Asthma and wheeze severity and the oropharyngeal microbiota in children and adolescents

Thorsen, Jonathan[1,2] (0000-0003-0200-0461); Stokholm, Jakob[1,3] (0000-0003-4989-9769); Rasmussen, Morten Arendt[1,3] (0000-0001-7431-5206); Roggenbuck-Wedemeyer, Michael[4,5]; Vissing, Nadja H[1] (0000-0003-0662-2727); Mortensen, Martin S[4] (0000-0001-5483-7533); Brejnrod, Asker D[4] (0000-0002-1327-2051); Fleming, Louise[6,7] (0000-0002-7268-7433); Bush, Andrew[6,7] (0000-0001-6756-9822); Roberts, Graham[8,9,10] (0000-0003-2252-1248); Singer, Florian[11] (0000-0003-3471-5664); Frey, Urs[12] (0000-0003-3773-2822); Hedlin, Gunilla[13]; Nordlund, Björn[13] (0000-0001-9888-1659); Murray, Clare S[14] (0000-0002-8961-8055); Abdel-Aziz, Mahmoud I.[15] (0000-0001-9263-7137); Hashimoto, Simone[15,16] (0000-0001-8995-3817); van Aalderen, Wim[16]; Maitland-van der Zee, Anke H[15,16] (0000-0002-0414-3442); Shaw, Dominick[17] (0000-0003-4106-8469); Fowler, Stephen J[14] (0000-0002-4524-1663); Sousa, Ana[18]; Sterk, Peter J[15]; Chung, Kian Fan[6,7] (0000-0001-7101-1426); Adcock, Ian M[6,19] (0000-0003-2101-8843); Djukanovic, Ratko[8,10] (0000-0001-6039-5612); Auffray, Charles[20] (0000-0003-2226-7411); Bansal, Aruna T[21] (0000-0002-7262-6534); Wagers, Scott[22]; Chawes, Bo[1] (0000-0001-6846-6243); Bønnelykke, Klaus[1] (0000-0003-2003-1018); Sørensen, Søren Johannes[4] (0000-0001-6227-9906); Bisgaard, Hans[1] (0000-0003-4131-7592)

Affiliations

1) COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Copenhagen, Denmark
2) Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
3) Department of Food Science, University of Copenhagen, Frederiksberg, Denmark
4) Section of Microbiology, Department of Biology, University of Copenhagen, Denmark
5) Novozymes, Bagsvaerd, Denmark

6) National Heart and Lung Institute, Imperial College, London, UK
7) Dept of Respiratory Paediatrics, Royal Brompton Hospital, London, UK
8) Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, UK
9) Human Development and Health, Faculty of Medicine, University of Southampton, UK
10) NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK
11) Pediatric Respiratory Medicine, Department of Pediatrics, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
12) University Children's Hospital (UKBB), University of Basel, Basel, Switzerland
13) Karolinska Institutet, Department of Women's and Children's Health and Astrid Lindgren Children's Hospital, Stockholm, Sweden
14) Division of Infection, Immunity and Respiratory Medicine, School of Biological Sciences, University of Manchester, Manchester Academic Health Science Centre and NIHR Manchester Biomedical Research Centre, Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom.
15) Department of Respiratory Medicine, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands
16) Department of Paediatric Respiratory Medicine and Allergy, Emma Children's Hospital, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands
17) Division of Respiratory Medicine, University of Nottingham, Nottingham, UK
18) Respiratory Therapeutic Unit, GlaxoSmithKline, Stockley Park, London, UK
19) NIHR Imperial Biomedical Research Centre, London, UK
20) European Institute for Systems Biology and Medicine, Vourles, France
21) Acclarogen Ltd, Cambridge, UK
22) BioSci Consulting, Maasmechelen, Belgium

Correspondence:

Professor Hans Bisgaard, MD, DMSc
E-mail: bisgaard@copsac.com
Website: www.copsac.com

Running title: Severe asthma/wheeze and oropharyngeal microbiota

Article type: Original article

Word count: 4157

Tables: 2

Figures: 5

Online Repository: Yes

Authors Contributions:

JT has written the first draft of the manuscript, performed the statistical analysis and prepared figures and tables. MR performed the sequencing. All co-authors have contributed substantially to the analyses and interpretation of the data, and have provided important intellectual input. All authors have agreed that the accuracy and integrity of any part of the work has been appropriately investigated and resolved and all have approved the final version of the manuscript. The corresponding author had full access to the data and had final responsibility for the decision to submit for publication. No honorarium, grant, or other form of payment was given to anyone to produce the manuscript.

Source of Funding: The research leading to these results has received support from the Innovative Medicines Initiative (IMI) Joint Undertaking, under grant agreement no. 115010, resources for which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and kind contributions from companies in the European Federation of Pharmaceutical Industries and Associations (EFPIA) (www.imi.europa.eu).
COPSAC is funded by private and public research funds all listed on www.copsac.com.

Conflict of interest:

G Roberts's institution received funding from EU IMI programme for undertaking this research. LF, SF & AB's institution received funding for undertaking this research. Dr. Fleming has received speakers fees or fees for expert consultation from Teva, Astra Zeneca, Sanofi, Respiri, Novartis; all fees paid direct to her institution and outside the submitted work. Dr Chung has received honoraria for participating in Advisory Board meetings of GSK, AZ, Roche, Novartis, Merck, BI, TEVA and Shionogi regarding treatments for asthma, chronic obstructive pulmonary disease and chronic cough and has also been renumerated for speaking engagements. AHM has received research grants outside the submitted work from GSK, Boehringer Íngelheim and Vertex, she is the PI of a P4O2 (Precision Medicine for

more Oxygen) public private partnership sponsored by Health Holland involving many private partners that contribute in cash and/or in kind (Boehringer Ingelheim, Breathomix, Fluidda, Ortec Logiqcare, Philips, Quantib-U, Roche, Smartfish, SODAQ, Thirona, TopMD and Novartis), and she has served in advisory boards for AstraZeneca, GSK and Boehringer Ingelheim with money paid to her institution. Dr. Singer reports personal fees from Vertex, personal fees from Novartis, outside the submitted work. Dr. Wagers reports other from Kings College Hospital NHS Foundation Trust, from Academic Medical Research, from AMC Medical Research BV, from Asthma UK, from Athens Medical School, from Boehringer Ingelheim International GmbH, from CHU de Toulouse, from CIRO, from DS Biologicals Ltd, from ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, from European Respiratory Society, from FISEVI, from Fluidic Analytics Ltd., from Fraunhofer IGB, from Fraunhofer ITEM, from GlaxoSmithKline Research & Dev Ltd, from Holland & Knight, from Karolinska Institutet Fakturor, from KU Leuven, from Longfonds, from National Heart & Lung Institute, from Novartis Pharma AG, from Owlstone Medical Limited, from PExA AB, from UCB Biopharma S.P.R.L., from Umeå University, from Univ. Hosptial Southampton NHS Foundation Trust, from Università Campus Bio-Medico di Roma, from Universita Cattolica Del Sacro Cuore, from Universität Ulm, from University of Bern, from University of Edinburgh, from University of Hull, from University of Leicester, from University of Loughborough, from University of Manchester, from University of Notthingham, from Vlaams Brabant, Dienst Europa, from Imperial College London, from Boehringer Ingelheim, from Breathomix, from Gossamer Bio, from Astrazeneca, from CIBER, from OncoRadiomics, from University of Leiden, from University of Wurzburg, from Chiesi Pharmaceutical, from University of Liege, from Teva Pharmacauticals, from Sanofi, from Pulmonary Fibrosis Foundation , from Three Lakes Foundation, outside the submitted work.
Besides the above mentioned, authors declare no potential, perceived, or real conflict of interest regarding the content of this manuscript.

Governance:
We are aware of and comply with recognized codes of good research practice, including the Danish Code of Conduct for Research Integrity. We comply with national and international rules on the safety and rights of patients and healthy subjects, including Good Clinical Practice (GCP) as defined in the EU's Directive on Good Clinical Practice, the International Conference on Harmonisation's (ICH) good clinical practice guidelines and the Helsinki

Declaration. We follow national and international rules on the processing of personal data, including the Danish Act on Processing of Personal Data and the practice of the Danish Data Inspectorate.

Key words: Asthma; Severe asthma; paediatrics; microbiome; microbiota; airway; oropharyngeal

Abbreviations:
16S rRNA = 16S ribosomal ribonucleic acid
ASV = Amplicon sequence variant
CV = Cross-validation
FDR = False Discovery Rate
FeNO = Fraction of exhaled nitric oxide
GTDB = Genome Taxonomy Database
ICS = Inhaled corticosteroid
LABA = Long-acting beta-2 agonist
U-BIOPRED = Unbiased BIOmarkers in PREDiction of respiratory disease outcomes
PERMANOVA = Permutational multivariate analysis of variance

Abstract

Background

There is a major unmet need for improving the care of children and adolescents with severe asthma and wheeze. Identification of factors contributing to disease severity may lead to improved diagnostics, biomarkers, or therapies. The airway microbiota may be such a key factor.

Objective

To compare the oropharyngeal airway microbiota of children and adolescents with severe and mild/moderate asthma/wheeze.

Methods

Oropharyngeal swab samples from school-age and pre-school children in the European U-BIOPRED multicenter study of severe asthma, all receiving severity-appropriate treatment, were examined using 16S rRNA gene sequencing. Bacterial taxa were defined as Amplicon Sequence Variants (ASVs).

Results

We analysed 241 samples from four cohorts; A) 86 school-age children with severe asthma, B) 39 school-age children with mild/moderate asthma, C) 65 pre-school children with severe wheeze and D) 51 pre-school children with mild/moderate wheeze. The most common bacteria were *Streptococcus* (mean relative abundance 33.5%), *Veillonella* (10.3%), *Haemophilus* (7.0%), *Prevotella* (5.9%) and *Rothia* (5.5%). Age group (school-age versus pre-school) was associated with the microbiota in beta-diversity analysis (F=3.32, p=0.011) and in a differential abundance analysis (28 significant ASVs). Among all children, we found no significant difference in the microbiota between children with severe and mild/moderate asthma/wheeze in a univariable beta-diversity analysis (F=1.99, p=0.08, n=241), but a significant difference in a multivariable model (F=2.66, p=0.035), including number of exacerbations in the previous year. Age was also significant when expressed as a Microbial Maturity Score (Spearman Rho 0.39, p=4.6e-10), however this score was not associated with asthma/wheeze severity.

Conclusion

There was a modest difference in the oropharyngeal airway microbiota between children with severe and mild/moderate asthma/wheeze across all children but not in individual age groups, and a strong association between the microbiota and age. This suggests the oropharyngeal

airway microbiota as an interesting entity in studying asthma severity, but probably without the strength to serve as a biomarker for targeted intervention.

Introduction

Asthma is one of the most common chronic diseases of childhood[1,2]. It originates from the interactions between multiple genes[3,4] and environmental factors[5–8], several of which are hypothesized to exert their effects in complex gene-environment interactions[9]. Many of these effects are potentially related to microbial exposures in early life[6,8,10]. This has led to hypotheses implicating our microbial environment as a key driver for the remarkable increase in incident asthma and related immune-mediated diseases over the past half-century[11–14].

The airways are now known to harbor a diverse array of microbes[15,16] However, the microbial impact on health and disease remains to be fully elucidated. Several studies have demonstrated striking differences in the airway microbiota of adult individuals with and without asthma[17–19]. In children, studies have shown that changes in the airway microbiota in infancy may precede asthma development[10,20,21], that children diagnosed with wheeze or asthma may have differences in their current airway microbiota compared with healthy children[22–25], and that changes in patterns of airway microbiota of children with asthma may precede loss of asthma control and exacerbations[26]. Despite variation between anatomical regions, studies have found an ecological continuity with high intra-subject consistency in the bacterial communities residing in the various parts of the respiratory tract, including the nose, mouth, pharynx and lower airways[15,27–31]. Furthermore, the airway microbiota develops in a highly dynamic and region-dependent manner over the course of infancy and childhood[32–35].

Although asthma is seen in patients of all ages, its prevalence peaks in childhood where it can manifest as mild recurring symptoms such as cough and wheeze, chronic symptoms requiring daily treatment, or even severe persistent symptoms and frequent exacerbations despite maximal treatment[36]. Treatment strategies for children with severe asthma and wheeze are still inadequate[37], which has prompted the search for new biomarkers and potential treatment avenues. The airway microbiota is intriguing in this context, since it has been shown to influence episodes of asthma-like symptoms in young children[38–40].

Most of the molecular mechanisms underpinning the differences in severity of asthma are unknown. The U-BIOPRED (Unbiased BIOmarkers in the PREDiction of respiratory disease outcomes) consortium is a multicenter public-private partnership study funded by the Innovative Medicines Initiative (IMI), designed around state-of-the-art comprehensive

profiling across many 'omics fields. The consortium is using systems biology and data driven discovery methods to enable characterization of multi-compartment physiological handprints to discover mechanistic differences between severe and mild/moderate asthma and wheeze[41,42].

In this study, our aim was to identify features of the oropharyngeal airway microbiota associated with asthma/wheeze severity. We analyzed oropharyngeal throat swab samples from the U-BIOPRED cohorts of children and adolescents with severe or mild/moderate asthma/wheeze, all receiving severity-appropriate treatment, using 16S rRNA gene amplicon sequencing. We hypothesized that the oropharyngeal airway microbiota differs between children and adolescents with severe and mild/moderate asthma/wheeze and that this might be age dependent.

