DispatchDate: 05.10.2020 · ProofNo: 1095, p.1

nature medicine

ARTICLES

https://doi.org/10.1038/s41591-020-1095-x

Maturation of the gut microbiome during the first year of life contributes to the protective farm effect on childhood asthma

Martin Depner[1], Diana Hazard Taft[2], Pirkka V. Kirjavainen[3,4], Karen M. Kalanetra[2], Anne M. Karvonen[3], Stefanie Peschel[1], Elisabeth Schmausser-Hechfellner[1], Caroline Roduit[5,6,7], Remo Frei[5,8], Roger Lauener[5,7,9,10], Amandine Divaret-Chauveau[11,12,13], Jean-Charles Dalphin[13,25], Josef Riedler[14], Marjut Roponen[15], Michael Kabesch[16], Harald Renz[17,18], Juha Pekkanen[3,19], Freda M. Farquharson[20], Petra Louis[20], David Mills[2], Erika von Mutius[1,18,21], PASTURE study group* and Markus J. Ege[18,21] ✉

Growing up on a farm is associated with an asthma-protective effect, but the mechanisms underlying this effect are largely unknown. In the Protection against Allergy: Study in Rural Environments (PASTURE) birth cohort, we modeled maturation using 16S rRNA sequence data of the human gut microbiome in infants from 2 to 12 months of age. The estimated microbiome age (EMA) in 12-month-old infants was associated with previous farm exposure ($\beta = 0.27$ (0.12–0.43), $P = 0.001$, $n = 618$) and reduced risk of asthma at school age (odds ratio (OR) = 0.72 (0.56–0.93), $P = 0.011$). EMA mediated the protective farm effect by 19%. In a nested case–control sample ($n = 138$), we found inverse associations of asthma with the measured level of fecal butyrate (OR = 0.28 (0.09–0.91), $P = 0.034$), bacterial taxa that predict butyrate production (OR = 0.38 (0.17–0.84), $P = 0.017$) and the relative abundance of the gene encoding butyryl-coenzyme A (CoA):acetate-CoA-transferase, a major enzyme in butyrate metabolism (OR = 0.43 (0.19–0.97), $P = 0.042$). The gut microbiome may contribute to asthma protection through metabolites, supporting the concept of a gut–lung axis in humans.

Many diseases in adults originate early in life[1]. In the prenatal period, environmental influences that affect disease development are filtered by the mother. After birth, however, the infant interacts directly with the environment, beginning with the colonization of body surfaces by microbiota within the first hours of life[2]. This process consists of mutual adaptation between host and microbiota and ultimately educates the host's immune system[3]. Studies in gnotobiotic mice support an essential role for microbial exposure in the development of the immune system[4]. The inverse relationship of microbial exposure and immune-mediated diseases, such as allergies and asthma, is the basis for the hygiene hypothesis and its amendments explaining the epidemic of inflammatory diseases in a world that has abandoned traditional lifestyles[5].

A proposed mechanism by which a traditional lifestyle may grant strong protective effects against asthma involves sustained microbial exposure on farms[6,7]. This protective effect has mainly been attributed to the consumption of farm milk and exposure to a variety of environmental microbiota in animal sheds[6,8].

A highly diverse microbial environment may influence the human microbiome and thus mitigate asthma risk, as shown in the microbiome of the upper airways[9,10]. For the gut microbiome, the effect on airway disease is less obvious. Murine models have suggested that protection from allergic inflammation in the lung is mediated by the production of metabolites, such as short-chain fatty acids (SCFAs), by certain gut bacteria[11].

The human gut microbiome undergoes profound changes during the first year of life and starts stabilizing soon thereafter[12,13]. Hence, we hypothesized that the first year of life, in particular, represents a time window during which exposures to the outer environment shape the development of the human microbiome with possible

[1]Institute for Asthma and Allergy Prevention, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [2]Viticulture and Enology, University of California, Davis, CA, USA. [3]Department of Health Security, Finnish Institute for Health and Welfare, Kuopio, Finland. [4]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. [5]Christine Kühne Center for Allergy Research and Education (CK-CARE), Davos, Switzerland. [6]Children's Hospital, University of Zürich, Zürich, Switzerland. [7]Children's Hospital of Eastern Switzerland, St. Gallen, Switzerland. [8]Swiss Institute of Allergy and Asthma Research (SIAF), University of Zürich, Davos, Switzerland. [9]University of Zürich, Zürich, Switzerland. [10]School of Medicine, University of St. Gallen, St. Gallen, Switzerland. [11]Pediatric Allergy Department, Children's Hospital, University Hospital of Nancy, Vandoeuvre les Nancy, France. [12]EA 3450 DevAH, Faculty of Medicine, University of Lorraine, Vandoeuvre les Nancy, France. [13]Department of Respiratory Disease, UMR CNRS 6249 Chrono-environnement, University Hospital of Besançon, Besançon, France. [14]Children's Hospital, Schwarzach, Austria. [15]Department of Environmental and Biological Sciences, University of Eastern Finland, Kuopio, Finland. [16]KUNO Children's University Hospital Regensburg, Department of Pediatric Pneumology and Allergy, Regensburg, Germany. [17]Department of Clinical Chemistry and Molecular Diagnostics, Philipps University of Marburg, Marburg, Germany. [18]German Center for Lung Research, Marburg, Germany. [19]Department of Public Health, University of Helsinki, Helsinki, Finland. [20]The Rowett Institute, University of Aberdeen, Foresterhill, Aberdeen, UK. [21]Dr. von Hauner Children's Hospital, Ludwig Maximilian University of Munich, Munich, Germany. [25]Deceased. *A list of authors and their affiliations appears at the end of the paper.
✉e-mail: markus.ege@med.uni-muenchen.de

lasting consequences. The large population-based PASTURE birth cohort provided the unique opportunity to comprehensively assess farm-related environmental effects on the early gut microbiome and, through the gut microbiome, on respiratory health.

Results

The PASTURE birth cohort followed children in European rural areas, of whom 50% were born to mothers living on a family-run farm. In the Austrian, Finnish, German and Swiss arms of this study ($n=930$, 49% female), fecal samples were taken at months 2 and 12. At both time points, we obtained sequence reads for bacterial 16S rRNA and the fungal internal transcribed spacer (ITS) region from samples from 618 (66%) and 189 (20%) children, respectively, which represented all four centers at similar shares between 22% and 33% (Extended Data Fig. 1 and Supplementary Table 1). Asthma was defined as a physician's diagnosis of asthma or recurrent obstructive bronchitis established until 6 years of age and was present in 8.1% of the 930 children.

Bacterial composition at months 2 and 12. At month 2 (Fig. 1a), the genus *Bifidobacterium* was predominant. Despite a positive association of the abundance of *Bifidobacterium* with breastfeeding ($\beta=0.43$ (0.23–0.64), $P<0.001$), this genus was not significantly linked with subsequent asthma. At month 12 (Fig. 1b), the relative abundance of *Bifidobacterium* was halved, whereas the genus *Blautia* of the family Lachnospiraceae increased substantially in relative abundance. In addition, various other genera, including *Coprococcus*, *Faecalibacterium* and *Roseburia*, became detectable. By clustering bacterial composition data from both time points by Dirichlet mixture modeling, we identified five Dirichlet clusters (DCs), with two clusters mainly representing month 2 samples, two clusters representing month 12 samples and one cluster shared by both time points (Fig. 1c,d). The first two clusters (DC1 and DC2) were dominated by *Bifidobacterium*, whereas the third cluster (DC3) revealed considerable heterogeneity between samples, with various different taxa accounting for at least 1% of relative abundance (Fig. 1c and Supplementary Table 2). DC4 and DC5 demonstrated more stabilized bacterial patterns with the emergence of *Firmicutes*. In month 12 samples, children in this cluster showed a higher prevalence of asthma, as compared to those in clusters DC4 and DC5 (Fig. 1e).

Microbial maturation. To better understand the physiological changes of the gut microbiome during the first year, we modeled the exact age of fecal sampling by random forest analysis of the composition of bacterial genera at months 2 and 12 in individuals with samples available at both time points. Because this model estimates the biological age of the healthy microbiome, we termed the resulting prediction score estimated microbiome age EMA. To exclude disease interference, we restricted the model building to 133 healthy individuals (67 farm children and 66 nonfarm children) who did not have diarrhea between months 2 and 12 and were never affected by wheeze or asthma.

The taxa that contributed most importantly to the prediction model were *Blautia* and *Coprococcus* (Fig. 2a). When applying the prediction model to the entire population ($n=618$), the composition of genera did not vary notably with EMA at month 2 (Fig. 2b), whereas at month 12, a clear pattern emerged, with increases particularly in *Ruminococcus*, *Roseburia* and *Coprococcus* (Fig. 2c). When stratifying for month 2 and 12 samples, the correlation of EMA with the exact sampling time point was largely removed (Fig. 2d; $\rho=0.10$ and $\rho=0.15$ for month 2 and 12 samples, respectively), thereby indicating that EMA essentially reflects maturation from 2 to 12 months. DC3 comprised month 2 samples with advanced EMA and month 12 samples with delayed EMA (Fig. 2d), thereby grouping individuals who did not follow the typical maturation pattern. As an alternative surrogate for maturation, we explored a principal-coordinate analysis (PCoA) for both time points (Extended Data Fig. 2a), the first axis of which correlated strongly with EMA (Extended Data Fig. 2b).

Children with any form of asthma had, on average, significantly lower EMA values at month 12 (Fig. 2e). The prevalence of asthma was 12% in children with incomplete maturation, as defined by having an EMA value in the lowest quartile (Fig. 2e). Children with higher EMA values had a lower risk of asthma (OR = 0.48 (0.25–0.93), $P=0.030$) and lung function impairment (OR = 0.48 (0.27–0.82), $P=0.008$), when compared to children from the lowest quartile of EMA values. Similarly, as a continuous variable, higher EMA values correlated with a reduced risk of asthma (OR = 0.72 (0.56–0.93), $P=0.011$) and particularly well with that of nonatopic asthma (Fig. 2f). The effect of EMA on asthma was unchanged when adjusted for DC3 at month 12, whereas the effect of DC3 on asthma was largely removed when adjusting for EMA (Fig. 2f). The effect of EMA was more pronounced in carriers of the non-risk genotype on chromosome 17q21 (Supplementary Table 3) and was also observed in a sensitivity analysis that excluded the 133 children on whom the prediction model was established (Extended Data Fig. 2c,d). When predicting EMA at month 2, there was no clear association with asthma (OR = 1.24 (0.93–1.65), $P=0.135$).

Microbial maturation versus composition. We analyzed microbial composition using a principal-component analysis (PCA) approach designed for compositional data. At month 2, the third PCA axis exerted an asthma-protective effect (OR = 0.68 (0.49–0.95), $P=0.024$) irrespective of concomitant atopy (Extended Data Fig. 3a). This axis correlated positively with the relative abundance of *Bacteroides* and *Parabacteroides* and negatively with that of *Enterococcus* (Extended Data Fig. 3b).

At month 12, the first PCA axis was inversely related to nonatopic asthma (OR = 0.62 (0.39–1.00), $P=0.048$) and correlated particularly well with *Roseburia*, *Ruminococcus* and *Faecalibacterium* (Extended Data Fig. 3c,d). A sensitivity analysis based on a PCoA using unweighted UniFrac as the distance measurement corroborated these patterns (Extended Data Fig. 4).

EMA correlated strongly with PCA axis 1 at month 12 ($\rho=0.75$) and α-diversity ($\rho=0.70$ for richness) but not with PCA axis 3 at month 2 (Extended Data Fig. 5). EMA and PCA axis 3 at month 2 emerged as independent determinants of asthma in a mutually adjusted model (Extended Data Fig. 3e–g), whereas the effect of PCA axis 1 at month 12 was explained by EMA.

EMA and the farm effect on asthma. The PASTURE study was designed to assess the farm effect on asthma (OR = 0.53 (0.30–0.92), $P=0.023$, $n=930$). In the present subpopulation ($n=618$), farm children also had a lower risk of asthma as compared to rural nonfarm children (center-adjusted OR = 0.56 (0.29–1.08), $P=0.082$). At month 2, no effect of farm exposure on the microbial composition was observed, while the asthma-protective PCA axis 3 was positively associated with breastfeeding and inversely associated with Cesarean sections and maternal smoking during pregnancy (Fig. 3a), independently of gestational age. In contrast, EMA was delayed by prolonged breastfeeding (Extended Data Fig. 6) but positively influenced by growing up on a farm ($\beta=0.27$ (0.12–0.43), $P=0.001$) and particular farm exposures, such as visits to animal sheds or the consumption of milk or eggs directly obtained from a farm (Fig. 3b). The latter variables also reflect a more diverse feeding pattern in farm children, as illustrated by more frequent consumption of all six main food items, cereals, meat, bread, yogurt, cake and vegetables or fruits (Supplementary Table 4). A sensitivity analysis showed independent effects on EMA by a diverse feeding pattern ($\beta=0.18$ (0.01–0.34), $P=0.034$) and prolonged breastfeeding ($\beta=-0.41$ (−0.62 to −0.21), $P<0.001$). Farm children were

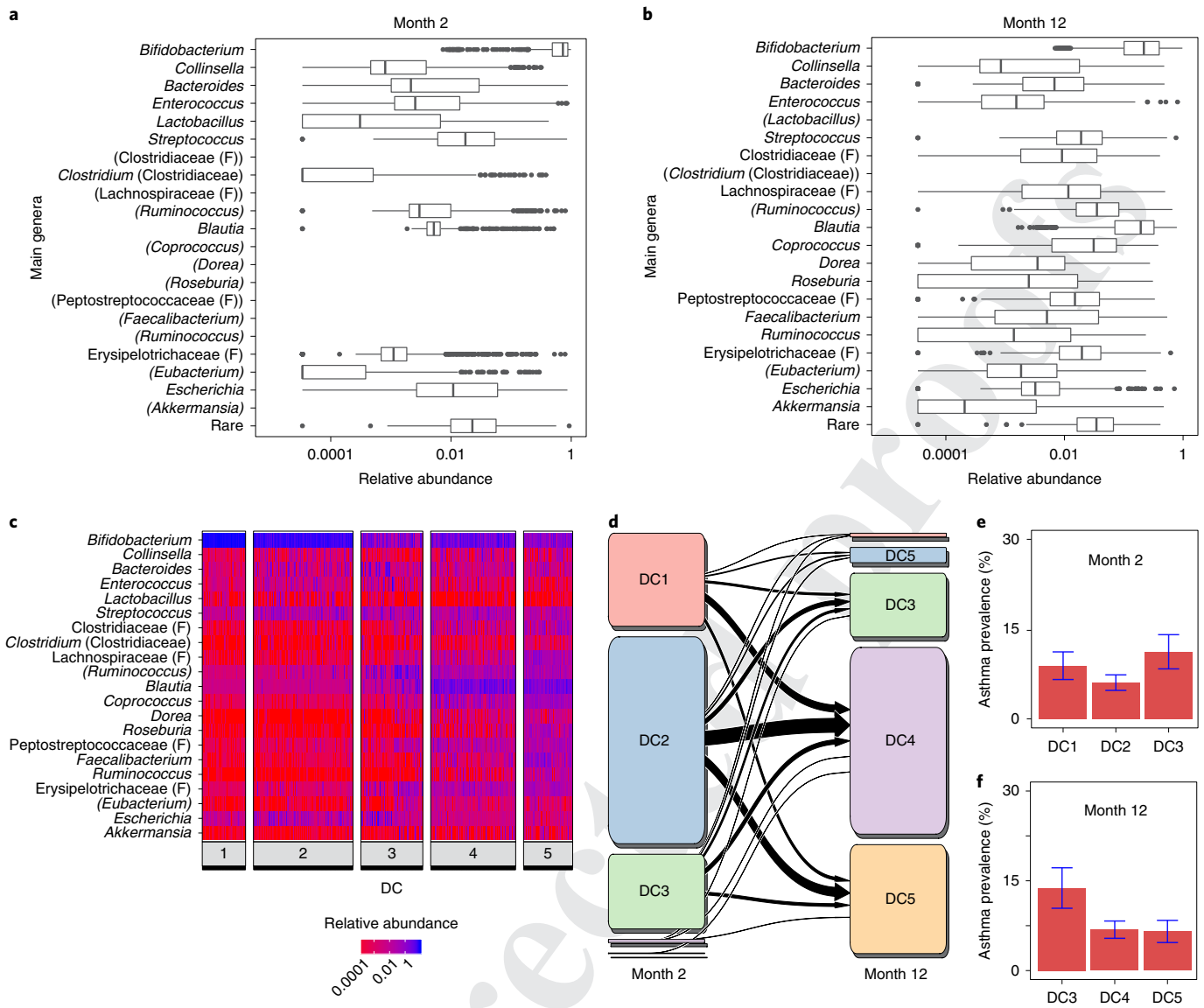


Fig. 1 | Composition of the bacterial gut microbiome at months 2 and 12. a,**b**, Log-scaled box plots of the relative abundance of different bacterial genera in 618 children at 2 months (**a**) and 12 months (**b**). Lower and upper hinges of the boxes denote the first and third quartiles, respectively; the bold central line represents the median; the whiskers extend to the most extreme data point within a distance of 1.5 times the interquartile range from the hinges; extreme values lie beyond the whiskers and are marked by circles. Names in brackets denote bacterial genera with a relative abundance <0.5% at the indicated time point. (F) indicates an unclassified genus from the specified bacterial family. **c**, Log-scaled heat map of the relative abundance of genera within the five clusters of a DMM modeling analysis across both time points, resulting in 2 × 618 samples. **d**, Transition of all 618 individuals between the DCs from months 2 to 12. **e**, Prevalence of asthma (with s.e. bars, $n = 618$) within the most prevalent clusters at month 2 (top) and month 12 (bottom).

allocated more frequently to the more advanced clusters DC4 and DC5 at month 12 ($P < 0.001$; Fig. 3c).

