**Report Summary**

This study aimed to assess the differences between the fecal microbiome of group 1 and 2. Only minor changes in the overall microbiome composition between group 1 and 2 were observed, with a substantial overlap of a subset of samples between the groups. Statistical analysis showed no significant separation of the fecal microbiome composition between the two groups. Differential abundance analysis revealed small differences on a species level between the groups. The diversity within the samples was significantly different between group 1 and group 2. Group 1 showed an increased species richness and evenness compared to group 2.

**Abstract**

The objective of this study was to assess the differences between the fecal microbiome of group 1 and 2. Shallow shotgun metagenomic sequencing was performed. Differential abundance testing was performed in R using Analysis of Composition of Microbiomes with Bias Correction (ANCOM-BC). Beta diversity, the diversity between samples, revealed minor differences between group 1 and group 2, demonstrated as two clusters in the Principal coordinate analysis (PCoA) based on Bray-Curtis dissimilarity index. However, a substantial subset of the group 2 samples overlapped with group 1. Statistical analysis ANOSIM (Analysis of Similarities) showed no significant separation between the two groups (adjusted *p*-value: 0.85). Differential abundance analysis revealed small differences on a species level between the groups. The relative abundance of the species *Clostridium P ventriculi* (*p*<0.0001), *Lactobacillus johnsonii* (*p*<0.0001)*, Lactiplantibacillus plantarum* (*p*<0.0001)*, Streptococcus lutetiensis* (*p*<0.0001), and *Enterococcus D casseliflavus* (*p*<0.0001) was significantly different between group 1 and group 2. Alpha diversity, the diversity within the samples, was significantly different between group 1 and group 2. Group 1 showed an increased species richness and evenness compared to group 2 (Shannon *p*=0.0022, Observed Features *p*=0.0011, Chao1 *p*=0.0012).

**Methods**

DNA isolation and shotgun metagenomic sequencing

The fecal samples were collected from both groups and were stored at -80 °C before being shipped on dry ice to the Gastrointestinal Laboratory of Texas A&M University, College Station, Texas. The microbial DNA from fecal samples was extracted and quantified using the DNeasy PowerSoil Pro Kit (QIAGEN Inc, Germantown, Maryland) and the Thermo Scientific NanoDrop Microvolume Spectrophotometers respectively.

The Nextera XT DNA Library Preparation Kit (Illumina Inc., San Diego, CA, USA) was used for sequencing library preparation, followed by pooling of the libraries. Afterward, SPRI bead purification and concentration were performed with SpeedBeads Magnetic Carboxylate Modified Particles (Cytiva Life Sciences, Marlborough, MA, USA). The pooled libraries were denatured using NaOH, diluted, and spiked with 2% PhiX. Sequencing was carried out on an Illumina NovaSeq6000 with a 2x150 flow cell to achieve a median of 2 million read pairs at the Diversigen laboratory (New Brighton, MN). The data were converted to FASTQ files and filtered for low-quality (Q-score <30) and short lengths (<50) sequences. Adapter sequences were removed, and all remaining sequences were trimmed to a maximum length of 100 base pairs before proceeding with the alignment process. Raw sequence data were uploaded to the NCBI Sequence Read Archive and analyzed using established pipelines. For taxonomic classification, FASTA sequences were aligned to a curated database containing all representative prokaryotic genomes from the NCBI RefSeq collection, as well as additional manually curated bacterial strains. Alignments were performed at 97% identity and compared with reference genomes. Taxonomy assignment was based on the lowest common ancestor approach, with compatibility to at least 80% of the reference sequences. Operational Taxonomic Units (OTUs) with fewer than one million species-level markers, or with less than 0.01% unique genome region matching and less than 0.1% of the entire genome, were discarded.

The normalized and filtered data tables were then imported into QIIME2 for subsequent downstream analysis. Prior to conducting diversity analyses, samples were rarefied to depth of 64000, ensuring uniform depth for analysis.

Statistical Methods

Statistical analysis was conducted to test the hypothesis that there is a difference in the fecal microbiome between the two groups. Data were tested against the hypothesis of normal distribution by conducting the Shapiro-Wilk’s test using GraphPad Prism v10.0.2 (GraphPad Software, Inc., La Jolla, CA, USA). The data did mainly follow a normal distribution. However, due to the zero inflated and compositional nature of microbiome data, non-parametric methods were applied.

Statistics on the Shannon, Chao1, and Observed Features indices were performed using Mann-Whitney tests. All alpha diversity statistical analyses were performed in GraphPad Prism v10.0.2 (GraphPad Software, Inc., La Jolla, CA, USA).

