	4.00E-03
	Age introduction solids (weeks)	0.905	0.744 - 1.102	3.21E-01
	Age at start day care (months)	1.009	0.925 - 1.099	8.44E-01

Supplementary Table S6: Multivariate Analysis by Linear Models (MaAsLin) on demographic, lifestyle and medical factors associated with genus level taxa (only results with statistically significant associations upon FDR-correction are presented)								
Age	Variable	Feature (genus)	Value	Coefficient	N	N.not .0	P-value	FDR-adjusted P-value (Q)
5 weeks	Birth Mode	Bacteroides	Vaginal delivery	0.175837645	297	297	1.38E-05	8.04E-03
5 weeks	Older siblings	Bifidobacterium	Older siblings in households	0.166120378	297	297	1.55E-05	8.04E-03
5 weeks	Older siblings	Corynebacterium	Older siblings in households	0.002526116	297	138	2.06E-05	8.04E-03
13 weeks	Cessation breastfeeding prior to sample collection	Pantoea	Breastfeeding ceased	0.00131711	286	79	2.96E-05	4.16E-02
21 weeks	Birth Mode	Bacteroides	Vaginal delivery	0.14714048	267	267	7.16E-04	6.28E-02
21 weeks	Birth Mode	Campylobacter	Vaginal delivery	-0.001529738	267	77	1.11E-03	8.25E-02
21 weeks	Cessation breastfeeding prior to sample collection	Other	Breastfeeding ceased	0.139999237	267	267	1.51E-09	1.72E-06
21 weeks	Cessation breastfeeding prior to sample collection	Lachnobacterium	Breastfeeding ceased	0.003857002	267	145	2.31E-09	1.72E-06
21 weeks	Cessation breastfeeding prior to sample collection	Pseudobutyrvibrio	Breastfeeding ceased	0.002832593	267	95	8.08E-08	4.02E-05
21 weeks	Cessation breastfeeding prior to sample collection	Bifidobacterium	Breastfeeding ceased	-0.197143826	267	267	7.22E-06	1.79E-03
21 weeks	Cessation breastfeeding prior to sample collection	Clostridium_C	Breastfeeding ceased	0.009197433	267	154	1.47E-05	3.14E-03
21 weeks	Cessation breastfeeding prior to sample collection	Oscillospira	Breastfeeding ceased	0.002240703	267	155	2.20E-05	4.10E-03
21 weeks	Cessation breastfeeding prior to sample collection	Eubacterium_A	Breastfeeding ceased	0.002266881	267	81	5.18E-05	8.59E-03
21 weeks	Cessation breastfeeding prior to sample collection	Ruminococcus	Breastfeeding ceased	0.007854753	267	194	6.28E-05	9.37E-03
21 weeks	Cessation breastfeeding prior to sample collection	Blautia	Breastfeeding ceased	0.002264146	267	203	2.04E-04	2.76E-02
21 weeks	Cessation breastfeeding prior to sample collection	Staphylococcus	Breastfeeding ceased	-0.004207764	267	195	2.93E-04	3.36E-02
21 weeks	Cessation breastfeeding prior to sample collection	Clostridium	Breastfeeding ceased	0.040367347	267	265	6.21E-04	6.28E-02
21 weeks	Cessation breastfeeding prior to sample collection	Megasphaera	Breastfeeding ceased	0.001435393	267	85	7.13E-04	6.28E-02
21 weeks	Cessation breastfeeding prior to sample collection	Enterococcus	Breastfeeding ceased	-0.015797733	267	263	1.11E-03	8.25E-02
21 weeks	Ethnicity	Proteus	non Caucasian	0.033164297	267	49	8.31E-04	6.88E-02
21 weeks	Older siblings	Eggerthella	Older siblings in households	0.006674972	267	124	4.98E-06	1.49E-03
21 weeks	Smoking_In_Pregnancy	Veillonella	Smoking_In_Pregnancy2	-0.088597495	267	267	2.60E-04	3.23E-02
21 weeks	Solids introduced prior to sample collection	Campylobacter	Solids introduced	0.002285898	267	77	2.94E-06	1.10E-03
21 weeks	Solids introduced prior to sample collection	Clostridium_C	Solids introduced	0.006088278	267	154	6.57E-04	6.28E-02
31 weeks	AD status at time of sample collection	Citrobacter	AD present	-0.0083787	304	288	1.19E-03	8.39E-02
31 weeks	Birth Mode	Bacteroides	Vaginal delivery	0.1860148	304	304	4.30E-05	5.14E-03
31 weeks	Birth Mode	Raoultella	Vaginal delivery	-0.0064118	304	294	1.05E-03	7.75E-02
31 weeks	Birth weight (in grams)	Corynebacterium	Birth weight (in grams)	1.22E-06	304	93	4.53E-04	4.14E-02
31 weeks	Cessation breastfeeding prior to sample collection	Oscillospira	Breastfeeding ceased	0.04809638	304	209	1.18E-17	1.83E-14
31 weeks	Cessation breastfeeding prior to sample collection	Ruminococcus	Breastfeeding ceased	0.06942166	304	259	1.04E-15	8.05E-13

