

# Urbanized microbiota in infants, immune constitution, and later risk of atopic diseases

Jenni Lehtimäki, PhD,<sup>a</sup> Jonathan Thorsen, MD, PhD,<sup>a,b</sup> Morten Arendt Rasmussen, PhD,<sup>a,c</sup> Mathis Hjelmsø, PhD,<sup>a</sup> Shiraz Shah, PhD,<sup>a</sup> Martin S. Mortensen, PhD,<sup>d,e</sup> Urvish Trivedi, PhD,<sup>d</sup> Gisle Vestergaard, PhD,<sup>d,f</sup> Klaus Bønnelykke, MD, PhD,<sup>a</sup> Bo Lund Chawes, MD, PhD,<sup>a</sup> Susanne Brix, PhD,<sup>g</sup> Søren J. Sørensen, PhD,<sup>d</sup> Hans Bisgaard, MD, PhD,<sup>a</sup> and Jakob Stokholm, MD, PhD<sup>a</sup>  
 Gentofte, Copenhagen, Frederiksberg, and Lyngby, Denmark, and  
 Hinxton, United Kingdom

**Background:** Urbanization is linked with an increased burden of asthma and atopic traits. A putative mechanism is insufficient exposure to beneficial microbes early in life, leading to immune dysregulation, as was previously shown for indoor microbial exposures.

**Objective:** Our aim was to investigate whether urbanization is associated with the microbiota composition in the infants' body and early immune function, and whether these contribute to the later risk of asthma and atopic traits.

**Methods:** We studied the prospective Copenhagen Prospective Studies on Asthma in Childhood 2010<sub>2010</sub> mother-child cohort of 700 children growing up in areas with different degrees of urbanization. During their first year of life, airway and gut microbiotas, as well as immune marker concentrations, were

defined. When the children were 6 years of age, asthma and atopic traits were diagnosed by pediatricians.

**Results:** In adjusted analyses, the risk of asthma and aeroallergen sensitization were increased in urban infants. The composition of especially airway but also gut microbiotas differed between urban and rural infants. The living environment-related structure of the airway microbiota was already associated with immune mediator concentrations at 1 month of age. An urbanized structure of the airway and gut microbiotas was associated with an increased risk of asthma coherently during multiple time points and also with the risks of eczema and sensitization.

**Conclusion:** Our findings suggest that urbanization-related changes in the infant microbiota may elevate the risk of asthma and atopic traits, probably via cross talk with the developing immune system. The airways may facilitate this effect, as they are open for colonization by environmental airborne microbes and serve as an immune interface. (*J Allergy Clin Immunol* 2021;■■■:■■■-■■■.)

**Key words:** Microbiome, residential environment, noncommunicable diseases, urbanization, childhood

From <sup>a</sup>the Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Gentofte; <sup>b</sup>the Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen; <sup>c</sup>the Section of Chemometrics and Analytical Technologies, Department of Food Science, University of Copenhagen, Frederiksberg; <sup>d</sup>the Section of Microbiology, Department of Biology, University of Copenhagen; <sup>e</sup>the Host-Microbiota Interactions Laboratory, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton; <sup>f</sup>the Section for Bioinformatics, Department of Health Technology, Technical University of Denmark, Lyngby, Denmark; and <sup>g</sup>the Department of Biotechnology and Biomedicine, Technical University of Denmark, Lyngby.

All funding received by the Copenhagen Prospective Studies on Asthma in Childhood (COPSAC) is listed at [www.copsac.com](http://www.copsac.com). The Lundbeck Foundation (grant R16-A1694), the Ministry of Health, Denmark (grant 903516), the Danish Council for Strategic Research (grant 0603-00280B), and the Capital Region Research Foundation have provided core support to the COPSAC research center. J. Stokholm is supported by the Danish Council for Independent Research's Sapere Aude programme (Grant no. 8045-00081B). J. Thorsen is supported by the BRIDGE-Translational Excellence Programme ([bridge.ku.dk](http://bridge.ku.dk)) at the Faculty of Health and Medical Sciences, University of Copenhagen, which is funded by the Novo Nordisk Foundation (grant agreement NNF18SA0034956). The funding agencies did not have any role in design and conduct of the study; collection, management, and interpretation of the data; or preparation, review, or approval of the article. No pharmaceutical company was involved in the study. Disclosure of potential conflict of interest: The authors declare that they have no relevant conflicts of interest.

We are aware of and comply with recognized codes of good research practice, including the Danish Code of Conduct for Research Integrity. We comply with national and international rules on the safety and rights of patients and healthy subjects, including Good Clinical Practice as defined in the European Union's Directive on Good Clinical Practice, the International Conference on Harmonization's good clinical practice guidelines, and the Declaration of Helsinki. Privacy is important to us, which is why we follow national and international legislation on the General Data Protection Regulation, the Danish Act on Processing of Personal Data, and the practice of the Danish Data Inspectorate.

Received for publication September 17, 2020; revised November 11, 2020; accepted for publication December 8, 2020.

Corresponding author: Hans Bisgaard, MD, DMSc, Ledreborg Alle 34, 2820, Gentofte, Denmark. E-mail: [bisgaard@copsac.com](mailto:bisgaard@copsac.com).  
 0091-6749/\$36.00

© 2020 American Academy of Allergy, Asthma & Immunology  
<https://doi.org/10.1016/j.jaci.2020.12.621>

Urban children tend to have a higher risk of developing asthma<sup>1,2</sup> and atopic traits such as allergic sensitization,<sup>3</sup> rhinitis,<sup>4</sup> and eczema<sup>5</sup> than rural children do. Several factors differ between these populations, including larger family size and increased exposure to animals in rural areas, which may contribute to the difference in disease risk.<sup>6-8</sup> Also, the microbial exposures are different between rural and urban areas, which has been shown for the composition of outdoor, indoor, and human microbiotas.<sup>9-12</sup> These microbial differences may mediate the protective effect of a rural environment, as demonstrated for the farming-related indoor microbiota, which has been repeatedly associated with asthma protection,<sup>13-15</sup> also in experimental settings.<sup>14,16,17</sup>

Microbial exposures are especially important during infancy and early childhood, when the constitution of the immune system is primed by extensive immune-microbiota cross talk.<sup>18,19</sup> Indeed, accumulating evidence shows that the infant microbiota is altered before development of allergic sensitization, rhinitis,<sup>20</sup> eczema,<sup>21</sup> and/or asthma,<sup>22-24</sup> potentially influencing this early cross talk. A recent report suggested that growing up on farms can support the maturation of infants' gut microbiota and thus reduce their risk of asthma when compared with the risk faced by other rural children.<sup>25</sup> However, how the constantly increasing urbanization of our living environments influences early immune-microbiota cross talk and later risk of disease remains unclear.

**Abbreviations used**

CCL: C-C motif chemokine ligand  
 COPSAC: Copenhagen Prospective Studies on Asthma in Childhood  
 CXCL8: C-X-C motif chemokine ligand 8  
 OR: Odds ratio  
 sPLS: Sparse partial least squares

To address the hypothesis that dissimilar microbiotas between urban and rural infants contribute to development of their immune systems and later risk of disease, we utilized longitudinal data from the Copenhagen Prospective Studies on Asthma in Childhood 2010 (COPSAC<sub>2010</sub>) mother-child cohort. We defined the degree of urbanization in their living environment during infancy and studied its associations with asthma and atopic traits at age 6 years. We further investigated whether the early microbiota composition of airways and gut, as well as the local airway and systemic immune function, may convey associations between living environment and disease.

**METHODS****Cohort**

The children in the COPSAC<sub>2010</sub> mother-child cohort (N = 700) have been followed prospectively since birth, with numerous scheduled visits at a dedicated research unit. Comprehensive descriptions of the cohort and the research approach have been published.<sup>26-28</sup> The COPSAC pediatricians were solely responsible for diagnosis and treatment of all respiratory, allergy, and skin-related symptoms.

**Definition of asthma and atopic traits**

*Asthma* by age 6 years was diagnosed prospectively on the basis of a previously detailed quantitative symptom algorithm<sup>26,29</sup> (see the [Methods](#) section in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)).

*Eczema* by age 6 years was diagnosed prospectively on the basis of the criteria of Hanifin and Rajka<sup>30</sup> requiring the presence of 3 of 4 major criteria and at least 3 of 23 minor criteria.

*Allergic rhinitis* at age 6 years was based on sensitization and clinical interviews of the parents regarding history of significant nasal congestion, sneezing, and/or runny nose outside periods with common cold and in the relevant period of the sensitized aeroallergen.

*Allergic sensitization* was determined at age 6 years; it was defined as any positive skin prick test result of at least 3 mm (ALK-Abelló, Hørsholm, Denmark) or a serum level of specific IgE against common inhalant and/or food allergens of at least 0.35 kUa/L (ImmunoCAP; Thermo Fisher Scientific, Allerod, Denmark). *Not sensitized* was defined as being both skin prick test- and specific IgE-negative (see the [Methods](#) section in the Online Repository).

**Definition of urbanization gradient**

We utilized the land cover database CORINE (<https://land.copernicus.eu/>) from the year 2012 for an unbiased definition of the living environment. Addresses at birth (n = 686) were translated to coordinates by using Google API services.<sup>31</sup> Land cover types around each coordinate point were extracted from the CORINE2012 raster map (100-m resolution) with a 3-km buffer (see [Table E1](#) in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). The correlation matrix of proportions of 5 major land cover types within the buffer region were analyzed by using principal components analysis.<sup>32</sup> The first axis explained 44% of the total variation and made clear the separation between artificial and natural (ie, agricultural and forested) land cover types (see [Fig E1, A](#) in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). The first axis was extracted for downstream analyses to represent a gradient of urbanization

running from the most rural to the most urban environments in the study region. Land cover information was also dichotomized into rural and urban groups ( $n_{\text{rural}} = 314$ ;  $n_{\text{urban}} = 372$  [see [Fig E1, B](#)]) by using partitioning around medoids clustering on euclidean distances.<sup>33</sup> Most children (n = 577) either did not move during their first year of life or moved later than 50 weeks after birth (n = 5). Of the 109 children who did move, 77% did not change between rural and urban groups (see the [Methods](#) section in the Online Repository).

**Covariates**

Differences between rural and urban living environments were tested for common disease-associated factors, including the following: pet ownership; day care attendance during the first year of life; length of the breast-feeding period; exposure to passive smoking; family income; parental education; number of older siblings; home type; mode of delivery; parental diagnosis of asthma, eczema and rhinitis; and use of antibiotics during the first year of life (see [Fig E2](#) in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)).

**Microbiota samples**

Airway microbiota samples were collected at 1 week (n = 549) and 1 (n = 647) and 3 months (n = 657) of age,<sup>34</sup> and gut microbiota samples were collected at 1 week (n = 544), 1 month (n = 615), and 1 year (n = 625) of age<sup>22</sup> (see the [Methods](#) section in the Online Repository).

DNA was extracted from airway and gut samples by using the PowerMag Soil DNA Isolation Kit optimized for epMotion (MO-BIO Laboratories, Inc, Carlsberg, Calif) by using the epMotion robotic platform model 5075 (Eppendorf, Hamburg, Germany). The manufacturer's protocol was modified slightly (described earlier<sup>34</sup>). Extracted DNA was stored at -20°C degrees before amplification of the 16S rRNA gene variable region 4 (V4) and sequencing with the Illumina MiSeq System (Illumina Inc, Calif), as previously described.<sup>34</sup> The bioinformatics analysis pipeline is detailed in the [Methods](#) section of the Online Repository.

**Immune mediators**

Airway immune mediator samples (upper airway mucosal lining fluid) were collected from 1-month-old infants (n = 620) with a synthetic absorptive matrix placed in both nostrils for 2 minutes.<sup>35</sup> Matrixes were stored immediately at -80°C degrees. Blood for systemic immune mediator analysis was drawn into an EDTA tube from a cubital vein when the child was 6 months old (n = 568) and centrifuged to separate plasma and cells, after which the plasma was immediately stored at -80°C until analysis.<sup>36</sup>

The levels of the *a priori*-selected cytokines and chemokines<sup>37,38</sup> were determined by using high-sensitivity immunoassays based on electrochemiluminescence in a multiplex setting. Samples were read in duplicates by using the Sector Imager 6000 (Meso Scale Discovery, Gaithersburg, Md).

**Statistics**

Odds ratios (ORs) for atopic traits in response to urban and rural (baseline) classes, as well as urbanization gradient, were calculated with logistic regression models adjusted for lifestyle variables. Adjustment included other covariates associated with urbanization (see [Fig E2](#)), whereas house type was excluded on account of collinearity with the explanatory variable of interest (urban/rural groups).

The R software package phyloseq<sup>39</sup> was used for handling the microbiota data. Observed richness and Shannon diversity were used as the  $\alpha$ -diversity metrics. The community compositions of microbiotas ( $\beta$ -diversity) were visualized with principal coordinates analysis for Bray-Curtis,<sup>40</sup> Weighted UniFrac, and Unweighted UniFrac<sup>41</sup> distances and tested with permutational multivariate ANOVA<sup>32</sup> with 1000 permutations. The dissimilarities within rural and urban individuals were defined with the distance of each sample from group mean.<sup>42</sup>

Immune mediators were log-transformed and centered log ratio-scaled to remove biologic variation due to differing amounts of protein per volume of

fluid from the data, set as well as to increase the comparability between immune mediators following methodology developed previously.<sup>38</sup> The normalized concentrations were tested between rural and urban groups in adjusted linear models.

Sparse partial least squares (sPLS) models with relative abundances of taxa agglomerated to the genus level, log-transformed, and cleaned to include those present at least in 5% of samples were used to predict the urbanization gradient.<sup>38,43</sup> The best model was selected on the basis of the root mean square error statistics, from repeated 10-fold cross-validation, to avoid overfitting. The predicted values from each model were extracted to represent a bacterial score for the urbanization gradient. All analyses were conducted in R software, version 3.5.2.<sup>44</sup>

## Ethics statement

This study was approved by the local ethics committee (H-B-2008-093), and the Danish Data Protection Agency (2015-41-3696). Both parents gave verbal and written informed consent before enrollment.

## RESULTS

### Asthma and atopic traits in rural and urban infants

All subjects spent their first year of life in a Westernized, densely populated, and geographically concise area in Zealand (7031 km<sup>2</sup>), Denmark (Fig 1, A). Few children lived on the farms. Several lifestyle variable, such as pet ownership, number of older siblings, income in the family, and education of the parents, differed between rural and urban infants (see Fig E2). Other screened variables, including mode of delivery, antibiotic use during the first year of life, vitamin D or fish oil intervention groups, and atopic diseases in the parents (all of which were previously associated with atopic diseases),<sup>27,45,46</sup> did not differ between rural and urban infants.

The prevalence of asthma at 6 years of age or earlier was 22.3% (n = 146), the prevalence of allergic rhinitis was 6.8% (n = 45), and the prevalence of eczema was 31.7% (n = 210). At 6 years of age, 31.7% of the children (n = 157) were sensitized; 23.7% of them (n = 117) had aeroallergen and 14.7% (n = 73) food sensitization. The prevalences of asthma, allergic rhinitis, and aeroallergen sensitization at 6 years of age were higher in children spending their first year of life in an urban environment than in children spending their first year of life in a rural environment (Fig 1, B and D). Allergic sensitization and eczema showed similar trends. Urban infants developed more specific sensitization to peanut, house dust mite, timothy grass, birch, and mugwort allergens (see Table E2 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). After adjustment for lifestyle features, the ORs for asthma (adjusted OR = 2.31; 95% CI = 1.47-3.68; *P* = .0003) and aeroallergen sensitization (adjusted OR = 1.77; 95% CI = 1.05-3.02; *P* = .0348) were increased in urban infants (Fig 1, C and E). The adjusted, predicted prevalence of asthma and aeroallergen sensitization showed a large increase along the urbanization gradient, whereas only food sensitization remained completely unrelated (see Fig E3 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)).

