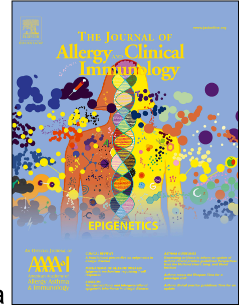
Journal Pre-proof



Indoor Bacterial Microbiota and the Development of Asthma by 10.5 years of age

Anne M. Karvonen, PhD, Pirkka V. Kirjavainen, PhD, Martin Täubel, PhD, Balamuralikrishna Jayaprakash, MSc, Rachel I. Adams, PhD, Joanne E. Sordillo, ScD, Diane R. Gold, MD, MPH, Anne Hyvärinen, PhD, Sami Remes, MD, MPH, Erika von Mutius, MD, Juha Pekkanen, MD

PII: S0091-6749(19)31033-4

DOI: https://doi.org/10.1016/j.jaci.2019.07.035

Reference: YMAI 14123

To appear in: *Journal of Allergy and Clinical Immunology*

Received Date: 7 September 2017

Revised Date: 6 June 2019

Accepted Date: 11 July 2019

Please cite this article as: Karvonen AM, Kirjavainen PV, Täubel M, Jayaprakash B, Adams RI, Sordillo JE, Gold DR, Hyvärinen A, Remes S, von Mutius E, Pekkanen J, Indoor Bacterial Microbiota and the Development of Asthma by 10.5 years of age, *Journal of Allergy and Clinical Immunology* (2019), doi: https://doi.org/10.1016/j.jaci.2019.07.035.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier Inc. on behalf of the American Academy of Allergy, Asthma & Immunology.

Journal Pre-proof

Indoor Bacterial Microbiota and the Development of Asthma by 10.5 years of age

Anne M. Karvonen, PhD,[a,b] Pirkka V. Kirjavainen, PhD,[a,c] Martin Täubel, PhD,[a] Balamuralikrishna Jayaprakash, MSc,[a] Rachel I. Adams, PhD,[d,e] Joanne E. Sordillo, ScD,[b,f] Diane R. Gold, MD, MPH,[b] Anne Hyvärinen, PhD,[a] Sami Remes, MD,MPH,[g] Erika von Mutius, MD,[h,i,j] and Juha Pekkanen, MD[a,k]

[a] Department of Health Security, Finnish Institute for Health and Welfare, Kuopio, Finland; P.O. Box 95, FIN-70701 Kuopio, Finland
[b] Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 181 Longwood Ave, Boston, MA 02115, USA
[c] Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland; Yliopistonranta 1, FIN-70210 Kuopio, Finland
[d] Plant & Microbial Biology, University of California, Berkeley, CA, USA; Koshland Hall, 111 Koshland Hall, Berkeley, CA 94720, USA
[e] California Department of Public Health, Environmental Health Laboratory Branch, Richmond, CA, USA; 850 Marina Bay Parkway, MS G365, Richmond, CA 94804, USA
[f] Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care, Boston, MA, USA; Landmark Center, 401 Park Drive, Suite 401 East Boston, MA 02215, USA
[g] Department of Pediatrics, Kuopio University Hospital, Kuopio, Finland; P.O.Box 100, FIN-70029 Kuopio, Finland
[h] Dr. von Hauner Children's Hospital, Ludwig Maximilians University, Munich, Germany; Lindwurmstraße 4, D-80337 München, Germany
[i] Member of the German Center for Lung Research, Germany
[j] Institute for Asthma and Allergy Prevention (IAP), Helmholtz Zentrum München, Munich, Germany; Ingolstaedter Landstrasse 1, D-85764 Neuherberg, Germany
[k] Department of Public Health, University of Helsinki, Helsinki, Finland; P.O.Box 20, 00014 University of Helsinki, Finland

e-mail addresses: anne.karvonen@thl.fi, pirkka.kirjavainen@thl.fi, martin.taubel@thl.fi, balamuralikrishna.jayaprakash@thl.fi, adamsri@berkeley.edu, rejoa@channing.harvard.edu, redrg@channing.harvard.edu, anne.hyvarinen@thl.fi, sami.remes@kuh.fi, Erika.Von.Mutius@med.uni-muenchen.de, and juha.pekkanen@helsinki.fi

Corresponding author:
Anne M. Karvonen
Department of Health Security
Finnish Institute for Health and Welfare
P.O. Box 95, FIN-70701 Kuopio, Finland
anne.karvonen@thl.fi
Tel: +358 (0) 29 524 6325, fax +358 (0) 29 524 6498

Journal Pre-proof

Funding

This study was supported by research grants from the Academy of Finland (grants 139021;287675; 296814; 296817; 308254); the Juho Vainio Foundation; the Yrjö Jahnsson Foundation; the Foundation for Pediatric Research; EVO/VTR-funding; Päivikki and Sakari Sohlberg Foundation; The Finnish Cultural Foundation; European Union QLK4-CT-2001-00250; and by the Finnish Institute for Health and Welfare, Finland.

Conflict of Interest: Dr Täubel, Dr. Jayaprakash, Dr. Adams, Dr. Sordillo, Prof. Gold, Dr. Hyvärinen and Dr. Remes have nothing to disclose. Dr. Karvonen and Prof. Pekkanen report grants from the Academy of Finland, grants from the Juho Vainio Foundation, grants from the Yrjö Jahnsson Foundation, grants from the Foundation for Pediatric Research, grants from EVO/VTR-funding, grants from Päivikki and Sakari Sohlberg Foundation, grants from The Finnish Cultural Foundation, grants from European Union, during the conduct of the study; Dr. Kirjavainen reports grants from Päivikki and Sakari Sohlberg Foundation, grants from Yrjö Jahnsson Foundation, grants from Nutricia Research Foundation, grants from Juho Vainio Foundation, grants from Academy of Finland, during the conduct of the study; Dr. von Mutius reports grants from European Commission, during the conduct of the study; personal fees from PharmaVentures, personal fees from OM Pharma, personal fees from Decision Resources, personal fees from Novartis Pharma SAS, personal fees from The Chinese University of Hongkong, personal fees from University of Copenhagen, personal fees from HAL Allergie GmbH, personal fees from Ökosoziales Forum Oberösterreich, personal fees from Mundipharma, personal fees from American Thoracic Society, personal fees from AbbVie Deutschland GmbH & Co. KG, personal fees from University of Tampere, personal fees from European Commission, personal fees from Massachassuetts Medical Society, personal fees from American Academy of Allergy, Asthma and Immunology, outside the submitted work. Furthermore, Erika von Mutius is listed as inventor on the following patents: Publication number EP 1411977: Composition containing bacterial antigens used for the prophylaxis and the treatment of allergic diseases; Publication number EP1637147: Stable dust extract for allergy protection; Publication number EP 1964570: Pharmaceutical compound to protect against allergies and inflammatory diseases; Application number LU101064, Barn dust extract for the prevention and treatment of diseases. (Pending); she is listed as inventor and has received royalties on the following patent: Publication number

Journal Pre-proof

EP2361632: Specific environmental bacteria for the protection from and/or the treatment of allergic, chronic inflammatory and/or autoimmune disorders.

Journal Pre-proof

Journal Pre-proof

1 **ABSTRACT**

2 **Background:** Early-life indoor bacterial exposure is associated with the risk of asthma but
3 the roles of specific bacterial genera are poorly understood.

4 **Objective:** To determine whether individual bacterial genera in indoor microbiota predict the
5 development of asthma.

6 **Methods:** Dust samples from living rooms were collected at 2 months of age. The dust
7 microbiota was characterized by Illumina MiSeq sequencing amplicons of bacterial 16S
8 ribosomal RNA gene. Children (N=373) were followed up for ever asthma until the age of
9 10.5 years.

10 **Results:** Richness was inversely associated with asthma after adjustments ($p=0.03$). The
11 phylogenetic microbiota composition in asthmatics' homes was characteristically different
12 from non-asthmatics' homes (weighted UniFrac, adjusted association, PERMANOVA-S,
13 $p=0.02$). The first two axis scores of principal coordinate analysis of the weighted UniFrac-
14 distance matrix were inversely associated with asthma. Out of 658 genera detected in the dust
15 samples, the relative abundances of 41genera correlated ($r>|0.4|$) with one of these axes.
16 *Lactococcus* genus was a risk factor for asthma (aOR 1.36, 95%CI 1.13-1.63 per IQR
17 change). The abundance of twelve bacterial genera (mostly from *Actinomycetales* order) was
18 associated with lower asthma risk ($p<0.10$), though not independently of each other. The sum
19 relative abundance of these12 intercorrelated genera was significantly protective and
20 explained majority of the association of richness with less asthma.

21 **Conclusion:** Our data confirms that phylogenetic differences in infant home microbiota are
22 associated with subsequent asthma risk and suggest that communities of selected bacteria are
23 more strongly linked to asthma protection than individual bacterial taxon or mere richness.

24

25 **Abstract word count**: 247

Journal Pre-proof

Journal Pre-proof

26 **Key messages**

27 • Childhood asthma risk is affected by bacterial composition of the early-life home
28 indoor microbiota.

29 • Communities of bacteria, rather than an individual taxon or overall bacterial diversity,
30 are most strongly linked to asthma protection.

31

32 **Capsule summary** Early-life home indoor bacterial exposures are associated with the risk of
33 asthma development. This study suggests that communities of protective environmental
34 bacteria rather than individual taxon or overall bacterial diversity may offer asthma
35 protection.

36

37 **Keywords** (3-5): Asthma development, children, diversity, environment, Lactococcus

38 **Abbreviations:**

39 BLAST Basic Local Alignment Search Tool

40 CE/m^2 Cell Equivalents/square meter

41 IQR Interquartile range

42 MaAsLin Multivariate Association with Linear model

43 NCBI National Center for Biotechnology Information

44 OR Odds ratio

45 aOR adjusted Odds ratio

46 OTU Operational Taxonomic Unit

47 PASTURE Protection against Allergy Study in Rural Environments

48 PCoA Principal Coordinate Analysis

49

50 **This article has an Online Repository at www.jacionline.org**

Journal Pre-proof

Journal Pre-proof

51 **Running head**: Environmental bacterial exposure and asthma

52 **Descriptor number that best classifies the subject of the manuscript**:

53 **Word count**: 3548

54

55 **Author Contributions:** Conception and design: AK, PK, MT, AH, EvM, JP; Analysis and
56 interpretation: AK, PK, MT, BJ, RA, JS, DG, AH, SR, EvM, JP; Drafting the manuscript for
57 important intellectual content**:** AK, PK, MT, BJ, RA, JS, DG, AH, SR, EvM, JP; Final
58 approval of the version to be published**:** AK, PK, MT, BJ, RA, JS, DG, AH, SR, EvM, JP;
59 Agreement to be accountable for all aspects of the work: AK, PK, MT, BJ, RA, JS, DG, AH,
60 SR, EvM, JP.

Journal Pre-proof

Journal Pre-proof

61 **Introduction**

62 Microbial exposures early in life may have a dual role in asthma development. Early -life
63 viral infections predispose to asthma and also bacterial infections and airway colonization by
64 potential respiratory bacterial pathogens may have similar influence.[1] On the other hand it is
65 well recognized that microbial exposure in *utero* and early life appear to be essential in
66 instructing adaptive and regulated immune system responses to other environmental elements
67 such as allergens, particles and viruses.[2]
68
69 Accordingly, intimate exposure to environments rich in microbes, such as associated with
70 traditional farming practices, may decrease the risk of asthma and other allergic diseases.[3] The
71 earlier epidemiologic studies on home microbial exposure and asthma were based on
72 characterization of exposure through measures of general microbial markers[4-7] such as
73 endotoxin in dust samples (reviewed in ref.[8]). We have previously shown in this cohort that
74 the quantity of exposure to bacterial and fungal cell wall components in early life has a bell-
75 shaped association with asthma at the age of 6 years.[9] Studies with DNA based methods have
76 indicated that the asthma protective characteristics may include diversity[10-13] or more
77 specifically diversity within certain taxon and lack of predisposing microbes.[14-17] However, it
78 remains unclear whether there are specific, individual taxa in indoor microbiome that are
79 independently associated with reduced asthma risk.
80
81 The overall objective of this study was to identify individual bacterial genera from the early-
82 life indoor environment that are associated with the development of asthma until the age of
83 10.5 years. We also tested whether the protective association between high bacterial diversity
84 and asthma is independent of the contributing microbes as has been hypothesized.
85

Journal Pre-proof

Journal Pre-proof

86 **Methods**

87 The study population consisted of children born in Middle and Eastern Finland: the first half
88 of the study population (N=214) belongs to a European birth cohort (PASTURE)[18] among
89 farmers and non-farmers, while the second half of the cohort consists of unselected children
90 (N=228).[19] Pregnant women who gave birth between September 2002 and May 2005 were
91 recruited. Selection procedure has been described earlier, and the study protocol was
92 approved by a local ethics committee in Finland.[19] Written informed consent was obtained
93 from the parents.

94

95 *Follow-up*

96 The children were followed up with questionnaires[19] as described in the Methods section in
97 the Supplemental Material. 'Ever asthma' was defined as first parent-reported doctor-
98 diagnosed asthma and/or second diagnoses of asthmatic (or obstructive) bronchitis. 'Current
99 asthma' was defined as 'ever asthma' with usage of asthma medication and/or reported
100 wheezing symptom in the past 12 months at 10.5 year follow-up. Wheezing phenotypes were
101 created using latent class analyses (see in the Supplemental Material).

102

103 *House dust samples*

104 House dust samples were sequenced from 394 living room floor dust samples. The protocols
105 for dust collection of 2 months of age and analyzes of general microbial markers have been
106 described previously.[7,9] The protocol for sequencing (V4 region of the 16S rRNA),[20]data
107 processing, and measuring the relative abundances and qPCRs (assay targeting the 16S rRNA
108 gene) are described in the Methods section of Supplemental Material. Bacterial richness (a
109 measure of the number of different Operational Taxonomic Units (OTU) in each sample) and
110 Shannon diversity (abundance and evenness of the taxa in each sample) indices were

Journal Pre-proof

Journal Pre-proof

111 calculated within samples. The 'load' of the bacterial genus (i.e. expressed as cell equivalents
112 per square meter (CE/m^2)) was calculated by multiplying relative abundance with total
113 bacteria(CE/mg) in that sample, as measured via qPCR and with the amounts of dust, and
114 dividing by sampling area(m^2).
115
116 *Statistical analyses*
117 Statistical analyses are described in more detail in Supplemental Material. Generalized
118 UniFrac based Principal Coordinate Analysis (PCoA) was performed using QIIME and the
119 first six axes scores (Eigen values >1) were used in the analyses. The adjusted association of
120 bacterial composition and ever asthma was studied using PERMANOVA-S.[21]
121
122 T- or Kruskal-Wallis tests were used for comparing relative abundances of taxa in homes of
123 asthmatics (ever asthma) and non-asthmatics children. For the multivariate models, the
124 variables were ln-transformed (natural logarithm +1, except diversity indices) and divided by
125 interquartile range (IQR). Discrete-time hazard models, generalized estimating equations, and
126 multinomial logistic regression were used for analyzing asthma, respiratory symptoms and
127 wheezing phenotypes, respectively. The results are presented as adjusted odds ratios (aORs)
128 and their 95% confidence intervals (95%CI).
129
130 In order to increase taxonomic resolution, oligotyping analysis was performed for
131 *Lactococcus* genus using entropy positions.[22] Multivariate association with linear models
132 (MaAsLin)[23] was run using all the most abundant taxa (mean relative abundance >0.1%) from
133 phylum to genus level.
134

Journal Pre-proof

Journal Pre-proof

135 All models were adjusted for follow-up time, study cohort, living on a farm and well-known
136 risk factors for asthma (maternal history of allergic diseases, gender, number of older siblings,
137 and smoking during pregnancy). Two selected models were carefully tested for 25 additional
138 confounding factors,[7] but none of these potential confounders changed the estimates of
139 exposure by >10%, and thus were not included in the analyses. At the age of 3 years, majority
140 (80%) of the children still lived in the same house. The data were analyzed using SAS 9.3 for
141 Windows (SAS Institute, Inc., Cary, NC).