31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Other	Breastfeeding ceased	0.1276393	304	304	1.13E-11	5.86E-09
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Pseudobutyrvibrio	Breastfeeding ceased	0.00440686	304	147	1.53E-10	5.95E-08
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Roseburia	Breastfeeding ceased	0.00412616	304	60	6.19E-10	1.92E-07
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Lachnobacterium	Breastfeeding ceased	0.00304302	304	182	2.20E-09	5.71E-07
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Coprobacillus	Breastfeeding ceased	0.00542691	304	31	4.50E-09	9.99E-07
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Eubacterium_A	Breastfeeding ceased	0.00411655	304	156	6.38E-09	1.24E-06
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Bifidobacterium	Breastfeeding ceased	-0.1679934	304	304	1.06E-07	1.82E-05
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Bacteroides	Breastfeeding ceased	0.20501827	304	304	6.47E-06	1.00E-03
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Clostridium_C	Breastfeeding ceased	0.01510415	304	223	8.10E-06	1.14E-03
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Staphylococcus	Breastfeeding ceased	-0.0034376	304	216	3.92E-05	5.07E-03
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Actinomyces	Breastfeeding ceased	-0.0055536	304	276	1.61E-04	1.67E-02
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Streptococcus	Breastfeeding ceased	-0.043623	304	304	3.06E-04	2.97E-02
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Blautia	Breastfeeding ceased	0.00395886	304	231	8.53E-04	6.98E-02
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Escherichia	Breastfeeding ceased	-0.135552	304	304	9.71E-04	7.54E-02
31 weeks	Cessation_breastfeeding_prior_to_sample_collectio	Veillonella	Breastfeeding ceased	-0.0786019	304	304	1.41E-03	9.50E-02
31 weeks	Smoking during pregnancy	Bilophila	Yes	-0.0073827	304	49	5.82E-04	5.02E-02
31 weeks	Solids introduced prior to sample collection	Veillonella	Solids introduced	0.14597588	304	304	4.96E-05	5.50E-03

Supplementary Table S7: PERMANOVA on demographic, dietary, lifestyle and medical factors associated with microbial community structure

Table S7: Permutational analysis of the variance at age 5 weeks (N = 238)					
Variable	Df	SumOfSqs	R2	F	P-value
Asthma Mother	1	0.049	0.002	0.556	9.59E-01
Ethnicity	1	0.063	0.003	0.716	8.13E-01
Birth Mode	2	0.291	0.014	1.654	1.70E-02
Day care	1	0.095	0.005	1.079	3.22E-01
Smoke DP	1	0.112	0.005	1.276	1.69E-01
Smoke AP	1	0.102	0.005	1.157	2.56E-01
Breastfed	1	0.094	0.004	1.066	3.90E-01
Residual	229	20.169	0.962		
Total	237	20.976	1.000		

Table S7: Permutational analysis of the variance at age 13 weeks (N = 233)					
Variable	Df	SumOfSqs	R2	F	P-value
Asthma Mother	1	0.082	0.004	0.997	4.47E-01
Ethnicity	1	0.074	0.004	0.903	5.69E-01
Birth Mode	2	0.169	0.009	1.025	4.07E-01
Day care	1	0.099	0.005	1.202	2.25E-01
Smoke DP	1	0.107	0.005	1.299	1.74E-01
Smoke AP	1	0.122	0.006	1.486	7.39E-02
Breastfed	1	0.305	0.016	3.717	9.99E-04
Solids	1	0.119	0.006	1.448	9.19E-02
AD	1	0.085	0.004	1.038	3.81E-01
Residual	222	18.241	0.940		
Total	232	19.403	1.000		

Table S7: Permutational analysis of the variance at age 21 weeks (N = 231)

Variable	Df	SumOfSqs	R2	F	P-value
Asthma Mother	1	0.103	0.006	1.360	1.25E-01
Ethnicity	1	0.113	0.006	1.498	8.19E-02
Birth Mode	2	0.171	0.009	1.132	2.36E-01
Day care	1	0.099	0.006	1.319	1.33E-01
Smoke DP	1	0.111	0.006	1.466	7.79E-02
Smoke AP	1	0.099	0.005	1.312	1.27E-01
Breastfed	1	0.468	0.026	6.212	9.99E-04
Solids	1	0.103	0.006	1.370	1.15E-01
AD	1	0.135	0.007	1.785	2.00E-02
Residual	220	16.589	0.922		
Total	230	17.991	1.000		

Table S7: Permutational analysis of the variance at age 31 weeks (N = 259)

Variable	Df	SumOfSqs	R2	F	P-value
Asthma Mother	1	0.046	0.002	0.516	9.84E-01
Ethnicity	1	0.148	0.006	1.647	4.80E-02
Birth Mode	2	0.244	0.010	1.359	4.90E-02
Day care	1	0.098	0.004	1.093	2.93E-01
Smoke DP	1	0.152	0.006	1.694	3.70E-02
Smoke AP	1	0.078	0.003	0.871	6.28E-01
Breastfed	1	0.663	0.027	7.377	9.99E-04
Solids	1	0.180	0.007	1.999	2.00E-03
AD	1	0.241	0.010	2.679	9.99E-04
Residual	248	22.287	0.923		
Total	258	24.137	1.000		

Supplementary Table S8: Multivariable linear regression analyses on demographic, dietary, lifestyle and medical factors associated with microbial maturity (Microbial Age z-scores)

Table S8: MAZ index at age 5 weeks (N = 235)							
	Unstandardized Coefficients		Standardized Coefficients	t	P-value	95,0% Confidence Interval for B	
		Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	-2.378	5.265		-0.452	6.520E-01	-12.757	8.000
Birhmode assisted vaginal	-0.476	0.680	-0.040	-0.701	4.840E-01	-1.816	0.863
Birhmode cesarean section	0.711	0.367	0.111	1.937	5.400E-02	-0.013	1.434
1 older sibling	-0.362	0.371	-0.060	-0.976	3.300E-01	-1.092	0.369
≥ 2 older siblings	0.763	0.620	0.078	1.230	2.200E-01	-0.459	1.984
Sex	-0.287	0.320	-0.051	-0.897	3.710E-01	-0.918	0.344
Birth Weight	-0.001	0.000	-0.077	-1.220	2.240E-01	-0.001	0.000
Gestational age	0.175	0.136	0.082	1.284	2.010E-01	-0.094	0.444
Breastfeeding duration	-0.010	0.007	-0.097	-1.507	1.330E-01	-0.024	0.003
Ethnicity	0.104	1.022	0.006	0.102	9.190E-01	-1.910	2.119
Maternal age delivery	-0.035	0.035	-0.060	-1.003	3.170E-01	-0.104	0.034
AD father	0.279	0.405	0.043	0.689	4.920E-01	-0.519	1.077
AD mother	-0.694	0.436	-0.103	-1.592	1.130E-01	-1.553	0.165
Asthma father	0.417	0.407	0.060	1.023	3.080E-01	-0.386	1.220
Asthma mother	0.894	0.366	0.152	2.444	1.500E-02	0.173	1.615
Smoking in pregnancy	0.364	0.393	0.056	0.928	3.540E-01	-0.410	1.138
Smoking after pregnancy	-0.168	0.401	-0.026	-0.418	6.760E-01	-0.959	0.623
AD	3.203	0.356	0.536	8.986	0.000E+00	2.500	3.905
Household with furry pets	0.179	0.339	0.030	0.528	5.980E-01	-0.490	0.848
Start of daycare in months	-0.014	0.021	-0.039	-0.647	5.180E-01	-0.055	0.028
Age of introduction solids (months)	-0.020	0.043	-0.030	-0.477	6.340E-01	-0.105	0.064
antibiotic use prior to sample collection	-0.116	1.717	-0.004	-0.068	9.460E-01	-3.500	3.268

