QIIME Analysis Pipeline - 16s rRNA (V3-V4)

```
# Create manifest file (tab-limited) with the columns
sample-id forward-absolute-filepath     reverse-absolute-filepath

# Import the FASTQ files into a QIIME artifact file (.qza file) that will contain the sequences and metadata
qiime tools import \
     --type 'SampleData[PairedEndSequencesWithQuality]' \
     --input-path manifest.txt  \
     --output-path paired-end-demux.qza \
     --input-format 'PairedEndFastqManifestPhred33V2'


# Remove amplicon primers with cutadapt
qiime cutadapt trim-paired \
     --i-demultiplexed-sequences paired-end-demux.qza \
     --p-cores  10 \
     --p-front-f CCTACGGGNGGCWGCAG \
     --p-front-r GACTACHVGGGTATCTAATCC  \
     --o-trimmed-sequences reads_trimmed.qza \
     --verbose > cutadapt_log.txt


# Generate quality plots and analyze sequence lengths, quality scores
qiime demux summarize \
     --i-data reads_trimmed.qza \
     --o-visualization reads_trimmed.qzv

# Denoising, length trimming, chimera and PhiX removal (DADA2)
qiime dada2 denoise-paired \
     --i-demultiplexed-seqs reads_trimmed.qza \
     --p-trim-left-f 0 \
     --p-trim-left-r 0 \
     --p-trunc-len-f 0 \
     --p-trunc-len-r 0 \
     --o-table featuretable_notrunc.qza \
     --o-representative-sequences repseqs_notrunc.qza \
     --o-denoising-stats denoisingstats.qza \
     --p-n-threads 10

# Create feature table summary
qiime feature-table summarize \
     --i-table table.qza \
     --o-visualization table_summary

# Tabulate representative sequences
```

```
qiime feature-table tabulate-seqs \
     --i-data representative_sequences.qza \
     --o-visualization rep_seqs

# Generate denoising stats
qiime metadata tabulate \
     --m-input-file denoising_stats.qza \
     --o-visualization denoising_stats

# Assign taxonomy to features

# Load SILVA database to Qiime2: get sequences
qiime tools import --type 'FeatureData[Sequence]' \
     --input-path \
     rep_set/rep_set_16S_only/99/silva_132_99_16S.fna \
     --output-path 99_otus_16S.qza

# Load SILVA database to Qiime2: get taxonomy strings
qiime tools import --type 'FeatureData[Taxonomy]' \
     --input-format HeaderlessTSVTaxonomyFormat \
     --input-path \
     taxonomy/16S_only/99/majority_taxonomy_7_levels.txt \
     --output-path 99_otus_16S_taxonomy.qza

# Extract V3-V4 region from the reference database
qiime feature-classifier extract-reads \
     --i-sequences 99_otus_16S.qza \
     --p-f-primer CCTACGGGNGGCWGCAG \
     --p-r-primer GACTACHVGGGTATCTAATCC \
     --p-min-length 250 \
     --p-max-length 500 \
     --o-reads ref_seqs_16S_V3-V4.qza \
     --verbose &> 16S_V3-V4_training.log

# Train the classifier on this region
qiime feature-classifier fit-classifier-naive-bayes \
     --i-reference-reads ref_seqs_16S_V3-V4.qza \
     --i-reference-taxonomy 99_otus_16S_taxonomy.qza \
     --o-classifier classifier_16S_V3-V4.qza \
     --verbose &> 16S_V3-V4_classifier.log

# Classify the representative sequences
qiime feature-classifier classify-sklearn \
     --i-classifier classifier_16S_V3-V4.qza \
     --i-reads representative_sequences.qza  \
     --o-classification classified_rep_seqs.qza
```

```
# Tabulate the features, their taxonomy and the confidence of taxonomy assignment
qiime metadata tabulate \
     --m-input-file classified_rep_seqs.qza \
     --o-visualization classified_rep_seqs.qzv

# Generate tree for phylogenetic diversity analysis
qiime phylogeny align-to-tree-mafft-fasttree \
     --i-sequences representative_sequences.qza \
     --output-dir phylogenetic_tree --p-n-threads 36 \
     --verbose &> phylogenetic_tree_generation.log


# Filter features which are present >=2 and with frequency >=10 for downstream differential
abundance analysis
qiime feature-table filter-features \
     --i-table table.qza \
     --p-min-samples 2 \
     --o-filtered-table temp2.qza

qiime feature-table filter-features \
     --i-table temp2.qza \
     --p-min-frequency 10 \
     --o-filtered-table temp3.qza

