Insights into the relationship between nasal bacterial composition and susceptibility to early-life respiratory disease: a pilot observational study.

Jose A. Caparrós-Martín[1,2,3], Elizabeth Kicic-Starcevich[3], Patricia Agudelo-Romero[3,4,5], David G Hancock[1,3,6], Thomas Iosifidis[1,3,7,8], Yuliya V Karpievitch[3,7,9], David J Martino[3], Guicheng Zhang[1,7], Desiree T Silva[1,7,10,11,12], Anthony Bosco[13,14], Susan L Prescott[1,10,15], Peter N LeSouëf[1,3], Anthony Kicic[3,7,8], Stephen M. Stick[6,8].

[1]UWA Medical School, UWA Centre for Child Health Research, The University of Western Australia, Perth, Western Australia, Australia.

[2]Curtin Health and Innovation Research Institute (CHIRI), Curtin University, Perth, Western Australia, Australia.

[3]Wal-yan Respiratory Research Centre, The Kids Research Institute Australia, The University of Western Australia, Perth, Western Australia, Australia.

[4]Australian Research Council Centre of Excellence in Plant Energy Biology, School of Molecular Sciences, The University of Western Australia, Perth, Western Australia, Australia.

[5]European Virus Bioinformatics Centre, Friedrich-Schiller-Universität Jena, 07737, Thuringia, Germany.

[6]Department of Sleep and Respiratory Medicine, Perth Children's Hospital, Perth, Western Australia, Australia.

[7]School of Population Health, Curtin University, Bentley, Western Australia, Australia.

[8]Centre for Cell Therapy and Regenerative Medicine, School of Medicine and Pharmacology, The University of Western Australia and Harry Perkins Institute of Medical Research, Perth, Western Australia, Australia.

[9]School of Biomedical Sciences, The University of Western Australia, Perth, Western Australia, Australia.

[10]The Kids Research Institute Australia, The University of Western Australia, Nedlands, WA, Australia.

[11]School of Medical and Health Sciences, Edith Cowan University, Perth, Western Australia, Australia.

[12]Department of Paediatrics and Neonatology, Joondalup Health Campus, Perth, Western Australia, Australia.
[13]Asthma and Airway Disease Research Center University of Arizona Tucson, Arizona.
[14]Department of Immunobiology, The University of Arizona College of Medicine Tucson, Arizona.
[15]Nova Institute for Health, Baltimore, Maryland, United States; Department of Family and Community Medicine, University of Maryland, Baltimore, Maryland, United States.

Correspondence to:

Dr Jose A Caparros-Martin; The University of Western Australia, Perth, Western Australia, Australia; Email: joseantonio.caparrosmartin@uwa.edu.au

Dr Thomas Iosifidis; Wal-yan Respiratory Research Centre, The Kids Research Institute Australia, The University of Western Australia, Perth, Western Australia, Australia; Email: thomas.iosifidis@thekids.org.au

Professor Stephen M. Stick; Department of Sleep and Respiratory Medicine, Perth Children's Hospital, Perth, Western Australia, Australia; Email: Stephen.stick@health.wa.gov.au

Abstract

Background. Early-life susceptibility to viral respiratory infections is associated with long-term respiratory morbidity in children. Currently, no reliable tools exist to predict susceptibility to these infections. Given its role in modulating pathogen virulence and airway inflammation, the endogenous microbiota represents a potential target for prevention. In this pilot study, we investigated whether distinct nasal microbial communities are associated with viral respiratory disease and respiratory phenotypes during the first year of life.

Methods. We analyzed 90 nasal swabs from 55 infants enrolled in the AERIAL study, representing background samples collected at schedule visits (~4 months old) and samples from symptomatic episodes. Bacterial profiling was done blinded to clinical data via full-length *16S rRNA* sequencing, and bacterial load was quantified using a pan-bacterial TaqMan® assay.

Results. Nasal bacterial diversity was similar between background and symptomatic swabs, with community composition differences largely driven by inter-individual variation. In paired samples, symptomatic swabs showed reduced diversity but no change in bacterial load. Virus-bacteria interactions were observed in rhinovirus-positive swabs, but not in SARS-CoV-2-positive swabs. Two distinct bacterial endotypes were identified, enriched in either *Moraxella* or *Streptococcus* species and differing in alpha diversity. In background swabs, microbial endotypes were not associated with the number of symptomatic swabs or bronchiolitis episodes. We observed a potential sex-specific association between background swabs endotypes and number of wheezing episodes, which warrants further investigation in follow-up studies.

Conclusion. Our pilot data suggest that the nasal microbiota might influence wheezing outcomes in a sex-specific manner, highlighting the need for larger, longitudinal investigations.

Keywords Viral infection, respiratory epithelia, nasal microbiota, ORIGINS.

Introduction

Acute respiratory infections are a leading cause of death in children under five and a significant source of morbidity, particularly among newborns and infants (1). Most symptomatic viral respiratory infections occur early in life, with recurrent infections associated with an increased risk of developing chronic wheezing disorders during childhood (2, 3). Wheezing disorders including recurrent wheeze and asthma, are among the most prevalent pediatric respiratory illnesses (4). Rates and severity are highest in vulnerable populations, such as Indigenous children, lower-socioeconomic backgrounds, or children with underlying chronic diseases (4). Despite this burden, no clinically validated biomarkers exist to reliably predict which children will develop severe or persistent respiratory disease (4).

Longitudinal studies have shown that nasopharyngeal colonization trajectories influence the risk of respiratory infections during infancy and the later development of asthma and related wheezing disorders (5,6). In early life, the nasopharyngeal microbiota transitions from a community dominated by maternal *Staphylococcus* species to one enriched with *Corynebacterium*, *Dolosigranulum*, or *Moraxella*, with transient colonization by *Haemophilus* or *Streptococcus* (6,7). Early acquisition of stable, pathogen-dominated communities has been linked to airway inflammation and increased respiratory morbidity (5,6,8-12). For example, a recent study found that profiles dominated by *Haemophilus* and *Moraxella* are associated with early interferon-related epithelial responses and pro-inflammatory airway environments (5), whereas *Corynebacterium*-enriched communities correlated with respiratory health (5,6). However, findings across studies vary, with some reporting that *Moraxella*-dominated communities are associated with fewer upper respiratory tract infections and greater protection against viral infections in both adults and children (13-15). These discrepancies may stem, in part, from the limitations of targeting fragments of the *16S rRNA* gene, which do not provide sufficient resolution to determine species-level taxonomy (16).

While the nasopharyngeal microbiota have been well studied, it remains unclear whether nasal microbial communities, which are at the frontline of viral exposure, are associated with viral infection outcomes. In this pilot study, we used full-length *16S rRNA* sequencing to profile the bacteria associated with

nasal swabs from the Airway Epithelium Respiratory Illnesses and Allergy (AERIAL) birth cohort (17) and explore whether specific bacterial communities are associated with susceptibility to symptomatic respiratory infections and respiratory phenotypes in the first year of life.

Material and methods

AERIAL study cohort

We profiled the microbiota associated with the nasal epithelium in nasal swabs collected from healthy infants participating in the Airway Epithelium Respiratory Illnesses and Allergy (AERIAL) birth cohort study (17). Ethical approval for AERIAL was obtained from Ramsey Health Care HREC WA-SA (reference number 1908). AERIAL is a sub study nested within the ORIGINS cohort (18,19), with the central objective of evaluating whether a dysfunctional epithelium exists at birth, and whether this increases the risk for developing respiratory pathologies later in life (17). A unique aspect of the AERIAL study is the use of a smart phone app for real-time monitoring of respiratory symptoms (17, 20). AERIAL also collected longitudinal nasal swabs (FLoQSwab®, Copan Group, Italy) during predefined asymptomatic collection timepoints at approximately 3, 6, and 9 months of age, as well as during symptomatic episodes triggered by app alerts (17). Symptomatic swabs were defined by either a temperature >37.5°C along with at least one symptom (runny/blocked nose, cough, sneeze, headache, myalgia, chills, rigors, tiredness, sore throat, dyspnea, loss of taste/smell, poor feeding, or diarrhea), or a temperature <37.5°C with at least two respiratory symptoms present (17).

Clinical data for AERIAL study participants collected as part of ORIGINS (18) included maternal and infant demographics, and medical history for both mothers and infants amongst others (17,18). Additional clinical information on respiratory disease collected as part of the AERIAL study protocol, covered respiratory illness events, wheezing, allergies, and relevant respiratory medications used during scheduled visits and symptomatic episodes (17).

Study size

This pilot study was designed as an exploratory investigation to examine potential associations and estimate effect sizes to inform future, larger-scale

studies. A formal power calculation was not performed, consistent with the pilot nature of the work. Instead, we analysed a subset of participants randomly selected from the cohort at the time, to ensure feasibility within available resources while maintaining representativeness of the broader population.

Sample collection

Both nostrils were sampled at scheduled visits and when a symptomatic episode was recorded through the smartphone app. One swab (FLoQSwab®, Copan Group, Italy) was placed in viral transport media (eSwab, Copan Group, Italy) and assayed against a qPCR panel for 8 common respiratory viruses at commercial laboratory Western Diagnostics Pathology Pty Ltd, Western Australia (17). The qPCR panel was designed to detect Influenza "A" and "B", respiratory syncytial virus, human metapneumovirus, parainfluenza, rhinovirus, adenovirus and SARS-CoV-2. The second swab was placed in 1 mL of eNAT® buffer (eNAT®, Copan Group, Italy), transported to the laboratory at 4ºC and subsequently stored at -80ºC until processing for microbiome analyses (17).

