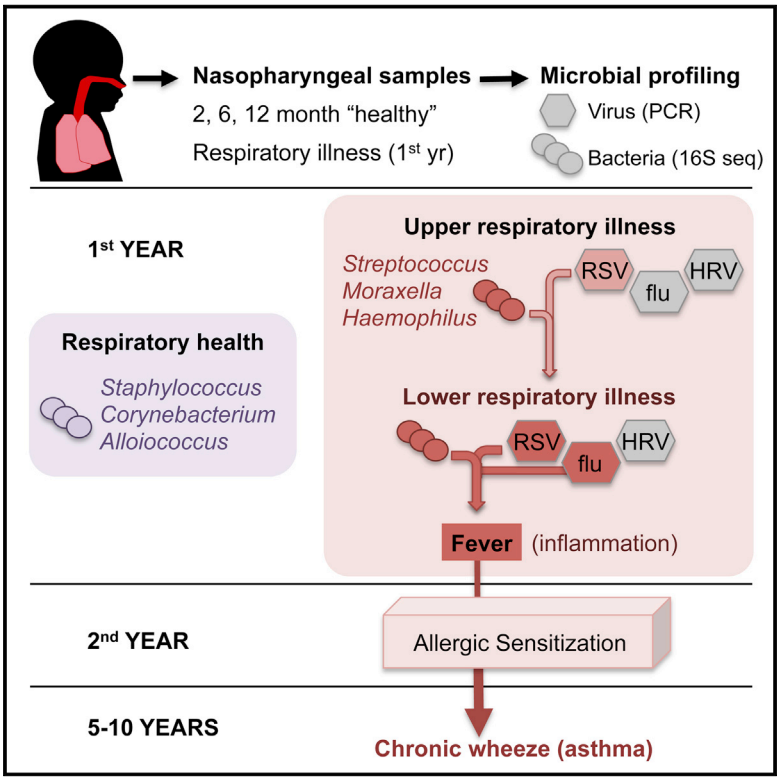
Resource

Cell Host & Microbe

The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development

Graphical Abstract



Highlights

- The nasopharynx microbiome of infants has a simple structure dominated by six genera
- Microbiome composition affects infection severity and pathogen spread to lower airways
- Early asymptomatic colonization with *Streptococcus* increases risk of asthma
- Antibiotic usage disrupts asymptomatic colonization patterns

Authors

Shu Mei Teo, Danny Mok, ..., Kathryn E. Holt, Michael Inouye

Correspondence

kholt@unimelb.edu.au (K.E.H.), minouye@unimelb.edu.au (M.I.)

In Brief

Teo et al. characterize bacterial and viral communities within the infant nasopharynx during the first year of life, comparing between asymptomatic colonization and episodes of acute respiratory infections. Microbiome composition affects infection severity and spread to lower airways and risk for future asthma development.

Teo et al., 2015, Cell Host & Microbe *17*, 1–12
May 13, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.chom.2015.03.008

CellPress

Please cite this article in press as: Teo et al., The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development, Cell Host & Microbe (2015), http://dx.doi.org/10.1016/j.chom.2015.03.008

Cell Host & Microbe
Resource

The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development

Shu Mei Teo,[1,2] Danny Mok,[3] Kym Pham,[4] Merci Kusel,[3] Michael Serralha,[3] Niamh Troy,[3] Barbara J. Holt,[3] Belinda J. Hales,[3] Michael L. Walker,[1] Elysia Hollams,[3] Yury A. Bochkov,[5] Kristine Grindle,[5] Sebastian L. Johnston,[7] James E. Gern,[5] Peter D. Sly,[6] Patrick G. Holt,[3,6] Kathryn E. Holt,[2,8,*] and Michael Inouye[1,8,*]
[1]Medical Systems Biology, Department of Pathology and Department of Microbiology & Immunology, The University of Melbourne, Parkville, VIC 3010, Australia
[2]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, VIC 3010, Australia
[3]Telethon Kids Institute, The University of Western Australia, West Perth, WA 6008, Australia
[4]Melbourne Translational Genomics Platform, Department of Pathology, The University of Melbourne, Parkville, VIC 3010, Australia
[5]Department of Internal Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI 53726, USA
[6]Queensland Children's Medical Research Institute, The University of Queensland, Brisbane, QLD 4059, Australia
[7]Airway Disease Infection Section and MRC & Asthma UK Centre in Allergic Mechanisms of Asthma, National Heart and Lung Institute, Imperial College London, Norfolk Place, London W2 1PG, UK
[8]Co-senior author
*Correspondence: kholt@unimelb.edu.au (K.E.H.), minouye@unimelb.edu.au (M.I.)
http://dx.doi.org/10.1016/j.chom.2015.03.008

SUMMARY

The nasopharynx (NP) is a reservoir for microbes associated with acute respiratory infections (ARIs). Lung inflammation resulting from ARIs during infancy is linked to asthma development. We examined the NP microbiome during the critical first year of life in a prospective cohort of 234 children, capturing both the viral and bacterial communities and documenting all incidents of ARIs. Most infants were initially colonized with *Staphylococcus* or *Corynebacterium* before stable colonization with *Alloiococcus* or *Moraxella*. Transient incursions of *Streptococcus*, *Moraxella*, or *Haemophilus* marked virus-associated ARIs. Our data identify the NP microbiome as a determinant for infection spread to the lower airways, severity of accompanying inflammatory symptoms, and risk for future asthma development. Early asymptomatic colonization with *Streptococcus* was a strong asthma predictor, and antibiotic usage disrupted asymptomatic colonization patterns. In the absence of effective anti-viral therapies, targeting pathogenic bacteria within the NP microbiome could represent a prophylactic approach to asthma.

INTRODUCTION

The human microbiome is now recognized as playing an important role in the etiology and pathogenesis of myriad diseases (Weinstock, 2012). However, elucidation of these complex roles requires targeted characterization of microbial communities present in relevant spatial niche(s) during critical periods of pathogenesis. The focus of this study is the respiratory tract, in particular the nasopharynx (NP), which is an accessible source of airway microbial communities (Hilty et al., 2010) and serves as a conduit for pathogens associated with lower respiratory illnesses (LRIs) that are responsible for substantial morbidity and mortality worldwide.

Of particular interest is asthma, a multi-factorial disease characterized by airway inflammation and associated smooth muscle hyperplasia. It is now recognized that the hallmark persistent wheeze of asthma is consolidated in childhood and, further, may progress to chronic asthma in adulthood (Holt and Sly, 2012; Sly et al., 2008) and potentially chronic obstructive pulmonary disease (Tai et al., 2014). We and others have previously shown that development of persistent atopic (allergic) wheeze in children is linked to the number of virus-associated febrile and/or wheezy LRIs experienced during infancy (Jackson et al., 2008; Kusel et al., 2007, 2012; Oddy et al., 2002). The principal virus type of current interest is human rhinoviruses (HRVs), particularly subtype C (HRV-C) (Bochkov and Gern, 2012); however, respiratory syncytial virus (RSV) is also recognized as a major cause of infant LRI (Wu and Hartert, 2011). The relative contributions of these viral pathogens in asthma initiation remain controversial (Stein and Martinez, 2010). Further complicating the picture, recent studies have also implicated bacterial pathogens as potential independent causal factors in infant LRIs and their long-term sequelae. Notably, culture of *S. pneumoniae*, *M. catarrhalis*, or *H. influenzae* from NP samples taken at 1 month of age has been linked to increased risk for subsequent diagnosis of asthma at 5 years of age (Bisgaard et al., 2007). These findings have fuelled debate around the use of antibiotics and vaccine strategies for respiratory illness in children (Penders et al., 2011; Rollins et al., 2010).

Several studies have investigated airway microbiota in children or adults with chronic respiratory illness, including asthma

(Bogaert et al., 2011; Hilty et al., 2010; Vissers et al., 2014); however, no study has investigated the airway microbiome during the critical infancy period (0–12 months). In this study, we investigated the NP microbiome during the first year of life using the Childhood Asthma Study (CAS), a prospective cohort of 234 children (Kusel et al., 2006, 2007, 2008, 2012), to elucidate the NP microbiome during respiratory health and illness, its longitudinal dynamics, susceptibility to exogenous factors such as antibiotics, and association with future asthma.

RESULTS

A total of 1,021 NP microbiome profiles were obtained from 234 infants using 16S rRNA gene deep sequencing (see Supplemental Experimental Procedures). These included 487 "healthy" NP samples collected in the absence of respiratory symptoms and 534 "infection" NP samples collected during episodes of acute respiratory illness (ARI) during the first year of life. Three quarters of the infants (n = 177) contributed at least two healthy NP samples at the age of ~2 months, ~6 months, and/or ~12 months. Eighty percent of the infants (n = 186) contributed a healthy sample before experiencing their first ARI. The 534 infection NP samples were from 184 infants who had experienced $\geq$ 1 ARI within the first year of life. NP samples were analyzed from all (380/381) recorded LRI in this period and a random selection of 20% (154/782) of recorded episodes of upper respiratory illness (URI). The characteristics of the infants are summarized in Table S1.

NP Microbiome Composition

Across all NP samples, > 193 million high-quality 16S rRNA sequences were classified into 14,131 operational taxonomic units (OTUs), of which 1,010 were supported by > 1,000 reads each. The dominant phyla were Proteobacteria (48%), Firmicutes (38%), Actinobacteria (13%), Bacteroidetes (1%), and Fusobacteria (0.5%) (Figure 1A). The NP microbiomes were dominated by six genera: *Moraxella* (31.2%), *Streptococcus* (15.5%), *Corynebacterium* (13.5%), *Staphylococcus* (10.3%), *Haemophilus* (9.7%), and *Alloiococcus* (8.8%; genus *Dolosigranulum* in some databases) (Figure 1A). Despite the inclusion of diverse OTUs of these genera in the reference database, our sequences were dominated by one OTU per genus (Figure S1), consistent with culture-based studies reporting NP colonization with the species *Moraxella catarrhalis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Haemophilus influenzae*, and *Alloiococcus otitidis*. Hierarchical clustering of NP microbiomes based on relative abundance of the six major genera identified six microbiome profile groups (MPGs, Figure 1B). Each MPG was dominated by one of the six genera, although some samples in the *Alloiococcus* MPG also had relatively high abundance of *Corynebacterium* (Figure 1B).

NP Microbiome Dynamics

Healthy NP samples collected around 2 months of age were dominated by *Staphylococcus* (41%) and *Corynebacterium* (22%) MPGs, but the frequency of these MPGs declined with age (11% and 10%, respectively, at 12 months old) (Figure 1C). In contrast, the prevalence of *Alloiococcus* and *Moraxella* MPGs in healthy samples increased with age (14% and 9% at 2 months, 26% and 41% at 12 months, respectively). Analysis of MPG transitions among consecutive healthy NP samples from the same individuals (Figures S3A and S3B) suggested that *Staphylococcus* carriage was unstable or transient, particularly where an ARI had occurred in the intervening period between sampling. *Alloiococcus* was a stable colonizer, but less so if an ARI occurred between sampling. Where an ARI occurred in the intervening period between healthy samples, the most common transitions were to the *Moraxella* MPG from other MPGs or the maintenance of stable colonization with *Moraxella* (Figure S3B). Almost all *Moraxella*-colonized infants experienced subsequent ARI before the next healthy sample was taken (explored further below). *Haemophilus* was very rarely detected in healthy NP microbiomes, while *Streptococcus* was present in 14% of healthy samples at each sampling time (Figure 1C). Similar age-related patterns were observed among infection samples, with a decline in *Staphylococcus* and *Corynebacterium* MPGs and increase in *Moraxella*, *Haemophilus*, and *Streptococcus* MPGs in older children (Figure 1D). Interestingly, males had significantly more *Moraxella* in healthy samples (OR 1.3 for log abundance; 95% CI 1.1–1.5, p = 0.0014, adjusted for age); no other gender effects were detected.

We next assessed the impact of environmental factors on the relative abundance of the common NP microbiome genera. We found no significant effects of delivery mode or breastfeeding on NP colonization at 2 months of age; the latter was unsurprising as nearly all infants (90%) were breastfed for at least 2 months. The abundance of *Streptococcus* in healthy NP samples was significantly lower among children whose parents reported having furry pets such as dogs or cats in the home (OR 0.84, 95% CI 0.70–1.0, p = 0.046; adjusted for age at NP sampling). No other significant associations were detected for pets. Children attending day care had significantly higher relative abundances of *Haemophilus* and *Moraxella* and lower relative abundances of *Corynebacterium* and *Staphylococcus* (Figure 2A) in both healthy and infection samples (note that very few children had commenced day care by 6 months of age, hence the impact of day care attendance was assessed at the 12-month time point only). Co-habiting with siblings was also associated with higher abundances of *Haemophilus*, *Streptococcus*, and *Moraxella* and lower abundance of *Staphylococcus* during health and ARI (adjusted for age at sampling, Figure 2B). Importantly, among healthy samples, antibiotic usage in the four weeks prior to sampling was associated with higher abundances of *Haemophilus*, *Streptococcus*, and *Moraxella* and lower abundances of *Alloiococcus* and *Corynebacterium* (adjusted for age at sampling, Figure 2C). The composition of the healthy NP microbiome was also affected by the number of prior respiratory infections experienced, with higher abundance of *Moraxella* and lower abundances of *Alloiococcus* or *Corynebacterium* in samples following increasing numbers of ARIs (Figure 2E). At the MPG level, ARIs dominated by *Haemophilus* increased in spring-summer, while those dominated by *Moraxella* peaked in autumn-winter (Figure 2F). We therefore tested for differences in relative abundance in autumn-winter and spring-summer, adjusting for age and number of prior infections; this confirmed significant seasonal effects on the abundance of *Haemophilus* (summer associated) and *Moraxella* (winter associated) among

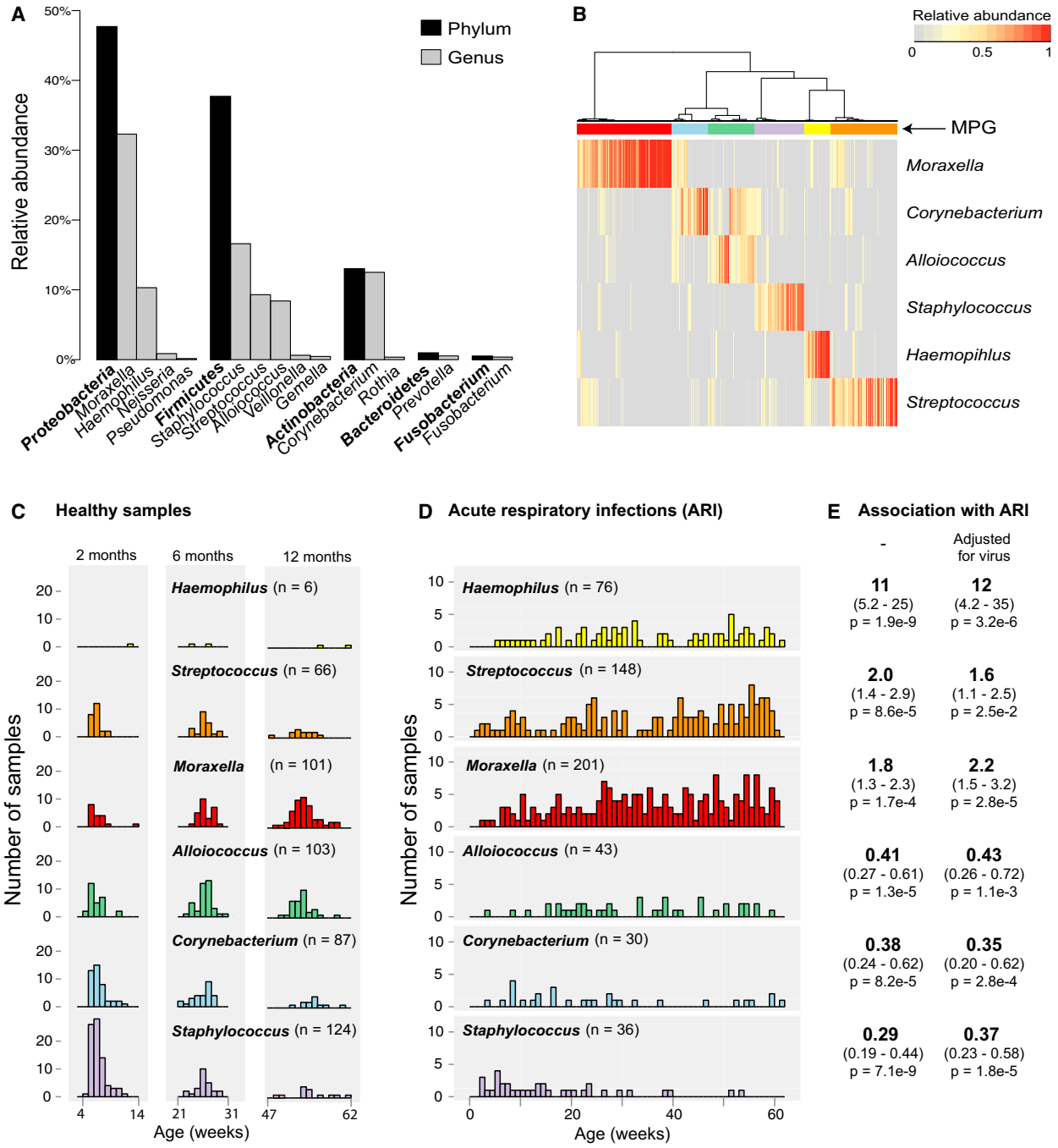


Figure 1. Bacterial Composition of 1,021 Nasopharyngeal Aspirates Collected from 234 Infants during Periods of Respiratory Health and Disease

(A) Frequency of the most abundant phyla and genera (comprising 99.9% of reads).

(B) Clustering of samples into microbiome profile groups (MPGs) based on relative abundance of the six most common genera. Colored bars indicate MPGs, labeled by their dominant genus: *Moraxella* (red), *Corynebacterium* (blue), *Alloiococcus* (green), *Staphylococcus* (purple), *Haemophilus* (yellow), and *Streptococcus* (orange).

(C) Weekly frequencies of MPGs among healthy samples, collected during planned visits at approximately 2, 6, and 12 months of age and following at least 4 weeks without symptoms of acute respiratory infection (ARI).

(D) Weekly frequencies of each MPG among ARI samples.

(E) Odds ratios for association of MPGs with ARI symptoms, adjusted for age, gender, season, number of prior infections, antibiotics intake, mother's antibiotics intake, delivery mode, and breastfeeding; with and without adjustment for detection of common viruses (RSV, HRV).

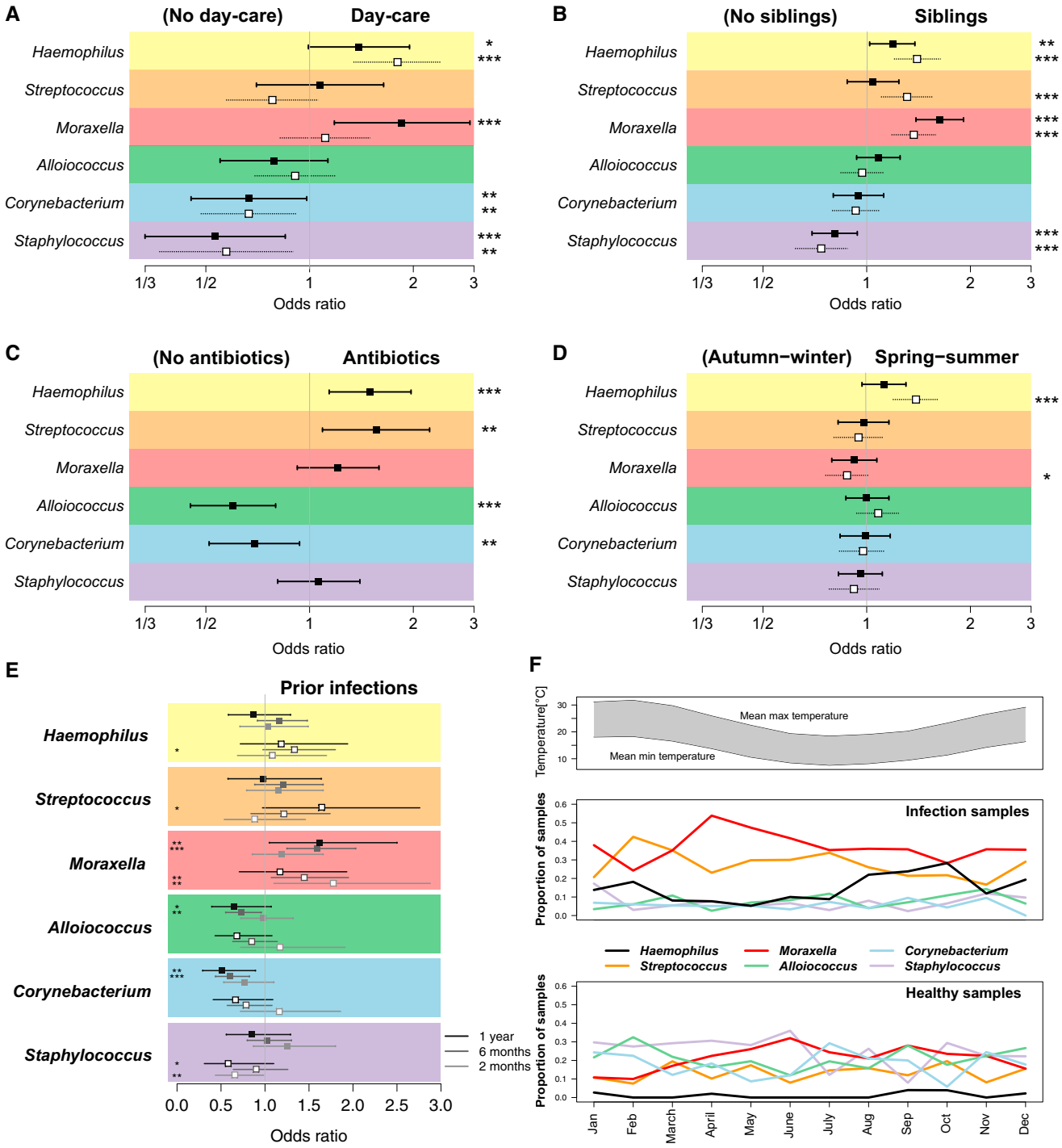


Figure 2. Impact of Environmental Factors on Relative Abundances of Major Genera of the NP Microbiome

(A–D) Squares, odds ratios; filled squares, healthy samples; empty squares, infection samples; bars, 95% confidence intervals; *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Associations are estimated using logistic regression and adjusted for age: (A) day care attendance (yes versus no, 12-month samples), (B) co-habiting with siblings, (C) antibiotics intake in the 4 weeks preceding NP sample collection, (D) season (spring-summer versus autumn-winter).

(E) Impact of prior ARI, estimated using proportional odds ordinal logistic regression and categorized as 0, 1, or $\geq$ 2 ARI.

(F) Seasonal patterns: top, mean maximum and minimum temperatures in study location (Perth); bottom, monthly proportions of samples in each microbiome profile group (MPG) for infection and healthy samples.

Table 1. Associations between ARI Symptom Severity and Risk-Associated MPGs

Symptom Group Comparison	Model	*Streptococcus*, *Moraxella*, or *Haemophilus* MPG			RSV	HRV (expanded screen)
		Streptococcus MPG	*Moraxella* MPG	*Haemophilus* MPG		
LRI versus URI	1A	OR = 2.2 (1.3–3.7), p = 0.0043			2.2 (1.1–4.7), p = 0.034	–
	1B	2.3 (1.2–4.5), p = 0.011	2.1 (1.1–3.8), p = 0.017	2.1 (1.0–4.6), p = 0.056	2.3 (1.1–4.8), p = 0.033	–
Febrile LRI versus non-febrile LRI	2A	OR = 3.8 (1.0–14), p = 0.047			4.0 (1.6–9.8), p = 0.0031	0.45 (0.21–0.95), p = 0.036
	2B	4.3 (1.0–18), p = 0.046	3.8 (0.98–15), p = 0.053	2.7 (0.53–14), p = 0.23	4.0 (1.6–10), p = 0.0031	0.46 (0.21–0.98), p = 0.043
Wheezy LRI versus non-wheezy LRI	3A	OR = 1.4 (0.60–3.1), p = 0.45			0.94 (0.41–2.1), p = 0.88	1.2 (0.62–2.2), p = 0.64
	3B	1.1 (0.39–3.0), p = 0.88	1.8 (0.74–4.1), p = 0.20	1.0 (0.34–3.2), p = 0.93	0.95 (0.42–2.2), p = 0.90	1.1 (0.60–2.2), p = 0.69

Odds ratio (OR) (95% confidence interval), p value. NP samples taken within a week of antibiotic use were excluded from analysis. The response variable (symptom group comparison) is shown in column 1. Two models (A, B) were fit for each comparison, labeled in column 2. Model A included *Streptococcus* or *Moraxella* or *Haemophilus* MPG as a single covariate; Model B included *Streptococcus*, *Moraxella*, and *Haemophilus* MPGs as separate covariates. Both models also include RSV, HRV, and the potential confounders age, gender, and season. Note: HRV was not included in the LRI versus URI comparison, as enhanced sensitivity re-screening for HRV was performed only in LRI samples.

ARI samples and a similar but non-significant trend among healthy NP samples (Figure 2D).

NP Microbial Determinants of ARI Symptoms

The *Moraxella*, *Streptococcus*, and *Haemophilus* MPGs were significantly more frequent in ARI compared to healthy NP samples, even after adjusting for a large set of potential confounders (age, gender, season, number of prior infections, antibiotic intake, mother's antibiotic intake, delivery mode, and breastfeeding) (Figure 1E). The *Staphylococcus*, *Corynebacterium*, and *Alloiococcus* MPGs were significantly less frequent in ARI (Figure 1E). The rare genus *Neisseria* was more common in infection samples, especially LRIs (of 28 NP samples with > 5% relative abundance of *Neisseria*, 4 were from URI and 20 were from LRI, though all were co-colonized with *Streptococcus*). It was not possible to confirm species from the 16S sequences; however, it is well known that the respiratory pathogens *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, and *N. meningitidis* are frequently cultured from respiratory infections in children. We have previously measured IgG to species-specific surface proteins of *S. pneumoniae* and *H. influenzae* at 12 months of age in the CAS cohort (Hales et al., 2012). Here we found that *H. influenzae*-specific IgG was significantly associated with the number of prior ARI samples testing positive for either of the two most common *Haemophilus* OTUs (Figures S2A–S2C); similar results were obtained for *S. pneumoniae*-specific IgG antibodies and the dominant *Streptococcus* OTU (Figures S2D–S2F). Healthy colonization with these genera was not associated with species-specific IgG.

A total of 138 children had $\geq$ 2 ARI samples profiled, and of these, 97 (70%) had $\geq$ 2 different MPGs among their ARIs. There was a clear temporal trend, with infections occurring closer in time more likely to be of the same MPG (Figure S3D). *Moraxella* and *Haemophilus* MPGs were particularly stable between consecutive infections, i.e., following an ARI in which the *Moraxella* or *Haemophilus* MPG was present, the next ARI was more likely to share the same MPG type than expected given the overall frequency of these MPGs among infection samples (Figure S3C).

We further considered the impact of the NP microbiome on infection severity and interactions with viral pathogens. All samples analyzed here were previously screened for a panel of viruses (Kusel et al., 2006). This screen detected viruses in 21% of healthy samples, 68% of URIs, and 69% of LRIs. The most common viruses detected were RSV (11% of ARI) and HRV (40% of ARI) (Figure S4), although subsequent expanded screening and subtyping of HRV in LRI samples suggests this is an underestimate (see below). For all virus groups except adenovirus and coronavirus, virus detection was significantly positively associated with ARI symptoms (i.e., ARI versus healthy samples, Figure S4). The association between ARI and *Streptococcus*, *Haemophilus*, and *Moraxella* MPGs remained after adjusting for detection of virus (OR 7.0, $p < 1 \times 10^{-15}$ for any of these MPGs; individual ORs in Figure 1E), indicating both viruses and bacteria contribute to ARI symptoms. Among the viruses analyzed, only RSV was significantly more frequent in LRIs versus URIs (16% versus 8.3%, OR 2.3; Figure S4, Table 1). The illness-associated MPGs *Streptococcus*, *Haemophilus*, and *Moraxella* were significantly associated with LRIs versus URIs (OR > 2), individually and collectively, adjusting for the effect of RSV (Table 1). As a group, the illness-associated MPGs increased in frequency from healthy to URI to LRI samples, regardless of the presence of RSV (Figure S5). Taken together, these analyses indicate that both viruses and bacteria independently contribute to ARI and that bacteria and RSV independently increase the risk of infection spread to the lower airways.

In our earlier studies with this cohort, we observed an association between LRIs (but not URIs) during infancy and risk for wheeze at age 5 years; moreover, this association was restricted to severe LRIs, i.e., those accompanied by fever and/or wheeze (Holt et al., 2010; Kusel et al., 2007, 2012). Thus we investigated the association of viruses and bacteria with the presence of fever and wheeze symptoms during LRI. To enable more accurate assessment of the role of HRV in this critical sample group, we

Table 2. Association between Microbial Events during Infancy and Chronic Wheeze at Age 5 and 10 Years

	Wheeze at 5 Years		Wheeze at 10 Years	
	All children	Atopics	All children	Atopics
Any febrile LRI	2.3 (1.2–4.5), p = 0.016*	2.7 (1.1–7), p = 0.034*	2.2 (0.93–5.2), p = 0.071	2.7 (0.89–8.7), p = 0.083
Any wheezy LRI	1.6 (0.79–3), p = 0.2	1.3 (0.49–3.3), p = 0.61	1.4 (0.58–3.3), p = 0.45	1.6 (0.53–4.8), p = 0.4
Any HRV wheezy LRI	2 (0.93–4.2), p = 0.073	2.5 (0.86–7.2), p = 0.092	2.1 (0.81–5.4), p = 0.11	1.9 (0.55–6.4), p = 0.29
Any HRV-C wheezy LRI	2.4 (0.93–6.1), p = 0.064	7.2 (1.7–35), p = 0.009*	3.5 (1.1–11), p = 0.026*	7.1 (1.6–40), p = 0.014*
Any HRV-A wheezy LRI	1.2 (0.42–3.1), p = 0.74	0.55 (0.11–2.2), p = 0.43	1.4 (0.37–4.7), p = 0.57	0.35 (0.02–2.2), p = 0.34
Any risk bacteria LRI	0.89 (0.45–1.8), p = 0.73	0.96 (0.39–2.4), p = 0.93	2 (0.82–5.2), p = 0.14	1.7 (0.59–5.2), p = 0.34
High-abundance *Streptococcus* colonization (≤7 weeks)	3.8 (1.3–12), p = 0.017*	4 (0.88–21), p = 0.077	2.7 (0.6–12), p = 0.18	3.9 (0.63–28), p = 0.15

Odds ratio (95% confidence interval), p value; estimated using logistic regression, adjusted for gender and maternal and paternal history of atopic disease; estimated separately for all children and those who were atopic by 2 years of age. LRI, lower respiratory illness, further classified according to microbes and symptoms; any risk bacteria LRI, any LRI with *Streptococcus*, *Moraxella*, or *Haemophilus* MPG. Early *Streptococcus* colonization was assessed in the first healthy NP sample, collected by 7 weeks of age and prior to any recorded infection; high abundance was classified as > 20% *Streptococcus* reads (based on distribution in Figure 5A).

performed an expanded screen for detection and subtyping of HRV within LRI samples (Supplemental Experimental Procedures). This assay proved more sensitive than the first screen and detected HRV in 66% of LRI samples (twice that in the earlier screen), with equal amounts of HRV-A and HRV-C (Figure S4). During LRI, the presence of any HRV, or HRV-C specifically, was negatively associated with fever (Table 1). The only viruses showing positive associations with fever were RSV (Table 1, Figure 3C, Figure S4) and influenza (8 influenza-positive fever LRI events only, all with illness-associated MPGs, thus not considered further). Among LRIs, the illness-associated MPGs were associated with fever, even after adjusting for RSV, HRV, age, season, and gender (Table 1). Interestingly, *Moraxella* MPG was also significantly positively associated with fever among RSV-positive LRIs (OR 9.2, 95% CI 1.1–73, p = 0.037; Figure S5), suggesting a possible interaction between *Moraxella* and RSV whereby their co-presence further enhances risk of fever. Overall, only 10 febrile LRIs (9.5%) could not be explained by the presence of RSV or illness-associated MPGs. Across the cohort, the presence of wheeze during LRI was not significantly associated with any viral or bacterial groups (Table 1), including HRV or HRV-C specifically. Of all ARIs analyzed, there were 61 with concomitant otitis media (OM) affecting 47 infants; however, ARIs with accompanying OM had similar NP microbiome profiles to those of LRIs with no OM diagnosis (Table S2).

