16S\_DuodenalMicrobiome\_MalZam Analysis Guidance Document

# Summary

This section details the analysis using the BEECH and Malnutrition Enteropathy datasets to compare the duodenal microbiome of stunted only vs SAM children. Almost all children with SAM were stunted but stunted only children were not severely wasted (i.e., WLZ < -3). The raw FASTQ files were pre-processed to generate ASVs, taxonomy tables and generate absolute abundance estimates of ASVs for both the BEECH and Malnutrition Enteropathy cohorts. These tables were then used for comparison between the malnutrition groups.

## Pre-processing:

ASVs were called using the DADA2 pipeline and taxonomic assignment was implemented using the DADA2 ‘AssignTaxonomy’ function based on the Silva database (v138.1) [silva\_nr99\_v138.1\_train\_set.fa.gz and silva\_species\_assignment\_v138.1.fa.gz].

To get these absolute values, the following steps were taken:

* Calculate bacterial load factor:
* Determine number of spike-in cells added to the sample in millilitres when the number of cells added to the samples in microliters is known.
  + 9.9 x 105 was added to 200uL of sample for BEECH so multiply this by 5 to get cells/ mL
  + 9.9 x 105 was added to 100uL of sample for ME so multiply this by 10 to get cells/ mL
* The number of cells was multiplied by the bacterial load factor to get the actual bacterial load
* The ASV counts were normalized for library size using Deseq2 and the output of this was scaled by bacterial load to get the absolute ASV counts.
* A phyloseq object was created with these counts, taxonomy, a phylogenetic tree for downstream analysis.

## Analysis

The input file used for all analysis is ‘Stunting\_SAM/RData/Zam\_phyloseqObj\_AbsASV\_Withtree.RData’ and its corresponding data dictionary with variables used in this analysis can be found in ‘Stunting\_SAM/ 16S\_DuodenalMicrobiome\_MalZam\_CodeBook\_Aug2025.xlsx’

### Summary Table

This code was used to generate the summary statistics for the Zambian cohorts and establish how many samples had missing data.

Code file: ‘Stunting\_SAM/Code/2\_SummaryTable\_Aug.R’

### Differential Abundance Analysis

This code was used to plot the relative abundance of top genera in each group and to determine the association of genera absolute abundance and malnutrition group and clinical features.

Code file: ‘Stunting\_SAM/Code/2\_Relative Abundance\_Aug.Rmd

### Alpha diversity

This code was used to estimate the alpha diversity (Shannon index and Faiths PD metric) and regression analysis with clinical features.

Code file: ‘Stunting\_SAM/Code/ 4\_Alpha Diversity\_Aug.Rmd

### Beta Diversity

This code was used to export the relative abundance of genera to be used as input for PERMANOVA analysis in the PRIMER7 software. This used the Anderson’s correction which accounts for imbalanced sample size between groups.

Code file: ‘Stunting\_SAM/Code/ 5\_Beta Diversity\_Aug.Rmd’

|  |  |  |  |
| --- | --- | --- | --- |
| **Current Document** | | | |
| Developed by: | Monica Mweetwa | Date: | 09/01/2025 |
| Reviewed by: | Nate McNulty | Date: | 01/04/2025 |
| Code revised by: | Monica Mweetwa | Date: | 14/08/2025 |