Table S8: MAZ index at age 13 weeks (N = 231)

	Unstandardized Coefficients	Std. Error	Standardized Coefficients	t	P-value	95,0% Confidence Interval for B	
			Beta			Lower Bound	Upper Bound
(Constant)	5.589	5.543		1.008	3.140E-01	-5.338	16.516
Birhmode assisted vaginal	0.262	0.734	0.025	0.357	7.220E-01	-1.185	1.709
Birhmode cesarean section	0.337	0.384	0.060	0.878	3.810E-01	-0.420	1.094
1 older sibling	0.172	0.382	0.033	0.449	6.540E-01	-0.581	0.925
≥ 2 older siblings	0.598	0.674	0.071	0.887	3.760E-01	-0.731	1.927
Sex	-0.138	0.336	-0.028	-0.410	6.820E-01	-0.799	0.524
Birth Weight	-0.001	0.000	-0.147	-1.937	5.400E-02	-0.002	0.000
Gestational age	-0.059	0.143	-0.031	-0.409	6.830E-01	-0.341	0.224
Breastfeeding duration	-0.008	0.007	-0.087	-1.111	2.680E-01	-0.022	0.006
Ethnicity	-0.764	1.009	-0.054	-0.757	4.500E-01	-2.753	1.225
Maternal age delivery	0.019	0.037	0.038	0.523	6.010E-01	-0.054	0.092
AD father	-0.043	0.410	-0.008	-0.105	9.160E-01	-0.852	0.766
AD mother	0.211	0.469	0.036	0.451	6.530E-01	-0.714	1.137
Asthma father	0.724	0.450	0.115	1.610	1.090E-01	-0.162	1.610
Asthma mother	0.293	0.379	0.057	0.774	4.400E-01	-0.453	1.039
Smoking in pregnancy	0.296	0.400	0.054	0.740	4.600E-01	-0.492	1.084
Smoking after pregnancy	-0.349	0.403	-0.064	-0.867	3.870E-01	-1.144	0.445
AD	0.052	0.366	0.010	0.142	8.870E-01	-0.669	0.773
Household with furry pets	-0.136	0.347	-0.027	-0.391	6.960E-01	-0.820	0.549
Start of daycare in months	0.023	0.018	0.085	1.229	2.210E-01	-0.014	0.059
Age of introduction solids (months)	-0.005	0.045	-0.009	-0.113	9.100E-01	-0.093	0.083
antibiotic use prior to sample collection	-0.021	0.054	-0.026	-0.388	6.990E-01	-0.128	0.086

Table S8: MAZ index at age 21 weeks (N - 228)

	Unstandardized Coefficients	Std. Error	Standardized Coefficients	t	P-value	95,0% Confidence Interval for B	
			Beta			Lower Bound	Upper Bound
(Constant)	8.762	5.692		1.539	1.250E-01	-2.463	19.986
Birhmode assisted vaginal	-0.628	0.734	-0.056	-0.856	3.930E-01	-2.076	0.820
Birhmode cesarean section	-0.652	0.395	-0.110	-1.651	1.000E-01	-1.430	0.126
1 older sibling	-0.134	0.384	-0.025	-0.349	7.270E-01	-0.891	0.623
≥ 2 older siblings	0.985	0.632	0.113	1.558	1.210E-01	-0.262	2.232
Sex	0.727	0.335	0.143	2.171	3.100E-02	0.067	1.386
Birth Weight	0.000	0.000	0.004	0.055	9.560E-01	-0.001	0.001
Gestational age	-0.105	0.150	-0.052	-0.700	4.850E-01	-0.400	0.190
Breastfeeding duration	-0.004	0.007	-0.040	-0.542	5.880E-01	-0.018	0.010
Ethnicity	-4.363	1.117	-0.255	-3.907	0.000E+00	-6.565	-2.161
Maternal age delivery	-0.047	0.036	-0.091	-1.289	1.990E-01	-0.118	0.025
AD father	-0.437	0.417	-0.076	-1.049	2.960E-01	-1.260	0.385
AD mother	-0.243	0.465	-0.039	-0.522	6.020E-01	-1.160	0.674
Asthma father	0.072	0.429	0.012	0.167	8.680E-01	-0.775	0.918
Asthma mother	0.838	0.382	0.156	2.191	3.000E-02	0.084	1.592
Smoking in pregnancy	-0.504	0.421	-0.083	-1.198	2.320E-01	-1.333	0.325
Smoking after pregnancy	1.106	0.419	0.189	2.637	9.000E-03	0.279	1.933
AD**	0.917	0.358	0.172	2.559	1.100E-02	0.210	1.623
Household with furry pets	0.215	0.362	0.039	0.593	5.540E-01	-0.499	0.929
Start of daycare in months	0.002	0.019	0.008	0.114	9.090E-01	-0.035	0.039
Age of introduction solids (months)	-0.018	0.044	-0.029	-0.411	6.820E-01	-0.104	0.068
antibiotic use prior to sample collection	-1.055	0.676	-0.107	-1.561	1.200E-01	-2.388	0.278

