

## Post-mortem Nasopharyngeal Microbiome Analysis of Zambian Infants with and without Respiratory Syncytial Virus Disease: A Nested Case Control Study

Jessica McClintock, BS,<sup>1</sup> Aubrey R. Odom-Mabey, MS,<sup>2</sup> Nitsueh Keber, MS,<sup>2</sup>  
Arshad Ismail, PhD,<sup>3,4</sup> Lawrence Mwananyanda, MD,<sup>5</sup> Christopher J. Gill, MD,<sup>5</sup>  
William B. MacLeod, Sc.D.,<sup>5</sup> Rachel C. Pieciak, MPH,<sup>5</sup> Rotem Lapidot, MD, MSCI,<sup>6,7</sup>  
W. Evan Johnson, PhD<sup>1,2</sup>

<sup>1</sup>Division of Infectious Disease, Center for Data Science, Rutgers New Jersey Medical School, Newark, NJ, 07103, USA

<sup>2</sup>Bioinformatics Program, Boston University, Boston, MA, 02118, USA

<sup>3</sup>Sequencing Core Facility, National Institute for Communicable Diseases of the National Health Laboratory Service, 2131 Johannesburg, South Africa

<sup>4</sup>Department of Biochemistry and Microbiology, University of Venda, Thohoyandou 0950, South Africa

<sup>5</sup>Department of Global Health, Boston University School of Public Health, Boston, MA, 02118, USA

<sup>6</sup>Pediatric Infectious Diseases, Boston Medical Center, Boston, MA, 02118, USA

<sup>7</sup>Pediatrics, Boston University School of Medicine, Boston, MA, 02118, USA

### ABSTRACT

**Background:** Respiratory Syncytial Virus (RSV) is the most common cause of bronchiolitis and lower respiratory tract infections in children in their first year of life, disproportionately affecting infants in developing countries. Previous studies have found that the nasopharyngeal microbiome of infants with RSV infection has specific characteristics that correlate with disease severity, including lower biodiversity, perturbations of the microbiota and differences in relative abundance. These studies have focused on infants seen in clinical or hospital settings, predominantly in developed countries.

**Methods:** We conducted a nested case control study within a random sample of 50 deceased RSV+ infants with age at death ranging from 4 days to 6 months and 50 matched deceased RSV- infants who were all previously enrolled in the Zambia Pertussis and RSV Infant Mortality Estimation (ZPRIME) study. All infants died within the community or within 48 hours of facility admittance. As part of the ZPRIME study procedures, all decedents underwent one-time, post-mortem nasopharyngeal sampling. The current analysis explored the differences between the nasopharyngeal microbiome profiles of RSV+ and RSV- decedents using 16S ribosomal DNA sequencing.

**Results:** We found that *Moraxella* was more abundant in the nasopharyngeal microbiome of RSV+ decedents than in RSV- decedents. Additionally, *Gemella* and *Staphylococcus* were less abundant in RSV+ decedents than in RSV- decedents.

**Conclusion:** These results support previously reported findings of the association between the nasopharyngeal microbiome and RSV and suggest that changes in the abundance of these microbes are likely specific to RSV and may correlate with mortality associated with the disease.

## INTRODUCTION

Respiratory syncytial virus (RSV) infection is the most common cause of bronchiolitis and pneumonia in children in the first year of life.<sup>1,2</sup> Symptoms of RSV range from mild upper respiratory infections to more severe infections such as bronchiolitis and pneumonia.<sup>3,4</sup> RSV has been established as the leading global cause of lower respiratory tract infections in infants, with an etiological fraction for pneumonia three times higher than the next highest named pathogen.<sup>2</sup> RSV continues to be a major health concern in both developed and developing nations.<sup>5,6</sup> The developing world experiences the highest burden of RSV infection, accounting for 99% of worldwide deaths from RSV-related lower respiratory tract infections in children under five.<sup>3</sup> We recently reported that in low-income countries, roughly two-thirds of infant RSV deaths occurred in the community.<sup>7</sup>

Several studies have looked at associations between the nasopharyngeal (NP) microbiome profiles of infants and RSV disease severity. Previous studies have shown that the NP microbiome in RSV-infected infants has characteristics that changed with severity of disease.<sup>2,8-13</sup> For instance, an overabundance of *Achromobacter*, *Haemophilus*, *Moraxella*, and *Streptococcus* microbes and a loss of *Corynebacterium*, *Staphylococcus*, and *Veillonella* have been associated with infants with more severe RSV.<sup>11-13</sup> Studies have also shown that antibiotic use leads to decreased alpha diversity and dysbiosis in the NP microbiome.<sup>14,15</sup>

RSV infection may include indirect effects mediated by its impact on other members of the respiratory ecosystem.<sup>16-18</sup> These effects seem to occur with bacteria such as *Haemophilus influenzae* and *Streptococcus pneumoniae*, suggesting that the impact of RSV on bacterial pathogenesis is relatively specific.<sup>19,20</sup>

Data from studies outside of high-income countries is very limited, and yet most RSV-associated mortality is concentrated in low- and middle-income countries.<sup>6</sup> Until recently, the burden of RSV in low- and middle-income countries was estimated with hospital-based surveillance data used to model RSV prevalence among community deaths. The Zambia Pertussis and RSV Infant Mortality Estimation (ZPRIME) study was a systematic, post-mortem surveillance study designed to address this knowledge gap by directly measuring facility and community deaths in Lusaka, Zambia.<sup>7</sup> The ZPRIME study found that RSV caused 2.8% of all infant deaths and 4.7% of all community deaths.

Our current analysis examines the NP microbiomes of a subset of RSV+ and RSV- deceased infants enrolled in the ZPRIME study. 16S ribosomal DNA sequencing was conducted on post-mortem infant NP samples with the goal of characterizing and comparing the NP microbiome profiles of RSV- and RSV+ decedents. Our study focuses on community deaths in a low-income county to address gaps from previous studies.

## METHODS

### Study Population and Study Design

We identified a nested cross-sectional sample of RSV+ and RSV- deceased infants from the ZPRIME study.<sup>7</sup> ZPRIME aimed to determine the proportion of infant deaths that could be

attributed to *Bordetella pertussis* and RSV infections. Decedents were between four days and six months old and enrolled in the ZPRIME study within 48 hours of death. We selected a subset of decedents for our analysis who were RSV+ with a PCR Ct<35. Demographic and clinical data, including sex, age at death, mother’s human immunodeficiency virus (HIV) status, location of death, and cause of death, were collected for each decedent (Table 1). All decedents underwent one-time, post-mortem NP sampling. NP samples were tested for RSV by reverse transcriptase quantitative PCR following the testing protocol developed by the Centers for Disease Control and Surveillance.<sup>21</sup>

We selected 50 RSV+ decedents via random sampling and matched those to 50 RSV- decedents. Due to age- and seasonality-related variations in RSV impact, we matched decedents by age at death (4–28 days (0+ months), 29–61 days (1+ month), 62–152 days (2+ months) and 152–182 days (4+ months)) and date of death (+/- 1.5 months) (Table 1).

