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

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

Affiliations

[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, 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] Childrens Hospital of Eastern Switzerland, St. Gallen, Switzerland

[8] Swiss Institute of Allergy and Asthma Research (SIAF), University of Zurich, Davos, Switzerland

[9] University of Zurich, Zurich, 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 Medecine, University of Lorraine, Vandoeuvre les Nancy, France

[13] Department of Respiratory Disease, UMR/CNRS 6249 Chrono-environnement, University Hospital of 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] Member of the German Center for Lung Research

[19] Department of Public Health, University of Helsinki, University of Helsinki, Helsinki, Finland

[20] The Rowett Institute, University of Aberdeen, Foresterhill, Aberdeen, UK

[21] Dr. von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany

† 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.

*A list of authors and their affiliations appears at the end of the paper.

Addresses for correspondence

Dr. Markus Ege

Department of Pediatrics, Dr. von Hauner Children's Hospital

University Hospital, LMU Munich

Lindwurmstr. 4

D-80333 Munich

phone: +49 (89) 4400 57709

fax: +49 (89) 4400 57103

email: markus.ege@med.uni-muenchen.de

Abstract

Growing up on a farm is associated with an asthma-protective effect but the underlying mechanisms are largely unknown. In the PASTURE birth cohort, we modeled maturation using 16S rRNA sequence data of the human gut microbiome in infants from 2 to 12 months. The estimated microbiome age (EMA) in infants at 12 months was associated with previous farm exposure (β=0.27 [0.12-0.43], p=0.001, n=618) and reduced risk of asthma at school age (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 measured level of fecal butyrate (OR=0.28 [0.09-0.91], p=0.034), bacterial taxa predicting butyrate production (OR=0.38 [0.17-0.84], p=0.017), and the relative abundance of the gene encoding butyryl-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.

Introduction

Many adult diseases originate early in life.[1] In the prenatal period, environmental influences on disease development are filtered by the maternal organism. After birth, however, the infant interacts directly with the environment, beginning with the colonization of the 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 relation of microbial exposure and immune-mediated diseases such as allergies and asthma has been the basis for the hygiene hypothesis and its amendments explaining the epidemics 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 the sustained microbial exposure on farms.[6,7] This protective effect has mainly been attributed to 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 for the microbiome of the upper airways.[9,10] For the gut microbiome, the effect on airway disease is less obvious. Murine models have suggested protection from allergic inflammation in the lung through metabolites such as short chain fatty acids (SCFA) produced by certain gut bacteria.[11]

The human gut microbiome undergoes profound changes during the first year of life and starts stabilizing thereafter.[12,13] Hence, we hypothesized that particularly the first year of life represented a time window for exposures of the outer environment to shape the development of the human microbiome with possibly 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 Protection against Allergy—Study in Rural Environments (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% females), fecal samples were taken at month 2 and month 12. At both time points, we obtained sequence reads for the 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, 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 month 2 and 12

At month 2 (Fig. 1a), the genus *Bifidobacterium* was predominant. Despite a positive association of the relative abundance of *Bifidobacterium* with breastfeeding (β=0.43 [0.23; 0.64], p<0.001), this genus was not significantly related to subsequent asthma. At month 12 (Fig. 1b), the relative abundance of *Bifidobacterium* 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 over both time points by Dirichlet mixture modeling, we identified 5 clusters (DCs), with two clusters mainly representing the month 2 samples, two representing the 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% relative abundance (Fig. 1c, Supplementary Table 2). DC4 and DC5 demonstrated more stabilized bacterial patterns with the emergence of *Firmicutes*. In samples of month 12, children in this cluster showed a higher prevalence of asthma as compared to clusters DC4 and DC5 (Fig. 1e).

Microbial maturation

To better understand the physiologic changes of the gut microbiome during the first year, we modeled the exact age of fecal sampling by a random forest of the composition of bacterial genera at month 2 and 12 in individuals with samples available at both time points. Since this model estimates the biological age of the healthy microbiome, we termed the resulting prediction score 'estimated microbiome age' (EMA). To exclude interference with disease, we restricted the model building to 133 healthy individuals (67 farm children and 66 non-farm children) without diarrhea between month 2 and 12 and never affected by wheeze or asthma.

The taxa most importantly contributing 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 an increase 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 the 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 and month 12 samples with delayed EMA (Fig. 2d) thereby describing individuals not following typical maturation. As an alternative surrogate for maturation we explored a PCoA over both time points (Extended Data Fig. 2a), whose first axis strongly correlated with EMA (Extended Data Fig. 2b).

Children with any asthma displayed 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 the lowest quartile of EMA (Fig. 2e). Children with more advanced EMA 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 against the lowest quartile of EMA. Similarly as a continuous variable, higher EMA reduced the risk of asthma (OR=0.72 [0.56-0.93], p=0.011) and particularly of non-atopic asthma (Fig. 2f). The effect of EMA on asthma was not changed 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 encoded on chromosome 17q21 (Supplementary Table 3) and was also seen in a sensitivity analysis excluding the 133 children in 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 principle components analysis (PCA) approach designed for compositional data. At month 2, the 3rd 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 with the relative abundance of *Bacteroides* and *Parabacteroides* and inversely with *Enterococcus* (Extended Data Fig. 3b).

At month 12, the 1st PCA-axis was inversely related to non-atopic asthma (OR=0.62 [0.39-1.00], p=0.048) and correlated particularly with *Roseburia*, *Ruminococcus*, and *Faecalibacterium* (Extended Data Fig. 3c,d). A sensitivity analysis based on principle coordinate analysis (PCoA) using unweighted UniFrac as 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.

Estimated microbiome age 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 seen, while the asthma-protective PCA-axis 3 was positively associated with breastfeeding and inversely with C-section and maternal smoking during pregnancy (Fig. 3a) independently from gestational age. In contrast, EMA was delayed by prolonged breastfeeding (Extended Data Fig. 6) but positively influenced by growing up on a farm (β=0.27 [0.12-0.43], p=0.001) and particular farm exposures such as stays in animal sheds or 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 a more frequent consumption of all six main food items cereals, meat, bread, yogurt, cake, and vegetables / fruits (Supplementary Table 4). A sensitivity analysis showed independent effects on EMA by a diverse feeding pattern (β=0.18 [0.01; 0.34], p=0.034) and prolonged breastfeeding (β=-0.41 [-0.62; -0.21], p<0.001). Farm children were allocated more commonly to the more advanced Dirichlet 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 PASTURE children, i.e. individuals without measurements at month 2 (Fig. 3d). This effect was also consistent over the 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 potential confounders (Supplementary Table 5).

A structure 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 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=0.100). Farm children were characterized by a more mature microbiome including *Coprococcus* and *Roseburia* (Fig. 3h), known producers of short chain fatty acids (SCFA).

Bacterial metabolites and estimated microbiome age

To assess bacterial taxa by their capacity to produce SCFA, 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).

Since the SCFA prediction scores were correlated mutually and partially 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 Dim 2, which was unrelated to asthma. Dim 3 represented the propionate score and partially EMA and was inversely related to atopic asthma. Dim 4 mainly reflected the difference between EMA and the butyrate score and had an additional protective effect on non-atopic asthma.

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

In a nested case-control sample (44 cases and 94 controls), we validated the butyrate score as the most relevant SCFA score. For this purpose we determined by qPCR the relative abundance of the gene encoding a major bacterial enzyme in the butyrate metabolism, i.e. butyryl-CoA:acetate CoA-transferase. When comparing the results of this gene assay with the originally measured butyrate levels and the estimated butyrate score we found particularly high asthma prevalence figures in the lowest quartiles of the respective measures (Fig. 4f). The corresponding associations with the different asthma phenotypes were similar for all measures (Fig. 4g), e.g. OR with asthma = 0.28 [0.09-0.91], p=0.034 for measured level, 0.38 [0.17-0.84], p=0.017 for butyrate score, and 0.43 [0.19-0.97], p=0.042 for the gene assay. Likewise, the 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 bacteria versus single taxa

Focusing on the result of the maturation process, i.e. 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 reflected *Roseburia* and *Ruminococcus* and belonged to the green module, whose first eigenvector was correlated with EMA (r=0.73) and the butyrate score (r=0.68). The yellow module was moderately related to EMA (r=0.35) and contained two main taxa of EMA, i.e. *Blautia* and *Coprococcus*; the latter formed the third hub.

When exploring the association of asthma with amplicon sequence variants (ASVs) related to the three hubs, two asthma-protective ASVs, one of genus *Roseburia* (OR=0.42 [0.18-1.01], p=0.053) and

one of the genus *Coprococcus* (OR=0.38 [0.16-0.92], p=0.032) emerged (Supplementary Table 6). Using 16S ribosomal sequences database BLAST, 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 *Coprococcus eutactus* (92.4%). Presence of these ASVs was strongly related to 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 with 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 from 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 in analogy to EMA and mainly depended on changes in *Saccharomyces*, *Alternaria*, and *Malassezia*. EFA was determined by consumption of starchy foods and unrelated to subsequent asthma (Extended Data Fig. 8). Though EFA and EMA were uncorrelated (rho=0.02), relative abundance of the highly prevalent genus *Alternaria* at 2 months was associated with subsequent bacterial maturation (β=0.05 [0.01-0.10], p=0.032).

Sensitivity analyses

EMA was also inversely associated with an asthma diagnosis established 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 unrelated to 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 and asthma phenotypes were homogeneous between farm and non-farm children with the exception of the association of EMA and atopic asthma, which was only seen in non-farm children (0.68 [0.45-1.02], p=0.060). Unlike microbial maturation and composition, the butyrate score did not significantly vary 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, estimated microbiome age (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 the subsequent colonization of the human gut, as exemplified for *Bacteroides*.[12] The relevance of mode of birth 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 the increasing variety of environmental exposures affecting the growing child. Indeed, the most influential change in the development of the mammalian gut microbiome is probably the transition from breastfeeding to solid food.[16,17] bifidobacteria, whose early predominance may be fostered by the bifidogenic properties of breastmilk,[18] decrease after weaning. In our analysis, the asthma-protective PCA-axis at month 2 was mainly influenced by mode of birth and correlated with *Bacteroides*. The positive association of this axis with breastfeeding and its asthma-protective effect was not explained by bifidobacteria.

Independently from this very early phenomenon, bacterial maturation between month 2 and 12 exerted a strong protective effect on asthma. Various shifts in bacterial composition including the bacterial families *Lachnospiraceae*, *Ruminococcaceae*, and *Bifidobacteriaceae* suggest high plasticity of the intestinal microbiome throughout 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,[22] which may explain the tendency towards asthma risk by higher EMA values at month 2, particularly in children assigned to DC3.

EMA predicted asthma better than DC3 at any time point, emphasizing the developmental aspect of a favorable microbiome. Possibly the bacterial composition is not beneficial by itself 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. Since the effect of prolonged breastfeeding on EMA was independent from the diversity of solid foods, this finding

supports the idea that cessation of breastfeeding is a key factor for microbial composition and maturation.[12,16,17]

The particular setting of the PASTURE study revealed other main determinants of maturation, which were all related to farm exposure, the epitome of the hygiene hypothesis. Farm children are known to be exposed to a larger variety of environmental microbiota[6] and potentially beneficial clusters of microorganisms.[23] There may be various mechanisms involved in the mediation of the protective effect of environmental microbiota on asthma: when playing in animal sheds, e.g., 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 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 feco-orally transmitted infections postulated as protective by the hygiene hypothesis.[5] The remaining 81% of the farm effect on asthma might be mediated by other mechanisms, operating also beyond the first year of life.

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

Inconsistencies between studies may arise from differences in time point of sampling. 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 even emerge subsequently as an asthma-risk taxon.[27] Though *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] Though *Bifidobacterium* has been suggested as a beneficial probiotic in other contexts,[29] it was increased in allergic children at later time points.[28] In our analysis, *Bifidobacterium* did not contribute to the asthma-protective effect. These inconsistencies were another reason for our integrative approach considering changes in bacterial composition over time.

A limitation of the current analysis might be seen in only two sampling time points, possibly missing fluctuations within this developmental window. Nevertheless, this drawback might emerge as an advantage, as comparing 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 the 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 has been validated by lung function measurements.[14]

To better understand how bacterial maturation may impact respiratory health we focused on the functional properties of the gut bacteria and modelled 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 shared with propionate production (Dim 3) was particularly relevant. Moreover, both aspects were involved in the mediation of the protective farm effect on asthma. In contrast, the likelihood of producing acetate, which predominantly represented Dim 2, was unrelated to asthma. This finding might argue in favor of a specific effect of butyrate and propionate in the human setting, where these SCFA, but not acetate, have been shown to impair 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 non-atopic asthma beyond butyrate production.

In order to validate the relevance of SCFA production we related asthma directly to SCFA levels as measured in the fecal samples[32] and found consistent associations with butyrate and a tendency for an association between atopic asthma and propionate. The gene assay targeting the main pathway of the bacterial butyrate metabolism, i.e. conversion of butyryl-CoA to butyrate via butyryl-CoA:acetate CoA-transferase,[33] corroborated the association of butyrate production and (non-atopic) asthma.

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

We found two promising asthma-protective ASV candidates of the genera *Roseburia* and *Coprococcus* (or *A. hallii*), which were also directly related to butyrate production and the butyryl-

CoA:acetate CoA-transferase gene. Adjustment for EMA, however, revealed that these taxa do not carry the asthma-protective effect themselves. Rather they may represent a network of bacteria with the capacity to produce SCFA. 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. The effect of butyrate- and propionate-producing bacteria reflects an aspect of healthy maturation of the gut microbiome and adds an independent component to the asthma-protective EMA effect, thereby extending the concept of the gut-lung axis[38,39] to the human setting.

The beneficial effect of gut microbiota may not be specific to respiratory health. Bacteria related to 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 lower abundance of *Roseburia* was also seen in rheumatoid arthritis.[41] Hence, the combination of the above taxa might be seen as a marker of a well-established host immune system and good general health in the absence of intestinal dysbiosis.[42]

The definition of phenotypes of asthma and atopy may vary over studies. Some studies focused on early outcomes;[20,21,26] few studies followed-up for atopic wheeze[43] or an asthma diagnosis[27] at age 5 years, when it 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 non-atopic 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 and asthma also in children of mothers without asthma. Therefore, studies focusing on atopic outcomes like atopic wheeze[20,43] or performed in high-risk populations[27] might reveal other facets of the microbial effect on asthma. The current analysis points towards an inflammatory pathology beyond atopy, supported by the anti-inflammatory properties of butyrate.[44]

Certainly butyrate is just an example and may be a marker of 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 implied in the health effects of the gut microbiome.[46] In addition, the vagus nerve can sense microbial metabolites with its afferent fibers and contribute to inflammation by a low tone of its efferent part, as illustrated for inflammatory bowel disease.[47] In analogy to 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.

Taken together, we found strong influences of an environment rich in microbial stimuli on the maturation of the gut microbiome. Maturation and prediction of butyrate production mediated partially 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.

Acknowledgment

We thank Prof. Bärbel Stecher-Letsch (Max-von-Pettenkofer Institute, Ludwig Maximilians University Munich, Germany) and Prof. Anne-Laure Boulesteix (Department of Medical Informatics, Biometry and Epidemiology of the Faculty of Medicine, Ludwig Maximilians University Munich, Germany) for valuable input in discussions about biological, respectively statistical issues. We are grateful to Dr. Giulia Pagani (Institute for Asthma and Allergy Prevention, Helmholtz Zentrum Munich, German Research Center for Environmental Health, Neuherberg, Germany) 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]). PL and FMF receive funding from the Scottish Government Rural and Environment Sciences and Analytical Services Division (RESAS).

PASTURE study group

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

From [1] Institute for Asthma and Allergy Prevention, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; [3] Department of Health Security, Finnish Institute for Health and Welfare, Kuopio, Finland; 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] Childrens Hospital of Eastern Switzerland, St. Gallen, Switzerland; [8] Swiss Institute of Allergy and Asthma Research (SIAF), University of Zurich, Davos, Switzerland; [9] University of Zurich, Zurich, 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 Medecine, University of Lorraine, Vandoeuvre les Nancy, France; [13] Department of Respiratory Disease, UMR/CNRS 6249 Chrono-environnement, University Hospital of 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] Member of the German Center for Lung Research; [19] Department of Public Health, University of Helsinki, University of Helsinki, Helsinki, Finland; [21] Dr. von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany; [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.

Author contributions

EvM, JCD, JR, and JP obtained funds, set up the PASTURE birth cohort and had responsibility for data collection and management of the study. DM, RL, MK, MR, HR, RF and CR were responsible for laboratory analyses. KMK and DM performed sequencing analyses. DH did bioinformatics, CR and RF performed SCFA-analyses. FMF and PL designed and performed the butyryl-CoA:acetate CoA-transferase assay. ESH was involved in data management and SP did statistical network analysis. PVK, AMK, and ADC were involved in acquisition, and interpretation of data.

MD performed statistical analyses and completed the background literature; MJE supervised statistical analyses; MD and MJE drafted the manuscript; all authors provided substantial revisions and approved the final version of the manuscript. The PASTURE study group were involved in 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 Statement

Dr. Mills is a co-founder of Evolve Biosystems and has stock and stock options therein; he has received payment for lectures from Nestle Nutrition Institute and the Abbott Nutrition Institute.

Dr. Kabesch has a patent share on the diagnostic use of SNPs in ORMDL3 on chromosome 17q21.

Dr. Renz has received research support from DFG, BMBF, EU, Land Hessen, DAAD, ALK, Stiftung Pathobiochemie, Ernst-Wendt-Stiftung, Mead Johnson Nutritional, Beckman Coulter, speakers honoraria from Allergopharma, Novartis, ThermoFisher, Danone, Mead Johnson Nutritional, Bencard, and consulting fees from Bencard and sterna-biologicals. He is a co-founder of sterna-biologicals.

Dr. von Mutius is listed as an inventor on the following patents: publication number EP 1411977, composition containing bacterial antigens used for the prophylaxis and the treatment of allergic diseases, granted on 18 April 2007; publication number EP1637147, 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. von Mutius is listed as inventor and has received royalties on the following patent: publication number EP2361632, specific environmental bacteria for the protection from and/or the treatment of allergic, chronic inflammatory and/or autoimmune disorders, granted on 19 March 2014. She has received Funding/Research Support from FrieslandCampina; she has received Consultation and Speaker fees: OM Pharma S. A., Böhringer Ingelheim International GmbH, Peptinnovate Ltd., Pharmaventures Ltd., Nestlé Deutschland AG (36 months prior to publication) and HiPP GmbH & Co KG (future).

Dr. Ege is a co-inventor of patents on the use of environmental bacteria for prevention from asthma (EP000001964570B1, US000009950017B2, and EP000002361632B1). His employer has received investigational products for an intervention study with minimally processed milk from FrieslandCampina.

