#May 3 - continue with cDNA data as in SOP exactly

#ie using trainset and silva.bacteria.fasta

# using start files from previous one where i aligned again v119

# max and min length 200 and 300 as per Erwin et al 2016

mothur > pcr.seqs(fasta=silva.bacteria.fasta, start=13862, end=23444, keepdots=F, processors=30)

Using 30 processors.

Output File Names:

silva.bacteria.pcr.fasta

It took 1 secs to screen 14956 sequences.

mothur > system(mv silva.bacteria.pcr.fasta silva.v4.fasta)

mothur > summary.seqs(fasta=silva.v4.fasta)

Using 30 processors.

Start End NBases Ambigs Polymer NumSeqs

Minimum: 1 8721 229 0 3 1

2.5%-tile: 8 9582 251 0 3 374

25%-tile: 8 9582 252 0 4 3740

Median: 8 9582 252 0 4 7479

75%-tile: 8 9582 252 0 5 11218

97.5%-tile: 8 9582 253 1 6 14583

Maximum: 9 9582 310 5 9 14956

Mean: 7.99719 9581.94 251.969 0.0520193 4.56646

# of Seqs: 14956

Output File Names:

silva.v4.summary

It took 0 secs to summarize 14956 sequences.

mothur > align.seqs(fasta=stability.trim.contigs.good.unique.fasta, reference=silva.v4.fasta, flip=T, processors=30)

Using 30 processors.

Reading in the silva.v4.fasta template sequences... DONE.

It took 3 to read 14956 sequences.

Aligning sequences from stability.trim.contigs.good.unique.fasta ...

Some of you sequences generated alignments that eliminated too many bases, a list is provided in stability.trim.contigs.good.unique.flip.accnos. If the reverse compliment proved to be better it was reported.

It took 683 secs to align 3797695 sequences.

Output File Names:

stability.trim.contigs.good.unique.align

stability.trim.contigs.good.unique.align.report

stability.trim.contigs.good.unique.flip.accnos

-------------------

# id and remove chimeras:

mothur > summary.seqs(fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.fasta, count=stability.trim.contigs.good.unique.good.filter.unique.precluster.count\_table, processors=30)

Using 30 processors.

Start End NBases Ambigs Polymer NumSeqs

Minimum: 1 724 226 0 3 1

2.5%-tile: 1 724 252 0 3 200366

25%-tile: 1 724 252 0 4 2003656

Median: 1 724 252 0 4 4007311

75%-tile: 1 724 252 0 5 6010966

97.5%-tile: 1 724 253 0 6 7814256

Maximum: 1 724 276 0 8 8014621

Mean: 1 724 252.048 0 4.45559

# of unique seqs: 2760617

total # of seqs: 8014621

Output File Names:

stability.trim.contigs.good.unique.good.filter.unique.precluster.summary

It took 69 secs to summarize 8014621 sequences.

mothur > chimera.uchime(fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.fasta, count=stability.trim.contigs.good.unique.good.filter.unique.precluster.count\_table, dereplicate=t)

Using 30 processors.

/home/femlab/Cam/silva SOP outputfile March 30/uchime file does not exist. Checking path...

Found uchime in your path, using /usr/bin/uchime

uchime by Robert C. Edgar

http://drive5.com/uchime

This code is donated to the public domain.

Checking sequences from stability.trim.contigs.good.unique.good.filter.unique.precluster.fasta ...

It took 1 secs to check 2 sequences from group D1Zeroplot1a.

It took 2147 secs to check 25786 sequences from group D1Highplot3b.

It took 2563 secs to check 29853 sequences from group D1Zeroplot2b.

It took 3292 secs to check 36011 sequences from group D1Highplot2b.

It took 3583 secs to check 37356 sequences from group D1Zeroplot2a.

It took 3754 secs to check 37323 sequences from group D2Zeroplot1a.

It took 3834 secs to check 38257 sequences from group D2Highplot3a.

It took 4223 secs to check 40437 sequences from group D1Highplot3a.