Materials and methods

Study cohorts

All participants from the U-BIOPRED paediatric cohorts were included in the present study, consisting of A) 97 school-age (6-17 years) children with severe asthma, B) 43 school-age children with mild/moderate asthma, C) 77 pre-school children (1-5 years) with severe wheeze and D) 54 pre-school children with mild/moderate wheeze. For the sake of brevity, we refer to these cohort participants as children throughout the manuscript despite also including adolescents. Children were recruited to these four cohorts using strict pre-defined inclusion and exclusion criteria, delineating phenotypes based on factors including age, diagnostic criteria for asthma, current treatment, and asthma control. See table S1 for a full list of criteria, which have also been published previously[37]. Briefly, the cohorts were characterized by the following: Cohort A – ongoing poorly controlled asthma (persistent symptoms, frequent exacerbations, or persistent airflow limitation) despite high-dose inhaled corticosteroids (ICS) and at least two other controller medications. Cohort B – controlled or partly controlled asthma, prescribed low-dose ICS and no other or one additional controller medication. Cohort C – persistent symptoms and frequent exacerbations despite current or failed high-dose ICS and a leukotriene receptor antagonist (LTRA). Cohort D – controlled or partially controlled symptoms prescribed no treatment or low-dose ICS and/or a LTRA, as previously described[37].

The study was registered on ClinicalTrials.gov (NCT01982162). The study was approved by local ethics committees and parents or caregivers gave written informed consent; children gave assent where appropriate.

Clinical examinations

Children were examined at planned visits to one of the 7 participating study centers in five European countries. At the time of visit, they had at least 4 weeks under stable medication and no clinical asthma exacerbation. Exacerbations were defined as the prescription of high dose OCS for at least three days or a doubling of OCS dose for participants on maintenance OCS following consultation with a medical practitioner or hospitalization or attendance in the Emergency Room due to asthma symptoms[43]. Clinical measurements of anthropometrics, skin prick tests of common aeroallergens (positive defined as ≥3mm) and lung function tests (spirometry with reversibility tests and whole-body plethysmography) were performed at the clinical visits, provided the children were of sufficient age to cooperate. In cohort A and B, fraction of exhaled nitric oxide (FeNO) was measured. Blood samples were analysed for specific IgE for common aeroallergens (positive defined as ≥ 0.35 kU/L). Information concerning home environment, comorbidities and use of medication, including antibiotics, inhaled corticosteroids (ICS), ICS/LABA (long-acting beta2 agonist) combinations, nasal corticosteroids, and oral corticosteroids was obtained by personal interview. Treatment was quantified as recent treatment if more frequent than once per month on average. Urine was analysed for cotinine. All procedures used detailed Standard Operating Procedures (SOPs) to ensure uniform practices across study centers. Full methods were previously published[37].

Microbiota sequencing

Throat swabs were collected by research assistants using dry flocked swabs (ESwab, BD, NJ, USA), placed in a Liquid Amies transport medium and stored at -20°C. Genomic DNA was extracted using the FastDNA™ SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA, USA). DNA extracts were checked for consistency and quality by gel electrophoresis. Then, the 16S rRNA gene was amplified using a two-step PCR process targeting the hypervariable V4 region and normalised for amplicon contents before undergoing sequencing using the Illumina MiSeq Desktop Sequencer (Illumina Inc., CA, USA) as previously described[33]. Paired-end reads (250 bp x 2) were pooled for subsequent processing. The reads were quality checked using FastQC[44] and MultiQC[45], and then cleaned by removing primers (515F-Y: 5'-

GTGCCAGCMGCCGCGGTAA and 806R: 5'-GGACTACHVGGGTWTCTAAT) using Cutadapt[46]. For designating Amplicon Sequence Variants (ASVs) we followed the DADA2 pipeline[47,48]. Briefly, the pipeline works by performing quality filtering and trimming, de-replicating sequences, learning dataset-specific error rates, de-noising by removing potentially containing errors sequences, merging paired-end reads while removing mismatches to reduce errors, constructing ASVs, removing chimera by implementing "bimera" method, and running taxonomic classification of ASVs using different public databases. The Silva database version 132[49] and the Genome Taxonomy Database (GTDB)[50] were used for taxonomic annotations. The Silva database was used for down-stream analysis, while GTDB was used to identify ASVs taxonomic annotation that could not be detected by the Silva database. Positive (mock community) and negative (sterile water) controls were added to each plate as previously described[33], and checked against real samples using relative abundance plots.

Statistical analysis

Data analysis was performed with the statistical software R v. 3.6.1 (R core team 2016), using the add-on packages {phyloseq}[51] for microbiota data handling and {ggplot2}[52] for visualizations. Alpha diversity was measured as Shannon diversity index and associations with clinical covariates were analyzed with linear models, which were adjusted for log(library size). Furthermore, library size, extraction plate ID (2 plates) and sequencing run (2 runs) were analyzed as potential technical covariates, but these were omitted from the alpha diversity analysis as they were not significant. Beta-diversity was assessed using weighted UniFrac[53] on log-transformed abundances adding a pseudocount of 1 and tested with the permutational multivariate analysis of variance (adonis2 PERMANOVA)[54], using a modified F statistic (between-group versus. within-group variance) as a measure of effect size and considering marginal effects of factors. Covariates were analyzed as single predictors in univariable models and all combined in a backward selection multivariable PERMANOVA model where terms were successively removed by highest p-value until all p-values were below 0.1. All beta-diversity models were adjusted for library size, log(library size), extraction plate ID and sequencing run, which were all statistically significant in most models. Differential abundance was tested using fitFeatureModel from the package {metagenomeSeq}[55]. In cases where the feature model failed (fewer than 2 positive samples in one of the groups), a Wilcoxon test was used instead. Only ASVs present in at least 10% of

the samples and with a mean relative abundance > 10^{-4} were analyzed. P-values were adjusted for false discovery rate (FDR) using the Benjamini-Hochberg approach and reported as q-values[56]. A raw p-value below 0.05 was considered nominally significant. We computed Microbial Maturity Scores by fitting a random forest regression model of age in years as a continuous outcome and using the relative abundance of all ASVs with at least 20% presence as predictors, resulting in 68 ASVs, using the default value of mtry ($n_{variables}$ / 3 ~ 23). This model was cross-validated using repeated 10-fold CV with 5 repeats, keeping the repeat with median performance (assessed by spearman correlations) as the final model. The score was then defined as the cross-validated predictions from that repeat. The Microbial Maturity Score was normalized for age by regressing out the actual age in a linear model, taking the residuals and z-scaling to produce the Microbiota-for-age Z-scores. Contribution of ASVs to maturity was quantified by the Variable Importance Score, extracted directly from the random forest model, and by correlating each individual ASV with the Microbial Maturity Score using Spearman's Rho correlation.

Results

Study participants

From the 271 children included in the 4 study cohorts, 241 throat swab samples were successfully sampled and sequenced to a depth of at least 1000 reads. Only one sample was taken per child. Important baseline characteristics of the four cohorts are summarized in **table 1** and study centers are visualized in **fig S1**. A full description of the participant characteristics can be found elsewhere[37].

Microbial composition

A median of 11,975 reads per sample [IQR 7,613-21,098] passed quality control. The dataset contained 4,491,120 quality filtered reads in 1177 Amplicon Sequence Variants (ASVs). The main microbial constituents belonged to genera *Streptococcus* (mean relative abundance 33.5%), *Veillonella* (10.3%), *Haemophilus* (7.0%), *Prevotella* (5.9%) and *Rothia* (5.5%), see **fig 1**. These rates varied considerably within each cohort, see **fig S2**. Notably, *Moraxella* was detected in low amounts (5.0% of samples, 0.01% mean relative abundance).

Microbiota, asthma/wheeze severity and age

We examined the differences between cohorts by Shannon diversity index (alpha diversity, within-sample), beta diversity (between-sample, compositional differences) and differential abundance analysis (considers the association with each ASV on its own).

When analyzing Shannon diversity index, we found no differences between the four cohorts, see **table S2**.

We found no differences between children with severe and mild/moderate asthma/wheeze within each age group (PERMANOVA; school-age, F=0.64, p=0.70, n=125; pre-school, F=1.08, p=0.33, n=116), or in the full cohort (F=1.99, p=0.08, n=241). We found a nominally significant difference in microbiota composition between school-age and pre-school children (F = 3.32, p = 0.011, n=241) as a categorical predictor. Analyzing age as a continuous predictor showed similar results (F=2.95, p=0.025), though we found no association with age as a continuous predictor within each age group. Mutually adjusting for severity and age group did not materially change the results (severity F=1.69 p=0.126; age group F=3.02, p=0.022). These differences are visualized in **Fig 2** and shown in **Table 2**.

We then examined if any ASVs were significantly associated with severity, in a differential abundance analysis. While nine ASVs were individually significant, none remained so after correction for multiple testing, see **Table S3**. Similarly, no ASVs remained significantly associated with severity within each age group.

We also examined the difference between age groups in a differential abundance analysis, see **Fig 3** and **Table S4**. Here, we found 28 significant ASVs ($p < 0.05$), eight of which were significant after FDR correction ($q < 0.05$), and an additional six with q-values between 0.05 and 0.10. These describe the age-group-related differences observed in the global beta-diversity analysis. The ASVs most strongly associated with the school-age group were *Megasphera micronuciformis*, *Actinomyces graevenitzii*, *Veillonella* and *Prevotella*, while the ASVs most strongly associated with the pre-school group were *Alloprevotella, Porphyromonas, Neisseria* and an unknown bacterial taxon (ASV150) which could not even be classified at phylum level using our default annotation database (Silva), while another (GTDB) classified it as a *Streptococcus.*

Phenotypic and environmental covariates and microbial diversity

Furthermore, we examined associations between the microbiota and phenotypic and environmental factors to identify potential covariates of high importance that may confound or cloud the main association with severity, or with age group.

When considering Shannon diversity index, we found that a concurrent diagnosis of eczema was associated with increased diversity, but only in school-age children, and that the number of exacerbations in the previous year was associated with decreased diversity, but only in pre-school children. None of these results were significant after FDR adjustment.

In the beta-diversity analysis, we found several factors which were significantly associated with microbiota composition, see **table 2**. Besides age group and age, described above, we found that study center, season of visit, siblings and having positive skin prick test or any sensitization were significantly associated with microbiota composition either overall or in one age group. None of these remained significant after FDR adjustment.

Multivariable beta-diversity model

Since some study design variables had strong associations, we finally performed a multivariable backward selection model in all children and each age group, see the lower part of **table 2**, to adjust for variables that may confound or cloud the association with our main variables of interest, severity and age group. We selected asthma/wheeze severity, study center and age group as compulsory variables. Many variables had stronger associations in the multivariable models. Of note, asthma/wheeze severity became significant among all children ($F=2.66$, $p=0.035$) when adjusting for the other covariates, but not in either age group. The number of exacerbations in the previous year, which was also associated with severity, became much stronger in the multivariable model ($F=4.98$, $p=0.001$). When omitting number of exacerbations from the multivariable model, severity was no longer significant ($F=1.93$, $p=0.088$), whereas omitting other factors in the model did not change the results.

Microbial maturation

Many studies in paediatric microbiota research now employ machine learning to show clinical impact of the degree of maturation, computing derived metrics such as Microbiota-for-age Z-scores[57,34,58,35]. We found that a Microbial Maturity Score, derived from a random forest model of ASV abundances, was strongly associated with actual age (Spearman

correlation 0.39, p=4.6e-10), even when employing cross-validation to prevent overfitting, see **fig 4A**. This score was subsequently normalized for actual age into a Microbiota-for-age Z-score to derive the relative level of maturation for each child. The ASVs contributing to the Microbial Maturity Score predictably echoed the results from the age-group differential abundance score; among the most important ASVs were *Alloprevotella*, *Veillonella* and several *Streptococcus* ASVs, see **fig S3**. We found that among school-age children, a trend for a slightly higher Microbiota-for-age Z-score was found among children with severe compared to mild/moderate asthma (mean difference 0.27, 95% confidence interval -0.67 to 0.14, p=0.20), but this association was not statistically significant. Similarly, we found no significant difference in Microbiota-by-age Z-scores between severe and mild/moderate pre-school wheeze (mean difference -0.09, -0.25 to 0.43, p=0.61), see **fig 4C**. These associations did not materially change after adjusting for study center, visit season and technical variables.

Discussion

Primary findings

In a large European multi-center study of children with severe and mild/moderate asthma/wheeze between the ages of 1 and 17 years, we found that the airway microbiota, sampled using oropharyngeal swabs, was similar overall between children with severe and mild/moderate asthma/wheeze. In crude analyses, we found no association between microbiota composition and severity, but after adjusting for the number of exacerbations and other covariates in a backward selection model, a significant difference was detected. The microbiota was strongly associated with age. Using a robust cross-validated random forest setup, we were able to express much of the microbiota development over time as a Microbial Maturity Score and identify key taxa contributing to this maturation. However, this score was not associated with asthma/wheeze severity.

Strengths and limitations

The major strength of the study is that the participating children were recruited based on strict predefined criteria of symptom burden and need for pharmacological therapy, after at least 6 months of treatment optimization and exclusion of differential diagnoses by a paediatrician. This ensures a uniform and comparable study populations as far as is possible when recruiting from multiple countries and age groups. Furthermore, all children underwent

detailed characterization. Other strengths of the study include a reasonable sample size and usage of Standard Operating Procedures (SOPs) to minimize sampling variation.

