ORIGINAL ARTICLE

Early nasal microbiota and acute respiratory infections during the first years of life

Laura Toivonen,[1,2] Kohei Hasegawa,[1] Matti Waris,[3] Nadim J Ajami,[4] Joseph F Petrosino,[4] Carlos A Camargo Jr,[1] Ville Peltola[2]

[1]Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA
[2]Department of Paediatrics and Adolescent Medicine, Turku University Hospital and University of Turku, Turku, Finland
[3]Virology Unit, Institute of Biomedicine, University of Turku, Turku, Finland
[4]Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, USA

Correspondence to
Dr Laura Toivonen, Department of Emergency Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA; ltoivonen@mgh.harvard.edu

ABSTRACT

Background Emerging evidence shows that airway microbiota may modulate local immune responses, thereby contributing to the susceptibility and severity of acute respiratory infections (ARIs). However, there are little data on the longitudinal relationships between airway microbiota and susceptibility to ARIs in children.

Objective We aimed to investigate the association of early nasal microbiota and the subsequent risk of ARIs during the first years of life.

Methods In this prospective population-based birth-cohort study in Finland, we followed 839 healthy infants for ARIs from birth to age 24 months. Nasal microbiota was tested using 16S rRNA gene sequencing at age 2 months. We applied an unsupervised clustering approach to identify early nasal microbiota profiles, and examined the association of profiles with the rate of ARIs during age 2–24 months.

Results We identified five nasal microbiota profiles dominated by *Moraxella*, *Streptococcus*, *Dolosigranulum*, *Staphylococcus* and *Corynebacteriaceae*, respectively. Incidence rate of ARIs was highest in children with an early *Moraxella*-dominant profile and lowest in those with a *Corynebacteriaceae*-dominant profile (738 vs 552/100 children years; unadjusted incidence rate ratio (IRR), 1.34; 95% CI 1.16 to 1.54; p < 0.001). After adjusting for nine potential confounders, the *Moraxella*-dominant profile-ARI association persisted (adjusted IRR (aIRR), 1.19; 95% CI 1.04 to 1.37; p = 0.01). Similarly, the incidence rate of lower respiratory tract infections (a subset of all ARIs) was significantly higher in children with an early *Moraxella*-dominant profile (aIRR, 2.79; 95% CI 1.04 to 8.09; p = 0.04).

Conclusion *Moraxella*-dominant nasal microbiota profile in early infancy was associated with an increased rate of ARIs during the first 2 years of life.

Key messages

What is the key question?
► Are early infant airway microbiota profiles associated with susceptibility to acute respiratory infections (ARIs) in young children?

What is the bottom line?
► In this prospective, population-based birth-cohort study, we identified five distinct nasal microbiota profiles in healthy 2-month-old infants, and found that the incidence rate of ARIs was highest in children with an early *Moraxella*-dominant nasal microbiota profile and lowest in those with a *Corynebacteriaceae*-dominant profile.

Why read on?
► This large prospective cohort study shows that early infant airway microbiota profiles may play an important role in susceptibility to ARIs in children during the first years of life.

INTRODUCTION

Acute respiratory infection (ARI) is a major cause of morbidity in infants and young children with a mean annual rate of 5–7 infections per child.[1–3] Clinical manifestations of ARI range from the self-limited common cold to severe (and even fatal) infections. Lower respiratory tract infections (LRTIs), including pneumonia and bronchiolitis, are the most frequent cause of hospitalisation in children, accounting for 50%–60% of hospitalisations before age 5 years.[4] Some children are more susceptible to ARIs but known risk factors (such as age, household siblings, daycare and prematurity) do not completely explain the variability in the incidence and severity of ARIs.

The recent advent of high-throughput sequencing methods has identified diverse microbial communities in the airways. Emerging evidence shows that airway microbiota may modulate local immune responses[5 6] thereby contributing to the susceptibility[7–9] and severity of ARIs.[10–12] Previous studies investigating the association between airway microbiota and susceptibility to ARIs in children have been small,[7 9 11] conducted in selected subjects (eg, infants at high risk for atopy),[11] have not investigated the rates of all types of ARIs,[7 9] or the airway microbiota has been concurrent with ARI.[8 10 12 13] Furthermore, studies have reported seemingly contradictory associations between airway microbiota and ARI incidence, specifically for *Moraxella*.[7 9 11 14] It remains to be elucidated how airway microbiota in the early infancy relates to the subsequent risks of ARIs (and, more specifically, LRTIs) in healthy children. Better understanding of this relationship would inform the development of preventive measures for children at high risk for ARIs.

To address this knowledge gap, we conducted a prospective population-based birth-cohort study in healthy children to identify nasal microbiota profiles in the early infancy and to examine the association of these profiles with the rates of ARIs during age 2–24 months.