Isolation of microbial DNA from nasal swabs, amplification of the full-length 16S rRNA gene and sequencing of the amplicon pools

Blinded to clinical data, we extracted bacterial DNA from 400 µL of eNAT® extract using the QIAamp DNA kit (QIAGEN). The recommended protocol was modified to include a bead-beating pre-processing step using 0.1 mm diameter zirconia/silica beads (BioSpec) as we previously described (21). Prior to the binding step, we added 100 µg of yeast tRNA carrier (ThermoFisher) to enhance DNA recovery. To identify reagent-related contaminants, negative extraction controls were processed alongside the swab specimens. We used a full-length *16S rRNA* gene amplicon sequencing approach to profile the bacterial component of the nasal epithelium-associated microbiota. Amplicon library preparation and sequencing were done at the Australian Genome Research Facility (AGRF Ltd, Melbourne, VIC, Australia). Briefly, for amplification of the full-length *16S rRNA* bacterial gene we used the degenerated forward 5'-*GCATC/barcode/AGRGTTYGATYMTGGCTCAG* and reverse 5'-*GCATC/barcode/RGYTACCTTGTTACGACTT* primers following the protocol recommended by PacBio (https://www.pacb.com/wp-

content/uploads/Procedure-checklist-Amplification-of-bacterial-full-length-16S-rRNA-gene-with-barcoded-primers.pdf). Individual PCR reactions showing the amplification of a single PCR product of the expected size are represented in Figure S1. Libraries were prepared using the SMRTbell® preparation kit 3.0, and sequenced using the Sequel II binding kit 3.1, on a single SMRT cell 8M, using a PacBio Sequel IIe instrument.

To minimize bias associated with amplifying and sequencing the *16SrRNA* gene in DNA extracts obtained from low bacterial biomass samples we followed two strategies: parallel DNA extraction and sequencing from eNAT® buffer negative controls, and bacterial load quantification of the DNA extracts using a pan bacteria TaqMan® assay (22). Details on the TaqMan® assay are described in our previous publication (23).

Processing of the sequencing data

Data was processed using *DADA2* (v1.28.0) (24), as implemented in the nfcore pipeline ampliseq (v2.8.0) (https://nfco.re/ampliseq), using Nextflow (v25.04.2) (25). For taxonomic classification of the generated amplicon sequencing variants (ASVs), we used the naïve Bayesian classifier implemented in *DADA2* (26), and a pretrained Genome Taxonomy Database (GTDB; vR09-RS220) as reference library (27). To prevent bias from potential contaminants and distinguish a true biological signal from background noise, we followed the analytical workflow outlined in our previous work (28). Briefly, only bacterial DNA extracts with qPCR result values higher than 3 standard deviations from the mean result in negative extraction controls were retained for further analyses (Figure S2). Then, the taxonomic profiles in the negative extraction controls were used to identify and remove potential contaminants using the R package *decontam* (v1.20.0) (29) (Table S1-S2). We performed Procrustes analysis to confirm that the described filtering steps did not significantly impact the overall structure of the original, unfiltered taxonomic data table (Table S3 and Figure S3). After filtering, we removed a sample with only 136 reads from further analyses.

Statistical analysis

Analyses were done in R statistical software (v4.3.0) within the RStudio environment (v2023.03.0). The R code and the deidentified datasets required to reproduce the findings of this study are available in the accompanying GitHub repository: https://github.com/jacapmar/MicrobiomeAERIAL. We used vegan (v2.6-4) for conducting permutational multivariate analysis of variance (PERMANOVA) and Procrustes analyses, and to calculate ecological diversity indexes. PERMANOVA was performed on a beta diversity distance matrix constructed using Aitchison distances. Principal Component Analysis (PCA) was done using the functions contained in the MixOmics package (v6.24.0) (30). To address the compositional nature of the microbiome data, a centered log-ratio (CLR) transformation was applied to the full taxonomy table before performing correlation-based analyses, including PCA and Spearman's rank correlation. For probabilistic modelling of microbial compositional profiles, we used Dirichlet Multinomial Mixtures (DMM) models as implemented in the R package DirichletMultinomial (v1.42.0) (31). Differential abundance analysis was performed using the Analysis of Compositions of Microbiomes with Bias Correction framework, as implemented in the *ancombc2* function from the ANCOMBC R package (v2.2.0) (32). According to the normality of the data and the hypotheses being tested, we also used Student's t-test and Wilcoxon Rank Sum Test. When multiple hypotheses were tested simultaneously, the false discovery rate method was used to control the type I error rate. Categorical data in contingency tables were compared using the Fisher exact test. Linear models were employed using the R built-in function *lm*. Logistic regression was fitted using the *glm* function in R. Count data was initially modelled using Poisson regression as implemented in the R built-in function *glm*. To evaluate the validity of the Poisson distributional assumption (that the mean equals the variance), we applied the *dispersiontest* function from the AER package in R (v1.2-14). This score test evaluates overdispersion based on Pearson residuals and the residual deviance. When overdispersion was detected (i.e., variance significantly greater than the mean), models were refitted using a negative binomial regression, which accounts for extra-Poisson variation by including a dispersion parameter. Negative binomial regression was fitted using the *glm.nb* function of the R package MASS (v7.3-60). Incidence risk ratios with 95% confidence intervals were calculated from the coefficients of the GLM and

GLM.nb models. Marginal effects were estimated using the *ggpredict* function from the ggeffects package (v2.2.1) in R. We used the R packages tidyverse (v2.0.0) for data manipulation and ggplot2 (v3.4.2) for data visualization. Figures were created using Adobe Illustrator (v24.1). The cut-off for statistical significance was set at $p<0.05$.

Results

Study cohort

A subset of 90 swabs from the AERIAL biobank was included in this study. Clinical metadata for the individuals represented in this swab cohort is provided in Table S4 and summarized in Table 1. These 90 swabs were collected from 55 randomly selected infants, of whom 30 (54%) were female. Forty-one swabs were obtained during scheduled visits (hereafter also referred to as background swabs), representing the nasal microbiota in the absence of respiratory symptoms. Background swabs were collected at 3-month (n=24, 58.5%), 6-month (n=15, 36.6%), or 9-month (n=2, 4.9%) visits (median age [IQR]: 131 days [104–200]). The median age for symptomatic swabs was 188 days (IQR [145-245]). Paired swabs consisting of a background swab and the first symptomatic episode collected after that background swab, were available for 33 infants. For one infant, two background and two symptomatic swabs were present in this cohort, but only one background and one symptomatic swab produced microbiome data.

In our cohort, 50 swabs tested positive for respiratory viruses by qPCR, including two background swabs and 48 symptomatic swabs (Table 1). Of the positive background swabs, one was positive for rhinovirus and one for SARS-CoV-2 (Table 1 and Table S4). Among symptomatic swabs, the most frequently detected viruses were rhinovirus (28, 56%) and SARS-CoV-2 (23, 46%) (Table 1 and Table S4). Five symptomatic swabs from five participants showed polyviral infections: SARS-CoV-2 and rhinovirus were detected together in three swabs; SARS-CoV-2, rhinovirus, and adenovirus in one swab; and rhinovirus, adenovirus, and parainfluenza in another (Table S4).

Characterization of the swab-associated nasal bacterial communities.

We obtained reliable microbial profiles from 71 (~79%) swab samples, of which 36 (~51%) represented background swabs. In line with previous findings from Western Australian cohorts (6,12), the swab-associated bacterial profiles included commensal taxa and opportunistic pathogens commonly found in the nasopharynx of healthy infants and children, including species from the genera *Moraxella* and *Streptococcus* (Figure S4). The most abundant features in our cohort were amplicon sequence variants (ASVs) assigned to the following taxa: *Moraxella nonliquefaciens* (mean [standard deviation (SD)]: 22.55% [36.42], prevalence 36.62%), *Moraxella catarrhalis* (13.71% [28.16], prevalence 43.66%), and *Streptococcus* (12.14% [16.94], prevalence 85.91%).

The nasal microbiota of some infants was characterized by the presence of respiratory pathogens (Figure S4). Given the established associations between early colonization by *Moraxella catarrhalis* or *Streptococcus pneumoniae* and increased susceptibility to respiratory infections and adverse respiratory phenotypes (5,9,11,12), we compared in our cohort the abundance of these pathogens between background and symptomatic swabs. Symptomatic swabs were characterized by higher relative abundance of *Streptococcus pneumoniae* (mean [SD]; background, 0.1% [0.5]; symptomatic, 4.3% [14.9]; Wilcoxon rank sum test (WRST) with false discovery rate (FDR) correction (r=0.27, p=0.044)) (Figure S5). However, we did not observe differences in the abundance of *Moraxella catarrhalis* (mean [SD]; background, 11.8% [28]; symptomatic, 15.6% [28.6]; WRST with FDR correction (r=0.12, p=0.32)) (Figure S5).

Interaction between respiratory viruses and bacterial taxa in the nose.

