Bacterial and viral respiratory tract microbiota and host characteristics in children with lower respiratory tract infections: a matched case-control study

Bacterial and viral microbiota, and host characteristics in children with lower respiratory tract infections: results from a matched case-control study

Wing Ho Man[1,2], Marlies A. van Houten[2,3], Marieke E. Mérelle[3], Arine M. Vlieger[4], Mei Ling J.N. Chu[1], Nicolaas J.G. Jansen[5], Elisabeth A.M. Sanders[1§], Debby Bogaert[1,6§]*

Affiliations:

[1] Department of Paediatric Immunology and Infectious Diseases, Wilhelmina Children’s Hospital / University Medical Center Utrecht, Utrecht, The Netherlands;

[2] Spaarne Gasthuis Academy, Hoofddorp and Haarlem, The Netherlands;

[3] Department of Pediatrics, Spaarne Gasthuis, Hoofddorp and Haarlem, The Netherlands;

[4] Department of Pediatrics, St. Antonius Ziekenhuis, Nieuwegein, The Netherlands;

[5] Department of Pediatric Intensive Care, Wilhelmina Children's Hospital/University Medical Center Utrecht, Utrecht, The Netherlands;

[6] Medical Research Council/University of Edinburgh Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom.

§ Joint senior authors

Preferred degree (one only):

W H Man MD, M A van Houten MD, M E Mérelle MD, A M Vlieger MD, M L J N Chu BSc, N J G Jansen MD, Prof E A M Sanders MD, Prof D Bogaert MD

*** Correspondence to:**

D. Bogaert, MD, PhD

Medical Research Council/University of Edinburgh Centre for Inflammation Research

Queen's Medical Research Institute, University of Edinburgh

47 Little France Crescent

EH16 4TJ, Edinburgh, United Kingdom
Email: D.Bogaert@ed.ac.uk
Tel: +44 131 2426582

Short title:

Microbiota as a marker for LRTIs.

Word count main text:

4,272 / 3,500

Word count abstract:

318 / 300

Summary

Background Lower respiratory tract infections (LRTIs) are a leading cause of childhood morbidity and mortality. Potentially pathogenic organisms are seen in both symptomatic and asymptomatic children and their presence does not per se indicate disease.

Methods To assess the concordance between upper and lower respiratory tract microbiota during LRTI, we first studied 29 children with a severe LRTI and obtained paired nasopharyngeal swabs and deep endotracheal aspirates (PICU cohort). In addition, we performed a case-control study on 154 children hospitalized with a LRTI, and 307 age-, gender-, and time-matched healthy children to study the use of nasopharyngeal microbiota in discriminating LRTI from health. Nasopharyngeal samples of cases were obtained at time of hospital admission and of controls during home visits. Child characteristics were obtained by questionnaires, pharmacy printouts and medical charts. We used qPCR and 16S rRNA-based sequencing to determine viral and bacterial microbiota profiles, respectively.

Findings In the PICU cohort, there was a high intra-individual concordance of viral (96% agreement; 95% CI 93-99%) and bacterial (Pearson's r=0·93; IQR 0·62-0·99, p<0·05) microbiota profiles between nasopharyngeal and endotracheal aspirate samples, supporting the use of nasopharyngeal samples as proxy for lung microbiota during LRTI. In our matched case-control cohort we found that either bacterial microbiota, viruses or child characteristics performed poorly in distinguishing health from disease (random forest classification, AUC 0·77, 0·70, and 0·80, respectively). However, a classification model based on combined bacterial and viral microbiota, plus child characteristics distinguished children with LRTI with high accuracy from their matched controls (AUC 0·92).

Interpretation Our data suggest that (i) nasopharyngeal microbiota may serve as a valid proxy for lower respiratory tract microbiota in childhood LRTI; (ii) clinical LRTI in children results from the interplay between microbiota and host characteristics, rather than a single microorganism; (iii) microbiota-based diagnostics may improve future diagnostic and treatment protocols.

Funding Spaarne Gasthuis, University Medical Center Utrecht, and The Netherlands Organization for Scientific Research (NWO-Vidi grant 91715359).

Research in context

Evidence before this study

Conventional culture-based studies have identified key pathogens for the development of lower respiratory tract infection (LRTI). While the overgrowth of these potential pathogens might partly explain the progression towards disease, pathogen colonization (the act of settlement and reproduction of pathogens) per se does not necessarily lead to infection (the acquisition of a microorganism that leads to damage to the host). We hypothesized that the entire nasopharyngeal microbial community might play a role in the susceptibility to and severity of LRTIs. We searched PubMed on May 1st, 2018, using the terms "(Child, Preschool[mh] OR Infant[mh] NOT Infant, Newborn[mh]) AND (Respiratory Tract Infections[mh] OR pneumonia[tiab] OR bronchiolitis[tiab] OR wheezing[tiab]) AND (microbiota[mh] OR microbiome[tiab]) AND (Case-Control Studies[mh] OR prospective[tiab])", with no language restrictions. We identified fourteen publications of which three reported on the role of the microbiota in acute LRTIs in children. One study demonstrated that specific microbiota profiles were associated with the development of respiratory infections in time and focused only on infants at high risk of atopy. A second study reported findings of a small (n=100) matched case-control study, but lacked matching on season, recruited a wide variety of children up to the age of 16 and more than half of their samples lacked sufficient bacterial DNA for further analysis. Therefore, it was underpowered to provide conclusive results. The third study included infants <1 year only and did not include a control group of healthy children. They showed that *Moraxella* was associated with less severe bronchiolitis and *Streptococcus* was associated with more severe bronchiolitis. Overall, the current evidence on the potential role of the microbial community in the pathogenesis and severity of LRTIs in children is limited.

Added value of this study

In our study, we used culture-independent techniques based on qPCR and 16S rRNA MiSeq sequencing of respiratory tract samples to ascertain first that the nasopharyngeal (viral and bacterial) microbiota can be used as proxy for lung microbiota in childhood LRTI, which is an important finding for future diagnostic approaches. Next, we demonstrate the relation between microbial community composition and susceptibility to and severity of LRTIs in children. Since we used a strictly matched case-control design in a well-powered cohort of 461 children, our study is the first to confidently demonstrate the association between microbiota, including viral presence, and LRTI in children. To our knowledge, the accuracy of our model in discriminating LRTI from health is unprecedented. Also, the phenotype-independent nature of the associations between respiratory microbiota and childhood LRTI has never been reported.

Implications of all the available evidence

Findings from this study broaden our knowledge on the likely important role of the complete nasopharyngeal ecosystem in the development and severity of LRTIs in young children. The excellent accuracy of our classifier model provides a premise for future microbiota-based diagnostic tools. Our data provide insights that may be critical for determining optimal therapeutic strategies, including targeted antibiotic treatment. The phenotype-independent associations during acute disease challenge conventional views on the role of viruses and bacteria in LRTI pathogenesis, especially the current dichotomy between bronchiolitis (viral origin) and pneumonia (bacterial origin). Our findings endorse studies to further explore microbiota-based diagnostics as a potential tool for clinical application in childhood LRTIs. This in turn may have major implications for the future treatment protocols.

Introduction

Lower respiratory tract infections (LRTIs) remain a major cause of morbidity and mortality in children worldwide.[1,2] While multiple host, environmental, and lifestyle factors have been recognized to increase susceptibility to LRTIs,[3] it remains unclear why some children remain asymptomatic upon pathogen exposure, while others develop severe disease.

Classically, a LRTI is caused by acquisition of potentially pathogenic viruses and bacteria (pathobionts) in the upper respiratory tract, that replicate and spread towards the lower respiratory tract where they invade the mucosa leading to inflammation and clinical disease.[4] Many pathobionts, however, are frequently encountered in the upper respiratory tract of healthy children.[5] We hypothesize that that a balanced microbial community protects against development of LRTI.[5]

Previous studies in children already demonstrated a relation between the bacterial microbiota composition of the nasopharynx and susceptibility to upper or lower respiratory infectious episodes over time.[6,7] We recently described that oral microbes like *Prevotella* and *Leptotrichia* spp. in the nasopharyngeal niche were strongly associated with subsequent development of upper respiratory tract infections (URTIs) in children and are more abundant at times of an URTI. In contrast, *Corynebacterium* and *Dolosigranulum* spp. were associated with resistance to symptomatic respiratory disease over time and less present during URTI episodes.[8] Additionally, in infants with respiratory syncytial virus (RSV) related LRTI, a strong correlation was observed between the presence of *Haemophilus influenzae* and *Streptococcus pneumoniae* and the severity of host inflammation, suggesting an important role for the complete microbial community in the upper respiratory tract and the symptomatology of clinical disease.[9]

No study has yet addressed the relationship between the nasopharyngeal microbiota community and the presence, clinical symptoms and severity of childhood LRTI in a proper case-controlled manner. Moreover, nasopharyngeal microbiota profiling was never studied in the context of classification of states of health and disease. We, therefore, conducted a prospective strictly matched case-control study in young children hospitalized for a LRTI. We set out to demonstrate (i) the association between upper and lower respiratory tract microbiota during childhood LRTI; (ii) the use of microbiota to predict LRTI presence and severity; and (iii) associations between microbiota and disease across different clinical presentations of LRTI, i.e. pneumonia, bronchiolitis and wheezing illness.