# Generate category-wise taxonomy plots for Figure 2.

# First, create a file called grouped_metadata.txt, which lists the category levels in the form of
a single column as below:
sample-id
Case
Control

qiime feature-table group --i-table  temp3.qza  \
     --p-axis 'sample' \
     --m-metadata-file sample_metadata.txt \
     --m-metadata-column Stunting \
     --p-mode 'mean-ceiling' \
     --o-grouped-table stunting_table.qza


qiime taxa barplot --i-table stunting_table.qza \
     --i-taxonomy classified_rep_seqs.qza \
     --m-metadata-file grouped_metadata.txt \
     --o-visualization stunting_taxabar.qzv


qiime feature-table group --i-table temp3.qza  \
```

```
     --p-axis 'sample' \
     --m-metadata-file sample_metadata.txt \
     --m-metadata-column Wasting \
     --p-mode 'mean-ceiling' \
     --o-grouped-table wasting_table.qza

qiime taxa barplot --i-table wasting_table.qza \
     --i-taxonomy classified_rep_seqs.qza \
     --m-metadata-file grouped_metadata.txt \
     --o-visualization wasting_taxabar.qzv

qiime feature-table group --i-table temp3.qza  \
     --p-axis 'sample' \
     --m-metadata-file sample_metadata.txt \
     --m-metadata-column UnderWt \
     --p-mode 'mean-ceiling' \
     --o-grouped-table underwt_table.qza

qiime taxa barplot --i-table underwt_table.qza \
     --i-taxonomy classified_rep_seqs.qza \
     --m-metadata-file grouped_metadata.txt \
     --o-visualization underwt_taxabar.qzv

# Generate rarefaction plots  (Figure S1)
qiime diversity alpha-rarefaction --i-table table.qza \
     --i-phylogeny rooted_tree.qza --p-max-depth 60623 \
     --m-metadata-file sample_metadata.txt \
     --o-visualization alpha_rarefaction.qzv

# Get core metrics including results of standard alpha and beta diversity metrics
qiime diversity core-metrics-phylogenetic \
     --i-phylogeny rooted_tree.qza \
     --i-table table.qza \
     --p-sampling-depth 66258 \
     --m-metadata-file sample_metadata.txt \
     --output-dir core_metrics --p-n-jobs 36 \
     --verbose  &> core_metrics_samples.log

# Export bray curtis ordination data for PCoA analysis (Figure 1). The file is exported as
ordination.txt inside the specified output folder. Further details explained in a later section
qiime  tools  export  --input-path  bray_curtis_pcoa_results.qza  --
output-path ./bray_pcoa/


# Test for category-wise differences in alpha diversity (generates tables for Figure S2)
for result in *vector.qza; do
outname=${result/_vector.qza/_group_significance.qzv};
```

```
qiime diversity alpha-group-significance \
     --i-alpha-diversity $result \
     --m-metadata-file sample_metadata.txt \
     --o-visualization $outname;
done
```

Generate beta diversity metrics

```
qiime diversity beta-rarefaction --i-table table.qza \
     --p-metric weighted_unifrac \
     --p-clustering-method nj \
     --m-metadata-file sample_metadata.txt \
     --p-sampling-depth 66258 \
     --i-phylogeny rooted_tree.qza \
     --o-visualization beta_rarefaction_w_unifrac.qzv
```

Test for category-wise differences in beta diversity metrics and generate plots (Figure S3). Repeat the same code for the remaining categories (stunting, wasting).

```
qiime diversity beta-group-significance \
     --i-distance-matrix unweighted_unifrac_distance_matrix.qza  \
     --m-metadata-file sample_metadata.txt \
     --m-metadata-column UnderWt  \
     --o-visualization unweighted-unifrac-underwt-significance.qzv \
     --p-pairwise

qiime diversity beta-group-significance \
     --i-distance-matrix weighted_unifrac_distance_matrix.qza  \
     --m-metadata-file sample_metadata.txt \
     --m-metadata-column UnderWt  \
     --o-visualization weighted-unifrac-underwt-significance.qzv \
     --p-pairwise

qiime diversity beta-group-significance \
     --i-distance-matrix bray_curtis_distance_matrix.qza  \
     --m-metadata-file sample_metadata.txt \
     --m-metadata-column UnderWt  \
     --o-visualization bray_curtis-underwt-significance.qzv \
     --p-pairwise

qiime diversity beta-group-significance \
     --i-distance-matrix jaccard_distance_matrix.qza  \
     --m-metadata-file  sample_metadata.txt \
     --m-metadata-column UnderWt  \