### Microbiota and immune mediator concentrations in rural and urban infants

The airway microbiota at all 3 time points showed significant differences between rural and urban infants regardless of the metric used to define the similarity of microbiotas between subjects (see Fig E4 and Table E3 in this article's Online

Repository at [www.jacionline.org](http://www.jacionline.org)). These differences remained significant after adjustment for each lifestyle variable, as well as in a fully adjusted model (permutational multivariate ANOVA [see Table E4 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)]). Part of the observed rural-urban differences were explained by the larger family size and more common pet ownership in rural environments. Shannon diversity and richness in the airway microbiota were higher in urban infants at 1 month of age (ANOVA; *P* = .004 and .005, respectively). Remarkably, urban infants had more homogenous airway microbiotas than did rural infants at all time points (ie, urban children had microbiotas more similar to each other than rural children had [see Fig E5 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)]). Among the 15 most abundant genera in the airways, several showed dissimilar relative abundances between rural and urban infants (see Fig E6 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). *Veillonella* showed a consistent association with an urban living environment through all time points. *Haemophilus* and *Rothia* were associated with an urban environment, whereas *Dolosigranulum* and *Moraxella* were associated with a rural environment at more than 1 time point (see Fig E6).

The composition of the gut microbiota was associated with the living environment only at 1 year of age regardless of distance metric (see Fig E4 and Table E3). Nevertheless, in the fully adjusted model, this association was attenuated (see Table E5 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)), mainly because of the influence of older siblings. The 15 most abundant genera in the gut microbiota at different time points did not show differences between rural and urban environments after adjustment for multiple testing (see Fig E7 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). However, the *Firmicutes*-to-*Bacteroidetes* ratio was lower in rural infants at 1 year of age (a mean of 3.4 in rural infants vs 4.1 in urban infants; ANOVA; *P* = .0004). A closer look at this ratio revealed that the richness within *Bacteroidetes* was higher in rural infants than in urban infants at the same time point (a mean of 13.4 in rural infants vs a mean of 11.5 in urban infants; ANOVA; *P* = .0003).

The relative concentrations of both airway (1 month) and systemic (6 months) immune mediators differed between urban and rural infants in adjusted analyses for both sample types (Fig 2 and see Table E6 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org)). In airways (Fig 2, A), type 2-related cytokine and chemokine concentrations showed a large disparity between urban and rural infants. For example, urban infants had higher relative concentrations of C-C motif chemokine ligands (CCLs) (CCL11, CCL13, and CCL17), whereas concentrations of CCL22 and CCL26 were higher in rural infants. Furthermore, IL-12p70 (type 1) was found in higher relative concentrations in rural infants, as was IL-10 (regulatory), which has anti-inflammatory properties. A different set of immune mediators was measured in the systemic samples (Fig 2, B). Level of C-X-C motif chemokine ligand 8 (CXCL8 [type 17]) was measured in both airway and systemic samples, with urban infants found to have a higher relative concentration in both sample types. However, other immune mediators, namely, IL-1β and TNF-α, measured from both sample types did not show similar responses.

### Urbanized infant microbiota

We defined the bacterial community profiles most indicative of urbanization gradient separately for each time point and sampling

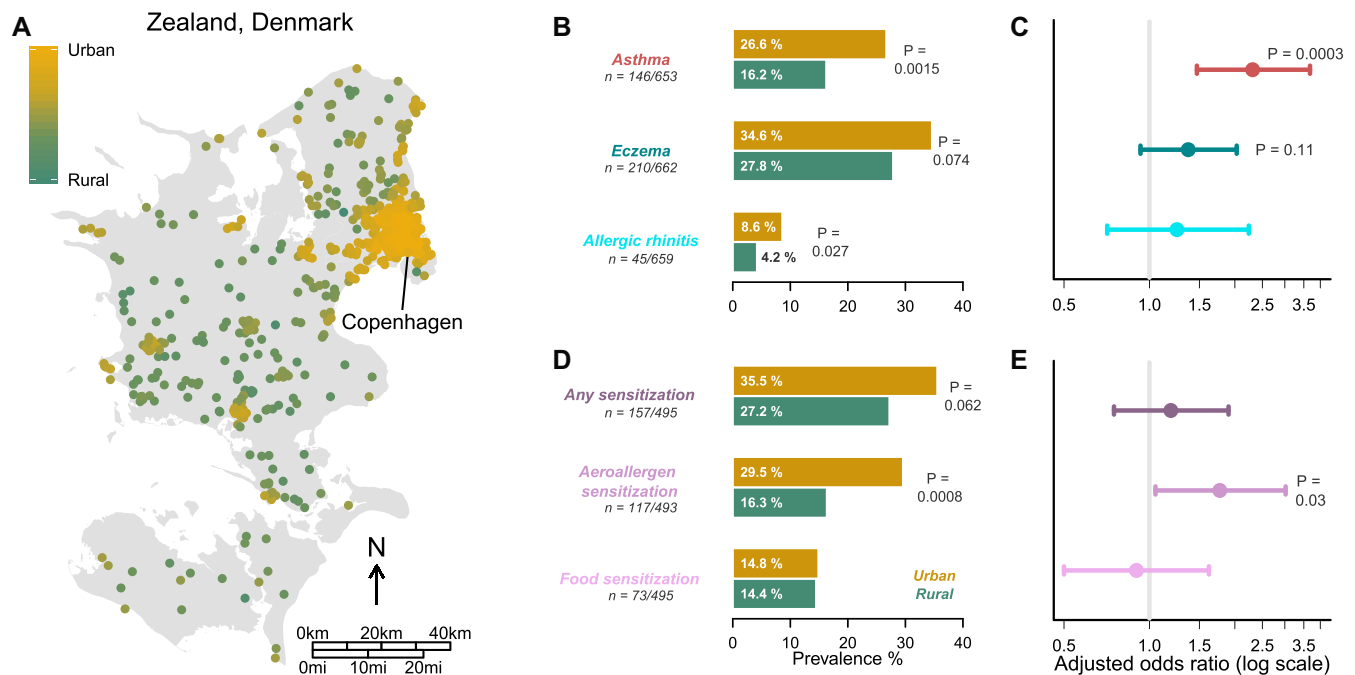


**FIG 1.** Prevalence of asthma and atopic diseases in rural and urban infants. The proportions of main land cover types within a radius of 3 km around the home at birth are condensed by principal components analysis to a gradient from mostly rural to highly urbanized living environments. **A**, Geographic distribution of children colored by the urbanization gradient. Urban children are mainly clustered around the capital, Copenhagen, as expected. Coordinate information is slightly jittered on the map to preserve anonymity. **B** and **D**, Crude prevalence of outcomes at 6 years of age in children living in urban and rural environments during infancy. Differences have been tested with Fisher exact tests. **C** and **E**, Adjusted ORs from logistic regression models for the outcomes in an urban versus rural (baseline) living environment. Urban infants have increased risk for asthma and aeroallergen sensitization. Analyses are adjusted for socioeconomic, pet ownership, older siblings, passive smoking exposure, and total length of breast-feeding.

site. We utilized cross-validated sPLS models to find a minimal (sparse) set of jointly contributing taxa that discriminated between urban and rural samples. The models predicted better during later time points in both airway and gut samples, indicating an accumulated influence of living environment with age (Fig 3).

Higher relative abundances of the genera *Veillonella*, *Rothia*, *Gemella*, *Bergyella*, and *Streptococcus* in airway microbial profiles were indicative of urban living environments, whereas *Moraxella*, *Paracoccus*, and *Dolosigranulum* were characteristic of a rural environment (Fig 3, A) at all time points corresponding to univariate analyses (see Fig E6). Some airway bacteria, such as *Bacillus* and *Listeria*, were associated with a rural environment at the first week of life, but the direction shifted later to association with urban environments.

*Moraxella* was also indicative of a rural environment and *Veillonella* was indicative of an urban environment in the gut microbial profiles (Fig 3, B), but only at a single time point each. *Bifidobacterium* in the gut was indicative of a rural environment already in the first week of life. In the gut profiles, only *Bacteroides* switched direction from an urban to a rural environment at the last time point. Overall, living environment-related members of gut communities did not match much between time points, primarily owing to the low predictability at 1 week and 1 month, as well as owing to the strong shifts in community composition during the first year of life (as described previously<sup>22</sup>).

### Urbanized microbial profiles, risk of disease, and concentrations of immune mediators

A bacterial score was extracted for each sample (how much did the composition of the sample resemble the sPLS trained bacterial signature?), with higher values indicating more urbanized bacterial profiles in infants.

Having an urbanized bacterial profile in the samples was associated with a higher risk of asthma for all but the 1-week gut and 3-month airway bacterial scores (for airway at 1 week: OR = 1.25; 95% CI = 1.01-1.55;  $P = .0473$ ; for airway at 1 month: OR = 1.22; 95% CI = 1.00-1.48;  $P = .0498$ ; for gut at 1 month: OR = 1.29; 95% CI = 1.05-1.59;  $P = .0175$ ; and for gut 1 year: OR = 1.24; 95% CI = 1.02-1.53;  $P = .0368$  [Fig 4]). The risk of eczema increased only in response to an urbanized gut bacterial profile at 1 week of age (OR = 1.24; 95% CI = 1.04-1.48;  $P = .0161$ ). Similarly, the risk of sensitization was restricted to 1 time point and compartment. An urbanized bacterial profile in the gut at 1 year of age was associated with the risk of any sensitization (OR = 1.28; 95% CI = 1.03-1.59;  $P = .0281$ ) and aeroallergen sensitization (OR = 1.24; 95% CI = 0.98-1.57;  $P = .0812$ ), although borderline significant (Fig 4).

The airway bacterial score at 1 month of age was correlated with immune mediator concentrations at the same time point (Fig 5 and see Table E6), and the direction of these associations was similar to that in our comparison between urban and rural groups

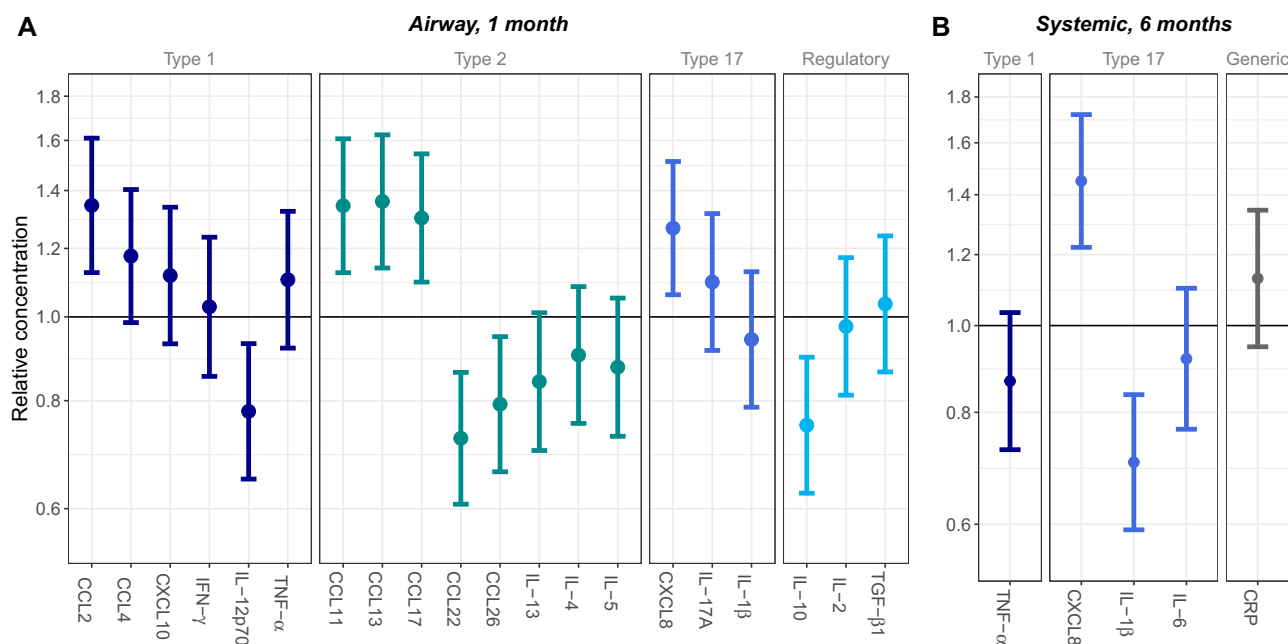


**FIG 2.** Relative concentration of airway and systemic immune mediators differ between urban and rural infants. The airway immune profile at 1 month of age ( $n = 586$  [A]) and systemic immune profile at 6 months of age ( $n = 665$  [B]) in response to urban and rural classes defined by linear models adjusted for living environment-associated lifestyle features. Positive estimates indicate higher relative concentration in urban infants, whereas negative estimates indicate a higher relative concentration in rural infants. Error bars represent the 95% CI. False-positive rate-corrected  $P$  values are presented in Table E6.

(Fig 2), but with important distinctions. The systemic CXCL8 concentration was higher with an urbanized bacterial profile (Fig 5) corresponding to results for living environment (Fig 2), but these associations were borderline after correction for multiple testing. In contrast to the living environment associations, airway IL-2 and IL-13 concentrations were shown to be lower in infants with an urbanized bacterial profile. Bacterial scores for the gut microbiota were not significantly associated with immune mediator profiles (see Fig E8 and Table E6).

## DISCUSSION

### Primary findings

The higher risk of asthma and atopic traits in urban than in rural children is a well-established phenomenon.<sup>13</sup> Our prospective, observational cohort study of 700 children supports the hypothesis that the early development of the human microbiota and its influential cross talk with the immune system shape the predisposition to these diseases in childhood. We found that an urban living environment in infancy was associated with a higher risk of asthma and aeroallergen sensitization in childhood. The urbanized microbiota both in the airways and in the gut in infancy was positively associated with the risk of having asthma. The urbanized gut microbiota profile at 1 week of age was also related to the risk of eczema, and at 1 year of age, with allergic sensitization. Furthermore, several immune markers showed significant associations with living environment-related airway bacteria, suggesting an influence of these bacteria in immune development.

