

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

Pirkka V. Kirjavainen,<sup>1,2</sup> Anne M. Karvonen,<sup>1</sup> Rachel I. Adams,<sup>3</sup> Martin Täubel,<sup>1</sup> Marjut Roponen,<sup>4</sup> Pauli Tuoresmäki,<sup>1,5</sup> Georg Loss,<sup>6</sup> Balamuralikrishna Jayaprakash,<sup>1</sup> Martin Depner,<sup>7</sup> Markus Johannes Ege,<sup>8,9</sup> Harald Renz,<sup>10</sup> Petra Ina Pfefferle,<sup>11</sup> Bianca Schaub,<sup>8</sup> Roger Lauener,<sup>12-15</sup> Anne Hyvärinen,<sup>1</sup> Rob Knight,<sup>16</sup> Dick J. J. Heederik,<sup>17</sup> Erika von Mutius<sup>7,8,9\*</sup> and Juha Pekkanen<sup>1,5\*</sup>

\*The authors contributed equally to this work

## Affiliations:

<sup>1</sup> Environmental Health Unit, National Institute for Health and Welfare, Kuopio, Finland

<sup>2</sup> Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland

<sup>3</sup> California Department of Public Health, California, United States

<sup>4</sup> Department of Environmental and Biological Sciences, University of Eastern Finland, Kuopio, Finland

<sup>5</sup> Department of Public Health, University of Helsinki, Helsinki, Finland

<sup>6</sup> University of California, San Diego, School of Medicine, Department of Pediatrics, CA, USA

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

<sup>8</sup> Dr. von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany

<sup>9</sup> Comprehensive Pneumology Center Munich, Member of the German Center for Lung Research, Munich, Germany

<sup>10</sup> Institute of Laboratory Medicine, Philipps University Marburg, Marburg, Germany

<sup>11</sup> Comprehensive Biomaterial Bank Marburg, Philipps-University Marburg, Germany

<sup>12</sup> Children's Hospital of Eastern Switzerland, St. Gallen, Switzerland

<sup>13</sup> University of Zurich, Zurich, Switzerland

<sup>14</sup> School of Medicine, University of St. Gallen, St. Gallen, Switzerland

<sup>15</sup> Christine Kühne-Center for Allergy Care and Education, CK-CARE, Davos, Switzerland

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

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

**Corresponding Author:** Pirkka V. Kirjavainen, [pirkka.kirjavainen@thl.fi](mailto:pirkka.kirjavainen@thl.fi)

## ABSTRACT

Asthma prevalence has increased in epidemic proportions with urbanization, but growing up on traditional farms offers protection even today.<sup>1</sup> The asthma-protective effect in farms appears to be associated with rich home dust microbiota,<sup>2,3</sup> 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 cohorts<sup>4</sup> 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 children<sup>2</sup> 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.

## MAIN TEXT

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.<sup>5,6</sup> 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.<sup>2,3</sup> If the indoor microbiota in farm homes causally protects from asthma, as suggested by experimental data,<sup>3,7,8</sup> 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,<sup>4,9</sup> 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.<sup>10,11</sup> 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).<sup>12</sup> 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).<sup>2,13,14,7</sup> This suggests that these general markers may be only proxies of more specific microbial composition such as described by FaRMI.<sup>15</sup> 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.<sup>16</sup>

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.<sup>17</sup>

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.<sup>18</sup> 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,<sup>19</sup> geographical location,<sup>20,21</sup> and other study-specific differences.<sup>20</sup> 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 ( $\text{FaRMI}_{\text{GABRIELA}}$ ) 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,<sup>15</sup> and living on the farm,<sup>1</sup> 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.<sup>2</sup> 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- $\gamma$ , interleukin (IL)-1 $\beta$ , 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.<sup>14</sup> Based on animal models, endotoxin tolerance may inhibit also allergen-induced airway inflammation.<sup>22,23</sup> 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.<sup>7</sup> 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 expression<sup>24</sup> 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.<sup>25</sup> pDCs are potential gate keepers that could direct immune response against harmless environmental antigens away from pro-asthmatic inflammatory responses such as airway eosinophilia.<sup>26</sup>

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.<sup>27</sup> This potential was visible in the predicted abundance of genes negatively associated with FaRMI (Supplementary Table 13).<sup>28-35</sup> 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.<sup>7,36,37</sup> 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.<sup>2,7,15,37</sup> 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.

## **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.vM; 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, Schindellegi, 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.

## **Auhor 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.



**Figure 1. Differences between farm and non-farm rural home indoor microbiota.** (a) Dissimilarity ( $\beta$ -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 Anova<sup>38</sup>,  $p < 0.001$ ).  $\alpha$ -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 ( $\blacktriangledown$ ) or positive ( $\blacktriangle$ ) 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 75<sup>th</sup> 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 25<sup>th</sup> and 80<sup>th</sup> percentile. Where suggestive association was observed at the 75<sup>th</sup> percentile, p-value<sup>1</sup> is presented in the superscript.

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

<sup>1</sup>Two-sided, uncorrected for multiple testing. <sup>2</sup>Adjusted for sex, maternal and paternal allergic disease, maternal smoking during pregnancy, number of older siblings, maternal education and cohort. <sup>3</sup>Cytokines measured in non-stimulated cell culture media and CRP from serum. <sup>4</sup>N=231 for IL-1 $\beta$  and IL-4; N=230 for IL-5, IL-10, IL-12p70, IL-13 and IL17A; N=208 for IFN- $\gamma$  and TNF- $\alpha$ ; and N=204 for IL-6;. <sup>5</sup>N=155; <sup>6</sup>N=234; <sup>7</sup>N=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.

## REFERENCES

1. von Mutius, E. & Vercelli, D. Farm living: effects on childhood asthma and allergy. *Nature reviews. Immunology* 10, 861-868 (2010).
2. Ege, M.J., *et al.* Exposure to environmental microorganisms and childhood asthma. *The New England journal of medicine* 364, 701-709 (2011).
3. Schuijs, M.J., *et al.* Farm dust and endotoxin protect against allergy through A20 induction in lung epithelial cells. *Science* 349, 1106-1110 (2015).
4. Karvonen, A.M., *et al.* Confirmed moisture damage at home, respiratory symptoms and atopy in early life: a birth-cohort study. *Pediatrics* 124, e329-338 (2009).
5. Liu, A.H. Revisiting the hygiene hypothesis for allergy and asthma. *Journal of Allergy and Clinical Immunology* 136, 860-865 (2015).
6. Reynolds, L.A. & Finlay, B.B. Early life factors that affect allergy development. *Nature reviews. Immunology* advance online publication(2017).
7. Stein, M.M., *et al.* Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children. *The New England journal of medicine* 375, 411-421 (2016).
8. Debarry, J., *et al.* Acinetobacter Iwoffii and Lactococcus lactis strains isolated from farm cowsheds possess strong allergy-protective properties. *The Journal of allergy and clinical immunology* 119, 1514-1521 (2007).
9. Karvonen, A.M., *et al.* Quantity and diversity of environmental microbial exposure and development of asthma: a birth cohort study. *Allergy* 69, 1092-1101 (2014).
10. USEPA. U.S. EPA. Child-Specific Exposure Factors Handbook (Final Report) (Washington, DC, 2008).
11. Hyytiainen, H.K., *et al.* Crawling-induced floor dust resuspension affects the microbiota of the infant breathing zone. *Microbiome* 6, 25 (2018).
12. Janssen, P.H. & Kirs, M. Structure of the archaeal community of the rumen. *Applied and environmental microbiology* 74, 3619-3625 (2008).
13. Birzele, L.T., *et al.* Environmental and mucosal microbiota and their role in childhood asthma. *Allergy* 72, 109-119 (2017).
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* 347, 869-877 (2002).
15. Lynch, S.V., *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* 134, 593-601 e512 (2014).
16. Valkonen, M., *et al.* Microbial characteristics in homes of asthmatic and non-asthmatic adults in the ECRHS cohort. *Indoor air* 28, 16-27 (2018).
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).
18. Genuneit, J., *et al.* The GABRIEL Advanced Surveys: study design, participation and evaluation of bias. *Paediatric and perinatal epidemiology* 25, 436-447 (2011).
19. Leppanen, H.K., *et al.* Quantitative assessment of microbes from samples of indoor air and dust. *Journal of exposure science & environmental epidemiology* (2017).
20. Adams, R.I., Bateman, A.C., Bik, H.M. & Meadow, J.F. Microbiota of the indoor environment: a meta-analysis. *Microbiome* 3, 49 (2015).
21. Barberan, A., *et al.* Continental-scale distributions of dust-associated bacteria and fungi. *Proc Natl Acad Sci U S A* 112, 5756-5761 (2015).
22. Natarajan, S., Kim, J., Bouchard, J., Cruikshank, W. & Remick, D.G. Pulmonary endotoxin tolerance protects against cockroach allergen-induced asthma-like inflammation in a mouse model. *International archives of allergy and immunology* 158, 120-130 (2012).
23. Kumar, S. & Adhikari, A. Dose-dependent immunomodulating effects of endotoxin in allergic airway inflammation. *Innate immunity* 23, 249-257 (2017).
24. Rochat, M.K., *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* 40, 786-794 (2010).

25. Wu, J. & Horuzsko, A. Expression and function of ILTs on tolerogenic dendritic cells. *Human immunology* 70, 353-356 (2009).
26. de Heer, H.J., *et al.* Essential Role of Lung Plasmacytoid Dendritic Cells in Preventing Asthmatic Reactions to Harmless Inhaled Antigen. *The Journal of Experimental Medicine* 200, 89-98 (2004).
27. Rosenblueth, M., Martinez-Romero, J.C., Reyes-Prieto, M., Rogel, M.A. & Martinez-Romero, E. Environmental mycobacteria: a threat to human health? *DNA and cell biology* 30, 633-640 (2011).
28. Deng, W., *et al.* Assembly, structure, function and regulation of type III secretion systems. *Nature reviews. Microbiology* 15, 323-337 (2017).
29. Wallden, K., Rivera-Calzada, A. & Waksman, G. Type IV secretion systems: versatility and diversity in function. *Cellular microbiology* 12, 1203-1212 (2010).
30. Shrivastava, R. & Miller, J.F. Virulence factor secretion and translocation by *Bordetella* species. *Current opinion in microbiology* 12, 88-93 (2009).
31. Pizarro-Cerdá, J. & Cossart, P. Bacterial Adhesion and Entry into Host Cells. *Cell* 124, 715-727 (2006).
32. Shan, L., He, P. & Sheen, J. Intercepting Host MAPK Signaling Cascades by Bacterial Type III Effectors. *Cell host & microbe* 1, 167-174 (2007).
33. Bakowski, M.A., Cirulis, J.T., Brown, N.F., Finlay, B.B. & Brumell, J.H. SopD acts cooperatively with SopB during *Salmonella enterica* serovar Typhimurium invasion. *Cellular microbiology* 9, 2839-2855 (2007).
34. Shao, F., Merritt, P.M., Bao, Z., Innes, R.W. & Dixon, J.E. A *Yersinia* effector and a *Pseudomonas* avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell* 109, 575-588 (2002).
35. Needham, B.D. & Trent, M.S. Fortifying the barrier: the impact of lipid A remodelling on bacterial pathogenesis. *Nature Reviews Microbiology* 11, 467 (2013).
36. Bisgaard, H., *et al.* Childhood asthma after bacterial colonization of the airway in neonates. *The New England journal of medicine* 357, 1487-1495 (2007).
37. O'Connor, G.T., *et al.* Early-life home environment and risk of asthma among inner-city children. *The Journal of allergy and clinical immunology* 141, 1468-1475 (2018).
38. Oksanen, J., *et al.* vgan: Community Ecology Package. in *R package* (R Foundation for Statistical Computing, Vienna, Austria, 2016).

## 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  $\geq 18$  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  $\geq 37$  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 study<sup>2</sup> 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 m<sup>2</sup> from a rug for two minutes or in the absence of a rug, an area of 4 m<sup>2</sup> 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.<sup>39</sup> 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 primers<sup>40</sup> and fungal ITS region by ITS1F/ITS2 primers.<sup>41</sup> 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,<sup>42</sup> while QIIME<sup>43</sup> 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 protocol<sup>44</sup> 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.<sup>45</sup> 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)<sup>46</sup> with cut-off  $\geq 3.5$  kU/L.