Journal Pre-proof

Journal Pre-proof

142 **Results**

143 Of the 442 children, 394 (89.1%) had data on the bacterial microbiota in dust samples and 373
144 of those (94.6%) had sufficient data to assess asthma until the age of 10.5 years and
145 information on covariates. By the age of 10.5 years, 69 children (18.5%) had developed ever
146 asthma and 29 (7.8%) current asthma at 10.5years.

147

148 *Bacterial diversity in asthmatics' and in non-asthmatics' homes*

149 The bacterial richness and Shannon diversity index were lower in the homes of children with
150 ever asthma than in homes of non-asthmatics (Figure 1). When the models were adjusted for
151 confounding factors, bacterial richness was inversely associated with ever asthma, and
152 Shannon index qualified as a trend (Table I, $p=0.03$ and 0.12, respectively).

153

154 *UniFrac based weighted PCoA axis scores and asthma*

155 The overall microbiota composition between asthmatics and non-asthmatics was significantly
156 different as indicated by weighted UniFrac beta-diversity analysis (adjusted association,
157 PERMANOVA-S, $p=0.02$). The first two principal coordinate analyses axes scores (PCoA1
158 and PCoA2) explained 36% of the variance in the weighted UniFrac dissimilarity distance
159 matrix (Figure 2). The PCoA1 and PCoA2 axes scores were inversely associated with ever
160 asthma (Table I). The first axis score appeared to reflect the ratio of *Firmicutes* and
161 *Proteobacteria* in the samples with negative correlation with *Firmicutes* and positive
162 correlation with *Proteobacteria* abundance at the phylum level. The second axis score
163 appeared to reflect diversity and *Actinobacteria* abundance seen as positive correlation with
164 both (Figure 3). The fourth most common phylum, *Bacteroidetes*, had a weak positive
165 correlation with both axis scores (Figure 3). There were no significant associations between

Journal Pre-proof

Journal Pre-proof

166 the other four PCoA axis scores (Eigen value >1) and ever asthma (see Table E1 in the Online
167 Repository).
168
169 *Phylum and genus levels in asthmatics' and in non-asthmatics' homes*
170 At the phylum level, the relative abundance of *Firmicutes* was higher and *Actinobacteria* was
171 lower in the homes of asthmatic children than in the homes of non-asthmatics (see Figure E1
172 in the Online Repository). At the genus level, the relative abundances of *Lactococcus*
173 (*Firmicutes*) and *Streptococcus* (*Firmicutes*) were higher, but relative abundance of
174 *Sphingomonas* (*Proteobacteria*) was lower in the homes of asthmatics than in the homes of
175 non-asthmatics (Figure 4). Consistent with results on richness, the combined relative
176 abundance of the rest of the genera (mean relative abundance < 1%) was lower in the homes
177 of asthmatic children than in the homes of non-asthmatics (43.0% *vs.* 47.8%, respectively,
178 $p<0.001$).
179
180 *Bacterial genera and asthma*
181 Out of detected 658 bacterial genera, 139 bacterial genera had the mean relative abundance
182 greater than 0.1%, and they were further studied. Forty one of the 139 genera correlated (r
183 $|>0.4|$) with either or both PCoA1 and PCoA2 axes scores (see Table E2 in the Online
184 Repository). After adjustments, the relative abundances of the 12 genera were inversely
185 ($p<0.1$) and *Lactococcus* positively ($p=0.001$) associated with the development of ever
186 asthma (Figure 5). *Lactococcus* (median relative abundance 3.9%) was the only genus that
187 was associated with higher risk of ever asthma after correction for multiple testing
188 (Bonferroni). High positive correlation coefficients (being mostly $r=0.5$ - 0.8) were found
189 within the relative abundances of these 12 genera, except for *Brevibacterium* and other genus
190 within *Dermabacteraceae* family, which had clearly lower correlation coefficients (see Figure

Journal Pre-proof

191 E2 and results in the Online Repository). When the negative associations of the 12 genera and
192 the positive association of *Lactococcus* were mutually adjusted in the model of ever asthma,
193 only the positive association of *Lactococcus* remained significant (see Figure E3 in the Online
194 Repository).
195
196 The relative abundances of the 12 protective genera were thus added up into a new variable
197 due to their high intercorrelation. The sum abundance of the 12 protective genera (median
198 relative abundance 5.2%) was dose-dependently associated with ever asthma [compared to the
199 lowest tertile, aOR 0.48 (95%CI 0.26, 0.85) for middle tertile $p=0.01$, and aOR 0.31 (95%CI
200 0.15, 0.63) for highest tertile $p=0.001$]. The sum abundance of the12 protective genera and
201 the *Lactococcus* genus were independent predictors for ever asthma (data not shown). The
202 associations with current asthma were largely similar (data not shown).
203
204 The predisposing association between the relative abundance of *Lactococcus*, and the inverse
205 association of the sum abundance of the 12 protective genera with ever asthma were
206 independent of bacterial richness, Shannon index, amounts of dust, endotoxin, $LPS_{10:0\text{-}16:0}$ and
207 muramic acids (see Table E3 in the Online Repository). The sum abundance of the 12
208 protective genera explained 61% of the association between richness and ever asthma (see
209 Table E4 in the Online Repository). Environmental and behavioral determinants associated
210 with reduced signals of asthma predisposing *Lactococcus* abundance and increase in asthma
211 protection associated microbes included animal and farm contacts, timber structures, age of
212 the house and natural ventilation (see results and Table E5 in the Online repository).
213
214 *Bacterial exposure and wheezing phenotypes*

Journal Pre-proof

Journal Pre-proof

215 In the analyses of wheezing phenotypes (based on latent class analyses) during the first 6
216 years of life, no associations were found between relative abundance of Lactococcus, the sum
217 abundance of the 12 protective genera, diversity indices and transient wheeze, which is
218 mostly related to infections in early age (see Table E6 in the Online Repository). Even the
219 number of cases in late onset and persistent wheeze groups were small, the associations were
220 towards the same directions than with ever asthma, but they were clearly weaker. However,
221 there were tendency of inverse associations between the sum of 12 protective genera and late
222 onset and persistent wheeze ($p<0.20$). When exploring associations with respiratory
223 symptoms, similar, but mostly non-significant, as with asthma ever associations were found,
224 except for *Lactococcus,* which had weaker associations with wheezing (see Table E7 in the
225 Online Repository).

226

227 *Oligotypes of Lactococcus and asthma*

228 In order to possibly increase taxonomic resolution for the finding of *Lactococcus* genus with
229 ever asthma, oligotyping analysis was performed with this taxon, and 10 oligotypes were
230 created (see Table E8 in the Online Repository). Most of the sequences belonged to the
231 GGCCAAGGA oligotype (95% of all sequences), which had the highest mean relative
232 abundance (7.2%) and correlated with the relative abundance of the *Lactococcus* genus and
233 OTU number 1100972 ($r=0.99$). For this oligotype, two best BLAST hits from NCBI
234 database were uncultured bacterial clone 1714 and *Lactococcus lactis* subsp.(lactis gene for
235 16S rRNA) with 100% similarity (identity and coverage). Relative abundances of each of the
236 10 oligotypes were positively associated with the development of ever asthma after
237 adjustments ($p<0.15$). The correlations between 10 oligotypes ranged from *0.28* to *0.70*
238 (mostly >0.45). When the relative abundances of 10 oligotypes were simultaneously adjusted,
239 none of them were significantly associated with ever asthma (data not shown).

Journal Pre-proof

Journal Pre-proof

240

241 *Loads of bacterial genera, total bacterial qPCR and asthma*

242 The associations with ever asthma were slightly weaker when the loads of the bacterial genera
243 (i.e. expressed as cell equivalents per square meter (CE/m^2)) were used instead of relative
244 abundances, except for the *Lactococcus* genus, for which the estimate was stronger (see Table
245 E9 in the Online Repository). The correlations between the relative abundances of sequences
246 of the 13 bacterial genera and their loads were fairly high ($r= 0.57\text{-}0.79$). Total bacterial
247 qPCR was not associated with ever asthma or current asthma (data not shown).

248

249 *MaAsLin*

250 MaAsLin identified the relative abundance of 9 taxa that were significantly associated with
251 ever asthma after multiple testing was taken into account (Q-value <0.05). The strongest
252 association was found with *Lactococcus* (see Table E10 in the Online Repository). For the
253 rest taxa, other genus within *Microbacteriaceae* family, which was one of 12 protective
254 genera, was also identified.

255

Journal Pre-proof

Journal Pre-proof

256 **Discussion**

257 The present study suggests that phylogenetic differences in the early-home indoor microbiota
258 composition precede asthma development and this association is not explained by bacterial
259 richness alone. Out of 658 genera detected in the dust samples, only the relative abundance of
260 *Lactococcus* genus was determined as an independent risk factor for asthma. Twelve bacterial
261 genera (mostly from *Actinomycetales* order) were identified as protective. The sum of the
262 relative abundance of these 12 protective genera was significantly protective and explained
263 majority of the association of richness with less asthma.

264

265 We found a similar inverse association between bacterial richness and asthma, as have been
266 reported in two recent cross-sectional studies from rural areas.[10,12] Another nested case-
267 control study with high allergy risk children from urban environment, by Lynch and co-
268 workers, found similar association between bacterial richness and atopy and recurrent wheeze
269 together with atopy, but not wheeze by itself at the age of 3 years.[14] In contrast, a study
270 among asthmatics showed that high levels of bacterial richness in homes was associated with
271 more severe asthma symptoms compared to homes with low bacterial richness in house
272 dust.[24] This might be explained by the notion that bacterial richness may have different
273 importance on asthma severity than on the development of asthma, something that has been
274 found earlier with high endotoxin exposure.[25] Thus, our findings support earlier observations
275 that a diverse environmental microbial exposure at early age via ingestion, inhalation and/or
276 skin may be essential for stimulating immune development to respond appropriately to other
277 environmental elements.[2]

278

279 The phylogenetic composition of the microbiota was significantly different in house dust of
280 asthmatics and non-asthmatics, as found with PERMANOVA-S analysis method that uses

Journal Pre-proof

Journal Pre-proof

281 UniFrac distance, a measure of similarity and dissimilarity of the bacterial composition
282 between samples. Of the 12 protective genera that were identified, seven were from the
283 *Actinomycetales* order, which are found in outdoor environmental sources (e.g. soil, fresh
284 water and compost).The 12 genera were intercorrelated and thus, it was not surprising that the
285 individual genera were not associated with asthma protection independently from each other
286 but their sum abundance was. Whether the 12 protective genera had a common source or
287 distinct functional influence on asthma development, remains unclear. Interestingly, the
288 association between bacterial richness and asthma was largely explained by the sum
289 abundance of the 12 protective genera, but not with the low relative abundance of
290 *Lactococcus*. This suggests that particular compositions of bacterial exposure, which source is
291 outdoors, better predict the development of asthma than overall bacterial richness. However,
292 the taxa that were identified and combined in the present study should be confirmed by other
293 studies in different environments and in different geographical areas, and their potential
294 protective functions should be explored.
295
296 This study revealed a genus of gram-positive bacteria, *Lactococcus* (belonging to the
297 *Firmicutes* phylum and *Streptococcaceae* family), which increased the risk of asthma
298 independently of microbial diversity. *Lactococcus* is the most prevalent genus in raw and
299 pasteurized cow's milk,[26] it is used in manufacturing of fermented dairy products and it is also
300 found in soil. In the oligotyping analyses, the vast majority of the sequences of *Lactococcus*
301 genus were allocated to one specific oligotype (GGCCAAGGA), which had the strongest
302 effect on asthma development and which, based on the BLAST analyses, might refer to
303 *Lactococcus lactis*. While there is a small but growing literature on early life environmental
304 microbial exposures and development of wheeze and asthma in children, no previous study
305 has shown association with *Lactococcus* genus. A previous study[27] found in a murine and

Journal Pre-proof

Journal Pre-proof

306 experimental model that exposure to *Lactococcus lactis* G121strain along with another
307 bacterial strain, *Acinetobacter lwoffii* F78, which were both isolated from farm stables,
308 prevented experimental allergic asthma in mice. The gram-positive *L. lactis* G121 especially
309 activated cells through NOD2 and TLR2. In our study, we observed that the *Lactococcus*
310 genus and its oligotypes were significant risk factors for asthma. Due to similar associations
311 between Lactococcus genus and ever asthma among children from farms and non-farms, it is
312 unlikely that farm milk is the source of Lactococcus genus in the present study. However, our
313 sequencing analyses method was not designed to enter into the strains/species level with
314 confidence, which is a general weakness of the amplicon sequencing method. Whether the
315 *Lactococcus* genus is a true risk factor for asthma, or a proxy of other predisposing factor, e.g.
316 particular life-style or nutrition, remains to be determined in experimental and other
317 epidemiological studies including quantitative and specific detection (e.g. using qPCR) of
318 *Lactococcus*.
319
320 We have previously shown in this cohort that farm-like bacterial relative abundance patterns
321 in indoor microbiota are associated with asthma protection by 6 years of age.[15] There were
322 little overlap between the specific genera identified in the current study to be associated with
323 asthma after adjustment for farming and the best predictors of the farm-like indoor microbiota
324 composition identified in our previous study. However, there were phylogenetic similarities
325 as both imply importance of high abundance of members within the *Actinobacteria* phylum.
326 In contrast, there was little or no overlap between the 13 genera identified in the present study
327 and taxa that have been associated with lower asthma risk in other previous studies.[10,12,14,16,17]
328 In these studies, minimal[10,12] to no[14,16,17] adjustments have been made for potential
329 confounders and comparability to our results is also influenced by other differences including
330 those in sampling material, microbiological determinations, outcomes and study designs (e.g.