The effect of EMA at month 12 on asthma was validated in 102 additional children in the PASTURE cohort, that is, individuals without measurements at month 2 (Fig. 3d). This effect was also consistent across the different study centers, as was the effect of farm exposure on EMA (Fig. 3e,f). The effect of EMA withstood adjustment for the childhood asthma locus on chromosome 17q21 and other potential confounders (Supplementary Table 5).

A structural equation model revealed that EMA mediated the asthma-protective effect of growing up on a farm by 19% ($P = 0.011$, Fig. 3g), also in the children of non-asthmatic mothers (25%, $P = 0.024$). Likewise, PCA axis 3 at month 2 tended to mediate the asthma-protective effect of breastfeeding by 18% ($P = 0.1$). Farm children were characterized by a more mature microbiome, including *Coprococcus* and *Roseburia* (Fig. 3h), known producers of SCFAs.

Bacterial metabolites and EMA. To assess bacterial taxa by their capacity to produce SCFAs, we modeled SCFA measurements obtained at month 12 in 209 children by the composition of bacterial genera using random forest models. Production of butyrate, propionate and acetate was most importantly predicted by *Roseburia*, *Bacteroides* and *Turicibacter*, respectively (Fig. 4a).

Because the SCFA prediction scores were mutually correlated as well as being partially correlated with EMA, we performed a four-dimensional PCA on EMA and the SCFA scores (Fig. 4b; $n = 720$). Both the butyrate score and EMA loaded on dimension (Dim) 1, which was inversely associated with asthma and non-atopic asthma (Fig. 4c). The acetate score loaded particularly on

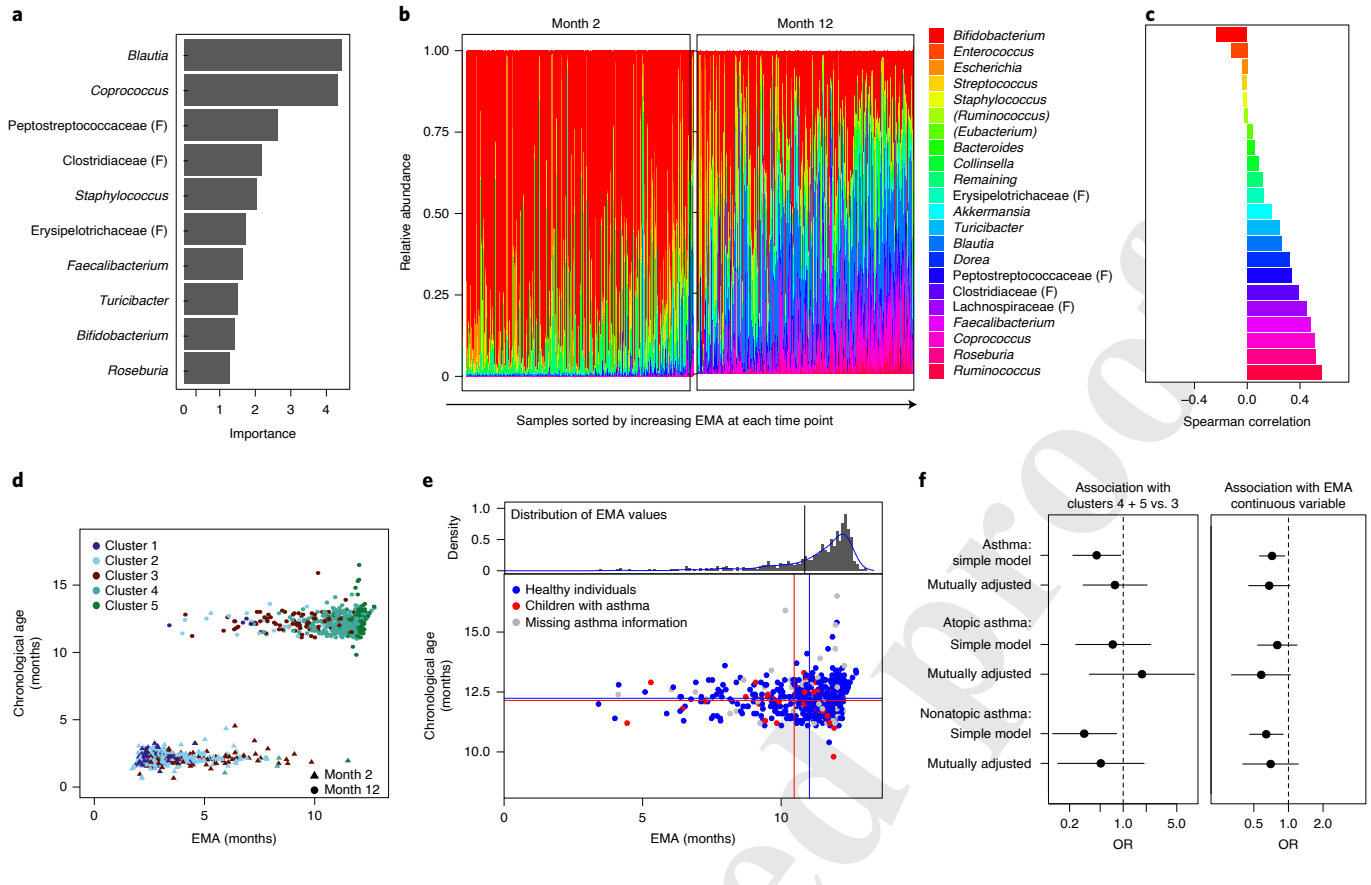


Fig. 2 | EMA as a measure of gut microbiome maturation. a, Variable importance in the prediction model of EMA in 133 healthy individuals. **b**, Stacked bar plots of the relative abundances of the main genera plotted against increasing EMA values. The individual samples ($n = 618$ for each time point) are ordered by increasing EMA and the genera are ordered by Spearman's correlation with changes in relative abundance between months 2 and 12. **c**, Spearman's correlation of EMA at month 12 with changes in relative abundance between months 2 and 12 in 618 children. The color code is the same as that in **b**. Correlation *P* values were below 0.05, except for those from *Escherichia* to *Bacteroides*. **d**, Scatterplot for chronological age at fecal sampling versus EMA at both time points ($n = 2 \times 618$). The color code reflects grouping by DC over both time points. All subsequent analyses relate only to EMA at month 12. **e**, Scatterplot for chronological age at fecal sampling versus EMA at month 12. The color code reflects asthma status at age 6 years. The red and blue lines denote average values for chronological age (horizontal) and EMA (vertical) by asthma status. The distribution of EMA values and an estimated density curve are given at the top of the scatterplot ($n = 618$); the vertical line denotes the lowest quartile, that is, 10.6 months. **f**, Association of asthma phenotypes with DCs (left) and EMA values (right). ORs were derived from logistic regression models ($n = 544$ children with data on asthma). Simple models were only adjusted for center; mutually adjusted models were adjusted for center and the other microbial variable, that is, EMA in the model for DCs and vice versa. Forest plots give point estimates with 95% confidence intervals.

Dim 2, which was not associated with asthma. Dim 3 represented the propionate score, partially represented the EMA score and was inversely associated with atopic asthma. Dim 4 mainly reflected the difference between EMA and the butyrate score and had an additional protective effect on nonatopic asthma.

Dim 1 and Dim 3 (representing EMA, butyrate and propionate scores) were positively associated with growing up on a farm and were correlated particularly well with the consumption of unprocessed farm milk (Fig. 4d). Dim 1 mediated 15% of the farm effect on asthma, and Dim 3 mediated an additional 6% (Fig. 4e).

In a nested case–control sample (44 affected and 94 healthy individuals), we validated the butyrate score as the most important SCFA score. For this purpose, we determined the relative abundance of the gene encoding a major bacterial enzyme in butyrate metabolism, butyryl–CoA:acetate–CoA-transferase, by qPCR. When comparing the results of this gene assay with the originally measured butyrate levels and the estimated butyrate score, we found a particularly high prevalence of asthma in the lowest quartiles of the different measures (Fig. 4f). The corresponding associations with the different asthma phenotypes were similar for all measures (Fig. 4g), for example, OR with asthma, 0.28 (0.09–0.91), $P = 0.034$ for the measured level, 0.38 (0.17–0.84), $P = 0.017$ for the butyrate score and 0.43 (0.19–0.97), $P = 0.042$ for the gene assay. Likewise, associations of the asthma phenotypes with the propionate score resembled those with the corresponding measured levels in the same case–control sample (Extended Data Fig. 7).

Network of bacterial single taxa. Focusing on the result of the maturation process, that is, the microbial composition and interrelation of the genera at 12 months, we performed a network analysis (Fig. 5a). This revealed five network modules with three hubs, which were closely connected. Two hubs were *Roseburia* and *Ruminococcus* and belonged to the green module, the first eigenvector of which was correlated with EMA ($r = 0.73$) and the butyrate score ($r = 0.68$). The yellow module was moderately associated with EMA ($r = 0.35$) and contained two main taxa of EMA, that is, *Blautia* and *Coprococcus*; the latter formed the third hub.

When exploring the association of asthma with amplicon sequence variants (ASVs) associated with the three hubs, two asthma-protective ASVs emerged, one associated with the genus

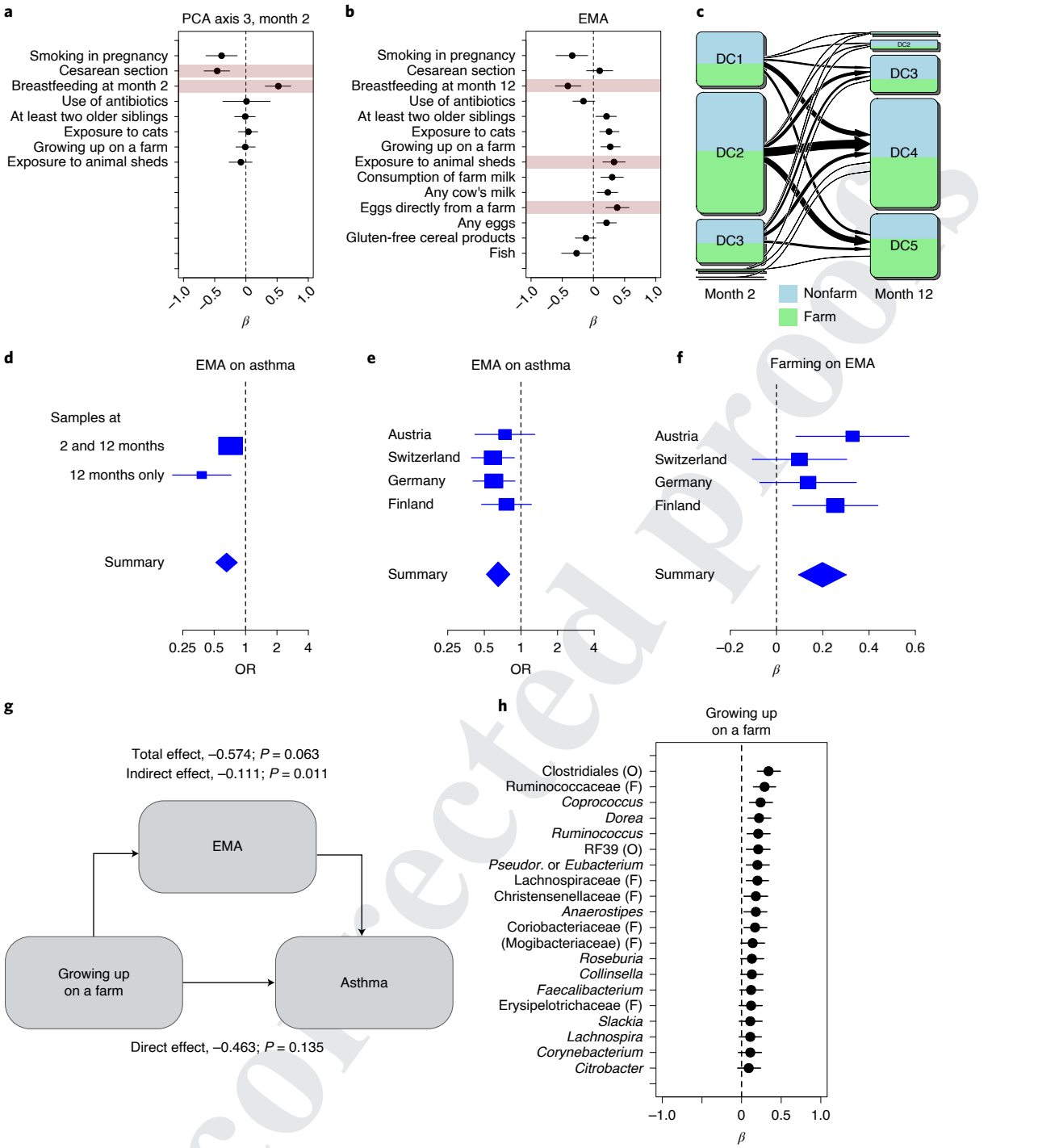


Fig. 3 | EMA and the farm effect on asthma. a, Bivariate associations of environmental determinants with the asthma-protective PCA axis 3 at month 2 ($n = 618$); colored bars denote determinants in the forward selection model. Forest plots give point estimates with 95% confidence intervals. **b**, Bivariate associations of environmental determinants with EMA; colored bars denote determinants in the forward selection model. Forest plots give point estimates with 95% confidence intervals. **c**, Transition of the 618 individuals between the previously defined DCs, stratified by farm status. **d**, Meta-analysis of the effect of EMA on asthma in the 618 children with fecal samples at 2 and 12 months and in the 102 children with fecal samples only at 12 months. For 626 of the 720 children, data on asthma status were available. Forest plots give point estimates with 95% confidence intervals. **e**, Meta-analysis of the effect of EMA on asthma in different study centers ($n = 720$ children; Austria, 173; Switzerland, 209; Germany, 176; Finland, 162). Forest plots give point estimates with 95% confidence intervals. **f**, Meta-analysis of the effect of growing up on a farm on EMA in different study centers ($n = 720$ children, same distribution as in **e**). Forest plots give point estimates with 95% confidence intervals. **g**, Mediation analysis of the protective effect of growing up on a farm on asthma mediated by EMA ($n = 626$). Shown are the estimates of the path model for direct and indirect effects; the proportion of the mediated (indirect) effect was 19%. **h**, Associations of growing up on a farm with the 20 topmost single genera ($n = 720$; *Pseudor.*, *Pseudoramibacter*). Forest plots give point estimates with 95% confidence intervals.

Roseburia (OR = 0.42 (0.18–1.01), $P = 0.053$) and one associated with the genus *Coprococcus* (OR = 0.38 (0.16–0.92), $P = 0.032$) (Supplementary Table 6). Using BLAST searches against the 16S ribosomal sequence database, the first ASV was more precisely assigned to *Roseburia inulinivorans* (100% identity), whereas the second ASV was more compatible with *Anaerobutyricum hallii* (98.4%) than with *Coprococcus eutactus* (92.4%). The presence of these ASVs was strongly associated with higher butyrate levels (GMR = 1.76 (1.34–2.32), $P < 0.001$ and 1.52 (1.12–2.05), $P = 0.008$, respectively) and relative abundance of the butyryl–CoA:acetate–CoA-transferase gene (GMR = 3.33 (1.55–7.15), $P = 0.003$ and 3.81 (1.74–8.34), $P = 0.001$, respectively). The associations of these ASVs with asthma, however, did not withstand adjustment for EMA. Likewise, no genus was found that had an independent protective effect on asthma (Supplementary Table 7), whereas *Eggerthella* (red module) exerted a particular risk effect on asthma (1.43 (1.07–1.92), $P = 0.016$) independently of EMA.

The mycobiome and fungal age. In addition to bacteria, we further explored fungal colonization using ITS data ($n = 189$; Extended Data Fig. 1). Estimated fungal age (EFA) was calculated analogously to EMA and mainly depended on changes in *Saccharomyces*, *Alternaria* and *Malassezia*. EFA was determined by consumption of starchy foods and was not associated with subsequent asthma (Extended Data Fig. 8). Although EFA and EMA were not correlated ($\rho = 0.02$), relative abundance of the highly prevalent genus *Alternaria* at 2 months was associated with subsequent bacterial maturation ($\beta = 0.05$ (0.01–0.10), $P = 0.032$).

Sensitivity analyses. EMA was also inversely associated with an asthma diagnosis after 3 years of age (Extended Data Fig. 9) and particularly with the less common (Supplementary Table 8) persistent wheeze phenotype (OR = 0.49 (0.35–0.70), $P < 0.001$)[14]. EMA was not associated with sensitization to seasonal, perennial or food allergens. Stratification for atopic sensitization in children or for maternal asthma did not reveal major differences in the associations of asthma phenotypes with EMA (Supplementary Table 9). Similarly, the associations of EMA with asthma phenotypes were homogeneous between farm and nonfarm children, with the exception of the association of EMA with atopic asthma, which was only observed in nonfarm children (0.68 (0.45–1.02), $P = 0.060$). Unlike microbial maturation and composition, the butyrate score did not vary significantly between centers ($P = 0.191$; Extended Data Fig. 10).