**Immune responsiveness.** Cytokine responsiveness was evaluated from cultured whole blood collected at 1 and 6 years as earlier described.<sup>47</sup> Overall cytokine production capacity profile of leukocytes was assessed from whole blood cultured 24h with phorbol 12-myristate 13-acetate and ionomycin (PI, 1  $\mu$ g/mL) and responsiveness to bacterial cell wall components from cultures stimulated 24h with lipopolysaccharide (LPS, 0.1  $\mu$ g/mL) or peptidoglycan (PPG, 10  $\mu$ 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- $\gamma$  were determined from the cell culture supernatants with multiplexed cytometric bead arrays (BD human CBAflex) with FACSArray bioanalyzer system (BD Biosciences).<sup>48</sup> 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.<sup>49</sup> 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  $\alpha$ -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 ( $\beta$ -diversity) in the bacterial/archaeal community was evaluated with Generalized UniFrac distances, calculated on using GUniFrac R-package<sup>50</sup> with midpoint rooted tree and using  $\alpha=0$  for abundance unweighted and  $\alpha=1$  for abundance weighted  $\beta$ -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.<sup>51</sup> 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.<sup>52</sup> 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  $\beta$ -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<sub>GABRIELA</sub>).

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.<sup>53</sup> 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 analysis<sup>54</sup> 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 1.35<sup>55</sup> 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>),<sup>56</sup> 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.<sup>57</sup> 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.<sup>57</sup> The data was analysed using STAMP software version 2.1.3.<sup>58</sup> Functional

predictions associated with farmhouse-like bacterial/archaeal relative abundance were analysed by White's non-parametric t-test<sup>59</sup> 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.<sup>60</sup> 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 levels<sup>9</sup> including loads of lipopolysaccharide10:0-16:0 (LPS10:0-16:0), 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.<sup>61</sup>

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).<sup>62</sup> 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.

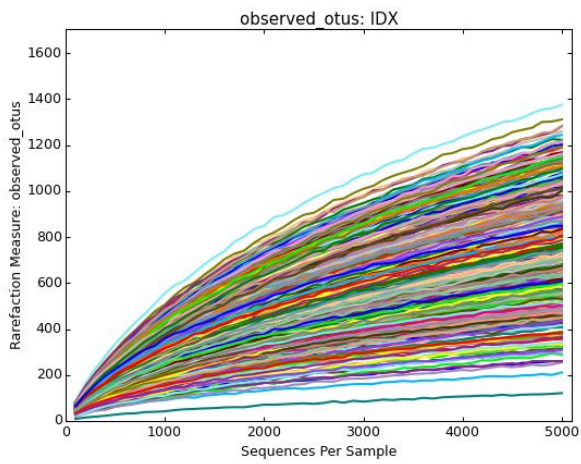
## REFERENCES FOR METHODS

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* 35, 1272-1278 (2005).
40. 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* 108, 4516-4522 (2011).
41. Smith, D.P. & Peay, K.G. Sequence Depth, Not PCR Replication, Improves Ecological Inference from Next Generation DNA Sequencing. *PLoS one* 9, e90234 (2014).
42. Magoc, T. & Salzberg, S.L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957-2963 (2011).
43. Caporaso, J.G., *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335-336 (2010).
44. Rideout, J.R., *et al.* Consistent, comprehensive and computationally efficient OTU definitions. *PeerJ PrePrints* 2, e411v411 (2014).
45. Amir, A., *et al.* Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *mSystems* 2(2017).
46. Herzum, I., Blumer, N., Kersten, W. & Renz, H. Diagnostic and analytical performance of a screening panel for allergy. *Clin. Chem. Lab. Med.* 43, 963-966 (2005).
47. Mustonen, K., *et al.* Moisture damage in home associates with systemic inflammation in children. *Indoor air* 26, 439-447 (2016).
48. Bomert, M., *et al.* Analytical performance of a multiplexed, bead-based cytokine detection system in small volume samples. *Clinical chemistry and laboratory medicine* 49, 1691-1693 (2011).
49. Martikainen, M.V., *et al.* Farm exposures are associated with lower percentage of circulating myeloid dendritic cell subtype 2 at age 6. *Allergy* 70, 1278-1287 (2015).
50. Chen, J., *et al.* Associating microbiome composition with environmental covariates using generalized UniFrac distances. *Bioinformatics* 28, 2106-2113 (2012).
51. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20, 289-290 (2004).
52. Mandal, S., *et al.* Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease* 26, 27663 (2015).

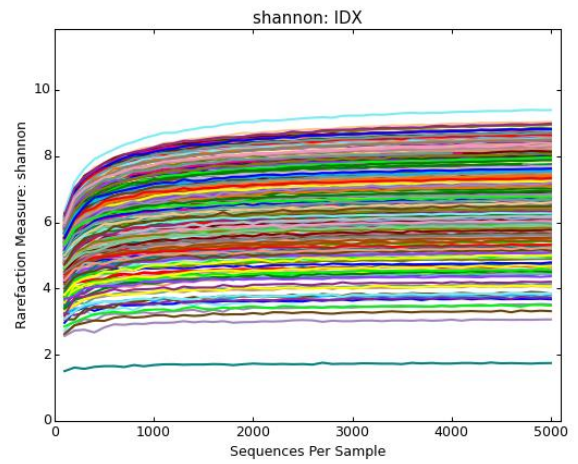
53. Pedregosa, F., *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2825-2830 (2011).
54. Eren, A.M., Borisy, G.G., Huse, S.M. & Mark Welch, J.L. Oligotyping analysis of the human oral microbiome. *Proc Natl Acad Sci U S A* 111, E2875-2884 (2014).
55. Schloss, P.D., *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology* 75, 7537-7541 (2009).
56. Gonzalez, A., *et al.* Qiita: rapid, web-enabled microbiome meta-analysis. *Nature methods* 15, 796-798 (2018).
57. Langille, M.G., *et al.* Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology* 31, 814-821 (2013).
58. Parks, D.H., Tyson, G.W., Hugenholtz, P. & Beiko, R.G. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123-3124 (2014).
59. White, J.R., Nagarajan, N. & Pop, M. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS computational biology* 5, e1000352 (2009).
60. Cohen, R.A. Introducing the GLMSELECT PROCEDURE for Model Selection (Paper 207-31). 1-18 (SAS Institute Inc., San Francisco, California 2006).
61. Firth, D. Bias reduction of maximum likelihood estimates. *Biometrika* 80, 27-38 (1993).
62. Uh, H.W., Hartgers, F.C., Yazdanbakhsh, M. & Houwing-Duistermaat, J.J. Evaluation of regression methods when immunological measurements are constrained by detection limits. *BMC immunology* 9, 59 (2008).

# Supplementary Figures

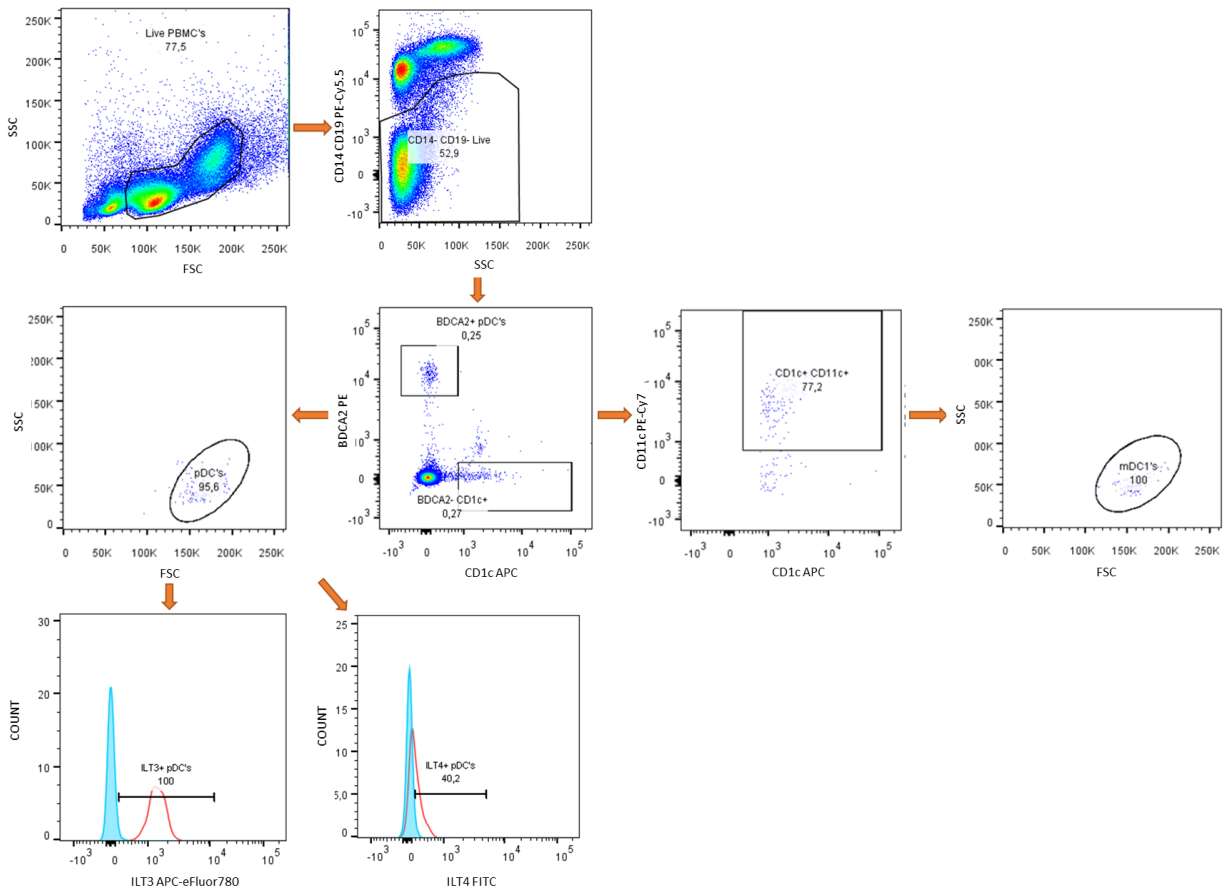
a)



b)



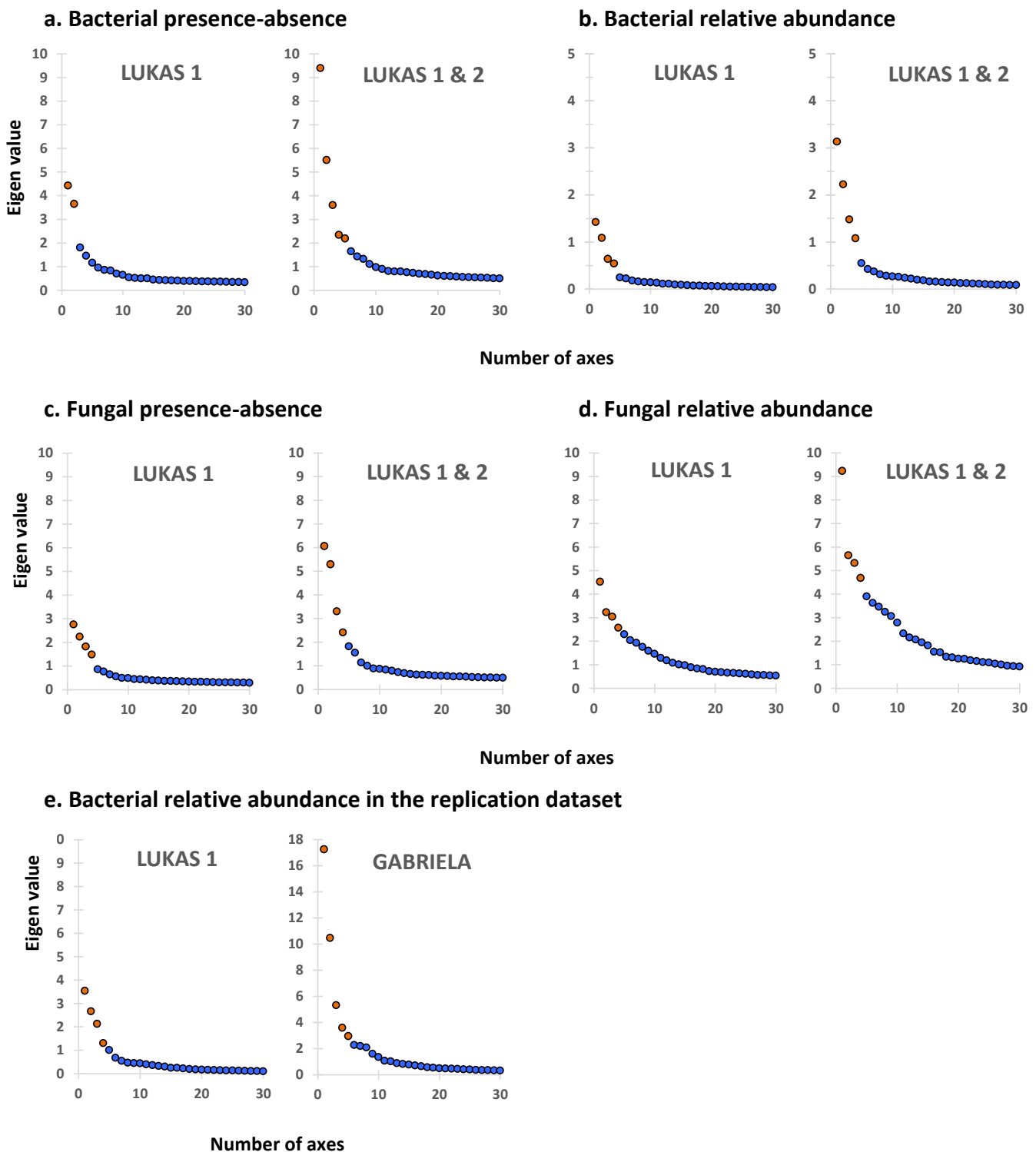
**Supplementary Figure 1.** Rarefaction curves for richness (observed OTUs) and Shannon index for each sample in LUKAS1 ( $n=203$ ) and LUKAS2 ( $n=191$ ). The figures indicate that alpha-diversity based sample rankings are largely established at  $\sim 1000$  sequences.