Additional details about our study population are available in Text, Supplemental Digital Content 1, as well as information about sample collection, processing, and storage, 16S ribosomal rDNA amplification and sequencing, and data processing.

<b>Characteristics of RSV+ and RSV- Decedents</b>				
	<b>RSV+ Infants</b>	<b>RSV- Infants</b>	<b>All Subjects</b>	<b>p-value</b>
<i>In study (n)</i>	50	50	100	
<b>Sex</b>				0.128
<i>Males % (n)</i>	46% (23/50)	56% (28/50)	51% (51/100)	
<i>Females % (n)</i>	38% (19/50)	40% (20/50)	39% (39/100)	
<i>Unknown Sex % (n)</i>	16% (8/50)	4% (2/50)	10% (10/100)	
<b>Age</b>				0.369
<i>Median Age (days)</i>	41	39	40.5	
<i>4-28 days % (n)</i>	38% (19/50)	38% (19/50)	38% (38/100)	
<i>29-61 days % (n)</i>	30% (15/50)	28% (14/50)	29% (29/100)	
<i>62-152 days % (n)</i>	24% (12/50)	26% (13/50)	25% (25/100)	
<i>153-182 days % (n)</i>	8% (4/50)	8% (4/50)	8% (8/100)	
<b>Mother’s HIV Status</b>				0.269
<i>HIV Exposed % (n)</i>	4% (2/50)	12% (6/50)	8% (8/100)	
<i>HIV Unexposed % (n)</i>	96% (48/50)	88% (44/50)	92% (92/100)	
<b>Location of Death</b>				1.000
<i>Early Facility Death % (n)</i>	14% (7/50)	16% (8/50)	15% (15/100)	
<i>Community Death % (n)</i>	86% (43/50)	84% (42/50)	85% (85/100)	
<b>Cause of Death</b>				<0.001
<i>Respiratory % (n)</i>	72% (36/50)	26% (13/50)	49% (49/100)	
<i>Non-Respiratory % (n)</i>	22% (11/50)	64% (32/50)	43% (43/100)	
<i>Unsure % (n)</i>	6% (3/50)	10% (5/50)	8% (8/100)	

**Table 1.** Characteristics of deceased RSV+ and RSV- infants. RSV+ infants were selected through random sampling and were matched to RSV- infants by age at death and date of death. Chi-squared tests were performed for each group. Note that the infant’s sex was not reported in all cases.

## Statistical Analysis

We analyzed microbe counts at both the genus and species levels, which is possible using PathoScope2.0.<sup>22–24</sup> We performed chi-squared tests to determine if the population characteristics of our RSV+ and RSV- groups were evenly distributed. We calculated relative abundances based on raw count data for the taxon grouped by disease state and constructed bar plots, boxplots, and heatmaps comparing microbe prevalence across the RSV+ and RSV- conditions.<sup>25,26</sup> Samples in the bar chart are ordered by the top 4 most abundant samples found in RSV+ samples to enhance visual organization. Differential abundance analysis was conducted based on relative abundance between genera using a nonparametric Wilcoxon rank sum test.<sup>27–29</sup> To correct for multiple testing, we used Benjamini-Hochberg p-value adjustments.<sup>30</sup> Significant results, defined as  $p < 0.05$  and adjusted  $p < 0.25$ , were visualized with boxplots.

To analyze alpha diversity, we computed Shannon and Simpson indices and visualized each index as a boxplot.<sup>31,32</sup> We compared indices across RSV state using Wilcoxon rank sum test as the data was not normally distributed, as determined by the Shapiro-Wilk test for normality.<sup>27–29,33–35</sup>

To analyze beta diversity, we computed the Bray–Curtis dissimilarity indices to observe differences in microbial composition between the samples.<sup>36</sup> The results were visualized using non-metric dimensional scaling (NMDS) and principal coordinate analysis (PCoA).<sup>26,37,38</sup> We performed NMDS with 20 stress test runs. Stress values are reported to indicate how well the model fits, where a lower stress value indicates a better model fit. Atchison distances were also computed and visualized via PCoA.<sup>39</sup> Ellipses on the NMDS and PCoA plots represent a 95% confidence level for a multivariate t-distribution. PERMANOVA was performed to determine if the centroids and dispersion was statistically significant between RSV states.<sup>40</sup>

We analyzed pathway abundances via a Wilcoxon rank sum test with Benjamini-Hochberg corrections to determine if there was a difference in inferred pathway abundances between RSV+ and RSV- samples.<sup>27–30,41–43</sup> Significant results have an adjusted  $p < 0.05$ .

## RESULTS

### Data Filtering

Due to the heavy presence of taxa with relatively low abundances in the raw data, taxa with average relative abundances  $< 0.1\%$  were filtered from the dataset and regrouped as a single “Other” category. This reduced our identified taxa assignments from 632 genera encompassing 1042 species to 36 genera encompassing 54 species, not including the “Other” grouping. We used these data for all analyses expect for visualization of relative abundance, which used a data set including genera with average relative abundances  $< 1\%$ . This smaller group consisted of 15 genera and 1 “Other” category.