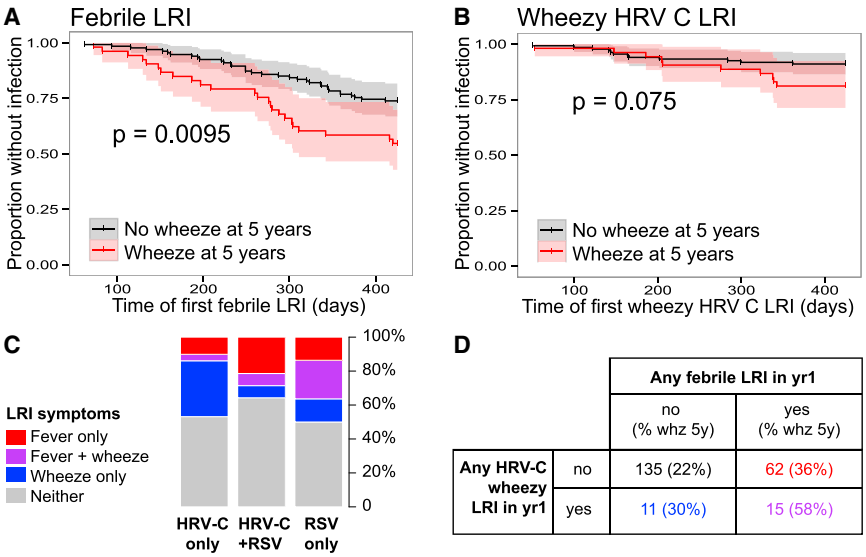
Impact of ARI on Later Chronic Wheeze

We next investigated associations between LRI in the first year of life and subsequent expression of chronic wheeze at age 5 or 10 years. Positive associations were found for two discrete classes of LRI. First, febrile LRI was significantly positively associated with later chronic wheeze and among children who were atopic by 2 years of age (Table 2). Furthermore, timing of first febrile LRI appeared to be important, with earlier febrile LRI occurring among children who had chronic wheeze at 5 years (p = 0.0095, Figure 3A). Second, HRV-C LRI accompanied by wheezing symptoms showed a positive association with later chronic wheeze among all children and particularly strongly so for those who were atopic by 2 years (OR ~7, Table 2), but not among non-atopics (OR 1, p = 0.95; Table S3). As we found no association between HRV-C and wheeze during LRI across the whole cohort (OR 1.2, 95% CI 0.63–2.2, p = 0.61; adjusting for illness-associated MPGs, RSV, age, season, and gender), we re-examined the association in those who developed atopy by age 2. Among this group of children (n = 65), the presence of HRV-C was significantly positively associated with increased risk of wheeze during LRI (OR 2.7, 95% CI 1.1–7, p = 0.035; adjusting for illness-associated MPGs, RSV, age, season, and gender).

Since febrile LRI and HRV-C wheezy LRI were both associated with later chronic wheeze (Table 2), and RSV was associated with febrile LRI (Table 1), we examined the interaction between RSV, HRV-C, LRI symptom severity, and chronic wheeze at age 5. Febrile LRI and wheezy HRV-C LRI in the first year of life appeared to exert independent effects on later chronic wheeze, since only a minority of HRV-C wheezy infections were also febrile (Figure 3C). Further, at the level of individual children, the frequency of chronic wheeze at 5 years was elevated among those with either one of febrile LRI (36%) or wheezy HRV-C LRI (30%) but was greatest among children who experienced both (58%, Figure 3D).

Impact of NP Colonization on ARI

Whereas *Streptococcus*, *Haemophilus*, or *Moraxella* MPGs were detected in healthy NP samples, subsequent infections tended to belong to the same MPG (Figure 4A). Critically, we investigated whether healthy colonization in early infancy was associated with subsequent episodes of ARI. Since prior infections also affect the healthy NP microbiome (Figure 2E), we restricted these analyses to healthy samples collected at 5–9 weeks of age and prior to each infant's first reported ARI (n = 160). Using Cox proportional hazards models, infants whose earliest healthy NP sample was of the *Moraxella* or *Streptococcus* MPG tended to experience ARI at a younger age than those with other MPGs (Figure 4B; note that *Haemophilus* MPGs were extremely rare in early healthy samples). When examining URIs and LRIs separately, we found that early *Moraxella* colonization was associated with earlier first URI, whereas early *Streptococcus* colonization was strongly associated with earlier first LRI (Figures 4C and 4D). Since our data suggested that *Alloiococcus* and *Moraxella*



		Any febrile LRI in yr1	
		no (% whz 5y)	yes (% whz 5y)
Any HRV-C wheezy LRI in yr1	no	135 (22%)	62 (36%)
	yes	11 (30%)	15 (58%)

Figure 3. Symptoms of Lower Respiratory Illness during the First Year of Life Are Associated with Viruses Present during the Infection and Predict Chronic Wheeze at 5 Years of Age

(A and B) Kaplan-Meier survival curves for age (days) at (A) first febrile LRI and (B) first HRV-C wheezy LRI, stratified by chronic wheeze status at 5 years. p values shown were estimated using Cox proportional hazards models, adjusted for gender and maternal and paternal history of atopic disease. Shaded areas indicate 95% confidence intervals.

(C) Frequencies of fever and wheeze symptoms during LRI, in which HRV-C and/or RSV were detected. Total numbers are: HRV-C only, n = 79; HRV-C and RSV, n = 14; RSV only, n = 22.

(D) Cross-tabulation of individuals according to their experience of LRI during infancy; percentages in brackets indicate frequency of chronic wheeze at 5 years.

were key stable colonizers of the NP microbiome, we also examined whether *Alloiococcus*-colonized infants differed from *Moraxella*-colonized infants (defined as those with $\geq$ 1 healthy sample of *Moraxella* MPG and none with *Alloiococcus* MPG, and vice versa). These groups did not differ in terms of overall numbers of ARI, LRI, or later wheezing phenotypes. However, compared to either *Moraxella*-colonized infants or those not in either group, *Alloiococcus*-colonized infants had fewer RSV infections, especially RSV LRIs (OR 0.27, Table S4).

Early NP Colonization Impacts Later Chronic Wheeze

We next assessed the association between early pre-ARI asymptomatic NP colonization and current wheeze at 5 and 10 years of age, stratified by atopic sensitization by 2 years of age (Supplemental Experimental Procedures). This analysis was restricted to the 160 infants (70%) who had an asymptomatic NP sample taken prior to their first ARI (and $\leq$ 9 weeks of age). *Haemophilus* MPG was not detected in pre-ARI healthy samples, and *Moraxella* MPG showed no evidence of association with future wheeze. The *Streptococcus* MPG showed a weak association; however, the relative abundance of *Streptococcus* in these samples was highly skewed, so we divided samples into those with high (> 20%) or low ($\leq$ 20%) relative abundance of *Streptococcus* reads (Figure 5A). High *Streptococcus* abundance in the first pre-ARI healthy NP sample was more frequent in infants who later displayed wheeze at 5 years, and this association was stronger when restricting the analysis to earlier NP samples (Figure 5B, Table 2). The same trend was evident for wheeze at 10 years of age, despite reduced sample size due to loss to follow-up (Table 2). Early *Streptococcus* colonization was associated with younger age at first LRI (Figure 4D), but not with presence of *Streptococcus* in the first LRI or with detection of *S. pneumoniae* antibodies at 12 months of age. No statistically significant associations were observed for future chronic wheeze when aggregating the 7-week healthy *Streptococcus*, *Haemophilus*, and *Moraxella* MPGs into a single predictor (with or without atopy by 2 years). Overall, infants who were atopic by age 2 and developed chronic wheeze at age 5 were twice as likely to have had early *Streptococcus* colonization, febrile LRI, and/or HRV-C wheezy LRI in the first year of life, compared to those that did not develop chronic wheeze (Figure 5C).

DISCUSSION

NP Microbiome Composition and Dynamics

Our data provide a detailed prospective characterization of bacterial communities within the human NP microbiome during the first year of life. The NP microbiome was qualitatively simple (Figure 1), dominated by six common genera: *Haemophilus*, *Streptococcus*, *Moraxella* (each more common in ARI), *Staphylococcus*, *Alloiococcus*, and *Corynebacterium* (more common in healthy samples). This is consistent with previous studies of NP microbiome composition in children aged 12–14 months (Biesbroek et al., 2014; Bogaert et al., 2011) and adults (Hilty et al., 2010), although intra- and inter-sample diversity was greater in these older groups, likely due in part to extreme seasonal variation in those study locations. In contrast, our study site in Perth, Australia has a very moderate climate, and we detected only limited seasonal effects (Figure 2D) that were readily adjusted for, enabling us to assess the dynamics of the NP microbiome at different stages of infancy and to examine its association with other factors.

We found that early NP colonization typically involved *Staphylococcus* or *Corynebacterium*, which was later replaced by *Moraxella* or *Alloiococcus*. *Staphylococcus* was the dominant colonizing bacteria in the early healthy NP microbiomes, but its presence declined rapidly with age (Figure 1). *S. aureus* is a common cause of neonatal sepsis, and high rates of *S. aureus* nasal colonization in the first months of life have been reported (Bisgaard et al., 2007, 2010). Few studies have examined longitudinal colonization; however, a recent study in African children reported *S. aureus* colonization in 42% of infants at 1 month and 12% at 12 months of age, a trend that was mirrored in maternal colonization rates (Schaumburg et al., 2014) and is strikingly similar to the patterns we observed (41% at 2 months,

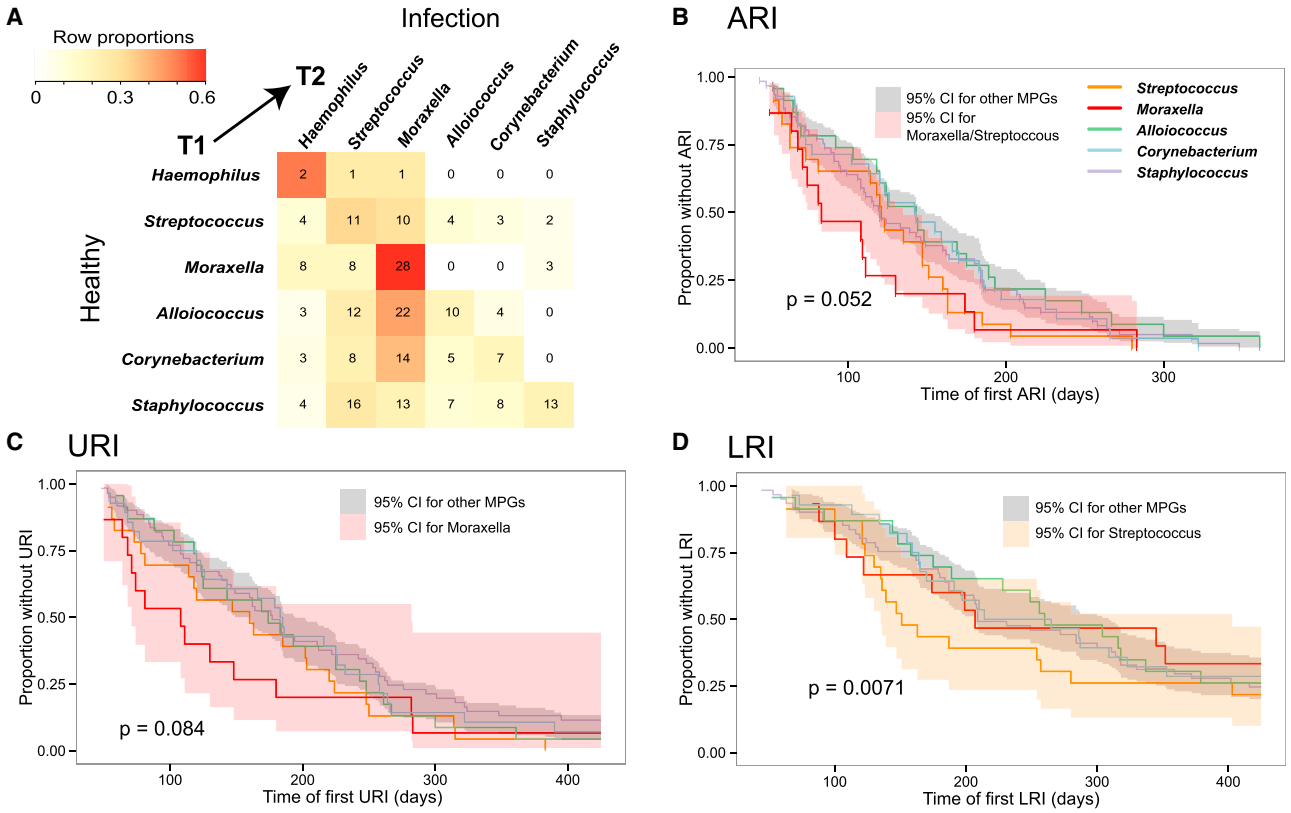


Figure 4. Impact of Early Colonization on Age at First Respiratory Infection

(A) Microbiome profile group (MPG) transitions between healthy samples (T1) and the next sequenced infection (T2). Cell numbers indicate the number of times the respective transition from T1 to T2 was observed in the dataset; cells are colored to indicate the row proportions as per legend.

(B–D) Kaplan-Meier survival curves for age (days) of first (B) acute respiratory illness (ARI), (C) upper respiratory illness (URI), and (D) lower respiratory illness (LRI), stratified according to the MPG of the first healthy sample (collected by 9 weeks of age and prior to any infection, n = 160). Cox proportional hazards models were adjusted for age, gender, season, virus status in the early healthy sample, and virus status at the first event. Shaded areas indicate 95% confidence intervals.

11% at 12 months). *Staphylococcus* and *Corynebacterium* are both common components of the human skin microbiome, and in our data they showed a comparable temporal pattern, with high rates in the early NP microbiome declining with age (Figure 1). We speculate that infants tend to be colonized initially with skin-dwelling bacteria (acquired from parents and others), which is replaced over time by stable colonization with *Moraxella* or *Alloiococcus* and punctuated by transient acquisition of *Streptococcus* or *Haemophilus* that is frequently accompanied by ARI symptoms. Our data further suggest the transition to *Moraxella* is associated with exposure to other children (in the home or at day care) and episodes of ARI, and that colonization with *Moraxella*, *Streptococcus*, and *Haemophilus* are selected for by antibiotic exposure (Figure 2).

Moraxella was the most common genus in our study population, dominating 21% of all healthy NP microbiomes and 38% of infection samples. *Moraxella* was represented in our sequence data by a single OTU matching that of *M. catarrhalis*, a human-restricted, unencapsulated, Gram-negative bacterium previously associated with both commensal NP colonization and pathogenicity in the respiratory tract and inner ear (de Vries et al., 2009). Increased rates of NP colonization with *M. catarrhalis* have been reported in children following pneumococcal vaccination (Revai et al., 2006); however, the samples in our study predate introduction of this vaccine. In our cohort the presence of *Moraxella* increased with age, and *Moraxella* was a particularly stable component of the NP microbiome during both health and disease (Figure S3). These findings are consistent with *Moraxella*'s known ability to form biofilms (de Vries et al., 2009), which offer protection from antibiotics and promote co-colonization with other common bacteria such as *S. pneumoniae* and *H. influenzae* (Verhaegh et al., 2011). *Moraxella* was more abundant during the cooler months (Figure 2), a seasonal trend consistent with recent reports that cold shock at 26°C stimulates growth, colonization, and expression of virulence-associated traits in *M. cattarhalis* (Spaniol et al., 2011). These data are consistent with a Dutch culture-based study of *M. catarrhalis* carriage in young children (Verhaegh et al., 2011), which found increasing prevalence of *M. catarrhalis* during the first year of life, with a strong seasonal effect (also observed in a Swedish cohort [Gisselsson-Solén et al., 2014]) and a significant positive association with day care attendance and siblings. However, these studies did not examine the stability of *Moraxella* colonization within individuals over time.

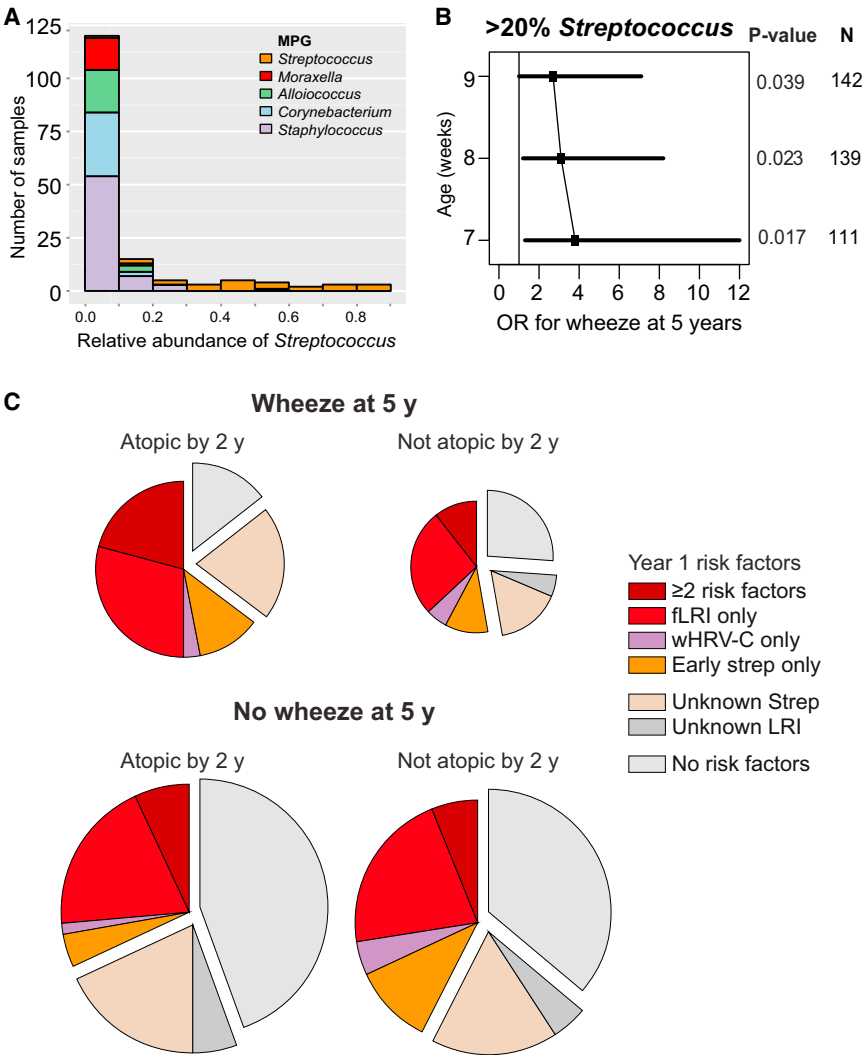


Figure 5. Predictors of Chronic Wheeze at Age 5

(A) *Streptococcus* abundance among healthy samples collected by 9 weeks of age, broken down by microbiome profile group (MPG).

(B) Adjusted odds ratio (OR, squares) and 95% confidence intervals (bars) for association between chronic wheeze at age 5 and high (> 20%) abundance of *Streptococcus* in the first healthy NP sample; p values and sample sizes (n) are indicated; individuals who experienced an infection prior to first healthy NP sample collection were excluded.

(C) Distribution of microbial events during infancy that were identified as risk factors for chronic wheeze at 5 years of age, stratified according to atopic status by age 2. fLRI, febrile LRI; wHRV-C, HRV-C LRI accompanied by wheeze; Strep, > 20% *Streptococcus* abundance in healthy NP sample taken in by 9 weeks old and prior to any ARI; unknown Strep, no such NP sample available (mainly due to ARI before 9 weeks of age); unknown LRI, incomplete viral/symptom profiling for LRI. Size of pie chart is proportional to the number of infants in each condition.

Importantly, our data show that the presence of *Moraxella* contributes to severity of RSV respiratory infections in infants (Figure S5, Table S3), which may be mechanistically related to RSV-*Moraxella* interactions reported in OM pathogenesis (Brockson et al., 2012).

Alloiococcus was a common and stable component of healthy NP samples and demonstrated enhanced stability in healthy NP microbiomes (Figure S3). *Alloiococcus* is a Gram-positive bacterium with one named species, *A. otitidis*, frequently detected in the ear canals of children with OM. Little is known about its mechanisms of colonization; however, the stable colonization patterns we observed may reflect a propensity to form stable biofilms, similar to *Moraxella*. The *Alloiococcus* MPG included some samples with high abundances of both *Alloiococcus* and *Corynebacterium*, indicating compatibility between these bacteria somewhat reminiscent of that between the *Moraxella* biofilm and *Streptococcus* or *Haemophilus*. Whether *Alloiococcus* is involved in OM pathology or is merely a healthy component of the microbiome is controversial (Tano et al., 2008); in our data, *Alloiococcus* was not associated with ARI or OM (Figure 1, Table S2).

Haemophilus was almost exclusively associated with ARI symptoms and was often found in consecutive infections, suggesting it can persist in the nasopharynx for some time (Figure 1, Table 1, Figure S3). Our analyses also suggest *Haemophilus* is highly transmissible between children and selected for by antibiotic exposure (Figure 2). Although our data include multiple *Haemophilus* OTUs, these epidemiological patterns are consistent with the human-adapted pathogen *H. influenzae*, which was supported by antibody data (Figure S2).

Role of the NP Microbiome in Asthma Development

We have previously reported that within the CAS cohort, the frequency during the first year of life of severe (wheezy and/or febrile) LRI was positively associated with subsequent risk for persistent wheeze at 5 and 10 years of age (Holt et al., 2010; Kusel et al., 2007, 2008), and similar findings have been reported in the U.S.-based COAST cohort (Gern, 2009). In both these study populations, the susceptible subgroup of children were those who developed early allergic sensitization (Gern, 2009; Holt et al., 2010, 2014; Jackson et al., 2008; Kusel et al., 2007, 2012; Oddy et al., 2002), and a variety of evidence suggests that the underlying mechanism involves synergistic interactions between atopic and anti-viral inflammatory pathways triggered within the infected airway mucosa (Holt and Sly, 2012). Of note, ARIs that remain restricted to the upper respiratory tract, and infections resulting in only mild LRI without wheezing or febrile symptoms, are relatively benign with respect to asthma risk in this cohort (Holt et al., 2010; Kusel et al., 2007, 2012)

(Table S3). Hence cofactors that can enhance the spread and ensuing severity of viral-initiated ARI are potentially central to asthma pathogenesis.

Here we found that two specific classes of LRI, namely febrile LRI and HRV-C-positive wheezy LRI, were independent risk factors for later chronic wheeze, especially among atopics (Table 2, Figure 3). Febrile LRI was common, occurring in one-third of all children and half of atopic chronic wheezers at age 5 (Figure 3 and Figure 5). Notably, the presence of illness-associated bacteria within the NP microbiome at the time of ARI increased both the risk for progression to the lower respiratory tract and the development of fever (Table 1). These mechanisms likely include myriad bacterial-viral interactions (recently reviewed by Vissers et al., 2014). In contrast, neither specific virus groups nor bacterial MPGs were associated with wheezy versus non-wheezy LRI within the overall population (Table 1), and the impact of wheezy LRI on later chronic wheeze was limited to LRI with HRV-C (Table 2), which occurred in just 11% of children. Re-examination of wheeze symptoms during LRI stratified by atopic status identified a positive association between HRV-C and wheeze during LRI among those children who later became sensitized (by 2 years of age), consistent with observations that HRV-C can induce wheeze in high-risk individuals (Lee et al., 2012). However, the presence of wheezy LRI alone, without further stratification by HRV-C detection and later atopic status, was not a significant predictor of later chronic wheeze, while febrile LRI per se was a significant predictor of chronic wheeze at 5 and 10 years, both among atopics and across the whole cohort (Figure 5, Table 2).

Early asymptomatic *Streptococcus* colonization at ~2 months of age, which occurred in 14% of children tested at that time, was significantly associated with chronic wheeze at 5 years of age (Figure 5, Table 2). Early *Streptococcus* colonization was not associated with incidence of infections with *Streptococcus* MPG, nor with detection of *S. pneumoniae*-specific antibodies at 12 months, suggesting that the mechanism is independent of innate immune response to *Streptococcus*. Rather, early *Streptococcus* colonization was associated with younger age of first LRI, and the level of subsequent asthma risk appears inversely related to age at initial *Streptococcus* colonization (Figure 4 and Figure 5). Collectively, these observations suggest that developing airway tissues may be maximally susceptible to the long-term effects of *Streptococcus*-mediated or LRI-mediated damage during the early postnatal period, when lung growth rates are most rapid.

To our knowledge, the only other study reporting the impact of early life NP colonization on wheeze/asthma at pre-school age is the 2007 report on the Copenhagen Prospective Study on Asthma in Childhood (n = 321, children born to mothers with asthma), in which a higher rate of wheeze at 5 years of age was detected among children colonized with *S. pneumoniae*, *M. catarrhalis*, and/or *H. influenzae* at 1 month post-birth (in culture-confirmed colonized, asthma prevalence was 33%, not colonized, 10%, OR 4.57) (Bisgaard et al., 2007). In that study, associations were not reported individually for the three species and colonization was measured as positive or negative by microbiological culture. In our cohort, wheeze at 5 years was not associated with early colonization with *Haemophilus* or *Moraxella*; however, *Haemophilus* was very rare in healthy NP samples, and *Moraxella* colonization was established later during infancy, which may be related to the warmer climate in Perth; these associations may differ in populations where asymptomatic carriage of these organisms is higher.

Conclusions and Implications

These findings collectively suggest that bacterial pathogens present in the NP microbiome at the time of upper respiratory viral infections during infancy are significant determinants of risk for the spread of infection to the lower airways and for the resultant expression of inflammatory symptoms marked by fever, and further, they may contribute directly and indirectly to the ensuing risk for development of persistent asthma, which itself is linked to these prior infectious episodes. Prevention of RSV or HRV infections in high-risk children, using immunoprophylaxis or vaccines, has been proposed as a mechanism for preventing the development and/or exacerbation of childhood asthma (Gern, 2009; Wu and Hartert, 2011). Our data suggest that in the absence of effective anti-viral therapies, targeting pathogenic bacteria present within the NP microbiome in this age group could represent an alternative approach toward the same goal. The pneumococcal vaccine, currently recommended from 2 months of age, could play a role; however, the niche created by eliminating vaccine-targeted *S. pneumoniae* serotypes can be readily filled by other serotypes and other bacteria such as *H. influenzae* and *M. catarrhalis* (Biesbroek et al., 2014; Revai et al., 2006). Manipulation of the microbiome by antibiotics may appear attractive; however, the association between antibiotic use in early childhood and subsequent asthma is controversial. In the CAS cohort (Kusel et al., 2008) and others (Semic-Jusufagic et al., 2014), this association has been proposed to arise through confounding and reverse causation, whereby genetic or clinical factors increase the likelihood of both antibiotic prescription (e.g., via genetic susceptibility to viral infection [Semic-Jusufagic et al., 2014]) and asthma development; others have proposed that antibiotic-induced disruption of the gut microbiota could explain the link (Arrieta and Finlay, 2014). Our analyses provide a potential causal pathway linking antibiotics to later asthma, whereby antibiotic use in infants selects for illness-associated bacteria in the NP microbiome, leading to increased risk of febrile LRI and later asthma development. Genetic analysis of the infants in our cohort could potentially help to unravel this in the future. Importantly, further studies are required to prospectively assess the impact of antibiotic use on NP bacterial colonization during health and infection, and the subsequent development of ARI, atopy, and wheeze.

EXPERIMENTAL PROCEDURES

Study Design

This study is an extension of the Childhood Asthma Study (CAS), a birth cohort of 234 infants at high risk of atopy as previously described (Kusel et al., 2007, 2008). Briefly, healthy NP aspirates (NPAs) were collected from subjects at planned visits at 2, 6, and 12 months of age. NPAs were also collected within 48 hr from the onset of an ARI; these were classified as either lower respiratory infection (LRI, if wheeze or rattly chest detected) or upper respiratory infection (URI, otherwise). NPAs were divided into four aliquots and stored at −80°C. One aliquot was used for bacterial 16S rRNA amplicon sequencing via Illumina MiSeq and another to screen for viruses including RSV, HRV, and other picornaviruses, influenza, parainfluenza, coronavirus, adenovirus, and human

metapneumovirus (hMPV). For LRI, a third aliquot was used in an expanded screen for detection and subtyping of HRV. Study ethics were approved by the ethics committees of King Edward Memorial and Princess Margaret Hospitals in Western Australia. Fully informed parental consent was obtained for all subjects. Details of sample and data collection, viral screening, DNA extraction, and sequencing are in Supplemental Experimental Procedures.