It took 5242 secs to check 49292 sequences from group D2Highplot1a.

It took 5496 secs to check 49238 sequences from group D1Zeroplot1b.

It took 5564 secs to check 47292 sequences from group D1Highplot4a.

It took 6216 secs to check 55094 sequences from group D1Highplot1b.

It took 6954 secs to check 57082 sequences from group D1Highplot4b.

It took 7237 secs to check 60674 sequences from group D3Zeroplot3a.

It took 7510 secs to check 62684 sequences from group D3Highplot1a.

It took 7791 secs to check 63348 sequences from group D3Zeroplot4a.

It took 7848 secs to check 65860 sequences from group D1Zeroplot3a.

It took 8542 secs to check 67256 sequences from group D1Highplot2a.

It took 8924 secs to check 67487 sequences from group D2Highplot2a.

It took 9172 secs to check 71648 sequences from group D2Zeroplot3a.

It took 9941 secs to check 75465 sequences from group D3Zeroplot2a.

It took 6566 secs to check 64651 sequences from group D2Highplot1b.

It took 11947 secs to check 85867 sequences from group D3Zeroplot1a.

It took 11954 secs to check 86200 sequences from group D2Zeroplot4a.

It took 4717 secs to check 54695 sequences from group D3Highplot1b.

It took 3401 secs to check 43725 sequences from group D2Highplot2b.

It took 12640 secs to check 88187 sequences from group D2Zeroplot2a.

It took 12723 secs to check 89834 sequences from group D3Highplot2a.

It took 5089 secs to check 59346 sequences from group D1Zeroplot3b.

It took 12988 secs to check 89726 sequences from group D3Highplot3a.

It took 9328 secs to check 75592 sequences from group D2Highplot3b.

It took 6590 secs to check 67929 sequences from group D3Zeroplot3b.

It took 13846 secs to check 93772 sequences from group D2Highplot4a.

It took 10383 secs to check 84029 sequences from group D2Zeroplot1b.

It took 14537 secs to check 101417 sequences from group D1Zeroplot4a.

It took 3855 secs to check 54622 sequences from group D3Zeroplot1b.

It took 6017 secs to check 64126 sequences from group D3Zeroplot2b.

It took 8637 secs to check 79300 sequences from group D3Zeroplot4b.

It took 16645 secs to check 109126 sequences from group D3Highplot4a.

It took 4785 secs to check 57648 sequences from group D2Zeroplot4b.

It took 4250 secs to check 59199 sequences from group D2Highplot4b.

It took 9788 secs to check 88867 sequences from group D2Zeroplot3b.

5279 here

It took 19604 secs to check 121188 sequences from group D1Highplot1a.

It took 6928 secs to check 80144 sequences from group D1Zeroplot4b.

It took 10328 secs to check 101137 sequences from group D3Highplot2b.

It took 10621 secs to check 100015 sequences from group D3Highplot3b.

It took 12688 secs to check 111307 sequences from group D2Zeroplot2b.

It took 14189 secs to check 127154 sequences from group D3Highplot4b.

Output File Names:

stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.count\_table

stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.chimeras

stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.accnos

mothur > remove.seqs(fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.fasta, accnos=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.accnos)

Removed 1988420 sequences from your fasta file.

Output File Names:

stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta

mothur > summary.seqs(fasta=current, count=current)

Using stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.count\_table as input file for the count parameter.

Using stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta as input file for the fasta parameter.

Using 30 processors.

Start End NBases Ambigs Polymer NumSeqs

Minimum: 1 724 226 0 3 1

2.5%-tile: 1 724 252 0 3 131856

25%-tile: 1 724 252 0 4 1318557

Median: 1 724 252 0 4 2637114

75%-tile: 1 724 252 0 5 3955671

97.5%-tile: 1 724 253 0 6 5142372

Maximum: 1 724 274 0 8 5274227

Mean: 1 724 252.051 0 4.44138

# of unique seqs: 772197

total # of seqs: 5274227

Output File Names:

stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.summary

---------- classify against mothur trainset------------------

mothur > classify.seqs(fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta, count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.count\_table, reference=trainset9\_032012.pds.fasta, taxonomy=trainset9\_032012.pds.tax, cutoff=80)

It took 513 secs to classify 772197 sequences.

