

Published in final edited form as:

Nat Med. 2019 July 01; 25(7): 1089–1095. doi:10.1038/s41591-019-0469-4.

Farm-like indoor microbiota in non-farm homes protects children from asthma development

Pirkka V. Kirjavainen^{1,2}, Anne M. Karvonen¹, Rachel I. Adams³, Martin Täubel¹, Marjut Roponen⁴, Pauli Tuoresmäki^{1,5}, Georg Loss⁶, Balamuralikrishna Jayaprakash¹, Martin Depner⁷, Markus Johannes Ege^{8,9}, Harald Renz¹⁰, Petra Ina Pfefferle¹¹, Bianca Schaub⁸, Roger Lauener^{12,13,14,15}, Anne Hyvärinen¹, Rob Knight¹⁶, Dick J. J. Heederik¹⁷, Erika von Mutius^{#7,8,9}, Juha Pekkanen^{#1,5}

¹Environmental Health Unit, National Institute for Health and Welfare, Kuopio, Finland

²Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland

³California Department of Public Health, California, United States

⁴Department of Environmental and Biological Sciences, University of Eastern Finland, Kuopio, Finland

⁵Department of Public Health, University of Helsinki, Helsinki, Finland

⁶University of California, San Diego, School of Medicine, Department of Pediatrics, CA, USA

⁷ Institute for Asthma and Allergy Prevention, Helmholtz Zentrum Munich, German Research Center for Environmental Health, Neuherberg, Germany

⁸Dr. von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany

⁹Comprehensive Pneumology Center Munich, Member of the German Center for Lung Research, Munich, Germany

¹⁰Institute of Laboratory Medicine, Philipps University Marburg, Marburg, Germany

¹¹Comprehensive Biomaterial Bank Marburg, Philipps-University Marburg, Germany

¹²Children's Hospital of Eastern Switzerland, St. Gallen, Switzerland

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Corresponding Author: Pirkka V. Kirjavainen, pirkka.kirjavainen@thl.fi.

Author Contributions:

Conception: P.V.K. Concept refinement and study design: P.V.K., E.v.M., D.J.J.H. and J.P.; Statistical modeling: P.V.K.; Statistical analyses: P.V.K., A.M.K. and P.T.; Bioinformatics and related computational analyses: R.I.A., M.T., G.L., P.T., B.J.; Microbiological laboratory work (supervision and infrastructure): M.T., A.H., R.K.; Immunological laboratory work (supervision, coordination and infrastructure): M.R., H.R., P.I.P., B.S., R.L. P.V.K. wrote the manuscript with important contributions to intellectual content from A.M.K., R.I.A., M.T., M.R., M.D., M.J.E., B.S., R.K., D.J.J.H., E.v.M. and J.P.

Competing interests:

P.V.K., A.M.K., R.I.A., M.T., M.R., P.T., G.L., B.J., M.D., H.R., P.I.P., B.S., R.L., A.H., D.J.J.H. and J.P. do not have competing interests to disclose. M.J.E. and E.v.M. report patents EP2361632B1 and EP1964570B1 held by their institution LMU. E.v.M. reports recipient of funds from the European Commission for the conduct of the LUKAS (EFRAIM) and GABRIEL study and declares personal fees from Pharma Ventures, Peptinnovate Ltd., OM Pharma SA, European Commission/European Research Council Executive Agency, Tampereen Yliopisto, University of Turku, HAL Allergie GmbH, Ökosoziales Forum Oberösterreich and Mundipharma Deutschland GmbH & Co. KG; R.K. is a on the Scientific Advisory Board of Commense, Inc.

¹³University of Zurich, Zurich, Switzerland

¹⁴School of Medicine, University of St. Gallen, St. Gallen, Switzerland

¹⁵Christine Kühne-Center for Allergy Care and Education, CK-CARE, Davos, Switzerland

¹⁶Center for Microbiome Innovation and Departments of Pediatrics and Computer Science & Engineering, University of California San Diego, La Jolla, California, USA

¹⁷Division of Environmental Epidemiology, Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The Netherlands

These authors contributed equally to this work.

Abstract

Asthma prevalence has increased in epidemic proportions with urbanization, but growing up on traditional farms offers protection even today.¹ The asthma-protective effect in farms appears to be associated with rich home dust microbiota,^{2,3} which could be used to model a health-promoting indoor microbiome. Here we show by modelling differences in house dust microbiota composition between farm and non-farm homes of Finnish birth cohorts⁴ that in children who grow up in non-farm homes asthma risk decreases as the similarity of their home bacterial microbiota composition to that of farm homes increases. The protective microbiota had a low abundance of Streptococcaceae relative to outdoor-associated bacterial taxa. The protective effect was independent of richness and total bacterial load and was associated with reduced proinflammatory cytokine responses against bacterial cell wall components *ex vivo*. We were able to reproduce these findings in a study among rural German children² and showed that children living in German non-farm homes with an indoor microbiota more similar to Finnish farm homes have decreased asthma risk. The indoor dust microbiota composition appears as a definable, reproducible predictor of asthma risk and a potential modifiable target for asthma prevention.

From ancient times, humans have adapted to rich microbial exposures in early life. Changes in these exposures in modern urbanized environments may drive the epidemic increases in asthma and allergies.^{5,6} Many studies describe and identify protective microbial exposures but with heterogeneity in the specific microbial signals. Thus microbial exposures that could be exploited for preventive interventions remain unidentified. Here, we tested whether it is possible to circumvent this issue with an anchor-based method, drawing on the well-characterized asthma-protective effect of growing up on animal farms that appears associated with their particular indoor dust microbiota composition.^{2,3} If the indoor microbiota in farm homes causally protects from asthma, as suggested by experimental data,^{3,7,8} similar microbiota in non-farm homes should also have a protective effect despite the different surrounding environment and life-style.

We characterized the indoor microbiota from living-room floor dust collected from the homes of Finnish birth cohorts, LUKAS1 and LUKAS2,^{4,9} at the index child age of 2 months. At this age infants who crawl are constantly exposed to floor dust via the respiratory tract, skin and mouth.^{10,11} The characteristics of the farm home microbiota were defined within LUKAS1, which includes only rural homes, half of which are on farms with livestock. LUKAS2 is a random cohort of mostly suburban children.

The microbial composition in farm homes was clearly distinct from that in non-farm homes (Figure 1). The farm home dust microbiota was characterized by high bacterial richness and low-abundance cattle-associated microbes that were typically absent from the non-farm homes such as members of the Bacteroidales, Clostridiales and Lactobacillales orders, and rumen-associated archaea of the *Methanobrevibacter* genus (Figure 1, Extended Data Figure 1 and 2, Supplementary Table 1a and Supplementary Table 2).¹² Several taxa within the Actinomycetales order were also more abundant in the farm than non-farm homes. In contrast, non-farm homes had higher proportions of human-associated bacteria, including members of the Streptococcaceae family and *Staphylococcus* genus (Extended Data Figure 2, Supplementary Table 1b). Several differences were also seen in the relative abundance of specific fungal taxa, but fungal richness did not differ significantly between farm and rural non-farm homes (Figure 1, Extended Data Figures 1 and 3, Supplementary Table 3).

We then modelled the farm home microbiota-like community composition in LUKAS1. Separate models were built for the farm-like bacterial/archaeal presence-absence, bacterial/archaeal relative abundance, fungal presence-absence and fungal relative abundance. The coefficients from these models were then applied to data of LUKAS2 non-farm homes.

The bacterial/archaeal presence-absence patterns were very different between farm and non-farm homes (Figure 1). Accordingly, the probability of farm-like presence/absence pattern in a non-farm home was very low, and was not associated with asthma risk among the LUKAS2 non-farm children (Supplementary Table 4). Farm-like fungal composition also had no association with asthma risk (Figure 2, Supplementary Table 4). In contrast, farm-like relative abundance of bacteria/archaea at age 2 months was associated with decreased risk of asthma development by 6 years of age (Figure 2). The association reached statistical significance also with active asthma at age 6 years when analyzed in a pooled sample of LUKAS1 and LUKAS2 (referred to as LUKAS from here onwards) non-farm children (Supplementary Table 4). The association between the farm-like relative abundance of bacteria/archaea and asthma was similar between children living or not living on farms, as indicated by non-significant interaction term ($p > 0.6$) and nearly equal odds ratio estimates in stratified analysis (Supplementary Table 4). However, in farms asthma was rare ($N=10$) and probably due to low statistical power the protective effect was not significant (except with dichotomous probability variable; $p=0.02$).

We named the probability variable based on the bacterial/archaeal relative abundance data FaRMI (Farm home Resembling Microbiota Index). Notable feature of FaRMI was its moderate classification accuracy in the training set (i.e. LUKAS1) (Extended Data Figure 4), which allows detection of farm-like features also in non-farm homes. From LUKAS farm homes 75.9% (88/116) and non-farm homes 32.7% (91/278) had more farm than rural non-farm like microbiota based on $\text{FaRMI} \geq 0.5$. The association between FaRMI and asthma in non-farm children was independent of markers of microbial exposure previously linked with reduced risk of asthma, including bacterial richness and total bacterial and endotoxin load (Supplementary Table 5).^{2,13,14,7} This suggests that these general markers may be only proxies of more specific microbial composition such as described by FaRMI.¹⁵ A minor part (17%) of the protective effect associated with FaRMI seemed to be explained by muramic acid concentration in dust which could indicate the importance of bacterial cell

wall structures. Muramic acid is a cell wall component characteristic (but not limited) to Gram-positive bacteria and its indoor levels were recently associated with asthma protection also in adults.¹⁶

Variables associated with increased FaRMI in non-farm homes included walking inside with shoes worn outdoors, which may reflect transfer of soil; presence of two or more older siblings; elevated indoor moisture; and increased age of the house (Supplementary Table 6). However, the asthma-protective association of FaRMI in non-farm homes was independent of these determinants, which indicates the importance of the microbial exposure over the environmental and lifestyle factors (Supplementary Table 5). Source tracking analysis confirmed that FaRMI was positively correlated with bacterial/archaeal OTUs of soil origin (Extended Data Figure 2). The beneficial influence of soil microbe exposure on asthma risk is supported by a recent study in mice.¹⁷

To test the reproducibility of the association between farm-like bacterial/archaeal relative abundance and asthma, we first tested the reproducibility using alternative, data reduction independent, methodological approach. For this purpose we trained a random decision forest in LUKAS1 and applied it to LUKAS2 non-farmers. Compared to our original approach, this is more specific approach, and does not take into account phylogenetic similarity and thus is, by default, less likely to detect farm-like features in non-farm homes. Nonetheless, the analysis supported the concept as asthma protective trend ($p < 0.1$) of farm-like features in non-farm home microbiota were also noted by this approach (Supplementary tables 7 and 8).

We then tested the reproducibility of the association between FaRMI and asthma in another study population. For this purpose we applied our approach to data from the cross-sectional, German study among rural children; GABRIELA.¹⁸ In GABRIELA, home indoor microbiota was characterized from mattress dust ($N=1031$) and animal shed dust microbiota from a subsample of farms ($N=50$). The microbial community membership structure in the GABRIELA mattress dust samples was clearly distinct from the LUKAS floor dust samples, as would be expected due to the different sample types,¹⁹ geographical location,^{20,21} and other study-specific differences.²⁰ However, the influence of farming on the home indoor microbiota was similar, characterized in both studies by clustering closer to the GABRIELA animal shed dust samples (Figure 3, Supplementary Table 9).