METHODS

Study design and conduct

We analysed data from the prospective, population-based birth-cohort study—the Steps to the Healthy Development and Well-being of Children (STEPS Study)—where families of Finnish infants born in January 2008 through April 2010 were enrolled during pregnancy or soon after birth.[15] As part of the STEPS study, 923 infants were enrolled in a follow-up for ARIs from birth to age 24 months.[1] The details of the study design, setting, subjects, testing and analysis may be found in the online supplement. Briefly, parents were asked to record all respiratory and other symptoms and physician visits with associated diagnoses and treatments into a daily symptom diary. Families were instructed to visit the study clinic during ARIs if they felt that an evaluation by a physician was needed. At the study clinic, children were examined by a study physician and clinical findings were documented using a structured form. Data on outpatient and emergency department visits at hospitals and hospitalisations from birth to age 24 months were retrieved from medical records of the Hospital District of Southwest Finland. Patient demographics, family history, prenatal, perinatal, and postnatal history, and environmental information were collected from the National Birth Registry and by structured questionnaires. Infants were examined at the study clinic at age 2 months at healthy state and nasal swab specimens were collected by trained study personnel using flocked nylon swabs (Copan, Brescia, Italy). Any current respiratory symptoms were documented. Parents of participating children gave their written, informed consent. The study complies with the Declaration of Helsinki.

Respiratory virus detection

The nasal swab samples were stored at −80°C after the collection. Swabs were suspended in phosphate-buffered saline and analysed by using real-time, quantitative PCR for rhinovirus, human enteroviruses and respiratory syncytial virus (RSV) as described earlier.[16 17] All specimens collected during influenza seasons were analysed by reverse transcriptase PCR for influenza A and B viruses.[18]

16S rRNA gene sequencing of nasal airway microbiota

16S rRNA gene sequencing methods were adapted from the methods developed for the National Institutes of Health Human Microbiome Project.[19 20] The details of the methods are described in the online supplemental methods. Briefly, bacterial genomic DNA was extracted with automated MagNA Pure 96 System (Roche Life Science). The 16S rDNA V4 region was amplified by PCR and sequenced on the MiSeq platform (Illumina, San Diego, California, USA) using the 2×250 bp paired-end protocol yielding pair-end reads that overlap almost completely. Sequencing reads were merged using USEARCH V.7.0.1090.[21] Samples with suboptimal amounts of sequencing reads were resequenced to ensure that most bacterial taxa were encompassed in our analyses. 16S rRNA gene sequences were clustered into operational taxonomic units (OTUs) at a similarity cut-off value of 97% using the UPARSE algorithm.[22] OTUs were determined by mapping the centroids to the SILVA database[23] containing only the 16S V4 region to determine taxonomies. Rarefaction curves of bacterial OTUs were constructed using sequence data. We utilised multiple quality control measures, including the use of non-template controls in the microbial DNA extraction, 16S rRNA gene amplification and amplicon sequencing processes. The details of the quality control measures are described in the online supplemental methods. Analyses were conducted at the genus level.

Outcomes

The primary outcome was the incidence rate of all ARIs (including upper ARIs and LRTIs). The secondary outcomes were (1) the rate of LRTIs during age 2–24 months and (2) recurrent wheezing by age 24 months. An ARI was defined as the presence of rhinitis or cough (with or without fever or wheezing) documented in the symptom diary by the parents, or as any ARI diagnosed by a physician.[1] The LRTIs (eg, bronchiolitis, pneumonia) were diagnosed by physicians. Recurrent wheezing was defined as ≥3 wheezing episodes during age 2–24 months as reported by the parents in the diary or diagnosed by a physician.

Statistical analysis

To identify nasal airway microbiota profiles at age 2 months, we performed unsupervised clustering by partitioning around medoids[24] based on the Bray-Curtis distance at the genus level.[25] The gap statistic method was used to select the number of profiles for the data (online supplementary figure E1).[26] The patient characteristics and outcomes were compared between the derived nasal microbiota profiles, using the χ^2, Fisher's exact and Kruskal-Wallis rank sum test as appropriate. The relative abundance of bacterial genera was compared between the profiles by using analysis of variance adjusting for multiple comparisons with the use of the Benjamini-Hochberg false discovery rate method.[27] Incidence rates of all ARIs and LRTIs were analysed by using unadjusted and multivariable-adjusted negative binomial regression models using subject-level data with natural logarithm of the individual completed follow-up time from age 2 month nasal sample to age 24 months as an offset. Risk of recurrent wheezing was analysed by modelling the time to event (ie, time from age 2 month nasal sample to the third wheezing episode) using Kaplan-Meier method, log-rank test and Cox proportional hazards models. We constructed a multivariable model adjusting for nine patient-level potential confounders (age, sex, household siblings, parental asthma, birth by caesarean section, breast feeding, history of systemic antibiotic use, respiratory symptoms at the time of nasal sampling and respiratory virus in age 2 month sample). These covariates were selected based on clinical plausibility and a priori knowledge.[9 11 28–30] In the sensitivity analyses, we repeated the analysis by fitting a negative binomial regression model with (1) excluding infants with respiratory symptoms at the time of age 2 month nasal sampling, (2) modelling the rate of non-LRTIs (ie, ARIs not including physician-diagnosed LRTIs) during age 2–24 months as the outcome and (3) modelling the number of days with ARI symptoms during age 2–24 months as the outcome. We reported p values as two tailed, with $p < 0.05$ considered statistically significant. Data were analysed using R version 3.5.1 with the *phyloseq* package for the microbiota analysis.[31]

RESULTS

Study population

Of 923 infants enrolled into the ARI follow-up subcohort in the STEPS Study, 6 (0.7%) withdrew from the study. A total of 900 (97.5%) infants underwent nasal swab sampling at age 2 months (online supplementary figure E2). 16S rRNA gene sequencing of the nasal airway samples obtained 16 970 216 high-quality merged sequences, of which 16 760 908 (98.8%) were mapped to the 16S reference database. Of the 900 nasal swabs, 839 (93.2%) met the quality control requirements and

had sufficient sequence depth for 16S rRNA gene sequencing (rarefaction cut-off, 1945 reads per sample) and were included in the current analyses. Among the analytic cohort, the median age at the time of nasal sampling was 2.6 (IQR 2.4–2.7) months, 52.8% were boys and 4.1% were born premature. The median age at the end of follow-up was 23.9 (IQR 17.0–24.0) months. There were no significant differences in the characteristics or clinical outcomes between infants in the analytical cohort and those without qualified microbiota data (non-analytical cohort) (online supplementary table E1).