Table S8: MAZ index at age 31 weeks (N = 253)

	Unstandardized Coefficients	Std. Error	Standardized Coefficients	t	P-value	95,0% Confidence Interval for B	
			Beta			Lower Bound	Upper Bound
(Constant)	-2.944	4.998		-0.589	5.560E-01	-12.792	6.904
Birhmode assisted vaginal	1.245	0.686	0.113	1.814	7.100E-02	-0.107	2.597
Birhmode cesarean section	0.254	0.354	0.045	0.719	4.730E-01	-0.443	0.952
1 older sibling	-0.361	0.353	-0.069	-1.024	3.070E-01	-1.056	0.334
≥ 2 older siblings	-0.273	0.581	-0.033	-0.470	6.390E-01	-1.418	0.871
Sex	-0.072	0.306	-0.015	-0.236	8.130E-01	-0.676	0.531
Birth Weight	-0.001	0.000	-0.097	-1.408	1.600E-01	-0.001	0.000
Gestational age	0.077	0.130	0.041	0.592	5.540E-01	-0.179	0.332
Breastfeeding duration**	-0.018	0.007	-0.192	-2.724	7.000E-03	-0.031	-0.005
Ethnicity	1.271	0.947	0.086	1.342	1.810E-01	-0.594	3.136
Maternal age delivery*	0.072	0.034	0.141	2.130	3.400E-02	0.005	0.139
AD father	-0.551	0.384	-0.099	-1.434	1.530E-01	-1.309	0.206
AD mother	-0.384	0.421	-0.064	-0.912	3.630E-01	-1.214	0.446
Asthma father	-0.329	0.398	-0.054	-0.827	4.090E-01	-1.114	0.455
Asthma mother	-0.526	0.348	-0.101	-1.511	1.320E-01	-1.211	0.160
Smoking in pregnancy	-0.195	0.373	-0.034	-0.523	6.020E-01	-0.930	0.540
Smoking after pregnancy	0.131	0.373	0.023	0.351	7.260E-01	-0.605	0.866
AD**	-0.977	0.332	-0.188	-2.945	4.000E-03	-1.630	-0.323
Household with furry pets	0.287	0.322	0.055	0.890	3.740E-01	-0.348	0.921
Start of daycare in months	-0.030	0.017	-0.113	-1.798	7.400E-02	-0.063	0.003
Age of introduction solids (months)	-0.012	0.040	-0.021	-0.308	7.590E-01	-0.092	0.067
antibiotic use prior to sample collection	0.431	0.502	0.055	0.860	3.910E-01	-0.557	1.420

Table S9: Shannon index at age 5 weeks (N = 235)

			Standardized Coefficients	t	P-value	95,0% Confidence Interval for B	
Unstandardized Coefficients	Std. Error	Beta				Lower Bound	Upper Bound
(Constant)	0.984	1.142		0.861	3.900E-01	-1.267	3.234
Birhmode assisted vaginal	0.075	0.153	0.032	0.487	6.270E-01	-0.227	0.376
Birhmode cesarean section	-0.025	0.081	-0.020	-0.314	7.540E-01	-0.184	0.134
1 older sibling	0.057	0.080	0.050	0.722	4.710E-01	-0.099	0.214
≥ 2 older siblings	0.006	0.134	0.003	0.047	9.620E-01	-0.257	0.270
Sex	0.045	0.070	0.042	0.636	5.250E-01	-0.093	0.183
Birth Weight	0.000	0.000	0.080	1.110	2.680E-01	0.000	0.000
Gestational age	0.023	0.030	0.056	0.765	4.450E-01	-0.036	0.082
Breastfeeding duration	-0.001	0.002	-0.065	-0.875	3.820E-01	-0.004	0.002
Ethnicity	0.017	0.220	0.005	0.078	9.380E-01	-0.417	0.451
Maternal age delivery	0.003	0.008	0.030	0.441	6.590E-01	-0.012	0.018
AD father	0.019	0.088	0.015	0.212	8.320E-01	-0.154	0.191
AD mother	-0.183	0.097	-0.140	-1.888	6.000E-02	-0.374	0.008
Asthma father	0.083	0.091	0.062	0.917	3.600E-01	-0.095	0.261
Asthma mother	-0.068	0.079	-0.060	-0.860	3.900E-01	-0.224	0.088
Smoking in pregnancy	-0.029	0.087	-0.023	-0.331	7.410E-01	-0.199	0.142
Smoking after pregnancy	0.014	0.084	0.012	0.166	8.680E-01	-0.152	0.180
AD	-0.004	0.076	-0.004	-0.055	9.560E-01	-0.154	0.145
Household with furry pets	-0.146	0.075	-0.127	-1.953	5.200E-02	-0.293	0.001
Start of daycare in months	0.002	0.004	0.027	0.413	6.800E-01	-0.006	0.010
Age of introduction solids (months)	-0.010	0.009	-0.078	-1.084	2.800E-01	-0.028	0.008
antibiotic use prior to sample collection	-0.025	0.281	-0.006	-0.090	9.280E-01	-0.580	0.529

Table S9: Shannon index at age 13 weeks (N = 231)

