#pipeline do qiime2 passo a passo

#para usar no pc do sage

#instalar o miniconda

wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86\_64.sh

bash Miniconda3-latest-Linux-x86\_64.sh

#dizer sim p tudo

#instalar o qiime2

wget https://data.qiime2.org/distro/core/qiime2-2019.4-py36-linux-conda.yml

#criando o ambiente do qiime

conda env create -n qiime2-2019.4 --file qiime2-2019.4-py36-linux-conda.yml

#criar o arquivo manifest-pe-dry.csv informando o nome que vc vai dar para as amostras, a localização delas no servidor ou no pc e quais são forward e reverse

#ativando o qiime2 no conda

source activate qiime2-2019.1

#ou no pc do diogo conda activate qiime2-2019.4

#estar na mesma pasta onde vc colocou o arquivo manifest, não precisa estar no mesmo lugar que as amostras

#dar imput nas amostras (paired ended manifest format - formato que sai do illumina R1 e R2)

qiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --input-path manifest-pe-dry.csv --output-path paired-end-demux.qza --input-format PairedEndFastqManifestPhred33

#cutting off primers (find and remove adapters in demultiplexed paired-end sequences - as sequencias são os primers usados no illumina)

qiime cutadapt trim-paired --i-demultiplexed-sequences paired-end-demux.qza --p-front-f TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG --p-front-r GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC --p-error-rate 0 --o-trimmed-sequences cutadapt.qza --p-cores 5

#visualizing

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

qiime tools view cutadapt.qzv

#joining pairs

#criar a pasta "joined"

qiime vsearch join-pairs --i-demultiplexed-seqs cutadapt.qza --output-dir /home/raphael/joined/

#no pc do diogo usar esse

#qiime vsearch join-pairs --i-demultiplexed-seqs cutadapt.qza --o-joined-sequences /home/rapha/joined/joined.qza

# q filter

#vc deve entrar na pasta que vc acabou de criar e rodar de lá

qiime quality-filter q-score-joined --i-demux joined\_sequences.qza --p-min-quality 20 --o-filtered-sequences filter.qza --o-filter-stats filter-stats.qza

#visualizing

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

#dereplicate (creates a feature table and feature representative sequences)

qiime vsearch dereplicate-sequences --i-sequences filter.qza --o-dereplicated-table table.qza --o-dereplicated-sequences seqs.qza

#clusterization (de novo method: cluster the features based on user-specified percent identity threshold of their sequences)

qiime vsearch cluster-features-de-novo --i-table table.qza --i-sequences seqs.qza --p-perc-identity 0.97 --o-clustered-table table-97.qza --o-clustered-sequences rep-seqs-97.qza

# removing chimeras (vsearch de novo method - pareia elas entre elas mesmas.Parametros em default)

uchime vsearch

qiime vsearch uchime-denovo --i-table table-97.qza --i-sequences rep-seqs-97.qza --output-dir uchime-dn-out

#exclude chimeras and borderlines chimeras

qiime feature-table filter-features --i-table table-97.qza --m-metadata-file uchime-dn-out/nonchimeras.qza --o-filtered-table uchime-dn-out/table-nonchimeric-wo-borderline.qza

qiime feature-table filter-seqs --i-data rep-seqs-97.qza --m-metadata-file uchime-dn-out/nonchimeras.qza --o-filtered-data uchime-dn-out/rep-seqs-nonchimeric-wo-borderline.qza

qiime feature-table summarize --i-table uchime-dn-out/table-nonchimeric-wo-borderline.qza --o-visualization uchime-dn-out/table-nonchimeric-wo-borderline.qzv

#taxonomy

qiime feature-classifier classify-sklearn --i-reads uchime-dn-out/rep-seqs-nonchimeric-wo-borderline.qza --i-classifier silva-132-99-nb-classifier.qza --p-reads-per-batch 100 --p-n-jobs 2 --o-classification taxonomy-dry.qza

qiime taxa barplot --i-table uchime-dn-out/table-nonchimeric-wo-borderline.qza --i-taxonomy taxonomy-dry.qza --m-metadata-file metadata.tsv --o-visualization barplot-dry.qzv

#rodar no mussismilia

# core-metric-phylogenetic

qiime phylogeny align-to-tree-mafft-fasttree --i-sequences uchime-dn-out/rep-seqs-nonchimeric-wo-borderline.qza --p-n-threads 5 --o-alignment aligned-sequence.qza --o-masked-alignment masked-aligned-sequence.qza --output-dir tree.qza

#rodar no mussismilia

#core phylogenetics

#valor de sampling depth vai variar de acordo com o tamano da amostra (ver em https://docs.qiime2.org/2018.2/tutorials/moving-pictures/#alpha-rarefaction-plotting na parte de "Alpha and beta diversity analysis¶")

qiime diversity core-metrics-phylogenetic --i-table uchime-dn-out/table-nonchimeric-wo-borderline.qza --i-phylogeny tree.qza/rooted\_tree.qza --p-sampling-depth 50000 --m-metadata-file metadata.tsv --p-n-jobs 2 --output-dir core-metrics-results-phylo

#FIM!