Dysbiosis signatures of fecal microbiota in South African infants with respiratory, gastrointestinal and other diseases

Srinivasan Krishnamoorthy, PhD,[a] Vinet Coetzee, PhD,[b] Johanita Kruger, PhD,[a] Hanneke Potgieter, MSc,[c] Elna M Buys, PhD[a]

[a]Department of Consumer and Food Sciences,

[b]Department of Genetics and Institute for Food, Nutrition, and Well-being, and

[c]Department of Consumer and Food Sciences and Institute for Food, Nutrition, and Well-being, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa.

Funding Source:

Supported by the Institute for Food, Nutrition and Well-being and Genomics Research Institute, University of Pretoria, South Africa.

Financial Disclosure: The authors do not have any financial relationships relevant to this article to disclose.

Conflict of Interest: The authors do not have any potential conflicts of interest to disclose. Also, Dr. Surendra Vikram mentioned in the acknowledgement section do not have any potential conflicts of interest to disclose.

Clinical Trial Registration: Not applicable

Abbreviations:

ANCOM – Analysis of compositional variance

OTU - Operational taxonomic unit

QIIME - Quantitative insights into microbial ecology

U5MR - Under the age of 5 mortality rate

Abstract

Objective

To determine the association between the fecal microbiota diversity of the infants with different disease conditions, and vitamin A supplementation, antibiotic, and deworming therapies.

Study design

In this case-control study, the bacterial community variations and the potential pathogens were identified through 16S ribosomal RNA gene-based amplicon sequencing and quantitative insights into microbial ecology (QIIME) pipeline in fecal samples. The participants were South African infants (Mean age: 16 ± 8 months; 17 male and 17 female) hospitalized and diagnosed with gastrointestinal, respiratory and other diseases.

Results

The top phyla of the infants with respiratory disease were Proteobacteria, followed by Firmicutes, which were equally abundant in gastrointestinal disease. A significant difference in Shannon (alpha) diversity index (95% CI, 2.6 – 4.4; P = .008), among the microbiota of the fecal samples categorized by disease conditions, was observed. In beta diversity analysis of fecal microbiota, remarkable variations were found within the groups of deworming therapy (95% CI, 0.40 – 0.90; P = .033), disease conditions (95% CI, 0.44 – 0.86; P < .012) through unweighted and antibiotic therapy (95% CI, 0.20 – 0.75; P = .007), vitamin A intake (95% CI, 0.10 – 0.80; P < .033) and disease conditions (95% CI, 0.10 – 0.79; P = .006) through weighted UniFrac distances. The candidate pathogen associated with the disease groups were identified through analysis of the composition of microbiomes analysis.

Conclusions

This study provides preliminary evidence for the fecal microbiome-derived dysbiosis signature and pathobiome concept that may be observed in young children during illness.

Keywords: pathobiome, microbiome, QIIME, taxonomy, Sub-Saharan Africa

Introduction

The global burden of child death, especially under the age of 5 mortality rate (U5MR), remains a challenge, and the highest U5MR was found in sub-Saharan Africa countries between 1990 and 2015. If this trend continues, 53% of the children under the age of 5 will die in the next 15 years in these countries.[1] Despite the global progress in U5MR scenario, an estimated 5.4 million children under 5 died in 2017, roughly half of which lived in sub-Saharan Africa countries.[2] One of the Sustainable Development Goals - 2019 targets is to end all forms of under-five child mortality globally by the year 2025.[3] The leading causes of U5MR in sub-Saharan Africa are respiratory and gastrointestinal diseases.[4] In critically ill pediatric patients, the gut and lung microbiome undergo profound changes.[5] The impact of the metabolites produced by the gut microbial community might modulate immunity and cause disorders in distant organs such as the upper and lower respiratory system.[6]

The gut microbiota dysbiosis is directly associated with intestinal disorders and numerous extra-intestinal diseases such as respiratory and neurological illness. The ratio of Firmicutes and Bacteroidetes is significantly lesser in infants with inflammatory bowel diseases.[7] Lower abundances of Lachnospiraceae, Faecalibacterium, Veillonellaceae, and Ruminococcaceae, which all belongs to Firmicutes phylum were observed in the fecal samples of the infants at the risk of asthma.[6] A non-invasive method in the assessment of diseases in the gut-lung axis among the infants under five years is typing the biomarker proteins and pathogens of their stool samples.[8]

These findings propel the hypothesis that the fecal microbiota composition and its variations pattern could represent the infant's disease conditions. The "one pathogen per disease" postulated by Robert Koch has been replaced by the pathobiome hypothesis, which states that human diseases are outcomes of a complex, interconnected network of disease-promoting microbial communities.[7] The identification of distinct pathobiome profiles

corresponding to a specific disease, therefore, serve as a novel therapeutic tool.[9] There is no study deciphering the fecal microbiota pattern in sub-Saharan Africa infants affected by respiratory and gastrointestinal diseases. Overall, this study aimed to identify and compare the fecal microbiome of 34 hospitalized South African infants diagnosed with respiratory disease (RD), gastrointestinal disease (GD) and other diseases (OD) using the 16s rRNA amplicon sequencing and quantitative insights into microbial ecology (QIIME) package.

Methods

Thirty-four South African infants (mean age, 16 ± 8 months; 17 male and 17 female) were recruited at the Cecilia Maki-wane Hospital, East London, South Africa. The infants were hospitalized for respiratory (n = 16 [47%]), gastrointestinal (n = 11 [33%]), and other diseases (n = 7 [20%]). Basic demographic information, medical history, blood parameters, and fecal samples were collected from each infant. Fecal samples were categorized and grouped based on the disease diagnosis report and the medical history from the hospital (Table I and Table II)[10,11]. Written consent was obtained from one of the parents or guardians of each infant. The study was approved by research ethics committees at the Faculty of Health Sciences (256/2016) and the Faculty of Natural and Agricultural Sciences (EC 160504-025), University of Pretoria, South Africa. Formal permission was also obtained from the Cecilia Makiwane Hospital, Eastern Cape, South Africa.