Methods

Details on the methods can be found in the appendix. Data have been deposited in the NCBI Sequence Read Archive database (BioProject ID PRJNA428382). This study conforms to the STROBE guidelines for reporting case-control studies (**Supplementary Table 3**).[10]

Study design and procedures

We first conducted a prospective study from September 2013 to June 2015 enrolling 29 patients aged 4 weeks to 5 years who became hospitalized at the pediatric intensive care unit of a university hospital (Wilhelmina Children's Hospital, Utrecht) for a WHO-defined LRTI[11] requiring mechanical ventilation (PICU cohort; **Supplementary Figure 1**; patient characteristics: **Supplementary Table 1**). Nasopharyngeal swabs and endotracheal aspirates were obtained within four hours after intubation by trained nurses.

Next to the PICU cohort, we conducted a prospective, matched case-control study from September 2013 to June 2015, recruiting 154 cases under the same inclusion and exclusion criteria as the PICU cohort. Cases were recruited from three Dutch teaching hospitals (Spaarne Hospital, Hoofddorp; Kennemer Gasthuis, Haarlem; and St. Antonius Hospital, Nieuwegein). For each case, two age, gender, and time-matched healthy controls from the community were recruited, with the exception of one case for whom only one matched healthy control could be recruited from the database (**Supplementary Figure 1**; baseline characteristics in **Table 1**). Nasopharyngeal swabs were taken of cases generally within 1 hour after admission, and of controls during a home visit.

Of all children, extensive data regarding medical history and data on demographic, lifestyle and environmental characteristics were obtained. Both studies were approved by the Dutch National Ethics Committee. Written informed parental consent was obtained from all participants.

Two expert pediatricians independently classified all cases of the case-control cohort in three major disease phenotypes, i.e. pneumonia, bronchiolitis, and wheezing illness. Cases with a mixed or unclear phenotype were deemed ‘mixed’ phenotype. The expert panelists classified based on the entire medical record, including all clinical notes at and during admission, laboratory assessments and imaging.

Microbiota analysis

Bacterial DNA was isolated from samples as previously described.[12] Amplification of the V4 hypervariable region of the 16S rRNA gene was performed using barcoded universal primer pair 533F/806R. Amplicon pools were sequenced using the Illumina MiSeq platform (San Diego, CA, USA) and processed in our bioinformatics pipeline as previously described.[8] To avoid OTUs with identical annotations, we refer to OTUs using their taxonomical annotations combined with a rank number based on the abundance of each given OTU. In addition, viral profiles were determined using qualitative multiplex realtime-PCR (RespiFinder® SMARTfast 22) and identification of *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Haemophilus influenzae, and Moraxella catarrhalis* was done by qPCR.

Statistical analysis

Data analysis was performed in R v3·2. All analyses assessing matched samples accounted for the matched nature of the samples. A p-value of less than 0·05 or a Benjamini-Hochberg (BH) adjusted q-value of less than 0·05 was considered statistically significant. Statistical significance of the differences in baseline characteristics and viral detection was calculated with conditional logistic regression. Nonmetric multidimensional scaling (NMDS) plots were based on Bray-Curtis dissimilarity matrices and statistical significance was calculated by *adonis* (vegan). Host characteristics associated with microbiota composition were evaluated with a stepwise selected distance-based redundancy analysis,[13] and projected in NMDS plots using *envfit* (vegan). Hierarchical clustering was performed as described previously.[9] Random forest analyses were used to determine biomarker species that most discriminate between clusters (*VSURF*).[14] We used *metagenomeSeq* and cross-validated *VSURF* analysis to identify specific microbial taxa associated with cases or controls.[15] Sparse random forest classifier analyses were performed on the bacterial data, viral data, metadata, and the combination of all three datasets. Performance of these classifiers was evaluated by calculating the area under the receiver operating characteristic (ROC) curve (AUC) using the out-of-bag predictions for classification (pROC) as previously described.[16] Since the potential real-world application of these classification models requires a more robust determination of biomarker bacteria, we also build the classification models using merged OTU on genus-level data. A cross-validated sparse random forest prediction model was built to investigate to what extend hospitalization duration could be predicted with all available data (caret). Above analyses were carried out for the entire case-control cohort and were in part repeated for each of the phenotypes independently. Additionally, in a post-hoc fashion, as a measure of disease severity, we stratified the cases according to the physicians' judgment whether antibiotics were needed during admission (Dutch pediatricians generally restrict antibiotic treatment to children with clinically more severe LRTI) and performed separate analyses accordingly.

Role of the funding source

The funding sources had no role in the design, execution, analyses, or interpretation of the data of this study. The corresponding author had full access to the data final responsibility for the decision to submit for publication.

Results

Nasopharyngeal microbiota profiles correlate with lower respiratory tract microbiota during childhood LRTI

To assess whether during acute LRTI in childhood the nasopharynx microbiota serves as a valid proxy for the lower respiratory tract microbiota, we first analyzed our PICU cohort. Viral presence in paired nasopharyngeal and endotracheal aspirates, were in almost full agreement (96%; 95% CI 93-99%).

Bacterial microbiota of paired samples showed good concordance in composition (median within Bray-Curtis similarity 0·61) versus low concordance for between individual findings (median inter-individual BC similarity 0·10, p<0·001; **Supplementary Figure 2A**). Moreover, we observed a significantly correlated Shannon diversity (Pearson's r 0·66, p<0·001). In addition, we observed that 58 taxa (combined relative abundance of 80·1%) were strongly correlated in the paired samples (median Pearson's r=0·93, IQR 0·62-0·99, p<0·05); only three common members of nasopharyngeal microbiota *Staphylococcus*, *Corynebacterium* and *Dolosigranulum* were almost exclusively present in nasopharyngeal samples while absent in endotracheal aspirates (Pearson's r<0·20, p>0·50; **Supplementary Figure 2B**). Vice versa, we could identify no taxa from the endotracheal samples that were not present in the nasopharynx. When assessing whether there were differences in the relative abundance for individual taxa between nasopharyngeal samples and endotracheal aspirates, we only found a significant result for *Corynebacterium propinquum* (Kruskal-Wallis test, Benjamini-Hochberg adjusted q=0·004), *Corynebacterium macginleyi/accolens* (q=0·019), *Dolosigranulum pigrum* (q=0·003), and three very low abundant taxa (median relative abundance <0·1%). The concordance did not depend on antibiotic treatment before sampling (n=5/29, 27%), the clinical suspicion of a bacterial infection (n= 20/29, 69%) or) or culture-confirmed bacterial infection (n=16/29, 55%).

Host, lifestyle and environmental factors are associated with risk of disease

In our separate, prospectively enrolled, matched case-control cohort, 40% were female and the median age was 13·6 months (IQR, 4·9 - 27·4). Cases had a history of more parental-reported RTIs, more wheezing symptoms, more recent antibiotic use, and more tobacco smoke exposure as compared to controls. Controls were breastfed for at least 3 months more often than cases, and the education level of parents of controls was higher than that of cases (all p<0·05, **Table 1**).

Host characteristics associated with microbial ecology in the healthy controls

In our control cohort, respiratory microbiota composition was significantly associated with month of sampling (*adonis*, R^2=6·2%) and age (R^2=4·2%), followed by day-care attendance, breastfeeding, a history of parental-reported RTIs, and previous antibiotic treatment within the last 6 months (all p<0·05; **Supplementary Figure 3**). Gender was not correlated with microbiota composition.

Viral and bacterial profile differences between cases and controls

We detected one or multiple viruses in 97·1% of cases and 82·5% of controls (p<0·001; **Figure 1**), with a mean of 1·6 and 1·4 viruses/sample in cases and controls, respectively (p=0·04). The most commonly detected viruses were rhinovirus (62%), coronaviruses (18%), respiratory syncytial virus (17%), and adenoviruses (17%). Influenza was relatively rare (8%). In LRTI cases, we observed 10 times more often RSV (49% vs. 4·0%, p<0·001), and more human metapneumovirus (hMPV; 6·1% vs. 1·7%, p=0·022). Rhinovirus was more often detected in controls (67·3% vs 49·7%, p<0·001).

With respect to bacterial microbiota, although cases did not have a higher bacterial biomass than controls (p=0·28), we observed a significant difference in overall microbiota composition between cases and controls (*adonis*, R^2=3·1%, p<0·001; **Figure 2A**). Projection of the vectors for host characteristics associated with microbiota composition showed that previous antibiotic use in the past six months, recent bronchodilator use, and a parental-reported history of RTIs pointed in the direction of disease (*envfit*; **Figure 2B**).