### Baseline characteristics of the study population

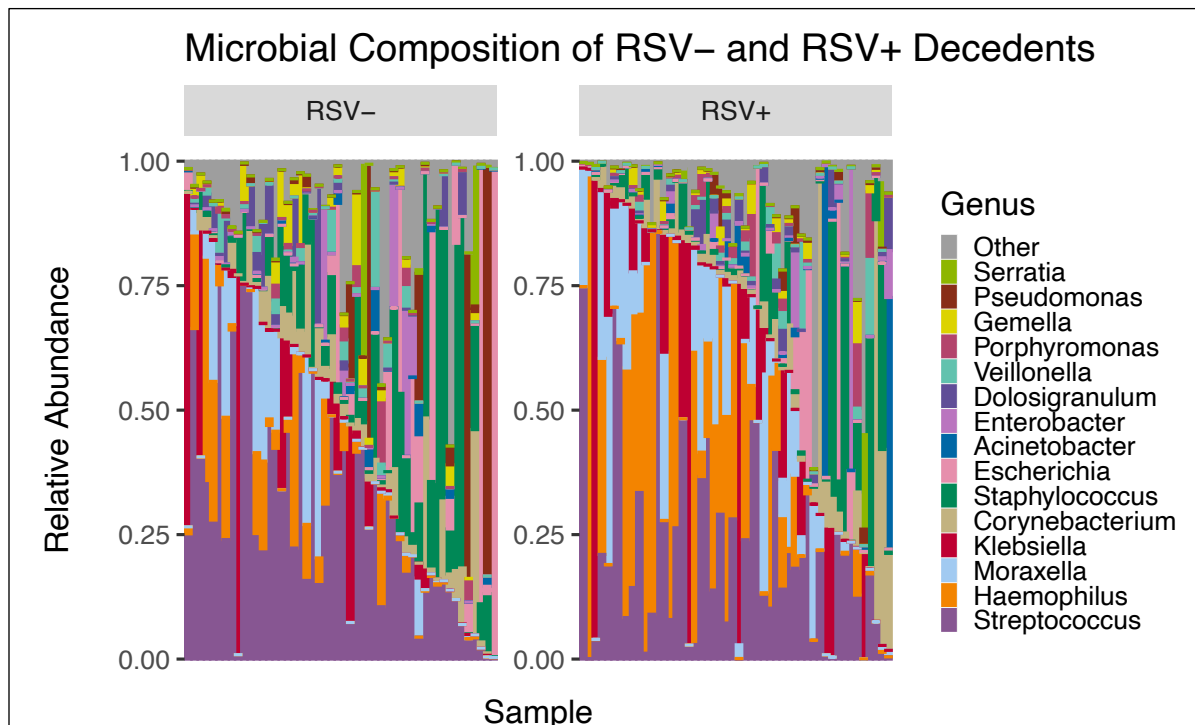
Table 1 displays the baseline characteristics for the 100 decedents. The groups were roughly even across sex (51% male, 39% female, 10% unknown). The median age of death was 40.5 days. The

majority (85%) of decedents were brought in dead from the community. A small proportion of deceased infants in both groups were known to be born to mothers infected with HIV (8%), although the final HIV status of the decedents themselves remains unknown. Among all decedents, there were 49 respiratory deaths, 43 non-respiratory deaths and 8 inconclusive deaths. As expected, most RSV+ deaths were respiratory (72%) and most RSV- deaths were non-respiratory (64%) in origin.

Chi-squared tests were performed to determine if the population characteristics were evenly distributed between the RSV+ and matched RSV- decedents. No significant difference was found for sex ( $p=0.128$ ), age ( $p=0.369$ ), HIV exposure status ( $p=0.269$ ), or location of death ( $p=1.000$ ) indicating an even decedent distribution among groups (Table 1). The cause of death (respiratory vs non-respiratory) differed significantly between RSV+ and RSV- deceased infants ( $p<0.001$ ).

### Relative Abundance

To compare the relative abundance of microbial taxa across RSV+ and RSV- samples, we generated a stacked bar plot composed of all genera with relative abundances greater than 1% (Figure 1). Figure 1 shows a total of 16 genera across all the samples with varying abundance levels. Each column represents the composite NP microbiome for a single participant and are ordered based on the abundance of the top 4 RSV+ microbes. When averaged across samples, the five most abundant microbes in RSV+ samples were *Streptococcus* (22.2%), *Haemophilus* (19.7%), *Moraxella* (11.0%), *Klebsiella* (10.2%), and *Corynebacterium* (6.8%). The five most



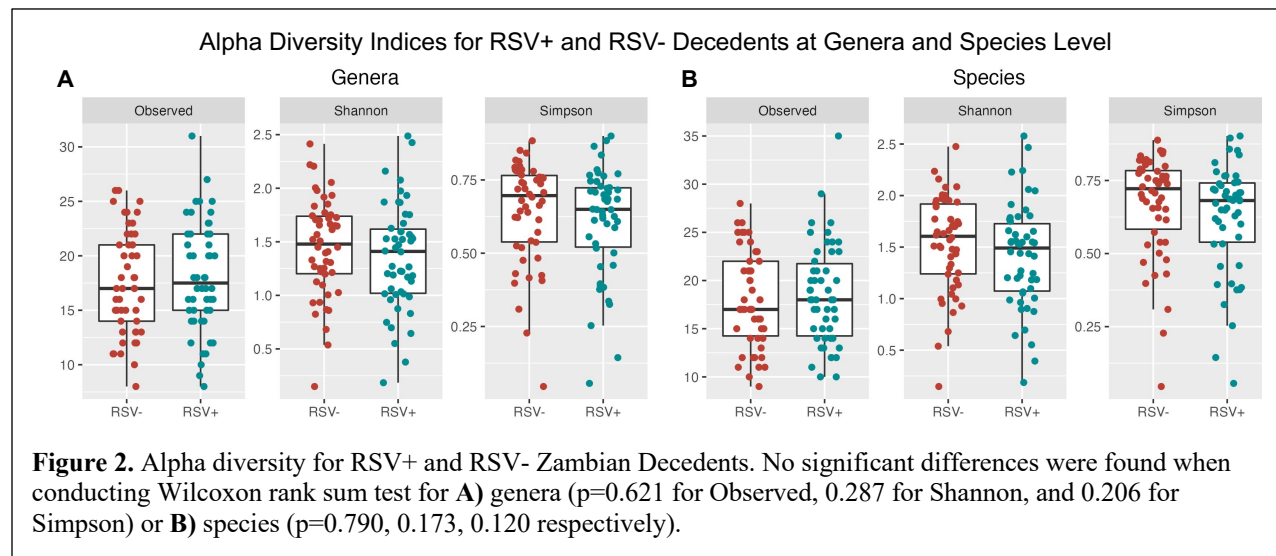
**Figure 1.** Microbial Composition of RSV- and RSV+ samples showing microbe diversity within each sample. On average, the five most abundant microbes in RSV+ samples were *Streptococcus* (22.2%), *Haemophilus* (19.7%), *Moraxella* (11.0%), *Klebsiella* (10.2%), and *Corynebacterium* (6.8%). The five most abundant microbes in RSV- samples were *Streptococcus* (29.0%), *Staphylococcus* (13.5%), *Escherichia* (8.5%), *Klebsiella* (6.4%) and *Haemophilus* (5.7%).

abundant microbes in RSV- samples were *Streptococcus* (29.0%), *Staphylococcus* (13.5%), *Escherichia* (8.5%), *Klebsiella* (6.4%) and *Haemophilus* (5.7%).

## Alpha Diversity

### *Shannon and Simpson Index*

Shannon and Simpson indices were calculated and a Shapiro-Wilk test was performed for each variable (Figure 2A). The Shannon indices for genera did not follow a normal distribution ( $p=0.985$  for RSV+ and  $p=0.860$  for RSV-), whereas the Simpson indices did follow a normal distribution ( $p=0.002$  for RSV+ and  $p<0.001$  for RSV-). A Wilcoxon rank sum test of significance was performed to determine significance of the differences in the Shannon and Simpson indices. We observed no significant difference in the alpha diversity between the RSV+ and RSV- decedents for genera ( $p=0.621$  for Observed,  $p=0.287$  for Shannon and  $p=0.206$  for Simpson). No significant alpha diversity differences were found when comparing species level taxa in the same manner ( $p=0.790$  for Observed, 0.173 for Shannon, 0.120 for Simpson) (Figure 2B).



