

The role of the microbiota in respiratory virus-bacterial pathobiont relationships in the upper respiratory tract

Matthew S. Kelly, MD, MPH,^{1,2,3} Pixu Shi, PhD,⁴ Sefelani C. Boiditswe, BNSc,¹ Emily Qin, BS,⁴ Andrew P. Steenhoff, MBBCh,^{1,5,6} Tiny Mazhani, MD,⁷ Mohamed Z. Patel, MBBCh,⁷ Coleen K. Cunningham, MD,⁸ John F. Rawls, PhD,³ Kathy Luinstra, HBSc,⁹ Jodi Gilchrist, MSc,⁹ Julia Maciejewski, MSc,⁹ Jillian H. Hurst, PhD,² Patrick C. Seed, MD, PhD,¹¹ David Bulir, MD, PhD,¹² Marek Smieja, MD, PhD^{9,10}

¹Botswana-University of Pennsylvania Partnership, Gaborone, Botswana; ²Division of Pediatric Infectious Diseases, Duke University, Durham, North Carolina, United States; ³Department of Molecular Genetics and Microbiology, Duke University, Durham, North Carolina, United States; ⁴Department of Biostatistics and Bioinformatics, Duke University, Durham, North Carolina, United States; ⁵Global Health Center, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States; ⁶Division of Pediatric Infectious Diseases, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States; ⁷University of Botswana School of Medicine, Gaborone, Botswana; ⁸Division of Pediatric Infectious Diseases, University of California, Irvine, Children's Hospital of Orange County, Orange, California, United States; ⁹Infectious Disease Research Group, Research Institute of St. Joe's Hamilton, Hamilton, Ontario, Canada; ¹⁰Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada; ¹¹Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA; ¹²Department of Chemical Engineering, McMaster University, Hamilton, Ontario, Canada

Keywords: *Haemophilus influenzae*, *Moraxella catarrhalis*, *Staphylococcus aureus*, *Streptococcus pneumoniae* carriage, children, respiratory microbiome, nasopharyngeal microbiota, childhood respiratory infections, sub-Saharan Africa

Corresponding author:

Matthew S. Kelly, MD, MPH (matthew.kelly@duke.edu)

MSRB3, Room 4128

3 Genome Court

Durham, NC 27278 USA

ABSTRACT

The mechanisms by which respiratory viruses predispose to secondary bacterial infections remain poorly characterized. Using 2,409 nasopharyngeal swabs from 300 infants in Botswana, we performed a detailed analysis of factors that influence the dynamics of bacterial pathobiont colonization during infancy. We quantify the extent to which viruses increase the acquisition of *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Streptococcus pneumoniae*. We provide evidence of cooperative interactions between these pathobionts while identifying host characteristics and environmental exposures that influence the odds of pathobiont colonization during early life. Using 16S rRNA gene sequencing, we demonstrate that respiratory viruses result in losses of putatively beneficial *Corynebacterium* and *Streptococcus* species that are associated with a lower odds of pathobiont acquisition. These findings provide novel insights into viral-bacterial relationships in the URT of direct relevance to respiratory infections and suggest that the URT bacterial microbiota is a potentially modifiable mechanism by which viruses promote bacterial respiratory infections.

INTRODUCTION

Acute respiratory infections (ARIs) are the most common infections across the lifespan and are associated with substantial morbidity and mortality, particularly during childhood. ARIs accounted for 2.2 million deaths in 2021, including more than 500,000 deaths among children less than 5 years of age.¹ Although respiratory viruses are detected in the majority of ARIs among children, most severe or fatal infections are caused by bacteria, including *Streptococcus pneumoniae* (pneumococcus), *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Staphylococcus aureus*.²⁻⁴ These bacteria are classified as “pathobionts” because they frequently colonize the nasal passages without causing clinical illness but can cause infection when the local microenvironment is altered, as may occur with respiratory virus infection. Recognition of the relationship between respiratory virus infections and ARIs caused by these bacterial pathobionts dates back to the 1918-1919 influenza pandemic, during which the majority of influenza-associated deaths were attributable to secondary bacterial pneumonia.^{5,6} Recent clinical studies have reported associations between respiratory virus infections and bacterial pneumonia caused by *S. aureus*^{7,8} and *S. pneumoniae*,^{9,10} while respiratory viruses have been shown in animal models to predispose to *H. influenzae* pneumonia¹¹ and acute otitis media caused by *M. catarrhalis*.¹² Moreover, during the COVID-19 pandemic, substantial reductions in invasive infections caused by *H. influenzae*¹³ and *S. pneumoniae*¹³⁻¹⁵ correlated with the reduced circulation of respiratory viruses other than severe acute respiratory virus coronavirus 2 (SARS-CoV-2), often in the context of stable prevalence of colonization by these pathobionts. However, despite these epidemiological studies, the precise impacts of respiratory virus infections on the colonization dynamics of bacterial respiratory pathobionts remain poorly characterized.

To colonize the upper respiratory tract (URT), bacterial pathobionts must evade myriad host defenses, including mechanical barriers, immune cells, antibodies, and host antimicrobial peptides. Additionally, these bacteria must co-colonize within the microbial communities that inhabit the URT mucosa. This

URT microbiota is increasingly recognized to resist pathobiont colonization and invasion through competition for resources, production of antimicrobial proteins and secondary metabolites, and modulation of host immune responses.¹⁶⁻¹⁹ Most prior studies of this resistance to pathobiont colonization in the URT were cross-sectional, used culture-based methods, and focused on relationships between bacterial pathobionts. These studies have generally reported that cocultivation of *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* is observed more frequently than would be expected based on the prevalence of each species,²⁰⁻²⁵ while a negative association between URT colonization by *S. aureus* and *S. pneumoniae* has been observed in many,²¹⁻²⁴ but not all,^{25,26} pediatric studies. Few studies investigated interactions between these bacterial pathobionts using a longitudinal study design and molecular detection methods. Moreover, surprisingly little is known regarding the extent to which other bacterial species in the URT microbiota influence colonization by these pathobionts. An improved understanding of these microbial relationships could lead to novel strategies to prevent ARIs by modifying the URT microbiota.

In this study, we investigated relationships between respiratory viruses, the URT bacterial microbiota, and bacterial respiratory pathobionts in 300 infants in Botswana who were followed during the first year of life. We characterize the impact of respiratory virus infections on the acquisition of bacterial pathobionts and identify cooperative relationships between *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* within the human URT. We demonstrate that the composition of the URT microbiota predicts the acquisition of bacterial pathobionts during infancy and provide evidence that respiratory viruses predispose to pathobiont acquisition, at least in part through modification of the URT microbiota. Finally, we identify potential relationships between bacterial pathobionts and other URT microbes that could be leveraged to develop novel biotherapeutics for ARI prevention.

RESULTS

Respiratory virus infections and bacterial pathobiont colonization occur frequently during infancy

As previously described, we collected nasopharyngeal swab samples (n=2,409) monthly (infant age 0-6 months) or bimonthly (infant age >6-12 months) from 300 mother-infant dyads recruited at urban and rural public clinics in southern Botswana.¹⁶ We tested infant nasopharyngeal samples for respiratory viruses and common bacterial respiratory pathobionts by PCR. We determined the composition of the bacterial microbiota by sequencing the V4 region of the 16S ribosomal RNA (rRNA) gene (**Figure 1a**). Infants were born vaginally, had a median [interquartile range (IQR)] birth weight of 3085g (2850g, 3395g), and were predominantly breastfed (**Table 1**). We identified one or more respiratory viruses in 639 of 2372 (27%) tested samples, with rhinovirus/enterovirus (sample prevalence of 21%), adenovirus (3%), and respiratory syncytial virus (RSV; 2%) being the most frequently detected viruses (**Supplementary Table 1**). The prevalence of respiratory virus detection increased with age during infancy, approaching 40% at 12 months of age (**Figure 1b**; logistic regression, $p < 0.0001$). The prevalence of colonization by bacterial respiratory pathobionts also varied during infancy (**Figure 1c**). Colonization by *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* increased progressively during infancy (logistic regression, $p < 0.0001$ for all comparisons). In contrast, the prevalence of *S. aureus* colonization declined with age (logistic regression, $p < 0.0001$) after peaking at one month of age.

ARI symptoms, identified by the presence of cough, rhinorrhea or nasal congestion, were reported at 610 of 2,409 (25%) study visits and were strongly associated with respiratory virus detection [mixed effect logistic regression; odds ratio (OR): 3.61, 95% confidence interval (CI): 2.89-4.51, $p < 0.0001$]. However, the prevalence of ARI symptoms varied based on the respiratory virus detected. ARI symptoms were reported at more than 70% of visits at which influenza viruses, human metapneumovirus, parainfluenza virus type 3, or RSV were detected, but at only 35% and 41% of visits with detection of adenovirus or rhinovirus/enterovirus, respectively (**Supplementary Table 1**). Acquisition of *H. influenzae* (mixed effect logistic regression; OR: 1.74, 95% CI: 1.31-2.33, $p = 0.0002$), *M. catarrhalis* (OR: 1.55, 95% CI: 1.18-

2.03, $p=0.002$), and *S. pneumoniae* (OR: 1.56, 95% CI: 1.18-2.07, $p=0.002$) were each associated with ARI symptoms after accounting for respiratory virus detection and the acquisition of other pathobionts, as was a higher pneumococcal carriage density among infants newly colonized with *S. pneumoniae* (per \log_{10} increase in copies/mL; OR: 1.19, 95% CI: 1.07-1.32, $p=0.002$). Infants acquiring *H. influenzae* (20%), *M. catarrhalis* (16%), or *S. pneumoniae* (15%) were only slightly more likely to receive antibiotics than infants who did not acquire one of these pathobionts (14%), suggesting that URT colonization by these bacterial respiratory pathobionts during infancy is often associated with ARI symptoms in the absence of established infection.

Respiratory viruses and environmental exposures associated with bacterial pathobiont colonization

We then evaluated the impact of coincident respiratory virus infections and other clinical factors on the acquisition of bacterial pathobionts. We found that respiratory virus detection was associated with higher odds of acquisition of *H. influenzae* (**Figure 2a, Table 2**; mixed effect logistic regression; OR: 1.44, 95% CI: 1.06-1.96, $p=0.02$), *M. catarrhalis* (OR: 1.33, 95% CI: 0.96-1.87, $p=0.09$), and *S. pneumoniae* (OR: 1.83, 95% CI 1.33-2.52, $p=0.0002$), and a lower odds of *S. aureus* acquisition (OR: 0.45, 95% CI: 0.29-0.70, $p=0.0004$). Acquisition of *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* tended to increase in association with infections involving rhinovirus/enterovirus only ($n=468$) and infections involving other respiratory viruses ($n=171$), although the odds of acquisition of *H. influenzae* and *M. catarrhalis* were higher with infections caused by other respiratory viruses (**Supplementary Table 2**). Among infants with pneumococcal carriage, respiratory virus detection was associated with an approximately four-fold increase in the *S. pneumoniae* colonization density [**Figure 2b**; median (IQR): 7.4 (6.6-7.8) vs. 7.0 (6.0-7.5) \log_{10} copies/mL; linear regression, $p=0.006$]. The density of pneumococcal colonization did not differ between infants with rhinovirus/enterovirus only compared to infants with other respiratory virus infections [median (IQR): 7.4 (6.6-7.8) vs. 7.3 (6.6-7.8) \log_{10} copies/mL; $p=0.73$].