	Unstandardized Coefficients Std. Error		Standardized Coefficients	t	P-value	95,0% Confidence Interval for B	
			Beta			Lower Bound	Upper Bound
(Constant)	4.139	1.268		3.264	1.000E-03	1.639	6.639
Birhmode assisted vaginal	0.184	0.168	0.076	1.098	2.730E-01	-0.147	0.515
Birhmode cesarean section	0.031	0.088	0.025	0.358	7.210E-01	-0.142	0.205
1 older sibling	0.005	0.087	0.004	0.054	9.570E-01	-0.168	0.177
≥ 2 older siblings	0.157	0.154	0.081	1.019	3.090E-01	-0.147	0.461
Sex	0.018	0.077	0.016	0.236	8.140E-01	-0.133	0.170
Birth Weight	0.000	0.000	-0.119	-1.573	1.170E-01	0.000	0.000
Gestational age	-0.034	0.033	-0.079	-1.035	3.020E-01	-0.099	0.031
Breastfeeding duration	-0.001	0.002	-0.046	-0.587	5.580E-01	-0.004	0.002
Ethnicity	0.100	0.231	0.031	0.431	6.670E-01	-0.356	0.555
Maternal age delivery	-0.004	0.008	-0.036	-0.493	6.230E-01	-0.021	0.013
AD father	0.020	0.094	0.016	0.208	8.350E-01	-0.166	0.205
AD mother	0.124	0.107	0.092	1.154	2.500E-01	-0.088	0.336
Asthma father	-0.012	0.103	-0.008	-0.113	9.100E-01	-0.214	0.191
Asthma mother	-0.006	0.087	-0.005	-0.064	9.490E-01	-0.176	0.165
Smoking in pregnancy	-0.053	0.091	-0.042	-0.583	5.610E-01	-0.234	0.127
Smoking after pregnancy	-0.035	0.092	-0.028	-0.380	7.040E-01	-0.217	0.147
AD	-0.074	0.084	-0.062	-0.888	3.760E-01	-0.239	0.091
Household with furry pets	-0.034	0.079	-0.029	-0.429	6.680E-01	-0.191	0.123
Start of daycare in months	-0.006	0.004	-0.101	-1.452	1.480E-01	-0.014	0.002
Age of introduction solids (months)	-0.006	0.010	-0.047	-0.605	5.460E-01	-0.026	0.014
antibiotic use prior to sample collection	0.001	0.012	0.006	0.094	9.250E-01	-0.023	0.026

Table S9: Shannon index at age 21 weeks (N = 228)

	Unstandardized Coefficients Std. Error		Standardized Coefficients	t	P-value	95,0% Confidence Interval for B	
			Beta			Lower Bound	Upper Bound
(Constant)	3.346	1.192		2.807	5.000E-03	0.996	5.696
Birhmode assisted vaginal	-0.156	0.154	-0.067	-1.009	3.140E-01	-0.460	0.148
Birhmode cesarean section	-0.074	0.082	-0.061	-0.902	3.680E-01	-0.236	0.088
1 older sibling	0.042	0.081	0.038	0.526	6.000E-01	-0.117	0.201
≥ 2 older siblings	0.005	0.133	0.003	0.037	9.710E-01	-0.257	0.267
Sex	0.083	0.070	0.079	1.183	2.380E-01	-0.055	0.222
Birth Weight	0.000	0.000	-0.056	-0.757	4.500E-01	0.000	0.000
Gestational age	0.004	0.031	0.009	0.122	9.030E-01	-0.058	0.065
Breastfeeding duration**	-0.004	0.001	-0.199	-2.685	8.000E-03	-0.007	-0.001
Ethnicity	-0.207	0.235	-0.058	-0.882	3.790E-01	-0.669	0.255
Maternal age delivery	0.000	0.008	-0.002	-0.024	9.810E-01	-0.015	0.015
AD father	-0.111	0.088	-0.094	-1.265	2.070E-01	-0.284	0.062
AD mother	-0.069	0.097	-0.054	-0.712	4.780E-01	-0.260	0.122
Asthma father	-0.012	0.089	-0.010	-0.137	8.910E-01	-0.188	0.164
Asthma mother	0.059	0.080	0.053	0.729	4.670E-01	-0.100	0.217
Smoking in pregnancy	-0.058	0.087	-0.046	-0.661	5.090E-01	-0.230	0.115
Smoking after pregnancy	-0.060	0.088	-0.050	-0.686	4.930E-01	-0.233	0.113
AD	-0.100	0.075	-0.091	-1.334	1.840E-01	-0.248	0.048
Household with furry pets	0.036	0.076	0.032	0.477	6.340E-01	-0.113	0.186
Start of daycare in months	0.003	0.004	0.060	0.882	3.790E-01	-0.004	0.011
Age of introduction solids (months)**	-0.028	0.009	-0.224	-3.086	2.000E-03	-0.046	-0.010
antibiotic use prior to sample collection	0.109	0.138	0.055	0.793	4.290E-01	-0.162	0.380

Table S9: Shannon index at age 31 weeks (N = 253)

	Unstandardized Coefficients		Standardized Coefficients	t	P-value	95,0% Confidence Interval for B	
			Beta			Lower Bound	Upper Bound
(Constant)	2.551	1.064		2.399	1.700E-02	0.456	4.647
Birhmode assisted vaginal	-0.012	0.142	-0.005	-0.082	9.340E-01	-0.292	0.269
Birhmode cesarean section	0.009	0.076	0.008	0.122	9.030E-01	-0.140	0.158
1 older sibling	0.104	0.075	0.091	1.386	1.670E-01	-0.044	0.251
≥ 2 older siblings*	0.246	0.124	0.136	1.978	4.900E-02	0.001	0.491
Sex	-0.029	0.065	-0.028	-0.448	6.550E-01	-0.157	0.099
Birth Weight	0.000	0.000	0.017	0.248	8.040E-01	0.000	0.000
Gestational age	0.021	0.028	0.051	0.757	4.500E-01	-0.033	0.075
Breastfeeding duration**	-0.005	0.001	-0.238	-3.449	1.000E-03	-0.008	-0.002
Ethnicity	-0.202	0.203	-0.063	-0.996	3.200E-01	-0.602	0.198
Maternal age delivery	-0.007	0.007	-0.060	-0.915	3.610E-01	-0.021	0.008
AD father	-0.073	0.082	-0.060	-0.885	3.770E-01	-0.235	0.089
AD mother	0.072	0.089	0.056	0.810	4.190E-01	-0.104	0.249
Asthma father	0.092	0.084	0.071	1.106	2.700E-01	-0.072	0.257
Asthma mother	0.068	0.074	0.061	0.918	3.600E-01	-0.078	0.215
Smoking in pregnancy	0.031	0.079	0.025	0.391	6.960E-01	-0.125	0.187
Smoking after pregnancy	0.002	0.079	0.002	0.025	9.800E-01	-0.155	0.158
AD	-0.178	0.070	-0.158	-2.525	1.200E-02	-0.317	-0.039
Household with furry pets	-0.010	0.069	-0.009	-0.148	8.820E-01	-0.145	0.125
Start of daycare in months	0.000	0.004	0.002	0.028	9.780E-01	-0.007	0.007
Age of introduction solids (months)**	-0.029	0.008	-0.232	-3.459	1.000E-03	-0.046	-0.013
antibiotic use prior to sample collection	-0.001	0.106	-0.001	-0.010	9.920E-01	-0.209	0.207

