ASSIGNMENT: Turn in fully commented QIIME2 code and a tutorial describing the analysis data and results. 1) Put all the QIIME2 code for the moving pictures tutorial into a text file and comment every line (#) explaining the purpose of the command or option.

mkdir qiime2-moving-pictures-tutorial

cd qiime2-moving-pictures-tutorial

#this step help with keeping everything organized in one place.

Sample metadata

Before starting the analysis, explore the sample metadata to familiarize yourself with the samples used in this study. The [sample metadata](https://data.qiime2.org/2024.10/tutorials/moving-pictures/sample_metadata) is available as a Google Sheet. You can download this file as tab-separated text by selecting File > Download as > Tab-separated values. Alternatively, the following command will download the sample metadata as tab-separated text and save it in the file sample-metadata.tsv. This sample-metadata.tsv file is used throughout the rest of the tutorial.

Please select a download option that is most appropriate for your environment:

* [Browser](https://docs.qiime2.org/2024.10/tutorials/moving-pictures/#browser-0)
* [wget](https://docs.qiime2.org/2024.10/tutorials/moving-pictures/#wget-0)
* [curl](https://docs.qiime2.org/2024.10/tutorials/moving-pictures/#curl-0)

wget \

-O "sample-metadata.tsv" \

"https://data.qiime2.org/2024.10/tutorials/moving-pictures/sample\_metadata.tsv"

#this command to download the sample metadata.

Obtaining and importing data

mkdir emp-single-end-sequences

wget \

-O "emp-single-end-sequences/barcodes.fastq.gz" \

"https://data.qiime2.org/2024.10/tutorials/moving-pictures/emp-single-end-sequences/barcodes.fastq.gz"

wget \

-O "emp-single-end-sequences/sequences.fastq.gz" \

"https://data.qiime2.org/2024.10/tutorials/moving-pictures/emp-single-end-sequences/sequences.fastq.gz"

#this command to work small subset of the complete sequence data, so command can run quickly.

qiime tools import \

--type EMPSingleEndSequences \ 🡺 this code means EMP = Earth Microbiome Project format, SingleEnd = Single-end reads (not paired-end), Sequences = importing raw FASTQ data, not processed sequences

--input-path emp-single-end-sequences \

--output-path emp-single-end-sequences.qza

#this code to imports raw sequence data in Earth Microbiome Project (EMP) format into QIIME 2 as a .qza artifact. The --type EMPSingleEndSequences flag tells QIIME that the input contains single-end reads that are multiplexed (barcoded but not yet assigned to samples). The input is a folder containing sequences.fastq.gz and barcodes.fastq.gz, and the output is an artifact named emp-single-end-sequences.qza.

Demultiplexing sequences

qiime demux emp-single \ 🡺 EMP single-end demultiplexing method

--i-seqs emp-single-end-sequences.qza \ 🡺 sequences have sample barcodes that need to be matched with metadata.

--m-barcodes-file sample-metadata.tsv \ 🡺 This is a TSV file mapping each sample ID to its barcode sequence

--m-barcodes-column barcode-sequence \ 🡺 Tells QIIME which column in the metadata contains the barcodes

--o-per-sample-sequences demux.qza \ 🡺 the **main output.**

--o-error-correction-details demux-details.qza 🡺 This file shows any barcode correction attempts that were made.

# this code for  knowing which barcode sequence is associated with each sample.

qiime demux summarize \

--i-data demux.qza \

--o-visualization demux.qzv

#this code to generate a summary for demultiplexing results. To determine how many sequences were obtained per sample.