A limitation to the study is the lack of a healthy control group, which may have precluded identification of associations between the microbiota and ongoing asthma/wheeze, which are not specific to the severe phenotype. Another limitation is the oropharyngeal sampling site, as this may only partially reflect lower airway microbiota which is presumed to be of high importance for asthma. Despite the systematic variation seen across different parts of the respiratory tract[59], there is a high exchange of microbes throughout it within an individual, making the upper respiratory tract microbiota an interesting and readily available compartment to study in lower airway diseases[15,27,29,60]. Of note, the oral microbiota has been reported to more closely resemble the lower respiratory tract than the nasal microbiota[28]. However, this intra-individual inter-compartment microbial exchange may differ between healthy and disease states, which may limit the relevance of the upper airway microbiota in studying lower airway diseases. In addition, it is a limitation that we do not have longitudinally collected samples, particularly given our findings of a strong association between the microbiota and age.

An important part of the design of the cohort is that most children were on regular medication which varied according to severity. This is a limitation of the design of the study in that we are unable to discern differences in severity from differences in treatment. In addition, it is a limitation that we do not have objective measurements of medication adherence and deposition which may vary significantly between children, and that we did not have access to samples taken before vs after initiation of treatment to estimate its impact on the airway microbiota. Similarly, there may be residual confounding by factors not included or inadequately captured in our study, including previous antibiotic use and environmental conditions.

The 16S rRNA gene amplicon sequencing technique applied is a major leap forward in microbial surveys compared to traditional culture and early molecular techniques. Whereas culturing is very specific in identifying species and strains, 16S sequencing has the advantage that nearly all bacteria present, whether living or dead, will be detected. Still more sophisticated approaches, such as metagenomic sequencing, offer even more detailed views into the composition and metabolic potential of the microbiome, but at substantially higher costs, making 16S rRNA gene amplicon sequencing a very attractive option for large-scale microbial surveys like the present study.

The strength of a diverse population of children with asthma and wheeze of varying degrees of severity from different countries is equally challenged by geographic and ethnic differences resulting in increased variation in the data. We found that geographical location strongly influenced beta-diversity, especially in pre-school children. There are multiple potential differences between treatment centers and the countries which could all contribute to this, including differences in genetics, treatment policies, environment, diet, and technical factors such as sampling and handling despite the usage of SOPs.

As clearly demonstrated by our results, children are different across age groups, representing a continuum of evolving states, both in terms of clinical phenotypes and indeed also their microbiotas. Our finding that age group was the largest source of systematic variation in the composition of the airway microbiota highlights the importance of appropriate stratification in recruitment and analysis of paediatric microbiome studies.

Interpretation

In our study, we found only a modest difference between children with severe and mild/moderate asthma/wheeze. This was dependent on controlling for number of exacerbations in the previous year, which was itself very different between severe and mild/moderate phenotypes. This could suggest that this feature of the severe phenotype has an independent effect on the airway microbiome, and that other latent features of the severe phenotype could each have distinct effects as well, which is consistent with reports of differential influence of phenotypic factors among adults with severe asthma[61]. Interestingly, while number of exacerbations had a strong association with the microbiota among all children, it was not significant within either age group. We did not find any individual-level features (single ASVs, diversity or other readouts) which would enable us to distinguish severity groups from each other and be used as a biomarker for such a discrimination or be considered as a potential novel drug target mechanistically contributing to disease severity.

These observed associations could be due to differences between phenotypes or differences in both maintenance treatment and exacerbation treatments, which complicates the interpretation in an observational study like this. The association with number of exacerbations might be due to antibiotic treatment, which is especially common in

preschoolers – an exposure which may cumulatively influence the airway microbiota. An example of this type of phenotype-treatment-microbiota relation has been shown for metformin treatment of type 2 diabetes[62]. Furthermore, severity classifications may differ between age groups, where phenotypic factors may differentially contribute to classification. Of note, number of exacerbations associated with severity but not age group in our study, whereas allergic sensitization associated with age group but not severity. Despite mutually adjusted models, disentangling these closely related clinical factors remains challenging and residual mutual confounding cannot be ruled out. This is important to consider when interpreting the association with severity which was only significant after adjusting for number of exacerbations.

In adults, the airway microbiota has been associated with asthma severity[19], specific inflammatory subtypes[63,64] and ICS responsiveness[65]. In children, the infant airway microbiota has been associated with asthma risk and severity of respiratory infections[10,20,66,67] as well as the efficacy of azithromycin treatment in recurrent episodes of asthma-like symptoms[40]. Studies in older children have shown that the microbiome is associated with having asthma[24] and exacerbation risk[26]. However, there is a paucity of such studies focusing on the clinical phenotype of severe asthma in children. Our findings indicate that the airway microbiome is not a strong determinant of asthma severity among both pre-school and school-age children, and thereby not an obvious target for interventions in these age groups for gaining better asthma control, but also that further study is warranted to clarify this relationship and shed light on the severe asthma phenotype in children.

We found a large impact of age and age group on the oropharyngeal airway microbiota composition and several significant taxa, but no significant changes in alpha diversity. This age effect is mirrored by previous reports, with most of the available studies focusing on the gut microbiota. Studies have suggested that an adult-like gut microbiota composition is established when a child reaches 2 years of age, but subtle changes may continue to develop[68–72]. The airway microbiota changes extensively from birth[33]; Teo et al[67] described the development of the nasopharyngeal microbiota during the first year of life, dominated early by *Staphylococcus, Corynebacterium, Alloiococcus, Moraxella* and *Streptococcus*, of which only the latter was present in high abundance in our samples, which could be due to differences in sampling sites[24]. Similarly, Biesbroek et al[32] showed that nasopharyngeal samples were co-dominated early on by *Moraxella, Staphylococcus, Streptococcus* and

Corynebacterium, which had significantly shifted by age 24 months to a high degree of *Moraxella* dominance, with *Streptococcus* and *Haemophilus* as the other primary constituents. Intriguingly, their 6- and 12-month samples were more similar to the 24-month time point than the 1.5-month time point, suggesting a rapid maturation in the nasopharyngeal microbiota. Both of these studies found major contributions from *Moraxella*, which was also found to dominate approximately half of samples in a study of BAL samples collected from 35 wheezers with a median age of 34 months, comparable to our pre-school cohorts[73]. *Moraxella* was also found in 4 week old infants by culture of hypopharyngeal aspirates, which was associated with later asthma[10]. However, *Moraxella* was not among the top taxa found in our study. This may be due to a difference in sampling location, since the employed technique and primers have been shown to detect *Moraxella* well in previous reports[33]. In addition, the prevalence of 5% in our study is comparable to that of adult throat swabs from the Human Microbiome Project, where the prevalence was also much lower than e.g. anterior nares[59] (7.5% throat vs 18.2% anterior nares, V35 data)[74]. Notably, detection of such a taxon will also depend on the sequencing depth.

We found similar overall compositions between the pre-school and school-age cohorts, and no significant association with age on composition in beta-diversity analyses within each age group but were able to define a strong Microbial Maturity Score using a random forest model, which takes into account potential interactions between multiple ASVs. This type of readout has previously been used to link delayed maturation in the gut microbiome with asthma development[58] and nasal microbiota with respiratory infections[34]. While the maturation score was strong in our data, we did not find any association to asthma/wheeze severity, i.e. delayed or accelerated maturation in children with severe compared with mild/moderate asthma/wheeze.

This report focuses on the potential of the oropharyngeal airway microbiota in distinguishing between children with severe and mild/moderate asthma/wheeze. However, combining multiple omics layers, such as transcriptomics, proteomics and metabolomics[37], could provide novel insights not present in each layer on its own[75].

Conclusion

We found that the airway microbiota, sampled using oropharyngeal swabs, was overall modestly different between children and adolescents with severe and mild/moderate

asthma/wheeze in a large multicenter study, across all children but not in individual age groups. This difference constituted a shift in microbiota composition in multivariable adjusted but not crude models. No specific taxa were significantly different between severity groups. Furthermore, we found strong associations between the microbiota and age, which we expressed as a microbial maturation score. However, this score could not distinguish between disease severity groups. This suggests oropharyngeal airway microbiota as interesting entity in studying asthma severity, but probably without the strength to serve as a biomarker for targeted intervention.

Acknowledgements

The authors would like to thank all the participating children and their families. In addition, the authors thank the entire U-BIOPRED study group, whose names are listed in a supplementary file.

Data availability

The sequencing data from this study has been deposited in the European Nucleotide Archive with accession number PRJEB47973. The data has been annotated with cohort it pertains to, which will allow reproduction of part of the main findings without compromising anonymity. Individual-level clinical data cannot be published due to privacy and to comply with the European Regulation 2016/679 of the European Parliament and of the Council (GDPR). Requests for access to the clinical data under a collaboration agreement can be directed to the corresponding author, and can be made available as a collaborative effort.

References

1. Martinez FD, Vercelli D. Asthma. The Lancet. 2013;382:1360–72.

2. World Health Organization. WHO | Asthma [Internet]. [cited 2022 Feb 10]. Available from: http://www.who.int/mediacentre/factsheets/fs307/en/

3. Moffatt MF, Gut IG, Demenais F, Strachan DP, Bouzigon E, Heath S, et al. A large-scale, consortium-based genomewide association study of asthma. N Engl J Med. 2010;363:1211–21.

4. Bønnelykke K, Sleiman P, Nielsen K, Kreiner-Møller E, Mercader JM, Belgrave D, et al. A genome-wide association study identifies CDHR3 as a susceptibility locus for early childhood asthma with severe exacerbations. Nat Genet. 2014;46:51–5.

5. Burke H, Leonardi-Bee J, Hashim A, Pine-Abata H, Chen Y, Cook DG, et al. Prenatal and passive smoke exposure and incidence of asthma and wheeze: systematic review and meta-analysis. Pediatrics. 2012;129:735–44.

595 6. Braun-Fahrländer C, Riedler J, Herz U, Eder W, Waser M, Grize L, et al. Environmental
596 exposure to endotoxin and its relation to asthma in school-age children. N Engl J Med.
597 2002;347:869–77.

598 7. von Mutius E, Vercelli D. Farm living: effects on childhood asthma and allergy. Nat
599 Rev Immunol. 2010;10:861–8.

600 8. Sevelsted A, Stokholm J, Bisgaard H. Risk of Asthma from Cesarean Delivery Depends
601 on Membrane Rupture. J Pediatr. 2016;171:38-42.e1-4.

602 9. Bønnelykke K, Ober C. Leveraging gene-environment interactions and endotypes for
603 asthma gene discovery. J Allergy Clin Immunol. 2016;137:667–79.

604 10. Bisgaard H, Hermansen MN, Buchvald F, Loland L, Halkjaer LB, Bønnelykke K, et al.
605 Childhood asthma after bacterial colonization of the airway in neonates. N Engl J Med.
606 2007;357:1487–95.

607 11. Strachan DP. Hay fever, hygiene, and household size. Br Med J. 1989;299:1259–60.

608 12. Bach JF. The Effect of Infections on Susceptibility to Autoimmune and Allergic
609 Diseases. N Engl J Med. 2002;347:911–20.

610 13. Eder W, Ege MJ, von Mutius E. The Asthma Epidemic. N Engl J Med. 2006;355:2226–
611 35.

612 14. Bisgaard H, Bønnelykke K, Stokholm J. Immune-mediated diseases and microbial
613 exposure in early life. Clin Exp Allergy J Br Soc Allergy Clin Immunol. 2014;44:475–
614 81.

615 15. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank I, Yadav A, et al.
616 Topographical Continuity of Bacterial Populations in the Healthy Human Respiratory
617 Tract. Am J Respir Crit Care Med. 2011;184:957–63.

618 16. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison
619 of the Respiratory Microbiome in Healthy Nonsmokers and Smokers. Am J Respir Crit
620 Care Med. 2013;187:1067–75.

17. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, et al. Disordered Microbial Communities in Asthmatic Airways. PLoS ONE. 2010;5:e8578.

18. Huang YJ, Nelson CE, Brodie EL, DeSantis TZ, Baek MS, Liu J, et al. Airway Microbiota and Bronchial Hyperresponsiveness in Patients with Sub-optimally Controlled Asthma. J Allergy Clin Immunol. 2011;127:372-381.e3.

19. Huang YJ, Nariya S, Harris JM, Lynch SV, Choy DF, Arron JR, et al. The airway microbiome in patients with severe asthma: Associations with disease features and severity. J Allergy Clin Immunol. 2015;136:874–84.

20. Thorsen J, Rasmussen MA, Waage J, Mortensen M, Brejnrod A, Bønnelykke K, et al. Infant airway microbiota and topical immune perturbations in the origins of childhood asthma. Nat Commun. 2019;10:1–8.

21. Toivonen L, Karppinen S, Schuez-Havupalo L, Waris M, He Q, Hoffman KL, et al. Longitudinal Changes in Early Nasal Microbiota and the Risk of Childhood Asthma. Pediatrics. 2020;146:e20200421.

22. Cardenas PA, Cooper PJ, Cox MJ, Chico M, Arias C, Moffatt MF, et al. Upper airways microbiota in antibiotic-naïve wheezing and healthy infants from the tropics of rural Ecuador. PloS One. 2012;7:e46803.