Journal Pre-proof

Journal Pre-proof

331 prospective vs. cross-sectional). A commonality in the present study and some of the previous
332 studies has been that rather than diversity as such, it is certain compositional aspects within
333 bacterial diversity that explain associations between indoor bacterial exposure and asthma.
334 Further studies aiming at functional profiling e.g. by metagenomics, metabolomics or
335 experimental studies are needed to characterize the potential asthma protective properties the
336 bacterial taxa identified in here may have.
337
338 Mechanisms behind the association between environmental microbial exposures in early life
339 and asthma protection are not well understood. Evidence from epidemiological and
340 experimental studies show that specific microbial exposures, such as those encountered in
341 farming environment or homes with dogs, triggers receptors of the innate immunity,[28] may
342 increase epithelial barrier function in airways, presence of immunosuppressive cells, suppress
343 responsiveness towards microbial immunogens and reduce allergen-induced airway
344 inflammation.[15,16,29,30] Exposure to rich and diverse microbiota may have a positive effect on
345 airway colonization, which may in turn defend against viral infections and thus, contribute to
346 the prevention of asthma.[31,32] In addition, there is evidence from murine models that exposure
347 to microbes in house dust modulates intestinal microbiota, and may, at least partially, mediate
348 the effect on immune responses in airways.[30,33] Whether this would also apply to humans
349 remains unknown. There is evidence that environmental factors such as dogs can influence
350 human gut microbiota composition,[30,34] but overall this influence is thought to be limited.[1]
351
352 In the present study, few environmental and behavioral determinants such as increased animal
353 contacts, natural as opposed to mechanical ventilation and timber structures were associated
354 with reduced signals of asthma predisposing *Lactococcus* abundance and increase in asthma
355 protection associated microbes. These and future findings from more focused studies could

Journal Pre-proof

Journal Pre-proof

356 direct public health initiatives for asthma prevention. Such initiatives may be efficient ways to
357 reduce the allergy and asthma burden as indicated e.g. by the Finnish Allergy Program that
358 provided practical recommendations also for behavior modification.[35]
359
360 The main strengths of the present study are the prospective birth cohort design with high
361 participation rate and extensive set of microbial exposure measurements, including high
362 resolution next generation sequencing data, DNA based targeted qPCR and general microbial
363 markers. Dust samples were collected from living room floors at early childhood, which has
364 shown to be an important time window for intensive maturation of the adaptive immunity.[3]
365 Long-term active air sampling, which is the best way to assess exposure, is logistically and
366 technically challenging in large cohorts, and, thus, surrogates of airborne microbial exposure
367 are practically exclusively used.[36] Floor dust better represents the overall environmental
368 exposures carried from outdoors to indoors than, for example, bed dust that likely reflects also
369 human associated microbiota.[37] However, dust from floors/rugs will only be partially
370 resuspended into the air with at a size that is inhalable and thus, only partially contribute to
371 inhalation exposure. We have recently shown that microbiota of floor dust are not fully
372 consistent with the microbiota of infant breathing zone air, but neither are the microbiota of
373 bulk air in a room fully representative of the particular infant breathing exposure upon near
374 floor activities.[38] As noted earlier, oral ingestion exposure or exposure through the skin during
375 the first years of life may be also relevant.[2] One weakness of our study is that the taxonomic
376 resolution of the sequencing approach did not, in general, allow species levels identification.
377 To overcome this restriction in taxonomic identification, future studies will have to
378 implement metagenomics (shot-gun sequencing) approaches or more targeted approaches –
379 such as qPCR or chip-based hybridization techniques – once knowledge on specific targets
380 has accumulated.

Journal Pre-proof

Journal Pre-proof

381

382 In conclusion, our data confirms that phylogenetic differences in home microbiota influence
383 asthma risk and suggest that communities of selected bacteria are more strongly linked to
384 asthma protection than individual bacterial taxon or richness.

385

386 **Acknowledgement**

387 We would like to thank the families for their participation in the study and study nurses Raija
388 Juntunen, Riikka Juola, Anneli Paloranta and Seija Antikainen for field work, Jutta Kauttio
389 for DNA extraction, Gert Doekes and Ulrike Gehring for analyses of general microbial
390 markers, Asko Vepsäläinen, Pekka Tiittanen, and Timo Kauppila for data management, Pauli
391 Tuoresmäki for creating images, Martin Depner for creating wheezing phenotypes, and John
392 Ziniti, and Vincent Carey for helping with advanced statistical analyses.

Journal Pre-proof

Journal Pre-proof

393 **References**

394 1. Kirjavainen PV, Hyytiäinen H, Täubel M. The Lung Microbiome (ERS Monograph).In:
395 Cox MJ, Ege MJ, von Mutius,E, editors. The environmental microbiota and asthma.
396 Sheffield: European Respiratory Society. 2019. p.216-39.

397 2. von Mutius E. The microbial environment and its influence on asthma prevention in early
398 life. J Allergy Clin Immunol. 2016;137(3):680-9. PubMed PMID: 26806048.

399 3. von Mutius E, Vercelli D. Farm living: Effects on childhood asthma and allergy. Nat Rev
400 Immunol. 2010;10:861-8. PubMed PMID: 21060319.

401 4. Braun-Fahrländer C, Riedler J, Herz U, Eder W, Waser M, Grize L, et al. Environmental
402 exposure to endotoxin and its relation to asthma in school-age children. N Engl J Med.
403 2002;347(12):869-77. PubMed PMID: 12239255.

404 5. Tischer C, Casas L, Wouters IM, Doekes G, Garcia-Esteban R, Gehring U, et al. Early
405 exposure to bio-contaminants and asthma up to 10 years of age: Results of the HITEA study.
406 Eur Respir J. 2015;45(2):328-37. PubMed PMID: 25186271.

407 6. Sordillo JE, Hoffman EB, Celedon JC, Litonjua AA, Milton DK, Gold DR. Multiple
408 microbial exposures in the home may protect against asthma or allergy in childhood. Clin Exp
409 Allergy. 2010;40(6):902-10. PubMed PMID: 20412140; PubMed Central PMCID:
410 PMC3730840.

411 7. Karvonen AM, Hyvärinen A, Gehring U, Korppi M, Doekes G, Riedler J, et al. Exposure to
412 microbial agents in house dust and wheezing, atopic dermatitis and atopic sensitization in
413 early childhood: A birth cohort study in rural areas. Clin Exp Allergy. 2012;42(8):1246-56.
414 PubMed PMID: 22805472.

Journal Pre-proof

Journal Pre-proof

415 8. Doreswamy V, Peden DB. Modulation of asthma by endotoxin. Clin Exp Allergy.
416 2011;41(1):9-19. PubMed PMID: 20977505.

417 9. Karvonen AM, Hyvärinen A, Rintala H, Korppi M, Täubel M, Doekes G, et al. Quantity
418 and diversity of environmental microbial exposure and development of asthma: A birth cohort
419 study. Allergy. 2014;69(8):1092-101. PubMed PMID: 24931137; PubMed Central PMCID:
420 PMC4143956.

421 10. Ege MJ, Mayer M, Normand AC, Genuneit J, Cookson WO, Braun-Fahrlander C, et al.
422 Exposure to environmental microorganisms and childhood asthma. N Engl J Med.
423 2011;364(8):701-9. PubMed PMID: 21345099.

424 11. Dannemiller KC, Mendell MJ, Macher JM, Kumagai K, Bradman A, Holland N, et al.
425 Next-generation DNA sequencing reveals that low fungal diversity in house dust is associated
426 with childhood asthma development. Indoor Air. 2014;24(3):236-47. PubMed PMID:
427 24883433; PubMed Central PMCID: PMC4048861.

428 12. Birzele LT, Depner M, Ege MJ, Engel M, Kublik S, Bernau C, et al. Environmental and
429 mucosal microbiota and their role in childhood asthma. Allergy. 2017;72(1):109-19. PubMed
430 PMID: 27503830.

431 13. Tischer C, Weikl F, Probst AJ, Standl M, Heinrich J, Pritsch K. Urban dust microbiome:
432 Impact on later atopy and wheezing. Environ Health Perspect. 2016;124:1919-23. PubMed
433 PMID:27232328; PubMed Central PMCID: PMC5132631.

434 14. Lynch SV, Wood RA, Boushey H, Bacharier LB, Bloomberg GR, Kattan M, et al. Effects
435 of early-life exposure to allergens and bacteria on recurrent wheeze and atopy in urban

Journal Pre-proof

Journal Pre-proof

436 children. J Allergy Clin Immunol. 2014;134(3):593-601.e12. PubMed PMID: 24908147;
437 PubMed Central PMCID: PMC4151305.

438 15. Kirjavainen PV, Karvonen AM, Adams RI, Täubel M, Roponen M, Tuoresmäki P, et al.
439 Farm-like indoor microbiota in non-farm homes protects children from asthma development.
440 Nat Med. 2019; 25(7):1089-95. PubMed PMID: 31209334.

441 16. Stein MM, Hrusch CL, Gozdz J, Igartua C, Pivniouk V, Murray SE, et al. Innate
442 immunity and asthma risk in amish and hutterite farm children. N Engl J Med. 2016;375:411-
443 21. PubMed PMID: 27518660; PubMed Central PMCID: PMC5137793.

444 17. O'Connor GT, Lynch SV, Bloomberg GR, Kattan M, Wood RA, Gergen PJ, et al. Early-
445 life home environment and risk of asthma among inner-city children. J Allergy Clin Immunol.
446 2018; 141(4): 1468-75. PubMed PMID: 28939248; PubMed Central PMCID: PMC6521865.

447 18. von Mutius E, Schmid S, PASTURE Study Group. The PASTURE project: EU support
448 for the improvement of knowledge about risk factors and preventive factors for atopy in
449 europe. Allergy. 2006;61:407-13. PubMed PMID: 16512801.

450 19. Karvonen AM, Hyvärinen A, Roponen M, Hoffmann M, Korppi M, Remes S, et al.
451 Confirmed moisture damage at home, respiratory symptoms and atopy in early life: A birth-
452 cohort study. Pediatrics. 2009;124:e329-38. PubMed PMID: 19651571.

453 20. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-
454 high-throughput microbial community analysis on the illumina HiSeq and MiSeq platforms.
455 ISME J. 2012;6(8):1621-4. PubMed PMID: 22402401; PubMed Central PMCID:
456 PMC3400413.

Journal Pre-proof

Journal Pre-proof

457 21. Tang ZZ, Chen G, Alekseyenko AV. PERMANOVA-S: Association test for microbial
458 community composition that accommodates confounders and multiple distances.
459 Bioinformatics. 2016;32:2618-25. PubMed PMID: 27197815; PubMed Central PMCID:
460 PMC5013911.

461 22. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, et al. Oligotyping:
462 Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods
463 Ecol Evol. 2013;4(12). PubMed PMID: 24358444; PubMed Central PMCID: PMC3864673.

464 23. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of
465 the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol.
466 2012;13(9):R79. PubMed PMID:23013615; PubMed Central PMCID: PMC3506950.

467 24. Dannemiller KC, Gent JF, Leaderer BP, Peccia J. Indoor microbial communities:
468 Influence on asthma severity in atopic and nonatopic children. J Allergy Clin Immunol. 2016;
469 138(1):76-83.e1. PubMed PMID: 26851966; PubMed Central PMCID: PMC5357886.

470 25. Kanchongkittiphon W, Mendell MJ, Gaffin JM, Wang G, Phipatanakul W. Indoor
471 environmental exposures and exacerbation of asthma: An update to the 2000 review by the
472 institute of medicine. Environ Health Perspect. 2015;123:6-20. PubMed PMID: 25303775;
473 PubMed Central PMCID: PMC4286274.

474 26. Quigley L, McCarthy R, O'Sullivan O, Beresford TP, Fitzgerald GF, Ross RP, et al. The
475 microbial content of raw and pasteurized cow milk as determined by molecular approaches. J
476 Dairy Sci. 2013;96:4928-37. PubMed PMID: 23746589.

477 27. Debarry J, Garn H, Hanuszkiewicz A, Dickgreber N, Blumer N, von Mutius E, et al.
478 Acinetobacter lwoffii and lactococcus lactis strains isolated from farm cowsheds possess

Journal Pre-proof

Journal Pre-proof

479 strong allergy-protective properties. J Allergy Clin Immunol. 2007;119:1514-21. PubMed
480 PMID: 17481709.

481 28. Lauener RP, Birchler T, Adamski J, Braun-Fahrlander C, Bufe A, Herz U, et al.
482 Expression of CD14 and toll-like receptor 2 in farmers' and non-farmers' children.
483 Lancet.2002;360:465-6. PubMed PMID: 12241724.

484 29. Schuijs MJ, Willart MA, Vergote K, Gras D, Deswarte K, Ege MJ, et al. Farm dust and
485 endotoxin protect against allergy through A20 induction in lung epithelial cells. Science.
486 2015;349:1106-10. PubMed PMID:26339029.

487 30. Fujimura KE, Demoor T, Rauch M, Faruqi AA, Jang S, Johnson CC, et al. House dust
488 exposure mediates gut microbiome lactobacillus enrichment and airway immune defense
489 against allergens and virus infection. Proc Natl Acad Sci U S A. 2014;111:805-10. PubMed
490 PMID: 24344318; PubMed Central PMCID: PMC3896155.

491 31. Holt PG. The mechanism or mechanisms driving atopic asthma initiation: The infant
492 respiratory microbiome moves to center stage. J Allergy Clin Immunol. 2015;136:15-22.
493 PubMed PMID:26145983.

494 32. Depner M, Ege MJ, Cox MJ, Dwyer S, Walker AW, Birzele LT, et al. Bacterial
495 microbiota of the upper respiratory tract and childhood asthma. J Allergy Clin Immunol.
496 2016; 139:826-34.e13. PubMed PMID: 27576124.

497 33. Ottman N, Ruokolainen L, Suomalainen A, Sinkko H, Karisola P, Lehtimäki J, et al. Soil
498 exposure modifies the gut microbiota and supports immune tolerance in a mouse model. J
499 Allergy Clin Immunol. 2019 ;143:1198-1206.e12. PubMed PMID: 30097187.

Journal Pre-proof

Journal Pre-proof

500 34. Tun HM, Konya T, Takaro TK, Brook JR, Chari R, Field CJ, et al. Exposure to household
501 furry pets influences the gut microbiota of infant at 3-4 months following various birth
502 scenarios. Microbiome. 2017;5:40. PubMed PMID: 28381231; PubMed Central PMCID:
503 PMC5382463.

504 35. Haahtela T, Valovirta E, Bousquet J, Mäkelä M and the Allergy Programme Steering
505 Group. The Finnish Allergy Programme 2008–2018 works. Eur Respir J. 2017; 49. PubMed
506 PMID:28642312.

507 36. Leppänen HK, Täubel M, Jayaprakash B, Vepsäläinen A, Pasanen P, Hyvärinen A.
508 Quantitative assessment of microbes from samples of indoor air and dust. J Expo Sci Environ
509 Epidemiol. 2017;28(3):231-41. PubMed PMID: 28975927.

510 37. Täubel M, Rintala H, Pitkäranta M, Paulin L, Laitinen S, Pekkanen J, et al. The occupant
511 as a source of house dust bacteria. J Allergy Clin Immunol. 2009;124:834-40.e47. PubMed
512 PMID: 19767077. 38. Hyytiäinen HK, Jayaprakash B, Kirjavainen PV, Saari SE, Holopainen
513 R, Keskinen J, et al. Crawling-induced floor dust resuspension affects the microbiota of the
514 infant breathing zone. Microbiome. 2018;6:25. PubMed PMID: 29394954; PubMed Central
515 PMCID: PMC5797336.