**Supplementary Figure 2.** Gating strategy for the identification of dendritic cell (DC) subsets.<sup>ref 49</sup>

Erythrocytes and debris were gated out according to the size (forward scatter) and cytoplasmic granularity (sideward scatter). 7-AAD positive dead cells, CD19+ B cells, and CD14+ monocytes were excluded by high PE-Cy 5.5 fluorescence. The main peripheral blood DC subsets were identified as BDCA2+ pDCs and BDCA1+CD11c+ mDCs. Gating adjustments were based on fluorescence minus one (FMO) tubes. The surface expression of ILT3 and ILT4 was analyzed after the DC subtype identification. Doublets were excluded using forward scatter area (FSC-A) and forward scatter height (FSC-H) to confirm that doublets did not interfere with analysis (not shown).





**Supplementary Figure 3.** Scree plots of principal coordinate analyses (PCoA) of beta-diversity matrices obtained at discovery stage from unweighted (a) or weighted (b) generalized UniFrac analysis of bacterial data or Bray-Curtis analysis of fungal presence absence (c) or relative abundance data (d); and at replication phase from PCoA of weighted UniFrac analysis of bacterial data. The PCoA axes included in the different farm-like microbiota analysis models are marked with orange. We selected axes above the point at which the variance explained by the additional axes levels off. In borderline cases four axes were selected for consistency.

# Supplementary Tables

**Supplementary Table 1.** Bacterial/archaeal taxa with significantly (two-sided corrected  $P$ -value<0.05) **(a)** higher or **(b)** lower relative abundance in the living room floor dust of the farm ( $n=107$ ) than non-farm rural homes ( $n=96$ ) in LUKAS1 as defined by ANCOM. Taxa in the tables are sorted first by taxonomic level, second by absolute difference in median proportions, or, where that is zero, by absolute difference in mean log abundance (number of sequences).

**a.)** Bacterial/archaeal taxa with higher relative abundance in the farm than non-farm rural homes in LUKAS1.

Taxon						Median Proportion (%)		Prevalence (%)	
Phylum	Class	Order	Family	Genus	Non-farm home	Farm home	Non-farm home	Farm home	
Chloroflexi					0.16	0.42	99	100	
Verrucomicrobia					0.06	0.12	95	100	
Euryarchaeota					0	0.04	19	91	
Spirochaetes					0	0.02	38	74	
Chloroflexi	Thermomicrobia				0.07	0.27	98	100	
Euryarchaeota	Methanobacteria				0	0.04	17	90	
Spirochaetes	Spirochaetes				0	0.02	38	74	
Verrucomicrobia	Verruco5				0	0	1	48	
Bacteroidetes	Sphingobacteriia	Sphingobacteriales			0.80	1.71	100	100	
Proteobacteria	Gammaproteobacteria	Xanthomonadales			0.43	0.82	100	100	
Chloroflexi	Thermomicrobia	JG30KFCM45			0.07	0.26	98	100	
Firmicutes	Bacilli	Turicibacterales			0.01	0.11	62	99	
Proteobacteria	Gammaproteobacteria	Alteromonadales			0.02	0.12	75	98	
Proteobacteria	Gammaproteobacteria	Oceanospirillales			0	0.04	29	83	
Euryarchaeota	Methanobacteria	Methanobacteriales			0	0.04	17	90	
Proteobacteria	Gammaproteobacteria	Aeromonadales			0	0.03	67	87	
Proteobacteria	Deltaproteobacteria	Desulfovibrionales			0	0.02	34	69	
Tenericutes	Mollicutes	RF39			0	0.01	22	65	
Verrucomicrobia	Verruco5	WCHB141			0	0	1	48	
Euryarchaeota	Methanomicrobia	Methanomicrobiales			0	0	1	44	
Firmicutes	Bacilli	Lactobacillales	Aerococcaceae		0.37	1.44	100	100	
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae		0.44	1.51	100	100	
Actinobacteria	Actinobacteria	Actinomycetales	Dermabacteraceae		0.25	0.96	100	100	
Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae		0.45	1.08	100	100	
Actinobacteria	Actinobacteria	Actinomycetales	Brevibacteriaceae		0.13	0.67	99	100	
Firmicutes	Clostridia	Clostridiales	Unassigned		0.19	0.66	99	100	
Firmicutes	Bacilli	Bacillales	Planococcaceae		0.20	0.52	100	100	
Firmicutes	Clostridia	Clostridiales	Clostridiaceae		0.25	0.55	100	100	
Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae		0.18	0.46	100	100	
Bacteroidetes	Bacteroidia	Bacteroidales	Unassigned		0.00	0.24	28	95	
Chloroflexi	Thermomicrobia	JG30KFCM45	Unassigned		0.07	0.26	98	100	
Actinobacteria	Actinobacteria	Actinomycetales	Dietziaceae		0.08	0.26	99	99	
Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae		0.04	0.21	85	100	
Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae		0.07	0.22	94	100	
Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae		0.10	0.25	98	100	
Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae		0.01	0.15	71	93	
Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae		0.08	0.21	98	100	
Bacteroidetes	Bacteroidia	Bacteroidales	RF16		0	0.13	13	88	
Firmicutes	Bacilli	Bacillales	Bacillaceae		0.10	0.23	100	100	
Firmicutes	Bacilli	Turicibacterales	Turicibacteraceae		0.01	0.11	62	99	
Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae		0.01	0.08	65	97	
Firmicutes	Clostridia	Clostridiales	Other		0.01	0.06	53	89	
Firmicutes	Clostridia	Clostridiales	Mogibacteriaceae		0.01	0.05	62	87	
Actinobacteria	Actinobacteria	Actinomycetales	Streptomycetaceae		0.02	0.06	87	97	

Taxon					Median Proportion (%)		Prevalence (%)	
Phylum	Class	Order	Family	Genus	Non-farm home	Farm home	Non-farm home	Farm home
Firmicutes	Bacilli	Bacillales	Unassigned		0.01	0.05	66	95
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae		0	0.04	29	83
Actinobacteria	Actinobacteria	Actinomycetales	Yaniellaceae		0	0.04	14	87
Actinobacteria	Actinobacteria	Actinomycetales	Nocardiopsaceae		0	0.04	48	81
Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae		0	0.04	17	90
Bacteroidetes	Bacteroidia	Bacteroidales	S247		0	0.03	33	85
Actinobacteria	Actinobacteria	Actinomycetales	Beutenbergiaceae		0	0.02	39	84
Firmicutes	Bacilli	Bacillales	Paenibacillaceae		0	0.02	49	79
Bacteroidetes	Bacteroidia	Bacteroidales	BS11		0	0.01	5	69
Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae		0	0.01	20	67
Tenericutes	Mollicutes	RF39	Unassigned		0	0.01	22	65
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae		0	0.01	17	65
Firmicutes	Bacilli	Bacillales	Thermoactinomycetaceae		0	0.01	32	70
Actinobacteria	Actinobacteria	Actinomycetales	Bogoriellaceae		0	0.01	18	72
Proteobacteria	Gammaproteobacteria	Alteromonadales	Idiomarinaceae		0	0.01	7	54
Bacteroidetes	Bacteroidia	Bacteroidales	p253418B5		0	0.01	3	52
Verrucomicrobia	Verruco5	WCHB141	RFP12		0	0	1	48
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Unassigned	0.13	0.99	99	100
Actinobacteria	Actinobacteria	Actinomycetales	Dermabacteraceae	<i>Brachy bacterium</i>	0.13	0.77	98	100
Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Unassigned	0.42	1.01	100	100
Actinobacteria	Actinobacteria	Actinomycetales	Brevibacteriaceae	<i>Brevibacterium</i>	0.13	0.67	99	100
Firmicutes	Clostridia	Clostridiales	Unassigned	Unassigned	0.19	0.66	99	100
Firmicutes	Bacilli	Bacillales	Staphylococcaceae	<i>Jeotgalicoccus</i>	0.10	0.52	99	100
Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	<i>Facklamia</i>	0.06	0.46	97	100
Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	<i>Arthrobacter</i>	0.08	0.46	99	100
Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	Unassigned	0.00	0.38	47	100
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<i>Erwinia</i>	0.16	0.48	99	100
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	<i>Psychrobacter</i>	0.01	0.32	69	100
Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	<i>Aerococcus</i>	0.07	0.35	94	99
Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Unassigned	0.12	0.37	100	100
Bacteroidetes	Bacteroidia	Bacteroidales	Unassigned	Unassigned	0.00	0.24	28	95
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Unassigned	0.07	0.28	95	100
Firmicutes	Clostridia	Clostridiales	Clostridiaceae	<i>Clostridium</i>	0.14	0.34	100	100
Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	Other	0.16	0.35	98	100
Chloroflexi	Thermomicrobia	JG30KFCM45	Unassigned	Unassigned	0.07	0.26	98	100
Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae	Unassigned	0.02	0.19	70	100
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	<i>57N15</i>	0.00	0.17	18	94
Actinobacteria	Actinobacteria	Actinomycetales	Dietziaceae	<i>Dietzia</i>	0.04	0.21	94	98
Firmicutes	Bacilli	Bacillales	Planococcaceae	<i>Planomicrobium</i>	0.02	0.17	75	98
Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	<i>Carnobacterium</i>	0.03	0.18	92	99
Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	<i>Devosia</i>	0.09	0.23	97	100
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	<i>Sphingobacterium</i>	0.08	0.22	93	99
Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Unassigned	0.01	0.15	71	93
Bacteroidetes	Bacteroidia	Bacteroidales	RF16	Unassigned	0.00	0.13	13	88
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	<i>Mycoplasma</i>	0.11	0.23	97	100
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Unassigned	0.04	0.16	89	100
Actinobacteria	Actinobacteria	Actinomycetales	Nocardiaceae	<i>Rhodococcus</i>	0.07	0.18	97	100
Firmicutes	Bacilli	Turicibacterales	Turicibacteraceae	<i>Turicibacter</i>	0.01	0.11	62	99
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Unassigned	0.01	0.10	54	95
Firmicutes	Bacilli	Bacillales	Bacillaceae	<i>Bacillus</i>	0.06	0.15	97	100
Firmicutes	Bacilli	Bacillales	Staphylococcaceae	<i>Salinicoccus</i>	0.02	0.11	79	95
Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	<i>Luteimonas</i>	0.02	0.10	76	100
Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	<i>Trichococcus</i>	0.00	0.08	27	98
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	<i>Dorea</i>	0.03	0.11	84	99
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Unassigned	0.00	0.07	10	86
Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	<i>Leucobacter</i>	0.01	0.07	75	97
Bacteroidetes	Bacteroidia	Bacteroidales	Paraprevotellaceae	<i>CF231</i>	0.00	0.06	10	90
Firmicutes	Clostridia	Clostridiales	Other	Other	0.01	0.06	53	89
Firmicutes	Bacilli	Bacillales	Planococcaceae	Other	0.01	0.06	60	93
Firmicutes	Bacilli	Lactobacillales	Leuconostocaceae	<i>Unassigned</i>	0.00	0.05	44	93
Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	<i>Lysobacter</i>	0.01	0.07	68	94
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	<i>Oscillospira</i>	0.01	0.06	71	94
Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	<i>Cellvibrio</i>	0.01	0.05	64	94
Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	Other	0.01	0.06	76	94