Discussion

In the PASTURE birth cohort, farm-related exposures influenced the maturation of the gut microbiome during the time window from 2 to 12 months. As a measure of maturation, EMA mediated a substantial proportion of the well-known farm effect on asthma. Bacterial communities with the potential of producing butyrate, such as *Roseburia* and *Coprococcus*, contributed to asthma protection (Fig. 5b).

Mode of birth has often been associated with subsequent colonization of the human gut, as exemplified by *Bacteroides*[12]. The importance of birth mode for future health[15] highlights the role of the maternal microbiome in the colonization of the neonatal gut. The current analyses, however, suggest that this maternal influence is gradually replaced by an increasing variety of environmental exposures that affect the growing child. Indeed, the most influential change in the development of the mammalian gut microbiome is probably the transition from breastfeeding to a solid food diet[16,17]. Bifidobacteria, whose early predominance may be fostered by the bifidogenic properties of breastmilk[18], decrease in abundance after weaning. In our analysis, the asthma-protective PCA axis at month 2 was mainly influenced by mode of birth and correlated with the abundance of *Bacteroides*. The positive association of this axis with breastfeeding and its asthma-protective effect was not explained by bifidobacteria.

Independently of this very early phenomenon, bacterial maturation between months 2 and 12 exerted a strong protective effect on asthma. Various shifts in bacterial composition, including in the bacterial families Lachnospiraceae, Ruminococcaceae and Bifidobacteriaceae, suggest high plasticity of the intestinal microbiome during the first year of life[12,13]. Evidently, the window of opportunity for the establishment of an asthma-protective microbiome extends substantially beyond the well-studied[19–21] period of the first 3 months of life. Early, precipitate maturation might even be unfavorable for asthma risk[22], which may explain the tendency for asthma risk in children with higher EMA values at month 2, particularly for those assigned to DC3.

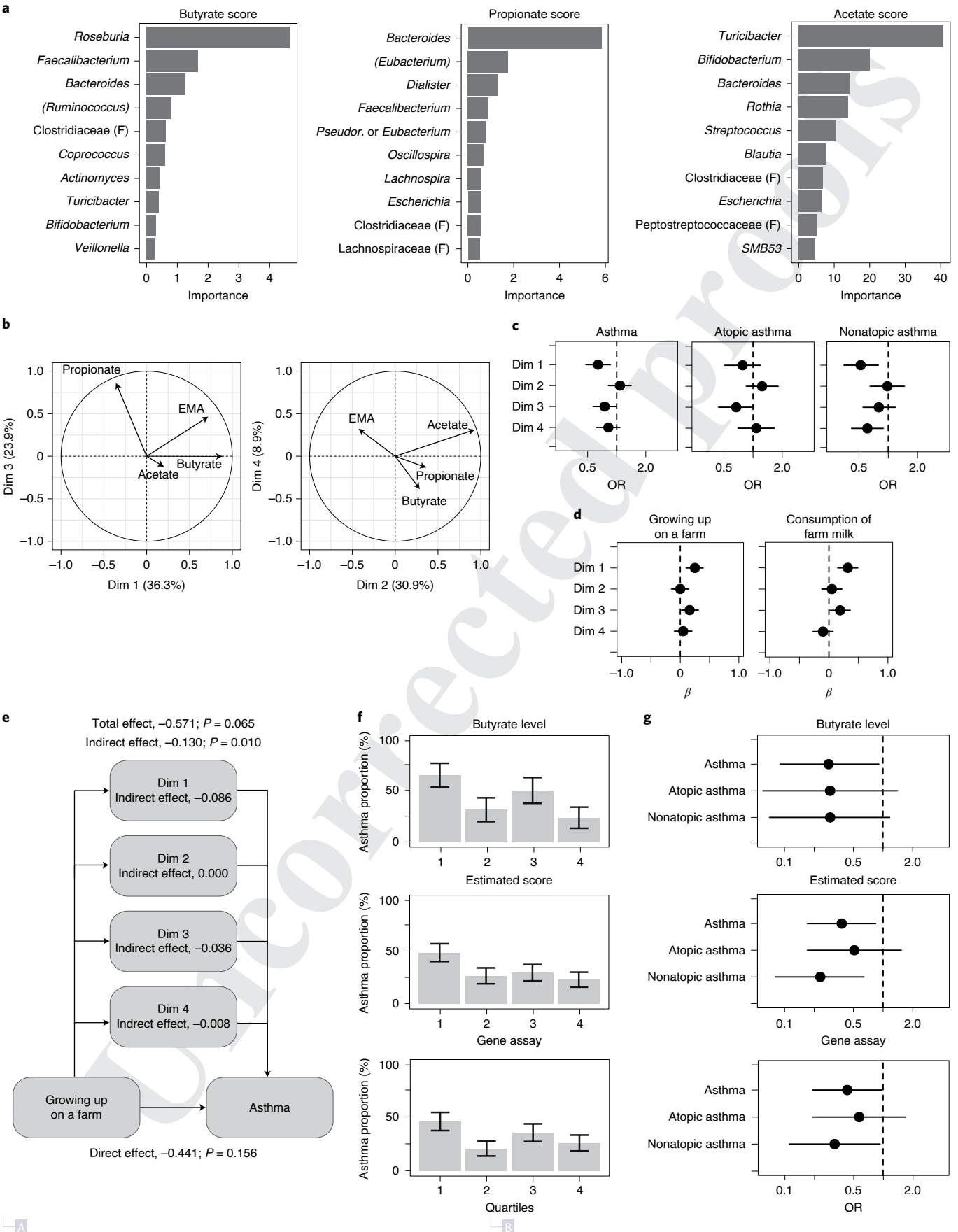
EMA predicted asthma better than DC3 at any time point, emphasizing the developmental aspect of a favorable microbiome. Possibly bacterial composition is not beneficial on its own but may indicate successful maturation. This notion has vast implications for prevention strategies, as the mere application of distinct probiotics or combinations thereof seems less promising for asthma prevention.

In contrast to the early microbiome, which was favorably influenced by breastfeeding, the subsequent maturation process was hampered by prolonged breastfeeding. Because the effect of prolonged breastfeeding on EMA was independent of the diversity of solid foods, this finding supports the idea that cessation of breastfeeding is a key factor that influences microbial composition and maturation[12,16,17].

The particular setting of the PASTURE study revealed other main determinants of maturation, which were all associated with farm exposure, the epitome of the hygiene hypothesis. Farm children are known to be exposed to a greater variety of environmental microbiota[6] and potentially beneficial clusters of microorganisms[23]. There may be various mechanisms involved in mediation of the protective effect of environmental microbiota on asthma. For example, when playing in animal sheds, children may inhale environmental microorganisms, which may exert their effects directly in the airways[9,10]. In the present mediation analysis, we demonstrate that 19% of the farm effect on asthma was mediated through the maturation

Fig. 4 | Bacterial metabolites and EMA. a, Variable importance for the prediction scores of the SCFAs butyrate, propionate and acetate, as modeled in 209 children with measured values. **b**, PCA for EMA and the three SCFA scores ($n = 720$). For illustrative purposes, Dim 3 is plotted against Dim 1 and Dim 4 against Dim 2. Explained variance is given in parentheses. **c**, Associations of asthma phenotypes with the four PCA dimensions. Forest plots give point estimates with 95% confidence intervals. **d**, Associations of growing up on a farm and consumption of farm milk with the four PCA dimensions. Forest plots give point estimates with 95% confidence intervals. **e**, Mediation analysis of the protective effect of growing up on a farm on asthma mediated by the four PCA dimensions ($n = 626$). Shown are the estimates of the path model for direct and indirect effects; the proportion of the mediated (indirect) effects was 23%. **f**, Validation of the butyrate score in a case–control subsample of 138 children (44 with asthma and 94 healthy individuals). Proportions of children with asthma (with s.e. bars) is given in quartiles of the originally measured butyrate level, the estimated butyrate score and the gene assay, that is, the relative abundance of the butyryl–CoA:acetate–CoA-transferase gene, encoding an enzyme in the bacterial metabolic pathway for butyrate production. **g**, Associations of asthma phenotypes with the originally measured butyrate level, the estimated butyrate score and the gene assay, all dichotomized at the lowest quartile. Forest plots give point estimates with 95% confidence intervals.

of the gut microbiome, suggesting that environmental microbiota are ingested and interact with the gut microbiome. At least, this notion is a reminder of the fecal–oral transmitted infections that are postulated by the hygiene hypothesis to be protective[5]. The remaining 81% of the farm effect on asthma might be mediated by other mechanisms, also operating beyond the first year of life.



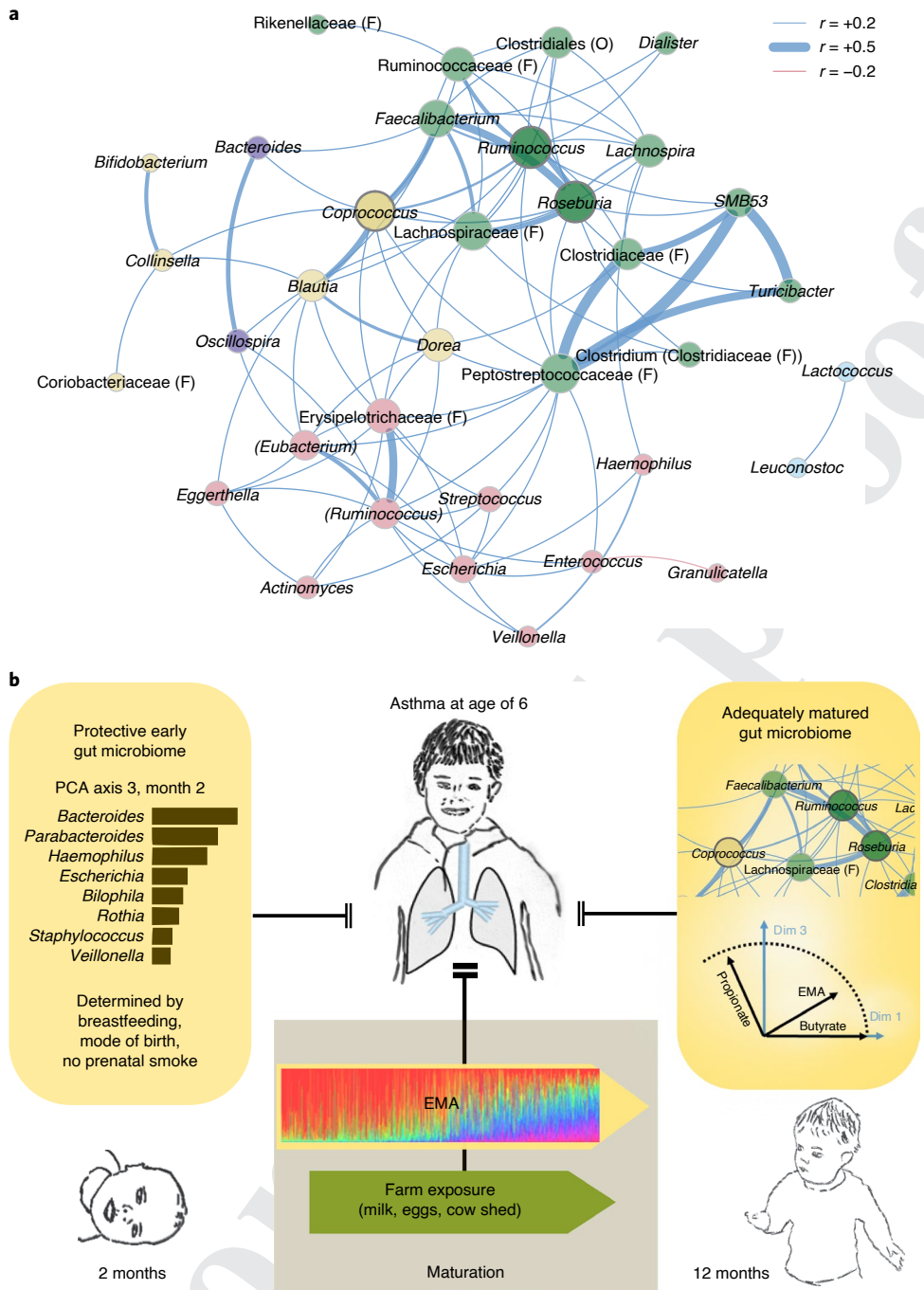


Fig. 5 | Network of single taxa and summary of findings. a, Network analysis of single genera ($n = 720$). The resulting five modules are marked by different node colors. Positive correlations are marked by blue edges and negative correlations by red edges. Edge thickness denotes SparCC correlation ranging from −0.2 to +0.5. Only correlations with an absolute value of at least 0.2 are shown. Network hubs, as defined by an eigenvector centrality value above the 99th percentile, are marked by black circles. **b**, Summary of findings. At month 2, the intestinal microbiome was mainly determined by prenatal, perinatal and early postnatal influences, such as prenatal smoke exposure, mode of birth and breastfeeding. An inverse association with asthma at school age was found for a principal component at month 2. The maturation of the microbiome from 2 to 12 months was quantified by EMA, a prediction score derived from random forest modeling of sampling time points in relation to changes in the composition of bacterial genera over time. This EMA mediated the protective farm effect on asthma by about 20%. At month 12, the resulting intestinal microbiome formed distinct network modules, with *Roseburia*, *Ruminococcus* and *Coprococcus* as hubs. EMA correlated with two dimensions of a PCA combining EMA and SCFA scores. These two dimensions almost exclusively explained the mediation of the protective farm effect on asthma and mainly represented bacterial producers of butyrate and propionate, respectively.

While exploring a single-taxon approach, we did not identify any taxon as protective in itself. This was unlikely to be caused by insufficient statistical power, as we detected a risk effect in the single taxon *Eggerthella*, which includes an emerging pathogen associated with asthma-risk effects in adults[24,25]. Likewise, in the nasal microbiome, we previously identified individual taxa, such as *Moraxella catharralis*, solely as being harmful[10]. Therefore, single taxa are more likely to exert harmful effects.

Inconsistencies between studies may arise due to differences in sampling time points. For example, risk of atopic wheeze was associated with relative abundance of the taxa *Faecalibacterium* sp., *Lachnospira* sp., *Veillonella* sp. and *Rothia* sp. at 3 months but not at 12 months[20]. The beneficial taxon *Veillonella* may not only lose its beneficial effect over time[26] but may even emerge subsequently as an asthma-risk taxon[27]. Although *Bacteroides*, *Prevotella* and *Coprococcus* were associated with allergic diseases from 6 months to 8 years, other taxa such as *Ruminococcus* have been shown to lose their beneficial effect within the first year[28]. Although *Bifidobacterium* was suggested to be a beneficial probiotic in other contexts[29], its abundance was increased in children with allergies at later time points[28]. In our analysis, *Bifidobacterium* did not contribute to the asthma-protective effect. These inconsistencies were another reason for using an integrative approach that considered changes in bacterial composition over time.

A limitation of the current analysis might be the fact that sampling was conducted at only two time points, possibly missing fluctuations within this developmental window. Nevertheless, this drawback might emerge as an advantage, as comparing the starting point and outcome of the core maturation process may highlight the essential changes of the microbiota in this time window. Further refinement occurs in subsequent years[12,13] and, on a smaller scale, throughout life[22,30]. Another limitation is that parent-reported diagnosis of asthma is susceptible to misclassification; in previous analyses, however, this outcome definition was validated by lung function measurements[14].

To better understand how bacterial maturation may impact respiratory health, we focused on the functional properties of gut bacteria and modeled communities with a high likelihood of producing distinct SCFAs. As all resulting SCFA scores were correlated with EMA, we tried to disentangle the different aspects of EMA and the three SCFA scores by a PCA. The component of EMA that was shared with the likelihood of producing butyrate (Dim 1) exerted the strongest asthma-protective effect, predominantly for the non-atopic phenotype. For the atopic phenotype, the aspect of EMA that was shared with propionate production (Dim 3) was particularly important. Moreover, both aspects were involved in mediating the protective farm effect on asthma. In contrast, the likelihood of producing acetate, which predominantly represented Dim 2, was not associated with asthma. This finding might support the concept of a specific effect of butyrate and propionate in humans, in which these SCFAs, but not acetate, have been shown to impair the viability of eosinophils[31]. Finally, Dim 4 denoted an aspect of EMA that was not shared with butyrate production; hence, bacterial maturation may exert an individual protective effect on nonatopic asthma beyond butyrate production alone.

To validate the importance of SCFA production, we linked asthma directly to SCFA levels measured in the fecal samples[32] and found consistent associations with butyrate levels and a tendency for an association between atopic asthma and propionate levels. The gene assay for butyryl–CoA:acetate–CoA-transferase, an enzyme that converts butyryl–CoA to butyrate and is part of the main pathway of bacterial butyrate metabolism[33], corroborated the association of butyrate production with (nonatopic) asthma.

Butyrate is the main source of energy for colonic epithelial cells, contributes to the maintenance of the epithelial gut barrier and has immunomodulatory and anti-inflammatory properties[34]. Various taxa, the composition of which varies considerably between individuals, can contribute to butyrate production directly and by cross-feeding[35]. Likewise, propionate has anti-inflammatory potential but is mainly produced by intestinal *Bacteroides* taxa, although some *Roseburia*, *Coprococcus* and *Blautia* taxa can also switch from butyrate to propionate production[36]. In particular, *Roseburia* has been suggested as a health biomarker because of its anti-inflammatory properties[37].