516

517

518

519

520

Journal Pre-proof

Journal Pre-proof

521

522

523

524

525

526 **Table**

527 **Table I. The associations between richness, Shannon index, the first two axes scores**

528 **(PCoA1 and PCoA2) and the development of ever asthma until the age of 10.5 years and**

529 **current asthma.**

530

	Ever asthma		**Current asthma**	
	aOR (95%CI)	*p-value*	aOR (95%CI)	*p-value*
Richness	0.61 (0.39, 0.95)	*0.03*	0.55 (0.27, 1.12)	*0.10*
Shannon index	0.77 (0.55, 1.07)	*0.12*	0.76 (0.45, 1.30)	*0.32*
PCoA1	0.74 (0.57, 0.98)	*0.03*	0.76 (0.50, 1.16)	*0.20*
PCoA2	0.75 (0.55, 1.02)	*0.07*	0.59 (0.36, 0.98)	*0.04*

531

532 aOR Adjusted Odds Ratios; 95%CI 95% Confidence Intervals; aORs are expressed as

533 interquartile range (IQR) change in the estimate (ln-transformed in axes scores); PCoA1 first

534 axis score of weighted UniFrac based Principal Coordinate Analyses; PCoA2 the second axis

535 score of weighted UniFrac based Principal Coordinate Analyses; Discrete-time hazard models

536 are adjusted for follow-up time, cohort, living on a farm, gender, maternal history of allergic

537 diseases, maternal smoking during pregnancy and number of older siblings. The number of

538 subjects at the beginning of the survey/ the total number of observations in the analyses/ the

Journal Pre-proof

Journal Pre-proof

539 number of the outcome in ever asthma 373/2387/69 and in current asthma 310/2333/29

540 models.

Journal Pre-proof

Journal Pre-proof

541 **Figure legends**

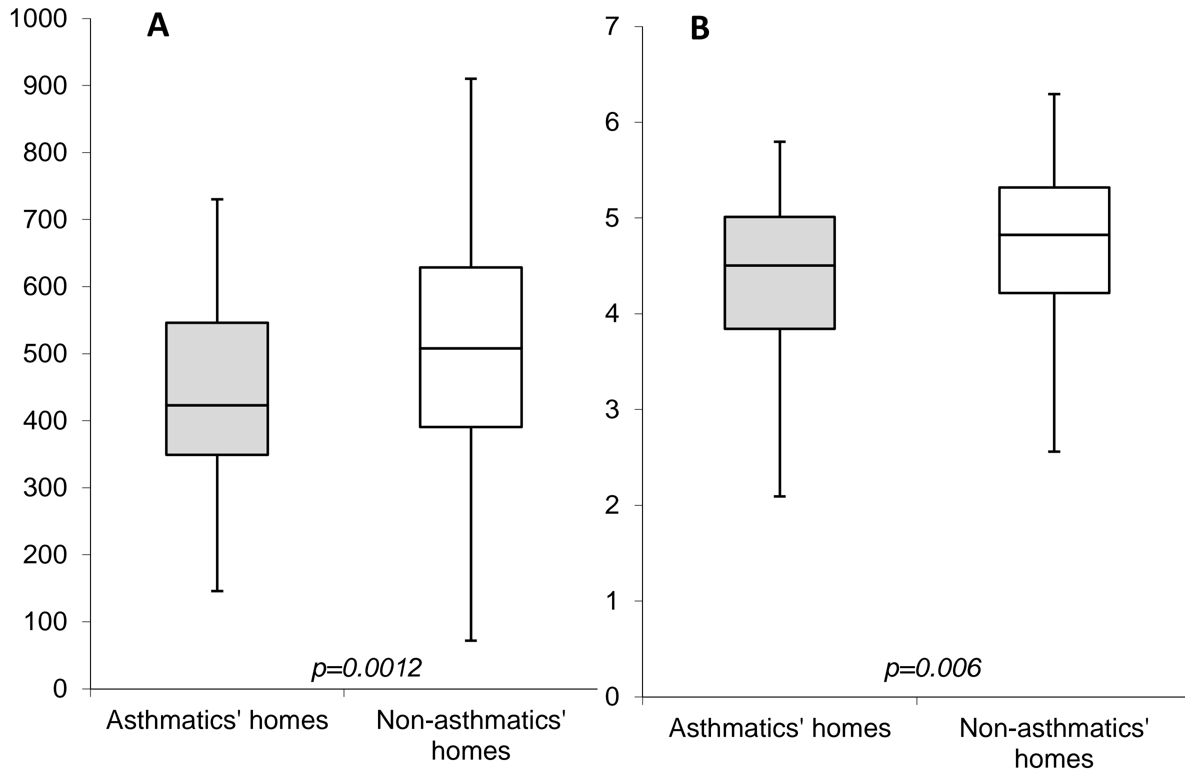
542 Figure 1**. Box-plots of A) bacterial richness and B) Shannon diversity index in homes of**
543 **children with asthma ever (grey) and in homes of non-asthmatics' children (white).**
544 Richness is the number of different OTUs in a sample. The boxplots present minimum, first
545 quartile, median, third quartile, and maximum values. *P*-values are from T-test.
546
547 Figure 2. **Plot of PCoA1 and PCoA2 axes scores by ever asthma status**. PCoA1 is the first
548 and PCoA2 the second axis scores from weighted UniFrac based Principal Coordinate
549 Analyses; children with ever asthma (red dots) and non-asthmatics (black dots); percentages
550 of the variance explained by the axis scores are in the parentheses. Red and black ellipses
551 represent 95%Confidence Intervals from T-test for children with ever asthma and non-
552 asthmatics children, respectively.
553
554 Figure 3. **Spearman Rank correlation coefficients between the first two axes scores,**
555 **PCoA1 (light grey) and PCoA2 (dark grey), and four the most abundant bacterial phyla,**
556 **richness and Shannon index.**
557
558 Figure 4. **Relative abundances of the bacterial genera in living room dust (at age 2**
559 **months) from homes of children A) with ever asthma and B) without asthma.** The 641
560 genera with mean relative abundance <1% in the whole dataset are combined in the sum
561 variable. Phylum names are given in parenthesis. U. denotes "unassigned" and O. "other"
562 genus within a family.*p-*value <0.05* from Kruskal-Wallis test.
563
564 Figure 5. **Adjusted odds ratios (95% confidence intervals) between the selected 41 genera**
565 **and asthma ever**. Genera have been ordered by phylum. aORs are expressed as interquartile

Journal Pre-proof

Journal Pre-proof

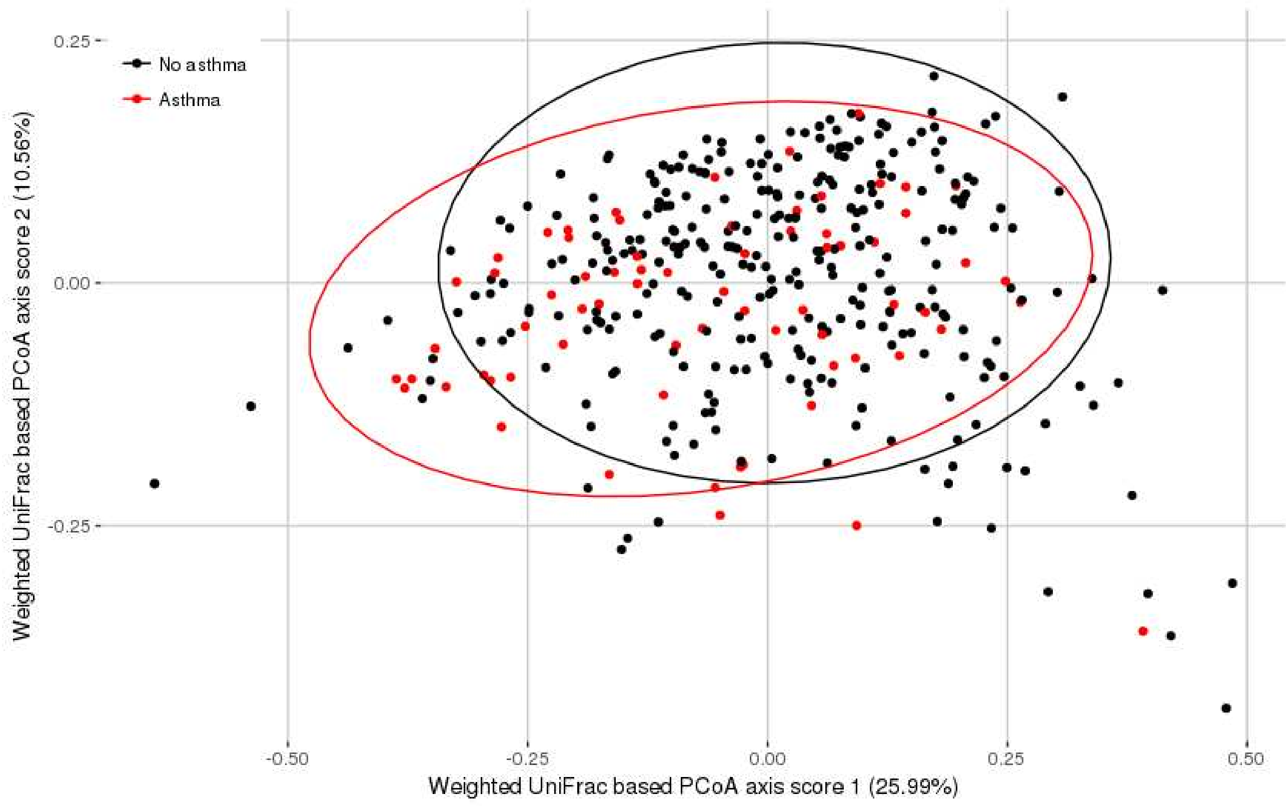
566 range (IQR) change in the estimate (ln-transformed). Models are adjusted for follow-up time,
567 living on a farm, cohort, gender, maternal history of allergic diseases, maternal smoking
568 during pregnancy and the number of older siblings. U. denotes “unassigned” and O. “other”
569 genus within a family, and C. denotes *Chloroflexi* (phylum).