23. Boutin S, Depner M, Stahl M, Graeber SY, Dittrich SA, Legatzki A, et al. Comparison of Oropharyngeal Microbiota from Children with Asthma and Cystic Fibrosis. Mediators Inflamm. 2017;2017:5047403.

24. Depner M, Ege MJ, Cox MJ, Dwyer S, Walker AW, Birzele LT, et al. Bacterial microbiota of the upper respiratory tract and childhood asthma. J Allergy Clin Immunol. 2017;139:826-834.e13.

25. Chiu CY, Chou HC, Chang LC, Fan WL, Dinh MCV, Kuo YL, et al. Integration of metagenomics-metabolomics reveals specific signatures and functions of airway microbiota in mite-sensitized childhood asthma. Allergy. 2020;75:2846–57.

26. Zhou Y, Jackson D, Bacharier LB, Mauger D, Boushey H, Castro M, et al. The upper-airway microbiota and loss of asthma control among asthmatic children. Nat Commun. 2019;10:1–10.

27. Yan M, Pamp SJ, Fukuyama J, Hwang PH, Cho DY, Holmes S, et al. Nasal Microenvironments and Interspecific Interactions Influence Nasal Microbiota Complexity and S. aureus Carriage. Cell Host Microbe. 2013;14:631–40.

28. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the Upper Respiratory Tract Microbiotas as the Source of the Lung and Gastric Microbiotas in Healthy Individuals. mBio. 2015;6:e00037-15.

29. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Beck JM, Huffnagle GB, et al. Spatial Variation in the Healthy Human Lung Microbiome and the Adapted Island Model of Lung Biogeography. Ann Am Thorac Soc. 2015;12:821–30.

30. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Falkowski NR, Huffnagle GB, et al. Bacterial Topography of the Healthy Human Lower Respiratory Tract. mBio. 2017;8:e02287-16.

31. Ahmed B, Cox MJ, Cuthbertson L, James PL, Cookson WOC, Davies JC, et al. Comparison of the upper and lower airway microbiota in children with chronic lung diseases. PLOS ONE. 2018;13:e0201156.

32. Biesbroek G, Tsivtsivadze E, Sanders EAM, Montijn R, Veenhoven RH, Keijser BJF, et al. Early Respiratory Microbiota Composition Determines Bacterial Succession Patterns and Respiratory Health in Children. Am J Respir Crit Care Med. 2014;190:1283–92.

33. Mortensen MS, Brejnrod AD, Roggenbuck M, Abu Al-Soud W, Balle C, Krogfelt KA, et al. The developing hypopharyngeal microbiota in early life. Microbiome. 2016;4:70.

34. Bosch AATM, Piters WAA de S, van Houten MA, Chu MLJN, Biesbroek G, Kool J, et al. Maturation of the Infant Respiratory Microbiota, Environmental Drivers, and Health Consequences. A Prospective Cohort Study. Am J Respir Crit Care Med. 2017;196:1582–90.

35. Man WH, Clerc M, de Steenhuijsen Piters WAA, van Houten MA, Chu MLJN, Kool J, et al. Loss of Microbial Topography between Oral and Nasopharyngeal Microbiota and Development of Respiratory Infections Early in Life. Am J Respir Crit Care Med. 2019;200:760–70.

36. Fitzpatrick AM, Teague WG, Meyers DA, Peters SP, Li X, Li H, et al. Heterogeneity of severe asthma in childhood: Confirmation by cluster analysis of children in the National Institutes of Health/National Heart, Lung, and Blood Institute Severe Asthma Research Program. J Allergy Clin Immunol. 2011;127:382-389.e13.

37. Fleming L, Murray C, Bansal AT, Hashimoto S, Bisgaard H, Bush A, et al. The burden of severe asthma in childhood and adolescence: results from the paediatric U-BIOPRED cohorts. Eur Respir J. 2015;

38. Stokholm J, Chawes BL, Vissing NH, Bjarnadóttir E, Pedersen TM, Vinding RK, et al. Azithromycin for episodes with asthma-like symptoms in young children aged 1–3 years: a randomised, double-blind, placebo-controlled trial. Lancet Respir Med. 2016;4:19–26.

39. Bacharier LB, Guilbert TW, Mauger DT, Boehmer S, Beigelman A, Fitzpatrick AM, et al. Early Administration of Azithromycin and Prevention of Severe Lower Respiratory Tract Illnesses in Preschool Children With a History of Such Illnesses: A Randomized Clinical Trial. JAMA. 2015;314:2034–44.

40. Thorsen J, Stokholm J, Rasmussen MA, Mortensen MS, Brejnrod AD, Hjelmsø M, et al. The Airway Microbiota Modulates Effect of Azithromycin Treatment for Episodes of Recurrent Asthma-like Symptoms in Preschool Children: A Randomized Clinical Trial. Am J Respir Crit Care Med. 2021;204:149–58.

41. Wheelock CE, Goss VM, Balgoma D, Nicholas B, Brandsma J, Skipp PJ, et al. Application of 'omics technologies to biomarker discovery in inflammatory lung diseases. Eur Respir J. 2013;42:802–25.

42. Loza MJ, Adcock I, Auffray C, Chung KF, Djukanovic R, Sterk PJ, et al. Longitudinally Stable, Clinically Defined Clusters of Patients with Asthma Independently Identified in

the ADEPT and U-BIOPRED Asthma Studies. Ann Am Thorac Soc. 2016;13 Suppl 1:S102-103.

43. Reddel HK, Taylor DR, Bateman ED, Boulet LP, Boushey HA, Busse WW, et al. An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice. Am J Respir Crit Care Med. 2009;180:59–99.

44. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Internet]. [cited 2022 Feb 10]. Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

45. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8.

46. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10–2.

47. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.

48. Callahan B. DADA2 Pipeline Tutorial (1.16) [Internet]. [cited 2022 Feb 10]. Available from: https://benjjneb.github.io/dada2/tutorial.html

49. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590-596.

50. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36:996–1004.

51. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE. 2013;8:e61217.

52. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016 [cited 2022 Feb 10]. Available from: https://ggplot2.tidyverse.org

53. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. Appl Environ Microbiol. 2007;73:1576–85.

54. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, et al. vegan: Community Ecology Package [Internet]. 2015 [cited 2022 Feb 10]. Available from: https://cran.r-project.org/web/packages/vegan/index.html

55. Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. Nat Methods. 2013;10:1200–2.

56. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B Methodol. 1995;57:289–300.

57. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. Nature. 2014;510:417–21.

58. Stokholm J, Blaser MJ, Thorsen J, Rasmussen MA, Waage J, Vinding RK, et al. Maturation of the gut microbiome and risk of asthma in childhood. Nat Commun. 2018;9:141.

59. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486:207–14.

60. Luna PN, Hasegawa K, Ajami NJ, Espinola JA, Henke DM, Petrosino JF, et al. The association between anterior nares and nasopharyngeal microbiota in infants hospitalized for bronchiolitis. Microbiome. 2018;6:2.

61. Abdel-Aziz MI, Brinkman P, Vijverberg SJH, Neerincx AH, Riley JH, Bates S, et al. Sputum microbiome profiles identify severe asthma phenotypes of relative stability at 12 to 18 months. J Allergy Clin Immunol. 2021;147:123–34.

62. Forslund K, Hildebrand F, Nielsen T, Falony G, Le Chatelier E, Sunagawa S, et al. Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. Nature. 2015;528:262–6.

63. Sverrild A, Kiilerich P, Brejnrod A, Pedersen R, Porsbjerg C, Bergqvist A, et al. Eosinophilic airway inflammation in asthmatic patients is associated with an altered airway microbiome. J Allergy Clin Immunol. 2017;140:407-417.e11.

64. Taylor SL, Leong LEX, Choo JM, Wesselingh S, Yang IA, Upham JW, et al. Inflammatory phenotypes in patients with severe asthma are associated with distinct airway microbiology. J Allergy Clin Immunol. 2018;141:94-103.e15.

65. Durack J, Lynch SV, Nariya S, Bhakta NR, Beigelman A, Castro M, et al. Features of the bronchial bacterial microbiome associated with atopy, asthma, and responsiveness to inhaled corticosteroid treatment. J Allergy Clin Immunol. 2017;140:63–75.

66. de Steenhuijsen Piters WAA, Heinonen S, Hasrat R, Bunsow E, Smith B, Suarez-Arrabal MC, et al. Nasopharyngeal Microbiota, Host Transcriptome and Disease Severity in Children with Respiratory Syncytial Virus Infection. Am J Respir Crit Care Med [Internet]. 2016 [cited 2016 May 9]; Available from: http://www.atsjournals.org.ep.fjernadgang.kb.dk/doi/abs/10.1164/rccm.201602-0220OC

67. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, et al. The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development. Cell Host Microbe. 2015;17:704–15.

68. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486:222–7.

69. Dominguez-Bello MG, Blaser MJ, Ley RE, Knight R. Development of the human gastrointestinal microbiota and insights from high-throughput sequencing. Gastroenterology. 2011;140:1713–9.

70. Hollister EB, Riehle K, Luna RA, Weidler EM, Rubio-Gonzales M, Mistretta TA, et al. Structure and function of the healthy pre-adolescent pediatric gut microbiome. Microbiome. 2015;3:36.

71. Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, et al. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. Cell Host Microbe. 2015;17:690–703.

786 72. Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen AM, et al.
787 The Dynamics of the Human Infant Gut Microbiome in Development and in
788 Progression toward Type 1 Diabetes. Cell Host Microbe. 2015;17:260–73.

789 73. Robinson PFM, Pattaroni C, Cook J, Gregory L, Alonso AM, Fleming LJ, et al. Lower
790 airway microbiota associates with inflammatory phenotype in severe preschool wheeze.
791 J Allergy Clin Immunol. 2019;143:1607-1610.e3.

792 74. NIH Human Microbiome Project - HMQCP [Internet]. [cited 2022 Feb 10]. Available
793 from: https://www.hmpdacc.org/hmp/HMQCP/

794 75. Abdel-Aziz MI, Neerincx AH, Vijverberg SJ, Kraneveld AD, Maitland-van der Zee AH.
795 Omics for the future in asthma. Semin Immunopathol. 2020;42:111–26.

796

797

Tables

Cohort	A	B	C	D	All
Severity	Severe	Mild/Moderate	Severe	Mild/Moderate	-
Age category	School age	School age	Pre-school age	Pre-school age	-
Number of participants	97	43	77	54	271
Number of samples included	86	39	65	51	241
Male sex, n (%)	44 (51.2%)	26 (66.7%)	42 (64.6%)	33 (64.7%)	145 (60.2%)
Age (yrs), mean (sd)	12.5 (2.8)	11.2 (3.2)	3.6 (1.2)	3.4 (1.2)	7.9 (4.9)
Caucasian ethnicity, n (%)	66 (76.7%)	30 (76.9%)	51 (78.5%)	45 (88.2%)	192 (79.7%)
Any siblings, n (%)	79 (91.9%)	33 (84.6%)	51 (78.5%)	43 (84.3%)	206 (85.5%)
C-section delivery, n (%)	23 (26.7%)	9 (23.1%)	13 (20.0%)	13 (25.5%)	58 (24.1%)
Height (cm), mean (sd)	153.4 (15.6)	147.9 (17.5)	103.2 (9.5)	103.1 (11.1)	128.5 (27.9)
Weight (kg), mean (sd)	52.9 (18.3)	43.2 (15.2)	18 (4.3)	17.2 (3.4)	34.4 (20.9)
Body Mass Index, mean (sd)	21.9 (5.1)	19 (3.1)	16.8 (2.2)	16 (1.1)	18.8 (4.3)
FEV1, % of predicted, mean (sd)	87.8 (21.5)	90.5 (17.3)	-	-	90.9 (20.3)
FEF 25-75, % of predicted, mean (sd)	69.5 (30.8)	78.4 (26.9)	-	-	77.1 (36.1)
FeNO ppb, median (IQR)	38.8 (16-61.8)	26 (14.5-41)	-	-	30 (15.6-57)
Recent inhaled corticosteroid treatment, n (%)	86 (100%)	38 (97.4%)	65 (100%)	23 (45.1%)	212 (88.0%)
Recent oral corticosteroid treatment, n (%)	25 (29.1%)	1 (2.6%)	14 (21.5%)	0 (0%)	40 (16.6%)
Recent nasal corticosteroid treatment, n (%)	12 (14.0%)	5 (12.8%)	1 (1.5%)	0 (0%)	18 (7.5%)
Recent antibiotic treatment, n (%)	15 (17.5%)	2 (5.1%)	8 (12.3%)	5 (9.8%)	30 (12.4%)
Number of exacerbations in the previous year, median (IQR)	3 (2-5)	1 (0-2)	4 (2-6)	1 (0-3)	2 (1-5)
Skin Prick Test positive, n (%)	66 (76.7%)	30 (76.9%)	19 (29.2%)	18 (35.3%)	133 (55.2%)
Specific IgE positivie, n (%)	54 (62.8%)	21 (53.8%)	16 (24.6%)	15 (29.4%)	106 (44.0%)
Urine cotinine positive, n (%)	7 (8.5%)	5 (14.7%)	10 (19.6%)	2 (4.5%)	24 (11.4%)

Table 1: Baseline characteristics of the participants from each of the four cohorts. Summaries correspond to children with included samples only. Sd – standard deviation. FEV1 – Forced expiratory volume in 1 second. FEF 25-75 – Forced expiratory flow at 25-75%. FeNO – Fraction of exhaled nitric oxide. IQR – Interquartile range.