In order to replicate our results independently of the LUKAS1 beta-diversity matrix, we built a linear model for FaRMI with relative abundance of bacterial/archaeal taxa in LUKAS1 and applied it to GABRIELA data (Supplementary Table 10a). This reproduced the asthma-protective effect of FaRMI in GABRIELA non-farm children (Figure 3, Supplementary Table 11). Based on $\text{FaRMI} \geq 0.5$, the microbiota in GABRIELA homes resembled more that in LUKAS1 farm than non-farm homes in 60.4% (241/399) of farm homes and in 36.8% (123/334) or 20.5% (61/298) of non-farm homes depending whether the children were regularly exposed to farms or not, respectively. This replication demonstrates that FaRMI as a model of the asthma-protective indoor microbiota composition, is not limited to a single geographical location, population or indoor dust sample type.

We then used the FaRMI-approach to model GABRIELA farmhouse-like microbiota (Supplementary Table 10b). The GABRIELA farm-house-like relative abundance pattern (FaRMI_{GABRIELA}) tended to be associated with lower asthma risk in GABRIELA and LUKAS non-farm children (Figure 3, Supplementary Table 11), which indicates that the FaRMI-approach is not limited to LUKAS1 as the anchor population. Common to both LUKAS1- and GABRIELA-based models was a negative association of FaRMI with the abundance of taxa in the Streptococcaceae family and positive associations with Sphingobacteriia and Alphaproteobacteria classes and Cyanobacteria phylum, which together explained over 60% of the total variance in both models (Figure 3, Extended Data Figure 5).

Early-life microbial exposures,¹⁵ and living on the farm,¹ may protect from atopic sensitization, which is a risk factor for asthma. The association between FaRMI and asthma was, however, independent of atopic sensitization among LUKAS non-farm children (Supplementary Tables 5 and 12). Similar observation has been previously made in relation to microbial richness in farm homes.² FaRMI at 2 months had some but no consistent association with total cytokine production capacity of blood leukocytes at 1 and 6 years, as determined by *ex vivo* mitogenic stimulation (Table 1, Extended Data Figure 6). Instead, high FaRMI was associated with suppression of the bacterial cell wall component induced secretion of type 1 immunity associated cytokines including interferon- γ , interleukin (IL)-1 β , IL-6 and IL-12. This indicates that the farm home –like microbiota may improve tolerance to microbial exposures; similarly as high early-life indoor endotoxin exposure may lead into endotoxin tolerance.¹⁴ Based on animal models, endotoxin tolerance may inhibit also allergen-induced airway inflammation.^{22,23} This hypothesis also parallels the findings on Amish and Hutterite farm children, known for low and high asthma prevalence, respectively. The Amish children appear to have higher proportion of immunosuppressive monocytes than the Hutterite children, and the dust from the Amish but not Hutterite homes inhibited ovalbumin-induced broncho-alveolar eosinophilia in mice.⁷ A higher proportion of immunosuppressive monocytes could also explain the cytokine pattern associated with high FaRMI but that could not be assessed with data available. Instead, we found in exploratory analyses a strong positive correlation between FaRMI at 2 months and immunoglobulin-like transcript (ILT) 4 expression²⁴ on peripheral blood plasmacytoid dendritic cells of non-asthmatic LUKAS1 non-farm children (pDC, $\rho=0.64$; $p=0.0015$; $n=26$) at 6 years. This correlation did not exist within LUKAS1 children who were diagnosed with asthma by 6 years ($\rho=-0.09$; $p=0.80$; $n=15$, Extended Data Figure 7). No correlations were seen between FaRMI and ILT4 expression on myeloid DC (mDC) or ILT3 expression on mDC or pDC (data not shown). ILTs are known inhibitory receptors and markers of tolerogenic DCs.²⁵ pDCs are potential gate keepers that could direct immune response against harmless environmental antigens away from pro-asthmatic inflammatory responses such as airway eosinophilia.²⁶

The route of exposure and mechanisms through which bacteria might mediate the immune suppressive effects and asthma protection require further research. We hypothesize that a key feature is the high relative abundance of environmental bacteria (including those of animal origin) relative to human associated bacteria. The human associated bacteria may be more likely to colonize, invade and infect us and thus release more proinflammatory

danger signals than the environmental bacteria.²⁷ This potential was visible in the predicted abundance of genes negatively associated with FaRMI (Supplementary Table 13).^{28–35} In previous studies *Streptococcus*, *Moraxella* and *Haemophilus* genera that include common opportunistic respiratory pathogens have been associated with increased risk for asthma when abundant in home dust or colonizing the airways early in life.^{7,36,37} These genera were also more abundant in the non-farm than farm homes, but did not seem to contribute considerably to FaRMI (Supplementary Table 10a). Unlike with asthma-predisposing taxa, there seems to be very little overlap between the different studies in low-asthma-risk associated indoor taxa.^{2,7,15,37} They may still share common beneficial properties, e.g. similar cell wall structures or functions, but it is also possible that the beneficial associations merely reflect lack of predisposing features and the relative abundance of these taxa compared to the potentially predisposing microbes. The lack of consistency could also indicate that the specific microbes are only proxies for less specific microbiota characteristics, such as richness or total load, or of environmental determinants, but this is not supported by our data.

The major strengths of our study are the prospective design, low dropout rate, long follow-up, adjustments for confounders, sample size, data on immunological responsiveness and replicability. Another advantage was the use of the anchor-based approach, which allowed us to study the health effects associated with microbiota typical to a protective environment (farming) independently of that environment. A limitation is that an observational study cannot establish causality, which needs to be established by an interventional study. Further studies, such as metagenomic, metabolomic and cell wall chemistry assessments, are needed to define the key features of asthma-protective microbial exposure. Comparing the relevance of fungal and bacterial exposures was limited by the modeling-associated differences and mechanistic studies to systemic immunity, which may differ from that in airways.

In conclusion, while the asthma-protective effect of farming is intriguing, it has little practical relevance unless the protective effect can be functionally transferred to non-farming environments. We have taken the ‘farm effect’ outside of farms by showing that compositionally similar indoor dust bacterial/archaeal microbiota is also protective in non-farm environments. This is in agreement with our hypothesis and consistent with (but not proof of) the possibility that bacteria could be causal mediators of the asthma-protective farm effect. Our results warrant translational studies to confirm the causal relationship through indoor microbial exposure-modifying intervention that may also form a novel strategy for primary asthma prevention. With our robust and straightforward approach for defining farm-like microbiota, it is now possible to evaluate the asthma-protective potential associated with a given indoor microbial community, select suitable donor microbiota for interventions, and monitor the changes induced in the recipient home microbiota.

Online Methods

The birth cohorts

In LUKAS1 (N=214) equal numbers of pregnant mothers living in farms with livestock and mother’s living in rural areas but not in farms were recruited in the major local hospitals in eastern and middle Finland (Kuopio, Iisalmi, Jyväskylä and Joensuu) between

September 2002–May 2004. The inclusion criteria were maternal age ≥ 18 years, singleton pregnancy, native language Finnish, no plans to move from the study area, expected delivery in one of the study hospitals, siblings of the study child not participating in the study, parturition at ≥ 37 weeks of gestation, no congenital abnormalities in newborn and successful cord blood sampling. In LUKAS2 all pregnant women with estimated delivery at Kuopio University Hospital between May 2004 and May 2005 were invited to join the study without selection by occupation or area of living. Mothers living in apartments were excluded to maintain housing conditions comparable with LUKAS1. In the current study LUKAS2 children living on farms were excluded ($n=11$). Written informed consent was acquired from all LUKAS mothers. Ethical permission was granted by the Research Ethics Committee, Hospital District of Northern Savo. The replication stage included 1031 children from the cross-sectional, Phase II GABRIELA study² with data on child home indoor microbiome. The children in the study were 6 to 12 years old (median 9) and they lived in the rural regions around Munich or Ulm in a farm home ($n=399$) or in non-farm home with ($n=334$) or without ($n=298$) regular exposure to farms.

Sample collection and processing

In the LUKAS study, the living room floor dust samples were collected at index child age of 2 months. The sample was collected by the occupants into a nylon sampling sock by vacuuming an area of 1 m² from a rug for two minutes or in the absence of a rug, an area of 4 m² from a smooth floor for two minutes. The living room was defined as the room where the family spent most time after dinner. The dust samples were homogenized by sieving through a sterile strainer, dried in a desiccator and stored at -20°C until DNA extraction. Genomic DNA was extracted from 20mg of dust using bead beating method and chemagic DNA plant kit (Perkin Elmer) on the KingFisher DNA extraction robot.

In the GABRIELA study, mattress dust was collected by the parents of the participating children using a standardized dust collection protocol.³⁹ The whole area of the child's mattress was vacuumed for a period of 2 min using a dust sampling nylon sock (Allied Filter Fabrics Pty Ltd, Australia) attached to the vacuum cleaner hose. Stable dust samples were collected with a brush from horizontal areas above 1.5 m. The dust samples were stored at -80°C after arrival at the study center. DNA extraction was performed using MoBio PowerSoil Extraction Kit (MO BIO Laboratories, Carlsbad, CA, USA) according to the Earth Microbiome Project Protocols (<http://www.earthmicrobiome.org/emp-standard-protocols/>).

Sequencing and bioinformatics

In LUKAS the bacterial/archaeal 16S rRNA gene V4 region was amplified using 515F/806R primers⁴⁰ and fungal ITS region by ITS1F/ITS2 primers.⁴¹ These DNA amplicons were sequenced as 300 base pair paired-end reads with Illumina MiSeq V3 chemistry. The amplifications and sequencing were performed by commercial provider (LGC Genomics GmbH, Berlin, Germany). In GABRIELA, the Earth Microbiome Project Protocols were used to create bacterial/archaeal amplicon libraries using identical primers (515F/806R) to LUKAS, followed by sequencing on the Illumina HiSeq platform. At the discovery stage, sequence reads were merged with FLASH,⁴² while QIIME⁴³ was used for quality filtering,

exclusion of chimeric sequences and further processing. Sequences were clustered into operational taxonomic units (OTUs) at 97% similarity using the open-reference protocol⁴⁴ against the 16S rRNA gene database, greengenes, or the ITS database UNITE. OTUs representing less than 0.001% of the total sequences (minimum count of 83 and 93 sequences for bacteria and fungi, respectively) were excluded. Chloroplast (n=93) and mitochondrial (n=23) sequences were removed from the bacteria OTU table. All samples with less than 2150 sequences were not included in the analyses. Rarefaction curves are presented in the Supplementary figure 1

In order to obtain compatible data for the replication stage, the reads from GABRIELA and LUKAS sequencing were processed together. The replication data was prepared using the Deblur software, based on sub-operational-taxonomic-units (sOTU). This achieves single-nucleotide resolution which supports the combination of different datasets.⁴⁵ Only forward R1 reads of Illumina paired end was used in this approach and R2 reverse reads were discarded in order to avoid noise from reverse reads during deblurring. Then low-quality reads, and artificial sequences such as primers and adapters, were removed. Each forward read file was trimmed to the lowest available length in any of the data sets: 115bp (from original length of about 260bp). After that, the sequences were clustered by the Deblur algorithm and the sOTU table was obtained. The minimum cutoff for a single sOTU was set as 50 reads, and sOTUs below that cutoff were filtered out. Taxonomy was assigned based on Ribosomal Database Project.

Asthma and atopic sensitization

Data on asthma outcome was obtained from parent-reported questionnaires. Doctor-diagnosed asthma at least once or asthmatic bronchitis more than once by the age of 6 years was termed as ‘asthma ever’, and asthma with medication or wheezing at the age of 6 years as ‘active asthma’. Immunological phenotype was assessed from venous blood samples collected at 1 and 6 years of age. Atopic sensitization was evaluated at 6 years by specific immunoglobulin E (sIgE) measurements to 13 inhaled (dust mites: *Dermatophagoides pteronyssinus* and *D. farinae*; pollens: alder, birch, European hazel, grass pollen mixture, rye, mugwort and plantain; and cat, horse and dog dander; as well as the mold *Alternaria alternata*) and 6 food allergens (hen’s egg, cow’s milk, peanut, hazelnut, carrot and wheat, Mediwiss Analytic, Moers, Germany)⁴⁶ with cut-off ≥ 3.5 kU/L.