Several infant and sample characteristics were also associated with the odds of acquisition of specific bacterial pathobionts (**Table 2**). Each additional child less than 5 years of age in the household was associated with higher odds of acquisition of *H. influenzae* (mixed effect logistic regression; OR: 1.21, 95% CI: 1.05-1.41, p=0.01), *M. catarrhalis* (OR: 1.14, 95% CI: 0.99-1.32, p=0.06), and *S. pneumoniae* (OR: 1.26, 95% CI: 1.09-1.46, p=0.002). Breastfeeding was associated with a lower odds of *S. aureus* acquisition (OR: 0.49, 95% CI: 0.27-0.87, p=0.01), while the odds of *S. aureus* acquisition was higher during the rainy season than during the dry season (OR: 1.58, 95% CI: 1.11-2.26, p=0.01). In contrast, the odds of *M. catarrhalis* acquisition was lower during the rainy season (OR: 0.55, 95% CI: 0.41-0.75, p=0.0002), but higher among infants living in households that use solid fuels for cooking or heating (OR: 1.47, 95% CI: 1.02-2.10, p=0.04). The odds of *S. pneumoniae* acquisition was higher in urban compared to rural settings (OR: 1.56, 95% CI: 1.09-2.23, p=0.01), but there was no association between receipt of 13-valent pneumococcal conjugate vaccine (PCV-13) doses and *S. pneumoniae* acquisition (OR: 1.15, 95% CI: 0.91-1.44, p=0.24). Among infants with pneumococcal carriage, a higher colonization density was observed with younger age (linear regression; p=0.007), a greater number of children in the household (p=0.002), and receipt of antibiotics since the prior study visit (p=0.004). No association was observed between PCV-13 vaccination status and the density of pneumococcal colonization [median (IQR) log₁₀ copies/mL; 0 doses: 7.5 (6.7-7.9); 1 dose: 7.3 (6.6-7.6); 2 doses: 7.1 (6.1-7.7); 3 doses: 7.0 (6.0-7.7); p=0.14]. Thus, in addition to confirming the substantial role of respiratory virus infections, our findings suggest that feeding practices, season, location of residence, and exposure to other children and smoke from solid fuels influence the acquisition of bacterial pathobionts during infancy.

Patterns of pathobiont colonization during infancy support cooperative interspecies relationships

We next sought to evaluate for interactive relationships between bacterial respiratory pathobionts within the infant URT. As reported by previous cross-sectional studies,²⁰⁻²³ we observed positive correlations for the co-detection of *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* within samples (**Figure 2c**; Fisher's

exact tests; *H. influenzae*-*M. catarrhalis*, OR: 4.46, 95% CI: 3.60-5.54, $p < 0.0001$; *M. catarrhalis*-*S. pneumoniae*, OR: 8.19, 95% CI: 6.77-9.92, $p < 0.0001$; *H. influenzae*-*S. pneumoniae*, OR: 5.67, 95% CI: 4.60-7.02, $p < 0.0001$), as well as negative correlations for the co-detection of *S. aureus* and other pathobionts (*S. aureus*-*H. influenzae*, OR: 0.44, 95% CI: 0.33-0.59, $p < 0.0001$; *S. aureus*-*M. catarrhalis*, OR: 0.33, 95% CI: 0.26-0.42, $p < 0.0001$; *S. aureus*-*S. pneumoniae*, OR: 0.46, 95% CI: 0.36-0.58, $p < 0.0001$). We then performed longitudinal analyses to evaluate if colonization by each bacterial pathobiont was associated with the subsequent acquisition of other bacterial pathobionts (**Figure 2a**, **Table 2**). We found that *M. catarrhalis* colonization was associated with higher odds of acquisition of *H. influenzae* (mixed effect logistic regression; OR: 1.82, 95% CI: 1.28-2.60, $p = 0.009$) and *S. pneumoniae* (OR: 1.58, 95% CI: 1.12-2.24, $p = 0.009$), and *S. pneumoniae* colonization was associated with a higher odds of *H. influenzae* acquisition (OR: 1.42, 95% CI: 1.01-1.98, $p = 0.04$). Additionally, among infants with pneumococcal carriage, concurrent *H. influenzae* colonization was associated with an approximately three-fold higher *S. pneumoniae* colonization density [**Figure 2b**; median (IQR): 7.3 (6.5-7.7) vs. 7.0 (5.9-7.6) \log_{10} copies/mL; linear regression; $p < 0.0001$]. The odds of *S. aureus* acquisition tended to be lower with preceding *S. pneumoniae* colonization (mixed effect logistic regression; OR: 0.66, 95% CI: 0.41-1.06, $p = 0.09$), but preceding *S. aureus* colonization was not associated with the acquisition of other pathobionts (**Figure 2a**). These findings provide further supportive evidence of potential cooperative interactions between *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* within the human URT. Moreover, the lack of associations between *S. aureus* and other pathobionts in longitudinal analyses suggests that the negative co-occurrence patterns observed between *S. aureus* and other bacterial pathobionts are unlikely to reflect direct antagonistic interspecies interactions.

Respiratory virus infections trigger shifts in the URT microbiota that promote pathobiont acquisition

Given the observed associations between respiratory virus infections and the acquisition of bacterial respiratory pathogens during infancy, we sought to evaluate the extent to which alterations of the URT bacterial microbiota may contribute to these viral-bacterial relationships. In order to investigate shifts in overall URT microbiota community structure, we used k-medoids clustering on Bray-Curtis distances to classify infant nasopharyngeal samples into eight distinct microbiota types (**Figure 3a,b**). URT microbiota types of similar composition were identified using k-means clustering based on Euclidean distances (**Supplementary Figure 1**), suggesting that the types identified by k-medoids clustering are robust to both distance metric and clustering algorithm. Excluding samples collected during the birth visit and with subject included as a random effect, the prevalence of microbiota types dominated by *Staphylococcus* (mixed effect logistic regression; STA, $p < 0.0001$) and *Corynebacterium* (COR, $p < 0.0001$) declined with age, while the prevalence of microbiota types dominated by *Haemophilus* (HAE, $p < 0.0001$), *Moraxella* (MOR, $p < 0.0001$), or *Corynebacterium/Dolosigranulum/Moraxella* (CDM, $p < 0.0001$) increased during infancy (**Figure 3c**). Compared to intervals during which a respiratory virus was not identified, intervals with respiratory virus infections were more likely to be associated with a shift in URT microbiota type (mixed effect logistic regression, $p = 0.046$). Specifically, respiratory virus infections were associated with declining prevalence of microbiota types dominated by *Staphylococcus* (STA, $p = 0.0002$), *Corynebacterium* (COR, $p = 0.03$), or *Corynebacterium/Dolosigranulum* (CD, $p = 0.007$), and increasing prevalence of microbiota types dominated by *Haemophilus* (HAE, $p < 0.0001$), *Moraxella* (MOR, $p = 0.10$), or *Streptococcus* (STR, $p < 0.0001$) (**Figure 3d**). Such transitions in URT microbiota community states correlated with the subsequent odds of acquiring bacterial pathogens. Compared to a *Corynebacterium*-dominant (COR) microbiota type, a *Haemophilus*-dominant (HAE) microbiota type was associated with a higher odds of acquisition of *M. catarrhalis* (mixed effect logistic regression; OR: 2.89, 95% CI: 1.29-6.45, $p = 0.01$) and *S. pneumoniae* (OR: 3.41, 95% CI: 1.45-8.00, $p = 0.005$), a *Moraxella*-dominant (MOR) microbiota type was associated with the acquisition of *H. influenzae* (OR: 2.58, 95% CI: 1.40-4.76, $p = 0.002$) and *S. pneumoniae* (OR: 3.10, 95% CI: 1.63-5.87, $p = 0.0005$), and a

Streptococcus-dominant (STR) microbiota type was associated with the acquisition of *H. influenzae* (OR: 2.26, 95% CI: 1.19-4.28, $p=0.01$). Acquisition of bacterial pathobionts was associated with shifts in URT microbiota composition towards microbiota types dominated by the genus containing that pathobiont (**Supplementary Figure 2**).

Next, we used MaAsLin2 to fit general linear mixed effect models to evaluate the impact of respiratory virus infection on the relative abundances of specific amplicon sequence variants (ASV) within the infant URT microbiota. For these and other analyses conducted at the ASV level, we aggregated all ASVs that did not have a relative abundance of at least 0.1% in 5% or more of samples into a single “other” category. Respiratory virus infection was associated with decreases in the relative abundances of an ASV classified as *Streptococcus thermophilus/vestibularis/salivarius* (ASV15, $q=0.18$) and several ASVs classified as *Corynebacterium* species (**Figure 3e**), including *C. propinquum/pseudodiphtheriticum* (ASV4, $q=0.19$; ASV18, $q=0.15$), *C. accolens/macginleyi* (ASV6, $q=0.10$), and *C. tuberculostearicum* (ASV86, $q=0.16$). Most other changes in the relative abundances of specific ASVs associated with respiratory virus infection are consistent with trends in pathobiont colonization that we previously identified by PCR, with increases in the abundances of *Haemophilus* sp. (ASV7, $q<0.0001$; ASV8, $p=0.0008$), *M. catarrhalis/nonliquefaciens* (ASV1, $q=0.10$), and *S. pneumoniae/pseudopneumoniae* (ASV5, $q<0.0001$), and a decrease in the abundance of *Staphylococcus* sp. (ASV3, $q<0.0001$) (**Figure 3e**). Each of these ASVs was more abundant (Wilcoxon rank-sum tests, $p<0.0001$ for all comparisons) in children colonized with the pathobiont from that genus than in noncolonized children, indicating that these ASVs are highly likely to contain the pathobiont belonging to that genus (**Supplementary Figure 3**).

The URT microbiota influences the acquisition of bacterial pathobionts during infancy

Finally, we assessed the extent to which the composition of the URT microbiota influenced the acquisition of pathobionts among infants using both discriminative (random forests) and generative (mixed-effect logistic regression) methods. We first constructed random forest models to determine the utility of the preceding URT microbiota in predicting the acquisition of bacterial respiratory pathobionts at the next study visit. We compared the accuracy of these models to those incorporating infant and sample characteristics such as age and respiratory virus infection. We found that models based on age alone predicted the acquisition of all bacterial pathobionts better than expected by chance (**Figure 4a**). However, the accuracy of these models varied by pathobiont, with the highest accuracy achieved for prediction of *S. aureus* acquisition [area under the receiver operating characteristic curve (AUC-ROC): 0.66, 95% CI: 0.61-0.71]. Adding other infant and sample characteristics improved the predictive accuracy of all models. Models incorporating only the taxonomic composition of the URT microbiota from the preceding sample performed as well as or better than models incorporating infant and sample characteristics for prediction of acquisition of *H. influenzae*, *S. aureus*, and *S. pneumoniae* (**Figure 4a**), but not *M. catarrhalis*. The accuracy of these taxonomic models was not substantively improved by the inclusion of infant and sample characteristics or data on the abundances of functional pathways within the URT microbiota as predicted by PICRUS2.²⁷ In models incorporating clinical variables and URT microbiota taxonomic composition, microbial taxa were the variables most predictive of pathobiont acquisition in all models, while age and the number of children in the household were the most discriminative clinical variables (**Figure 4b**).