### Strengths and limitation

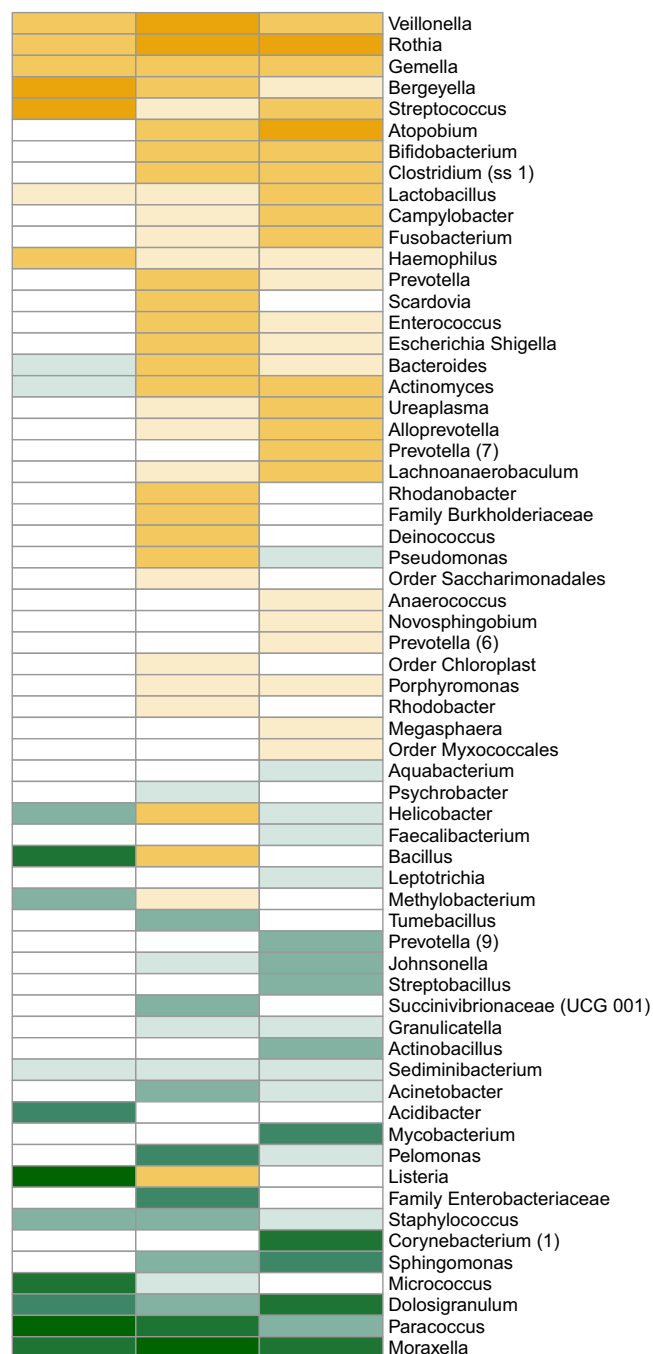
A major strength of this study was the extensive and prospective clinical assessment of outcomes and environmental

exposures. All children were uniformly diagnosed and repeatedly monitored for development of symptoms of asthma and atopic traits at the COPSAC clinic, in which the microbial and immunologic samples utilized in the study were also collected. Our definition of the living environment was based on land cover data providing unbiased characterization of rural and urban individuals. However, our observational study is unable to confirm causal relationships and depends on future studies with interventional or experimental approaches to validate the mechanisms behind our findings. Further, the methodology that we used for microbiota analysis can hide some details, such as the importance of bacterial strains.

Even though our findings support the role of the urbanized microbiota in disease development, other explanatory variables can shape the risk. The examined diseases are believed to originate from multiple triggers acting via a myriad of gene-environment interactions.<sup>47</sup> These can include exposures other than bacteria, such as indoor fungi, which has been associated with childhood asthma.<sup>48</sup> Additionally, exposure to other molecules; for example, allergens may be crucial, as both allergens and bacterial exposure can be needed for normal immune development.<sup>49</sup> Further, our analysis lacks information on (air) pollution exposure, which has previously been associated with the development of asthma<sup>50</sup> and is known to differ between rural and urban areas of Denmark.<sup>51,52</sup> Therefore, potential differences in the exposure to nonbacterial items could partly explain the identified differences between living in an urban or rural environment. Finally, we did not have information about mothers' behavior, environment, and health during pregnancy, which can be important. Moreover, our cohort has limited ethnic diversity, which reduces the generality of our findings.

**A Airway microbiota**

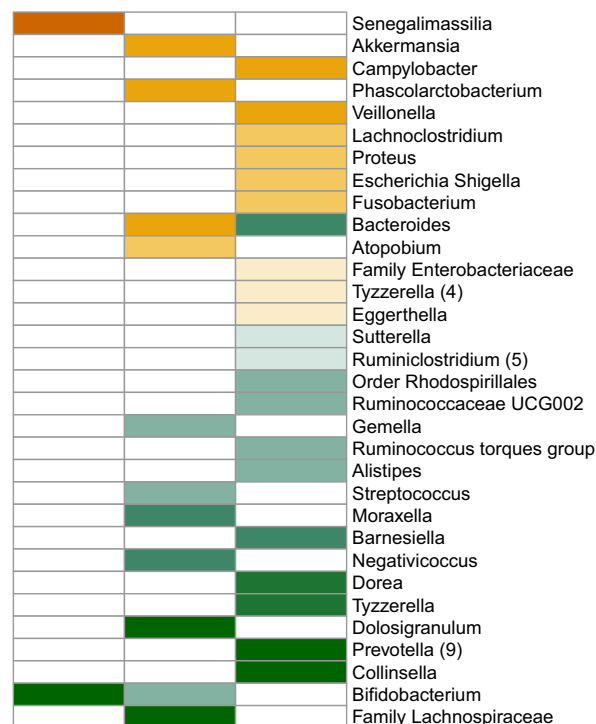
1 week 1 month 3 months



$R^2 = 0.038$   $R^2 = 0.069$   $R^2 = 0.062$   
 $P = 2.35e-09$   $P = 3.21e-14$   $P = 1.26e-10$

**B Gut microbiota**

1 week 1 month 1 year



$R^2 = 0.020$   $R^2 = 0.028$   $R^2 = 0.057$   
 $P = 0.0004$   $P = 1.01e-07$   $P = 1.61e-10$

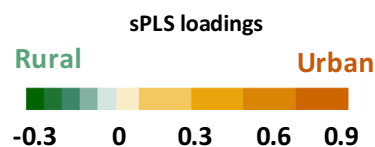


**FIG 3.** The composition of living environment-related microbial profiles for each time point. sPLS models were run for (sub)genera present in at least 5% of samples. Selection of the best model for each time point was done by comparing root mean square error values between models with differing numbers of taxa and components. Each selected model for the airway (A) and gut (B) microbiotas had 1 component. Included taxa were ordered in the heatmap by their summed loadings across all time points. Taxa with positive loadings are more common in urban (gold) living environments, whereas taxa with negative loadings are more common in rural (green) living environments. Taxa colored white are not included in the present model.

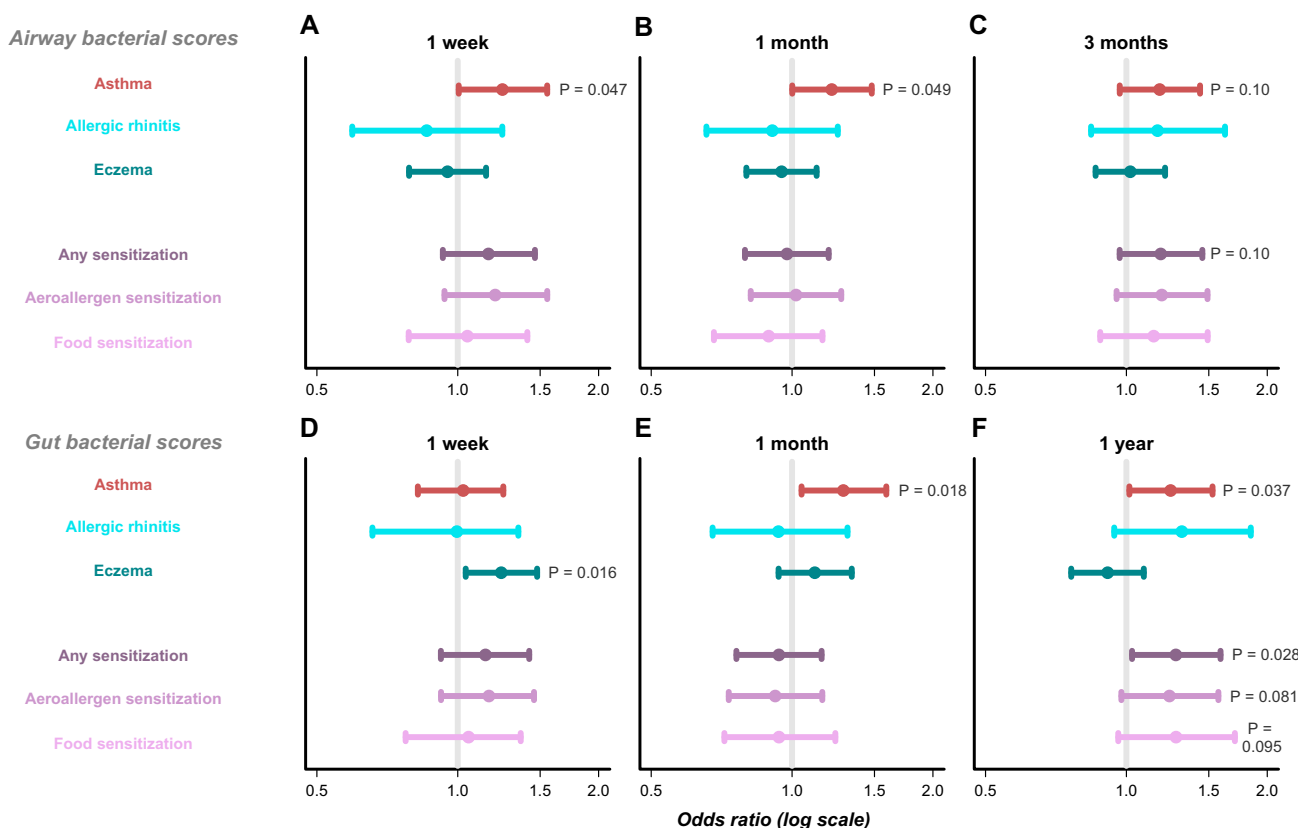


**FIG 4.** Airway and gut bacterial scores indicative of living environment and the risk of asthma and atopic traits at 6 years of age. Bacterial scores were scaled to an SD of 1 to ease interpretation. For airways (A–C), the risk of development of asthma increases with higher values of bacterial score (ie, urban-like bacterial profiles at 1 week and 1 month of age). For gut (D–F), eczema is associated with the bacterial score from 1 week (D), whereas the 1-month and 1-year time points are important for asthma development (E). The risk of sensitization increases with higher values of the gut bacterial score at 1 year of age (F). Error bars represent the 95% CI.

## Interpretation

Exposure to farming-associated indoor microbes (ie, the “external” microbiota early in life) is associated with asthma protection.<sup>13–16,53</sup> Here, we focused on the “internal” microbiota (ie, the infant’s own microbiota) in rural and urban children and showed corresponding findings, especially for the airway microbiota. Recently, living on a farm has been shown to support the development of gut microbiota in infants, which can decrease the risk of asthma.<sup>25</sup> These associations are also suggested by the biodiversity and old friends hypotheses.<sup>54,55</sup> External and internal microbiotas are likely to have intimate relations. The airway microbiota is more open for colonization by airborne microbes than the gut is (because inhaled microbes continuously travel to the lungs), and it thus reflects the external microbial exposures. Active mucosal layers in the airways and lungs can respond to microbial exposures and shape the function of the immune system. However, as we did not sample from the gut mucosa, we cannot exclude its potential importance.

The bacteria associated with predisposing and protective effects vary between studies.<sup>56</sup> Members of urbanized bacterial communities probably act together, but *Veillonella* was an interesting genus associated with urbanization both in the airways and the gut. *Veillonella* in the gut at 1 year of age was previously associated with an increased risk of asthma in this cohort, whereas another study found a protective association.<sup>22,23</sup> In contrast to

our findings, airway colonization with *Moraxella*, an opportunistic pathogen,<sup>57</sup> was previously associated with increased risk of asthma,<sup>57–59</sup> including in the previous COPSAC<sub>2000</sub> cohort, in which culture-based methodology was applied.<sup>59</sup> We suspect that the opportunistic (pathogenic) nature of some urban and rural indicator bacteria, such as rural environment–associated *Moraxella*, *Acinetobacter*,<sup>60,61</sup> and *Mycobacterium*,<sup>62</sup> may explain their potential importance in atopic diseases, as well as their cohort specific roles in diseases.

The hypothesized window of susceptibility (ie, the period in which extrinsic microbial stimuli can permanently influence immune responsiveness) is suggested to be during the first 100 days after birth<sup>23</sup> or during the entire infancy period, including the prenatal period.<sup>18</sup> In this study, the window associated with asthma development was the first year of life, whereas the window was much narrower for eczema and allergic sensitization. Our study was not designed for defining such a window of susceptibility, however. Previously, the window associated with asthma development had been restricted to shorter periods after birth (<100 days and <12 months)<sup>23,24</sup> and a longer period for eczema (≤12 month of age),<sup>21</sup> leaving the answer to the question of when the exposure is most important unclear.

We found large disparities in the immune mediator concentrations between urban and rural children, as well as between urbanized and rural bacterial profiles. Therefore, we conclude that

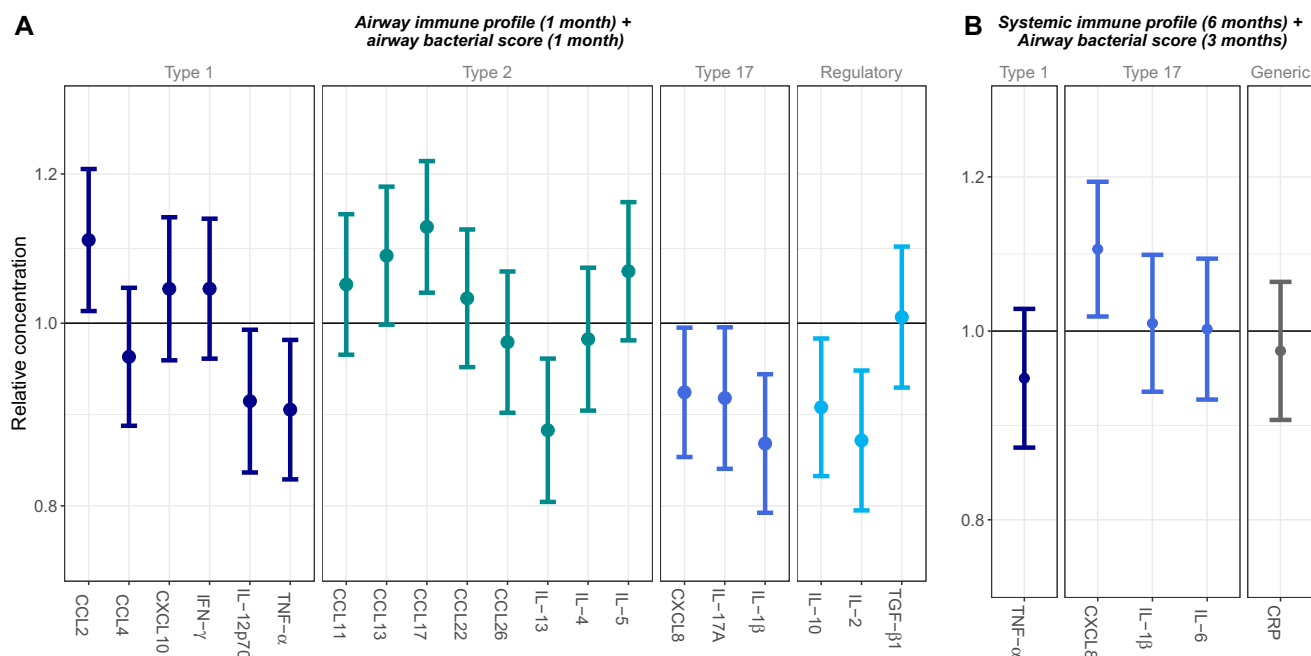


**FIG 5.** Airway bacterial score versus the relative concentration of airway and systemic immune mediators. The figure shows normalized and scaled relative concentrations of immune mediators in relation to airway bacterial score per SD. Higher values indicate association with an urban-like bacterial community composition. **A**, Topical immune mediators at 1 month of age. **B**, Systemic immune mediators at 6 months of age in response to bacterial scores. Linear models were adjusted for living environment-associated lifestyle features. Error bars represent the 95% CI. False positive rate-corrected *P* values are presented in Table E6.

the living environment, as well as associated bacteria, may influence immune function during infancy. Further, our findings showed interesting similarities to our previous study, which discovered that asthma-associated bacteria at 1 month of age were positively associated with CCL2 and CCL17 but negatively associated with TNF- $\alpha$  and IL-1 $\beta$ .<sup>38</sup> These same associations were found in response to urbanized airway bacteria at 1 month of age. However, *in vivo* cytokine and chemokine concentrations do not reflect easily delimited individual immune pathways. For example, both type 2- and type 17-related cytokines differed between rural and urban infants in a nonorchestrated manner. Some individual mediators showed expected patterns, such as an increased level of CXCL8 (a neutrophil chemotactic type 17 mediator that is enriched in moderate-to-severe asthma<sup>63</sup>) in urban children and in response to the urbanized airway microbiota. Deficiency in regulatory responses is suspected to be important for the trajectory toward hyperreactivity against allergens and development of manifest clinical allergy.<sup>64</sup> Accordingly, urban infants had lower concentrations of the anti-inflammatory, regulatory-type mediator IL-10. Having an urbanized airway microbiota was associated with decreased IL-2 levels in the airways, which is interesting, as suppressed IL-2 expression in infants has been associated with food allergy.<sup>65</sup>

There are several potential explanations for some immune mediator concentrations differing from the expected values. First, the influence of an individual mediator depends on other mediators. For example, a higher concentration of IL-13 (type 2) in response to rural airway bacteria can indicate that when combined with IL-12p70, TNF- $\alpha$ , IL-1 $\beta$ , IL-10, and IL-2, it does not provide increased susceptibility to type 2-based symptoms. Second, proinflammatory immune mediators can be beneficial

during certain developmental stages (eg, higher IL-4, IL-5, and IL-13 levels in 1-month-old rural infants). Our results underscore the fact that the current understanding of an adequately stimulated immune system during different developmental stages is not complete.