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. *Nature medicine* **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. *Annals of the American Thoracic Society* **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 e1476 (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 e813 (2017).
11. Trompette, A.*, et al.* Gut microbiota metabolism of dietary fiber influences allergic airway disease and hematopoiesis. *Nature medicine* **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. *Science translational medicine* **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. *Journal of Biosciences* **44**(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. *Science translational medicine* **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. *Science translational medicine* **7**, 307ra152 (2015).
21. Fujimura, K.E.*, et al.* Neonatal gut microbiota associates with childhood multisensitized atopy and T cell differentiation. *Nature medicine* **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. *Journal of clinical microbiology* **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 (Lond)* **130**, 2199-2207 (2016).
27. Stokholm, J.*, et al.* Maturation of the gut microbiome and risk of asthma in childhood. *Nature communications* **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 (Baltimore)* **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. *Environmental microbiology* **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. *Frontiers in microbiology* **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. *The ISME journal* **12**, 610-622 (2018).
36. Reichardt, N.*, et al.* Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *The ISME journal* **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**(2018).
39. Marsland, B.J., Trompette, A. & Gollwitzer, E.S. The Gut-Lung Axis in Respiratory Disease. *Annals of the American Thoracic Society* **12 Suppl 2**, S150-156 (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 e410 (2018).
44. Zhai, S.*, et al.* Dietary butyrate suppresses inflammation through modulating gut microbiota in high-fat diet-fed mice. *FEMS Microbiol Lett* **366**(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* (2019).
47. Bonaz, B., Bazin, T. & Pellissier, S. The Vagus Nerve at the Interface of the Microbiota-Gut-Brain Axis. *Frontiers in neuroscience* **12**, 49 (2018).
48. Osadchiy, V., Martin, C.R. & Mayer, E.A. The Gut-Brain Axis and the Microbiome: Mechanisms and Clinical Implications. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* **17**, 322-332 (2019).
49. Kabata, H. & Artis, D. Neuro-immune crosstalk and allergic inflammation. *J Clin Invest* **130**, 1475-1482 (2019).

Figure legends

Figure 1: Composition of the bacterial gut microbiome at month 2 and 12

a,b. Log-scaled boxplots for relative abundance in 618 children at 2 and 12 months, respectively. 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. If names are in brackets, the respective bacterial genus shows a relative abundance < 0.5% at the respective time point. '(F)' stands for unclassified genus of the respective bacterial family. c. Log-scaled heat map for relative abundance of genera within the 5 clusters of a Dirichlet multinomial mixture modeling analysis across both time points resulting in 2 times 618 samples. d. Transition of all 618 individuals between the Dirichlet clusters (DC) from month 2 to month 12 e. Prevalence of asthma (with standard error bars, n=618) within the most prevalent clusters at month 2 and at month 12, respectively.

Figure 2: Estimated microbiome age (EMA) as a measure of maturation of the gut microbiome

a. Variable importance in the prediction model of estimated microbiome age (EMA) in the 133 healthy individuals. b. Stacked bar plots of relative abundance of main genera plotted against increasing EMA values. The individual samples (n=618 for each time point) are ordered by increasing EMA, the genera are ordered by Spearman's correlation with change in relative abundance between month 2 and 12. c. Spearman's correlation of EMA at month 12 with change in relative abundance between month 2 and 12 in 618 children. The color code corresponds to panel b. Correlation p-values were below 0.05 except for Escherichia to Bacteroides. d. Scatter plot for chronological age at fecal sampling against EMA at both time points (n=2 x 618). The color code reflects grouping by Dirichlet clusters over both time points. All subsequent analyses relate only to EMA at month 12. e. Scatter plot for chronological age at fecal sampling against 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. Distribution of EMA values and estimated density curve is given on top of the scatter plot (n=618); the vertical line denotes the lowest quartile, i.e. 10.6 months. f. Association of asthma phenotypes with Dirichlet clusters (left panel) and EMA values (right panel). Odds ratios are derived from logistic regression models (n=544 children with data on asthma). Simple models are only adjusted for center, mutually adjusted models are adjusted for center and the other microbial variable, i.e. EMA in the model for Dirichlet clusters and vice versa (A. = Asthma).

Figure 3: Estimated microbiome age and the farm effect on asthma

a. Bivariate associations of environmental determinants and the asthma-protective PCA-axis 3 at month 2 (n=618); colored bars denote determinants in the forward selection model. b. Bivariate associations of environmental determinants and the estimated microbiome age (EMA); colored bars denote determinants in the forward selection model. c. Transition of the 618 individuals between the previously defined Dirichlet clusters (DC) stratified by farming status. d. Meta-analysis for 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 these 720 children data on asthma were available. e. Meta-analysis of the effect of EMA on asthma over the study centers (n=720 children; Austria 173, Switzerland 209, Germany 176, Finland 162). f. Meta-analysis of the effect of growing up on a farm on EMA over the study centers (n=720 children, center distribution as in e). g. Mediation analysis of the protective effect of farming on asthma mediated by the estimated microbiome age (EMA, n=626). Shown are the estimates of the path model for indirect and direct 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).

Figure 4: Bacterial metabolites and estimated microbiome age

a. Variable importance for the prediction scores of the short chain fatty acids (SCFA) butyrate, propionate, and acetate, as modeled in 209 children with measured values. b. Principle component analysis (PCA) for estimated microbiome age (EMA) and the three SCFA scores (n=720). For illustrative purposes, Dimension (Dim) 3 is plotted against Dim 1 and Dim 4 against Dim 2. Explained variance is given in brackets. c. Associations of asthma phenotypes with the four PCA dimensions. d. Associations of growing up on a farm and consumption of farm milk with the four PCA dimensions. e. Mediation analysis of the protective effect of farming on asthma mediated by the four PCA dimensions (n=626). Shown are the estimates of the path model for indirect and direct effects; the proportion of the mediated (indirect) effect was 23%. f. Validation of the butyrate score in a case-control subsample of 138 children (44 asthma cases and 94 healthy controls). Proportion of asthma cases (with standard error bars) is given within quartiles of the originally measured butyrate level, the estimated butyrate score, and the gene assay, i.e. the relative abundance values of the gene for butyryl-CoA:acetate CoA-transferase, an enzyme of 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.

Figure 5: Network of single taxa and summary of findings

a. Network analysis of single genera (n=720). The resulting 5 modules are marked by different colors of the nodes. Positive correlations are marked by blue, negative correlations by red edges. Thickness of edges denotes strength of 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 pre-, peri-, and early postnatal influences such as prenatal smoke exposure or its absence, mode of birth, and breastfeeding. An inverse association with asthma at school age was found for a principle component at month 2. The maturation of the microbiome from 2 to 12 months was quantified by ‘estimated microbiome age’ (EMA), a prediction score derived from a random forest of sampling time point in relation to changes in the composition of bacterial genera over time. This EMA mediated the protective farm effect on asthma by about a fifth. 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 principle component analysis combining EMA and short-chain-fatty acid 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.

Tables

No tables in main text.

Online Methods

Study population

The Protection against Allergy—Study in Rural Environments (PASTURE) birth cohort has been recruited in rural areas of Austria, Finland, France, Germany, and Switzerland with the aim to explore protective effects of growing up on a farm on asthma.[50] Detailed information on the study design can be found in the Life Sciences Reporting Summary. Briefly, adult pregnant 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 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 child's napkin at 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 until 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 4th, 5th, or 6th year of life. Wheeze phenotypes were derived from a latent class analysis as described previously.[14]

Allergen-specific IgE was assessed at 6 years. Seasonal IgE was defined as at least one specific IgE to alternaria, alder, birch, hazel, grass pollen, rye, mugwort, or plantain≥0.7 IU/ml at age 6 years. Perennial IgE (D. pteronyssinus, D. farinae, cat, horse, dog) and food IgE (hen's egg, cow's milk, peanut, hazelnut, carrot and wheat flour) were defined in analogy. Assessment was done in

peripheral blood by using the semi-quantitative Allergy Screen test panel for atopy (Mediwiss Analytic, Moers, Germany) in a central laboratory.[53] The atopic and non-atopic phenotypes of asthma were defined by presence or absence of concomitant sensitization to inhalant (seasonal or perennial) allergens with specific IgE ≥0.7 IU/ml at age 6 years while the reference always were children without asthma.

Spirometry was performed as described previously,[54] and spirometric indices were calculated according 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 premature birth (less than 37 weeks of gestation), low birthweight (below 2500g) as well as variables for birthweight and gestational age dichotomized at the median, APGAR-score at 5 minutes, and delivery mode including natural vaginal birth, vaginal birth with forceps or vacuum extraction or cesarean section. Data on treatment with systemic antibiotics was available for the first 2 months and the first year beyond two months; maternal treatment with antibiotics during the first two months of lactation was also considered.

Breastfeeding was defined at any breastfeeding until 2 months or current breastfeeding at month 2. Duration of breast feeding was dichotomized at various cut-off levels from 2 to 12 months. Food diversity was defined as introduction of 5 out of 6 main food items (vegetables or fruits, cereals, meat, bread, yogurt, 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 plus egg, fish, nut, soy, margarine, chocolate, other milk products, cow's milk, butter), which were dichotomized at at least 11 items. Furthermore, the children's diet was assessed with respect to the type of supplemental food and its introduction in terms of at least weekly consumption.[56] Farm milk consumption was defined as weekly consumption of any milk obtained directly from a farm irrespectively of boiling or skimming.

Farm exposure was assessed as growing up on a farm or more specifically by regular contact to hay or stays in 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), presence of pets (cats, dogs) in the respective time periods, smoking in pregnancy, and environmental smoke exposure defined as at least one cigarette smoked at home by any person per day. In addition, parental history of atopy, which is a combination of asthma, hay fever, and atopic

eczema, or only asthma (maternal, paternal or of both) and high parental education (at least completion of secondary school) were included.

DNA-extraction of fecal samples

Fecal samples were frozen within 10 minutes from collection and stored at -20°C until further processing. At a central laboratory (THL Kuopio, Finland), DNA was extracted from the fecal samples in batches as follows: Partially defrosted fecal samples were homogenized using Stomacher® 80 micro-Biomaster (Seward Ltd, UK) laboratory paddle blender (2 min at high speed). DNA was extracted from 150 mg of ice-cold homogenized fecal sample, using bead-beating method with Zymo Research fecal DNA MiniPrep™ kit (Catalog No. D6010, Zymo Research, Irvine, CA) according to the manufacturer's instructions. The bead-beating step was done using FastPrep® FP120 homogenizer (2 min at full speed 6.5 m/s). Finally, the samples were eluted with 100 µL of elution buffer D3004-4-10 (Zymo Research, Freiburg, Germany). The sample extracts 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 x GoTaq Green Master Mix (Promega), 1 mM MgCl2, and 2 pmol of each primer. Samples were amplified in triplicate PCR reactions. Conditions consisted of an initial 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 of 72 °C for 10 min.

Primers BITS (5'–NNNNNNNNCTACCTGCGGARGGATCA–3') and B58S3 (5'–GAGATCCRTTGYTRAAAGTT–3') were used to amplify fungal internal transcribed spacer region 1.[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 x GoTaq Green Master Mix (Promega), 1 mM MgCl2, and 2 pmol of each primer. Reaction conditions consisted of an initial 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, a final extension of 72 °C for 5 min.

Amplicons were run on an 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 University of California Davis Genome Center DNA Technologies Core 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 was demultiplexed using sabre.[60] Demultiplexed data was imported into QIIME2-2018.6[61] and quality trimmed. Reads were denoised using DADA2[62] as implemented in QIIME2. Taxonomy was assigned to representative sequences using a naïve 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.

Taken sequences from 2 and 12 months together, 5,915 amplicon sequence variants (ASVs) were retrieved from 16S rRNA sequences, after excluding Chloroplasts. For ITS 54,459 ASVs were retrieved when restricted to fungi. Our 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 named the taxon 'completely unidentified'. Samples with <1000 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 the iPLEX Gold technology, a matrix assisted laser desorption / ionization – time of flight (MALDI-TOF) mass spectrometry system from SEQUENOM.[67] Technical errors were minimized by comparing genotype 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, Hercules, Calif) and 18S as reference gene. Quality checks were passed by samples of 896 children (95%). Of these, 512 children were included in the present analysis for 16S rRNA.

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, rs8076131 related to the *ORMDL3* gene and rs2290400 / rs7216389 related to the *GSDMB* gene at this locus were found to interact for childhood asthma with environmental tobacco smoking[71] and viral infections.[72] SNP rs8076131 was selected for the current analysis because it has been described in the context of functional regulation.[73]

Short chain fatty acids

Metabolite levels of short-chain fatty acids (SCFA) were measured in fecal samples obtained from 301 children of the PASTURE study at the age of 12 months.[32] These fecal samples were processed as previously described.[74] Briefly, 1 ml of 0.15 mM H2SO4 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, Darmstadt, Deutschland). The resultant fecal homogenates were analyzed by High Performance Liquid Chromatography (Merck Hitachi, Schaumburg, USA) using an Rezex ROA-Organic Acid H+ ion exchange column together with a SecurityGuard Cartidges Carbo-H from Phenomenex (Torrance, USA) at a flow rate of 0.4 ml at 40 °C with 10 mM H_2SO_4 as eluent solution. The samples were quantified in relation to standards.[75] Of these 301 children, 209 (69%) were included in the present 16S rRNA analysis (Figure S1).

Relative abundance of 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 asthmatics and 106 healthy controls with available fecal samples at 12 months. Quantitative PCR (qPCR) primers and annealing temperatures against 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 (2009).[76] qPCR equipment and reagents are described in Reichardt et al (2018).[35] DNA samples were used without further dilution unless their concentration (determined with a Qubit dsDNA HS assay kit, Thermo Fisher Scientific, Renfrew, UK) exceeded 4 ng/µl in the qPCR assay. The data are expressed as percent butyryl-CoA:acetate CoA-transferase gene copies of total bacterial 16S rRNA gene copies as detailed in Ramirez-Farias et al, 2009.[76] 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%). Data were stored in EXCEL.

Statistical analysis

Statistical analysis was performed with R 3.4.3 and 3.6.1 (https://www.R-project.org), particularly with package phyloseq, and MPlus (Muthén & Muthén, 1998-2012). Upon request, computer code will be made available to readers.

Relative abundance was used to describe taxonomic distribution of bacteria and fungi. Rare taxa were defined by a relative abundance below 0.5% in the respective population and subsumed in a

818 category termed “rare”. For logarithmic presentations, values were transformed by decadic
819 logarithm; where necessary zero values were replaced by the lowest measured value. All statistical
820 tests were performed two-sided, and an effective p-value <0.05 was considered statistically
821 significant.

822 ***Dirichlet Clustering***

823 Dirichlet multinomial mixtures (DMM) modelling was performed with R package
824 DirichletMultinomial. We clustered the samples over both time points and as a sensitivity analysis
825 separately for both time points. DMM bins samples on the basis of microbial community structure.[77]
826 The number of clusters was determined by a local minimum of a Laplace approximation score, i.e. 5
827 clusters over both time points and 3 clusters for the separate models. Transition between clusters
828 was illustrated by R package Gmisc.

829 ***Random Forests***

830 Random Forests regression was performed by R-package ranger to model sampling age on the basis
831 of the relative abundance of bacterial or fungal genera present at 2 and 12 months in a subsample of
832 133 (in case of fungi 35) healthy individuals, i.e. children without asthma, wheeze ever, or diarrhea
833 during the first year. The models were estimated using 2000 trees and the ceiling of the square root
834 of the number of selected variables per level. The resulting prediction model, mainly defined by
835 alterations in relative abundance of all genera were subsequently applied to the entire population,
836 using the function *predict* of ranger. These estimates were used as a proxy for bacterial or fungal
837 maturation and subsequently called estimated microbiome age (EMA) or estimated fungal age (EFA).
838 To confirm that results were independent of the training sets, we performed sensitivity analyses
839 restricting the models to children not included in model building. Taxa were ranked by their variable
840 importance in random forest models for EMA and EFA, respectively, which discriminate best
841 between the two measurement time points.

842 A similar approach was used to estimate short chain fatty acids (SCFA) scores for Butyrate,
843 Propionate, and Acetate. SCFA levels were modeled by the relative abundance of bacterial genera in
844 children with SCFA measures available. These prediction models were applied to predict SCFA
845 production scores in the entire population. Taxa were ranked by their variable importance in random
846 forest models for SCFA production. The number of randomly picked variables was optimized by 10-
847 fold cross-validation. As a member of the out-of-bag methods, random forest carries the advantage
848 over classical cross-validation that it yields an unbiased error estimate, i.e. high validity.[78] As random
849 forest integrates all information on microbial taxa in a single model no correction for multiple
850 comparisons was necessary. Besides continuous variables, the estimated microbiome age as well as

the butyrate score were also dichotomized at the lowest quartile in respective 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, i.e. 1029 16s rRNA sequences and 1000 ITS sequences. Rarefaction and calculation of species richness and Shannon's diversity index was iterated 1000 times and the resulting measures of alpha-diversity were subsequently averaged. An independent rarefying step was done in the sample of month 12 only for analyses of presence vs. absence of specific ASVs. For the assessment of the bacterial composition, the R package composition was used to perform 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.[79] The clr transformed abundance values were entered in principal component analyses (PCA) for assessing beta-diversity. In addition, beta-diversity was assessed by principal coordinate analyses (PCoA) on the ASV-level using unweighted UniFrac as distance measure calculated by R package GUnifrac. Samples taken at 2 months and at 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 only adjusted for center were termed "simple models", whereas the term "mutually adjusted models" refers to models where two exposures were compared and forced in the same model. Interaction was analyzed by including a product term in the regression models.

Confirmatory analyses on estimated microbiome age and farming or asthma were replicated over study centers and assessed by a meta-analysis with fixed effects (R package rmeta).

To compare indirect and direct effects, mediation models were calculated in Mplus (Muthén & Muthén, 1998-2012) and validated with the R package mediation. The mediated effect is reported as the proportion of the estimated indirect effect of the total effect.

To test for associations of single taxa with asthma we first tested for differences in relative abundance by Wilcoxon test; main associations ($p<0.1$) were then confirmed in the clr-transformed variables with logistic regression models. These models were adjusted for center only or additional for EMA to determine single taxa effects 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 performed two-sided.

Box and whiskers 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 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 correlations between different measures except for network analyses.

Network analyses

Correlations between pairs of bacterial genera were estimated using the SparCC approach.[80] The corresponding correlation network was visualized using the R package qgraph. In the network plot, only correlations with an absolute value greater than or equal to 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 function *eigen_centrality* 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 igraph function *cluster_fast_greedy*, which is a fast greedy algorithm for determining clusters via maximizing the modularity measure over all possible clusterings.[81]

To relate the composition of the network modules to EMA and the butyrate score, we used an approach similar to the eigengene analysis,[82] i.e. we used the first eigenvector of a PCA using the clr-transformed taxa passing the threshold-criteria to build the network plots as representative for the respective module.

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.

Methods-only 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 e1912 (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 e386 (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. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E139-148 (2014).
58. Caporaso, J.G.*, et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America* **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. Journal of Machine Learning Research* **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. *The ISME journal* **6**, 610-618 (2012).