The baseline characteristics of the infants were derived through descriptive statistics using IBM SPSS statistics version 25 (SPSS Inc, Chicago, Illinois), and the categorical variables were summarized as a median, percentage, and IQR owing to a skewed distribution (a = 0.05). The clinical parameter ranges were adapted from commonly used cut-offs for pediatric patients based on the unweighted normal ranges of biochemical parameters by sex and race-standardized by Laboratory Reference Intervals in Africa, 2012.[12]

Table I. Criteria for categorizing the disease groups

Respiratory diseases[10]
- Pneumonia
- Lower respiratory tract infection
- Pneumonia meningitis
- Bronchiolitis
- Pulmonary tuberculosis
- Lobar pneumonia
- Tuberculous meningitis

Gastrointestinal diseases[11]
- Acute gastroenteritis
- Abdominal obstruction
- Diarrhea

Other diseases
- Severe acute malnutrition
- Anemia
- Trabeculotomy
- Congenital strabismus
- Cardiomegaly
- Failure to thrive

Table II. Subset data for the infants hospitalized with various disease conditions

Samples	Age (mo)	Sex*	Vitamin A intake†	Deworming therapy†	Vomiting†	Diarrhea†	Antibiotic intake†	Clinically diagnosed diseases
1	17	2	0	0	1	0	1	Pneumonia meningitis (respiratory disease)
2	13	1	0	0	0	0	0	Pulmonary tuberculosis (respiratory disease)
3	14	1	1	1	0	0	1	Acute gastroenteritis (gastrointestinal disease)
4	13	2	0	1	1	1	1	Bronchiolitis (respiratory disease)
5	14	2	1	1	0	0	0	Lobar pneumonia (respiratory disease)
6	11	2	1	0	0	0	0	Abdominal obstruction (gastrointestinal disease)
7	8	2	1	0	1	0	1	Lower respiratory tract infection (respiratory disease)
8	8	2	1	0	1	1	1	Acute gastroenteritis (gastrointestinal disease)
9	6	1	1	0	0	0	0	Bronchiolitis (respiratory disease)
10	18	2	0	1	0	0	1	Pneumonia (respiratory disease)
11	11	1	1	0	0	0	1	Pulmonary tuberculosis (respiratory disease)
12	6	2	1	0	1	1	0	Acute gastroenteritis (gastrointestinal disease)
13	23	2	1	1	0	0	1	Pulmonary tuberculosis (respiratory disease)
14	6	1	1	0	0	0	1	Acute gastroenteritis (gastrointestinal disease)
15	17	1	1	1	0	0	0	Lower respiratory tract infection (respiratory disease)
16	15	1	1	1	0	0	1	Pneumonia (respiratory disease)
17	10	1	0	0	0	1	1	Acute gastroenteritis (gastrointestinal disease)
18	12	1	0	0	1	0	1	Lobar pneumonia (respiratory disease)
19	6	1	1	1	0	1	1	Abdominal obstruction (gastrointestinal disease)
20	5	2	1	1	0	0	0	Tuberculous meningitis (respiratory disease)
21	17	1	0	1	0	0	0	Congenital strabismus (other diseases)
22	6	2	0	0	0	0	1	Cardiomegaly (other diseases)
23	12	2	0	0	0	0	1	Pulmonary tuberculosis (respiratory disease)
24	21	2	1	0	0	0	0	Pneumonia (respiratory disease)
25	19	1	0	1	0	1	1	Acute gastroenteritis (gastrointestinal disease)
26	20	1	1	1	0	1	1	Anemia (other diseases)
27	22	2	0	1	1	0	1	Acute gastroenteritis (gastrointestinal disease)
28	12	2	0	0	1	1	1	Acute gastroenteritis (gastrointestinal disease)
29	6	1	1	0	0	1	1	Lower respiratory tract infection (respiratory disease)
31	24	1	1	1	0	0	0	Congenital strabismus (other diseases)
32	15	1	0	1	1	0	0	Cardiomegaly (other diseases)
33	21	2	1	0	0	0	1	Failure to thrive (other diseases)
34	8	2	0	0	0	1	1	Diarrhea (gastrointestinal disease)
35	12	1	0	1	0	0	1	Severe acute malnutrition (other diseases)

CRP, C-reactive protein, *Hb*, Hemoglobin, *WBC*, White blood cell count.
*1, Male; 2, female.
†0, No; 1, yes.

Fecal sample collection and metagenomics analysis

Fecal samples were collected through a standardized protocol at the time of hospitalization in a sterile 13 ml stool tube with DNA/RNA shield. The samples were frozen (−20 °C) immediately after collection. The frozen stool samples were shipped with dry ice to Inqaba Biotechnical Industries (Pty) Ltd., South Africa, where the microbiota was identified and characterized via the V3-V4 16S rRNA gene fragments through Illumina Miseq sequencing platform.

Taxonomic composition and 16s rRNA sequence diversity analysis through QIIME software

The fecal microbiome data were processed through the software package "QIIME2" version 2018.11.[13] Using the ribosomal database project classifier algorithm in QIIME, the 16S rRNA gene sequence-based taxonomy was classified. Briefly, the paired-end cassava format sequences were imported into QIIME prior to quality filtering. A total of 3,00,019 valid sequences were generated from the 34 infants fecal DNA samples, and after filtering and trimming, 2,56,429 high-quality reads were obtained. The sequences were demultiplexed and clustered using the dada2 plugin[14]. A sequencing depth cut-off value of 2600 sequences from each sample was set and designed into OTUs. The alpha diversity (microbial diversity within samples) and richness indices of the samples were measured by observed species OTUs and the Shannon diversity index and were statistically tested using the Kruskal-Wallis method. The beta diversity (community diversity divergence between samples) was evaluated using the UniFrac principal coordinate analysis (PCoA) to explore the potential factors that could explain the grouping pattern of similar communities[15] and tested through non-parametric permutational multivariate analysis of variance (PERMANOVA). Analysis of the composition of microbiomes (ANCOM) test was run among the disease conditions to determine for any significant differences in the relative abundance of any taxa.

Results

Study population and baseline characteristics

Study subjects were grouped into three groups: the respiratory disease group (n = 16), the gastrointestinal disease group (n = 11) and the other diseases group (n = 7) according to the criteria defined in Table 1. Table III shows the baseline characteristic data for each participant group. Among the three disease groups, the occurrence of diarrhea in the GD group (64%) was higher than in the RD group (12%) and the OD group (14%). In line with the concept that the gut microbiota and its metabolites are playing a role on the gut-lung axis, around 36% of the GD group infants also reported with upper and lower respiratory tract infection. Mean C-reactive protein (CRP) level was higher, with 7.7 mg/dL in the GD group and a maximum value of 27.3 mg/dL.