	All children				School-age children				Pre-school children			
Univariable	N	R^2	F	P-value	N	R^2	F	P-value	N	R^2	F	P-value
Severity (Severe vs Mild/Moderate)	241	0.007	1.99	0.080	125	0.005	0.64	0.695	116	0.009	1.08	0.327
Age group (School-age vs. Pre-school)	241	0.012	3.32	**0.011**								
Age (Per year)	241	0.011	2.95	**0.025**	125	0.003	0.49	0.802	116	0.020	0.60	0.931
Sex (Male vs. Female)	241	0.004	0.96	0.388	125	0.004	0.50	0.787	116	0.006	0.81	0.544
Ethnicity (Caucasian vs. Other)	241	0.003	0.92	0.434	125	0.009	1.26	0.221	116	0.001	0.15	0.994
Delivery (Cesarean sction vs. Vaginal)	241	0.005	1.22	0.282	125	0.004	0.63	0.678	116	0.006	0.69	0.658
Study center (Overall effect)	241	0.040	1.84	**0.011**	125	0.036	1.03	0.369	116	0.068	1.75	**0.017**
Season of visit (Overall effect)	241	0.022	2.01	**0.015**	125	0.039	1.87	**0.035**	116	0.038	1.60	0.066
Siblings (Any)	241	0.003	0.82	0.519	125	0.013	1.84	0.109	116	0.019	2.39	**0.035**
Number of Siblings	241	0.024	1.07	0.352	125	0.047	1.12	0.292	116	0.054	1.37	0.103
Breastfeeding (Per month)	241	0.005	1.27	0.243	125	0.006	0.82	0.524	116	0.008	0.94	0.427
Cat in the home	241	0.004	1.20	0.265	125	0.009	1.29	0.266	116	0.007	0.81	0.546
Dog in the home	241	0.002	0.63	0.695	125	0.007	1.01	0.365	116	0.011	1.41	0.204
Cotinine in Urine (Positive vs Negative)	211	0.003	0.59	0.728	116	0.006	0.85	0.508	95	0.003	0.29	0.953
Number of Exacerbations, previous year	241	0.008	2.24	0.057	125	0.009	1.28	0.228	116	0.012	1.50	0.171
Specific IgE (Any positive)	141	0.007	1.08	0.302	82	0.008	0.76	0.538	59	0.003	0.20	0.989
Skin Prick Test (Any positive)	204	0.013	2.85	**0.016**	104	0.006	0.71	0.613	100	0.006	0.70	0.611
Any sensitization (sIgE or SPT)	226	0.011	2.90	**0.020**	118	0.006	0.85	0.470	108	0.008	0.93	0.436
Diagnosed Eczema	239	0.008	2.17	0.073	125	0.009	1.28	0.253	114	0.016	2.01	0.068
Diagnosed Allergic Rhinitis	233	0.002	0.54	0.766	122	0.001	0.09	0.998	111	0.008	0.99	0.386
Recent antibiotic treatment	241	0.002	0.59	0.718	125	0.003	0.37	0.909	116	0.008	0.97	0.396
Recent inhaled corticosteroid treatment	241	0.008	2.08	0.065					116	0.010	1.28	0.234
Recent nasal corticosteroid treatment	241	0.003	0.69	0.629	125	0.005	0.74	0.558	116	0.004	0.47	0.813
Recent oral corticosteroid treatment	241	0.004	1.08	0.316	125	0.007	1.04	0.359	116	0.014	1.70	0.125
Multivariable Backward Selection												
Severity (Severe vs Mild/Moderate)	241	0.009	2.66	**0.035**	125	0.004	0.64	0.709	116	0.009	1.19	0.275
Age group (School-age vs. Pre-school)	241	0.009	2.71	**0.025**								
Delivery (Cesarean section vs Vaginal)	241	0.006	1.83	0.097								
Study center (Overall effect)	241	0.047	2.23	**0.002**	125	0.040	1.15	0.247	116	0.072	1.90	**0.012**
Season of visit (Overall effect)	241	0.021	2.00	**0.019**	125	0.041	2.01	**0.025**				
Siblings (Any)									116	0.023	3.00	**0.011**
Breastfeeding (Per month)	241	0.007	1.93	0.095								
Number of Exacerbations, previous year	241	0.017	4.98	**0.001**								
Current oral corticosteroid treatment	241	0.006	1.78	0.095								

Table 2: Associations between environmental and phenotypic covariates and microbial composition (log-transformed weighted UniFrac distances) from PERMANOVA tests of beta diversity. Results from Univariable tests correspond to models with one term at a time, whereas Multivariable Backward Selection refers to a model with all terms presented above,

successively dropping the term with the highest p-value until all terms have p-values < 0.1. In addition, all models, including those labeled univariable, were adjusted for the technical variables library size, log(library size), DNA extraction plate and sequencing run. Similarly, age group and study center were kept in all models. F statistics measure distance between / within groups and can be interpreted as a ratio where a value of 1 corresponds to no association. R^2 measures variance explained. P-values from non-parametric permutation tests from the adonis PERMANOVA model.

Figures

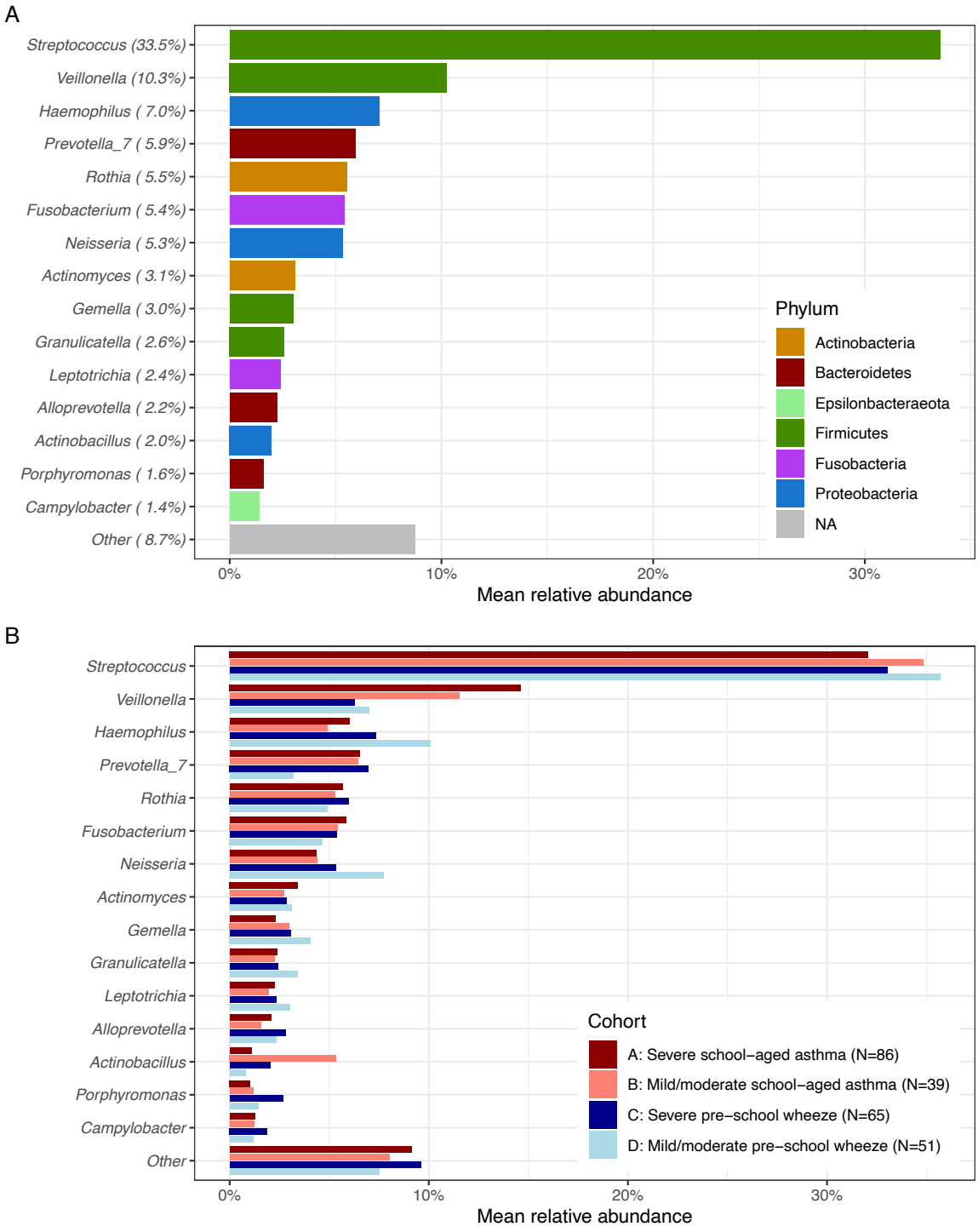

Fig 1: A) Mean relative abundances of most common genera across all samples (N=241), colored by phylum. B) Mean relative abundances of these genera within each of the four cohorts.

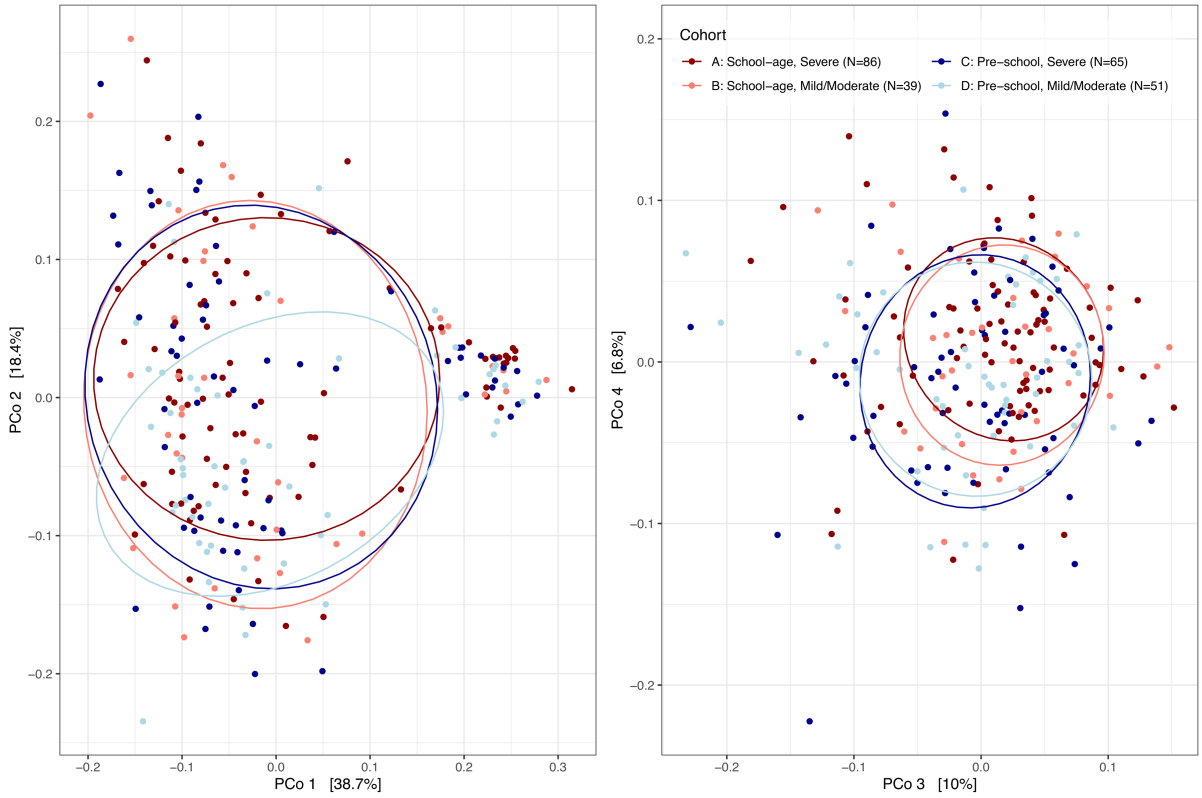

Fig 2: Principal Coordinates plot from weighted UniFrac distances on log-transformed counts between all samples (N=241), colored by cohort. Ellipses demonstrate +/- 1 SD within each cohort. A shift between the school-age (cohorts A and B) and pre-school (cohorts C and D) children are observed (adonis2, F = 3.32, p = 0.011, n=241), but not between severe and mild/moderate cases within each age group or in the full cohort. This shift was more pronounced in PCo3 and PCo4 than the two first axes, which indicates that the majority of the compositional variation in the microbiota did not differ between cohorts.