Finally, we fit mixed-effect logistic regression models to identify specific URT microbiota features associated with the subsequent acquisition of bacterial pathobionts. As opposed to random forests, this approach enabled us to account for repeated sampling from the same subject and adjust for a non-linear effect of age at the time of sampling. These analyses identified several microbial taxa associated with pathobiont acquisition (**Figure 4c**). First, consistent with analyses based on the detection of pathobionts

by PCR, higher relative abundances of *M. catarrhalis/nonliquefaciens* (ASV1) were associated with higher odds of *H. influenzae* ($q=0.0002$) and *S. pneumoniae* ($q=0.006$) acquisition. Similarly, *S. pneumoniae/pseudopneumoniae* (ASV5) was positively associated with the acquisition of *H. influenzae* ($q<0.0001$), while higher abundances of *Haemophilus* sp. (ASV8) were associated with *M. catarrhalis* acquisition ($q=0.05$). Additionally, we identified several ASVs classified as commensal bacterial species that were negatively associated with bacterial pathobiont acquisition (**Figure 4c**), including ASVs that declined in abundance in association with respiratory virus infection. Specifically, higher relative abundances of *C. propinquum/pseudodiphtheriticum* (ASV4) were associated with lower odds of *S. aureus* acquisition ($q=0.03$), and *C. accolens/macginleyi* (ASV6) was negatively associated with *M. catarrhalis* acquisition ($q=0.05$). Both *C. tuberculostearicum* (ASV10, $q=0.02$) and *C. amycolatum* (ASV53, $q=0.06$) were negatively associated with *S. pneumoniae* acquisition. Higher relative abundances of *Streptococcus thermophilus/vestibular/salivarius* (ASV15) were associated with lower odds of *H. influenzae* acquisition ($q=0.10$), while *Streptococcus mitis/oralis* (ASV9) was negatively associated with the acquisition of *H. influenzae* ($q=0.03$), *M. catarrhalis* ($q=0.14$), and *S. pneumoniae* ($q=0.009$). Finally, higher abundances of *D. pigrum* (ASV2) were associated with lower odds of *S. aureus* acquisition ($q=0.03$), and higher abundances of *Veillonella massiliensis* (ASV50) were associated with lower odds of *S. pneumoniae* acquisition ($q=0.02$). Taken together, these data suggest that respiratory viruses promote bacterial pathobiont acquisition through both the cooperative relationships that exist between *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* and detrimental effects on commensal microbes that contribute to colonization resistance to these respiratory pathobionts.

DISCUSSION

In this study of 300 infants in Botswana, we describe the impact of host and environmental factors, respiratory virus infection, and the URT microbiota on the dynamics of bacterial respiratory pathobiont

colonization during the first year of life. Our findings characterize in detail the impact of respiratory virus infections on the acquisition of bacterial pathobionts and provide further supportive evidence of cooperative relationships between *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* within the human URT. We also demonstrate that shifts in URT microbiota composition occurring in association with respiratory virus infection promote pathobiont acquisition. Finally, we identify potential interspecies relationships that, if confirmed through future mechanistic studies, could lead to an improved understanding of microbial ecology in the URT with direct relevance for the prevention of ARIs among both children and adults.

The risk of infections caused by bacterial respiratory pathobionts peaks at the extremes of age. The highest incidences of pneumonia¹ and invasive infections caused by *H. influenzae*²⁸ and *S. pneumoniae*²⁹ occur among young children and older adults, while most cases of acute otitis media occur during early childhood.^{30,31} URT colonization by bacterial pathobionts is a necessary prerequisite both for the development of ARIs caused by these pathobionts and for person-to-person transmission.³² Thus, studies evaluating pathobiont colonization can provide important insights into their population burden and identify modifiable factors that influence the risks of pathobiont acquisition and infection. In this study, we collected serial URT samples from a large cohort of infants in a setting in which risk factors for ARI are highly prevalent.^{33,34} We found that respiratory virus infections increased the odds of acquisition of *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* by 44%, 34%, and 83%, respectively. Further, respiratory virus infections were additionally associated with higher colonization densities among infants with *S. pneumoniae* carriage. Children with pneumonia have higher URT pneumococcal colonization densities than healthy children with nasopharyngeal carriage,^{35,36} suggesting that overgrowth of *S. pneumoniae* in the URT may facilitate invasion of the lower respiratory tract. Moreover, higher URT densities were shown to promote *S. pneumoniae* transmission in a murine model.³⁷ In contrast, we found that respiratory virus infections were actually associated with lower odds of *S. aureus* acquisition. Prior studies suggest

that the risk of secondary *S. aureus* pneumonia is likely to vary based on the inciting virus. Influenza viruses^{7,8} and SARS-CoV-2^{38,39} are the viruses that have been most strongly associated with this complication, while rhinoviruses/enteroviruses accounted for the vast majority of the viral infections identified in our cohort.

Using longitudinal analyses and molecular methods for detecting bacterial pathobionts, we found evidence of cooperative relationships between *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae*. These findings align with those of prior studies, most of which were cross-sectional and used culture-based methods,²⁰⁻²⁵ and suggest that these pathobionts may act in concert in the human URT to promote colonization and infection. Several potential mechanisms for cooperation between these species have previously been identified. For example, these pathobionts are known to form multi-species biofilms in several chronic respiratory diseases, including in the middle ears of children with chronic otitis media.^{40,41} Moreover, the production of virulence factors by these pathobionts may be altered in these biofilms. For example, Cope and colleagues found that *S. pneumoniae* promoted transcription of genes corresponding to type IV pili by *H. influenzae* when these species were co-cultured, while pyruvate oxidase gene expression was upregulated by *S. pneumoniae*.⁴² Interestingly, Schaar and colleagues demonstrated that *M. catarrhalis* strains harboring beta-lactamases can release outer membrane vesicles containing these enzymes that prevent antibiotic-induced killing of susceptible strains of *H. influenzae* and *S. pneumoniae*.⁴³ Consistent with several previous studies of young children,²¹⁻²⁴ we identified a negative co-occurrence pattern between *S. aureus* and *S. pneumoniae* in cross-sectional analyses. However, we found no clear evidence of antagonism between these species in longitudinal analyses, suggesting that the negative co-occurrence of *S. aureus* and *S. pneumoniae* reported in studies of young children is unlikely to be explained by direct bacterial interference. Instead, this observation could reflect age-related changes in host mucosal immunity, local micronutrients, or URT microbiota composition, particularly as the antagonism between these species has generally not been observed in older children⁴⁴ or adults.²²

We also identified infant characteristics and environmental exposures that were associated with pathobiont colonization during infancy in this setting. First, we found that younger age was associated with a higher pneumococcal colonization density. Prior studies have reported conflicting findings for this association, with higher densities observed with decreasing age among infants in Indonesia.⁴⁵ However, a recent study conducted in the United States reported lower densities among children 6 months of age or younger compared to children 7-24 months of age.⁴⁶ Additionally, we found that receipt of PCV-13 was not associated with the colonization density among infants with pneumococcal carriage. O'Brien and colleagues reported that vaccination was associated with a lower colonization density for only vaccine serotypes,⁴⁷ while the majority of pneumococci that we previously isolated from infants in this cohort were non-vaccine serotypes.⁴⁸ However, other studies have reported no association between vaccination and pneumococcal colonization density, and some studies have observed higher densities among vaccine recipients.⁴⁹⁻⁵¹ Finally, we found that URT colonization by *S. aureus* among infants in Botswana is highly seasonal, with acquisition being much higher during the warm months of the rainy season, consistent with seasonal patterns of *S. aureus* skin and soft tissue infections reported in previous pediatric cohorts.⁵²⁻⁵⁴

With increasing recognition of the role the URT microbiota in modifying the risk and severity of ARIs, targeted modification of the URT microbiota is emerging as a promising strategy for ARI prevention. This study demonstrated that respiratory virus infection is associated with shifts in URT microbiota composition and the subsequent acquisition of bacterial pathobionts. More specifically, respiratory virus infection is associated with the acquisition of *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae*, with the acquisition of one of these pathobionts often increasing the odds of acquisition of the other two pathobionts. In addition, respiratory virus infection is associated with losses of commensal microbes that appear to contribute to colonization resistance to bacterial pathobionts in the URT. In particular, we found that ASVs classified as *C. propinquum/pseudodiphtheriticum* (ASV4), *C. accolens/macginleyi* (ASV6),

and *C. tuberculostearicum* (ASV10) declined in abundance in association with respiratory virus infection and were associated with lower odds of acquisition of *M. catarrhalis*, *S. aureus*, or *S. pneumoniae*. These findings are consistent with those from prior studies suggesting *Corynebacterium* species play a central role in excluding pathobionts from the URT.^{16,55-57} For example, Hardy and colleagues reported that *C. pseudodiphtheriticum* strains demonstrate contact-independent killing of *S. aureus in vitro*,⁵⁷ while Kiryukhina and colleagues found that nasal inoculation of a *C. pseudodiphtheriticum* strain eradicated or reduced the burden of *S. aureus* carriage in healthy volunteers.⁵⁸ We also found that several streptococcal ASVs were associated with lower odds of pathobiont acquisition, including an ASV classified as *S. thermophilus/vestibularis/salivarius* (ASV15) that declined in abundance with respiratory virus infection and an ASV classified as *Streptococcus mitis/oralis* (ASV9) that was negatively associated with the acquisition of *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae*. Interestingly, antibody responses to *S. mitis* are known to cross-react with *S. pneumoniae*, and intranasal administration of *S. mitis* resulted in lower bacterial loads in a murine model of pneumococcal pneumonia.⁵⁹ Moreover, a recent study reported *in vitro* growth inhibition of nontypeable *H. influenzae* strains by a *S. mitis* strain, although nasal inoculation of this strain did not protect mice from *H. influenzae* pneumonia.⁶⁰ We also found that higher abundances of an ASV classified as *D. pigrum* (ASV2) were negatively associated with *S. aureus* acquisition, consistent with findings from several epidemiological studies and *in vitro* experiments demonstrating inhibition of *S. aureus* growth by *D. pigrum* strains.^{17,61-63} Finally, our analyses suggest a potential antagonistic relationship between *Veillonella massiliensis* and *S. pneumoniae* that warrants further investigation.