Immune responsiveness

Cytokine responsiveness was evaluated from cultured whole blood collected at 1 and 6 years as earlier described.⁴⁷ Overall cytokine production capacity profile of leukocytes was assessed from whole blood cultured 24h with phorbol 12-myristate 13-acetate and ionomycin (PI, 1 μ g/mL) and responsiveness to bacterial cell wall components from cultures stimulated 24h with lipopolysaccharide (LPS, 0.1 μ g/mL) or peptidoglycan (PPG, 10 μ g/ml, at 6 years only). At 6 years, also non-stimulated cultures were assessed to analyse spontaneous cytokine secretion. All stimulants were from Sigma, Deisenhofen, Germany. The concentration of interleukins (IL) 1b, 4, 5, 6, 10, 12p70, 13, 17A and interferon- γ were determined from the cell culture supernatants with multiplexed cytometric bead arrays (BD human CBAflex) with FACSArray bioanalyzer system (BD Biosciences).⁴⁸ The

concentrations were standardized by leukocyte counts. Non-detects were set to the detection limit standardized by leukocyte count. The CRP values at the six years were measured by SYNCHRON® System(s) (Beckman Coulter Inc., Fullerton, CA, USA).

Expression of the tolerance associated surface receptors ILT3 and ILT4 on circulating mDCs and pDCs was analyzed by flow cytometry from cryopreserved peripheral blood mononuclear cell (PBMC) samples of a subpopulation of LUKAS1 children as described earlier.⁴⁹ The mean viability of thawed cells was 93.9%. The antibodies used for staining are described in Supplementary Table 14 and the gating strategy in Supplementary Figure 2. The LUKAS1 subpopulation data represented 1:2 asthma – non-asthma design where all children with asthma ever by 6 years were included if their PBMC sample and specific IgE data was available at age 6. A double number of non-asthmatic children with one half living on a farm were randomly selected of children with available PBMC sample and IgE data at age 6 with priority on children with dendritic cell data at age 4.5 years. Only data from children not living on a farm was included in the analyses for this study.

Microbial diversity

Measures of α -diversity in LUKAS1 and 2 samples, richness (defined as OTUs observed) and Shannon entropy, were calculated with QIIME from 2150 resampled sequences and presented as a mean of ten iterations. Phylogenetically-informed variation between pairs of samples (β -diversity) in the bacterial/archaeal community was evaluated with Generalized UniFrac distances, calculated on using GUniFrac R-package⁵⁰ with midpoint rooted tree and using $\alpha=0$ for abundance unweighted and $\alpha=1$ for abundance weighted β -diversity. Due to the lack of conserved sequences within ITS that would allow reliably establishing phylogenetic relatedness between taxa, we did not calculate phylogenetically informed beta-diversity within fungal microbiota. Instead, the beta-diversity within fungal community was evaluated by a Bray-Curtis distance matrix calculated on binary (presence/absence) or relative abundance data. Principal coordinate analyses were carried out using R pcoa function in ape-package with default settings.⁵¹ Statistical significance of group-wise differences in the beta-diversity matrix were analyzed by PERMANOVA using adonis function from vegan R-package with default settings (999 permutations).

Defining microbial taxa associated with farm- or non-farm environments

Differences in phylum, class, order, family and genus level relative abundance between farm and non-farm homes was assessed using ANCOM v1.1-3 with false discovery rate (FDR) of 0.05.⁵² The association to farm- or non-farm environment was assigned to that with higher median abundance.

Defining farm home microbiota-like community composition

Farm home microbiota-like community composition was modelled in LUKAS1 with logistic regression analysis (PROC LOGISTIC statement, SAS version 9.3). The home location on a farm or non-farm rural environment was the dependent variable and the main components of Principal Coordinate Analysis (PCoA) axis scores of beta-diversity matrices were the predictor variables. Bacterial and fungal microbiota were investigated separately. For both bacteria and fungi separate models were built using axis scores from PCoA of abundance

unweighted and weighted β -diversity matrices. The PCoA axes were selected based on the scree plot method including axes above the point at which the variance explained by the additional axes levels off (Supplementary Figure 3). The models give an estimate of the probability that the sample is from a farm home. The farm home likeness of the microbial composition in the LUKAS2 non-farm homes was then estimated by applying the regression coefficients obtained from the LUKAS1-based models to the corresponding microbial data from LUKAS2 samples.

Some analyses were done in non-farm homes of both LUKAS2 and LUKAS1 to obtain increased sample size and power if results remain comparable as was observed. Due to the discovered association with asthma, the probability that was modelled based on the relative abundance weighted bacterial/archaeal beta-diversity was named Farm home Resembling Microbiota Index (FaRMI) and studied further in greater detail. In this model the first four PCoA axes were used as the predictor variables (Supplementary Figure 3). In the replication stage analogous logistic regression equation was built in GABRIELA to obtain probability that a given microbial composition represents (is more similar to) GABRIELA farm homes as opposed to homes of children neither living nor regularly exposed to farms (FaRMI_{GABRIELA}).

To test alternative methodological approach, we calculated equivalent to FaRMI (i.e. probability predicted based on microbiota composition that the sample is from a farm home as opposed to non-farm home) directly from the OTU-table with random forest analysis using RandomForestClassifier from scikit-learn python module.⁵³ Supervised training of the model was done in LUKAS1 dataset, which was randomly split to training and test set so that the test set had 25% of the samples. The classifier was trained using the training set and tested using the test set. The trained classifier was then used to calculate probability scores also for LUKAS2 samples.

Oligotyping *Methanobrevibacter* genus

Oligotyping analysis⁵⁴ with Oligotyping pipeline script version 1.7 was used to obtain indicative species level identification within *Methanobrevibacter* genus. This was done to determine whether the farm home associated increase in *Methanobrevibacter* was more likely of ruminant or human origin. Within LUKAS1 and 2 samples there were in total 1085 sequence reads, within 181 samples, that were assigned to *Methanobrevibacterium* genus based on OTU IDs. These sequences were aligned against the greengenes 16S rRNA database reference alignment using mothur version 1.3555 and trimmed for equal length of 253 base pairs. Uninformative positions due to insertion or deletion were removed with O-trim-uninformative-columns-from-alignment script. To obtain taxonomic assignments the oligotype reads were blasted with 16S Ribosomal RNA Bacteria/Archaeal database (NCBI blastn tool) and the results were filtered for minimum 99% identity. The results were optimized for highly similar sequences with megablast.

Source tracking

Using the Qiita web portal (<https://qiita.ucsd.edu>),⁵⁶ sequences from the LUKAS sample data were clustered into OTUs against the Greengenes database using the closed reference

workflow. These samples were combined with publicly available studies selected as potential source environments (bovine, human and soil) for the target LUKAS sequences. The Qiita IDs for the studies used are provided in the Supplementary Table 15. The combined data table was run through SourceTracker in the R platform to predict the sources of bacteria.

Predicted functional metagenomics

The metagenome functional content was predicted using PICRUSt software following the standard pipeline (<http://picrust.github.io/picrust/>) with the LUKAS OTU table, from which de-novo OTUs were removed, as the input file.⁵⁷ The functional predictions were based on Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology database. The quality of the PICRUSt predictions was good as indicated by low Nearest Sequenced Taxon Index (NSTI) median 0.049 (interquartile range 0.042-0.58) describing the availability of nearby genome representatives.⁵⁷ The data was analysed using STAMP software version 2.1.3.⁵⁸ Functional predictions associated with farmhouse-like bacterial/archaeal relative abundance were analysed by White's non-parametric t-test⁵⁹ using Benjamini-Hochberg false discovery rate for multiple testing correction. FaRMI was studied as a dichotomous variable, with probability of 0.5 as the cut-off.

Determinants with possible influence on indoor microbiome

To find determinants of farmhouse-like bacterial/archaeal relative abundance (FaRMI) in non-farm homes several environmental, building and occupancy associated determinants were tested including: the number of older siblings, type of living area (city centre, suburban, rural community, rural sparse), environmental biodiversity (based on land use information), home construction year, presence of basement, ground (slab vs. other), home frame material, moisture (condensation on windows, relative humidity and indoor specific absolute humidity), ventilation, heating with wood, contact with farm-animals, closeness of dunghill, having pets, walking with outdoor shoes inside. The effect of determinants on FaRMI were analysed by chi-square test using categorized variables. FaRMI was used as a dichotomous variable using probability of 0.5 as the cut-off. Independence of found significant associations from other, highly collinear determinants, was assessed by logistic regression analysis (data not shown).

Replication

Contribution of specific taxa to the PCoA axes will vary depending on the samples included in the beta-diversity matrix. Therefore the "true" replicability of the findings in LUKAS (LUKAS1 and LUKAS2) with LUKAS1 farm home-like microbiota (FaRMI) were tested using generalized linear models (GLM) analysis based estimates of FaRMI (PROC GLMSELECT statement, SAS 9.3). The estimates were calculated with the relative abundances of sOTUs under a particular taxa from phylum down to genus level as predictor variables. The taxa variables were ranked by the relative abundance percentile to standardize data and to avoid issues with influential data points. The variables were entered into the GLM model based on their significance until minimum predicted residual error sum of squares (PRESS) was reached using 10-fold cross validation.⁶⁰ We also tested the replicability in the other direction, i.e. from GABRIELA to LUKAS. For this purpose we

first computed $\text{FaRMI}_{\text{GABRIELA}}$, i.e. beta-diversity derived probability, predicted using beta-diversity based modelling of bacterial/archaeal relative abundance, a given sample represents GABRIELA farmhouse rather than a home of children neither living nor regularly exposed to farms. The $\text{FaRMI}_{\text{GABRIELA}}$ was then modelled using the GLM-based analysis.

Statistical analyses on health outcomes

The associations between the probability variables (including FaRMI) and asthma were studied with logistic regression models (PROC LOGISTIC statement, SAS version 9.3). These models were adjusted as per a priori decision for basic confounders including (as applicable) living on a farm, cohort, gender, the maternal history of allergic diseases (asthma, atopic dermatitis or allergic rhinitis), number of older siblings and smoking during pregnancy (never, only before pregnancy, during pregnancy). In addition the effect of several other confounding factors on the association between FaRMI and asthma were tested using the change-in-estimate criterion with 10% cutoff. The tested variables included paternal history of atopic disease (hay fever, atopic dermatitis and/or asthma) and asthma, maternal and paternal education levels, birth weight, mode of delivery, indoor exposure to dog and/or cat ownership at the age of 2 months, distance to farm, breast-feeding, consumption of farm milk, day care attendance and regular exposure to passive tobacco smoke at the age of 1 year as well as house type (detached vs. row house) and age, season (winter, spring/autumn, summer), type of vacuumed floor and time from last vacuuming with reference to dust sampling. Based on these analyses, all asthma association models were further adjusted by paternal allergic disease and maternal education level. Additionally, the independence of the observed association of markers of microbial diversity, proxies of total microbial levels⁹ including loads of lipopolysaccharide^{10:0-16:0} (LPS^{10:0-16:0}), muramic acid and endotoxin, and atopic sensitization were tested as indicated in the results. Overall, the additional adjustments had little influence to the observed effects. Where indicated, model instabilities due to small number of cases in stratified analyses were solved by combining categories in the instability causing variable and/or by using Firth's penalized likelihood-based bias-adjusted estimates.⁶¹

The association between the FaRMI and cytokine responses *ex vivo* were studied using quantile regression that is suitable for skewed data and found to be robust against heteroscedastic errors (PROC QUANTREG, SAS 9.3).⁶² The cytokine analyses were adjusted for the a priori decided basic confounders. Data on cytokine stimulations where the cytokine concentrations were below the detection limit in over 25% of samples were not analyzed. In the replication stage, with GABRIELA data, the models were adjusted with comparable variables as in LUKAS including gender, first degree relative with allergic disease (asthma, atopic dermatitis or hay fever), number of older sibling, smoking during pregnancy (yes/no) and maternal education level as well as study design related variables including age of the child, study center and in models with non-farm children alone the strata (i.e. not living but regularly exposed to farms, neither exposed nor living on a farm). Based on stratified analysis and non-significant interaction term, there was no evidence that strata had significant influence on the reported results (data not shown). Associations between FaRMI and dendritic cells inhibitory molecule expression were assessed within subsample of LUKAS¹ non-farm children stratified by asthma ever using Spearman

rank's correlation test (PROC CORR statement, SAS 9.3) with the a priori decided basic confounders as partial variables. To test whether such association were different between asthmatic and non-asthmatics was determined based on interaction term significance in logistic regression modeling adjusted with the a priori decided basic confounders.