We observed seven distinct microbiota profiles within the cases and controls (hierarchical clustering; **Supplementary Figure 4A**). The profiles dominated by *Haemophilus influenzae/haemolyticus* (30·0% of samples) and *Streptococcus pneumoniae* (6·1%) were significantly related with LRTI cases, whereas profiles dominated by *Moraxella catarrhalis*/*nonliquefaciens* (47·3%), and *Corynebacterium propinquum/Dolosigranulum pigrum* (9·3%) were significantly associated with health (chi-square tests, p<0·05; **Supplementary Figure 4B**). A posteriori plotting of the biomarker species of these clusters in the NMDS ordination further supported the above associations between profiles and health or disease (**Figure 2A**). The profile dominated by *H. influenzae/haemolyticus* (median 126 pg/ul, **Supplementary**

Figure 4C) had a significant higher bacterial load than the other profiles (Wilcoxon rank-sum test, p<0·05), and a trend towards higher loads compared to *S. aureus/epidermidis* dominated profile (median 82 pg/ul, p= 0·25). The bacterial load of the *Streptococcus pneumoniae* dominated profile (median 67 pg/ul) was significantly higher than that of the *C. propinquum* & *D. pigrum* dominated profile (median 15 pg/ul, p=0·002) though did not differ from that of the *M. catarrhalis/nonliquefaciens* dominated profile (median 35 pg/ul, p=0·40).

On individual bacterial taxon level, we observed 49 taxa that differentiated cases from controls (*metagenomeSeq*, mean combined relative abundance 83·5%), which was confirmed for 17 of these bacteria by cross-validated random forest analysis (**Supplementary Figure 5A**). Among these, we observed a higher abundance of *H. influenzae/haemolyticus*, *S. pneumoniae*, *Actinomyces* spp., and *Prevotella* spp. in LRTI cases, while we observed a higher abundance of different *Moraxella* spp., *C. propinquum*, *D. pigrum*, and *Helcococcus* in controls.

Clinical presentation independent viral and bacterial differences between cases and controls

The classification of clinical phenotypes by an expert panel resulted in 37 cases of pneumonia, 57 cases of bronchiolitis, and 34 cases of wheezing illness. The remaining 26 cases were regarded as mixed phenotype (patient characteristics stratified per phenotype: **Supplementary Table 2**). RSV presence was predominant among all LRTI cases irrespective of phenotype, i.e. in 62% of bronchiolitis cases versus 3·6% in their matched controls (p<0·001), in 56% of pneumonia cases versus 4·1% in controls (p<0·001), in 58% of mixed-phenotype cases versus 1·9% in controls (p<0·001), and in 15% of wheezing illness cases versus 6·1% in controls (p=0·15; **Supplementary Figure 6A**). Rhinovirus was equally or less frequently detected in cases relative to controls. hMPV was only found in pneumonia and bronchiolitis cases.

When we stratified per clinical phenotype, we again showed that the overall bacterial microbiota composition was significantly different between cases and controls for each phenotype (*adonis*, all p<0·01; **Supplementary Figure 7**). The differential abundance of individual microbes between cases and controls was highly similar for each phenotype (**Supplementary Figure 5C-E**). In all phenotypes there was overrepresentation of *Haemophilus*, *Neisseria* and oral taxa, such as *Actinomyces*, and underrepresentation of multiple *Moraxella*, *Dolosigranulum*, and *Helcococcus* spp. The phenotype-independent differences in microbiota composition between cases and controls were further strengthened by the results of the mixed-phenotype group, which largely overlapped that of the three other phenotypes (**Supplementary Figure 5F**).

Combined importance for disease

Combining viral and bacterial biomarkers with host factors in a sparse random forest analysis resulted in a very high classification accuracy of LRTI versus health (AUC 0·92; **Figure 3A**). The most important set of predictors of disease were the presence of RSV, a high abundance of *H. influenzae/haemolyticus* and *S. pneumoniae*, and low abundance of several *Moraxella* spp., together with recent antibiotic treatment, lack of breastfeeding and history of RTIs (**Figure 3B)**. The combined classifier outperformed the models built on bacterial microbiota alone (AUC 0·77), viruses alone (AUC 0·70), child characteristics alone (AUC 0·80) or the model including only the two classically most important pathobionts, i.e. RSV and *S. pneumoniae* (AUC 0·75). External validation of our classifier model on the samples of the PICU cohort, demonstrated a correct classification in 92% when testing on nasopharyngeal samples, and a correct classification in 100% when testing on endotracheal aspirates.

Separate models for each of the phenotypes showed equally high accuracy in classifying LRTI (AUC 0·90-0·94; **Figure 3A, C-F**).

To test more broad and universally applicable classification models using bacterial microbiota data clustered on genus-level instead of OTU-level, we demonstrated again a very high classification accuracy of the presence of LRTI versus health (entire cohort AUC 0·92; phenotype-specific AUC 0·86-0·94; **Supplementary Figure 8**).

Microbiota and severity of disease

In post-hoc analyses we attempted to see whether a similar classification model could also predict severity of disease, which was a performed as stratified analyses within the LRTI group only. As a first measure, we studied whether or not the physician decided to start antibiotic treatment, after sampling of the nasopharynx, which occurred in 43/154 cases (28%); most for pneumonia cases (29/37, 78%) and few for bronchiolitis (4/57, 7%), wheezing illness (4/37, 12%), and mixed infection cases (6/26, 23%). Upon admission, the to-be-treated cases showed no differences in viral presence compared to the not-to-be-treated cases (**Supplementary Figure 6B**). With respect to bacterial ecology, we observed similar though slightly more pronounced differences in microbiota compositions when compared to the matched controls in the to-be-treated cases compared to the not-to-be-treated cases (*adonis*, R^2=5·8% and R^2=2·6%, respectively, both p<0·001; **Supplementary Figure 9**). Antibiotic treatment prescription (following sampling), however, was not associated with increased abundance of pathobionts such as *H. influenzae/haemolyticus* or *S. pneumoniae*. Instead, there was a higher abundance of oral taxa, such as *Veillonella*, *Prevotella*, and *Actinomyces* spp. in the to-be-treated cases compared to the not-to-be-treated cases (**Supplementary Figure 5G-H**).

As a second measure of severity, we studied hospitalization duration, as a second measure of disease severity: this could be predicted fairly accurately at admission by a random forest model including 14 viral, bacterial and host characteristics (Pearson’s r 0·50, p<0·001; **Supplementary Figure 10**). Predictors from highest to lowest importance were younger age, abundance of *C. propinquum*, *Neisseria*, *S. aureus/epidermidis*, *S. thermophilus*, *Veillonella*, *P. melaninogenica*, and other *Streptococci*, followed by disease phenotype, abundance of *Atopobium*, *Lactobacillales*, presence of RSV, absence of HRV, and abundance of *Leptotrichia* (**Supplementary Figure 10A**). When stratifying these data only for the not-to-be-treated group, the prediction of hospitalization duration at admission became stronger (Pearson’s r 0·55, p<0·001; **Supplementary Figure 10C**). For the to-be-treated group separately, the predictive capacity of the model was lost (p=0·73) suggesting interference of antibiotics with natural recovery.

As a third measure of severity, we analyzed the nasopharyngeal data of our PICU cohort in relation to matched with an age and season-matched subset of our case-control cohort. As expected, the overall microbiota compositions of the PICU cases demonstrated a similar but more pronounced shift from healthy controls compared to that of the (moderate-severe) cases from the case-control cohort (*adonis*, R^2=5·6% and R^2=4·2% for PICU versus case-control cohort, respectively; both p<0·001; **Supplementary Figure 11A**). Moreover, the PICU cases demonstrated an even more pronounced overrepresentation of several *Haemophilus*, *Streptococcus* (including *S. pneumoniae*), *Veillonella* and *Actinomyces* spp., and a more pronounced underrepresentation of multiple *Moraxella*, and especially *Dolosigranulum* and *Corynebacterium* spp. when compared to healthy controls (**Supplementary Figure 5B** and **Supplementary Figure 11B**).

Discussion

The upper respiratory tract microbiome is generally regarded the source community for the lower respiratory tract during LRTI in childhood,[5] although this has rarely been proven, certainly not in young children with LRTI. Here, we show that in line with literature there is a high intra-individual concordance of viral[17,18] and bacterial[19,20] microbiota profiles between nasopharyngeal and endotracheal aspirate samples in LRTI cases admitted to a PICU. The Bray-Curtis similarity of 0·61 approximates that of biological replicates of microbiota profiles of the lungs (i.e. two sequentially obtained lavages from the same lung lobe of the same child).[20] This suggests that the upper respiratory microbiota is not only the source community of the lower respiratory tract, but also that, except for a few commensal species, microbial colonization and proliferation in the nasopharynx parallels that of the lower airways during childhood LRTI. Therefore, our findings support the idea that upper respiratory tract samples can be used as proxy for lung microbiota in childhood LRTI.

Next, in our unselected, strictly matched case-control cohort, we demonstrate a strong association between nasopharyngeal microbiota composition and the presence of childhood LRTIs. Viral presence was ubiquitous in both cases and controls, with in particular RSV and to a lesser extent hMPV highly overrepresented in cases, in line with studies evaluating the viral etiology of childhood LRTIs.[21] The presence and abundance of *Haemophilus* spp., *S. pneumoniae* and oral species were strongly associated with disease, in line with previous reports linking these taxa to susceptibility to and severity of RTIs in children.[6,7,22,23] In contrast, the abundance of potentially beneficial bacteria like *Moraxella*, *Corynebacterium*, *Dolosigranulum*, and *Helcococcus* spp. were underrepresented in cases, in line with previous reports connecting these genera with prevention of infections.[6,7,12,24] By combining viral, bacterial and host related predictors, we found that children with LRTIs can be uniquely differentiated from strictly matched healthy controls, while far less by the individual predictors. This underlines the multifactorial pathophysiology of childhood LRTI. The contribution of the nasopharyngeal microbiota, both bacterial and viral, appears largely independent of the clinical presentation, and even holds for bronchiolitis and wheezing illness that are generally assumed to be of viral etiology.