We found that the acquisition of *H. influenzae*, *M. catarrhalis*, and *S. pneumoniae* were associated with ARI symptoms independent of respiratory virus detection, with a higher carriage density being additionally associated with ARI symptoms among infants with pneumococcal colonization. URT carriage of these bacterial pathobionts has historically been viewed as an asymptomatic state, although

several recent studies have begun to question this traditional dogma. Most notably, experimental human pneumococcal carriage is associated with a robust nasal inflammatory response and symptoms in many participants in the absence of secondary infection.⁶⁴ For example, among 96 healthy adults inoculated intranasally with a *S. pneumoniae* serotype 3 strain, more than half (53%) of the 38 colonized participants developed mild symptoms, while symptoms were reported by only 16% of noncolonized participants.⁶⁵ Additionally, periods of *H. influenzae*, *M. catarrhalis*, or *S. pneumoniae* colonization were associated with an increase in daily symptoms among adults with chronic obstructive pulmonary disease in the absence of clinical exacerbation.⁶⁶ The symptoms associated with the acquisition of bacterial respiratory pathobionts could also be more pronounced in young children than in adults, given that young age is associated with the upregulation of innate and adaptive immune pathways in the URT mucosa of healthy individuals.^{67,68}

This study has several limitations. First, it was conducted among infants at urban and rural sites in southern Botswana, and our findings may not be generalizable to other age groups or infants in other geographical areas. Additionally, infants were seen for scheduled visits every 1-2 months, and additional visits were not conducted for ARIs between these study visits. Thus, it is likely that some respiratory virus infections occurring among these infants were not identified. Moreover, because most respiratory virus infections were caused by rhinoviruses/enteroviruses, we lacked the statistical power to evaluate associations between other specific respiratory viruses and bacterial pathobiont colonization. Data from a quantitative PCR assay were only available for *S. pneumoniae*, precluding analyses of the impact of respiratory virus infection on the URT colonization density of the other bacterial pathobionts. Additionally, comparative analyses of URT microbiota composition were based on sample relative abundances, and we are thus unable to determine if the observed associations represent differences in the absolute amounts of bacterial taxa within the URT across samples. Finally, this study employed amplicon sequencing of the V4 region of the 16S rRNA gene to characterize the URT bacterial microbiota, which

precluded the classification of most sequencing reads to the species level and did not provide data for other microbial kingdoms.

In summary, this study provides novel insights into factors that influence URT colonization by bacterial pathobionts that are major causes of ARIs across the lifespan. Our findings contribute to knowledge of microbial ecology in the human URT and demonstrate the influence of viral-bacterial and bacterial-bacterial relationships on the colonization dynamics of these bacterial pathobionts during infancy. An improved understanding of these interspecies relationships could inform the development of novel biotherapeutic strategies that could reduce the global burden of ARIs among both children and adults.

METHODS

Study design and procedures

Study procedures and participant characteristics were previously published in detail.¹⁶ Briefly, mother-infant dyads were recruited within 72 hours of delivery at three urban or rural clinics in or near Gaborone, Botswana. Infants were born between February 2016 and January 2020 and were seen for monthly study visits during the first six months of life and every other month thereafter until the infants were 12 months of age. At all visits, nasopharyngeal swab samples were collected from mothers and infants and a caregiver questionnaire was administered to evaluate for ARI symptoms, defined by the presence of cough, rhinorrhea, or nasal congestion. Testing of infant nasopharyngeal samples for respiratory viruses was performed using real-time PCR assays, as described previously.² Testing for bacterial pathobionts (*H. influenzae*, *M. catarrhalis*, *S. aureus*, and *S. pneumoniae*) was similarly performed using a custom multiplex real-time PCR assay developed by DB (**Supplementary Table 3**). Pneumococcal colonization density was estimated using a quantitative PCR assay targeting the autolysin (*lytA*) gene, as previously

described.⁵⁵ All PCR assays were performed by the Infectious Disease Research Group in the Research Institute of St. Joe's Hamilton (Hamilton, Ontario, Canada).

Processing of samples for 16S ribosomal RNA gene sequencing

Infant nasopharyngeal samples underwent amplicon sequencing of the V4 region of the 16S rRNA gene in eight batches at Duke University (Durham, North Carolina, United States). DNA was extracted using Powersoil Pro extraction kits (Qiagen). DNA concentrations were determined using Qubit dsDNA high-sensitivity assay kits (Thermo Fisher Scientific). Negative extraction and PCR controls were amplified with all eight batches of samples included in analyses to evaluate for background contamination. For the first two sample batches, these negative controls were verified to not have visible bands on gel electrophoresis. For the remaining six batches, sequencing was performed on negative control samples. PCR amplification of the V4 region of the 16S rRNA gene was performed using primers 515F and 806R, and amplicon pools were sequenced using a MiSeq instrument (Illumina) configured for 2 x 250 bp sequencing.

Bioinformatic processing of 16S rRNA sequencing data

Paired-end reads were filtered and trimmed ($\text{maxEE} = 2$; $\text{truncLen} = 175/125$ bp), merged, denoised, chimera filtered, and binned into ASVs using DADA2 v1.16.0.⁶⁹ ASVs were given provisional taxonomic assignments using the DADA2 implementation of the naïve Bayesian classifier ($\text{minBoot}=80$) and based on alignment to the expanded Human Oral Microbiome Database v15.1.⁷⁰ Identification and removal of potential reagent contaminants and batch variability correction were performed using a two-tier strategy as proposed. First, contaminant ASVs were identified and removed based on the presence in negative control samples or negative correlation with DNA concentration using the frequency method ($\text{threshold}=0.10$) implemented in the *decontam* R package v1.24.0.⁷¹ Second, contaminant ASVs were identified and removed based on differential prevalence across the eight sequencing batches ($k=0.05$) as

described by Moossavia and colleagues.⁷² Samples with less than 1,000 sequencing reads after contaminant removal were excluded from further analyses. Sequencing reads were classified into 18,773 ASVs representing 248 genera from 12 phyla. ASVs were assigned numbers in ascending order based on mean relative abundance (ASV1 was the most abundant ASV across all samples in the final dataset). PICRUSt2 v2.5.2²⁷ was used to predict the abundances of metabolic pathways and enzymes in the MetaCyc database v27.1.⁷³

Statistical models for analyses of molecular data on respiratory viruses and bacterial pathobionts

Logistic regression was used to evaluate associations between infant age and the prevalence of bacterial pathobiont colonization as detected by PCR. To evaluate associations between the detection of respiratory viruses or bacterial pathobionts and the presence of ARI symptoms, a mixed effect logistic regression model was fit with subject as a random effect to account for repeated sampling of individuals. Mixed effect logistic regression was similarly used to evaluate associations between respiratory virus infection, preceding pathobiont colonization, and infant and sample characteristics and the acquisition of specific bacterial pathobionts. These models included subject as a random effect and the following infant and sample characteristics as fixed effects: age, sex, low birth weight (<2500g), maternal HIV infection, location of residence (urban vs. rural), number of children under 5 years of age in the household, season (time-varying; dry vs. rainy), breastfeeding (time-varying), receipt of antibiotics since the prior study visit (time-varying), and PCV-13 doses (time-varying; for *S. pneumoniae* only). Linear regression adjusting for these same variables was used to evaluate the impact of respiratory virus infection, pathobiont colonization, and infant and sample characteristics on the pneumococcal colonization density.

Statistical methods for analyzing infant upper respiratory microbiota composition

K-medoids clustering on Bray-Curtis distances was used to group samples into distinct microbiota types, with the optimal number of microbiota types chosen using the Calinski-Harabasz index. To assess the

robustness of the eight identified microbiota types to clustering algorithm and distance metric, samples were also grouped into eight microbiota types using k-means clustering and the default squared Euclidean distance metric. Mixed effect logistic regression was used to evaluate associations between infant age and the prevalence of each microbiota type. Similarly, mixed effect logistic regression models were fit to evaluate associations between microbiota type and the acquisition of specific bacterial pathobionts with subject as a random effect and adjusting for infant and sample characteristics as fixed effects. Analyses of URT microbiota composition conducted at the ASV level were limited to 31 ASVs present at a relative abundance at or above 0.1% in at least 5% of samples. To assign species information to each of these ASVs, we performed a standard nucleotide REFSEQ BLAST search using the National Center for Biotechnology Information's Bacteria and Archaea 16S ribosomal RNA project database⁷⁴ and used a best-hit approach based on the E value with a minimum percent identity of 97%. We used MaAsLin2 v1.18.0⁷⁵ to fit linear mixed effect models evaluating associations between respiratory virus detection by PCR and the centered log ratio-transformed relative abundances of ASVs within the URT microbiota. Models included subject as a random effect and the following fixed effects: age (cubic b-spline of the age in months), sex, low birth weight, maternal HIV infection, location of residence, the number of children under 5 years of age in the household, season (time-varying; dry vs. rainy), breastfeeding (time-varying), receipt of antibiotics since the prior study visit (time-varying), and PCV-13 doses (time-varying; for *S. pneumoniae* only). We evaluated the utility of the composition of the preceding URT microbiota in predicting the acquisition of bacterial pathobionts at the current study visit by applying the random forest algorithm with *randomForest* R package v4.7.1.1, restricting analyses to preceding samples testing negative for the pathobiont of interest by PCR. Prediction accuracy was evaluated using AUC-ROC with leave-one-subject-out cross-validation. Total sum scaling was used to normalize ASV count data to relative abundances before natural log transformation, with zeros replaced by 0.5 before conversion to relative abundances. For analyses of microbial pathway relative abundances, zeros were replaced by values corresponding to half of the sample-specific minimum. To address the unequal distribution of

events of pathobiont acquisition, we configured the random forests to use stratified sampling. The reported feature importance represents the average calculated across all leave-one-subject-out cross-validation runs. To identify specific ASVs that offered the highest marginal association with bacterial pathobiont acquisition while adjusting for infant age, we employed mixed effect logistic regression with the log relative abundance of each ASV and the cubic b-spline of age in months as predictors, pathobiont acquisition as the outcome, and a random effect to account for repeated measurements from the same subject. For all analyses, listwise deletion was used to exclude samples or time intervals with missing data for any variable. The Benjamini-Hochberg procedure was used to control the false discovery rate for analyses involving multiple comparisons ($q < 0.20$). All analyses were conducted using R v4.4.1.

DECLARATIONS

Acknowledgements: We would like to thank Copan Italia (Brescia, Italy) for the donation of the MSwab media and flocked swabs used in collecting nasopharyngeal specimens. We also thank the Duke University School of Medicine for using the Microbiome Core Facility, which performed the DNA extractions and library preparations for this research. We additionally thank the staff at the Sequencing and Genome Technologies core facility at Duke University for library sequencing. We offer sincere gratitude to the children and families who participated in this research.

Ethics approval and consent to participate

This study was approved by the Botswana Health Research & Development Committee, the Princess Marina Hospital ethics committee, and institutional review boards at the University of Pennsylvania, Duke University Health System, and McMaster University. Written informed consent was obtained from all participants or their legal guardians.

Funding

This research was supported in part by a research grant from Investigator-Initiated Studies Program of Merck & Co., Inc. The opinions expressed in this paper are those of the authors and do not necessarily represent those of Merck & Co., Inc. Additional support for this work was provided by a grant from the Burroughs Wellcome Fund and the American Society of Tropical Medicine and Hygiene, Children's Hospital of Philadelphia and the Pincus Family Foundation, and through core services from the Penn Center for AIDS Research, a National Institutes of Health (NIH)-funded program (P30-AI045008). MSK was supported by a NIH Career Development Award (K23-AI135090) and a research grant from the Society for Pediatric Research.

