

ii) *mothur* Pipeline

Laura Mary Pugh

22/11/2019

The following shell script is the pipeline that was used to run the *mothur* portion of the analysis from the command line as a job on the HPC cluster.

```
#!/bin/bash
#$ -cwd
#$ -V
#$ -M lmpugh@dundee.ac.uk
#$ -m e
#$ -m a

cp /homes/lmpugh/cluster_home/raw_data/16S_samples/trimmed/files/silva.nr_v132.pcr.align $TMPDIR
cp /homes/lmpugh/cluster_home/raw_data/16S_samples/trimmed/files/silva.nr_v132.tax $TMPDIR
cp /homes/lmpugh/cluster_home/raw_data/16S_samples/trimmed/*.fastq.gz $TMPDIR

mothur "#set.dir(input=$TMPDIR/);
make.file(inputdir=$TMPDIR, type=gz, prefix=16S);
make.contigs(file=16S.files);
summary.seqs(fasta=current);
screen.seqs(fasta=current, group=16S.contigs.groups, summary=current, minlength=100,
    maxlength=550, maxhomop=10);
unique.seqs(fasta=current);
count.seqs(name=current, group=current);
summary.seqs(count=current);
align.seqs(candidate=16S.trim.contigs.good.unique.fasta, reference=silva.nr_v132.pcr.align);
summary.seqs(fasta=current, count=current);
;screen.seqs(fasta=current, count=current, summary=current, start=1, end=11549);
filter.seqs(fasta=current, trump=.)
;screen.seqs(fasta=current, count=current, maxambig=0);
unique.seqs(fasta=current, count=current);
pre.cluster(fasta=current, count=current, diffs=2);
chimera.vsearch(fasta=current, count=current, dereplicate=t);
remove.seqs(fasta=current, accnos=current, count=current);
summary.seqs(fasta=current, count=current);
classify.seqs(fasta=current, count=current, reference=silva.nr_v132.pcr.align,
    taxonomy=silva.nr_v132.tax, cutoff=80);
remove.lineage(fasta=current, count=current, taxonomy=current,
    taxon=Chloroplast-Mitochondria-unknown-Archaea-Eukaryota);
summary.tax(taxonomy=current, count=current);
cluster.split(fasta=current, count=current, taxonomy=current, splitmethod=classify, taxlevel=4,
    cutoff=0.03);
make.shared(list=current, count=current, label=0.03);
classify.otu(list=current, count=current, taxonomy=current, label=0.03);
rename.file(taxonomy=current, shared=current);
count.groups(shared=current)"

cp /$TMPDIR/16S.* /homes/lmpugh/cluster_home/raw_data/16S_samples/trimmed/files/
```