Nasal microbiota profiles

The 16S rRNA gene sequencing of nasal microbiota at age 2 months detected 29 phyla and 911 genera. The dominant phyla were *Firmicutes* (52.2%), *Proteobacteria* (32.6%) and *Actinobacteria* (14.1%). The nasal airway microbiota was dominated by four genera: *Moraxella* (26.1%), *Dolosigranulum* (16.6%), *Staphylococcus* (16.0%) and *Streptococcus* (15.2%), followed by *Corynebacteriaceae* genus 1 (7.6%) and genus 2 (5.0%) and *Neisseriaceae* genus 1 (3.0%).

Clustering of nasal airway microbiota identified five distinct microbiota profiles in infants at age 2 months: (1) *Moraxella*-dominant (30.4%), (2) *Streptococcus*-dominant (22.4%), (3) *Dolosigranulum*-dominant (22.4%), (4) *Staphylococcus*-dominant (17.9%) and (5) *Corynebacteriaceae*-dominant (6.9%) profiles (table 1 and figure 1). The bacterial richness and alpha diversity (Shannon Index) were lowest in infants with a *Moraxella*-dominant profile (both p<0.001). Some of patient characteristics differed by the nasal microbiota profiles (table 2). For example, infants with a *Moraxella*-dominant profile were more likely to have household siblings and mild respiratory symptoms at the age 2 month visit (both p<0.05).

Table 1 Richness, alpha diversity and relative abundance by nasal airway microbiota profiles at age 2 months

	Nasal microbiota profile at age 2 months					
	Moraxella dominant, n=255 (30.4%)	Streptococcus dominant, n=188 (22.4%)	Dolosigranulum dominant, n=188 (22.4%)	Staphylococcus dominant, n=150 (17.9%)	Corynebacteriaceae dominant, n=58 (6.9%)	P value
Richness						
Number of OTUs, median (IQR)	10 (6–16)	33 (25–44)	19 (14–26)	24 (16–33)	25 (20–37)	<0.001
Alpha diversity						
Shannon Index, median (IQR)	0.45 (0.23–0.75)	1.66 (1.23–2.07)	1.01 (0.76–1.31)	0.99 (0.72–1.35)	1.15 (0.79–1.47)	<0.001
Relative abundance of 10 most abundant genera						
Mean (SD)						
Moraxella	0.84 (0.16)	0.01 (0.03)	0.01 (0.06)	0.01 (0.05)	0.00 (0.01)	<0.001*
Dolosigranulum	0.05 (0.08)	0.02 (0.06)	0.63 (0.18)	0.01 (0.05)	0.04 (0.09)	<0.001*
Staphylococcus	0.02 (0.04)	0.10 (0.10)	0.03 (0.08)	0.68 (0.19)	0.05 (0.07)	<0.001*
Streptococcus	0.04 (0.08)	0.45 (0.23)	0.07 (0.08)	0.11 (0.10)	0.09 (0.09)	<0.001*
Corynebacteriaceae genus 1	0.02 (0.03)	0.01 (0.04)	0.15 (0.12)	0.02 (0.07)	0.46 (0.32)	<0.001*
Corynebacteriaceae genus 2	0.01 (0.02)	0.09 (0.17)	0.03 (0.07)	0.05 (0.10)	0.17 (0.25)	<0.001*
Neisseriaceae genus 1	0.01 (0.06)	0.06 (0.18)	0.02 (0.07)	0.02 (0.09)	0.06 (0.19)	<0.001*
Gemella	0.00 (0.01)	0.04 (0.05)	0.01 (0.01)	0.02 (0.04)	0.02 (0.06)	<0.001*
Haemophilus	0.01 (0.06)	0.03 (0.11)	0.01 (0.02)	0.01 (0.04)	0.03 (0.13)	0.05*
Veillonella	0.00 (0.00)	0.02 (0.04)	0.00 (0.01)	0.01 (0.02)	0.01 (0.02)	<0.001*
Range						
Moraxella	0.35–1.00	0.00–0.34	0.00–0.42	0.00–0.43	0.00–0.09	
Dolosigranulum	0.00–0.43	0.00–0.37	0.19–0.94	0.00–0.34	0.00–0.38	
Staphylococcus	0.00–0.28	0.00–0.47	0.00–0.44	0.20–0.99	0.00–0.28	
Streptococcus	0.00–0.48	0.00–0.96	0.00–0.44	0.00–0.39	0.00–0.39	
Corynebacteriaceae genus 1	0.00–0.21	0.00–0.39	0.00–0.51	0.00–0.43	0.00–0.97	
Corynebacteriaceae genus 2	0.00–0.14	0.00–0.78	0.00–0.46	0.00–0.57	0.00–0.94	
Neisseriaceae genus 1	0.00–0.50	0.00–0.92	0.00–0.53	0.00–0.64	0.00–0.78	
Gemella	0.00–0.09	0.00–0.48	0.00–0.06	0.00–0.43	0.00–0.47	
Haemophilus	0.00–0.64	0.00–0.99	0.00–0.23	0.00–0.47	0.00–0.97	
Veillonella	0.00–0.03	0.00–0.25	0.00–0.09	0.00–0.09	0.00–0.08	