     --o-visualization jaccard-underwt-significance.qzv
     --p-pairwise
```

```
# Generate OTU counts for each taxonomic level by setting --p-level from 1 to 7
qiime taxa collapse --i-table temp3.qza \
     --i-taxonomy classified_rep_seqs.qza \
     --p-level 6 \
     --o-collapsed-table feature_table_level_6.qza

# Add pseudocount of 1 to all values to avoid zero values
qiime composition add-pseudocount \
     --i-table feature_table_for_ANCOM.qza \
     --o-composition-table Final_OTU_table.qza

# PICRUSt analysis
qiime picrust2 full-pipeline --i-table temp3.qza \
     --i-seq representative_sequences.qza \
     --output-dir picrust_out \
     --p-threads 20

qiime feature-table summarize \
     --i-table picrust_out/pathway_abundance.qza \
     --o-visualization picrust_out/pathway_abundance.qzv

qiime feature-table summarize \
     --i-table picrust _out/ec_metagenome.qza \
     --o-visualization picrust_out/ec_metagenome.qzv

qiime feature-table summarize \
     --i-table picrust_out/ko_metagenome.qza \
     --o-visualization picrust_out/ko_metagenome.qzv

qiime tools export \
     --input-path picrust_out/pathway_abundance.qza \
     --output-path pathabun_exported

qiime tools export \
     --input-path picrust_out/ec_metagenome.qza \
     --output-path ec_exported

qiime tools export \
     --input-path pic_out_1/ko_metagenome.qza \
     --output-path ko_exported

biom convert -i pathabun_exported/feature-table.biom \
     -o pathabun_exported/table41_onrnar_pathabund.tsv  --to-tsv
```

```
biom convert -i ec_exported/feature-table.biom \
     -o ec_exported/table41_nonrar_ec.tsv  --to-tsv

biom convert -i ko_exported/feature-table.biom \
     -o ko_exported/table41_nonrar_ko.tsv  --to-tsv
```

R code for Effect size and PCoA analysis (Figure 1)

```
library(vegan)
library("mixOmics")
library("ggpubr")

# The last table in the “ordination.txt” that was exported in the earlier section need to be saved as Bray_PCoA.txt with appropriate column headers.
bray_pcoa=read.table("Bray_PCoA.txt",sep="\t",header=1,row.names=1)

metadata=read.table("sample_metadata.txt",sep="\t",header=1,row.names
=1)

# Match sample order using match command if required.
metadata_reordered = metadata[match(rownames(bray_pcoa),
rownames(metadata))]

envfit_res=envfit(bray_pcoa, env= metadata_reordered[ ,c("Age_d",
"Sex","LAZ","WAZ","WLZ","Height_cm","Weight_kg","FFM%","FM%","TEE","L
RR","Hb","Mat_Edu_y")])

df_envfit=rbind(scores(envfit_res,display=c("vectors")),scores(envfit
_res,display=c("factors")))

df_envfit<-df_envfit*vegan:::ordiArrowMul(df_envfit)

df_envfit<-as.data.frame(df_envfit)

bray_pcoa_comb <- merge(bray_pcoa, metadata, by="row.names")

ggplot() + geom_point(data=bray_pcoa_comb, aes(Axis1,Axis2)) +
geom_segment(data=df_envfit*0.5, aes(x = 0, y = 0, xend = Axis1, yend
= Axis2), arrow = arrow(length = unit(0.2,"cm")), color="#4C005C",
alpha=0.5) + geom_text(data=as.data.frame(df_envfit*0.4),aes(Axis1,
Axis2, label = rownames(df_envfit)), color="#4C005C", alpha=0.5)+
theme_classic()
```

 # PLSDA Analysis

```
# Load the table, which contains the relative abundance data for taxa belonging to each taxonomic level in a single file

data_taxa=read.table("nonrar_all_taxa_RA.txt",sep="\t", header=T, row.names=1)

data_taxa_comb <- merge(data_taxa, metadata, by="row.names")

plsda.uw=plsda(data_taxa_comb[,2:457],data_taxa_comb[,c("Underwt")],ncomp=10)

plsda.st=plsda(data_taxa_comb[,2:457],data_taxa_comb[,c("Stunting"],ncomp=10)

plsda.wt=plsda(data_taxa_comb[,2:457],data_taxa_comb[,c("Wasting"],ncomp=10)

plotIndiv(plsda.st, col = brewer.pal(n = 3,name="Set1")[0:2], style = "graphics", plot.ellipse = T, ind.names =F, ellipse=TRUE,title = 'PLS-DA - Stunting', size.axis=1.3, size.title=1.5, size.xlabel=1.4, size.ylabel=1.4)
plotIndiv(plsda.wt, col = brewer.pal(n = 3,name="Set1")[0:2], style = "graphics", plot.ellipse = T, ind.names =F, ellipse=TRUE,title = 'PLS-DA - Wasting', size.axis=1.3, size.title=1.5, size.xlabel=1.4, size.ylabel=1.4)
plotIndiv(plsda.uw, col = brewer.pal(n = 3,name="Set1")[0:2], style = "graphics", plot.ellipse = T, ind.names =F, ellipse=TRUE,title = 'PLS-DA - Underweight', size.axis=1.3, size.title=1.5, size.xlabel=1.4, size.ylabel=1.4)
```