Taxon					Median Proportion (%)		Prevalence (%)	
Phylum	Class	Order	Family	Genus	Non-farm home	Farm home	Non-farm home	Farm home
Actinobacteria	Actinobacteria	Actinomycetales	Streptomycetaceae	<i>Streptomyces</i>	0.02	0.06	87	96
Firmicutes	Bacilli	Bacillales	Unassigned	<i>Unassigned</i>	0.01	0.05	66	95
Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae	<i>Unassigned</i>	0.02	0.05	84	97
Actinobacteria	Actinobacteria	Actinomycetales	Yaniellaceae	<i>Yaniella</i>	0.00	0.04	14	87
Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	<i>Other</i>	0.01	0.04	60	94
Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae	<i>Methanobrevibacter</i>	0.00	0.04	17	90
Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	<i>Oerskovia</i>	0.00	0.03	41	90
Firmicutes	Bacilli	Bacillales	Planococcaceae	<i>Sporosarcina</i>	0.01	0.04	74	94
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	<i>Butyrivibrio</i>	0.00	0.03	5	79
Bacteroidetes	Bacteroidia	Bacteroidales	S247	<i>Unassigned</i>	0.00	0.03	33	85
Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	<i>Unassigned</i>	0.00	0.03	47	82
Firmicutes	Clostridia	Clostridiales	Mogibacteriaceae	<i>Unassigned</i>	0.01	0.04	55	80
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Erythrobacteraceae	<i>Unassigned</i>	0.02	0.05	75	87
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	<i>Comamonas</i>	0.01	0.04	64	92
Firmicutes	Clostridia	Clostridiales	Veillonellaceae	<i>Phascolarctobacterium</i>	0.00	0.03	29	87
Bacteroidetes	Flavobacteriia	Flavobacteriales	Weeksellaceae	<i>Wautersiella</i>	0.00	0.03	37	80
Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	<i>Knoellia</i>	0.01	0.04	72	88
Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	<i>Paludibacter</i>	0.00	0.03	30	76
Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	<i>Microbispora</i>	0.01	0.03	52	89
Actinobacteria	Actinobacteria	Actinomycetales	Beutenbergiaceae	<i>Salana</i>	0.00	0.02	39	84
Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	<i>Unassigned</i>	0.00	0.02	3	80
Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	<i>Desemzia</i>	0.00	0.02	39	80
Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	<i>Akkermansia</i>	0.00	0.02	39	77
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	<i>Myroides</i>	0.00	0.02	17	75
Actinobacteria	Actinobacteria	Actinomycetales	Nocardioseae	<i>Prauseria</i>	0.00	0.02	28	74
Actinobacteria	Actinobacteria	Actinomycetales	Intrasporangiaceae	<i>Janibacter</i>	0.00	0.02	37	75
Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	<i>Unassigned</i>	0.00	0.02	15	79
Actinobacteria	Actinobacteria	Actinomycetales	Pseudonocardiaceae	<i>Saccharopolyspora</i>	0.00	0.02	23	69
Proteobacteria	Epsilonproteobacteria	Campylobacteriales	Campylobacteriaceae	<i>Arcobacter</i>	0.00	0.02	12	72
Firmicutes	Bacilli	Bacillales	Planococcaceae	<i>Rummeliibacillus</i>	0.00	0.02	17	79
Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	<i>Other</i>	0.01	0.02	63	90
Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	<i>Oligella</i>	0.00	0.02	13	69
Firmicutes	Bacilli	Bacillales	Planococcaceae	<i>Solibacillus</i>	0.00	0.02	18	74
Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	<i>Pediococcus</i>	0.00	0.02	23	74
Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	<i>Unassigned</i>	0.00	0.02	32	69
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	<i>CandidatusPortiera</i>	0.00	0.02	3	69
Bacteroidetes	Bacteroidia	Bacteroidales	BS11	<i>Unassigned</i>	0.00	0.01	5	69
Firmicutes	Clostridia	Clostridiales	Mogibacteriaceae	<i>Mogibacterium</i>	0.00	0.01	28	75
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	<i>Halomonas</i>	0.00	0.01	15	63
Tenericutes	Mollicutes	RF39	Unassigned	<i>Unassigned</i>	0.00	0.01	22	65
Actinobacteria	Actinobacteria	Actinomycetales	Cellulomonadaceae	<i>Actinotalea</i>	0.00	0.01	23	65
Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	<i>Unassigned</i>	0.00	0.01	1	62
Actinobacteria	Actinobacteria	Actinomycetales	Bogoriellaceae	<i>Georgenia</i>	0.00	0.01	18	72
Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	<i>Trueperella</i>	0.00	0.01	8	62
Bacteroidetes	Bacteroidia	Bacteroidales	Paraprevotellaceae	<i>YRC22</i>	0.00	0.01	4	63
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	<i>Fluviicola</i>	0.00	0.01	15	61
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	<i>Alkanindiges</i>	0.00	0.01	17	61
Bacteroidetes	Flavobacteriia	Flavobacteriales	Weeksellaceae	<i>Weeksella</i>	0.00	0.01	6	57
Bacteroidetes	Bacteroidia	Bacteroidales	p253418B5	<i>Unassigned</i>	0.00	0.01	3	52
Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	<i>Erysipelothrix</i>	0.00	0.00	6	50
Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	<i>CandidatusEndobugula</i>	0.00	0.00	4	52
Verrucomicrobia	Verruco5	WCHB141	RFP12	<i>Unassigned</i>	0.00	0.00	1	48
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	<i>Aequorivita</i>	0.00	0.00	2	51
Firmicutes	Clostridia	Clostridiales	Clostridiaceae	<i>Proteiniclasticum</i>	0.00	0.00	3	51
Proteobacteria	Gammaproteobacteria	Alteromonadales	Idiomarinaceae	<i>Unassigned</i>	0.00	0.00	4	48
Proteobacteria	Gammaproteobacteria	Alteromonadales	Chromatiaceae	<i>Unassigned</i>	0.00	0.00	2	44

**b.) Bacterial taxa with lower relative abundance in the farm than non-farm rural homes in LUKAS1.**

Taxon					Median Proportion (%)		Prevalence (%)	
					Non-farm home	Farm home	Farm home	Non-farm home
Phylum	Class	Order	Family	Genus				
Fusobacteria					0.45	0.25	99	99
Fusobacteria	Fusobacteriia				0.45	0.25	99	99
					0	0		
Proteobacteria	Gammaproteobacteria	Pasteurellales			1.21	0.59	100	100
Proteobacteria	Betaproteobacteria	Neisseriales			0.57	0.34	99	100
Fusobacteria	Fusobacteriia	Fusobacteriales			0.45	0.25	99	99
Firmicutes	Bacilli	Gemellales			0.21	0.12	100	99
Proteobacteria	Betaproteobacteria	Methylophilales			0.25	0.16	100	100
Proteobacteria	Gammaproteobacteria	Cardiobacteriales			0.02	0.01	58	76
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae		11.36	6.08	100	100
Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae		1.21	0.59	100	100
Firmicutes	Clostridia	Clostridiales	Tissierellaceae		0.87	0.54	100	100
Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae		0.57	0.34	99	100
Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae		0.20	0.07	95	98
Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae		0.24	0.12	96	97
Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae		0.25	0.16	100	100
Firmicutes	Bacilli	Gemellales	Gemellaceae		0.20	0.12	100	99
Proteobacteria	Gammaproteobacteria	Cardiobacteriales	Cardiobacteriaceae		0.02	0.01	58	76
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	<i>Streptococcus</i>	5.03	2.33	100	100
Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	<i>Lactococcus</i>	4.89	2.51	100	100
Firmicutes	Bacilli	Bacillales	Staphylococcaceae	<i>Staphylococcus</i>	5.14	4.13	100	100
Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	<i>Micrococcus</i>	1.44	1.02	100	100
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	<i>Bacteroides</i>	0.74	0.34	95	100
Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	<i>Haemophilus</i>	0.65	0.36	99	100
Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	<i>Porphyromonas</i>	0.52	0.25	99	99
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	<i>Moraxella</i>	0.29	0.10	93	93
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Other	0.76	0.59	100	100
Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	<i>Propionibacterium</i>	0.19	0.04	84	96
Firmicutes	Clostridia	Clostridiales	Tissierellaceae	<i>Anaerococcus</i>	0.36	0.23	100	99
Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	<i>Fusobacterium</i>	0.24	0.12	96	97
Actinobacteria	Actinobacteria	Actinomycetales	Actinomycetaceae	<i>Actinomyces</i>	0.31	0.20	99	99
Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Unassigned	0.25	0.15	100	100
Firmicutes	Clostridia	Clostridiales	Veillonellaceae	<i>Veillonella</i>	0.21	0.12	97	100
Firmicutes	Bacilli	Gemellales	Gemellaceae	Unassigned	0.18	0.11	98	99
Firmicutes	Clostridia	Clostridiales	Tissierellaceae	<i>Peptoniphilus</i>	0.16	0.10	97	99
Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	<i>Granulicatella</i>	0.11	0.05	89	97
Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	<i>Rothia</i>	0.17	0.12	98	99
Firmicutes	Clostridia	Clostridiales	Tissierellaceae	<i>Fingoldia</i>	0.18	0.13	98	100
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	<i>Capnocytophaga</i>	0.11	0.06	92	95
Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	Unassigned	0.09	0.05	88	96
Proteobacteria	Epsilonproteobacteria	Campylobacteriales	Campylobacteraceae	<i>Campylobacter</i>	0.05	0.02	82	93
Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	<i>Lautropia</i>	0.08	0.05	91	98
Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	<i>Ralstonia</i>	0.03	0.02	84	91
Firmicutes	Clostridia	Clostridiales	Peptostreptococcaceae	<i>Peptostreptococcus</i>	0.01	0.01	60	70
Bacteroidetes	Saprospirae	Saprospirales	Chitinophagaceae	<i>Sediminibacterium</i>	0.03	0.02	85	89

**Supplementary Table 2. Oligotyping analysis of *Methanobrevibacter* genus.**<sup>a</sup> The final number of quality controlled oligotypes within *Methanobrevibacter* genus was 7 of which 6 had the best match with a typical rumen associated species and one with human gut commensal. Within the LUKAS1 non-farm homes where *Methanobrevibacter* was detected (15/96; 15.6%), *M. smithii* oligotype was found and was also the dominant oligotype in 53% (8/15) of homes while in farm homes where *Methanobrevibacter* was detected (93/107; 86.9%), the same oligotype was found in 6% (6/93) and was the dominant oligotype only in 5% (5/93) of the homes (Fisher's exact test, two-sided  $P$ -value =  $1.7 \times 10^{-5}$ ).

Reads per oligotype	Best match	Identity	E-value	Typical source <sup>ref 12</sup>
209	<i>M.olleyae</i>	99%	4e-129	Rumen
187	<i>M.millerae</i>	99%	2e-127	Rumen
164	<i>M.millerae</i>	100%	2e-132	Rumen
142	<i>M.millerae</i>	99%	9e-131	Rumen
139	<i>M.millerae</i> / <i>M.thaueri</i>	99%	4e-129	Rumen
89	<i>M.smithii</i>	100%	2e-132	Human gut
38	<i>M.olleyae</i>	99%	2e-127	Rumen

<sup>a</sup>The oligotype analysis had total purity score of 0.92.

**Supplementary Table 3.** Fungal taxa with significantly (two-sided corrected  $P$ -value<0.05) **(a)** higher or **(b)** lower relative abundance in the living room floor dust of the farm ( $n=101$ ) than non-farm ( $n=97$ ) rural homes in LUKAS1 as defined by ANCOM. Taxa in the tables are sorted first by taxonomic level, second by absolute difference in median proportions, or, where that is zero, by absolute difference in mean log abundance (number of sequences).