*Benjamini-Hochberg false-discovery rate adjusted p value accounting for multiple comparisons.
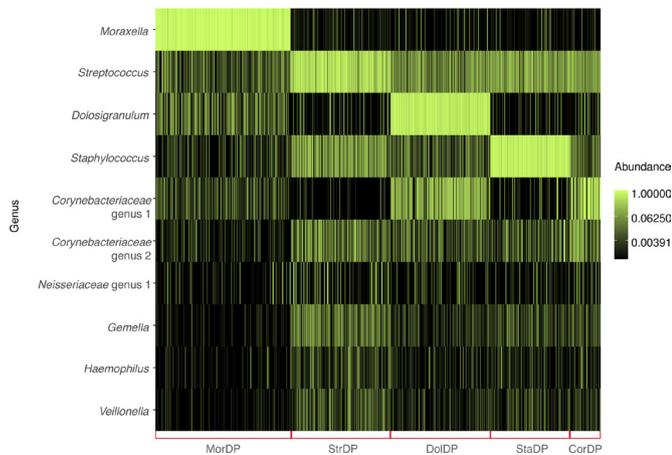OTU, operational taxonomic unit.

Figure 1 Clustering and composition of nasal microbiota in 839 healthy infants at age 2 months. Nasal microbiota profiles were clustered using the partitioning around medoids clustering method based on Bray-Curtis distance. The optimal number of clusters was identified by the gap statistics. Red brackets indicate five microbiota profiles: *Moraxella*-dominant profile (MorDP); *Streptococcus*-dominant profile (StrDP); *Dolosigranulum*-dominant profile (DolDP); *Staphylococcus*-dominant profile (StaDP) and *Corynebacteriaceae*-dominant profile (CorDP). Ten most abundant genera are shown to describe the composition of samples within the microbiota profiles.

NASAL MICROBIOTA PROFILES AND INCIDENCE RATES OF ARIS

A total of 7810 episodes of ARIs (including upper ARIs and LRTIs) were documented during age 2–24 months with an incidence rate of 656 (95%CI 635 to 679) per 100 children years. During these 7810 ARI episodes, a total of 4055 physician visits and 67 hospitalisations for ARIs were documented. Children were evaluated by a physician and diagnosed with an ARI during 3233 (41%) ARIs and 59% of the ARIs were based on diary data. Altogether 285 episodes of LRTI were documented with an incidence rate of 24 (95% CI 20 to 29) per 100 children years. The duration of ARI symptoms was 4982 (95% CI 4698 to 5288) days per 100 children years. Recurrent wheezing was documented in 11.6% of children.

The incidence rate of ARIs differed between the nasal microbiota profiles. Children with an early *Moraxella*-dominant profile had a significantly higher incidence rate of ARIs during age 2–24 months compared with those with a *Corynebacteriaceae*-dominant profile (738 vs 552 per 100 children years; unadjusted incidence rate ratio (IRR), 1.34; 95% CI 1.16 to 1.54; p<0.001, table 3). After adjusting for nine potential confounders, the significant association persisted (adjusted IRR (aIRR), 1.19; 95% CI 1.04 to 1.37; p=0.01). Likewise, the incidence rate of LRTIs was significantly higher in children with an early *Moraxella*-dominant (aIRR, 2.79; 95% CI 1.04 to 8.09; p=0.04) and *Staphylococcus*-dominant (aIRR, 2.90; 95% CI 1.07 to 8.49; p=0.04) profiles, compared with those with a *Corynebacteriaceae*-dominant profile (table 3). In the sensitivity analyses, excluding infants with respiratory symptoms at the time of age 2 months nasal sampling (n=135), these associations persisted (online supplementary table E2). The rate of non-LRTIs (ie, ARIs not including physician-diagnosed LRTIs) was also significantly higher in children with an early *Moraxella*-dominant profile (online supplementary table E3). Also, the duration of days with ARI symptoms was significantly higher in children with an early *Moraxella*-dominant profile (aIRR, 1.61;

95% CI 1.25 to 2.06; p<0.001) compared with those with an early *Corynebacteriaceae*-dominant profile (online supplementary table E4). With regard to the risk of recurrent wheezing, children with an early *Moraxella*-dominant profile had non-significantly higher risk compared with those with an early *Streptococcus*-dominant profile (adjusted HR, 1.82; 95% CI 0.94 to 3.52; p=0.08; table 4 and online supplementary figure E3). Full results of the above models are shown in online supplementary tables E5-E7.

DISCUSSION

In this prospective, population-based, birth-cohort study of 839 healthy children, we identified five nasal airway microbiota profiles at age 2 months with differential risks of subsequent ARIs. Indeed, the incidence rate of all ARIs was highest in children with an early *Moraxella*-dominant profile and lowest in those with an early *Corynebacteriaceae*-dominant profile. Similarly, incidence rates of LRTIs were highest in children with early *Moraxella*-dominant and *Staphylococcus*-dominant profiles. Our data build on previous smaller studies (n<235)[9 11] reporting that certain early nasal microbiota profiles are associated with increased susceptibility to ARIs ('risk microbiota'), and extend them by demonstrating robust longitudinal microbiota-ARI associations in a large well-characterised cohort study.