It took 30 secs to create the summary file for 772197 sequences.

Output File Names:

stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.taxonomy

stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.tax.summary

--------------remove seqs from wrong lineages and unknowns----------------------

remove.lineage(fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta, count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.count\_table, taxonomy=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.taxonomy, taxon=Chloroplast-Mitochondria-unknown-Eukaryota)

stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy

stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta

stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.pick.count\_table

summary.seqs()

#Using 30 processors.

# Start End NBases Ambigs Polymer NumSeqs

#Minimum: 1 724 226 0 3 1

#2.5%-tile: 1 724 252 0 3 131732

#25%-tile: 1 724 252 0 4 1317315

#Median: 1 724 252 0 4 2634629

#75%-tile: 1 724 252 0 5 3951943

#97.5%-tile: 1 724 253 0 6 5137525

#Maximum: 1 724 274 0 8 5269256

#Mean: 1 724 252.051 0 4.44127

## of unique seqs: 771131

#total # of seqs: 5269256

-------------cluster----------------

cluster.split(fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta, count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.pick.count\_table, taxonomy=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy, splitmethod=classify, taxlevel=4, cutoff=0.15, processors=3)

#took 3 days

#Output File Names:

#stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.an.unique\_list.list

#MOnday 9 May

make.shared(list=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.an.unique\_list.list, count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.pick.count\_table, label=0.03)

#Output File Names:

#stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.an.unique\_list.shared

classify.otu(list=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.an.unique\_list.list, count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.uchime.pick.pick.count\_table, taxonomy=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy, label=0.03)

#Output File Names:

#stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.an.unique\_list.0.03.cons.taxonomy

#stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.an.unique\_list.0.03.cons.tax.summary

# rename files

system(mv stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.an.unique\_list.shared stability.an.shared)

system(mv stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.an.unique\_list.0.03.cons.taxonomy stability.an.cons.taxonomy)

# count how many seqs in each sample.

count.groups(shared=stability.an.shared)

# output:

D1Highplot1a contains 202032.

D1Highplot1b contains 57501.

D1Highplot2a contains 101444.

D1Highplot2b contains 54982.

D1Highplot3a contains 77943.

D1Highplot3b contains 55764.

D1Highplot4a contains 114519.

D1Highplot4b contains 126208.

D1Zeroplot1a contains 2.

D1Zeroplot1b contains 81390.

D1Zeroplot2a contains 60722.

D1Zeroplot2b contains 55459.

D1Zeroplot3a contains 102400.

D1Zeroplot3b contains 87169.

D1Zeroplot4a contains 120280.

D1Zeroplot4b contains 91851.

D2Highplot1a contains 74283.

D2Highplot1b contains 105300.

D2Highplot2a contains 117221.

D2Highplot2b contains 67330.

D2Highplot3a contains 66912.

D2Highplot3b contains 138886.

D2Highplot4a contains 117592.

D2Highplot4b contains 83163.

D2Zeroplot1a contains 78431.

D2Zeroplot1b contains 156563.

D2Zeroplot2a contains 157469.

D2Zeroplot2b contains 193654.

D2Zeroplot3a contains 106489.

D2Zeroplot3b contains 159562.

D2Zeroplot4a contains 138404.

D2Zeroplot4b contains 93478.

D3Highplot1a contains 101190.

D3Highplot1b contains 76071.

D3Highplot2a contains 157174.

D3Highplot2b contains 164944.

D3Highplot3a contains 124162.

D3Highplot3b contains 152775.

D3Highplot4a contains 157467.

D3Highplot4b contains 202575.

D3Zeroplot1a contains 136668.