66. Abarenkov, K*., et al.* The UNITE database for molecular identification of fungi--recent updates and future perspectives. *New 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. *Nature genetics* **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).

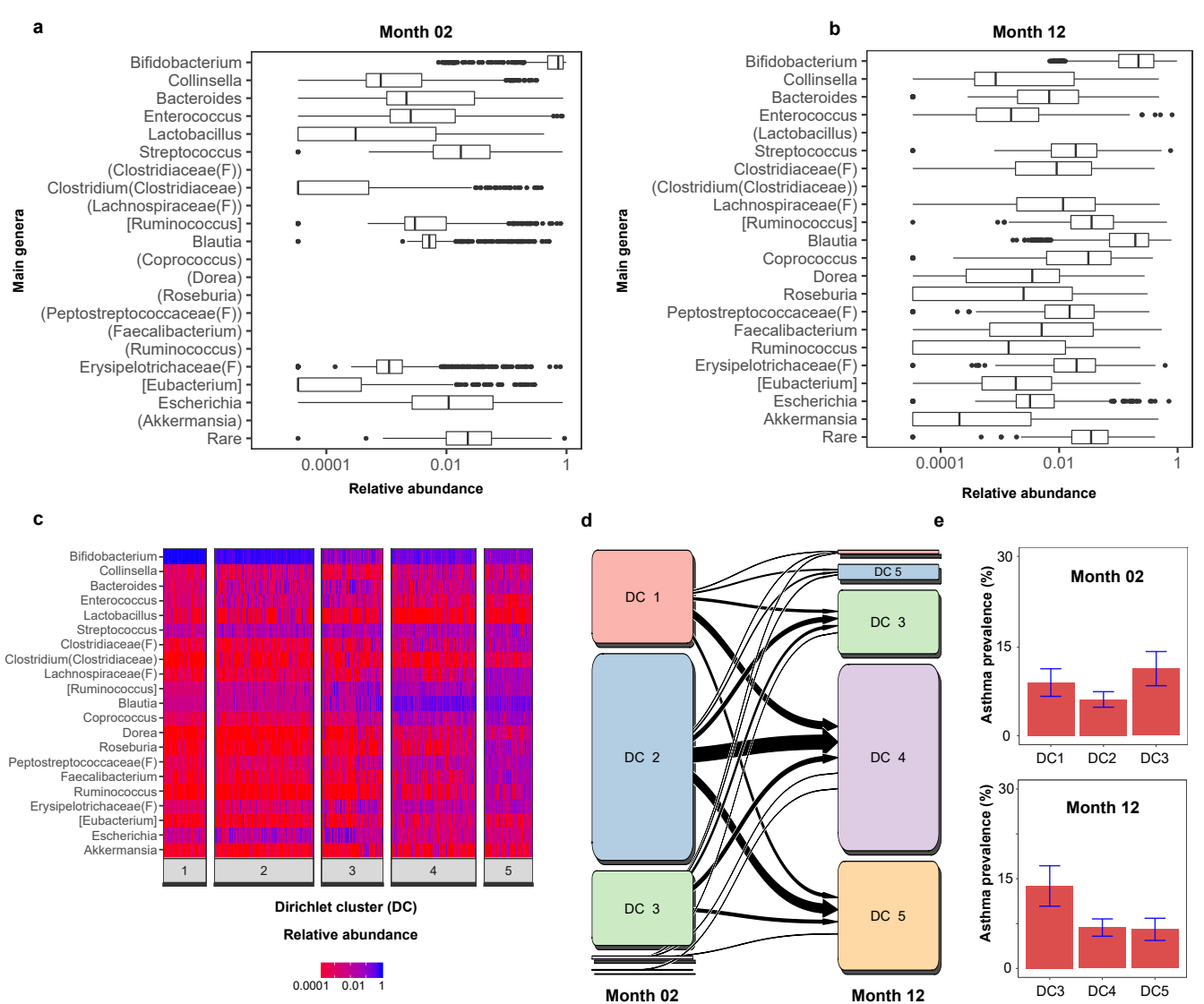
72. Caliskan, M*., et al.* Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med* **368**, 1398-1407 (2013).

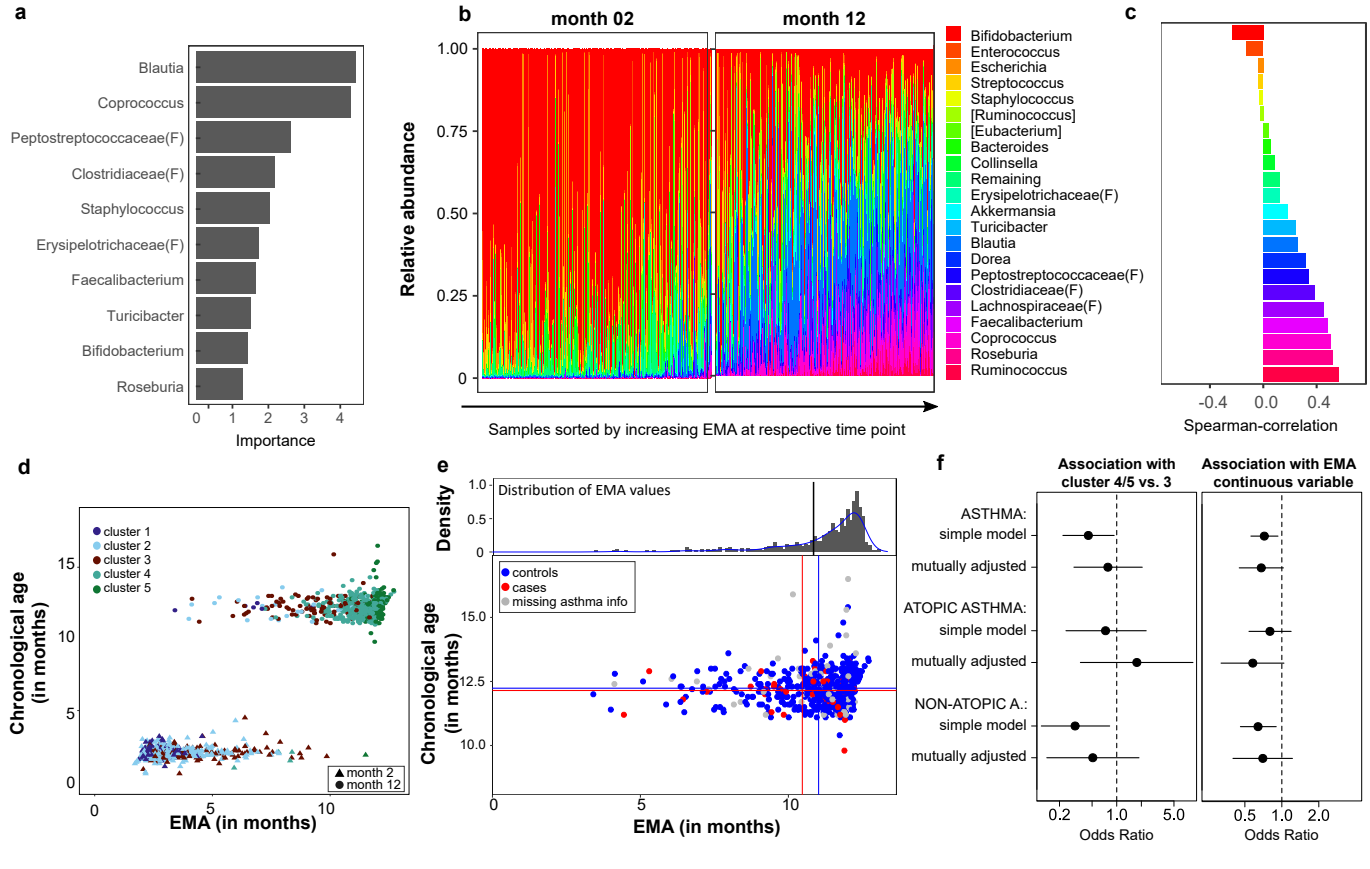
73. Schedel, M*., et al.* Polymorphisms related to ORMDL3 are associated with asthma susceptibility, alterations in transcriptional regulation of ORMDL3, and changes in TH2 cytokine levels. *J Allergy Clin Immunol* **136**, 893-903 e814 (2015).

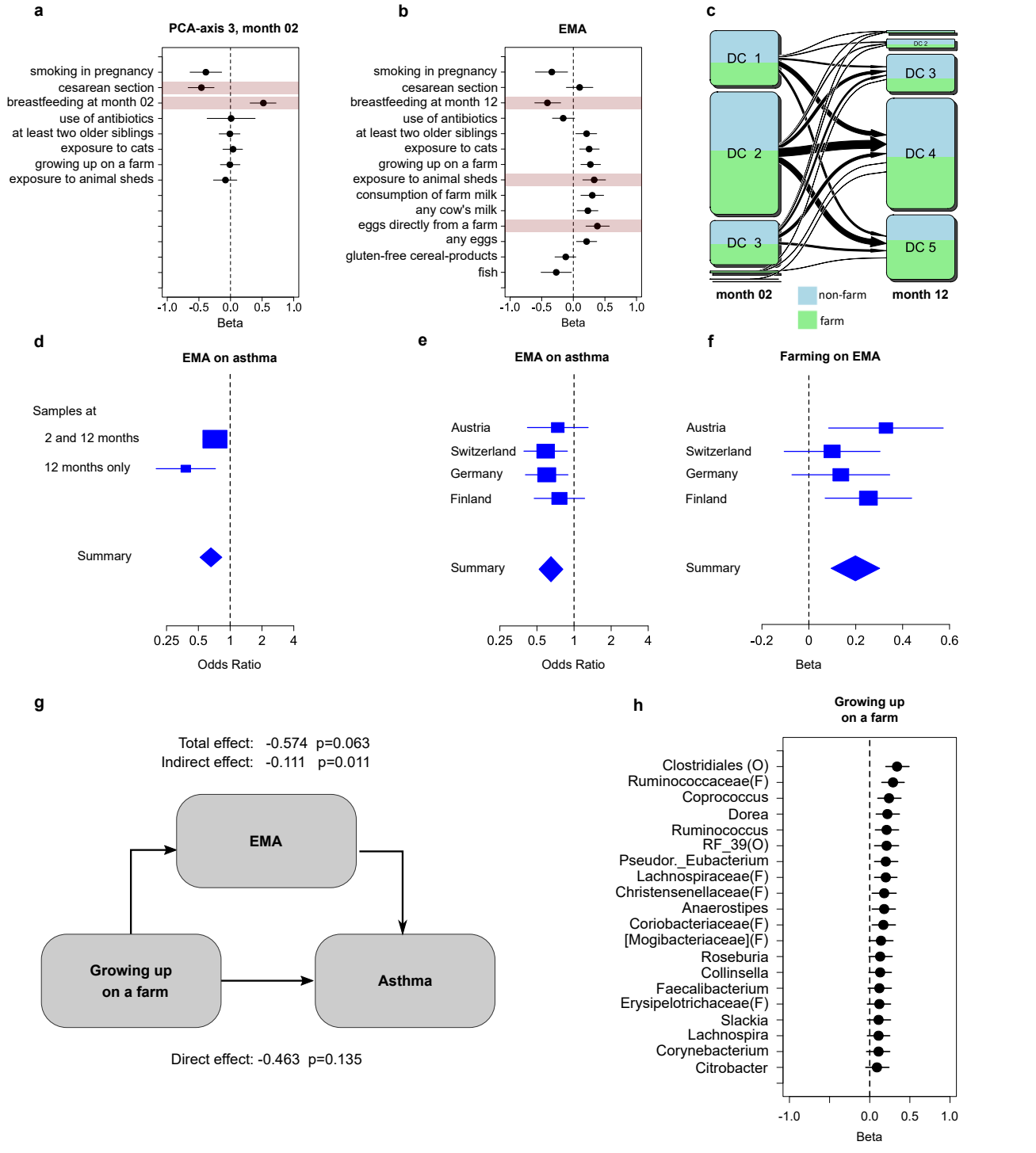
74. Dostal, A*., et al.* Effects of iron supplementation on dominant bacterial groups in the gut, faecal SCFA and gut inflammation: a randomised, placebo-controlled intervention trial in South African children. *Br J Nutr* **112**, 547-556 (2014).

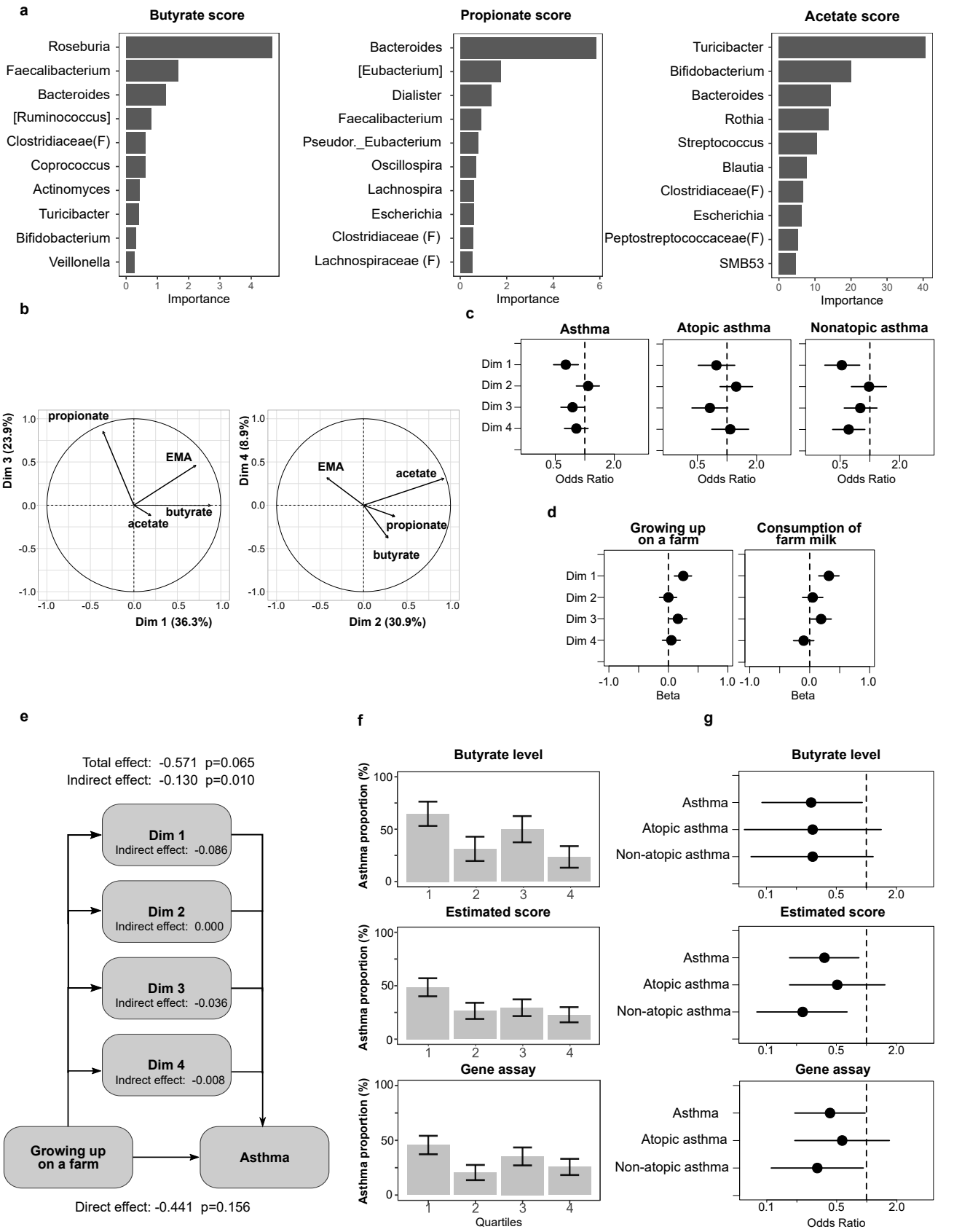
75. Pham, V.T., Lacroix, C., Braegger, C.P. & Chassard, C. Early colonization of functional groups of microbes in the infant gut. *Environmental microbiology* **18**, 2246-2258 (2016).

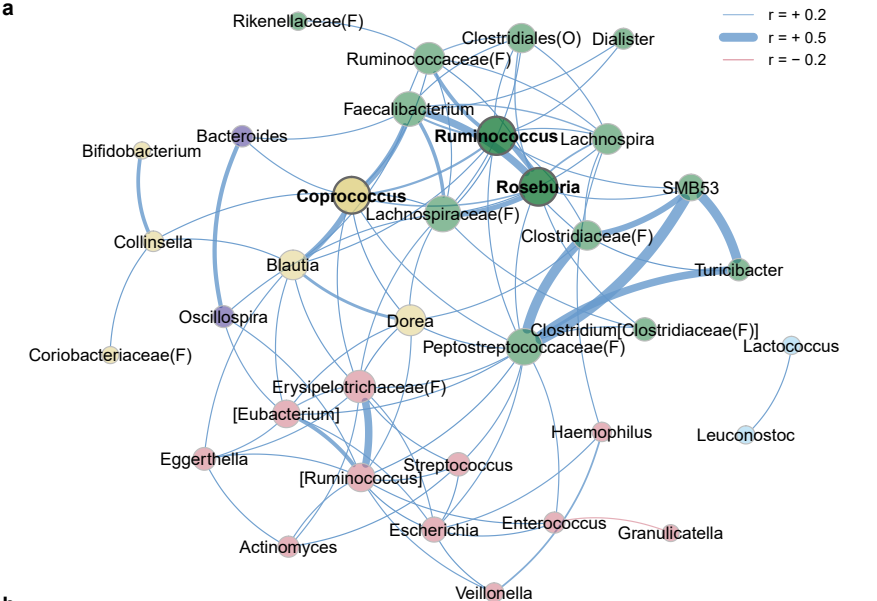
76. Ramirez-Farias, C*., et al.* Effect of inulin on the human gut microbiota: stimulation of Bifidobacterium adolescentis and Faecalibacterium prausnitzii. *Br J Nutr* **101**, 541-550 (2009).
77. Holmes, I., Harris, K. & Quince, C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. *PloS one* **7**, e30126 (2012).
78. Breiman, L. *Machine Learning* **45**, 5-32 (2001).
79. Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V. & Egozcue, J.J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in microbiology* **8**, 2224 (2017).
80. Friedman, J. & Alm, E.J. Inferring correlation networks from genomic survey data. *PLoS Comput Biol* **8**, e1002687 (2012).
81. Brandes, U*., et al.* On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering* **20**, 172-188 (2008).
82. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *BMC Syst Biol* **1**, 54 (2007).