To quantify the dissimilarity in microbial structure between groups, a non-parametric ANOSIM (Analysis of Similarities) was performed on the Bray-Curtis distance matrix in Quantitative Insights Into Microbial Ecology 2 (QIIME2 2024.2). Groups were defined based on the individual’s respective treatment and the time point at collection. To compute p-values, 999 permutations were calculated, and pairwise comparisons were performed against baseline. The significance level was set at a *p*-value of <.05, and the Benjamini-Hochberg method was used to correct for multiple comparisons.

Differential abundance analysis between the groups was identified using Composition of Microbiomes with Bias Correction (ANCOM-BC) at phylum, family, genus, species, and strain levels. Briefly, data obtained from QIIME2 (v2024.05) was imported into R statistical software v4.2.3 (R Core Team, 2020) with the R package ANCOMBC (v 2.2.0) and phyloseq (v3.20) package.

Identified significant taxa were visualized using GraphPad Prism v10.0.2 (GraphPad Software, Inc., La Jolla, CA, USA).

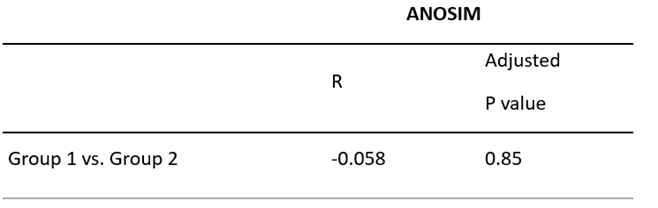
A stacked relative abundance table on a phylum level was created utilizing R statistical software v4.2.3 (R Core Team, 2020) with the phyloseq v3.20 package.

**Results**

Beta diversity

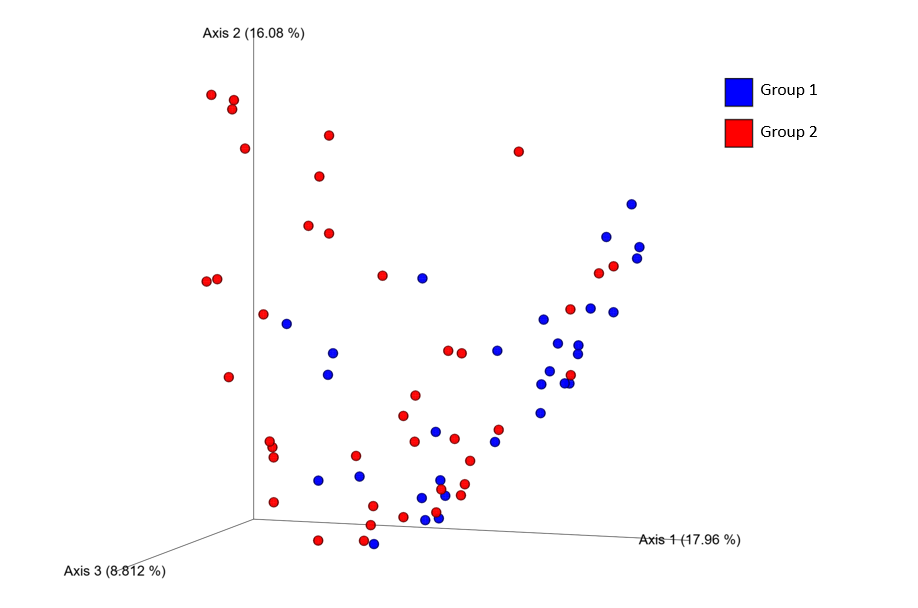
*Beta diversity is a metric used to measure the microbiome diversity between different samples or groups. It provides information about how similar or different the microbial communities are between the different groups.*

Beta diversity, the diversity between samples, revealed minor differences in the microbial composition between group 1 and group 2, demonstrated as two clusters in the Principal coordinate analysis (PCoA) based on Bray-Curtis dissimilarity index (Figure 1). However, a substantial subset of the group 2 samples overlapped with group 1. Statistical analysis ANOSIM showed no significant separation between the two groups (adjusted *p*-value: 0.85) (Table 1).



**Table 1:** Statistics for beta diversity based on the Bray-Curtis dissimilarity metric. The ANOSIM (Analysis of Similarities) results indicate a R-value of -0.058 (a measure of group distinction between -1 to 1), with an adjusted *p*-value of 0.85 suggesting no significant separation between group 1 and 2.