D3Zeroplot1b contains 72287.

D3Zeroplot2a contains 111083.

D3Zeroplot2b contains 126709.

D3Zeroplot3a contains 97894.

D3Zeroplot3b contains 103583.

D3Zeroplot4a contains 122907.

D3Zeroplot4b contains 117364.

Total seqs: 5269256.

Output File Names:

stability.an.count.summary

# therefore lowest no seqs is D1Highplot2b - contains 54982.

## we will therefore subsample to this number

sub.sample(shared=stability.an.shared, size=54982)

# Output File Names:

stability.an.0.03.subsample.shared

# sample D1Zeroplot1a gets removed coz its wrong.

# Alpha Diversity

# get data for rarefaction curves (a measure of diversity but not richness)

rarefaction.single(shared=stability.an.shared, calc=sobs, freq=100)

# Output File Names:

stability.an.groups.rarefaction

# plot this in R using script in data analysis tools folder.

# next, get a table containing the number of sequences, the sample coverage, the number of observed OTUs, and the Inverse Simpson diversity estimate using the summary.single command

summary.single(shared=stability.an.shared, calc=nseqs-coverage-sobs-invsimpson, subsample=54982)

# Output File Names:

stability.an.groups.ave-std.summary

stability.an.groups.summary

# Beta diversity

# calculate similarity of the membership and structure found in the various samples

dist.shared(shared=stability.an.shared, calc=thetayc-jclass-braycurtis, subsample=54982)

# Output File Names:

stability.an.thetayc.0.03.lt.dist

stability.an.jclass.0.03.lt.dist

stability.an.braycurtis.0.03.lt.dist

stability.an.thetayc.0.03.lt.ave.dist

stability.an.thetayc.0.03.lt.std.dist

stability.an.jclass.0.03.lt.ave.dist

stability.an.jclass.0.03.lt.std.dist

stability.an.braycurtis.0.03.lt.ave.dist

stability.an.braycurtis.0.03.lt.std.dist

#make a tree to discribe similarity of samples to each other

## 1. first with theta yue&clayton calc

tree.shared(phylip=stability.an.thetayc.0.03.lt.ave.dist)

#Output File Names:

stability.an.thetayc.0.03.lt.ave.tre

# 2. Next with braycurtis calc

tree.shared(phylip=stability.an.braycurtis.0.03.lt.ave.dist)

# Output File Names:

stability.an.braycurtis.0.03.lt.ave.tre

#3. with Jaccard coeff

tree.shared(phylip=stability.an.jclass.0.03.lt.ave.dist)

# Output File Names:

stability.an.jclass.0.03.lt.ave.tre

# Testing to determine whether the clustering within the tree is statistically significant

# with parsimony on tayc tree

parsimony(tree=stability.an.thetayc.0.03.lt.ave.tre, group=lime.design, groups=all)

#output

Tree# Groups ParsScore ParsSig

1 high-zero 15 0.371

# Output File Names:

stability.an.thetayc.0.03.lt.ave.tre.parsimony

stability.an.thetayc.0.03.lt.ave.tre.psummary

parsimony(tree=stability.an.thetayc.0.03.lt.ave.tre, group=growthstage.design, groups=all)

# output

Tree# Groups ParsScore ParsSig

1 elongation-heading 7 0.022

1 elongation-ripening 5 <0.001

1 heading-ripening 5 <0.001

Output File Names:

stability.an.thetayc.0.03.lt.ave.tre.parsimony

stability.an.thetayc.0.03.lt.ave.tre.psummary

parsimony(tree=stability.an.braycurtis.0.03.lt.ave.tre, group=growthstage.design, groups=all)

parsimony(tree=stability.an.braycurtis.0.03.lt.ave.tre, group=growthstage2.design, groups=all)

Using 30 processors.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#

Comparing to random:|||||||||||||||||||||||||||||||||||||||||||||||||||

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Tree# Groups ParsScore ParsSig