16S rRNA Sequence Analysis

Paired end reads were merged using Flash v1.2.7 (Magoč and Salzberg, 2011) with read length 151 base pairs (bp) and expected fragment length 253 bp. Merged sequences were quality filtered as follows: $\leq$ 3 low-quality bp (Phred score < 3) allowed before trimming, $\geq$ 189 consecutive high-quality bp with no uncalled bases (Ns) (Bokulich et al., 2013). Thirty-three million sequences were filtered out (13%), leaving 219 million for analysis. These were assigned to OTUs using the closed reference method in QIIME v1.8 (Caporaso et al., 2010) with the Greengenes 99% OTU reference set, version 13_5 (McDonald et al., 2012) (which consists of > 200,000 representative sequences obtained from clustering all sequences from the Greengenes reference database at 99% sequence similarity). Briefly, this method uses UCLUST (Edgar, 2010) to search each sequence against the reference set and assigns it to an OTU based on the best hit at $\geq$ 99% sequence identity. Sequences not matching the reference database (3%) were excluded from analysis; these sequences could be chimeras, sequencing errors, or novel sequences that are not well characterized in the database. This left > 200,000 taxonomy-assigned sequences on average for each NPA (interquartile range: 108,000–255,000; eight samples with < 1,000 reads). Compared to NPAs, negative control samples (see Supplemental Experimental Procedures) had clearly lower total reads (interquartile range: 1,044–2,346) with a different community structure (Figure S6).

Clustering into MPGs

For each NPA, the relative abundance of each OTU was calculated. Most analyses were summarized at genus level, whereby all OTUs assigned to the same genus were collapsed into a single group for reporting. Samples were assigned to microbiome profile groups (MPGs) based on hierarchical clustering of the relative abundances of the six most common genera (distance metric: 1-Pearson's correlation; clustering method: Ward's minimum variance, implemented in *R hclust*). Previous microbiome studies have reported clinically or environmentally meaningful associations with community profiles or types (e.g., enterotypes) identified using similar clustering approaches (Ravel et al., 2011; Zhou et al., 2014).

Statistical Methods

All statistical analyses were performed using *R*. ORs were estimated using generalized estimating equations (GEE) logistic regression with unstructured correlation and robust standard errors, to take into account multiple samples from the same subject. Survival analyses were done using Cox proportional hazards models. Potential confounders were adjusted for by including them in the regression models. Full details of each analysis and variable definitions are given in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

Short read 16S data for this study have been deposited in NCBI GenBank under accession number SRP056779.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2015.03.008.

ACKNOWLEDGMENTS

This study was supported by the NHMRC of Australia (Project Grant #1049539; Fellowships #1061409 [K.E.H.] and #1061435 [M.I., co-funded with the Australian Heart Foundation]) and the Victorian Life Sciences Computation Initiative (VLSCI) (#VR0082). S.L.J. is supported in part by a Chair from Asthma UK (CH11SJ) and by Medical Research Council Centre grant G1000758. J.E.G., Y.A.B., and K.G. are supported by NIH grants U19 AI104317 and P01 HL070831.

Received: January 6, 2015
Revised: February 9, 2015
Accepted: March 9, 2015
Published: April 9, 2015

REFERENCES

Arrieta, M.C., and Finlay, B. (2014). The intestinal microbiota and allergic asthma. J. Infect. *69* (1), S53–S55.

Biesbroek, G., Wang, X., Keijser, B.J., Eijkemans, R.M., Trzciński, K., Rots, N.Y., Veenhoven, R.H., Sanders, E.A., and Bogaert, D. (2014). Seven-valent pneumococcal conjugate vaccine and nasopharyngeal microbiota in healthy children. Emerg. Infect. Dis. *20*, 201–210.

Bisgaard, H., Hermansen, M.N., Buchvald, F., Loland, L., Halkjaer, L.B., Bønnelykke, K., Brasholt, M., Heltberg, A., Vissing, N.H., Thorsen, S.V., et al. (2007). Childhood asthma after bacterial colonization of the airway in neonates. N. Engl. J. Med. *357*, 1487–1495.

Bisgaard, H., Hermansen, M.N., Bønnelykke, K., Stokholm, J., Baty, F., Skytt, N.L., Aniscenko, J., Kebadze, T., and Johnston, S.L. (2010). Association of bacteria and viruses with wheezy episodes in young children: prospective birth cohort study. BMJ *341*, c4978.

Bochkov, Y.A., and Gern, J.E. (2012). Clinical and molecular features of human rhinovirus C. Microbes Infect. *14*, 485–494.

Bogaert, D., Keijser, B., Huse, S., Rossen, J., Veenhoven, R., van Gils, E., Bruin, J., Montijn, R., Bonten, M., and Sanders, E. (2011). Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. PLoS ONE *6*, e17035.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A., and Caporaso, J.G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods *10*, 57–59.

Brockson, M.E., Novotny, L.A., Jurcisek, J.A., McGillivary, G., Bowers, M.R., and Bakaletz, L.O. (2012). Respiratory syncytial virus promotes Moraxella catarrhalis-induced ascending experimental otitis media. PLoS ONE *7*, e40088.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

de Vries, S.P., Bootsma, H.J., Hays, J.P., and Hermans, P.W. (2009). Molecular aspects of Moraxella catarrhalis pathogenesis. Microbiol. Mol. Biol. Rev. *73*, 389–406.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

Gern, J.E. (2009). Rhinovirus and the initiation of asthma. Curr. Opin. Allergy Clin. Immunol. *9*, 73–78.

Gisselsson-Solén, M., Henriksson, G., Hermansson, A., and Melhus, A. (2014). Risk factors for carriage of AOM pathogens during the first 3 years of life in children with early onset of acute otitis media. Acta Otolaryngol. *134*, 684–690.

Hales, B.J., Chai, L.Y., Elliot, C.E., Pearce, L.J., Zhang, G., Heinrich, T.K., Smith, W.A., Kusel, M.M., Holt, P.G., Sly, P.D., and Thomas, W.R. (2012). Antibacterial antibody responses associated with the development of asthma in house dust mite-sensitised and non-sensitised children. Thorax *67*, 321–327.

Hilty, M., Burke, C., Pedro, H., Cardenas, P., Bush, A., Bossley, C., Davies, J., Ervine, A., Poulter, L., Pachter, L., et al. (2010). Disordered microbial communities in asthmatic airways. PLoS ONE *5*, e8578.

Holt, P.G., and Sly, P.D. (2012). Viral infections and atopy in asthma pathogenesis: new rationales for asthma prevention and treatment. Nat. Med. *18*, 726–735.

Holt, P.G., Rowe, J., Kusel, M., Parsons, F., Hollams, E.M., Bosco, A., McKenna, K., Subrata, L., de Klerk, N., Serralha, M., et al. (2010). Toward improved prediction of risk for atopy and asthma among preschoolers: a prospective cohort study. J. Allergy Clin. Immunol. *125*, 653–659, e1, e7.

Holt, P.G., Strickland, D.H., Hales, B.J., and Sly, P.D. (2014). Defective respiratory tract immune surveillance in asthma: a primary causal factor in disease onset and progression. Chest *145*, 370–378.

Jackson, D.J., Gangnon, R.E., Evans, M.D., Roberg, K.A., Anderson, E.L., Pappas, T.E., Printz, M.C., Lee, W.M., Shult, P.A., Reisdorf, E., et al. (2008). Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children. Am. J. Respir. Crit. Care Med. *178*, 667–672.

Kusel, M.M., de Klerk, N.H., Holt, P.G., Kebadze, T., Johnston, S.L., and Sly, P.D. (2006). Role of respiratory viruses in acute upper and lower respiratory tract illness in the first year of life: a birth cohort study. Pediatr. Infect. Dis. J. *25*, 680–686.

Kusel, M.M., de Klerk, N.H., Kebadze, T., Vohma, V., Holt, P.G., Johnston, S.L., and Sly, P.D. (2007). Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma. J. Allergy Clin. Immunol. *119*, 1105–1110.

Kusel, M.M., de Klerk, N., Holt, P.G., and Sly, P.D. (2008). Antibiotic use in the first year of life and risk of atopic disease in early childhood. Clin. Exp. Allergy *38*, 1921–1928.

Kusel, M.M., Kebadze, T., Johnston, S.L., Holt, P.G., and Sly, P.D. (2012). Febrile respiratory illnesses in infancy and atopy are risk factors for persistent asthma and wheeze. Eur. Respir. J. *39*, 876–882.

Lee, W.M., Lemanske, R.F., Jr., Evans, M.D., Vang, F., Pappas, T., Gangnon, R., Jackson, D.J., and Gern, J.E. (2012). Human rhinovirus species and season of infection determine illness severity. Am. J. Respir. Crit. Care Med. *186*, 886–891.

Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. *6*, 610–618.

Oddy, W.H., de Klerk, N.H., Sly, P.D., and Holt, P.G. (2002). The effects of respiratory infections, atopy, and breastfeeding on childhood asthma. Eur. Respir. J. *19*, 899–905.

Penders, J., Kummeling, I., and Thijs, C. (2011). Infant antibiotic use and wheeze and asthma risk: a systematic review and meta-analysis. Eur. Respir. J. *38*, 295–302.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S., McCulle, S.L., Karlebach, S., Gorle, R., Russell, J., Tacket, C.O., et al. (2011). Vaginal microbiome of reproductive-age women. Proc. Natl. Acad. Sci. USA *108* (1), 4680–4687.

Revai, K., McCormick, D.P., Patel, J., Grady, J.J., Saeed, K., and Chonmaitree, T. (2006). Effect of pneumococcal conjugate vaccine on nasopharyngeal bacterial colonization during acute otitis media. Pediatrics *117*, 1823–1829.

Rollins, D.R., Beuther, D.A., and Martin, R.J. (2010). Update on infection and antibiotics in asthma. Curr. Allergy Asthma Rep. *10*, 67–73.

Schaumburg, F., Alabi, A.S., Mombo-Ngoma, G., Kaba, H., Zoleko, R.M., Diop, D.A., Mackanga, J.R., Basra, A., Gonzalez, R., Menendez, C., et al. (2014). Transmission of Staphylococcus aureus between mothers and infants in an African setting. Clin. Microbiol. Infect. *20*, O390–O396.

Semic-Jusufagic, A., Belgrave, D., Pickles, A., Telcian, A.G., Bakhsoliani, E., Sykes, A., Simpson, A., Johnston, S.L., and Custovic, A. (2014). Assessing the association of early life antibiotic prescription with asthma exacerbations, impaired antiviral immunity, and genetic variants in 17q21: a population-based birth cohort study. Lancet Respir Med *2*, 621–630.

Sly, P.D., Boner, A.L., Björksten, B., Bush, A., Custovic, A., Eigenmann, P.A., Gern, J.E., Gerritsen, J., Hamelmann, E., Helms, P.J., et al. (2008). Early identification of atopy in the prediction of persistent asthma in children. Lancet *372*, 1100–1106.

Spaniol, V., Troller, R., Schaller, A., and Aebi, C. (2011). Physiologic cold shock of Moraxella catarrhalis affects the expression of genes involved in the iron acquisition, serum resistance and immune evasion. BMC Microbiol. *11*, 182.

Stein, R.T., and Martinez, F.D. (2010). Respiratory syncytial virus and asthma: still no final answer. Thorax *65*, 1033–1034.

Tai, A., Tran, H., Roberts, M., Clarke, N., Wilson, J., and Robertson, C.F. (2014). The association between childhood asthma and adult chronic obstructive pulmonary disease. Thorax *69*, 805–810.

Tano, K., von Essen, R., Eriksson, P.O., and Sjöstedt, A. (2008). Alloiococcus otitidis—otitis media pathogen or normal bacterial flora? APMIS *116*, 785–790.

Verhaegh, S.J., Snippe, M.L., Levy, F., Verbrugh, H.A., Jaddoe, V.W., Hofman, A., Moll, H.A., van Belkum, A., and Hays, J.P. (2011). Colonization of healthy children by Moraxella catarrhalis is characterized by genotype heterogeneity, virulence gene diversity and co-colonization with Haemophilus influenzae. Microbiology *157*, 169–178.

Vissers, M., de Groot, R., and Ferwerda, G. (2014). Severe viral respiratory infections: are bugs bugging? Mucosal Immunol. *7*, 227–238.

Weinstock, G.M. (2012). Genomic approaches to studying the human microbiota. Nature *489*, 250–256.

Wu, P., and Hartert, T.V. (2011). Evidence for a causal relationship between respiratory syncytial virus infection and asthma. Expert Rev. Anti Infect. Ther. *9*, 731–745.

Zhou, Y., Mihindukulasuriya, K.A., Gao, H., La Rosa, P.S., Wylie, K.M., Martin, J.C., Kota, K., Shannon, W.D., Mitreva, M., Sodergren, E., and Weinstock, G.M. (2014). Exploration of bacterial community classes in major human habitats. Genome Biol. *15*, R66.

Cell Host & Microbe

Supplemental Information

The Infant Nasopharyngeal Microbiome Impacts the Incidence of Lower Respiratory Infection and Risk of Asthma Development

Shu Mei Teo, Danny Mok, Kym Pham, Merci Kusel, Michael Serralha, Niamh Troy, Barbara J. Holt, Belinda J. Hales, Michael L. Walker, Elysia Hollams, Yury A Bochkov, Kristine Grindle, Sebastian L. Johnston, James E Gern, Peter D. Sly, Patrick G. Holt, Kathryn E. Holt, and Michael Inouye

Supplemental Data

Figure S1, related to Figure 1. Subtree of the Greengenes 99% sequence similarity OTU tree for genus (**A**) *Haemophilus.1,* (**B**) *Moraxella,* (**C**) *Streptococcus,* (**D**) *Alloiococcus,* (**E***) Corynebacterium,* (**F**) *Staphylococcus*. OTUs present at >0.1% frequency across our 1,021 NP samples are highlighted, and relative frequencies within these samples are shown with blue bars (right).

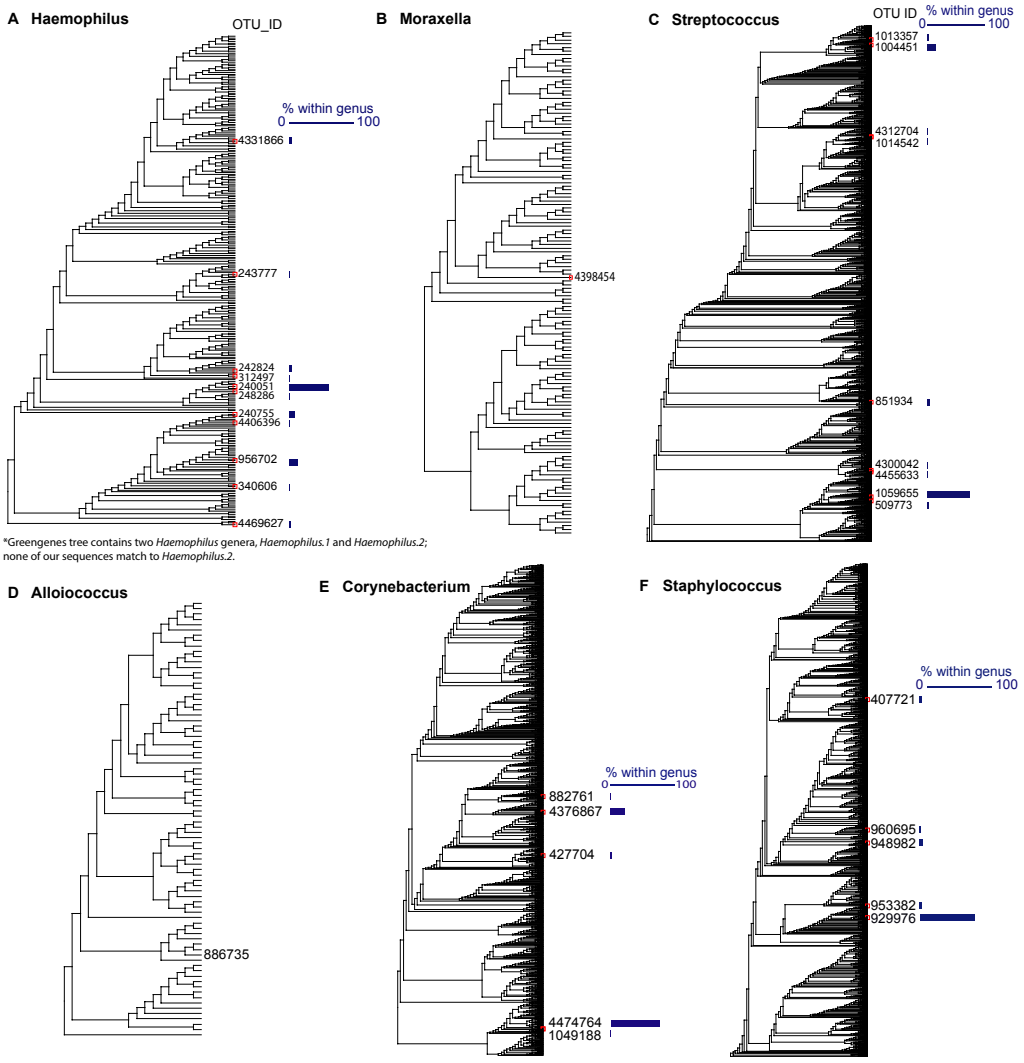
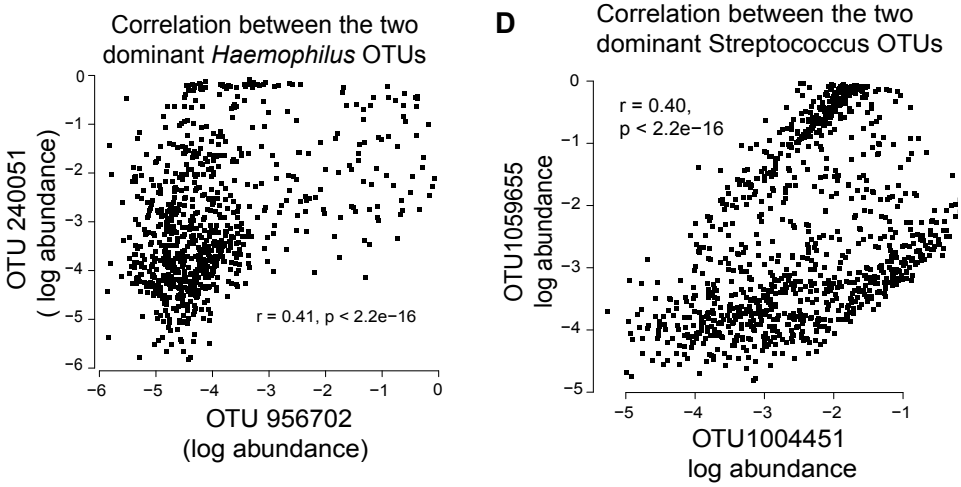


Figure S2, related to Figure 1. Association with species-specific antibodies.
(A) Correlation between the two most frequent *Haemophilus* OTUs. (**B-C**) Association between detectable IgG1/IgG4 antibodies to *H. influenzae* P4/P6 surface proteins (measured at 12 months of age) and number of (**B**) prior ARI NP samples or (**C**) prior asymptomatic NP samples containing the dominant *Haemophilus* OTUs. **(D)** Correlation between the two most frequent *Streptococcus* OTUs. **(E-F)** Association between detectable IgG1 antibodies to *S. pneumoniae* pneumococcal surface protein A1, A2, or C (measured at 12 months of age) and number of **(E)** prior ARI NP samples or **(F)** prior asymptomatic NP samples containing the dominant *Streptococcus* OTUs.



B Association between detectable *H. influenzae* antibodies at 12 months and number of prior infections with *Haemophilus* OTUs.

Presence of OTU	OR (95% CI)
240051	1.9 (1.3 - 2.9)**
956702	3.6 (1.8 - 6.9)**

** p < 0.001

C Association between detectable *H. influenzae* antibodies at 12 months and any prior asymptomatic NP samples with *Haemophilus* OTUs.

Presence of OTU	OR (95% CI)
240051	1.2 (0.67 - 2.0)
956702	2.1 (0.97 - 4.5)

E Association between detectable *S.pneumoniae* antibodies at 12 months and number of prior infections with *Streptococcus* OTUs.

Presence of OTU	OR (95% CI)
1059655	2.1 (1.4 - 3.2)**
1004451	1.3 (0.81 - 2.0)

** p < 0.001

F Association between detectable *S.pneumoniae* antibodies at 12 months and any prior asymptomatic NP samples with *Streptococcus* OTUs.

Presence of OTU	OR (95% CI)
1059655	1.1 (0.61 - 1.9)
1004451	0.50 (0.24 - 1.0)

Figure S3, related to section "*NP microbiome dynamics*". Transitions within subjects. (A-C) Transitions between **(A)** consecutive healthy samples, with no intervening infections, **(B)** with intervening infections, and **(C)** between consecutive infection samples. T1, 1st timepoint; T2, next timepoint. Cell numbers indicate the number of cases in which the respective transition from T1 to T2 was observed; cells are coloured to indicate the row proportions as per legend. Stable transitions were defined as those with the same MPG at T1 and T2 (i.e. diagonals in the matrix). For each panel, the table at the right shows observed and expected frequencies of stable transitions for each genus MPG (note *Haemophilus* is not plotted in (A) and (B) because it was extremely rare in healthy samples). The expected frequency for each stable transition was taken to be the square of the proportion of samples in that MPG at T1 (*i.e.,* assuming constant frequencies and random transitions). Observed values are the proportion of stable transitions out of all observed transitions; 95% confidence intervals were calculated from 1,000 bootstraps of the real data, sampled with replacement. **(D)** Frequency of stable transitions between MPGs in consecutive ARIs, stratified by time between ARIs. Stable transitions were defined as pairs of consecutive ARIs from the same individual that shared the same MPG.

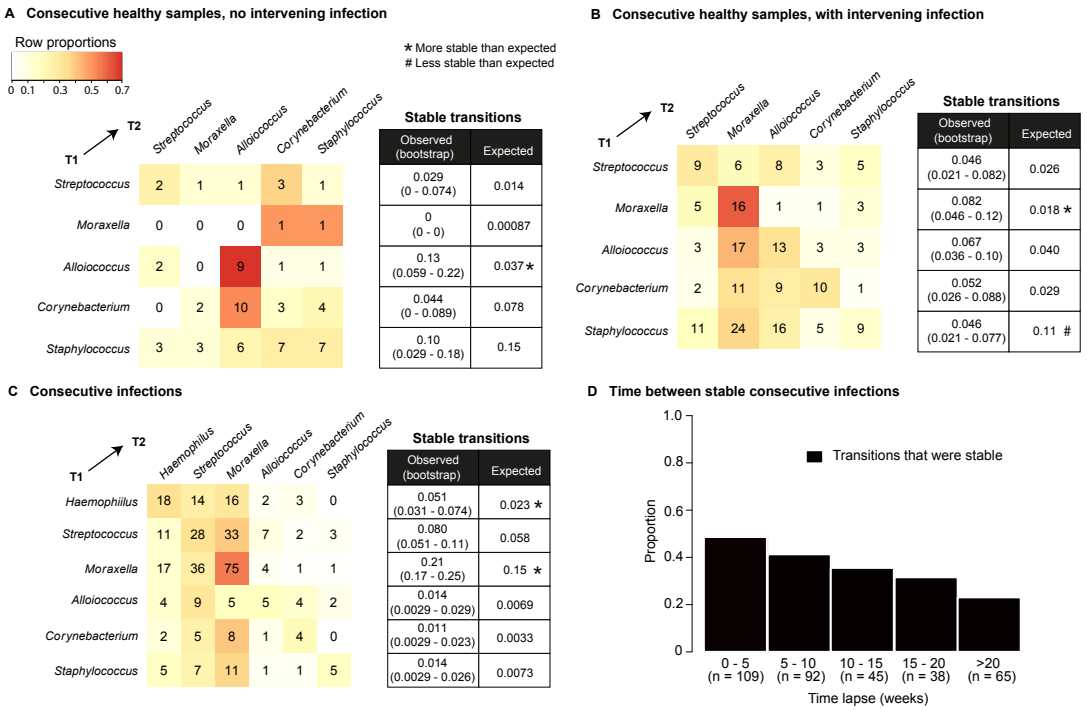


Figure S4, related to section "*NP microbial determinants of ARI symptoms*". Rates of bacterial microbiome profile groups (MPGs) and viruses for different types of NP samples. ARI, acute respiratory illness; URI, upper respiratory illness; LRI, lower respiratory illness; fLRI, febrile LRI; wLRI, wheezy LRI. Note HRV subtyping was performed for LRI samples only.

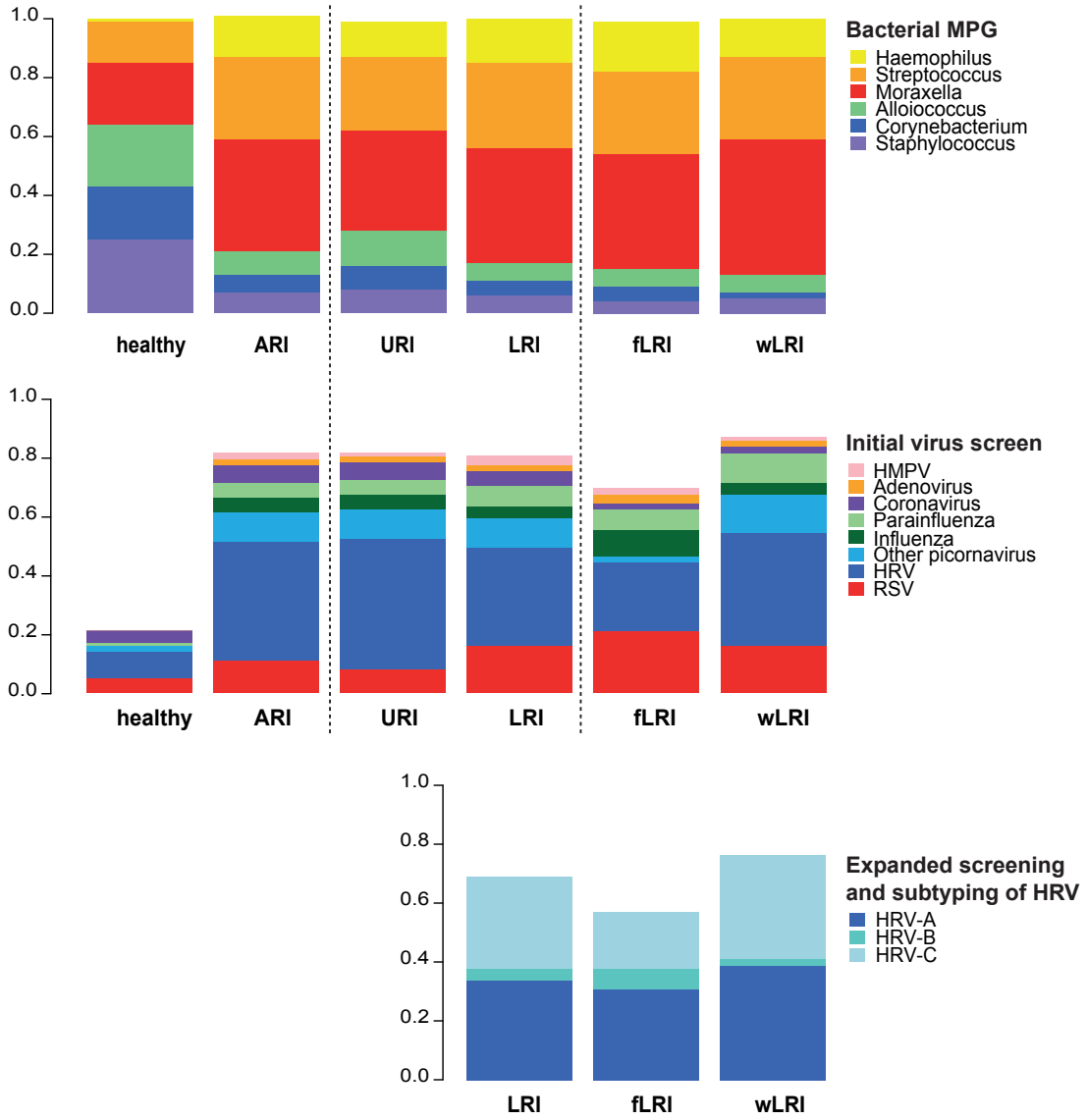


Figure S5, related to Table 1. Distribution of samples by MPG, amongst (A) RSV positive and (B) RSV negative samples. Logistic regression was used to assess, amongst (A) RSV-positive LRIs, and (B) RSV-negative LRIs, the association of fever status with *Moraxella*, *Streptococcus* and *Haemophilus* MPGs as independent covariates in the same model, adjusted for age, gender and season (p-values shown). Samples with antibiotics intake within the last week were excluded from this analysis.