We found two promising asthma-protective ASV candidates in the genera *Roseburia* and *Coprococcus* (or *A. hallii*), which were also directly associated with butyrate production and the butyryl–CoA:acetate–CoA-transferase gene. Adjustment for EMA, however, revealed that these taxa did not carry the asthma-protective effect themselves. Rather, they may represent a network of bacteria with the capacity to produce SCFAs. This notion is supported by the role of the genera *Roseburia* and *Coprococcus* as hubs in the network analysis.

Altogether, higher SCFA levels may reduce inflammation at various body sites, including the airway mucosa. Butyrate- and propionate-producing bacteria represent an aspect of healthy maturation of the gut microbiome and add an independent component to the asthma-protective EMA effect, thereby extending the concept of the gut–lung axis[38,39] to humans.

The beneficial effect of gut microbiota may not be specific to respiratory health. Bacteria associated with maturation of the gut microbiome (*Ruminococcus*, *Faecalibacterium*, *Roseburia* and Lachnospiraceae) were also identified in children with a low prevalence of enteric infections[40]. Moreover, a low abundance of *Roseburia* was also observed in patients with rheumatoid arthritis[41]. Hence, the combination of the above taxa might represent a marker for well-established host immune systems and good general health in the absence of intestinal dysbiosis[42].

The definition of asthma and atopy phenotypes may vary from study to study. Some studies have focused on early outcomes[20,21,26]; few studies followed up for atopic wheeze[43] or an asthma diagnosis[27] at an age of 5 years, when diagnoses can be established with reasonable certainty. Therefore, we assessed various asthma phenotypes defined by wheezing patterns or concomitant atopy. EMA was consistently associated with the nonatopic phenotype of asthma and persistent wheeze but not with atopic sensitization per se. A family history of atopy, however, did not influence the susceptibility to the microbiome-associated farm effect on asthma. In contrast to the $COPSAC_{2010}$ study[27], we found an inverse association of EMA with asthma also in children whose mothers did not have asthma. Therefore, studies that focus on atopic outcomes, such as atopic wheeze[20,43], or are performed in high-risk populations[27] might reveal other facets of the microbial effect on asthma. The current analysis points toward an inflammatory pathology behind atopy, supported by the anti-inflammatory properties of butyrate[44].

Certainly butyrate is only an example and may be a marker for other metabolites that might be directly involved in signaling between intestinal and respiratory mucosa, such as D-tryptophan[45]. Microbial carbohydrate metabolism has also been implicated in health effects exerted by the gut microbiome[46]. In addition, the vagus nerve can sense microbial metabolites with its afferent fibers and can contribute to inflammation by a low tone in its efferent part, as shown in inflammatory bowel disease[47]. By analogy with the so-called gut–brain axis[48], the vagus nerve may also be involved in neuro-immune crosstalk[49] and in communication between the gut microbiome and the airway tone, as suggested by the EMA effect on lung function.

In sum, we found strong influences on the maturation of the gut microbiome by an environment rich in microbial stimuli. Maturation and prediction of butyrate production partially mediated the well-known asthma-protective farm effect, thereby suggesting a gut–lung axis in humans. In contrast, atypical microbial maturation may contribute to the pathogenesis of inflammatory diseases. This emphasizes the need for prevention strategies in the first year of life, when the gut microbiome is highly plastic and amenable to modification.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of

data and code availability are available at https://doi.org/10.1038/s41591-020-1095-x.

Received: 28 September 2019; Accepted: 8 September 2020;

References

1. Barker, D. J. The fetal and infant origins of adult disease. *BMJ* **301**, 1111 (1990).
2. Tamburini, S., Shen, N., Wu, H. C. & Clemente, J. C. The microbiome in early life: implications for health outcomes. *Nat. Med.* **22**, 713–722 (2016).
3. de Steenhuijsen Piters, W. A. et al. Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. *Am. J. Respir. Crit. Care Med.* **194**, 1104–1115 (2016).
4. Chung, H. et al. Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149**, 1578–1593 (2012).
5. Ege, M. J. The hygiene hypothesis in the age of the microbiome. *Ann. Am. Thorac. Soc.* **14**, S348–S353 (2017).
6. Ege, M. J. et al. Exposure to environmental microorganisms and childhood asthma. *N. Engl. J. Med.* **364**, 701–709 (2011).
7. Stein, M. M. et al. Innate immunity and asthma risk in Amish and Hutterite farm children. *N. Engl. J. Med.* **375**, 411–421 (2016).
8. Illi, S. et al. Protection from childhood asthma and allergy in Alpine farm environments—the GABRIEL Advanced Studies. *J. Allergy Clin. Immunol.* **129**, 1470–1477 (2012).
9. Birzele, L. T. et al. Environmental and mucosal microbiota and their role in childhood asthma. *Allergy* **72**, 109–119 (2017).
10. Depner, M. et al. Bacterial microbiota of the upper respiratory tract and childhood asthma. *J. Allergy Clin. Immunol.* **139**, 826–834 (2017).
11. Trompette, A. et al. Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nat. Med.* **20**, 159–166 (2014).
12. Stewart, C. J. et al. Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* **562**, 583–588 (2018).
13. Yassour, M. et al. Natural history of the infant gut microbiome and impact of antibiotic treatment on bacterial strain diversity and stability. *Sci. Transl. Med.* **8**, 343ra381 (2016).
14. Depner, M. et al. Clinical and epidemiologic phenotypes of childhood asthma. *Am. J. Respir. Crit. Care Med.* **189**, 129–138 (2014).
15. Kumbhare, S. V., Patangia, D. V., Patil, R. H., Shouche, Y. S. & Patil, N. P. Factors influencing the gut microbiome in children: from infancy to childhood. *J. Biosci.* **44**, 49 (2019).
16. Backhed, F. et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* **17**, 690–703 (2015).
17. Bokulich, N. A. et al. Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci. Transl. Med.* **8**, 343ra382 (2016).
18. Borewicz, K. et al. The effect of prebiotic fortified infant formulas on microbiota composition and dynamics in early life. *Sci. Rep.* **9**, 2434 (2019).
19. Abrahamsson, T. R. et al. Low gut microbiota diversity in early infancy precedes asthma at school age. *Clin. Exp. Allergy* **44**, 842–850 (2014).
20. Arrieta, M. C. et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci. Transl. Med.* **7**, 307ra152 (2015).
21. Fujimura, K. E. et al. Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nat. Med.* **22**, 1187–1191 (2016).
22. Galazzo, G. et al. 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* **158**, 1584–1596 (2020).
23. Ege, M. J. et al. Environmental bacteria and childhood asthma. *Allergy* **67**, 1565–1571 (2012).
24. Gardiner, B. J. et al. Clinical and microbiological characteristics of *Eggerthella lenta* bacteremia. *J. Clin. Microbiol.* **53**, 626–635 (2015).
25. Wang, Q. et al. A metagenome-wide association study of gut microbiota in asthma in UK adults. *BMC Microbiol.* **18**, 114 (2018).
26. Stiemsma, L. T. et al. Shifts in *Lachnospira* and *Clostridium* sp. in the 3-month stool microbiome are associated with preschool age asthma. *Clin. Sci.* **130**, 2199–2207 (2016).
27. Stokholm, J. et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat. Commun.* **9**, 141 (2018).
28. Simonyte Sjodin, K. et al. Temporal and long-term gut microbiota variation in allergic disease: a prospective study from infancy to school age. *Allergy* **74**, 176–185 (2019).
29. Yang, B. et al. A meta-analysis of the effects of probiotics and synbiotics in children with acute diarrhea. *Medicine* **98**, e16618 (2019).
30. Derrien, M., Alvarez, A. S. & de Vos, W. M. The gut microbiota in the first decade of life. *Trends Microbiol.* **27**, 997–1010 (2019).
31. Theiler, A. et al. Butyrate ameliorates allergic airway inflammation by limiting eosinophil trafficking and survival. *J. Allergy Clin. Immunol.* **144**, 764–776 (2019).
32. Roduit, C. et al. High levels of butyrate and propionate in early life are associated with protection against atopy. *Allergy* **74**, 799–809 (2019).
33. Louis, P., Young, P., Holtrop, G. & Flint, H. J. Diversity of human colonic butyrate-producing bacteria revealed by analysis of the butyryl–CoA:acetate CoA-transferase gene. *Environ. Microbiol.* **12**, 304–314 (2010).
34. Riviere, A., Selak, M., Lantin, D., Leroy, F. & De Vuyst, L. Bifidobacteria and butyrate-producing colon bacteria: importance and strategies for their stimulation in the human gut. *Front. Microbiol.* **7**, 979 (2016).
35. Reichardt, N. et al. Specific substrate-driven changes in human faecal microbiota composition contrast with functional redundancy in short-chain fatty acid production. *ISME J.* **12**, 610–622 (2018).
36. Reichardt, N. et al. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J.* **8**, 1323–1335 (2014).
37. Tamanai-Shacoori, Z. et al. *Roseburia* spp.: a marker of health? *Future Microbiol.* **12**, 157–170 (2017).
38. Frati, F. et al. The role of the microbiome in asthma: the gut–lung axis. *Int. J. Mol. Sci.* **20**, 123 (2018).
39. Marsland, B. J., Trompette, A. & Gollwitzer, E. S. The gut–lung axis in respiratory disease. *Ann. Am. Thorac. Soc.* **12**(Suppl. 2), S150–S156 (2015).
40. Singh, P. et al. Intestinal microbial communities associated with acute enteric infections and disease recovery. *Microbiome* **3**, 45 (2015).
41. Forbes, J. D. et al. A comparative study of the gut microbiota in immune-mediated inflammatory diseases—does a common dysbiosis exist? *Microbiome* **6**, 221 (2018).
42. Kemter, A. M. & Nagler, C. R. Influences on allergic mechanisms through gut, lung, and skin microbiome exposures. *J. Clin. Invest.* **130**, 1483–1492 (2019).
43. Arrieta, M. C. et al. Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting. *J. Allergy Clin. Immunol.* **142**, 424–434 (2018).
44. Zhai, S. et al. Dietary butyrate suppresses inflammation through modulating gut microbiota in high-fat diet-fed mice. *FEMS Microbiol. Lett.* **366**, fnz153 (2019).
45. Kepert, I. et al. D-Tryptophan from probiotic bacteria influences the gut microbiome and allergic airway disease. *J. Allergy Clin. Immunol.* **139**, 1525–1535 (2017).
46. Cait, A. et al. Reduced genetic potential for butyrate fermentation in the gut microbiome of infants who develop allergic sensitization. *J. Allergy Clin. Immunol.* **144**, 1638–1647 (2019).
47. Bonaz, B., Bazin, T. & Pellissier, S. The vagus nerve at the interface of the microbiota–gut–brain axis. *Front. Neurosci.* **12**, 49 (2018).
48. Osadchiy, V., Martin, C. R. & Mayer, E. A. The gut–brain axis and the microbiome: mechanisms and clinical implications. *Clin. Gastroenterol. Hepatol.* **17**, 322–332 (2019).
49. Kabata, H. & Artis, D. Neuro-immune crosstalk and allergic inflammation. *J. Clin. Invest.* **130**, 1475–1482 (2019).

PASTURE study group

Martin Depner[1], Amandine Divaret-Chauveau[11,12,13], Markus J. Ege[18,21], Remo Frei[5,8], Jon Genuneit[22,23], Anne Hyvärinen[3], Sabina Illi[1], Michael Kabesch[16], Anne M. Karvonen[3], Pirkka V. Kirjavainen[3,4], Roger Lauener[5,7,9,10], Lucie Laurent[13], Juha Pekkanen[3,19], Petra I. Pfefferle[18,24], Harald Renz[17,18], Josef Riedler[14], Caroline Roduit[5,6,7], Marjut Roponen[15], Bianca Schaub[18,21] and Erika von Mutius[1,18,21]

[22]Institute of Epidemiology and Medical Biometry, Ulm University, Ulm, Germany. [23]Pediatric Epidemiology, Department of Pediatrics, Medical Faculty, Leipzig University, Leipzig, Germany. [24]Comprehensive Biomaterial Bank Marburg (CBBM), Philipps University of Marburg, Marburg, Germany.

Methods

Study population. The PASTURE birth cohort has been recruited in rural areas of Austria, Finland, France, Germany and Switzerland with the aim of exploring the asthma-protective effects of growing up on a farm[50]. Detailed information on the study design can be found in the Nature Research Reporting Summary. Briefly, pregnant adult women were invited to participate during their third trimester; half of them lived on family-run livestock farms. Their children were recruited at birth and seen at the ages of 2, 12, 56 and 72 months during home visits. Additional questionnaires were completed at 2, 12, 18, 24, 36, 48, 60 and 72 months.

Additional detailed information on the children's health, nutrition and farm-related exposures was gathered by using weekly diaries and monthly questionnaires covering the 9th to 52nd weeks of life[51,52]. Stool samples were collected from the children's napkins during the 2- and 12-month home visits. Because fecal sampling was not performed by design in the French arm, these children were excluded from the current analyses a priori. All aspects of the study were approved by the local institutional review boards in each country (Austria, Ethikkommission für das Land Salzburg; Finland, the Research Ethics Committee, Hospital District of Northern Savo; Germany, Ethik-Kommission der Bayerischen Landesärztekammer; Switzerland, Kantonale Ethik-Kommission St. Gallen; France, Comité Consultatif pour la Protection des Personnes en Recherche Biomédicale (CCPPRB), Commission Informatique et Libertés (CNIL)). Written informed consent was obtained from the parents or guardians.

Definition of health outcomes. Asthma was defined as a physician's diagnosis of asthma at least once up to age 6 years or recurrent diagnoses of spastic, obstructive or asthmatic bronchitis, as reported by the parents at age 6 years[14].

For a sensitivity analysis, we defined 'asthma after 3 years' as an asthma diagnosis established in the fourth, fifth or sixth year of life. Wheeze phenotypes were derived from a latent class analysis as described previously[14].

Allergen-specific IgE was assessed at age 6 years. Seasonal IgE was defined as at least one IgE specific to alternaria, alder, birch, hazel, grass pollen, rye, mugwort or plantain at levels ≥0.7 IU ml^{-1} at age 6 years. Perennial IgE (to *Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*, cat, horse and dog) and food IgE (to chicken eggs, cow's milk, peanuts, hazelnuts, carrots and wheat flour) were defined analogously. Assessment was performed on peripheral blood by using the semiquantitative AllergyScreen test panel for atopy (Mediwiss Analytic) in a central laboratory[53]. The atopic and nonatopic phenotypes of asthma were defined by the presence or absence of concomitant sensitization to inhalant allergens (seasonal or perennial) with specific IgE levels ≥0.7 IU ml^{-1} at age 6 years, while the references were always children without asthma.

Spirometry was performed as previously described[54], and spirometric indices were calculated according to the equations of a task force of the European Respiratory Society (https://www.ers-education.org/guidelines/global-lung-function-initiative/spirometry-tools.aspx). FEV_1 values were determined as *z* scores, and the lower quintile was defined as children with 'bad lung function'.

Assessment of exposures. Assessment of environmental exposures by questionnaires covered pregnancy and the first year of life and included information from premature births (less than 37 weeks of gestation) and infants with low birth weight (<2,500 g), as well as variables for birth weight and gestational age dichotomized at the median, APGAR score at 5 min and delivery mode, including natural vaginal birth, vaginal birth with forceps, vacuum extraction or Cesarean section. Data on treatment with systemic antibiotics were available for the first 2 months and the first year beyond 2 months; maternal treatment with antibiotics during the first 2 months of lactation was also considered.

Breastfeeding was defined as any breastfeeding until 2 months of age or current breastfeeding at month 2. Duration of breastfeeding was dichotomized at various cutoff levels from 2 to 12 months. Food diversity was defined as the introduction of five of six main food items (vegetables or fruits, cereals, meat, bread, yogurt and cake) within the first year as previously described[55]. In a sensitivity analysis, we explored an extended list of 15 food items (main food items with the addition of eggs, fish, nuts, soy, margarine, chocolate, other milk products, cow's milk and butter), which were dichotomized into at least 11 items. Furthermore, the children's diet was assessed with respect to the kind of supplemental food and its introduction in terms of at least weekly consumption[56]. Farm milk consumption was defined as the weekly consumption of any milk obtained directly from a farm, irrespective of boiling or skimming.

Farm exposure was assessed as growing up on a farm or more specifically by regular contact with hay or visits to animal sheds, including sheds with bigger animals such as cows, poultry sheds or barns. As further environmental determinants, we assessed the number of siblings (at least two siblings) and presence of pets (cats or dogs) in each time period, smoking during pregnancy and environmental smoke exposure, defined as at least one cigarette smoked at home per day by any person. In addition, information on parental history of atopy, which is a combination of asthma, hay fever and atopic eczema, or asthma alone (maternal, paternal or both) and degree of parental education (at least completion of secondary school) were included.

DNA extraction from fecal samples. Fecal samples were frozen within 10 min of collection and stored at −20 °C until further processing. At a central laboratory (THL Kuopio), DNA was extracted from the fecal samples in batches as follows. Partially defrosted fecal samples were homogenized using a Stomacher 80 microBiomaster laboratory paddle blender (Seward; 2 min at high speed). DNA was extracted from 150 mg of ice-cold homogenized fecal samples, using a bead-beating method with a fecal DNA Miniprep kit (D6010, Zymo Research) according to the manufacturer's instructions. The bead-beating step was performed with a FastPrep FP120 homogenizer (2 min at full speed, 6.5 m s^{-1}). Finally, the samples were eluted with 100 µl of elution buffer D3004-4-10 (Zymo Research). The samples for extraction were kept on ice throughout the entire procedure. The extracted DNA was immediately frozen at −20 °C and stored at −80 °C.