## Beta Diversity

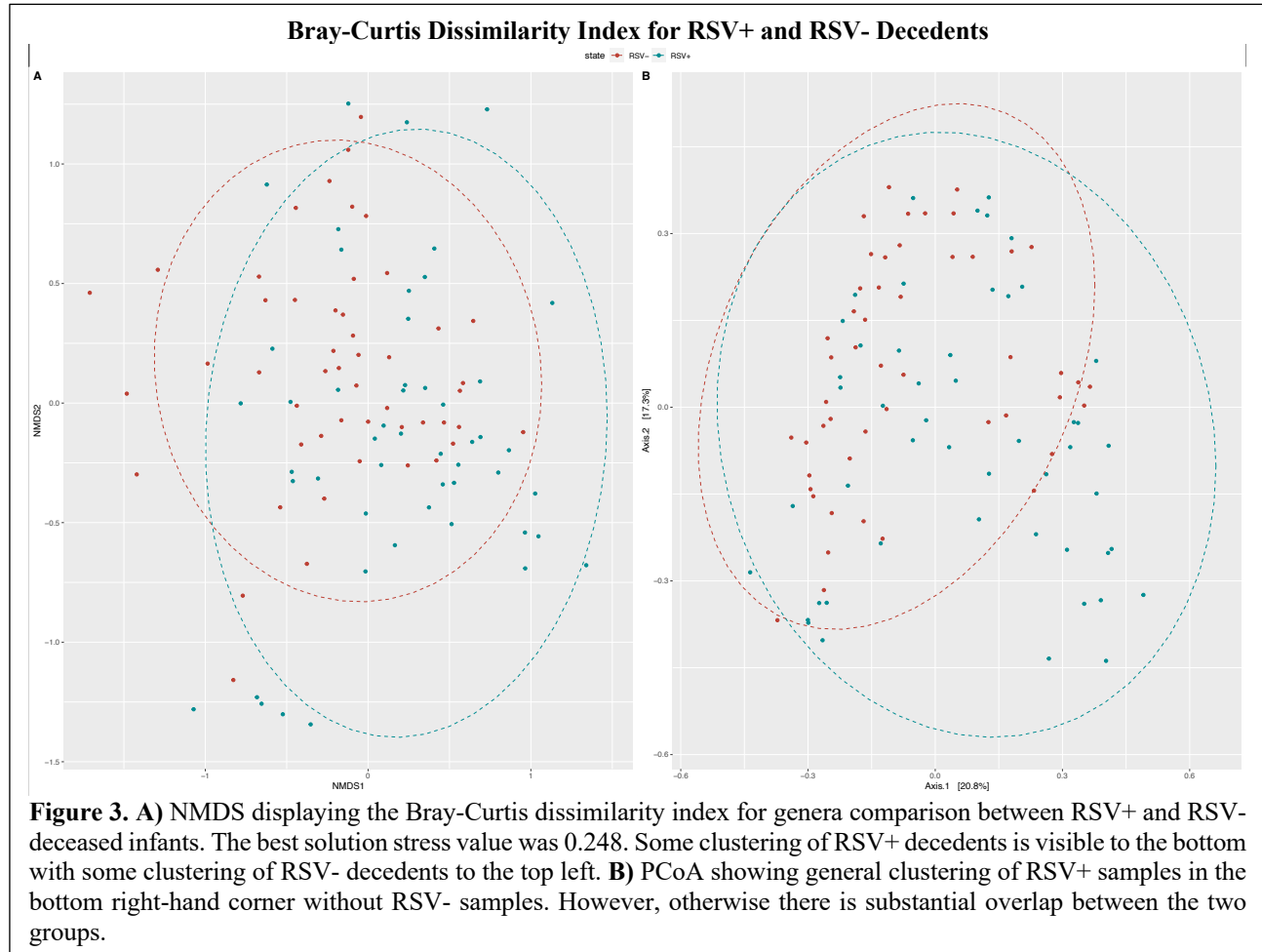
### *Bray-Curtis Dissimilarity Index*

We computed Bray-Curtis dissimilarity indices and visualized the results using NMDS and PCoA (Figure 3). Using PERMANOVA, we found that there was a significant difference in the centroids and dispersion between RSV groups ( $p<0.001$ ).

In the NMDS plot for genera, we observed clustering of RSV+ decedent samples to the bottom and RSV- decedent samples clustering to the upper left (Figure 3A). However, there was little spatial separation of clusters, and the stress value was high at 0.248. This indicates some differentiation between groups, but a high degree of overlap. NMDS for species also showed very little clustering by RSV status with a stress value of 0.262 (See Figure, Supplemental Digital Content 2A).

From the PCoA plot for genera, we see that the first two axes account for 38% of the variability