All statistical analyses with health outcomes were performed using SAS Enterprise Guide 5.1 (SAS Institute Inc., Cary, NC, USA) unless stated otherwise.

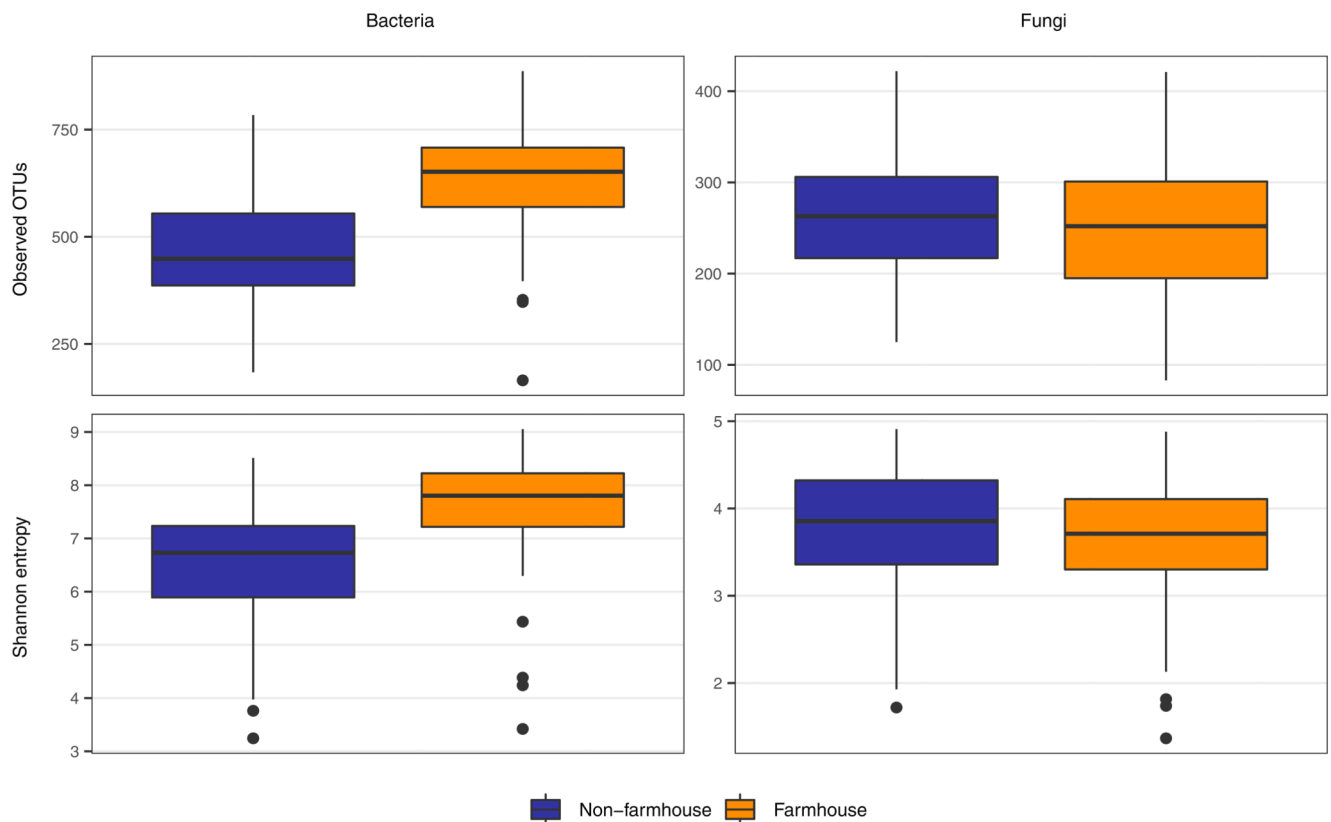
Data availability

The bacterial and fungal sequences from LUKAS have been deposited in European Bioinformatics Institute European Nucleotide Archive database under accession number PRJEB29081. Other data supporting the findings of this study are available through direct communication with the corresponding author. Limitations apply to variables where too small subgroups may compromise research participant privacy/consent. In these cases amendment to the ethical approval will be required prior to data transfer.

Code availability

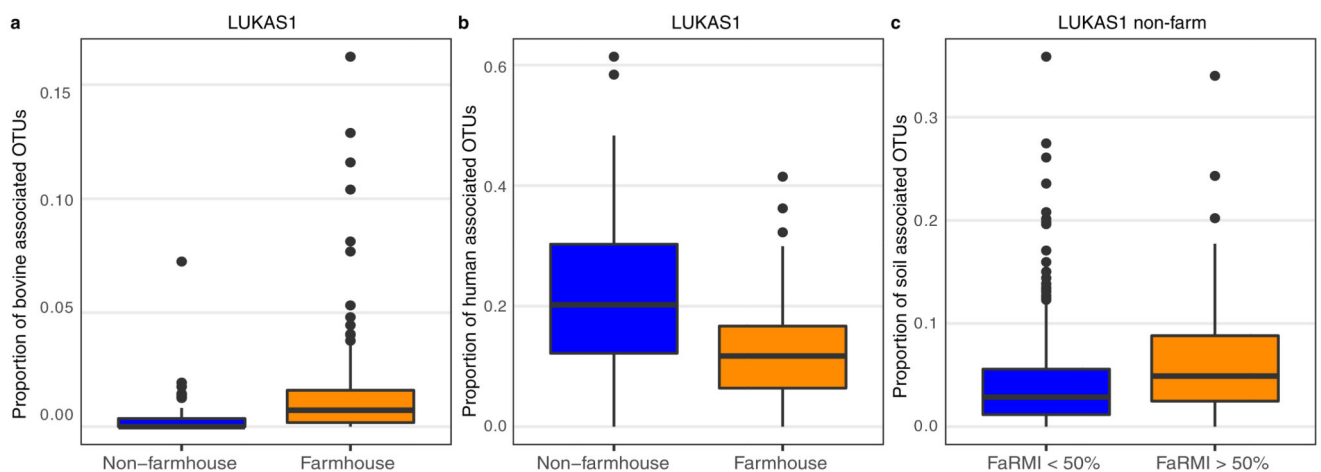
All codes used in the study are available on the public repository [<https://github.com/PirkkaKirjavainen/FaRMI>]. Contact the corresponding author for more information.

Extended Data



Extended Data Figure 1. Bacterial and fungal diversity in farm- and non-farmhomes.

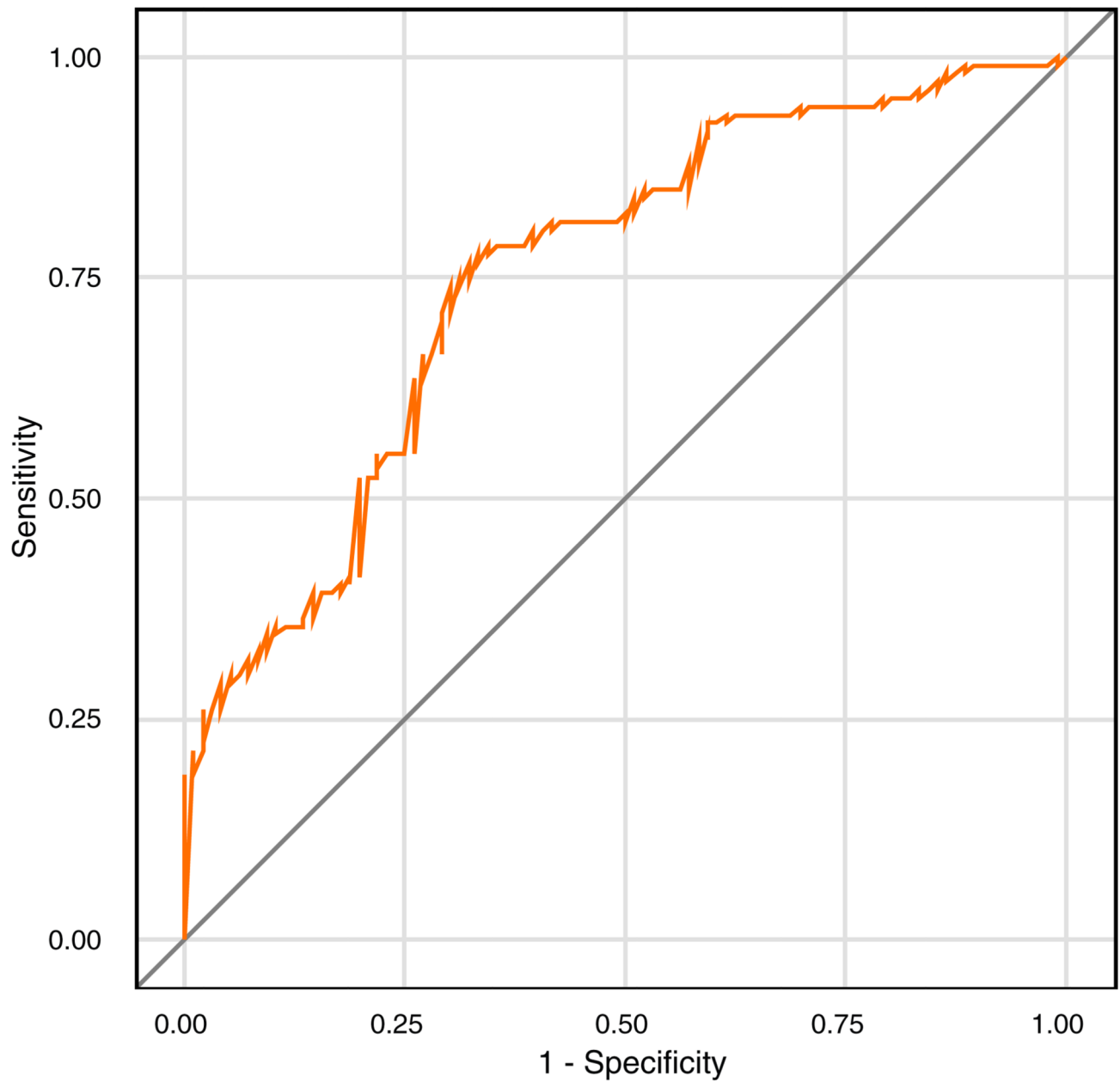
In the LUKAS1 farm homes the bacterial/archaeal richness (N=107), with median 652 operational taxonomic units (OTUs, interquartile range (IQR) 567-708) and Shannon entropy, with median 7.8 (IQR 7.2-8.2), were consistently higher than in the majority of the rural control homes (N=96) where the respective values were 449 (IQR 384-555) for richness and 6.7 (IQR 5.9-7.2) for Shannon (Wilcoxon, two-sided $p < 0.0001$). In fungal microbiota, there was a tendency for higher richness, with median 263 (217-306) fungal OTUs, and Shannon entropy, with median 3.9 (3.4-4.3), in the rural control (N=97) than farm homes (N=101) where the respective values were 252 (195-301; $p = 0.12$) for richness and 3.7 (3.3-4.1; Wilcoxon, two-sided $p = 0.07$) for Shannon. The boxes represent IQR with median marked within the box, the whiskers represent minimum/maximum value within $1.5 \times$ IQR below the lower quartile/above the upper quartile, respectively, the dots represent outliers.