Journal Pre-proof

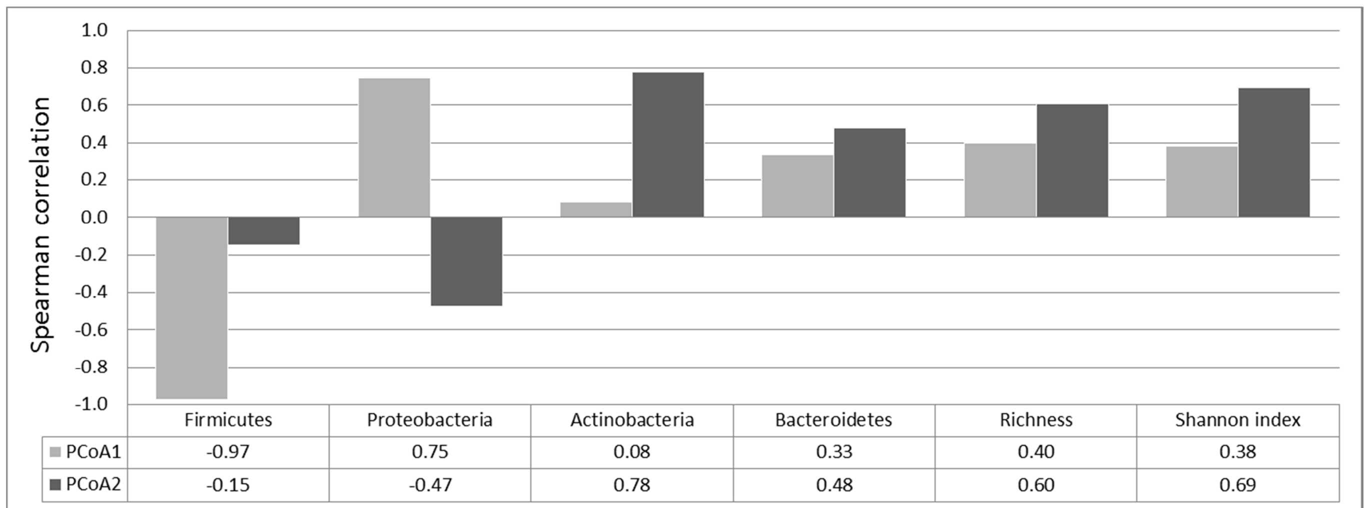


Journal Pre-proof

Journal Pre-proof



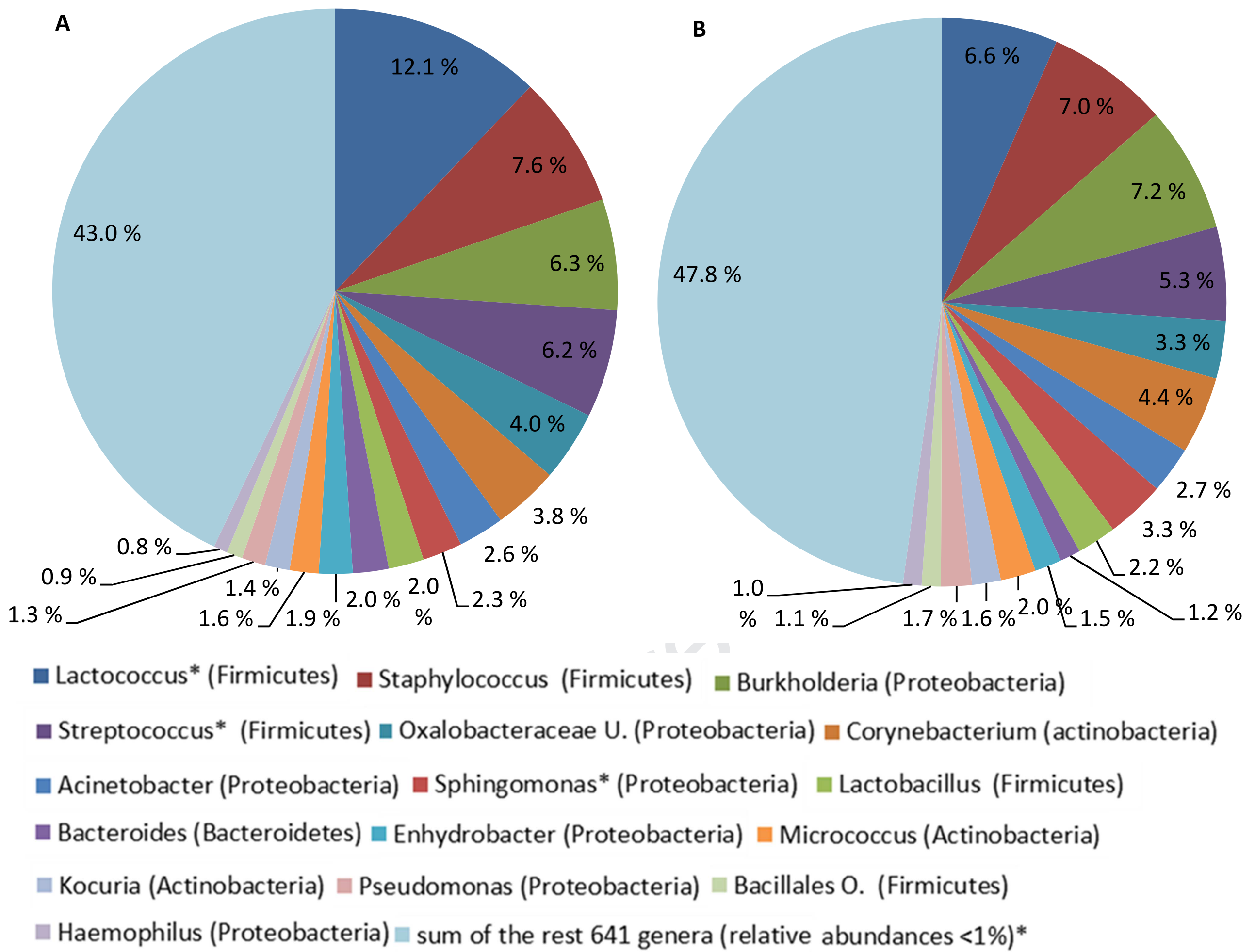
Journal



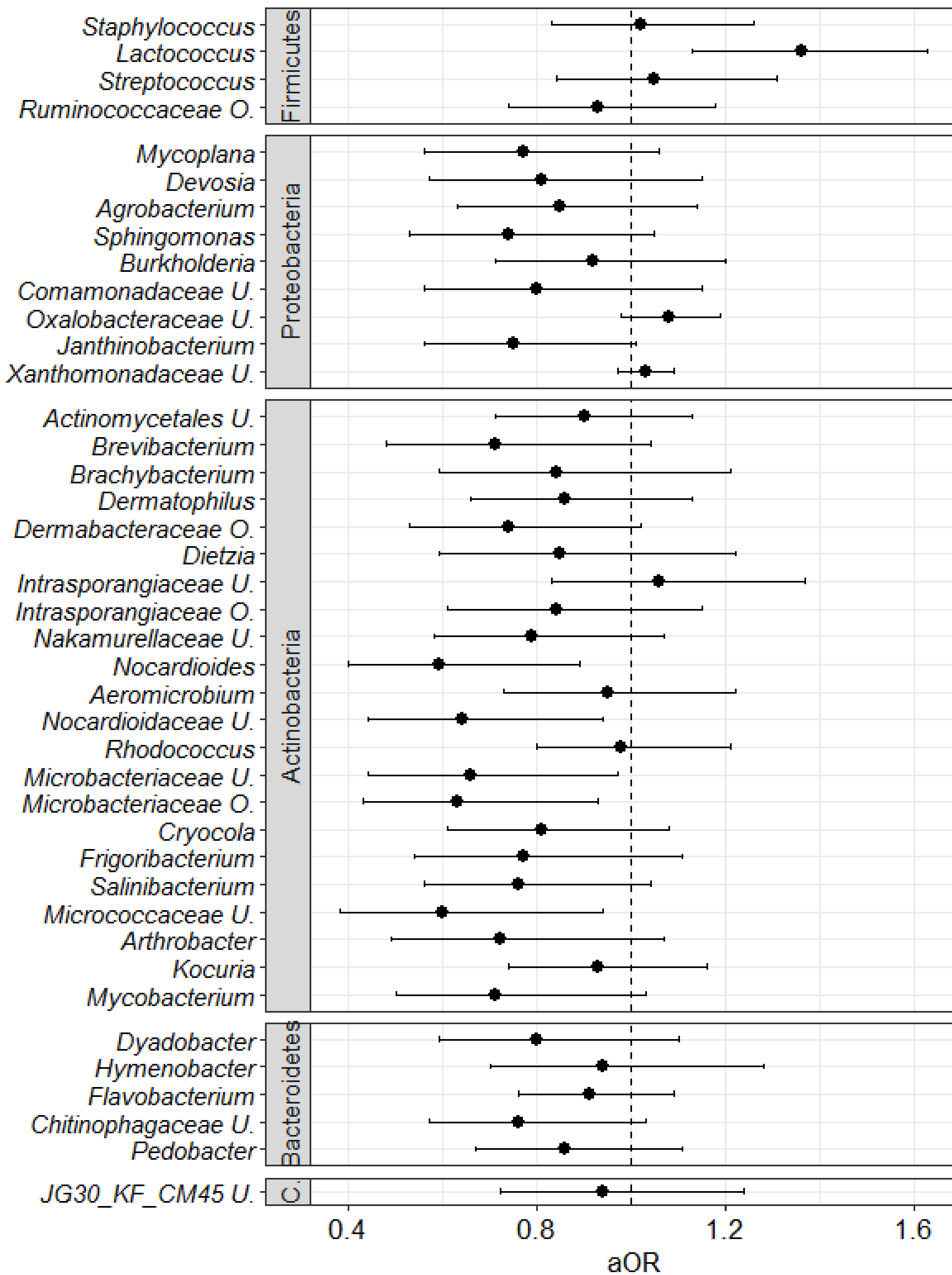
	Firmicutes	Proteobacteria	Actinobacteria	Bacteroidetes	Richness	Shannon index
PCoA1	-0.97	0.75	0.08	0.33	0.40	0.38
PCoA2	-0.15	-0.47	0.78	0.48	0.60	0.69

Journal Pre-proof

Journal Pre-proof



Journal Pre-proof



Journal Pre-proof

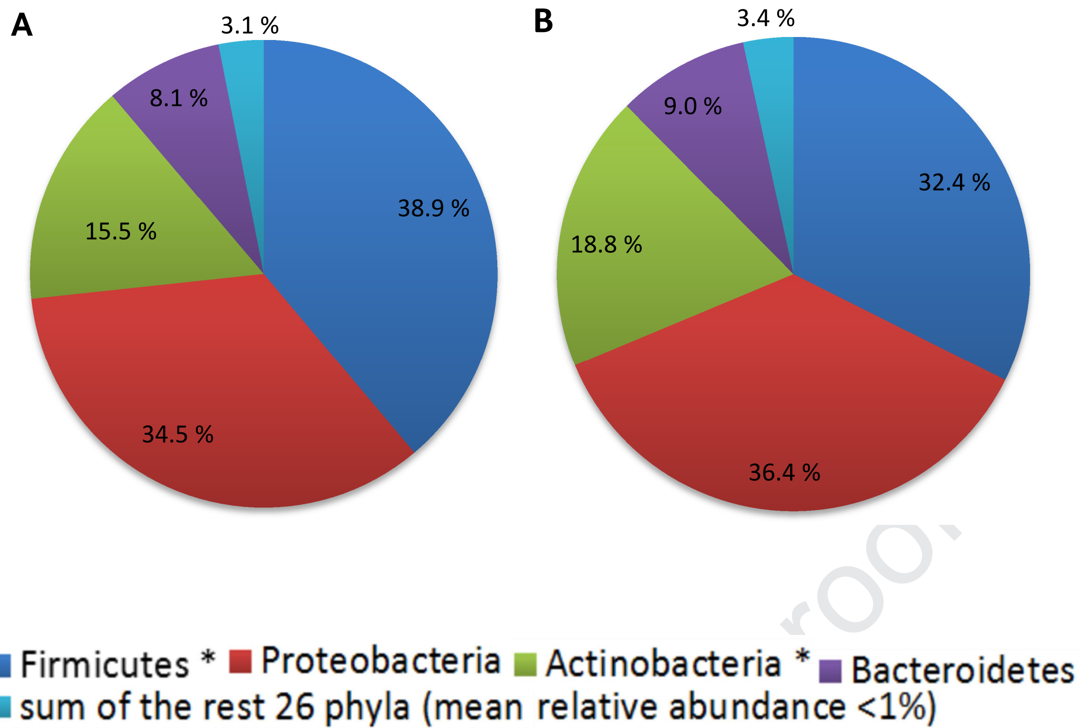


Figure E1. Relative abundances of the bacterial phyla in living room dust (at age 2 months) from homes of A) asthmatics and B) non-asthmatic children. The 26 phyla with mean relative abundance <1% in the whole dataset are combined into 'rest 26 phyla' category. **p-value <0.05* from Kruskal-Wallis test.

Journal Pre-proof

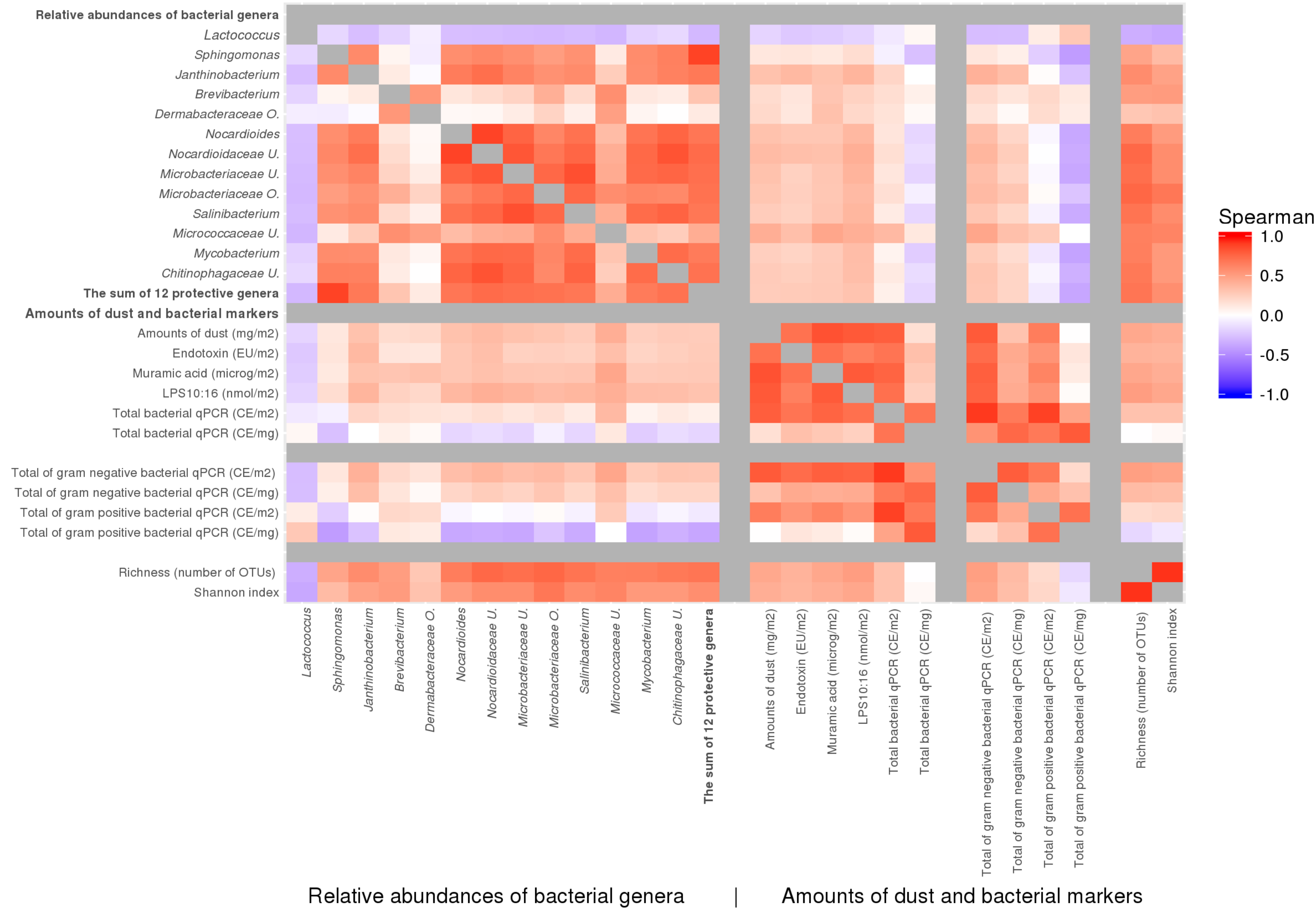


Figure E2. Heat map of Spearman correlations between the relative abundances of 13 bacterial genera, amounts of dust, bacterial markers, qPCRs, richness and Shannon index.

Journal Pre-proof

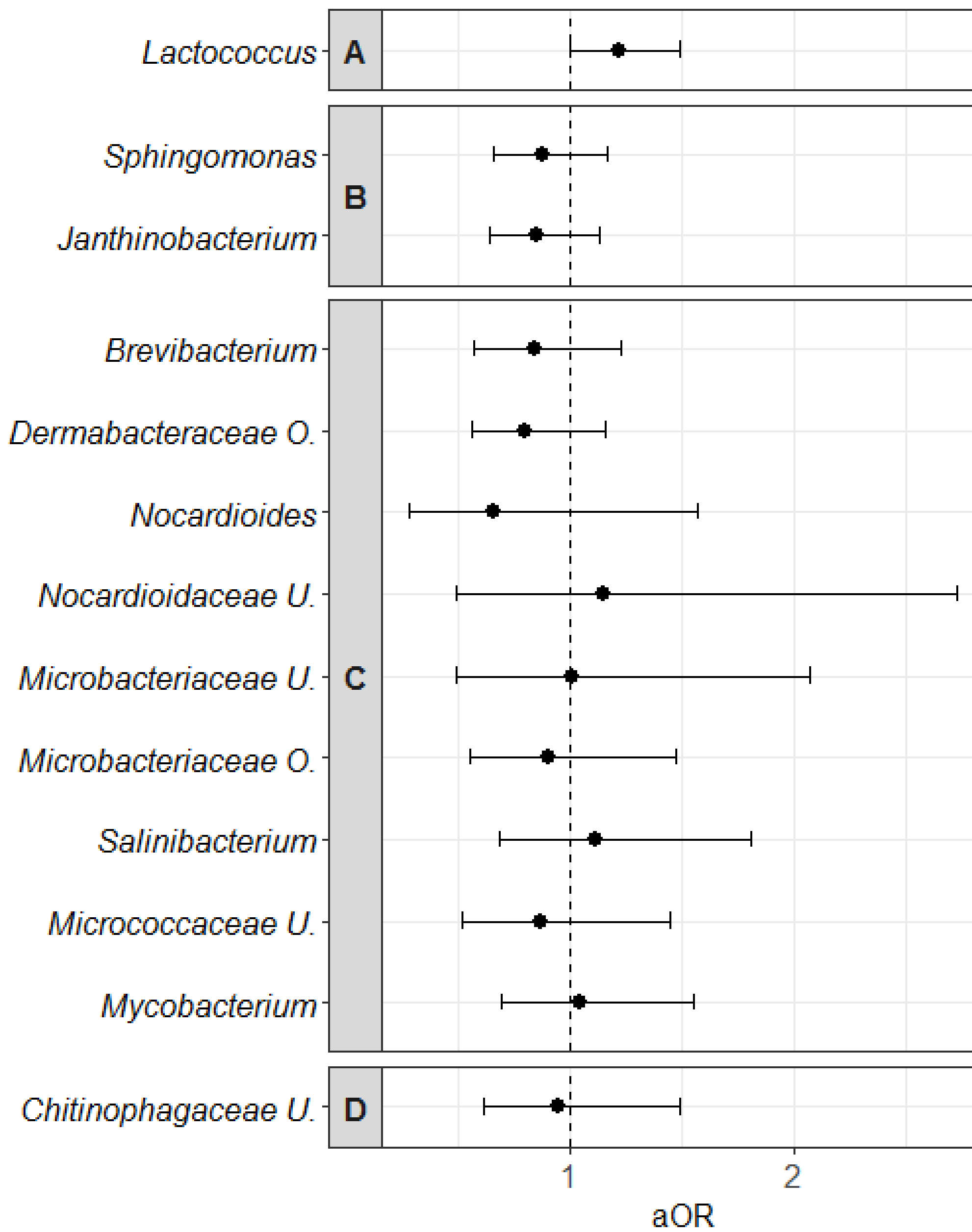


Figure E3. Mutually adjusted odds ratios (95% confidence intervals) between 13 taxa and ever asthma. Models are adjusted for follow-up time, living on a farm, cohort, gender, maternal history of allergic diseases, maternal smoking during pregnancy, the number of older siblings, and all genera simultaneously. U. denotes “unassigned” and O. “other” genus within a family. A=Firmicutes, B=Proteobacteria, C=Actinobacteria, D=Chloroflex.