Our results in the case-control study were confirmed independently in a second (PICU) cohort, especially showing nearly absent *Corynebacterium* and *Dolosigranulum*, suggesting that these children especially, had reduced resistance against overgrowth and dissemination of pathobionts to the lungs resulting in subsequent symptoms of LRTI.[5] Also, the fact that in our post-hoc analyses the same oral species were associated with both the decision to treat with antibiotics and with hospitalization duration, suggests a causal role for these bacteria in the severity of LRTIs.[25] A possible mechanism is that gram-negative oral bacteria promote a pro-inflammatory mucosal response,[26] leading to an increase in catecholamines that in turn accelerate the growth of these same gram-negative oral species, as well as that of potential pathogens such as *Haemophilus* spp. and *S. pneumoniae*.[27,28] Therefore, hypothetically it seems interesting to study whether determining the abundance of oral bacteria in respiratory specimens and letting that result drive the decision to treat with antibiotics, would improve our outcome.

So, what could be the implications of our finding? First, the unprecedented accuracy of our model in discriminating LRTI from health, makes microbiota-based diagnostics including viruses and bacteria, interesting as a potential tool for clinical application. Current diagnostics for detecting potentially pathogenic viruses and bacteria cover only a limited range of pathobionts and discriminate poorly between asymptomatic colonization or the cause of symptomatic disease. A recent proof-of-principle study using rapid microbiota-based diagnosis (<12 hours) for severe pneumonia in adults, underlines that such diagnostic tools improve diagnostic accuracy and could be within reach.[29] If the cost such technology reduces further and becomes available for pediatric use, we might be able to refrain from broad-spectrum antibiotics more often, and could instead specifically target the most abundant or overgrowing species by small-spectrum agents.[30] Although our microbiota-based approach has to be validated in independent cohorts, the non-inferior performance of the genus-level model suggests potential for future development of universal or country/region-based models, also in the context of prediction of severity and duration of disease by combined microbiota and host characteristics. This would potentially allow the physician to increase or decrease the threshold for antimicrobial treatment depending on the predicted outcome.

A second implication of our findings results from the observation that specific consortia of microorganisms are associated with health. Given these data are in line with multiple recent studies across the globe,[6,8,31–33] our findings urge for new studies to obtain mechanistic insight into their potential role in prevention of respiratory disease. For *Corynebacterium* spp. it was already reported to reduce virulence of *S. aureus* and inhibit *S. pneumoniae* growth *in vitro*.[34,35] Moreover, nasal application of *Corynebacterium* spp. induced resistance against RSV and secondary pneumococcal pneumonia in infant mice.[36] Together, all studies prompt for future research efforts to assess the (combined) effects of these commensal bacteria in modulation of the respiratory ecosystem, especially the containment of potential pathogens such as RSV, *Haemophilus* and *Streptococcus* spp. and host immune responses underlying respiratory symptoms.

A third potential implication follows from the observed phenotype-independent relation of viral and bacterial microbiota with LRTIs. This parallels the highly overlapping clinical presentations of these phenotypes in children, resulting in the lack of a robust gold standard for accurate classification and treatment.[37] Our findings contribute to the paradigm shift that is currently arising, demonstrating that viruses contribute to presumed bacterial pneumonia[38] and vice versa that bacteria seem to have an important role in pathogenesis and severity of presumed viral bronchiolitis[9] and wheezing illness,[39] suggesting the inappropriateness of these conventional single bacteria- and virus-centric views following Koch's postulates. Our findings also allude to the hypothesis that there is a universal pathway for the development of clinical LRTIs, linked to microbial dysbiosis, where clinical phenotypes are driven more by host (e.g. age, anatomy, baseline mucosal inflammation, status of innate and adaptive immunity, and genetic background) and environment rather than by single pathogen characteristics. This also underlines that treatment decisions for the time being should not be made on clinical phenotype, but rather on severity of disease. We fully realize we are only at the start of this scientific debate, and many discussions among and between clinicians, microbiologists, and biologists need to take place, as well as confirmatory studies of our results. However, technically and practically tools are there to adapt diagnostic and treatment protocols within the coming 5 years if the community finds this suitable.

The major strength of our study is the strictly matched case-control design, which precludes bias from the confounding effects of age, time, and gender. Moreover, the unselected recruitment of cases provides conclusive evidence in a cohort that highly represents the patients seen by pediatric clinicians. Last, the consistent patterns in our unsupervised and supervised analyses contribute to the robustness of our results.

Our study also has limitations. First, case-control designs could theoretically introduce selection bias. Second, only known respiratory viruses were detected by qPCR-based assays, but not the entire respiratory virome. However, virome studies report a high concordance between the results of metagenomic sequencing and qPCR-based assays.[40] Third, as with any observational study, our findings do not necessarily prove causality. Longitudinal analyses are underway to address cause-consequence analyses. Fourth, the endotracheal aspirate may not provide a perfect reflection of the lower respiratory tract microbiota extending into the bronchi and alveoli. That said, clinical evidence based on conventional microbiology data up till now have suggested that tracheal aspirates are a good proxy for the lower respiratory tract, and therefore an appropriate proxy for the clinical diagnosis of cause of disease in children with severe LRTI.[41] Furthermore, recent evidence showed a strong concordance with negligible differences between bacterial microbiota from endotracheal samples and bronchial lavages.[42] Finally, fifth, it should be underlined that 16S rRNA sequencing only permits annotation up to in between genus- and species-level identification of bacteria, and does not provide the resolution of metagenomic techniques such as shotgun sequencing, especially regarding closely related species such as streptococcal species. We tried to provide also some more species-level data by qPCR for confirmation of the four common and potentially pathogen OTUs, supporting our conclusions. Future studies might therefore be needed on multiple levels to further confirm our data, and refine the conclusions.

In conclusion, our findings urge for further exploration of microbiota-based diagnostics, as well as for further validation of our prediction model for severity of disease in different settings and countries, to explore their usefulness in optimizing treatment, and improve antimicrobial stewardship.

References

1 Liu L, Oza S, Hogan D, *et al.* Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet* 2016; **388**: 3027–35.

2 Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015; **386**: 743–800.

3 Rudan I. Epidemiology and etiology of childhood pneumonia. *Bull World Health Organ* 2008; **86**: 408–16.

4 Tsolia MN, Psarras S, Bossios A, *et al.* Etiology of Community-Acquired Pneumonia in Hospitalized School-Age Children: Evidence for High Prevalence of Viral Infections. *Clin Infect Dis* 2004; **39**: 681–6.

5 Man WH, de Steenhuijsen Piters WAA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol* 2017; **15**: 259–70.

6 Teo SM, Mok D, Pham K, *et al.* The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. *Cell Host Microbe* 2015; **17**: 704–15.

7 Biesbroek G, Tsivtsivadze E, Sanders EAM, *et al.* Early Respiratory Microbiota Composition Determines Bacterial Succession Patterns and Respiratory Health in Children. *Am J Respir Crit Care Med* 2014; **190**: 1283–92.

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

9 De Steenhuijsen Piters WAA, Heinonen S, Hasrat R, *et al.* Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. *Am J Respir Crit Care Med* 2016; **194**: 1104–15.

10 von Elm E, Altman DG, Egger M, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *PLoS Med* 2007; **4**: e296.

11 WHO. IMCI chart booklet. World Health Organization, 2014 http://www.who.int/maternal_child_adolescent/documents/IMCI_chartbooklet/en/ (accessed April 8, 2015).

12 Prevaes SMPJ, de Winter-de Groot KM, Janssens HM, *et al.* Development of the Nasopharyngeal Microbiota in Infants with Cystic Fibrosis. *Am J Respir Crit Care Med* 2016; **193**: 504–15.

13 Blanchet FG, Legendre P, Borcard D. Forward selection of explanatory variables. *Ecology* 2008; **89**: 2623–32.

14 Genuer R, Poggi J-M, Tuleau-Malot C. Variable selection using random forests. *Pattern Recognit Lett* 2010; **31**: 2225–36.

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

16 Vatanen T, Kostic AD, D'Hennezel E, *et al.* Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell* 2016; **165**: 842–53.

17 van de Pol AC, Wolfs TFW, van Loon AM, *et al.* Molecular quantification of respiratory syncytial virus in respiratory samples: reliable detection during the initial phase of infection. *J Clin Microbiol* 2010; **48**: 3569–74.

18 Perkins SM, Webb DL, Torrance SA, *et al.* Comparison of a Real-Time Reverse Transcriptase PCR Assay and a Culture Technique for Quantitative Assessment of Viral Load in Children Naturally Infected with Respiratory Syncytial Virus. *J Clin Microbiol* 2005; **43**: 2356–62.

19 Bassis CM, Erb-Downward JR, Dickson RP, *et al.* Analysis of the Upper Respiratory Tract Microbiotas as the Source of the Lung and Gastric Microbiotas in Healthy Individuals. *MBio*

2015; **6**: e00037-15.