Previous research has demonstrated alterations in the composition of the nasal microbiota associated with respiratory viruses and suggested that these viruses can influence bacterial pathogen virulence (6,9,33). To explore potential multilevel interactions in our dataset, we focused our correlation analyses on the top three ASVs, which together accounted for 60% of the cumulative abundance (Figure 1A), and the two most frequently detected viruses; rhinovirus and SARS-CoV-2 (Table 1). Detection of rhinovirus was positively correlated with *Moraxella catarrhalis* (ρ = 0.37, FDR-adjusted p = 0.004), and negatively correlated with *Streptococcus* (ρ = -0.47, FDR-adjusted p = 0.0002), while SARS-CoV-2 showed no significant correlations with any of the bacterial

features analyzed (Figure 1B). Notably, bacterial alpha diversity in rhinovirus-positive symptomatic swabs was lower than in symptomatic swabs positive for SARS-CoV-2 (WRST, p = 0.021; Figure 1C). Likewise, differential abundance analysis at the species level using ANCOM-BC2 revealed that *Moraxella catarrhalis* was the only feature significantly enriched in rhinovirus-infected samples, with a fold change of 7.44 (95% CI: 2.39–23.18; adjusted p = 0.01). Conversely, bacterial load in symptomatic swabs did not vary with the virus detected (Figure 1D).

Microbial community analysis across symptomatic and background swabs.

Next, we evaluated community structural variation between samples collected during scheduled visits and those taken during a symptomatic episode. The PERMANOVA results suggest that the type of swab (whether background or symptomatic) was not a primary contributor to bacterial community composition variation (PERMANOVA: pseudo-F=0.99, R^2=0.01, p=0.38). This is further supported by the principal component analysis represented in Figure S6, which shows a high overlap between bacterial communities from background and symptomatic swabs. Similarly, we did not observe differences between swabs collected at scheduled visits or during symptomatic episodes in bacterial load (median [IQR]: background: 0.038 [0.036-0.041]; symptomatic: 0.038 [0.036-0.042]; WRST (r=0.01, p=0.92)) or alpha diversity (median [IQR]: background: 1.39 [0.7-1.79]; symptomatic: 1.03 [0.57-1.92]; WRST (r=0.07, p=0.60)) (Figure 2A-B). In contrast, PERMANOVA analysis demonstrated that most of the variance in our dataset (83%) was explained by interindividual differences (pseudo-F=2.53, R^2=0.83, p=0.0001), while sex accounted for little variation (pseudo-F=1.12, R^2=0.02, p=0.31), suggesting that other environmental covariates may play a more prominent role in shaping bacterial community composition in the nose.

To address the high inter-individual variation in our swab cohort, we focused our next analyses on longitudinal data, specifically comparing background and subsequent symptomatic swabs obtained from the same infant. We observed a reduction in bacterial alpha diversity in symptomatic swabs (paired WRST p=0.039)), while bacterial load remained unaffected (mean difference [95%

confidence interval (CI)]: -0.001 [-0.003-0.0009], paired t-test *p*=0.23) (Figure 2C-D).

Associations between bacterial community structure and future respiratory outcomes.

To evaluate the existence of bacterial community structures associated with risk of respiratory disease, we performed a probabilistic modelling of the bacterial community data using Dirichlet Multinomial Mixtures (31). We identified two microbial community types, hereafter referred to as endotype 1 and endotype 2. Endotype 1 was defined by four ASVs assigned to *Moraxella nonliquefaciens*, *Moraxella catarrhalis*, *Streptococcus*, and *Corynebacterium pseudodiphtheriticum* (Figure 3B and Figure S7). In contrast, endotype 2 was primarily defined by an ASV assigned to the *Streptococcus* genus (Figure 3B and Figure S7). Alpha diversity in endotype 2 was higher than in endotype 1 (median [IQR], endotype 1: 0.67 [0.34-0.96]; endotype 2: 1.86[1.52-2.40]; WRST (r=0.77, p=8.56x10^{-14})) (Figure 3C). Neither of the two endotypes was associated with sex (p=0.19, Fisher exact test FET) or seasonality (p=0.93, FET) (Figure S8). Similarly, the microbial community type present in background swabs was not associated with childcare attendance (p=0.58, FET). (Figure S8C).

In our dataset, children classified in endotype 1 had an expected rate of 2.53 wheezing episodes during the first year of life (95% CI: 1.34–5.23, p=0.007). Compared to endotype 1, those in endotype 2 exhibited a non-significant reduction in wheezing frequency (IRR=0.56, 95% CI: 0.22–1.40, p=0.22), suggesting a potential but statistically inconclusive association between nasal microbial endotypes and the number of wheezing episodes (Figure S9A). Similarly, Individuals in endotype 2 had a later onset of wheeze by an estimated 2.39 months compared to endotype 1, although this association was not significant (β=2.39, 95% CI: -0.55-5.34, F(1, 21)=2.85, p=0.106) (Figure 3D).

Building on prior research that identified sex-specific early-life risk factors for wheeze, we investigated whether nasal microbial endotypes showed a similar sex-specific pattern (34). Among females with endotype 1 (*Moraxella*-dominated), the rate of wheezing episodes in the first year of life was highest (IRR=4.00, 95% CI: 1.57–12.83, p=0.008; Figure 3E). In contrast, females with

endotype 2 (*Streptococcus*-dominated) had a significantly lower rate of episodes (IRR=0.18, 95% CI: 0.04–0.69, *p*=0.016), indicating an approximate 82% reduction compared to endotype 1 (Figure 3E, Figure S9B). Although male sex was not significantly associated with wheezing within endotype 1 (IRR=0.45, 95% CI: 0.12–1.65, *p*=0.23), we observed a significant interaction between male sex and endotype 2 (IRR=6.72, 95% CI: 1.03–43.72, *p*=0.036), suggesting that the effect of endotype 2 differs by sex. Specifically, while endotype 2 was protective in female, this effect was absent in males. Classification of samples into endotypes using a different probabilistic mixed model (Latent Dirichlet Allocation) yielded results comparable to those obtained with the DMM approach (Figure S10).

While the sex–endotype interaction was statistically significant, we identified a single influential data point with high leverage: a female participant with endotype 1 who experienced 14 wheezing episodes in the first year (Table S4, infant_11). Although this data point is valid and reflects natural variation in our cohort, its removal rendered the association non-significant, indicating sensitivity to individual cases (Figure S11). Thus, our findings suggesting a sex-specific link between the nasal microbiota endotypes and wheezing outcomes should be interpreted with caution and warrant validation in larger, independent datasets.

We found no association between background swab endotypes and the number of symptomatic swabs or bronchiolitis episodes during the first year of life (Figure S12A-B). Similar results were observed after adjusting for sex (Figure S12C-D). Similarly, compared to endotype 1, endotype 2 was not associated with the odds of wheezing (Odds Ratio [OR]=0.48, 95% CI: 0.11–1.96, *p*=0.32), recurrent wheezing (OR=0.47, 95% CI: 0.11–1.98, *p*=0.30), or bronchiolitis (OR=0.86, 95% CI: 0.18–4.16, *p*=0.84) during the first year of life, even after adjusting the models by sex (Figure S13).

To account for interindividual responses, we examined temporal transitions in nasal microbial endotypes between paired background and symptomatic swab samples collected from the same individuals, and linked these transition patterns to respiratory phenotypes, including the number of wheezing episodes and symptomatic swabs during the first year of life (Figure S14). We observed three types of transitions: from endotype 1 to endotype 1, from endotype 2 to

endotype 1, and from endotype 2 to endotype 2. Notably, no participants transitioned from endotype 1 to endotype 2, suggesting a high degree of stability for endotype 1 during symptomatic episodes. Despite these patterns, we observed no significant differences in the number of wheezing episodes or symptomatic swabs across the transition groups (Figure S14).

Discussion

In this observational study, we characterized the nasal microbiota of infants naturally exposed to respiratory viruses during their first year of life in a well-phenotyped infant cohort. We compared samples collected during scheduled visits with those taken during episodes of respiratory symptoms to identify microbial patterns linked to respiratory disease phenotypes.

We observed a complex and variable structure in the nasal microbiota, with several ASVs assigned to species from multiple genera, including *Streptococcus* and *Moraxella*. This contrasts with previous studies of the nasopharynx, which have reported that early-life microbial communities are typically simpler and dominated by a single taxon, either a commensal or a potential pathogen, such as *Corynebacterium* or *Moraxella* (5,6,11–13). While technical factors, such as the use of full-length *16S rRNA* gene sequencing (which provides improved taxonomic resolution), may in part explain these differences, it is also plausible that the anterior nares, being more exposed to the external environment, support greater microbial diversity and turnover than the more protected nasopharynx. This might also help explain the high interindividual variation in bacterial community structure observed in our swab cohort, a pattern similarly reported in other body sites directly exposed to the environment, such as the oral cavity (35).

Consistent with previous studies in Western Australian children (6,12), the nasal bacterial profiles in our cohort comprised both commensal and opportunistic taxa, particularly various species from the genera *Moraxella* and *Streptococcus*. Notably, *Streptococcus pneumoniae* was significantly enriched in symptomatic swabs, although its overall prevalence in the cohort was low (~11%) but comparable to the detection rates reported in previous research using nasal swabs (9). In contrast, the abundance of *Moraxella catharralis*, which has previously been shown to increase during viral respiratory infections (9,12,36),

did not differ between background and symptomatic swabs. The prevalence of *Moraxella* species was high in our cohort, both in background and symptomatic swabs, which contrasts with previous research conducted in Perth where early-life nasopharyngeal colonization by *Moraxella* was not observed (12). This difference might, in part, reflect the introduction of pneumococcal conjugate vaccines into Australia's National Immunization Program for all children since 2005, although associations between pneumococcal vaccination and upper airway colonization by *Moraxella* remain inconsistent across studies (37,38).