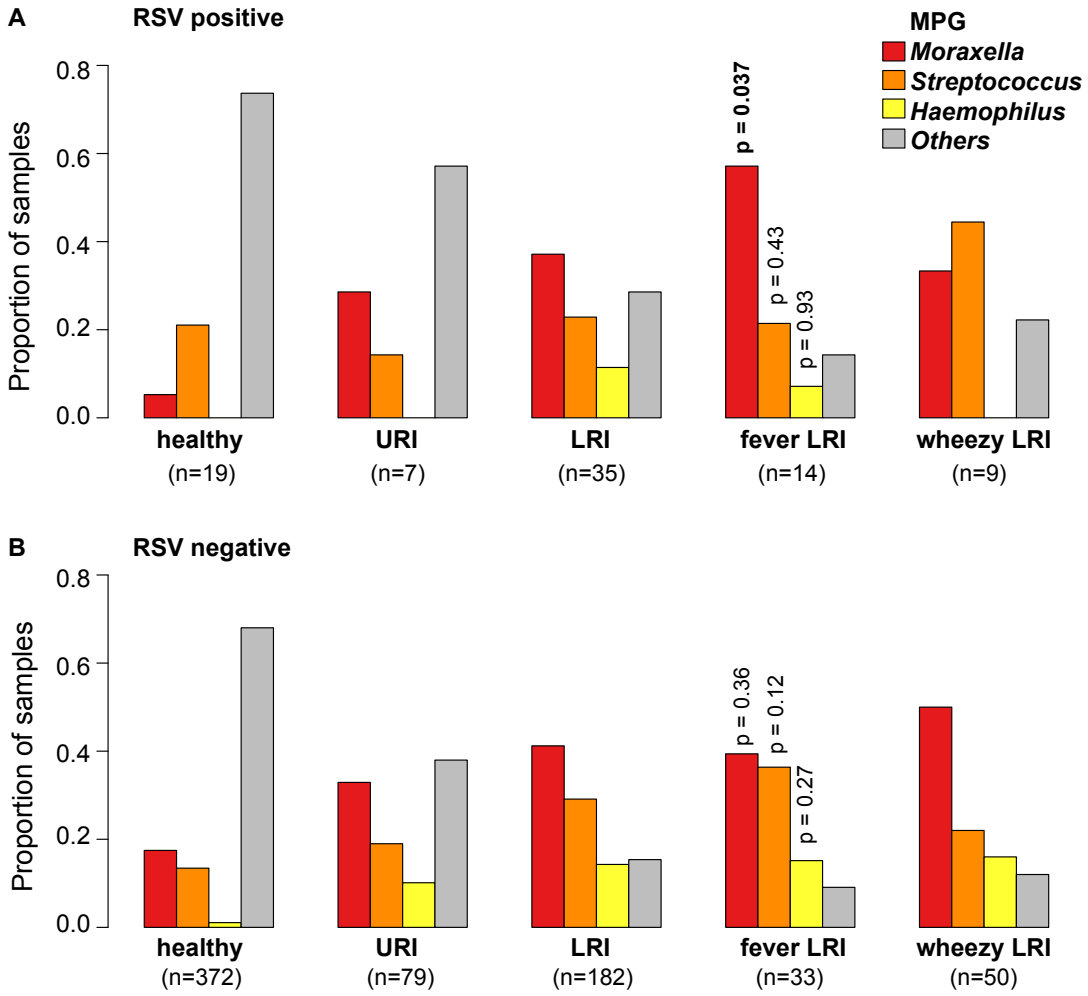
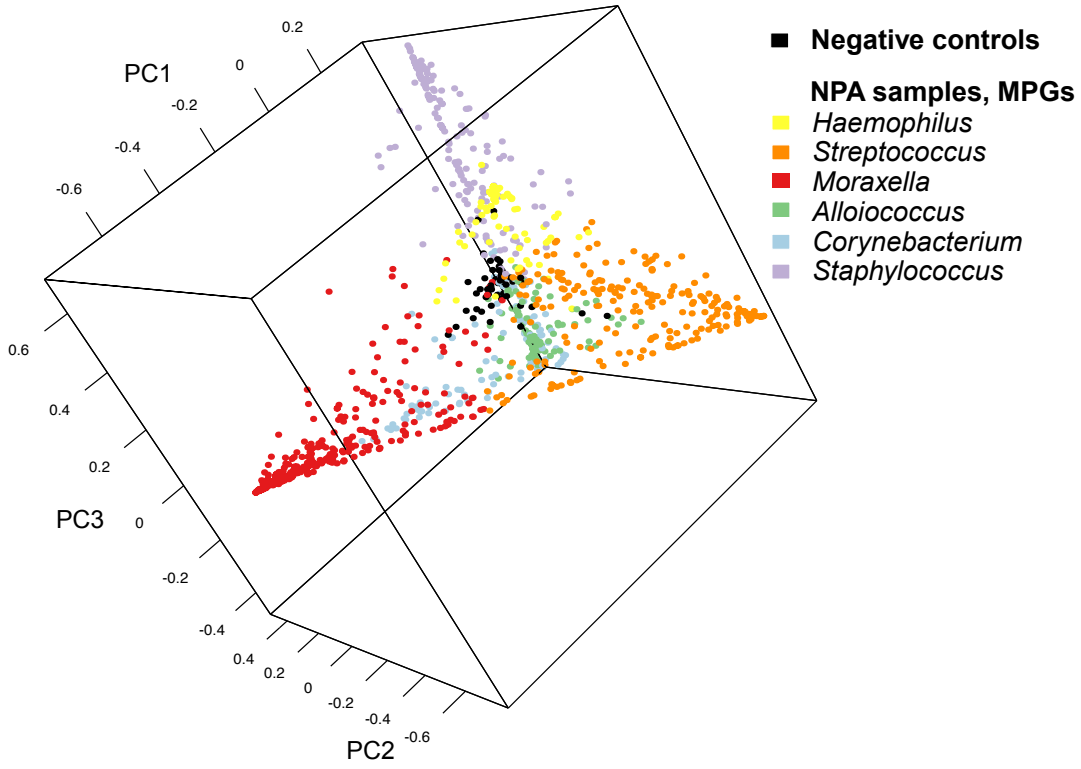


Figure S6, related to "Experimental procedures". Principal components analysis of OTU relative abundances for NPA samples and negative controls



Characteristics of subjects	Number	Proportion
Total participants	234	100%
Male gender	132	56%
Paternal history of atopic disease	156	67%
Maternal history of atopic disease	196	84%
Vaginal delivery	167	72%
Breastfed at 3 months	196	84%
Smoking exposure at 1 year	44	19%
Day-care attendance at 1 year	67	29%
Current wheeze at 1 year	76	33%
Other children in house	108	47%
Furry pet in the first year	109	48%
≥1 LRI in first 425 days of life	167	75%
Atopic at 6 months of age	47	21%
Atopic at 2 years of age	88	41%
Sensitised to inhaled allergens[#] by 1 year of age	23	10%
Current wheeze at 5 years of age	56	28%

Table S1. Characteristics of infants in the study. [#] Inhaled allergens include house dust mite, cat epithelium and dander, couch grass, rye grass, or mould mix.

MPG	Number of OM ARI (LRI)	OM vs. non-OM ARI OR (95% CI)	OM vs. non-OM LRI OR (95% CI)
Staphylococcus	3 (3)	2.2 (0.52-8.9) p = 0.29	2.0 (0.51-8.0) p = 0.32
Corynebacterium	1 (0)	0.27 (0.036-2.1) p = 0.21	0.32 (0.041-2.4) p = 0.27
Alloiococcus	2 (2)	0.36 (0.089-1.5) p = 0.16	0.44 (0.1-1.9) p = 0.28
Moraxella	25 (16)	1.1 (0.61-2.0) p = 0.71	1.0 (0.58-1.8) p = 0.92
Streptococcus	21 (17)	1.2 (0.66-2.3) p = 0.50	1.2 (0.64-2.3) p = 0.55
Haemophilus	9 (5)	0.90 (0.33-2.5) p = 0.84	0.83 (0.31-2.2) p = 0.70

Table S2, related to section *"NP microbial determinants of ARI symptoms"*. Distribution of microbiome profile groups (MPGs) among ARI with otitis media (OM) diagnosis. Odds ratios (OR) for association of each MPG with OM, estimated using generalized estimating equations (GEE) logistic regression with unstructured correlation and robust standard errors, adjusted for age at infection. ARI, acute respiratory infection; LRI, lower respiratory illness.

	All children		Atopic by 2 years		Not atopic by 2 years	
	Wheeze at 5 years	**Wheeze at 10 years**	**Wheeze at 5 years**	**Wheeze at 10 years**	**Wheeze at 5 years**	**Wheeze at 10 years**
Any URI	0.81 (0.21-4) p = 0.78	0.71 (0.14-5.3) p = 0.69	0.42 (0.067-2.6) p = 0.33	0.48 (0.068-4.1) p = 0.46	NA	NA
Any LRI	0.95 (0.47-2) p = 0.89	1.9 (0.74-5.6) p = 0.2	1.4 (0.51-3.8) p = 0.55	1.9 (0.61-6.6) p = 0.29	0.57 (0.18-1.9) p = 0.35	2.6 (0.38-52) p = 0.41
Any fever LRI	**2.3 (1.2-4.5) p = 0.016**	2.2 (0.93-5.2) p = 0.071	**2.7 (1.1-7) p = 0.034**	2.7 (0.89-8.7) p = 0.083	1.5 (0.48-4.6) p = 0.46	1.9 (0.33-9.8) p = 0.45
Any wheezy LRI	1.6 (0.79-3) p = 0.2	1.4 (0.58-3.3) p = 0.45	1.3 (0.49-3.3) p = 0.61	1.6 (0.53-4.8) p = 0.4	1.9 (0.66-5.7) p = 0.23	1.2 (0.22-5.7) p = 0.83
Any mild LRI	0.48 (0.24-0.94) p = 0.035	0.92 (0.4-2.1) p = 0.84	0.51 (0.2-1.2) p = 0.14	0.76 (0.27-2.1) p = 0.61	0.44 (0.13-1.3) p = 0.16	1.6 (0.35-9.1) p = 0.54
Any HRV LRI	1.1 (0.59-2.2) p = 0.69	1.6 (0.69-3.8) p = 0.28	1.2 (0.49-2.9) p = 0.7	1.2 (0.43-3.3) p = 0.75	0.93 (0.31-2.7) p = 0.9	2.8 (0.59-16) p = 0.21
Any RSV LRI	1.1 (0.51-2.4) p = 0.76	1.4 (0.51-3.5) p = 0.5	2.2 (0.76-6.2) p = 0.14	2.7 (0.8-8.9) p = 0.11	0.56 (0.11-2.1) p = 0.42	0.33 (0.016-2.5) p = 0.35
Any risk bacteria LRI	0.89 (0.45-1.8) p = 0.73	2 (0.82-5.2) p = 0.14	0.96 (0.39-2.4) p = 0.93	1.7 (0.59-5.2) p = 0.34	0.73 (0.25-2.3) p = 0.57	4.9 (0.72-98) p = 0.16
Any HRV wheezy LRI	2 (0.93-4.2) p = 0.073	2.1 (0.81-5.4) p = 0.11	2.5 (0.86-7.2) p = 0.092	1.9 (0.55-6.4) p = 0.29	1.8 (0.49-6.3) p = 0.34	2.7 (0.46-14) p = 0.24
Any HRV-C wheezy LRI	2.4 (0.93-6.1) p = 0.064	**3.5 (1.1-11) p = 0.026**	**7.2 (1.7-35) p = 0.009**	**7.1 (1.6-40) p = 0.014**	1.1 (0.15-5) p = 0.95	1.2 (0.058-11) p = 0.86
Any HRV-A wheezy LRI	1.2 (0.42-3.1) p = 0.74	1.4 (0.37-4.7) p = 0.57	0.55 (0.11-2.2) p = 0.43	0.35 (0.018-2.2) p = 0.34	3.5 (0.74-16) p = 0.10	6.3 (0.94-44) p = 0.051
Any RSV wheezy LRI	2.6 (0.71-9.5) p = 0.15	1.7 (0.34-7.5) p = 0.47	7.2 (1-150) p = 0.084	2.5 (0.28-20) p = 0.37	0.95 (0.045-8) p = 0.97	1.2 (0.052-12) p = 0.89
***Streptococcus* colonization at 7 weeks**	**3.8 (1.3-12) p = 0.017**	2.7 (0.66-12) p = 0.18	4.0 (0.88-21) p = 0.077	3.9 (0.63-28) p = 0.15	3.4 (0.55-22) p = 0.17	NA
***Streptococcus* colonization at 8 weeks**	**3.1 (1.2-8.2) p = 0.023**	1.9 (0.46-7.1) p = 0.33	**4.4 (1.1-20) p = 0.040**	2.4 (0.46-12) p = 0.29	2.7 (0.47-14) p = 0.24	NA
***Streptococcus* colonization at 9 weeks**	**2.7 (1.0-7.1) p = 0.039**	1.9 (0.45-6.9) p = 0.36	3.3 (0.86-13) p = 0.080	2.4 (0.47-13) p = 0.28	2.6 (0.45-14) p = 0.26	NA

Table S3, related to Table 2. Association between chronic wheeze at age 5 and 10 years, by LRI subtypes and early asymptomatic colonization with *Streptococcus*. Any mild LRI: Any LRI without wheeze or fever symptoms; Any risk bacteria LRI: any LRI in *Streptococcus*, *Moraxella* or *Haemophilus* MPG. *Streptococcus* colonization in the first healthy sample (collected by 7/8/9 weeks of age and prior to any respiratory infection) was assessed using >20% abundance cut-off. Odds ratios, 95% confidence intervals and p-values shown were estimated using logistic regression, adjusted for gender and maternal and paternal history of atopic disease, calculated separately for all children, those who were atopic by 2 years and those not atopic by 2 years.

Variable	*Alloiococcus* colonized (n = 63)	*Moraxella* colonized (n = 64)	Others (n = 107)	*Alloiococcus* vs. *Moraxella* (p-value)	*Alloiococcus* vs. all others (p-value)
Mean no. ARI	4.6	5.0	5.4	0.31	0.077
Mean no. LRI	1.3	1.7	1.9	0.32	0.10
Mean no. fLRI	0.44	0.49	0.51	0.53	0.47
Mean no. wLRI	0.32	0.62	0.53	0.14	0.11
Mean no. OM	0.44	0.57	0.47	0.29	0.54
Any RSV+ ARI	18	30	43	0.094	0.12
Any RSV+ LRI	5	17	27	0.017*	0.0050*
Any RSV+ fLRI	1	8	10	0.034*	0.050*
Wheeze at age 5 y **Atopic** **Non-atopic** **No wheeze**	 8 4 44	 11 6 38	 15 10 56	0.45#	0.41#
Wheeze at age 10 y **Atopic** **Non-atopic** **No wheeze**	 5 1 36	 4 2 35	 13 5 43	1#	0.46#
Atopic by age 2 y	33	32	51	1	0.88

Table S4, related to section *"Impact of NP colonization on ARI"*. Infections and wheeze phenotypes for infants by colonization status. Groups are defined by microbiome profile group (MPG) clustering of healthy samples. '*Alloiococcus* colonized', ≥1 healthy sample with *Alloiococcus* MPG but none with *Moraxella* MPG; '*Moraxella* colonized', ≥1 healthy sample with *Moraxella* MPG but none with *Alloiococcus* MPG; 'Others', no *Moraxella* or *Alloiococcus* healthy sample or ≥1 of each. Infection types: ARI, acute respiratory infection; URI, upper respiratory illness; LRI, lower respiratory illness; fLRI, febrile LRI; wLRI, wheezy LRI; OM, otitis media; RSV+, PCR detection of respiratory syncytial virus; atopy status is defined by IgE > 0.35 kU/L for any antigen at 6, 12 or 24 months. Comparisons of rates of infection were assessed using Wilcoxon rank sum test; other comparisons are binary variables and were calculated using Fisher's exact test; # atopic wheeze vs no wheeze; *$p < 0.05$.

Supplemental Experimental Procedures:

Study design and participants

This study is an extension of the Childhood Asthma Study (CAS) – a prospective community-based birth cohort study of 234 infants at high risk of atopy (at least 1 parent with a doctor diagnosed history of hay fever, asthma or eczema). Parents of the subjects completed questionnaires prior to birth, and when each child reached 1, 2, 3, 4, 5 and 10 years of age. Parents were also asked to complete a daily diary for the first five years, recording information on breastfeeding, respiratory symptoms (fever, runny nose, cough, wheeze), and use of any medications by both mother and child. Healthy nasopharyngeal aspirates (NPAs) were collected from subjects by study clinicians during planned visits at approximately 2 months, 6 months and 12 months of age, after the child had been free from any symptoms of respiratory illness for a period of at least 4 weeks. Parents were also asked to report to the study clinicians whenever the child showed symptoms of an acute respiratory illness (ARI), at which point the family was visited within 48 hours by a study nurse who recorded clinical details of the infection and collected an NPA from the child. Clinical data recorded included the presence of fever, wheeze or rattly chest and any medications taken (including antibiotics). Each ARI was classified as either a lower respiratory illness (LRI; if wheeze or a rattly chest was present), or an upper respiratory illness (URI; otherwise). The material in each NPA was divided into four aliquots and stored at -80°C. Blood was collected from each child at age 6 months and 1, 2, 3, 4, 5 and 10 years.

Initial viral screen

One aliquot of each NPA (healthy and ARI) was screened for common causative agents via reverse transcriptase polymerase chain reactions (PCR) as previously reported (Kusel et al., 2006). Target organisms were: human rhinoviruses (HRV); other picornaviruses (coxsackie, echo and enteroviruses); coronaviruses 229E and OC43; respiratory syncytial virus (RSV); influenza A and B; parainfluenzaviruses 1-3; adenoviruses and human metapneumovirus (HMPV). Viral profile data was successfully generated for 451 healthy, 327 LRI and 649 URI NPAs.

Expanded screen for detection and subtyping of HRV

Following the recent identification of HRV-C (Lee et al., 2012), and expansion in availability of HRV sequence information, a second aliquot of each LRI NPA was screened for picornaviruses using seminested PCR (Bochkov et al., 2014), which incorporated a wider diversity of HRV primers than the initial screen, and then partial sequencing was used to classify HRV into A, B and C species and type.

DNA extraction and bacterial 16S rRNA amplicon sequencing

One aliquot each of 1,021 NPAs was used for 16S rRNA microbiome profiling. These include 487/561 healthy NPAs collected from visits at 2 months, 6 months and 12 months of age; 380/381 LRI reported during the same period; and 154/782 URI (random selection of 0-2 URI per infant). Overall, 397 healthy NPAs, 326 LRIs and 101 URIs were profiled for both virus and bacteria.

Total DNA was extracted using a method combining homogenization and chemical lysis of cells. Extractions were performed in biosafety cabinets that were UV-sterilised, including all plastic-ware, for 30 min prior to the procedure. The NPA were

thawed from -80°C storage, transferred into 1.5 mL sterile screw-capped tubes and briefly micro-centrifuged. The saline storage buffer was removed and pellets were resuspended in 400 µL of lysis solution supplied with the Wizard SV Genomic DNA System (Promega, Victoria, Australia). Samples were mixed vigorously by pipetting and then transferred into a labelled Lysing Matrix B tube (MP Biomedicals, New South Wales, Australia). Suspensions were homogenized using a FastPrep-24 homogenizer for 40 s at 6.5 m/s. Following micro-centrifugation, homogenates were transferred into a 1.5 mL screw-capped tube. A further 200 µL of lysis solution was added into each lysing matrix tube and vortexed to wash off any residual homogenate, then transferred to the respective homogenate tube to retain the original lysis volume. Homogenates were then treated with nuclei lysis buffer/RNase A and DNA extraction was carried out using the Wizard SV Genomic DNA System as per manufacturer's instructions. Purified DNA was eluted in 100 µL of pre-warmed sterile low 1 X TE (Fisher Biotec, WA, Australia), aliquoted and stored at -80°C.

Amplicons were prepared for MiSeq sequencing using primers (prepared by Integrated DNA Technologies, Iowa, USA) spanning the V4 region of the 16S rRNA gene and containing barcoded reverse primers as published by Caporaso *et al.* (Caporaso et al., 2012). The forward universal primer included the 5´ Illumina adapter sequence, forward primer pad, linker and the 515F 16S rRNA sequence: 5´-AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTGCCAGC MGCCGCGGTAA-3´. The reverse primer included the 3´ Illumina adapter sequence, a 12-mer Golay barcode (denoted as N), reverse primer pad, linker and the 806R 16S rRNA sequence: 5´-CAAGCAGAAGACGGCATACGAGATNNNNNNNNNNNN AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3´. All laboratory equipment used was wiped with DNA Away (MBP, Mexico) before conducting each PCR procedure. Master mixes were prepared in a UV-treated PCR chamber before they were dispensed into 96-well plates using a Multiprobe II liquid handling system (Perkin Elmer, Victoria, Australia), followed by addition of samples and controls using the robot. Amplification of each sample was performed in quadruplicate to obtain enough amplicon for sequencing. To each plate, a positive control (gDNA from *S. enterica* strain LT2, a bacterium not normally associated with the respiratory system (ATCC#700720D-5, USA)) and water and TE negative controls (obtained from each extraction procedure) were included and assessed for amplification by agarose gel electrophoresis. All controls were as expected: *S. enterica* controls were positive and the water and TE negative controls were negative. Positive control samples were not analysed further, while the negative controls were prepared for sequencing in the same way as NPA samples.

Due to the high throughput nature of this study, we did not quantify and normalize sample DNA that was added into each PCR reaction, rather a fixed volume (4 µL) of DNA template was used per well. Amplification was conducted on a GeneAmp 9700 PCR System (Perkin Elmer) using the following conditions: an initial 94°C denaturation step for 2 min, followed by 30 cycles of 94°C denaturation for 30 s, 58°C annealing for 30 s and 72°C extension for 1 min.

Quadruplicate sample amplicons were combined into a single well on the PCR reaction plate, then transferred to a fresh round-bottom polystyrene plate where they were purified using Agencourt AMPure XP beads as directed by the manufacturer, with slight modifications (Beckman Coulter, USA). Purified amplicons were eluted in

25 µL sterile low 1 X TE buffer (Fisher Biotec). Quantitation of amplicon was performed using the Quant-iT PicoGreen dsDNA quantitation kit (Life Technologies, Victoria, Australia) and fluorescence was determined on a Wallac Victor^3 Multilabel counter (Perkin Elmer). PCR samples were equalized to 2 nM concentration (a neat aliquot was used where a sample fell below this concentration) and pools of 48, 60 or 96 barcoded samples were generated and sent for sequencing.

Primer adaptors were removed from library pools using a 0.8x ratio of Agencourt AMPure XP beads (Beckman Coulter, USA). Library quantitation was determined by the high sensitivity Qubit kit (Life Technologies, USA) whilst library quality and average size distribution was assessed by the Bioanalyser (Agilent Technologies, USA) high sensitivity kit. Library pools were diluted to 2nM followed by NaOH denaturation as per manufacter's instructions (Illumina Inc., USA). Sequencing primers read 1: 5'- TATGGTAATTGTGTGCCAGCMGCCGCGGTAA -3', read 2: 5'-AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3' and index: 5'-ATTAGAWACCCBDGTAGTCCGGCTGACTGACT-3' (Sigma, Australia) were spiked into the MiSeq Cartridge at a final concentration of 0.5µM. Denatured libraries were loaded at 6.5pM with a 5% PhiX spike for diversity and sequencing control, onto a v2 300 cycle cartridge for sequencing on the Illumina MiSeq. All sequencing runs yielded approximately 1100k clusters/mm^2 with an average of 80% clusters passing filter and an average 19M reads passing filter.

Statistical Analysis

Classification of NPAs. Unless otherwise stated, the following definitions were used:

- Infection: Collected during a physician verified episode of symptoms of acute respiratory illness (ARI).
- Healthy: Collected in the absence of such symptoms for at least 4 weeks.
- LRI: Lower respiratory illness, defined as ARI with wheeze or rattly chest.
- URI: Upper respiratory illness, defined as ARI with no wheeze or rattly chest.
- 2-month samples: Age at collection 6-100 days (range 6-100 days for infections, 34-97 days for healthy).
- 6-month samples: Age at collection 150-210 days (interquartile range 165-196 days for infections, 179-190 days for healthy).
- 1-year samples: Age at collection 330-425 days (interquartile range 357-398 days for infections, 363-382 days for healthy).

Variable definitions. Unless otherwise stated, the following definitions were used in the statistical analyses:

- Gender (binary): Female (coded 0), Male (coded 1).
- Season (binary): Spring-summer (September to February, coded 1); autumn-winter (March to August, coded 0).
- Antibiotics intake (binary): Coded 1 if the child had taken antibiotics within the last 4 weeks before sample collection; 0 otherwise. Antibiotics data was curated from two sources of information; a daily diary kept by the mother, which reports any medication taken per day, and a questionnaire taken during each infection sampling. Positive antibiotics intake from any one source was sufficient. We considered only oral/intravenous antibiotics and excluded Flagyl (Metronidazole) and any topical antibiotics.

- Mother's antibiotics intake (binary): Coded 1 if the mother had taken antibiotics within the last 4 weeks before sample collection and she was breastfeeding on the day of antibiotics intake; 0 otherwise. Data was extracted from the daily diary, using the same group of antibiotics as for subjects.
- Breastfeeding (binary): Coded 1 if the child was breastfed on the day of sample collection; 0 otherwise.
- Siblings (binary): Coded 1 if any children less than 16 years of age were residing in the same house; 0 otherwise. Data was extracted from the 1-year questionnaire.
- Furry pets (binary): Coded 1 if household had any cat, dog, rabbit or guinea pig for the whole first year; 0 otherwise. Data was extracted from the 1-year questionnaire.
- Day-care (binary): Coded 1 if child has started attending day-care by sample collection; 0 otherwise.
- Delivery mode (binary): Coded 1 caesarean section delivery; 0 otherwise.
- Maternal/paternal history of atopic disease: Coded 1 if the mother/father had doctor diagnosed allergic hayfever, eczema or asthma.
- Atopy by 2 years (binary): Atopic status was assessed using serum IgE levels measured at 6 months, 1 year and 2 years. Positive atopy status was defined as any specific IgE to house dust mite, cat epithelium and dander, peanut, foodmix, couch grass, rye grass, mould mix, or infant phadiatop > 0.35 kU/L (*9*). Atopy by two years was defined as IgE above this cut-off at any of the three time points (6 months, 1 year, 2 years). Children atopic by 2 years are referred to as atopics; non-atopics otherwise.
- Chronic wheeze (binary): Presence of wheeze in the last 12 months, assessed during face-to-face interviews with the parents of subjects at 5 and 10 years of age.

Association of MPGs with healthy vs infection status. Odds ratios for each MPG with infection status (NPA taken during reported episode of respiratory illness, vs. NPA taken in the absence of such symptoms for at least 4 weeks prior) were estimated using generalized estimating equations (GEE) logistic regression with unstructured correlation and robust standard errors, to take into account samples from the same subject. The following variables were adjusted for by inclusion in the model: age at sample collection (days), gender, season, number of prior infections, antibiotics intake, mother's antibiotics intake, delivery mode and breastfeeding. ORs further adjusted for detection of common viruses (RSV, HRV) by PCR were also reported (**Figure 1E**).

Association between microbes and classes of ARI. Amongst ARI samples, odds ratios for MPGs and markers of infection severity (LRI vs. URI, febrile vs. non-febrile LRI, and wheezy vs. non-wheezy LRI) were estimated using generalized estimating equations (GEE) logistic regression with unstructured correlation and robust standard errors. Models were adjusted for presence of the most common viruses (RSV and HRV), age, gender, and season. Samples with antibiotics intake within the last week were excluded from this analysis (**Table 1**). Febrile vs. non-febrile LRI comparison was also done separately for RSV-positive and RSV-negative subsets using logistic regression and adjusting for age, gender and season (**Figure S5**).

Association of MPGs with otitis media symptoms. For each MPG, odds ratios for otitis media symptoms within (a) ARIs, and (b) LRIs were estimated using GEE logistic regression, adjusting for age at sample collection (**Table S2**).

Stability of MPGs within individuals. NP microbiome transitions were assessed by analysing the MPG assignments of consecutive samples within individual subjects (time point T_1 to next time point T_2). Transition into the same MPG (T_1 MPG = T_2 MPG) was considered a stable transition. Healthy and infection samples were analysed separately. The expected frequency of a stable transition for a given MPG, *i*, was calculated as the squared proportion of samples in that MPG at T_1, $(p_i(T_1))^2$. One thousand bootstrapped estimates of observed stability were generated by sampling with replacement from the observed transitions and calculating, for each bootstrapped sample, the proportion of stable transitions for each MPG. The estimated 95% confidence intervals for the observed frequency of respective stable transitions were taken to be the 2.5th and 97.5th quantiles of the 1,000 bootstrapped statistics. The MPG was considered significantly more stable than expected if the expected frequency of stable transitions $(p_i(T_1))^2$ fell below the lower limits of the 95% CI of observed transitions (and less stable if the expected value fell above the upper limits) (**Figure S3**).

Association of infection and wheeze phenotypes with Alloiococcus/Moraxella stable colonization groups. Children were grouped by their colonization patterns: '*Alloiococcus* colonized', ≥1 healthy sample with *Alloiococcus* MPG and none with *Moraxella* MPG; '*Moraxella* colonized', ≥1 healthy sample with *Moraxella* MPG and none with *Alloiococcus* MPG; 'Others', no *Moraxella* or *Alloiococcus* healthy sample or ≥1 of each. Association of the groups with number of infections was assessed using Wilcoxon rank sum test; other associations were assessed using Fisher's exact test (**Table S4**).

Association of genus abundance with environmental factors. We investigated the relationships between relative abundance of the six main genera and various environmental factors: attendance at day-care, living with siblings, antibiotics intake within the last 4 weeks, breastfeeding, number of prior infections (0, 1, ≥2; based on recorded ARI) and gender. As relative abundance values were highly skewed, they were log-transformed (base 10) prior to analysis. Odds ratios (ORs) for binary variables were estimated using logistic regression; odds ratios for association with prior infections were assessed using proportional odds ordinal logistic regression. Healthy and infection samples were analysed separately. Association with day-care was assessed for 12-month samples only, as few infants had started day-care by the time of their 6-month NPA; ORs for siblings, furry pets, antibiotics, season, and gender were adjusted for age at collection by inclusion in the model. Association with antibiotics was assessed for healthy samples only, as prescription due to infection severity could confound association with infection samples. Prior infections were assessed separately for each of the three age strata for healthy NPA sampling (2, 6 and 12 months) (**Figure 2**). Temperature data for Perth, Western Australia was obtained from the Australian Government Bureau of Meteorology website (http://www.bom.gov.au/climate/averages/tables/cw_009225.shtml).

Association between early Streptococcus colonization and subsequent chronic wheeze. Odds ratios were calculated separately for atopics, non-atopics and all children, using logistic regression. Gender, age at sample collection, maternal history of atopic disease and paternal history of atopic disease were adjusted for by inclusion in the model. Samples with any prior infection or antibiotics intake within the last week were excluded from this analysis (**Table 2, Figure 5B**).

Association between early colonization and time of first ARI. This analysis was restricted to children who contributed an asymptomatic NP sample between 5-9 weeks of age and prior to their first reported ARI (n=160). Children were grouped according to the MPG of the first pre-ARI healthy sample. Kaplan-Meier curves were plotted separately for each group, showing time of first infection (age in days) as recorded in the daily symptom diaries for (a) first ARI, (b) first URI and (c) first LRI. Cox proportional hazards models were fit to assess the significance of apparent differences in time of infection, adjusting for age, gender, season, virus status in the early healthy sample, and virus status in the subsequent infection (**Figure 4**).

Association between LRI subtypes and subsequent chronic wheeze. Odds ratios were calculated separately for atopics, non-atopics and all children, using logistic regression. Gender, maternal history of atopic disease and paternal history of atopic disease were adjusted for by including them as covariates in the model (**Table 2, Table S3**).

Association between time of first LRI subtype and subsequent chronic wheeze. Children were grouped according to their chronic wheeze status at 5 years. Kaplan-Meier curves were plotted separately for each group, showing time of first infection (age in days) for (a) first fever LRI, (b) first wheezy LRI with detection of HRV-C. Cox proportional hazards models were fit to assess the significance of apparent differences in time of infection, adjusting for gender, maternal history of atopic disease and paternal history of atopic disease (**Figure 3A-B**).

Association between antibodies to species-specific surface proteins and number of NP samples containing specific OTUs. Logistic regression was used to estimate odds ratios for association of detectable IgG1/IgG4 antibodies (assessed from blood during collection at 12 months) to *H. influenza* P4/P6 surface proteins with (a) number of infection NP samples and (b) number of healthy NP samples with presence of the two most abundant *Haemophilus* OTUs (**Figure S2A**). The same was done for *S. pneumoniae* A1, A2 or C surface proteins and the two most abundant *Streptococcus* OTUs (**Figure S2B**).

Supplemental References

Bochkov, Y.A., Grindle, K., Vang, F., Evans, M.D., and Gern, J.E. (2014). Improved molecular typing assay for rhinovirus species A, B, and C. Journal of clinical microbiology *52*, 2461-2471.