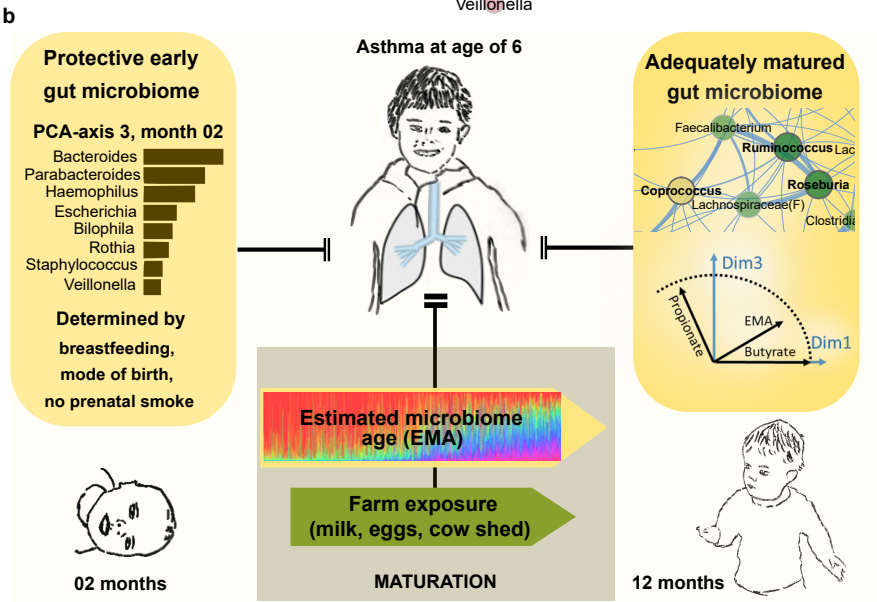

a
Month 02
Bifidobacterium
Collinsella
Bacteroides
Enterococcus
Lactobacillus
Streptococcus
(Clostridiaceae(F))
Clostridium(Clostridiaceae)
(Lachnospiraceae(F))
[Ruminococcus]
Blautia
(Coprococcus)
(Dorea)
(Roseburia)
(Peptostreptococcaceae(F))
(Faecalibacterium)
(Ruminococcus)
Erysipelotrichaceae(F)
[Eubacterium]
Escherichia
(Akkermansia)
Rare
Main genera
0.0001
0.01
1
Relative abundance
b
Month 12
Bifidobacterium
Collinsella
Bacteroides
Enterococcus
(Lactobacillus)
Streptococcus
Clostridiaceae(F)
(Clostridium(Clostridiaceae))
Lachnospiraceae(F)
[Ruminococcus]
Blautia
Coprococcus
Dorea
Roseburia
Peptostreptococcaceae(F)
Faecalibacterium
Ruminococcus
Erysipelotrichaceae(F)
[Eubacterium]
Escherichia
Akkermansia
Rare
Main genera
0.0001
0.01
1
Relative abundance
c
Bifidobacterium
Collinsella
Bacteroides
Enterococcus
Lactobacillus
Streptococcus
Clostridiaceae(F)
Clostridium(Clostridiaceae)
Lachnospiraceae(F)
[Ruminococcus]
Blautia
Coprococcus
Dorea
Roseburia
Peptostreptococcaceae(F)
Faecalibacterium
Ruminococcus
Erysipelotrichaceae(F)
[Eubacterium]
Escherichia
Akkermansia
1
2
3
4
5
Dirichlet cluster (DC)
Relative abundance
0.0001 0.01 1
d
DC 1
DC 2
DC 3
DC 5
DC 3
DC 4
DC 5
Month 02
Month 12
e
Month 02
Asthma prevalence (%)
30
15
0
DC1
DC2
DC3
Month 12
Asthma prevalence (%)
30
15
0
DC3
DC4
DC5


a
Blautia
Coprococcus
Peptostreptococcaceae(F)
Clostridiaceae(F)
Staphylococcus
Erysipelotrichaceae(F)
Faecalibacterium
Turicibacter
Bifidobacterium
Roseburia
0 1 2 3 4
Importance
b
month 02
month 12
Relative abundance
1.00
0.75
0.50
0.25
0
Samples sorted by increasing EMA at respective time point
Bifidobacterium
Enterococcus
Escherichia
Streptococcus
Staphylococcus
[Ruminococcus]
[Eubacterium]
Bacteroides
Collinsella
Remaining
Erysipelotrichaceae(F)
Akkermansia
Turicibacter
Blautia
Dorea
Peptostreptococcaceae(F)
Clostridiaceae(F)
Lachnospiraceae(F)
Faecalibacterium
Coprococcus
Roseburia
Ruminococcus
c
-0.4 0.0 0.4
Spearman-correlation
d
cluster 1
cluster 2
cluster 3
cluster 4
cluster 5
Chronological age (in months)
0 5 10 15
EMA (in months)
month 2
month 12
e
Distribution of EMA values
Density
1.0
0.5
0
controls
cases
missing asthma info
Chronological age (in months)
15.0
12.5
10.0
0 5 10
EMA (in months)
f
Association with cluster 4/5 vs. 3
Association with EMA continuous variable
ASTHMA:
simple model
mutually adjusted
ATOPIC ASTHMA:
simple model
mutually adjusted
NON-ATOPIC A.:
simple model
mutually adjusted
0.2 1.0 5.0
Odds Ratio
0.5 1.0 2.0
Odds Ratio


a
PCA-axis 3, month 02
smoking in pregnancy
cesarean section
breastfeeding at month 02
use of antibiotics
at least two older siblings
exposure to cats
growing up on a farm
exposure to animal sheds
-1.0 -0.5 0.0 0.5 1.0
Beta
b
EMA
smoking in pregnancy
cesarean section
breastfeeding at month 12
use of antibiotics
at least two older siblings
exposure to cats
growing up on a farm
exposure to animal sheds
consumption of farm milk
any cow's milk
eggs directly from a farm
any eggs
gluten-free cereal-products
fish
-1.0 -0.5 0.0 0.5 1.0
Beta
c
DC 1
DC 2
DC 3
DC 2
DC 3
DC 4
DC 5
month 02
month 12
non-farm
farm
d
EMA on asthma
Samples at
2 and 12 months
12 months only
Summary
0.25 0.5 1 2 4
Odds Ratio
e
EMA on asthma
Austria
Switzerland
Germany
Finland
Summary
0.25 0.5 1 2 4
Odds Ratio
f
Farming on EMA
Austria
Switzerland
Germany
Finland
Summary
-0.2 0 0.2 0.4 0.6
Beta
g
Total effect: -0.574 p=0.063
Indirect effect: -0.111 p=0.011
EMA
Growing up on a farm
Asthma
Direct effect: -0.463 p=0.135
h
Growing up on a farm
Clostridiales (O)
Ruminococcaceae(F)
Coprococcus
Dorea
Ruminococcus
RF_39(O)
Pseudor._Eubacterium
Lachnospiraceae(F)
Christensenellaceae(F)
Anaerostipes
Coriobacteriaceae(F)
[Mogibacteriaceae](F)
Roseburia
Collinsella
Faecalibacterium
Erysipelotrichaceae(F)
Slackia
Lachnospira
Corynebacterium
Citrobacter
-1.0 0.0 0.5 1.0
Beta


a
Butyrate score
Roseburia
Faecalibacterium
Bacteroides
[Ruminococcus]
Clostridiaceae(F)
Coprococcus
Actinomyces
Turicibacter
Bifidobacterium
Veillonella
Importance
Propionate score
Bacteroides
[Eubacterium]
Dialister
Faecalibacterium
Pseudor._Eubacterium
Oscillospira
Lachnospira
Escherichia
Clostridiaceae (F)
Lachnospiraceae (F)
Importance
Acetate score
Turicibacter
Bifidobacterium
Bacteroides
Rothia
Streptococcus
Blautia
Clostridiaceae(F)
Escherichia
Peptostreptococcaceae(F)
SMB53
Importance
b
propionate
EMA
butyrate
acetate
Dim 3 (23.9%)
Dim 1 (36.3%)
EMA
acetate
propionate
butyrate
Dim 4 (8.9%)
Dim 2 (30.9%)
c
Asthma
Atopic asthma
Nonatopic asthma
Dim 1
Dim 2
Dim 3
Dim 4
Odds Ratio
d
Growing up on a farm
Consumption of farm milk
Beta
e
Total effect: -0.571 p=0.065
Indirect effect: -0.130 p=0.010
Dim 1
Indirect effect: -0.086
Dim 2
Indirect effect: 0.000
Dim 3
Indirect effect: -0.036
Dim 4
Indrect effect: -0.008
Growing up on a farm
Asthma
Direct effect: -0.441 p=0.156
f
Butyrate level
Estimated score
Gene assay
Asthma proportion (%)
Quartiles
g
Butyrate level
Estimated score
Gene assay
Asthma
Atopic asthma
Non-atopic asthma
Odds Ratio


a
r = + 0.2
r = + 0.5
r = − 0.2
Rikenellaceae(F)
Clostridiales(O)
Dialister
Ruminococcaceae(F)
Faecalibacterium
Bacteroides
Ruminococcus
Lachnospira
Bifidobacterium
Coprococcus
Roseburia
SMB53
Lachnospiraceae(F)
Clostridiaceae(F)
Collinsella
Blautia
Turicibacter
Oscillospira
Dorea
Clostridium[Clostridiaceae(F)]
Peptostreptococcaceae(F)
Lactococcus
Coriobacteriaceae(F)
Erysipelotrichaceae(F)
[Eubacterium]
Haemophilus
Leuconostoc
Eggerthella
Streptococcus
[Ruminococcus]
Escherichia
Enterococcus
Granulicatella
Actinomyces
Veillonella


b
Protective early gut microbiome
PCA-axis 3, month 02
Bacteroides
Parabacteroides
Haemophilus
Escherichia
Bilophila
Rothia
Staphylococcus
Veillonella
Determined by breastfeeding, mode of birth, no prenatal smoke
Asthma at age of 6
Adequately matured gut microbiome
Faecalibacterium
Ruminococcus
Coprococcus
Roseburia
Lachnospiraceae(F)
Dim3
Propionate
EMA
Butyrate
Dim1
Estimated microbiome age (EMA)
Farm exposure (milk, eggs, cow shed)
MATURATION
02 months
12 months

PASTURE | AT | CH | DE | SF | FR

N=930

Microbiome (N=720)

16S rRNA at month 2 and 12

N=618

Asthma	N=43
Controls	N=501
NA	N=74

Healthy controls: N=133 **

Remaining: N=480

Fig. 1-2
Fig. 3a-d

16S rRNA at month 12 only

N=102

Asthma	N=10
Controls	N=72
NA	N=20

Fig. 3d

16S rRNA at month 12

N=720

Asthma	N=53
Controls	N=573
NA	N=94

Fig. 3e-h
Fig. 4a-e
Fig. 5a

Genetics

N=512

Asthma	N=34
Controls	N=419
NA	N=59

ITS at month 2 and 12 *

N=189

Asthma	N=10
Controls	N=150
NA	N=29

Healthy controls: N=35 **

Remaining: N=153

SCFA

Asthma	N=30
Controls	N=172
NA	N=7

Case-control

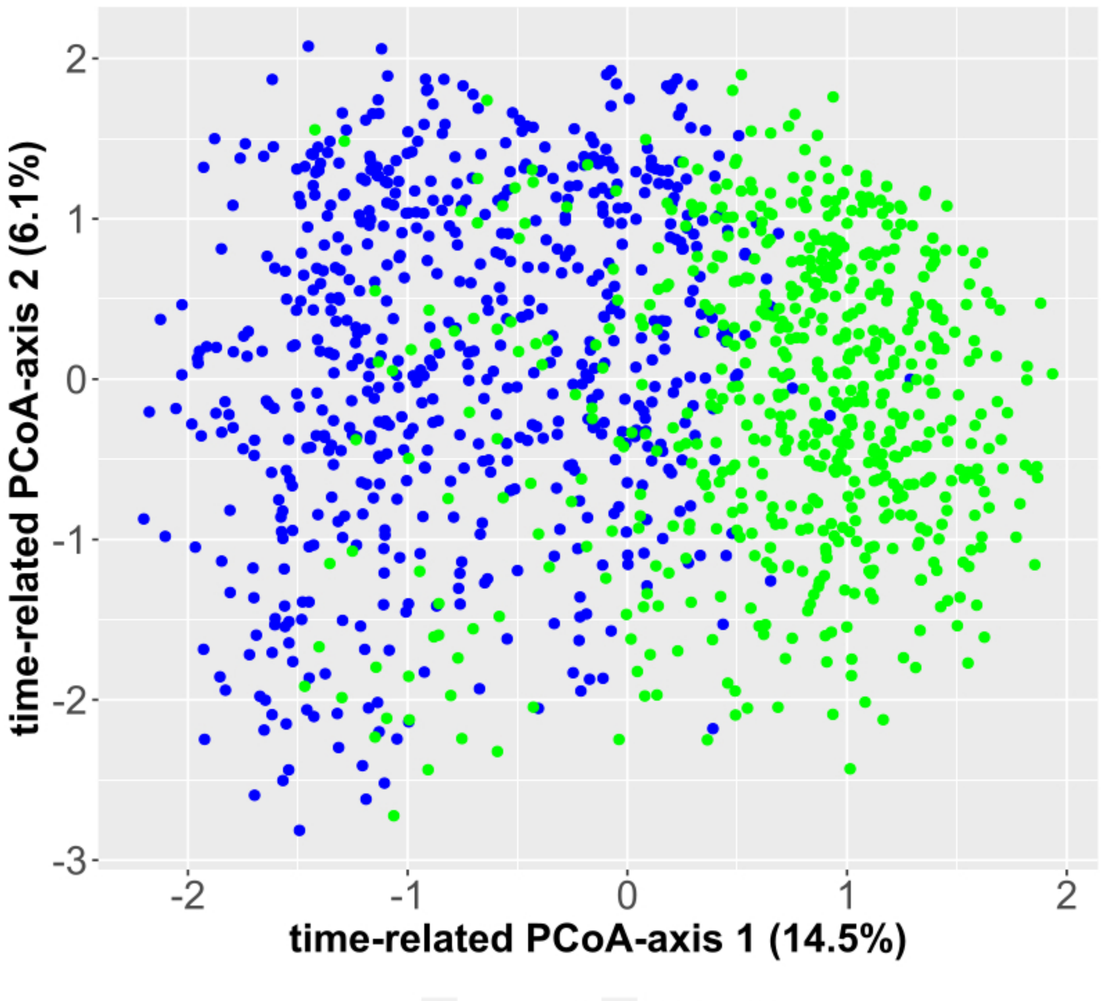
Asthma	N=51	N=44
Controls	N=106	N=94
NA	N=0	

Working*** samples

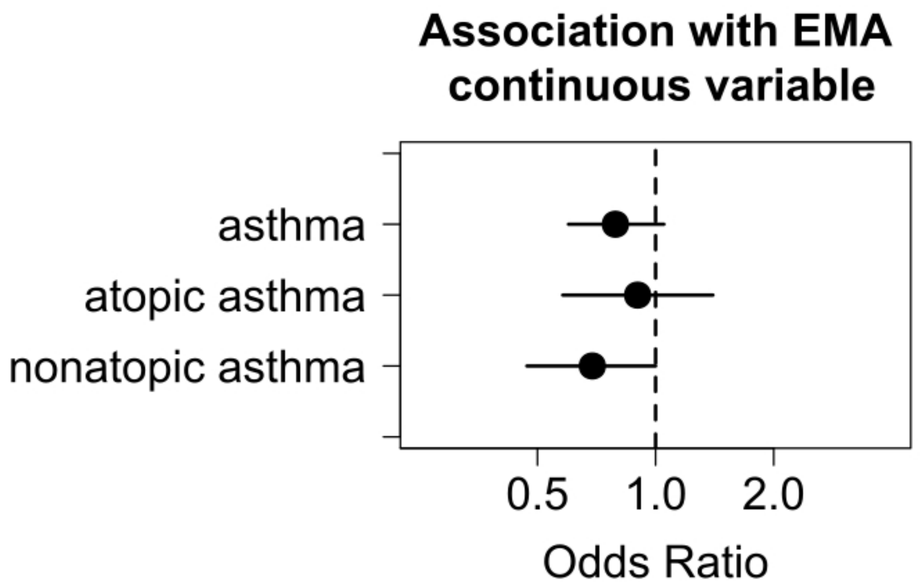
N=209
N=157

Fig. 4f-g

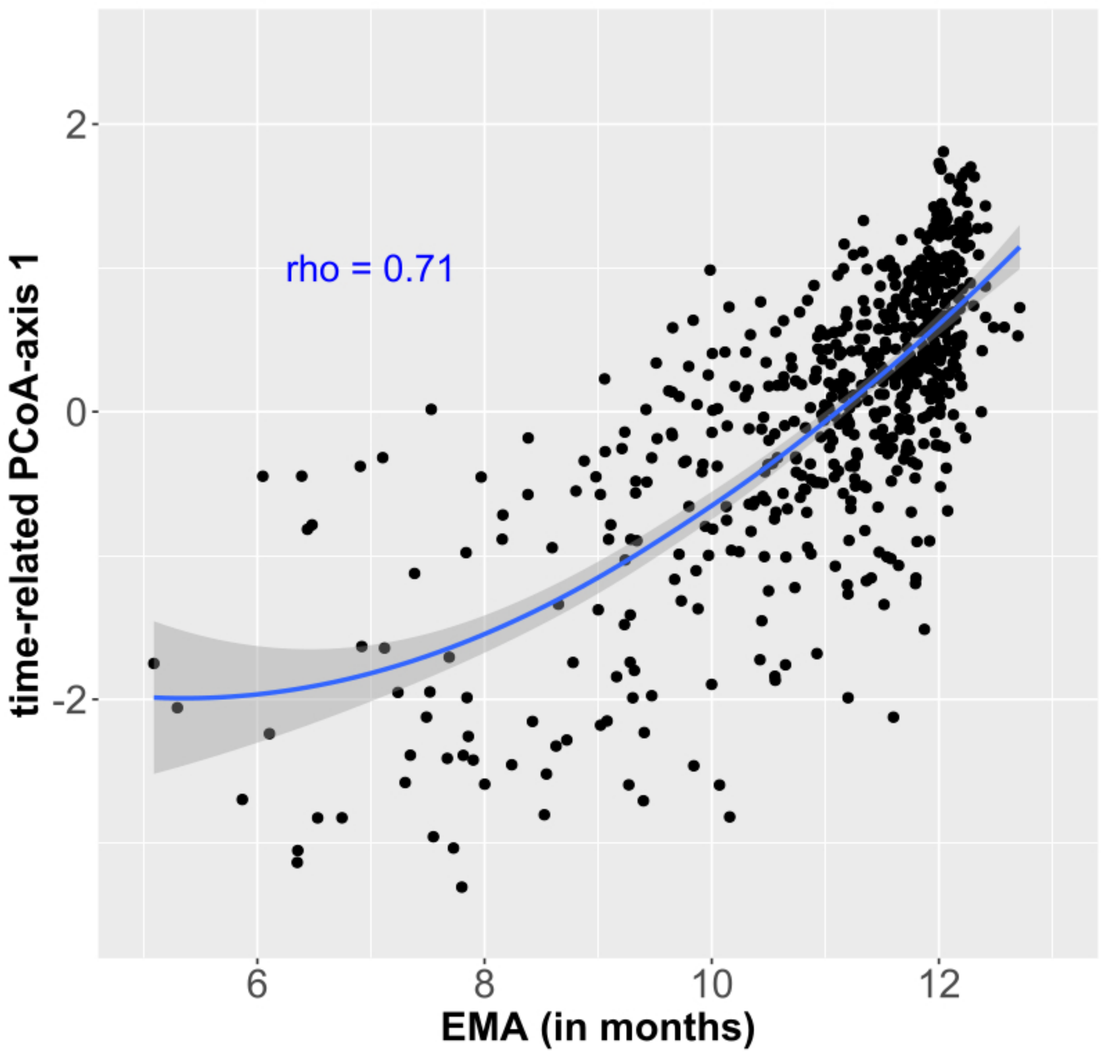
a


time-related PCoA-axis 2 (6.1%)
2
1
0
-1
-2
-3
-2
-1
0
1
2
time-related PCoA-axis 1 (14.5%)
month2
month12