Author contributions:

Conceptualization: MSK, PCS, MS

Methodology: MSK, PS, APS, CKC, JFR, JHH, PCS, DB, MS

Investigation: MSK, SCB, EQ, TM, MZP, KL, JG, JM

Visualization: MSK, PS

Funding acquisition: MSK, APS

Project administration: MSK

Supervision: MSK

Writing – original draft: MSK

Writing – review & editing: PS, SCB, EQ, APS, TM, MZP, CKC, JFR, KL, JG, JM, JHH, PCS, DB, MS

Competing interests: MSK is a consultant for Merck & Co, Inc. and Invivyd. All other authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials. The 16S ribosomal RNA sequencing dataset supporting the conclusions of this study is available in the Sequence Read Archive (PRJNA698366). The statistical files and scripts used for data analyses are also publicly available (https://github.com/mskelly7/Pathobiont_colonization).

List of Supplementary Materials

Supplementary Tables 1-3

Supplementary Figures 1-3

REFERENCES

1. Bender RG, Sirota SB, Swetschinski LR, Dominguez R-MV, Novotney A, Wool EE, et al. Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990–2021: a systematic analysis from the Global Burden of Disease Study 2021. *The Lancet Infectious Diseases*. 2024.
2. Kelly MS, Smieja M, Luinstra K, Wirth KE, Goldfarb DM, Steenhoff AP, et al. Association of respiratory viruses with outcomes of severe childhood pneumonia in Botswana. *PloS one*. 2015;10(5):e0126593.
3. O'Brien KL, Baggett HC, Brooks WA, Feikin DR, Hammitt LL, Higdon MM, et al. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *The Lancet*. 2019;394(10200):757-79.
4. Wahl B, O'Brien KL, Greenbaum A, Majumder A, Liu L, Chu Y, et al. Burden of *Streptococcus pneumoniae* and *Haemophilus influenzae* type b disease in children in the era of conjugate vaccines: global, regional, and national estimates for 2000–15. *The Lancet Global Health*. 2018;6(7):e744-e57.
5. Chien Y-W, Klugman KP, Morens DM. Bacterial pathogens and death during the 1918 influenza pandemic. *New England Journal of Medicine*. 2009;361(26):2582-3.
6. Morens DM, Taubenberger JK, Fauci AS. Predominant Role of Bacterial Pneumonia as a Cause of Death in Pandemic Influenza: Implications for Pandemic Influenza Preparedness. *The Journal of infectious diseases*. 2008;198(7):962-70.
7. Finelli L, Fiore A, Dhara R, Brammer L, Shay DK, Kamimoto L, et al. Influenza-associated pediatric mortality in the United States: increase of *Staphylococcus aureus* coinfection. *Pediatrics*. 2008;122(4):805-11.
8. Kallen AJ, Brunkard J, Moore Z, Budge P, Arnold KE, Fosheim G, et al. *Staphylococcus aureus* community-acquired pneumonia during the 2006 to 2007 influenza season. *Annals of emergency medicine*. 2009;53(3):358-65.
9. Zhou H, Haber M, Ray S, Farley MM, Panozzo CA, Klugman KP. Invasive pneumococcal pneumonia and respiratory virus co-infections. *Emerging infectious diseases*. 2012;18(2):294.
10. Kim PE, Musher DM, Glezen WP, Barradas MCR, Nahm WK, Wright CE. Association of invasive pneumococcal disease with season, atmospheric conditions, air pollution, and the isolation of respiratory viruses. *Clinical infectious diseases*. 1996;22(1):100-6.
11. Lee LN, Dias P, Han D, Yoon S, Shea A, Zakharov V, et al. A mouse model of lethal synergism between influenza virus and *Haemophilus influenzae*. *Am J Pathol*. 2010;176(2):800-11.
12. Brockson ME, Novotny LA, Jurchisek JA, McGillivray G, Bowers MR, Bakaletz LO. Respiratory syncytial virus promotes *Moraxella catarrhalis*-induced ascending experimental otitis media. *PloS one*. 2012;7(6):e40088.
13. Brueggemann AB, van Rensburg MJJ, Shaw D, McCarthy ND, Jolley KA, Maiden MC, et al. Changes in the incidence of invasive disease due to *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Neisseria meningitidis* during the COVID-19 pandemic in 26 countries and territories in the Invasive

Respiratory Infection Surveillance Initiative: a prospective analysis of surveillance data. *The Lancet Digital Health*. 2021;3(6):e360-e70.

14. Willen L, Ekinçi E, Cuypers L, Theeten H, Desmet S. Infant pneumococcal carriage in Belgium not affected by COVID-19 containment measures. *Frontiers in cellular and infection microbiology*. 2022;11:825427.
15. Dagan R, van der Beek BA, Ben-Shimol S, Greenberg D, Shemer-Avni Y, Weinberger DM, et al. The COVID-19 pandemic as an opportunity for unravelling the causative association between respiratory viruses and pneumococcus-associated disease in young children: a prospective study. *EBioMedicine*. 2023;90.
16. Kelly MS, Plunkett C, Yu Y, Aquino JN, Patel SM, Hurst JH, et al. Non-diphtheriae *Corynebacterium* species are associated with decreased risk of pneumococcal colonization during infancy. *The ISME journal*. 2021:1-11.
17. Accorsi EK, Franzosa EA, Hsu T, Cordy RJ, Maayan-Metzger A, Jaber H, et al. Determinants of *Staphylococcus aureus* carriage in the developing infant nasal microbiome. *Genome biology*. 2020;21(1):1-24.
18. Stubbendieck RM, Dissanayake E, Burnham PM, Zelasko SE, Temkin MI, Wisdorf SS, et al. *Rothia* from the human nose inhibit *Moraxella catarrhalis* colonization with a secreted peptidoglycan endopeptidase. *mBio*. 2023;14(2):e00464-23.
19. Stubbendieck RM, May DS, Chevrette MG, Temkin MI, Wendt-Pienkowski E, Cagnazzo J, et al. Competition among nasal bacteria suggests a role for siderophore-mediated interactions in shaping the human nasal microbiota. *Appl Environ Microbiol*. 2019;85(10):e02406-18.
20. Jacoby P, Watson K, Bowman J, Taylor A, Riley TV, Smith DW, et al. Modelling the co-occurrence of *Streptococcus pneumoniae* with other bacterial and viral pathogens in the upper respiratory tract. *Vaccine*. 2007;25(13):2458-64.
21. Pettigrew MM, Gent JF, Revai K, Patel JA, Chonmaitree T. Microbial interactions during upper respiratory tract infections. *Emerging infectious diseases*. 2008;14(10):1584.
22. Shiri T, Nunes MC, Adrian PV, Van Niekerk N, Klugman KP, Madhi SA. Interrelationship of *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Staphylococcus aureus* colonization within and between pneumococcal-vaccine naïve mother-child dyads. *BMC infectious diseases*. 2013;13(1):483.
23. Nzenze S, Shiri T, Nunes M, Klugman K, Kahn K, Twine R, et al. Temporal association of infant immunisation with pneumococcal conjugate vaccine on the ecology of *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Staphylococcus aureus* nasopharyngeal colonisation in a rural South African community. *Vaccine*. 2014;32(42):5520-30.
24. Chien YW, Vidal JE, Grijalva CG, Bozio C, Edwards KM, Williams JV, et al. Density interactions among *Streptococcus pneumoniae*, *Haemophilus influenzae* and *Staphylococcus aureus* in the nasopharynx of young Peruvian children. *The Pediatric infectious disease journal*. 2013;32(1):72-7.
25. Dunne EM, Murad C, Sudigdoadi S, Fadlyana E, Tarigan R, Indriyani SAK, et al. Carriage of *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Staphylococcus aureus* in Indonesian children: A cross-sectional study. *PloS one*. 2018;13(4):e0195098.

26. McNally LM, Jeena PM, Gajee K, Sturm AW, Tomkins AM, Coovadia HM, et al. Lack of association between the nasopharyngeal carriage of *Streptococcus pneumoniae* and *Staphylococcus aureus* in HIV-1–infected South African children. *The Journal of infectious diseases*. 2006;194(3):385-90.
27. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, et al. PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology*. 2020;38(6):685-8.
28. Soeters HM, Blain A, Pondo T, Doman B, Farley MM, Harrison LH, et al. Current Epidemiology and Trends in Invasive *Haemophilus influenzae* Disease—United States, 2009–2015. *Clinical Infectious Diseases*. 2018;67(6):881-9.
29. Mohanty S, Done N, Liu Q, Song Y, Wang T, Gaburo K, et al. Incidence of pneumococcal disease in children \leq 48 months old in the United States: 1998–2019. *Vaccine*. 2024;42(11):2758-69.
30. Mohanty S, Podmore B, Cuñado Moral A, Weiss T, Matthews I, Sarpong E, et al. Incidence of acute otitis media from 2003 to 2019 in children \leq 17 years in England. *BMC public health*. 2023;23(1):201.
31. Hu T, Podmore B, Barnett R, Beier D, Galetzka W, Qizilbash N, et al. Incidence of acute otitis media in children $<$ 16 years old in Germany during 2014–2019. *BMC pediatrics*. 2022;22(1):204.
32. Bogaert D, de Groot R, Hermans P. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. *The Lancet infectious diseases*. 2004;4(3):144-54.
33. Kelly M, Wirth K, Madrigano J, Feemster K, Cunningham C, Arscott-Mills T, et al. The effect of exposure to wood smoke on outcomes of childhood pneumonia in Botswana. *The International Journal of Tuberculosis and Lung Disease*. 2015;19(3):349-55.
34. Kelly MS, Wirth KE, Steenhoff AP, Cunningham CK, Arscott-Mills T, Boiditswe SC, et al. Treatment failures and excess mortality among HIV-exposed, uninfected children with pneumonia. *Journal of the Pediatric Infectious Diseases Society*. 2014:piu092.
35. Vu HTT, Yoshida LM, Suzuki M, Nguyen HAT, Nguyen CDL, Nguyen ATT, et al. Association between nasopharyngeal load of *Streptococcus pneumoniae*, viral coinfection, and radiologically confirmed pneumonia in Vietnamese children. *The Pediatric infectious disease journal*. 2011;30(1):11-8.
36. Baggett HC, Watson NL, Deloria Knoll M, Brooks WA, Feikin DR, Hammitt LL, et al. Density of upper respiratory colonization with *Streptococcus pneumoniae* and its role in the diagnosis of pneumococcal pneumonia among children aged $<$ 5 years in the PERCH study. *Clinical Infectious Diseases*. 2017;64(suppl_3):S317-S27.
37. Short KR, Reading PC, Wang N, Diavatopoulos DA, Wijburg OL. Increased nasopharyngeal bacterial titers and local inflammation facilitate transmission of *Streptococcus pneumoniae*. *mBio*. 2012;3(5):10.1128/mbio.00255-12.
38. Cusumano JA, Dupper AC, Malik Y, Gavioli EM, Banga J, Berbel Caban A, et al., editors. *Staphylococcus aureus* bacteremia in patients infected with COVID-19: a case series. *Open forum infectious diseases*; 2020: Oxford University Press US.