Sequencing analyses. Amplification and sequencing of fecal samples were performed as described previously for bacterial and fungal communities[57].

Primers F515 (5′-NNNNNNNNGTGTGCCAGCMGCCGCGGTAA-3′) and R806 (5′-GGACTACHVGGGTWTCTAAT-3′)[58] were used to amplify the V4 region of the 16S rRNA gene. The forward primers had unique 8-bp barcodes (indicated by 'N') and a 2-bp linker sequence at the 5′ end. PCR reactions contained DNA template, 1× GoTaq Green Master Mix (Promega), 1 mM $MgCl_2$ and 2 pmol of each primer. Samples were amplified in triplicate PCR reactions. Conditions consisted of an initial step at 94 °C for 3 min, followed by 25 cycles of 94 °C for 45 s, 50 °C for 60 s and 72 °C for 90 s and a final extension at 72 °C for 10 min.

Primers BITS (5′-NNNNNNNNCTACCTGCGGARGGATCA-3′) and B58S3 (5′-GAGATCCRTTGYTRAAAGTT-3′) were used to amplify the fungal ITS region 1 (ref. [59]). Again, the forward primers had unique 8-bp barcodes and a linker sequence (bold portion) at the 5′ end. PCR reactions contained DNA template, 1× GoTaq Green Master Mix (Promega), 1 mM $MgCl_2$ and 2 pmol of each primer. Reaction conditions consisted of an initial step at 95 °C for 2 min, followed by 40 cycles of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 60 s and a final extension at 72 °C for 5 min.

Amplicons were run on a 0.8% agarose gel to verify amplification by gel electrophoresis. Bacterial and fungal amplicons were combined into two separately pooled samples, purified using the QIAquick spin kit (Qiagen) and submitted to the DNA Technologies Core at the University of California Davis Genome Center for Illumina paired-end library preparation, cluster generation and 250-bp paired-end sequencing on an Illumina MiSeq instrument in separate runs.

Raw sequencing data from each run were demultiplexed using Sabre[60]. Demultiplexed data were imported into QIIME2-2018.6 (ref. [61]) and quality trimmed. Reads were denoised using DADA2 (ref. [62]) as implemented in QIIME2. Taxonomy was assigned to representative sequences using a naive Bayes classifier[63,64] pre-built from the 99% Greengenes database[65] specific to the 515F–806R region for bacterial data. For fungal sequences, a classifier was built from the UNITE dynamic database for taxonomic assignment[66]. For fungal data, no tree was created because there is currently no valid taxonomy available with respect to ITS sequences.

Combining sequences from 2 and 12 months together, 5,915 ASVs were retrieved from 16S rRNA sequences, after the exclusion of chloroplasts. For ITS, 54,459 ASVs were retrieved when restricted to fungi. Analyses are reported on the genus level. If genera were not identified, we used the name of the lowest identified level. If information was only available on the kingdom level, we designated the taxon as 'completely unidentified'. Samples with <1,000 reads were removed.

Genetics. Genotyping for 939 children with available DNA samples in the PASTURE study was performed at the Centre National de Génotypage, Evry, France, using iPLEX Gold technology and a matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) mass spectrometry system from Sequenom[67]. Technical errors were minimized by comparing genotypic frequencies with the expected allelic population equilibrium based on the Hardy–Weinberg equilibrium test. cDNA was amplified in duplicate using an iCycler (Bio-Rad Laboratories) and 18S as a reference gene. Quality checks were passed by samples from 896 children (95%). Of these, 512 children were included in the present 16S rRNA analysis.

Selection of single-nucleotide polymorphisms (SNPs) was based on previous reports and included polymorphisms at the chromosome 17q21 childhood-onset asthma-risk locus[68–70]. Specifically, SNPs rs8076131, associated with the *ORMDL3* gene, and rs2290400 and rs7216389, associated with the *GSDMB* gene, at this locus were found to interact with environmental tobacco smoking[71] and viral infections[72] for childhood asthma. SNP rs8076131 was selected for the current analysis because it has been described in the context of functional regulation[73].

SCFAs. Metabolite levels of SCFAs were measured in fecal samples obtained from 301 children in the PASTURE study at the age of 12 months[32]. Fecal samples were processed as previously described[74]. Briefly, 1 ml of 0.15 mM H_2SO_4 was added to 0.3 g feces to generate a fecal suspension. After rigorous vortexing, the samples were centrifuged twice (14,000*g* for 30 min) and sequentially filtered through a 0.45-µm Millex-HA filter and a 0.2-µm Millex-LG filter (Merck). The resultant fecal homogenates were analyzed by HPLC (Merck Hitachi) using a

Rezex ROA-Organic Acid H^+ ion exchange column together with a SecurityGuard Cartidges, Carbo-H from Phenomenex at a flow rate of 0.4 ml at 40 °C with 10 mM H_2SO_4 as the eluent. The samples were quantified in relation to standards[75]. Of the 301 children, 209 (69%) were included in the present 16S rRNA analysis (Supplementary Fig. 1).

Relative abundance of the butyryl–CoA:acetate–CoA-transferase gene. Relative abundance of the butyryl–CoA:acetate–CoA-transferase gene was measured in a nested 1:2 case–control sample of 51 children with asthma and 106 healthy individuals with available fecal samples at 12 months of age. qPCR primers and annealing temperatures used for samples with total bacteria (primers UniF and UniR, 0.5 µM final concentration) and the butyryl–CoA:acetate–CoA-transferase gene (primers BCoATscrF and BCoATscrR, 2.5 µM final concentration) are described in Ramirez-Farias et al.[76]. Equipment and reagents for qPCR are described in Reichardt et al.[35]. DNA samples were used without further dilution, unless their concentration (determined with a Qubit dsDNA HS assay kit, Thermo Fisher Scientific) exceeded 4 ng µl^{-1} in the qPCR assay. The data are expressed as a percent of butyryl–CoA:acetate–CoA-transferase gene copies normalized to total bacterial 16S rRNA gene copies, as described in Ramirez-Farias et al.[76]. Measurement of the relative abundance of the butyryl–CoA:acetate–CoA-transferase gene was performed in duplicate, and measurements were considered valid if the standard deviation was below 0.4. This was necessary to eliminate technical artifacts, such as air bubbles, that could interfere with the optical fluorescence reading. Of all 157 children who were selected for the case–control study, valid results were obtained in 138 individuals (88%). Data were stored in Excel.

Statistical analysis. Statistical analysis was performed with R versions 3.4.3 and 3.6.1 (https://www.r-project.org/), particularly with the phyloseq package, and Mplus[77]. Upon request, computer code will be made available to readers.

Relative abundance was used to describe the taxonomic distribution of bacteria and fungi. Rare taxa were defined as having a relative abundance below 0.5% in each population and were subsumed in a category termed 'rare'. For logarithmic presentations, values were transformed by decadic logarithm, in which necessary zero values were replaced by the lowest measured value. All statistical tests were two sided, and an effective P value <0.05 was considered statistically significant.

Dirichlet clustering. Dirichlet multinomial mixtures (DMM) modeling was performed with the R package DirichletMultinomial. We clustered the samples over both time points and, as a sensitivity analysis, separately for both time points. DMM bins samples on the basis of microbial community structure[78]. The number of clusters was determined by a local minimum of a Laplace approximation score, that is, five clusters over both time points and three clusters for the separate models. Transitions between clusters were illustrated with the R package Gmisc.

Random forests. Random forest regression was performed with the R package ranger to model sampling age based on the relative abundance of bacterial or fungal genera present at 2 and 12 months in a subsample of 133 (for fungi, 35) healthy individuals, that is, children without asthma, who never wheezed or had diarrhea during the first year. The models were estimated using 2,000 trees and a ceiling of the square root of the number of selected variables per level. The resulting prediction model, mainly defined by alterations in relative abundance of all genera, was subsequently applied to the entire population, using the 'predict' function of ranger. These estimates were used as a proxy for bacterial or fungal maturation and subsequently called EMA or EFA. To confirm that results were independent of the training sets, we performed sensitivity analyses by restricting the models to children who were not included in model building. Taxa were ranked by their variable importance in random forest models for EMA and EFA, which discriminate best between the two measurement time points.

A similar approach was used to estimate SCFA scores for butyrate, propionate and acetate. SCFA levels were modeled by the relative abundance of bacterial genera in children with available SCFA measurements. These prediction models were applied to predict SCFA production scores in the entire population. Taxa were ranked by their variable importance for SCFA production in random forest models. The number of randomly picked variables was optimized by tenfold cross-validation. As a member of the out-of-bag methods, random forest modeling has an advantage over classical cross-validation in that it yields an unbiased error estimate, that is, high validity[79]. As random forest modeling integrates all information on microbial taxa in a single model, no correction for multiple comparisons was necessary. Besides continuous variables, the EMA as well as the butyrate score was also dichotomized at the lowest quartile in subsamples to give an estimate for a threshold phenomenon.

Microbial diversity and composition. Samples were rarefied at the minimum sequence numbers in the available biosamples, that is, 1,029 16S rRNA sequences and 1,000 ITS sequences. Rarefaction and calculation of species richness and Shannon diversity index was iterated 1,000 times, and the resulting measures of α-diversity were subsequently averaged. An independent rarefying step was performed only for month 12 samples to analyze the presence or absence of specific ASVs. For the assessment of bacterial composition, the R package composition was used to perform a centered-log ratio (clr) transformation, after adding a pseudocount of 0.5 to abundance values. This approach, developed by Aitchison, was shown to be essential in microbiome analyses[80]. The clr-transformed abundance values were used in PCA for assessing β-diversity. In addition, β-diversity was assessed by PCoA on the ASV level, using unweighted UniFrac as a distance measure, calculated by the R package GUniFrac. Samples taken at 2 and 12 months were evaluated separately by PCA and PCoA. In addition, PCoA was also applied in a sensitivity analysis combining all samples from both time points. Associations of indices of maturation, butyrate production or microbial composition (as determined by PCA) with asthma or determinants were based on regression models, in which the microbial variables were usually z standardized to render them comparable against each other. All analyses were adjusted for center. Models adjusted only for center were termed 'simple models', whereas the term 'mutually adjusted models' refers to models in which two exposures were compared and forced in the same model. Interaction was analyzed by including a product term in the regression models.

Confirmatory analyses for the associations of EMA with growing up on a farm or asthma were replicated across study centers and assessed by a meta-analysis with fixed effects (R package rmeta).

To compare direct and indirect effects, mediation models were calculated in Mplus[77] and validated with the R package mediation. The mediated effect is reported as the proportion of the estimated indirect effect to the total effect.

To test for associations of single taxa with asthma, we first tested for differences in relative abundance by the Wilcoxon test; main associations ($P < 0.1$) were then confirmed in the clr-transformed variables with logistic regression models. These models were initially adjusted only for center, or additionally for EMA, to determine single-taxon effects that were independent of the general maturation process. Single taxa were assessed in an exploratory approach; therefore, adjustment for multiple comparisons was not performed. All statistical tests were two sided.

Box-and-whisker plots were used as follows: lower and upper hinges denote the first and third quartiles, respectively; the bold central line represents the median; the whiskers extend to the most extreme data point within a distance of 1.5 times the interquartile range from the hinges; extreme values lie beyond the whiskers and are marked by circles. Forest plots give point estimates with 95% confidence intervals.

Spearman coefficient was used to calculate any kind of correlation between different measures, except for network analyses.

Network analyses. Correlations between pairs of bacterial genera were estimated using the strength of sparse correlations for compositional data (SparCC) approach[81]. The corresponding correlation network was visualized using the R package qgraph. In the network plot, only correlations with an absolute value ≥0.2 are shown. For readability, nodes without any connections were removed. Node sizes were scaled on the eigenvector centrality measure, which was determined via the 'eigen_centrality' function from the R package igraph.

Hubs were defined as nodes with an eigenvector centrality value greater than the 99th percentile of all eigenvector centrality values in the network. Groups of highly connected nodes, also called clusters or modules, were identified via the 'cluster_fast_greedy' igraph function, which is a fast greedy algorithm for determining clusters by maximizing the modularity measure over all possible clusterings[82].

To relate the composition of the network modules to EMA and the butyrate score, we used an approach similar to the eigengene analysis[83], that is, we used the first eigenvector of a PCA with the clr-transformed taxa passing the threshold criteria to build the network plots representing each module.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Taxonomy was assigned using the Greengenes database (http://greengenes.lbl.gov) for 16S rRNA sequences and the UNITE dynamic database (https://unite.ut.ee/) for ITS sequences. All 16S rRNA and ITS sequences are deposited in the Supplementary Information without metadata. PASTURE is an ongoing birth cohort with fieldwork still being executed. As long as the study is not anonymized, European data protection legislation prohibits sharing of individual data, even when pseudonymized. Upon request, the authors will share aggregate data that do not allow identification of individuals.

References

50. von Mutius, E., Schmid, S. & Group, P. S. The PASTURE project: EU support for the improvement of knowledge about risk factors and preventive factors for atopy in Europe. *Allergy* **61**, 407–413 (2006).
51. Loss, G. et al. Consumption of unprocessed cow's milk protects infants from common respiratory infections. *J. Allergy Clin. Immunol.* **135**, 56–62 (2015).

52. Loss, G. J. et al. The early development of wheeze. Environmental determinants and genetic susceptibility at 17q21. *Am. J. Respir. Crit. Care Med.* **193**, 889–897 (2016).
53. Hose, A. J. et al. Latent class analysis reveals clinically relevant atopy phenotypes in 2 birth cohorts. *J. Allergy Clin. Immunol.* **139**, 1935–1945 (2017).
54. Fuchs, O. et al. Farming environments and childhood atopy, wheeze, lung function, and exhaled nitric oxide. *J. Allergy Clin. Immunol.* **130**, 382–388 (2012).
55. Roduit, C. et al. Increased food diversity in the first year of life is inversely associated with allergic diseases. *J. Allergy Clin. Immunol.* **133**, 1056–1064 (2014).
56. Depner, M. et al. Atopic sensitization in the first year of life. *J. Allergy Clin. Immunol.* **131**, 781–788 (2013).
57. Bokulich, N. A., Thorngate, J. H., Richardson, P. M. & Mills, D. A. Microbial biogeography of wine grapes is conditioned by cultivar, vintage, and climate. *Proc. Natl Acad. Sci. USA* **111**, E139–E148 (2014).
58. Caporaso, J. G. et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**(Suppl. 1), 4516–4522 (2011).
59. Bokulich, N. A. & Mills, D. A. Improved selection of internal transcribed spacer-specific primers enables quantitative, ultra-high-throughput profiling of fungal communities. *Appl. Environ. Microbiol.* **79**, 2519–2526 (2013).
60. Joshi, N. Sabre—a barcode demultiplexing and trimming tool for FastQ files. Vol. 2019 (2011).
61. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335–336 (2010).
62. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
63. Bokulich, N. A. et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).
64. Pedregosa, F. et al. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
65. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
66. Abarenkov, K. et al. The UNITE database for molecular identification of fungi—recent updates and future perspectives. *N. Phytol.* **186**, 281–285 (2010).
67. Kormann, M. S. et al. G-protein-coupled receptor polymorphisms are associated with asthma in a large German population. *Am. J. Respir. Crit. Care Med.* **171**, 1358–1362 (2005).
68. Moffatt, M. F. et al. A large-scale, consortium-based genomewide association study of asthma. *N. Engl. J. Med.* **363**, 1211–1221 (2010).
69. Moffatt, M. F. et al. Genetic variants regulating *ORMDL3* expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473 (2007).
70. Torgerson, D. G. et al. Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat. Genet.* **43**, 887–892 (2011).
71. Bouzigon, E. et al. Effect of 17q21 variants and smoking exposure in early-onset asthma. *N. Engl. J. Med.* **359**, 1985–1994 (2008).
72. Caliskan, M. et al. Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N. Engl. J. Med.* **368**, 1398–1407 (2013).
73. Schedel, M. et al. Polymorphisms related to *ORMDL3* are associated with asthma susceptibility, alterations in transcriptional regulation of *ORMDL3*, and changes in T_H2 cytokine levels. *J. Allergy Clin. Immunol.* **136**, 893–903 (2015).
74. Dostal, A. et al. Effects of iron supplementation on dominant bacterial groups in the gut, faecal SCFA and gut inflammation: a randomised, placebo-controlled intervention trial in South African children. *Br. J. Nutr.* **112**, 547–556 (2014).
75. Pham, V. T., Lacroix, C., Braegger, C. P. & Chassard, C. Early colonization of functional groups of microbes in the infant gut. *Environ. Microbiol.* **18**, 2246–2258 (2016).
76. Ramirez-Farias, C. et al. Effect of inulin on the human gut microbiota: stimulation of *Bifidobacterium adolescentis* and *Faecalibacterium prausnitzii*. *Br. J. Nutr.* **101**, 541–550 (2009).
77. Muthén, L. K. & Muthén, B. O. *Mplus User's Guide* 7th edn (Muthén & Muthén, 1998–2012).
78. Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PloS ONE* **7**, e30126 (2012).
79. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
80. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome datasets are compositional: and this is not optional. *Front. Microbiol.* **8**, 2224 (2017).
81. Friedman, J. & Alm, E. J. Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* **8**, e1002687 (2012).
82. Brandes, U. et al. On modularity clustering. *IEEE Trans. Knowl. Data Eng.* **20**, 172–188 (2008).
83. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst. Biol.* **1**, 54 (2007).