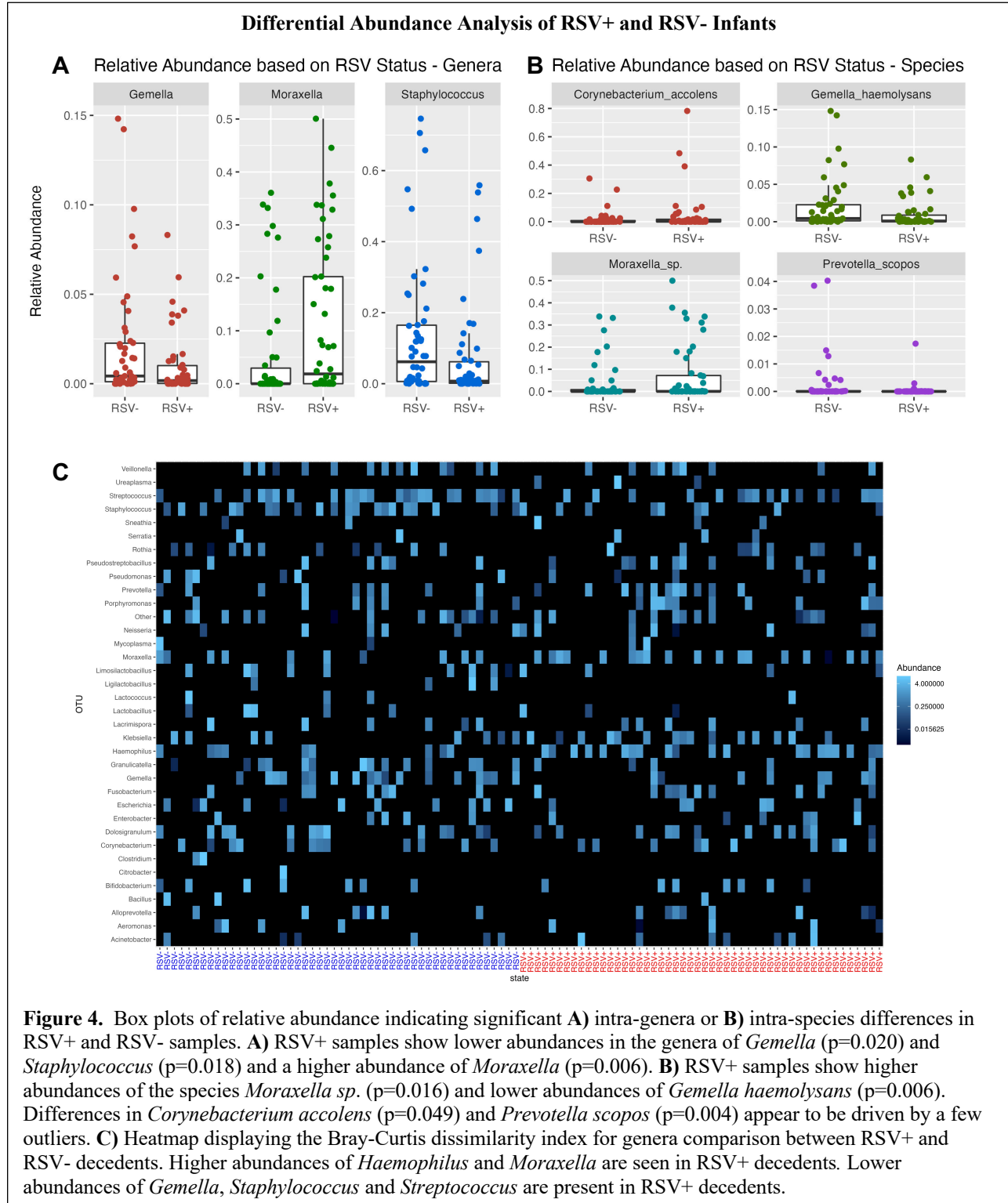
in the data (Figure 2B). Some clustering is visible in the RSV+ samples, hedging away from the RSV- samples in the bottom right-hand corner with overlap between samples throughout the remainder. As with the NMDS analysis, signs of distinct clustering are largely absent. PCoA for species showed mostly overlap between RSV+ and RSV- samples with some individually clustered RSV+ samples in the bottom left (See Figure, Supplemental Digital Content 2B).



### Differential Abundance Analysis

Differential abundance analysis via the non-parametric Wilcoxon rank sum test was performed to compare differences in relative microbial composition between RSV+ and RSV- samples. From this test, we determined that *Gemella* ( $p=0.020$ ), *Moraxella* ( $p=0.006$ ) and *Staphylococcus* ( $p=0.018$ ) all had marginally significant differences in median abundances between RSV+ and RSV- samples (See Table, Supplemental Digital Content 3).

We created box plots of genera abundances to visualize the inter-group differences. We observed that RSV+ decedents have higher abundances of *Gemella* and *Staphylococcus* and a lower abundance of *Moraxella* compared to RSV- decedents (Figure 4A). We also created a heatmap based on relative abundances. In the heatmap, we see subtle trends in abundance differences between the RSV+ and RSV- decedents consistent with our differential abundance results and box plots (See Figure 3A).



Differential abundance analysis at the species level, which is possible using PathoScope2.0, indicated significant differences for *Corynebacterium accolens* ( $p=0.049$ ), *Gemella haemolysans* ( $p=0.006$ ), *Moraxella sp.* ( $p=0.016$ ), and *Prevotella scopos* ( $p=0.004$ ).<sup>22-24</sup> The box plots in Figure



4B illustrate higher abundances of *Moraxella sp.* in RSV+ decedents and lower abundances of *Gemella haemolysans* when compared with RSV- decedents. Differences in *Corynebacterium accolens* ( $p=0.049$ ) and *Prevotella scopos* appear to be driven by a few outliers. Heatmap visualization revealed similar visible differences (See Figure, Supplemental Digital Content 4).

## Pathway Analysis

Using PICRUSt2, we inferred pathway abundances for each sample based on the microbial abundances computed with PathoScope.<sup>22–24,44</sup> We analyzed these abundances using a Wilcoxon rank sum test to determine which pathways were differentially present between RSV+ and RSV- decedents. This provided 20 different pathways; of these, 3 pathways were known to be correlated with *Haemophilus influenzae* (See Table, Supplemental Digital Content 5). Of the remaining 17 pathways, 10 were correlated with *E. Coli*, 4 correlated to a variety of plant species, and 3 correlated to other infectious diseases (*Mycobacterium tuberculosis* and various species of *Chlamydia*).

## DISCUSSION

Previous studies have shown that the microbial ecosystem in the nasopharynx, in which RSV acquisition occurs, may influence the host response to infection and certainly varies with disease severity.<sup>2,8–13</sup> In contrast with prior studies, which focused mainly on infants who were newly diagnosed with RSV and were acutely ill, our study focused on deceased infants who died of a respiratory cause, namely RSV. This focuses our results on infants who had experienced the most extreme outcome of RSV infection, death. Previous studies have also shown that antibiotic use leads to decreased alpha diversity and dysbiosis in the NP microbiome.<sup>14,15</sup> To avoid confounding by exposure to antibiotics, our analysis focused on comparing the microbiome profile of infants who died in the community. This includes infants who died in a facility with a stay less than 48 hours who would not be affected by nosocomial exposure due to limited time in the hospital.