Table III. Baseline characteristics of participants in the respiratory, gastrointestinal, and other disease groups

Characteristics	Respiratory disease group (n = 16)	Gastrointestinal disease group (n = 11)	Other diseases group (n = 7)
Sex			
Male (n)	7	5	5
Female (n)	9	6	2
Age (months)	13 [5-21]	13 [6-24]	12 [6-23]
Clinical findings			
Diarrhea, n (%)	2 (12)	7 (64)	1 (14)
Vomiting, n (%)	4 (25)	4 (25)	1 (14)
WBC ($\times 10^3/\mu L$)			
5.0-10.0*	13.4 [5.6-26.8]	15.7 [5.3-51.3]	12.1 [5.2-21.0]
Albumin, (g/dL)			
3.5-5.5*	3.4 [2.7-4.2]	3.7 [3.4-4.2]	3.6 [2.8-4.5]
CRP, (mg/dL)			
4.0-8.0*	6.5 [0.3-27.0]	7.7 [0.3-27.3]	4.5 [1.4-16.8]
Hemoglobin, (g/dL)			
11.0-14.0*	10.3 [8.1-11.9]	9.9 [6.6-11.7]	10.5 [9.2-11.7]

CRP, C-reactive protein; *WBC*, white blood cell count.
Continuous variables are given as median [IQR] unless specified otherwise. In the statistical comparison between the disease groups, no significant differences were observed (details not provided here).
*Normal ranges in a healthy pediatric patient.[16]

Taxonomic analysis of infant fecal microbiota

The 16s rRNA OTU clustering analysis showed that Proteobacteria was the most abundant phyla across all samples, followed by Firmicutes (Figure 1, A). Alphaproteobacteria, Bacilli, Clostridia, Gammaproteobacteria, Betaproteobacteria, Mammalia, Deinococci, Chloroflexi, Sphingobacteria, Liliopsida, Bryopsida, Deltaproteobacteria, Gymnostomatea, and Thermotogae were found at the class level. The phyla level distribution of OTUs across the categories of diseases is given in Table IV (available at www.jpeds.com). The dominant phyla in the order of Proteobacteria, Firmicutes, Bacteroidetes, and Actinobacteria were found in the respiratory disease infants group, and Firmicutes, Proteobacteria, Actinobacteria, and Bacteroidetes were found in the gastrointestinal disease infants group. In the other diseases group, Proteobacteria followed by Firmicutes were the dominant phyla. Comparing with the other groups, the other diseases group have a greater relative abundance of Bacteroidetes (11%) and Actinobacteria (7%), which are 5% and 4% in the respiratory disease group and 0.2% and 2.0% in the gastrointestinal disease group, respectively. Although observable fluctuations were low in the relative abundance of phyla in the disease categorized group, results from the ANCOM tests indicated that at the genus level, *Escherichia* (W = 231), *Klebsiella* (W = 250), and *Enterococcus* (W = 222) differed significantly in its abundance levels among the groups (Figure 1, B).

Table IV. Percentage of classified OTUS at the phylum level distribution across the samples

Samples	Firmicutes	Proteobacteria	Bacteroidetes	Actinobacteria	Synergistetes	Verrucomicrobia	Fusobacteria
Respiratory disorder							
1	38.06	29.23	0.10	0.10	0.80	0.10	ND
2	2.35	22.69	0.17	0.15	0.12	ND	ND
4	2.74	24.80	0.30	0.80	ND	0.66	1.67
5	6.73	8.72	0.33	0.86	4.58	ND	2.12
7	1.40	41.73	5.91	0.20	0.48	ND	ND
9	10.12	2.60	0.93	0.66	3.01	ND	ND
10	1.65	30.69	1.88	0.10	0.58	ND	1.58
11	22.76	7.17	3.18	0.10	0.25	0.33	ND
13	45.12	10.99	1.43	0.48	0.28	0.10	0.89
15	1.65	4.63	0.40	2.50	3.25	0.10	0.25
16	12.04	16.13	5.41	0.55	ND	ND	35.87
18	5.53	39.11	0.13	0.23	0.24	0.11	0.98
20	3.02	6.41	1.19	1.73	ND	ND	ND
23	20.77	42.63	7.96	0.69	ND	ND	ND
24	28.53	3.47	0.06	0.47	34.00	ND	ND
29	3.10	10.52	0.19	14.03	3.25	ND	ND
Gastrointestinal disorder							
3	87.36	2.49	0.10	0.63	7.12	ND	ND
6	7.30	1.48	0.39	1.78	0.12	ND	ND
8	41.95	3.16	0.10	3.20	7.21	ND	ND
12	11.75	33.71	0.10	3.39	1.82	0.10	0.11
14	43.79	9.82	0.25	0.44	2.12	0.12	0.10
17	40.97	24.8	0.28	0.22	1.02	0.10	ND
19	25.39	47.61	0.10	5.04	2.87	0.10	ND
25	8.10	3.61	0.92	0.10	ND	ND	ND
27	3.17	90.04	0.14	0.12	0.12	ND	ND
28	12.13	48.29	0.10	0.38	ND	ND	ND
34	35.91	38.52	0.10	0.15	3.87	ND	ND
Other disorder							
21	18.47	2.52	0.10	0.15	ND	ND	ND
22	8.72	25.97	0.10	0.10	0.24	ND	ND
26	15.61	35.34	9.71	0.10	ND	ND	ND
31	2.70	3.89	0.58	12.67	ND	ND	1.25
32	7.84	12.73	4.73	0.22	1.54	0.10	ND
33	5.31	16.68	7.24	0.24	0.14	ND	0.98
35	0.20	4.63	0.17	0.15	ND	0.32	ND

ND, not detected.
Only bacterial members with an OTU relative abundance of more than 0.1% are shown.