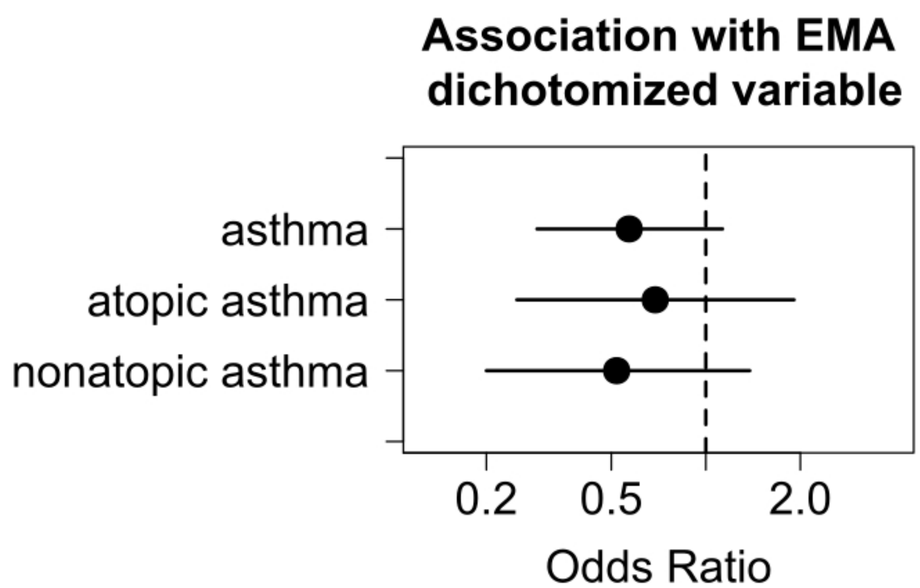
b


Association with EMA
continuous variable
asthma
atopic asthma
nonatopic asthma
0.5
1.0
2.0
Odds Ratio

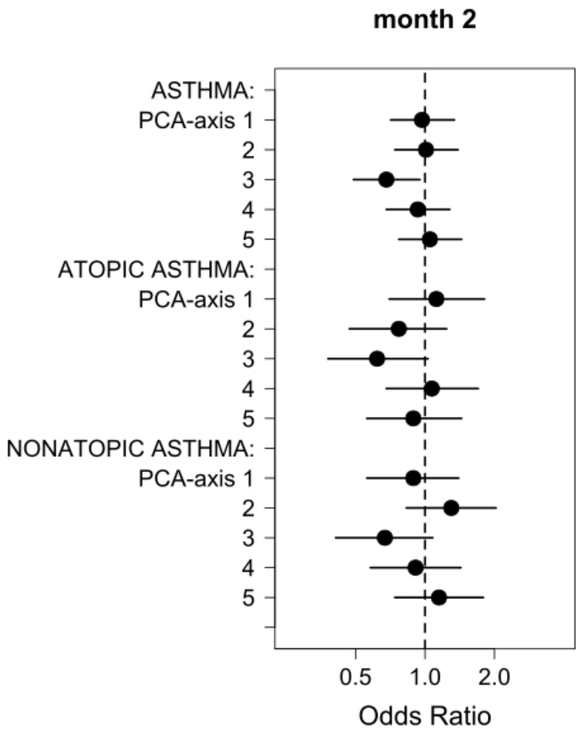
c


time-related PCoA-axis 1
2
0
-2
rho = 0.71
6
8
10
12
EMA (in months)

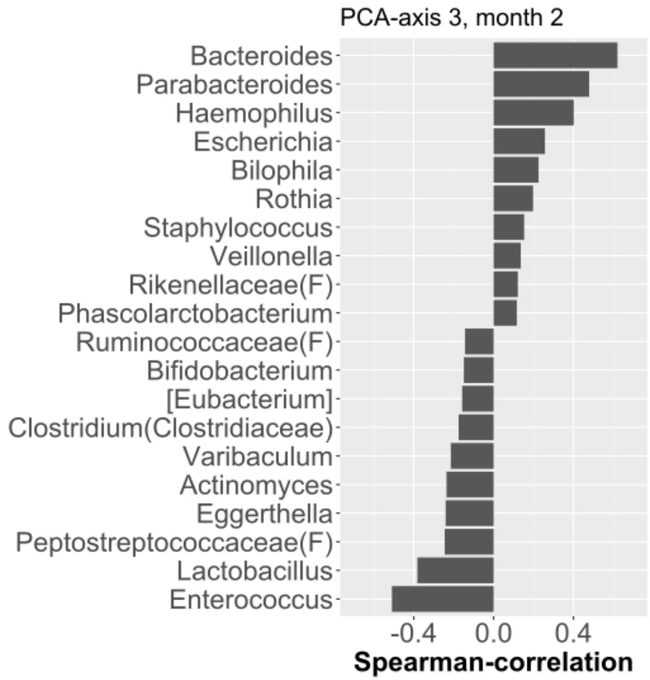
d


Association with EMA
dichotomized variable
asthma
atopic asthma
nonatopic asthma
0.2
0.5
2.0
Odds Ratio

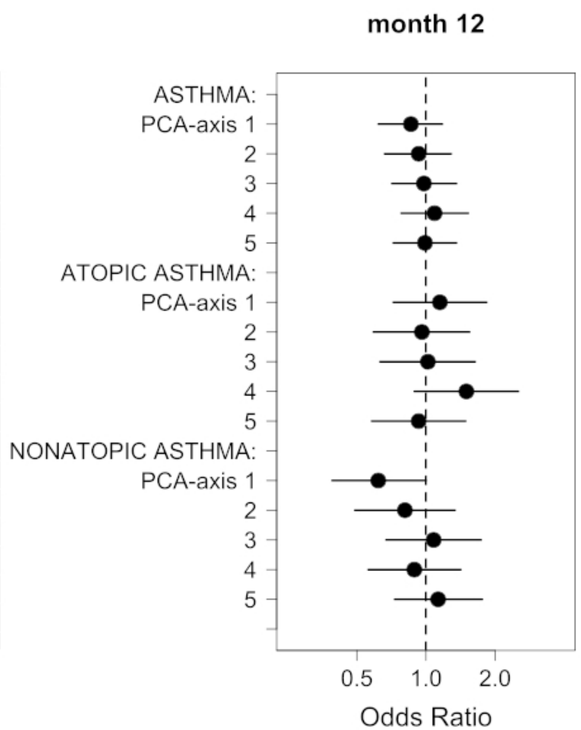
a


month 2
ASTHMA:
PCA-axis 1
2
3
4
5
ATOPIC ASTHMA:
PCA-axis 1
2
3
4
5
NONATOPIC ASTHMA:
PCA-axis 1
2
3
4
5
0.5
1.0
2.0
Odds Ratio

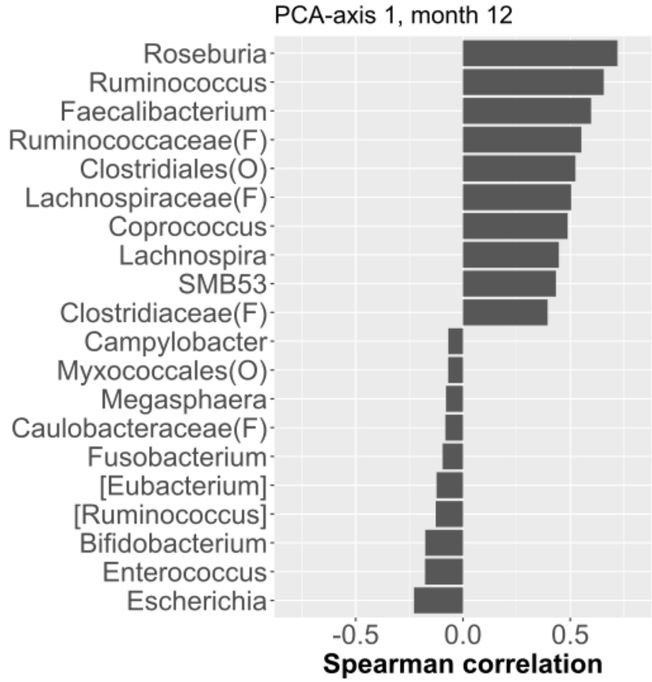
b


PCA-axis 3, month 2
Bacteroides
Parabacteroides
Haemophilus
Escherichia
Bilophila
Rothia
Staphylococcus
Veillonella
Rikenellaceae(F)
Phascolarctobacterium
Ruminococcaceae(F)
Bifidobacterium
[Eubacterium]
Clostridium(Clostridiaceae)
Varibaculum
Actinomyces
Eggerthella
Peptostreptococcaceae(F)
Lactobacillus
Enterococcus
-0.4
0.0
0.4
Spearman-correlation

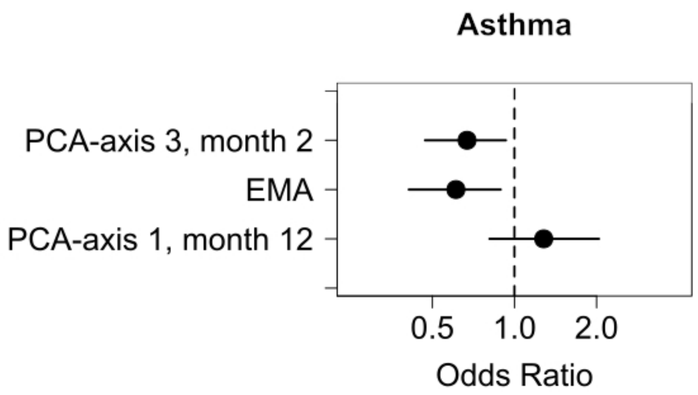
c


month 12
ASTHMA:
PCA-axis 1
2
3
4
5
ATOPIC ASTHMA:
PCA-axis 1
2
3
4
5
NONATOPIC ASTHMA:
PCA-axis 1
2
3
4
5
0.5
1.0
2.0
Odds Ratio

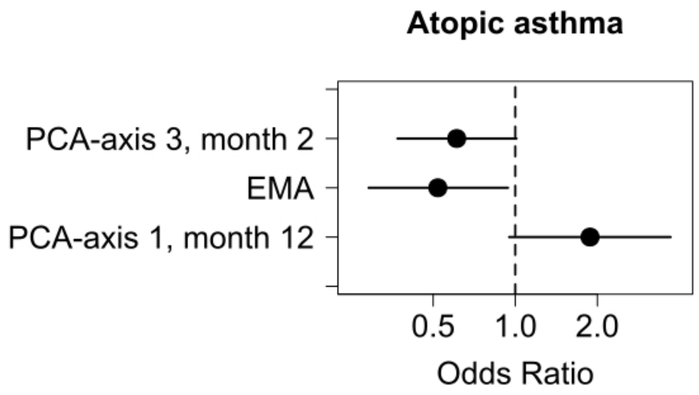
d


PCA-axis 1, month 12
Roseburia
Ruminococcus
Faecalibacterium
Ruminococcaceae(F)
Clostridiales(O)
Lachnospiraceae(F)
Coprococcus
Lachnospira
SMB53
Clostridiaceae(F)
Campylobacter
Myxococcales(O)
Megasphaera
Caulobacteraceae(F)
Fusobacterium
[Eubacterium]
[Ruminococcus]
Bifidobacterium
Enterococcus
Escherichia
-0.5
0.0
0.5
Spearman correlation

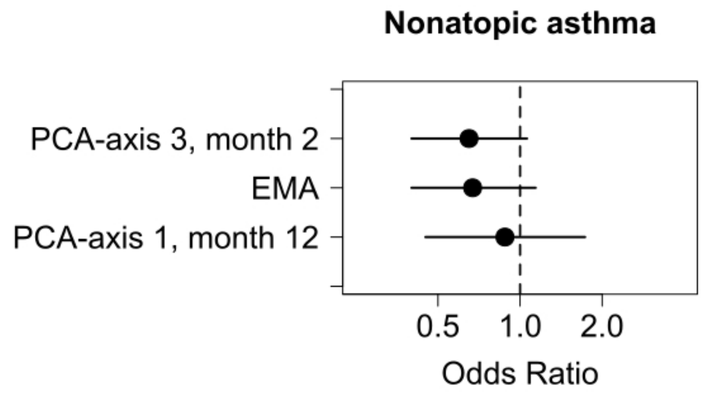
e

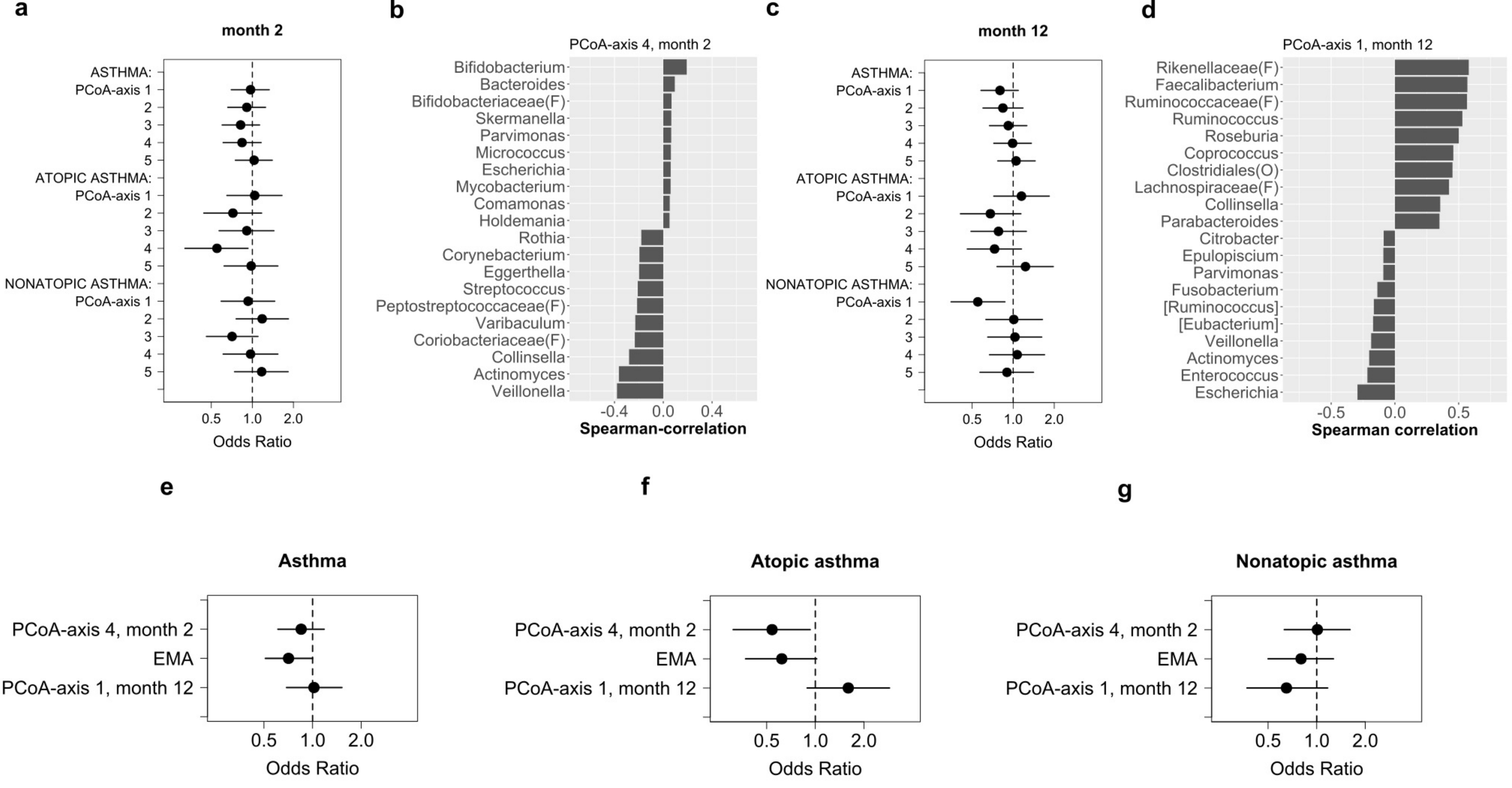

Asthma
PCA-axis 3, month 2
EMA
PCA-axis 1, month 12
0.5
1.0
2.0
Odds Ratio

f

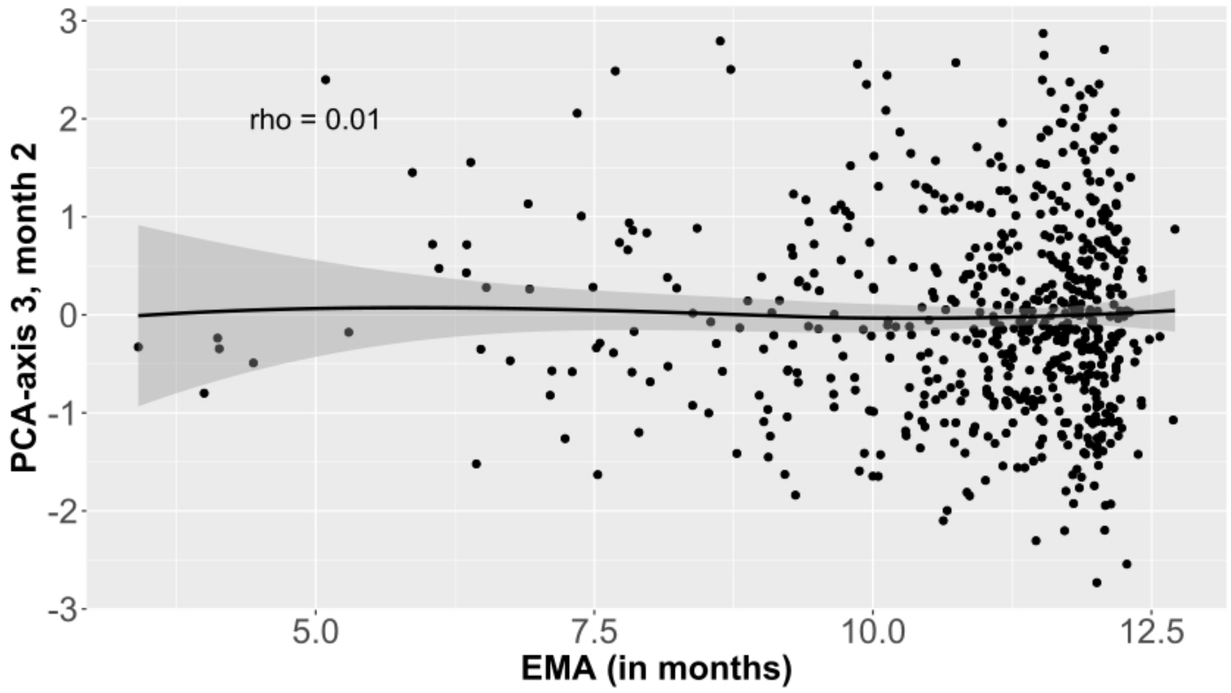

Atopic asthma
PCA-axis 3, month 2
EMA
PCA-axis 1, month 12
0.5
1.0
2.0
Odds Ratio