Recent studies have suggested that early *Moraxella*-dominant airway microbiota may be associated with increased risk of ARIs but the findings have not been consistent. In a cohort of 234 Australian infants with a high risk of atopy, Teo and colleagues reported that early *Moraxella*-dominant nasopharyngeal microbiota was associated with an earlier first upper respiratory infection.[11] Similarly, in a cohort of 112 children in the Netherlands, Bosch and colleagues reported an association between enrichment of *Moraxella* in early life in children and a high frequency of ARIs during the first year of life.[9] In contrast, in an analysis of trial data in the Netherlands (n=60), Biesbroek and colleagues found that infants with *Moraxella*-dominant nasopharyngeal microbiota had a lower risk of subsequent ARI.[7] However, the study by Biesbroek and colleagues used retrospective parent-reported ARI as a dichotomised outcome, and the rate of ARIs—a frequent outcome in infants and young children—was not assessed. These differences in methodology may explain the inconsistent results. Earlier studies have also linked detection and dominance of airway microbiota by *Moraxella* at the time of ARI to the symptoms[11 13] and severity[11] of ARIs in children. However, we show that the longitudinal association between early *Moraxella*-dominant airway microbiota and increased risk of ARIs and LRTIs persisted when excluding infants with respiratory symptoms at the time of nasal sampling. Based on our results and previous studies, colonisation with *Moraxella* in the airway is likely to play an important role in the susceptibility to ARIs in young children.

We also demonstrated that children with an early *Moraxella*-dominant nasal microbiota had the highest rate of LRTIs. Previous data on early airway microbiota assessed by next-generation sequencing methods and rates of LRTIs are scarce. The Australian study found that *Moraxella* enrichment at the time of ARI was significantly associated with higher risks of LRTI.[11] In contrast, in case-control studies, *Moraxella*-dominant airway microbiota has been found to be associated with lower risks of severe bronchiolitis[8] and RSV hospitalisation,[12] and not to be associated with pneumonia in children.[13] However, in these studies, the nasal microbiota was investigated at the time of severe infection (with possible dysbiosis), which may explain

Table 2 Characteristics and clinical outcomes in healthy children by nasal microbiota profiles at age 2 months

Characteristic	Nasal microbiota profile at age 2 months					
	Moraxella dominant, n=255 (30.4%)	Streptococcus dominant, n=188 (22.4%)	Dolosigranulum dominant, n=188 (22.4%)	Staphylococcus dominant, n=150 (17.9%)	Corynebacteriaceae dominant, n=58 (6.9%)	P value
Age at 2 months scheduled visit, months, median (IQR)	2.6 (2.4–2.7)	2.5 (2.4–2.7)	2.6 (2.4–2.7)	2.5 (2.3–2.7)	2.7 (2.4, 2.8)	0.07
Male sex	135 (52.9)	101 (53.7)	105 (55.9)	78 (52.0)	24 (41.4)	0.43
Household sibling	188 (73.7)	53 (28.2)	66 (35.1)	34 (22.7)	9 (15.5)	<0.001
Maternal asthma	20 (7.8)	16 (8.5)	14 (7.4)	8 (5.3)	6 (10.3)	0.71
Parental asthma	31 (12.2)	31 (16.5)	23 (12.2)	15 (10.0)	9 (15.5)	0.44
Maternal smoking during pregnancy	9 (3.5)	10 (5.3)	15 (8.0)	7 (4.7)	5 (8.6)	0.22
Birth by Caesarean section	21 (8.2)	28 (14.9)	22 (11.7)	31 (20.7)	9 (15.5)	0.008
Prematurity (<37 weeks)	6 (2.4)	9 (4.8)	6 (3.2)	8 (5.3)	5 (8.6)	0.16
Small for gestational age	4 (1.6)	2 (1.1)	5 (2.7)	3 (2.0)	2 (3.4)	0.60
Intrapartum antibiotics	27 (10.6)	25 (13.3)	28 (14.9)	19 (12.7)	6 (10.3)	0.70
Breast feeding during the first 2 months of life*	170 (66.7)	129 (68.6)	128 (68.1)	103 (68.7)	41 (70.7)	0.68
Systemic antibiotic use before age 2 months nasal sampling	40 (15.7)	42 (22.3)	24 (12.8)	30 (20.0)	5 (8.6)	0.03
Age 2 month scheduled visit						
Mild respiratory symptoms	72 (28.2)	22 (11.7)	22 (11.7)	13 (8.7)	6 (10.3)	<0.001
Respiratory virus in nasal sample†	61 (23.9)	23 (12.2)	35 (18.6)	8 (5.3)	7 (12.1)	<0.001
Age at the end of the follow-up, months, median (IQR)	24.0 (16.3–24.0)	24.0 (20.2–24.0)	24.0 (19.3–24.0)	23.9 (16.5–24.0)	23.2 (14.6–24.0)	0.08
Clinical outcomes during age 2–24 months‡						
Total number of all ARIs	2675	1714	1619	1339	463	
Total number of LRTIs	112	55	52	59	7	
Total duration of follow-up, children years	358.7	262.7	259.2	211.8	82.5	
ARIs, incidence rate/100 children years (95% CI)						
All ARIs	738 (696–782)	643 (599–690)	615 (573–660)	628 (580–680)	552 (485–628)	<0.001
Days with ARI symptoms	5991 (5406–6662)	4947 (4378–5618)	4540 (4027–5142)	4584 (4009–5274)	3180 (2578–3980)	0.002
LRTIs	32 (23–44)	21 (14–33)	19 (13–29)	28 (18–44)	8 (3–20)	0.007
Recurrent wheezing§	33 (12.9)	15 (8.0)	22 (11.7)	17 (11.3)	5 (8.6)	0.60

Data are number (%) of children unless otherwise indicated. Percentages may not equal 100 because of missingness.