## Conclusions

We found that an urbanized living environment, an urbanized microbiota in infants, *in vivo* immune mediators, and the risk of childhood asthma and atopic traits were associated with one another in a geographically concise cohort of 700 unselected children. Our findings support an important role of the infant microbiota in the association between urbanization and disease, probably via cross talk with the developing immune system.

We express our deepest gratitude to the children and families of the COPSAC<sub>2010</sub> cohort study for all their support and commitment. We acknowledge and appreciate the unique efforts of the COPSAC research team.

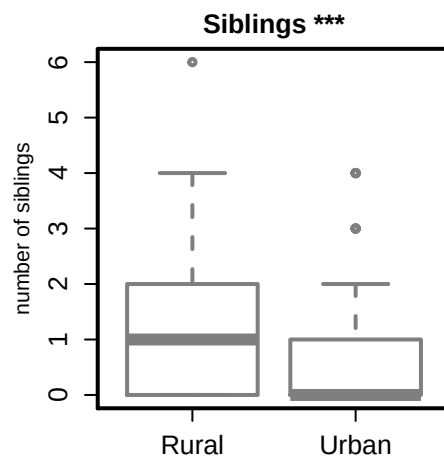
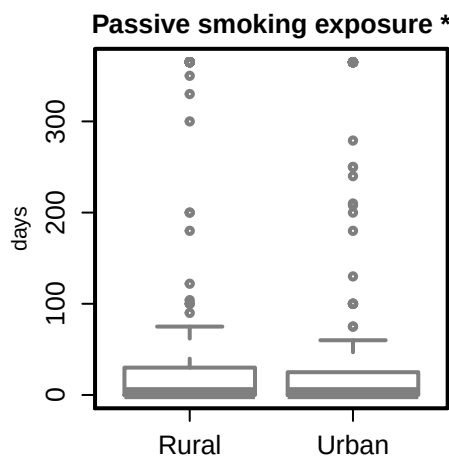
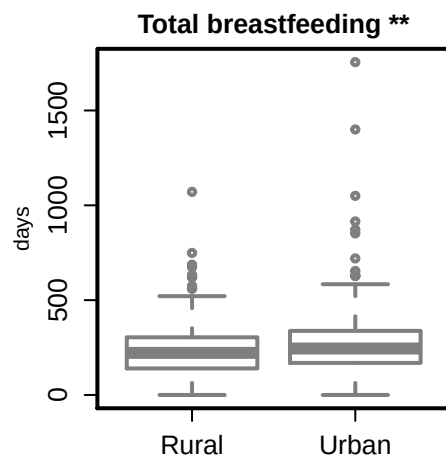
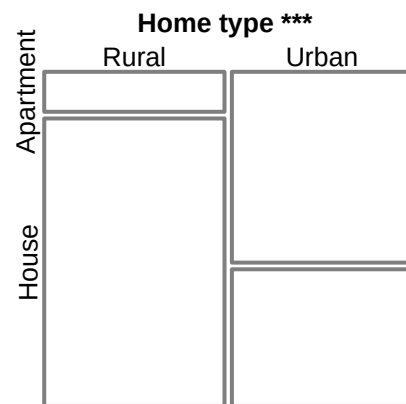
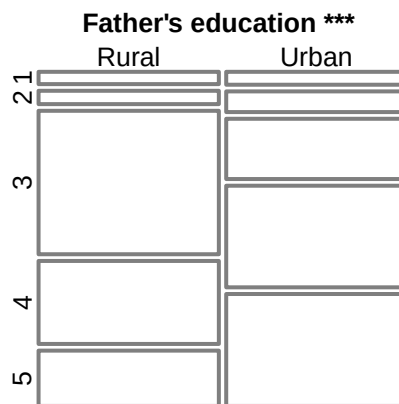
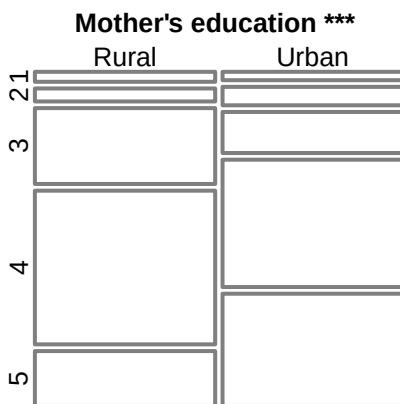
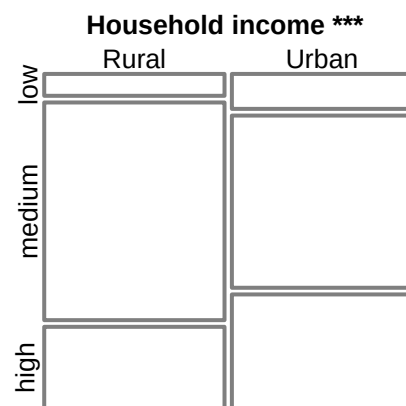
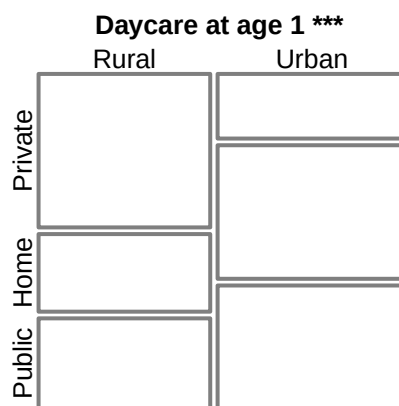
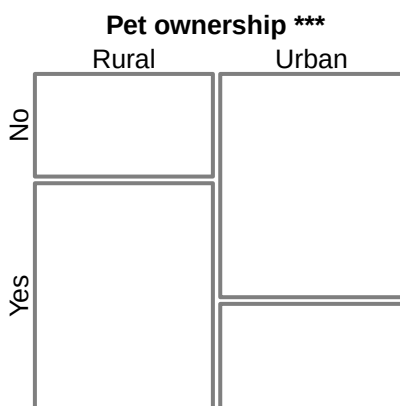
## Key messages

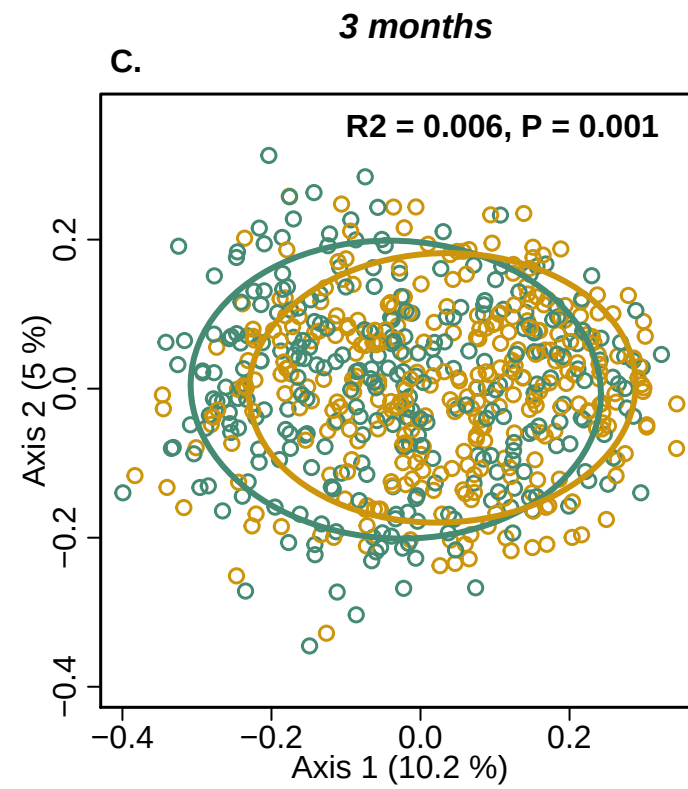
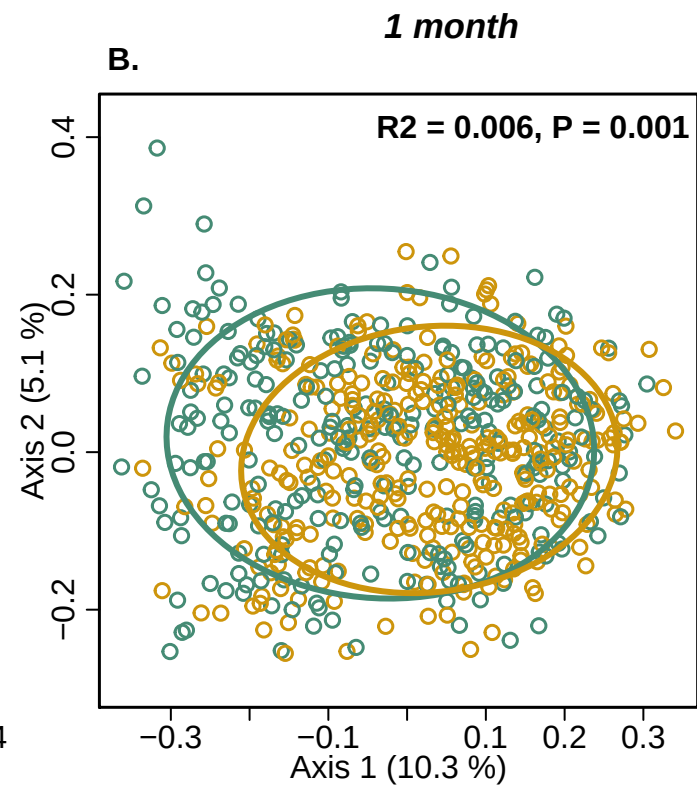
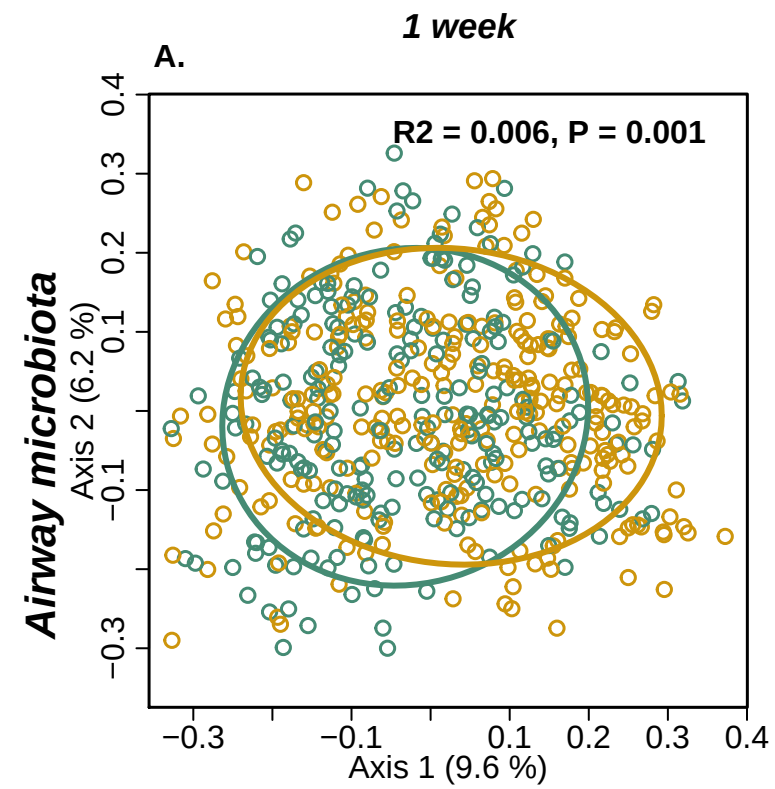
- Airway and gut microbiotas, as well as immune marker concentrations, differ between rural and urban infants.
- An urbanized composition of the microbiota in infancy increases the risk of asthma, eczema, and allergic sensitization at 6 years of age.
- Urban environments may predispose to the development of a disease-promoting composition of the infant microbiota.