**Extended Data Figure 2. Key microbial sources in floor dust microbiota in farm- and non-farm homes**

The relative abundance of (a) bovine- and (b) human associated bacterial/archaeal operational taxonomic units (OTUs) in living room floor dust in farm- (N=107) and non-farm homes (N=96) within LUKAS1 as determined by source tracking. Both comparisons (a-b) were significantly different with Wilcoxon test at two-sided $p < 0.0001$. (c) Relative abundance of soil-associated bacterial/archaeal OTUs was higher in the LUKAS1 (LUKAS1 and 2) non-farm homes that resembled more LUKAS1 farm- (N=179) than non-farm homes (N=215) as defined by FaRMI (Wilcoxon test, two-sided $p = 0.0003$). The boxes represent IQR with median marked within the box, the whiskers represent minimum/maximum value within $1.5 \times$ IQR below the lower quartile/above the upper quartile, respectively, the dots represent outliers.

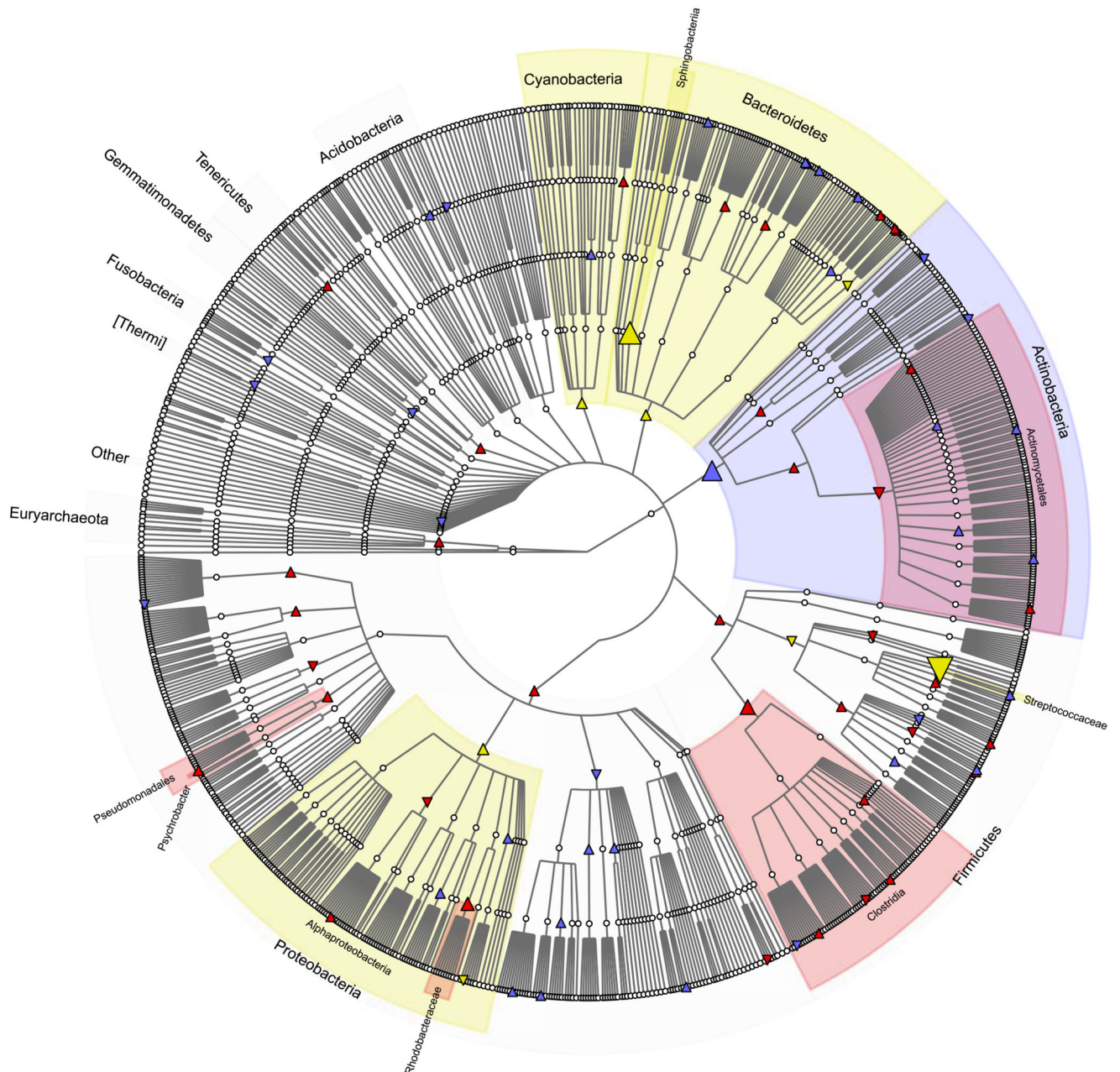


Nat Med. Author manuscript; available in PMC 2024 December 07.



Extended Data Figure 4. Classification accuracy of LUKAS1 farm-like microbiota model in the data it was trained in (LUKAS1)

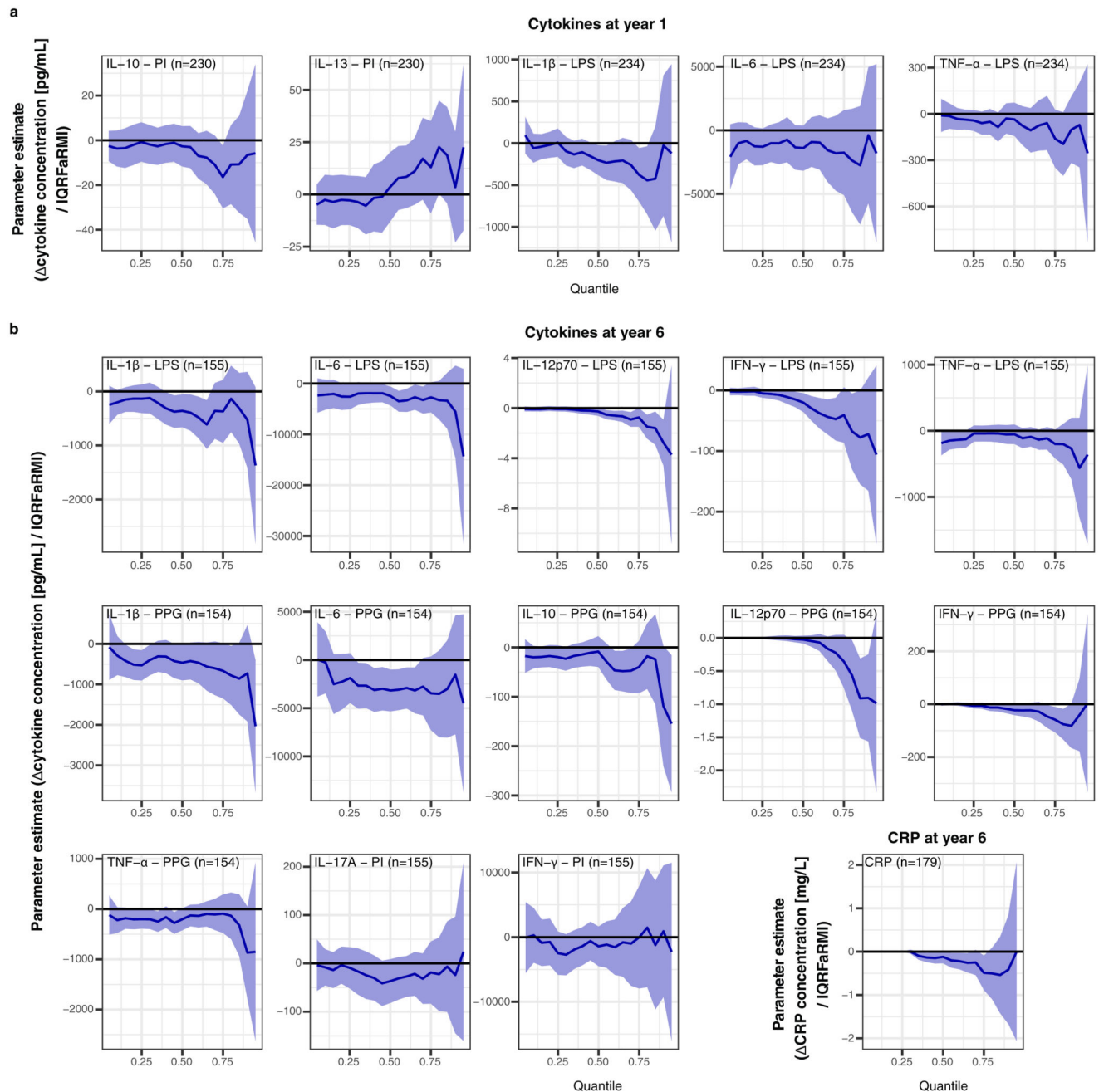
Based on the receiver operating characteristics (ROC) curve FaRMI had only moderate classification accuracy with area under the curve 0.74. This is a critical feature of FaRMI as it enables the detection of farm-like features also in non-farm homes.



Extended Data Figure 5. Taxa included in models of LUKAS1 and GABRIELA farm-like indoor microbiota

That is variables contributing to FaRMI and FaRMI_{GABRIELA}, respectively (see also Supplementary Table 9). Taxa in both models marked with yellow, taxa in LUKAS only with blue and taxa in GABRIELA only with red triangles. The direction of the triangle indicates negative (▼) or positive (▲) association with FaRMI and FaRMI_{GABRIELA}. The size of the triangles are proportional to the variance explained by the taxa (adjusted partial R^2); for the common taxa the higher adjusted R^2 . In three cases the direction was opposite

between the two models, in these cases the triangles represent the model where the taxa had higher adjusted R^2 . Clades are colored where the adjusted R^2 was $>1\%$.