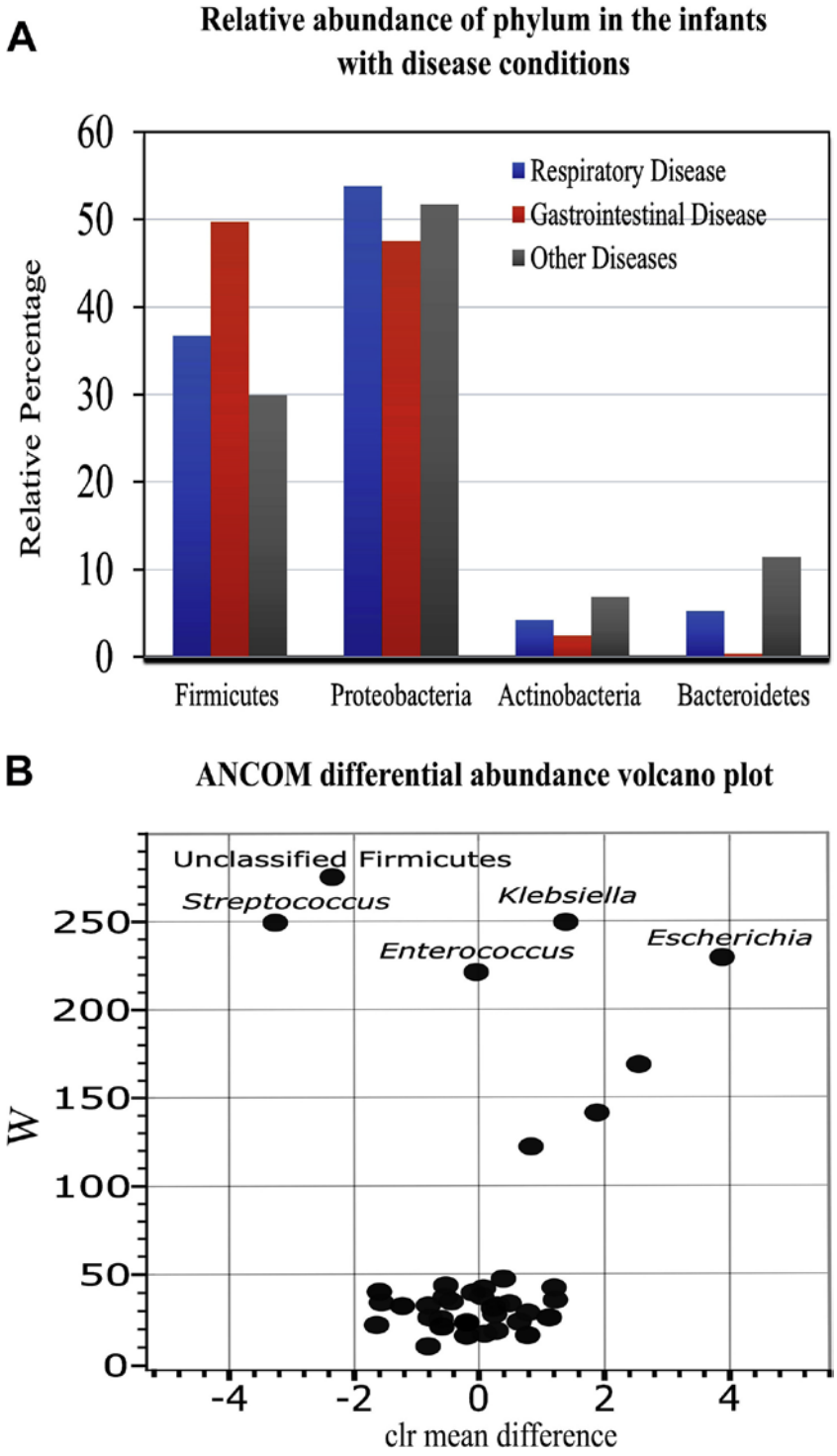

Figure 1. **A,** Bar chart showing percentage relative abundance of phylum in the infants with respiratory (respiratory disease), gastrointestinal (gastrointestinal disease), and other diseases (other diseases). **B**, The ANCOM differential abundance volcano plot. The centered log-ratio (clr) transformed the OTU table at the genus level with 0 values modified to 1 was used. The number of times the null hypothesis (all groups have an equal abundance of species) was rejected. Only species which reject the null hypothesis more than 200 times are labelled.

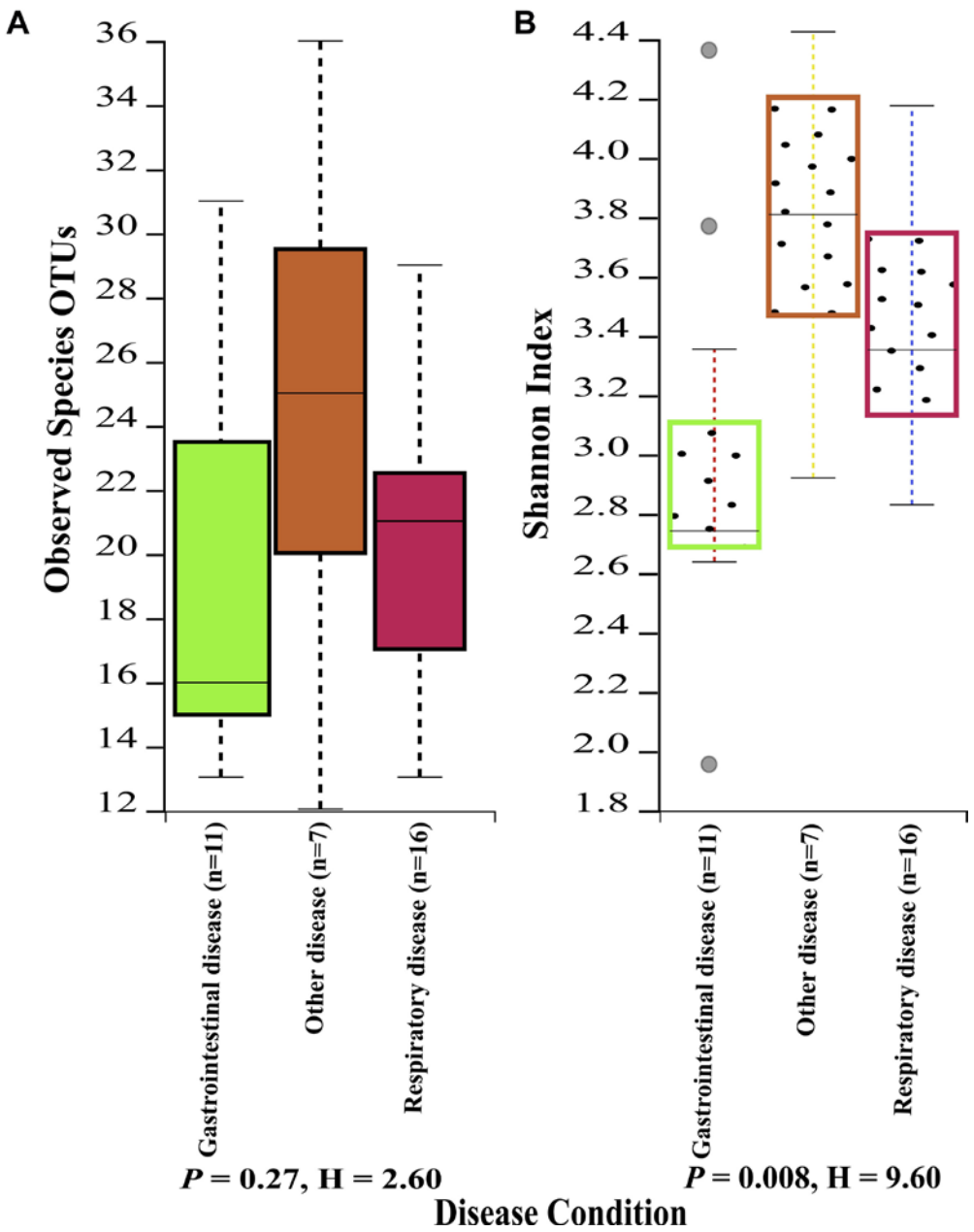

Figure 2. Alpha diversity boxplots between infant fecal microbiota and disease conditions. **A**, Observed species OTUs and **B**, Shannon index. The boxes denote IQRs. The *P* and H values between the group categories (gastrointestinal disease, respiratory disease, and other diseases) are indicated below each boxplot. The *dotted line* inside the box represents the median. Outliers are shown with open circles.

Diversity of the infant fecal microbiota

There were no significant variations in the observed OTU richness from 16s rRNA gene sequencing data among the infant fecal microbiota (IFM) of the categorized disease groups ($P > .05$, Figure 2A). However, in the Shannon diversity analysis, the greater

diversity, including better evenness, was observed between the category of disease conditions and the IFM (95% CI, 2.6 – 4.4; *P* = .008, Figure 2B). The H value was 9.60, which indicated varying microbial diversity among the samples in the cohort of infants. In the alpha diversity comparison of groups based on deworming, diarrhea, vitamin A, and antibiotic intake, there were no significant differences at an alpha level of 0.05.

