

All commands in this document were run in a QIIME 2 2019.7 environment by Greg Caporaso.

Full data provenance can be viewed at <https://view.qiime2.org> for all data artifacts contained in this directory.

## Data preparation and denoising

```
sqiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --input-path run1-manifest.tsv
--output-path run1-demux.qza --input-format PairedEndFastqManifestPhred33V2
```

```
sqiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --input-path run2-manifest.tsv
--output-path run2-demux.qza --input-format PairedEndFastqManifestPhred33V2
```

```
sqiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --input-path run3-manifest.tsv
--output-path run3-demux.qza --input-format PairedEndFastqManifestPhred33V2
```

```
sqiime demux summarize --i-data run1-demux.qza --o-visualization run1-demux.qzv
sqiime demux summarize --i-data run2-demux.qza --o-visualization run2-demux.qzv
sqiime demux summarize --i-data run3-demux.qza --o-visualization run3-demux.qzv
```

```
sbatch -t 2880 -c 32 --mem 300000 qiime dada2 denoise-single --i-demultiplexed-seqs run1-demux.qza
--output-dir dada2-single-run1 --p-trunc-len 150 --p-trim-left 13 --p-n-threads 0
```

```
sbatch -t 2880 -c 32 --mem 300000 qiime dada2 denoise-single --i-demultiplexed-seqs run2-demux.qza
--output-dir dada2-single-run2 --p-trunc-len 150 --p-trim-left 13 --p-n-threads 0
```

```
sbatch -t 2880 -c 32 --mem 300000 qiime dada2 denoise-single --i-demultiplexed-seqs run3-demux.qza
--output-dir dada2-single-run3 --p-trunc-len 150 --p-trim-left 13 --p-n-threads 0
```

```
sbatch -t 2880 -c 32 --mem 300000 qiime dada2 denoise-paired --i-demultiplexed-seqs run1-demux.qza
--output-dir dada2-paired-run1 --p-trunc-len-f 150 --p-trunc-len-r 143 --p-trim-left-f 13 --p-trim-left-r 13
--p-n-threads 0
```

```
sbatch -t 2880 -c 32 --mem 300000 qiime dada2 denoise-paired --i-demultiplexed-seqs run2-demux.qza
--output-dir dada2-paired-run2 --p-trunc-len-f 150 --p-trunc-len-r 143 --p-trim-left-f 13 --p-trim-left-r 13
--p-n-threads 0
```

```
sbatch -t 2880 -c 32 --mem 300000 qiime dada2 denoise-paired --i-demultiplexed-seqs run3-demux.qza
--output-dir dada2-paired-run3 --p-trunc-len-f 150 --p-trunc-len-r 143 --p-trim-left-f 13 --p-trim-left-r 13
--p-n-threads 0
```

```
sqiime feature-table summarize --i-table dada2-single-run1/table.qza --o-visualization
dada2-single-run1/table.qzv
sqiime feature-table summarize --i-table dada2-single-run2/table.qza --o-visualization
dada2-single-run2/table.qzv
```

```
sqiime feature-table summarize --i-table dada2-single-run3/table.qza --o-visualization  
dada2-single-run3/table.qzv
```

```
sqiime feature-table summarize --i-table dada2-paired-run1/table.qza --o-visualization  
dada2-paired-run1/table.qzv
```

```
sqiime feature-table summarize --i-table dada2-paired-run2/table.qza --o-visualization  
dada2-paired-run2/table.qzv
```

```
sqiime feature-table summarize --i-table dada2-paired-run3/table.qza --o-visualization  
dada2-paired-run3/table.qzv
```

*Losing too many sequences during paired-end read joining so proceeding with single-end reads.*

```
mkdir dada2-single
```

```
sqiime feature-table merge --i-tables dada2-single-run1/table.qza --i-tables dada2-single-run2/table.qza  
--i-tables dada2-single-run3/table.qza --o-merged-table dada2-single/table.qza
```

```
sqiime feature-table merge-seqs --i-data dada2-single-run1/representative_sequences.qza --i-data  
dada2-single-run2/representative_sequences.qza --i-data dada2-single-run3/representative_sequences.qza  
--o-merged-data dada2-single/representative_sequences.qza
```

```
qiime feature-table summarize --i-table dada2-single/table.qza --o-visualization dada2-single/table.qzv  
--m-sample-metadata-file sample-metadata.tsv
```

```
sqiime feature-classifier classify-sklearn --i-classifier  
/projects/microbiome/greg-analysis/gg-13-8-99-nb-classifier.qza --i-reads  
dada2-single/representative_sequences.qza --o-classification dada2-single/taxonomy-gg.qza
```

```
sbatch --time=24:00:00 --mem 64000 qiime feature-classifier classify-sklearn --i-classifier  
/projects/microbiome/greg-analysis/silva-132-99-nb-classifier.qza --i-reads  
dada2-single/representative_sequences.qza --o-classification dada2-single/taxonomy-silva.qza
```

```
sqiime fragment-insertion sepp --i-representative-sequences dada2-single/representative_sequences.qza  
--o-tree dada2-single/rooted-tree-sepp.qza --o-placements dada2-single/insertion-placements.qza
```

```
sbatch --time=24:00:00 --mem 16000 --dependency afterok:22535466 qiime fragment-insertion filter-features  
--i-table dada2-single/table.qza --i-tree dada2-single/rooted-tree-sepp.qza --o-filtered-table  
dada2-single/table-sepp.qza --o-removed-table dada2-single/junk.qza
```