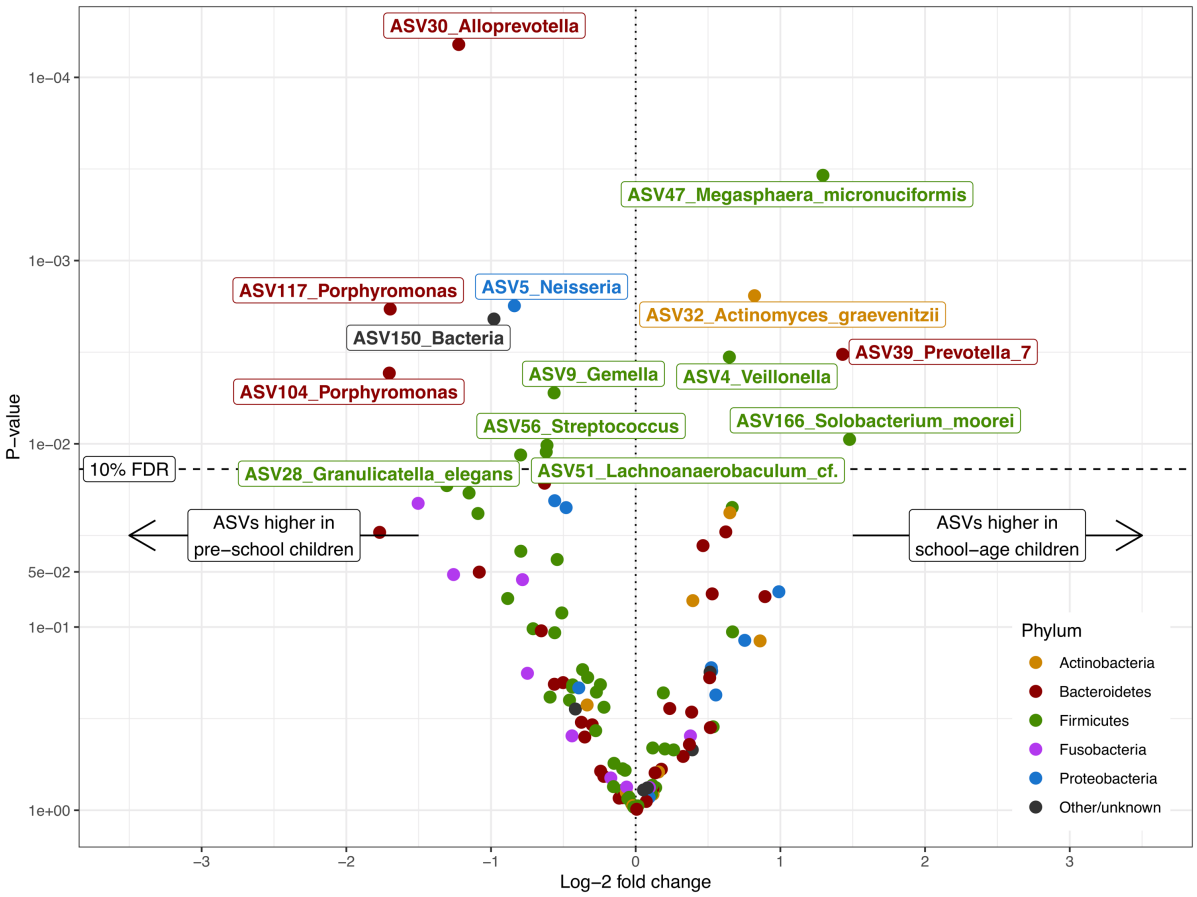

Fig 3: Volcano plot showing effect size (Log-2 fold change, x-axis) and p-values (y-axis) of differential abundance analysis of Amplicon Sequence Variants (ASVs) between school-age and pre-school children, colored by taxonomic phylum and labeled by highest taxonomic level identifiable. Differences were tested using the feature model from metagenomeSeq. N=241.

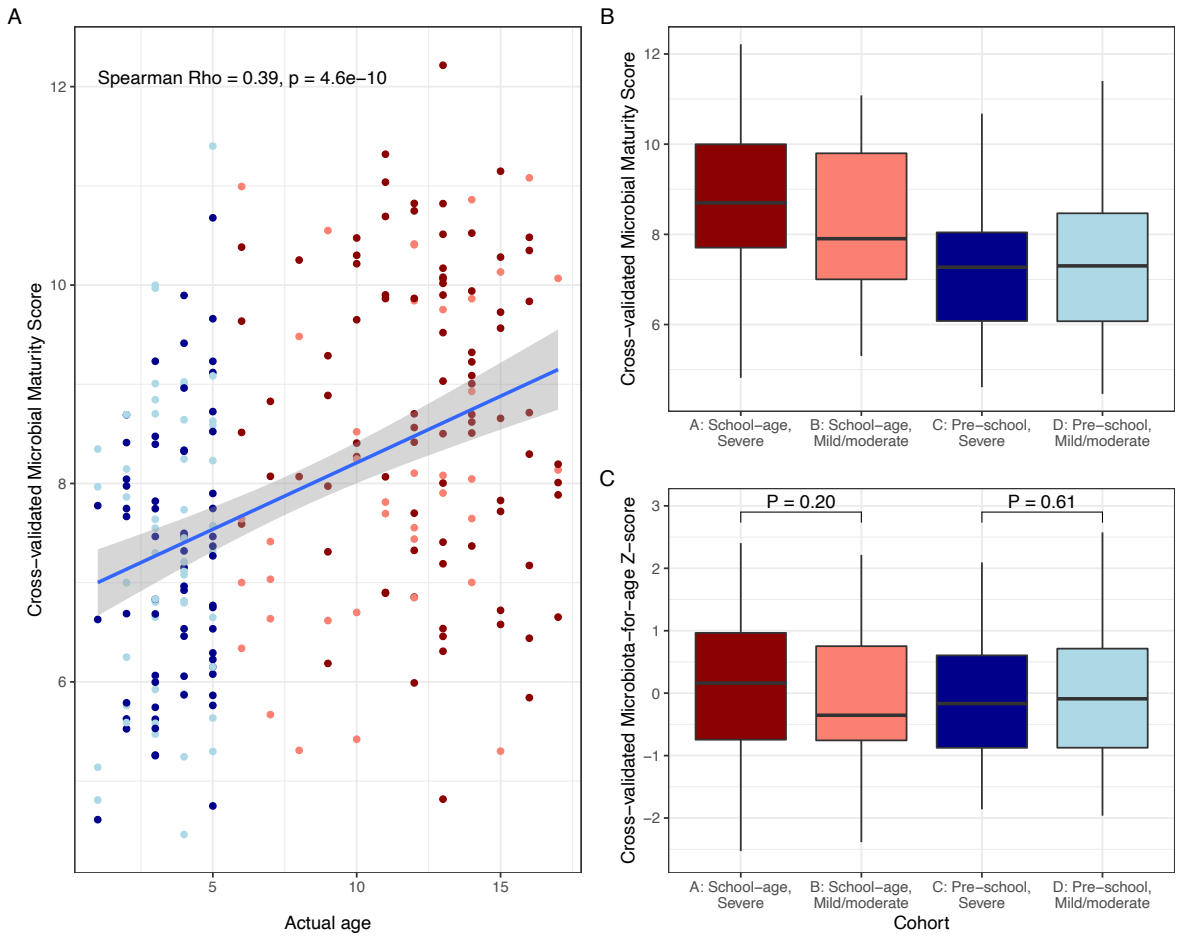

Fig 4: Predictive Microbial Maturity score based on a cross-validated Random Forest model of Amplicon Sequence Variant (ASV) relative abundances predicting actual ages of the participants. A) The relationship between the actual age of participants and the Microbial Maturity score. B) The distribution of Microbial Maturity score across cohorts. C) Distribution of the relative microbial maturity adjusted for age, termed Microbiota-for-age Z-score, across cohorts, showing no significant association between relative maturity and disease severity in either age group.

Supplemental information

	A: Severe school-aged asthma (Cohort SA)	B: Mild - moderate school-aged asthma (Cohort MMA)	C: Severe pre-school wheeze (Cohort SW)	D: Mild - moderate pre-school wheeze (Cohort MMW)
	1. Parent / legal guardian able to give written informed consent 2. Assent obtained from all children where appropriate 3. Parent / legal guardian, or where appropriate the child, able to complete the study and all measurements 4. Parent / legal guardian, or where appropriate the child, able to read, comprehend, and write at a sufficient level to complete study related materials			
Inclusion criteria	1. Aged between 6 – 17 years at screening 2. Under follow up with a respiratory specialist for at least 6 months prior to enrolling in the study **3.** Previous assessment to exclude other diagnoses and optimise asthma control including treatment of co-morbidities, assessment of adherence and reduction in allergen exposure in sensitised individuals	1. Aged between 6 – 17 years at screening	1. Aged between 1 – 5 years at the time of screening 2. Under follow up with a respiratory specialist for at least 6 months prior to enrolling in the study **3.** Previous assessment to exclude other diagnoses and optimise asthma control including treatment of co-morbidities, assessment of adherence and reduction in allergen exposure in sensitised individuals	1. Aged between 1 – 5 years at the time of screening
	1. As a result of medical interview, physical examination or screening investigation the physician responsible considers the child unfit for the study 2. History of allergy, which, in the opinion of the responsible physician, contra-indicates participation 3. Participant is female who is pregnant or lactating or up to 6 weeks post partum or within 6 weeks cessation of breast feeding. 4. Participation within 3 months of the first dose in a study using a new molecular entity, or the first dose in any other study investigating drugs or having participated within three months in a study with invasive procedures (permission to enroll required from Scientific Board) 5. Risk of non-compliance with study procedures 6. Prematurity ≤35 weeks gestation 7. Change in asthma medication within 4 weeks of the screening assessment (except those using the Symbicort maintenance and reliever therapy (SMART) regime (assessment deferred)			

	8. History or current evidence of an upper or lower respiratory infection or symptoms (including common cold) within 2 weeks of baseline assessment (assessment deferred) 9. Severe exacerbation within 4 weeks of the baseline assessment (assessment deferred)	
Exclusion criteria	1. Significant alternative diagnoses that may mimic or complicate asthma, in particular dysfunctional breathing, panic attacks, and overt psychosocial problems (if these are thought to be the *major* problem rather than *in addition* to severe asthma) 2. Significant other primary pulmonary disorders in particular cystic fibrosis, interstitial lung disease 3. Patients with bronchiectasis should only be excluded if this is thought to be the major pulmonary disorder rather than *in addition* to severe asthma 4. Diagnosis of chronic inflammatory diseases other than asthma (inflammatory bowel disease, rheumatoid arthritis)	1. Significant other primary pulmonary disorders in particular cystic fibrosis, interstitial lung disease 2. Patients with bronchiectasis should only be excluded if this is thought to be the major pulmonary disorder rather than *in addition* to severe asthma 3. Diagnosis of chronic inflammatory diseases other than asthma (inflammatory bowel disease, rheumatoid arthritis)
Asthma diagnosis	**Any one or more of the following:** • Improvement in $FEV_1 \geq 12\%$ or 200ml predicted after inhalation of 400 mcg salbutamol • Airway hyper-responsiveness (PC_{20} <8mg/ml) • Spontaneous variability in FEV_1 of $\geq$ 12% or 200 ml over the past year • Diurnal variability in peak flow in a properly conducted trial of $\geq$ 15%, with a compatible clinical picture • Persistent airflow limitation: post steroid trial (either 2 weeks of high dose OCS or intramuscular triamcinolone) post bronchodilator z score <-1.96 (cohort a only)	**History of breathlessness and wheeze**

Asthma treatment	High dose ICS (≥500mcg FP or ≥800mcg BUD daily or equivalent) ± OCS plus a trial of at least two other controller medications (long acting beta agonists, leukotriene receptor antagonists or theophylline).	Low to moderate dose ICS (≤250mcg FP daily or ≤400mcg BUD or equivalent) plus none or one other controller medication (long acting beta agonists, leukotriene receptor antagonists or theophylline)	High dose ICS (≥200mcg FP or ≥400mcg BUD daily or equivalent) and a leukotriene receptor antagonist, or a failed trial of these	No treatment or low dose ICS (≤100mcg FP or ≤200mcg BUD daily or equivalent) and/or a leukotriene receptor antagonist
Asthma control	**Any one or more of the following:** **1. Persistent symptoms**, both of the following for 3 of the past 6 months: • Persistent symptoms (at least 50% of days) • Need for reliever treatment ≥3 times per week because of asthma symptoms **2. Frequent severe exacerbations** • ≥2 in the past year or ≥3 in the past 2 years requiring hospital attendance or high dose OCS • or ≥1 in the past year requiring PICU admission) **3. z score FEV_1 <-1.96** • post bronchodilator post steroid trial **4. Prescribed daily or alternate day oral steroids and/or omalizumab** • irrespective of level of symptom control	**Controlled asthma. All of the following in the last 4 weeks:** • Daytime symptoms twice per week or less • No limitation to activities • No nocturnal symptoms • Need for reliever treatment twice per week or less • $FEV_1 \geq 80\%$ predicted **OR** **Partially controlled asthma. One or two of the following in the previous 4 weeks:** • Daytime symptoms more than twice per week • Any limitation of activities • Any nocturnal symptoms • Need for reliever treatment twice per week or more • Pre bronchodilator $FEV_1 < 80\%$ predicted	**Any one or more of the following:** **1. Persistent symptoms, both of the following for 3 of the past 6 months:** • Persistent symptoms of wheeze, cough or breathlessness (at least 50% of days) • Need for reliever treatment ≥3 times per week because of asthma symptoms **2. Frequent severe exacerbations** • (≥2 in the past year or ≥3 in the past 2 years requiring hospital attendance or course of high dose ICS or oral prednisolone • or ≥1 in the past year requiring PICU admission; **3. Prescription of daily or alternate day OCS** • irrespective of level of symptom control	**Controlled symptoms (wheeze, cough or breathlessness). All of the following in the last 4 weeks:** • Daytime symptoms twice per week or less • No limitations to activities • No nocturnal symptoms • Need for reliever treatment twice per week or less **OR** **Partially controlled symptoms. One or two of the following in the previous 4 weeks:** • Daytime symptoms more than twice per week • Any limitation of activities • Any nocturnal symptoms • Need for reliever treatment twice per week or more

Table S1: Inclusion and exclusion criteria for the four cohorts. Reproduced from the original paper describing the baseline characteristics of the cohort: Fleming L, Murray C, Bansal AT, Hashimoto S, Bisgaard H, Bush A, et al. The burden of severe asthma in childhood and adolescence: results from the paediatric U-BIOPRED cohorts. Eur Respir J. 2015. DOI: 10.1183/13993003.00780-2015.