## REFERENCES

- Kozyrskyj AL, Becker AB. Rural-urban differences in asthma prevalence: possible explanations. *J Allergy Clin Immunol* 2004;113:S306.
- Rodriguez A, Vaca M, Oviedo G, Erazo S, Chico ME, Teles C, et al. Urbanisation is associated with prevalence of childhood asthma in diverse, small rural communities in Ecuador. *Thorax* 2011;66:1043-50.
- Ruokolainen L, von Hertzen L, Fyhrquist N, Laatikainen T, Lehtimäki J, Auvinen P, et al. Green areas around homes reduce atopic sensitization in children. *Allergy* 2015;70:195-202.
- Christensen SH, Timm S, Janson C, Benediktsdóttir B, Forsberg B, Holm M, et al. A clear urban-rural gradient of allergic rhinitis in a population-based study in Northern Europe. *Eur Clin Respir J* 2016;3:33463.
- Hajar T, Simpson EL. The rise in atopic dermatitis in young children: what is the explanation? *JAMA Netw Open* 2018;1:e184205.
- von Mutius E, Smits HH. Primary prevention of asthma: from risk and protective factors to targeted strategies for prevention. *Lancet* 2020;396:854-66.
- Strachan DP, Ait-Khaled N, Foliaki S, Mallol J, Odhiambo J, Pearce N, et al. Siblings, asthma, rhinoconjunctivitis and eczema: a worldwide perspective from the International Study of Asthma and Allergies in Childhood. *Clin Exp Allergy* 2015;45:126-36.
- Ball TM, Castro-Rodriguez JA, Griffith KA, Holberg CJ, Martinez FD, Wright AL. Siblings, day-care attendance, and the risk of asthma and wheezing during childhood. *N Engl J Med* 2000;343:538-43.
- Barberán A, Ladau J, Leff JW, Pollard KS, Menninger HL, Dunn RR, et al. Continental-scale distributions of dust-associated bacteria and fungi. *Proc Natl Acad Sci U S A* 2015;112:5756-61.
- Gupta S, Hjelmsø MH, Lehtimäki J, Li X, Mortensen MS, Russel J, et al. Environmental shaping of the bacterial and fungal community in infant bed dust and correlations with the airway microbiota. *Microbiome* 2020;8:115.
- Lehtimäki J, Karkman A, Laatikainen T, Paalanen L, von Hertzen L, Haahtela T, et al. Patterns in the skin microbiota differ in children and teenagers between rural and urban environments. *Sci Rep* 2017;7:45651.
- Ruokolainen L, Paalanen L, Karkman A, Laatikainen T, von Hertzen L, Vlasoff T, et al. Significant disparities in allergy prevalence and microbiota between the young people in Finnish and Russian Karelia. *Clin Exp Allergy* 2017;47:665-74.
- Ege MJ, Mayer M, Normand A-C, Genuit J, Cookson WOCM, Braun-Fahrlander C, et al. Exposure to environmental microorganisms and childhood asthma. *N Engl J Med* 2011;364:701-9.
- Stein MM, Hrusch CL, Gozdz J, Igartua C, Pivniouk V, Murray SE, et al. Innate immunity and asthma risk in Amish and Hutterite farm children. *N Engl J Med* 2016;375:411-21.
- Kirjavainen PV, Karvonen AM, Adams RI, Täubel M, Roponen M, Tuoresmäki P, et al. Farm-like indoor microbiota in non-farm homes protects children from asthma development. *Nat Med* 2019;1.
- Schuijs MJ, Willart MA, Vergote K, Gras D, Deswarte K, Ege MJ, et al. Farm dust and endotoxin protect against allergy through A20 induction in lung epithelial cells. *Science* 2015;349:1106-10.
- Ottman N, Ruokolainen L, Suomalainen A, Sinkko H, Karisola P, Lehtimäki J, et al. Soil exposure modifies the gut microbiota and supports immune tolerance in a mouse model. *J Allergy Clin Immunol* 2019;143:1198-206.e12.
- Renz H, Adkins BD, Bartfeld S, Blumberg RS, Farber DL, Garssen J, et al. The neonatal window of opportunity—early priming for life. *J Allergy Clin Immunol* 2018;141:1212-4.
- Belkaid Y, Hand T. Role of the microbiota in immunity and inflammation. *Cell* 2014;157:121-41.
- Bisgaard H, Li N, Bonnelykke K, Chawes BLK, Skov T, Paludan-Müller G, et al. Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. *J Allergy Clin Immunol* 2011;128:646-52.e5.
- Abrahamsson TR, Jakobsson HE, Andersson AF, Björkstén B, Engstrand L, Jenmalm MC. Low diversity of the gut microbiota in infants with atopic eczema. *J Allergy Clin Immunol* 2012;129:434-40.e1-2.
- Stokholm J, Blaser MJ, Thorsen J, Rasmussen MA, Waage J, Vinding RK, et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat Commun* 2018;9:141.
- Arrieta M-C, Stiemsma LT, Dimitriu PA, Thorsen L, Russell S, Yurist-Doutsch S, et al. Early infancy microbial and metabolic alterations affect risk of childhood asthma. *Sci Transl Med* 2015;7:307ra152.
- Abrahamsson TR, Jakobsson HE, Andersson AF, Björkstén B, Engstrand L, Jenmalm MC. Low gut microbiota diversity in early infancy precedes asthma at school age. *ClinExp Allergy* 2014;44:842-50.
- Depner M, Taft DH, Kirjavainen PV, Kalanetra KM, Karvonen AM, Peschel S, et al. Maturation of the gut microbiome during the first year of life contributes to the protective farm effect on childhood asthma. *Nat Med* 2020;26:1766-75.
- Bisgaard H, Vissing NH, Carson CG, Bischoff AL, Følsgaard NV, Kreiner-Møller E, et al. Deep phenotyping of the unselected COPSAC2010 birth cohort study. *Clin Exp Allergy* 2013;43:1384-94.
- Bisgaard H, Stokholm J, Chawes BL, Vissing NH, Bjarnadóttir E, Schoos A-MM, et al. Fish oil-derived fatty acids in pregnancy and wheeze and asthma in offspring. *N Engl J Med* 2016;375:2530-9.
- Chawes BL, Bonnelykke K, Stokholm J, Vissing NH, Bjarnadóttir E, Schoos A-MM, et al. Effect of vitamin D3 supplementation during pregnancy on risk of persistent wheeze in the offspring: a randomized clinical trial. *JAMA* 2016;315:353-61.
- Bisgaard H, Phipps CB, Bonnelykke K. Endotyping early childhood asthma by quantitative symptom assessment. *J Allergy Clin Immunol* 2011;127:1155-64.e2.
- Hanifin J, Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol* 1980;92:44-7.
- Cooley D. googleway: accesses Google Maps APIs to Retrieve Data and Plot Maps. R package version 2.7.1 2018. Available at: <https://CRAN.R-project.org/package=googleway>. Accessed October 1, 2018.
- Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGinn D, et al. vegan: community ecology package 2019. Available at: <https://CRAN.R-project.org/package=vegan>. Accessed November 1, 2018.
- Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: cluster analysis basics and extensions. R package version 2.0.7-1 2018. Available at <https://cran.r-project.org/web/packages/cluster/>. Accessed December 1, 2018.
- Mortensen MS, Breyer AD, Roggenbuck M, Abu Al-Soud W, Balle C, Krogfelt KA, et al. The developing hypopharyngeal microbiota in early life. *Microbiome* 2016;4:70.
- Chawes BLK, Edwards MJ, Shamji B, Walker C, Nicholson GC, Tan AJ, et al. A novel method for assessing unchallenged levels of mediators in nasal epithelial lining fluid. *J Allergy Clin Immunol* 2010;125:1387-9.e3.
- Chawes BL, Stokholm J, Bonnelykke K, Brix S, Bisgaard H. Neonates with reduced neonatal lung function have systemic low-grade inflammation. *J Allergy Clin Immunol* 2015;135:1450-66.
- Wolsk HM, Chawes BL, Thorsen J, Stokholm J, Bonnelykke K, Brix S, et al. Noninvasive sampling of mucosal lining fluid for the quantification of in vivo upper airway immune-mediator levels. *J Vis Exp* 2017;126:55800.
- Thorsen J, Rasmussen MA, Waage J, Mortensen M, Breyer AD, Bonnelykke K, et al. Infant airway microbiota and topical immune perturbations in the origins of childhood asthma. *Nat Commun* 2019;10:1-8.
- McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;8:e61217.
- Bray JR, Curtis JT. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs* 1957;27:325-49.
- Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative  $\beta$  diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol* 2007;73:1576-85.
- Lahti L. microbiome R package. Available at: <http://microbiome.github.io>. Accessed February 1, 2019.
- Rohart F, Gautier B, Singh A, Cao K-AL. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol* 2017;13:e1005752.
- R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>. Accessed September 1, 2018.
- van Nimwegen FA, Penders J, Stobberingh EE, Postma DS, Koppelman GH, Kerkhof M, et al. Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy. *J Allergy Clin Immunol* 2011;128:948-55.e1-3.
- Russell SL, Gold MJ, Hartmann M, Willing BP, Thorson L, Wlodarska M, et al. Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma. *EMBO Rep* 2012;13:440-7.
- Cecchi L, D'Amato G, Annesi-Maesano I. External exposome and allergic respiratory and skin diseases. *J Allergy Clin Immunol* 2018;141:846-57.
- Sharpe RA, Bearman N, Thornton CR, Husk K, Osborne NJ. Indoor fungal diversity and asthma: a meta-analysis and systematic review of risk factors. *J Allergy Clin Immunol* 2015;135:110-22.
- Lynch SV, Wood RA, Boushey H, Bacharier LB, Bloomberg GR, Kattan M, et al. Effects of early-life exposure to allergens and bacteria on recurrent wheeze and atopy in urban children. *J Allergy Clin Immunol* 2014;134:593-601.e12.
- Carlsten C, Rider CF. Traffic-related air pollution and allergic disease: an update in the context of global urbanization. *Curr Opin Allergy Clin Immunol* 2017;17:85-9.
- Brønnum-Hansen H, Bender AM, Andersen ZJ, Sørensen J, Bønløkke JH, Boshuizen H, et al. Assessment of impact of traffic-related air pollution on morbidity and mortality in Copenhagen Municipality and the health gain of reduced exposure. *Environ Int* 2018;121:973-80.

52. Fauser P, Ketzel M, Becker T, Plejdrup MS, Brandt J, Gidhagen L, et al. Human exposure to carcinogens in ambient air in Denmark, Finland and Sweden. *Atmos Environ* 2017;167:283-97.
53. von Mutius E, Vercelli D. Farm living: effects on childhood asthma and allergy. *Nat Rev Immunol* 2010;10:861-8.
54. von Hertzen L, Hanski I, Haahtela T. Natural immunity. Biodiversity loss and inflammatory diseases are two global megatrends that might be related. *EMBO Rep* 2011;12:1089-93.
55. Rook GA. Regulation of the immune system by biodiversity from the natural environment: an ecosystem service essential to health. *Proc Natl Acad Sci U S A* 2013;110:18360-7.
56. Karkman A, Lehtimäki J, Ruokolainen L. The ecology of human microbiota: dynamics and diversity in health and disease. *Ann N Y Acad Sci* 2017;1399:78-92.
57. Larsen JM, Brix S, Thyssen AH, Birch S, Rasmussen MA, Bisgaard H. Children with asthma by school age display aberrant immune responses to pathogenic airway bacteria as infants. *J Allergy Clin Immunol* 2014;133:1008-13.
58. Chung KF. Airway microbial dysbiosis in asthmatic patients: a target for prevention and treatment? *J Allergy Clin Immunol* 2017;139:1071-81.
59. Bisgaard H, Hermansen MN, Buchvald F, Loland L, Halkjaer LB, Bønnelykke K, et al. Childhood asthma after bacterial colonization of the airway in neonates. *N Engl J Med* 2007;357:1487-95.
60. Hanski I, von Hertzen L, Fyhrquist N, Koskinen K, Torppa K, Laatikainen T, et al. Environmental biodiversity, human microbiota, and allergy are interrelated. *Proc Natl Acad Sci U S A* 2012;109:8334-9.
61. Debarry J, Hanuszkiewicz A, Stein K, Holst O, Heine H. The allergy-protective properties of *Acinetobacter lwoffii* F78 are imparted by its lipopolysaccharide. *Allergy* 2010;65:690-7.
62. Hopkin JM. Atopy, asthma, and the mycobacteria. *Thorax* 2000;55:443-5.
63. Gras D, Martinez-Anton A, Bourdin A, Garulli C, de Senneville L, Vachier I, et al. Human bronchial epithelium orchestrates dendritic cell activation in severe asthma. *Eur Respir J* 2017;49.
64. Lambrecht BN, Hammad H. The immunology of asthma. *Nat Immunol* 2015;16:45-56.
65. Zhang Y, Collier F, Naselli G, Saffery R, Tang ML, Allen KJ, et al. Cord blood monocyte-derived inflammatory cytokines suppress IL-2 and induce nonclassic "TH2-type" immunity associated with development of food allergy. *Sci Transl Med* 2016;8:321ra8.

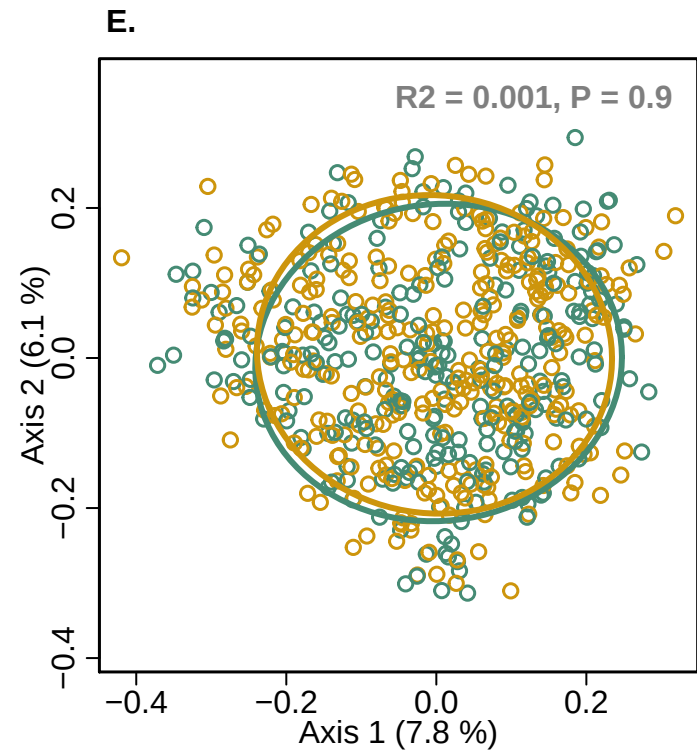
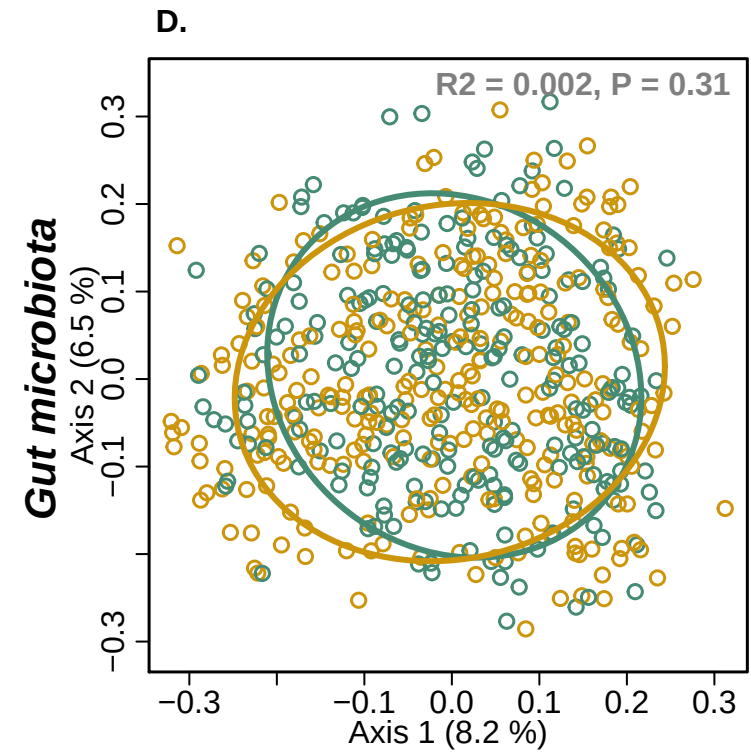