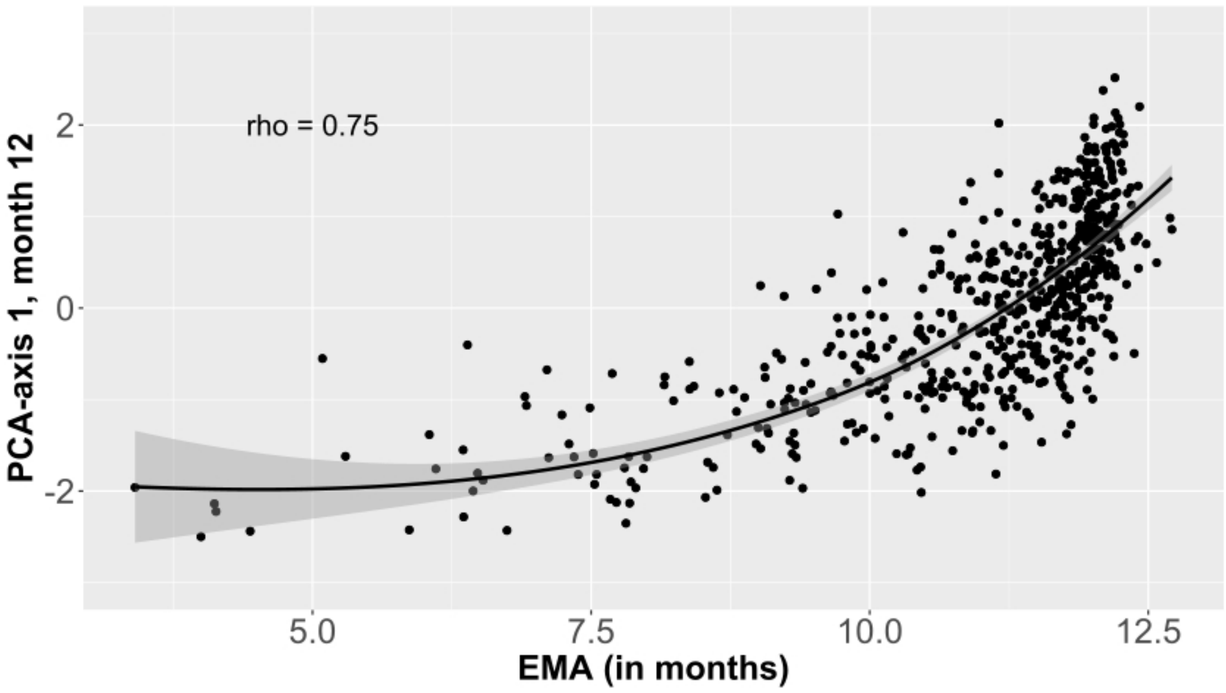
g


Nonatopic asthma
PCA-axis 3, month 2
EMA
PCA-axis 1, month 12
0.5
1.0
2.0
Odds Ratio

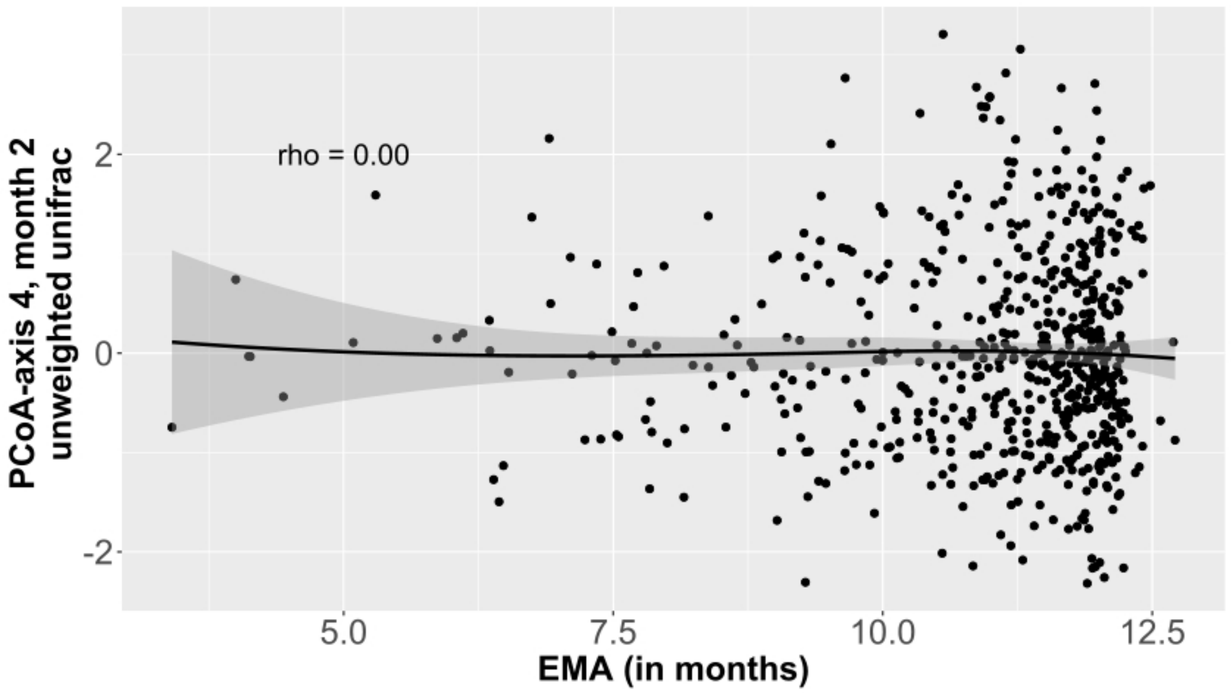

a
month 2
ASTHMA:
PCoA-axis 1
2
3
4
5
ATOPIC ASTHMA:
PCoA-axis 1
2
3
4
5
NONATOPIC ASTHMA:
PCoA-axis 1
2
3
4
5
0.5
1.0
2.0
Odds Ratio
b
PCoA-axis 4, month 2
Bifidobacterium
Bacteroides
Bifidobacteriaceae(F)
Skermanella
Parvimonas
Micrococcus
Escherichia
Mycobacterium
Comamonas
Holdemania
Rothia
Corynebacterium
Eggerthella
Streptococcus
Peptostreptococcaceae(F)
Varibaculum
Coriobacteriaceae(F)
Collinsella
Actinomyces
Veillonella
-0.4
0.0
0.4
Spearman-correlation
c
month 12
ASTHMA:
PCoA-axis 1
2
3
4
5
ATOPIC ASTHMA:
PCoA-axis 1
2
3
4
5
NONATOPIC ASTHMA:
PCoA-axis 1
2
3
4
5
0.5
1.0
2.0
Odds Ratio
d
PCoA-axis 1, month 12
Rikenellaceae(F)
Faecalibacterium
Ruminococcaceae(F)
Ruminococcus
Roseburia
Coprococcus
Clostridiales(O)
Lachnospiraceae(F)
Collinsella
Parabacteroides
Citrobacter
Epulopiscium
Parvimonas
Fusobacterium
[Ruminococcus]
[Eubacterium]
Veillonella
Actinomyces
Enterococcus
Escherichia
-0.5
0.0
0.5
Spearman correlation
e
Asthma
PCoA-axis 4, month 2
EMA
PCoA-axis 1, month 12
0.5
1.0
2.0
Odds Ratio
f
Atopic asthma
PCoA-axis 4, month 2
EMA
PCoA-axis 1, month 12
0.5
1.0
2.0
Odds Ratio
g
Nonatopic asthma
PCoA-axis 4, month 2
EMA
PCoA-axis 1, month 12
0.5
1.0
2.0
Odds Ratio

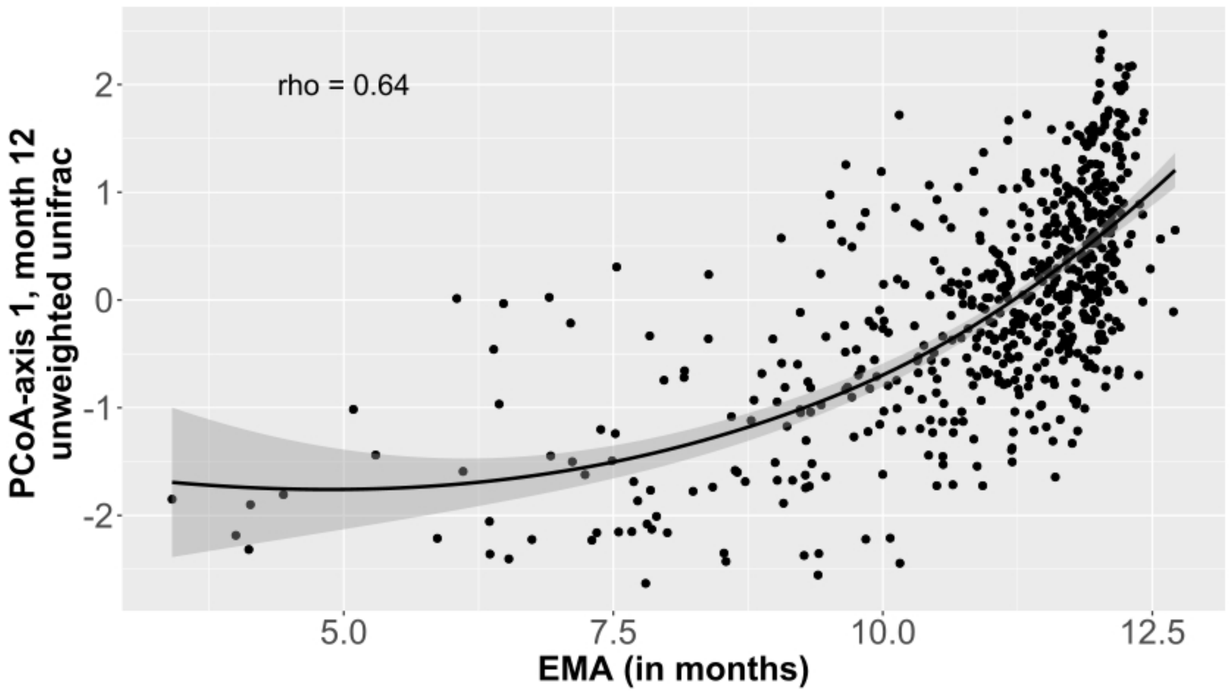
a


rho = 0.01
PCA-axis 3, month 2
EMA (in months)

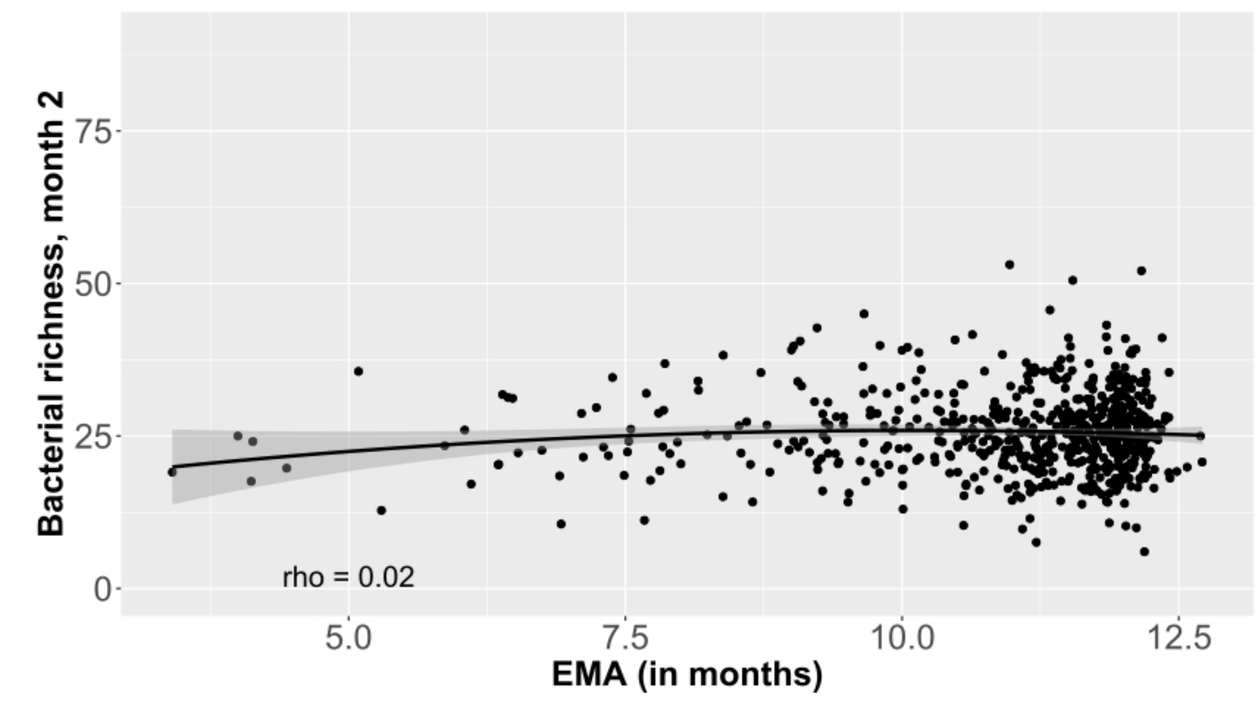

rho = 0.75
PCA-axis 1, month 12
EMA (in months)

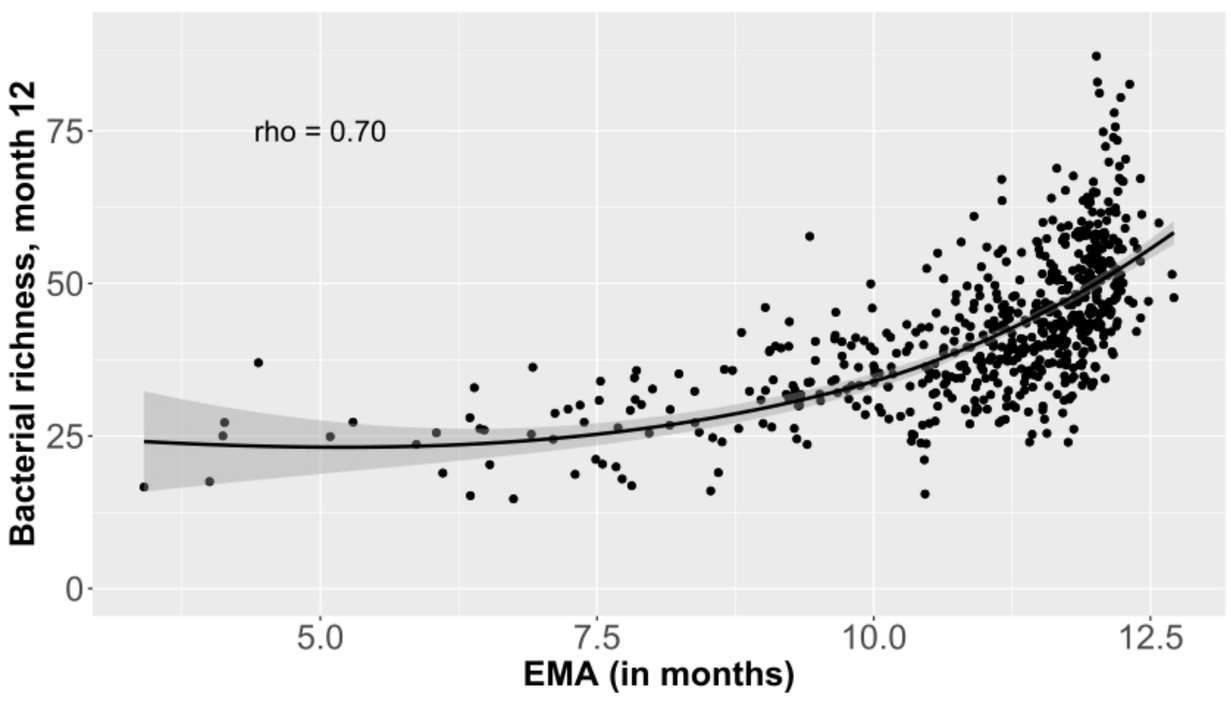
b


rho = 0.00
PCoA-axis 4, month 2
unweighted unifrac
EMA (in months)


rho = 0.64
PCoA-axis 1, month 12
unweighted unifrac
EMA (in months)