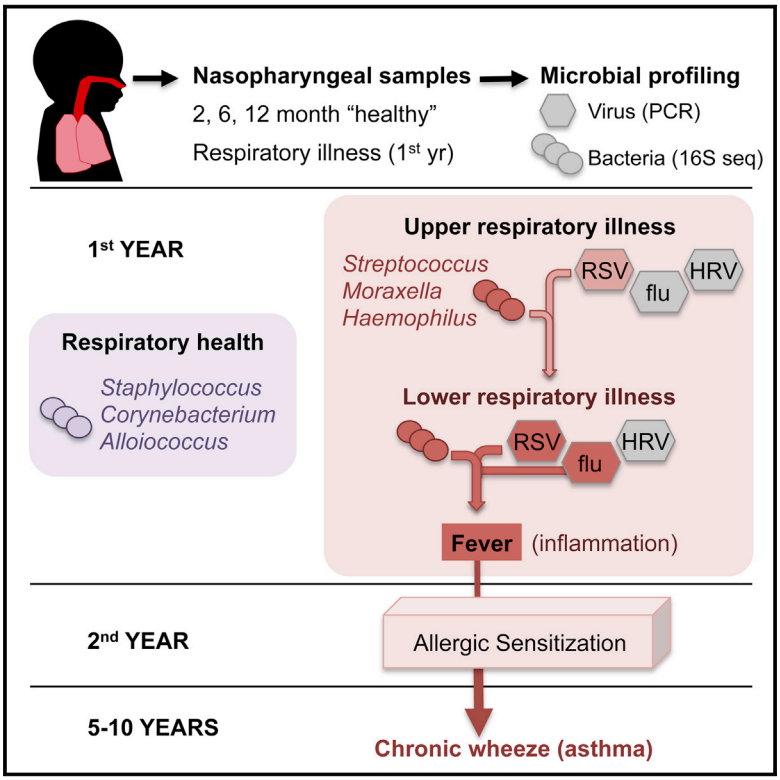
Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., *et al.* (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J *6*, 1621-1624.

Resource

Cell Host & Microbe

The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development

Graphical Abstract



Highlights

- The nasopharynx microbiome of infants has a simple structure dominated by six genera
- Microbiome composition affects infection severity and pathogen spread to lower airways
- Early asymptomatic colonization with *Streptococcus* increases risk of asthma
- Antibiotic usage disrupts asymptomatic colonization patterns

Authors

Shu Mei Teo, Danny Mok, ..., Kathryn E. Holt, Michael Inouye

Correspondence

kholt@unimelb.edu.au (K.E.H.), minouye@unimelb.edu.au (M.I.)

In Brief

Teo et al. characterize bacterial and viral communities within the infant nasopharynx during the first year of life, comparing between asymptomatic colonization and episodes of acute respiratory infections. Microbiome composition affects infection severity and spread to lower airways and risk for future asthma development.

Teo et al., 2015, Cell Host & Microbe *17*, 1–12
May 13, 2015 ©2015 Elsevier Inc.
http://dx.doi.org/10.1016/j.chom.2015.03.008

CellPress

Please cite this article in press as: Teo et al., The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development, Cell Host & Microbe (2015), http://dx.doi.org/10.1016/j.chom.2015.03.008

Cell Host & Microbe
Resource

The Infant Nasopharyngeal Microbiome Impacts Severity of Lower Respiratory Infection and Risk of Asthma Development

Shu Mei Teo,[1,2] Danny Mok,[3] Kym Pham,[4] Merci Kusel,[3] Michael Serralha,[3] Niamh Troy,[3] Barbara J. Holt,[3] Belinda J. Hales,[3] Michael L. Walker,[1] Elysia Hollams,[3] Yury A. Bochkov,[5] Kristine Grindle,[5] Sebastian L. Johnston,[7] James E. Gern,[5] Peter D. Sly,[6] Patrick G. Holt,[3,6] Kathryn E. Holt,[2,8,*] and Michael Inouye[1,8,*]
[1]Medical Systems Biology, Department of Pathology and Department of Microbiology & Immunology, The University of Melbourne, Parkville, VIC 3010, Australia
[2]Department of Biochemistry and Molecular Biology, Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, VIC 3010, Australia
[3]Telethon Kids Institute, The University of Western Australia, West Perth, WA 6008, Australia
[4]Melbourne Translational Genomics Platform, Department of Pathology, The University of Melbourne, Parkville, VIC 3010, Australia
[5]Department of Internal Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI 53726, USA
[6]Queensland Children's Medical Research Institute, The University of Queensland, Brisbane, QLD 4059, Australia
[7]Airway Disease Infection Section and MRC & Asthma UK Centre in Allergic Mechanisms of Asthma, National Heart and Lung Institute, Imperial College London, Norfolk Place, London W2 1PG, UK
[8]Co-senior author
*Correspondence: kholt@unimelb.edu.au (K.E.H.), minouye@unimelb.edu.au (M.I.)
http://dx.doi.org/10.1016/j.chom.2015.03.008

SUMMARY

The nasopharynx (NP) is a reservoir for microbes associated with acute respiratory infections (ARIs). Lung inflammation resulting from ARIs during infancy is linked to asthma development. We examined the NP microbiome during the critical first year of life in a prospective cohort of 234 children, capturing both the viral and bacterial communities and documenting all incidents of ARIs. Most infants were initially colonized with *Staphylococcus* or *Corynebacterium* before stable colonization with *Alloiococcus* or *Moraxella*. Transient incursions of *Streptococcus*, *Moraxella*, or *Haemophilus* marked virus-associated ARIs. Our data identify the NP microbiome as a determinant for infection spread to the lower airways, severity of accompanying inflammatory symptoms, and risk for future asthma development. Early asymptomatic colonization with *Streptococcus* was a strong asthma predictor, and antibiotic usage disrupted asymptomatic colonization patterns. In the absence of effective anti-viral therapies, targeting pathogenic bacteria within the NP microbiome could represent a prophylactic approach to asthma.

INTRODUCTION

The human microbiome is now recognized as playing an important role in the etiology and pathogenesis of myriad diseases (Weinstock, 2012). However, elucidation of these complex roles requires targeted characterization of microbial communities present in relevant spatial niche(s) during critical periods of pathogenesis. The focus of this study is the respiratory tract, in particular the nasopharynx (NP), which is an accessible source of airway microbial communities (Hilty et al., 2010) and serves as a conduit for pathogens associated with lower respiratory illnesses (LRIs) that are responsible for substantial morbidity and mortality worldwide.

Of particular interest is asthma, a multi-factorial disease characterized by airway inflammation and associated smooth muscle hyperplasia. It is now recognized that the hallmark persistent wheeze of asthma is consolidated in childhood and, further, may progress to chronic asthma in adulthood (Holt and Sly, 2012; Sly et al., 2008) and potentially chronic obstructive pulmonary disease (Tai et al., 2014). We and others have previously shown that development of persistent atopic (allergic) wheeze in children is linked to the number of virus-associated febrile and/or wheezy LRIs experienced during infancy (Jackson et al., 2008; Kusel et al., 2007, 2012; Oddy et al., 2002). The principal virus type of current interest is human rhinoviruses (HRVs), particularly subtype C (HRV-C) (Bochkov and Gern, 2012); however, respiratory syncytial virus (RSV) is also recognized as a major cause of infant LRI (Wu and Hartert, 2011). The relative contributions of these viral pathogens in asthma initiation remain controversial (Stein and Martinez, 2010). Further complicating the picture, recent studies have also implicated bacterial pathogens as potential independent causal factors in infant LRIs and their long-term sequelae. Notably, culture of *S. pneumoniae*, *M. catarrhalis*, or *H. influenzae* from NP samples taken at 1 month of age has been linked to increased risk for subsequent diagnosis of asthma at 5 years of age (Bisgaard et al., 2007). These findings have fuelled debate around the use of antibiotics and vaccine strategies for respiratory illness in children (Penders et al., 2011; Rollins et al., 2010).

Several studies have investigated airway microbiota in children or adults with chronic respiratory illness, including asthma

(Bogaert et al., 2011; Hilty et al., 2010; Vissers et al., 2014); however, no study has investigated the airway microbiome during the critical infancy period (0–12 months). In this study, we investigated the NP microbiome during the first year of life using the Childhood Asthma Study (CAS), a prospective cohort of 234 children (Kusel et al., 2006, 2007, 2008, 2012), to elucidate the NP microbiome during respiratory health and illness, its longitudinal dynamics, susceptibility to exogenous factors such as antibiotics, and association with future asthma.

RESULTS

A total of 1,021 NP microbiome profiles were obtained from 234 infants using 16S rRNA gene deep sequencing (see Supplemental Experimental Procedures). These included 487 "healthy" NP samples collected in the absence of respiratory symptoms and 534 "infection" NP samples collected during episodes of acute respiratory illness (ARI) during the first year of life. Three quarters of the infants (n = 177) contributed at least two healthy NP samples at the age of ~2 months, ~6 months, and/or ~12 months. Eighty percent of the infants (n = 186) contributed a healthy sample before experiencing their first ARI. The 534 infection NP samples were from 184 infants who had experienced $\geq$ 1 ARI within the first year of life. NP samples were analyzed from all (380/381) recorded LRI in this period and a random selection of 20% (154/782) of recorded episodes of upper respiratory illness (URI). The characteristics of the infants are summarized in Table S1.

NP Microbiome Composition

Across all NP samples, > 193 million high-quality 16S rRNA sequences were classified into 14,131 operational taxonomic units (OTUs), of which 1,010 were supported by > 1,000 reads each. The dominant phyla were Proteobacteria (48%), Firmicutes (38%), Actinobacteria (13%), Bacteroidetes (1%), and Fusobacteria (0.5%) (Figure 1A). The NP microbiomes were dominated by six genera: *Moraxella* (31.2%), *Streptococcus* (15.5%), *Corynebacterium* (13.5%), *Staphylococcus* (10.3%), *Haemophilus* (9.7%), and *Alloiococcus* (8.8%; genus *Dolosigranulum* in some databases) (Figure 1A). Despite the inclusion of diverse OTUs of these genera in the reference database, our sequences were dominated by one OTU per genus (Figure S1), consistent with culture-based studies reporting NP colonization with the species *Moraxella catarrhalis*, *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Haemophilus influenzae*, and *Alloiococcus otitidis*. Hierarchical clustering of NP microbiomes based on relative abundance of the six major genera identified six microbiome profile groups (MPGs, Figure 1B). Each MPG was dominated by one of the six genera, although some samples in the *Alloiococcus* MPG also had relatively high abundance of *Corynebacterium* (Figure 1B).

NP Microbiome Dynamics

Healthy NP samples collected around 2 months of age were dominated by *Staphylococcus* (41%) and *Corynebacterium* (22%) MPGs, but the frequency of these MPGs declined with age (11% and 10%, respectively, at 12 months old) (Figure 1C). In contrast, the prevalence of *Alloiococcus* and *Moraxella* MPGs in healthy samples increased with age (14% and 9% at 2 months, 26% and 41% at 12 months, respectively). Analysis of MPG transitions among consecutive healthy NP samples from the same individuals (Figures S3A and S3B) suggested that *Staphylococcus* carriage was unstable or transient, particularly where an ARI had occurred in the intervening period between sampling. *Alloiococcus* was a stable colonizer, but less so if an ARI occurred between sampling. Where an ARI occurred in the intervening period between healthy samples, the most common transitions were to the *Moraxella* MPG from other MPGs or the maintenance of stable colonization with *Moraxella* (Figure S3B). Almost all *Moraxella*-colonized infants experienced subsequent ARI before the next healthy sample was taken (explored further below). *Haemophilus* was very rarely detected in healthy NP microbiomes, while *Streptococcus* was present in 14% of healthy samples at each sampling time (Figure 1C). Similar age-related patterns were observed among infection samples, with a decline in *Staphylococcus* and *Corynebacterium* MPGs and increase in *Moraxella*, *Haemophilus*, and *Streptococcus* MPGs in older children (Figure 1D). Interestingly, males had significantly more *Moraxella* in healthy samples (OR 1.3 for log abundance; 95% CI 1.1–1.5, p = 0.0014, adjusted for age); no other gender effects were detected.

We next assessed the impact of environmental factors on the relative abundance of the common NP microbiome genera. We found no significant effects of delivery mode or breastfeeding on NP colonization at 2 months of age; the latter was unsurprising as nearly all infants (90%) were breastfed for at least 2 months. The abundance of *Streptococcus* in healthy NP samples was significantly lower among children whose parents reported having furry pets such as dogs or cats in the home (OR 0.84, 95% CI 0.70–1.0, p = 0.046; adjusted for age at NP sampling). No other significant associations were detected for pets. Children attending day care had significantly higher relative abundances of *Haemophilus* and *Moraxella* and lower relative abundances of *Corynebacterium* and *Staphylococcus* (Figure 2A) in both healthy and infection samples (note that very few children had commenced day care by 6 months of age, hence the impact of day care attendance was assessed at the 12-month time point only). Co-habiting with siblings was also associated with higher abundances of *Haemophilus*, *Streptococcus*, and *Moraxella* and lower abundance of *Staphylococcus* during health and ARI (adjusted for age at sampling, Figure 2B). Importantly, among healthy samples, antibiotic usage in the four weeks prior to sampling was associated with higher abundances of *Haemophilus*, *Streptococcus*, and *Moraxella* and lower abundances of *Alloiococcus* and *Corynebacterium* (adjusted for age at sampling, Figure 2C). The composition of the healthy NP microbiome was also affected by the number of prior respiratory infections experienced, with higher abundance of *Moraxella* and lower abundances of *Alloiococcus* or *Corynebacterium* in samples following increasing numbers of ARIs (Figure 2E). At the MPG level, ARIs dominated by *Haemophilus* increased in spring-summer, while those dominated by *Moraxella* peaked in autumn-winter (Figure 2F). We therefore tested for differences in relative abundance in autumn-winter and spring-summer, adjusting for age and number of prior infections; this confirmed significant seasonal effects on the abundance of *Haemophilus* (summer associated) and *Moraxella* (winter associated) among

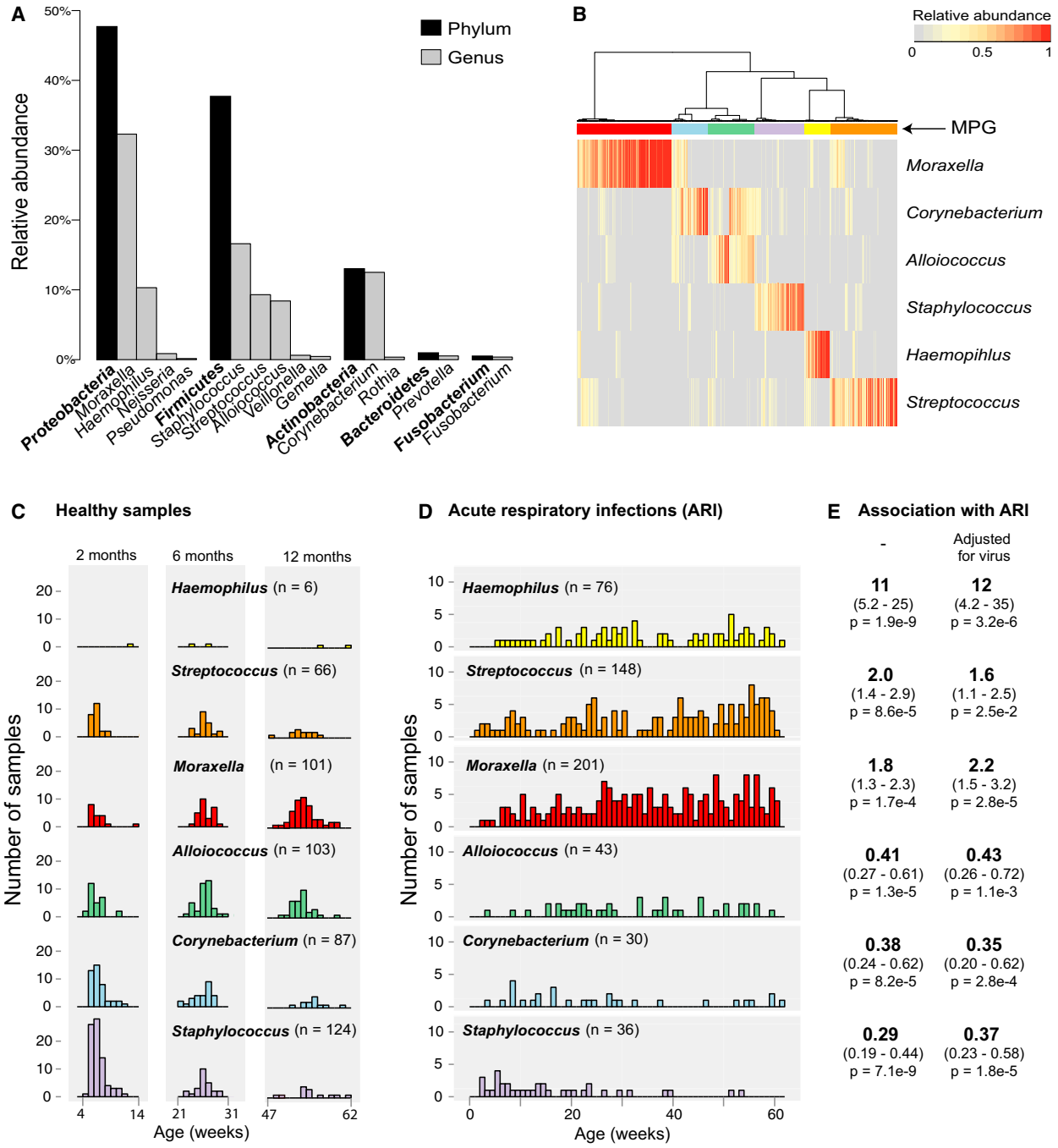


Figure 1. Bacterial Composition of 1,021 Nasopharyngeal Aspirates Collected from 234 Infants during Periods of Respiratory Health and Disease

(A) Frequency of the most abundant phyla and genera (comprising 99.9% of reads).

(B) Clustering of samples into microbiome profile groups (MPGs) based on relative abundance of the six most common genera. Colored bars indicate MPGs, labeled by their dominant genus: *Moraxella* (red), *Corynebacterium* (blue), *Alloiococcus* (green), *Staphylococcus* (purple), *Haemophilus* (yellow), and *Streptococcus* (orange).

(C) Weekly frequencies of MPGs among healthy samples, collected during planned visits at approximately 2, 6, and 12 months of age and following at least 4 weeks without symptoms of acute respiratory infection (ARI).

(D) Weekly frequencies of each MPG among ARI samples.

(E) Odds ratios for association of MPGs with ARI symptoms, adjusted for age, gender, season, number of prior infections, antibiotics intake, mother's antibiotics intake, delivery mode, and breastfeeding; with and without adjustment for detection of common viruses (RSV, HRV).

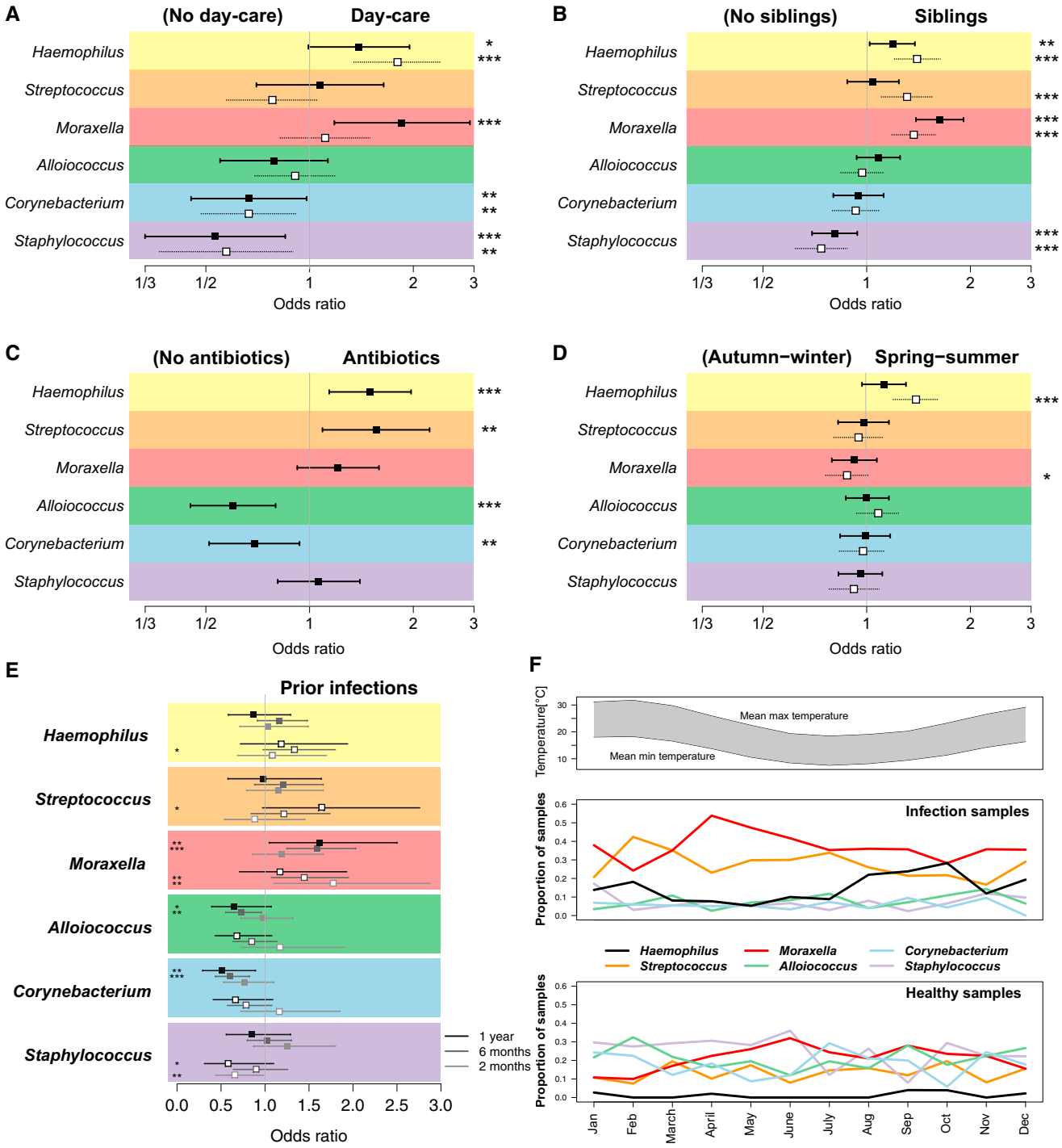


Figure 2. Impact of Environmental Factors on Relative Abundances of Major Genera of the NP Microbiome

(A–D) Squares, odds ratios; filled squares, healthy samples; empty squares, infection samples; bars, 95% confidence intervals; *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Associations are estimated using logistic regression and adjusted for age: (A) day care attendance (yes versus no, 12-month samples), (B) co-habiting with siblings, (C) antibiotics intake in the 4 weeks preceding NP sample collection, (D) season (spring-summer versus autumn-winter).

(E) Impact of prior ARI, estimated using proportional odds ordinal logistic regression and categorized as 0, 1, or $\geq$ 2 ARI.

(F) Seasonal patterns: top, mean maximum and minimum temperatures in study location (Perth); bottom, monthly proportions of samples in each microbiome profile group (MPG) for infection and healthy samples.

Table 1. Associations between ARI Symptom Severity and Risk-Associated MPGs

Symptom Group Comparison	Model	*Streptococcus*, *Moraxella*, or *Haemophilus* MPG			RSV	HRV (expanded screen)
		Streptococcus MPG	*Moraxella* MPG	*Haemophilus* MPG		
LRI versus URI	1A	OR = 2.2 (1.3–3.7), p = 0.0043			2.2 (1.1–4.7), p = 0.034	–
	1B	2.3 (1.2–4.5), p = 0.011	2.1 (1.1–3.8), p = 0.017	2.1 (1.0–4.6), p = 0.056	2.3 (1.1–4.8), p = 0.033	–
Febrile LRI versus non-febrile LRI	2A	OR = 3.8 (1.0–14), p = 0.047			4.0 (1.6–9.8), p = 0.0031	0.45 (0.21–0.95), p = 0.036
	2B	4.3 (1.0–18), p = 0.046	3.8 (0.98–15), p = 0.053	2.7 (0.53–14), p = 0.23	4.0 (1.6–10), p = 0.0031	0.46 (0.21–0.98), p = 0.043
Wheezy LRI versus non-wheezy LRI	3A	OR = 1.4 (0.60–3.1), p = 0.45			0.94 (0.41–2.1), p = 0.88	1.2 (0.62–2.2), p = 0.64
	3B	1.1 (0.39–3.0), p = 0.88	1.8 (0.74–4.1), p = 0.20	1.0 (0.34–3.2), p = 0.93	0.95 (0.42–2.2), p = 0.90	1.1 (0.60–2.2), p = 0.69

Odds ratio (OR) (95% confidence interval), p value. NP samples taken within a week of antibiotic use were excluded from analysis. The response variable (symptom group comparison) is shown in column 1. Two models (A, B) were fit for each comparison, labeled in column 2. Model A included *Streptococcus* or *Moraxella* or *Haemophilus* MPG as a single covariate; Model B included *Streptococcus*, *Moraxella*, and *Haemophilus* MPGs as separate covariates. Both models also include RSV, HRV, and the potential confounders age, gender, and season. Note: HRV was not included in the LRI versus URI comparison, as enhanced sensitivity re-screening for HRV was performed only in LRI samples.

ARI samples and a similar but non-significant trend among healthy NP samples (Figure 2D).

NP Microbial Determinants of ARI Symptoms

The *Moraxella*, *Streptococcus*, and *Haemophilus* MPGs were significantly more frequent in ARI compared to healthy NP samples, even after adjusting for a large set of potential confounders (age, gender, season, number of prior infections, antibiotic intake, mother's antibiotic intake, delivery mode, and breastfeeding) (Figure 1E). The *Staphylococcus*, *Corynebacterium*, and *Alloiococcus* MPGs were significantly less frequent in ARI (Figure 1E). The rare genus *Neisseria* was more common in infection samples, especially LRIs (of 28 NP samples with > 5% relative abundance of *Neisseria*, 4 were from URI and 20 were from LRI, though all were co-colonized with *Streptococcus*). It was not possible to confirm species from the 16S sequences; however, it is well known that the respiratory pathogens *S. pneumoniae*, *H. influenzae*, *M. catarrhalis*, and *N. meningitidis* are frequently cultured from respiratory infections in children. We have previously measured IgG to species-specific surface proteins of *S. pneumoniae* and *H. influenzae* at 12 months of age in the CAS cohort (Hales et al., 2012). Here we found that *H. influenzae*-specific IgG was significantly associated with the number of prior ARI samples testing positive for either of the two most common *Haemophilus* OTUs (Figures S2A–S2C); similar results were obtained for *S. pneumoniae*-specific IgG antibodies and the dominant *Streptococcus* OTU (Figures S2D–S2F). Healthy colonization with these genera was not associated with species-specific IgG.

A total of 138 children had $\geq$ 2 ARI samples profiled, and of these, 97 (70%) had $\geq$ 2 different MPGs among their ARIs. There was a clear temporal trend, with infections occurring closer in time more likely to be of the same MPG (Figure S3D). *Moraxella* and *Haemophilus* MPGs were particularly stable between consecutive infections, i.e., following an ARI in which the *Moraxella* or *Haemophilus* MPG was present, the next ARI was more likely to share the same MPG type than expected given the overall frequency of these MPGs among infection samples (Figure S3C).

We further considered the impact of the NP microbiome on infection severity and interactions with viral pathogens. All samples analyzed here were previously screened for a panel of viruses (Kusel et al., 2006). This screen detected viruses in 21% of healthy samples, 68% of URIs, and 69% of LRIs. The most common viruses detected were RSV (11% of ARI) and HRV (40% of ARI) (Figure S4), although subsequent expanded screening and subtyping of HRV in LRI samples suggests this is an underestimate (see below). For all virus groups except adenovirus and coronavirus, virus detection was significantly positively associated with ARI symptoms (i.e., ARI versus healthy samples, Figure S4). The association between ARI and *Streptococcus*, *Haemophilus*, and *Moraxella* MPGs remained after adjusting for detection of virus (OR 7.0, $p < 1 \times 10^{-15}$ for any of these MPGs; individual ORs in Figure 1E), indicating both viruses and bacteria contribute to ARI symptoms. Among the viruses analyzed, only RSV was significantly more frequent in LRIs versus URIs (16% versus 8.3%, OR 2.3; Figure S4, Table 1). The illness-associated MPGs *Streptococcus*, *Haemophilus*, and *Moraxella* were significantly associated with LRIs versus URIs (OR > 2), individually and collectively, adjusting for the effect of RSV (Table 1). As a group, the illness-associated MPGs increased in frequency from healthy to URI to LRI samples, regardless of the presence of RSV (Figure S5). Taken together, these analyses indicate that both viruses and bacteria independently contribute to ARI and that bacteria and RSV independently increase the risk of infection spread to the lower airways.

In our earlier studies with this cohort, we observed an association between LRIs (but not URIs) during infancy and risk for wheeze at age 5 years; moreover, this association was restricted to severe LRIs, i.e., those accompanied by fever and/or wheeze (Holt et al., 2010; Kusel et al., 2007, 2012). Thus we investigated the association of viruses and bacteria with the presence of fever and wheeze symptoms during LRI. To enable more accurate assessment of the role of HRV in this critical sample group, we

Table 2. Association between Microbial Events during Infancy and Chronic Wheeze at Age 5 and 10 Years

	Wheeze at 5 Years		Wheeze at 10 Years	
	All children	Atopics	All children	Atopics
Any febrile LRI	2.3 (1.2–4.5), p = 0.016*	2.7 (1.1–7), p = 0.034*	2.2 (0.93–5.2), p = 0.071	2.7 (0.89–8.7), p = 0.083
Any wheezy LRI	1.6 (0.79–3), p = 0.2	1.3 (0.49–3.3), p = 0.61	1.4 (0.58–3.3), p = 0.45	1.6 (0.53–4.8), p = 0.4
Any HRV wheezy LRI	2 (0.93–4.2), p = 0.073	2.5 (0.86–7.2), p = 0.092	2.1 (0.81–5.4), p = 0.11	1.9 (0.55–6.4), p = 0.29
Any HRV-C wheezy LRI	2.4 (0.93–6.1), p = 0.064	7.2 (1.7–35), p = 0.009*	3.5 (1.1–11), p = 0.026*	7.1 (1.6–40), p = 0.014*
Any HRV-A wheezy LRI	1.2 (0.42–3.1), p = 0.74	0.55 (0.11–2.2), p = 0.43	1.4 (0.37–4.7), p = 0.57	0.35 (0.02–2.2), p = 0.34
Any risk bacteria LRI	0.89 (0.45–1.8), p = 0.73	0.96 (0.39–2.4), p = 0.93	2 (0.82–5.2), p = 0.14	1.7 (0.59–5.2), p = 0.34
High-abundance *Streptococcus* colonization (≤7 weeks)	3.8 (1.3–12), p = 0.017*	4 (0.88–21), p = 0.077	2.7 (0.6–12), p = 0.18	3.9 (0.63–28), p = 0.15

Odds ratio (95% confidence interval), p value; estimated using logistic regression, adjusted for gender and maternal and paternal history of atopic disease; estimated separately for all children and those who were atopic by 2 years of age. LRI, lower respiratory illness, further classified according to microbes and symptoms; any risk bacteria LRI, any LRI with *Streptococcus*, *Moraxella*, or *Haemophilus* MPG. Early *Streptococcus* colonization was assessed in the first healthy NP sample, collected by 7 weeks of age and prior to any recorded infection; high abundance was classified as > 20% *Streptococcus* reads (based on distribution in Figure 5A).

performed an expanded screen for detection and subtyping of HRV within LRI samples (Supplemental Experimental Procedures). This assay proved more sensitive than the first screen and detected HRV in 66% of LRI samples (twice that in the earlier screen), with equal amounts of HRV-A and HRV-C (Figure S4). During LRI, the presence of any HRV, or HRV-C specifically, was negatively associated with fever (Table 1). The only viruses showing positive associations with fever were RSV (Table 1, Figure 3C, Figure S4) and influenza (8 influenza-positive fever LRI events only, all with illness-associated MPGs, thus not considered further). Among LRIs, the illness-associated MPGs were associated with fever, even after adjusting for RSV, HRV, age, season, and gender (Table 1). Interestingly, *Moraxella* MPG was also significantly positively associated with fever among RSV-positive LRIs (OR 9.2, 95% CI 1.1–73, p = 0.037; Figure S5), suggesting a possible interaction between *Moraxella* and RSV whereby their co-presence further enhances risk of fever. Overall, only 10 febrile LRIs (9.5%) could not be explained by the presence of RSV or illness-associated MPGs. Across the cohort, the presence of wheeze during LRI was not significantly associated with any viral or bacterial groups (Table 1), including HRV or HRV-C specifically. Of all ARIs analyzed, there were 61 with concomitant otitis media (OM) affecting 47 infants; however, ARIs with accompanying OM had similar NP microbiome profiles to those of LRIs with no OM diagnosis (Table S2).