*Data on breast feeding available for 653 children.

†Rhinovirus was detected in 120 (14.3%) children, respiratory syncytial virus in 11 (1.3%) and other virus in 5 (0.6%) children.

‡Data on ARIs during age 2–24 months available for 798 children. Incidence rates were calculated using negative binomial distribution and log link with natural logarithm of follow-up time from age 2 months visit to age 24 months as an offset.

§≥3 wheezing episodes at age 2–24 months.

ARI, acute respiratory infection; LRTI, lower respiratory tract infection.

the difference in results. Our finding that early *Moraxella*-dominant airway microbiota may play a role in the risk of developing LRTIs is supported by previous reports using bacterial culture or PCR. By using culture, in the Copenhagen Prospective Study on Asthma in Childhood (COPSAC) cohort of 321 Danish children born to mothers with asthma, early colonisation at age 1 month with *Moraxella catarrhalis* was associated with first wheeze and hospitalisation for wheeze.[32] In a later subanalysis of the same study (n=265), early airway colonisation with *M. catarrhalis* was associated with increased risks of pneumonia and bronchiolitis in the first 3 years of life.[14] Similarly, in a Danish study using PCR (n=218), neonatal nasal colonisation with *Streptococcus pneumoniae*, *Haemophilus influenzae* or *M. catarrhalis* was associated with cough and wheezing episodes in the first year of life.[33] Potential reasons for discrepancies across studies

include differences in patient populations, settings (eg, hospital vs community setting), sampling methods (eg, nasal swab vs nasopharyngeal aspirate) and laboratory techniques (eg, culture vs next-generation sequencing). The validity of our results is buttressed by the use of 16S rRNA gene sequencing of the nasal microbiota from the population-based cohort of healthy children and a sample size that is much larger than any prior study on this topic.

There are several potential mechanisms that explain the observed association of early nasal airway microbiota with susceptibility to ARIs and LRTIs. It is possible that there is a causal relationship—for example, *Moraxella* alters the mucosal immune responses and increases susceptibility to ARIs and LRTIs. Indeed, *M. catarrhalis* colonisation of the airways in asymptomatic neonates has been shown to induce species-specific

Table 3 Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with rate of acute respiratory infections (ARI) and lower respiratory infections (LRTI) during age 2–24 months*

Outcome by nasal microbiota profile	Unadjusted analysis		Multivariable-adjusted analysis	
	Incidence rate ratio (95% CI)	P value	Incidence rate ratio (95% CI)	P value
ARIs				
Moraxella dominant	1.34 (1.16 to 1.54)	<0.001	1.19 (1.04 to 1.37)	0.01
Streptococcus dominant	1.16 (1.00 to 1.35)	0.04	1.14 (0.99 to 1.31)	0.08
Dolosigranulum dominant	1.11 (0.96 to 1.29)	0.15	1.08 (0.94 to 1.24)	0.31
Staphylococcus dominant	1.14 (0.98 to 1.32)	0.10	1.14 (0.98 to 1.31)	0.09
Corynebacteriaceae dominant	Reference	–	Reference	–
LRTIs				
Moraxella dominant	3.89 (1.47 to 11.06)	0.008	2.79 (1.04 to 8.09)	0.04
Streptococcus dominant	2.62 (0.95 to 7.67)	0.07	1.89 (0.69 to 5.55)	0.22
Dolosigranulum dominant	2.38 (0.87 to 6.95)	0.10	1.95 (0.72 to 5.69)	0.19
Staphylococcus dominant	3.42 (1.24 to 10.13)	0.02	2.90 (1.07 to 8.49)	0.04
Corynebacteriaceae dominant	Reference	–	Reference	–

*Incidence rates of ARIs and LRTIs were compared by using unadjusted and multivariable-adjusted negative binomial regression models with follow-up time from age 2 months visit to age 24 months as an offset variable. Multivariable analysis adjusted for nine patient-level covariates (age, sex, household siblings, parental asthma, birth by caesarean section, breast feeding, history of systemic antibiotic use, respiratory symptoms at the time of nasal sampling and respiratory virus in age 2 months sample). *Corynebacteriaceae*-dominant microbiota profile was used as the reference.