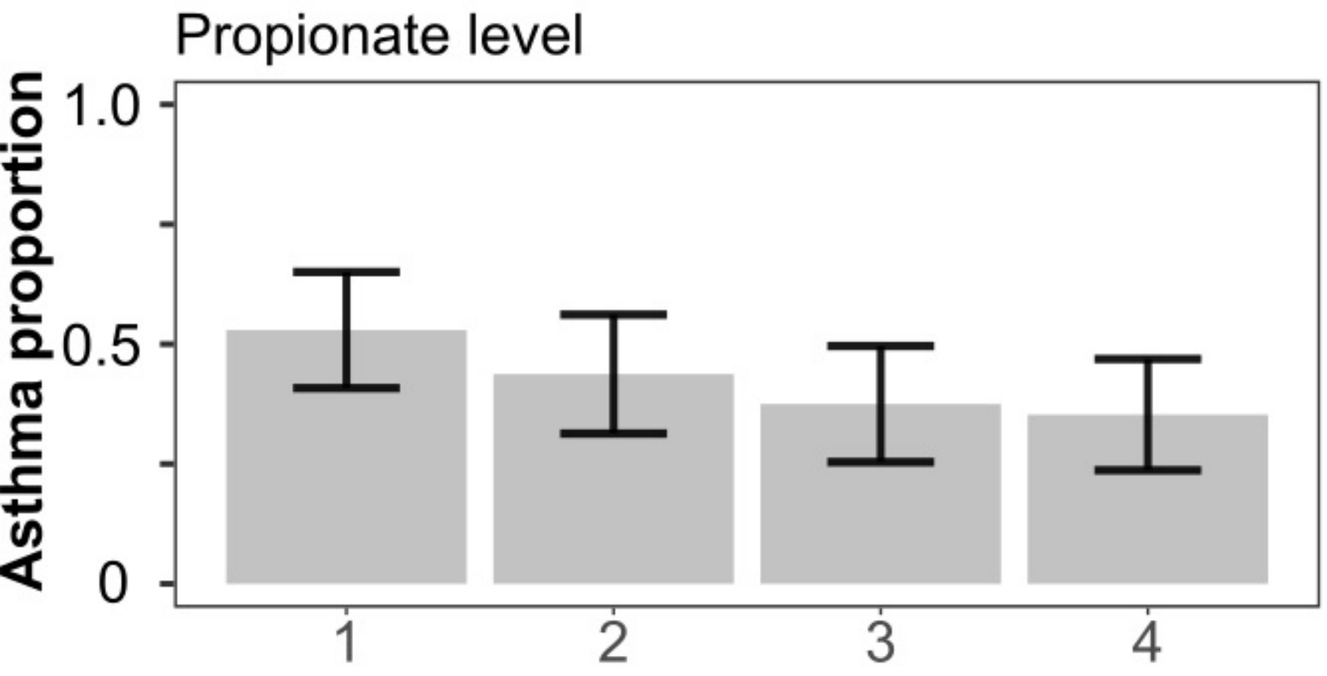
c

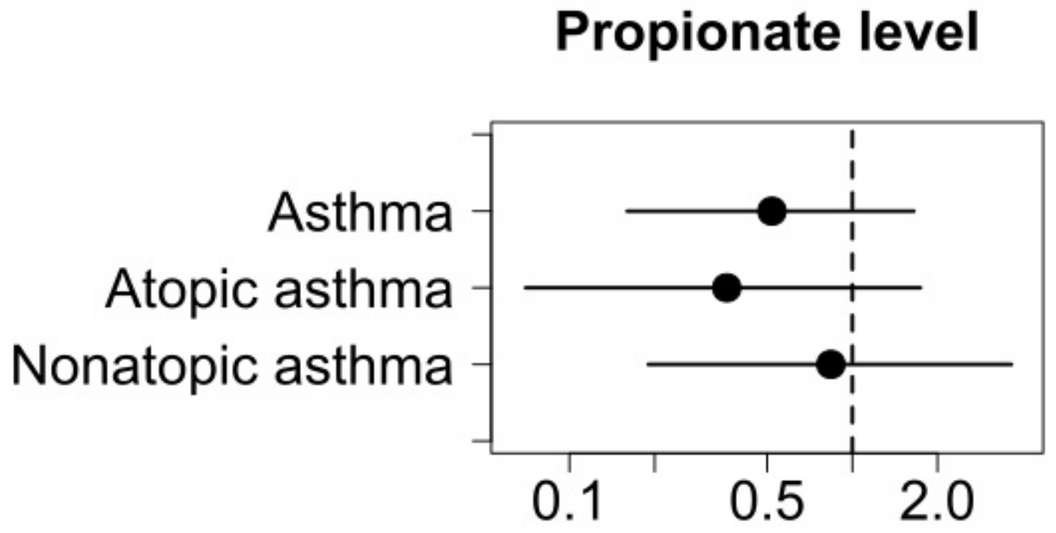

rho = 0.02
Bacterial richness, month 2
EMA (in months)

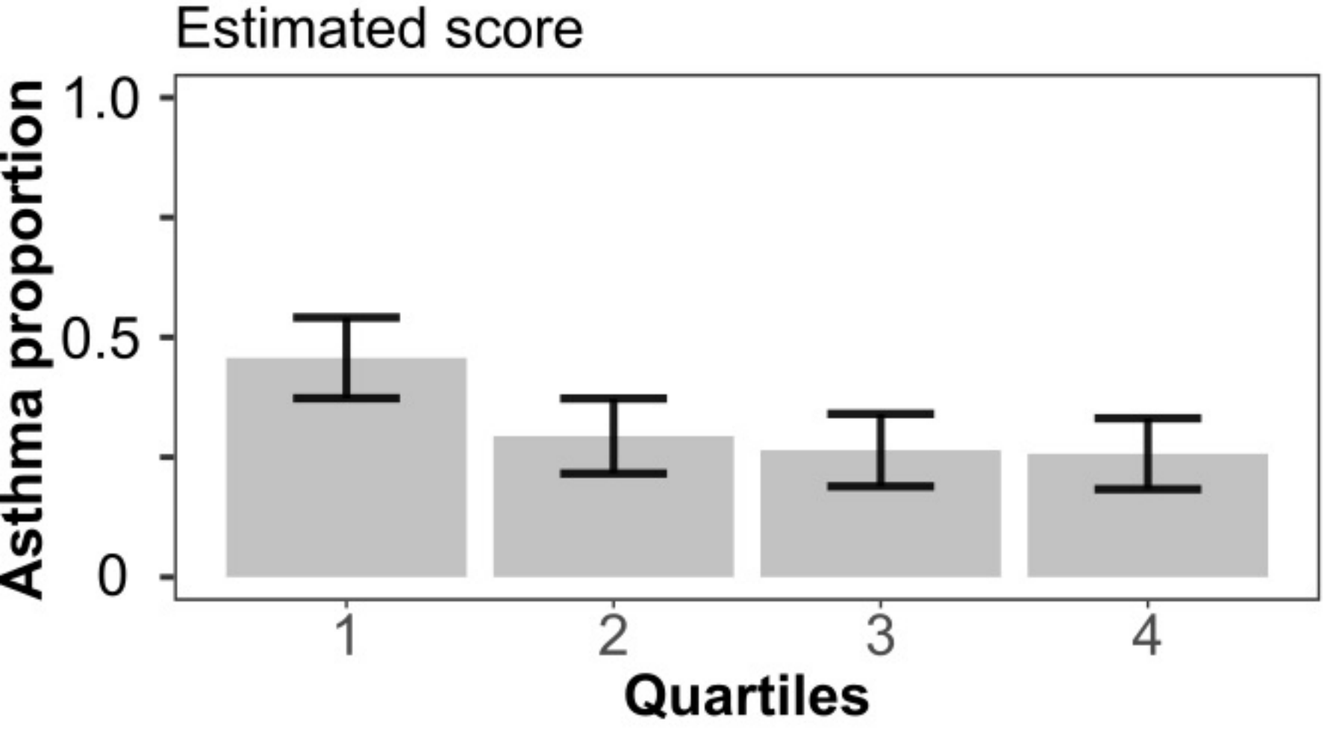

rho = 0.70
Bacterial richness, month 12
EMA (in months)

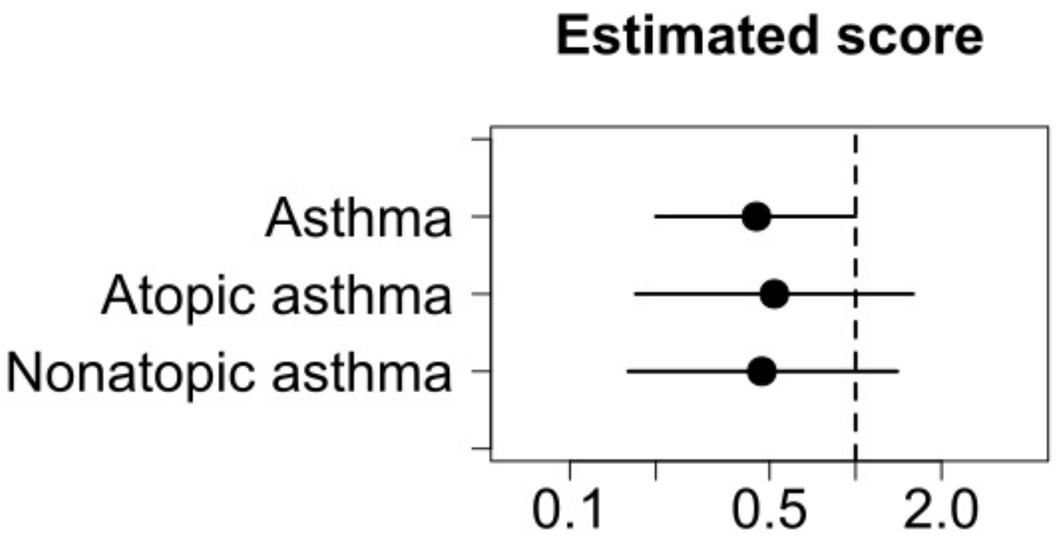

0.5
0.0
-0.5
-1.0
Beta (95% CI)
month 3
month 4
month 5
month 6
month 7
month 8
month 9
month 10
month 11
month 12
Duration of breastfeeding