39. Bassetti M, Magnasco L, Vena A, Portunato F, Giacobbe DR. Methicillin-resistant *Staphylococcus aureus* lung infection in coronavirus disease 2019: how common? Current opinion in infectious diseases. 2022;35(2):149-62.
40. Hall-Stoodley L, Hu FZ, Gieseke A, Nistico L, Nguyen D, Hayes J, et al. Direct detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media. *JAMA : the journal of the American Medical Association*. 2006;296(2):202-11.
41. Thornton RB, Rigby PJ, Wiertsema SP, Filion P, Langlands J, Coates HL, et al. Multi-species bacterial biofilm and intracellular infection in otitis media. *BMC pediatrics*. 2011;11:1-10.
42. Cope EK, Goldstein-Daruech N, Kofonow JM, Christensen L, McDermott B, Monroy F, et al. Regulation of virulence gene expression resulting from *Streptococcus pneumoniae* and nontypeable *Haemophilus influenzae* interactions in chronic disease. *PloS one*. 2011;6(12):e28523.
43. Schaar V, Nordström T, Mörgelin M, Riesbeck K. *Moraxella catarrhalis* outer membrane vesicles carry β -lactamase and promote survival of *Streptococcus pneumoniae* and *Haemophilus influenzae* by inactivating amoxicillin. *Antimicrobial agents and chemotherapy*. 2011;55(8):3845-53.
44. Esposito S, Terranova L, Ruggiero L, Ascolese B, Montinaro V, Rios WP, et al. *Streptococcus pneumoniae* and *Staphylococcus aureus* carriage in healthy school-age children and adolescents. *Journal of medical microbiology*. 2015;64(Pt 4):427-31.
45. Murad C, Dunne EM, Sudigdoadi S, Fadlyana E, Tarigan R, Pell CL, et al. Pneumococcal carriage, density, and co-colonization dynamics: A longitudinal study in Indonesian infants. *International Journal of Infectious Diseases*. 2019;86:73-81.
46. Kaur R, Pichichero M. Colonization, Density, and Antibiotic Resistance of *Streptococcus pneumoniae*, *Haemophilus Influenzae*, and *Moraxella catarrhalis* among PCV13-Vaccinated Infants in the First Six Months of Life in Rochester, New York: A Cohort Study. *Journal of the Pediatric Infectious Diseases Society*. 2023;12(3):135-42.
47. O'Brien KL, Millar EV, Zell ER, Bronsdon M, Weatherholtz R, Reid R, et al. Effect of Pneumococcal Conjugate Vaccine on Nasopharyngeal Colonization among Immunized and Unimmunized Children in a Community-Randomized Trial. *The Journal of infectious diseases*. 2007;196(8):1211-20.
48. Hurst JH, Shaik-Dasthagirisahab YB, Truong L, Boiditswe SC, Patel SM, Gilchrist J, et al. Serotype epidemiology and antibiotic resistance of pneumococcal isolates colonizing infants in Botswana (2016-2019). *PloS one*. 2024;19(5):e0302400.
49. Dunne EM, Satzke C, Ratu FT, Neal EFG, Boelsen LK, Matanitobua S, et al. Effect of ten-valent pneumococcal conjugate vaccine introduction on pneumococcal carriage in Fiji: results from four annual cross-sectional carriage surveys. *Lancet Glob Health*. 2018;6(12):e1375-e85.
50. Roca A, Bottomley C, Hill PC, Bojang A, Egere U, Antonio M, et al. Effect of Age and Vaccination With a Pneumococcal Conjugate Vaccine on the Density of Pneumococcal Nasopharyngeal Carriage. *Clinical Infectious Diseases*. 2012;55(6):816-24.
51. von Mollendorf C, Dunne EM, La Vincente S, Ulziibayar M, Suuri B, Luvsantseren D, et al. Pneumococcal carriage in children in Ulaanbaatar, Mongolia before and one year after the introduction of the 13-valent pneumococcal conjugate vaccine. *Vaccine*. 2019;37(30):4068-75.

52. Chen AE, Goldstein M, Carroll K, Song X, Perl TM, Siberry GK. Evolving epidemiology of pediatric *Staphylococcus aureus* cutaneous infections in a Baltimore hospital. *Pediatric emergency care*. 2006;22(10):717-23.
53. Kakar N, Kumar V, Mehta G, Sharma RC, Koranne RV. Clinico-bacteriological study of pyodermas in children. *The Journal of dermatology*. 1999;26(5):288-93.
54. Rogers M, Dorman DC, Gapes M, Ly J. A three-year study of impetigo in Sydney. *Medical journal of Australia*. 1987;147(2):63-5.
55. Kelly MS, Surette MG, Smieja M, Rossi L, Luinstra K, Steenhoff AP, et al. Pneumococcal Colonization and the Nasopharyngeal Microbiota of Children in Botswana. *The Pediatric infectious disease journal*. 2018;37(11):1176-83.
56. Bomar L, Brugger SD, Yost BH, Davies SS, Lemon KP. *Corynebacterium accolens* releases antipneumococcal free fatty acids from human nostril and skin surface triacylglycerols. *mBio*. 2016;7(1):e01725-15.
57. Hardy BL, Dickey SW, Plaut RD, Riggins DP, Stibitz S, Otto M, et al. *Corynebacterium pseudodiphtheriticum* exploits *Staphylococcus aureus* virulence components in a novel polymicrobial defense strategy. *mBio*. 2019;10(1):10.1128/mbio.02491-18.
58. Kiryukhina N, Melnikov V, Suvorov A, Morozova YA, Ilyin V. Use of *Corynebacterium pseudodiphtheriticum* for elimination of *Staphylococcus aureus* from the nasal cavity in volunteers exposed to abnormal microclimate and altered gaseous environment. *Probiotics and antimicrobial proteins*. 2013;5(4):233-8.
59. Shekhar S, Khan R, Schenck K, Petersen FC. Intranasal Immunization with the Commensal *Streptococcus mitis* Confers Protective Immunity against Pneumococcal Lung Infection. *Applied and environmental microbiology*. 2019;85(6):e02235-18.
60. Mathieu E, Marquant Q, Chain F, Bouguyon E, Saint-Criq V, Le-Goffic R, et al. An Isolate of *Streptococcus mitis* Displayed In Vitro Antimicrobial Activity and Deleterious Effect in a Preclinical Model of Lung Infection. *Nutrients*. 2023;15(2):263.
61. Khamash DF, Mongodin EF, White JR, Voskertchian A, Hittle L, Colantuoni E, et al. The Association Between the Developing Nasal Microbiota of Hospitalized Neonates and *Staphylococcus aureus* Colonization. *Open Forum Infectious Diseases*. 2019;6(4).
62. Brugger SD, Eslami SM, Pettigrew MM, Escapa IF, Henke MT, Kong Y, et al. *Dolosigranulum pigrum* cooperation and competition in human nasal microbiota. *Msphere*. 2020;5(5).
63. Khamash DF, Mongodin EF, White JR, Voskertchian A, Hittle L, Colantuoni E, et al., editors. The association between the developing nasal microbiota of hospitalized neonates and *Staphylococcus aureus* colonization. *Open forum infectious diseases*; 2019: Oxford University Press US.
64. Mitsi E, Nikolaou E, Goncalves A, Blizard A, Hill H, Farrar M, et al. RSV and rhinovirus increase pneumococcal carriage acquisition and density, whereas nasal inflammation is associated with bacterial shedding. *Cell host & microbe*. 2024.

65. Robinson RE, Mitsi E, Nikolaou E, Pojar S, Chen T, Reiné J, et al. Human infection challenge with serotype 3 pneumococcus. *American journal of respiratory and critical care medicine*. 2022;206(11):1379-92.
66. Desai H, Eschberger K, Wrona C, Grove L, Agrawal A, Grant B, et al. Bacterial colonization increases daily symptoms in patients with chronic obstructive pulmonary disease. *Annals of the American Thoracic Society*. 2014;11(3):303-9.
67. Yoshida M, Worlock KB, Huang N, Lindeboom RG, Butler CR, Kumasaka N, et al. Local and systemic responses to SARS-CoV-2 infection in children and adults. *Nature*. 2022;602(7896):321-7.
68. Hurst JH, Mohan AA, Dalapati T, George IA, Aquino JN, Lugo DJ, et al. Differential host responses within the upper respiratory tract and peripheral blood of children and adults with SARS-CoV-2 infection. *medRxiv*. 2023.
69. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods*. 2016;13(7):581.
70. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. New insights into human nostril microbiome from the expanded human oral microbiome database (eHOMD): a resource for the microbiome of the human aerodigestive tract. *Msystems*. 2018;3(6):e00187-18.
71. Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*. 2018;6(1):226.
72. Moossavi S, Fehr K, Khafipour E, Azad MB. Repeatability and reproducibility assessment in a large-scale population-based microbiota study: case study on human milk microbiota. *Microbiome*. 2021;9(1):41.
73. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research*. 2016;44(D1):D471-80.
74. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
75. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, et al. Multivariable Association Discovery in Population-scale Meta-omics Studies. *bioRxiv*. 2021.

TABLES

Table 1. Characteristics of the 300 infants in the study population

Subject Characteristics	N	%
Infant sex		
Female	167	56%
Male	133	44%
Median (IQR) infant birth weight, g	3085	(2850, 3395)
Infant HIV status		
HIV-unexposed, uninfected	212	71%
HIV-exposed, uninfected	87	29%
HIV-exposed, infected	1	<1%
Location of residence		
Rural	115	38%
Urban	185	62%
Electricity in the home	210	70%
Household use of solid fuels	164	55%
Maternal educational level		
None or primary	21	7%
Secondary	238	79%
Tertiary	41	14%
Median (IQR) number of household members		
<5 years	0	(0, 1)
5-17 years	1	(0, 3)
≥18 years	3	(2, 4)
Enrollment season		
Dry (April to October)	190	63%
Rainy (November to March)	110	37%
One or more PCV-13 doses during follow-up	248	83%
Infant breastfeeding, any	250	83%
Infant antibiotic exposure, any		
Amoxicillin	118	39%
Metronidazole	31	10%
Trimethoprim-sulfamethoxazole	75	25%

IQR, interquartile range; PCV-13, 13-valent pneumococcal conjugate

Table 2. Associations between respiratory virus detections and the odds of acquisition of specific bacterial pathogens during infancy