stimulation of inflammatory mediators (eg, a mixed T helper cell (TH) type 1/TH2/TH17 response with high levels of interleukin 1b, tumour necrosis factor-α and macrophage inflammatory protein-1β).[5] One possibility is that early *Moraxella*-dominant nasal microbiota may be only a marker of children with increased susceptibility to ARIs or LRTIs. Furthermore, a recent study reported that *Moraxella* interacts with respiratory viruses (eg, rhinovirus), contributing to the symptoms and severity.[34] Children with an early *Moraxella*-dominant nasal microbiota profile were more likely to have household siblings. It is possible that this pattern of nasal microbiota was transferred from the sibling. Nevertheless, in the multivariable models, the early *Moraxella*-dominant nasal microbiota profile was associated with increased rates of ARIs and LRTIs independent from sibling status. Additionally, as the early *Corynebacteriaceae*-dominant profile was associated with the lowest rates of ARIs and LRTIs, it is possible that *Corynebacteriaceae* confer resistance against respiratory pathogens and thus against ARIs ('resilience microbiota'). Previous reports have also linked airway *Corynebacteria* to lower risks of ARIs and wheeze in children.[7–9 13 35] Ultimately, we may learn that these different mechanisms are not mutually exclusive. Our data, in conjunction with the previous studies, suggest a complex interplay between airway microbiota, respiratory viruses and host immune response in the development and morbidity of ARIs.[12 36]

Our study has potential limitations. First, while ARI may involve the lower airways, our data relied on the nasal airway microbiota. However, the nasal sampling is a non-invasive procedure in young infants. Previous studies have also reported a correlation between upper and lower airway microbiota.[37–39] Second, as we used nasal samples of healthy infants with potentially low amounts of bacterial DNA, part of the nasal samples had insufficient sequence depth and were excluded from the analyses. Nonetheless, nasal samples were analysed from >90% of the infants and the analytical cohort did not differ from the infants without microbiota data. Third, the study design precluded us from examining succession of the airway microbiota, and its association with respiratory health in children. To address this question, the cohort is currently being followed longitudinally with multiple specimen collections; we would like to test for nasal microbiota at different time points in childhood. Fourth, the definition of ARI was based on diary in 59% of ARIs. Symptoms of ARI may be subjective with some

Table 4 Unadjusted and multivariable-adjusted associations of nasal microbiota profiles at age 2 months with risk of recurrent wheezing by age 24 months*

Outcome by nasal microbiota profile	Unadjusted analysis		Multivariable-adjusted analysis	
	HR (95% CI)	P value	HR (95%CI)	P value
Recurrent wheezing				
Moraxella dominant	1.75 (0.95 to 3.21)	0.07	1.82 (0.94 to 3.52)	0.08
Streptococcus dominant	Reference	–	Reference	–
Dolosigranulum dominant	1.58 (0.82 to 3.04)	0.17	1.75 (0.90 to 3.41)	0.10
Staphylococcus dominant	1.45 (0.73 to 2.91)	0.29	1.59 (0.78 to 3.21)	0.20
Corynebacteriaceae dominant	1.13 (0.41 to 3.11)	0.81	1.21 (0.43 to 3.36)	0.72

*Risk of recurrent wheezing (≥3 wheezing episodes at age 2–24 months) was compared by using unadjusted and multivariable-adjusted Cox proportional hazards models. Multivariable analysis adjusted for nine potential patient-level confounders (age, sex, household siblings, parental asthma, birth by caesarean section, breast feeding, history of systemic antibiotic use, respiratory symptoms at the time of nasal sampling and respiratory virus in age 2 months sample). *Streptococcus*-dominant microbiota profile was used as the reference.

parents being more cautious in reporting milder symptoms than others. However, the documented rate of ARIs in this study is consistent with the literature.[2][3] Furthermore, we also examined physician-diagnosed LRTIs, and found a larger effect size in the association between the early *Moraxella*-dominant microbiota profile and the LRTI outcome, which supports the validity of our inferences. Fifth, despite the statistical adjustment in the multivariable models, the observed associations do not prove causality and may be explained by unmeasured confounders, such as environmental microbiota. Sixth, the number of patients with recurrent wheezing in this healthy child cohort was relatively low, leading to a limited statistical power to examine the association with recurrent wheezing. Lastly, with 16S rRNA gene sequencing methods, identification at the species level was not possible in the present study. However, by applying metagenomic whole-genome shotgun sequencing to nasopharyngeal samples of infants hospitalised for bronchiolitis, we previously demonstrated that *Moraxella* genus was dominated by *M. catarrhalis* in infant airways.[36]

CONCLUSIONS

In this prospective, population-based birth cohort study, we identified five distinct nasal microbiota profiles in healthy 2-month-old infants. Early *Moraxella*-dominant nasal microbiota was associated with a significantly increased rate of all ARIs as compared with those with an early *Corynebacteriaceae*-dominant profile. Rates of LRTIs were significantly higher in children with early *Moraxella*-dominant and *Staphylococcus*-dominant profiles. These findings suggest that early nasal microbiota may play an important role in the development and severity of ARIs in children. Our data should advance research into understanding the complex interplay among the immunity, airway microbiome and pathobiology of ARI in young children, and characterising the children at higher risk for ARIs.