To understand if disease conditions, antibiotic therapies, and vitamin A intake influenced the infant's microbial ecosystem, we analyzed the IFM community through the beta diversity UniFrac distance analysis. A one-way PERMANOVA test through unweighted UniFrac analysis at 95% CI showed significant differences between the IFM of infants among the disease conditions (95% CI, 0.44 – 0.86; *P* < .012) (Figure 3A) and deworming therapy (95% CI, 0.40 – 0.90; *P* = .033) (Figure 3B). However, no other comparisons significantly differed.

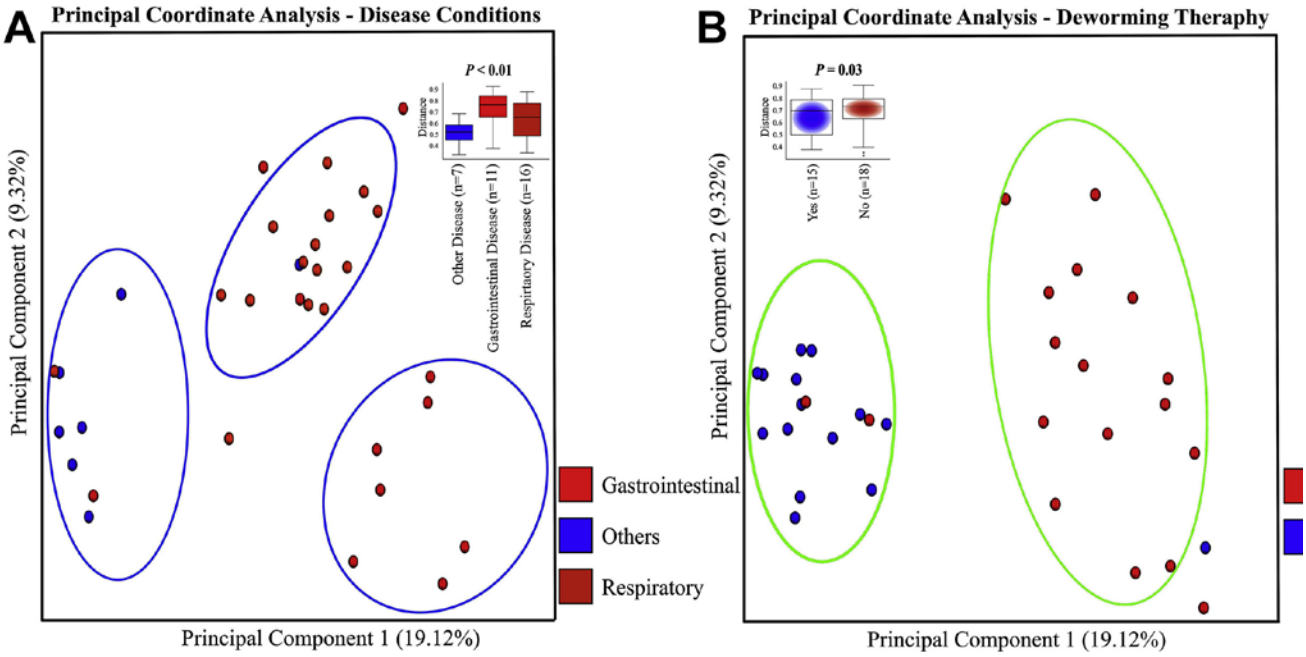

Figure 3. Two-dimensional principal coordinate analysis (PCoA) plots of beta diversity based on the unweighted UniFrac distance matrix of the 34 infant fecal microbiota. Each dot represents a sample point. *Box* plot next to the principal coordinate analysis graph shows the significance. **A,** Groups categorized by disease conditions (gastrointestinal disease, respiratory disease, and other diseases). **B,** Groups categorized based on deworming therapy.

In the case of weighted UniFrac distance matrix analysis at 95% CI, the significant difference was observed among the IFM in regards to antibiotic therapy (95% CI, 0.20 – 0.75; P = .007), vitamin A intake (95% CI, 0.10 – 0.80; P < .033), and disease conditions (95% CI, 0.10 – 0.79; P = .006) (Figure 4). All other categorical comparisons were insignificant (data not shown here). These observations provide significant insights into the influence of deworming therapy, vitamin A supplementation, and disease conditions on IFM and their potential underlying causes of the variations.