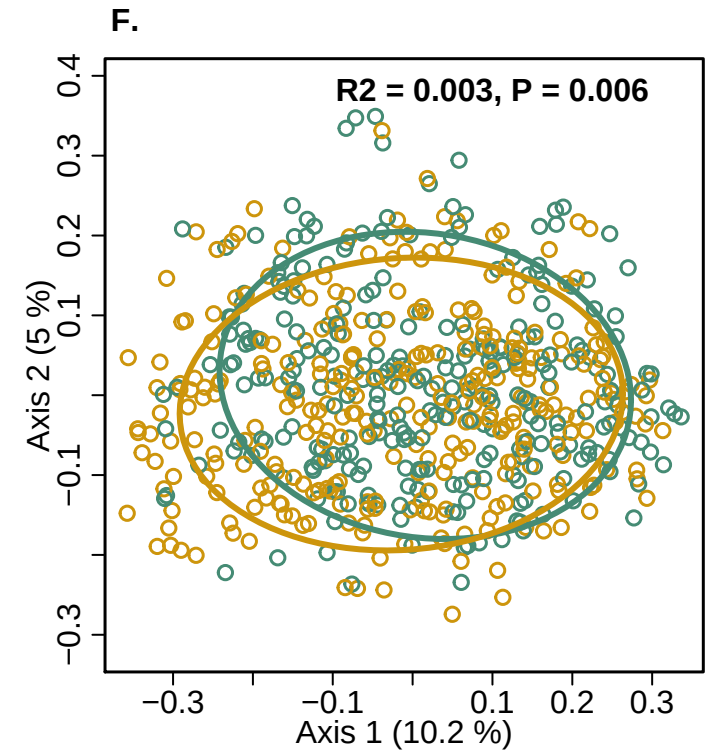
**1 year**

Rural  
Urban

NA

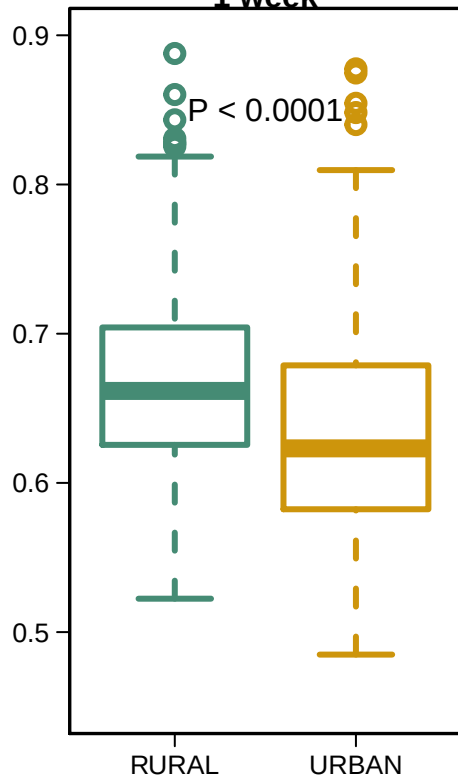


NA

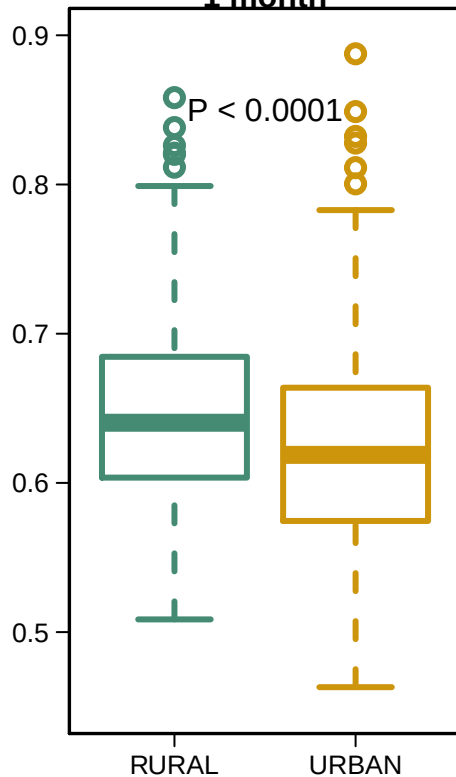


*Bray-curtis dissimilarity from the group mean*

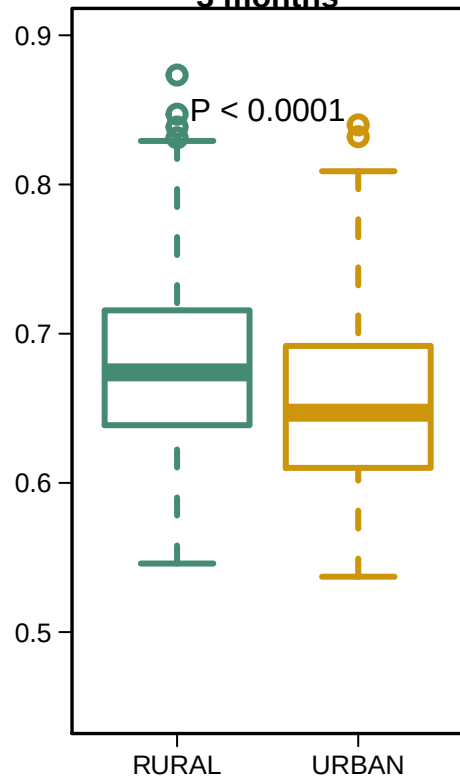
**1 week**

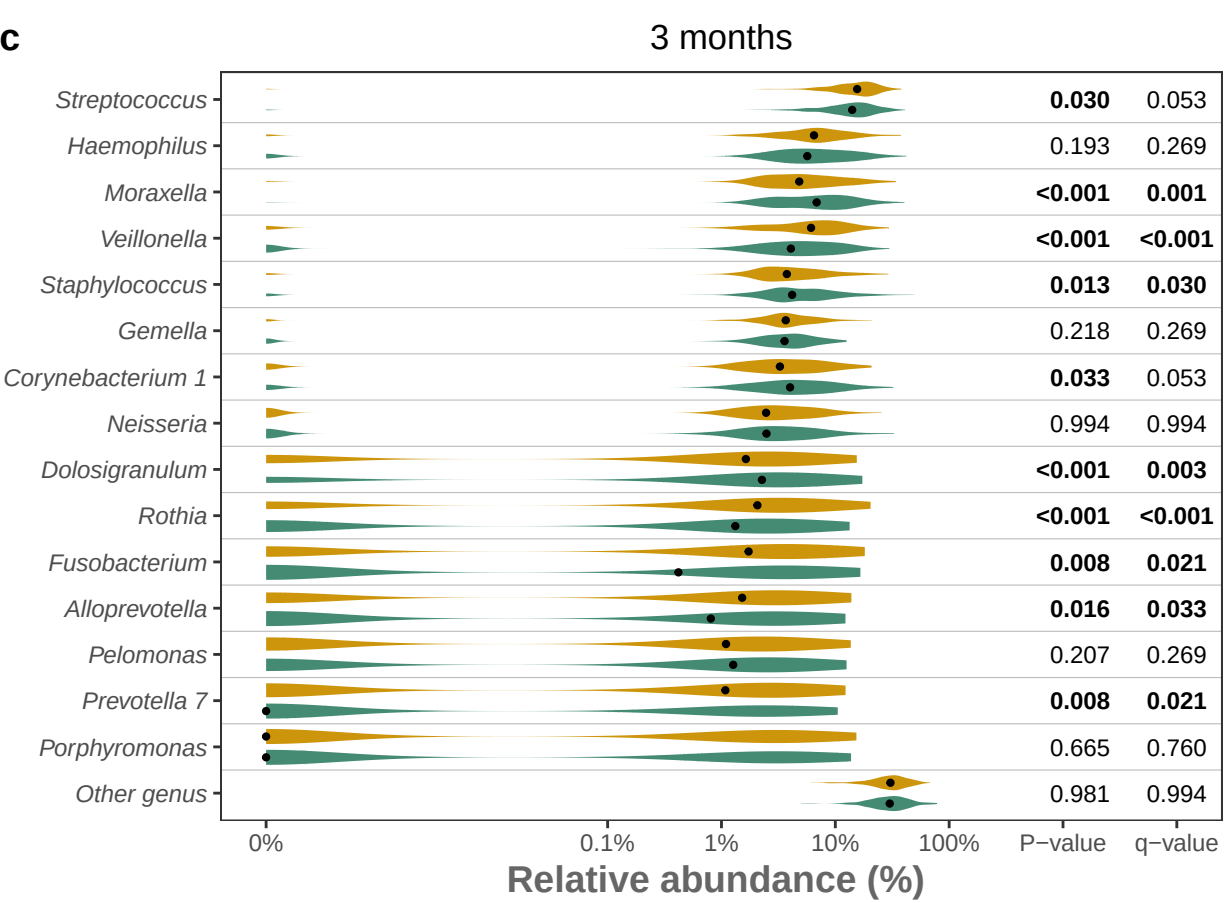
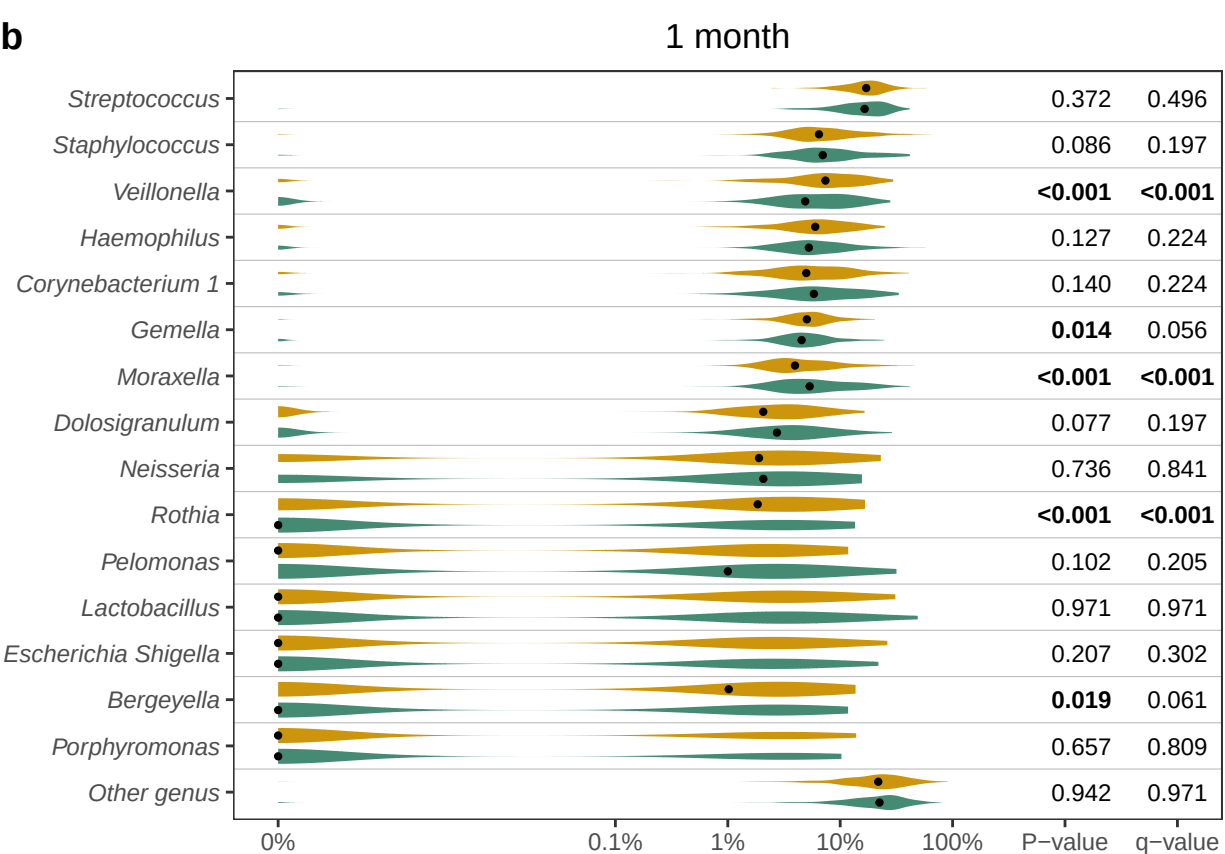
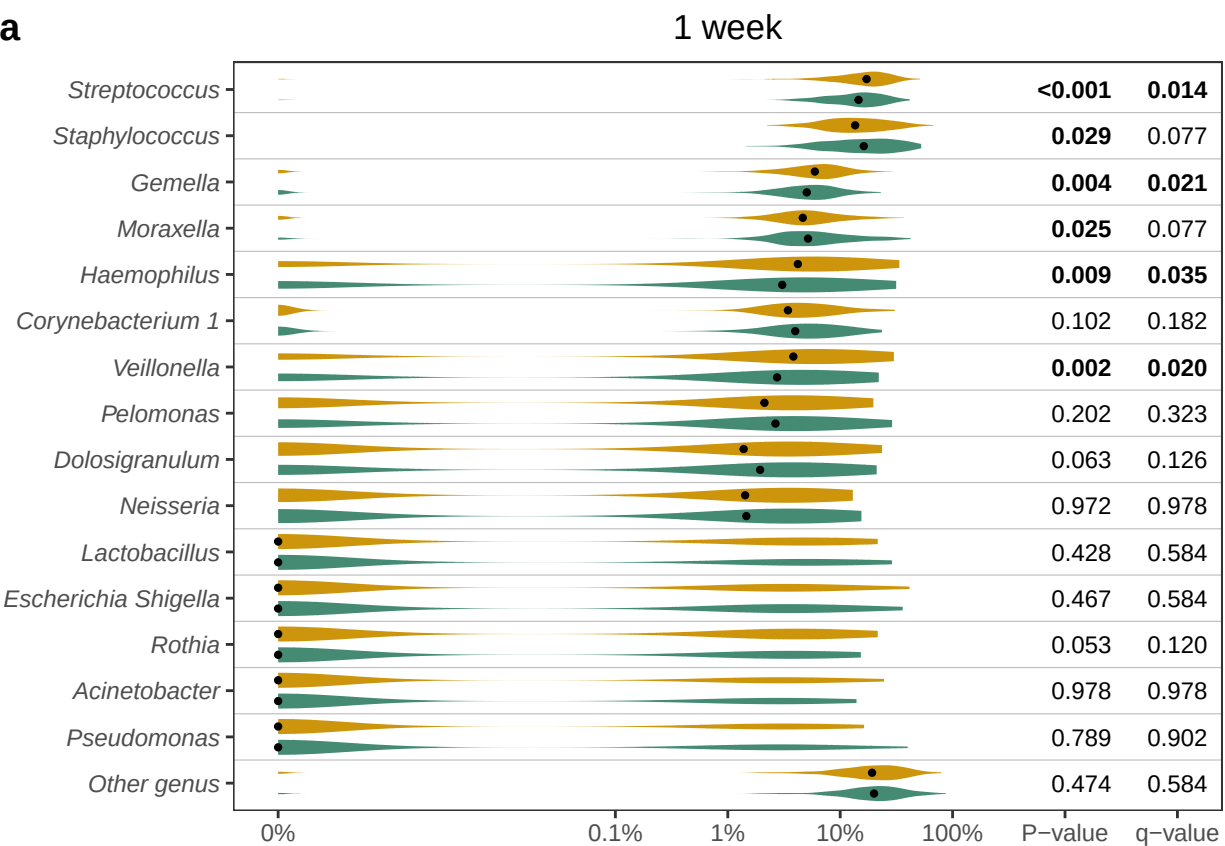