**a.) Fungal taxa with higher relative abundance in the farm than non-farm rural homes in LUKAS1.**

Taxon					Median proportion (%)		Prevalence (%)	
Phylum	Class	Order	Family	Genus	Non-farm home	Farm home	Non-farm home	Farm home
unidentified	unidentified				0	0	12	46
Zygomycota	unidentified				0	0	4	27
Ascomycota	Sordariomycetes	Microascales			0.04	0.26	83	99
Ascomycota	Leotiomycetes	Thelebolales			0.01	0.05	63	88
Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae		0.53	1.07	100	100
Ascomycota	Sordariomycetes	Microascales	Microasceae		0.03	0.24	78	99
Ascomycota	Leotiomycetes	Thelebolales	Other		0.01	0.04	62	87
Ascomycota	Lecanoromycetes	Lecanorales	Lecanoraceae		0.01	0.03	56	80
Ascomycota	Sordariomycetes	Hypocreales	Clavicipitaceae		0	0.02	48	71
Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae		0	0.01	41	66
Ascomycota	Sordariomycetes	Uncertaeedis	Apiosporaceae		0	0.01	32	60
Ascomycota	Eurotiomycetes	Eurotiales	Uncertaeedis		0	0	13	49
unidentified	unidentified	unidentified	unidentified		0	0	12	46
Ascomycota	Lecanoromycetes	Teloschistales	Teloschistaceae		0	0	18	48
Ascomycota	Eurotiomycetes	Onygenales	Onygenaceae		0	0	4	33
Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	<i>Aspergillus</i>	1.37	5.73	100	100
Ascomycota	Sordariomycetes	Hypocreales	Uncertaeedis	<i>Acremonium</i>	0.28	0.47	96	100
Ascomycota	Sordariomycetes	Hypocreales	Nectriaceae	<i>Gibberella</i>	0.08	0.25	92	92
Ascomycota	Sordariomycetes	Microascales	Microasceae	unidentified	0.01	0.15	58	98
Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	Other	0.14	0.25	99	100
Basidiomycota	Microbotryomycetes	Sporidiobolales	Uncertaeedis	<i>Sporobolomyces</i>	0.05	0.16	98	98
Ascomycota	Sordariomycetes	Hypocreales	Uncertaeedis	<i>Sarocladium</i>	0.02	0.08	66	94
Basidiomycota	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	Other	0.01	0.07	62	92
Basidiomycota	Tremellomycetes	Cystofilobasidiales	Cystofilobasidiaceae	<i>Udeniomyces</i>	0.05	0.09	85	97
Ascomycota	Leotiomycetes	Thelebolales	Other	Other	0.01	0.04	62	87
Ascomycota	Sordariomycetes	Sordariales	Chaetomiaceae	<i>Humicola</i>	0	0.03	45	75
Ascomycota	Sordariomycetes	Sordariales	Lasiophaeriaceae	Other	0	0.02	44	78
Ascomycota	Dothideomycetes	Capnodiales	Uncertaeedis	<i>Pseudotaeniolina</i>	0	0.02	30	69
Ascomycota	Lecanoromycetes	Lecanorales	Lecanoraceae	unidentified	0	0.02	46	77
Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	<i>Bipolaris</i>	0	0.02	23	66
Ascomycota	Sordariomycetes	Hypocreales	Clavicipitaceae	<i>Claviceps</i>	0	0.01	46	70
Ascomycota	Sordariomycetes	Xylariales	Uncertaeedis	<i>Microdochium</i>	0	0.01	23	66
Ascomycota	Sordariomycetes	Microascales	Microasceae	<i>Wardomyces</i>	0	0.01	22	66
Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	<i>Dissoconium</i>	0	0.01	32	64
Ascomycota	Sordariomycetes	Microascales	Microasceae	Other	0	0.01	35	64
Ascomycota	Sordariomycetes	Uncertaeedis	Apiosporaceae	<i>Arthrinium</i>	0	0.01	32	60
Ascomycota	Sordariomycetes	Sordariales	Lasiophaeriaceae	unidentified	0	0.01	12	55
Ascomycota	Saccharomycetes	Saccharomycetales	Uncertaeedis	<i>Wickerhamomyces</i>	0	0	14	51
Ascomycota	Eurotiomycetes	Eurotiales	Trichocomaceae	<i>Paecilomyces</i>	0	0	5	51
Ascomycota	Eurotiomycetes	Eurotiales	Uncertaeedis	<i>Thermomyces</i>	0	0	13	49
Ascomycota	Dothideomycetes	Pleosporales	Sporormiaceae	<i>Preussia</i>	0	0	10	45
Ascomycota	Lecanoromycetes	Teloschistales	Teloschistaceae	Other	0	0	6	38
unidentified	unidentified	unidentified	unidentified	unidentified	0	0	12	46
Ascomycota	Eurotiomycetes	Onygenales	Onygenaceae	<i>Chrysosporium</i>	0	0	3	29

**b.) Fungal taxa with lower relative abundance in the farm than non-farm rural homes in LUKAS1.**

Taxon					Median proportion (%)		Prevalence (%)	
Phylum	Class	Order	Family	Genus	Non-farm home	Farm home	Non-farm home	Farm home
Other	Other				0.81	0.41	100	100
Basidiomycota	Other				0.30	0.12	98	98
Ascomycota	Taphrinomycetes				0.23	0.14	100	98
Ascomycota	Eurotiomycetes	Chaetothyriales			1.37	0.66	100	100
Ascomycota	Dothideomycetes	Other			0.83	0.31	100	100
Other	Other	Other			0.81	0.41	100	100
Ascomycota	Dothideomycetes	Dothideales			0.52	0.28	100	100
Basidiomycota	Other	Other			0.30	0.12	98	98
Ascomycota	Dothideomycetes	Botryosphaerales			0.26	0.08	96	95
Ascomycota	Leotiomycetes	Other			0.28	0.11	98	97
Basidiomycota	Agaricomycetes	Agaricales			0.08	0.03	94	94
Ascomycota	Dothideomycetes	Capnodiales	Other		1.65	0.85	100	98
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae		1.06	0.44	100	100
Ascomycota	Dothideomycetes	Other	Other		0.83	0.31	100	100
Other	Other	Other	Other		0.81	0.41	100	100
Ascomycota	Dothideomycetes	Pleosporales	Montagnulaceae		0.23	0.04	90	82
Basidiomycota	Other	Other	Other		0.30	0.12	98	98
Ascomycota	Leotiomycetes	Other	Other		0.28	0.11	98	97
Ascomycota	Dothideomycetes	Pleosporales	Venturiaceae		0.22	0.08	93	97
Ascomycota	Dothideomycetes	Pleosporales	Cucurbitariaceae		0.18	0.06	92	80
Ascomycota	Taphrinomycetes	Taphrinales	Taphrinaceae		0.20	0.12	100	98
Ascomycota	Dothideomycetes	Dothideales	Other		0.08	0.01	91	68
Ascomycota	Dothideomycetes	Botryosphaerales	unidentified		0.03	0.01	82	67
Ascomycota	Lecanoromycetes	Other	Other		0.02	0.01	77	63
Basidiomycota	Tremellomycetes	Cystofilobasidiales	unidentified		0.01	0.00	73	35
Ascomycota	Dothideomycetes	Capnodiales	Other	Other	1.65	0.85	100	98
Ascomycota	Dothideomycetes	Other	Other	Other	0.83	0.31	100	100
Other	Other	Other	Other	Other	0.81	0.41	100	100
Ascomycota	Dothideomycetes	Pleosporales	Montagnulaceae	Other	0.23	0.04	90	82
Basidiomycota	Other	Other	Other	Other	0.30	0.12	98	98
Ascomycota	Leotiomycetes	Other	Other	Other	0.28	0.11	98	97
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	<i>Aureobasidium</i>	0.27	0.14	100	100
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	<i>Phaeococcomyces</i>	0.22	0.09	98	97
Ascomycota	Dothideomycetes	Pleosporales	Cucurbitariaceae	<i>Pyrenochaetopsis</i>	0.18	0.06	92	80
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	<i>Exophiala</i>	0.18	0.06	100	97
Ascomycota	Dothideomycetes	Botryosphaerales	Botryosphaeriaceae	<i>Dothiorella</i>	0.15	0.04	81	82
Ascomycota	Dothideomycetes	Pleosporales	Venturiaceae	Other	0.16	0.06	89	89
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	<i>Cladophialophor</i>	0.14	0.05	98	93
Basidiomycota	Microbotryomycete	Other	Other	Other	0.10	0.03	81	72
Ascomycota	Dothideomycetes	Dothideales	Other	Other	0.08	0.01	91	68
Ascomycota	Taphrinomycetes	Taphrinales	Taphrinaceae	<i>Taphrina</i>	0.10	0.04	94	90
Ascomycota	Dothideomycetes	Capnodiales	Mycosphaerellaceae	Other	0.08	0.03	87	85
Ascomycota	Dothideomycetes	Pleosporales	Venturiaceae	<i>Venturia</i>	0.05	0.02	82	82
Ascomycota	Eurotiomycetes	Chaetothyriales	Herpotrichiellaceae	unidentified	0.10	0.07	97	94
Ascomycota	Leotiomycetes	Helotiales	Incertaesedis	<i>Cadophora</i>	0.06	0.03	83	79
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	unidentified	0.03	0.00	91	61
Ascomycota	Leotiomycetes	Helotiales	Phacidiaceae	<i>Phacidium</i>	0.03	0.01	77	63
Ascomycota	Dothideomycetes	Botryosphaerales	unidentified	unidentified	0.03	0.01	82	67
Ascomycota	Lecanoromycetes	Other	Other	Other	0.02	0.01	77	63
Basidiomycota	Tremellomycetes	Cystofilobasidiales	unidentified	unidentified	0.01	0.00	73	35
Ascomycota	Taphrinomycetes	Taphrinales	Taphrinaceae	Other	0.01	0.00	68	46
Ascomycota	Saccharomycetes	Saccharomycetale	Incertaesedis	<i>Cyberlindnera</i>	0.00	0.00	46	28



**Supplementary Table 4.** The modelled effect of farm home -like indoor microbiota composition at 2 months and asthma ever by 6 years of age or active asthma at 6 years of age.

Predicted cohort	Type of farm home-like microbiota composition	IP <sup>a</sup>	N-total %-asthma ever		Asthma ever		N-total %-asthma active		Active asthma	
			n	%	aOR <sup>b</sup> (95% CI)	P-value	n	%	aOR (95% CI)	P-value
LUKAS2 non-farm	Bacterial <sup>c</sup> presence-absence	0.01	164	19.5	0.98 (0.94 - 1.02)	0.22	156	12.2	0.98 (0.94 - 1.03)	0.44
	Bacterial relative abundance	0.35	164	19.5	<b>0.40 (0.19 - 0.82)</b>	<b>0.01</b>	156	12.2	0.63 (0.27 - 1.44)	0.27
	Fungal presence-absence	0.03	157	19.1	0.99 (0.92 - 1.07)	0.75	149	11.4	1.05 (0.96 - 1.14)	0.28
	Fungal relative abundance	0.29	157	19.1	1.01 (0.55- 1.90)	0.96	149	11.4	1.28 (0.53 - 3.09)	0.58
LUKAS 1 & 2 non-farm	Bacterial presence-absence	0.01	251	19.1	0.96 (0.92 - 1.01)	0.12	227	11.9	0.98 (0.93 - 1.03)	0.39
	Bacterial relative abundance	0.34	251	19.1	<b>0.47 (0.27 - 0.81)</b>	<b>0.007</b>	227	11.9	<b>0.48 (0.23 - 0.98)</b>	<b>0.04</b>
	Fungal presence-absence	0.03	244	18.9	0.98 (0.93 - 1.04)	0.55	220	11.4	1.01 (0.95 - 1.08)	0.68
	Fungal relative abundance	0.28	244	18.9	0.87 (0.55 - 1.38)	0.55	220	11.4	1.19 (0.63 - 2.27)	0.59
LUKAS 1 & 2 farm <sup>d,e</sup>	Bacterial presence-absence	I <sup>f</sup>	37	10.8	1		33	9.1		
		II	39	5.1	0.42 (0.07 - 2.55)	0.51	36	0	- <sup>g</sup>	
		III	35	11.4	1.23 (0.26 - 5.74)		32	9.4		
	Bacterial relative abundance <sup>h</sup>	2.74	111	9.0	0.37 (0.10 - 1.33)	0.13 <sup>i</sup>	101	5.9	0.38 (0.11-1.30)	0.12 <sup>e,j</sup>
	Fungal presence-absence	0.08	105	9.5	0.91 (0.71-1.16)	0.44	96	6.3	0.88 (0.72 – 1.13)	0.29 <sup>e</sup>
	Fungal relative abundance <sup>h</sup>	0.62	105	9.5	1.22 (0.38 – 3.90)	0.74	96	6.3	1.81 (0.46 – 8.32)	0.35 <sup>e</sup>
LUKAS 1 & 2	Bacterial relative abundance	0.40	362	16.0	<b>0.47 (0.27 - 0.83)</b>	<b>0.009</b>	328	10.1	<b>0.48 (0.23 - 0.99)</b>	<b>0.046</b>

<sup>a</sup>Logistic regression analysis based adjusted odds ratios (aOR) presented per interquartile range (IQR) or in reference to the first tertile of predicted individual probability (IP) that the microbiota in the child's home, predicted based on microbial community beta-diversity based modelling, represents more of a LUKAS1 farm than non-farm home. <sup>b</sup>Unless specifically marked, all models adjusted for sex, maternal and paternal allergic disease, maternal smoking during pregnancy, number of older siblings and maternal education as well as cohort and living at a farm where relevant. *P* - values are two-sided. <sup>c</sup>Bacterial = bacterial/archaeal. <sup>d</sup>These models were adjusted with maternal smoking during pregnancy as two category variable (yes/no) due to model instabilities with the three category variable accounting for cessation of smoking prior to pregnancy. <sup>e</sup>Where marked, Firth's penalized estimation was used, if maximum likelihood estimate did not exist. <sup>f</sup>Continuous IP variable could not be used because the association with asthma was not linear. Comparison between tertiles with the first tertile as the reference (odds ratio = 1). <sup>g</sup>Unable to calculate reliable adjusted or unadjusted estimates due to quasi-complete separation of data points. <sup>h</sup>IP value was exponentially transformed because the variable was strongly skewed to the left but transferable to normal distribution. <sup>i</sup>The asthma predisposing trend of non-farm like microbiota among farmers is statistically significant when analyzed as a dichotomous variable (FaRMI below vs above median in LUKAS1 and 2 farm homes); With Firth's penalized maximum likelihood estimation aOR for asthma ever 6.93 (1.51 – 62.3); *p*=0.02; <sup>j</sup>for active asthma this could not be estimated due to low number of cases.