*Group replicates by site by summing the counts*

```
qiime feature-table group --i-table dada2-single/table.qza --o-grouped-table dada2-single/site-table.qza  
--m-metadata-file sample-metadata.tsv --m-metadata-column site-id --p-axis sample --p-mode sum
```

```
qiime feature-table summarize --i-table dada2-single/site-table.qza --o-visualization dada2-single/site-table.qzv  
--m-sample-metadata-file site-metadata.tsv
```

```
sbatch --time=24:00:00 --mem 16000 --dependency afterok:22535466 qiime fragment-insertion filter-features  
--i-table dada2-single/site-table.qza --i-tree dada2-single/rooted-tree-sepp.qza --o-filtered-table  
dada2-single/site-table-sepp.qza --o-removed-table dada2-single/junk2.qza
```

## Diversity analyses

```
qiime diversity core-metrics-phylogenetic --i-table site-table-sepp.qza --i-phylogeny rooted-tree-sepp.qza  
--m-metadata-file ../site-metadata.tsv --p-sampling-depth 7164 --output-dir cm7164
```

```
qiime diversity alpha-group-significance --i-alpha-diversity faith_pd_vector.qza --m-metadata-file  
../site-metadata.tsv --o-visualization faith-pd-group-significance.qza
```

```
qiime diversity alpha-group-significance --i-alpha-diversity observed_otus_vector.qza --m-metadata-file  
../site-metadata.tsv --o-visualization oo-group-significance.qza
```

```
qiime diversity beta-group-significance --i-distance-matrix unweighted_unifrac_distance_matrix.qza  
--m-metadata-file ../site-metadata.tsv --m-metadata-column saprophytic-positive --o-visualization  
uu-group-significance-saprophytic.qzv
```

```
qiime diversity beta-group-significance --i-distance-matrix unweighted_unifrac_distance_matrix.qza  
--m-metadata-file ../site-metadata.tsv --m-metadata-column infectious-positive --o-visualization  
uu-group-significance-infectious.qzv
```

```
qiime diversity beta-group-significance --i-distance-matrix unweighted_unifrac_distance_matrix.qza  
--m-metadata-file ../site-metadata.tsv --m-metadata-column Positivity_100copies --o-visualization  
uu-group-significance-infectious-or-saprophytic.qzv
```

```
qiime diversity beta-group-significance --i-distance-matrix weighted_unifrac_distance_matrix.qza  
--m-metadata-file ../site-metadata.tsv --m-metadata-column saprophytic-positive --o-visualization  
wu-group-significance-saprophytic.qzv
```

```
qiime diversity beta-group-significance --i-distance-matrix weighted_unifrac_distance_matrix.qza  
--m-metadata-file ../site-metadata.tsv --m-metadata-column infectious-positive --o-visualization  
wu-group-significance-infectious.qzv
```

```
qiime diversity beta-group-significance --i-distance-matrix weighted_unifrac_distance_matrix.qza  
--m-metadata-file ../site-metadata.tsv --m-metadata-column Positivity_100copies --o-visualization  
wu-group-significance-infectious-or-saprophytic.qzv
```

*Run Benjamini-Hochberg FDR correction on all reported p-values.*

```
>>> import statsmodels.stats.multitest  
>>> statsmodels.stats.multitest.multipletests([0.959, 0.004, 0.013, 0.838,  
0.001, 0.006, 0.018, 0.004, 0.127, 0.006, 0.064, 0.525], alpha=0.01,  
method='fdr_bh')[1]
```

```
array([0.959      , 0.0144      , 0.026      , 0.91418182, 0.012      ,  
       0.0144      , 0.03085714, 0.0144      , 0.16933333, 0.0144      ,  
       0.096      , 0.63          ])
```

## Differential abundance testing (genera)

```
qiime taxa collapse --i-table ../site-table-sepp.qza --i-taxonomy ../taxonomy-gg.qza --p-level 6  
--o-collapsed-table ../site-table-sepp-genus.qza
```

```
qiime composition add-pseudocount --i-table ../site-table-sepp-genus.qza --o-composition-table  
../site-table-sepp-composition-genus.qza
```

```
sqiime composition ancom --i-table site-table-sepp-composition-genus.qza --m-metadata-file  
../../site-metadata.tsv --m-metadata-column Positivity_100copies --o-visualization  
ancom-genus-infectious-or-saprophytic.qzv
```

```
sqiime composition ancom --i-table site-table-sepp-composition-genus.qza --m-metadata-file  
../../site-metadata.tsv --m-metadata-column saprophytic-positive --o-visualization  
ancom-genus-saprophytic.qzv
```

```
sqiime composition ancom --i-table site-table-sepp-composition-genus.qza --m-metadata-file  
../../site-metadata.tsv --m-metadata-column infectious-positive --o-visualization ancom-genus-infectious.qzv
```