Acknowledgements

This paper is dedicated to Jean-Charles Dalphin, one of the founders of the PASTURE study, who passed away shortly after the submission of this manuscript. We are very grateful for his constant support of this successful collaboration. We thank B. Stecher-Letsch and A.-L. Boulesteix for valuable input in discussions about biological and statistical issues, respectively. We are grateful to G. Pagani for technical support with graphs. The PASTURE study was supported by the European Commission (research grants QLK4-CT-2001-00250, FOOD-CT-2006-31708 and KBBE-2007-2-2-06) and the European Research Council (grant 250268). The current analyses were supported by the German Federal Ministry of Education and Research (BMBF; project German Center for Lung Research (DZL)). P.L. and F.M.F. receive funding from the Scottish Government's Rural and Environment Sciences and Analytical Services Division (RESAS).

Author contributions

E.v.M., J.-C.D., J.R. and J.P. obtained funds, set up the PASTURE birth cohort and were responsible for data collection and management of the study. D.M., R.L., M.K., M.R., H.R., R.F. and C.R. were responsible for laboratory analyses. K.M.K. and D.M. performed sequencing analyses. D.H.T. performed bioinformatics, and C.R. and R.F. performed SCFA analyses. F.M.F. and P.L. designed and performed the butyryl–CoA:acetate–CoA-transferase assay. E.S.-H. was involved in data management and S.P. performed statistical network analysis. P.V.K., A.M.K. and A.D.-C. were involved in the acquisition and interpretation of data. M.D. performed statistical analyses and completed the background literature search, M.J.E. supervised statistical analyses, M.D. and M.J.E. drafted the manuscript, and all authors provided substantial revisions and approved the final version of the manuscript. The PASTURE study group was involved in the acquisition, management and interpretation of data in Austria, Finland, France, Germany and Switzerland. The members of the PASTURE study group contributed substantially to the design, conception and conduct of the study or the acquisition or analysis of data.

Competing interests

D.M. is a cofounder of Evolve Biosystems and has stock and stock options therein; he has received payment for lectures from the Nestlé Nutrition Institute and the Abbott Nutrition Institute. M.K. has a patent share on the diagnostic use of SNPs in *ORMDL3* on chromosome 17q21. H.R. has received research support from DFG, BMBF, EU, Land Hessen, DAAD, ALK, Stiftung Pathobiochemie, Ernst-Wendt-Stiftung, Mead Johnson Nutritional and Beckman Coulter; speaker's honoraria from Allergopharma, Novartis, Thermo Fisher, Danone, Mead Johnson Nutritional and Bencard; and consulting fees from Bencard and Sterna Biologicals. He is a cofounder of Sterna Biologicals. E.v.M. is listed as an inventor on the following patents: publication number EP 1411977, composition containing bacterial antigens used for the prophylaxis and treatment of allergic diseases, granted on 18 April 2007; publication number EP 1637147, stable dust extract for allergy protection, granted on 10 December 2008; publication number EP 1964570, pharmaceutical compound to protect against allergies and inflammatory diseases, granted on 21 November 2012. E.v.M. is listed as an inventor and has received royalties on the following patent: publication number EP 2361632, specific environmental bacteria for the protection from and/or treatment of allergic, chronic inflammatory and/or autoimmune disorders, granted on 19 March 2014. She has received funding and research support from FrieslandCampina; she has received consultation and speaker fees from OM Pharma, Böhringer Ingelheim International, Peptinnovate, Pharmaventures, Nestlé Deutschland (36 months before publication) and HiPP (future). M.J.E. is a co-inventor on patents for the use of environmental bacteria to prevent asthma (EP000001964570B1, US000009950017B2 and EP000002361632B1). His employer has received investigational products for an intervention study with minimally processed milk from FrieslandCampina.

Additional information

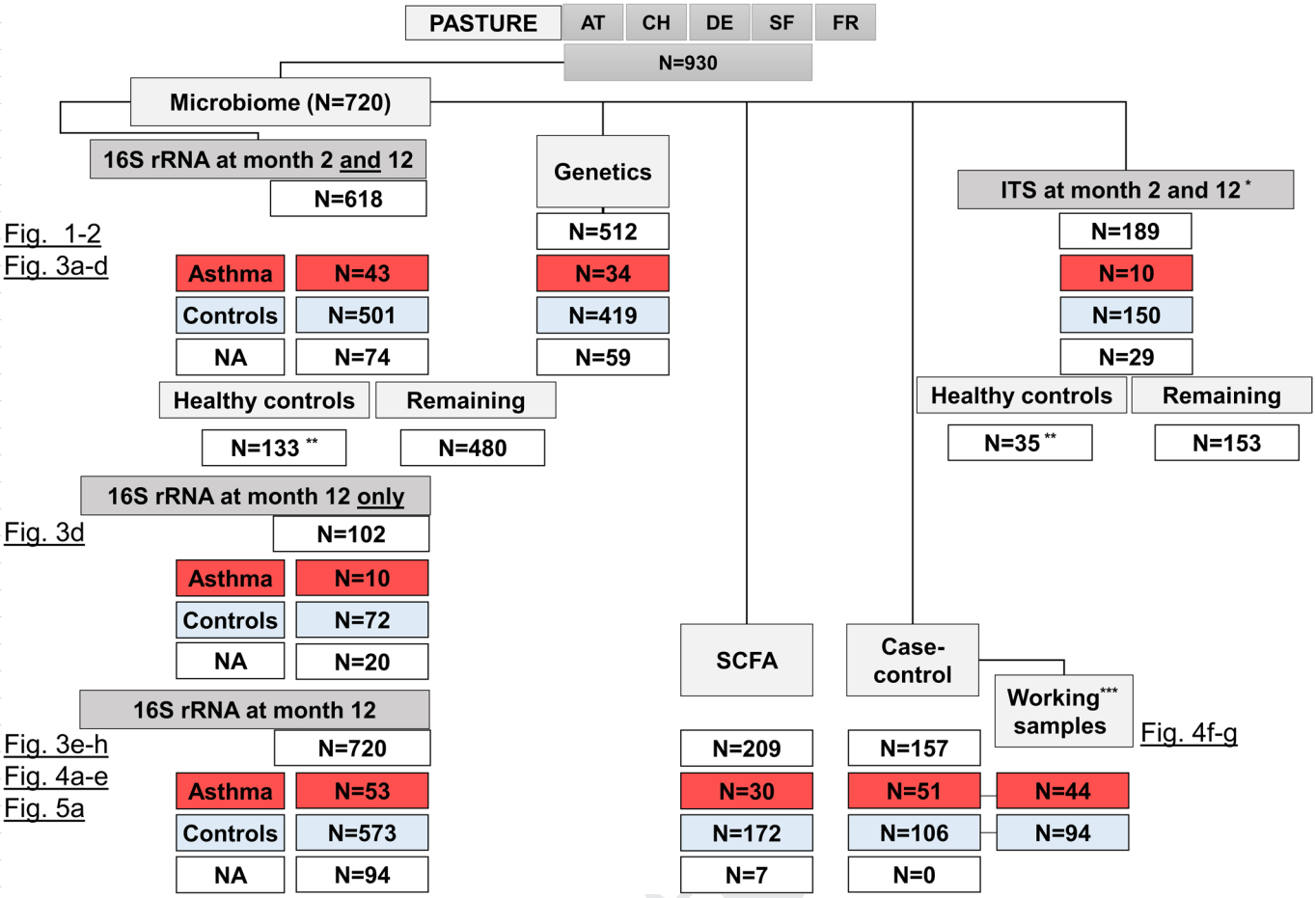
Extended data is available for this paper at https://doi.org/10.1038/s41591-020-1095-x.

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-020-1095-x.

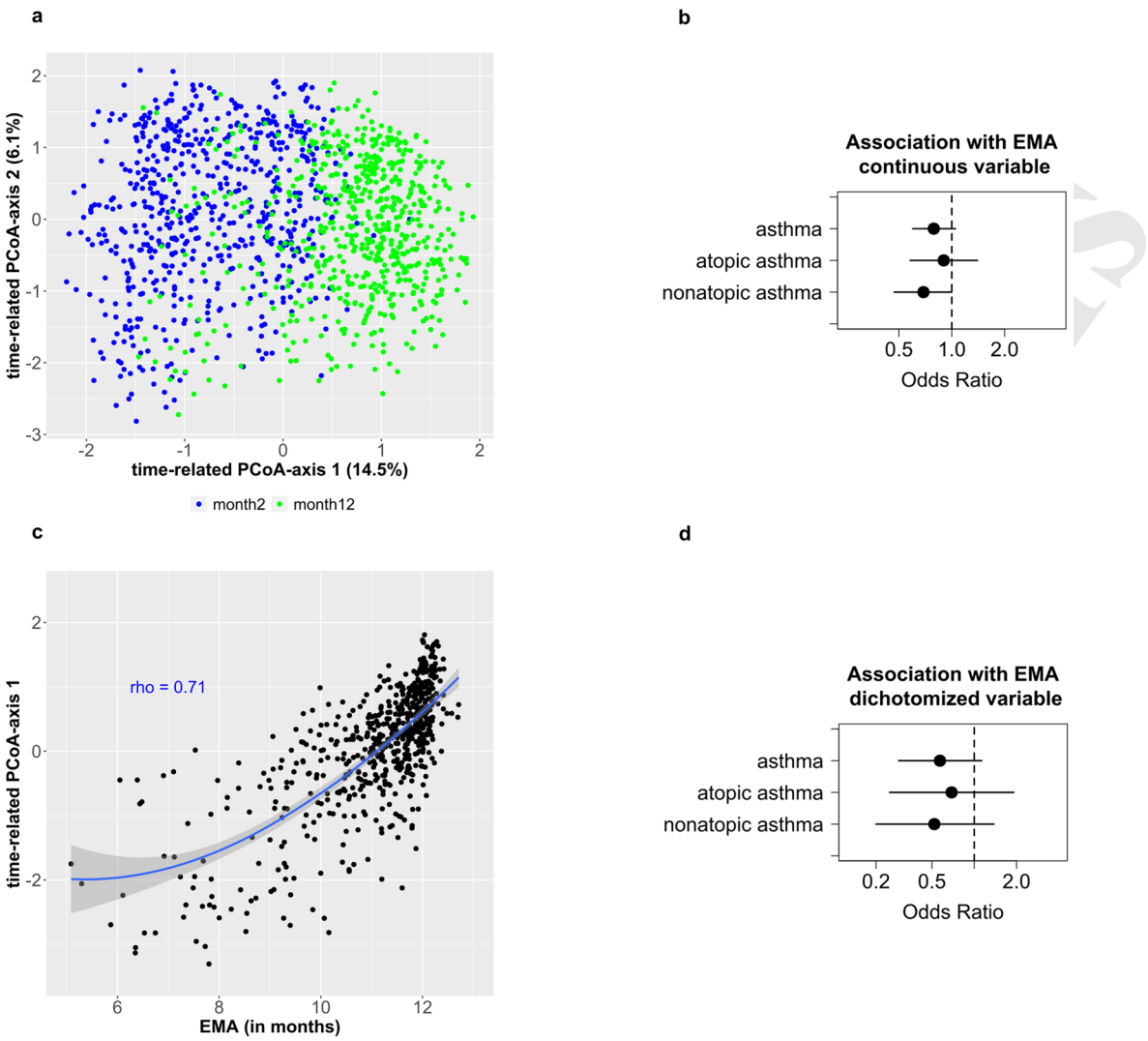
Correspondence and requests for materials should be addressed to M.J.E.

Peer review information Alison Farrell was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

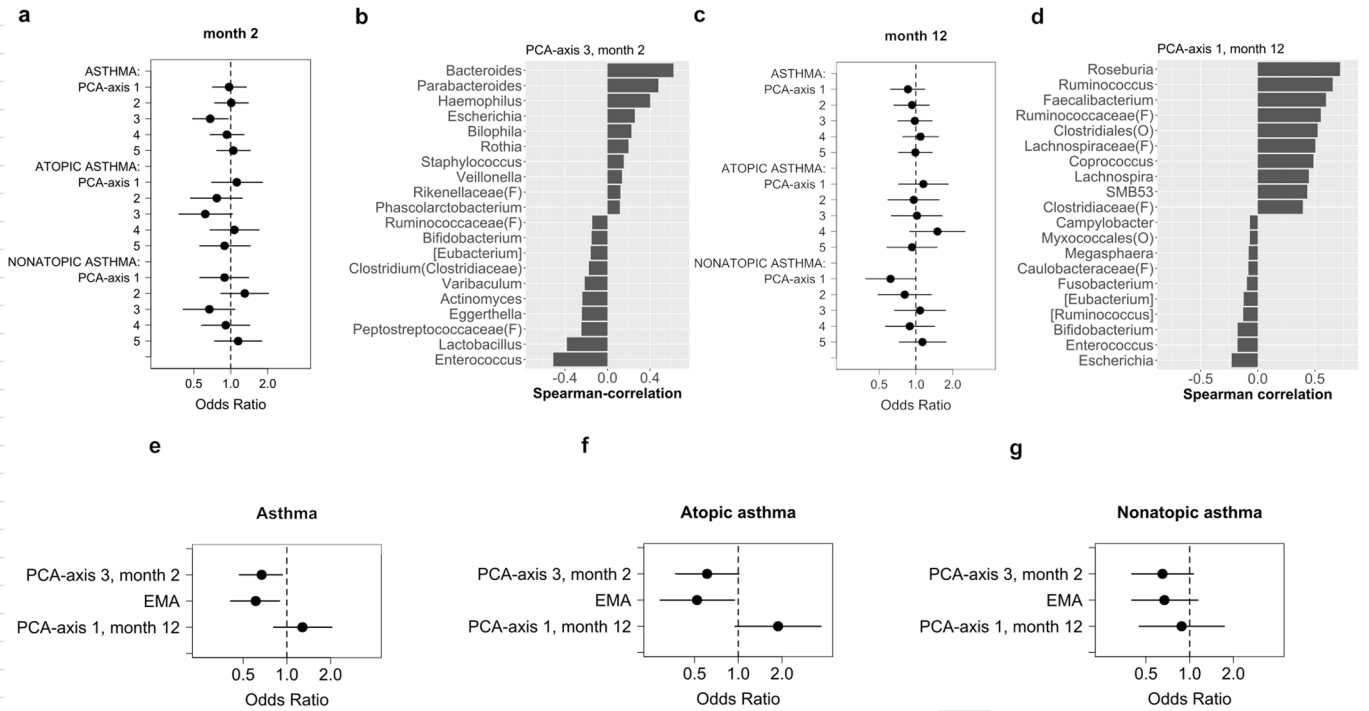
Reprints and permissions information is available at www.nature.com/reprints.



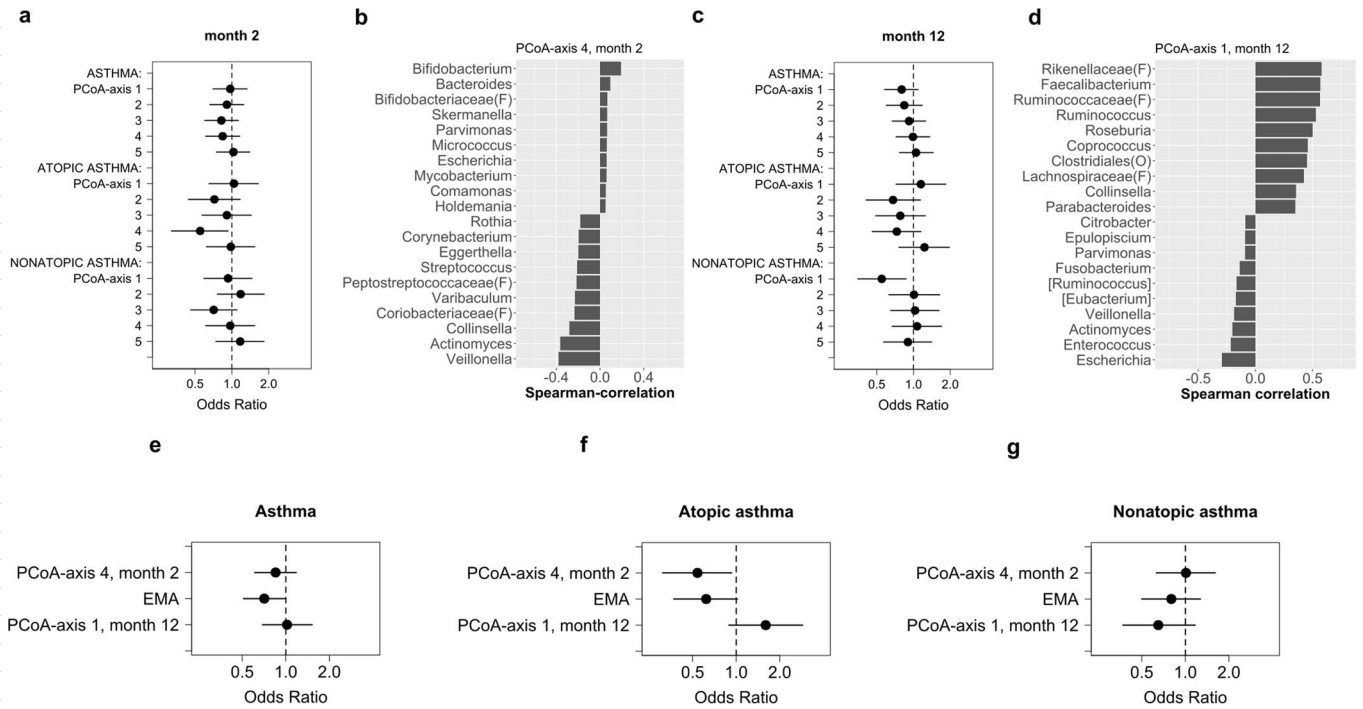
Extended Data Fig. 1 | Participant flow. The current microbiome analysis population (n = 720) was selected based on available microbiome data at month 12. The French arm was not included because by design microbiome samples were not taken at month 2. The children with samples available at month 2 and 12 (n = 618) were quite equally distributed over the centers (Austria N = 139; Switzerland N = 205; Germany N = 136; Finland N = 138). For n = 102 children only 12 months samples were available. Subsamples are colored in red, blue, and white according to asthma status (yes, no, not available, respectively). The different subsamples colored in grey represent the basis of the respective figures of the main manuscript as indicated*. Samples with fungi data are a subsample of the 618 samples with bacteria, and only children with ITS samples at both time points available were analyzed**. Healthy controls were defined by no diarrhea between 2 and 12 months and no asthma / wheeze anytime; individuals with missing or implausible values for sampling time point were excluded (5 for bacteria, 1 for fungi)***. Butyryl-CoA:acetate CoA-transferase gene assay failed in 19 of the 157 samples (12%) for technical reasons.