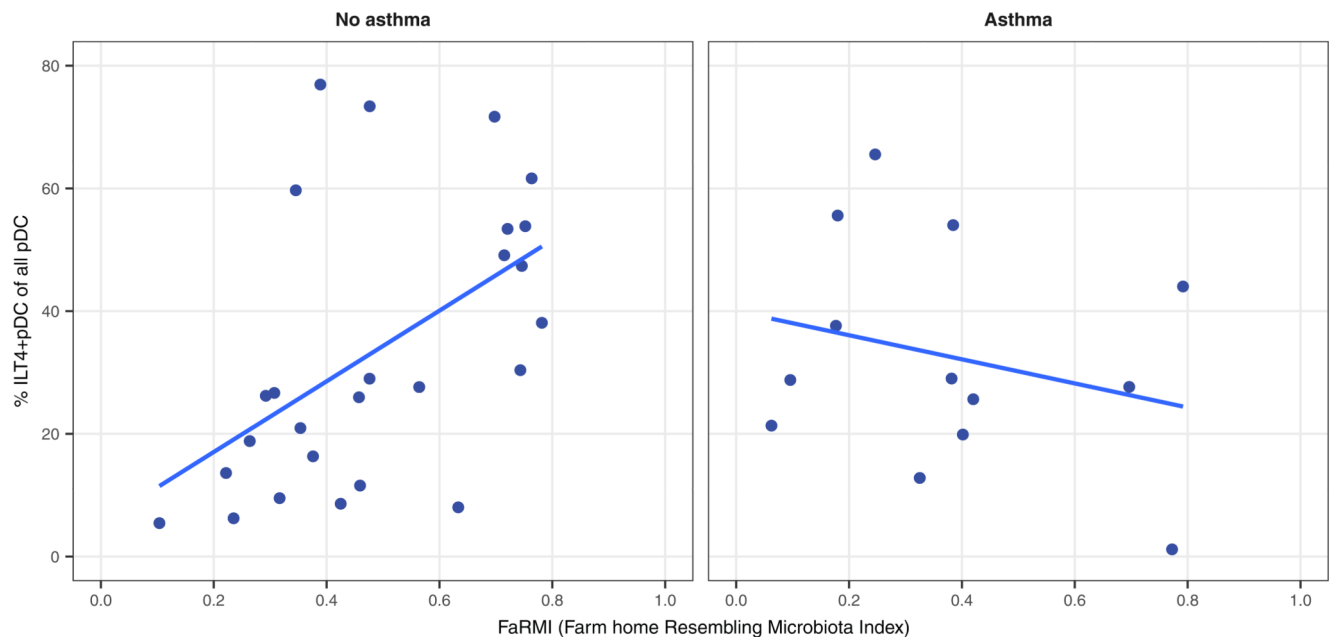


Extended Data Figure 6. Cytokine responses and serum CRP-levels in association to farm-like indoor microbiota

Quantile process plots of the quantile regression analysis showing the estimated change in cytokine concentration (pg/mL) at given percentile per one interquartile change in Farm home Resembling Microbiota Index (FaRMI) at year 1 **(a)** and year 6 **(b)**.

The shaded areas show the 95% confidence intervals based on resampling with 1000

repetitions and where these do not overlap with the horizontal zero-change line the decrease/increase at that percentile is statistically significant (two-sided p-value without correction for multiple testing <0.05). Plots are presented for all cytokines that show tendency for significant ($p < 0.1$) association with FaRMI between the 25th and 80th percentile without correction for multiple testing. Cytokines were measured from blood cultures stimulated with phorbol 12-myristate 13-acetate and ionomycin (PI), lipopolysaccharide (LPS), peptidoglycan (PPG). CRP was measured from serum. IL=interleukin, TNF=tumour necrosis factor, IFN=interferon and CRP=C-reactive protein.



Extended Data Figure 7. Proportion of immunoglobulin-like transcript (ILT) 4 expressing plasmacytoid dendritic cells (pDC) is correlated with FaRMI.

Proportion of immunoglobulin-like transcript (ILT) 4 expressing plasmacytoid dendritic cells (pDC) increased with increasing FaRMI within LUKAS1 children living in a non-farm home who were not diagnosed with asthma by 6 years of age ($N=26$). In those children who had been diagnosed with asthma ($N=16$) such a correlation did not exist. In logistic regression analysis modelling the association between the ILT4 expression on pDCs and FaRMI the interaction term asthma ever * FaRMI was significant ($p=0.03$). The scatterplots fitted with simple linear regression lines

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank J. Kauttio for the processing of LUKAS dust samples, G. Humphrey and J. Gaffney for the processing of GABRIELA samples, A. Amir for assistance in bioinformatics; U. Naukkarinen and R. Tiihonen for cytokine stimulations; S. Illi for cytokine data processing; M. Martikainen for dendritic cell analyses; the participating families of LUKAS and GABRIELA studies; the field workers in LUKAS and GABRIELA study and the

LUKAS and GABRIELA study groups. We acknowledge funding by the Academy of Finland grants 139021 J.P., 256375 M.R., 287675 A.K., 296814 J.P., 296817 M.T. and 308254 P.V.K.; the Juho Vainio Foundation P.V.K., M.T.; Päivikki and Sakari Sohlberg Foundation P.V.K., J.P.; The Finnish Cultural Foundation J.P.; Yrjö Jahnsson Foundation P.V.K., J.P., Kuopion seudun hengityssäätiö B.J. European Union QLK4-CT-2001-00250 J.P., H.R., P.I.P., R.L. E.v.M; the National Institute for Health and Welfare, Finland P.V.K., A.M.K., M.T., B.J., A.H., J.P.; Alfred P. Sloan Foundation G-2016-7076 R.A.; Deutsche Zentrum für Lungenforschung grants M.E. and 82DZL00502 H.R.; Deutsche Forschungsgemeinschaft (DFG)-funded SFB 1021 H.R.; SCHA 997/8-1 B.S.; GILKUIJ-39 B.S.; Kühne Foundation, Schindlegli, Switzerland R.L.; MU 891/5-1 Leibniz Prize, German Research Foundation E.v.M., ERC2009-AdG_20090506_250268 E.v.M.; LSHB-CT-2006-018996 E.v.M.

References

1. von Mutius E, Vercelli D. Farm living: effects on childhood asthma and allergy. *Nature reviews. Immunology*. 2010; 10: 861–868. [PubMed: 21060319]
2. Ege MJ, et al. Exposure to environmental microorganisms and childhood asthma. *The New England journal of medicine*. 2011; 364: 701–709. [PubMed: 21345099]
3. Schuijs MJ, et al. Farm dust and endotoxin protect against allergy through A20 induction in lung epithelial cells. *Science*. 2015; 349: 1106–1110. [PubMed: 26339029]
4. Karvonen AM, et al. Confirmed moisture damage at home, respiratory symptoms and atopy in early life: a birth-cohort study. *Pediatrics*. 2009; 124: e329–338. [PubMed: 19651571]
5. Liu AH. Revisiting the hygiene hypothesis for allergy and asthma. *Journal of Allergy and Clinical Immunology*. 2015; 136: 860–865. [PubMed: 26449798]
6. Reynolds LA, Finlay BB. Early life factors that affect allergy development. *Nature reviews. Immunology*. 2017. [PubMed: 28504257]
7. Stein MM, et al. Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children. *The New England journal of medicine*. 2016; 375: 411–421. DOI: 10.1056/NEJMoa1508749 [PubMed: 27518660]
8. Debarry J, et al. *Acinetobacter lwoffii* and *Lactococcus lactis* strains isolated from farm cowsheds possess strong allergy-protective properties. *The Journal of allergy and clinical immunology*. 2007; 119: 1514–1521. [PubMed: 17481709]
9. Karvonen AM, et al. Quantity and diversity of environmental microbial exposure and development of asthma: a birth cohort study. *Allergy*. 2014; 69: 1092–1101. DOI: 10.1111/all.12439 [PubMed: 24931137]
10. USEPA. Child-Specific Exposure Factors Handbook (Final Report). U.S. EPA; Washington, DC: 2008.
11. Hyytiäinen HK, et al. Crawling-induced floor dust resuspension affects the microbiota of the infant breathing zone. *Microbiome*. 2018; 6: 25. doi: 10.1186/s40168-018-0405-8 [PubMed: 29394954]
12. Janssen PH, Kirs M. Structure of the archaeal community of the rumen. *Applied and environmental microbiology*. 2008; 74: 3619–3625. DOI: 10.1128/AEM.02812-07 [PubMed: 18424540]
13. Birzele LT, et al. Environmental and mucosal microbiota and their role in childhood asthma. *Allergy*. 2017; 72: 109–119. [PubMed: 27503830]
14. Braun-Fahrlander C, et al. Environmental exposure to endotoxin and its relation to asthma in school-age children. *The New England journal of medicine*. 2002; 347: 869–877. [PubMed: 12239255]
15. Lynch SV, et al. Effects of early-life exposure to allergens and bacteria on recurrent wheeze and atopy in urban children. *The Journal of allergy and clinical immunology*. 2014; 134: 593–601 e512. DOI: 10.1016/j.jaci.2014.04.018 [PubMed: 24908147]
16. Valkonen M, et al. Microbial characteristics in homes of asthmatic and non-asthmatic adults in the ECRHS cohort. *Indoor air*. 2018; 28: 16–27. [PubMed: 28960492]
17. Ottman N, et al. Soil exposure modifies the gut microbiota and supports immune tolerance in a mouse model. *Journal of Allergy and Clinical Immunology*. 2018. [PubMed: 30097187]
18. Genuneit J, et al. The GABRIEL Advanced Surveys: study design, participation and evaluation of bias. *Paediatric and perinatal epidemiology*. 2011; 25: 436–447. [PubMed: 21819425]
19. Leppanen HK, et al. Quantitative assessment of microbes from samples of indoor air and dust. *Journal of exposure science & environmental epidemiology*. 2017. [PubMed: 28975927]

20. Adams RI, Bateman AC, Bik HM, Meadow JF. Microbiota of the indoor environment: a meta-analysis. *Microbiome*. 2015; 3: 49. doi: 10.1186/s40168-015-0108-3 [PubMed: 26459172]
21. Barberan A, et al. Continental-scale distributions of dust-associated bacteria and fungi. *Proc Natl Acad Sci U S A*. 2015; 112: 5756–5761. DOI: 10.1073/pnas.1420815112 [PubMed: 25902536]
22. Natarajan S, Kim J, Bouchard J, Cruikshank W, Remick DG. Pulmonary endotoxin tolerance protects against cockroach allergen-induced asthma-like inflammation in a mouse model. *International archives of allergy and immunology*. 2012; 158: 120–130. DOI: 10.1159/000330896 [PubMed: 22269653]
23. Kumar S, Adhikari A. Dose-dependent immunomodulating effects of endotoxin in allergic airway inflammation. *Innate immunity*. 2017; 23: 249–257. [PubMed: 28142300]
24. Rochat MK, et al. Maternal vitamin D intake during pregnancy increases gene expression of ILT3 and ILT4 in cord blood. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*. 2010; 40: 786–794. [PubMed: 20030662]
25. Wu J, Horuzsko A. Expression and function of ILTs on tolerogenic dendritic cells. *Human immunology*. 2009; 70: 353–356. DOI: 10.1016/j.humimm.2009.01.024 [PubMed: 19405174]
26. de Heer HJ, et al. Essential Role of Lung Plasmacytoid Dendritic Cells in Preventing Asthmatic Reactions to Harmless Inhaled Antigen. *The Journal of Experimental Medicine*. 2004; 200: 89–98. DOI: 10.1084/jem.20040035 [PubMed: 15238608]
27. Rosenblueth M, Martinez-Romero JC, Reyes-Prieto M, Rogel MA, Martinez-Romero E. Environmental mycobacteria: a threat to human health? *DNA and cell biology*. 2011; 30: 633–640. [PubMed: 21595554]
28. Deng W, et al. Assembly, structure, function and regulation of type III secretion systems. *Nature reviews. Microbiology*. 2017; 15: 323–337. [PubMed: 28392566]
29. Wallden K, Rivera-Calzada A, Waksman G. Type IV secretion systems: versatility and diversity in function. *Cellular microbiology*. 2010; 12: 1203–1212. DOI: 10.1111/j.1462-5822.2010.01499.x [PubMed: 20642798]
30. Shrivastava R, Miller JF. Virulence factor secretion and translocation by *Bordetella* species. *Current opinion in microbiology*. 2009; 12: 88–93. DOI: 10.1016/j.mib.2009.01.001 [PubMed: 19186097]
31. Pizarro-Cerdá J, Cossart P. Bacterial Adhesion and Entry into Host Cells. *Cell*. 2006; 124: 715–727. [PubMed: 16497583]
32. Shan L, He P, Sheen J. Intercepting Host MAPK Signaling Cascades by Bacterial Type III Effectors. *Cell host & microbe*. 2007; 1: 167–174. [PubMed: 18005696]
33. Bakowski MA, Cirulis JT, Brown NF, Finlay BB, Brumell JH. SopD acts cooperatively with SopB during *Salmonella enterica* serovar Typhimurium invasion. *Cellular microbiology*. 2007; 9: 2839–2855. [PubMed: 17696999]
34. Shao F, Merritt PM, Bao Z, Innes RW, Dixon JE. A *Yersinia* effector and a *Pseudomonas* avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell*. 2002; 109: 575–588. [PubMed: 12062101]
35. Needham BD, Trent MS. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nature Reviews Microbiology*. 2013; 11: 467. doi: 10.1038/nrmicro3047 [PubMed: 23748343]
36. Bisgaard H, et al. Childhood asthma after bacterial colonization of the airway in neonates. *The New England journal of medicine*. 2007; 357: 1487–1495. [PubMed: 17928596]
37. O'Connor GT, et al. Early-life home environment and risk of asthma among inner-city children. *The Journal of allergy and clinical immunology*. 2018; 141: 1468–1475. DOI: 10.1016/j.jaci.2017.06.040 [PubMed: 28939248]
38. Oksanen, J, , et al. R package. R Foundation for Statistical Computing; Vienna, Austria: 2016.
39. Schram-Bijkerk D, et al. Bacterial and fungal agents in house dust and wheeze in children: the PARSIFAL study. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*. 2005; 35: 1272–1278. [PubMed: 16238785]
40. Caporaso JG, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*. 2011; 108: 4516–4522. DOI: 10.1073/pnas.1000080107 [PubMed: 20534432]

