**1. Activate the qiime environment before running any qiime commands**

source /home/opt/Miniconda3/miniconda3/bin/activate qiime2-2023.9

The above gives you the latest version of qiime2. Sometimes you can't run files from a previous version of qiime2 in a new version. If you want to work with older files you can activate the previous version that is still on the server (below).

source /opt/Miniconda2/miniconda2/bin/activate qiime2-2021.2

**2. Import sequencing files into QIIME2**

Make a directory in a specific projects folder to contain reads and barcode files

mkdir reads

cp Undetermined\_S0\_L001\_R1\_001.fastq.gz ./reads

cp Undetermined\_S0\_L001\_R2\_001.fastq.gz ./reads

cp Undetermined\_S0\_L001\_I1\_001.fastq.gz ./reads

When amplicon sequencing is performed at CU Boulder it will come back in this format. Other sequencing facilities may send it back as forward and reverse read files for every sample file. If you receive sample files that way then you must import as a manifest file. You can visit this <https://docs.qiime2.org/2017.12/tutorials/importing/> page on the qiime2 website to get examples of manifest files or to read more about importing using the EMP format we will use in this tutorial.

**3. Rename forward, reverse, and barcode fastq.gz files**

cd reads

mv Undetermined\_S0\_L001\_I1\_001.fastq.gz barcodes.fastq.gz

mv Undetermined\_S0\_L001\_R1\_001.fastq.gz forward.fastq.gz

mv Undetermined\_S0\_L001\_R2\_001.fastq.gz reverse.fastq.gz

cd..

First, enter the new directory that your read files are in, rename the read files as shown, and then go up one directory

**4. Import Sequence files**

If using EMP sequences that have one file for forward reads, one file for reverse reads, and one file with associated barcodes and is still multiplexed:

qiime tools import --type EMPPairedEndSequences --input-path reads/ --output-path emp-paired-end-sequences.qza

**5. Demultiplex the sequences**

qiime demux emp-paired \

--m-barcodes-file metadata.txt \

--m-barcodes-column BarcodeSequence \

--p-no-rev-comp-mapping-barcodes \

--p-no-golay-error-correction\

--i-seqs emp-paired-end-sequences.qza \

--o-per-sample-sequences demux.qza \

--o-error-correction-details demux-details.qza

--i-seqs: .qza file from step 4 --m-barcodes-file: a text file that contains barcodes for the samples we want in the feature table. It can also include other important metadata. Make sure the barcode column is named barcode-sequence (or matches whatever is indicted in next --m-barcodes-column)

--m-barcodes-column: barcode-sequence indicates that in the mapping file, barcodes are in a column labeled 'barcode-sequence'

--o-per-sample-sequences : output file containing reads separated by sample

--o-error-correction-details: provides stats about the error correction process on your samples

--p-no-rev-comp-mapping-barcodes: specifies barcodes don't need to be reverse complemented but vary based on sequencing center. If sequences per sample are low this can be adjusted. For 16S samples from CU Boulder you want to include this.

--p-no-golay-error-correction: doesn't perform 12nt Golay error correction on barcode reads. This is similar to the step above, and will not be used for samples from every facility but will be used for 16S samples from from CU Boulder

**6. Visualize sequence reads and sequence quality**

qiime demux summarize --i-data demux.qza --o-visualization demux.qzv

After obtaining .qzv file can visualize it in the QIIME2 view tools: <https://view.qiime2.org/> The overview tab allows you to see the number of forward and reverse reads for each sample Interactive Quality Plot allows you to see the quality scores of the forward and reverse reads based on sequence base pairs. This is useful for deciding where to trim in the following step. Frequently, quality scores will drop before the end of the sequencing run. You will want to choose trim lengths to remove low-quality reads but still allow enough overlap that forward and reverse reads can be merged.

If you are unsure of your sequencing length this visualization allows you to see how long the forward and reverse reads are as well as look at the average quality of each base pair based on its position in the sequence. We generally see poorer read quality at the beginning and end of reads so we trim that portion away.

**7. Run DADA2 to denoise and merge sequence reads**

DADA2 is a pipeline that detects and removes Illumina sequencing errors when possible and merges forward and reverse reads. The output is called a feature table (table.qza) consisting of amplicon sequence variants (ASVs) and their count per sample. Additionally, the rep-seqs.qza file contains each ASV and its entire sequence (this file does not have any count or sample info, only sequences matched with ASV id)

qiime dada2 denoise-paired \

--i-demultiplexed-seqs demux.qza \

--p-trim-left-f 0 \

--p-trim-left-r 0 \

--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

If you have paired-end reads you can trim the forward and reverse reads separately from each other, each at the beginning and end of the sequence. Where you trim depends on quality reads which you can determine using the demux.qzv file. --p-trim-left-f: the trim length at the beginning of the forward read

--p-trim-left-r: the trim length at the beginning of the reverse read

--p-trunc-len-f: the truncation length at the end of the forward read

--p-trunc-len-r: the truncation length at the end of the reverse read

--o-table: output table that has the number of each ASV per sample

--o-representative-sequences: a file that contains every unique ASV

Run the following code to get a qzv table of the output. This will allow you to see how many reads were retained during the dada2 filtering step.

qiime metadata tabulate \

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

--o-visualization denoising-stats.qzv

**8. Visualize your feature table**

qiime feature-table summarize \

--i-table table.qza \

--o-visualization table.qzv \

--m-sample-metadata-file metadata.txt

**9. Visualize representative sequences**

qiime feature-table tabulate-seqs \

--i-data rep-seqs.qza \

--o-visualization rep-seqs.qzv

After obtaining .qzv file can visualize it in the QIIME2 view tools: <https://view.qiime2.org/>