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

21 Jain S, Self WH, Wunderink RG, *et al.* Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N Engl J Med* 2015; : 150714140110004.

22 Laufer AS, Metlay JP, Gent JF, Fennie KP, Kong Y, Pettigrew MM. Microbial communities of the upper respiratory tract and otitis media in children. *MBio* 2011; **2**: e00245-10.

23 Brook I. Prevotella and Porphyromonas infections in children. *J Med Microbiol* 1995; **42**: 340–7.

24 Pettigrew MM, Laufer AS, Gent JF, Kong Y, Fennie KP, Metlay JP. Upper respiratory tract microbial communities, acute otitis media pathogens, and antibiotic use in healthy and sick children. *Appl Environ Microbiol* 2012; **78**: 6262–70.

25 Dickson RP, Erb-Downward JR, Huffnagle GB. Towards an ecology of the lung: New conceptual models of pulmonary microbiology and pneumonia pathogenesis. *Lancet Respir Med* 2014; **2**: 238–46.

26 Thompson LR, Sanders JG, McDonald D, *et al.* A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* 2017; **551**: 457.

27 Marks LR, Davidson BA, Knight PR, Hakansson AP. Interkingdom signaling induces Streptococcus pneumoniae biofilm dispersion and transition from asymptomatic colonization to disease. *MBio* 2013; **4**: e00438-13.

28 O'Donnell PM, Aviles H, Lyte M, Sonnenfeld G. Enhancement of in vitro growth of pathogenic bacteria by norepinephrine: importance of inoculum density and role of transferrin. *Appl Environ Microbiol* 2006; **72**: 5097–9.

29 Pendleton KM, Erb-Downward JR, Bao Y, *et al.* Rapid Pathogen Identification in Bacterial Pneumonia Using Real-Time Metagenomics. *Am J Respir Crit Care Med* 2017; **196**: 1610–2.

30 Bogaert D, van Belkum A. Antibiotic treatment and stewardship in the era of microbiota-oriented diagnostics. *Eur J Clin Microbiol Infect Dis* 2018; **37**: 795–8.

31 Biesbroek G, Bosch AATM, Wang X, *et al.* The Impact of Breastfeeding on Nasopharyngeal Microbial Communities in Infants. *Am J Respir Crit Care Med* 2014; **190**: 140612135546007.

32 Luna PN, Hasegawa K, Ajami NJ, *et al.* The association between anterior nares and nasopharyngeal microbiota in infants hospitalized for bronchiolitis. *Microbiome* 2018; **6**: 2.

33 Salter SJ, Turner C, Watthanaworawit W, *et al.* A longitudinal study of the infant nasopharyngeal microbiota: The effects of age, illness and antibiotic use in a cohort of South East Asian children. *PLoS Negl Trop Dis* 2017; **11**: e0005975.

34 Ramsey MM, Freire MO, Gabrilska RA, Rumbaugh KP, Lemon KP. Staphylococcus aureus Shifts toward Commensalism in Response to Corynebacterium Species. *Front Microbiol* 2016; **7**: 1230.

35 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**: e01725-15.

36 Kanmani P, Clua P, Vizoso-Pinto MG, *et al.* Respiratory Commensal Bacteria Corynebacterium pseudodiphtheriticum Improves Resistance of Infant Mice to Respiratory Syncytial Virus and Streptococcus pneumoniae Superinfection. *Front Microbiol* 2017; **8**: 1613.

37 Scott JAG, Wonodi C, Moïsi JC, *et al.* The definition of pneumonia, the assessment of severity, and clinical standardization in the Pneumonia Etiology Research for Child Health study. *Clin Infect Dis* 2012; **54 Suppl 2**: S109-16.

38 Bosch AATM, Biesbroek G, Trzcinski K, Sanders EAM, Bogaert D. Viral and bacterial interactions in the upper respiratory tract. *PLoS Pathog* 2013; **9**: e1003057.

39 Beigelman A, Bacharier LB. Early-life respiratory infections and asthma development. *Curr Opin Allergy Clin Immunol* 2016; **16**: 172–8.

40 Lysholm F, Wetterbom A, Lindau C, *et al.* Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. *PLoS One* 2012;

7: e30875.

41 McCauley LM, Webb BJ, Sorensen J, Dean NC. Use of Tracheal Aspirate Culture in Newly Intubated Patients with Community-Onset Pneumonia. *Ann Am Thorac Soc* 2016; **13**: 376–81.

42 Dickson RP, Erb-Downward JR, Freeman CM, *et al*. Bacterial Topography of the Healthy Human Lower Respiratory Tract. *MBio* 2017; **8**: e02287-16.

Acknowledgements

The authors are indebted to all the participating children and their families. We thank all the members of the research team of the Spaarne Gasthuis Academie, the laboratory staff, the Streeklaboratorium Haarlem, the GGD Kennemerland, and JGZ Kennemerland. We are grateful to Olga Krabbe for her primary exploration of the data. This work is dedicated to the memory of Dr. Reinier H. Veenhoven who contributed to the planning and design of the study.

Declaration of Interests

EAMS declares to have received unrestricted research support from Pfizer, grant support for vaccine studies from Pfizer and GSK. DB declares to have received unrestricted fees paid to the institution for advisory work for Friesland Campina and well as research support from Nutricia and MedImmune. None of the fees or grants listed here was received for the research described in this paper. No other authors reported financial disclosures. None of the other authors report competing interests.

Author contributions

D.B., M.A. van H., and E.A.M.S conceived and designed the experiments. W.H.M., M.A. van H., M.E.M., A.M.V., and N.J.G.J. included the participants. M.L.J.N.C. were responsible for the execution and quality control of the laboratory work. W.H.M. and D.B. analyzed the data. W.H.M., M.A. van H., E.A.M.S, and D.B. wrote the paper. All authors significantly contributed to interpreting the results, critically revised the manuscript for important intellectual content, and approved the final manuscript.

Data availability

Sequence data that support the findings of this study have been deposited in the NCBI Sequence Read Archive (SRA) database with BioProject ID PRJNA428382.

Table and Figures

Table 1. Baseline characteristics for the cases and their matched controls.

Data on medication use was acquired by pharmacy printouts, whereas the rest of the data was acquired by parent questionnaires. Breastfeeding was nonexclusive. Educational level was classified into three categories: low level (primary school education or pre-vocational education as highest qualification), intermediate (selective secondary education or vocational education) and high level (university of applied sciences and research university). Smoke exposure included children who were exposed to second-hand tobacco smoke. P values were determined by univariate conditional logistic regression. Matching factors were not tested. IQR = interquartile range; RTI = respiratory tract infection; LRTI = any parental-reported lower RTI.

	Cases	Controls	P value
n	154	307	
Basics			
Girl (%)	61 (39·6)	122 (39·7)	
Age (months) (median [IQR])	13·6 [4·9, 27·4]	14·1 [5·3, 28·4]	
Born at term (%)	142 (92·2)	294 (95·8)	0·111
Mode of delivery (%)			0·457
vaginal	124 (80·5)	260 (84·7)	
elective C-section	15 (9·7)	26 (8·5)	
emergency C-section	15 (9·7)	21 (6·8)	
Season of sampling (%)			
Spring	49 (32·0)	91 (29·6)	
Summer	22 (14·4)	44 (14·3)	
Autumn	8 (5·2)	19 (6·2)	
Winter	74 (48·4)	153 (49·8)	
Medical History			
LRTI (%)	38 (25·0)	22 (7·2)	<0·001
Wheezing (%)	41 (26·6)	22 (7·2)	<0·001
Otitis (%)	38 (24·7)	46 (15·0)	0·008
Hospitalization for RTI (%)	33 (21·7)	10 (3·3)	<0·001
Medication			
Antibiotics past 6 months (%)	41 (27·2)	19 (6·2)	<0·001
Feeding			
Breastfeeding >3 months (%)	58 (37·7)	169 (55·0)	<0·001
Family			
Education level parents (%)			<0·001
high	99 (64·7)	262 (85·3)	
intermediate	49 (32·0)	42 (13·7)	
low	5 (3·3)	3 (1·0)	
Siblings (median [IQR])	1·0 [1·0, 2·0]	1·0 [0·0, 1·0]	0·002
Environment			
Smoke exposure (%)	36 (23·4)	44 (14·3)	0·015

Figure 1. Viral PCR positivity in cases and controls.

The proportions of qPCR respiratory virus detections for cases (red, n=148) and controls (green, n=302).