Despite the high viral burden observed in symptomatic swabs, most of which tested positive for either rhinovirus or SARS-CoV-2, community-level analyses revealed no significant differences in overall microbial composition, alpha diversity, or bacterial load between background and symptomatic swabs. This contrasts with several previous studies that reported reduced microbial diversity and increased dominance of pathogenic taxa from the genera *Streptococcus*, *Moraxella*, or *Haemophilus* during viral respiratory infections (6,36,39). We suggest this discrepancy might be due to the high interindividual variability in microbial composition within our cohort and the limited sample size. Notably, when comparing longitudinal samples from the same participants, we observed a significant reduction in bacterial alpha diversity in symptomatic swabs. Additionally, different viruses might also interact with nasal bacterial communities in distinct ways. Consistent with this, multilevel virus–bacteria interactions were observed in rhinovirus-positive swabs but not in those positive for SARS-CoV-2. Rhinovirus-positive symptomatic swabs exhibited lower bacterial alpha diversity compared to those infected with SARS-CoV-2; however, this difference was not observed in co-infected samples. Additionally, rhinovirus detection was positively correlated with the abundance of *Moraxella catarrhalis* and negatively with *Streptococcus*. Differential abundance analysis further revealed a significant enrichment in *Moraxella catarrhalis* in rhinovirus-positive swabs. The interaction between *Moraxella catarrhalis* and rhinovirus has been previously reported, with evidence suggesting that rhinovirus might enhance the pathogenicity of *Moraxella catarrhalis* and increase its abundance following infection (9,33,40). Interestingly, we did not observe significant correlations between bacterial taxa and SARS-CoV-2 detection, possibly reflecting differences in evolutionary history. While rhinoviruses might have co-

evolved with humans, fostering stable interactions with the respiratory microbiota, the recent emergence of SARS-CoV-2 might not yet support such ecological associations.

Using probabilistic modelling of the bacterial community structure, we identified two endotypes: a low-diversity, *Moraxella*-dominated endotype (endotype 1) and a *Streptococcus*-dominated endotype (endotype 2). These endotypes were not associated with demographic or environmental factors such as sex, seasonality, or childcare attendance. However, they were differentially associated with respiratory outcomes in a sex-specific manner. Endotype 1, when detected in background swabs, was linked to an increased rate of wheezing episodes, particularly among females. Additionally, detection of endotype 1 in background swabs showed a trend toward earlier wheezing onset. Although sex-specific predictors of wheeze have been described (34), a sex-specific association with the nasal microbiota has not been reported. Our results linking endotype 1 with wheeze are consistent with previous studies showing that early colonization with *Moraxella catarrhalis* increases the risk for recurrent wheeze and asthma in early life (11, 41). Notably, in addition to the abundance of *Moraxella catarrhalis*, endotype 1 was also characterized by the abundance of *Moraxella nonliquefaciens*. While *Moraxella catarrhalis* has been strongly implicated in asthma pathogenesis, the roles of other *Moraxella* species remain less well-defined, with some studies reporting conflicting associations regarding the presence of *Moraxella* in the nasopharynx and respiratory health (13-15). For example, a recent study identified *Moraxella lincolnii* in association with asthma (39). Thus, although both *Moraxella lincolnii* and *Moraxella nonliquefaciens* are generally considered commensals of the respiratory tract, their presence in the airways during early life might still be relevant in the etiology of wheezing phenotypes.

Temporal analysis of endotype stability in longitudinal samples revealed that most individuals retained their initial endotype during symptomatic episodes, particularly those with background swab communities classified as endotype 1. Interestingly, we observed transitions from endotype 2 to endotype 1 during a symptomatic episode, but not the reverse, suggesting that endotype 1 is relatively stable over the time. Similar findings have been reported in the nasopharynx, where *Moraxella*-dominated communities showed the highest

stability between two consecutive healthy samples (6). A comparable pattern has also been observed in patients with chronic obstructive pulmonary disease (COPD), where the acquisition of a *Haemophilus*-dominated airway microbiota appears to limit microbial community reconfiguration over the time (42). While different transition groups in our cohort were not associated with respiratory phenotypes in the first year of life, they could reflect distinct inflammatory profiles in response to symptomatic episodes, as it has been observed in COPD (42).

Our pilot study differs from previous research in the field in several key aspects. We amplified and sequenced the full length *16S rRNA* locus, which provides more accurate species-level taxonomic resolution, unlike studies that focus on shorter fragments of the *16S rRNA* gene (16). We minimized the impact of contaminating sequences in the bacterial profiles of our cohort by quantifying bacterial load and retaining only samples with significant bacterial DNA presence compared to extraction controls. Additionally, we utilized a purpose-designed smartphone app for real-time monitoring of respiratory symptoms, thereby minimizing bias associated with parent-reported data and recall limitations (17,20). We also acknowledge several limitations in our study. First, AERIAL is a single-center study (17), which might limit the generalizability of our findings to broader populations. Second, sample collection occurred between November 2021 and January 2023, a period following the COVID-19 lockdowns in Western Australia. To enable timely sampling despite public health restrictions, parents were trained to collect nasal swabs at home. However, these public health measures might have influenced the circulation and prevalence of respiratory viruses in the cohort (43). Third, our pilot study involved a relatively small sample size and was limited to two consecutive timepoints in the longitudinal analysis. Given the high individual variability, the small sample might have reduced statistical power to detect complex interactions and increased sensitivity to outliers. For instance, the observed sex-endotype interaction was influenced by a single participant with an unusually high number of wheezing episodes. While biologically plausible, such findings require confirmation in larger cohorts. Lastly, as with all amplicon-based studies, our microbial profiling approach does not provide functional

insights into how the nasal microbiota might modulate responses to viral infections.

Conclusion

In conclusion, this pilot observational study provides new evidence that early-life nasal bacterial communities, particularly when stratified into compositional endotypes, might be linked to wheezing risk in a sex-specific manner. Although preliminary, these findings highlight the potential utility of microbial profiling for early risk stratification of respiratory disease susceptibility and warrant validation in larger, independent cohorts.

Ethics approval.

The AERIAL study was approved by Ramsey Health Care HREC WA-SA, reference number 1908. Informed consent was obtained from the parents or caregivers of all children participating in the AERIAL study.

Data availability.

The taxonomy table and all de-identified datasets used for analysis are available as supplemental material and in the accompanying GitHub repository: https://github.com/jacapmar/MicrobiomeAERIAL. Due to ethical considerations regarding participant confidentiality, raw sequencing files cannot be made publicly available, however the supplemental de-identified datasets are sufficient for reproducing all reported findings. The R code is also available in the GitHub repository.

Conflict of interests.

AB is a co-founder, equity holder, and director of the startup company Respiradigm Pty Ltd that is related to this work. AB is the founder of the startup company INSiGENe Pty Ltd that is unrelated to this work. No other authors have conflict of interest to disclose.

Acknowledgments.

We are grateful to all participants in the AERIAL study and their families. Their willingness to adapt to COVID pandemic protocols and self-collect nasal swab

samples enabled our study to continue. We acknowledge the AERIAL clinical staff: Bailee Renouf, Amy Greenly, Angela Fuery, Jodie Leslie, Tanya Betty, Suzie Cooper, AERIAL laboratory staff: Ashleigh Heng-Chin, Luke Berry, Courtney Kidd, Minda Amin, and ORIGINS staff at Western Diagnostics Pathology: Fiona Rynne and Allan Albornoz for their continuous support and contribution towards the AERIAL study.

AERIAL is a sub-project of ORIGINS. This unique long-term study, a collaboration between The Kids Research Institute Australia and Joondalup Health Campus, is one of the most comprehensive studies of pregnant women and their families in Australia to date, recruiting 10,000 families over a decade from the Joondalup and Wanneroo communities of Western Australia. We are grateful to all the ORIGINS families who support the project. We would also like to acknowledge and thank the following teams and individuals who have made ORIGINS possible: ORIGINS team; Joondalup Health Campus (JHC); members of the ORIGINS Community Reference and Participant Reference Groups; Research Interest Groups and the ORIGINS Scientific Committee; The Kids Research Institute Australia; City of Wanneroo; City of Joondalup; and Professor Fiona Stanley. ORIGINS has received core funding support from the Telethon Perth Children's Hospital Research Fund, Joondalup Health Campus, the Paul Ramsay Foundation, and the Commonwealth Government of Australia through the Channel 7 Telethon Trust. Substantial in-kind support has been provided by The Kids Research Institute Australia and Joondalup Health Campus.

S.M.S. received funding from the National Health and Medical Research Council (project grant number NHMRC115648) to support the AERIAL study and is also recipient of a NHMRC Investigator grant (NHMRC 2007725). A.K. is a Rothwell Family Fellow. T.I. is supported by a Stan Perron Charitable Foundation People Fellowship. S.P.A.-R. received funding from the Branchi family. A.B. is supported by the NIH (R21 AI176305-01A1, R01AI099108-11A1).

Author Contributions: J.A.C.-M., E.K.-S., P.A.-R., D.G.H., T.I., Y.V.K., D.J.M., G.Z., L.T., D.T.S., A.B., S.L.P., P.N.L., A.K. and S.M.S participated in the conceptualization of the study. J.A.C.-M. optimized the extraction protocol and

performed the extraction and quantification of the microbial DNA. J.A.C.-M. and P.A.-R. carried out the computational analyses. J.A.C.-M. analyzed the data. E.K.-S. coordinated the logistics of AERIAL study including collection of the swab samples and associated metadata. E.K.-S. coordinated accession to the stored samples and extracted the associated metadata. D.G.H. curated the metadata associated with this specific study. S.M.S. conceived the AERIAL study and obtained the funding to support this study. J.A.C.-M. prepared the figures and the original draft for publication. All the authors participated in critical discussions about the results presented in this manuscript as well as in the revision and proof reading of the manuscript.

