#Coding Assignment Final Project Basic Steps

#all files must be saved as .txt or tab delineated file

Echo "Importing the FAstq Data"

# \* anything in that folder

Echo "Creating the manifest file in Excel"

#Make sure to copy sequences

cd #current working directory where files are

ls -1 "$PWD/"\*R1\*

ls -1 "$PWD/"\*R2\*

# print working directory and then print files that have any thing before R2 and anything after but must contain the R2

Echo "importing Meta Data and manifest for samples and positive controls to the depot and scratch space; depot is permanent storage while scratch is 30 day working storage"

cd \Users\gfg\OneDrive - purdue.edu\Microbiome Analysis Class

scp metadata\_henk.txt horn83.rcac.purdue:/depot/microbiome/data/ANSC516/Connor

scp metadata\_henk.txt horn83@bell.rcac.purdue:/depot/microbiome/data/ANSC516/Connor

scp Meta\_Data\_Control\_Samples.txt roy174@bell.rcac.purdue:/depot/microbiome/data/ANSC516/Kayla

scp Meta\_Data\_Control\_Samples.txt horn83@bell.rcac.purdue:/scratch/bell/horn83/RS\*/

scp Manifest-Project-Revised.txt horn83@bell.rcac.purdue:/depot/microbiome/data/ANSC516/Connor

scp Manifest-Project-Revised.txt horn83@bell.rcac.purdue:/depot/microbiome/data/ANSC516/Connor

Echo "part of code that can be made into a slurm; creation of the demux file"

# Creating a Demux.qza file in order to determine how much sequence to remove based on quality scores you need to known amplicon length - foward keep + reverse keep= bp overlap

#you want at least 50bp overall and you want 13 bp removal and to keep quality scores 25% tile above a 20

module load bioinfo

module load Qiime/2-2022.8

cd /scratch/bell/horn83/qiime2-moving-pictures-tutorial/

#list the version of qiime you want to load

#qiime tools import \

#--type 'SampleData[PairedEndSequencesWithQuality]' \

#--input-path <manifest\_henk.txt> \

#--input-format PairedEndFastqManifestPhred33V2 \

#--output-path demux.qza

# example code want paired because we have foward and reverse although there is single and \ and then without running the command

Echo "demux samples and demux controls"

qiime tools import \

--type 'SampleData[PairedEndSequencesWithQuality]' \

--input-path manifest\_henk.txt \

--input-format PairedEndFastqManifestPhred33V2 \

--output-path demux.qza

#Convert to qza so can read and determine the sequence lengths to cut

#qiime demux summarize \

# --i-data demux.qza \

# --o-visualization demux.qzv

# order is program , package, function for line one

#in powershell

# cd where you want files to be deposited

scp horn83@bell.rcac.purdue:/scratch/bell/horn83/qiime2-moving-pictures-tutorial \*

scp horn83@bell.rcac.purdue:/scratch/bell/horn83/qiime2-moving-pictures-tutorial \*

#Echo DAD2 step

qiime dada2 denoise-paired \--i-demultiplexed-seqs demux.qza \

--p-trim-left-f 13 \

--p-trim-left-r 9 \

--p-trunc-len-f 296 \

--p-trunc-len-r 223 \

--o-table tableC.qza \

--o-representative-sequences rep-seqsC.qza \

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

# want to do the filter step on your table before you read out the table but you need a classifer

# convert to qzv for controls

qiime metadata tabulate \

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

--o-visualization denoising-statsC.qzv

qiime feature-table summarize \

--i-table tableC.qza \

--o-visualization tableC.qzv \

--m-sample-metadata-file Meta\_Data\_Control\_Samples.txt

qiime feature-table tabulate-seqs \

--i-data rep-seqsC.qza \

--o-visualization rep-seqsC.qzv

qiime taxa filter-table \

--i-table table.qza \

--i-taxonomy taxonomy.qza \

--p-exclude mitochondria,chloroplast \

--o-filtered-table table-no-mitochondria-no-chloroplast.qza

#Rename files so that the filtered one is table.qza

mv table.qza table-with-mitochondria-and-chloroplast.qza

mv table-no-mitochondria-no-chloroplast.qza table.qza

#Classifier

cd /depot/microbiome/data/ANSC516/Connor

mkdir classifier

cd classifier

wget <https://data.qiime2.org/2022.8/common/silva-138-99-seqs.qza>

wget <https://data.qiime2.org/2022.8/common/silva-138-99-tax.qza>

qiime feature-classifier extract-reads \

--i-sequences silva-138-99-seqs.qza \

--p-f-primer GTGYCAGCMGCCGCGGTAA \

--p-r-primer GGACTACNVGGGTWTCTAAT \

--p-min-length 100 \

--p-max-length 500 \

--p-n-jobs 12 \

--o-reads ref-seqs.qza

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

--i-reference-reads ref-seqsC.qza \

--i-reference-taxonomy silva-138-99-tax.qza \

--o-classifier silva-classifier.qza

qiime feature-classifier classify-sklearn \

--i-classifier gg-13-8-99-515-806-nb-classifier.qza \

--i-reads rep-seqsC.qza \

--o-classification taxonomy.qza

qiime picrust2 full-pipeline \

--i-table tableC.qza \

--i-seq rep-seqsC.qza \

--output-dir q2-picrust2\_output \

--p-placement-tool sepp \

--p-threads 1 \

--p-hsp-method pic \

--p-max-nsti 2 \

--verbose