

## Codes used in Microbiome Section Report (G2)

### **a. START in Ubuntu**

```
# import my files into qiime2 and create the artefact
qiime tools import \
  --input-path demux.qza \
  --output-path sequences
```

### **b. In R Studio**

```
# Load required packages
```{r}
library(dada2)
```

# Load sequences
```{r}
path <- "sequences"
list.files(path)
```

# Read in file names
```{r}
# Forward and reverse fastq filenames have format: SAMPLENAME_R1_001.fastq
# and SAMPLENAME_R2_001.fastq
fnFs <- sort(list.files(path, pattern="_R1_001.fastq", full.names = TRUE))
fnRs <- sort(list.files(path, pattern="_R2_001.fastq", full.names = TRUE))
# Extract sample names, assuming filenames have format:
# SAMPLENAME_XXX.fastq
sample.names <- sapply(strsplit(basename(fnFs), "_"), `[`, 1)
```

# Inspect read quality
```{r}
plotQualityProfile(fnFs[1:2])
```

```{r}
plotQualityProfile(fnRs[1:2])
```
```

### c. In Ubuntu using qiime2

```
# quality control of forward and reverse reads based on histogram
# it uses the demux.qza file to trim and truncate reads
# the output a rep-seqs file, a table of the sequences, and stats
qiime dada2 denoise-single \
qiime dada2 denoise-paired --i-demultiplexed-seqs demux.qza --p-trim-left-f 15 --
p-trunc-len-f 150 --p-trim-left-r 0 --p-trunc-len-r 120 --o-representative-sequences
rep-seqs.qza --o-table table.qza --o-denoising-stats stats.qza
```

```
# the code creates a visualizations for the Table and Sequences
# it tabulates the stats file
# an output would be a visualization of the stats file
qiime metadata tabulate \
--m-input-file stats.qza \
--o-visualization stats.qzv
```

```
# creating a table summarizing the data
# it uses the table files
# results in an output of a metadata file
qiime feature-table summarize \
--i-table table.qza \
--o-visualization table.qzv \
--m-sample-metadata-file metadata.tsv
```

```
# creating a table of the sequences, output a visualization
# it uses the rep-seqs file
# results in a visualization of the rep-seqs file
qiime feature-table tabulate-seqs \
--i-data rep-seqs.qza \
--o-visualization rep-seqs.qzv
```

```
# adjusted the following code by adjusting the names of the files I personalized
qiime feature-table summarize \
--i-table table.qza \
--o-visualization table.qzv \
```

```
--m-sample-metadata-file metadata.tsv
qiime feature-table tabulate-seqs \
--i-data rep-seqs.qza \
--o-visualization rep-seqs.qzv
```

```
# qiime uses a specific database as a classifier
# input the reads
# output of the taxonomy
qiime feature-classifier classify-sklearn \
--i-classifier gg-13-8-99-515-806-nb-classifier.qza \
--i-reads rep-seqs.qza \
--o-classification taxonomy.qza
```

```
# making a visualization of the taxonomy.qza file to a qzv
qiime metadata tabulate \
--m-input-file taxonomy.qza \
--o-visualization taxonomy.qzv
```

```
# filters the taxonomy, excluding contaminants of mitochondria and chloroplast
genetic material
qiime taxa filter-table \
--i-table table.qza \
--i-taxonomy taxonomy.qza \
--p-exclude mitochondria,chloroplast \
--o-filtered-table table.qza
```

```
# making a visualization of the table.qza file to a qzv (now excluding mitochondria
and chloroplast DNA)
qiime metadata tabulate \
--m-input-file table.qza \
--o-visualization table.qzv
```

```
# making a visualization of the taxonomy.qza file to a qzv (now excluding
mitochondria and chloroplast DNA)
qiime metadata tabulate \
--m-input-file taxonomy.qza \
--o-visualization taxonomy.qzv
```

```

# generating a taxa bar plot (excluding mitochondria and chloroplast DNA)
# input the table.qza of the data
# input the metadata file
# output of the taxonomy bar plots through a .qzv file
qiime taxa barplot \
  --i-table table.qza \
  --i-taxonomy taxonomy.qza \
  --m-metadata-file metadata.tsv \
  --o-visualization taxa-bar-plots.qzv

# generating a phylogenetic tree for diversity analysis
# input sequences
# output of alignment, masked aligned, unrooted and rooted phylo trees.
qiime phylogeny align-to-tree-mafft-fasttree \
  --i-sequences rep-seqs.qza \
  --o-alignment aligned-rep-seqs.qza \
  --o-masked-alignment masked-aligned-rep-seqs.qza \
  --o-tree unrooted-tree.qza \
  --o-rooted-tree rooted-tree.qza

#generating the phylogenetic core metrics
# input the phylogenetic tree
# input the table excluding mitochondria and chloroplast DNA
# setting parameters for the sampling-depth based on the visualization of the
table.qzv file
# using the metadata file to make an output of the phylogenetic core metrics
qiime diversity core-metrics-phylogenetic \
  --i-phylogeny rooted-tree.qza \
  --i-table table.qza \
  --p-sampling-depth \
  --m-metadata-file metadata.tsv \
  --output-dir core-metrics-results

# generate alpha/beta diversity numbers using a sampling of 6741, determined by
the table.qza file
# input phylogeny-tree and table
# parameters of the sampling depth 6741 using the metadata file

```

#output directory of the file “core-metric-results”

qiime diversity core-metrics-phylogenetic \

--i-phylogeny rooted-tree.qza \

--i-table table.qza \

--p-sampling-depth 6741 \

--m-metadata-file metadata.tsv \

--output-dir core-metrics-results

# analysis of the alpha-group-significance, specific to Shannon Diversity and Observed Features

# input specific core-metric outputs

# output visualizations of core-metric outputs

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/shannon\_vector.qza \

--m-metadata-file metadata.tsv \

--o-visualization core-metrics-results/shannon\_vector.qzv

qiime diversity alpha-group-significance \

--i-alpha-diversity core-metrics-results/observed\_features\_vector.qza \

--m-metadata-file metadata.tsv \

--o-visualization core-metrics-results/observed\_features\_vector.qzv

# analysis of beta-group-significance, specific to the metadata columns sex, population, and flock

# input core-metric output from beta analysis

# specific metadata file

# specify what column on the metadata file should be used as data

# output of visualization of the specific core-metric result

# pair-wise parameters

qiime diversity beta-group-significance \

--i-distance-matrix core-metrics-results/bray\_curtis\_distance\_matrix.qza \

--m-metadata-file metadata.tsv \

--m-metadata-column sex \

--o-visualization core-metrics-results/bray\_curtis\_distance\_sex.qzv \

--p-pairwise

qiime diversity beta-group-significance \

```
--i-distance-matrix core-metrics-results/bray_curtis_distance_matrix.qza \  
--m-metadata-file metadata.tsv \  
--m-metadata-column population \  
--o-visualization core-metrics-results/bray_curtis_distance_population.qzv \  
--p-pairwise
```

```
qiime diversity beta-group-significance \  
--i-distance-matrix core-metrics-results/bray_curtis_distance_matrix.qza \  
--m-metadata-file metadata.tsv \  
--m-metadata-column flock \  
--o-visualization core-metrics-results/bray_curtis_distance_flock.qza \  
--p-pairwise
```