Figure Legends.

Figure 1. A. The bar plot shows the overall relative abundance (%) of the top 10 ASVs in our dataset. **B**. Heatmap summarizing the Spearman correlation analysis between the centered log-ratio transformed counts of the top three ASVs and the two most frequently detected viruses. Heatmap colors indicate the direction of the correlation—positive (red) or negative (blue)—with color intensity reflecting the Spearman's ρ (rho) coefficient. Both the correlation coefficient and the FDR-adjusted *p*-value are shown within each cell. P-values were corrected for multiple comparisons using the false discovery rate (FDR) method. **C-D**. Boxplots representing alpha diversity (**C**, Shannon index) or bacterial load (**D**) in symptomatic swabs based on the detected virus. RHV: rhinovirus. Boxplots are overlaid with density plots (blue), and individual samples are shown as red dots. P-values for the indicated comparisons were calculated using the Wilcoxon rank-sum test.

Figure 2. Boxplots representing the bacterial load (**A**, **C**) and alpha diversity ecological parameters (**B**, **D**) associated with DNA extracts from nasal swabs in this study. In panels **C** and **D**, boxplots represent paired nasal swab specimens, with samples obtained from the same infant at scheduled visits (background) and during the subsequent symptomatic episode connected by straight lines.

Figure 3. A. Principal component analysis illustrating the linear projection of the compositional profiles of the microbiota onto the first and third component of the

model, with each sample labelled according to its DMM-model based endotype membership. **B.** Heatmap summarizing the microbial profiles grouped by Dirichlet component. Rows represent the indicated ASVs ordered by their contribution to each component from top to bottom. For simplicity only the 10 top ASVs are represented. The color intensity in the heatmap corresponds to the square root of each ASV's relative abundance. Narrow columns represent individual swab samples, while wide columns show the mean abundance of each ASV within the Dirichlet components. **C.** Boxplots representing alpha diversity (Shannon index) in relation to the two microbial community types or endotypes represented in (**A**). **D.** Predicted age at first wheeze (in months) with 95% confidence intervals, stratified by nasal microbial endotypes. The slope estimate, adjusted R^2, and *F*-test results of the fitted linear model are shown at the top of the figure. **E.** Number of wheezing episodes in relation to the predicted values from the negative binomial model. The grey points represent individual observed data points, while the bars display the mean observed values for each endotype by sex, with error bars indicating the standard error of the mean. The red diamonds represent the predicted number of wheezing episodes from the negative binomial model. Sample size for each group (N) is shown at the base of each bar. Means were compared using z-tests with FDR adjustment for multiple comparisons. A sensitivity analysis associated with this panel is shown in supplemental Figure S11.

Table 1. Patient demographics and swab characteristics. HMPV, human metapneumovirus; RSV, respiratory syncytial virus. *, four SARS-CoV-2–positive swabs also tested positive for other viruses. Similarly, four rhinovirus-positive swabs and one parainfluenza-positive swab were co-infected with additional viruses. All adenovirus-positive swabs were detected in combination with other viruses.

Characteristics	**Total number of infants n=55**
Sex, n (%)	
Male	25 (46%)
Female	30 (54%)
Nasal swabs type, n (%)	

Scheduled visits (background)	41 (45%)
Symptomatic swabs	49 (55%)
Background swabs age (months), n (%)	
3 months of age	24 (58.5%)
6 months of age	15 (36.6%)
9 months of age	2 (9.4%)
Viral detection by swab type, n (%)	
Scheduled visits (background)	2 (4.8%)
Symptomatic swabs	48 (98%)
Number of swabs by virus detected, n (%)	
*Rhinovirus **	28 (56%)
*SARS-CoV-2 **	23 (46%)
*Adenovirus **	2 (4%)
*Parainfluenza **	2 (4%)
HMPV	1 (2%)
RSV	1 (2%)

References:

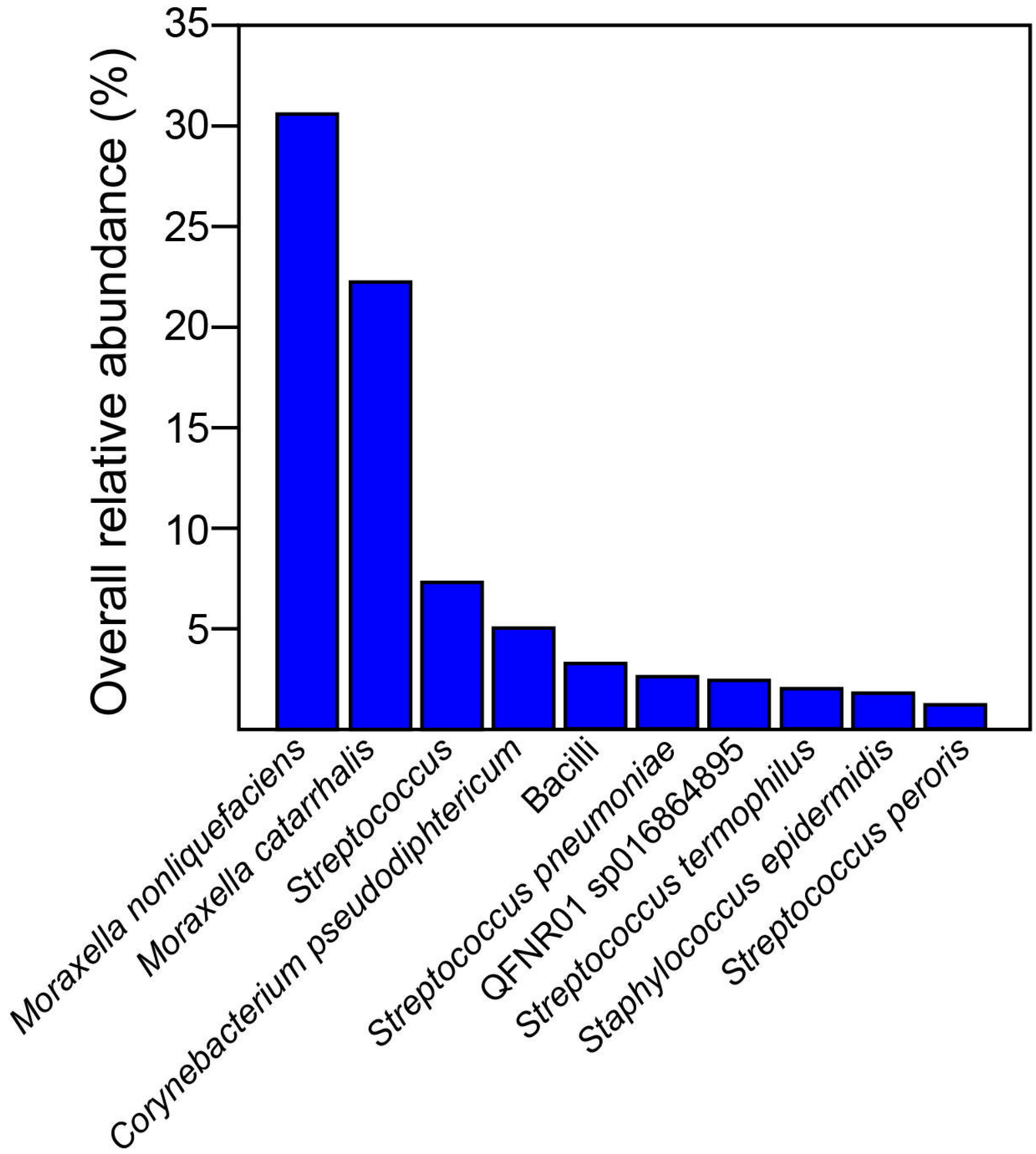
1. United Nations Inter-agency Group for Child Mortality Estimation (UN IGME), 'Levels & Trends in Child Mortality: Report 2019, Estimates developed by the United Nations Inter-agency Group for Child Mortality Estimation' [Internet]. United Nations Children's Fund. 2019 [cited August 2024]. Available from: https://www.unicef.org/media/60561/file/UN-IGME-child-mortality-report-2019.pdf.
2. van Meel ER, Mensink-Bout SM, den Dekker HT, Ahluwalia TS, Annesi-Maesano I, Arshad SH, et al. Early-life respiratory tract infections and the risk of school-age lower lung function and asthma: a meta-analysis of 150 000 European children. Eur Respir J. 2022;60(4).
3. Kusel MM, Kebadze T, Johnston SL, Holt PG, Sly PD. Febrile respiratory illnesses in infancy and atopy are risk factors for persistent asthma and wheeze. Eur Respir J. 2012;39(4):876-82.
4. Global Initiative for Asthma. Global Strategy for Asthma Management and Prevention, 2023. Updated July 2023. Available from: www.ginasthma.org
5. de Steenhuijsen Piters WAA, Watson RL, de Koff EM, Hasrat R, Arp K, Chu M, et al. Early-life viral infections are associated with disadvantageous immune and microbiota profiles and recurrent respiratory infections. Nat Microbiol. 2022;7(2):224-37.
6. Teo SM, Tang HHF, Mok D, Judd LM, Watts SC, Pham K, et al. Airway Microbiota Dynamics Uncover a Critical Window for Interplay of Pathogenic Bacteria and Allergy in Childhood Respiratory Disease. Cell Host Microbe. 2018;24(3):341-52 e5.