Download Path: ProjectName/1\_Final\_Data\_Summary/taxa-summary/ANCOMBC\_Group1vsGroup2.xlsx



**Figure 1:** Principal coordinate analysis (PCoA) plot based on the Bray-Curtis beta diversity measure, depicting relatedness of microbiome composition between samples. Statistical analysis is shown in table 1. The PCoA plot reveals some clustering between group 1 and 2 with substantial overlap. Suggesting a shift in composition in a subset of individuals in group 2.

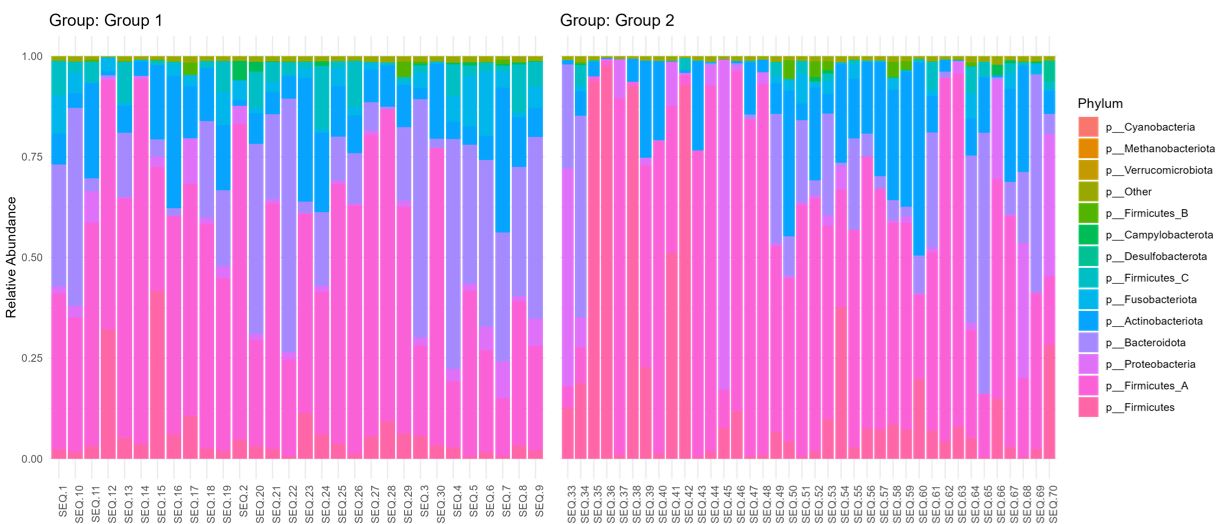
Download Path: ProjectName/1\_Final\_Data\_Summary/beta-diversity/bray\_curtis\_emperor.qzv

Taxonomic difference

The shotgun metagenomic sequencing showed small changes in fecal bacterial taxa at phylum, family, genus, and species levels between the two groups.

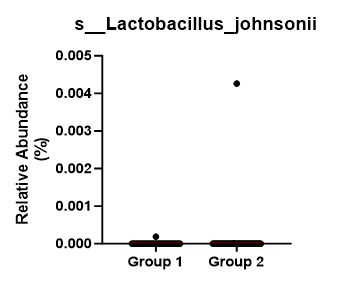
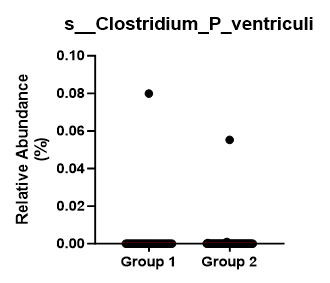
The overall changes in the relative abundance at the phylum level between the groups are graphically shown in Figure 2.

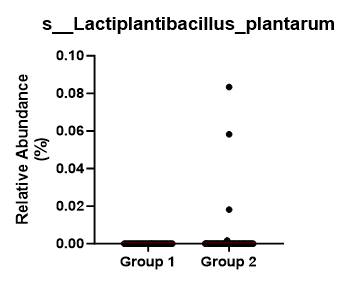
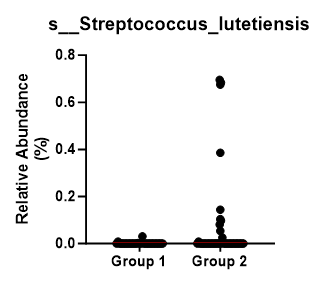
There was a significant difference in relative abundance of five species between group 1 and group 2. The species (A) *Clostridium P ventriculi* (*p*<0.0001),(B) *Lactobacillus johnsonii* (*p*<0.0001)*,* (C) *Lactiplantibacillus plantarum* (*p*<0.0001)*,* (D) *Streptococcus lutetiensis* (*p*<0.0001) and (E) *Enterococcus D casseliflavus* (*p*<0.0001) differed significantly between the two groups (Figure 3).