Microbiological testing	<i>H. influenzae</i> acquisition		<i>M. catarrhalis</i> acquisition		<i>S. aureus</i> acquisition		<i>S. pneumoniae</i> acquisition	
	Odds ratio (95% CI)	P	Odds ratio (95% CI)	P	Odds ratio (95% CI)	P	Odds ratio (95% CI)	P
Coincident respiratory virus infection	1.44 (1.06-1.96)	0.02	1.34 (0.96-1.87)	0.09	0.45 (0.29-0.70)	0.0004	1.83 (1.33-2.52)	0.0002
Preceding <i>H. influenzae</i> colonization	□	□	1.29 (0.80-2.08)	0.30	0.92 (0.55-1.51)	0.72	1.42 (0.90-2.24)	0.13
Preceding <i>M. catarrhalis</i> colonization	1.82 (1.28-2.60)	0.0009	□	□	0.75 (0.48-1.18)	0.22	1.58 (1.12-2.24)	0.009
Preceding <i>S. aureus</i> colonization	0.94 (0.64-1.40)	0.77	1.27 (0.89-1.80)	0.19	□	□	1.02 (0.71-1.48)	0.91
Preceding <i>S. pneumoniae</i> colonization	1.42 (1.01-1.98)	0.04	1.07 (0.71-1.60)	0.75	0.66 (0.41-1.06)	0.09	□	□
Infant and sample characteristics								
Male sex	0.79 (0.55-1.14)	0.21	0.99 (0.70-1.40)	0.95	1.62 (1.05-2.49)	0.03	1.09 (0.77-1.53)	0.63
Low birth weight (<2500g)	1.09 (0.57-2.11)	0.79	0.88 (0.48-1.61)	0.67	0.75 (0.34-1.67)	0.48	1.31 (0.71-2.43)	0.38
Maternal HIV infection	1.02 (0.62-1.66)	0.95	0.85 (0.53-1.37)	0.51	0.67 (0.37-1.21)	0.18	1.27 (0.80-2.02)	0.31
Urban residence	0.98 (0.67-1.42)	0.90	1.06 (0.74-1.51)	0.76	1.17 (0.75-1.82)	0.50	1.56 (1.09-2.23)	0.01
Household use of solid fuels	1.06 (0.72-1.55)	0.77	1.47 (1.02-2.10)	0.04	1.22 (0.77-1.91)	0.40	1.07 (0.75-1.55)	0.70
Number of children in the home	1.21 (1.05-1.41)	0.01	1.14 (0.99-1.32)	0.06	0.95 (0.79-1.15)	0.59	1.26 (1.09-1.46)	0.002
Rainy season	1.04 (0.77-1.39)	0.80	0.55 (0.41-0.75)	0.0002	1.58 (1.11-2.26)	0.01	0.78 (0.57-1.05)	0.10
Number of PCV-13 doses received	□	□	□	□	□	□	1.15 (0.91-1.44)	0.24
Current breastfeeding	0.87 (0.55-1.37)	0.55	0.81 (0.51-1.30)	0.39	0.49 (0.27-0.87)	0.01	1.16 (0.73-1.84)	0.52
Antibiotics since prior study visit	1.00 (0.68-1.47)	0.99	1.16 (0.76-1.78)	0.49	0.48 (0.27-0.85)	0.01	0.75 (0.50-1.14)	0.18

Odds ratios, 95% confidence intervals, and p values were estimated using mixed effect logistic regression models with inclusion of age as a fixed effect. CI, confidence interval; PCV-13, 13-valent pneumococcal conjugate vaccine

FIGURE LEGENDS

Figure 1. Respiratory viral infections and bacterial pathobiont colonization during infancy. a.

Overview of the number of infants (N) and samples (n) from which data were available for respiratory virus testing by PCR, identification of bacterial pathobiont colonization by PCR, and characterization of the upper respiratory tract microbiota through sequencing of the 16S ribosomal RNA gene. The number of samples with available data is further shown by age of sample collection in months (m0-m12). **b.** Stacked histogram showing the prevalence of detection of specific respiratory viruses by age of sample collection. **c.** Line plots depicting the prevalence of nasopharyngeal colonization by bacterial pathobionts by age of sample collection. URT, upper respiratory tract; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

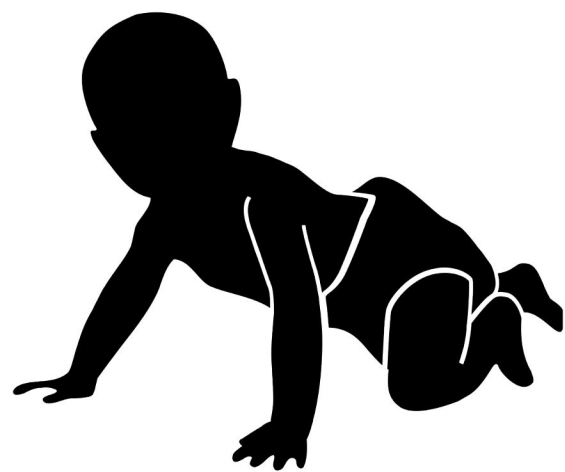
Figure 2. Viral-bacterial and bacterial-bacterial relationships and the dynamics of bacterial pathobiont colonization during infancy. a.

Forest plots presenting results from mixed effect logistic regression models evaluating associations between preceding bacterial pathobiont colonization or respiratory virus infection and the subsequent acquisition of specific bacterial pathobionts. **b.** Box and whisker plots show the \log_{10} copies/mL of *Streptococcus pneumoniae* DNA detected by quantitative PCR in samples from infants with pneumococcal colonization with and without concurrent respiratory virus detection or *H. influenzae* colonization. Lines splitting the boxes correspond to median values, box edges represent the 25th and 75th percentiles, and single points represent data for individual samples. p values were estimated using linear regression. **c.** Venn diagram depicting the co-occurrence of bacterial respiratory pathobionts within infant upper respiratory samples, including the number (percentage) of samples with detection of specific combinations of these pathobionts. There were 586 (25%) samples from which no bacterial pathobionts were detected (not shown).

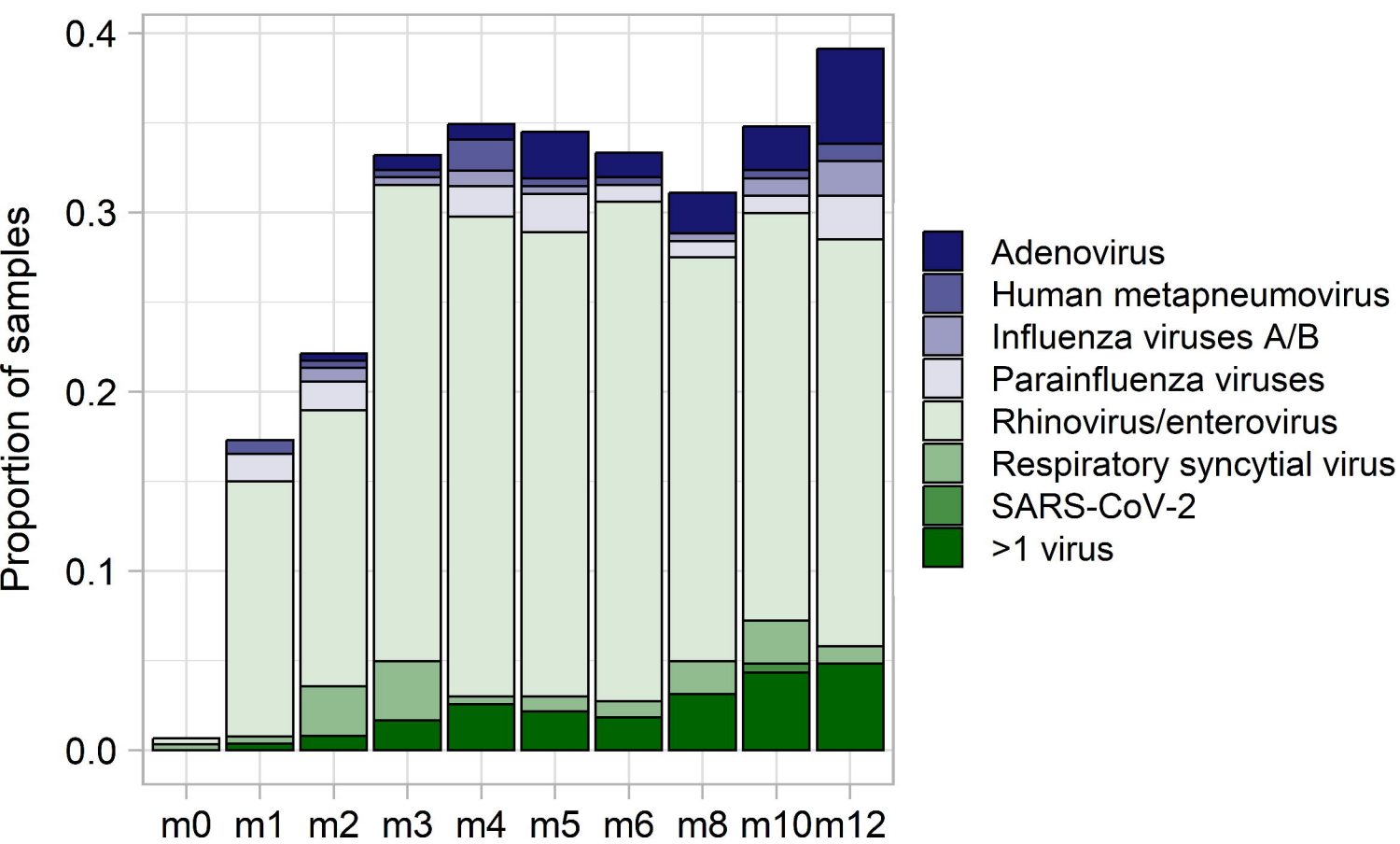
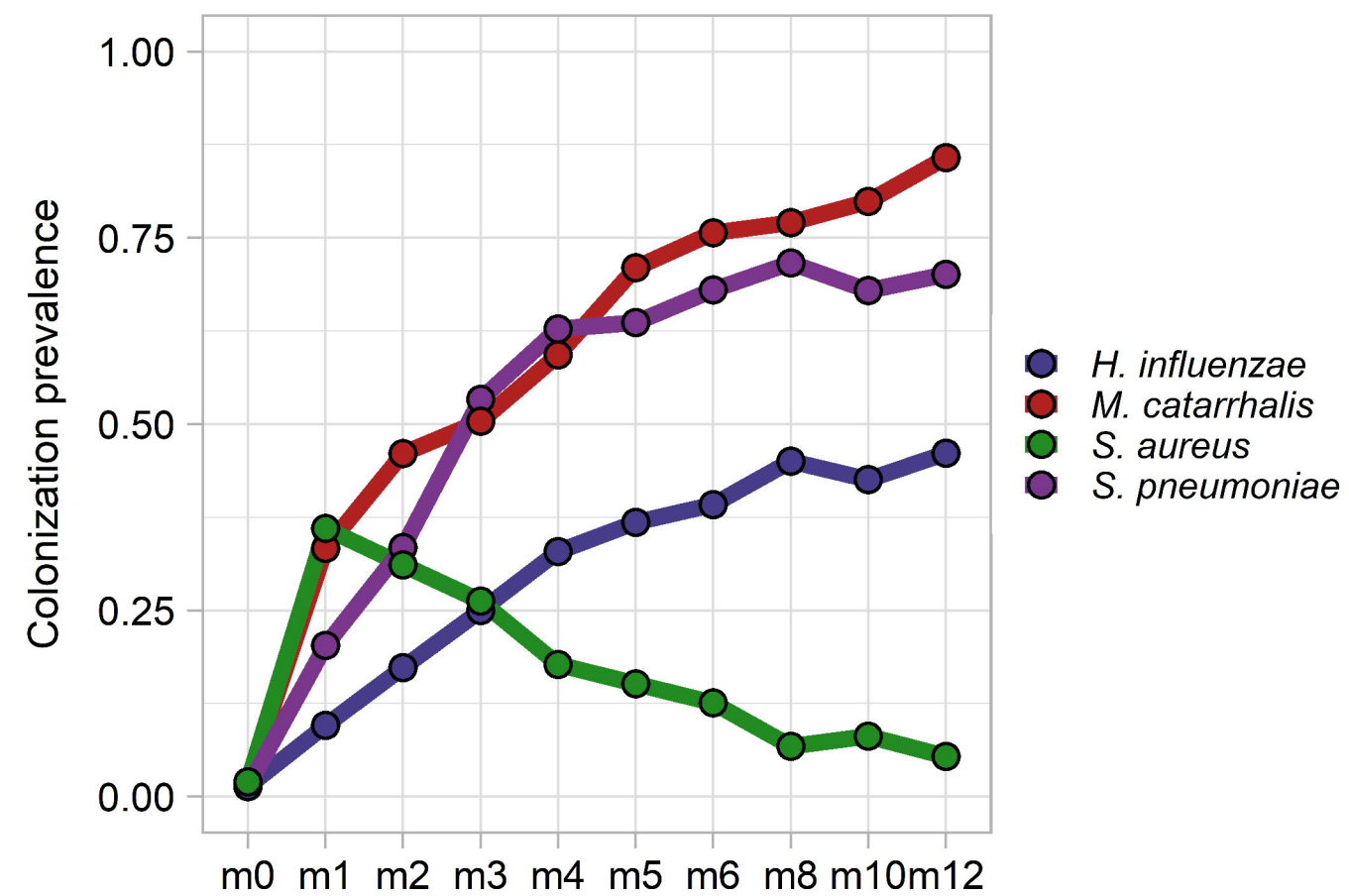
Figure 3. The dynamics of the upper respiratory microbiota among infants in Botswana. a. NMDS plot based on Bray-Curtis distances depicts microbiota composition based on 16S rRNA sequencing of nasopharyngeal swabs collected from infants in Botswana during the first year of life. k-medoids clustering was used to identify 8 distinct microbiota types. Ellipses define the regions containing 80% of all samples that can be drawn from the underlying multivariate t distribution. **b.** Relative abundances of highly abundant genera in upper respiratory samples from infants (n=2,235 samples) by microbiota type. **c.,d.** Alluvial diagram depicting upper respiratory microbiota community state transitions during the first year of life (**c.**) and with respiratory virus infection (**d.**). **e.** MaAsLin2 was used to fit a generalized linear mixed model evaluating the association between respiratory virus infection and the relative abundances of ASVs within the infant upper respiratory microbiota. The coefficients from these models, which correspond to the relative effect sizes of associations, are shown for significant associations ($q < 0.20$). ASVs that decreased/increased in abundance in association with respiratory virus infection are shown as blue/green bars. The taxonomic classification of each ASV based on BLAST searches is shown followed by a number in parentheses corresponding to the mean relative abundance of this ASV in infant nasopharyngeal samples (ASV1 was the most abundant ASV across all samples). URT microbiota types are abbreviated CD, *Corynebacterium/Dolosigranulum*-dominant; CDM, *Corynebacterium/Dolosigranulum/Moraxella*-dominant; COR, *Corynebacterium*-dominant; HAE, *Haemophilus*-dominant; MOR, *Moraxella*-dominant; STA, *Staphylococcus*-dominant; STR, *Streptococcus*-dominant; and OTH, “other.” ASV, amplicon sequence variant; non-metric multidimensional scaling