41. Smith DP, Peay KG. Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. *PloS one*. 2014; 9: e90234. doi: 10.1371/journal.pone.0090234 [PubMed: 24587293]
42. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*. 2011; 27: 2957–2963. DOI: 10.1093/bioinformatics/btr507 [PubMed: 21903629]
43. Caporaso JG, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010; 7: 335–336. DOI: 10.1038/nmeth.f.303 [PubMed: 20383131]
44. Rideout JR, et al. Consistent, comprehensive and computationally efficient OTU definitions. *PeerJ PrePrints*. 2014; 2: e411v411. doi: 10.7717/peerj.545 [PubMed: 25177538]
45. Amir A, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems*. 2017; 2 doi: 10.1128/mSystems.00191-16 [PubMed: 28289731]
46. Herzum I, Blumer N, Kersten W, Renz H. Diagnostic and analytical performance of a screening panel for allergy. *Clin Chem Lab Med*. 2005; 43: 963–966. [PubMed: 16176178]
47. Mustonen K, et al. Moisture damage in home associates with systemic inflammation in children. *Indoor air*. 2016; 26: 439–447. [PubMed: 25924948]
48. Bomert M, et al. Analytical performance of a multiplexed, bead-based cytokine detection system in small volume samples. *Clinical chemistry and laboratory medicine*. 2011; 49: 1691–1693. [PubMed: 21663461]
49. Martikainen MV, et al. Farm exposures are associated with lower percentage of circulating myeloid dendritic cell subtype 2 at age 6. *Allergy*. 2015; 70: 1278–1287. [PubMed: 26119336]
50. Chen J, et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics*. 2012; 28: 2106–2113. DOI: 10.1093/bioinformatics/bts342 [PubMed: 22711789]
51. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 2004; 20: 289–290. [PubMed: 14734327]
52. Mandal S, et al. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease*. 2015; 26: 27663. doi: 10.3402/mehd.v26.27663 [PubMed: 26028277]
53. Pedregosa F, et al. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 2011; 2825
54. Eren AM, Borisy GG, Huse SM, Mark Welch JL. Oligotyping analysis of the human oral microbiome. *Proc Natl Acad Sci U S A*. 2014; 111: E2875–2884. DOI: 10.1073/pnas.1409644111 [PubMed: 24965363]
55. Schloss PD, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*. 2009; 75: 7537–7541. DOI: 10.1128/AEM.01541-09 [PubMed: 19801464]
56. Gonzalez A, et al. Qiita: rapid, web-enabled microbiome meta-analysis. *Nature methods*. 2018; 15: 796–798. DOI: 10.1038/s41592-018-0141-9 [PubMed: 30275573]
57. Langille MG, et al. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology*. 2013; 31: 814–821. DOI: 10.1038/nbt.2676 [PubMed: 23975157]
58. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*. 2014; 30: 3123–3124. DOI: 10.1093/bioinformatics/btu494 [PubMed: 25061070]
59. White JR, Nagarajan N, Pop M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS computational biology*. 2009; 5: e1000352. doi: 10.1371/journal.pcbi.1000352 [PubMed: 19360128]
60. Cohen, RA. Introducing the GLMSELECT PROCEDURE for Model Selection (Paper 207-31). SAS Institute Inc; San Francisco, California: 2006. 1–18.
61. Firth D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993; 80: 27–38.
62. Uh HW, Hartgers FC, Yazdanbakhsh M, Houwing-Duistermaat JJ. Evaluation of regression methods when immunological measurements are constrained by detection limits. *BMC immunology*. 2008; 9: 59. doi: 10.1186/1471-2172-9-59 [PubMed: 18928527]

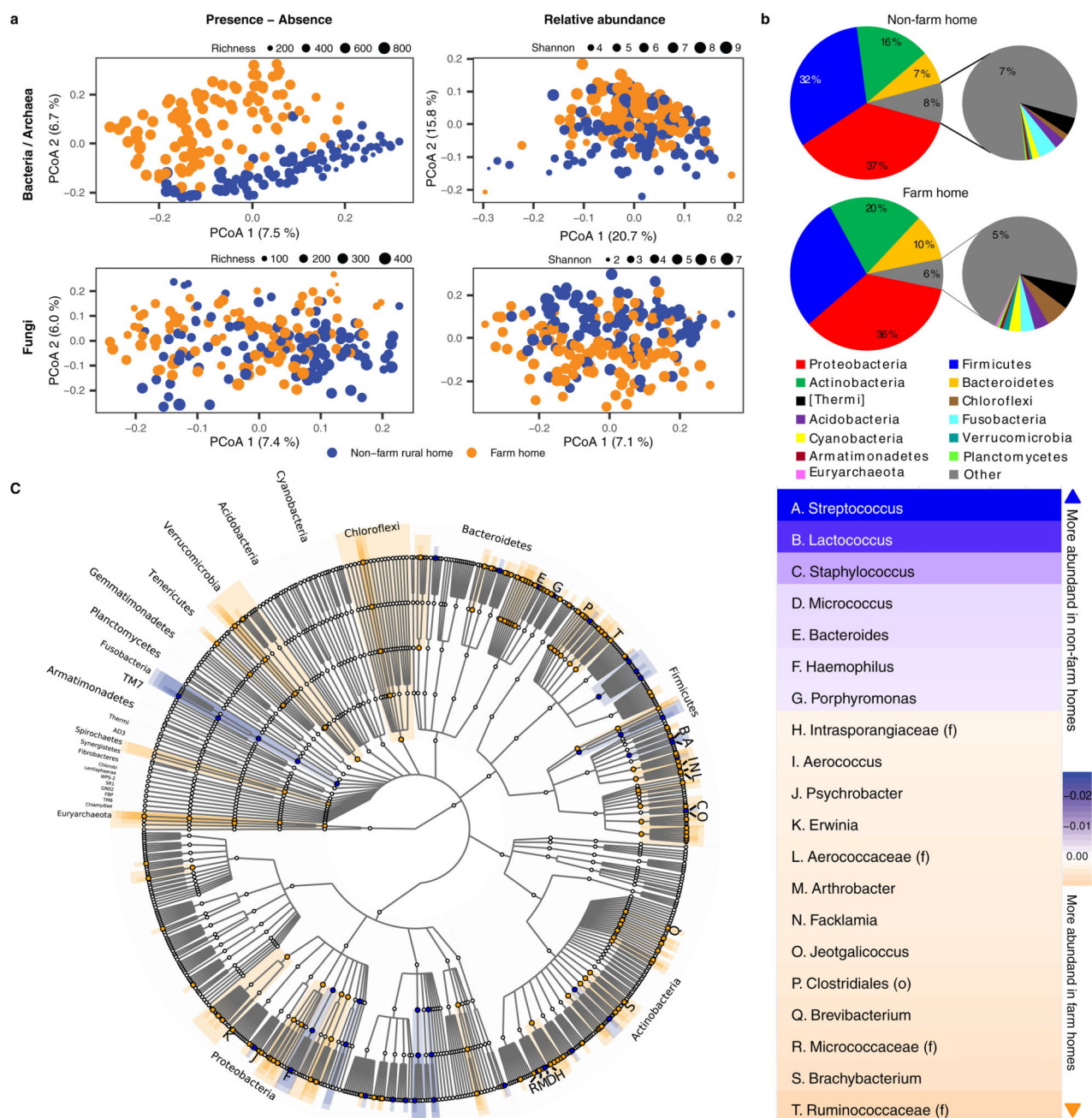


Figure 1. Differences between farm and non-farm rural home indoor microbiota.

(a) Dissimilarity (β –diversity) of bacterial/archaeal (phylogenetically informed) and fungal presence-absence and relative abundance patterns in living-room floor dust of farm (orange) and non-farm (blue) homes in LUKAS1. The first two PCoA axes with %-variance explained are presented. The differences between farm homes (n=107 with bacteria; n=101 with fungi) vs non-farm homes (n=96 with bacteria; n=97 with fungi) were significant in all the four distance matrices (Permutational Multivariate of Anova38, $p < 0.001$). α -Diversity of each sample is illustrated by the size of the points on the plot, which are directly

proportional to richness (number of OTUs) or index of Shannon entropy as indicated.

(b) Relative abundance of predominant bacterial phyla in non-farm and farm homes. **(c)** Bacterial/archaeal taxa with significantly higher relative abundance in farm than non-farm (orange circles) or in non-farm than farm homes (blue circles) as determined with ANCOM. Clades are coloured respectively up to family level. Top 20 bacterial genera with the greatest absolute difference in median relative abundance between farm and non-farm homes are indicated with a letter (A-T). For unassigned genera the highest assigned taxonomic name is presented (f=family, o=order).