We found no significant difference in genera or species richness or diversity in individual samples as measured by the Shannon and Simpson indexes. This indicates that the number of different species was roughly similar between groups but does not provide insight into whether the specific compositions differed. To explore how the specific compositions of the NP microbiome differed, we turned to examine beta diversity via Bray-Curtis analysis. NMDS and PCoA plots showed high degrees of overlap, but PERMANOVA indicated a significant difference between the two groups ( $p<0.001$ ). To explore which microbes influenced the difference in beta diversity, we analyzed differences in the relative abundance of microbial taxa between conditions. RSV+ decedents had higher abundances of *Moraxella* ( $p=0.006$ ) and lower relative abundances of *Gemella* ( $p=0.020$ ) and *Staphylococcus* ( $p=0.018$ ) when compared to RSV- decedents. We also noticed increased relative abundances of *Haemophilus* in RSV+ decedents, although these were not significant. We identified significant differential abundance in 3 pathways associated with *Haemophilus*. This confirms findings reported previously by Ederveen et al., Rosas-Salazar et al. and de Steenhuijsen Pijters et al.<sup>11–13</sup>

Ederveen et al. reported that an overabundance of *Haemophilus* and *Achromobacter* microbes and a loss of microbes like *Veillonella* are associated with infants who had been hospitalized with

RSV.<sup>12</sup> They also indicated that infants who recover from RSV have higher *Moraxella* abundances. Additionally, Rosas-Salazar et al. observed that infants diagnosed with RSV had higher abundances of *Haemophilus*, *Moraxella*, and *Streptococcus* whereas healthy patients had high abundances of *Corynebacterium* and *Staphylococcus*.<sup>11</sup> Also, de Steenhuijsen Piters et al. reported that in infants seen in a clinic or hospitalized for RSV, *H. influenzae* and *Streptococcus* are seen in higher abundance and correlated to disease severity.<sup>13</sup> Our study found similar results in infants who died with RSV in *Haemophilus* and *Moraxella*. Differences in our study and those by these mentioned studies may be attributed to the fact their study participants received medical care in a clinical or hospital setting. While our infants represent the most severe disease status (resulting in death), none of the infants in our study were treated for more than 48 hours in a clinical setting before passing away and most of the cases in our study did not receive any exposure to a clinical setting in the course of disease. Differences in some of the microbes, such as *Streptococcus*, may be due to exposure to clinical settings and/or usage of antibiotics. Our results confirm that higher abundances of *Haemophilus* and *Moraxella* and decreases in *Staphylococcus* are correlated with severe RSV infection.

Our study had several key limitations. For one, we did not have complete data on how the infants in either population died. Despite the adjudication process (See Text, Supplemental Digital Content 1), there is a chance that the RSV+ group may have died from issues unrelated to RSV or respiratory disease which could skew our results. Compounding this, information about the final HIV status of the decedents is not available and HIV status of the mothers may not be accurately captured since mother's HIV status was not always noted in medical charts or death records and was not inquired about during verbal autopsies. HIV exposure has been shown to cause NP microbiome dysbiosis and may affect our results.<sup>45,46</sup> Furthermore, many of the decedents, particularly those who had died in a facility, may have been treated with antibiotics. While our focus on community deaths was intended to reduce the confounding effect of antibiotics, we could not exclude the possibility that some decedents were exposed to antibiotics without our knowledge. We also do not have information regarding use of antibiotics earlier in life which may also disrupt the nasopharyngeal microbiome.<sup>15</sup> Additionally, verbal autopsies did not yield accurate information about antibiotic exposure. Lastly, the impact of RSV on microbial ecology may not necessarily be confined to the acute RSV infection period but could occur at some time removed from the initial infection. The cross-sectional nature of our data set, focusing on time point shortly after death, could not detect those delayed effects.

With those caveats, our observations are consistent with what has been identified in previous research by Ederveen et al., Rosas-Salazar et al. and de Steenhuijsen Piters et al.

## CONCLUSION

Our findings support those findings previously published and indicate that RSV may be correlated with higher abundances of *Moraxella* and *Haemophilus* and lower abundances of *Staphylococcus*. Our study also points to decreased abundances of *Gemella*.

Further research should be done to confirm these findings and specifically to determine if the change in microbiome proceeds RSV infection or is an effect of infection. Additional research looking at community-acquired RSV would also allow greater understanding of microbe

differences without exposure to clinical settings or antibiotics. Better understanding may allow for better identification of at-risk infants and lead to better treatment and care.

**Author Contribution Statement:** CJG and LM were the co-principal investigators for the ZPRIME study. WBM was the ZPRIME study statistician and provided all needed data from the ZPRIME study for our analysis. AI oversaw the lab that generated the microbiome samples from ZPRIME for our study. RP was the ZPRIME project manager and aided in collaboration between our teams. RL contributed to the implementation of the ZPRIME study, including with the adjudication process. WEJ oversaw all data and statistical analysis for the current study and provided recommendations on best practices. NK completed initial data exploration and analysis. JM revised and completed the data visualization, data analysis and statistical analysis. She also took the lead on writing the manuscript. AROM provided support for data and statistical analysis, template code for sequencing, animalcules and PICRUST analysis, and provided critical feedback on the manuscript. JM, AROM, CJG, WBM, RP, RL and WEJ discussed the results, contributed to the ideas contained in the final manuscript, and helped prepare the manuscript. All authors reviewed the final manuscript.

**Data Access Statement:** Raw data is available on the sequence read archive, accession number PRJNA913857. All processed data and code used for data preprocessing, statistical analysis, and figure generation is available on GitHub ([https://github.com/jessmcc22/ZPRIME\\_RSV](https://github.com/jessmcc22/ZPRIME_RSV)). Sample IDs are consistent across microbiome and demographic information on the sequence read archive and GitHub. Full results for all analyses are also available on GitHub.

**Ethics Statement:** The ZPRIME study, where our data originates, was approved by the ethical review boards at Boston University Medical Center and the University of Zambia. Written informed consent was obtained from the decedents' next of kin or guardian.

**Sources of Support:** The work was supported by the Bill and Melinda Gates Foundation; W.E.J and J.M. were funded in part by NIH grant R01 GM127430; W.E.J. and A.R.O.M were supported in part by the NIH under grant R21 AI154387

**Conflicts of Interest Statement:** None declared

### **List of Supplemental Digital Content**

Supplemental Digital Content 1. Text: Methods continued

Supplemental Digital Content 2. Figure: Species Level Bray-Curtis Dissimilarity Index

Supplemental Digital Content 3. Table: p-values and adjusted p-values for RSV Differential Abundance Analysis

Supplemental Digital Content 4. Figure: Species Level Differential Abundance Heatmap

Supplemental Digital Content 5. Table: PICRUST Pathway Analysis Significant Results