Supplementary Table S10: Joint modelling on shannon index and atopic dermatitis (N = 312)

Effect	Value	Std. Err.	z-value	p-value
Treatment (reference = placebo)	-0.084281561	0.232406141	-0.362647735	7.17E-01
Gender (reference = boy)	0.005917078	0.212608897	0.027830811	9.78E-01
AD father (reference = no)	0.806733579	0.28429961	2.837617606	4.55E-03
Breastfeeding duration (weeks)	0.00393551	0.003819348	1.030414062	3.03E-01
1 older sibling (reference = no sibs)	0.207578182	0.240966362	0.861440495	3.89E-01
≥ 2 olders siblings	0.002798779	0.378946852	0.007385678	9.94E-01
Furry pets in household (reference = no)	-0.098288506	0.234112117	-0.419835193	6.75E-01
Cesarean section birth mode (reference = natural vaginal delivery)	-0.398465081	0.500715684	-0.795791091	4.26E-01
Assisted vaginal delivery	0.009940472	0.450870722	0.022047279	9.82E-01
Birth weight (in grams)	0.126044132	0.269191957	0.468231419	6.40E-01
AD mother (reference = no)	0.651357836	0.301932622	2.157295331	3.10E-02
ASSOCIATION (Shannon index)	-1.578448198	0.40938711	-3.855637266	1.15E-04
log(xi.1)	-3.77880888	1.427664555	-2.646846465	8.12E-03
log(xi.2)	-3.912696513	1.556174201	-2.514304961	1.19E-02
log(xi.3)	-2.749007557	1.857350729	-1.480069173	1.39E-01
log(xi.4)	-0.188366114	1111.772811	-0.000169429	1.00E+00

Supplementary Table S11: Joint modelling on MAZ scores and atopic dermatitis (N = 312)

Effect	Value	Std. Err.	z-value	p-value
Treatment (reference = placebo)	-0.06407407	0.211695794	-0.30267049	7.62E-01
Gender (reference = boy)	0.034598927	0.211291403	0.16374981	8.70E-01
AD father (reference = no)	0.906944883	0.290099079	3.126328035	1.77E-03
Breastfeeding duration (weeks)	0.007557587	0.00364736	2.072070356	3.83E-02
1 older sibling (reference = no sibs)	0.154546605	0.238565208	0.647817031	5.17E-01
> 2 older siblings	-0.034512157	0.375489343	-0.091912482	9.27E-01
Furry pets in household (reference = no)	-0.078684706	0.23136654	-0.340086798	7.34E-01
Cesarean section birth mode (reference = natural vaginal delivery)	-0.12821753	0.474555339	-0.27018457	7.87E-01
Assisted vaginal delivery	-0.343478873	0.267486101	-1.284099886	1.99E-01
Birth weight (in grams)	0.510627325	0.264140586	1.933164954	5.32E-02
AD mother (reference = no)	0.766246518	0.305654581	2.506903427	1.22E-02
ASSOCIATION (MAZ scores)	0.127799617	0.029913897	4.272249057	1.94E-05
log(xi.1)	-4.595666187	1.097884853	-4.185927309	2.84E-05
log(xi.2)	-4.671636806	1.098344847	-4.253342489	2.11E-05
log(xi.3)	-2.512362383	1.302212452	-1.929302995	5.37E-02
log(xi.4)	-7.065394817	8240.266544	-0.000857423	9.99E-01

Supplementary Table S12: MetaLonDa Analysis Output in association to Atopic Dermatitis

Bacterial Genera	start (days)	end (days)	dominant	FDR adjusted p-value
Corynebacterium	126.110	151.240	HC	9.68E-03
Atopobium	25.590	79.440	HC	7.65E-03
Prevotella	104.570	133.290	HC	0.00E+00
Lachnobacterium	25.590	61.490	HC	1.18E-02
Lachnobacterium	108.160	129.700	HC	6.49E-03
Lachnobacterium	208.680	269.710	HC	5.00E-03
Faecalibacterium	47.130	201.500	HC	5.08E-03

Supplementary Table S13: Logistic regression model of microbial diversity (Shannon index) in association to allergic sensitization at school-age