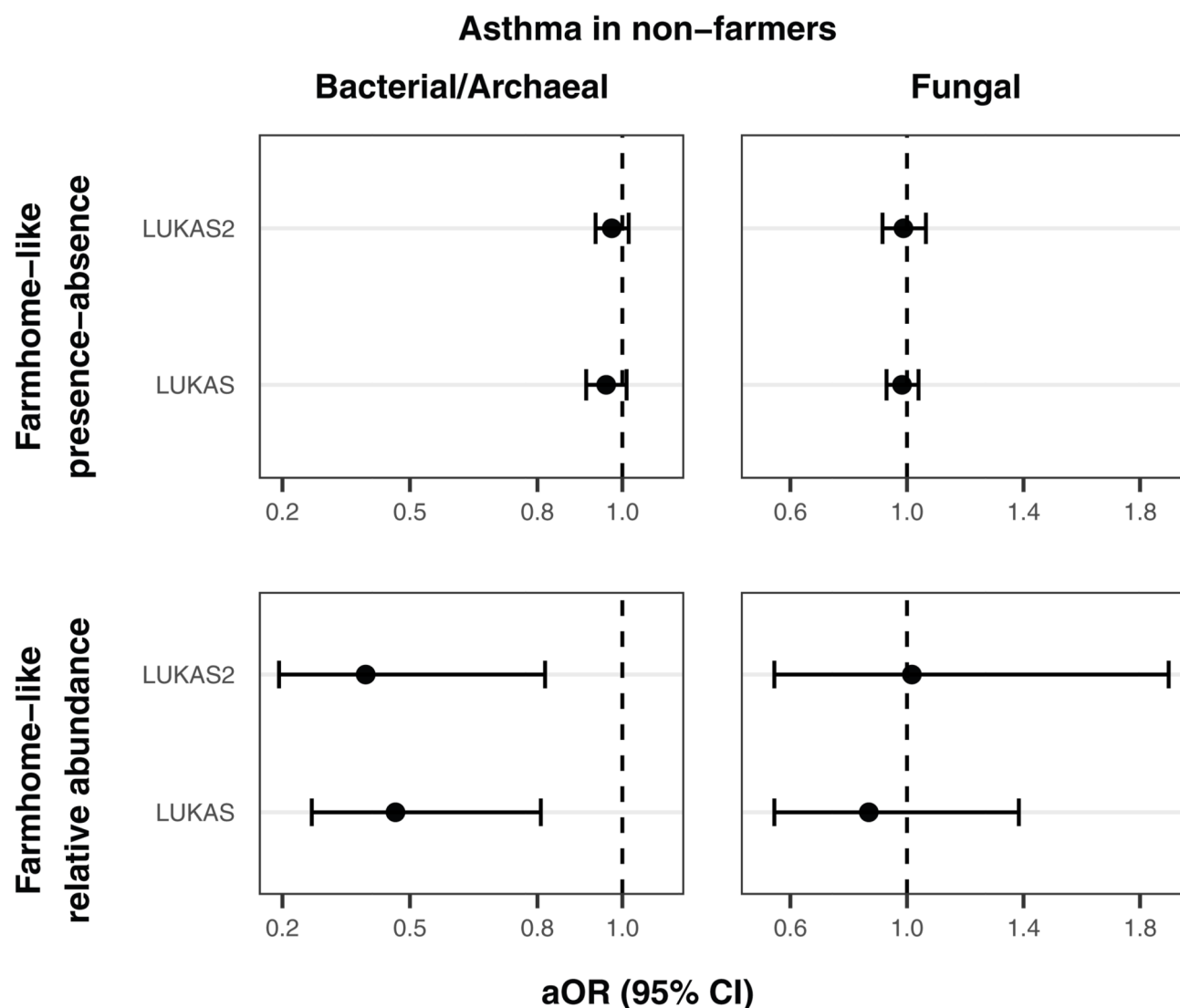


Figure 2. Farm home-like indoor microbiota is associated with asthma protection in non-farm children.

Association between asthma during the first 6 years of life and compositional similarity of home indoor dust bacterial/archaeal or fungal microbiota at age 2 months to that in farm homes in the suburban LUKAS2 (n=164) and the pooled LUKAS1 and LUKAS2 (LUKAS, n=251) studies. The compositional similarity was defined as beta-diversity-derived predicted probability that the sample would be from a LUKAS1 farm as opposed to LUKAS1 non-farm home. The association with asthma is shown as adjusted odds ratio per interquartile range (IQR) of the probability. The center values represent the odds ratios and the error bars 95% confidence intervals.

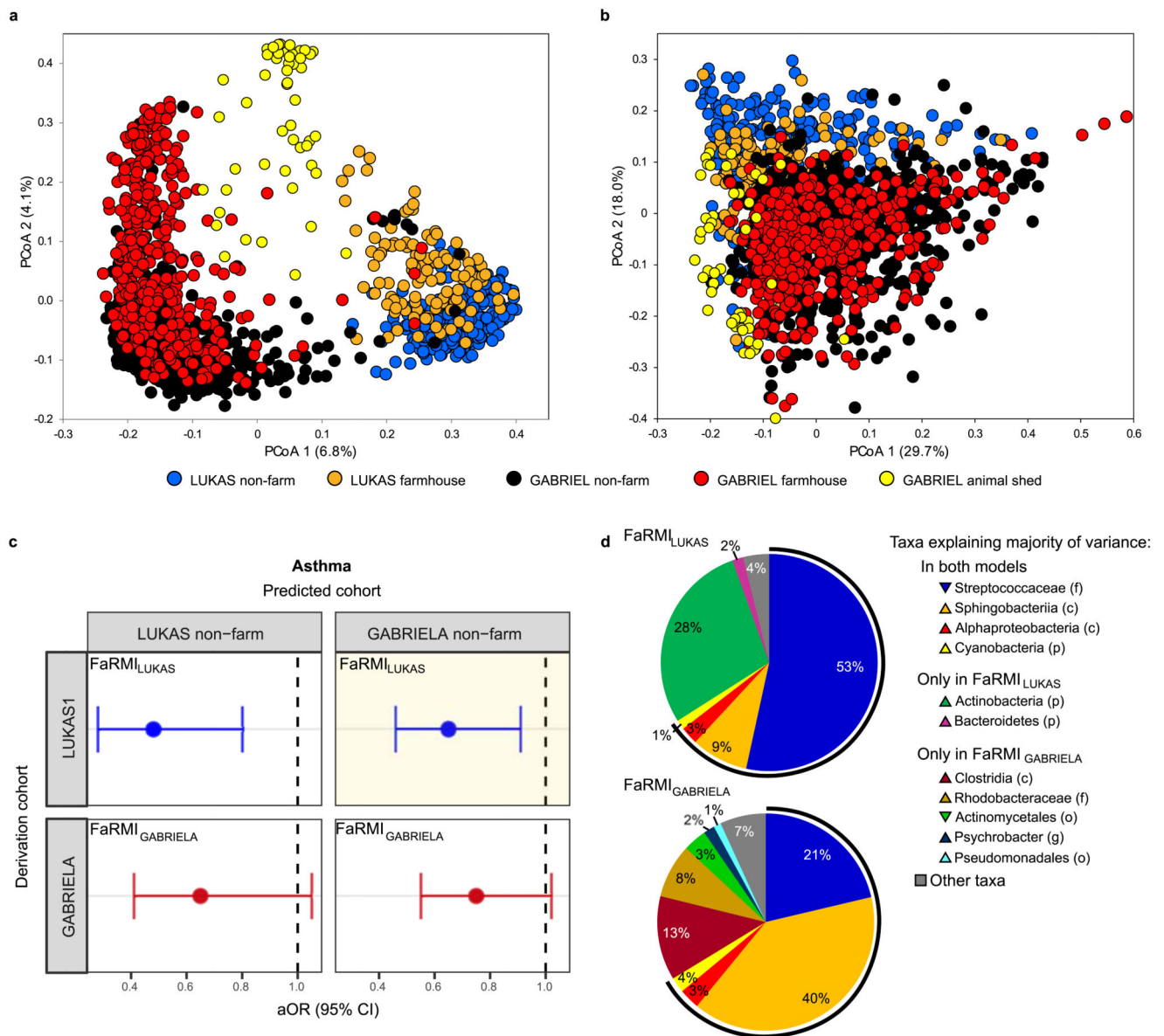


Figure 3. Replication of the asthma protective effect of growing up in a home with a farm home-like indoor bacterial microbiota.

Principal coordinate analysis of (a) unweighted and (b) weighted Generalized UniFrac analysis of GABRIELA mattress dust and LUKAS floor dust bacterial/archaeal microbiota.

$N_{(\text{LUKAS non-farm})}=278$, $N_{(\text{LUKAS farm home})}=116$, $N_{(\text{GABRIELA non-farm})}=632$, $N_{(\text{GABRIELA farm home})}=399$, $N_{(\text{GABRIELA animal shed})}=50$. While the cohort specific differences load primarily to the first axis (horizontal) in the presence-absence microbial data, the farm-effect on microbiota is visible in the second axis (vertical) where the farmhouses from both cohorts cluster closer to the animal shed microbiota. In the weighted analysis, this clustering pattern is also present but less pronounced. (c) The association between farm home-like indoor microbiota and asthma by 6 years of age among LUKAS ($n=244$) and by 6 to 12 years of age among GABRIELA ($N=603$) children living in a non-farm

home. The farm home-like microbiota was defined as modeled FaRMI_{LUKAS} (derived in LUKAS1) and FaRMI_{GABRIELA} (derived in GABRIELA) and adjusted odds ratios (aOR) are presented per interquartile range of modeled FaRMI_{LUKAS}(blue)/FaRMI_{GABRIELA} (red). The center values represent the odds ratios and the error bars 95% confidence intervals (CI). The primary replication from LUKAS to GABRIELA is highlighted with light orange shade **(d)** The same 4 taxa (Streptococcaceae, Sphingobacteriia, Alphaproteobacteria and Cyanobacteria) marked with a black arch explained nearly two-thirds of the variance of FaRMI in both LUKAS (FaRMI_{LUKAS}) and in GABRIELA (FaRMI_{GABRIELA}). The pie charts present all taxa with the adjusted $R^2 > 1\%$. The direction of the triangle indicates negative (▼) or positive (▲) association with FaRMI and/or FaRMI_{GABRIELA}. The letters after the taxa names stand for p=phylum, c=class, o=order, f= family, and g=genus.

Table 1
FaRMI was associated with reduced proinflammatory responsiveness.

The association between FaRMI and serum CRP levels and cytokine responses to *ex vivo* microbial stimulants in LUKAS non-farm children. Adjusted estimates of quantile regression analysis for cytokine/CRP concentrations at the 75th percentile per one interquartile range (IQR) change in FaRMI. The adjusted estimates are presented as a percentage of IQR of the respective cytokine/CRP concentration to allow comparison between the different analytes. *Quantile regression process plots are shown in the Extended Data Figure 6 for all analytes with suggestive associations ($p < 0.1$) to FaRMI anywhere between the 25th and 80th percentile. Where suggestive association was observed at the 75th percentile, p -value¹ is presented in the superscript.

Cytokine	Adjusted relative estimates ² at 75 th percentile in relation to FaRMI					
	Capacity to produce		Response to bacterial exposure			In vivo activation ³
	PI		LPS		PPG	
	1 year ⁴	6 years ⁵	1 year ⁶	6 years ⁵	6 years ⁷	
IL-1β	7.7	22.7	-27.5* ^{0.021}	-25.0*	-33.8* ^{0.056}	–
IL-4	19.7	-16.5	–	–	–	–
IL-5	9.7	11.4	–	–	–	–
IL-6	-15.9	10.5	-16.6*	-25.5* ^{0.073}	-30.7* ^{0.076}	24
IL-10	-28.9* ^{0.022}	-13.1	-12.1	1.6	-27.8*	13.1
IL-12p70	–	–	5.5	-43.0* ^{0.050}	-33.2* ^{0.084}	–
IL-13	18.1*	6.8	–	–	–	–
IL-17A	-3.8	-9.1*	–	–	–	–
IFN-γ	-7.5	-0.1*	-5.2	-46.2* ^{0.081}	-44.8* ^{0.017}	–
TNF-α	-4.7	-16.1	-22.9* ^{0.095}	-31.8* ^{0.073}	-8.3*	–
CRP	ND	ND	ND	ND	ND	-69.2* ^{0.016}

¹Two-sided, uncorrected for multiple testing.

²Adjusted for sex, maternal and paternal allergic disease, maternal smoking during pregnancy, number of older siblings, maternal education and cohort.

³Cytokines measured in non-stimulated cell culture media and CRP from serum.

⁴N=231 for IL-1 β and IL-4; N=230 for IL-5, IL-10, IL-12p70, IL-13 and IL17A; N=208 for IFN- γ and TNF- α ; and N=204 for IL-6;.

⁵N=155;

⁶N=234;

⁷N=154 for cytokines and N=179 for CRP.

ND = not determined, “–” = Data not analysed because over 25% of observations below the detection limit, PI= phorbol 12-myristate 13-acetate and ionomycin, LPS=lipopolysaccharide, PPG=peptidoglycan, IL=interleukin and IFN=interferon, CRP=C-reactive protein, FaRMI=Farm house Resembling Microbiota Index a probability variable describing the similarity of home indoor dust microbiota bacterial/archaeal relative abundance to that in LUKAS1 farm homes.