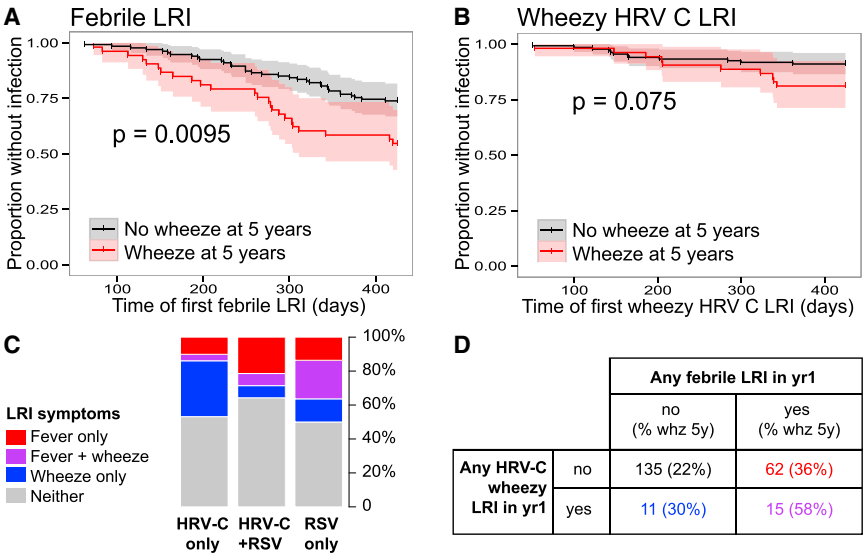
Impact of ARI on Later Chronic Wheeze

We next investigated associations between LRI in the first year of life and subsequent expression of chronic wheeze at age 5 or 10 years. Positive associations were found for two discrete classes of LRI. First, febrile LRI was significantly positively associated with later chronic wheeze and among children who were atopic by 2 years of age (Table 2). Furthermore, timing of first febrile LRI appeared to be important, with earlier febrile LRI occurring among children who had chronic wheeze at 5 years (p = 0.0095, Figure 3A). Second, HRV-C LRI accompanied by wheezing symptoms showed a positive association with later chronic wheeze among all children and particularly strongly so for those who were atopic by 2 years (OR ~7, Table 2), but not among non-atopics (OR 1, p = 0.95; Table S3). As we found no association between HRV-C and wheeze during LRI across the whole cohort (OR 1.2, 95% CI 0.63–2.2, p = 0.61; adjusting for illness-associated MPGs, RSV, age, season, and gender), we re-examined the association in those who developed atopy by age 2. Among this group of children (n = 65), the presence of HRV-C was significantly positively associated with increased risk of wheeze during LRI (OR 2.7, 95% CI 1.1–7, p = 0.035; adjusting for illness-associated MPGs, RSV, age, season, and gender).

Since febrile LRI and HRV-C wheezy LRI were both associated with later chronic wheeze (Table 2), and RSV was associated with febrile LRI (Table 1), we examined the interaction between RSV, HRV-C, LRI symptom severity, and chronic wheeze at age 5. Febrile LRI and wheezy HRV-C LRI in the first year of life appeared to exert independent effects on later chronic wheeze, since only a minority of HRV-C wheezy infections were also febrile (Figure 3C). Further, at the level of individual children, the frequency of chronic wheeze at 5 years was elevated among those with either one of febrile LRI (36%) or wheezy HRV-C LRI (30%) but was greatest among children who experienced both (58%, Figure 3D).

Impact of NP Colonization on ARI

Whereas *Streptococcus*, *Haemophilus*, or *Moraxella* MPGs were detected in healthy NP samples, subsequent infections tended to belong to the same MPG (Figure 4A). Critically, we investigated whether healthy colonization in early infancy was associated with subsequent episodes of ARI. Since prior infections also affect the healthy NP microbiome (Figure 2E), we restricted these analyses to healthy samples collected at 5–9 weeks of age and prior to each infant's first reported ARI (n = 160). Using Cox proportional hazards models, infants whose earliest healthy NP sample was of the *Moraxella* or *Streptococcus* MPG tended to experience ARI at a younger age than those with other MPGs (Figure 4B; note that *Haemophilus* MPGs were extremely rare in early healthy samples). When examining URIs and LRIs separately, we found that early *Moraxella* colonization was associated with earlier first URI, whereas early *Streptococcus* colonization was strongly associated with earlier first LRI (Figures 4C and 4D). Since our data suggested that *Alloiococcus* and *Moraxella*



		Any febrile LRI in yr1	
		no (% whz 5y)	yes (% whz 5y)
Any HRV-C wheezy LRI in yr1	no	135 (22%)	62 (36%)
	yes	11 (30%)	15 (58%)

Figure 3. Symptoms of Lower Respiratory Illness during the First Year of Life Are Associated with Viruses Present during the Infection and Predict Chronic Wheeze at 5 Years of Age

(A and B) Kaplan-Meier survival curves for age (days) at (A) first febrile LRI and (B) first HRV-C wheezy LRI, stratified by chronic wheeze status at 5 years. p values shown were estimated using Cox proportional hazards models, adjusted for gender and maternal and paternal history of atopic disease. Shaded areas indicate 95% confidence intervals.

(C) Frequencies of fever and wheeze symptoms during LRI, in which HRV-C and/or RSV were detected. Total numbers are: HRV-C only, n = 79; HRV-C and RSV, n = 14; RSV only, n = 22.

(D) Cross-tabulation of individuals according to their experience of LRI during infancy; percentages in brackets indicate frequency of chronic wheeze at 5 years.

were key stable colonizers of the NP microbiome, we also examined whether *Alloiococcus*-colonized infants differed from *Moraxella*-colonized infants (defined as those with $\geq$ 1 healthy sample of *Moraxella* MPG and none with *Alloiococcus* MPG, and vice versa). These groups did not differ in terms of overall numbers of ARI, LRI, or later wheezing phenotypes. However, compared to either *Moraxella*-colonized infants or those not in either group, *Alloiococcus*-colonized infants had fewer RSV infections, especially RSV LRIs (OR 0.27, Table S4).

Early NP Colonization Impacts Later Chronic Wheeze

We next assessed the association between early pre-ARI asymptomatic NP colonization and current wheeze at 5 and 10 years of age, stratified by atopic sensitization by 2 years of age (Supplemental Experimental Procedures). This analysis was restricted to the 160 infants (70%) who had an asymptomatic NP sample taken prior to their first ARI (and $\leq$ 9 weeks of age). *Haemophilus* MPG was not detected in pre-ARI healthy samples, and *Moraxella* MPG showed no evidence of association with future wheeze. The *Streptococcus* MPG showed a weak association; however, the relative abundance of *Streptococcus* in these samples was highly skewed, so we divided samples into those with high (> 20%) or low ($\leq$ 20%) relative abundance of *Streptococcus* reads (Figure 5A). High *Streptococcus* abundance in the first pre-ARI healthy NP sample was more frequent in infants who later displayed wheeze at 5 years, and this association was stronger when restricting the analysis to earlier NP samples (Figure 5B, Table 2). The same trend was evident for wheeze at 10 years of age, despite reduced sample size due to loss to follow-up (Table 2). Early *Streptococcus* colonization was associated with younger age at first LRI (Figure 4D), but not with presence of *Streptococcus* in the first LRI or with detection of *S. pneumoniae* antibodies at 12 months of age. No statistically significant associations were observed for future chronic wheeze when aggregating the 7-week healthy *Streptococcus*, *Haemophilus*, and *Moraxella* MPGs into a single predictor (with or without atopy by 2 years). Overall, infants who were atopic by age 2 and developed chronic wheeze at age 5 were twice as likely to have had early *Streptococcus* colonization, febrile LRI, and/or HRV-C wheezy LRI in the first year of life, compared to those that did not develop chronic wheeze (Figure 5C).

DISCUSSION

NP Microbiome Composition and Dynamics

Our data provide a detailed prospective characterization of bacterial communities within the human NP microbiome during the first year of life. The NP microbiome was qualitatively simple (Figure 1), dominated by six common genera: *Haemophilus*, *Streptococcus*, *Moraxella* (each more common in ARI), *Staphylococcus*, *Alloiococcus*, and *Corynebacterium* (more common in healthy samples). This is consistent with previous studies of NP microbiome composition in children aged 12–14 months (Biesbroek et al., 2014; Bogaert et al., 2011) and adults (Hilty et al., 2010), although intra- and inter-sample diversity was greater in these older groups, likely due in part to extreme seasonal variation in those study locations. In contrast, our study site in Perth, Australia has a very moderate climate, and we detected only limited seasonal effects (Figure 2D) that were readily adjusted for, enabling us to assess the dynamics of the NP microbiome at different stages of infancy and to examine its association with other factors.

We found that early NP colonization typically involved *Staphylococcus* or *Corynebacterium*, which was later replaced by *Moraxella* or *Alloiococcus*. *Staphylococcus* was the dominant colonizing bacteria in the early healthy NP microbiomes, but its presence declined rapidly with age (Figure 1). *S. aureus* is a common cause of neonatal sepsis, and high rates of *S. aureus* nasal colonization in the first months of life have been reported (Bisgaard et al., 2007, 2010). Few studies have examined longitudinal colonization; however, a recent study in African children reported *S. aureus* colonization in 42% of infants at 1 month and 12% at 12 months of age, a trend that was mirrored in maternal colonization rates (Schaumburg et al., 2014) and is strikingly similar to the patterns we observed (41% at 2 months,

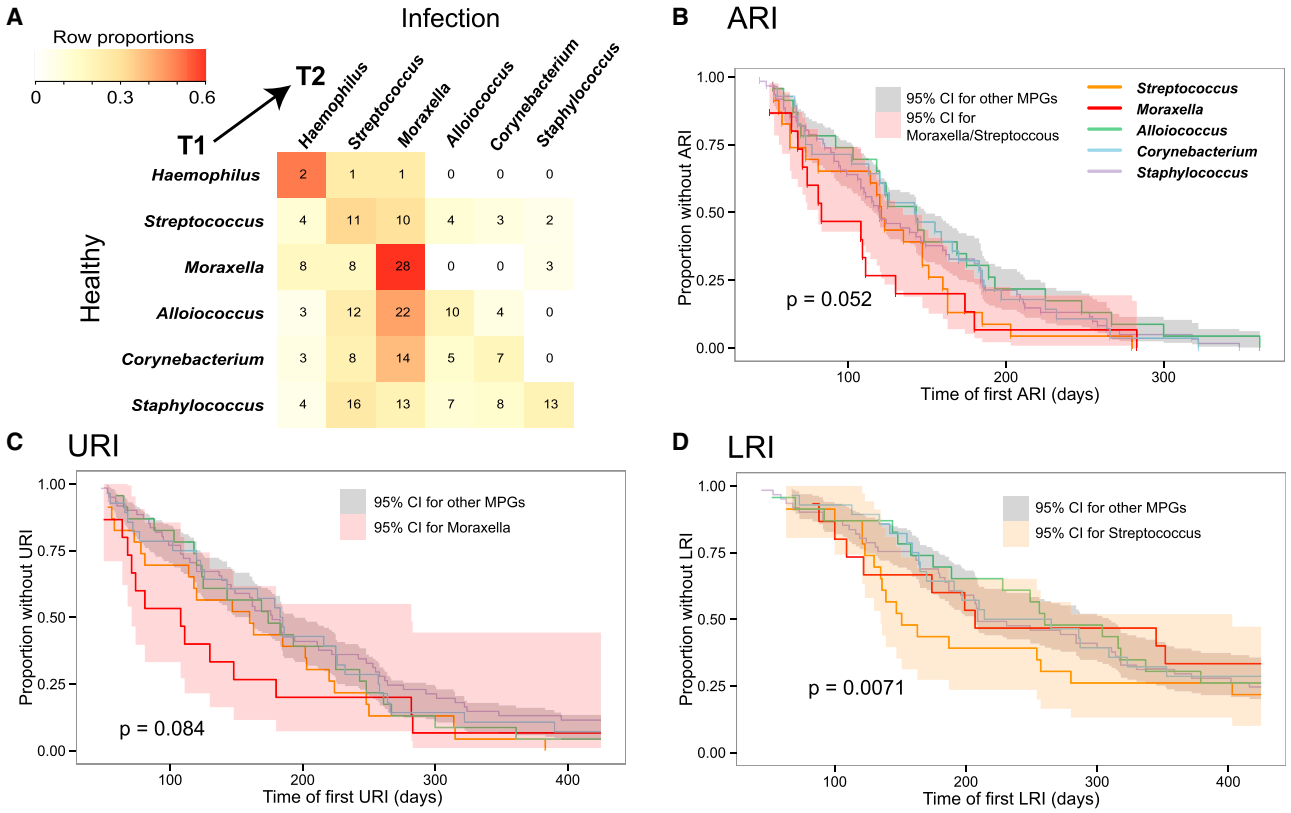


Figure 4. Impact of Early Colonization on Age at First Respiratory Infection

(A) Microbiome profile group (MPG) transitions between healthy samples (T1) and the next sequenced infection (T2). Cell numbers indicate the number of times the respective transition from T1 to T2 was observed in the dataset; cells are colored to indicate the row proportions as per legend.

(B–D) Kaplan-Meier survival curves for age (days) of first (B) acute respiratory illness (ARI), (C) upper respiratory illness (URI), and (D) lower respiratory illness (LRI), stratified according to the MPG of the first healthy sample (collected by 9 weeks of age and prior to any infection, n = 160). Cox proportional hazards models were adjusted for age, gender, season, virus status in the early healthy sample, and virus status at the first event. Shaded areas indicate 95% confidence intervals.

11% at 12 months). *Staphylococcus* and *Corynebacterium* are both common components of the human skin microbiome, and in our data they showed a comparable temporal pattern, with high rates in the early NP microbiome declining with age (Figure 1). We speculate that infants tend to be colonized initially with skin-dwelling bacteria (acquired from parents and others), which is replaced over time by stable colonization with *Moraxella* or *Alloiococcus* and punctuated by transient acquisition of *Streptococcus* or *Haemophilus* that is frequently accompanied by ARI symptoms. Our data further suggest the transition to *Moraxella* is associated with exposure to other children (in the home or at day care) and episodes of ARI, and that colonization with *Moraxella*, *Streptococcus*, and *Haemophilus* are selected for by antibiotic exposure (Figure 2).

Moraxella was the most common genus in our study population, dominating 21% of all healthy NP microbiomes and 38% of infection samples. *Moraxella* was represented in our sequence data by a single OTU matching that of *M. catarrhalis*, a human-restricted, unencapsulated, Gram-negative bacterium previously associated with both commensal NP colonization and pathogenicity in the respiratory tract and inner ear (de Vries et al., 2009). Increased rates of NP colonization with *M. catarrhalis* have been reported in children following pneumococcal vaccination (Revai et al., 2006); however, the samples in our study predate introduction of this vaccine. In our cohort the presence of *Moraxella* increased with age, and *Moraxella* was a particularly stable component of the NP microbiome during both health and disease (Figure S3). These findings are consistent with *Moraxella*'s known ability to form biofilms (de Vries et al., 2009), which offer protection from antibiotics and promote co-colonization with other common bacteria such as *S. pneumoniae* and *H. influenzae* (Verhaegh et al., 2011). *Moraxella* was more abundant during the cooler months (Figure 2), a seasonal trend consistent with recent reports that cold shock at 26°C stimulates growth, colonization, and expression of virulence-associated traits in *M. cattarhalis* (Spaniol et al., 2011). These data are consistent with a Dutch culture-based study of *M. catarrhalis* carriage in young children (Verhaegh et al., 2011), which found increasing prevalence of *M. catarrhalis* during the first year of life, with a strong seasonal effect (also observed in a Swedish cohort [Gisselsson-Solén et al., 2014]) and a significant positive association with day care attendance and siblings. However, these studies did not examine the stability of *Moraxella* colonization within individuals over time.

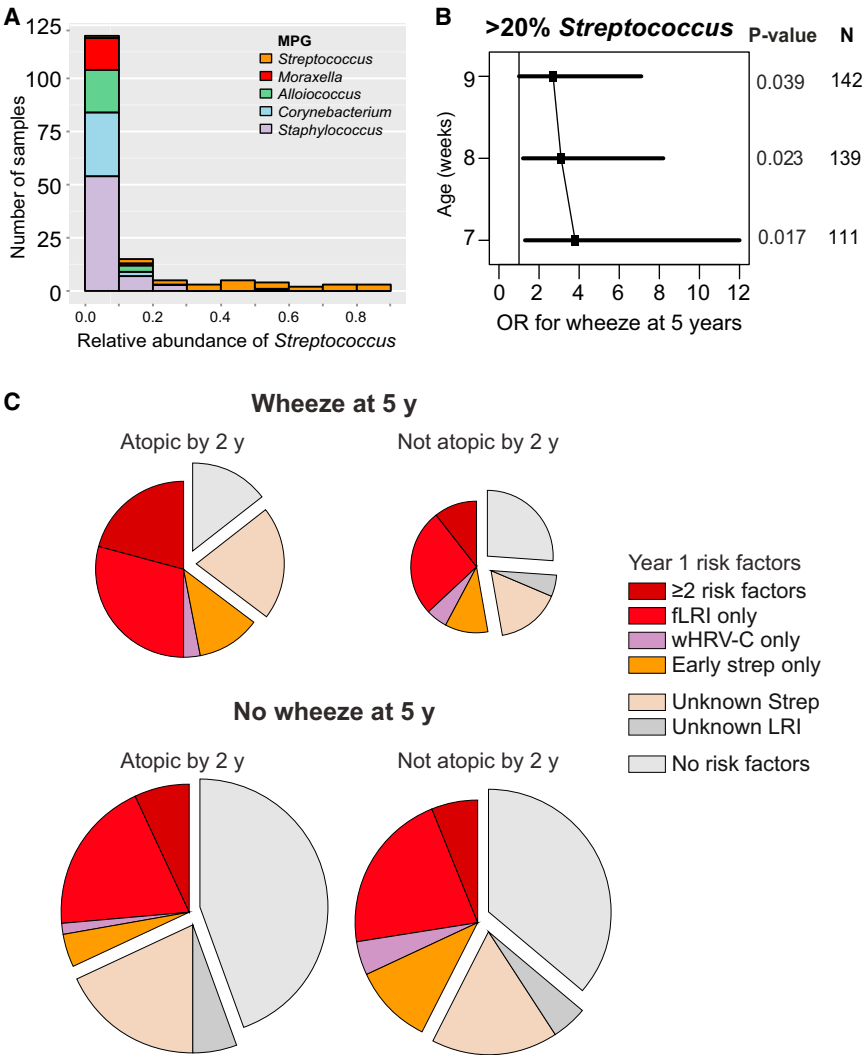


Figure 5. Predictors of Chronic Wheeze at Age 5

(A) *Streptococcus* abundance among healthy samples collected by 9 weeks of age, broken down by microbiome profile group (MPG).

(B) Adjusted odds ratio (OR, squares) and 95% confidence intervals (bars) for association between chronic wheeze at age 5 and high (> 20%) abundance of *Streptococcus* in the first healthy NP sample; p values and sample sizes (n) are indicated; individuals who experienced an infection prior to first healthy NP sample collection were excluded.

(C) Distribution of microbial events during infancy that were identified as risk factors for chronic wheeze at 5 years of age, stratified according to atopic status by age 2. fLRI, febrile LRI; wHRV-C, HRV-C LRI accompanied by wheeze; Strep, > 20% *Streptococcus* abundance in healthy NP sample taken in by 9 weeks old and prior to any ARI; unknown Strep, no such NP sample available (mainly due to ARI before 9 weeks of age); unknown LRI, incomplete viral/symptom profiling for LRI. Size of pie chart is proportional to the number of infants in each condition.

Importantly, our data show that the presence of *Moraxella* contributes to severity of RSV respiratory infections in infants (Figure S5, Table S3), which may be mechanistically related to RSV-*Moraxella* interactions reported in OM pathogenesis (Brockson et al., 2012).

Alloiococcus was a common and stable component of healthy NP samples and demonstrated enhanced stability in healthy NP microbiomes (Figure S3). *Alloiococcus* is a Gram-positive bacterium with one named species, *A. otitidis*, frequently detected in the ear canals of children with OM. Little is known about its mechanisms of colonization; however, the stable colonization patterns we observed may reflect a propensity to form stable biofilms, similar to *Moraxella*. The *Alloiococcus* MPG included some samples with high abundances of both *Alloiococcus* and *Corynebacterium*, indicating compatibility between these bacteria somewhat reminiscent of that between the *Moraxella* biofilm and *Streptococcus* or *Haemophilus*. Whether *Alloiococcus* is involved in OM pathology or is merely a healthy component of the microbiome is controversial (Tano et al., 2008); in our data, *Alloiococcus* was not associated with ARI or OM (Figure 1, Table S2).

Haemophilus was almost exclusively associated with ARI symptoms and was often found in consecutive infections, suggesting it can persist in the nasopharynx for some time (Figure 1, Table 1, Figure S3). Our analyses also suggest *Haemophilus* is highly transmissible between children and selected for by antibiotic exposure (Figure 2). Although our data include multiple *Haemophilus* OTUs, these epidemiological patterns are consistent with the human-adapted pathogen *H. influenzae*, which was supported by antibody data (Figure S2).

Role of the NP Microbiome in Asthma Development

We have previously reported that within the CAS cohort, the frequency during the first year of life of severe (wheezy and/or febrile) LRI was positively associated with subsequent risk for persistent wheeze at 5 and 10 years of age (Holt et al., 2010; Kusel et al., 2007, 2008), and similar findings have been reported in the U.S.-based COAST cohort (Gern, 2009). In both these study populations, the susceptible subgroup of children were those who developed early allergic sensitization (Gern, 2009; Holt et al., 2010, 2014; Jackson et al., 2008; Kusel et al., 2007, 2012; Oddy et al., 2002), and a variety of evidence suggests that the underlying mechanism involves synergistic interactions between atopic and anti-viral inflammatory pathways triggered within the infected airway mucosa (Holt and Sly, 2012). Of note, ARIs that remain restricted to the upper respiratory tract, and infections resulting in only mild LRI without wheezing or febrile symptoms, are relatively benign with respect to asthma risk in this cohort (Holt et al., 2010; Kusel et al., 2007, 2012)

(Table S3). Hence cofactors that can enhance the spread and ensuing severity of viral-initiated ARI are potentially central to asthma pathogenesis.

Here we found that two specific classes of LRI, namely febrile LRI and HRV-C-positive wheezy LRI, were independent risk factors for later chronic wheeze, especially among atopics (Table 2, Figure 3). Febrile LRI was common, occurring in one-third of all children and half of atopic chronic wheezers at age 5 (Figure 3 and Figure 5). Notably, the presence of illness-associated bacteria within the NP microbiome at the time of ARI increased both the risk for progression to the lower respiratory tract and the development of fever (Table 1). These mechanisms likely include myriad bacterial-viral interactions (recently reviewed by Vissers et al., 2014). In contrast, neither specific virus groups nor bacterial MPGs were associated with wheezy versus non-wheezy LRI within the overall population (Table 1), and the impact of wheezy LRI on later chronic wheeze was limited to LRI with HRV-C (Table 2), which occurred in just 11% of children. Re-examination of wheeze symptoms during LRI stratified by atopic status identified a positive association between HRV-C and wheeze during LRI among those children who later became sensitized (by 2 years of age), consistent with observations that HRV-C can induce wheeze in high-risk individuals (Lee et al., 2012). However, the presence of wheezy LRI alone, without further stratification by HRV-C detection and later atopic status, was not a significant predictor of later chronic wheeze, while febrile LRI per se was a significant predictor of chronic wheeze at 5 and 10 years, both among atopics and across the whole cohort (Figure 5, Table 2).

Early asymptomatic *Streptococcus* colonization at ~2 months of age, which occurred in 14% of children tested at that time, was significantly associated with chronic wheeze at 5 years of age (Figure 5, Table 2). Early *Streptococcus* colonization was not associated with incidence of infections with *Streptococcus* MPG, nor with detection of *S. pneumoniae*-specific antibodies at 12 months, suggesting that the mechanism is independent of innate immune response to *Streptococcus*. Rather, early *Streptococcus* colonization was associated with younger age of first LRI, and the level of subsequent asthma risk appears inversely related to age at initial *Streptococcus* colonization (Figure 4 and Figure 5). Collectively, these observations suggest that developing airway tissues may be maximally susceptible to the long-term effects of *Streptococcus*-mediated or LRI-mediated damage during the early postnatal period, when lung growth rates are most rapid.

To our knowledge, the only other study reporting the impact of early life NP colonization on wheeze/asthma at pre-school age is the 2007 report on the Copenhagen Prospective Study on Asthma in Childhood (n = 321, children born to mothers with asthma), in which a higher rate of wheeze at 5 years of age was detected among children colonized with *S. pneumoniae*, *M. catarrhalis*, and/or *H. influenzae* at 1 month post-birth (in culture-confirmed colonized, asthma prevalence was 33%, not colonized, 10%, OR 4.57) (Bisgaard et al., 2007). In that study, associations were not reported individually for the three species and colonization was measured as positive or negative by microbiological culture. In our cohort, wheeze at 5 years was not associated with early colonization with *Haemophilus* or *Moraxella*; however, *Haemophilus* was very rare in healthy NP samples, and *Moraxella* colonization was established later during infancy, which may be related to the warmer climate in Perth; these associations may differ in populations where asymptomatic carriage of these organisms is higher.

Conclusions and Implications

These findings collectively suggest that bacterial pathogens present in the NP microbiome at the time of upper respiratory viral infections during infancy are significant determinants of risk for the spread of infection to the lower airways and for the resultant expression of inflammatory symptoms marked by fever, and further, they may contribute directly and indirectly to the ensuing risk for development of persistent asthma, which itself is linked to these prior infectious episodes. Prevention of RSV or HRV infections in high-risk children, using immunoprophylaxis or vaccines, has been proposed as a mechanism for preventing the development and/or exacerbation of childhood asthma (Gern, 2009; Wu and Hartert, 2011). Our data suggest that in the absence of effective anti-viral therapies, targeting pathogenic bacteria present within the NP microbiome in this age group could represent an alternative approach toward the same goal. The pneumococcal vaccine, currently recommended from 2 months of age, could play a role; however, the niche created by eliminating vaccine-targeted *S. pneumoniae* serotypes can be readily filled by other serotypes and other bacteria such as *H. influenzae* and *M. catarrhalis* (Biesbroek et al., 2014; Revai et al., 2006). Manipulation of the microbiome by antibiotics may appear attractive; however, the association between antibiotic use in early childhood and subsequent asthma is controversial. In the CAS cohort (Kusel et al., 2008) and others (Semic-Jusufagic et al., 2014), this association has been proposed to arise through confounding and reverse causation, whereby genetic or clinical factors increase the likelihood of both antibiotic prescription (e.g., via genetic susceptibility to viral infection [Semic-Jusufagic et al., 2014]) and asthma development; others have proposed that antibiotic-induced disruption of the gut microbiota could explain the link (Arrieta and Finlay, 2014). Our analyses provide a potential causal pathway linking antibiotics to later asthma, whereby antibiotic use in infants selects for illness-associated bacteria in the NP microbiome, leading to increased risk of febrile LRI and later asthma development. Genetic analysis of the infants in our cohort could potentially help to unravel this in the future. Importantly, further studies are required to prospectively assess the impact of antibiotic use on NP bacterial colonization during health and infection, and the subsequent development of ARI, atopy, and wheeze.

EXPERIMENTAL PROCEDURES

Study Design

This study is an extension of the Childhood Asthma Study (CAS), a birth cohort of 234 infants at high risk of atopy as previously described (Kusel et al., 2007, 2008). Briefly, healthy NP aspirates (NPAs) were collected from subjects at planned visits at 2, 6, and 12 months of age. NPAs were also collected within 48 hr from the onset of an ARI; these were classified as either lower respiratory infection (LRI, if wheeze or rattly chest detected) or upper respiratory infection (URI, otherwise). NPAs were divided into four aliquots and stored at −80°C. One aliquot was used for bacterial 16S rRNA amplicon sequencing via Illumina MiSeq and another to screen for viruses including RSV, HRV, and other picornaviruses, influenza, parainfluenza, coronavirus, adenovirus, and human

metapneumovirus (hMPV). For LRI, a third aliquot was used in an expanded screen for detection and subtyping of HRV. Study ethics were approved by the ethics committees of King Edward Memorial and Princess Margaret Hospitals in Western Australia. Fully informed parental consent was obtained for all subjects. Details of sample and data collection, viral screening, DNA extraction, and sequencing are in Supplemental Experimental Procedures.

