**USEARCH in command line**

Start with sequences trimmed with cutadapt script. Sometimes, with the free 32-bit version of USEARCH, data files are too large to run these commands all at once. If this is the case, I break them up into several sub-folders for the first few steps of the process and then combine them into concatenated files for the actual clustering commands later.

Unzip the files

gunzip \*.gz

Merge forward and reverse sequences with a minimum merge length of 100. Output a merged sequence file.

usearch -fastq\_mergepairs \*\_R1\*.fastq -relabel @ -fastq\_minmergelen 100 -fastqout merged.fq

Filter this file with a maximum ee of 1.0, and output a fasta file of the filtered sequences

usearch -fastq\_filter merged.fq -fastq\_maxee 1.0 -fastaout filtered.fa

For large datasets, may need to run the fastx\_uniques command before creating one filtered file, but we can use the one filtered file to create the ASV by sample table later with the usearch\_global command, so we can create it at this step anyway.

cat filtered\_a.fa filtered\_b.fa … > filtered.fa

With large datasets, perform this fastx\_uniques command on each separate group of files and then concatenate afterward. This command looks through the filtered dataset and finds unique sequences, then ouputting them with their abundance (-sizeout command) indicated at the end of each

usearch -fastx\_uniques filtered.fa -fastaout uniques.fa -sizeout

If working with separate sub-folders, now is the time to concatenate these unique sets of sequences together for the clustering step

cat uniques\_a.fa uniques \_b.fa uniques \_c.fa uniques \_d.fa uniques\_e.fa > sep\_uniques.fa

The unoise3 command relies on abundance and will output different results depending on the order of sequences in the dataset, therefore, it is key to sort them by size prior to clustering

usearch -sortbysize sep\_uniques.fa -fastaout seqs\_sorted.fa -minsize 4

This is the clustering/denoising algorithm which then creates a fasta file of just the unique ASVs in our dataset

usearch -unoise3 seqs\_sorted.fa -tabbedout out.txt -zotus denoised.fa

Then we can create a table by sample of ASV abundances to use in subsequent analyses

usearch -usearch\_global filtered.fq -db denoised.fa -strand plus -id 0.97 -otutabout zotu\_table.txt