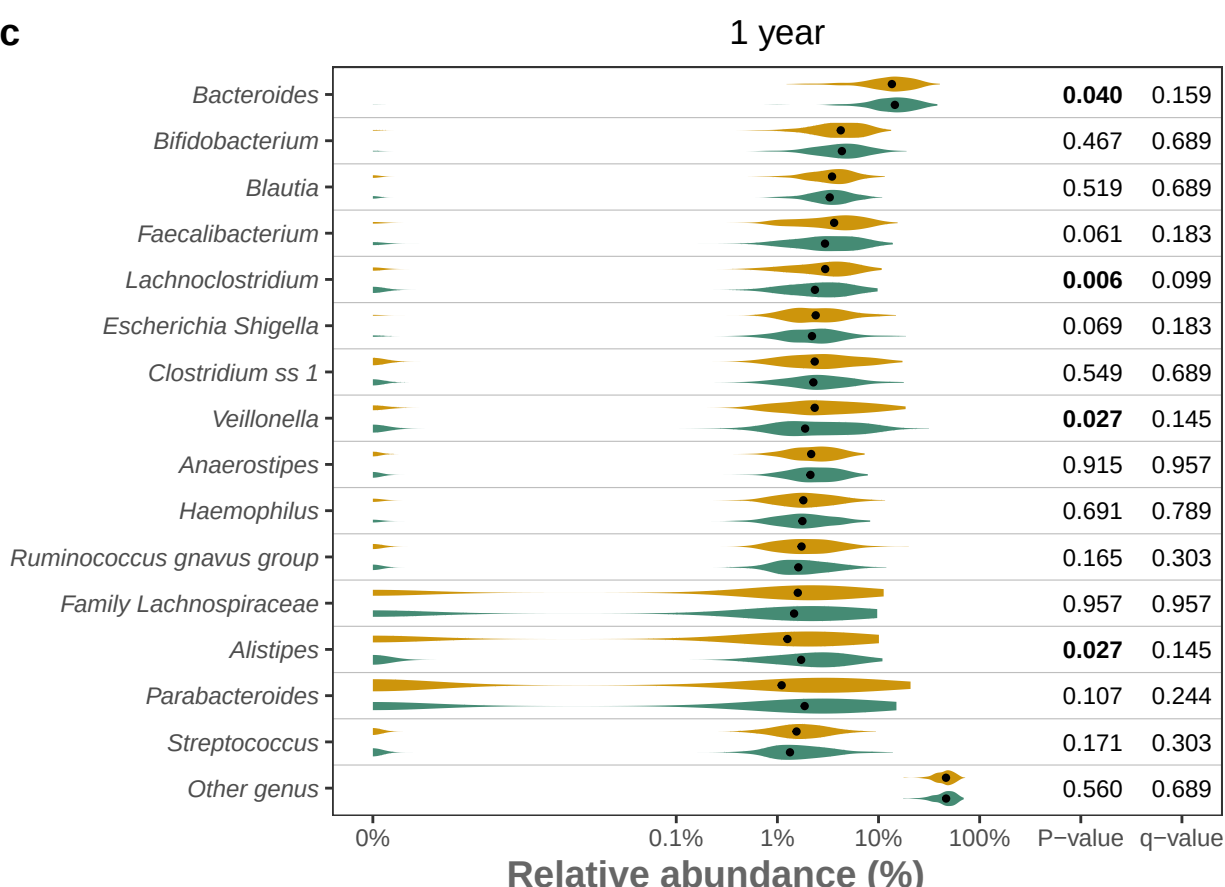
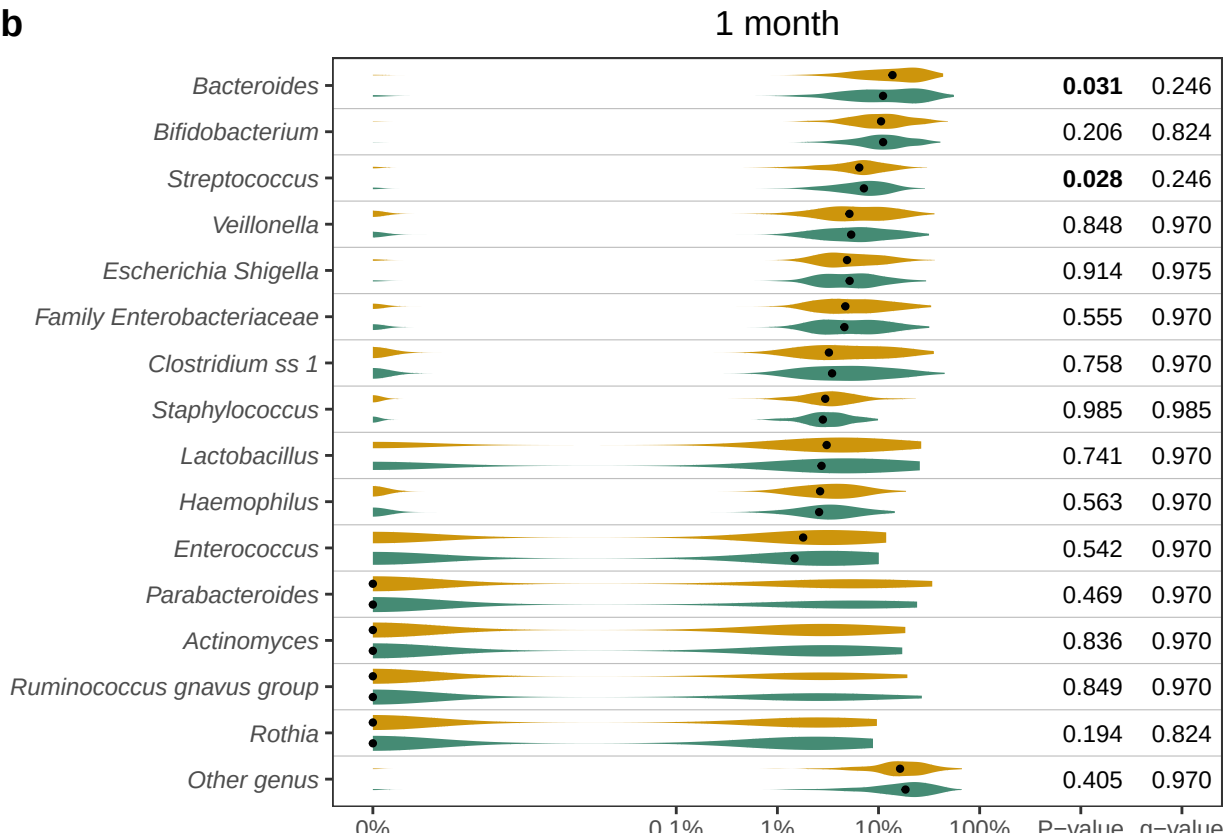
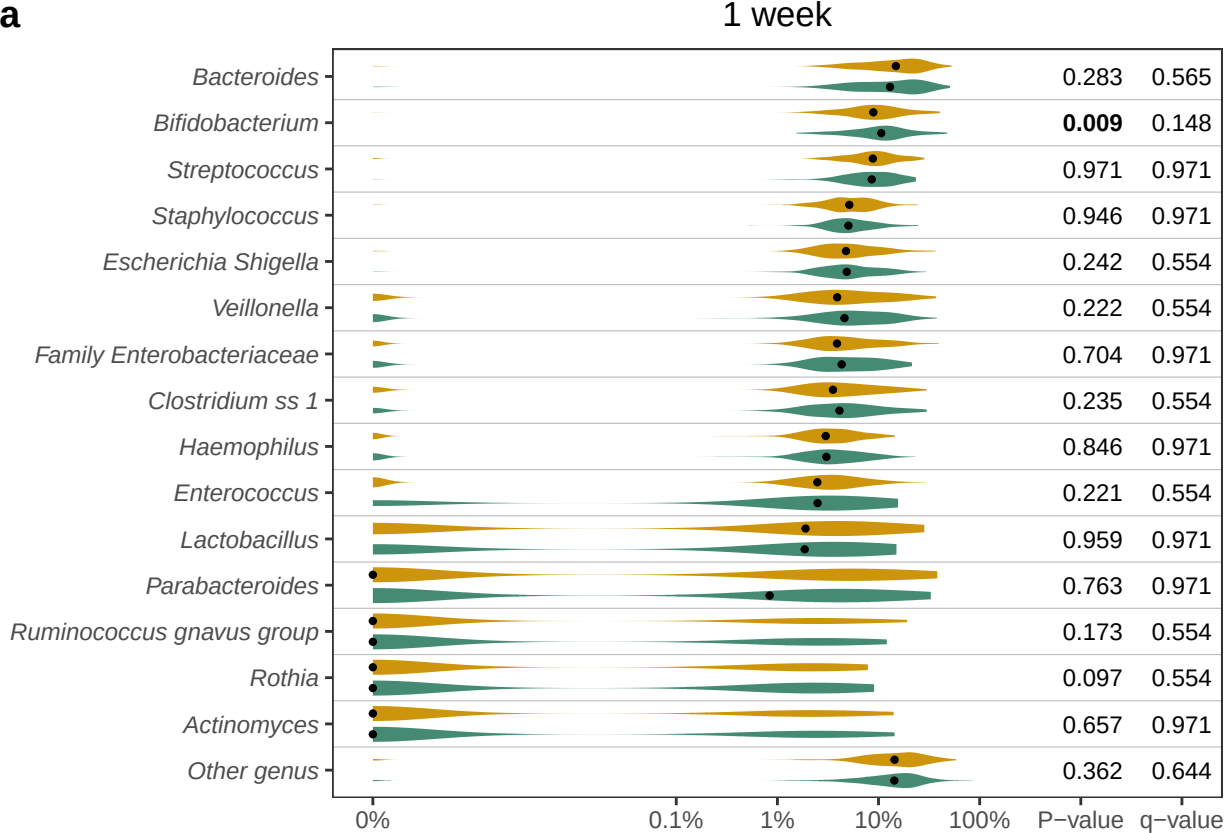
**1 month**

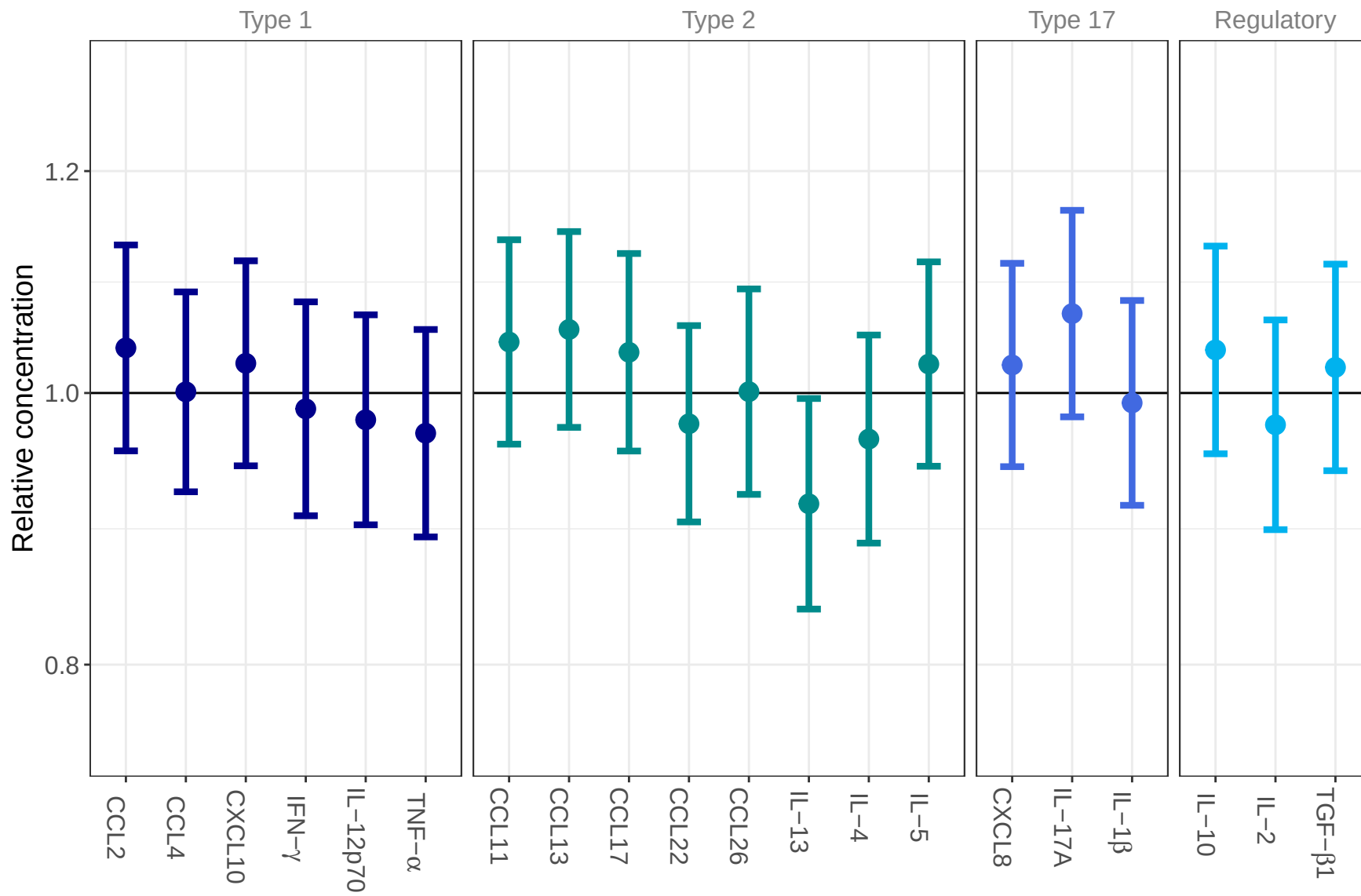
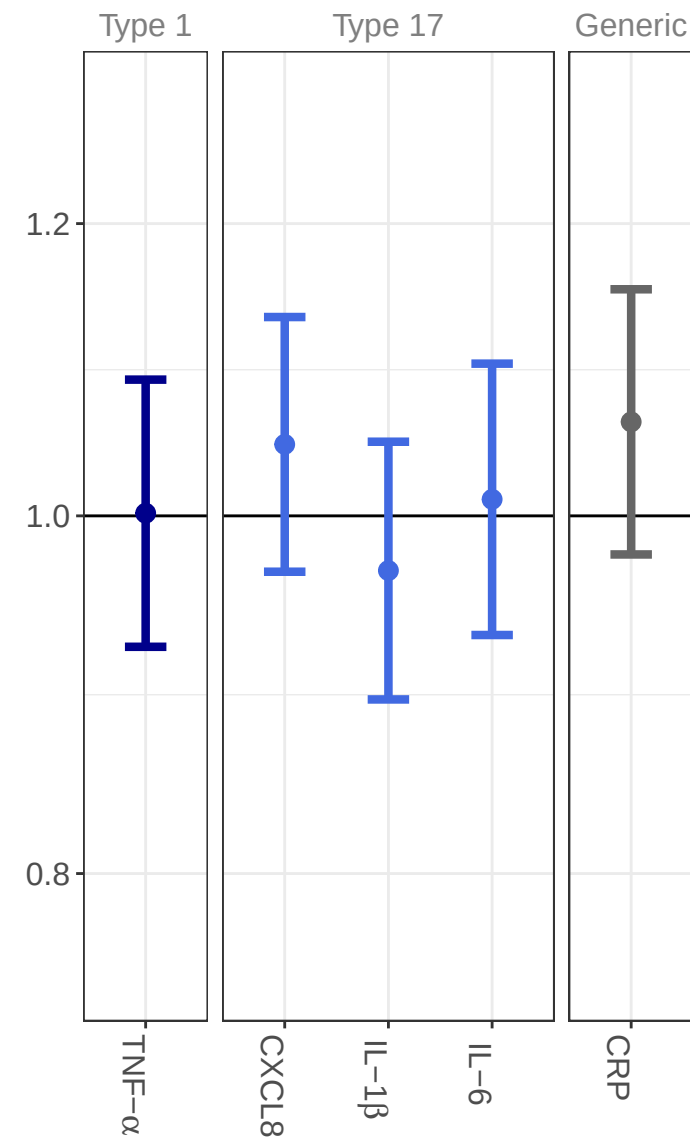


**3 months**







**a****Airway immune profile (1 month) +  
gut bacterial score (1 month)****b****Systemic immune profile (6 months) +  
gut bacterial score (1 month)**

## ONLINE REPOSITORY INFORMATION

### ONLINE REPOSITORY METHODS

#### Definition of diseases

**Asthma** before or during age 6 years was diagnosed prospectively based on a previously detailed quantitative symptom algorithm<sup>1,2</sup>, requiring all of the following: (1) recurrent episodes of troublesome lung symptoms (5 episodes within 6 months, each lasting at least 3 consecutive days, 4 weeks of continuous symptoms, or an acute severe exacerbation requiring oral prednisone or hospital admission); (2) symptomatology typical of asthma, including exercise induced symptoms, prolonged nocturnal cough, and persistent cough outside common cold; (3) need for intermittent rescue use of inhaled  $\beta_2$ -agonist; and (4) response to a 3-month course of inhaled corticosteroids and relapse upon ended treatment.

**Allergic sensitization** was determined at age 6 years, defined as any positive skin prick test (SPT)  $\geq 3$  mm (ALK-Abello, Horsholm, Denmark) or specific IgE (sIgE) in serum  $\geq 0.35$  kUa/L against common inhalant (birch, grass, mugwort, horse, dog, cat, *D. pteronyssinus*, *D. farinae*, *Aspergillus fumigatus*, *Cladosporium herbarum*, *Penicillium notatum*, *Alternaria alternata*) and/or food allergens (milk, egg, wheat flour, soybean, peanut, cod, shrimp, rye flour, pork, and beef) (ImmunoCAP; Thermo Fisher Scientific, Allerod, Denmark). Children classified as “not sensitized” were both SPT and sIgE negative for all tested allergens.

#### Definition of urbanization gradient

The European Environment Agency maintains the land cover database CORINE (<https://land.copernicus.eu/>), covering European Union. Database describes the land cover at three specification levels from which we utilized the highest hierarchical level including agricultural land, artificial areas, forested and semi-natural areas, water bodies and wetlands in order to characterize the main patterns in the land cover. Data from the year 2012 was selected due to the close correspondence to the birth years of children (2008-11).