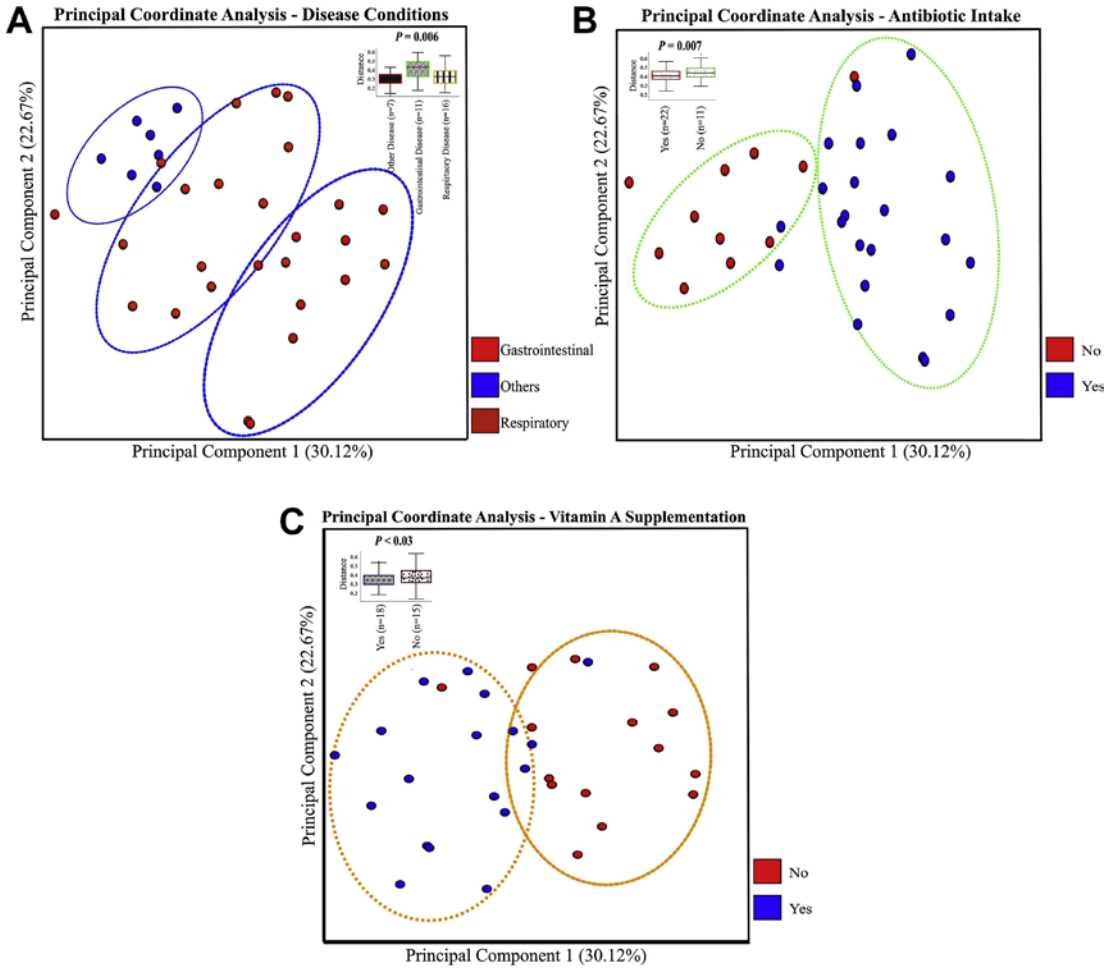

Figure 4. Two-dimensional principal coordinate analysis (PCoA) plots of beta diversity based on the weighted UniFrac distance matrix of the 34 infant fecal microbiota categorized by **A,** disease conditions (gastrointestinal disease, respiratory disease, and other diseases). **B,** Antibiotic intake and **C,** vitamin A supplementation. Each dot represents a sample point. *Box* plot next to the principal coordinate analysis plot given to show the significance.

Candidate pathogens based on OTU abundances in diseased infants

The list of potentially pathogenic bacteria with top OTUs was identified through the relative abundance of the respective OTUs and ANCOM analysis. This test compares the significant differences in the abundances of genera among the disease groups. The "W" value in the volcano plot indicates the number of times the null hypothesis (abundance of species across the groups are same) is rejected. Hence, the genus "*Escherichia*" (W=229), "*Klebsiella*" (W=248), "*Streptococcus*" (W=250) and "*Enterococcus*" (W=220) were detected to be significantly different across the disease groups. A stacked bar diagram representing the pathobiome of the disease groups is given in Figure 5; online only. The OTU counts of *Escherichia coli* found abundant across all the groups. Next, to *E. coli*, *Klebsiella pneumonia* was the most abundant pathogen in RD group infants with a maximum of 33.06% and *Enterococcus faecium* with a maximum of 32.34% abundancy of OTUs in the GD group among the pathobiome. Interestingly, one of the rare pathogens *Cronobacter condiment* was found significant (11.27% relative abundance) in one of the GD infants. No common candidate pathogen associated with OD group was found. The mean relative abundance of *E. coli* (90%) OTUs was present in most of the OD subjects pathobiome. Also, the *Listeria* was identified in 10 IFM diagnosed with respiratory and gastrointestinal disease conditions. Based on the relative abundance of OTUs, the percentage of *Listeria innocua* (30.52% of the total bacteria in sample 1) is more compared to *Listeria monocytogenes* (maximum 2.49% in sample 1).

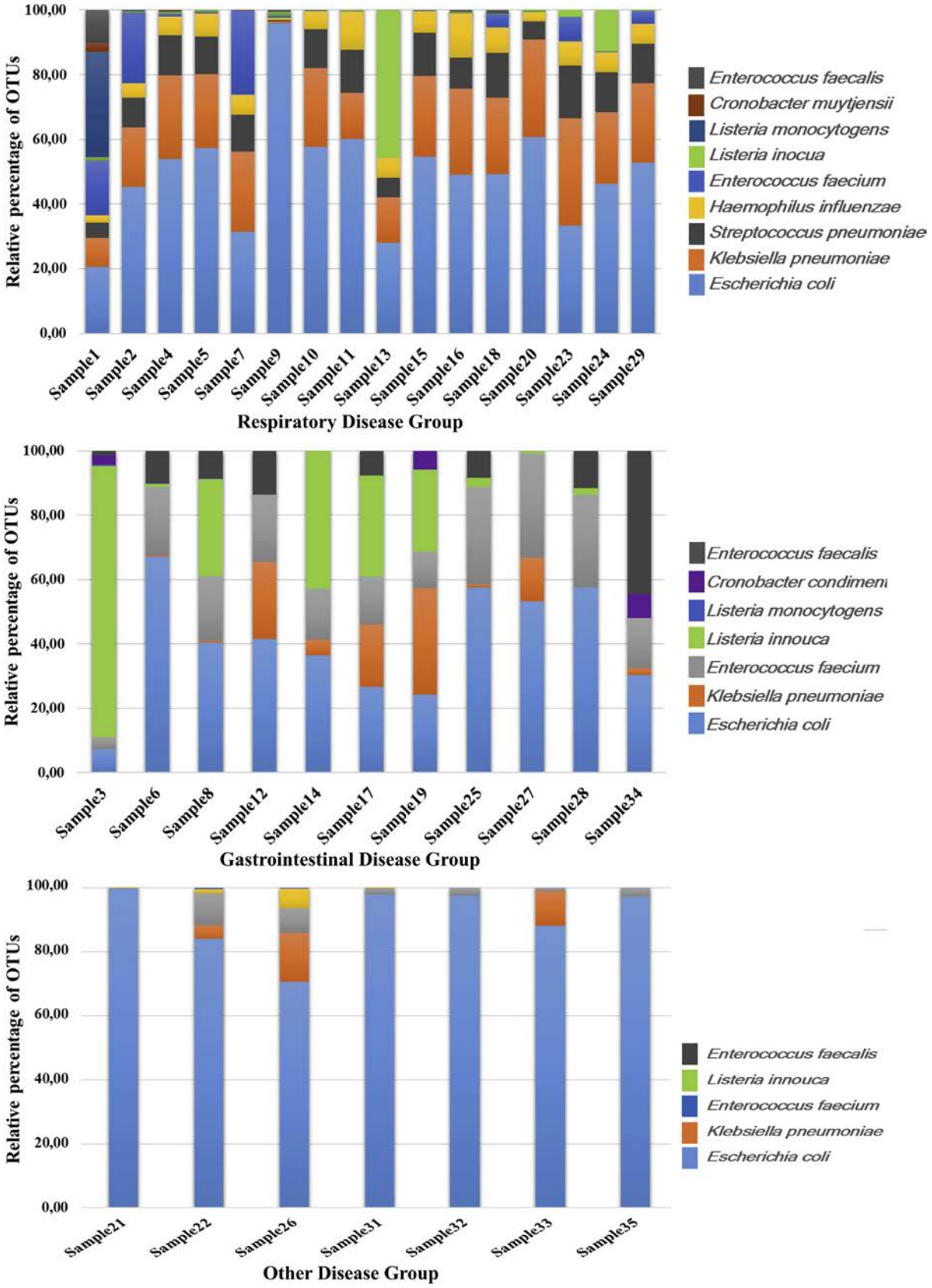

Figure 5. Pathobiome analysis of infants fecal microbiome with disease conditions. Pathogens with more than 1% were picked, and the percentage of pathogen cluster was calculated from the OTUs of total microbiota.