**Supplementary Table 5.** Independence of the association between asthma and FaRMI (farm home - like relative abundance of bacteria/archaea) in non-farm LUKAS homes on proxies of total bacterial exposure and potential confounders, including environmental determinants.

Adjustments	<i>n</i> -total %-asthma ever		Effect of FaRMI on Asthma ever by 6 years	
	<i>n</i>	%	aOR <sup>a</sup> (95% CI) per IQR <sub>FaRMI</sub>	<i>P</i> -value
None	260	18.8	0.51 (0.31 – 0.86)	0.011
Basic adjustments <sup>b</sup>	251	19.1	0.47 (0.27 - 0.81)	0.007
<b>The basic adjustments + adjustment for:<sup>c</sup></b>				
Atopic sensitization <sup>d</sup>	192	20.3	0.47 (0.25 - 0.89)	0.020
Amount of dust	251	19.1	0.52 (0.29 - 0.92)	0.024
Bacterial richness	251	19.1	0.43 (0.21 - 0.87)	0.018
Total bacterial concentration	251	19.1	0.45 (0.25 - 0.79)	0.005
LPS <sub>10-16</sub> / m <sup>2</sup>	246	19.5	0.49 (0.27 - 0.87)	0.015
Endotoxin / m <sup>2</sup>	226	18.6	0.45 (0.25 - 0.81)	0.008
Muramic acid / m <sup>2</sup>	236	17.8	0.56 (0.31 - 1.02)	0.059
Indoors with outdoor shoes <sup>e</sup>	164	19.5	0.40 (0.19 - 0.85)	0.017
Construction year	244	19.3	0.50 (0.28 - 0.88)	0.017
Condensation on windows	244	19.3	0.47 (0.27 - 0.83)	0.009
Fish tank	245	18.8	0.44 (0.25 - 0.78)	0.005

<sup>a</sup>Logistic regression analysis based adjusted odds ratios (aOR) presented per interquartile range of FaRMI (0.34). <sup>b</sup>Basic adjustments include sex, cohort, maternal and paternal allergic disease, maternal smoking during pregnancy, number of older siblings and maternal education and with the additional variables as indicated. All confounders tested are listed in the methods. *P*-values are two-sided. <sup>c</sup>aOR for model with all adjustments in the same model 0.03 (0.001 - 0.84); *P*=0.04, *n*=92, 15.2% with asthma. However, the model includes quasi-complete separation of data points and thus result is based on the last maximum likelihood iteration, unadjusted OR within this subpopulation 0.51 (0.20 – 1.27); *P*=0.15. <sup>d</sup>Atopic sensitization defined by presence of any specific IgE ≥ 3.5kU/L at 6 years. <sup>e</sup>Data collected from 1 year questionnaire from LUKAS2 only.

**Supplementary Table 6.** Determinants associated with farm home –like relative abundance of bacteria/archaea (FaRMI) in LUKAS non-farm homes.<sup>a</sup>

Determinant	Class	<i>n</i>	%-With FaRMI>50%	<i>P</i> -value <sup>b</sup>
<b>Siblings (older)</b>	None	103	34.0	<i>0.04</i>
	1	96	24.0	
	2 or more	79	41.8	
<b>Construction year</b>	Before 1970	58	44.8	<i>0.04</i>
	1970-1989	103	35.0	
	After 1989	100	25.0	
<b>Condensation on windows</b>	No	194	28.9	<i>0.01</i>
	Yes, interior	56	42.9	
	Yes, exterior	11	63.6	
<b>Having a fish tank</b>	No	253	30.4	<i>0.01</i>
	Yes	19	57.9	
<b>Walking indoors with outdoor shoes<sup>c</sup></b>	Almost never	128	28.1	<i>0.02</i>
	Sometimes /often	44	47.7	
	Missing (LUKAS1)	214		

<sup>a</sup>Full list of potential determinants tested listed in the methods section, <sup>b</sup>Two-sided *p*-value from a chi-square test, not corrected for multiple testing, <sup>c</sup>from 1 year questionnaire and LUKAS2 only.

**Supplementary Table 7.** The modelled effect of farm home -like indoor microbiota composition as defined by random forest scoring based Farm Resembling Microbiota Index (FaRMI) at 2 months and asthma ever by 6 years of age.

Predicted cohort	N-total %-asthma ever		IP <sup>a</sup>	Effect of random forest based FaRMI on Asthma ever by 6 years	
	<i>n</i>	%		aOR <sup>b</sup> (95% CI)	<i>P</i> -value
LUKAS2 non-farm	164	19.5	0.08	0.71 (0.12 -1.07)	0.0995
LUKAS 1 & 2 non-farm	251	19.1	0.01	0.72 (0.51 - 1.03)	0.07
LUKAS 1 & 2	362	16.0	0.40	0.78 (0.60 - 1.01)	0.06

<sup>a</sup>Logistic regression analysis based adjusted odds ratios (aOR) presented per interquartile range (IQR) of predicted individual probability (IP) that the microbiota in the child's home, predicted based on Random decision forest represents more of a LUKAS1 farm than non-farm home. P-values are two-sided. <sup>b</sup>All models adjusted for sex, maternal and paternal allergic disease, maternal smoking during pregnancy, number of older siblings and maternal education as well as cohort and living at a farm where relevant.

**Supplementary Table 8.** Thirty operational taxonomic units (OTUs) contributing most in the random decision forest based classification of the house dust microbiota compositions to farm or non-farm –like.

Importance	#OTU ID	Taxonomy
0.01993	4331090	p_Firmicutes; c_Bacilli; o_Lactobacillales; f_Aerococcaceae; g_; s_
0.01775	112158	p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Corynebacteriaceae;
0.01750	4404401	p_Firmicutes; c_Bacilli; o_Bacillales; f_Staphylococcaceae; g_Jeotgalicoccus;
0.01358	4317034	p_Bacteroidetes; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_5-7N15; s_
0.01253	517565	p_Firmicutes; c_Bacilli; o_Lactobacillales; f_Aerococcaceae; g_Facklamia; s_
0.01156	4474641	p_Firmicutes; c_Bacilli; o_Lactobacillales; f_Aerococcaceae; g_; s_
0.01143	295070	p_Firmicutes; c_Clostridia; o_Clostridiales; f_; g_; s_
0.01138	310510	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Clostridiaceae; g_Clostridium; s_
0.01130	722376	p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Corynebacteriaceae;
0.01083	347529	p_Firmicutes; c_Bacilli; o_Turicibacterales; f_Turicibacteraceae; g_Turicibacter; s_
0.01003	1505939	p_Firmicutes; c_Bacilli; o_Lactobacillales; f_Carnobacteriaceae; g_Trichococcus; s_
0.00994	1142110	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Peptostreptococcaceae; g_; s_
0.00981	277352	p_Proteobacteria; c_Gammaproteobacteria; o_Pseudomonadales; f_Moraxellaceae;
0.00889	290588	p_Firmicutes; c_Bacilli; o_Bacillales; f_Planococcaceae
0.00872	4339901	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Ruminococcaceae; g_; s_
0.00845	359533	p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Corynebacteriaceae;
0.00823	4324041	p_Bacteroidetes; c_Bacteroidia; o_Bacteroidales; f_RF16; g_; s_
0.00790	4306004	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Ruminococcaceae; g_; s_
0.00787	2459569	p_Firmicutes; c_Bacilli; o_Lactobacillales; f_Aerococcaceae; g_; s_
0.00758	310786	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Ruminococcaceae; g_Ruminococcus; s_
0.00746	4395530	p_Bacteroidetes; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_5-7N15; s_
0.00639	180060	p_Proteobacteria; c_Gammaproteobacteria; o_Pseudomonadales;
0.00636	348622	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Lachnospiraceae; g_Dorea; s_
0.00628	161687	p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Dermabacteraceae; g_Brachybacterium; s_conglomeratum
0.00627	720093	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Ruminococcaceae; g_; s_
0.00626	353194	p_Bacteroidetes; c_Bacteroidia; o_Bacteroidales; f_Rikenellaceae; g_; s_
0.00615	4482376	p_Firmicutes; c_Bacilli; o_Lactobacillales; f_Carnobacteriaceae; g_Trichococcus; s_
0.00605	4387469	p_Firmicutes; c_Clostridia; o_Clostridiales; f_Ruminococcaceae; g_; s_
0.00605	519367	p_Bacteroidetes; c_Flavobacteriia; o_Flavobacteriales; f_Flavobacteriaceae; g_; s_
0.00601	518065	p_Actinobacteria; c_Actinobacteria; o_Actinomycetales; f_Corynebacteriaceae;

p=phylum, c=class, o=order, f=family, g=genus, s=species

**Supplementary Table 9.** Generalized Unifrac distance of the bacterial microbiota in GABRIELA and LUKAS farm ( $n_{\text{GABRIELA}}=399$ ,  $n_{\text{LUKAS}}=116$ ) and non-farm homes ( $n_{\text{GABRIELA}}=632$ ,  $n_{\text{LUKAS}}=278$ ) from the microbiota in the GABRIELA animal shed dust samples ( $n=50$ ).

UniFrac distance metrics	Cohort	Living on a farm	Median	Interquartile range	p-value <sup>a</sup>
Unweighted <sup>b</sup>	GABRIELA	No	0.87	(0.86-0.88)	1.11E-87
		Yes	0.83	(0.81-0.86)	
	LUKAS	No	0.92	(0.91-0.93)	9.85E-29
		Yes	0.89	(0.87-0.90)	
Weighted <sup>c</sup>	GABRIELA	No	0.48	(0.45-0.52)	3.62E-32
		Yes	0.44	(0.41-0.48)	
	LUKAS	No	0.51	(0.47-0.55)	9.81E-12
		Yes	0.46	(0.44-0.50)	

<sup>a</sup>Continuity corrected two-sided p-value from Wilcoxon test comparing differences in UniFrac distance between animal shed microbiota and microbiota from farm or non-farm home. <sup>b</sup>Generalized UniFrac unweighted ( $\alpha=0$ ), or <sup>c</sup>weighted for relative abundance ( $\alpha=1$ ).

**Supplementary Table 10.** Selected taxa -based models of **a)** LUKAS1 (FaRMI) and **b)** GABRIELA (FaRMI<sub>GABRIELA</sub>) farm home -like bacterial relative abundance. Taxa were entered into the generalized linear models (GLM) based on significance level until minimum predicted residual error sum of squares (PRESS) was reached using 10-fold cross validation.<sup>ref 60</sup> The taxonomic level included in the model is bolded. The taxa are presented in the order they were included in the model. The few taxa explaining most of the variance in the models are highlighted with orange to indicate positive and with blue to indicated negative association with FaRMI.