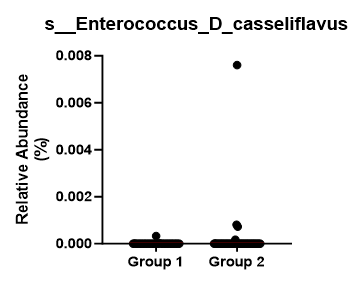
**Figure 2:** Relative abundance taxonomy bar plot at the phylum level. Overall phylum abundance between group 1 and group 2.

Download Path: ProjectName/1\_Final\_Data\_Summary/taxa-summary/barplot.png

(A) (B)john



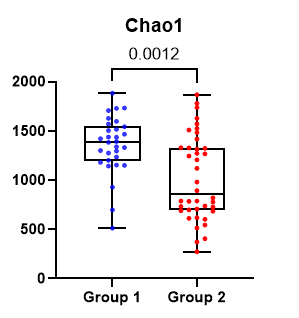
(C) (D)

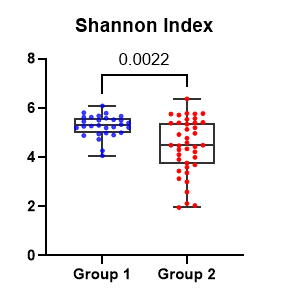
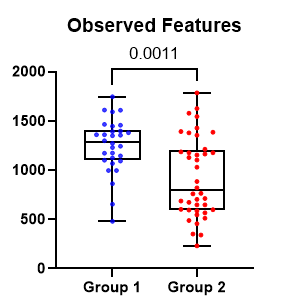
(E)

**Figure 3:** Differential abundance analysis for group 1 and group 2. The relative abundance of the species (A) *Clostridium P ventriculi* (*p*<0.0001),(B) *Lactobacillus johnsonii* (*p*<0.0001)*,* (C) *Lactiplantibacillus plantarum* (*p*<0.0001)*,* (D) *Streptococcus lutetiensis* (*p*<0.0001) and (E) *Enterococcus D casseliflavus* (*p*<0.0001) was significantly different between group 1 and group 2. Differentially abundant taxa between the two groups at species levels were identified using Analysis of Composition of Microbiomes with Bias Correction (ANCOM-BC).

Alpha diversity

*Another metric to measure microbiome diversity is alpha diversity, which measures the diversity within samples. For example, alpha diversity can describe the species diversity within a single community or sample. It provides information about how diverse or rich the microbial community is within each sample.*

Alpha diversity, or diversity within samples, was significantly different in all alpha-diversity variables between group 1 and group 2 (Figure 4). Three different metrics were used to assess alpha diversity, each providing slightly different information. The Shannon index is an alpha diversity index that reflects both richness and evenness within the samples. Richness is defined as the total number of different species present in a sample (number of different species). A high richness indicates the abundance of many different species within a sample. Evenness describes how evenly distributed the abundance of those species within a sample is (equality of different species abundance). A high evenness reflects that most species are present in an equal abundance. Whereas a low evenness shows that a few species are more abundant than others. The Shannon index was significantly increased in group 1 compared to group 2, indicating an increased richness and evenness in group 1 (Shannon *p*=0.0022). Observed Features is another metric to evaluate alpha diversity. It refers to the number of different species that are detected and counted within a sample, reflecting species richness. Chao1 is also a metric that evaluates species richness in alpha diversity. In contrast to Observed Features it is a statistical estimator used to estimate the total species richness within a sample. It is particularly useful when low-abundant species are present that might be missed in a typical analysis. Observed Features is a direct count of detected species, while Chao1 attempts to estimate the total richness by accounting for potentially missed rare species. Observed Features and Chao1 were significantly increased in group 1 compared to group 2, indicating an increase in species richness in group 1 (Observed features *p*=0.0011, Chao1 *p*=0.0012).

(A) (B) (C)

**Figure 4:** Alpha diversity metrics for group 1 and group 2. Both groups were significantly different in all alpha-diversity variables. (A) The Shannon index reflects the evenness and richness within samples, while (B) Observed Features and (C) Chao1 evaluate the richness within samples. The Shannon index, Observed Features, and Chao1 were significantly increased in group 1 compared to group 2, indicating a higher richness and evenness of species within the samples of group 1 (Shannon *p*=0.0022, Observed Features *p*=0.0011, Chao1 *p*=0.0012). The Mann-Whitney test was used to compare alpha diversity parameters between group 1 and 2.