16S\_DuodenalMicrobiome\_MalZam Analysis Guidance Document

# Summary

This section details the analysis using the BEECH and EEBiomarker datasets to compare the duodenal microbiome of stunted vs SAM children. Almost all children with SAM were stunted but stunted children were not servery malnourished i.e severe wasting (WLZ < -2). The raw fastQ files were analysed by the WashU team and we received an RData file ‘Zambia\_EE\_BEECH\_16s\_absQuant\_forMonica.RData’ which contained a mapping file (map) and a phyloseq object (ps2) with data from both BEECH and EEBiomarker available.

The corresponding data dictionary with variables used in this analysis can be found in Stunting\_SAM/16S\_DuodenalMicrobiome\_MalZam\_CodeBook.xlsx

## Data Cleaning:

For data cleaning, a phylogenetic tree and clinical metadata i.e morphometry, enteropathy marker and anthropometry data were added to the phyloseq object (ps2) for downstream analysis.

Data cleaning also included excluding ASVs that had no taxonomic assignment, assigned as mitochondria family or chloroplast class.

Code file: ‘Stunting\_SAM/Code/1\_Data Cleaning\_30122024.Rmd’

Inputs:

* Stunting\_SAM/Data/RData/Zambia\_EE\_BEECH\_16s\_absQuant\_forMonica.RData
* Stunting\_SAM/Data/Metadata/SamBeechMetadata22.csv
* Stunting\_SAM/Data/Metadata/Morphometry/BEECHmorphometry\_91\_edited.csv
* Stunting\_SAM/Data/Metadata/Morphometry/SAM\_morphometry.csv

Output:

* Stunting\_SAM/Data/Metadata/Final\_Metadata.csv
* Stunting\_SAM/Data/Sequences.fasta
* Stunting\_SAM/Data/RData/phyloseq\_dataset\_species.RData
* Stunting\_SAM/Data/Tree.tre

## Analysis

### Summary Table

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

Inputs: Stunting\_SAM/Data/RData/phyloseq\_dataset\_species.RData

Output: Stunting\_SAM/Output/DescriptiveTable\_all.csv

### Comparison Part 1 – Alpha and Beta diversity and, relative abundance visualization

Code file: ‘Stunting\_SAM/Code/3\_Analysis\_SAMvsStunting1.Rmd’

Inputs: Stunting\_SAM/Data/RData/phyloseq\_dataset\_species.RData

Outputs folder: Stunting\_SAM/Output

Lines 13 – 22: Input files were imported into R and necessary libraries loaded

Lines 24 – 83: Visualization of relative abundance. The output for this is in the ‘RelAbund’ folder

Lines 85 – 129: Visualization of alpha diversity measures.

Lines 132 – 340: Linear models of Faith’s PD metric with clinical features i.e. anthropometry, biomarkers and morphometry.

* 140 – 207: All samples analysed together
* 210 – 273: Stunted children only from the BEECH cohort
* 275 – 338: stunted children (LAZ < -2 from both cohorts)
* 341 – 405: Children with SAM only (WLZ < -2) i.e the EEBiomarker group only
* 407 – 429: Plotting scatter plots of the significant models

Lines 431 – 468: Beta diversity analysis and visualization. The output for this is in the ‘BetaDiv’ folder

Lines 470 – 479: combined plot i.e Figure 1 in the manuscript stored in ‘Output’ folder.

### Comparison Part 2 – Differential abundance analysis

Code file: ‘Stunting\_SAM/Code/3\_Analysis\_SAMvsStunting2.Rmd’

Inputs: Stunting\_SAM/Data/RData/phyloseq\_dataset\_species.RData

Outputs folder: Stunting\_SAM/Output

* SAM vs stunting:

The first section of the notebook (lines 26 -228) details linear models on the log2 transformed relative abundances of taxa at different ranks form phylum to genus with malnutrition type. Scatter plots of significant associations at each rank were plotted and prevalence extracted.

* Genera associations with clinical data:

The second half of the notebook (lines 231 – 480) has linear models of taxa as all taxonomic ranks with clinical features i.e anthropometry, biomarkers and morphometry.

|  |  |  |  |
| --- | --- | --- | --- |
| **Current Document** | | | |
| Developed by: | Monica Mweetwa | Date: | 30/12/2024 |
| Reviewed by: |  | Date: |  |
| Code tested by: |  | Date: |  |
|  |  | | |