7. Wang L, Quinn B, Duliman J, Denn E. Maternal nasal microbiota seed the offspring nasal microbiota in early infancy. Journal of Allergy and Clinical Immunology. 2024;153(2).
8. McCauley KE, DeMuri G, Lynch K, Fadrosh DW, Santee C, Nagalingam NN, et al. Moraxella-dominated pediatric nasopharyngeal microbiota associate with upper respiratory infection and sinusitis. PLoS One. 2021;16(12):e0261179.
9. Kloepfer KM, Lee WM, Pappas TE, Kang TJ, Vrtis RF, Evans MD, et al. Detection of pathogenic bacteria during rhinovirus infection is associated with increased respiratory symptoms and asthma exacerbations. J Allergy Clin Immunol. 2014;133(5):1301-7, 7 e1-3.
10. Luna PN, Hasegawa K, Ajami NJ, Espinola JA, Henke DM, Petrosino JF, et al. The association between anterior nares and nasopharyngeal microbiota in infants hospitalized for bronchiolitis. Microbiome. 2018;6(1):2.
11. Bisgaard H, Hermansen MN, Buchvald F, Loland L, Halkjaer LB, Bonnelykke K, et al. Childhood asthma after bacterial colonization of the airway in neonates. N Engl J Med. 2007;357(15):1487-95.
12. Teo SM, Mok D, Pham K, Kusel M, Serralha M, Troy N, et al. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. Cell Host Microbe. 2015;17(5):704-15.
13. Biesbroek G, Tsivtsivadze E, Sanders EA, Montijn R, Veenhoven RH, Keijser BJ, et al. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. Am J Respir Crit Care Med. 2014;190(11):1283-92.
14. Yu X, Wang L, Zheng, X, Wen Y, Zhang Z, Fan L, et al. Moraxella occupied the largest proportion in the nasal microbiome in healthy children, which potential protect them from COVID-19. Microb Pathog. 2022; 170:105685.
15. van den Munckhof EHA, Hafkamp HC, de Kluijver J, Kuijper EJ, de Koning MNC, Quint WGV et al. Nasal microbiota dominated by Moraxella spp. Is associated with respiratory health in the elderly population: a case control study. Respir Res. 2020; 21(1):181.
16. Johnson JS, Spakowicz DJ, Hong BY, Petersen LM, Demkowicz P, Chen L, et al. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat Commun. 2019;10(1):5029.
17. Kicic-Starcevich E, Hancock DG, Iosifidis T, Agudelo-Romero P, Caparros-Martin JA, Karpievitch YV, et al. Airway epithelium respiratory illnesses and allergy (AERIAL) birth cohort: study protocol. Front Allergy. 2024;5:1349741.
18. Silva DT, Hagemann E, Davis JA, Gibson LY, Srinivasjois R, Palmer DJ, et al. Introducing the ORIGINS project: a community-based interventional birth cohort. Reviews on Environmental Health. 2020;35(3):281-93.
19. D'Vaz N, Kidd C, Miller S, Amin M, David JA, Talati Z, et al. The ORIGINS project biobank: a collaborative bio resource for investigating the developmental origins of health and disease. Int J Environ Res Public Health. 2023; 20(13):6297.
20. Hancock DG, Kicic-Starcevich E, Sondag T, Rivers R, McGee K, Karpievitch YV, et al. Real time monitoring of respiratory viral infections in cohort studies using a smartphone app. iScience. 2024; 27(10):110912.

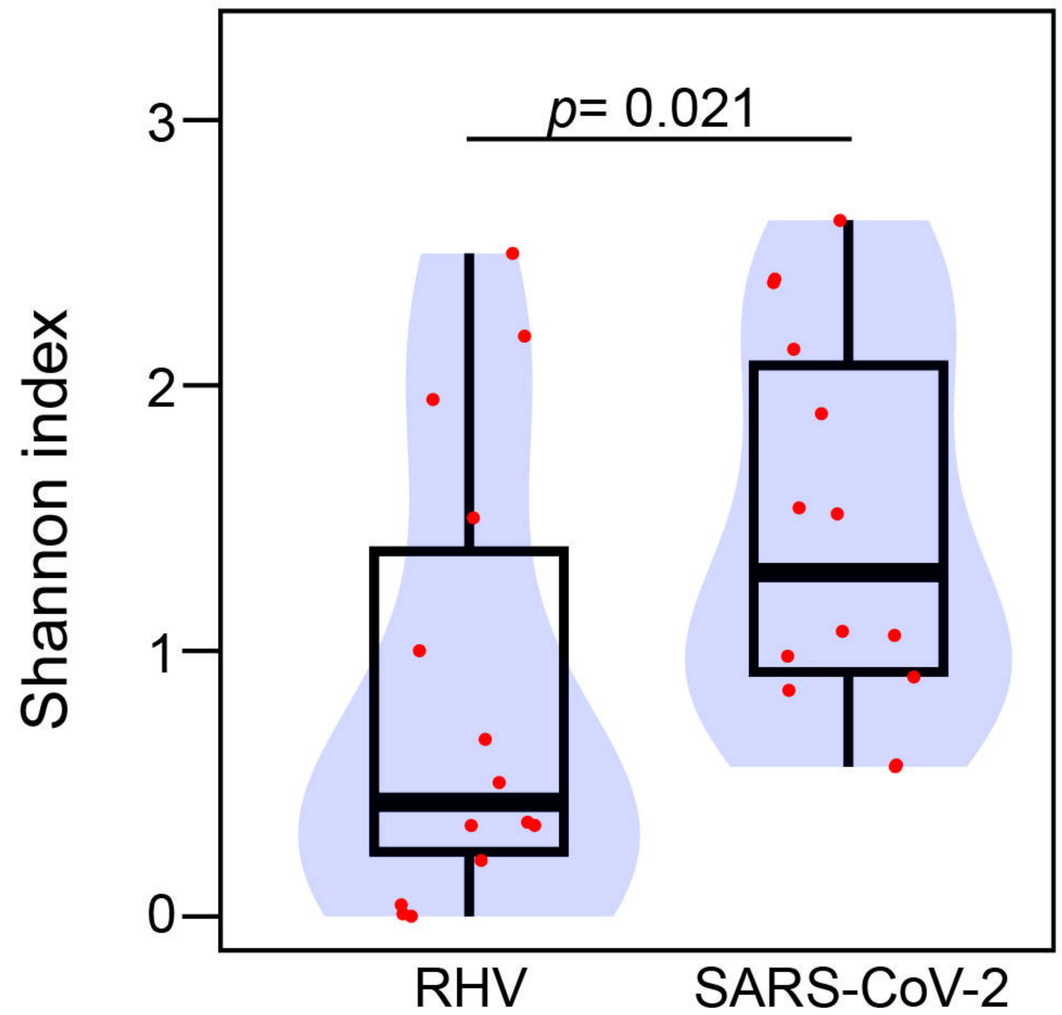
21. Caparros-Martin JA, Lareu RR, Ramsay JP, Peplies J, Reen FJ, Headlam HA, et al. Statin therapy causes gut dysbiosis in mice through a PXR-dependent mechanism. Microbiome. 2017;5(1):95.
22. Eisenhofer R, Minich JJ, Marotz C, Cooper A, Knight R, Weyrich LS. Contamination in Low Microbial Biomass Microbiome Studies: Issues and Recommendations. Trends Microbiol. 2019;27(2):105-17.
23. Saladie M, Caparros-Martin JA, Agudelo-Romero P, Wark PAB, Stick SM, O'Gara F. Microbiomic analysis on low abundant respiratory biomass samples; improved recovery of microbial DNA from bronchoalveolar lavage fluid. Front Microbiol. 2020;11:572504.
24. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581-3.
25. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017;35(4):316-9.
26. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol. 2007;73(16):5261-7.
27. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018;36(10):996-1004.
28. Caparros-Martin JA, Saladie M, Agudelo-Romero SP, Reen FJ, Ware SW, Sly PD, et al. Detection of bile acids in bronchoalveolar lavage fluid defines the inflammatory and microbial landscape of the lower airways in infants with cystic fibrosis. Microbiome. 2023;11:132.
29. 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.
30. Rohart F, Gautier B, Singh A, Le Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol. 2017;13(11):e1005752.
31. Holmes I, Harris K, Quince C. Dirichlet multinomial mixtures: generative models for microbial metagenomics. PLoS One. 2012;7(2):e30126.
32. Lin H, Peddada SD. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. Nat Methods. 2024;21(1):83-91.
33. Bosch AA, Biesbroek G, Trzcinski K, Sanders EA, Bogaert D. Viral and bacterial interactions in the upper respiratory tract. PLoS Pathog. 2013;9(1):e1003057.
34. Tse SM, Rifas-Shiman SL, Coul BA, Litonjua AA, Oken E, Gold DR. J Allergy Clin Immunol 2016; 138(6):1561-1568.
35. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012. 486(7402):207-214.
36. Allen EK, Koeppel AF, Hendley JO, Turner SD, Winther B, Sale MM. Characterization of the nasopharyngeal microbiota in health and during rhinovirus challenge. Microbiome. 2014;2:22.
37. Revai K, McCormick DP, Patel J, Grady JJ, Saeed K, Chonmaitree T. Effect of pneumococcal conjugate vaccine on nasopharyngeal bacterial colonization during acute otitis media. Pediatrics. 2006; 117(5):1823-1829.