**10. Assign taxonomy to reads** Classifiers for the Earth Microbiome Primers (EMP) (515f/806R) have been trained and are available on the server for both Silva 138, GTDB 207, and GTDB X. Make sure to check you are using the appropriate version. If you need to train a classifier because of a new database release or you are using different primers, detailed instructions can be found <https://docs.qiime2.org/2019.10/tutorials/feature-classifier/>

Classifying with Silva 138

qiime feature-classifier classify-sklearn \

--i-classifier /home/Database/qiime2/classifiers/qiime2-2021.2/silva-138-99-515-806-nb-classifier.qza \

--i-reads rep-seqs.qza \

--o-classification taxonomy\_silva138\_16S.qza

Classifying with GTDB 214.1

qiime feature-classifier classify-sklearn \

--i-classifier /home/projects-wilkins/PPH\_2022/amplicon/qiime2-2023.9\_classifiers/GTDBclassifier214.1\_EMP.qza \

--i-reads rep-seqs.qza \

--o-classification taxonomy\_gtdb\_214.qza

Classifying with GTDB 207

qiime feature-classifier classify-sklearn \

--i-classifier /home/Database/qiime2/classifiers/qiime2-2021.2/GTDBclassifier\_emp\_r207.qza \

--i-reads rep-seqs.qza \

--o-classification taxonomy\_gtdb\_207.qza

The follwing step converts the qza file to a qvz file to visualize. From the qzv window, you can download a csv file with each ASVs taxonomic assignment

qiime metadata tabulate \

--m-input-file taxonomy\_gtdb\_214.qza \

--o-visualization taxonomy\_gtdb\_214.qzv

**11. Filter out Chloroplast and Mitochondria reads**

qiime taxa filter-table \

--i-table table.qza \

--i-taxonomy taxonomy\_gtdb\_214.qza \

--p-exclude mitochondria,chloroplast,Unassigned,Eukaryota \

--o-filtered-table table-gtdb214\_no\_chlorplast.qza

**12. Convert raw feature table to text file**

unzip table-gtdb214\_no\_chlorplast.qza

This will produce a .biom file in a new folder. The new folder will have a name that is a string of letters and numbers, I rename the file to be more descriptive. To convert the biom file to a text file

biom convert -i table.biom -o table.from\_biom.txt --to-tsv

Can merge feature table with taxonomy file based on ASV identity Useful link for file conversion: <https://biom-format.org/documentation/biom_conversion.html>

**Optional**

Create Tables Based on Different Taxonomic Levels

1= Domain; 2= Phylum; 3= Class; 4= Order; 5= Family; 6= Genus; 7= Species

qiime taxa collapse \

--i-table table-gtdb214\_no\_chlorplast.qza\

--i-taxonomy taxonomy\_gtdb214\_16S.qza \

--p-level 3\

--o-collapsed-table table-class\_gtdb214.qza

unzip table-class\_gtdb214.qza

This will produce a .biom file in a new folder. to convert the biom file to a text file

biom convert -i table.biom -o table.from\_biom.txt --to-tsv

**13. Rarefy feature table so each sample has uniform reads**

Rarefying amplicon data continues to be a controversial topic. Arguments against rarefying data include that important sequencing info may be lost especially concerning rare species. However, assumptions of many statistical tests for alpha and beta diversity are not met unless a form of normalization is applied to the data. No normalization method is perfect but normalization is an important process for many statistical analyses and rarefying data continues to be the widest applied method. This video discusses the pros and cons of different normalization methods (<https://riffomonas.org/code_club/2022-02-24-normalization>).

Paper Discussion: <https://journals.asm.org/doi/10.1128/msphere.00355-23>

First, you must generate a tree that can be used for phylogenetic diversity analysis such as Faith's Phylogenetic Diversity and weighted/unweighted UniFrac distances.

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

The level you decide to rarefy your samples to is based on several factors including your study system and sequencing depth. Several tests can be run to help you decide what depth to rarefy to, given here in this R code. Additionally, it is useful to look at the table.qzv file to understand how many reads each sample has. Another useful parameter is rarefaction which allows you to understand at what sampling depth you have captured most of the community's diversity. This can be done in R or in QIIME2.

qiime diversity alpha-rarefaction \

--i-table table-gtdb214\_no\_chlorplast.qza \

--i-phylogeny rooted-tree.qza \

--p-max-depth 15000 \

--m-metadata-file metadata.txt \

--o-visualization alpha-rarefaction.qzv

Next, you will run QIIME2's diversity analysis. This process produces a rarefied table that can be exported and used in R. It also produces outputs for many different diversity analyses for both alpha and beta diversity that can be directly imported into R, however, we will rerun these analyses in R. The step called --p-sampling depth is the number to which you will rarefy your samples, in this example I rarefied to 15,000 reads. QIIME2 rarefies using a single sub-sampling event.

qiime diversity core-metrics-phylogenetic \

--i-phylogeny rooted-tree.qza \

--i-table table-gtdb214\_no\_chlorplast.qza \

--p-sampling-depth 15000 \

--m-metadata-file metadata.txt \

--output-dir core-metrics-results-15000

You can export rarefied tables similarly to the above step 12. Your rarefied table will be in the core-metrics-results folder created in the previous step.

unzip core-metrics-results-15000/rarefied\_table.qza

This will produce a .biom file in a new folder. The new folder will have a name that is a string of letters and numbers, I rename the file to be more descriptive. Next, convert the biom file to a text file.

biom convert -i table.biom -o table.from\_biom.txt --to-tsv

You can merge the rarefied feature table with the taxonomy file based on ASV identity. You can also produce collapsed taxa tables of your rarefied tables following the same steps given above.