Coefficients	Estimate	Std. Error	z value	P-value
(Intercept)	5.627791598	3.674612075	1.531533529	1.26E-01
Shannon at 5 weeks	-0.410285084	0.454296315	-0.903122193	3.66E-01
Shannon at 13 weeks	-0.208920785	0.527504064	-0.396055309	6.92E-01
Shannon at 21 weeks	0.279310503	0.601239738	0.464557623	6.42E-01
Shannon at 31 weeks	-1.620866258	0.604467106	-2.681479675	7.33E-03
Breastfeeding duration (weeks)	-3.030233351	0.873149718	-3.470462497	5.20E-04
Solid food introduction (week)	6.294484319	3.381347875	1.861531126	6.27E-02
Cesarean section	-0.238619825	1.081714761	-0.220594036	8.25E-01
Assisted vaginal delivery	-0.94531093	0.626875226	-1.507973023	1.32E-01
Birth Weight (g)	-0.000366139	0.00065891	-0.555673949	5.78E-01
Sex Female (reference = male)	-0.527185935	0.500811149	-1.052664135	2.92E-01
Treatment Control (reference = yes)	0.742100392	0.503761986	1.473117091	1.41E-01
No AD Father (reference=Yes)	-1.635728509	0.645015214	-2.535953375	1.12E-02
No AD Mother (reference=Yes)	-0.655147341	0.663647229	-0.987192159	3.24E-01
1 older sibling (reference = no sibs)	0.45591163	0.568647631	0.801747172	4.23E-01
≥ 2 older siblings	-6.44E-05	0.999108354	-6.44E-05	1.00E+00
Furry pets in household (reference = no)	-0.72774705	0.548034524	-1.327921908	1.84E-01

Supplementary Table S14: Logistic regression model of microbial maturity (MAZ-score) in association to allergic sensitization at school-age

Coefficients	Estimate	Std. Error	z value	P-value
(Intercept)	-0.087464637	2.640818493	-0.033120276	9.74E-01
MAZ at 5 weeks	0.382210056	0.136202238	2.806195127	5.01E-03
MAZ at 13 weeks	-0.164301672	0.140623631	-1.16837882	2.43E-01
MAZ at 21 weeks	0.050549523	0.182113301	0.277571833	7.81E-01
MAZ at 31 weeks	-0.016145267	0.133545263	-0.120897342	9.04E-01
Breastfeeding duration (weeks)	-3.026775739	0.884920175	-3.420394091	6.25E-04
Solid food introduction (week)	8.847895913	3.301639509	2.679849175	7.37E-03
Birth Weight (g)	-0.00039128	0.000649369	-0.602553926	5.47E-01
Treatment Control (reference = yes)	0.957059287	0.534603349	1.790223143	7.34E-02
Sex Female (reference = male)	-0.402875364	0.508750997	-0.791891054	4.28E-01
Cesarean section	-0.236701605	1.019008403	-0.232286215	8.16E-01
Assisted vaginal delivery	-0.842345125	0.656241797	-1.283589569	1.99E-01
No AD Father (reference=Yes)	-1.43835435	0.64026914	-2.246483953	2.47E-02
No AD Mother (reference=Yes)	-0.52709629	0.640499938	-0.822945108	4.11E-01
1 older sibling (reference = no sibs)	0.119864913	0.552340617	0.217012672	8.28E-01
≥ 2 older siblings	-0.745760306	1.048383644	-0.711342942	4.77E-01
Furry pets in household (reference = no)	-0.840188796	0.565073526	-1.486866323	1.37E-01

Supplementary Table S15: Logistic regression model of microbial diversity (Shannon index) in association to asthma at school-age				
Coefficients	Estimate	Std. Error	z value	P-value
(Intercept)	-9.519296394	5.319877537	-1.789382618	7.36E-02
Shannon at 5 weeks	-0.431558541	0.692669049	-0.623037137	5.33E-01
Shannon at 13 weeks	0.065089728	0.751687765	0.086591443	9.31E-01
Shannon at 21 weeks	-0.02811964	0.661640302	-0.04249989	9.66E-01
Shannon at 31 weeks	0.164692463	0.801999013	0.205352451	8.37E-01
Breastfeeding duration (weeks)	-0.254586729	0.764690582	-0.332927768	7.39E-01
Solid food introduction (week)	6.194863693	5.019759796	1.234095643	2.17E-01
Cesarean section	0.344889489	1.456595944	0.236777735	8.13E-01
Birth Weight (g)	0.001270255	0.00090951	1.396637011	1.63E-01
Assisted vaginal delivery	-0.825554535	0.948850266	-0.870057758	3.84E-01
Treatment Control (reference = yes)	-0.166071147	0.657007741	-0.252768936	8.00E-01
Sex Female (reference = male)	-2.051878298	0.859550668	-2.387152234	1.70E-02
No AD Father (reference=Yes)	0.085116272	0.823378904	0.103374365	9.18E-01
No AD Mother (reference=Yes)	-0.313860614	1.003653848	-0.31271799	7.54E-01
1 older sibling (reference = no sibs)	0.643148333	0.754705643	0.852184343	3.94E-01
≥ 2 older siblings	-15.68707851	1926.712049	-0.00814189	9.94E-01
Furry pets in household (reference = no)	1.483880523	0.934270057	1.588277941	1.12E-01

Supplementary Table S16: Logistic regression model of microbial maturity (MAZ-score) in association to asthma at school-age				
Coefficients	Estimate	Std. Error	z value	P-value
(Intercept)	-12.40821269	5.470392851	-2.268248923	2.33E-02
MAZ at 5 weeks	0.363442479	0.136558324	2.661445086	7.78E-03
MAZ at 13 weeks	-0.415956399	0.187305674	-2.220735705	2.64E-02
MAZ at 21 weeks	0.254280302	0.228921442	1.110775385	2.67E-01
MAZ at 31 weeks	-0.104036691	0.157184716	-0.661875362	5.08E-01
Breastfeeding duration (weeks)	-0.486254724	0.839838341	-0.578986099	5.63E-01
Solid food introduction (week)	10.08291775	5.955334706	1.69309002	9.04E-02
Cesarean section	1.480313854	1.561995355	0.94770695	3.43E-01
Assisted vaginal delivery	-1.1528046	1.0910663	-1.056585287	2.91E-01
Birth Weight (g)	0.001070665	0.000970668	1.103018401	2.70E-01
Treatment Control (reference = yes)	0.126012384	0.750777261	0.167842569	8.67E-01
Sex Female (reference = male)	-3.129532598	1.16397764	-2.688653536	7.17E-03
No AD Father (reference=Yes)	0.731989105	0.882583537	0.829370903	4.07E-01
No AD Mother (reference=Yes)	-1.282273155	1.392172865	-0.921058862	3.57E-01
1 older sibling (reference = no sibs)	0.961383693	0.768278099	1.25134856	2.11E-01
≥ 2 older siblings	-14.55804809	1838.156717	-0.007919917	9.94E-01
Furry pets in household (reference = no)	1.829645625	1.082872207	1.68962285	9.11E-02