Discussion

This study assessed the fecal microbiota signature of infants under two years of age with different diseases in sub-Saharan Africa. This study hypothesized that shifts in composition, distribution, and diversity in infant fecal microbiota and distinct pathogens might represent specific disease conditions. The signature fecal microbiota, the marker pathogen based on the ANCOM, and the significant diversity between the clinical observation-based groups, supports the potential of microbiome-based interventions for treating diseases and monitoring the pediatric health.

In the bacterial community structure analysis, overall, 14 phyla were recovered from the samples. However, 7 major bacterial phyla, which accounted for 90% of the total sequences are Firmicutes, Proteobacteria, Actinobacteria, Bacteroidetes, Verrucomicrobia, Fusobacteria, and Synergistetes. O'Dwyer et al., reported six key phyla that colonize healthy human gut cells in the order of Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria, Fusobacteria, and Cyanobacteria.[16] At the phylum level, the mean relative abundance percentage of Proteobacteria (54%) is higher than Firmicutes (34%) in the RD group, which were almost in the equal ratio in the GD group. Proteobacteria (52%), Firmicutes (30%), Bacteroidetes (11%) and Actinobacteria (7%) comprised most of the OD group. The low abundant phylum like Verrucomicrobia which constitutes *Akkermansia muciniphila* like mucus degrading bacteria[17] and the Synergistetes which was investigated for inflammatory bowel disease[18] were generally reported as fecal associated microbiota in diseased gut microbiota studies. Based on the ANCOM results on bacterial genera among the disease groups, it is confirmed that a few taxa contribute to the observed differences among the RD, GD, and OD groups. *Klebsiella* genus which belongs to Proteobacteria found significantly abundant in the RD group, which is in line with the overall taxonomical classification. The earlier report suggested that the infants with Bacteroides-dominant microbiota cluster

exhibited the lowest incidence of respiratory disease, and Proteobacteria-dominant profiles exhibited the highest incidence of respiratory disease.[9]

The significant outcomes of this study in alpha and beta diversity among the classified groups emphasized the importance of factors such as antibiotic intake and vitamin A supplementation history to consider prior to the treatment plan. In the comparison of IFM diversity through OTU richness, no significant difference was observed among the study groups, but a remarkable difference was seen within the subjects of disease category group through the Shannon index (95% CI 2.6 – 4.4; P = .008). This finding may be due to a constant level of microbiota immigration and elimination through host mucosal clearance in respiratory diseases such as bronchiolitis.[16]

Interestingly, taking the phylogenetic distance into account, the beta-diversity through UniFrac distance metrics across the study groups has led to the identification of surprising relationships between IFM and disease conditions and deworming therapy. Variation was significantly diverse through the PCoA plots of weighted (abundance of observed OTUs) distance matrix in IFM comparing to the disease, vitamin A supplementation and antibiotic therapy group categories. This finding was concordant with Fallani et al., that the antibiotic usage in infants younger than two years can cause rapid changes in the gut microbiota, which may also lead to antibiotic resistance.[19] Likewise, in unweighted UniFrac (presence or absence) matrix, the association between the IFM, disease conditions, and deworming therapy was remarkable. It was reported that the deworming practice among the infants could shift the gut microbiota beneficial to the host and may influence the immune response, which was unclear.[20] Overall, these diversity analysis results underscore the importance of considering these categorical factors such as deworming, antibiotic intake, and vitamin supplementation while designing the treatment strategy.

The pathobiome, the causative pathogen cluster for infectious disease, is gaining considerable attention to study and design broad antibiotic therapies for various diseases.[21] The top pathogens of the fecal microbiome based on their increased abundance derived through the dominant OTU percentages in disease groups (Figure 5; online only) is explored to highlight the types of pathogen community in those pediatric diseases. Determination of candidate pathogens as a bioindicator for infectious diseases through gut bacterial OTU proportion is a successful approach in the pre-diagnosis process.[22] In the current study, *K. pneumonia*, *E. coli, K. variicola,* and *L. innocua* were identified as dominant pathogens with highest OTU percentage in the RD group and clinically proven to cause similar symptoms in infants. In line with this study findings, it has been found that the leading bacterial species associated with respiratory disease were *K. pneumonia, E. coli,* and *K. variicola.*[23] Likewise, the observation of *E. faecium, E. faecalis,* and *E. coli* as dominant pathogens in the GD group is in line with pathogens investigated in the gut microbiota of the infants with intestinal diseases.[24] In the OD group, *E. coli* was the most common pathogen, and various pathogens were associated with different diagnosed diseases. *Cronobacter condiment* which found in GD group was reported to associate with neonatal meningitis of newborn.[25] Work from this study has provided insights into signature pathogens in disease groups which is crucial to develop pathogenic biomarkers for designing an antibiotic or probiotic therapy strategy.[26]

The presence of *Listeria* in fecal samples of many infants in Eastern Cape of South Africa might be due to the Listeriosis outbreaks in that area during the study period. *L. ivanovii, L. monocytogenes*, and a few *L. innocua* strains are pathogenic and cause abnormalities in the gastrointestinal tract.[27] *L. innocua* has the most resistance to a selected group of antibiotics and has the potential to transfer resistance to the low-resistance *L. monocytogenes.*[28] Hence, understanding the molecular mechanisms behind the *L. innocua* and

L. monocytogenes in the infant gut environment critical to developing targeted antimicrobial therapy.

Several limitations of our study exist. First, the study subjects consisted of infants hospitalized with severe gastrointestinal, respiratory and other diseases and were categorized based on the disease diagnosed. Hence, our results might not be inferred to those with milder disease. Secondly, through the 16s rRNA sequencing platform, we were unable to explain the genetic and biochemical background behind the pathogenesis of the diagnosed candidate pathogens. Moreover, the very small sample size prevented us from adjusting for potential confounders.

In this case-control study among South African infants, we showed potential links between the fecal microbiota and clinical parameters, disease-based signature microbiota and the marker pathogens. Our data may facilitate interventional investigations to disentangle the impacts of IFM with the complex web of the gastrointestinal and respiratory diseases, vitamin A supplementation, deworming therapy, and antibiotic medication. Although the causal assumptions remain premature, the identification of fecal microbiota, the dominant phylum (Firmicutes, Proteobacteria, Actinobacteria, Bacteroidetes) and pathogens with the previous medical history of the pediatric patients could help in inferring the prevalence of disease conditions and its treatment strategies. A detailed mechanistic investigation is needed to link the fecal microbial community with clinical outcomes, which may lead to the discovery of novel therapeutics.