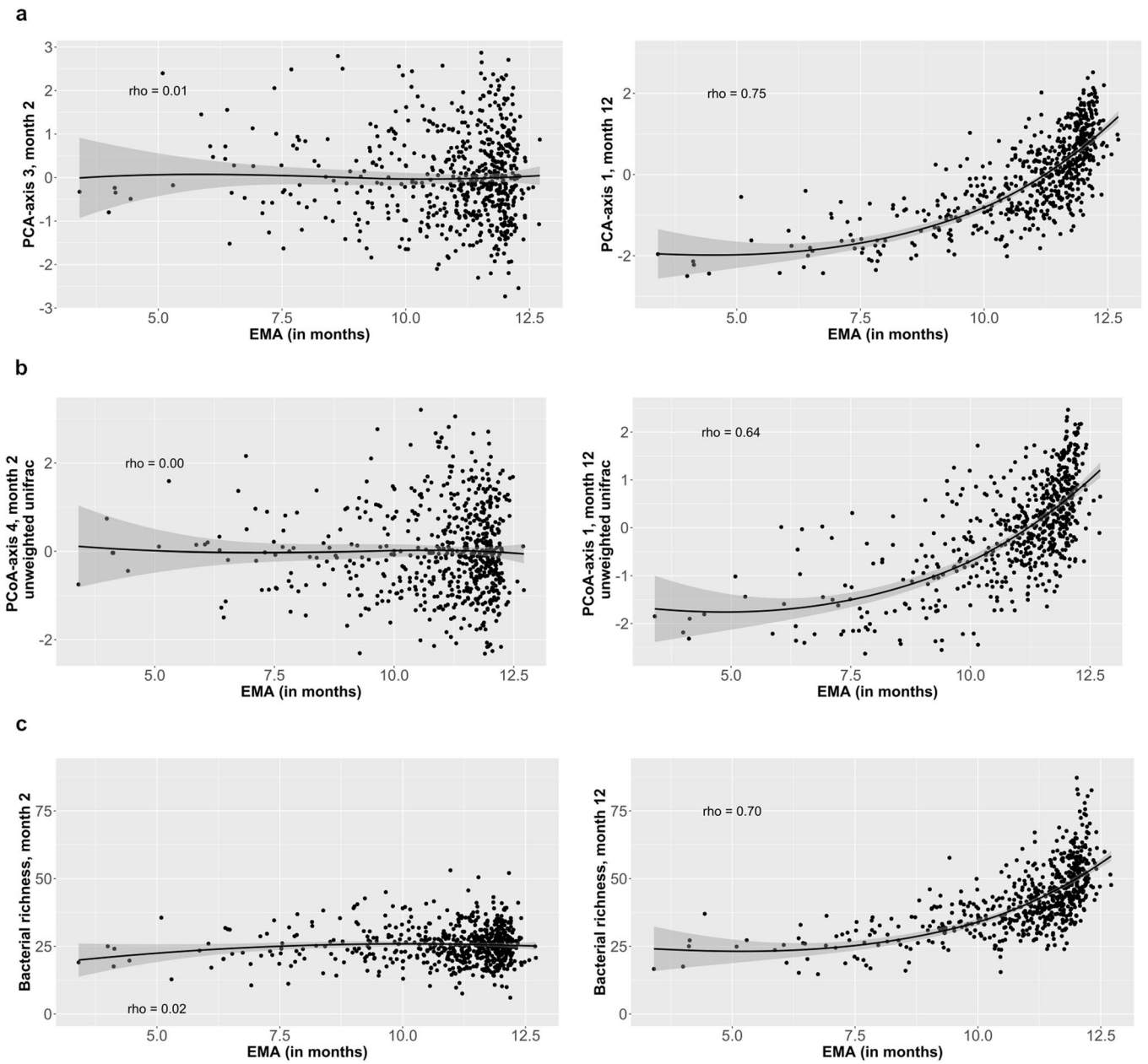
Extended Data Fig. 2 | Sensitivity analysis on the effect of estimated microbiome age (EMA). a, Scatter plot of the first two axes of a principal coordinate analysis (PCoA) over both time points on ASV (amplicon sequence variants) levels. The values in brackets represent percentage of variance explained corrected by negative eigenvalues. **b**, Scatterplot of the first PCoA-axis against EMA. **c**,**d**, Associations of asthma phenotypes with EMA restricted to individuals not included when establishing the prediction model (n = 480 children), that is the 618 children with measurements at both time points minus the 138 healthy individuals. EMA is used as z-standardized continuous variable (**c**) and dichotomized at the lowest quartile (**d**).



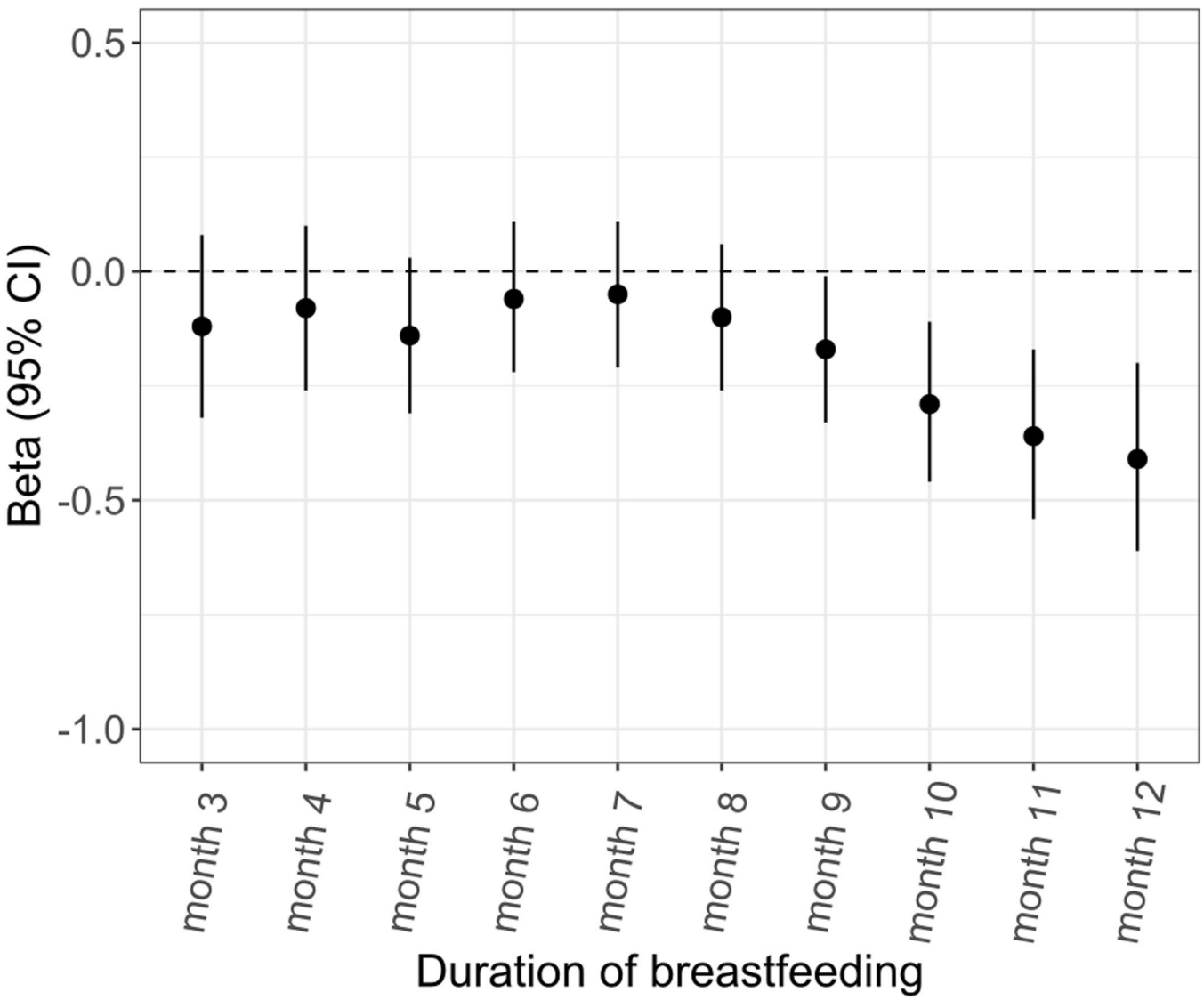
Extended Data Fig. 3 | Principal components analysis for microbial composition at 2 and 12 months and estimated microbiome age (EMA). a, Associations of asthma phenotypes with the first five axes of a principle component analysis (PCA) at month 2. **b**, Correlation of the asthma-protective PCA-axis 3 at month 2 (7% variance explained) with single genera. **c**, Associations of asthma phenotypes with PCA-axes at month 12. **d**, Correlation of the asthma-protective PCA-axis 1 at month 12 (14% variance explained) with single genera. **e**–**g**, Mutually adjusted associations of EMA and the asthma-protective axes at both time points with asthma (**e**), atopic asthma (**f**), and non-atopic asthma (**g**).



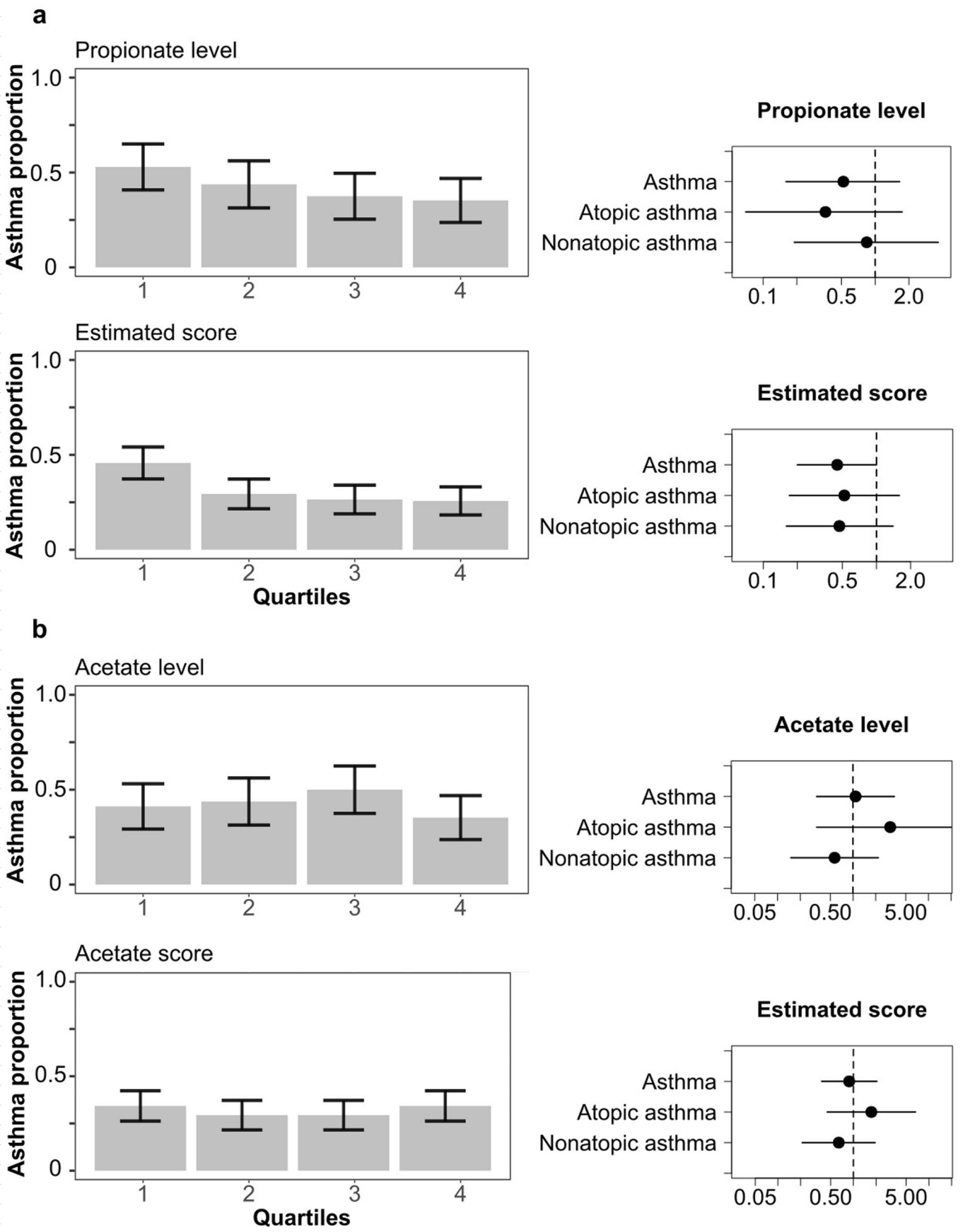
Extended Data Fig. 4 | Association of principle coordinate axes with asthma phenotypes and bacterial genera. Associations of asthma phenotypes of the first five axes of a principle coordinate analysis (PCoA) at month 2 **a**, and 12 **c**, using unweighted UniFrac as distance measure. Spearman correlations of the 10 most positively and 10 most negatively correlated individual genera with the asthma-protective PCoA-axes at month 2 **b**, and 12 **d**. Mutually adjusted models for EMA and the asthma-protective PCoA-axes at month 2 and 12 for asthma **e**, atopic asthma **f**, and nonatopic asthma **g**. Associations are shown as odds ratios for the z-standardized variables.



Extended Data Fig. 5 | Correlation of estimated microbiome age (EMA) with asthma-protective axes and richness. Relationship between EMA (x-axis) and various microbial measures (y-axis) including asthma-protective PCA- **a**, and PCoA-axes **b**, and bacterial richness **c**. The left column relates to 2 months, the right column to 12 months. As correlation coefficient Spearman's rho is given.

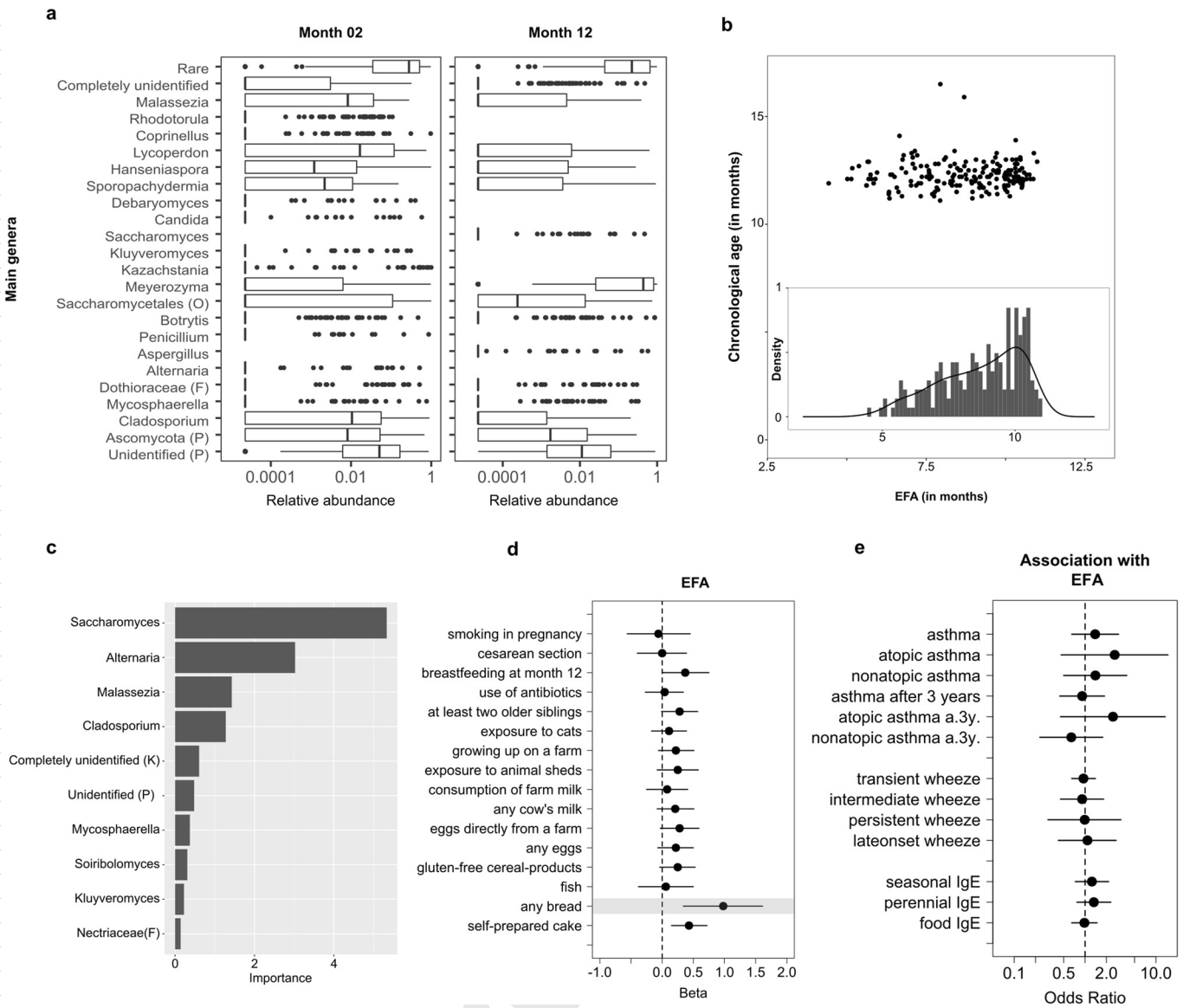


Extended Data Fig. 6 | Association of duration of breastfeeding with estimated microbiome age (EMA). Beta estimates of linear regression model of EMA versus duration of breastfeeding dichotomized at the indicated time points.