Table E1. The associations between the first six axes scores of weighted UniFrac based Principal Coordinate Analyses (PCoA) and the development of asthma ever until the age of 10.5 years and current asthma.

	Ever asthma (n=69)					Current Asthma (n=29)					
Axes scores[ǁ]	N[†]	N	%	aOR (95%CI)	*p* -value[‡]	N[†]	N	%	aOR (95%CI)	*p* -value[‡]	%[§]
PCoA1											
lowest tertile	127	817	3.9	1		108	806	1.7	1		
middle tertile	124	798	2.6	0.70 (0.39, 1.25)		101	768	1.3	0.72 (0.31, 1.70)		
highest tertile	122	772	2.1	0.57 (0.30, 1.08)^	*0.08*	99	759	0.7	0.38 (0.13, 1.13)	*0.07*	25.99
PCoA2											
lowest tertile	124	808	3.5	1		109	816	1.2	1		
middle tertile	126	776	3.7	1.02 (0.59, 1.78)		101	750	1.7	1.19 (0.49, 2.87)		
highest tertile	123	803	1.5	0.44 (0.21, 0.90)*	***0.02***	98	767	0.8	0.63 (0.21, 1.86)	*0.39*	10.56
PCoA3											
lowest tertile	127	837	2.5	1		105	804	0.9	1		
middle tertile	123	795	2.5	1.05 (0.56, 1.96)		104	794	1.3	1.51 (0.56, 4.05)		
highest tertile	123	755	3.7	1.55 (0.86, 2.79)	*0.14*	99	735	1.6	1.72 (0.66, 4.85)	*0.26*	5.53
PCoA4											
lowest tertile	119	761	3.2	1		101	761	1.5	1		
middle tertile	128	836	2.8	0.94 (0.52, 1.70)		106	807	1.1	0.83 (0.34, 2.04)		
highest tertile	126	790	2.8	1.01 (0.55, 1.85)	*0.97*	101	765	1.2	1.04 (0.42, 2.61)	*0.93*	4.67
PCoA5											
lowest tertile	126	817	3.3	1		106	794	1.5	1		
middle tertile	124	781	2.8	0.87 (0.48, 1.56)		101	766	1.3	0.86 (0.36, 2.08)		
highest tertile	123	789	2.5	0.77 (0.42, 1.40)	*0.39*	101	773	0.9	0.58 (0.22, 1.54)	*0.27*	3.44
PCoA6											
lowest tertile	126	771	3.6	1		102	759	1.7	1		
middle tertile	124	820	1.8	0.50 (0.26, 0.95)*		103	797	0.8	0.52 (0.19, 1.39)		
highest tertile	123	796	3.3	0.86 (0.49, 1.51)	*0.60*	103	777	1.3	0.79 (0.33, 1.87)	*0.59*	2.68

Journal Pre-proof

[‖]Axes scores are divided into tertiles, N† the number of subjects at the beginning of the survey; N the total number of observations in the analyses; % percentage of the outcome in the given group; aOR Adjusted Odds Ratios; 95%CI 95% Confidence Intervals; ǂp-value for linear trend test; §percentage of the component explained the variance in PCoA analyses. Models are adjusted for follow-up time, living on a farm, cohort, gender, maternal history of allergic diseases, maternal smoking during pregnancy, and the number of older siblings. *p* -value * *<0.05, ^<0.1* .

Journal Pre-proof

Table E2. Description of relative abundances of 41 bacterial genera (sorted by phylum, class, order and family), which correlated either axis scores 1 or 2 (r >|0.4|) in 394 dust samples.

class	order	family	genus	r^1	r^2	median	Interquartile range	MAX	<DL
Firmicutes (phylum)									
Bacilli	*Bacillales*	*Staphylococcaceae*	*Staphylococcus*	-0.50	-0.08	5.17%	(3.20, 8.50)	45.02%	0
Bacilli	*Lactobacillales*	*Streptococcaceae*	*Lactococcus*	-0.59	-0.24	3.94%	(1.37, 10.01)	56.87%	0
Bacilli	*Lactobacillales*	*Streptococcaceae*	*Streptococcus*	0.46	0.41	3.80%	(1.77, 7.04)	62.47%	0
Clostridia	*Clostridiales*	*Ruminococcaceae*	unassigned	0.01	0.41	0.20%	(0.06, 0.72)	9.19%	15
Proteobacteria (phylum)									
Alphaproteobacteria	*Caulobacterales*	*Caulobacteraceae*	*Mycoplana*	0.46	0.41	0.14%	(0.05, 0.28)	3.69%	19
Alphaproteobacteria	*Rhizobiales*	*Hyphomicrobiaceae*	*Devosia*	0.43	0.41	0.12%	(0.05, 0.24)	1.55%	13
Alphaproteobacteria	*Rhizobiales*	*Rhizobiaceae*	*Agrobacterium*	0.42	0.30	0.17%	(0.08, 0.30)	1.53%	8
Alphaproteobacteria	*Sphingomonadales*	*Sphingomonadaceae*	*Sphingomonas*	0.45	0.12	2.08%	(1.08, 3.99)	34.51%	0
Betaproteobacteria	*Burkholderiales*	*Burkholderiaceae*	*Burkholderia*	0.15	-0.48	3.93%	(1.55, 9.50)	58.27%	0
Betaproteobacteria	*Burkholderiales*	*Comamonadaceae*	unassigned	0.47	0.26	0.39%	(0.19, 0.67)	2.45%	0
Betaproteobacteria	*Burkholderiales*	*Oxalobacteraceae*	unassigned	0.43	-0.16	1.82%	(0.86, 3.21)	56.68%	0
Betaproteobacteria	*Burkholderiales*	*Oxalobacteraceae*	*Janthinobacterium*	0.45	0.12	0.17%	(0.07, 0.43)	3.80%	14
Gammaproteobacteria	*Xanthomonadales*	*Xanthomonadaceae*	unassigned	0.40	0.25	0.19%	(0.10, 0.33)	14.48%	4
Actinobacteria (phylum)									
Actinobacteria	*Actinomycetales*	unassigned	unassigned	0.36	0.45	0.18%	(0.09, 0.38)	5.46%	4
Actinobacteria	*Actinomycetales*	*Brevibacteriaceae*	*Brevibacterium*	0.03	0.47	0.21%	(0.08, 0.52)	6.79%	5
Actinobacteria	*Actinomycetales*	*Dermabacteraceae*	*Brachybacterium*	0.11	0.50	0.20%	(0.06, 0.58)	6.48%	12
Actinobacteria	*Actinomycetales*	*Dermabacteraceae*	*Dermatophilus*	0.01	0.41	0.10%	(0.04, 0.22)	4.29%	22
Actinobacteria	*Actinomycetales*	*Dermabacteraceae*	other	-0.03	0.46	0.10%	(0.04, 0.18)	2.65%	15
Actinobacteria	*Actinomycetales*	*Dietziaceae*	*Dietzia*	0.08	0.48	0.05%	(0.02, 0.15)	1.14%	38
Actinobacteria	*Actinomycetales*	*Intrasporangiaceae*	unassigned	0.27	0.45	0.16%	(0.06, 0.34)	3.03%	10
Actinobacteria	*Actinomycetales*	*Intrasporangiaceae*	other	0.28	0.50	0.17%	(0.08, 0.35)	2.87%	4
Actinobacteria	*Actinomycetales*	*Nakamurellaceae*	unassigned	0.43	0.32	0.08%	(0.03, 0.21)	1.40%	24
Actinobacteria	*Actinomycetales*	*Nocardioidaceae*	*Nocardioides*	0.43	0.30	0.07%	(0.03, 0.18)	0.77%	24
Actinobacteria	*Actinomycetales*	*Nocardioidaceae*	*Aeromicrobium*	0.41	0.43	0.06%	(0.03, 0.14)	0.92%	27

Journal Pre-proof

Actinobacteria	*Actinomycetales*	*Nocardioidaceae*	unassigned	0.45	0.32	0.28%	(0.11, 0.70)	3.04%	7
Actinobacteria	*Actinomycetales*	*Nocardioidaceae*	*Rhodococcus*	0.43	0.40	0.09%	(0.04, 0.17)	1.37%	10
Actinobacteria	*Actinomycetales*	*Microbacteriaceae*	unassigned	0.44	0.43	0.16%	(0.06, 0.32)	2.06%	7
Actinobacteria	*Actinomycetales*	*Microbacteriaceae*	other	0.40	0.46	0.10%	(0.04, 0.17)	1.46%	12
Actinobacteria	*Actinomycetales*	*Microbacteriaceae*	*Cryocola*	0.44	0.38	0.07%	(0.03, 0.17)	1.37%	24
Actinobacteria	*Actinomycetales*	*Microbacteriaceae*	*Frigoribacterium*	0.43	0.37	0.11%	(0.05, 0.25)	1.94%	17
Actinobacteria	*Actinomycetales*	*Microbacteriaceae*	*Salinibacterium*	0.44	0.43	0.09%	(0.03, 0.21)	1.97%	20
Actinobacteria	*Actinomycetales*	*Micrococcaceae*	unassigned	0.25	0.56	0.48%	(0.22, 0.94)	7.40%	2
Actinobacteria	*Actinomycetales*	*Micrococcaceae*	*Arthrobacter*	0.35	0.46	0.11%	(0.04, 0.34)	3.22%	14
Actinobacteria	*Actinomycetales*	*Micrococcaceae*	*Kocuria*	0.01	0.49	1.08%	(0.51, 2.03)	18.63%	0
Actinobacteria	*Actinomycetales*	*Mycobacteriaceae*	*Mycobacterium*	0.41	0.34	0.09%	(0.04, 0.17)	1.02%	13
Bacteroidetes (phylum)									
Cytophagia	*Cytophagales*	*Cytophagaceae*	*Dyadobacter*	0.43	0.40	0.07%	(0.02, 0.17)	0.97%	26
Cytophagia	*Cytophagales*	*Cytophagaceae*	*Hymenobacter*	0.42	0.22	0.37%	(0.13, 0.86)	3.69%	1
Flavobacteriia	*Flavobacteriales*	*Flavobacteriaceae*	*Flavobacterium*	0.44	0.30	0.24%	(0.09, 0.58)	13.26%	11
[Saprospirae]	*[Saprospirales]*	*Chitinophagaceae*	unassigned	0.40	0.30	0.16%	(0.06, 0.33)	2.01%	14
Sphingobacteriia	*Sphingobacteriales*	*Sphingobacteriaceae*	*Pedobacter*	0.49	0.31	0.44%	(0.16, 0.93)	14.40%	5
Chloroflexi (phylum)									
Thermomicrobia	*JG30-KF-CM45*	unassigned	unassigned	0.31	0.49	0.10%	(0.04, 0.24)	3.81%	11

r^1 Spearman's correlations with axis score 1 (PCoA1), r^2 Spearman correlations with axis score 2 (PCoA2), median relative abundance (median), 25th and 75th percentile of relative abundance (Interquartile range), maximum value of the relative abundance (MAX) and the number of samples under detection limit (DL).

Table E3. Associations between relative abundance of Lactococcus, sum of 12 protective genera and ever asthma additionally adjusted for bacterial diversity indices, amount of dust, endotoxin, LPS and muramic acids.

	Ever asthma						
		Additionally adjusted for					
		Richness	Shannon index	Amount of dust	Endotoxin	LPS $_{10:0\text{-}16:0}$	Muramic acids
	aOR (95%CI)	aOR (95%CI)	aOR (95%CI)	aOR (95%CI)	aOR (95%CI)	aOR (95%CI)	aOR (95%CI)
Lactococcus	1.36 (1.13, 1.63)	1.29 (1.06, 1.58)	1.35 (1.08, 1.67)	1.32 (1.09, 1.59)	1.42 (1.16, 1.75)	1.36 (1.13, 1.64)	1.36 (1.10, 1.69)
The sum of 12 protective genera	0.50 (0.32, 0.79)	0.53 (0.30, 0.96)	0.50 (0.30, 0.85)	0.53 (0.33, 0.84)	0.52 (0.32, 0.86)	0.54 (0.34, 0.86)	0.63 (0.40, 1.00)

Models are adjusted for follow-up time, living on a farm, cohort, gender, maternal history of allergic diseases, maternal smoking during pregnancy, and the number of older siblings, and additionally for diversity indices (continues), amount of dust and general microbial markers (units/m^2 in tertiles); aOR Adjusted Odds Ratios; 95%CI 95% Confidence Intervals; aORs are expressed as IQR change in the estimate (ln-transformed).

Journal

Journal Pre-proof

Table E4. Associations between bacterial richness, Shannon index and ever asthma additionally adjusted for relative Lactococcus and the sum of 12 protective genera.

	Ever asthma		
		Additionally adjusted for	
		Lactococcus	The sum of 12 protective genera
	aOR (95%CI)	aOR (95%CI)	aOR (95%CI)
Richness	0.61 (0.39, 0.95)	0.74 (0.46, 1.19)	0.91 (0.49, 1.69)
Shannon index	0.77 (0.55, 1.07)	0.97 (0.65, 1.45)	0.99 (0.66, 1.50)

Models are adjusted for follow-up time, living on a farm, cohort, gender, maternal history of allergic diseases, maternal smoking during pregnancy, and the number of older siblings; aOR Adjusted Odds Ratios; 95%CI 95% Confidence Intervals; aORs are expressed as IQR change in the estimate.

Journal Pre-proof

Journal Pre-proof

Supplemental Table E5. Determinants of relative abundance of *Lactococcus* genus and the sum of 12 protective genera.