38. Biesbroek G, Wang X, Keijser BJF, Eijkemans RMJ, Trzcinski K, Rots NY, et al. Seven-valent pneumococcal conjugate vaccine and nasopharyngeal microbiota in healthy children. Emerg Infect Dis. 2014; 20(2):201-210.
39. Korten I, Mika M, Klenja S, Kieninger E, Mack I, Barbani MT, et al. Interactions of Respiratory Viruses and the Nasal Microbiota during the First Year of Life in Healthy Infants. mSphere. 2016;1(6).
40. Dissanayake E, Brockman-Schneider RA, Stubbendieck RM, Helling BA, Zhang Z, Bochkov YA et al. Rhinovirus increases Moraxella catarrhalis adhesion to the respiratory epithelium. Front Cell Infect Microbiol. 2023. 12:1060748.
41. Thorsen J, Li XJ, Peng S, Sunde RB, Shah SA, Bhattacharyya M et al. The airway microbiota of neonates colonized with asthma-associated pathogenic bacteria. Nat Commun. 14(1):6668.
42. Wang Z, Locantore N, Haldar K, Ramsheh MY, Beech AS, Ma W et al. Inflammatory endotype-associated airway microbiome in chronic obstructive pulmonary disease clinical stability and exacerbations: a multicohort longitudinal analysis. Am J Respir Crit Care Med. 2021. 203(12):1488-1502.
43. Chow EJ, Uyeki TM, Chu HY. The effects of the COVID-19 pandemic on community respiratory virus activity. Nat Rev Microbiol. 2023;21(3):195-210.

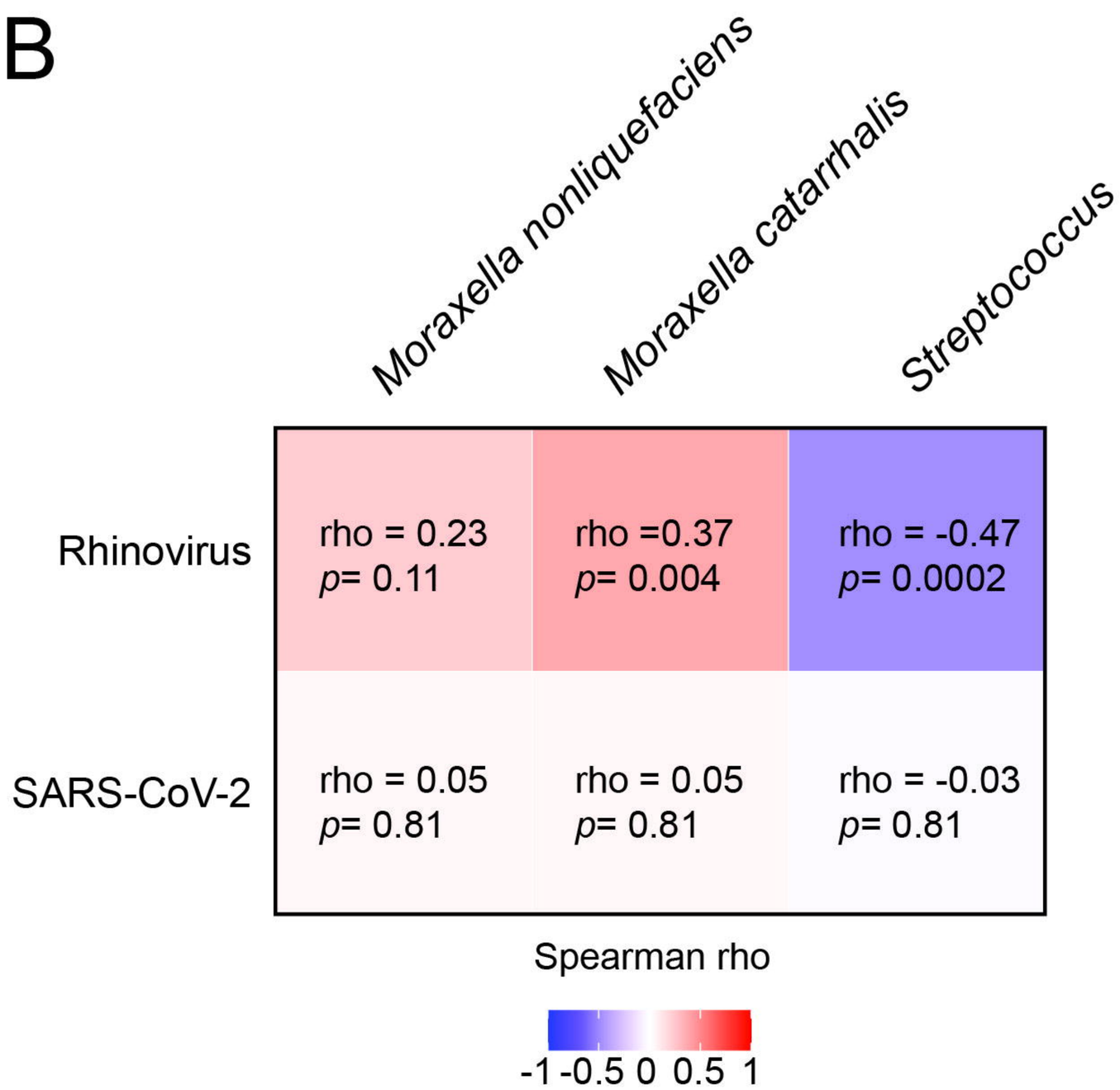
A


Overall relative abundance (%)
35
30
25
20
15
10
5
Moraxella nonliquefaciens
Moraxella catarrhalis
Streptococcus
Corynebacterium pseudodiphtericum
Bacilli
Streptococcus pneumoniae
QFNR01 sp016864895
Streptococcus termophilus
Staphylococcus epidermidis
Streptococcus peroris

