Ecological Genomics HW #1

Hannah Shafer

10/04/2021

Set up and activate the qiime2 package that I’ll use for my analysis, and make sure I’m in the right directory:

cd myresults  
conda activate qiime2-2021.8  
export TMPDIR="/data/project\_data/16S/tmptmpdir"  
echo $TMPDIR

Create the initial file, visualize the data, and create summary plots:

qiime tools import \  
 --type 'SampleData[PairedEndSequencesWithQuality]' \  
 --input-path /data/project\_data/16S/pyc\_manifest \  
 --input-format PairedEndFastqManifestPhred33V2 \  
 --output-path demux-paired-end\_full.qza  
qiime metadata tabulate \  
 --m-input-file pyc\_subset\_manifest \  
 --o-visualization tabulated-pyc\_sub-metadata.qzv  
qiime demux summarize \  
 --i-data demux-paired-end\_full.qza \  
 --o-visualization demux-pyc-sub.qzv

In a new terminal window not connected to the server, use secure file transfer to get .qzv files onto my computer:

scp hrshafer@pbio381.uvm.edu:~/myresults/\*.qzv

Deniose the data and create summary plots:

qiime dada2 denoise-paired \  
 --i-demultiplexed-seqs demux-paired-end.qza \  
 --p-n-threads 20 \  
 --p-trim-left-f 16 \  
 --p-trim-left-r 0 \  
 --p-trunc-len-f 289 \  
 --p-trunc-len-r 257 \  
 --o-table table.qza \  
 --o-representative-sequences rep-seqs.qza \  
 --o-denoising-stats denoising-stats.qza  
  
#looking at summary of entire dataset, and the quality scores for the forward & reverse reads of the sequences; the trim/truncate commands tell where to cut the data off on each side (left & right) and in each direction (forward & reverse), and the numbers were chosen based on when the quality score looked like they evened out (Callahan et al., 581)  
  
qiime feature-table summarize \  
 --i-table table.qza \  
 --o-visualization table.qzv \  
 --m-sample-metadata-file /data/project\_data/16S/pyc\_manifest  
qiime feature-table tabulate-seqs \  
 --i-data rep-seqs.qza \  
 --o-visualization rep-seqs.qzv  
qiime metadata tabulate \  
 --m-input-file denoising-stats.qza \  
 --o-visualization denoising-stats.qzv  
   
#based on the above DADA2 denoising command, we generate our summary tables of our denoised sequence data, which we can then use below to generate our phylogenetic tree (McDonald et al., 7)

Create a phylogenetic tree to eventually use for our diveristy calculations:

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  
  
#once we have this step completed, we can see how related the different taxa in the microbiome are, which will let us calculate our weighted unifrac diversity metrics (Price et al., e9490)  
  
qiime diversity core-metrics-phylogenetic \  
 --i-phylogeny rooted-tree.qza \  
 --i-table table.qza \  
 --p-sampling-depth 13000 \  
 --m-metadata-file /data/project\_data/16S/pyc\_manifest \  
 --output-dir core-metrics-results  
  
#here, we use a sampling depth of 13,000 - this number was chosen for us in the tutorial, and it was likely chosen to maximize the number of samples used in analysis while maintaining an acceptable quality score for the sequences; my investigation is into this number  
#the output of this command is several diversity metrics, including the weighted unifrac metric, which I use in my analysis

Download this file onto my computer to visualize it on view.qiime2.org:

scp hrshafer@pbio381.uvm.edu:~/myresults/core-metrics-results/weighted\_unifrac\_emperor.qzv

Complete the same function again but with the higher sampling depth, and transfer new file to my computer for visualization and comparison:

qiime diversity core-metrics-phylogenetic \  
 --i-phylogeny rooted-tree.qza \  
 --i-table table.qza \  
 --p-sampling-depth 60000 \  
 --m-metadata-file /data/project\_data/16S/pyc\_manifest \  
 --output-dir core-metrics-results2  
  
#changed sampling depth from 13,000 to 60,000, and also made sure to update the output to have a slightly altered name so the original file doesn't get overwritten  
  
scp hrshafer@pbio381.uvm.edu:~/myresults/core-metrics-results2/weighted\_unifrac\_emperor.qzv