**a.) Taxa best predicting the LUKAS1 farm home -like relative abundance defining FaRM-index.**

Step	Phylum	Class	Order	Family	Genus	Estimate	Partial adjusted R-square
0	<b>Intercept</b>	-	-	-	-	1.1E-01	0.000
1	Firmicutes	Bacilli	Lactobacillales	<b>Streptococcaceae</b>	-	-2.9E-03	0.530
2	<b>Actinobacteria</b>	-	-	-	-	3.4E-03	0.282
3	Bacteroidetes	<b>Sphingobacteriia</b>	-	-	-	5.4E-06	0.084
4	Proteobacteria	<b>Alphaproteobacteria</b>	-	-	-	8.3E-04	0.026
5	<b>Bacteroidetes</b>	-	-	-	-	1.7E-03	0.016
6	<b>Cyanobacteria</b>	-	-	-	-	8.9E-04	0.014
7	Proteobacteria	Alphaproteobacteria	Sphingomonadales	<b>Sphingomonadaceae</b>	-	1.1E-03	0.005
8	Bacteroidetes	Flavobacteriia	Flavobacteriales	[Weeksellaceae]	<b>Chryseobacterium</b>	3.8E-04	0.004
9	Firmicutes	Bacilli	Bacillales	<b>Planococcaceae</b>	-	6.6E-05	0.003
10	Firmicutes	<b>Bacilli</b>	-	-	-	-1.7E-03	0.002
11	Proteobacteria	Betaproteobacteria	<b>Rhodocyclales</b>	-	-	6.1E-04	0.002
12	<b>Unassigned</b>	-	-	-	-	-8.0E-04	0.001
13	Actinobacteria	Actinobacteria	Bifidobacteriales	Bifidobacteriaceae	<b>Gardnerella</b>	-4.4E-04	0.001
14	Cyanobacteria	Chloroplast	<b>Chlorophyta</b>	-	-	6.9E-04	0.001
15	Bacteroidetes	Bacteroidia	Bacteroidales	<b>Bacteroidaceae</b>	-	2.0E-04	0.001
16	Firmicutes	Bacilli	Bacillales	<b>Listeriaceae</b>	-	-7.4E-04	0.001
17	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	<b>Staphylococcus</b>	4.0E-04	<0.001
18	Proteobacteria	Deltaproteobacteria	Myxococcales	Myxococcaceae	<b>Coralloccoccus</b>	1.1E-03	<0.001
19	Bacteroidetes	Flavobacteriia	Flavobacteriales	[Weeksellaceae]	<b>Ornithobacterium</b>	9.3E-04	<0.001
20	Fusobacteria	Fusobacteriia	Fusobacteriales	<b>Fusobacteriaceae</b>	-	-4.9E-04	<0.001
21	Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	<b>Dactylosporangium</b>	9.0E-04	<0.001
22	Acidobacteria	Acidobacteria-6	iii1-15	<b>mb2424</b>	-	-1.1E-03	<0.001
23	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	<b>Alcaligenes</b>	5.9E-04	<0.001
24	Proteobacteria	Alphaproteobacteria	unassigned	unassigned	<b>unassigned</b>	-7.2E-04	<0.001
25	Proteobacteria	Alphaproteobacteria	<b>Caulobacterales</b>	-	-	6.6E-04	<0.001
26	Actinobacteria	Actinobacteria	Actinomycetales	Dermabacteraceae	<b>Brachybacterium</b>	3.2E-04	<0.001
27	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	<b>Brevundimonas</b>	-4.0E-04	<0.001
28	Gemmatimonadetes	<b>Gemm-3</b>	-	-	-	-6.6E-04	<0.001
29	Proteobacteria	Betaproteobacteria	<b>Methylophilales</b>	-	-	4.6E-04	<0.001
30	Bacteroidetes	Bacteroidia	Bacteroidales	<b>Porphyromonadaceae</b>	-	-3.8E-04	<0.001
31	Proteobacteria	<b>Betaproteobacteria</b>	-	-	-	-5.4E-04	<0.001
32	Proteobacteria	Betaproteobacteria	Burkholderiales	<b>Oxalobacteraceae</b>	-	3.8E-04	<0.001
33	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	<b>Lachnobacterium</b>	-4.1E-04	<0.001
34	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	<b>Winogradskyella</b>	6.5E-04	<0.001
35	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	<b>Streptococcus</b>	3.4E-04	<0.001
36	Actinobacteria	Actinobacteria	Actinomycetales	<b>Micrococcaceae</b>	-	3.6E-04	<0.001
37	Bacteroidetes	Bacteroidia	Bacteroidales	[Odoribacteraceae]	<b>Butyricimonas</b>	7.1E-04	<0.001
38	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Dickeya</b>	-1.1E-03	<0.001
39	Actinobacteria	Actinobacteria	Actinomycetales	<b>Yaniellaceae</b>	-	4.7E-04	<0.001
40	Firmicutes	Bacilli	Bacillales	<b>Staphylococcaceae</b>	-	2.0E-04	<0.001
41	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	<b>Eggerthella</b>	-5.6E-04	<0.001
42	Bacteroidetes	[Saprosipirae]	[Saprosipirales]	Chitinophagaceae	<b>Flaviumibacter</b>	1.0E-03	<0.001
43	Acidobacteria	Solibacteres	Solibacterales	<b>Solibacteraceae</b>	-	4.3E-04	<0.001
44	[Thermi]	Deinococci	Deinococcales	<b>Trueperaceae</b>	-	-2.2E-04	<0.001
45	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	<b>Comamonas</b>	2.3E-04	<0.001

b.) Taxa best predicting the GABRIELA farm home -like relative abundance defining FaRM<sub>GABRIELA</sub>.

Step	Phylum	Class	Order	Family	Genus	Estimate	Partial adjusted R-square
0	<b>Intercept</b>	-	-	-	-	-2.86E-02	0.000
1	Bacteroidetes	<b>Sphingobacteriia</b>	-	-	-	5.09E-04	0.374
2	Firmicutes	Bacilli	Lactobacillales	<b>Streptococcaceae</b>	-	-1.68E-03	0.201
3	Firmicutes	<b>Clostridia</b>	-	-	-	1.76E-03	0.118
4	Proteobacteria	Alphaproteobacteria	Rhodobacterales	<b>Rhodobacteraceae</b>	-	3.64E-04	0.076
5	Actinobacteria	Actinobacteria	<b>Actinomycetales</b>	-	-	-7.99E-04	0.033
6	Proteobacteria	<b>Alphaproteobacteria</b>	-	-	-	9.67E-04	0.029
7	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	<b>Psychrobacter</b>	1.63E-04	0.016
8	Proteobacteria	Gammaproteobacteria	<b>Pseudomonadales</b>	-	-	4.02E-04	0.011
9	<b>Cyanobacteria</b>	-	-	-	-	1.85E-03	0.022
10	Proteobacteria	Gammaproteobacteria	<b>Enterobacteriales</b>	-	-	2.28E-04	0.009
11	Bacteroidetes	Cytophagia	Cytophagales	<b>Cytophagaceae</b>	-	7.39E-04	0.006
12	<b>Tenericutes</b>	-	-	-	-	2.71E-04	0.006
13	Proteobacteria	-	-	-	-	1.54E-03	0.005
14	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	<b>Jeotgalicoccus</b>	3.70E-04	0.005
15	Firmicutes	Clostridia	Clostridiales	<b>[Mogibacteriaceae]</b>	-	2.73E-04	0.003
16	<b>Bacteroidetes</b>	-	-	-	-	3.24E-04	0.002
17	Firmicutes	Bacilli	<b>Bacillales</b>	-	-	2.06E-03	0.002
18	Firmicutes	<b>Bacilli</b>	-	-	-	-1.47E-03	0.002
19	Firmicutes	Bacilli	Lactobacillales	<b>Lactobacillaceae</b>	-	2.69E-04	0.002
20	Proteobacteria	Gammaproteobacteria	<b>Pasteurellales</b>	-	-	-5.27E-04	0.002
21	<b>Actinobacteria</b>	-	-	-	-	-1.18E-03	0.002
22	Firmicutes	Bacilli	Lactobacillales	Aerococcaceae	<b>Facklamia</b>	3.66E-04	0.002
23	Firmicutes	Bacilli	Bacillales	<b>Staphylococcaceae</b>	-	-1.33E-03	0.002
24	Cyanobacteria	Nostocophycideae	Nostocales	<b>Nostocaceae</b>	-	2.47E-04	0.001
25	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	<b>Paludibacter</b>	2.58E-04	0.001
26	<b>Unassigned</b>	-	-	-	-	3.74E-04	<0.001
27	Firmicutes	Bacilli	<b>Other</b>	-	-	-2.30E-04	<0.001
28	Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae	<b>Porphyromonas</b>	8.36E-04	<0.001
29	Firmicutes	Clostridia	Clostridiales	Clostridiaceae	<b>Clostridium</b>	3.20E-04	<0.001
30	Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	<b>Oribacterium</b>	1.78E-04	<0.001
31	Proteobacteria	Alphaproteobacteria	<b>Rickettsiales</b>	-	-	-1.03E-03	<0.001
32	Proteobacteria	Alphaproteobacteria	Rickettsiales	mitochondria	<b>unassigned</b>	7.58E-04	<0.001
33	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	<b>Bacteroides</b>	2.82E-04	<0.001
34	Actinobacteria	Actinobacteria	Actinomycetales	<b>Corynebacteriaceae</b>	-	3.68E-04	<0.001
35	Tenericutes	Mollicutes	Mycoplasmatales	<b>Mycoplasmataceae</b>	-	4.70E-04	<0.001
36	<b>Euryarchaeota</b>	-	-	-	-	2.59E-04	<0.001
37	Actinobacteria	<b>Thermoleophilia</b>	-	-	-	3.69E-04	<0.001
38	Bacteroidetes	Flavobacteriia	Flavobacteriales	<b>[Weeksellaceae]</b>	-	2.02E-04	<0.001
39	<b>Firmicutes</b>	-	-	-	-	6.04E-04	<0.001
40	Firmicutes	Clostridia	Clostridiales	Veillonellaceae	<b>Mitsuokella</b>	-2.66E-04	<0.001
41	Bacteroidetes	Bacteroidia	Bacteroidales	<b>Porphyromonadaceae</b>	-	-5.34E-04	<0.001
42	Actinobacteria	<b>Actinobacteria</b>	-	-	-	3.06E-03	<0.001
43	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	<b>Agromyces</b>	2.12E-04	<0.001
44	Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	<b>Wolinella</b>	-4.50E-04	<0.001
45	Proteobacteria	Gammaproteobacteria	<b>Xanthomonadales</b>	-	-	2.17E-04	<0.001
46	Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	<b>Brevundimonas</b>	-1.72E-04	<0.001



**Supplementary Table 11.** The modelled effects of LUKAS1 and GABRIELA farm home -like bacterial composition on asthma development in non-farm homes of each cohort. The farm home -like composition defined by the selected taxa-based models of FaRMI and FaRMI<sub>GABRIELA</sub>, respectively. The primary replication from LUKAS1 to GABRIELA is indicated with bold font.

Derivation cohort	Predicted cohort	N-total %-Asthmatics		Taxa-based models of FaRMI/ FaRMI <sub>GABRIELA</sub>	Doctor diagnosed asthma by 6 to 12 years of age <sup>a</sup>	
		<i>n</i>	%	IQR	aOR <sup>b</sup> (95% CI)	<i>P</i> -value
LUKAS1	LUKAS non-farm	244	19.3	0.32	0.48 (0.28 - 0.80)	0.005
	GABRIELA non-farm	603	12.3	0.27	<b>0.65 (0.46 - 0.91)</b>	<b>0.01</b>
GABRIELA	GABRIELA non-farm	603	12.3	0.21	0.75 (0.55 - 1.02)	0.07
	LUKAS non-farm	244	19.3	0.25	0.65 (0.41 - 1.05)	0.08

<sup>a</sup>In LUKAS1 the indoor microbiota composition was determined at 2 months and asthma refers to diagnosis by the 6 years of age (prospective design), in GABRIELA the microbiota composition was determined at 6 to 12 years of age and asthma refers to diagnosis by the respective time point (cross-sectional design). <sup>b</sup>Logistic regression analysis based adjusted odds ratios (aOR) presented per interquartile range (IQR) of selected taxa-based models of FaRMI or FaRMI<sub>GABRIELA</sub>. All models adjusted for sex, maternal and paternal allergic disease (LUKAS) or first degree relative with allergic disease (GABRIELA), maternal smoking during pregnancy, number of older siblings, maternal education and age (GABRIELA). *P*-values are two-sided.

**Supplementary Table 12.** Association between farm home -like relative abundance of bacteria/archaea (FaRMI) in LUKAS non-farm homes at 2 months and atopic sensitization at 6 years or asthma ever by 6 years stratified<sup>a</sup> by atopic sensitization<sup>b,c</sup>.