Supplemental Digital Content 6. Table: STORMS Checklist

## REFERENCES

1. Navarro Alonso JA, Bont LJ, Bozzola E, Herting E, Lega F, Mader S, Nunes MC, Ramilo O, Valiotis G, Olivier CW, Yates A, Faust SN. RSV: perspectives to strengthen the need for protection in all infants. *Emerging Themes in Epidemiology*. 2021.
2. O'Brien KL, Baggett HC, Brooks WA, Feikin DR, Hammitt LL, Higdon MM, Howie SRC, Deloria Knoll M, Kotloff KL, Levine OS, Madhi SA, Murdoch DR, Prospero C, Scott JAG, Shi Q, Thea DM, Wu Z, Zeger SL, Adrian P v., Akarasewi P, Anderson TP, Antonio M, Awori JO, Baillie VL, Bunthi C, Chipeta J, Chisti MJ, Crawley J, DeLuca AN, Driscoll AJ, Ebruke BE, Endtz HP, Fancourt N, Fu W, Goswami D, Groome MJ, Haddix M, Hossain L, Jahan Y, Kagucia EW, Kamau A, Karron RA, Kazungu S, Kourouma N, Kuwanda L, Kwenda G, Li M, Machuka EM, Mackenzie G, Mahomed N, Maloney SA, McLellan JL, Mitchell JL, Moore DP, Morpeth SC, Mudau A, Mwananyanda L, Mwansa J, Silaba Ominde M, Onwuchekwa U, Park DE, Rhodes J, Sawatwong P, Seidenberg P, Shamsul A, Simões EAF, Sissoko S, Wa Somwe S, Sow SO, Sylla M, Tamboura B, Tapia MD, Thamthitiwat S, Toure A, Watson NL, Zaman K, Zaman SMA. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *The Lancet*. 2019;394(10200).
3. Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, Singleton RJ, O'Brien KL, Roca A, Wright PF, Bruce N, Chandran A, Theodoratou E, Sutanto A, Sedyaningsih ER, Ngama M, Munywoki PK, Kartasasmita C, Simões EA, Rudan I, Weber MW, Campbell H. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. *The Lancet*. 2010;375(9725).
4. Midulla F, Nenna R, Scagnolari C, Petrarca L, Frassanito A, Viscido A, Arima S, Antonelli G, Pierangeli A. How Respiratory Syncytial Virus Genotypes Influence the Clinical Course in Infants Hospitalized for Bronchiolitis. *Journal of Infectious Diseases*. 2019;219(4).
5. Staadegaard L, Caini S, Wangchuk S, Thapa B, de Almeida WAF, de Carvalho FC, Njouom R, Fasce RA, Bustos P, Kyncl J, Novakova L, Caicedo AB, de Mora Coloma DJ, Meijer A, Hooiveld M, Huang S, Wood T, Guiomar R, Rodrigues AP, Danilenko D, Stolyarov K, Lee VJM, Ang LW, Cohen C, Moyes J, Larrauri A, Delgado-Sanz C, Le MQ, Hoang PVM, Demont C, Bangert M, van Summeren J, Dückers M, Paget J. The Global Epidemiology of RSV in Community and Hospitalized Care: Findings from 15 Countries. *Open Forum Infect Dis*. 2021;8(7).
6. Shi T, McAllister DA, O'Brien KL, Simoes EAF, Madhi SA, Gessner BD, Polack FP, Balsells E, Acacio S, Aguayo C, Alassani I, Ali A, Antonio M, Awasthi S, Awori JO, Azziz-Baumgartner E, Baggett HC, Baillie VL, Balmaseda A, Barahona A, Basnet S, Bassat Q, Basualdo W, Bigogo G, Bont L, Breiman RF, Brooks WA, Broor S, Bruce N, Bruden D, Buchy P, Campbell S, Carosone-Link P, Chadha M, Chipeta J, Chou M, Clara W, Cohen C, de Cuellar E, Dang DA, Dash-yandag B, Deloria-Knoll M, Dherani M, Eap T, Ebruke BE, Echavarria M, de Freitas Lázaro Emediato CC, Fasce RA, Feikin DR, Feng L, Gentile A, Gordon A, Goswami D, Goyet S, Groome M, Halasa N, Hirve S, Homaira N, Howie SRC, Jara J, Jroundi I, Kartasasmita CB, Khuri-Bulos N, Kotloff KL, Krishnan A, Libster R, Lopez O, Lucero MG, Lucion F, Lupisan SP, Marccone DN, McCracken JP, Mejia M, Moisi JC, Montgomery JM, Moore DP, Moraleda C, Moyes J, Munywoki P,

- Mutyara K, Nicol MP, Nokes DJ, Nymadawa P, da Costa Oliveira MT, Oshitani H, Pandey N, Paranhos-Baccalà G, Phillips LN, Picot VS, Rahman M, Rakoto-Andrianarivelo M, Rasmussen ZA, Rath BA, Robinson A, Romero C, Russomando G, Salimi V, Sawatwong P, Scheltema N, Schweiger B, Scott JAG, Seidenberg P, Shen K, Singleton R, Sotomayor V, Strand TA, Sutanto A, Sylla M, Tapia MD, Thamthitawat S, Thomas ED, Tokarz R, Turner C, Venter M, Waicharoen S, Wang J, Watthanaworawit W, Yoshida LM, Yu H, Zar HJ, Campbell H, Nair H. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *The Lancet*. 2017;390(10098).
7. Gill CJ, Mwananyanda L, MacLeod WB, Kwenda G, Pieciak R, Mupila Z, Murphy C, Chikoti C, Forman L, Berklein F, Lapidot R, Chimoga C, Ngoma B, Larson A, Lungu J, Nakazwe R, Nzara D, Pemba L, Yankonde B, Chirwa A, Mwale M, Thea DM. Infant deaths from respiratory syncytial virus in Lusaka, Zambia from the ZPRIME study: a 3-year, systematic, post-mortem surveillance project. *Lancet Glob Health*. 2022;10(2).
  8. Schippa S, Frassanito A, Marazzato M, Nenna R, Petrarca L, Neroni B, Bonfiglio G, Guerrieri F, Frasca F, Oliveto G, Pierangeli A, Midulla F. Nasal microbiota in RSV bronchiolitis. *Microorganisms*. 2020;8(5).
  9. Weinberger DM, Klugman KP, Steiner CA, Simonsen L, Viboud C. Association between Respiratory Syncytial Virus Activity and Pneumococcal Disease in Infants: A Time Series Analysis of US Hospitalization Data. *PLoS Med*. 2015;12(1).
  10. Man WH, van Houten MA, Mérelle ME, Vlieger AM, Chu MLJN, Jansen NJG, Sanders EAM, Bogaert D. Bacterial and viral respiratory tract microbiota and host characteristics in children with lower respiratory tract infections: a matched case-control study. *Lancet Respir Med*. 2019;7(5).
  11. Rosas-Salazar C, Shilts MH, Tovchigrechko A, Chappell JD, Larkin EK, Nelson KE, Moore ML, Anderson LJ, Das SR, Hartert T v. Nasopharyngeal microbiome in respiratory syncytial virus resembles profile associated with increased childhood asthma risk. *American Journal of Respiratory and Critical Care Medicine*. 2016.
  12. Ederveen THA, Ferwerda G, Ahout IM, Vissers M, de Groot R, Boekhorst J, Timmerman HM, Huynen MA, van Hijum SAFT, de Jonge MI. Haemophilus is overrepresented in the nasopharynx of infants hospitalized with RSV infection and associated with increased viral load and enhanced mucosal CXCL8 responses. *Microbiome*. 2018;6(1).
  13. de Steenhuijsen Piters WAA, Heinonen S, Hasrat R, Bunsow E, Smith B, Suarez-Arrabal MC, Chaussabel D, Cohen DM, Sanders EAM, Ramilo O, Bogaert D, Mejias A. Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. *Am J Respir Crit Care Med*. 2016;194(9).
  14. Flynn M, Dooley J. The microbiome of the nasopharynx. *Journal of Medical Microbiology*. 2021.
  15. Raita Y, Toivonen L, Schuez-Havupalo L, Karppinen S, Waris M, Hoffman KL, Camargo CA, Peltola V, Hasegawa K. Maturation of nasal microbiota and antibiotic exposures during early childhood: a population-based cohort study. *Clinical Microbiology and Infection*. 2021;27(2).
  16. Hament JM, Aerts PC, Fleer A, van Dijk H, Harmsen T, Kimpfen JLL, Wolfs TFW. Enhanced adherence of *Streptococcus pneumoniae* to human epithelial cells infected with respiratory syncytial virus. *Pediatr Res*. 2004;55(6).