1 elongation-heading 6 0.005

1 elongation-ripening 2 0.001

1 heading-ripening 2 0.001

Output File Names:

stability.an.braycurtis.0.03.lt.ave.tre.parsimony

stability.an.braycurtis.0.03.lt.ave.tre.psummary

# redid with growth stage as one of samples in wrong group

mothur> parsimony(tree=stability.an.thetayc.0.03.lt.ave.tre, group=growthstage3.design, groups=all)

Removing group: growthStage because all sequences have been removed.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#\*\*\*\*#

Comparing to random:||||||||||||||||||||||||||||||||||||||||||||||||||||

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Tree# Groups ParsScore ParsSig

1 elongation-heading 6 0.007

1 elongation-ripening 4 <0.001

1 heading-ripening 5 <0.001

Output File Names:

stability.an.thetayc.0.03.lt.ave.tre.parsimony

stability.an.thetayc.0.03.lt.ave.tre.psummary

--------- the following was done on little lab computer on 15/08/2016-----------

**nmds(phylip=stability.an.thetayc.0.03.lt.ave.dist)**

Number of dimensions: 2

Lowest stress : 0.207155

R-squared for configuration: 0.814574

Output File Names:

stability.an.thetayc.0.03.lt.ave.nmds.iters

stability.an.thetayc.0.03.lt.ave.nmds.stress

stability.an.thetayc.0.03.lt.ave.nmds.axes

**nmds(phylip=stability.an.thetayc.0.03.lt.ave.dist, mindim=3, maxdim=3)**

Number of dimensions: 3

Lowest stress : 0.110842

R-squared for configuration: 0.923162

Output File Names:

stability.an.thetayc.0.03.lt.ave.nmds.iters

stability.an.thetayc.0.03.lt.ave.nmds.stress

stability.an.thetayc.0.03.lt.ave.nmds.axes

**amova(phylip=stability.an.thetayc.0.03.lt.ave.dist, design=growth.design)**

elongation-heading-ripening Among Within Total

SS 1.46473 2.38644 3.85117

df 2 44 46

MS 0.732363 0.0542373

Fs: 13.5029

p-value: <0.001\*

elongation-heading Among Within Total

SS 0.295227 1.41142 1.70665

df 1 29 30

MS 0.295227 0.0486696

Fs: 6.06594

p-value: <0.001\*

elongation-ripening Among Within Total

SS 0.974667 1.52978 2.50445

df 1 29 30

MS 0.974667 0.052751

Fs: 18.4767

p-value: <0.001\*

heading-ripening Among Within Total

SS 0.921106 1.83168 2.75279

df 1 30 31

MS 0.921106 0.0610562

Fs: 15.0862

p-value: <0.001\*

Experiment-wise error rate: 0.05

Pair-wise error rate (Bonferroni): 0.0166667

If you have borderline P-values, you should try increasing the number of iterations

Output File Names:

stability.an.thetayc.0.03.lt.ave.amova

homova(phylip=stability.an.thetayc.0.03.lt.ave.dist, design=growth.design)

HOMOVA BValue P-value SSwithin/(Ni-1)\_values

elongation-heading-ripening 0.877926 1 0.0396255 0.0571108 0.0650015

Experiment-wise error rate: 0.05

If you have borderline P-values, you should try increasing the number of iterations

Output File Names:

stability.an.thetayc.0.03.lt.ave.homova

**corr.axes(axes=stability.an.thetayc.0.03.lt.ave.nmds.axes, shared=stability.an.0.03.subsample.shared, method=spearman, numaxes=3)**

Output File Names:

stability.an.0.03.subsample.spearman.corr.axes

**get.communitytype(shared=stability.an.0.03.subsample.shared)**

didn’t work

**indicator(shared=stability.an.0.03.subsample.shared, design=growth.design)**

 lefse(shared=stability.an.0.03.subsample.shared, **design=growth.design)**

dfgdf

-------------anosim

anosim(phylip= stability.an.thetayc.0.03.lt.ave.dist, design=growth.design)