**16S amplicon analysis using QIIME2 (ubuntu)**

1. **About data**

In this tutorial, we use manifest type of data.

Download data: <https://drive.google.com/file/d/1D78AsU8T8EoDp1fAYLB8HhIIrHYpys29/view?usp=sharing>

* Extract to: …/reads/…

Two feacal sample of mouses was collected (MT28,MT66).

Each sample has two fastq files: Forward and reverse read.

(raw\_1,raw\_2).

1. **Importing data**

**2.1.Generate a manifest file:**

An example described here: <https://github.com/Anegin24/Microbiome-analysis/blob/main/Import-paired-end-sequences-qiime2%20(Ubuntu)>

echo -e 'sample-id\tforward-absolute-filepath\treverse-absolute-filepath' > manifest.tsv

for FOR in reads/\*\_raw\_1\*gz; do echo $FOR; done

for FOR in reads/\*\_raw\_1\*gz;

do ID=$(basename $FOR | cut -f1 -d\_);

REV=${FOR/\_raw\_1/\_raw\_2};

echo -e "${ID}\t${PWD}/${FOR}\t${PWD}/${REV}";

done >>manifest.tsv

**2.2. Import data to QIIME2**

qiime tools import \

--type 'SampleData[PairedEndSequencesWithQuality]' \

--input-path manifest.tsv \

--output-path reads.qza \

--input-format PairedEndFastqManifestPhred33V2

1. **Demultiplex sequences**

Because the data already have barcodes sequences so we don’t have to demux. Just use demux summarize and see the information of quality control inside.

qiime demux summarize **\**

--i-data reads.qza **\**

--o-visualization reads.qzv

1. **Denoising**

**Use dada2 to paire sequences with exactly contigs and trim barcodes-linkerprimersequences.**

**Because forward sequences have 6 nucleotides in barcodes and 17 nucleotides in linkerprimer so we p-trim-left-f 23. Same with reverse sequences.**

qiime dada2 denoise-paired **\**

--i-demultiplexed-seqs reads.qza **\**

--p-trim-left-f 23 **\**

--p-trim-left-r 26 **\**

--p-trunc-len-f 250 **\**

--p-trunc-len-r 250 **\**

--o-table table.qza **\**

--o-representative-sequences rep-seqs.qza **\**

--o-denoising-stats denoising-stats.qza

**Visualization output:**

qiime feature-table summarize **\**

--i-table table.qza **\**

--o-visualization table.qzv **\**

--m-sample-metadata-file sample-metadata.tsv

qiime feature-table tabulate-seqs **\**

--i-data rep-seqs.qza **\**

--o-visualization rep-seqs.qzv

qiime metadata tabulate **\**

--m-input-file denoising-stats.qza **\**

--o-visualization denoising-stats.qzv

## Generate a tree for phylogenetic diversity analyses

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

1. **Alpha/Beta diversity**

qiime diversity core-metrics-phylogenetic **\**

--i-phylogeny rooted-tree.qza **\**

--i-table table.qza **\**

--p-sampling-depth 1103 **\**

--m-metadata-file sample-metadata.tsv **\**

--output-dir core-metrics-results

1. **Taxonomy classifier**

**7.1. Database preparing**

**Step 1:** Download silva 138.1

qiime rescript get-silva-data \

*--p-version '138.1’ \*

*--p-target 'SSURef\_NR99’ \*

*--p-include-species-labels \*

*--o-silva-sequences silva-138.1-ssu-nr99-seqs.qza \*

*--o-silva-taxonomy silva-138.1-ssu-nr99-tax.qza*

**Step 2:** Remove low quality sequences

qiime rescript cull-seqs \

--i-sequences silva-138.1-ssu-nr99-seqs.qza \

--o-clean-sequences silva-138.1-ssu-nr99-seqs-cleaned.qza

**Step 3:** Remove Eukaryota sequences

qiime taxa filter-seqs \

--i-sequences silva-138.1-ssu-nr99-seqs-cleaned.qza \

--i-taxonomy silva-138.1-ssu-nr99-tax.qza \

--p-exclude 'd\_\_Eukaryota’ \

--p-mode 'contains’ \

--o-filtered-sequences silva138\_noEuk\_seqs.qza

**Step 4:** Filtering sequences by length and taxonomy

qiime rescript filter-seqs-length-by-taxon \

--i-sequences silva-138.1-ssu-nr99-seqs-cleaned.qza \

--i-taxonomy silva-138.1-ssu-nr99-tax.qza \

--p-labels Archaea Bacteria Eukaryota \

--p-min-lens 900 1200 1400 \

--o-filtered-seqs silva-138.1-ssu-nr99-seqs-filt.qza \

--o-discarded-seqs silva-138.1-ssu-nr99-seqs-discard.qza

**Step 5:** Dereplicate of sequences and taxonomy

qiime rescript dereplicate \

--i-sequences silva-138.1-ssu-nr99-seqs-filt.qza \

--i-taxa silva-138.1-ssu-nr99-tax.qza \

--p-rank-handles 'silva’ \

--p-mode 'uniq’ \

--o-dereplicated-sequences silva-138.1-ssu-nr99-seqs-derep-uniq.qza \

--o-dereplicated-taxa silva-138.1-ssu-nr99-tax-derep-uniq.qza

**7.2. Generate V1-V2 database**

**Step 6:** Primer 27f & 338r

qiime feature-classifier extract-reads

--i-sequences silva138.1\_noEuk\_AB\_seqs\_uniq.qza

--p-f-primer AGAGTTTGATCMTGGCTCAG

--p-r-primer TGCTGCCTCCCGTAGGAGT

--p-n-jobs 12

--o-reads silva138.1\_AB\_V1-V2seqs.qza

**Step 6:** Dereplicate of sequences and taxonomy

qiime rescript dereplicate \

--i-sequences silva138.1\_AB\_V1-V2seqs.qza \

--i-taxa silva138.1\_noEuk\_AB\_tax\_uniq.qza \

*--p-rank-handles 'silva' \*

*--p-mode 'uniq’ \*

--o-dereplicated-sequences silva138.1\_AB\_V1-V2seqs\_uniq.qza \

--o-dereplicated-taxa silva138.1\_AB\_V1-V2taxa\_uniq.qza

**Step 8:** Train taxonomy

qiime feature-**classifier** fit-**classifier**-naive-bayes \

*--i-reference-reads* silva138.1\_AB\_V1-V2seqs\_uniq.qza *\*

*--i-reference-taxonomy* silva138.1\_AB\_V1-V2taxa\_uniq.qza \

*--o-classifier* silva138.1\_AB\_V1-V2\_classifier.qza

**7.3. Taxonomy analysis**

qiime feature-classifier classify-sklearn **\**

--i-classifier V34classifier.qza **\**

--i-reads rep-seqs.qza **\**

--o-classification taxonomy.qza

qiime metadata tabulate **\**

--m-input-file taxonomy.qza **\**

--o-visualization taxonomy.qzv

qiime taxa barplot **\**

--i-table table.qza **\**

--i-taxonomy taxonomy.qza **\**

--m-metadata-file sample-metadata.tsv **\**

--o-visualization taxa-bar-plots.qzv