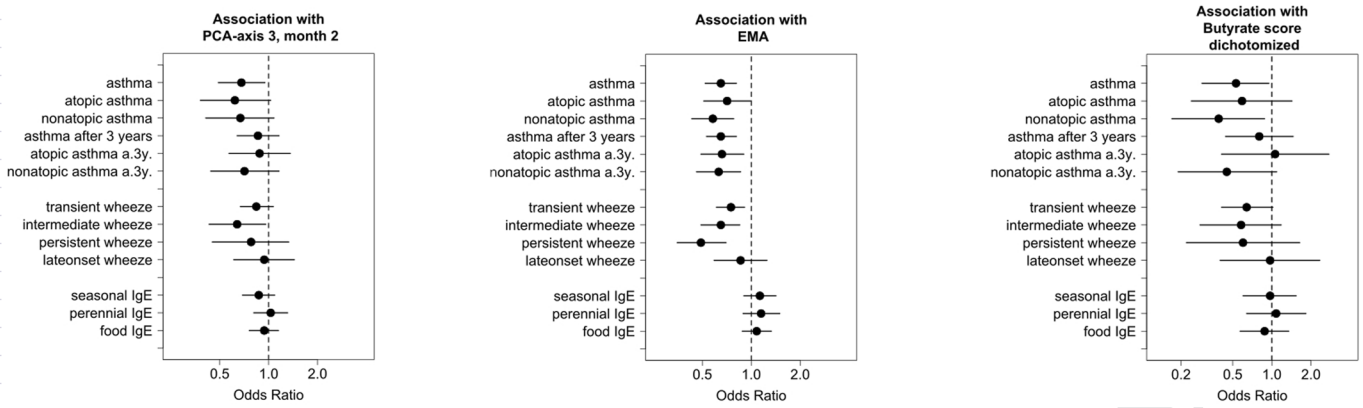


Extended Data Fig. 7 | See next page for caption.

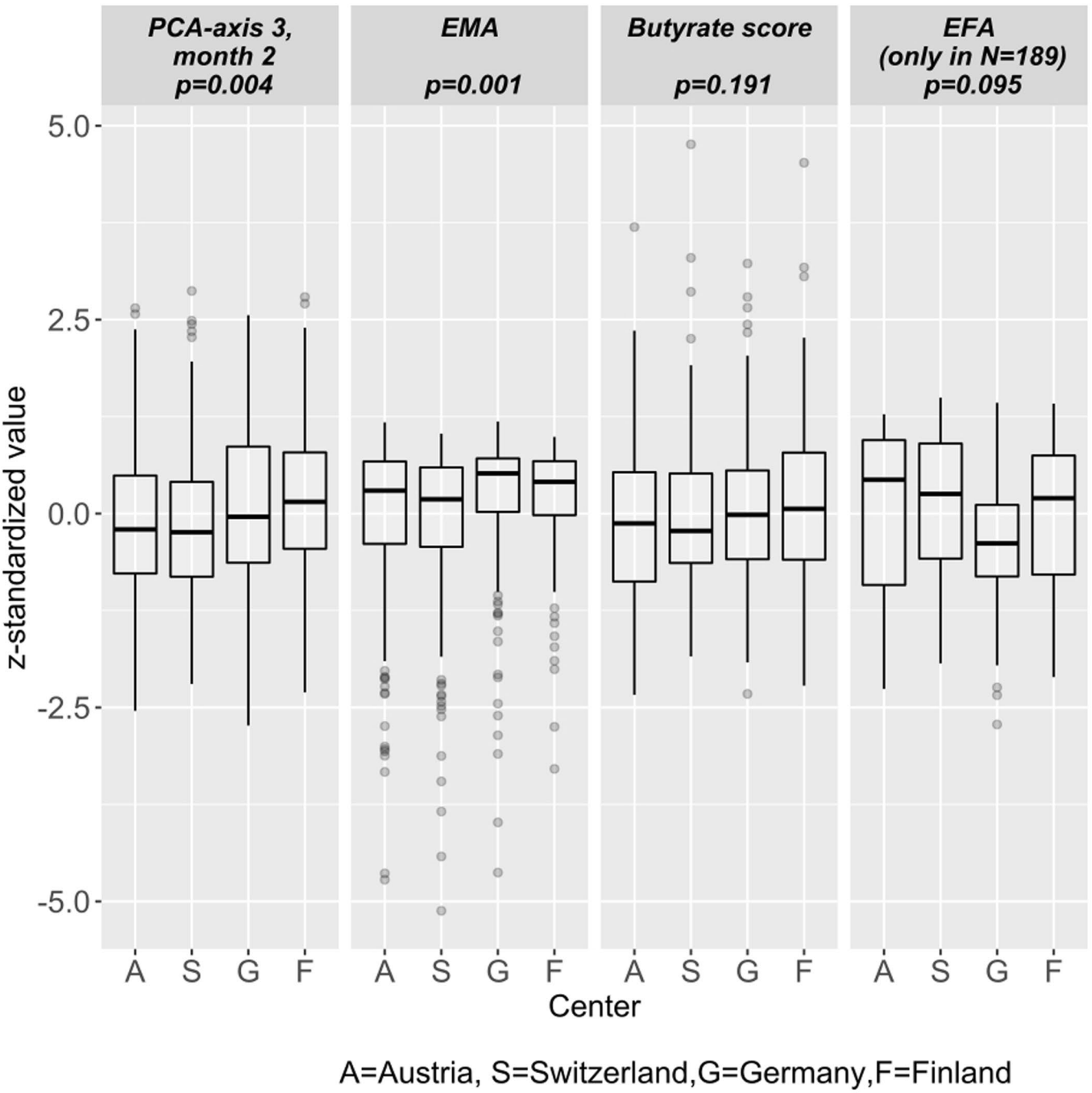
Extended Data Fig. 7 | Acetate score and propionate score in the case-control sample. The upper panels **a**, refer to propionate, the lower panels **b**, to acetate. The left column gives proportion of asthma cases within quartiles of the respective short-chain fatty acid (SCFA) variables. The right column gives odds ratios with 95%-confidence intervals for the associations of asthma phenotypes with the respective dichotomous SCFA variables (upper quartiles versus lowest quartile). Propionate and acetate level designate measured SCFA levels, whereas the estimated scores refer to the prediction models of measured SCFA levels by the microbial composition.



Extended Data Fig. 8 | The gut mycobiome and estimated fungal age (EFA). a, Log-scaled box and whiskers plots of relative abundance of most common fungal taxa at month 2 and month 12 in 189 children. Lower and upper hinges of the boxes denote the first and third quartiles, respectively; the bold central line represents the median; the whiskers extend to the most extreme data point within 1.5 times the interquartile range from the hinges; extreme values lie beyond the whiskers and are marked by circles. Missing boxes indicate relative abundance < 0.5% at the respective time point. '(F)', '(O)', or '(P)' stand for unclassified genera of the respective fungal family, order or phylum. **b**, Chronological age, that is the exact sampling time point in months plotted against estimated fungal age (EFA) illustrates that all chronologic information is largely removed from EFA. The density plot included in panel b reveals a skewed distribution of EFA. **c**, Fungal taxa most importantly predicting fungal age in the 35 healthy individuals. **d**, Determinants of EFA in the population with ITS data. Odds ratios are given with 95%-confidence intervals. Listed are determinants with p-values <0.01 in bivariate analyses; only consumption of any bread (marked in red) remains in a multivariable model. **e**, Associations of asthma phenotypes with EFA.



Extended Data Fig. 9 | Association of asthma phenotypes with microbial measures. Asthma was defined as a doctor's diagnosis of asthma or recurrent obstructive bronchitis. Asthma after 3 years was defined as a doctor's diagnosis of asthma or recurrent obstructive bronchitis after the age of 3 years. The atopic and nonatopic phenotypes of asthma were defined by presence or absence of concomitant sensitization to inhalant allergens with specific IgE ≥ 0.7 IU/ml at age 6 years. Wheeze phenotypes were defined by a latent class analysis as previously performed[14]. Transient and intermediate wheeze were milder forms with better lung function and less medication. Persis-tent wheeze was related to genetic risk encoded on chromosome 17q21 and displayed reduced lung function. Lateonset wheeze was particularly associated with atopic sensitization and fraction of exhaled nitric oxide. Seasonal IgE was defined as at least one specific IgE to alder, birch, hazel, grass pollen, rye, mugwort, plantain, or alternaria ≥ 0.7 IU/ml at age 6 years. Perennial IgE (D. pteronyssinus, D. farinae, cat, horse, or dog) and food IgE (hen's egg, cow's milk, peanut, hazelnut, carrot or wheat flour) were defined in analogy.



Extended Data Fig. 10 | Distribution of the microbial variables over the study centers. PCA=principal component analysis, EMA=estimated microbiome age, EFA=estimated fungal age; p-values are derived from two-sided Kruskal-Wallis tests. The analyses were performed in all 618 individuals with data available for the respective measures, except for EFA, where data was available only in 189 individuals. Lower and upper hinges of the boxes denote the first and third quartiles, respectively; the bold central line represents the median; the whiskers extend to the most extreme data point within 1.5 times the interquartile range from the hinges; extreme values lie beyond the whiskers and are marked by circles.

QUERY FORM

Nature Medicine	
Manuscript ID	**[Art. Id: 1095]**
Author	**Martin Depner**

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	*Nature of Query*
Q1:	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.
Q2:	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.
Q3:	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly. Note that changes here will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding email address(es).
Q4:	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published.
Q5:	Your paper has been copy edited. Please review every sentence to ensure that it conveys your intended meaning; if changes are required, please provide further clarification rather than reverting to the original text. Please note that formatting (including hyphenation, Latin words, and any reference citations that might be mistaken for exponents) has been made consistent with our house style.
Q6:	In the sentence beginning ‘The estimated microbiome age' please confirm the definition for the abbreviation OR.
Q7:	In the sentence beginning ‘At month 2 (Fig. 1a)', please confirm if these numbers are correct as a range (0.23-0.64) as elsewhere in this text, rather than separated by a semicolon. Similarly, please confirm the same for the sentence beginning ‘A sensitivity analysis showed independent'.
Q8:	In the sentence beginning ‘In month 12 samples' please specify which cluster is meant by ‘this cluster'.
Q9:	In the legend for Fig. 1a,b, please indicate what is meant by the square brackets.
Q10:	In the legend for Fig. 1a, the word ‘importance' seems vague, is there a more quantitative description for this?
Q11:	In the legend for Fig. 2c, please confirm that it is correct to say ‘except for those from Escherichia to Bacteroides'. Please also indicate the meaning of the square brackets.
Q12:	In the legends for Figs. 2f, 3a,b,d-f,h and 4c,d,g, a description of the dots or squares and bars in the figures was included from the Methods ‘Forest plots give point estimates with 95% confidence intervals.' Please confirm if this is correct.

QUERY FORM

Nature Medicine	
Manuscript ID	**[Art. Id: 1095]**
Author	**Martin Depner**

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	*Nature of Query*
Q13:	In the sentence beginning 'In contrast, EMA was delayed by prolonged', please confirm that it is correct to say 'visits to animal sheds'.
Q14:	In the legend for Fig. 3h, please indicate what is meant by (O).
Q15:	In the sentence beginning 'The presence of these ASVs was strongly associated', please provide a definition for GMR.
Q16:	Please check the sentence beginning 'Butyrate- and propionate-producing bacteria' for accuracy.
Q17:	In the sentence beginning 'The current analysis points toward', please check that it is correct to say 'pathology behind atopy'.
Q18:	In the sentence beginning 'Primers BITS', please indicate where the linker sequence is located.
Q19:	In the sentence beginning 'The resultant fecal homogenates', please check the units provided for flow rate.
Q20:	In the sentence beginning 'qPCR primers', please check that it is correct to say 'used for samples'.
Q21:	In the sentence beginning 'The data are expressed as a percent', please check that it is correct to say 'normalized to'.
Q22:	Please provide a Code availability statement, with a brief description of the code, if relevant.
Q23:	In the sentence beginning 'Random forests regression was performed', please check that it is correct to say the children 'never wheezed or had diarrhea'.
Q24:	In the sentence beginning 'This approach, developed by Aitchison', please provide a reference for the approach developed by Aitchison.
Q25:	In the sentence beginning 'To relate the composition of the network modules', please check for accuracy.
Q26:	Please check that all funders have been appropriately acknowledged and that all grant numbers are correct.
Q27:	If applicable, please ensure accession codes are scheduled for release on or before this article's scheduled publication date, and update the database record with publication details from this article once available.
Q28:	Please check that the Competing Interests declaration is correct as stated. If you declare competing interests, please check the full text of the declaration for accuracy and completeness.

QUERY FORM

Nature Medicine	
Manuscript ID	**[Art. Id: 1095]**
Author	**Martin Depner**

AUTHOR:

The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	*Nature of Query*
Q29:	Please provide the page range for reference 60.
Q30:	For reference 64 (Pedregosa et al. 2011), please provide the title of the article.

nature research

Corresponding author(s): Markus Ege

Last updated by author(s): Aug 21, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	R 3.4.3; Excel 2016
Data analysis	Mplus 8.1; R 3.4.3; R 3.6.1; R packages phyloseq, DirichletMultinomial, Gmisc, ranger, composition, GUnifrac, rmeta, qgraph. igraph; QIIME2-2018.6; sabre; dada2 1.16.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Taxonomy was assigned using the GreenGenes database (greengenes.lbl.gov) for 16S rRNA sequences and the UNITE dynamic database (unite.ut.ee) for ITS sequences. All 16S rRNA and ITS sequences are deposited in the Supplementary Information without metadata. PASTURE is an ongoing birth cohort with fieldwork still being executed. As long as the study is not yet anonymized, European data protection legislation prohibits sharing of individual data, also when pseudonymized. Upon request, the authors will share aggregate data that do not allow identification of individuals.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Participant flow is illustrated by Extended Data Figure 1. All subsamples relevant for the respective figures are identified in Extended Data Figure 1. The PASTURE birth cohort aimed at including 1000 individuals (von Mutius & Schmid, Allergy 2006). This number was not derived from a sample size calculation based on an expected effect size.
Data exclusions	The French subsample of PASTURE was excluded because collection of fecal samples at month 2 was not scheduled by design. For the calculation of bacterial age, n=5 samples were excluded, for whom exact sampling time point was not given or implausible. This is explained in detail in the manuscript.
Replication	The key findings, i.e. association of farm exposure with bacterial age and bacterial age with asthma, were replicated over the 4 centers of the PASTURE study, and a common estimate was derived by a meta-analysis with fixed effects. Measurement of relative abundance of the butyryl-CoA:acetate CoA-transferase gene was performed in duplicates, and measurements were considered valid if standard deviation was below 0.4. This was necessary to eliminate technical artefacts such as air bubbles that may interfere with the optical fluorescence reading. Of all 157 children selected into the case-control study, valid results were obtained in 138 individuals (88%).
Randomization	Not applicable, since no intervention was performed. The PASTURE study is an observational birth cohort.
Blinding	Not applicable, since no intervention was performed. The PASTURE study is an observational birth cohort.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The main characteristics such as age and sex are given in the text, further population characteristics are described in Supplementary Table 1. Asthma phenotypes are detailed by study group in Supplementary Table 8. Articles with more detailed population characteristics on PASTURE are referenced.
Recruitment	Pregnant mothers were recruited through birth clinics and midwife offices in rural areas during the last trimenon of pregnancy. Women living on a farm were selected to a meet a proportion of 50%. Self-selection by affected families may have occurred; the resulting bias is likely to be mild as the asthma prevalence at 6 years (8.1%) is in the range expected from population based studies (8% - 11%, Ege et al. NEJM 2011, ref. 6). Because of the stratification for farming, self-selection of farmers did not play a role.
Ethics oversight	All aspects of the study were approved by the local institutional review boards in each country (Austria: Ethikkommission für das Land Salzburg; Finland: The Research Ethics Committee, Hospital District of Northern Savo; Germany: Ethik-Kommission der Bayerischen Landesärztekammer; Switzerland: Kantonale Ethik-Kommission St. Gallen; France: Comité Consultatif pour la Protection des Personnes en Recherche Biomédicale (CCPPRB) Commission Informatique et Libertés (CNIL)). Written informed consent was obtained from the parents or guardians.

Note that full information on the approval of the study protocol must also be provided in the manuscript.