Sequence quality control and feature table construction

#this include multiple quality control methods such as DADA2 and Deblur which are interchangeable. dada2 denoise-single

**qiime dada2 denoise-single \**

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

**--p-trim-left 0 \**

**--p-trunc-len 120 \**

**--o-representative-sequences rep-seqs-dada2.qza \**

**--o-table table-dada2.qza \**

**--o-denoising-stats stats-dada2.qza**

**-------------------------------------------------------------**

**Differential abundance testing with ANCOM-BC**

**#To identify features that are differentially abundant features across groups.**

**qiime feature-table filter-samples \**

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

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

**--p-where "[body-site]='gut'" \**

**--o-filtered-table gut-table.qza**

**Deblur**

**qiime quality-filter q-score \**

**--i-demux demux.qza \**

**--o-filtered-sequences demux-filtered.qza \**

**--o-filter-stats demux-filter-stats.qza**

**#This code for associate erroneous sequence reads with the true biological sequence from which they are derived.**

**-** **qiime deblur denoise-16S method**

**qiime deblur denoise-16S \**

**--i-demultiplexed-seqs demux-filtered.qza \ 🡺 to read the demultiplexed and quality-filtered single-end 16s reads.**

**--p-trim-length 120 \ 🡺 to be sure no bases get trimmed from the beginning.**

**--o-representative-sequences rep-seqs-deblur.qza \ 🡺 Denoised representative sequences (ASVs), like in DADA2.**

**--o-table table-deblur.qza \ 🡺** **A feature table and frequency of ASVs for a sample.**

**--p-sample-stats \ 🡺 Tells Deblur to output per-sample stats**

**--o-stats deblur-stats.qza 🡺 read filtering, error removal, and ASV assignment for each sample.**

**#This code/method helps with truncating the sequences at position.**

**qiime metadata tabulate \**

**--m-input-file demux-filter-stats.qza \**

**--o-visualization demux-filter-stats.qzv**

**qiime deblur visualize-stats \**

**--i-deblur-stats deblur-stats.qza \**

**--o-visualization deblur-stats.qzv**

**#The code above will trim the initial bases**

**#This code is for reading sequence error profiles to associate with sequence reads with rue biological sequence.**

**qiime quality-filter q-score \**

**--i-demux demux.qza \ 🡺 this one to demultiplexed single-end in Qiime2 format.**

**--o-filtered-sequences demux-filtered.qza \**

**--o-filter-stats demux-filter-stats.qza**

**#This code is for knowing the sequence error profiles to associate erroneous sequence reads.**

**qiime deblur denoise-16S \ 🡺** **sing the deblur plugin and the denoise-16S method, designed for 16S data**

**--i-demultiplexed-seqs demux-filtered.qza \ 🡺 the input file**

**--p-trim-length 120 \ 🡺 Truncates all sequences to 120 bp. Deblur requires all sequences to be the same length, so you choose a length that keeps most of your high-quality reads.**

**--o-representative-sequences rep-seqs-deblur.qza \ 🡺 ASVs (deblurred sequences)**

**--o-table table-deblur.qza \ 🡺 feature table showing how many times each ASV was found in each sample.**

**--p-sample-stats \ 🡺 collect per-sample statistics**

**--o-stats deblur-stats.qza 🡺 The statistics file summarizing**

**To  performing a differential abundance test at a specific taxonomic level**

**qiime taxa collapse \ 🡺 merges features into higher level**

**--i-table gut-table.qza \ 🡺 original feature table (counts of ASVs per sample)**

**--i-taxonomy taxonomy.qza \ 🡺 rtifact with taxonomy annotations for each ASV**

**--p-level 6 \**

**--o-collapsed-table gut-table-l6.qza 🡺 Collapsed feature table at level 6 (genus)**

**qiime composition ancombc \ 🡺 method from the composition plugin**

**--i-table gut-table-l6.qza \ 🡺 genus-level table from the previous step.**

**--m-metadata-file sample-metadata.tsv \ 🡺 A TSV file that contains sample groupings**

**--p-formula 'subject' \ 🡺 he column(s) in your metadata used to compare groups.**

**--o-differentials l6-ancombc-subject.qza 🡺 le with the ANCOM-BC results**

**#To Visualize Differential Abundance**

**qiime composition da-barplot \ 🡺 barplot visualizer,**

**--i-data l6-ancombc-subject.qza \ 🡺 differential abundance results you just got from ANCOM-BC**

**--p-significance-threshold 0.001 \ 🡺 how taxa with q-value < 0.001**

**--p-level-delimiter ';' \ 🡺 Tells QIIME how taxonomic levels are separated in the taxonomy strings.**

**--o-visualization l6-da-barplot-subject.qzv 🡺 barplot showing differentially abundant taxa at genus level.**