	Lactococcus (>median)					Sum of 12 protective genera (>median)			
	N	%	*p-value* [1]	cOR (95%CI)	*p-value* [2]	%	*p-value* [1]	cOR (95%CI)	*p-value* [2]
Living on a farm, no (ref.)	278	57.2		1		40.7		1	
yes	116	32.8	***<0.0001***	0.36 (0.23, 0,57)	***<0.0001***	72.4	***<0.0001***	3.83 (2.39, 6.15)	***<0.0001***
	N	%	*p-value* [1]	aOR (95%CI)	*p-value* [2]	%	*p-value* [1]	aOR (95%CI)	*p-value* [2]
Construction year, before 1970 (ref.)	127	30.7		1		69.3		1	
1970-1989	135	61.5		2.98 (1.75, 5.06)	***<0.0001***	43.7		0.45 (0.26, 0.77)	***0.003***
after 1989	115	58.3	***<0.0001***	2.39 (1.37, 4.19)	***0.002***	36.5	***<0.0001***	0.37 (0.21, 0.66)	***0.001***
Floors, 1 (ref.)	230	55.2		1		45.2		1	
≥2	147	42.2	***0.01***	0.68 (0.44, 1.05)	*0.08*	57.8	***0.02***	1.39 (0.90, 2.16)	*0.14*
Basement, no (ref.)	264	56.8		1		44.3		1	
yes	113	34.5	***<0.0001***	0.54 (0.33, 0.88)	***0.01***	63.7	***0.001***	1.42 (0.86, 2.34)	*0.17*
Ground, slab (ref.)	329	52.9		1		45.9		1	
other	48	31.3	***0.005***	0.54 (0.27, 1.06)	*0.07*	79.2	***<0.0001***	3.28 (1.54, 7.00)	***0.002***
Log frame, timber (ref.)	319	48.6		1		54.6		1	
brick/ concrete	58	58.6	*0.16*	1.19 (0.66, 2.13)	*0.56*	25.9	***<0.0001***	0.37 (0.19, 0.71)	***0.003***
Exterior wall (siding) material									
timber (ref.)	236	46.6		1		57.2		1	
brick	119	59.7		1.51 (0.95, 2.39)	*0.08*	39.5		0.55 (0.35, 0.88)	***0.01***
other	22	36.4	***0.03***	0.59 (0.23, 1.49)	*0.27*	31.8	***0.002***	0.36 (0.13, 0.94)	***0.04***
Ventilation, natural (ref.)	165	39.4		1		63.0		1	
mechanical exhaust ventilation (from kitchen/bathroom)	137	60.6		1.92 (1.18, 3.11)	***0.01***	38.7		0.47 (0.29, 0.78)	***0.003***
house-specific mechanical supply and exhaust ventilation	75	54.7	***0.001***	1.49 (0.84, 2.64)	*0.17*	42.7	***<0.0001***	0.57 (0.32, 1.03)	*0.06*
Heating with wood, no (ref.)	80	57.5		1		22.5		1	
yes, occasionally	40	47.5		0.96 (0.43, 2.13)	*0.91*	62.5		3.94 (1.67, 9.31)	***0.002***

yes, regularly	250	46.8	*0.24*	0.84 (0.43, 1.43)	*0.53*	55.6	***<0.0001***	3.25 (1.78, 5.91)	***0.0001***
Having dog(s) and keeping indoors									
no dog/ never indoors (ref.)	261	56.7		1		42.5		1	
sometimes/often indoors	40	50.0		1.08 (0.53, 2.19)	*0.84*	77.5		3.53 (1.57, 7.96)	***0.002***
mostly indoors	88	30.7	***0.0001***	0.32 (0.19, 0.55)	***<0.0001***	61.4	***<0.0001***	2.24 (1.34, 3.74)	***0.002***
Having cat(s) and keeping indoors									
no cat/ never indoors (ref.)	301	55.2		1		47.8		1	
sometimes/often indoors	30	26.7		0.46 (0.19, 1.13)	*0.09*	60.0		0.73 (0.29, 1.65)	*0.41*
mostly indoors	62	35.5	***0.001***	0.55 (0.31, 1.00)	***0.05***	56.5	*0.25*	0.94 (0.52, 1.72)	*0.85*
Number of pet species (dog,cat,rabbit,turtle, rodents, fish,bird)									
0 (ref.)	154	66.2		1		31.8		1	
1	162	44.4		0.46 (0.29, 0.74)	***0.001***	58.0		2.43 (1.51, 3.91)	***0.0003***
2-6	78	29.5	***<0.0001***	0.29 (0.15, 0.57)	***0.0003***	69.2	***<0.0001***	2.82 (1.46, 5.45)	***0.002***
Contact with horses during first trimester									
never/seldom (ref.)	350	50.9		1		46.9		1	
monthly/weekly	24	41.7		0.76 (0.32, 1.80)	*0.53*	79.2		4.15 (1.48, 11.65)	***0.007***
daily	20	45.0	*0.62*	1.32 (0.50, 3.44)	*0.57*	70.0	***0.002***	1.47 (0.52, 4.16)	*0.47*
Last vacuumed (days), 0-1 (ref.)	142	48.6		1		53.5		1	
2-3	114	42.1		0.68 (0.41, 1.13)	*0.14*	52.6		1.12 (0.67, 1.88)	*0.67*
≥4	138	58.0	***0.04***	1.20 (0.73, 1.95)	*0.48*	44.2	*0.24*	0.89 (0.54, 1.47)	*0.66*

Median levels are used as a cut-off values for define high levels of exposure; N Total number of observations in given class; % percentage of the high (>median) level of the sum of relative abundances of 12 genera in given class; *p* - value[1] *p* -values are from X^2 test; cOR (95%CI) crude odds ratio (farming vs. *Lactococcus* or the sum of 12 genera) and its confidence intervals; *p* -value[2] *p* -values are from logistic regression models. aOR (95%CI) adjusted odds ratios and its confidence intervals; the models are adjusted for farming; ref. denotes refrence category.

Journal Pre-proof

Table E6. Adjusted associations between relative abundance of *Lactococcus* , the sum of 12 protective genera, diversity indices, and wheezing phenotypes.

	Transient wheeze		**Intermediate wheeze**		**Late onset wheeze**		**Persistent wheeze**	
	aOR (95%CI)	*p*-value	aOR (95%CI)	*p*-value	aOR (95%CI)	*p*-value	aOR (95%CI)	*p*-value
Lactococcus	1.02 (0.78, 1.33)	*0.90*	0.88 (0.44, 1.77)	*0.72*	1.19 (0.71, 1.99)	*0.52*	1.14 (0.75, 1.72)	*0.54*
The sum of 12 protective genera	1.05 (0.73, 1.50)	*0.81*	0.84 (0.31, 2.28)	*0.73*	0.46 (0.15, 1.45)	*0.18*	0.57 (0.25, 1.32)	*0.19*
Richness	1.14 (0.69, 1.89)	*0.61*	0.86 (0.26, 2.83)	*0.80*	0.68 (0.21, 2.17)	*0.51*	0.87 (0.36, 2.07)	*0.75*
Shannon index	0.96 (0.64, 1.45)	*0.84*	1.21 (0.42, 3.46)	*0.73*	0.80 (0.33, 1.98)	*0.64*	0.90 (0.45, 1.81)	*0.78*

The total number of observations in the analyses 322; Never/infrequent wheeze was used as a reference group, n=201 (62.4%); Transient wheeze, n=77 (23.9%); Intermediate wheeze, n=12 (3.7%); Late onset wheeze, n=11 (3.0%); Persistent wheeze, n=21 (6.5%); aOR Adjusted Odds Ratios; 95%CI 95% Confidence Intervals; aORs are expressed as IQR change in the estimate; Multinomial logistic regression models are adjusted for cohort, living on a farm, gender, maternal history of allergic diseases, maternal smoking during pregnancy and the number of older siblings.

Journ

Table E7. Associations between relative abundance of *Lactococcus* genus, the sum of 12 protective genera, bacterial richness, Shannon index and respiratory symptoms.

	Wheezing without cold			**Any wheezing**			**Nocturnal cough wihout cold**		
	N/n	aOR (95%CI)	*p-value*	N/n	aOR (95%CI)	*p-value*	N/n	aOR (95%CI)	*p-value*
	2716/213			2736/462			2711/475		
Lactococcus		1.02 (0.84, 1.24)	*0.84*		1.07 (0.92, 1.24)	*0.39*		1.12 (0.98, 1.28)	*0.10*
The sum of 12 protective genera		0.71 (0.51, 0.99)	*0.04*		0.79 (0.64, 0.99)	*0.04*		0.86 (0.69, 1.07)	*0.17*
Richness		0.85 (0.56, 1.27)	*0.42*		0.89 (0.66, 1.20)	*0.43*		0.87 (0.65, 1.16)	*0.34*
Shannon index		0.90 (0.67, 1.20)	*0.46*		0.95 (0.76, 1.20)	*0.69*		0.88 (0.69, 1.12)	*0.29*

N total number of observations in the analyses; n total number of symptoms; Models are adjusted for follow-up time, living on a farm, cohort, gender, maternal history of allergic diseases, maternal smoking during pregnancy, and the number of older siblings; aOR Adjusted Odds Ratios; 95%CI 95% Confidence Intervals; aORs are expressed as IQR change in the estimate.

Journal

Journal Pre-proof

Table E8. Associations between relative abundances of *Lactococcus* oligotypes and the development of asthma.

				Ever asthma		Current asthma	
Oligotype¤	number of sequences	mean relative abundance	<DL	aOR (95%CI)	*p* -value	aOR (95%CI)	*p* -value
GGCCAAGGA	355,480	0.07240	1	**1.33 (1.12, 1.58)**	***0.001***	1.20 (0.92, 1.58)	*0.17*
ATACCATGA	9,241	0.00188	22	1.12 (0.96, 1.31)	*0.14*	1.18 (0.91, 1.52)	*0.15*
GGCTAGGGA	2,042	0.00043	108	**1.13 (1.03, 1.23)**	***0.006***	1.06 (0.92, 1.21)	*0.41*
GGCCAAGGG	1,790	0.00038	88	**1.25 (1.11, 1.41)**	***0.0003***	**1.31 (1.08, 1.57)**	***0.005***
ATACAATGA	1,741	0.00036	110	1.10 (0.98, 1.22)	*0.10*	1.14 (0.98, 1.31)	*0.08*
GGTCAGGGG	1,511	0.00033	155	**1.08 (1.03, 1.14)**	***0.002***	**1.21 (1.06, 1.37)**	***0.004***
GGCCAAGTA	1,124	0.00024	149	**1.07 (1.00, 1.15)**	***0.05***	1.05 (0.94, 1.17)	*0.37*
GGCCAATGA	1,102	0.00023	155	**1.14 (1.03, 1.25)**	***0.008***	1.10 (0.95, 1.27)	*0.21*
GGCCAGGGA	992	0.00021	134	**1.25 (1.11, 1.41)**	***0.0003***	**1.21 (1.02, 1.43)**	***0.03***
GGTCAGGGA	815	0.00017	172	**1.15 (1.03, 1.29)**	***0.015***	**1.19 (1.01, 1.40)**	***0.04***

Models are adjusted for follow-up time, living on a farm, cohort, gender, maternal history of allergic diseases, maternal smoking during pregnancy, and the number of older siblings; aOR Adjusted Odds Ratios; 95%CI 95% Confidence Intervals, ORs are expressed as IQR change in the estimate (ln-transformed relative abundance); ¤ Letters referes to nucleotides in 68, 127, 208, 209, 213, 220, 241, 243, 245 positions. Significant results (p<0.05) are in boldface.

Table E9. Associations between relative abundances or loads of 13 bacterial genera and ever asthma.

	Ever asthma			
	Relative abundance		Load (CE/m^2)	
Bacterial genus	aOR (95%CI)	*p* - value	aOR (95%CI)	*p* - value
Lactococcus	**1.36 (1.13, 1.63)**	***0.001***	**1.49 (1.03, 2.16)**	***0.03***
Sphingomonas	0.74 (0.53, 1.05)	*0.09*	0.77 (0.56, 1.06)	*0.11*
Janthinobacterium	0.75 (0.55, 1.01)	*0.06*	0.90 (0.72, 1.10)	*0.28*
Brevibacterium	0.71 (0.48, 1.04)	*0.08*	0.81 (0.61, 1.08)	*0.16*
Dermabacteraceae O.	0.74 (0.53, 1.02)	*0.07*	0.85 (0.70, 1.02)	*0.09*
Nocardioides	**0.59 (0.40, 0.89)**	***0.01***	0.87 (0.73, 1.03)	*0.10*
Nocardioidaceae U.	**0.64 (0.44, 0.94)**	***0.02***	0.81 (0.64, 1.02)	*0.07*
Microbacteriaceae U.	**0.66 (0.44, 0.97)**	***0.04***	0.85 (0.68, 1.08)	*0.18*
Microbacteriaceae O.	**0.63 (0.43, 0.93)**	***0.02***	**0.81 (0.66, 0.99)**	***0.04***
Salinibacterium	0.76 (0.56, 1.04)	*0.09*	**0.82 (0.69, 0.98)**	***0.03***
Micrococcaceae U.	**0.60 (0.38, 0.94)**	***0.03***	0.85 (0.64, 1.13)	*0.26*
Mycobacterium	0.71 (0.50, 1.03)	*0.07*	0.87 (0.72, 1.04)	*0.13*
Chitinophagaceae U.	0.76 (0.57, 1.03)	*0.08*	0.88 (0.71, 1.08)	*0.22*

Load of a bacterial genus (CE/m^2) was calculated by multiplying relative abundance with total bacterial load measured via qPCR in a sample; aOR Adjusted Odds Ratios; 95%CI 95% Confidence Intervals; aORs are expressed as IQR change in the estimate (ln-transformed relative abundance or CE/m^2); U denotes 'unassigned' and O denotes 'other' genus within family; Models are adjusted for follow-up time, living on a farm, cohort, gender, maternal history of allergic diseases, maternal smoking during pregnancy, and the number of older siblings; Significant results ($p<0.05$) are in boldface.

Journal Pre-proof

Table E10. Top 9 of the results (Q value <0.05) of MaAsLin analysis with bacterial/archaeal taxa and ever asthma.

Variable	**Feature**	**Value**	**Coefficient**	**N**	**N.not.0**	**P. value**	**Q. value**
asthma_ever	k_Bacteria_p_Firmicutes_c_Bacilli_o_Lactobacillales_f_Streptococcaceae_g_Lactococcus	1	0.08480	394	394	0.00001	0.00034
asthma_ever	k_Bacteria_p_Firmicutes_c_Bacilli_o_Lactobacillales_f_Streptococcaceae	1	0.07675	394	394	0.00015	0.00289
asthma_ever	k_Bacteria_p_Proteobacteria_c_Betaproteobacteria_o_Burkholderiales_f_Comamonadaceae_g_Roseateles	1	-0.00348	394	176	0.00046	0.00790
asthma_ever	k_Bacteria_p_Proteobacteria_c_Alphaproteobacteria_o_Rhizobiales_f_Hyphomicrobiaceae_g_Rhodoplanes	1	-0.00415	394	248	0.00090	0.01396
asthma_ever	k_Bacteria_p_Firmicutes	1	0.06472	394	394	0.00117	0.01772
asthma_ever	k_Bacteria_p_Actinobacteria_c_Actinobacteria_o_Actinomycetales_f_Microbacteriaceae_g_Other	1	-0.00603	394	382	0.00329	0.04267
asthma_ever	k_Bacteria_p_Chloroflexi_c_Chloroflexi	1	-0.00468	394	256	0.00339	0.04359
asthma_ever	k_Bacteria_p_Bacteroidetes_c_Rhodothermi_o_Rhodothermales_f_Rhodothermaceae_g_Rubricoccus	1	-0.00252	394	148	0.00352	0.04494
asthma_ever	k_Bacteria_p_Proteobacteria_c_Alphaproteobacteria_o_Rhizobiales_f_Bradyrhizobiaceae_g_Unassigned	1	-0.00568	394	389	0.00386	0.04866

Variable= ever asthma until the age of 10.5 years, Feature= the taxonomic hierarchical level of the given taxa; Value= asthma ever vs no asthma, Coefficient=coefficient of the asthma ever and given taxa in the linear model; N the total number of samples; N.not.0= the number of samples, which relative abundance is above detection limit; P.value= p-value for linear association; Q-value=multiple testing corrections with Benjamini and Hochberg (BH) procedure.