Addresses at birth ( $n=686$ ) were translated to coordinates using Google API services through the `googleway` package for R<sup>3</sup>. Most children ( $n=577$ ) did not move during their first year of life or only moved later than 50 weeks after birth ( $n=5$ ). From the 109 children who did move, 77% did not change between rural and urban groups (see below). In order to remove any noise arising from moving, we excluded all children who changed between rural and urban areas prior to sample collection ( $n_{1\text{week}}=1$ ,  $n_{1\text{month}}=3$ ,  $n_{3\text{months}}=8$ ,  $n_{6\text{months}}=15$ ,  $n_{1\text{year}}=24$ ). Land cover types around each coordinate point were extracted from

CORINE2012 raster map (having 100 m resolution) for three different buffers with radiuses of three kilometers, one kilometer and 500 meters.

The proportions of land cover types within buffer regions were analyzed with Principal Coordinate Analysis (PCA, vegan R-package<sup>4</sup>) utilizing scaling built in to function rda (produces correlation matrix). The first axis explained 44% of the total variation and made clear separation between artificial and natural i.e. agricultural and forested land cover types (**Fig S1a**). The first axis was extracted for downstream analyses to represent a gradient of urbanization running from the most rural to the most urban environments. Land cover information was also classified to two groups ( $n_{\text{rural}}=324$ ,  $n_{\text{urban}}=376$ ; **Fig S1b**) with Partitioning Around Medoids clustering (PAM, cluster R-package<sup>5</sup>). A silhouette width of 0.61 indicated clear separation between rural and urban classes. Five hundred meters and one kilometer buffer radiuses decreased the proportion of rural infants compared to three kilometer buffer radius indicating that in highly populated study region, areas close to home are often built (**Table S1**). As we were interested about comparing rural and urban children, we considered that larger buffer size makes more truthful distinction between areas as noted earlier<sup>6</sup>.

### **Collection and analysis of microbiota samples**

Airway microbiota samples were collected at one week ( $n=549$ ), one ( $n=647$ ) and three months ( $n=657$ ) of age<sup>7</sup>. Fluid was aspirated with a soft catheter passed through the nose to the hypopharyngeal region. Catheters were immediately flushed with one ml of sterile 0.9 % NaCl solution and stored at -80 °C degrees.

Gut microbiota samples were collected at one week ( $n=544$ ), one month ( $n=615$ ) and one year ( $n=625$ ) of age<sup>8</sup>. Sampling happened either at the COPSAC clinic during visits or at home by the parents following detailed instructions. Each sample was mixed with 1 mL of 10 % vol/vol glycerol broth (Statens Serum Institut, Copenhagen, Denmark) and stored at -80 °C degrees.

DNA was extracted from airway and gut samples using the PowerMag® Soil DNA Isolation Kit optimized for epMotion® (MO-BIO Laboratories, Inc., Carlsberg, CA, USA) using the epMotion® robotic platform model 5075 (Eppendorf, Hamburg, Germany). The manufacturer's protocol was mildly modified (described earlier<sup>7</sup>). Extracted DNA was stored at -20 °C degrees prior to amplification of the 16S rRNA gene variable region 4 (V4) and sequencing with the Illumina MiSeq System (Illumina Inc., CA, USA) as previously described<sup>9</sup>.

### **Bioinformatics**

Sequencing adapters were removed using Cutadapt v1.15<sup>10</sup>. Reads were analyzed using QIIME 2 v2018.2.0<sup>11</sup> and denoised using DADA2<sup>12</sup>, including removal of the first 8 bp from both reads as well as trimming of forward reads to 180 bp and reverse reads to 160 bp. Resulting amplicon sequence variants (ASVs) were compared to the 99% identity clustered SILVA database v132<sup>13</sup>

using a naive Bayes classifier<sup>14</sup> trained on the amplified region. All sequencing data is deposited at the Sequence Read Archive (SRA) with the accession numbers PRJNA340273 and PRJNA417357.

Prior to downstream analyses, taxa with less than 5 reads in the data were removed. All microbiota samples were further log transformed to reduce bias from sequencing depth variation.

## Statistics

Odds ratios for atopic traits in response to rural (baseline) and urban classes were calculated with logistic regression models adjusted for lifestyle variables. Principal Coordinate Analysis (PCoA) with Gower's distance (cluster R-package<sup>5</sup>) was used to simplify the collinear variables household income, maternal and paternal education levels to a single variable (socio-economics) for adjustment. Adjustment also included pet ownership, older siblings, passive smoking exposure, daycare attendance at age one year, and duration of breastfeeding while house type was excluded due to collinearity with the explanatory variable of interest (urban / rural groups). The sensitization and clinical outcomes were also tested against the urbanization gradient with an adjusted logistic regression model. The predicted values were extracted from the model and plotted against the urbanization gradient.

R-package phyloseq<sup>15</sup> was used for handling the microbiota and meta data. Observed richness and Shannon diversity was used as the alpha diversity metric. The community compositions of microbiotas (beta diversity) were visualized with PCoA (vegan R-package) for Bray-Curtis<sup>16</sup>, Weighted UniFrac and Unweighted UniFrac<sup>17</sup> distances. Permutational Multivariate Analysis of variance (PERMANOVA, vegan R-package) with 1000 permutations was used for testing significance. Additionally, PERMANOVA (with Bray-Curtis dissimilarities) was adjusted separately for each living-environment associated lifestyle variable as well as a combined model with all of them. The group-wise similarity was defined with the distance of each sample from group mean (microbiome R-package<sup>18</sup>).

Immune markers were log transformed and centered log ratio scaled in order to remove biological variation due to differing amount of protein per volume of fluid from the dataset as well as to increase the comparability between immune markers. The normalized concentrations were tested against rural and urban groups in adjusted linear models.

Sparse Partial Least Squares (sPLS) models with relative abundances of taxa agglomerated to genus level, log-transformed, and cleaned to include those present at least in 5% of samples were used to predict the urbanization gradient (mixOmics<sup>19</sup> and caret R-packages<sup>20</sup>). Also, pseudocount was added due to log-transformation. The optimum model was selected based on the Root Mean Squared Error (RMSE) statistics, from repeated 10-fold cross-validation, in order to avoid overfitting. The predicted values from each model were extracted to

represent a bacterial score for the urbanization gradient. The odds ratios for outcomes as well as the relative concentrations of immune markers in response to bacterial scores scaled to a standard deviation of one were defined with logistic and linear regression models, respectively.

## ONLINE REPOSITORY REFERENCES

1. Bisgaard H, Vissing NH, Carson CG, Bischoff AL, Følsgaard NV, Kreiner-Møller E, et al. Deep phenotyping of the unselected COPSAC2010 birth cohort study. *Clin Exp Allergy*. 2013;43:1384–94.
2. Bisgaard H, Pipper CB, Bønnelykke K. Endotyping early childhood asthma by quantitative symptom assessment. *J Allergy Clin Immunol*. 2011;127:1155-1164.e2.
3. Cooley D. googleway: Accesses Google Maps APIs to Retrieve Data and Plot Maps. R package version 2.7.1. 2018; Available from: <https://CRAN.R-project.org/package=googleway>
4. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. R package version 2.5-4. 2019; Available from: <https://CRAN.R-project.org/package=vegan>
5. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1. 2018;
6. Ruokolainen L, von Hertzen L, Fyhrquist N, Laatikainen T, Lehtomäki J, Auvinen P, et al. Green areas around homes reduce atopic sensitization in children. *Allergy*. 2015;70:195–202.
7. Mortensen MS, Brejnrod AD, Roggenbuck M, Abu Al-Soud W, Balle C, Krogfelt KA, et al. The developing hypopharyngeal microbiota in early life. *Microbiome*. 2016;4:70.
8. Stokholm J, Blaser MJ, Thorsen J, Rasmussen MA, Waage J, Vinding RK, et al. Maturation of the gut microbiome and risk of asthma in childhood. *Nat Commun*. 2018;9:141.
9. Mortensen MS, Brejnrod AD, Roggenbuck M, Abu Al-Soud W, Balle C, Krogfelt KA, et al. The developing hypopharyngeal microbiota in early life. *Microbiome*. 2016;4:70.
10. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011;17:10–2.
11. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet C, Al-Ghalith GA, et al. QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science [Internet]. PeerJ Inc.; 2018 Oct [cited 2018 Nov 28]. Report No.: e27295v1. Available from: <https://peerj.com/preprints/27295>
12. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods*. 2016;13:581–3.
13. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:D590-596.
14. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*. 2018;6:90.
15. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*. 2013;8:e61217.
16. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*. 1957;27:325–49.

17. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and Qualitative  $\beta$  Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. *Appl Environ Microbiol.* 2007;73:1576–85.
18. Lahti L. microbiome R package. Available from: <http://microbiome.github.io>
19. Rohart F, Gautier B, Singh A, Cao K-AL. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLOS Computational Biology.* 2017;13:e1005752.
20. Thorsen J. Implementation and utility functions for using the mixOmics package in the caret framework: jonathanh/mixOmicsCaret [Internet]. 2017 [cited 2018 Dec 9]. Available from: <https://github.com/jonathanh/mixOmicsCaret>

## Online repository figure legends

**Figure E1. Urban and rural living environments.** The urbanization gradient is the first PCA axis (PC-1) of the land cover data (**a**). The first axis explained 44% of the total variation in land cover data. The loading vectors show the correlation of the different land cover types, e.g. increasing values on the first axis indicate higher proportions of artificial surfaces. Artificial surfaces include land modified by humans such as built areas, roads, rails, ports and factories. Eigenvectors of different land-use types were multiplied with 0.2 in order to fit them to figure. Decreasing values indicate higher proportions of agricultural and forested land covers. Colors specify the rural (green) and urban (golden) categories defined by Partitioning Around Medoids clustering (**b**). These categories are also shown in a density plot, which shows the number of children in rural and urban groups along the living environment gradient.

**Figure E2. The living environment associates strongly with several lifestyle features.** These features included pet ownership during the first year of life, attendance to daycare at age of 1 year, total length of breastfeeding in days, exposure to passive smoke in days, household income, education of both mother and father, number of siblings and the type of home. Stars in the title of each figure indicate the significance level of the associations, defined with Fisher's exact test for categorical variables and ANOVA for continuous variables, such that \* indicates p-values lower than 0.05, \*\* indicates p-values lower than 0.01 and \*\*\* indicate p-values lower than 0.001. For further analyses, highly collinear variables i.e. income and education were simplified to one variable (socio-economic features). Home type was left out from further analyses due to high collinearity with living environment. Daycare types: Public – Public daycare institution, usually several groups of 10-15 children each; Private – Private daycare, usually 2-4 children in a private home. Education: 1 – elementary school, 2 – college or high school, 3 – tradesman certification, 4 – medium academic education, or 5 – completed university degree prior to infant's first birthday.

**Fig E3.** Figures **a** and **b** show that the predicted prevalence (from adjusted logistic regression models) of asthma and atopic traits at six years of age increase with increasing urbanization of living environment during infancy.

**Figure E4. Airway and gut microbiota in urban and rural infants.** Principal Coordinate Analysis on Bray-Curtis distances. Significance is tested with PERMANOVA. Ellipsoids indicate the location of the 75 % of the samples in the group. The airway samples had 6,647 unique amplicon sequence variants (ASVs) with a median richness of 20 (range 8-223) at one week, 26 (8-220) at one month and 33 (8-112) at three months of age. The gut samples had 5,183 unique ASVs with a median richness of 30 (9-272) at one week, 30 (6-171) at one month and 68 (12-175) at one year of age. As suggested by these richness estimates, 32% of all ASVs in airways and 39% in gut existed only in a single sample indicating rather high time- and individual-dependent variation in the microbiota.

**Figure E5. The intra-group dissimilarity in the airway microbial communities is smaller in urban than rural infants.** This is indicated by the average dissimilarity of each sample from the group mean. Significance is tested with Analysis of variance.

**Figure E6. Differences in the relative abundance of the fifteen most abundant *airway* (sub-) genera between rural and urban infants at all time points.** The most abundant genera have been selected independently for each time point, and organized by mean relative abundance. Green indicates rural and gold urban infants. Relative abundances were tested with non-parametric Wilcoxon signed-rank test for significance. q-values indicate False Positive Rate corrected p-values (Benjamini-Hochberg approach was utilized).

**Figure E7. Differences in the relative abundance of the fifteen most abundant *gut* (sub-) genera between rural and urban infants at all time points.** The most abundant genera have been selected independently for each time point, and organized by mean relative abundance. Green indicates rural and gold urban infants. Relative abundances were tested with non-parametric Wilcoxon signed-rank test for significance. q-values indicate False Positive Rate corrected p-values (Benjamini-Hochberg approach was utilized).

**Figure E8. Gut bacterial score versus the relative concentration of airway and systemic immune markers.** The figure shows normalized and scaled relative concentrations of immune markers in relation to airway bacterial score per standard deviation. Higher values indicate association with urban-like bacterial community composition. Panel A shows the topical immune markers at one month of age, and panel B the systemic immune markers at six months of age in response to bacterial scores. Linear models were adjusted for living environment associated lifestyle features. Error bars represent 95% CI. False positive rate corrected p-values are presented in Table E6.

## Online repository table legends

**Table E1. Influence of the buffer radius for the proportion of rural and urban infants.** The living environment was defined at radius of 3km, 1km and 500m for all children. Partitioning around medoids clustering tends to produce clusters with uneven sizes with smaller buffer radius sizes. Thus, 3 km buffer was selected for analyses to both even the number of rural and urban children, as well as to follow the recommendation from previous research exploring the suitable buffer size of living environment (Ruokolainen, L. et al. Allergy 2014). Silhouette widths indicate good clustering for all tested buffers.

**Table E2. Specific Immunoglobulin E levels in children born in rural and urban environments.** “% >0.35” indicates the proportion of children sensitized to specific allergen while “n >0.35” indicates the number of children sensitized to specific allergen. Difference in specific sensitization was tested with Fisher’s exact test.

**Table E3. Living environment gradient associates with the airway and gut microbiota in unadjusted analyses.** Models for the microbiota were done with Permutational Multivariate Analysis of Variance (PERMANOVA) as implemented in the function `adonis` (vegan R package). Bray-Curtis, Weighted UniFrac and Unweighted UniFrac distance metrics were used.

**Table E4. Living environment gradient remains significantly associated with airway microbiota after adjustment for lifestyle variables.** Figure legend S1 explains the selection of living environment related lifestyle variables. Adjusted models for the airway microbiota were done with Permutational Multivariate Analysis of Variance (PERMANOVA) as implemented in the function `adonis` (vegan R package). Bray-Curtis dissimilarity was used for the airway microbiota. Living environment gradient was adjusted separately for each living environment related lifestyle variable and finally to all of them. Living environment gradient was added as the last parameter to the model (PERMANOVA adds variables sequentially).

**Table E5. Living environment gradient has significant association with gut microbiota only in the 1-year samples after adjustment for single lifestyle variables.** Figure legend S1 explains the selection of living environment –related lifestyle variables. Adjusted models for gut microbiota were done with Permutational Multivariate Analysis of Variance (PERMANOVA) as implemented in function adonis (vegan R package). Bray-Curtis dissimilarity was used for gut microbiota. Living environment gradient was adjusted separately for each living environment –related lifestyle variable and finally to them all. Living environment gradient was added as the last parameter to the model (PERMANOVA adds variables sequentially)

**Table E6. Immune marker concentrations in response to three explanatory variables.** The immune marker concentrations in airways at one month of age as well as in blood at ix months of age were tested against categorical living environment variable (Rural/urban), continuous airway bacterial score as well as continuous gut bacterial score. q-values indicate False Positive Rate corrected p-values (Benjamini-Hochberg approach was utilized).