Acknowledgements

We thank Dr Surendra Vikram for helping in metagenome analysis. We thank the Vice Chancellor’s Postdoctoral Fellowship Programme, University of Pretoria, South Africa.

References

1. You D, Hug L, Ejdemyr S, Idele P, Hogan, D, Mathers C, et al. Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency group for child mortality estimation. Lancet 2015;386(10010):2275-2286. doi: 10.1016/S0140-6736(15)00120-8
2. United Nations Children's Fund. [https://data.unicef.org/] Under-five mortality. Child mortality. [updated 2018; cited 2019 May 10]. Available from:https://data.unicef.org/topic/child-survival/under-five-mortality.
3. International Institute for Sustainable Development. [https://sdg.iisd.org/] End hunger, achieve food security and improved nutrition and promote sustainable agriculture. [updated 2019; cited 2019 May 5] Available from: https://sdg.iisd.org/sdgs/goal-2-zero-hunger.
4. Fletcher, Stephanie M, Ellis J. Prevalence of gastrointestinal pathogens in Sub-Saharan Africa: systematic review and meta-analysis. J Public Health Afr 2011;2(2):e30. doi:10.4081/jphia.2011.e30
5. Mukherjee S, Hanidziar D. More of the gut in the lung: how two microbiomes meet in ARDS. Yale J Biol Med 2018;91(2):143-149. pmid:29955219
6. de Steenhuijsen Piters WAA, Sanders EAM, Bogaert D. The role of the local microbial ecosystem in respiratory health and disease. Philos Trans R Soc Lond B Biol Sci 2015;370(1675):20140294. doi: 10.1098/rstb.2014.0294
7. Scotti E, Boue S, Sasso GL, Zanetti F, Belcastro V, Poussin C, et al. Exploring the microbiome in health and disease: Implications for toxicology. Toxicol Res Appl 2017;1:1-37. doi:10.1177/2397847317741884.

8. Young VB. The intestinal microbiota in health and disease. Curr Opin Gastroenterol 2012;1(1):63-69. doi: 10.1097/MOG.0b013e32834d61e9.
9. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat Commun 2017;8(1):1784. doi: 10.1038/s41467-017-01973-8
10. Agrawal V, David RJ, Harris VJ. Classification of acute respiratory disorders of all newborns in a tertiary care centre. J Natl Med Assoc 2003;95(7):585-595. pmid: 12911256
11. Rasquin-Weber A, Hyman P, Cucchiara S, Fleisher DR, Hyams JS, Milla PJ, et al. Childhood functional gastrointestinal diseases. Gut 1999;45(Suppl 2):II60-II68. doi: 10.1053/j.gastro.2005.08.063
12. Zeh CE, Odhiambo CO, Mills LA. Laboratory reference intervals in Africa. 2012. https://cdn.intechopen.com/pdfs/39110. Accessed September 14, 2018.
13. QIIME2, 2018. https://docs.qiime2.org/2018.11/. Accessed September 10, 2018.
14. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from illumina amplicon data. Nat Methods 2016;13(7):581. doi: 10.1038/nmeth.3869
15. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol 2007; 73(5):1576-85. doi: 10.1128/AEM.01996-06
16. O'Dwyer DN, Dickson RP, Moore BB. The lung microbiome, immunity, and the pathogenesis of chronic lung disease. J Immunol 2016;196(12):4839-47. doi: 10.4049/jimmunol.1600279
17. Fujio-Vejar S, Vasquez Y, Morales P, Magne F, Vera-Wolf P, Ugalde JA, et al. the gut microbiota of healthy chilean subjects reveals a high abundance of the phylum Verrucomicrobia. Front Microbiol 2017;30(8):1221. doi: 10.3389/fmicb.2017.01221.

18. Chen L, Wang W, Zhou R, Ng SC, Li J, Huang M, et al. Characteristics of fecal and mucosa-associated microbiota in Chinese patients with inflammatory bowel disease. Medicine (Baltimore) 2014;93(8):e51. doi: 10.1097/MD.0000000000000051.

19. Fallani M, Amarri S, Uusijarvi A, Adam R, Khanna S, Aguilera M, et al. Determinants of the human infant intestinal microbiota after the introduction of first complementary foods in infant samples from five European centres. Microbiology 2011;157(5):1385-92. doi: 10.1099/mic.0.042143-0.

20. Yang CA, Liang C, Lin CL, Hsiao CT, Peng CT, Lin HC, et al. Impact of *Enterobius vermicularis* infection and mebendazole treatment on intestinal microbiota and host immune response. PLOS Negl Trop Dis 2017;11(9):e0005963. doi:10.1371/journal.pntd.0005963.

21. Alshawaqfeh M, Bashairehs A, Serpedin E, Suchodolski. Reliable biomarker discovery from metagenomic data via RegLRSD algorithm. BMC Bioinf 2017;18(1):328. doi:10.1186/s12859-017-1738-1.

22. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic analysis of the human distal gut microbiome. Science 2006; 312(5778):1355-1359. doi: 10.1126/science.1124234.

23. Zhong Q, Feng H, Lu Q, Liu X, Zhao Q, Du Y et al. Recurrent wheezing in neonatal pneumonia is associated with combined infection with respiratory syncytial virus and *Staphylococcus aureus* or *Klebsiella pneumoniae*. Sci Rep 2018;8(1):995. doi: 10.1038/s41598-018-19386-y.

24. Rizzatti G, Lopetuso LR, Gibiino G, Binda C, Gasbarrini A. Proteobacteria: A common factor in human diseases. BioMed Res Int 2017; 2017(9351507):1-7. doi:10.1155/2017/9351507

25. Wang M, Cao H, Wang Q, Xu T, Guo X, Liu B. The roles of two type VI secretion systems in *Cronobacter sakazakii* ATCC 12868. Front Microbiol 2018;9:2499. doi: 10.3389/fmicb.2018.02499.
26. Urushiyama D, Suda W, Ohnishi E, Araki R, Kiyoshima C, Kurakazu M, et al. Microbiome profile of the amniotic fluid as a predictive biomarker of perinatal outcome. Sci Rep 2017;7(1):12171. doi: 10.1038/s41598-017-11699-8.
27. Johnson, J, Jinneman K, Stelma G, Smith BG, Lye D, Messer J, et al. Natural atypical *Listeria innocua* strains with *Listeria monocytogenes* pathogenicity island 1 genes. Appl Environ Microbiol 2004;70(7):4256-66. doi: 10.1128/AEM.70.7.4256-4266.2004.
28. Lungu B, O'Bryan CA, Muthaiyan A, Milillo SR, Johnson MG, Crandall PG, et al. *Listeria monocytogenes*: Antibiotic resistance in food production. Foodborne Pathog Dis 2011;8(5):569–578. doi:10.1089/fpd.2010.0718.