Supplementary Table S17: MetaLonDa Analysis Output in association to Asthma

Bacterial Genera	start (days)	end (days)	dominant	FDR adjusted p-value
Actinomyces	183.040	192.000	HC	1.809E-02
Enterococcus	167.360	194.240	HC	3.307E-02
Streptococcus	44.160	129.280	Asthma	1.741E-02
Streptococcus	183.040	236.800	HC	3.156E-03
Sarcina	28.480	57.600	HC	8.062E-03
Sarcina	84.480	86.720	HC	4.844E-02
Sarcina	88.960	93.440	HC	4.521E-02
Lachnobacterium	28.480	100.160	HC	1.365E-02
Lachnobacterium	138.240	176.320	HC	4.618E-03
Lachnobacterium	221.120	236.800	HC	2.783E-02
Lachnospira	48.640	80.000	HC	1.717E-02
Lachnospira	91.200	200.960	HC	2.878E-03
Lachnospira	218.880	236.800	HC	1.040E-03
Dialister	28.480	122.560	HC	6.209E-04
Dialister	131.520	144.960	Asthma	1.378E-02
Dialister	151.680	176.320	HC	1.200E-02
Dialister	203.200	207.680	HC	4.823E-02
Dialister	216.640	236.800	HC	1.935E-02
Klebsiella	28.480	102.400	HC	6.922E-03
Leclercia	28.480	187.520	HC	2.746E-03
Haemophilus	115.840	236.800	HC	1.038E-02

Supplementary methods

Microbiota profiling

PCR reactions for amplicon sequencing consisted of 5pmol of each primer, 50ng template DNA, 200µM dNTPs, 1.5mM MgCl₂, 4mg/mL BSA, 1x reaction buffer, and 1 U Taq polymerase. The PCR protocol was as follows: 94°C for 5 minutes, followed by 30 cycles of 94°C for 30 seconds, 50°C for 30 seconds, and 72°C for 30 seconds, with a final 72°C for 7 minutes. Presence of a PCR product was verified by electrophoresis (2% agarose gel). PCR products were normalized for quantity using the SequalPrep Normalization kit (ThermoFisher #A10510-01).

Sequencing data were processed using the short-read library 16S rRNA gene sequencing pipeline (sl1p)¹. Briefly, primers and/or read-through of the variable 3 region was trimmed using cutadapt, forward and reverse reads were assembled using PANDAseq², low-quality reads were culled using sickle with a quality threshold of 30 (<https://github.com/najoshi/sickle>), and chimeras were removed using USEARCH as written into QIIME.^{3,4} Operational taxonomic units (OTUs) were generated by closed reference picking against the Greengenes reference database (February 4th 2011 release)⁵ using UCLUST.

Operationalisation of microbiota determinants

The potential determinants of microbiota composition that were examined included: maternal smoking during pregnancy (yes/no), maternal age at delivery (in years), gestational age (in weeks), ethnicity (Caucasian/other), birth weight (in grams), sex (boy/girl), birth mode (natural vaginal/assisted vaginal/cesarean section), maternal smoking after pregnancy (yes/no), number of older siblings (0, 1, ≥ 2), paternal/maternal eczema (yes/no), paternal/maternal asthma (yes/no), presence of furry pets in household (yes/no), duration of breastfeeding (in weeks), age at introduction of solid feeding (in weeks), antibiotic use before the age of 31 weeks (yes/no), age at start of day-care (in months).

Joint Models

A Joint Model (JM) is a powerful method that takes into account the dependency and association between longitudinal data and time-to-event data. The JM is basically composed of a Linear Mixed-effect Model (LMM) part and a survival part. The LMM was built using the Shannon index and the MAZ score, respectively, and using the age of the child at sampling, the duration of the breastfeeding, and the age of introduction of solid food as predicting variables. A Cox model was then built for the second part of the JM using the survival time as outcome and adjusting for treatment, gender, maternal and paternal AD, birth weight and mode, number of older siblings and presence of household furry pets.

These two models were then together used as input to the JM function of the *JM*package⁶ to train the joint model.

References

1. Whelan FJ, Surette MG. A comprehensive evaluation of the sl1p pipeline for 16S rRNA gene sequencing analysis. *Microbiome* 2017;5:100.
2. Masella AP, Bartram AK, Truszkowski JM, et al. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* 2012;13:31.

3. **Caporaso JG, Kuczynski J, Stombaugh J**, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335-6.
4. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460-1.
5. DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;72:5069-72.
6. Rizopoulos D. JM: An R Package for the Joint Modelling of Longitudinal and Time-to-event Data. *Journal of Statistical Software* 2010;35.

Author names in bold designate shared co-first authorship

What you need to know:

BACKGROUND AND CONTEXT: Establishment of the gastrointestinal microbiota during infancy affects immune system development and oral tolerance induction.

NEW FINDINGS: In a longitudinal study of fecal microbiota of children from 5 weeks through 6–11 years, we tracked changes in microbiota diversity and composition associated with birth mode, diet, and development of allergies and asthma.

LIMITATIONS: This was a post-hoc analysis of participants in an interventional study. Studies are needed to determine whether these microbes cause or are a result of the observed associations

IMPACT: Based on analyses of fecal samples, the intestinal microbiota increases in diversity from age 13 weeks through 6–11 years. Changes associate with vaginal vs cesarean section delivery, the end of breastfeeding, and development of allergies and asthma.

Lay Summary: We analyzed stool samples collected from children 5 weeks old through 6–11 years. We identified microbes that change in abundance based on vaginal vs cesarean birth, when breastfeeding ends, and in children that develop allergies or asthma.