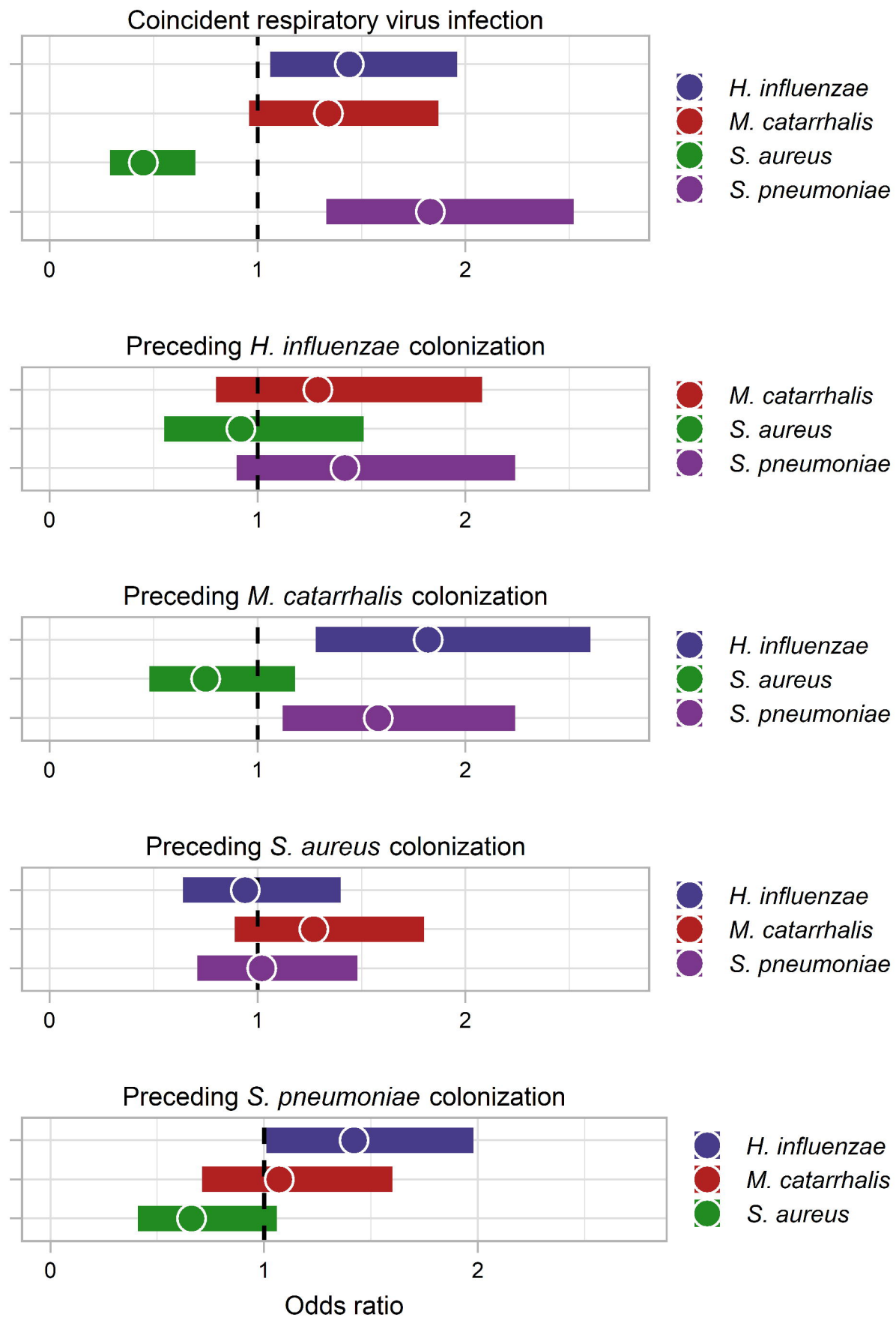
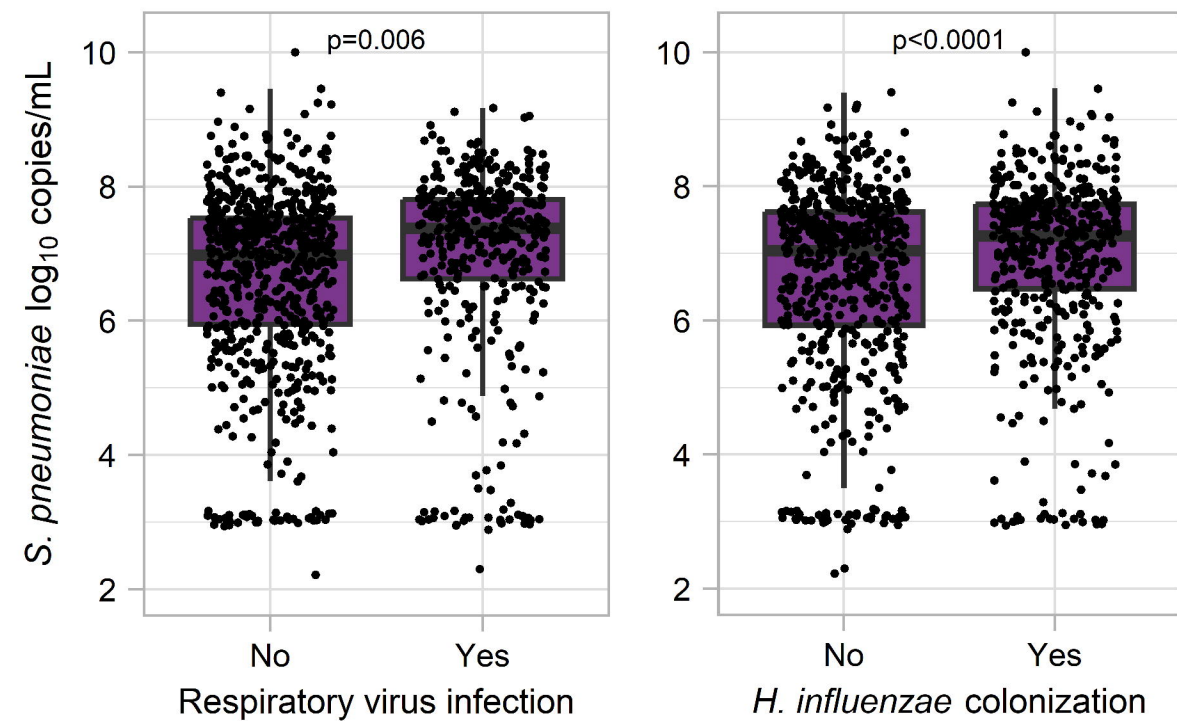
Figure 4. Associations between the upper respiratory microbiota and the acquisition of bacterial pathobionts. a. Forest plots depict the accuracy of random forest models for the prediction of acquisition of bacterial pathobionts. Clinical variables included in the models were age, number of children in the household (<5 years of age), respiratory virus infection, maternal HIV infection, breastfeeding, receipt of

antibiotics, season, low birth weight, use of solid fuels, location of residency (urban or rural), and 13-valent pneumococcal conjugate vaccine doses (for *S. pneumoniae* only). Data on URT microbiota taxonomy and microbial pathways were from the study visit preceding the visit at which the sample was collected to evaluate for bacterial pathobiont acquisition. For time-varying clinical variables, the data included in the models were collected at the study visit at which the sample was collected to evaluate for pathobiont acquisition. **b.** Heatmap depicting variable of importance scores for features in random forest models predicting pathobiont colonization and containing URT microbiota taxonomic data and clinical variables. The taxonomic classification of each ASV based on a BLAST search is shown, followed by a number in parentheses corresponding to the mean relative abundance of this ASV in infant nasopharyngeal samples (ASV1 was the most abundant ASV across all samples). **c.** Mixed effect logistic regression was used to identify URT microbiota features from the preceding sample that predicted pathobiont acquisition. The coefficients from these models, which correspond to the relative effect sizes of associations, are shown for significant associations ($q < 0.20$). ASVs for which higher/lower relative abundances were associated with lower odds of pathobiont acquisition are shown as blue/green bars. AUC-ROC, area under the receiving operating characteristic curve; URT, upper respiratory tract; ASV, amplicon sequence variant

a

	N	n	m0	m1	m2	m3	m4	m5	m6	m8	m10	m12
Respiratory viruses	300	2372	299	260	253	241	232	232	219	222	207	207
Bacterial pathobionts	298	2372	298	261	254	240	231	231	222	222	209	204
URT microbiota	291	2235	202	251	244	229	227	226	219	222	207	208

b**c**

a**b****c**