REFERENCES

1 Toivonen L, Schuez-Havupalo L, Karppinen S, *et al*. Rhinovirus infections in the first 2 years of life. *Pediatrics* 2016;138:e20161309.
2 Monto AS, Sullivan KM. Acute respiratory illness in the community. Frequency of illness and the agents involved. *Epidemiol Infect* 1993;110:145–60.
3 Sarna M, Ware RS, Sloots TP, *et al*. The burden of community-managed acute respiratory infections in the first 2-years of life. *Pediatr Pulmonol* 2016;51:1336–46.
4 Goto T, Tsugawa Y, Mansbach JM, *et al*. Trends in infectious disease hospitalizations in US children, 2000 to 2012. *Pediatr Infect Dis J* 2016;35:e158–63.
5 Følsgaard NV, Schjørring S, Chawes BL, *et al*. Pathogenic bacteria colonizing the airways in asymptomatic neonates stimulates topical inflammatory mediator release. *Am J Respir Crit Care Med* 2013;187:589–95.
6 Olszak T, An D, Zeissig S, *et al*. Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* 2012;336:489–93.
7 Biesbroek G, Tsivtsivadze E, Sanders EAM, *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.
8 Hasegawa K, Linnemann RW, Mansbach JM, *et al*. Nasal airway microbiota profile and severe bronchiolitis in infants: a case-control study. *Pediatr Infect Dis J* 2017;36:1044–51.
9 Bosch AATM, de Steenhuijsen Piters WAA, van Houten MA, *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.
10 Hasegawa K, Mansbach JM, Ajami NJ, *et al*. Association of nasopharyngeal microbiota profiles with bronchiolitis severity in infants hospitalised for bronchiolitis. *Eur Respir J* 2016;48:1329–39.
11 Teo SM, Mok D, Pham K, *et al*. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host Microbe* 2015;17:704–15.
12 de Steenhuijsen Piters WAA, Heinonen S, Hasrat R, *et al*. Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. *Am J Respir Crit Care Med* 2016;194:1104–15.
13 Kelly MS, Surette MG, Smieja M, *et al*. The nasopharyngeal microbiota of children with respiratory infections in Botswana. *Pediatr Infect Dis J* 2017;36:e211–8.
14 Vissing NH, Chawes BLK, Bisgaard H. Increased risk of pneumonia and bronchiolitis after bacterial colonization of the airways as neonates. *Am J Respir Crit Care Med* 2013;188:1246–52.
15 Lagström H, Rautava P, Kaljonen A, *et al*. Cohort profile: Steps to the healthy development and well-being of children (the STEPS Study). *Int J Epidemiol* 2013;42:1273–84.
16 Osterback R, Tevaluoto T, Ylinen T, *et al*. Simultaneous detection and differentiation of human rhino- and enteroviruses in clinical specimens by real-time PCR with locked nucleic acid probes. *J Clin Microbiol* 2013;51:3960–7.
17 Toivonen L, Schuez-Havupalo L, Rulli M, *et al*. Blood MxA protein as a marker for respiratory virus infections in young children. *J Clin Virol* 2015;62:8–13.
18 Jokela P, Vuorinen T, Waris M, *et al*. Performance of the Alere i influenza A&B assay and mariPOC test for the rapid detection of influenza A and B viruses. *J Clin Virol* 2015;70:72–6.
19 Human Microbiome Project Consortium. A framework for human microbiome research. *Nature* 2012;486:215–21.
20 Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–14.
21 Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460–1.
22 Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 2013;10:996–8.
23 Quast C, Pruesse E, Yilmaz P, *et al*. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;41:D590–D596.
24 Wu GD, Chen J, Hoffmann C, *et al*. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 2011;334:105–8.
25 Bray JR, Curtis JT. An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* 1957;27:325–49.
26 Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J Royal Statistical Soc B* 2001;63:411–23.
27 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc* 1995;57:289–300.
28 Hasegawa K, Linnemann RW, Mansbach JM, *et al*. Household siblings and nasal and fecal microbiota in infants. *Pediatr Int* 2017;59:473–81.

29 DeMuri GP, Gern JE, Eickhoff JC, *et al*. Dynamics of bacterial colonization with *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis* during symptomatic and asymptomatic viral upper respiratory tract infection. *Clin Infect Dis* 2018;66:1045–53.

30 Bosch AATM, Levin E, van Houten MA, *et al*. Development of upper respiratory tract microbiota in infancy is affected by mode of delivery. *EBioMedicine* 2016;9:336–45.

31 McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 2013;8:e61217.

32 Bisgaard H, Hermansen MN, Buchvald F, *et al*. Childhood asthma after bacterial colonization of the airway in neonates. *N Engl J Med* 2007;357:1487–95.

33 von Linstow M-L, Schønning K, Hoegh AM, *et al*. Neonatal airway colonization is associated with troublesome lung symptoms in infants. *Am J Respir Crit Care Med* 2013;188:1041–2.

34 Kloepfer KM, Lee WM, Pappas TE, *et al*. Detection of pathogenic bacteria during rhinovirus infection is associated with increased respiratory symptoms and asthma exacerbations. *J Allergy Clin Immunol* 2014;133:07.e1-3:1301–7.

35 Ta LDH, Yap GC, Tay CJX, *et al*. Establishment of the nasal microbiota in the first 18 months of life: Correlation with early-onset rhinitis and wheezing. *J Allergy Clin Immunol* 2018;142:86–95.

36 Stewart CJ, Mansbach JM, Wong MC, *et al*. Associations of nasopharyngeal metabolome and microbiome with severity among infants with bronchiolitis. A multiomic analysis. *Am J Respir Crit Care Med* 2017;196:882–91.

37 Marsh RL, Kaestli M, Chang AB, *et al*. The microbiota in bronchoalveolar lavage from young children with chronic lung disease includes taxa present in both the oropharynx and nasopharynx. *Microbiome* 2016;4.

38 Charlson ES, Bittinger K, Haas AR, *et al*. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med* 2011;184:957–63.

39 Kloepfer KM, Deschamp AR, Ross SE, *et al*. In children, the microbiota of the nasopharynx and bronchoalveolar lavage fluid are both similar and different. *Pediatr Pulmonol* 2018;53:475–82.