Journal Pre-proof

Journal Pre-proof

Online Repository materials

Indoor Bacterial Microbiota and the Development of Asthma by 10.5 years of age

Anne M. Karvonen, PhD, Pirkka V. Kirjavainen, PhD, Martin Täubel, PhD, Balamuralikrishna Jayaprakash, MSc, Rachel I. Adams, PhD, Joanne E. Sordillo, ScD, Diane R. Gold, MD, MPH, Anne Hyvärinen, PhD, Sami Remes, MD,MPH, Erika von Mutius, MD, and Juha Pekkanen, MD

Contents

2 Methods 2
3 Results 8

Journal Pre-proof

Journal Pre-proof

2 Methods

Follow-up

The first questionnaire was administered during the third trimester of pregnancy.[E1] The follow-up questionnaires (2, 12, 18 and 24 months old, and thereafter annually until the age of 6 years and then at the age of 10.5 years), except the questionnaire at two months, enquired about respiratory symptoms, doctor diagnosed asthma and asthmatic (obstructive) bronchitis, usage of medication and confounders for the time period after the preceding questionnaire and previous 12 months. 'Ever asthma' was defined as first parent-reported doctor-diagnosed asthma and/or second diagnoses of asthmatic (or obstructive) bronchitis. 'Current asthma' was defined as 'ever asthma' with usage of asthma medication and/or reported wheezing symptom in the past 12 months at 10.5-year follow-up. Among non-asthmatics, one missing follow-up was accepted and coded as no disease (n=25), if parents had answered all the remaining follow-up questionnaires and in the 6 and/or 10.5yrs' questionnaires they reported that their child had not had asthma and/or asthmatic bronchitis ever in life. Other children were excluded from the analyses. In total, 373 and 310 children had information on outcome, confounders and sequencing data and thus, were included into the ever asthma and current asthma models, respectively.

House dust sample

The protocol for dust collection of the age of 2 months has been explained in detail previously.[E2] Briefly, in the Finnish PASTURE study, fieldworkers took a sample from living room floors for the present report, while in the extended cohort, parents collected the floor dust. The samples were processed and analyzed at National Institute for Health and Welfare (THL), where they were sieved to remove larger particles, dried, split and then stored at -80ºC before analyses.

Journal Pre-proof

Journal Pre-proof

Sequencing

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. The extracted DNA was stored at -80^{0}C and shipped to the sequencing laboratory on dry ice. LGC Genomics (Germany) did the library preparation and sequencing. The V4 region of the 16S marker was amplified using 515F/806R primers.[E3] This primer set amplifies not only bacterial, but also archaeal taxa; however, the latter at limited phylogenetic coverage.[E4] Thus, we refer in the text mostly to 'bacterial', rather than to 'bacterial/archaeal' sequencing/microbiota, even if both kingdoms were at least partially covered. Amplicons were sequenced with 300 base pair paired-end reads using Illumina's MiSeq V3 chemistry.

Individual samples with a low read number after sequencing were re-sequenced (n=148). Sequences of two re-sequenced samples were used instead of the original sequences as the number of reads in the original sequencing efforts were below the rarefaction value of 2,150 sequences.

LGC Genomics performed preprocessing with the following steps. Demultiplexing relied on the bcl2fastq software provided by Illumina. Inline barcode sorting/clipping, adapter clipping, and primer detection/clipping steps relied on in-house scripts by LGC Genomics. Forward and reverse reads were merged using FLASH.[E5] After this preprocessing, the vast majority of reads were retained (mean 96.9%). All subsequent analysis relied on QIIME.[E6] Sequences were quality filtered using the split_libraries_fastq.py command with a minimum quality score of 20. Chimeric sequences were identified using usearch 61 algorithm[E7] and excluded. Sequences were clustered into Operational Taxonomic Units (OTUs) at 97% similarity using the open-reference OTU picking strategy[E8] with 100% subsampling of those sequences that failed to cluster to greengenes database.[E9] OTUs representing less than 0.001% of the total sequences (minimum count of 83

Journal Pre-proof

sequences) were excluded. Chloroplast (n=93) and mitochondrial (n=23) sequences were removed from the bacteria OTU table. Tree building used pynast to align sequences,[E10] and the resulting alignment was filtered to remove gaps prior to building a phylogeny using the fasttree method.

In total, 4,248 bacterial OTUs were identified at the 97% clustering approach, comprised of 658 different bacterial genera in 394 dust samples in bacterial data. Relative bacterial richness (a measure of the number of OTUs) in each sample (rarified to 2,150) and Shannon diversity (abundance and evenness of the taxa in each sample) indices were calculated within samples. Abundance weighted and unweighted UniFrac distance between samples, a measure of phylogenetic similarity and dissimilarity between samples based on their bacterial community composition, was calculated in QIIME.[E11] Beta diversity calculations based on UniFrac metrics were calculated between samples using beta_diversity.py script available in QIIME. Both alpha and beta diversity calculations were calculated at the same rarefaction value of 2,150.

Relative abundance and load of taxa

The average (SD) number of raw reads across all samples was 12,876 (mean ± SD = 12,876 ± 5,223). The measurements of bacteria were based on relative abundance i.e. the percentage of reads of a taxon per the total reads in a sample. The 139 bacterial genera with mean relative abundances of greater than 0.1%, were used in the analyses of asthma.

Total bacterial biomass was calculated from the sum of gram negative and gram positive bacteria using a previously developed qPCR assay[E12] and was expressed as total bacterial concentration in the samples (cell equivalents (CE) per mg of dust). The amount of total bacteria and taxa were also expressed as loads, i.e. the cell equivalents per square meter of sampling surface (CE/m^2). The 'genus load' was calculated by multiplying the relative abundance with total bacteria (CE/mg) in

Journal Pre-proof

Journal Pre-proof

that sample, as measured via qPCR and with the amounts of dust, and dividing by sampling area (m^2).

General microbial markers

Earlier analyzed bacterial biomass or cell-wall components[E2] (general microbial markers: endotoxin and $LPS_{10:0\text{-}16:0}$ for gram-negative bacteria; muramic acid for gram-positive bacteria) and amount of dust from the 2 months dust samples were used for testing whether the identified taxa were independent predictors for asthma adjusting models additionally with these markers.

Oligotyping for Lactococcus genus

Oligotyping analysis[E13] of *Lactococcus* genus (375,838 reads; 98.0% of all reads analyzed) was performed using 10 entropy positions (nucleotides in 68, 127, 208, 209, 213, 220, 241, 243, 245 positions). Total purity score of the analysis was 0.92. Most of the reads (355,480 reads) belonged to the GGCCAAGGA –oligotype. BLAST hits from the NCBI library was performed using accession numbers (December 1st, 2016). Relative abundances of individual oligotypes were calculated by multiplying the relative abundance of *Lactococcus* genus by the percentage of an oligotype divided by 100.

Journal Pre-proof

Wheezing phenotypes

Wheezing phenotypes were based on the repeated parent-reported any wheezing until the age of 6 years. Due the respective time period varied between questionnaires, each follow-up period was recoded to cover 12 months. Only those who had the maximum of one missing information on wheezing, were included into the analyses. Firstly, the wheezing phenotypes were estimated using latent class analyses (LCA)[E14] in the whole PASTURE cohort (N=953)[E15] where the first half of the study population belongs. Then, LCA was rerun with children from the Finnish extended cohort (N=185), which gave us similar wheezing patterns as seen by Depner et al.[E15] The correlations

Journal Pre-proof

between two LCAs were high. Five phenotypes were created: Never/ Infrequent wheeze, which was used as a reference category, Transient, Infrequent, Late onset and Persistent wheeze.

Statistical analyses

UniFrac, a measure of similarity and dissimilarity between samples based on their bacterial composition, was calculated based on the abundance ('weighted') of individual bacterial OTUs, including also their phylogenetic relatedness.[E11] For community data, Generalized UniFrac at an alpha of 1.0 based Principal Coordinate Analysis (PCoA), which takes into account phylogenetic relatedness of the biological community, was performed using QIIME and the first 6 axes scores were analyzed (Eigen values >1). The adjusted association of bacterial composition and ever asthma was studied using PERMANOVA-S.[E16] Spearman's correlations were calculated. Multinomial logistic regression was used for analyses of wheezing phenotypes.

In the multivariate models, the ln-transformed (natural logarithm+1) divided by interquartile range (IQR) variables of the exposure were used, except alpha-diversity metrics. In addition, variables were divided into three groups using the tertile as cut-offs.

Survival analyses (discrete-time hazard models) were used in analyzing ever asthma and current asthma. Generalized estimating equations (GEE) with an exchangeable correlation structure to account for correlation between repeated measures within subjects, were used to determine associations between different genera and repeated measures of parent-reported wheezing and cough at different ages. The results are presented as adjusted odds ratios (aORs) and their 95% confidence intervals (95%CI).

Journal Pre-proof

Journal Pre-proof

Multivariate association with linear models (MaAsLin)[E17] was run using all the most abundant taxa (mean relative abundance >0.1%) from phylum to genus level. Models were performed using defaults, which include for example arcsine square root transformation, Grubbs' test for outliers test, multiple testing corrections with Benjamini and Hochberg (BH) procedure and feature boosting in order simplify models. The boosting refers to the inclusion of confounders (all the variables that were used in the asthma models in the present study) that are prior tested and selected into the model with a given taxon, if they showed any potential association with the given microbial taxon.

All models were adjusted for time of follow-up, study cohort, living on a farm and well-known risk factors for asthma (maternal history of allergic diseases, gender, number of older siblings, and smoking during pregnancy). The models of bacterial richness, *Lactococcus* and ever asthma were carefully tested for 25 additional confounding factors (for example dogs or cats ownership and staying indoors; attendance to day care; breastfeeding; mode of delivery; birth weight; maternal age; paternal allergic disease; maternal or paternal educational level; house characteristics e.g. mold or moisture damage and ventilation system; and season of dust sampling), which have been described earlier.[E1] None of these potential confounders changed the estimates of exposure by >10%, and thus were not included in the analyses. Due to skewed distributions and outliers, the highest 5% of relative abundances of 13 genera were excluded in the sensitivity analyses: the results were mostly the same or the estimates were even stronger (data not shown). The data were analyzed using SAS 9.3 for Windows. Plots were created by using ggplot2 in RStudio and box-plots and diagrams with Excel.

Journal Pre-proof

Journal Pre-proof

3 Results

At the phylum level, *Proteobacteria* (36%) and *Firmicutes* (34%) had the highest relative abundances of the 30 phyla, which were detected in the floor dust, followed by *Actinobacteria* (18%) and *Bacteroidetes* (9%) (Figure E1).

Correlations between relative abundance of genera and diversity indices

Relative abundance of *Lactococcus* correlated negatively with the relative abundances of 12 genera (mostly $r= -0.25$) (Figure E2). Diversity indices were highly intercorrelated ($r=0.93$), but were less correlated with the amounts of dust (mg/m^2) or total bacterial qPCR. Correlation coefficients between the sum abundance of the 12 protective genera and richness, Shannon diversity index, amounts of dust, total bacterial qPCR and endotoxin were 0.69, 0.58, 0.27, 0.08 and 0.27, respectively (Figure E2).

Determinants for Lactococcus and the sum abundance of the 12 protective genera

Houses that were newer (build in 1970 or after) and had mechanical exhaust ventilation (from kitchen/bathroom) had higher levels of *Lactococcus* (Table E5). In contrast, lower levels were observed in farm homes and homes that had a basement, or homes where dog(s) or cat(s) were kept mostly indoors. The determinants for higher levels of the sum abundance of the 12 protective genera were foundation type, heating with wood, dog ownership and frequent contact with horses during first trimester of pregnancy. On the other hand, newer houses (build 1970 or after), construction material and other ventilation system in the house than natural ventilation decreased the levels of the sum abundance of the 12 genera (Table E10).

Journal Pre-proof

Journal Pre-proof

Sensitivity analyses

At the age of 3 years, majority (80.2%) of the children still lived in the same house where the dust sample was collected at the age of 2 months. When children (n=74, 19.8%), who had moved at least once before the age of 3 years and children who still lived in the same house, analyzed separately, the main results (asthma ever vs. *Lactococcus*, the sum abundance of the 12 protective genera and diversity indices) did not change.

Journal Pre-proof

Journal Pre-proof

References

E1. Karvonen AM, Hyvärinen A, Roponen M, Hoffmann M, Korppi M, Remes S, et al. Confirmed moisture damage at home, respiratory symptoms and atopy in early life: a birth-cohort study. Pediatrics. 2009; 124(2): e329-38.

E2. Karvonen AM, Hyvärinen A, Rintala H, Korppi M, Täubel M, Doekes G, et al. Quantity and diversity of environmental microbial exposure and development of asthma: a birth cohort study. Allergy 2014; 69(8): 1092-101.

E3. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012; 6(8): 1621-4.

E4. Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. Examining the global distribution of dominant archaeal populations in soil. ISME J. 2011; 5(5): 908-17.

E5. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011; 27(21): 2957-63.

E6. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010; 7(5): 335-56. E7. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010; 26(19): 2460-1.

E8. Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, et al. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. PeerJ. 2014; 2: e545.

Journal Pre-proof

Journal Pre-proof

E9. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006; 72(7): 5069-72.

E10. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010; 26(2): 266-7.

E11. Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol. 2005; 71(12): 8228-35E12. Kärkkäinen PM, Valkonen M, Hyvärinen A, Nevalainen A, Rintala H. Determination of bacterial load in house dust using qPCR, chemical markers and culture. J Environ Monit. 2010; 12(3): 759-68. E13. Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, et al.. Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol. 2013; 4(12): 1111-9.

E14. Rabe-Hesketh S, Skrondal A. Classical latent variable models for medical research. Stat Methods Med Res. 2008; 17(1): 5-32. E15. Depner M, Fuchs O, Genuneit J, Karvonen AM, Hyvärinen A, Kaulek V, et al. Clinical and epidemiologic phenotypes of childhood asthma. Am J Respir Crit Care Med. 2014; 189(2): 129-38.

E16. Tang ZZ, Chen G, Alekseyenko AV. PERMANOVA-S: association test for microbial community composition that accommodates confounders and multiple distances. Bioinformatics. 2016; 32(17): 2618-25.

E17. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV, et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012; 13(9): R79.

Journal Pre-proof