	All children					School-age children					Pre-school children				
Variable	N	Estimate	95% CI	P-value	P-value (overall)	N	Estimate	95% CI	P-value	P-value (overall)	N	Estimate	95% CI	P-value	P-value (overall)
Severity (Severe vs. Mild/Moderate)	241	0.075	[-0.07;0.22]	0.306	0.303	125	0.166	[-0.04;0.37]	0.107	0.102	116	-0.024	[-0.23;0.19]	0.825	0.822
Age group (School-age vs. Pre-school)	241	0.088	[-0.05;0.23]	0.212	0.208										
Age (Per year)	240	0.006	[-0.01;0.02]	0.445	0.442	124	-0.017	[-0.05;0.01]	0.286	0.279	116	0.039	[-0.05;0.13]	0.374	0.367
Sex (Male vs. Female)	241	-0.056	[-0.20;0.09]	0.440	0.436	125	-0.005	[-0.20;0.19]	0.961	0.961	116	-0.092	[-0.31;0.13]	0.402	0.395
Ethnicity (Caucasian vs. Other)	241	-0.045	[-0.22;0.13]	0.609	0.607	125	0.031	[-0.19;0.25]	0.787	0.784	116	-0.125	[-0.40;0.15]	0.369	0.362
Delivery (Cesarean section vs Vaginal)	241	-0.029	[-0.19;0.13]	0.728	0.726	125	0.010	[-0.21;0.23]	0.930	0.929	116	-0.068	[-0.32;0.18]	0.592	0.586
Study centers (B vs A)	241	0.195	[-0.13;0.52]	0.235	0.476						116	0.115	[-0.28;0.51]	0.565	0.447
Study centers (C vs A)	241	0.033	[-0.25;0.31]	0.815	0.476	125	-0.308	[-0.70;0.08]	0.118	0.412	116	0.114	[-0.26;0.48]	0.544	0.447
Study centers (D vs A)	241	-0.051	[-0.64;0.54]	0.864	0.476	125	-0.306	[-0.94;0.32]	0.338	0.412					
Study centers (E vs A)	241	0.093	[-0.17;0.36]	0.489	0.476	125	-0.135	[-0.51;0.24]	0.479	0.412	116	0.038	[-0.27;0.35]	0.809	0.447
Study centers (F vs A)	241	0.016	[-0.27;0.30]	0.915	0.476	125	-0.141	[-0.53;0.25]	0.479	0.412	116	-0.205	[-0.59;0.18]	0.294	0.447
Study centers (G vs A)	241	-0.106	[-0.40;0.18]	0.472	0.476	125	-0.336	[-0.75;0.08]	0.109	0.412	116	-0.160	[-0.52;0.20]	0.377	0.447
Season of visit (Spring vs Autumn)	241	0.007	[-0.19;0.21]	0.944	0.596	125	-0.053	[-0.34;0.23]	0.710	0.788	116	0.113	[-0.18;0.41]	0.447	0.220
Season of visit (Summer vs Autumn)	241	0.097	[-0.09;0.29]	0.315	0.596	125	-0.054	[-0.29;0.18]	0.651	0.788	116	0.332	[-0.00;0.67]	0.051	0.220
Season of visit (Winter vs Autumn)	241	0.102	[-0.09;0.29]	0.301	0.596	125	0.085	[-0.21;0.38]	0.571	0.788	116	0.187	[-0.08;0.46]	0.171	0.220
Siblings (Any)	241	0.041	[-0.16;0.24]	0.684	0.681	125	0.127	[-0.18;0.44]	0.416	0.409	116	-0.032	[-0.30;0.24]	0.812	0.809
Number of Siblings	241	0.039	[-0.03;0.11]	0.252	0.249	125	0.068	[-0.02;0.15]	0.119	0.114	116	-0.025	[-0.14;0.09]	0.656	0.651
Breastfeeding (Per month)	241	0.010	[-0.00;0.02]	0.053	0.051	125	0.009	[-0.00;0.02]	0.176	0.170	116	0.014	[-0.01;0.03]	0.153	0.147
Cat in the home	241	0.022	[-0.16;0.21]	0.814	0.813	125	0.079	[-0.18;0.34]	0.550	0.544	116	-0.014	[-0.29;0.26]	0.921	0.920
Dog in the home	241	0.085	[-0.10;0.27]	0.359	0.356	125	0.085	[-0.13;0.30]	0.437	0.430	116	0.013	[-0.36;0.39]	0.944	0.943
Cotinine in Urine (Positive vs Negative)	211	-0.102	[-0.34;0.14]	0.400	0.396	116	-0.240	[-0.57;0.09]	0.153	0.147	95	0.067	[-0.29;0.43]	0.714	0.709
Number of Exacerbations, previous year	241	-0.015	[-0.04;0.01]	0.207	0.204	125	0.017	[-0.02;0.05]	0.326	0.319	116	-0.044	[-0.08;-0.01]	**0.012**	**0.010**
Specific IgE (Any positive)	141	0.110	[-0.08;0.30]	0.261	0.255	82	-0.049	[-0.45;0.35]	0.810	0.806	59	0.138	[-0.11;0.39]	0.273	0.259
Skin Prick Test (Any positive)	204	0.063	[-0.10;0.22]	0.433	0.429	104	-0.115	[-0.51;0.28]	0.566	0.559	100	0.072	[-0.16;0.30]	0.537	0.530

Any sensitization (sIgE or SPT)	226	0.108	[-0.04;0.26]	0.165	0.162	118	-0.038	[-0.40;0.32]	0.835	0.832	108	0.128	[-0.08;0.34]	0.230	0.222
Diagnosed Eczema	239	0.134	[-0.01;0.28]	0.078	0.075	125	0.222	[0.01;0.43]	**0.039**	**0.036**	114	0.035	[-0.18;0.25]	0.752	0.748
Diagnosed Allergic Rhinitis	233	0.173	[-0.10;0.45]	0.216	0.212	122	0.205	[-0.14;0.55]	0.239	0.232	111	0.108	[-0.35;0.56]	0.639	0.633
Recent antibiotic treatment	241	-0.015	[-0.23;0.20]	0.887	0.887	125	0.126	[-0.15;0.40]	0.368	0.362	116	-0.201	[-0.54;0.14]	0.240	0.233
Recent inhaled corticosteroid treatment	241	-0.113	[-0.33;0.10]	0.299	0.295	125	-0.595	[-1.65;0.46]	0.267	0.260	116	-0.161	[-0.41;0.08]	0.192	0.185
Recent nasal corticosteroid treatment	241	0.107	[-0.16;0.37]	0.426	0.423	125	0.123	[-0.15;0.40]	0.377	0.371	116	-0.748	[-1.87;0.38]	0.191	0.185
Recent oral corticosteroid treatment	241	0.023	[-0.16;0.21]	0.807	0.806	125	0.103	[-0.13;0.34]	0.385	0.378	116	-0.134	[-0.45;0.19]	0.409	0.402

Table S2: Association between Shannon Diversity Index and clinical covariates in all children and within each age group.

Age group	ASV	Phylum	Family	Genus	Lowest known taxonomy	Mean Relative Abundance	Presence	Log2FC	P-value	Q-value
All children	ASV79	Patescibacteria			Saccharimonadales	0.0022	0.23	1.02	0.014	0.53
All children	ASV41	Bacteroidetes	Prevotellaceae	Prevotella_6	Prevotella_6_salivae	0.0034	0.39	0.83	0.005	0.53
All children	ASV125	Firmicutes	Ruminococcaceae	Ruminococcaceae_UCG-014	Ruminococcaceae_UCG-014	0.0006	0.25	0.75	0.028	0.53
All children	ASV28	Firmicutes	Carnobacteriaceae	Granulicatella	Granulicatella_elegans	0.0056	0.38	-0.68	0.026	0.53
All children	ASV68	Proteobacteria	Pasteurellaceae	Haemophilus	Haemophilus	0.0015	0.37	-0.66	0.038	0.53
All children	ASV56	Firmicutes	Streptococcaceae	Streptococcus	Streptococcus	0.0017	0.48	-0.58	0.017	0.53
All children	ASV14	Proteobacteria	Pasteurellaceae	Haemophilus	Haemophilus	0.0140	0.57	-0.52	0.033	0.53
All children	ASV4	Firmicutes	Veillonellaceae	Veillonella	Veillonella	0.0755	0.66	0.45	0.050	0.62
All children	ASV2	Firmicutes	Streptococcaceae	Streptococcus	Streptococcus	0.0862	0.84	-0.43	0.037	0.53
School-age	ASV11	Proteobacteria	Pasteurellaceae	Actinobacillus	Actinobacillus	0.0169	0.27	-1.58	0.006	0.36
School-age	ASV29	Proteobacteria	Pasteurellaceae	Actinobacillus	Actinobacillus	0.0043	0.20	-1.22	0.025	0.40
School-age	ASV28	Firmicutes	Carnobacteriaceae	Granulicatella	Granulicatella_elegans	0.0031	0.38	-1.17	0.016	0.40
School-age	ASV68	Proteobacteria	Pasteurellaceae	Haemophilus	Haemophilus	0.0012	0.37	-1.13	0.006	0.36
School-age	ASV94	Firmicutes			Lactobacillales	0.0008	0.40	-0.85	0.026	0.40
School-age	ASV56	Firmicutes	Streptococcaceae	Streptococcus	Streptococcus	0.0012	0.48	-0.81	0.026	0.40
School-age	ASV2	Firmicutes	Streptococcaceae	Streptococcus	Streptococcus	0.0828	0.84	-0.69	0.015	0.40
Pre-school	ASV79	Patescibacteria			Saccharimonadales	0.0023	0.23	1.35	0.021	0.73
Pre-school	ASV41	Bacteroidetes	Prevotellaceae	Prevotella_6	Prevotella_6_salivae	0.0025	0.39	1.08	0.027	0.73
Pre-school	ASV36	Firmicutes	Lachnospiraceae	Lachnoanaerobaculum	Lachnoanaerobaculum	0.0056	0.41	0.97	0.021	0.73
Pre-school	ASV7	Bacteroidetes	Prevotellaceae	Prevotella_7	Prevotella_7_melaninogenica	0.0344	0.60	0.74	0.047	0.73

Table S3: Results from differential abundance analyses of OTUs between severe and mild/moderate cases, for all children together and stratified by age group. Differences were tested using the feature model from metagenomeSeq. Only nominally significant ASVs are shown, but no ASVs

were significant after FDR correction. Presence is the fraction of samples where an ASV was found. Log2FC is log2 fold change, an effect size with a positive value indicating association with school-age children. N=241.

ASV	Phylum	Family	Genus	Lowest known taxonomy	Mean Relative Abundance	Presence	Log2FC	P-value	Q-value
ASV30	Bacteroidetes	Prevotellaceae	Alloprevotella	Alloprevotella	0.0077	0.46	-1.22	6.61×10^{-05}	**0.007**
ASV47	Firmicutes	Veillonellaceae	Megasphaera	Megasphaera_micronuciformis	0.0030	0.24	1.29	0.00034	**0.019**
ASV32	Actinobacteria	Actinomycetaceae	Actinomyces	Actinomyces_graevenitzii	0.0052	0.39	0.82	0.00155	**0.039**
ASV5	Proteobacteria	Neisseriaceae	Neisseria	Neisseria	0.0440	0.54	-0.84	0.00176	**0.039**
ASV117	Bacteroidetes	Porphyromonadaceae	Porphyromonas	Porphyromonas	0.0004	0.13	-1.70	0.00184	**0.039**
ASV150	Unknown	Unknown	Unknown	Bacteria	0.0003	0.25	-0.98	0.00208	**0.039**
ASV39	Bacteroidetes	Prevotellaceae	Prevotella_7	Prevotella_7	0.0022	0.20	1.43	0.00325	**0.047**
ASV4	Firmicutes	Veillonellaceae	Veillonella	Veillonella	0.0755	0.66	0.65	0.00336	**0.047**
ASV104	Bacteroidetes	Porphyromonadaceae	Porphyromonas	Porphyromonas	0.0007	0.13	-1.70	0.00411	0.051
ASV9	Firmicutes	Family_XI	Gemella	Gemella	0.0291	0.64	-0.56	0.00527	0.059
ASV166	Firmicutes	Erysipelotrichaceae	Solobacterium	Solobacterium_moorei	0.0005	0.14	1.48	0.00946	0.092
ASV56	Firmicutes	Streptococcaceae	Streptococcus	Streptococcus	0.0017	0.48	-0.61	0.01019	0.092
ASV51	Firmicutes	Lachnospiraceae	Lachnoanaerobaculum	Lachnoanaerobaculum_cf.	0.0026	0.39	-0.62	0.01106	0.092
ASV28	Firmicutes	Carnobacteriaceae	Granulicatella	Granulicatella_elegans	0.0056	0.38	-0.80	0.01150	0.092
ASV13	Bacteroidetes	Porphyromonadaceae	Porphyromonas	Porphyromonas_pasteri	0.0134	0.60	-0.63	0.01641	0.117
ASV214	Firmicutes	Streptococcaceae	Streptococcus	Streptococcus	0.0002	0.12	-1.31	0.01691	0.117
ASV127	Firmicutes	Ruminococcaceae	Ruminococcaceae_UCG-014	Ruminococcaceae_UCG-014	0.0007	0.17	-1.15	0.01855	0.117
ASV14	Proteobacteria	Pasteurellaceae	Haemophilus	Haemophilus	0.0140	0.57	-0.56	0.02044	0.117
ASV96	Fusobacteria	Leptotrichiaceae	Streptobacillus	Streptobacillus	0.0016	0.14	-1.50	0.02113	0.117
ASV33	Firmicutes	Veillonellaceae	Veillonella	Veillonella	0.0063	0.34	0.67	0.02224	0.117
ASV3	Proteobacteria	Pasteurellaceae	Haemophilus	Haemophilus	0.0519	0.68	-0.48	0.02229	0.117
ASV24	Actinobacteria	Micrococcaceae	Rothia	Rothia_mucilaginosa	0.0123	0.37	0.65	0.02378	0.117
ASV210	Firmicutes	Unknown	Unknown	Bacillales	0.0002	0.13	-1.09	0.02400	0.117
ASV41	Bacteroidetes	Prevotellaceae	Prevotella_6	Prevotella_6_salivae	0.0034	0.39	0.62	0.03024	0.136
ASV110	Bacteroidetes	Porphyromonadaceae	Porphyromonas	Porphyromonas	0.0008	0.12	-1.77	0.03042	0.136
ASV7	Bacteroidetes	Prevotellaceae	Prevotella_7	Prevotella_7_melaninogenica	0.0360	0.60	0.47	0.03594	0.155