17. Hament JM, Aerts PC, Fleer A, van Dijk H, Harmsen T, Kimpen JLL, Wolfs TFW. Direct binding of respiratory syncytial virus to pneumococci: A phenomenon that enhances both pneumococcal adherence to human epithelial cells and pneumococcal invasiveness in a murine model. *Pediatr Res*. 2005;58(6).
18. Sande CJ, Njunge JM, Mwangeli Ngoi J, Mutunga MN, Chege T, Gicheru ET, Gardiner EM, Gwela A, Green CA, Drysdale SB, Berkley JA, Nokes J, Pollard AJ. Airway response to respiratory syncytial virus has incidental antibacterial effects. *Nat Commun*. 2019;10(1).
19. Zar HJ, Nduru P, Stadler JAM, Gray D, Barnett W, Lesosky M, Myer L, Nicol MP. Early-life respiratory syncytial virus lower respiratory tract infection in a South African birth cohort: epidemiology and effect on lung health. *Lancet Glob Health*. 2020;8(10).
20. Verhoeven D, Xu Q, Pichichero ME. Differential impact of respiratory syncytial virus and parainfluenza virus on the frequency of acute otitis media is explained by lower adaptive and innate immune responses in otitis-prone children. *Clinical Infectious Diseases*. 2014;59(3).
21. Wang L, Piedra PA, Avadhanula V, Durigon EL, Machabishvili A, López MR, Thornburg NJ, Peret TCT. Duplex real-time RT-PCR assay for detection and subgroup-specific identification of human respiratory syncytial virus. *J Virol Methods*. 2019;271.
22. Faits T, Odom-Mabey AR, Castro-Nallar E, Crandall KA, Johnson WE. Metagenomic profiling pipelines improve taxonomic classification for 16S amplicon sequencing data. *bioRxiv [Internet]*. 2022 Jan 1;2022.07.27.501757. Available from: <http://biorxiv.org/content/early/2022/07/29/2022.07.27.501757.abstract>
23. Byrd AL, Perez-Rogers JF, Manimaran S, Castro-Nallar E, Toma I, McCaffrey T, Siegel M, Benson G, Crandall KA, Johnson WE. Clinical PathoScope: Rapid alignment and filtration for accurate pathogen identification in clinical samples using unassembled sequencing data. *BMC Bioinformatics*. 2014;15(1).
24. Hong C, Manimaran S, Shen Y, Perez-Rogers JF, Byrd AL, Castro-Nallar E, Crandall KA, Johnson WE. PathoScope 2.0: A complete computational framework for strain identification in environmental or clinical sequencing samples. *Microbiome*. 2014;2(1).
25. Rajaram S, Oono Y. NeatMap - non-clustering heat map alternatives in R. *BMC Bioinformatics*. 2010;11.
26. McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One*. 2013;8(4).
27. Hollander M, Wolfe DA, Chicken E. *Nonparametric statistical methods*. Nonparametric Statistical Methods. 2015.
28. Bauer DF. Constructing confidence sets using rank statistics. *J Am Stat Assoc*. 1972;67(339).
29. Team RC. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. 2021.
30. Shaffer JP. Multiple hypothesis testing. *Annu Rev Psychol*. 1995;46(1).
31. Shannon CE. A Mathematical Theory of Communication. *Bell System Technical Journal*. 1948;27(3).
32. Simpson EH. Measurement of diversity [16]. *Nature*. 1949.
33. Royston P. Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. *Appl Stat*. 1995;44(4).
34. Royston JP. Algorithm AS 181: The W Test for Normality. *Appl Stat*. 1982;31(2).

35. Royston JP. An Extension of Shapiro and Wilk's W Test for Normality to Large Samples. *Appl Stat.* 1982;31(2).
36. Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr.* 1957;27(4).
37. Minchin PR. An evaluation of the relative robustness of techniques for ecological ordination. *Vegetatio.* 1987;69(1-3).
38. Faith DP, Minchin PR, Belbin L. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio.* 1987;69(1-3).
39. Jones MC, Aitchison J. The Statistical Analysis of Compositional Data. *J R Stat Soc Ser A.* 1987;150(4).
40. McArdle BH, Anderson MJ. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology.* 2001;82(1).
41. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-Like Differential Expression (ALDEx) Analysis for Mixed Population RNA-Seq. *PLoS One.* 2013 Jul 2;8(7):e67019.
42. Fernandes AD, Reid JN, Macklaim JM, McMurrough TA, Edgell DR, Gloor GB. Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome.* 2014 Dec 5;2(1):15.
43. Gloor GB, Macklaim JM, Fernandes AD. Displaying Variation in Large Datasets: Plotting a Visual Summary of Effect Sizes. *Journal of Computational and Graphical Statistics.* 2016;25(3).
44. Douglas GM, Maffei VJ, Zaneveld JR, Yurgel SN, Brown JR, Taylor CM, Huttenhower C, Langille MGI. PICRUSt2 for prediction of metagenome functions. *Nature Biotechnology.* 2020.
45. Odom-Mabey AR, Gill CJ, Pieciak R, Ismail A, Thea D, MacLeod WB, Johnson WE, Lapidot R. Characterization of longitudinal nasopharyngeal microbiome patterns in maternally HIV-exposed Zambian infants. *Gates Open Res.* 2022 Nov 11;6:143.
46. Bender JM, Li F, Martelly S, Byrt E, Rouzier V, Leo M, Tobin N, Pannaraj PS, Adisetiyo H, Rollie A, Santiskulvong C, Wang S, Autran C, Bode L, Fitzgerald D, Kuhn L, Aldrovandi GM. Maternal HIV infection influences the microbiome of HIV-uninfected infants. *Sci Transl Med.* 2016;8(349).