16S rRNA Sequence Analysis

Paired end reads were merged using Flash v1.2.7 (Magoč and Salzberg, 2011) with read length 151 base pairs (bp) and expected fragment length 253 bp. Merged sequences were quality filtered as follows: $\leq$ 3 low-quality bp (Phred score < 3) allowed before trimming, $\geq$ 189 consecutive high-quality bp with no uncalled bases (Ns) (Bokulich et al., 2013). Thirty-three million sequences were filtered out (13%), leaving 219 million for analysis. These were assigned to OTUs using the closed reference method in QIIME v1.8 (Caporaso et al., 2010) with the Greengenes 99% OTU reference set, version 13_5 (McDonald et al., 2012) (which consists of > 200,000 representative sequences obtained from clustering all sequences from the Greengenes reference database at 99% sequence similarity). Briefly, this method uses UCLUST (Edgar, 2010) to search each sequence against the reference set and assigns it to an OTU based on the best hit at $\geq$ 99% sequence identity. Sequences not matching the reference database (3%) were excluded from analysis; these sequences could be chimeras, sequencing errors, or novel sequences that are not well characterized in the database. This left > 200,000 taxonomy-assigned sequences on average for each NPA (interquartile range: 108,000–255,000; eight samples with < 1,000 reads). Compared to NPAs, negative control samples (see Supplemental Experimental Procedures) had clearly lower total reads (interquartile range: 1,044–2,346) with a different community structure (Figure S6).

Clustering into MPGs

For each NPA, the relative abundance of each OTU was calculated. Most analyses were summarized at genus level, whereby all OTUs assigned to the same genus were collapsed into a single group for reporting. Samples were assigned to microbiome profile groups (MPGs) based on hierarchical clustering of the relative abundances of the six most common genera (distance metric: 1-Pearson's correlation; clustering method: Ward's minimum variance, implemented in *R hclust*). Previous microbiome studies have reported clinically or environmentally meaningful associations with community profiles or types (e.g., enterotypes) identified using similar clustering approaches (Ravel et al., 2011; Zhou et al., 2014).

Statistical Methods

All statistical analyses were performed using *R*. ORs were estimated using generalized estimating equations (GEE) logistic regression with unstructured correlation and robust standard errors, to take into account multiple samples from the same subject. Survival analyses were done using Cox proportional hazards models. Potential confounders were adjusted for by including them in the regression models. Full details of each analysis and variable definitions are given in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

Short read 16S data for this study have been deposited in NCBI GenBank under accession number SRP056779.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.chom.2015.03.008.

ACKNOWLEDGMENTS

This study was supported by the NHMRC of Australia (Project Grant #1049539; Fellowships #1061409 [K.E.H.] and #1061435 [M.I., co-funded with the Australian Heart Foundation]) and the Victorian Life Sciences Computation Initiative (VLSCI) (#VR0082). S.L.J. is supported in part by a Chair from Asthma UK (CH11SJ) and by Medical Research Council Centre grant G1000758. J.E.G., Y.A.B., and K.G. are supported by NIH grants U19 AI104317 and P01 HL070831.

Received: January 6, 2015
Revised: February 9, 2015
Accepted: March 9, 2015
Published: April 9, 2015

REFERENCES

Arrieta, M.C., and Finlay, B. (2014). The intestinal microbiota and allergic asthma. J. Infect. *69* (1), S53–S55.

Biesbroek, G., Wang, X., Keijser, B.J., Eijkemans, R.M., Trzciński, K., Rots, N.Y., Veenhoven, R.H., Sanders, E.A., and Bogaert, D. (2014). Seven-valent pneumococcal conjugate vaccine and nasopharyngeal microbiota in healthy children. Emerg. Infect. Dis. *20*, 201–210.

Bisgaard, H., Hermansen, M.N., Buchvald, F., Loland, L., Halkjaer, L.B., Bønnelykke, K., Brasholt, M., Heltberg, A., Vissing, N.H., Thorsen, S.V., et al. (2007). Childhood asthma after bacterial colonization of the airway in neonates. N. Engl. J. Med. *357*, 1487–1495.

Bisgaard, H., Hermansen, M.N., Bønnelykke, K., Stokholm, J., Baty, F., Skytt, N.L., Aniscenko, J., Kebadze, T., and Johnston, S.L. (2010). Association of bacteria and viruses with wheezy episodes in young children: prospective birth cohort study. BMJ *341*, c4978.

Bochkov, Y.A., and Gern, J.E. (2012). Clinical and molecular features of human rhinovirus C. Microbes Infect. *14*, 485–494.

Bogaert, D., Keijser, B., Huse, S., Rossen, J., Veenhoven, R., van Gils, E., Bruin, J., Montijn, R., Bonten, M., and Sanders, E. (2011). Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. PLoS ONE *6*, e17035.

Bokulich, N.A., Subramanian, S., Faith, J.J., Gevers, D., Gordon, J.I., Knight, R., Mills, D.A., and Caporaso, J.G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nat. Methods *10*, 57–59.

Brockson, M.E., Novotny, L.A., Jurcisek, J.A., McGillivary, G., Bowers, M.R., and Bakaletz, L.O. (2012). Respiratory syncytial virus promotes Moraxella catarrhalis-induced ascending experimental otitis media. PLoS ONE *7*, e40088.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods *7*, 335–336.

de Vries, S.P., Bootsma, H.J., Hays, J.P., and Hermans, P.W. (2009). Molecular aspects of Moraxella catarrhalis pathogenesis. Microbiol. Mol. Biol. Rev. *73*, 389–406.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

Gern, J.E. (2009). Rhinovirus and the initiation of asthma. Curr. Opin. Allergy Clin. Immunol. *9*, 73–78.

Gisselsson-Solén, M., Henriksson, G., Hermansson, A., and Melhus, A. (2014). Risk factors for carriage of AOM pathogens during the first 3 years of life in children with early onset of acute otitis media. Acta Otolaryngol. *134*, 684–690.

Hales, B.J., Chai, L.Y., Elliot, C.E., Pearce, L.J., Zhang, G., Heinrich, T.K., Smith, W.A., Kusel, M.M., Holt, P.G., Sly, P.D., and Thomas, W.R. (2012). Antibacterial antibody responses associated with the development of asthma in house dust mite-sensitised and non-sensitised children. Thorax *67*, 321–327.

Hilty, M., Burke, C., Pedro, H., Cardenas, P., Bush, A., Bossley, C., Davies, J., Ervine, A., Poulter, L., Pachter, L., et al. (2010). Disordered microbial communities in asthmatic airways. PLoS ONE *5*, e8578.

Holt, P.G., and Sly, P.D. (2012). Viral infections and atopy in asthma pathogenesis: new rationales for asthma prevention and treatment. Nat. Med. *18*, 726–735.

Holt, P.G., Rowe, J., Kusel, M., Parsons, F., Hollams, E.M., Bosco, A., McKenna, K., Subrata, L., de Klerk, N., Serralha, M., et al. (2010). Toward improved prediction of risk for atopy and asthma among preschoolers: a prospective cohort study. J. Allergy Clin. Immunol. *125*, 653–659, e1, e7.

Holt, P.G., Strickland, D.H., Hales, B.J., and Sly, P.D. (2014). Defective respiratory tract immune surveillance in asthma: a primary causal factor in disease onset and progression. Chest *145*, 370–378.

Jackson, D.J., Gangnon, R.E., Evans, M.D., Roberg, K.A., Anderson, E.L., Pappas, T.E., Printz, M.C., Lee, W.M., Shult, P.A., Reisdorf, E., et al. (2008). Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children. Am. J. Respir. Crit. Care Med. *178*, 667–672.

Kusel, M.M., de Klerk, N.H., Holt, P.G., Kebadze, T., Johnston, S.L., and Sly, P.D. (2006). Role of respiratory viruses in acute upper and lower respiratory tract illness in the first year of life: a birth cohort study. Pediatr. Infect. Dis. J. *25*, 680–686.

Kusel, M.M., de Klerk, N.H., Kebadze, T., Vohma, V., Holt, P.G., Johnston, S.L., and Sly, P.D. (2007). Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma. J. Allergy Clin. Immunol. *119*, 1105–1110.

Kusel, M.M., de Klerk, N., Holt, P.G., and Sly, P.D. (2008). Antibiotic use in the first year of life and risk of atopic disease in early childhood. Clin. Exp. Allergy *38*, 1921–1928.

Kusel, M.M., Kebadze, T., Johnston, S.L., Holt, P.G., and Sly, P.D. (2012). Febrile respiratory illnesses in infancy and atopy are risk factors for persistent asthma and wheeze. Eur. Respir. J. *39*, 876–882.

Lee, W.M., Lemanske, R.F., Jr., Evans, M.D., Vang, F., Pappas, T., Gangnon, R., Jackson, D.J., and Gern, J.E. (2012). Human rhinovirus species and season of infection determine illness severity. Am. J. Respir. Crit. Care Med. *186*, 886–891.

Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963.

McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., and Hugenholtz, P. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. *6*, 610–618.

Oddy, W.H., de Klerk, N.H., Sly, P.D., and Holt, P.G. (2002). The effects of respiratory infections, atopy, and breastfeeding on childhood asthma. Eur. Respir. J. *19*, 899–905.

Penders, J., Kummeling, I., and Thijs, C. (2011). Infant antibiotic use and wheeze and asthma risk: a systematic review and meta-analysis. Eur. Respir. J. *38*, 295–302.

Ravel, J., Gajer, P., Abdo, Z., Schneider, G.M., Koenig, S.S., McCulle, S.L., Karlebach, S., Gorle, R., Russell, J., Tacket, C.O., et al. (2011). Vaginal microbiome of reproductive-age women. Proc. Natl. Acad. Sci. USA *108* (1), 4680–4687.

Revai, K., McCormick, D.P., Patel, J., Grady, J.J., Saeed, K., and Chonmaitree, T. (2006). Effect of pneumococcal conjugate vaccine on nasopharyngeal bacterial colonization during acute otitis media. Pediatrics *117*, 1823–1829.

Rollins, D.R., Beuther, D.A., and Martin, R.J. (2010). Update on infection and antibiotics in asthma. Curr. Allergy Asthma Rep. *10*, 67–73.

Schaumburg, F., Alabi, A.S., Mombo-Ngoma, G., Kaba, H., Zoleko, R.M., Diop, D.A., Mackanga, J.R., Basra, A., Gonzalez, R., Menendez, C., et al. (2014). Transmission of Staphylococcus aureus between mothers and infants in an African setting. Clin. Microbiol. Infect. *20*, O390–O396.

Semic-Jusufagic, A., Belgrave, D., Pickles, A., Telcian, A.G., Bakhsoliani, E., Sykes, A., Simpson, A., Johnston, S.L., and Custovic, A. (2014). Assessing the association of early life antibiotic prescription with asthma exacerbations, impaired antiviral immunity, and genetic variants in 17q21: a population-based birth cohort study. Lancet Respir Med *2*, 621–630.

Sly, P.D., Boner, A.L., Björksten, B., Bush, A., Custovic, A., Eigenmann, P.A., Gern, J.E., Gerritsen, J., Hamelmann, E., Helms, P.J., et al. (2008). Early identification of atopy in the prediction of persistent asthma in children. Lancet *372*, 1100–1106.

Spaniol, V., Troller, R., Schaller, A., and Aebi, C. (2011). Physiologic cold shock of Moraxella catarrhalis affects the expression of genes involved in the iron acquisition, serum resistance and immune evasion. BMC Microbiol. *11*, 182.

Stein, R.T., and Martinez, F.D. (2010). Respiratory syncytial virus and asthma: still no final answer. Thorax *65*, 1033–1034.

Tai, A., Tran, H., Roberts, M., Clarke, N., Wilson, J., and Robertson, C.F. (2014). The association between childhood asthma and adult chronic obstructive pulmonary disease. Thorax *69*, 805–810.

Tano, K., von Essen, R., Eriksson, P.O., and Sjöstedt, A. (2008). Alloiococcus otitidis—otitis media pathogen or normal bacterial flora? APMIS *116*, 785–790.

Verhaegh, S.J., Snippe, M.L., Levy, F., Verbrugh, H.A., Jaddoe, V.W., Hofman, A., Moll, H.A., van Belkum, A., and Hays, J.P. (2011). Colonization of healthy children by Moraxella catarrhalis is characterized by genotype heterogeneity, virulence gene diversity and co-colonization with Haemophilus influenzae. Microbiology *157*, 169–178.

Vissers, M., de Groot, R., and Ferwerda, G. (2014). Severe viral respiratory infections: are bugs bugging? Mucosal Immunol. *7*, 227–238.

Weinstock, G.M. (2012). Genomic approaches to studying the human microbiota. Nature *489*, 250–256.

Wu, P., and Hartert, T.V. (2011). Evidence for a causal relationship between respiratory syncytial virus infection and asthma. Expert Rev. Anti Infect. Ther. *9*, 731–745.

Zhou, Y., Mihindukulasuriya, K.A., Gao, H., La Rosa, P.S., Wylie, K.M., Martin, J.C., Kota, K., Shannon, W.D., Mitreva, M., Sodergren, E., and Weinstock, G.M. (2014). Exploration of bacterial community classes in major human habitats. Genome Biol. *15*, R66.

Cell Host & Microbe

Supplemental Information

The Infant Nasopharyngeal Microbiome Impacts the Incidence of Lower Respiratory Infection and Risk of Asthma Development

Shu Mei Teo, Danny Mok, Kym Pham, Merci Kusel, Michael Serralha, Niamh Troy, Barbara J. Holt, Belinda J. Hales, Michael L. Walker, Elysia Hollams, Yury A Bochkov, Kristine Grindle, Sebastian L. Johnston, James E Gern, Peter D. Sly, Patrick G. Holt, Kathryn E. Holt, and Michael Inouye

Supplemental Data

Figure S1, related to Figure 1. Subtree of the Greengenes 99% sequence similarity OTU tree for genus (**A**) *Haemophilus.1,* (**B**) *Moraxella,* (**C**) *Streptococcus,* (**D**) *Alloiococcus,* (**E***) Corynebacterium,* (**F**) *Staphylococcus*. OTUs present at >0.1% frequency across our 1,021 NP samples are highlighted, and relative frequencies within these samples are shown with blue bars (right).

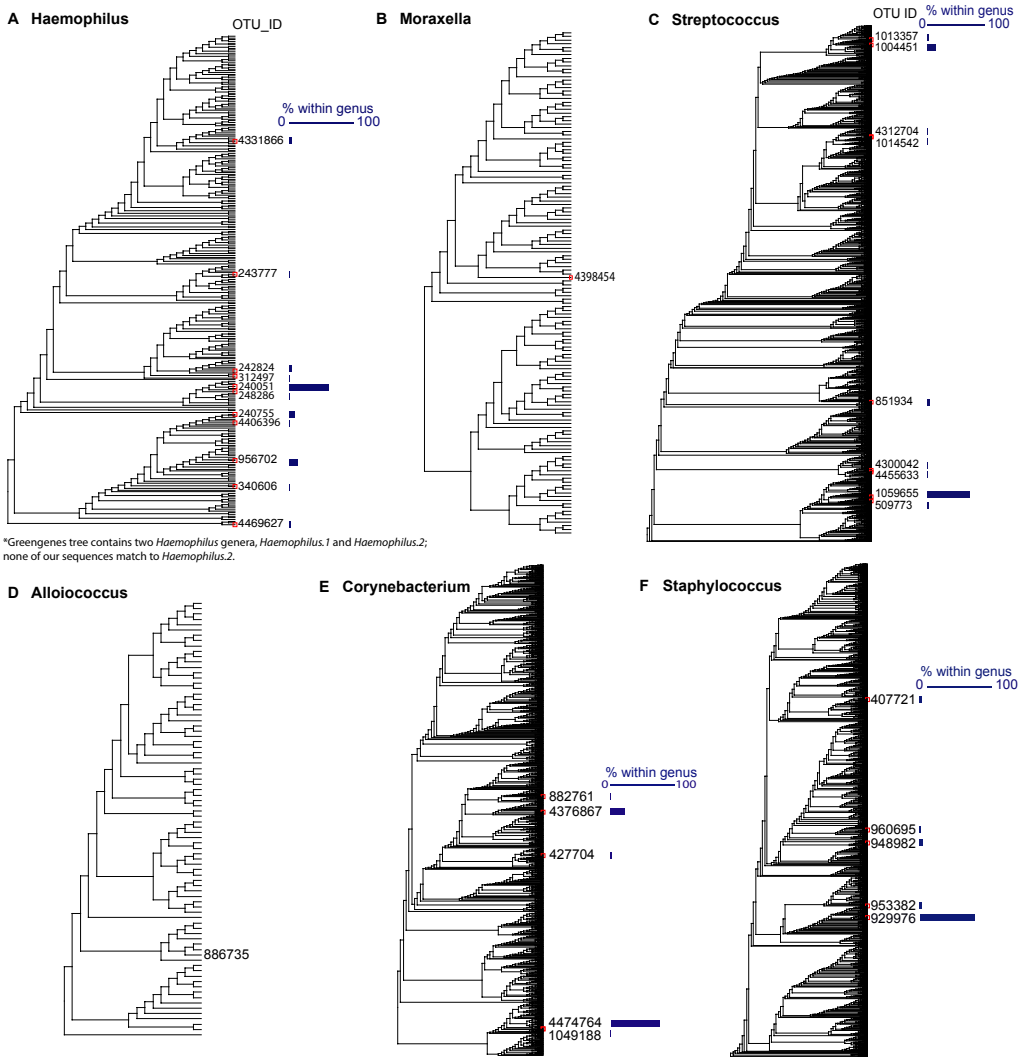
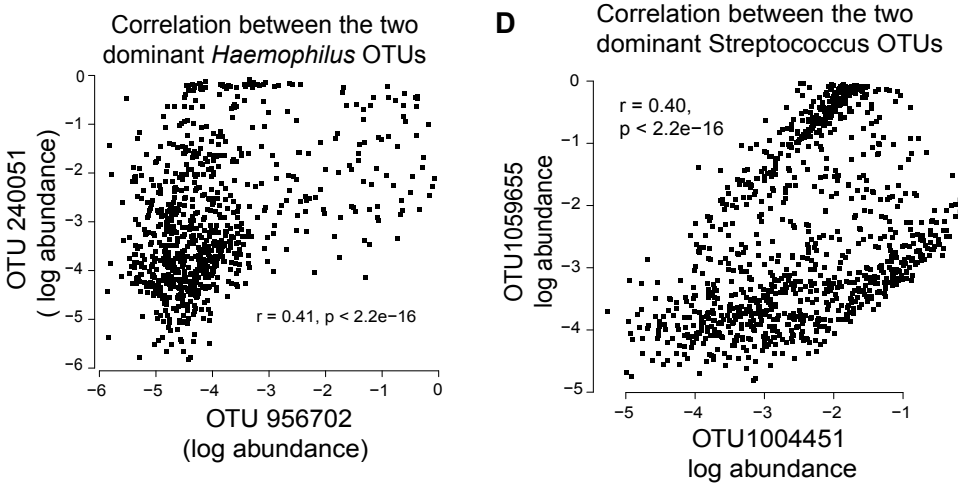


Figure S2, related to Figure 1. Association with species-specific antibodies.
(A) Correlation between the two most frequent *Haemophilus* OTUs. (**B-C**) Association between detectable IgG1/IgG4 antibodies to *H. influenzae* P4/P6 surface proteins (measured at 12 months of age) and number of (**B**) prior ARI NP samples or (**C**) prior asymptomatic NP samples containing the dominant *Haemophilus* OTUs. **(D)** Correlation between the two most frequent *Streptococcus* OTUs. **(E-F)** Association between detectable IgG1 antibodies to *S. pneumoniae* pneumococcal surface protein A1, A2, or C (measured at 12 months of age) and number of **(E)** prior ARI NP samples or **(F)** prior asymptomatic NP samples containing the dominant *Streptococcus* OTUs.



B Association between detectable *H. influenzae* antibodies at 12 months and number of prior infections with *Haemophilus* OTUs.

Presence of OTU	OR (95% CI)
240051	1.9 (1.3 - 2.9)**
956702	3.6 (1.8 - 6.9)**

** p < 0.001

C Association between detectable *H. influenzae* antibodies at 12 months and any prior asymptomatic NP samples with *Haemophilus* OTUs.

Presence of OTU	OR (95% CI)
240051	1.2 (0.67 - 2.0)
956702	2.1 (0.97 - 4.5)

E Association between detectable *S.pneumoniae* antibodies at 12 months and number of prior infections with *Streptococcus* OTUs.

Presence of OTU	OR (95% CI)
1059655	2.1 (1.4 - 3.2)**
1004451	1.3 (0.81 - 2.0)

** p < 0.001

F Association between detectable *S.pneumoniae* antibodies at 12 months and any prior asymptomatic NP samples with *Streptococcus* OTUs.

Presence of OTU	OR (95% CI)
1059655	1.1 (0.61 - 1.9)
1004451	0.50 (0.24 - 1.0)

Figure S3, related to section "*NP microbiome dynamics*". Transitions within subjects. (A-C) Transitions between **(A)** consecutive healthy samples, with no intervening infections, **(B)** with intervening infections, and **(C)** between consecutive infection samples. T1, 1st timepoint; T2, next timepoint. Cell numbers indicate the number of cases in which the respective transition from T1 to T2 was observed; cells are coloured to indicate the row proportions as per legend. Stable transitions were defined as those with the same MPG at T1 and T2 (i.e. diagonals in the matrix). For each panel, the table at the right shows observed and expected frequencies of stable transitions for each genus MPG (note *Haemophilus* is not plotted in (A) and (B) because it was extremely rare in healthy samples). The expected frequency for each stable transition was taken to be the square of the proportion of samples in that MPG at T1 (*i.e.,* assuming constant frequencies and random transitions). Observed values are the proportion of stable transitions out of all observed transitions; 95% confidence intervals were calculated from 1,000 bootstraps of the real data, sampled with replacement. **(D)** Frequency of stable transitions between MPGs in consecutive ARIs, stratified by time between ARIs. Stable transitions were defined as pairs of consecutive ARIs from the same individual that shared the same MPG.

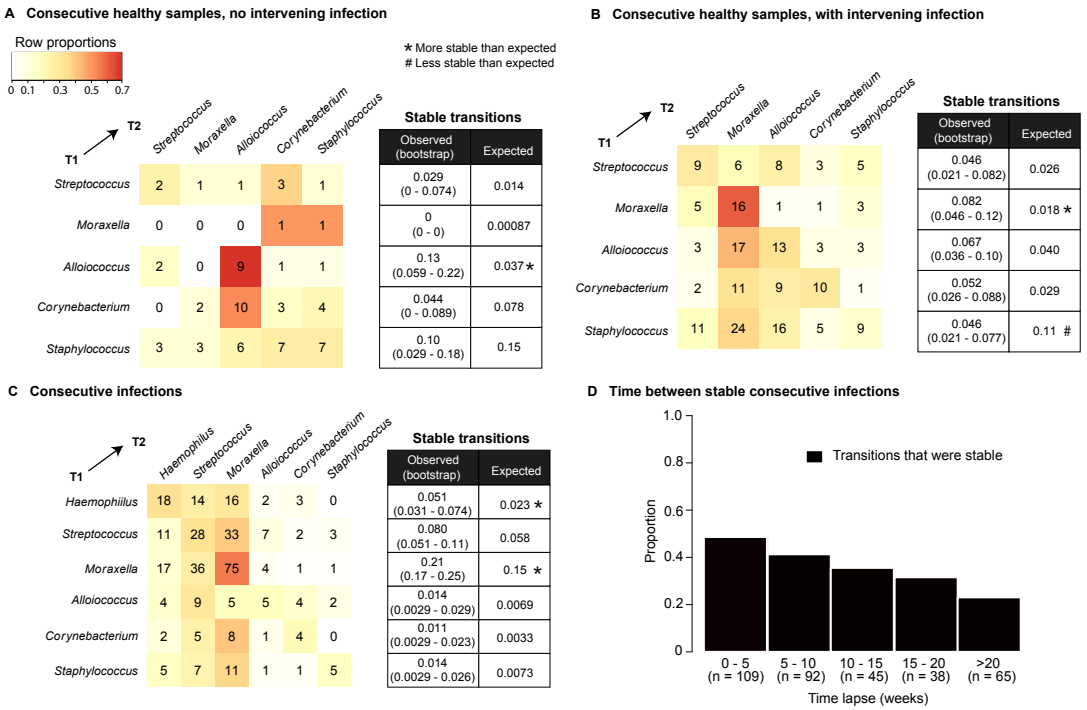


Figure S4, related to section "*NP microbial determinants of ARI symptoms*". Rates of bacterial microbiome profile groups (MPGs) and viruses for different types of NP samples. ARI, acute respiratory illness; URI, upper respiratory illness; LRI, lower respiratory illness; fLRI, febrile LRI; wLRI, wheezy LRI. Note HRV subtyping was performed for LRI samples only.

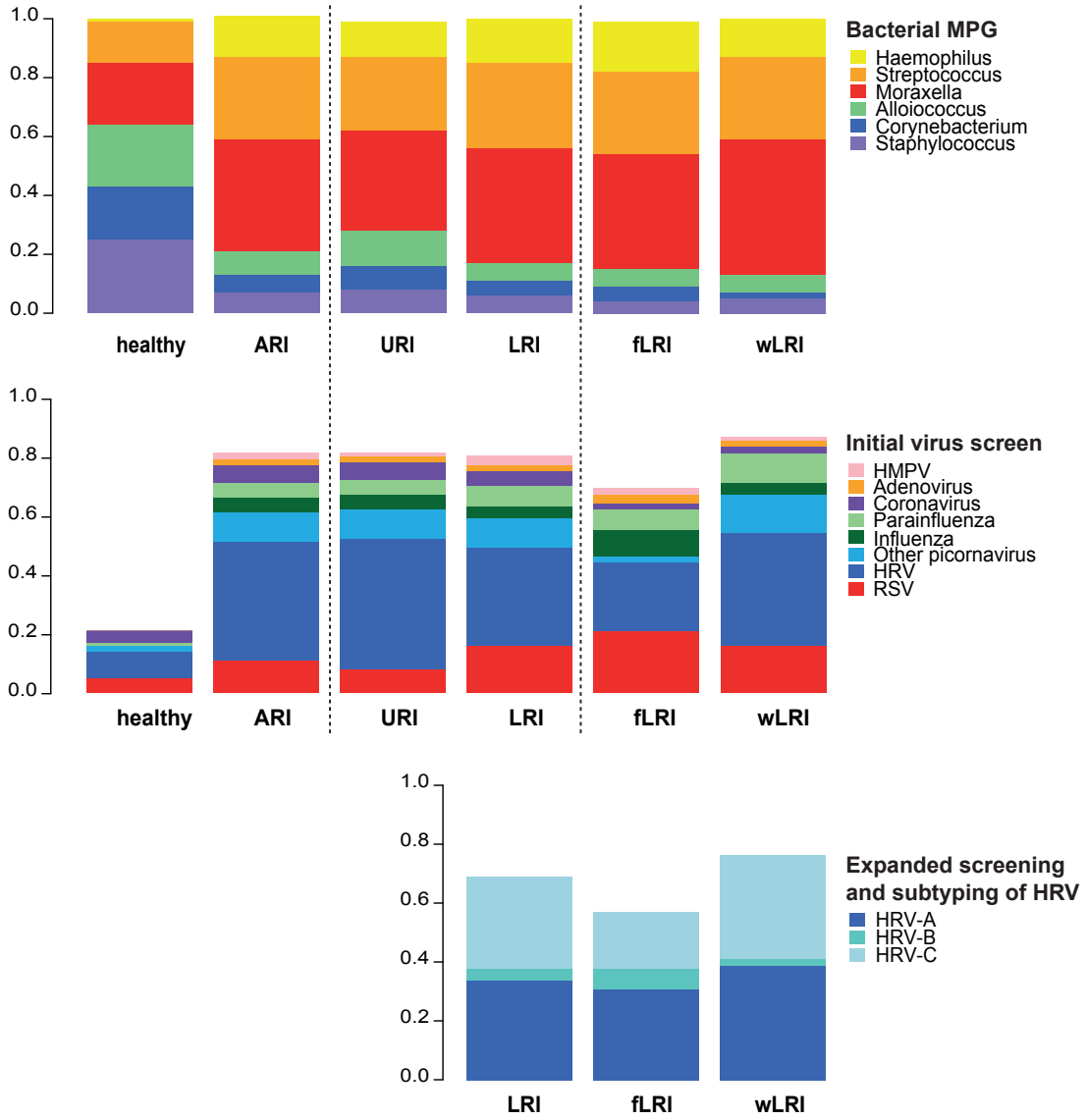


Figure S5, related to Table 1. Distribution of samples by MPG, amongst (A) RSV positive and (B) RSV negative samples. Logistic regression was used to assess, amongst (A) RSV-positive LRIs, and (B) RSV-negative LRIs, the association of fever status with *Moraxella*, *Streptococcus* and *Haemophilus* MPGs as independent covariates in the same model, adjusted for age, gender and season (p-values shown). Samples with antibiotics intake within the last week were excluded from this analysis.