ASV124	Firmicutes	Lachnospiraceae	Oribacterium	Oribacterium_parvum	0.0006	0.17	-0.79	0.03856	0.160
ASV81	Firmicutes	Streptococcaceae	Streptococcus	Streptococcus	0.0011	0.37	-0.54	0.04281	0.171

Table S4: Results from differential abundance analyses of OTUs between school-age and pre-school children, corresponding to fig. 4. Differences were tested using the feature model from metagenomeSeq. Only nominally significant ASVs are shown, some of which remained significant after FDR correction (bold). Presence is the fraction of samples where an ASV was found. Log2FC is log fold change, an effect size with a positive value indicating association with school-age children, and a negative value indicating association with pre-school children. N=241.

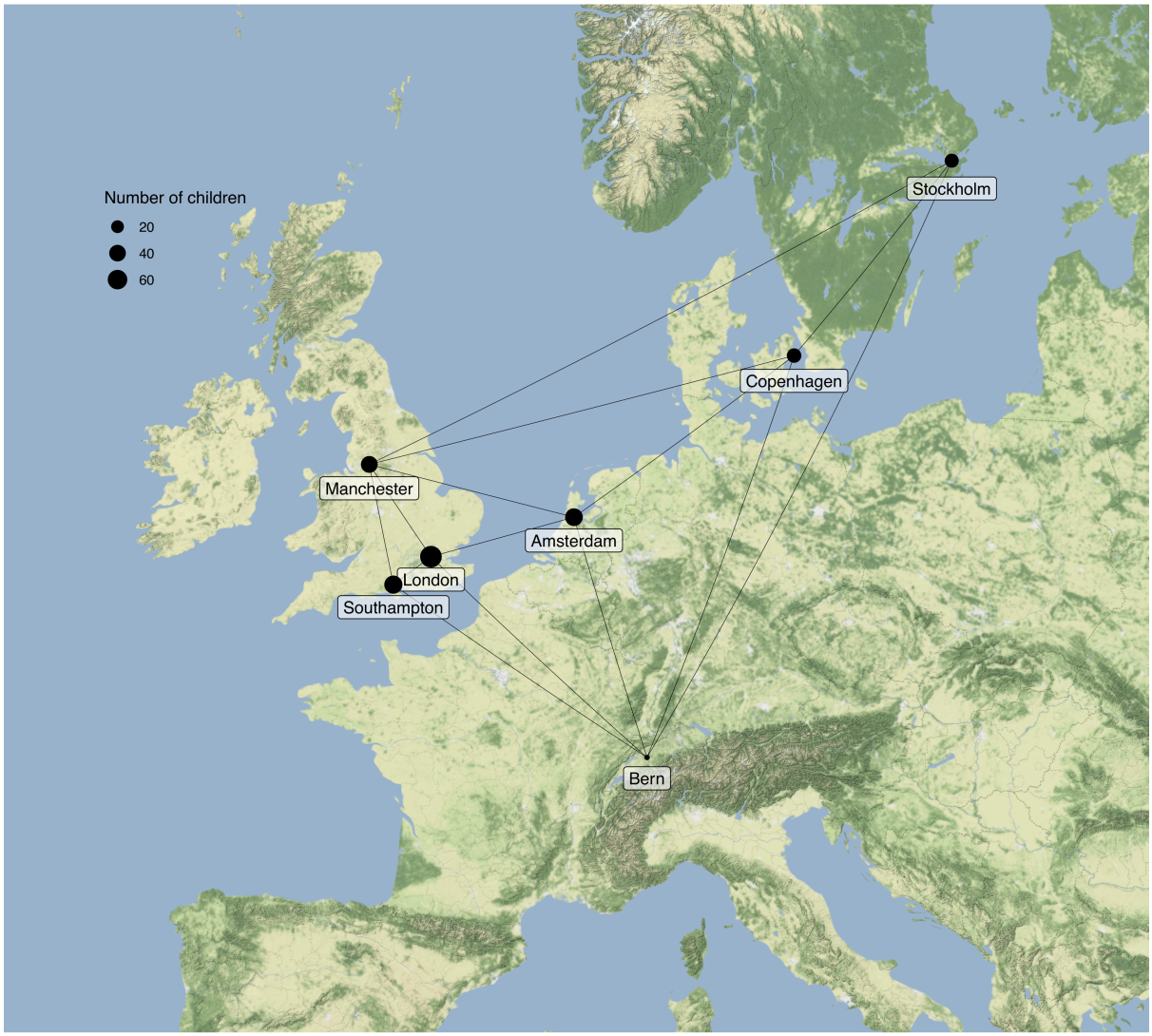

Fig S1: Geographical distribution and distribution of children between study centers.

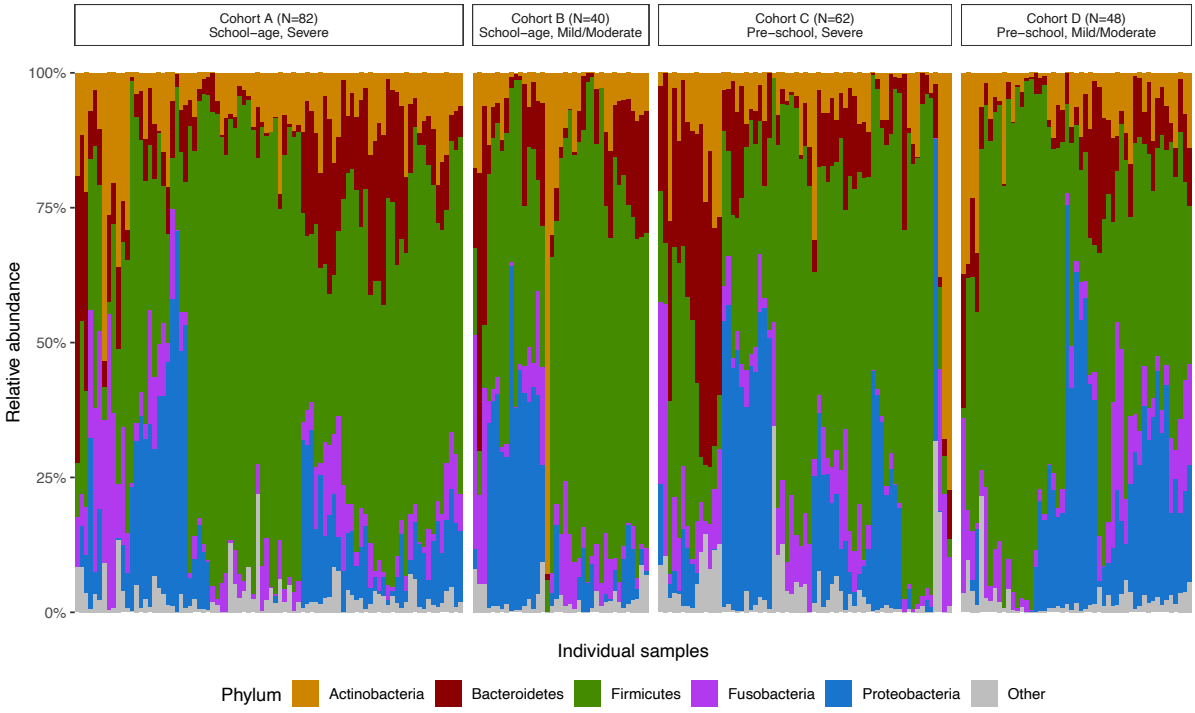

Fig S2: Phylum-level relative abundances for each sample, grouped by cohort, ordered by hierarchical clustering within each cohort. N=241.

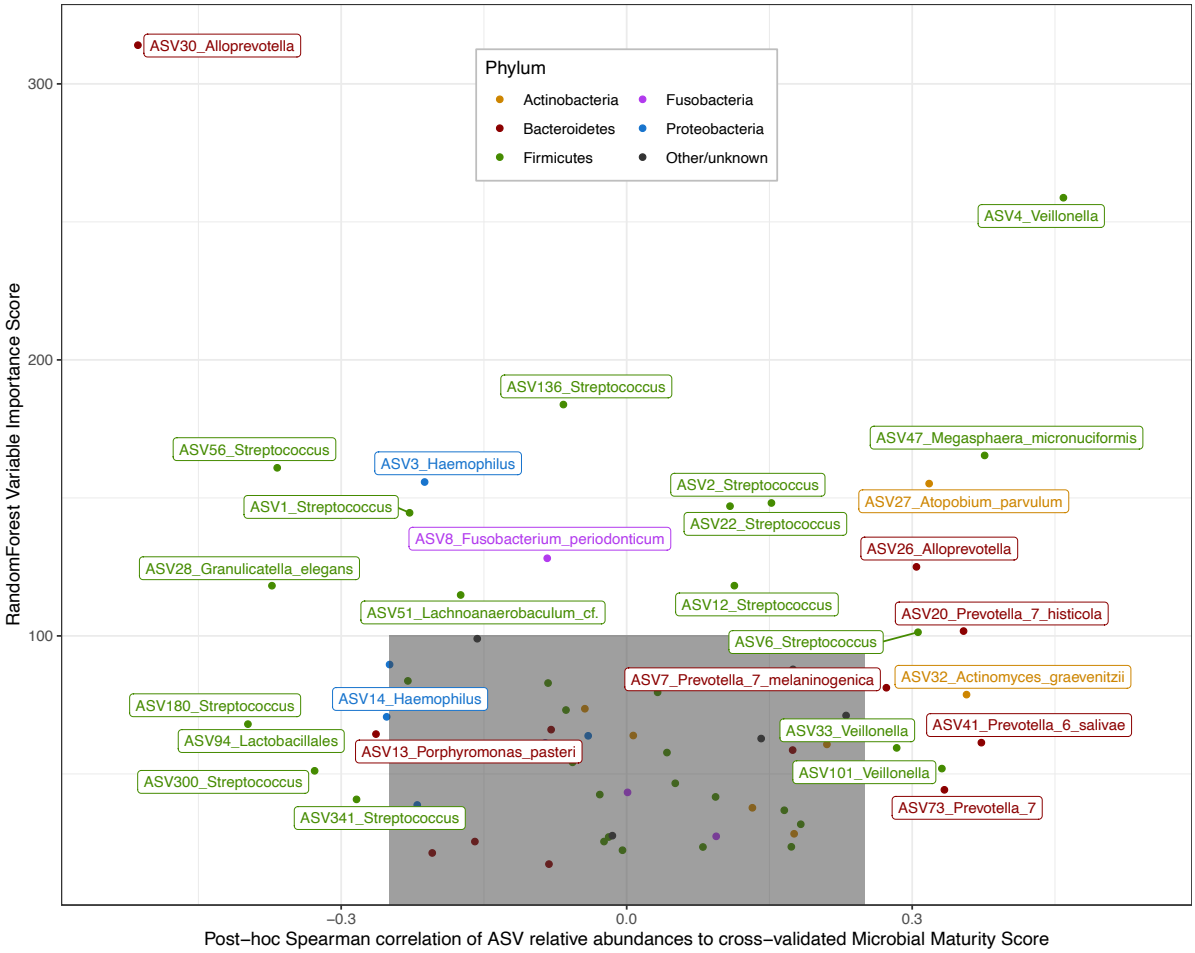

Fig S3: An overview of the most important taxa in the Microbial Maturity Score random forest model, as assessed by the Variable Importance Score (y axis) and the directionality and magnitude of correlation with the score for each taxon (x-axis).