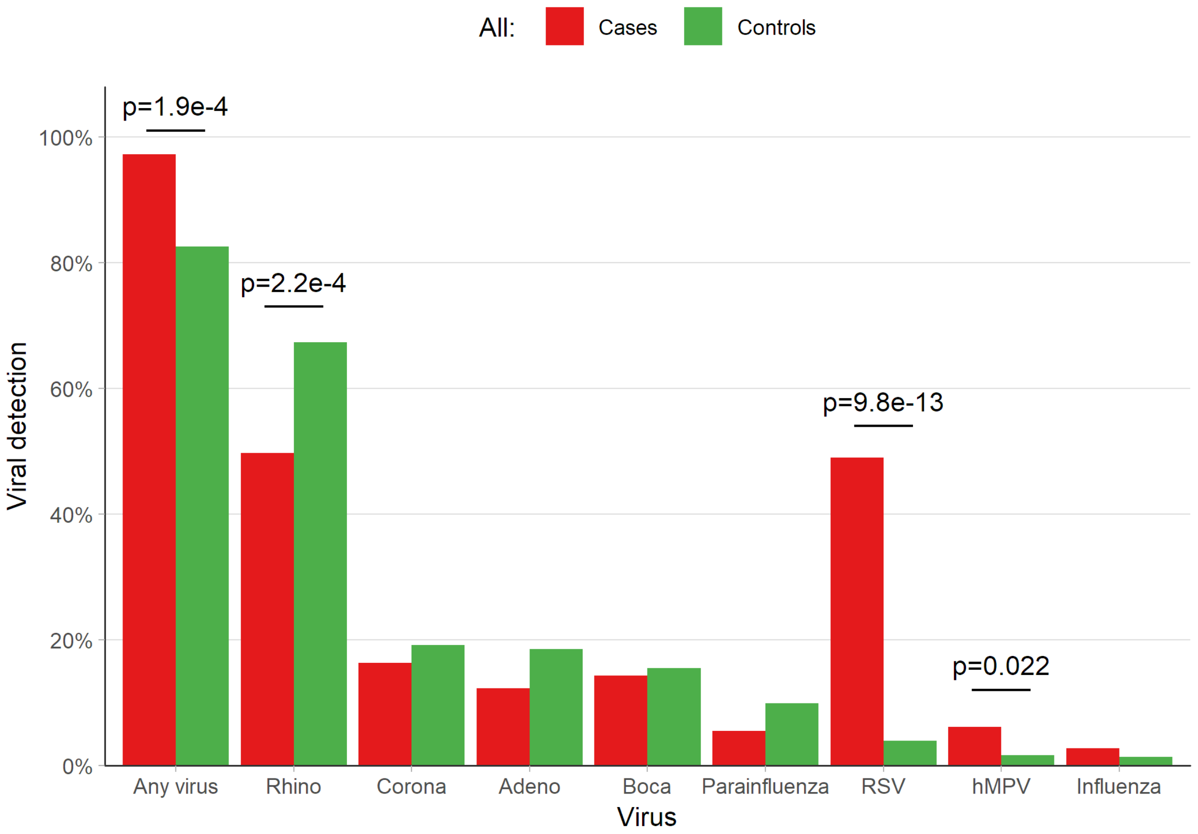
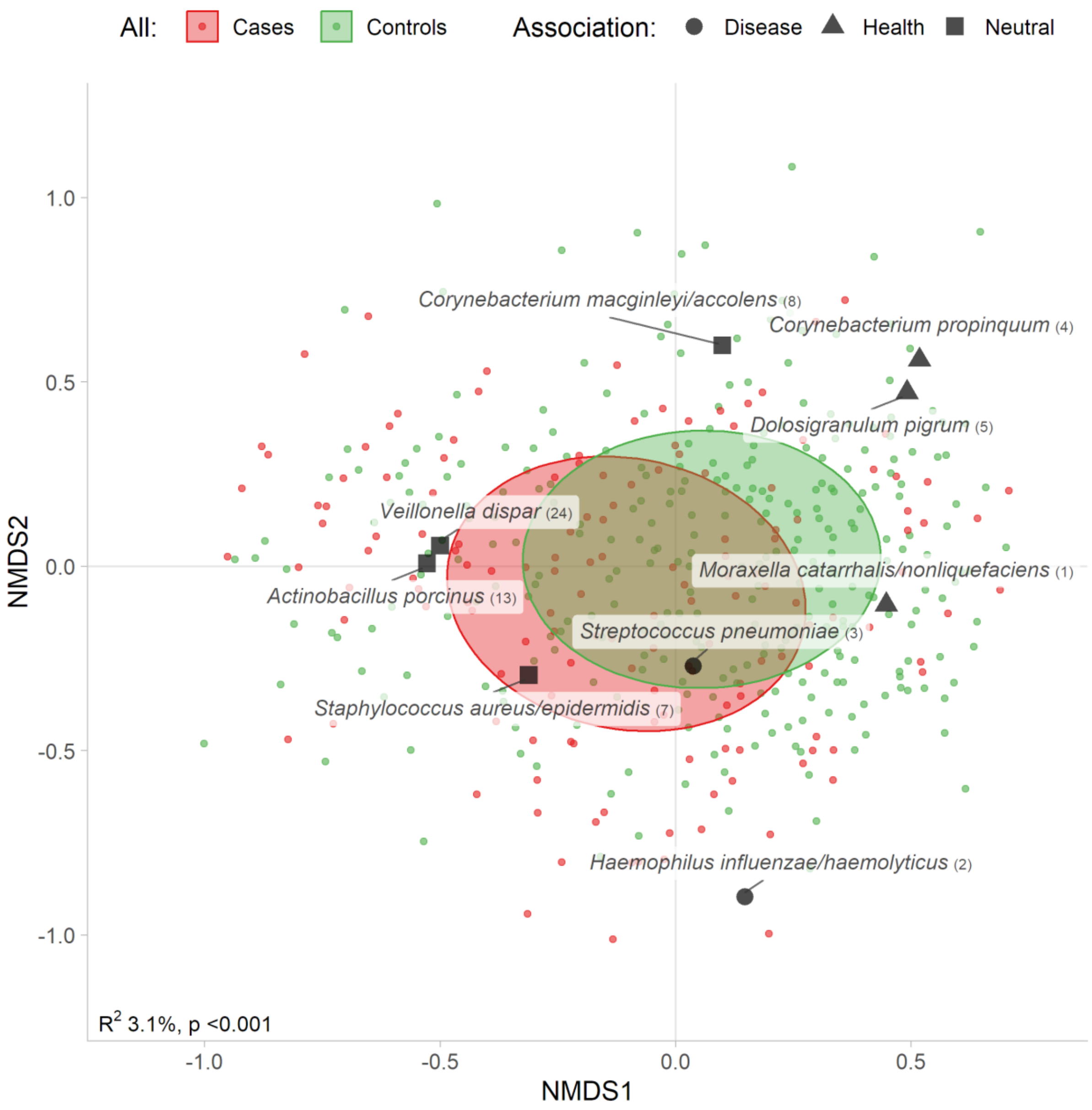


Figure 2. NMDS biplot.

NMDS biplot depicting the individual nasopharyngeal microbiota composition (data points, n=457) colored by subcohort: LRTI-cases at admission (red, n=151) and matched controls (green, n=306). Ellipses represent the standard deviation of all points within a cohort. In addition, figure **A** depicts the 9 bacterial species biomarkers (determined by Random Forest analysis on hierarchical clustering results). Figure **B** adds a posteriori projection of covariates that significantly explained the compositional variation between cases and controls (grey = significant in univariable analysis, black = significant in multivariable analysis). For readability, only a selection of the covariates explaining the largest variation are displayed. In addition, the association with age (purple) has been included to demonstrate that the age-effect (vertical orientation for younger vs. older subjects) was perpendicular (~90° angle) to the disease-health axis (horizontal orientation), showing that age-related differences in microbiota composition per se are not associated with disease. Stress: 0·269.

A


All:
Cases
Controls
Association:
Disease
Health
Neutral
Corynebacterium macginleyi/accolens (8)
Corynebacterium propinquum (4)
Dolosigranulum pigrum (5)
Veillonella dispar (24)
Moraxella catarrhalis/nonliquefaciens (1)
Actinobacillus porcinus (13)
Streptococcus pneumoniae (3)
Staphylococcus aureus/epidermidis (7)
Haemophilus influenzae/haemolyticus (2)
NMDS2
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
NMDS1
R2 3.1%, p <0.001