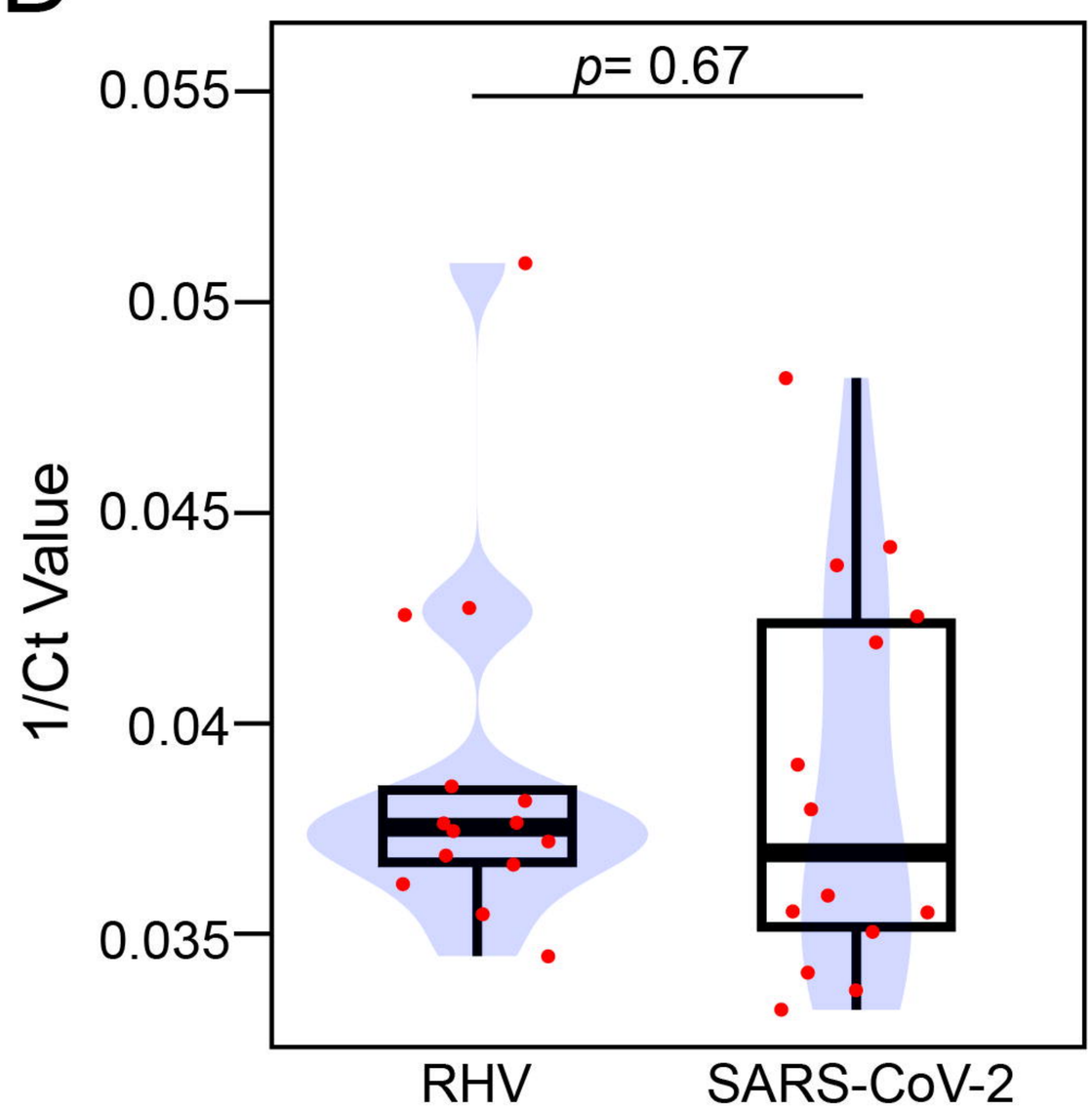
C

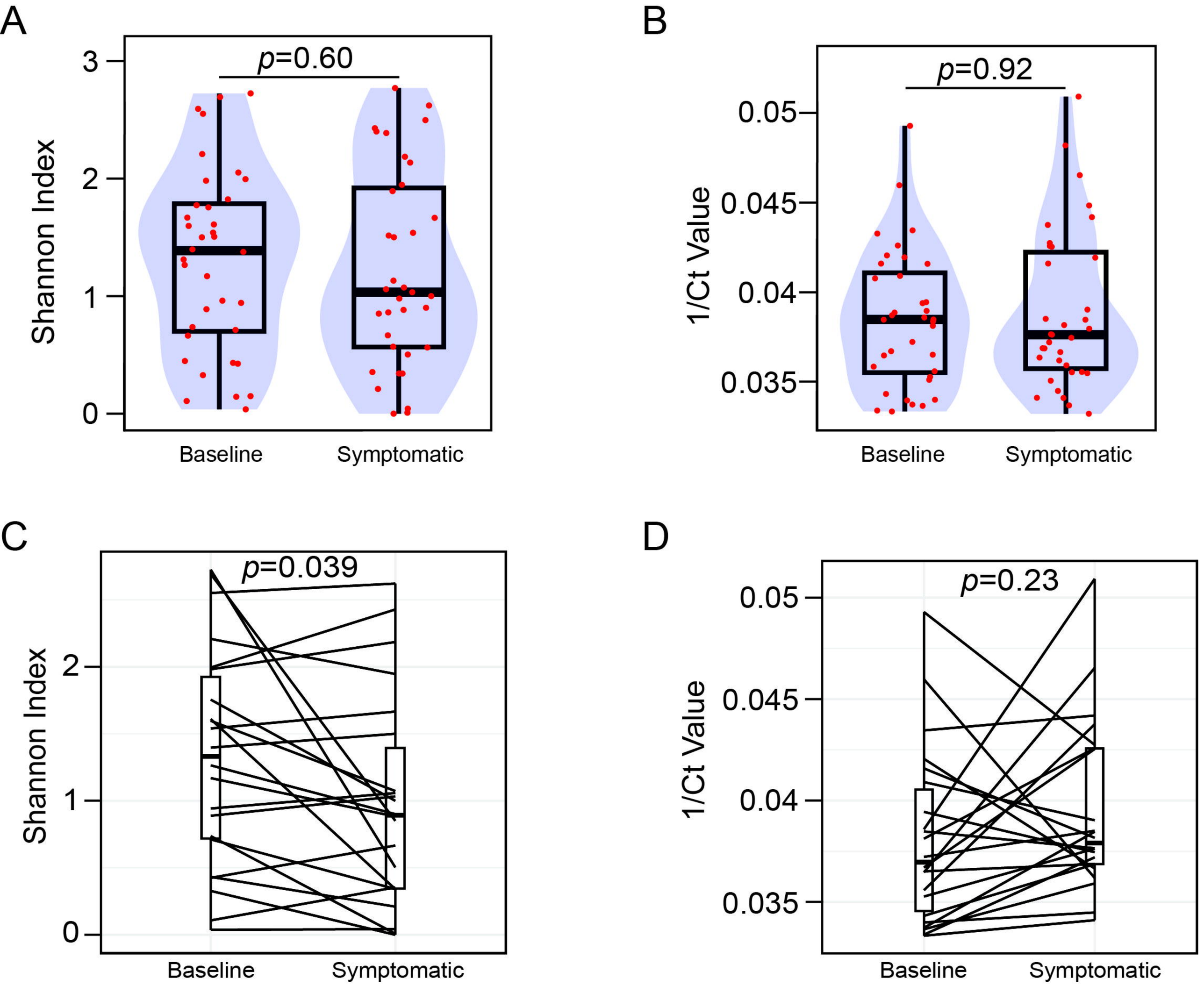

Shannon index
3
2
1
0
p= 0.021
RHV
SARS-CoV-2

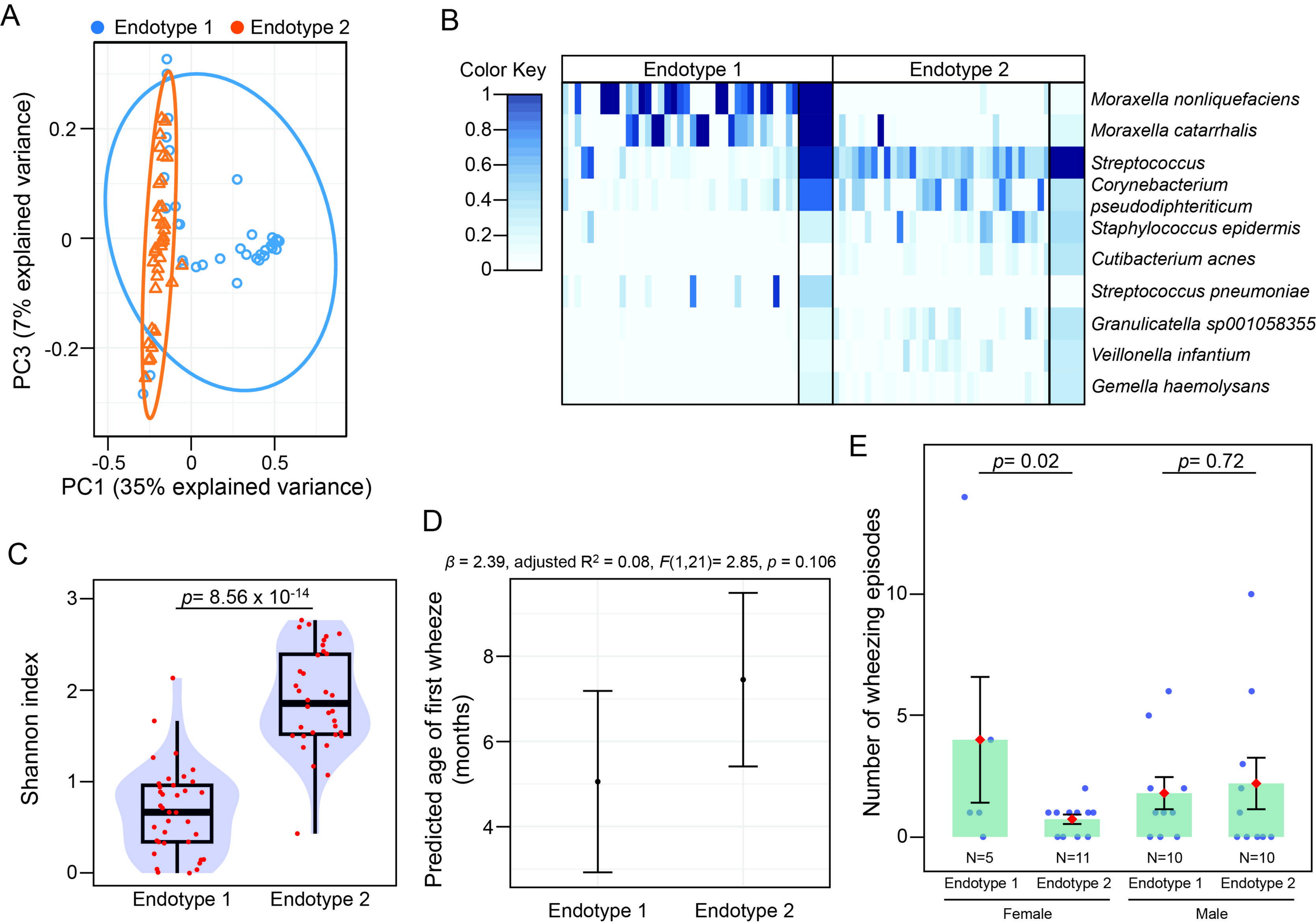
B


Moraxella nonliquefaciens
Moraxella catarrhalis
Streptococcus
Rhinovirus
rho = 0.23
p= 0.11
rho =0.37
p= 0.004
rho = -0.47
p= 0.0002
SARS-CoV-2
rho = 0.05
p= 0.81
rho = 0.05
p= 0.81
rho = -0.03
p= 0.81
Spearman rho
-1 -0.5 0 0.5 1

D


1/Ct Value
0.055
0.05
0.045
0.04
0.035
p= 0.67
RHV
SARS-CoV-2


A
p=0.60
Shannon Index
3
2
1
0
Baseline
Symptomatic
B
p=0.92
1/Ct Value
0.05
0.045
0.04
0.035
Baseline
Symptomatic
C
p=0.039
Shannon Index
2
1
0
Baseline
Symptomatic
D
p=0.23
1/Ct Value
0.05
0.045
0.04
0.035
Baseline
Symptomatic


A
Endotype 1
Endotype 2
PC3 (7% explained variance)
PC1 (35% explained variance)
0.2
0
-0.2
-0.5
0
0.5
B
Color Key
1
0.8
0.6
0.4
0.2
0
Endotype 1
Endotype 2
Moraxella nonliquefaciens
Moraxella catarrhalis
Streptococcus
Corynebacterium pseudodiphteriticum
Staphylococcus epidermis
Cutibacterium acnes
Streptococcus pneumoniae
Granulicatella sp001058355
Veillonella infantium
Gemella haemolysans
C
p= 8.56 x 10-14
Shannon index
3
2
1
0
Endotype 1
Endotype 2
D
β = 2.39, adjusted R2 = 0.08, F(1,21)= 2.85, p = 0.106
Predicted age of first wheeze (months)
8
6
4
Endotype 1
Endotype 2
E
p= 0.02
p= 0.72
Number of wheezing episodes
10
5
0
N=5
N=11
N=10
N=10
Endotype 1
Endotype 2
Endotype 1
Endotype 2
Female
Male