a

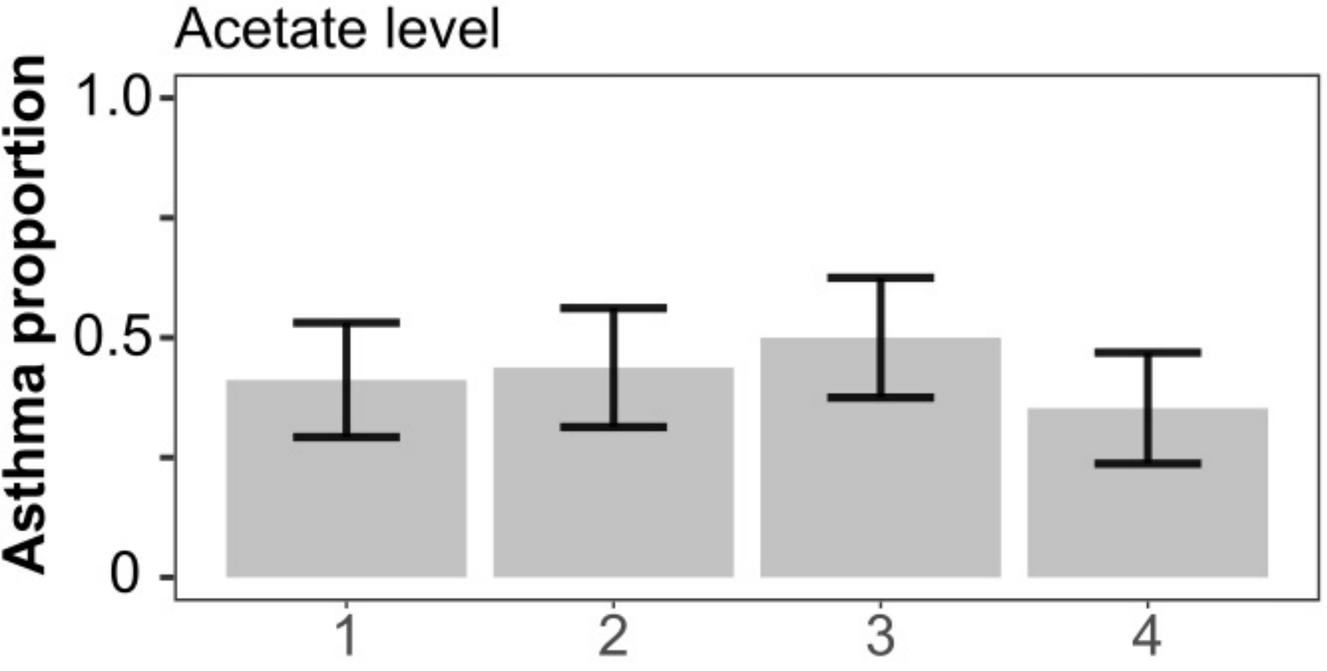

Propionate level
Asthma proportion
1.0
0.5
0
1
2
3
4

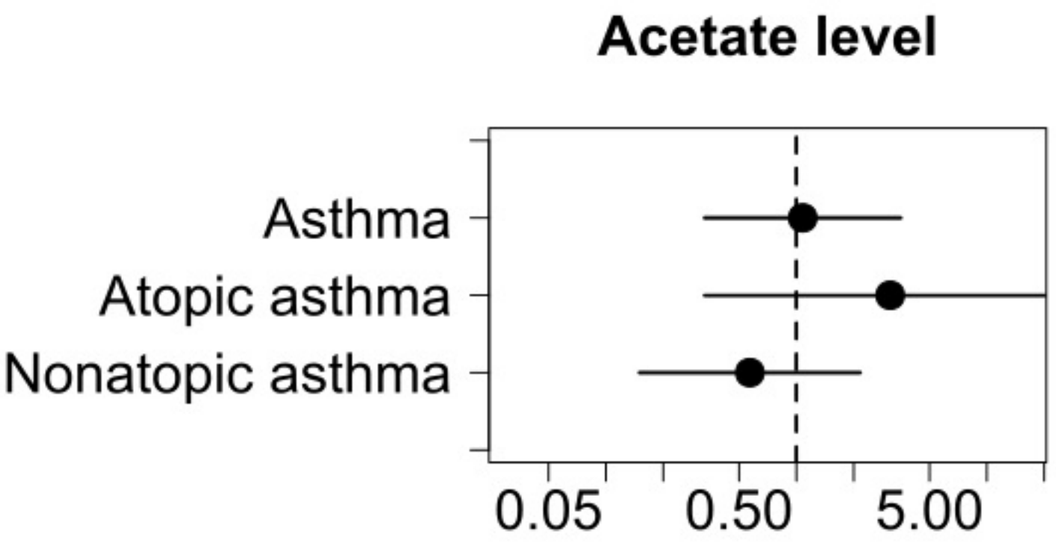

Propionate level
Asthma
Atopic asthma
Nonatopic asthma
0.1
0.5
2.0

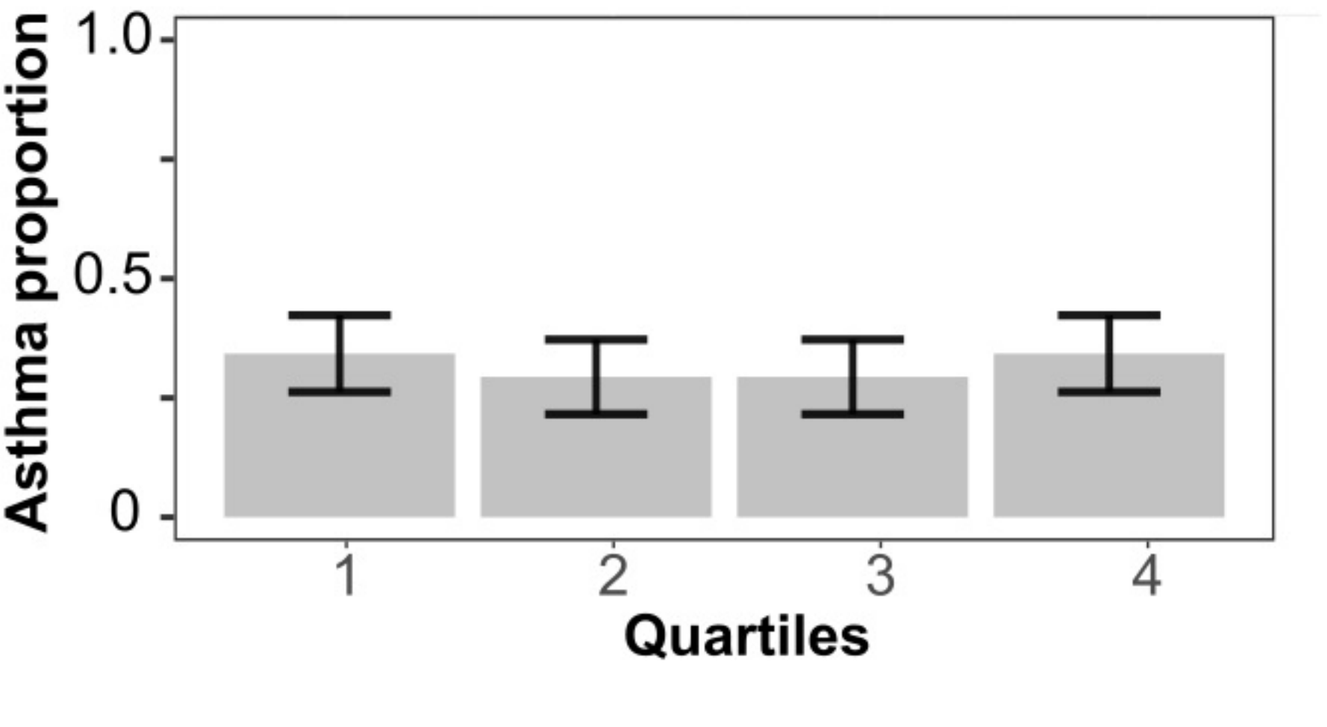

Estimated score
Asthma proportion
1.0
0.5
0
1
2
3
4
Quartiles

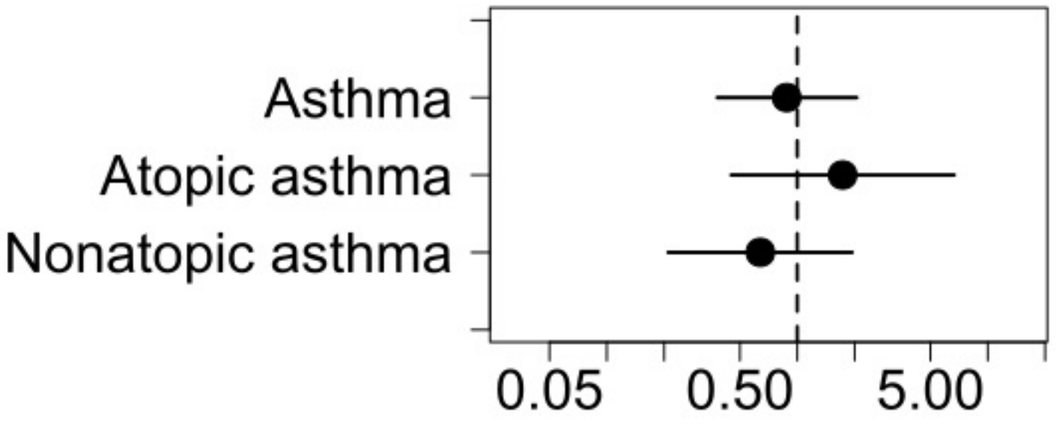

Estimated score
Asthma
Atopic asthma
Nonatopic asthma
0.1
0.5
2.0

b

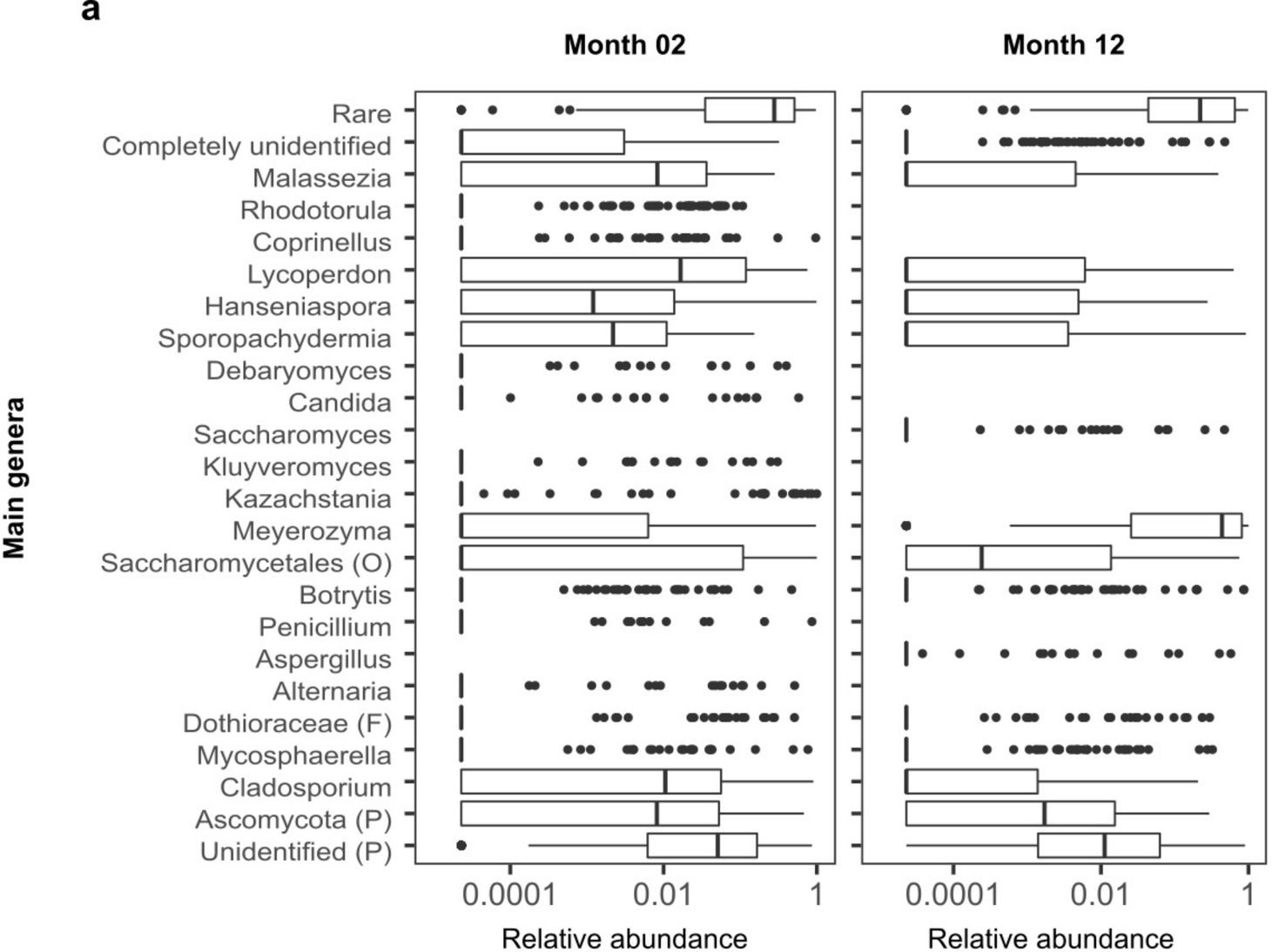

Acetate level
Asthma proportion
1.0
0.5
0
1
2
3
4


Acetate level
Asthma
Atopic asthma
Nonatopic asthma
0.05
0.50
5.00

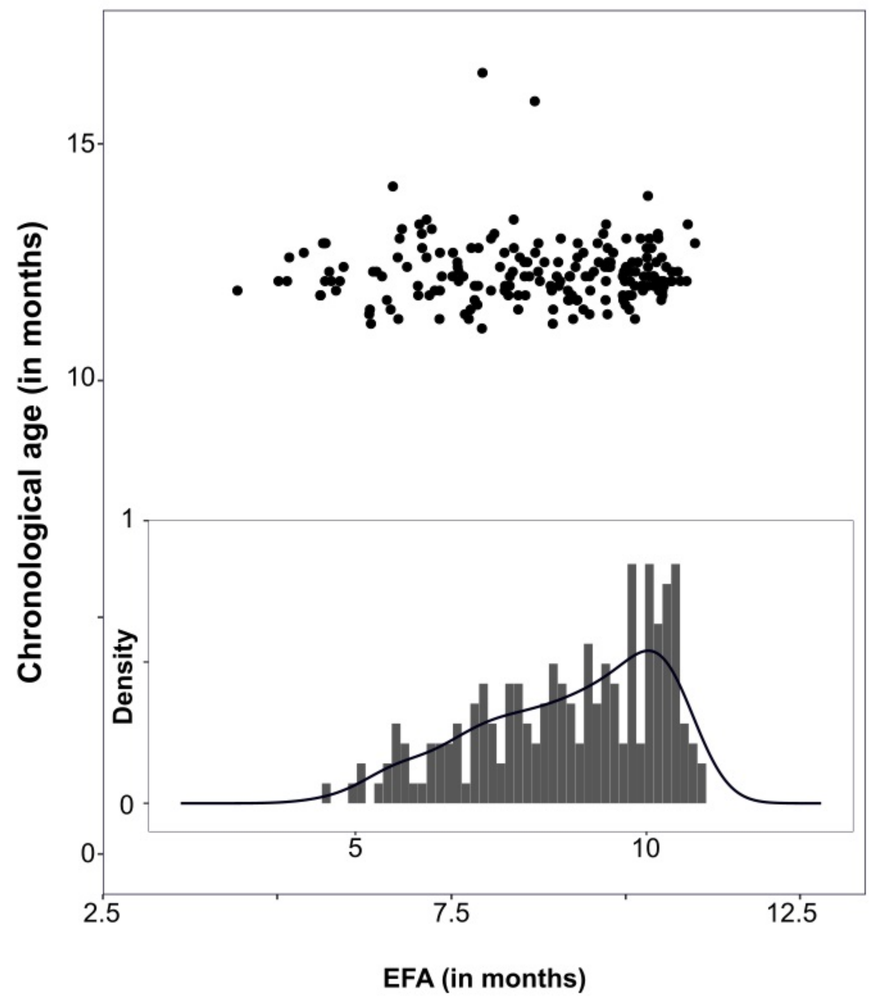

Acetate score
Asthma proportion
1.0
0.5
0
1
2
3
4
Quartiles


Estimated score
Asthma
Atopic asthma
Nonatopic asthma
0.05
0.50
5.00

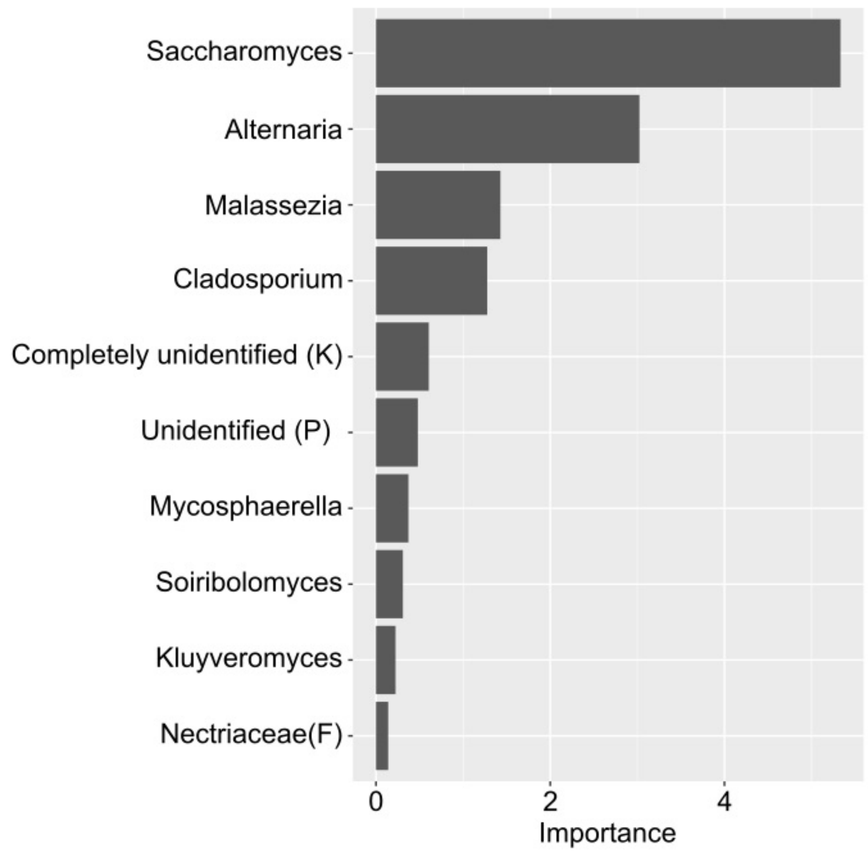
a


Month 02
Month 12
Rare
Completely unidentified
Malassezia
Rhodotorula
Coprinellus
Lycoperdon
Hanseniaspora
Sporopachydermia
Debaryomyces
Candida
Saccharomyces
Kluyveromyces
Kazachstania
Meyerozyma
Saccharomycetales (O)
Botrytis
Penicillium
Aspergillus
Alternaria
Dothioraceae (F)
Mycosphaerella
Cladosporium
Ascomycota (P)
Unidentified (P)
Main genera
0.0001
0.01
1
Relative abundance

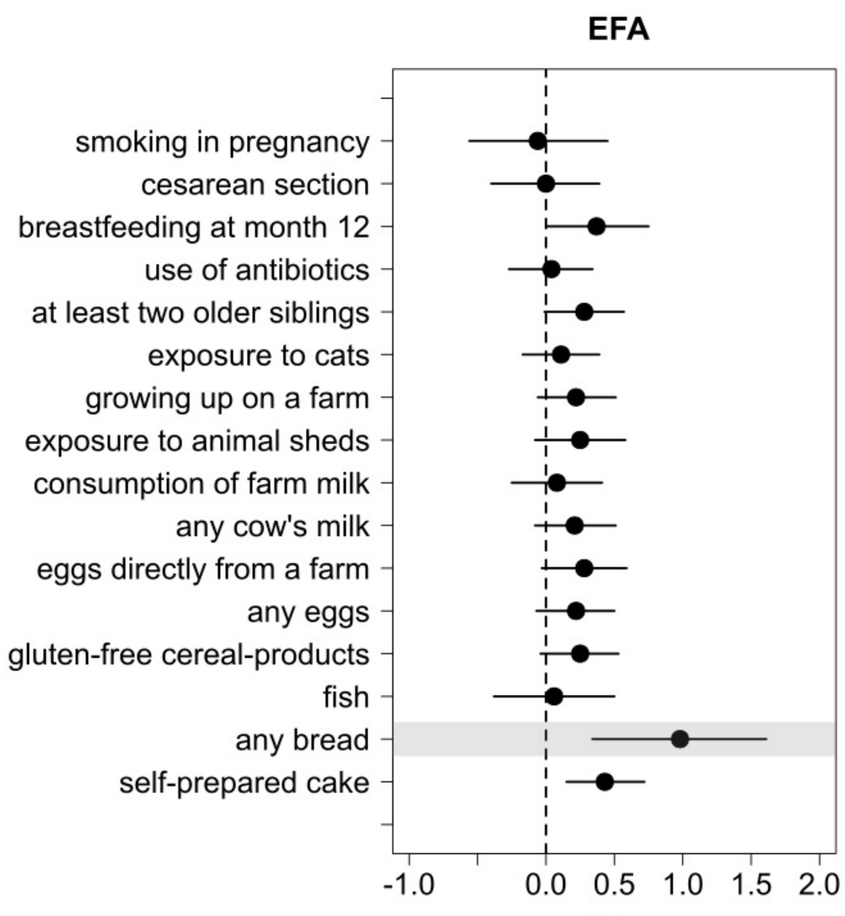
b


Chronological age (in months)
15
10
0
1
0
Density
5
10
2.5
7.5
12.5
EFA (in months)

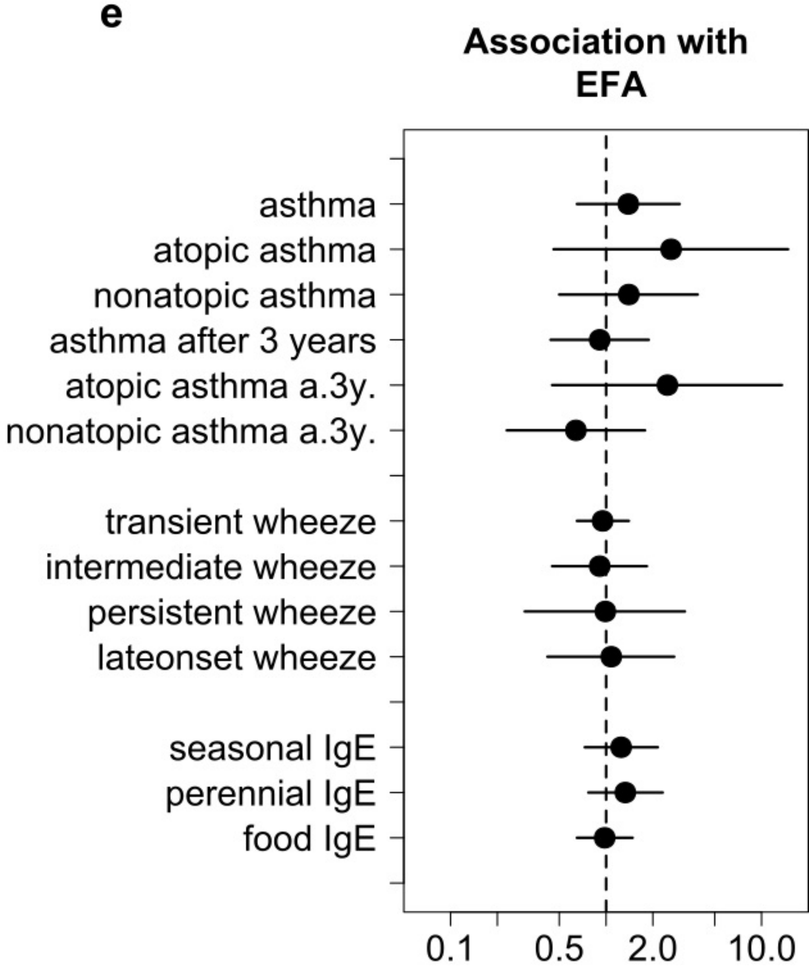
c

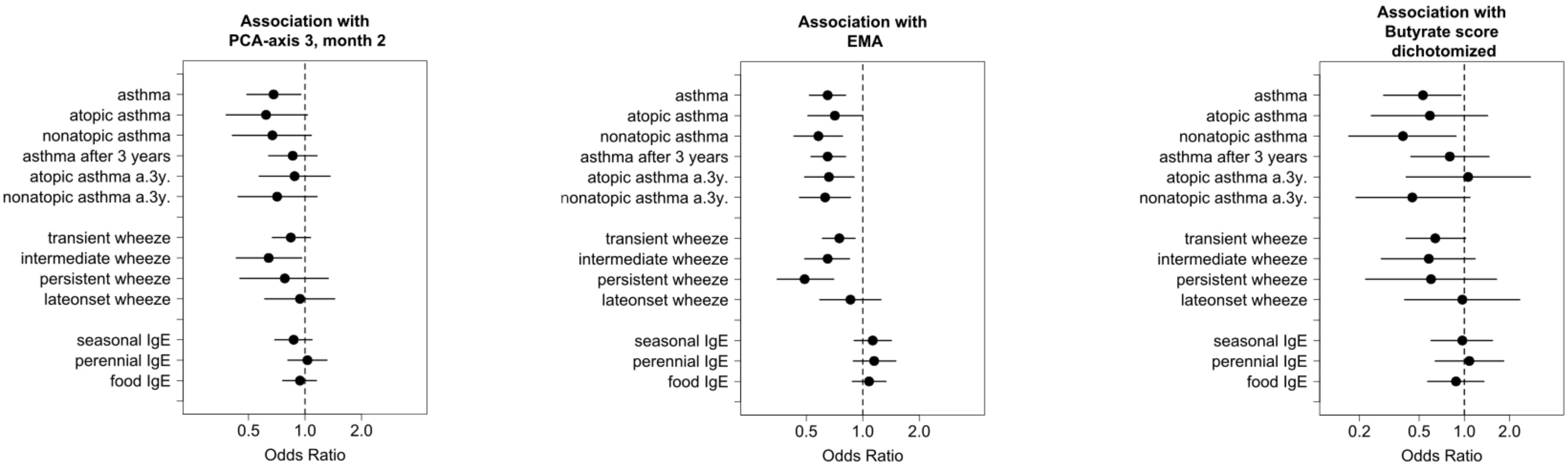

Saccharomyces
Alternaria
Malassezia
Cladosporium
Completely unidentified (K)
Unidentified (P)
Mycosphaerella
Soiribolomyces
Kluyveromyces
Nectriaceae(F)
0
2
4
Importance

d


EFA
smoking in pregnancy
cesarean section
breastfeeding at month 12
use of antibiotics
at least two older siblings
exposure to cats
growing up on a farm
exposure to animal sheds
consumption of farm milk
any cow's milk
eggs directly from a farm
any eggs
gluten-free cereal-products
fish
any bread
self-prepared cake
-1.0
0.0
0.5
1.0
1.5
2.0
Beta

e


Association with EFA
asthma
atopic asthma
nonatopic asthma
asthma after 3 years
atopic asthma a.3y.
nonatopic asthma a.3y.
transient wheeze
intermediate wheeze
persistent wheeze
lateonset wheeze
seasonal IgE
perennial IgE
food IgE
0.1
0.5
2.0
10.0
Odds Ratio


Association with
PCA-axis 3, month 2
asthma
atopic asthma
nonatopic asthma
asthma after 3 years
atopic asthma a.3y.
nonatopic asthma a.3y.
transient wheeze
intermediate wheeze
persistent wheeze
lateonset wheeze
seasonal IgE
perennial IgE
food IgE
0.5
1.0
2.0
Odds Ratio
Association with
EMA
asthma
atopic asthma
nonatopic asthma
asthma after 3 years
atopic asthma a.3y.
nonatopic asthma a.3y.
transient wheeze
intermediate wheeze
persistent wheeze
lateonset wheeze
seasonal IgE
perennial IgE
food IgE
0.5
1.0
2.0
Odds Ratio
Association with
Butyrate score
dichotomized
asthma
atopic asthma
nonatopic asthma
asthma after 3 years
atopic asthma a.3y.
nonatopic asthma a.3y.
transient wheeze
intermediate wheeze
persistent wheeze
lateonset wheeze
seasonal IgE
perennial IgE
food IgE
0.2
0.5
1.0
2.0
Odds Ratio

PCA-axis 3, month 2 p=0.004

EMA p=0.001

Butyrate score p=0.191

EFA (only in N=189) p=0.095

z-standardized value

5.0

2.5

0.0

-2.5

-5.0

A S G F A S G F A S G F A S G F

Center

A=Austria, S=Switzerland,G=Germany,F=Finland