B

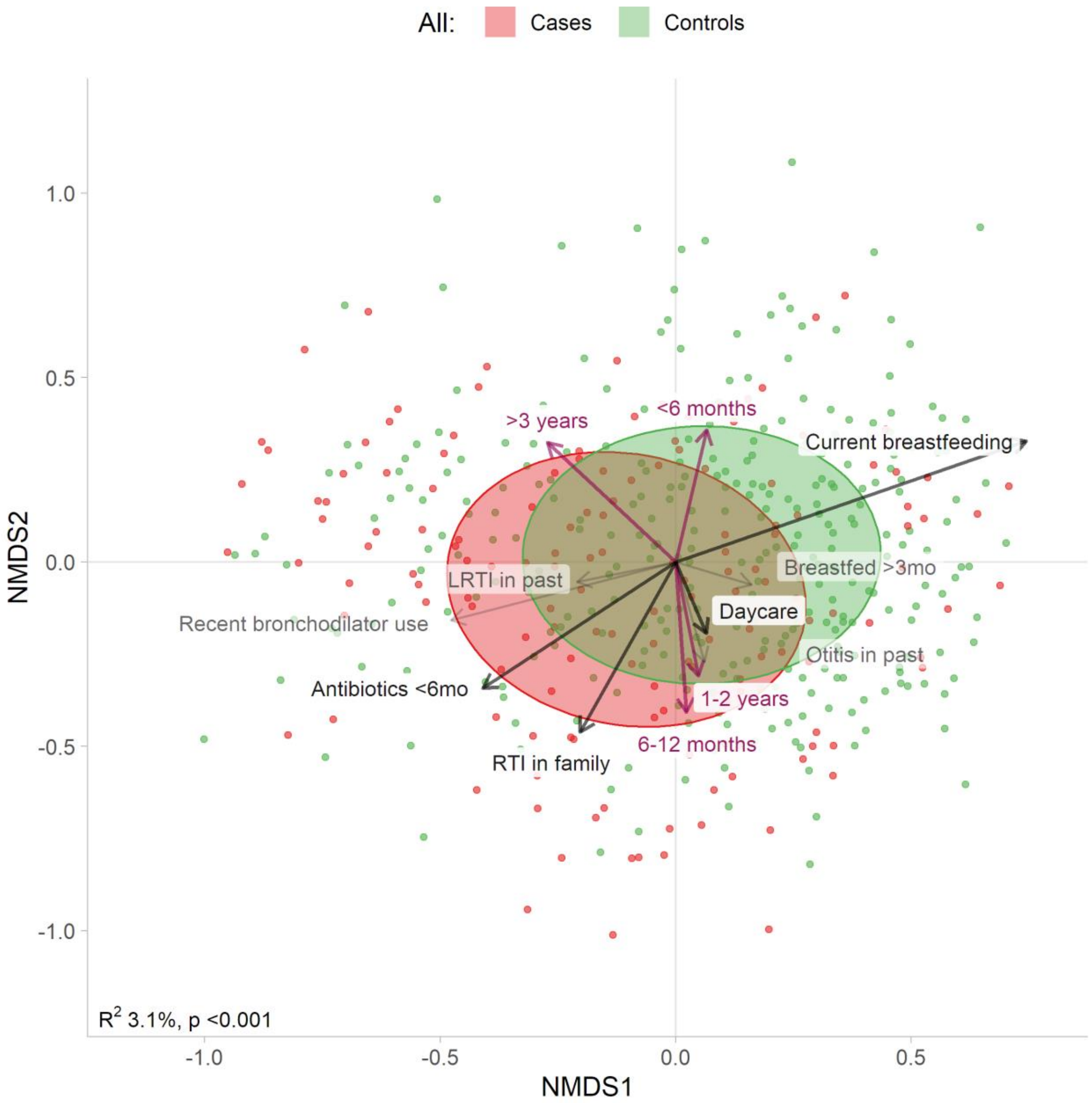
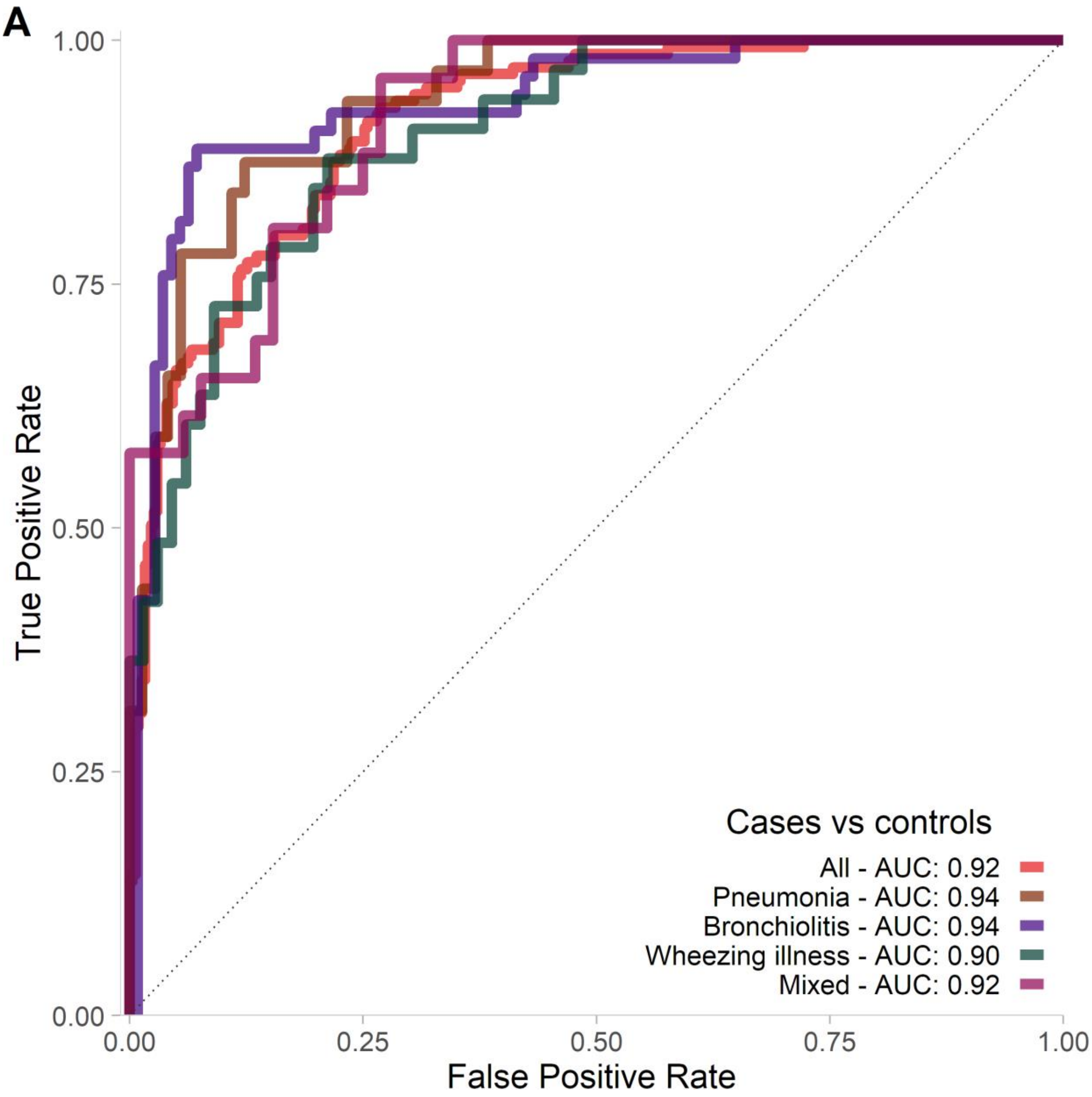
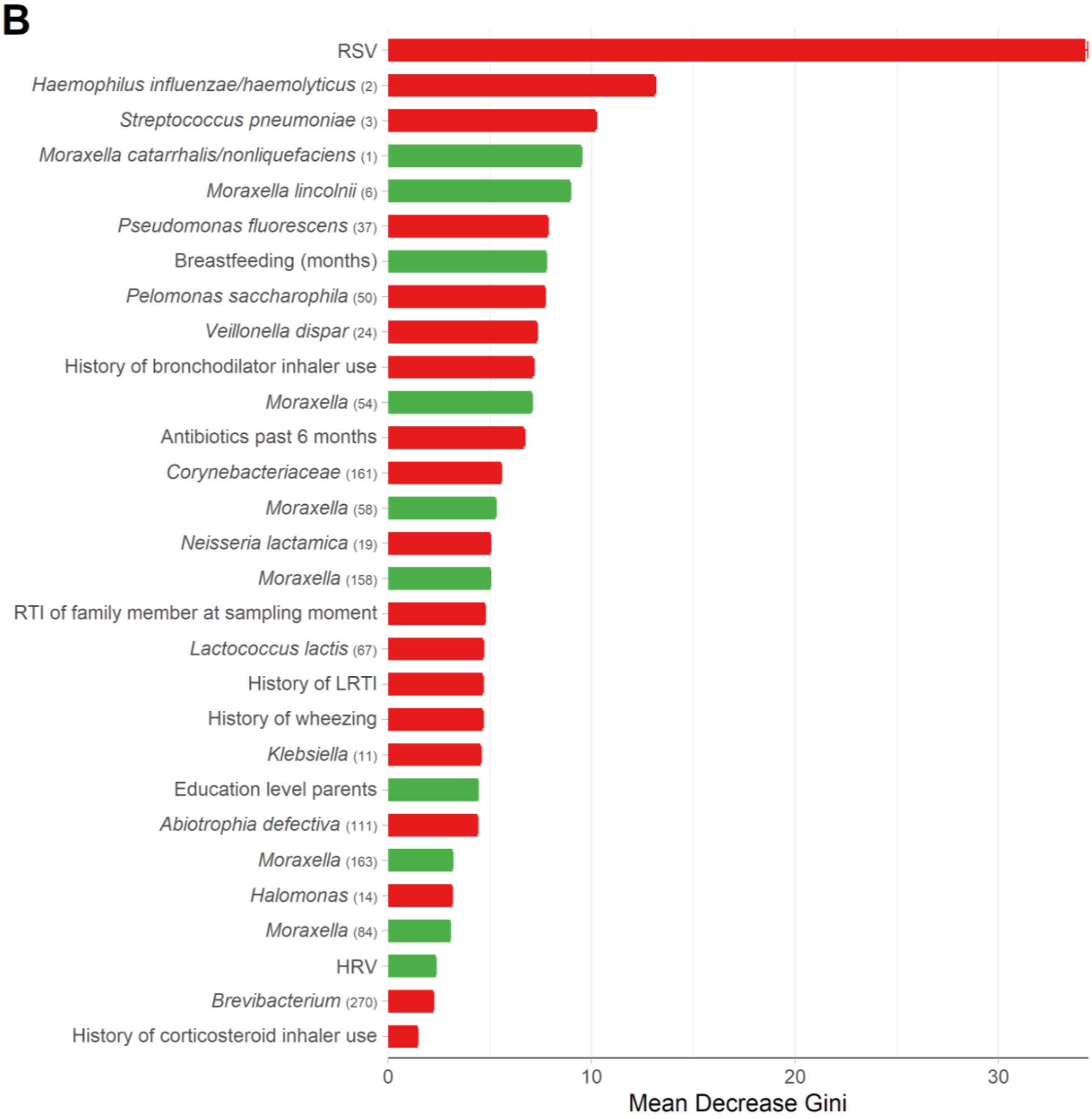