Health outcome	N-total		Risk of adverse outcome		
	%sensitized / -asthmatic		IQR	aOR <sup>d</sup> (95% CI)	P-value
	n	%			
Sensitized to any allergen <sup>b</sup>	195	29.7	0.35	1.07 (0.64 - 1.80)	0.80
Sensitized to inhaled allergens <sup>c</sup>	195	31.8	0.35	1.11 (0.64 - 1.92)	0.72
Asthma in non-sensitized	134	18.7	0.33	0.68 (0.32-1.55)	0.31
Asthma in sensitized <sup>b</sup>	58	24.1	0.41	0.22 (0.05-1.02)	0.05

<sup>a</sup>P-value for interaction term FaRMI\*sensitized to any allergen was 0.19. Atopic sensitization was defined by concentration of <sup>b</sup>any or <sup>c</sup>inhalant allergen specific IgE  $\geq 3.5$  kU/L in venous blood at 6 years. <sup>d</sup>aOR=logistic regression analysis based adjusted odds ratios presented per interquartile range (IQR) of FaRMI for the indicated adverse health outcome. All models adjusted for sex, cohort, maternal and paternal allergic disease, maternal smoking during pregnancy, number of older siblings and maternal education. P-values are two-sided.

**Supplementary Table 13.** Example of differences<sup>a</sup> in the relative frequency of PICRUSt<sup>ref 57</sup>-predicted genes (KEGG orthologs) between LUKAS non-farm homes with more farm home than non-farm home-like indoor microbiota (FaRMI>50%) and those with more non-farm home than farm home-like indoor microbiota (FaRMI<50%). Top 50 significant (two-sided p<0.05) differences selected based on magnitude of the relative difference in means. Blue color indicates negative and orange positive association with FaRMI. The gene orthologs are underlined where PubMed literature search for the gene product indicated association with bacterial virulence mechanism including secretion systems, which pathogenic bacteria may utilize in translocation of virulence factors into host cells.<sup>refs 28,29</sup>

KEGG Ortholog	Relative difference in means (%) <sup>b</sup>	Mean relative frequency (%)		Corrected P-value <sup>c</sup>
		FaRMI<50%	FaRMI>50%	
K00078: dihydrodiol dehydrogenase / D-xylose 1-dehydrogenase (NADP)	-200	4.79E-06	0	4.2E-06
K02137: F-type H <sup>+</sup> -transporting ATPase oligomycin sensitivity conferral protein [EC:3.6.3.14]	-200	2.66E-06	0	3.2E-02
<u>K04049: type III secretion protein SctB</u> <sup>ref 28 (reference for all type III secretion system associated proteins)</sup>	-200	2.26E-06	0	9.1E-03
<u>K04050: type III secretion protein SctE</u>	-200	2.26E-06	0	8.8E-03
<u>K04051: type III secretion protein SctG</u>	-200	2.26E-06	0	9.0E-03
<u>K04052: type III secretion protein SctH</u>	-200	2.26E-06	0	9.0E-03
<u>K04057: type III secretion protein SctP</u>	-200	2.26E-06	0	8.9E-03
<u>K04060: type III secretion protein SctY</u>	-200	2.26E-06	0	8.9E-03
K06125: 4-hydroxybenzoate hexaprenyltransferase [EC:2.5.1.-]	-200	2.66E-06	0	3.2E-02
K06931: None	200	0	8.00E-07	2.4E-02
K08658: prenyl protein peptidase [EC:3.4.22.-]	200	0	9.51E-06	3.0E-24
K12688: autotransporter serine protease [EC:3.4.21.-]	-200	2.37E-06	0	1.3E-02
K00213: 7-dehydrocholesterol reductase [EC:1.3.1.21]	198	1.13E-07	2.06E-05	8.1E-34
K00630: glycerol-3-phosphate O-acyltransferase [EC:2.3.1.15]	198	1.13E-07	2.06E-05	1.6E-33
K00654: serine palmitoyltransferase [EC:2.3.1.50]	174	2.54E-07	3.58E-06	5.9E-07
K00960: DNA-directed RNA polymerase [EC:2.7.7.6]	-174	1.15E-05	8.15E-07	1.4E-09
K03955: NADH dehydrogenase (ubiquinone) 1 alpha/beta subcomplex 1 [EC:1.6.5.3 1.6.99.3]	-169	4.40E-06	3.66E-07	7.5E-03
K03278: UDP-D-galactose:(glucosyl)LPS alpha-1,3-D-galactosyltransferase [EC:2.4.1.44]	-163	2.33E-05	2.35E-06	1.3E-02
K07805: putative membrane protein PagD	-163	2.33E-05	2.35E-06	1.3E-02
K08477: outer membrane protease E [EC:3.4.21.-]	-163	2.33E-05	2.35E-06	1.3E-02
<u>K11023: pertussis toxin subunit 1</u> [EC:2.4.2.-] <sup>ref 30</sup>	-163	2.33E-05	2.35E-06	1.3E-02
K11191: PTS system, N-acetylmuramic acid-specific IIB component [EC:2.7.1.69]	-163	2.33E-05	2.35E-06	1.3E-02
K13085: phosphatidylinositol-4,5-bisphosphate 4-phosphatase [EC:3.1.3.78]	-163	2.33E-05	2.35E-06	1.4E-02
<u>K13284: invasin A</u> <sup>ref 31 (reference for all invasins)</sup>	-163	2.33E-05	2.35E-06	1.3E-02
<u>K13286: invasin C</u>	-163	2.33E-05	2.35E-06	1.3E-02
<u>K13450: phosphothreonine lyase</u> [EC:4.2.3.-] <sup>ref 32</sup>	-163	2.33E-05	2.35E-06	1.3E-02
<u>K13739: secreted effector protein SopD</u> <sup>ref 33</sup>	-163	2.33E-05	2.35E-06	1.3E-02
<u>K13740: secreted effector protein SptP</u> <sup>ref 31</sup>	-163	2.33E-05	2.35E-06	1.3E-02
<u>K13741: guanine nucleotide exchange factor SopE</u> <sup>ref 31</sup>	-161	2.73E-05	2.91E-06	4.1E-03
K12227: TraL protein	-156	8.67E-06	1.06E-06	6.5E-05
<u>K08599: YopT peptidase</u> [EC:3.4.22.-] <sup>ref 34</sup>	-148	3.80E-06	5.62E-07	1.7E-02
K13318: dTDP-4-keto-6-deoxy-L-hexose 4-reductase	148	1.74E-05	1.17E-04	4.6E-02
<u>K12228: TrbB protein</u> <sup>ref 29 (reference for all the Type IV secretion system associated proteins)</sup>	-148	8.79E-06	1.32E-06	1.1E-04
K07538: 6-hydroxycyclohex-1-ene-1-carboxyl-CoA dehydrogenase [EC:1.1.1.-]	141	2.90E-06	1.69E-05	3.4E-09
K03280: UDP-N-acetylglucosamine:(glucosyl)LPS alpha-1,2-N-acetylglucosaminyltransferase [EC:2.4.1.56]	-138	2.76E-05	5.07E-06	8.7E-03
<u>K09953: lipid A 3-O-deacylase</u> <sup>ref 35</sup>	-138	2.76E-05	5.07E-06	8.9E-03
K11735: GABA permease	-133	2.87E-05	5.76E-06	8.8E-03
K05924: None	-131	2.90E-06	6.07E-07	1.7E-02
<u>K12203: defect in organelle trafficking protein DotB</u> <sup>ref 29</sup>	-129	1.31E-05	2.84E-06	2.2E-06
<u>K12204: defect in organelle trafficking protein DotC</u>	-129	1.31E-05	2.84E-06	2.5E-06
<u>K12205: defect in organelle trafficking protein DotD</u>	-129	1.31E-05	2.84E-06	1.9E-06
K00188: 2-oxoisovalerate ferredoxin oxidoreductase, delta subunit [EC:1.2.7.7]	128	6.69E-07	3.02E-06	1.9E-02
K00319: methylenetetrahydromethanopterin dehydrogenase [EC:1.5.99.9]	128	6.69E-07	3.02E-06	1.9E-02
K00440: coenzyme F420 hydrogenase alpha subunit [EC:1.12.98.1]	128	6.69E-07	3.02E-06	1.8E-02
K00442: coenzyme F420 hydrogenase delta subunit	128	6.69E-07	3.02E-06	1.8E-02
K00443: coenzyme F420 hydrogenase gamma subunit [EC:1.12.98.1]	128	6.69E-07	3.02E-06	2.0E-02
K00578: tetrahydromethanopterin S-methyltransferase subunit B [EC:2.1.1.86]	128	6.69E-07	3.02E-06	2.0E-02
K00579: tetrahydromethanopterin S-methyltransferase subunit C [EC:2.1.1.86]	128	6.69E-07	3.02E-06	2.0E-02
K00580: tetrahydromethanopterin S-methyltransferase subunit D [EC:2.1.1.86]	128	6.69E-07	3.02E-06	1.8E-02
K00581: tetrahydromethanopterin S-methyltransferase subunit E [EC:2.1.1.86]	128	6.69E-07	3.02E-06	2.0E-02
K00582: tetrahydromethanopterin S-methyltransferase subunit F [EC:2.1.1.86]	128	6.69E-07	3.02E-06	2.0E-02

<sup>a</sup>From the predicted LUKAS non-farm home dust bacterial/archaeal metagenome, nearly half (48%) of the 6909 genes (KEGG orthologs) were present in significantly (p<0.05) different proportions between homes with more farm home than non-farm home-like indoor microbiota (FaRMI>50%) and those with more non-farm home than farm home-like indoor microbiota (FaRMI<50%). <sup>b</sup>Difference between means as a percentage of the mean of the means of the two groups. <sup>c</sup>White's non-parametric two-sided t-test with Benjamini-Hochberg false discovery rate.

**Supplementary table 14.** Antibodies used for studying the expression of inhibitory receptors on blood dendritic cells (adapted from Martikainen *et al.*, 2015<sup>ref 49</sup>).

Antibody	Fluorochrome	Cellular target / usage	Clone	Isotype	Manufacturer
CD14	PE-Cy5.5	monocytes	TüK4	Mouse IgG2a	Invitrogen <sup>d</sup>
CD19	PE-Cy5.5	B cells	SJ25-C1	Mouse IgG1	Invitrogen
CD11c	PE-Cy7	mDC1 and mDC2	3.9	Mouse IgG1 κ	eBioscience <sup>e</sup>
CD1c <sup>a</sup>	APC	mDC1	AD5-8E7	Mouse IgG2a	Miltenyi Biotec <sup>f</sup>
CD303 <sup>b</sup>	PE	pDC	AC144	Mouse IgG1	Miltenyi Biotec
ILT4 <sup>c</sup>	FITC	tolerogenic potential	27D6	Rat IgM	eBioscience
ILT3	Biotin	tolerogenic potential	ZM4.1	Mouse IgG1, κ	Biolegend <sup>g</sup>
IgG1	FITC	isotype control	X40	Mouse IgG1, κ	BD Biosciences
Rat IgM	FITC	isotype control	eBRM	Rat IgM κ	eBioscience
IgG1	biotin	isotype control	MOPC-21	Mouse IgG1, κ	Biolegend
IgG1	biotin	isotype control	P3.6.2.8.1	Mouse IgG1, κ	eBioscience
Streptavidin	APC-eFluor780	secondary antibody for biotin-conjugated antibodies	-	-	eBioscience

<sup>a</sup>Alias BDCA-1, <sup>b</sup>alias BDCA-2, <sup>c</sup>alias CD85d, <sup>d</sup>Camarillo, CA, USA, <sup>e</sup>San Diego, CA, USA, <sup>f</sup>Bergisch Gladbach, Germany, <sup>g</sup>San Diego, CA, USA.

**Supplementary table 15.** QIITA IDs for studies used in the source tracking analysis.

QIITA ID	Study title
<i>Soil microbiota</i>	
ID 619	Neon soils
ID 103	Pyrosequencing Based Assessment of Soil pH as a Predictor of Soil Bacterial Community Structure at the Continental Scale
ID 1578	Changes in microbial communities along redox gradients in polygonized Arctic wet tundra soils
ID 990	Spatial scale drives patterns in soil bacterial diversity.
ID 104	Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes
ID 1024	The Soil Microbiome Influences Grapevine Associated Microbiota MiSeq
ID 928	Examining the global distribution of dominant archaeal populations in soil
ID 94	Soil bacterial and fungal communities across a pH gradient in an arable soil
<i>Human microbiota:</i>	
ID 449	Bacterial community variation in human body habitats across space and time
ID 232	Forensic identification using skin bacterial communities - keyboard
ID 850	Human gut microbiome differentiation viewed across cultures, ages and families illumina
ID 1928	Structure, function and diversity of the healthy human microbiome (V35)
<i>Bovine microbiota:</i>	
ID 1621	Time series to determine the effect of Monensin on microbial hindgut bacterial composition