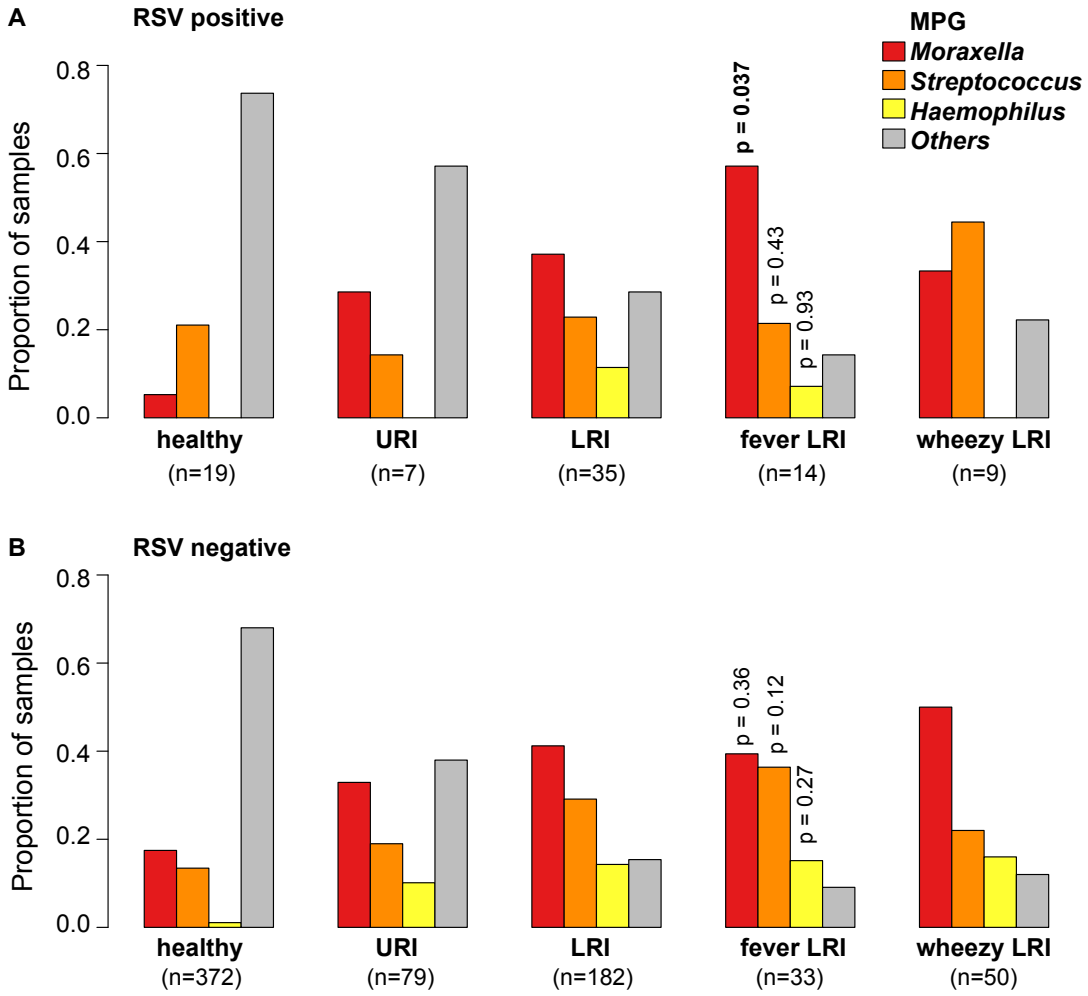
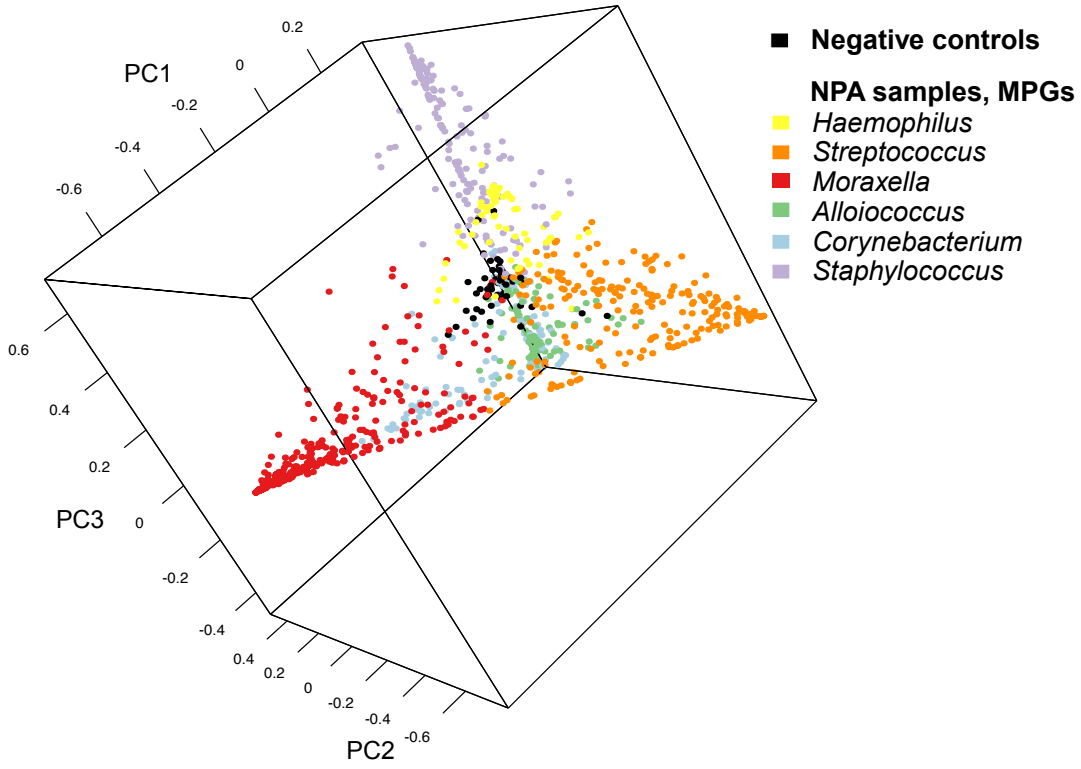


Figure S6, related to "Experimental procedures". Principal components analysis of OTU relative abundances for NPA samples and negative controls



Characteristics of subjects	Number	Proportion
Total participants	234	100%
Male gender	132	56%
Paternal history of atopic disease	156	67%
Maternal history of atopic disease	196	84%
Vaginal delivery	167	72%
Breastfed at 3 months	196	84%
Smoking exposure at 1 year	44	19%
Day-care attendance at 1 year	67	29%
Current wheeze at 1 year	76	33%
Other children in house	108	47%
Furry pet in the first year	109	48%
≥1 LRI in first 425 days of life	167	75%
Atopic at 6 months of age	47	21%
Atopic at 2 years of age	88	41%
Sensitised to inhaled allergens[#] by 1 year of age	23	10%
Current wheeze at 5 years of age	56	28%

Table S1. Characteristics of infants in the study. [#] Inhaled allergens include house dust mite, cat epithelium and dander, couch grass, rye grass, or mould mix.

MPG	Number of OM ARI (LRI)	OM vs. non-OM ARI OR (95% CI)	OM vs. non-OM LRI OR (95% CI)
Staphylococcus	3 (3)	2.2 (0.52-8.9) p = 0.29	2.0 (0.51-8.0) p = 0.32
Corynebacterium	1 (0)	0.27 (0.036-2.1) p = 0.21	0.32 (0.041-2.4) p = 0.27
Alloiococcus	2 (2)	0.36 (0.089-1.5) p = 0.16	0.44 (0.1-1.9) p = 0.28
Moraxella	25 (16)	1.1 (0.61-2.0) p = 0.71	1.0 (0.58-1.8) p = 0.92
Streptococcus	21 (17)	1.2 (0.66-2.3) p = 0.50	1.2 (0.64-2.3) p = 0.55
Haemophilus	9 (5)	0.90 (0.33-2.5) p = 0.84	0.83 (0.31-2.2) p = 0.70

Table S2, related to section *"NP microbial determinants of ARI symptoms"*. Distribution of microbiome profile groups (MPGs) among ARI with otitis media (OM) diagnosis. Odds ratios (OR) for association of each MPG with OM, estimated using generalized estimating equations (GEE) logistic regression with unstructured correlation and robust standard errors, adjusted for age at infection. ARI, acute respiratory infection; LRI, lower respiratory illness.

	All children		Atopic by 2 years		Not atopic by 2 years	
	Wheeze at 5 years	**Wheeze at 10 years**	**Wheeze at 5 years**	**Wheeze at 10 years**	**Wheeze at 5 years**	**Wheeze at 10 years**
Any URI	0.81 (0.21-4) p = 0.78	0.71 (0.14-5.3) p = 0.69	0.42 (0.067-2.6) p = 0.33	0.48 (0.068-4.1) p = 0.46	NA	NA
Any LRI	0.95 (0.47-2) p = 0.89	1.9 (0.74-5.6) p = 0.2	1.4 (0.51-3.8) p = 0.55	1.9 (0.61-6.6) p = 0.29	0.57 (0.18-1.9) p = 0.35	2.6 (0.38-52) p = 0.41
Any fever LRI	**2.3 (1.2-4.5) p = 0.016**	2.2 (0.93-5.2) p = 0.071	**2.7 (1.1-7) p = 0.034**	2.7 (0.89-8.7) p = 0.083	1.5 (0.48-4.6) p = 0.46	1.9 (0.33-9.8) p = 0.45
Any wheezy LRI	1.6 (0.79-3) p = 0.2	1.4 (0.58-3.3) p = 0.45	1.3 (0.49-3.3) p = 0.61	1.6 (0.53-4.8) p = 0.4	1.9 (0.66-5.7) p = 0.23	1.2 (0.22-5.7) p = 0.83
Any mild LRI	0.48 (0.24-0.94) p = 0.035	0.92 (0.4-2.1) p = 0.84	0.51 (0.2-1.2) p = 0.14	0.76 (0.27-2.1) p = 0.61	0.44 (0.13-1.3) p = 0.16	1.6 (0.35-9.1) p = 0.54
Any HRV LRI	1.1 (0.59-2.2) p = 0.69	1.6 (0.69-3.8) p = 0.28	1.2 (0.49-2.9) p = 0.7	1.2 (0.43-3.3) p = 0.75	0.93 (0.31-2.7) p = 0.9	2.8 (0.59-16) p = 0.21
Any RSV LRI	1.1 (0.51-2.4) p = 0.76	1.4 (0.51-3.5) p = 0.5	2.2 (0.76-6.2) p = 0.14	2.7 (0.8-8.9) p = 0.11	0.56 (0.11-2.1) p = 0.42	0.33 (0.016-2.5) p = 0.35
Any risk bacteria LRI	0.89 (0.45-1.8) p = 0.73	2 (0.82-5.2) p = 0.14	0.96 (0.39-2.4) p = 0.93	1.7 (0.59-5.2) p = 0.34	0.73 (0.25-2.3) p = 0.57	4.9 (0.72-98) p = 0.16
Any HRV wheezy LRI	2 (0.93-4.2) p = 0.073	2.1 (0.81-5.4) p = 0.11	2.5 (0.86-7.2) p = 0.092	1.9 (0.55-6.4) p = 0.29	1.8 (0.49-6.3) p = 0.34	2.7 (0.46-14) p = 0.24
Any HRV-C wheezy LRI	2.4 (0.93-6.1) p = 0.064	**3.5 (1.1-11) p = 0.026**	**7.2 (1.7-35) p = 0.009**	**7.1 (1.6-40) p = 0.014**	1.1 (0.15-5) p = 0.95	1.2 (0.058-11) p = 0.86
Any HRV-A wheezy LRI	1.2 (0.42-3.1) p = 0.74	1.4 (0.37-4.7) p = 0.57	0.55 (0.11-2.2) p = 0.43	0.35 (0.018-2.2) p = 0.34	3.5 (0.74-16) p = 0.10	6.3 (0.94-44) p = 0.051
Any RSV wheezy LRI	2.6 (0.71-9.5) p = 0.15	1.7 (0.34-7.5) p = 0.47	7.2 (1-150) p = 0.084	2.5 (0.28-20) p = 0.37	0.95 (0.045-8) p = 0.97	1.2 (0.052-12) p = 0.89
***Streptococcus* colonization at 7 weeks**	**3.8 (1.3-12) p = 0.017**	2.7 (0.66-12) p = 0.18	4.0 (0.88-21) p = 0.077	3.9 (0.63-28) p = 0.15	3.4 (0.55-22) p = 0.17	NA
***Streptococcus* colonization at 8 weeks**	**3.1 (1.2-8.2) p = 0.023**	1.9 (0.46-7.1) p = 0.33	**4.4 (1.1-20) p = 0.040**	2.4 (0.46-12) p = 0.29	2.7 (0.47-14) p = 0.24	NA
***Streptococcus* colonization at 9 weeks**	**2.7 (1.0-7.1) p = 0.039**	1.9 (0.45-6.9) p = 0.36	3.3 (0.86-13) p = 0.080	2.4 (0.47-13) p = 0.28	2.6 (0.45-14) p = 0.26	NA

Table S3, related to Table 2. Association between chronic wheeze at age 5 and 10 years, by LRI subtypes and early asymptomatic colonization with *Streptococcus*. Any mild LRI: Any LRI without wheeze or fever symptoms; Any risk bacteria LRI: any LRI in *Streptococcus*, *Moraxella* or *Haemophilus* MPG. *Streptococcus* colonization in the first healthy sample (collected by 7/8/9 weeks of age and prior to any respiratory infection) was assessed using >20% abundance cut-off. Odds ratios, 95% confidence intervals and p-values shown were estimated using logistic regression, adjusted for gender and maternal and paternal history of atopic disease, calculated separately for all children, those who were atopic by 2 years and those not atopic by 2 years.

Variable	*Alloiococcus* colonized (n = 63)	*Moraxella* colonized (n = 64)	Others (n = 107)	*Alloiococcus* vs. *Moraxella* (p-value)	*Alloiococcus* vs. all others (p-value)
Mean no. ARI	4.6	5.0	5.4	0.31	0.077
Mean no. LRI	1.3	1.7	1.9	0.32	0.10
Mean no. fLRI	0.44	0.49	0.51	0.53	0.47
Mean no. wLRI	0.32	0.62	0.53	0.14	0.11
Mean no. OM	0.44	0.57	0.47	0.29	0.54
Any RSV+ ARI	18	30	43	0.094	0.12
Any RSV+ LRI	5	17	27	0.017*	0.0050*
Any RSV+ fLRI	1	8	10	0.034*	0.050*
Wheeze at age 5 y **Atopic** **Non-atopic** **No wheeze**	 8 4 44	 11 6 38	 15 10 56	0.45#	0.41#
Wheeze at age 10 y **Atopic** **Non-atopic** **No wheeze**	 5 1 36	 4 2 35	 13 5 43	1#	0.46#
Atopic by age 2 y	33	32	51	1	0.88

Table S4, related to section *"Impact of NP colonization on ARI"*. Infections and wheeze phenotypes for infants by colonization status. Groups are defined by microbiome profile group (MPG) clustering of healthy samples. '*Alloiococcus* colonized', ≥1 healthy sample with *Alloiococcus* MPG but none with *Moraxella* MPG; '*Moraxella* colonized', ≥1 healthy sample with *Moraxella* MPG but none with *Alloiococcus* MPG; 'Others', no *Moraxella* or *Alloiococcus* healthy sample or ≥1 of each. Infection types: ARI, acute respiratory infection; URI, upper respiratory illness; LRI, lower respiratory illness; fLRI, febrile LRI; wLRI, wheezy LRI; OM, otitis media; RSV+, PCR detection of respiratory syncytial virus; atopy status is defined by IgE > 0.35 kU/L for any antigen at 6, 12 or 24 months. Comparisons of rates of infection were assessed using Wilcoxon rank sum test; other comparisons are binary variables and were calculated using Fisher's exact test; # atopic wheeze vs no wheeze; *$p < 0.05$.

Supplemental Experimental Procedures:

Study design and participants

This study is an extension of the Childhood Asthma Study (CAS) – a prospective community-based birth cohort study of 234 infants at high risk of atopy (at least 1 parent with a doctor diagnosed history of hay fever, asthma or eczema). Parents of the subjects completed questionnaires prior to birth, and when each child reached 1, 2, 3, 4, 5 and 10 years of age. Parents were also asked to complete a daily diary for the first five years, recording information on breastfeeding, respiratory symptoms (fever, runny nose, cough, wheeze), and use of any medications by both mother and child. Healthy nasopharyngeal aspirates (NPAs) were collected from subjects by study clinicians during planned visits at approximately 2 months, 6 months and 12 months of age, after the child had been free from any symptoms of respiratory illness for a period of at least 4 weeks. Parents were also asked to report to the study clinicians whenever the child showed symptoms of an acute respiratory illness (ARI), at which point the family was visited within 48 hours by a study nurse who recorded clinical details of the infection and collected an NPA from the child. Clinical data recorded included the presence of fever, wheeze or rattly chest and any medications taken (including antibiotics). Each ARI was classified as either a lower respiratory illness (LRI; if wheeze or a rattly chest was present), or an upper respiratory illness (URI; otherwise). The material in each NPA was divided into four aliquots and stored at -80°C. Blood was collected from each child at age 6 months and 1, 2, 3, 4, 5 and 10 years.

Initial viral screen

One aliquot of each NPA (healthy and ARI) was screened for common causative agents via reverse transcriptase polymerase chain reactions (PCR) as previously reported (Kusel et al., 2006). Target organisms were: human rhinoviruses (HRV); other picornaviruses (coxsackie, echo and enteroviruses); coronaviruses 229E and OC43; respiratory syncytial virus (RSV); influenza A and B; parainfluenzaviruses 1-3; adenoviruses and human metapneumovirus (HMPV). Viral profile data was successfully generated for 451 healthy, 327 LRI and 649 URI NPAs.

Expanded screen for detection and subtyping of HRV

Following the recent identification of HRV-C (Lee et al., 2012), and expansion in availability of HRV sequence information, a second aliquot of each LRI NPA was screened for picornaviruses using seminested PCR (Bochkov et al., 2014), which incorporated a wider diversity of HRV primers than the initial screen, and then partial sequencing was used to classify HRV into A, B and C species and type.

DNA extraction and bacterial 16S rRNA amplicon sequencing

One aliquot each of 1,021 NPAs was used for 16S rRNA microbiome profiling. These include 487/561 healthy NPAs collected from visits at 2 months, 6 months and 12 months of age; 380/381 LRI reported during the same period; and 154/782 URI (random selection of 0-2 URI per infant). Overall, 397 healthy NPAs, 326 LRIs and 101 URIs were profiled for both virus and bacteria.

Total DNA was extracted using a method combining homogenization and chemical lysis of cells. Extractions were performed in biosafety cabinets that were UV-sterilised, including all plastic-ware, for 30 min prior to the procedure. The NPA were

thawed from -80°C storage, transferred into 1.5 mL sterile screw-capped tubes and briefly micro-centrifuged. The saline storage buffer was removed and pellets were resuspended in 400 µL of lysis solution supplied with the Wizard SV Genomic DNA System (Promega, Victoria, Australia). Samples were mixed vigorously by pipetting and then transferred into a labelled Lysing Matrix B tube (MP Biomedicals, New South Wales, Australia). Suspensions were homogenized using a FastPrep-24 homogenizer for 40 s at 6.5 m/s. Following micro-centrifugation, homogenates were transferred into a 1.5 mL screw-capped tube. A further 200 µL of lysis solution was added into each lysing matrix tube and vortexed to wash off any residual homogenate, then transferred to the respective homogenate tube to retain the original lysis volume. Homogenates were then treated with nuclei lysis buffer/RNase A and DNA extraction was carried out using the Wizard SV Genomic DNA System as per manufacturer's instructions. Purified DNA was eluted in 100 µL of pre-warmed sterile low 1 X TE (Fisher Biotec, WA, Australia), aliquoted and stored at -80°C.

Amplicons were prepared for MiSeq sequencing using primers (prepared by Integrated DNA Technologies, Iowa, USA) spanning the V4 region of the 16S rRNA gene and containing barcoded reverse primers as published by Caporaso *et al.* (Caporaso et al., 2012). The forward universal primer included the 5´ Illumina adapter sequence, forward primer pad, linker and the 515F 16S rRNA sequence: 5´-AATGATACGGCGACCACCGAGATCTACACTATGGTAATTGTGTGCCAGC MGCCGCGGTAA-3´. The reverse primer included the 3´ Illumina adapter sequence, a 12-mer Golay barcode (denoted as N), reverse primer pad, linker and the 806R 16S rRNA sequence: 5´-CAAGCAGAAGACGGCATACGAGATNNNNNNNNNNNN AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3´. All laboratory equipment used was wiped with DNA Away (MBP, Mexico) before conducting each PCR procedure. Master mixes were prepared in a UV-treated PCR chamber before they were dispensed into 96-well plates using a Multiprobe II liquid handling system (Perkin Elmer, Victoria, Australia), followed by addition of samples and controls using the robot. Amplification of each sample was performed in quadruplicate to obtain enough amplicon for sequencing. To each plate, a positive control (gDNA from *S. enterica* strain LT2, a bacterium not normally associated with the respiratory system (ATCC#700720D-5, USA)) and water and TE negative controls (obtained from each extraction procedure) were included and assessed for amplification by agarose gel electrophoresis. All controls were as expected: *S. enterica* controls were positive and the water and TE negative controls were negative. Positive control samples were not analysed further, while the negative controls were prepared for sequencing in the same way as NPA samples.

Due to the high throughput nature of this study, we did not quantify and normalize sample DNA that was added into each PCR reaction, rather a fixed volume (4 µL) of DNA template was used per well. Amplification was conducted on a GeneAmp 9700 PCR System (Perkin Elmer) using the following conditions: an initial 94°C denaturation step for 2 min, followed by 30 cycles of 94°C denaturation for 30 s, 58°C annealing for 30 s and 72°C extension for 1 min.

Quadruplicate sample amplicons were combined into a single well on the PCR reaction plate, then transferred to a fresh round-bottom polystyrene plate where they were purified using Agencourt AMPure XP beads as directed by the manufacturer, with slight modifications (Beckman Coulter, USA). Purified amplicons were eluted in

25 µL sterile low 1 X TE buffer (Fisher Biotec). Quantitation of amplicon was performed using the Quant-iT PicoGreen dsDNA quantitation kit (Life Technologies, Victoria, Australia) and fluorescence was determined on a Wallac Victor^3 Multilabel counter (Perkin Elmer). PCR samples were equalized to 2 nM concentration (a neat aliquot was used where a sample fell below this concentration) and pools of 48, 60 or 96 barcoded samples were generated and sent for sequencing.

Primer adaptors were removed from library pools using a 0.8x ratio of Agencourt AMPure XP beads (Beckman Coulter, USA). Library quantitation was determined by the high sensitivity Qubit kit (Life Technologies, USA) whilst library quality and average size distribution was assessed by the Bioanalyser (Agilent Technologies, USA) high sensitivity kit. Library pools were diluted to 2nM followed by NaOH denaturation as per manufacter's instructions (Illumina Inc., USA). Sequencing primers read 1: 5'- TATGGTAATTGTGTGCCAGCMGCCGCGGTAA -3', read 2: 5'-AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3' and index: 5'-ATTAGAWACCCBDGTAGTCCGGCTGACTGACT-3' (Sigma, Australia) were spiked into the MiSeq Cartridge at a final concentration of 0.5µM. Denatured libraries were loaded at 6.5pM with a 5% PhiX spike for diversity and sequencing control, onto a v2 300 cycle cartridge for sequencing on the Illumina MiSeq. All sequencing runs yielded approximately 1100k clusters/mm^2 with an average of 80% clusters passing filter and an average 19M reads passing filter.

Statistical Analysis

Classification of NPAs. Unless otherwise stated, the following definitions were used:

- Infection: Collected during a physician verified episode of symptoms of acute respiratory illness (ARI).
- Healthy: Collected in the absence of such symptoms for at least 4 weeks.
- LRI: Lower respiratory illness, defined as ARI with wheeze or rattly chest.
- URI: Upper respiratory illness, defined as ARI with no wheeze or rattly chest.
- 2-month samples: Age at collection 6-100 days (range 6-100 days for infections, 34-97 days for healthy).
- 6-month samples: Age at collection 150-210 days (interquartile range 165-196 days for infections, 179-190 days for healthy).
- 1-year samples: Age at collection 330-425 days (interquartile range 357-398 days for infections, 363-382 days for healthy).

Variable definitions. Unless otherwise stated, the following definitions were used in the statistical analyses:

- Gender (binary): Female (coded 0), Male (coded 1).
- Season (binary): Spring-summer (September to February, coded 1); autumn-winter (March to August, coded 0).
- Antibiotics intake (binary): Coded 1 if the child had taken antibiotics within the last 4 weeks before sample collection; 0 otherwise. Antibiotics data was curated from two sources of information; a daily diary kept by the mother, which reports any medication taken per day, and a questionnaire taken during each infection sampling. Positive antibiotics intake from any one source was sufficient. We considered only oral/intravenous antibiotics and excluded Flagyl (Metronidazole) and any topical antibiotics.

- Mother's antibiotics intake (binary): Coded 1 if the mother had taken antibiotics within the last 4 weeks before sample collection and she was breastfeeding on the day of antibiotics intake; 0 otherwise. Data was extracted from the daily diary, using the same group of antibiotics as for subjects.
- Breastfeeding (binary): Coded 1 if the child was breastfed on the day of sample collection; 0 otherwise.
- Siblings (binary): Coded 1 if any children less than 16 years of age were residing in the same house; 0 otherwise. Data was extracted from the 1-year questionnaire.
- Furry pets (binary): Coded 1 if household had any cat, dog, rabbit or guinea pig for the whole first year; 0 otherwise. Data was extracted from the 1-year questionnaire.
- Day-care (binary): Coded 1 if child has started attending day-care by sample collection; 0 otherwise.
- Delivery mode (binary): Coded 1 caesarean section delivery; 0 otherwise.
- Maternal/paternal history of atopic disease: Coded 1 if the mother/father had doctor diagnosed allergic hayfever, eczema or asthma.
- Atopy by 2 years (binary): Atopic status was assessed using serum IgE levels measured at 6 months, 1 year and 2 years. Positive atopy status was defined as any specific IgE to house dust mite, cat epithelium and dander, peanut, foodmix, couch grass, rye grass, mould mix, or infant phadiatop > 0.35 kU/L (*9*). Atopy by two years was defined as IgE above this cut-off at any of the three time points (6 months, 1 year, 2 years). Children atopic by 2 years are referred to as atopics; non-atopics otherwise.
- Chronic wheeze (binary): Presence of wheeze in the last 12 months, assessed during face-to-face interviews with the parents of subjects at 5 and 10 years of age.

Association of MPGs with healthy vs infection status. Odds ratios for each MPG with infection status (NPA taken during reported episode of respiratory illness, vs. NPA taken in the absence of such symptoms for at least 4 weeks prior) were estimated using generalized estimating equations (GEE) logistic regression with unstructured correlation and robust standard errors, to take into account samples from the same subject. The following variables were adjusted for by inclusion in the model: age at sample collection (days), gender, season, number of prior infections, antibiotics intake, mother's antibiotics intake, delivery mode and breastfeeding. ORs further adjusted for detection of common viruses (RSV, HRV) by PCR were also reported (**Figure 1E**).

Association between microbes and classes of ARI. Amongst ARI samples, odds ratios for MPGs and markers of infection severity (LRI vs. URI, febrile vs. non-febrile LRI, and wheezy vs. non-wheezy LRI) were estimated using generalized estimating equations (GEE) logistic regression with unstructured correlation and robust standard errors. Models were adjusted for presence of the most common viruses (RSV and HRV), age, gender, and season. Samples with antibiotics intake within the last week were excluded from this analysis (**Table 1**). Febrile vs. non-febrile LRI comparison was also done separately for RSV-positive and RSV-negative subsets using logistic regression and adjusting for age, gender and season (**Figure S5**).

Association of MPGs with otitis media symptoms. For each MPG, odds ratios for otitis media symptoms within (a) ARIs, and (b) LRIs were estimated using GEE logistic regression, adjusting for age at sample collection (**Table S2**).

Stability of MPGs within individuals. NP microbiome transitions were assessed by analysing the MPG assignments of consecutive samples within individual subjects (time point T_1 to next time point T_2). Transition into the same MPG (T_1 MPG = T_2 MPG) was considered a stable transition. Healthy and infection samples were analysed separately. The expected frequency of a stable transition for a given MPG, *i*, was calculated as the squared proportion of samples in that MPG at T_1, $(p_i(T_1))^2$. One thousand bootstrapped estimates of observed stability were generated by sampling with replacement from the observed transitions and calculating, for each bootstrapped sample, the proportion of stable transitions for each MPG. The estimated 95% confidence intervals for the observed frequency of respective stable transitions were taken to be the 2.5th and 97.5th quantiles of the 1,000 bootstrapped statistics. The MPG was considered significantly more stable than expected if the expected frequency of stable transitions $(p_i(T_1))^2$ fell below the lower limits of the 95% CI of observed transitions (and less stable if the expected value fell above the upper limits) (**Figure S3**).

Association of infection and wheeze phenotypes with Alloiococcus/Moraxella stable colonization groups. Children were grouped by their colonization patterns: '*Alloiococcus* colonized', ≥1 healthy sample with *Alloiococcus* MPG and none with *Moraxella* MPG; '*Moraxella* colonized', ≥1 healthy sample with *Moraxella* MPG and none with *Alloiococcus* MPG; 'Others', no *Moraxella* or *Alloiococcus* healthy sample or ≥1 of each. Association of the groups with number of infections was assessed using Wilcoxon rank sum test; other associations were assessed using Fisher's exact test (**Table S4**).

Association of genus abundance with environmental factors. We investigated the relationships between relative abundance of the six main genera and various environmental factors: attendance at day-care, living with siblings, antibiotics intake within the last 4 weeks, breastfeeding, number of prior infections (0, 1, ≥2; based on recorded ARI) and gender. As relative abundance values were highly skewed, they were log-transformed (base 10) prior to analysis. Odds ratios (ORs) for binary variables were estimated using logistic regression; odds ratios for association with prior infections were assessed using proportional odds ordinal logistic regression. Healthy and infection samples were analysed separately. Association with day-care was assessed for 12-month samples only, as few infants had started day-care by the time of their 6-month NPA; ORs for siblings, furry pets, antibiotics, season, and gender were adjusted for age at collection by inclusion in the model. Association with antibiotics was assessed for healthy samples only, as prescription due to infection severity could confound association with infection samples. Prior infections were assessed separately for each of the three age strata for healthy NPA sampling (2, 6 and 12 months) (**Figure 2**). Temperature data for Perth, Western Australia was obtained from the Australian Government Bureau of Meteorology website (http://www.bom.gov.au/climate/averages/tables/cw_009225.shtml).

Association between early Streptococcus colonization and subsequent chronic wheeze. Odds ratios were calculated separately for atopics, non-atopics and all children, using logistic regression. Gender, age at sample collection, maternal history of atopic disease and paternal history of atopic disease were adjusted for by inclusion in the model. Samples with any prior infection or antibiotics intake within the last week were excluded from this analysis (**Table 2, Figure 5B**).

Association between early colonization and time of first ARI. This analysis was restricted to children who contributed an asymptomatic NP sample between 5-9 weeks of age and prior to their first reported ARI (n=160). Children were grouped according to the MPG of the first pre-ARI healthy sample. Kaplan-Meier curves were plotted separately for each group, showing time of first infection (age in days) as recorded in the daily symptom diaries for (a) first ARI, (b) first URI and (c) first LRI. Cox proportional hazards models were fit to assess the significance of apparent differences in time of infection, adjusting for age, gender, season, virus status in the early healthy sample, and virus status in the subsequent infection (**Figure 4**).

Association between LRI subtypes and subsequent chronic wheeze. Odds ratios were calculated separately for atopics, non-atopics and all children, using logistic regression. Gender, maternal history of atopic disease and paternal history of atopic disease were adjusted for by including them as covariates in the model (**Table 2, Table S3**).

Association between time of first LRI subtype and subsequent chronic wheeze. Children were grouped according to their chronic wheeze status at 5 years. Kaplan-Meier curves were plotted separately for each group, showing time of first infection (age in days) for (a) first fever LRI, (b) first wheezy LRI with detection of HRV-C. Cox proportional hazards models were fit to assess the significance of apparent differences in time of infection, adjusting for gender, maternal history of atopic disease and paternal history of atopic disease (**Figure 3A-B**).

Association between antibodies to species-specific surface proteins and number of NP samples containing specific OTUs. Logistic regression was used to estimate odds ratios for association of detectable IgG1/IgG4 antibodies (assessed from blood during collection at 12 months) to *H. influenza* P4/P6 surface proteins with (a) number of infection NP samples and (b) number of healthy NP samples with presence of the two most abundant *Haemophilus* OTUs (**Figure S2A**). The same was done for *S. pneumoniae* A1, A2 or C surface proteins and the two most abundant *Streptococcus* OTUs (**Figure S2B**).

Supplemental References

Bochkov, Y.A., Grindle, K., Vang, F., Evans, M.D., and Gern, J.E. (2014). Improved molecular typing assay for rhinovirus species A, B, and C. Journal of clinical microbiology *52*, 2461-2471.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., *et al.* (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J *6*, 1621-1624.