All:
Cases
Controls
NMDS2
NMDS1
1.0
0.5
0.0
-0.5
-1.0
-1.0
-0.5
0.0
0.5
>3 years
<6 months
Current breastfeeding
Breastfed >3mo
LRTI in past
Recent bronchodilator use
Daycare
Otitis in past
Antibiotics <6mo
1-2 years
6-12 months
RTI in family
R^2 3.1%, p <0.001

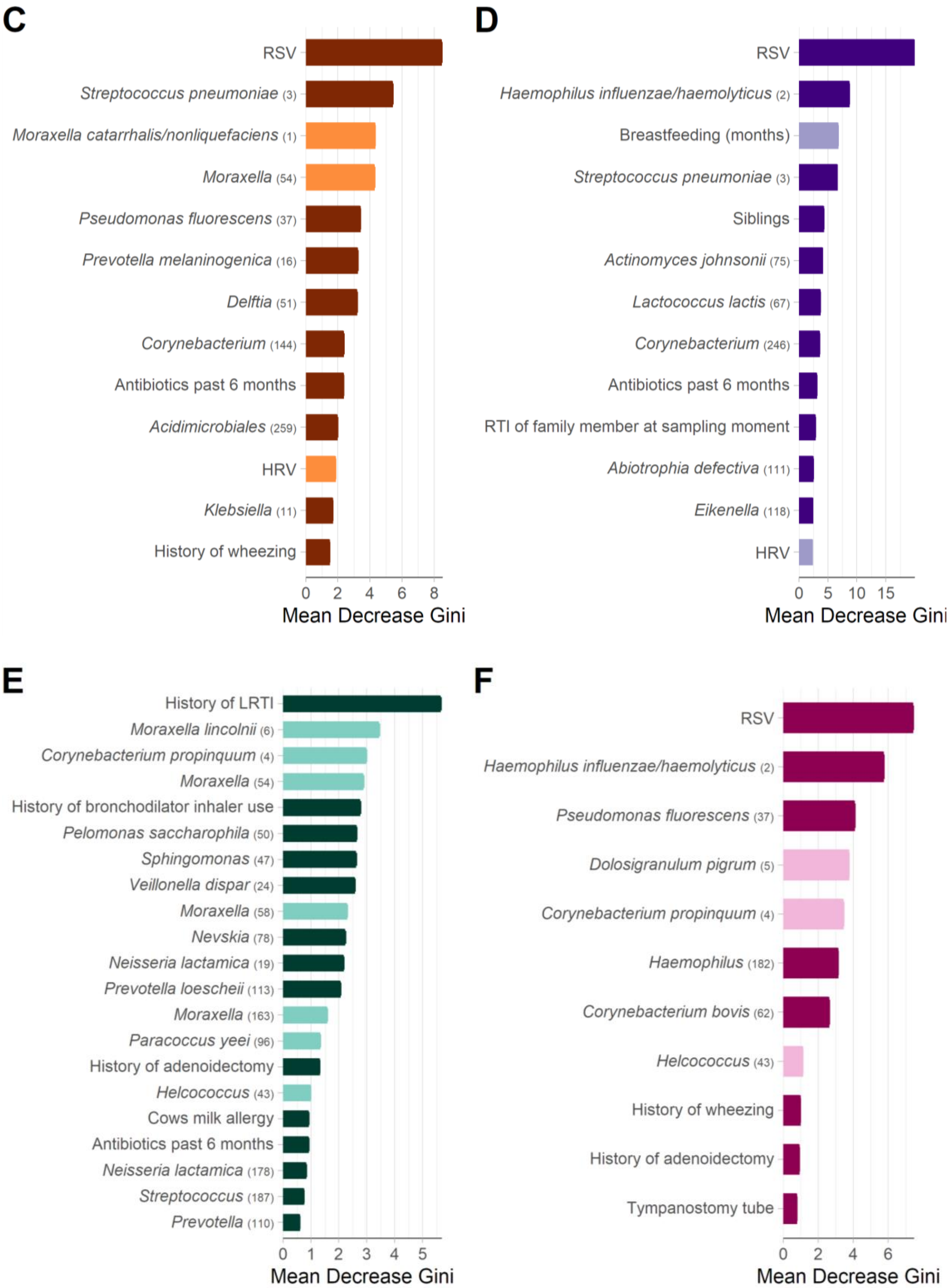
Figure 3. Random forest models classifying disease and health based on 16S rRNA data, viral presence and patient characteristics combined.

Twenty-four variables were discriminating cases from controls in the unstratified cohort (n=457; **B**) leading to a sparse classification model with an AUC of 0·92 (**A**). Variables are ranked in descending order based on their importance to the accuracy of the model. Variable importance was estimated by calculating the mean decrease in Gini after randomly permuting the values of each given variable (mean ± standard deviation, 100 replicates). The direction of the associations was estimated *post-hoc* using point biserial correlations (green = associated with health; red = associated with disease). The disease-discriminatory variables for the pneumonia cases (brown, n=108; **C**), bronchiolitis cases (purple, n=171; **D**), wheezing illness cases (dark green, n=100; **E**), and mixed-phenotype cases (pink, n=78; **F**) versus their matched controls are depicted in figures **C-F** (light colored bars are positively associated with health). The ROC curves for distinguishing disease from health of these stratified sparse random forest classifying models are depicted in **A**.


A
1.00
0.75
0.50
0.25
0.00
True Positive Rate
0.00
0.25
0.50
0.75
1.00
False Positive Rate
Cases vs controls
All - AUC: 0.92
Pneumonia - AUC: 0.94
Bronchiolitis - AUC: 0.94
Wheezing illness - AUC: 0.90
Mixed - AUC: 0.92


B
RSV
Haemophilus influenzae/haemolyticus (2)
Streptococcus pneumoniae (3)
Moraxella catarrhalis/nonliquefaciens (1)
Moraxella lincolnii (6)
Pseudomonas fluorescens (37)
Breastfeeding (months)
Pelomonas saccharophila (50)
Veillonella dispar (24)
History of bronchodilator inhaler use
Moraxella (54)
Antibiotics past 6 months
Corynebacteriaceae (161)
Moraxella (58)
Neisseria lactamica (19)
Moraxella (158)
RTI of family member at sampling moment
Lactococcus lactis (67)
History of LRTI
History of wheezing
Klebsiella (11)
Education level parents
Abiotrophia defectiva (111)
Moraxella (163)
Halomonas (14)
Moraxella (84)
HRV
Brevibacterium (270)
History of corticosteroid inhaler use
0
10
20
30
Mean Decrease Gini

616


C
RSV
Streptococcus pneumoniae (3)
Moraxella catarrhalis/nonliquefaciens (1)
Moraxella (54)
Pseudomonas fluorescens (37)
Prevotella melaninogenica (16)
Delftia (51)
Corynebacterium (144)
Antibiotics past 6 months
Acidimicrobiales (259)
HRV
Klebsiella (11)
History of wheezing
0 2 4 6 8
Mean Decrease Gini
D
RSV
Haemophilus influenzae/haemolyticus (2)
Breastfeeding (months)
Streptococcus pneumoniae (3)
Siblings
Actinomyces johnsonii (75)
Lactococcus lactis (67)
Corynebacterium (246)
Antibiotics past 6 months
RTI of family member at sampling moment
Abiotrophia defectiva (111)
Eikenella (118)
HRV
0 5 10 15
Mean Decrease Gini
E
History of LRTI
Moraxella lincolnii (6)
Corynebacterium propinquum (4)
Moraxella (54)
History of bronchodilator inhaler use
Pelomonas saccharophila (50)
Sphingomonas (47)
Veillonella dispar (24)
Moraxella (58)
Nevskia (78)
Neisseria lactamica (19)
Prevotella loescheii (113)
Moraxella (163)
Paracoccus yeei (96)
History of adenoidectomy
Helcococcus (43)
Cows milk allergy
Antibiotics past 6 months
Neisseria lactamica (178)
Streptococcus (187)
Prevotella (110)
0 1 2 3 4 5
Mean Decrease Gini
F
RSV
Haemophilus influenzae/haemolyticus (2)
Pseudomonas fluorescens (37)
Dolosigranulum pigrum (5)
Corynebacterium propinquum (4)
Haemophilus (182)
Corynebacterium bovis (62)
Helcococcus (43)
History of wheezing
History of adenoidectomy
Tympanostomy tube
0 2 4 6
Mean Decrease Gini