When using, please cite:

Schloss, P.D., et al., Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol, 2009. 75(23):7537-41.

Distributed under the GNU General Public License

#Modified MiSeqSOP tutorial by Pankaj Singh (pankaj.singh@uwa.edu.au)

#this is a streamlined version of official MiSeqSOP tutorial which can be used for ITS as well with some changes

<https://mothur.org/wiki/miseq_sop/>

#Listing forward and reverse reads in order to combine them. It is a good step which also serves as a checkpoint to ensure you have matching names for forward and reverse reads

#Copy codes post mothur> . I have highlighted the required codes

mothur >

make.file(inputdir=., type=fastq, prefix=stability)

Setting input directories to:

C:\Users\00113286\OneDrive - UWA\Desktop\mothur\

Output File Names:

C:\Users\00113286\OneDrive - UWA\Desktop\mothur\stability.files

#Making contigs from reads

#Here all reads from input files are combined and converted into a single and simple FASTA file

mothur >

make.contigs(file=stability.files)

Using 32 processors.

>>>>> Processing file pair F3D0\_R1.fastq - F3D0\_R2.fastq (files 1 of 10) <<<<<

Making contigs...

Done.

It took 0 secs to assemble 7793 reads.

>>>>> Processing file pair F3D141\_R1.fastq - F3D141\_R2.fastq (files 2 of 10) <<<<<

Making contigs...

Done.

It took 0 secs to assemble 5958 reads.

>>>>> Processing file pair F3D1\_R1.fastq - F3D1\_R2.fastq (files 3 of 10) <<<<<

Making contigs...

Done.

It took 1 secs to assemble 5869 reads.

>>>>> Processing file pair F3D2\_R1.fastq - F3D2\_R2.fastq (files 4 of 10) <<<<<

Making contigs...

Done.

It took 0 secs to assemble 19620 reads.

>>>>> Processing file pair F3D3\_R1.fastq - F3D3\_R2.fastq (files 5 of 10) <<<<<

Making contigs...

Done.

It took 1 secs to assemble 6758 reads.

>>>>> Processing file pair F3D5\_R1.fastq - F3D5\_R2.fastq (files 6 of 10) <<<<<

Making contigs...

Done.

It took 0 secs to assemble 4448 reads.

>>>>> Processing file pair F3D6\_R1.fastq - F3D6\_R2.fastq (files 7 of 10) <<<<<

Making contigs...

Done.

It took 0 secs to assemble 7989 reads.

>>>>> Processing file pair F3D7\_R1.fastq - F3D7\_R2.fastq (files 8 of 10) <<<<<

Making contigs...

Done.

It took 0 secs to assemble 5129 reads.

>>>>> Processing file pair F3D8\_R1.fastq - F3D8\_R2.fastq (files 9 of 10) <<<<<

Making contigs...

Done.

It took 1 secs to assemble 5294 reads.

>>>>> Processing file pair F3D9\_R1.fastq - F3D9\_R2.fastq (files 10 of 10) <<<<<

Making contigs...

Done.

It took 0 secs to assemble 7070 reads.

Group count:

F3D0 7793

F3D1 5869

F3D141 5958

F3D2 19620

F3D3 6758

F3D5 4448

F3D6 7989

F3D7 5129

F3D8 5294

F3D9 7070

Total of all groups is 75928

It took 3 secs to process 75928 sequences.

Output File Names:

stability.trim.contigs.fasta #Your good quality sequences

stability.scrap.contigs.fasta # Low quality sequences

stability.contigs\_report #summary of this step

stability.contigs.count\_table # No. of sequences which occurred repeatedly

#extracting what information and sequence distribution you have in your good quality contigs

mothur >

summary.seqs(fasta=stability.trim.contigs.fasta, count=stability.contigs.count\_table)

Using 32 processors.

Start End NBases Ambigs Polymer NumSeqs

Minimum: 1 248 248 0 3 1

2.5%-tile: 1 252 252 0 4 1899

25%-tile: 1 252 252 0 4 18983

Median: 1 252 252 0 4 37965

75%-tile: 1 253 253 0 5 56947

97.5%-tile: 1 253 253 6 6 74030

Maximum: 1 502 502 246 240 75928

Mean: 1 252 252 0 4

# of unique seqs: 75928

total # of seqs: 75928

It took 0 secs to summarize 75928 sequences.

Output File Names:

stability.trim.contigs.summary

#Removing ambiguties and poylmers maxambig=0, sequence length and parameters will vary based on the quality of data, here we are selecting 253 as it our amplicon length was around 251. and 75 % of reads are 253 nucleotides long

#Read on your own regarding ambiguous nucleotide codes

<https://www.bioinformatics.org/sms/iupac.html>

mothur >

screen.seqs(fasta=stability.trim.contigs.fasta, count=stability.contigs.count\_table, maxlength=253, maxhomop=8, maxambig=0)

Using 32 processors.

It took 0 secs to screen 75928 sequences, removed 11937.

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Running command: remove.seqs(accnos=stability.trim.contigs.bad.accnos.temp, count=stability.contigs.count\_table)

Removed 11937 sequences from stability.contigs.count\_table.

Output File Names:

stability.contigs.pick.count\_table

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Output File Names:

stability.trim.contigs.good.fasta

stability.trim.contigs.bad.accnos

stability.contigs.good.count\_table

It took 1 secs to screen 75928 sequences.

# give you the list of current files which you are working on, computation you are using and directory as well

mothur >

get.current()

Current RAM usage: 11.6151 Gigabytes. Total Ram: 127.696 Gigabytes.

Current files saved by mothur:

accnos=stability.trim.contigs.bad.accnos

fasta=stability.trim.contigs.good.fasta

contigsreport=stability.contigs\_report

count=stability.contigs.good.count\_table

processors=32

summary=stability.trim.contigs.summary

file=C:\Users\00113286\OneDrive - UWA\Desktop\mothur\stability.files

Current input directories saved by mothur:

C:\Users\00113286\OneDrive - UWA\Desktop\mothur\

Current default directories saved by mothur:

C:\Users\00113286\OneDrive - UWA\Desktop\mothur\

Current working directory: C:\Users\00113286\OneDrive - UWA\Desktop\mothur\

Output File Names:

current\_files.summary

#Selecting unique sequences (relevant DADA2 step), here we only select the sequences which are unique , we avoid repetitive sequences

#It saves time and is more accurate way

mothur >

unique.seqs(fasta=current, count=current)

Using stability.contigs.good.count\_table as input file for the count parameter.

Using stability.trim.contigs.good.fasta as input file for the fasta parameter.

63991 8789

Output File Names:

stability.trim.contigs.good.unique.fasta

stability.trim.contigs.good.count\_table

# identify chimeric reads which originate due to sequencing errors, it has overlaps with two different regions or can have homology with more than one target in taxonomic database

mothur >

chimera.vsearch(fasta=current, count=current)

Using stability.trim.contigs.good.count\_table as input file for the count parameter.

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

Using 32 processors.

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

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Splitting by sample:

Using 32 processors.

Reducing processors to 10.

Selecting sequences for groups F3D7

Selecting sequences for groups F3D8

Selecting sequences for groups F3D5

Selecting sequences for groups F3D141

Selecting sequences for groups F3D1

Selecting sequences for groups F3D3

Selecting sequences for groups F3D0

Selecting sequences for groups F3D6

Selecting sequences for groups F3D9

Selecting sequences for groups F3D2

Selected 858 sequences from F3D7.

Selected 892 sequences from F3D5.

Selected 992 sequences from F3D8.

Selected 1096 sequences from F3D141.

Selected 1174 sequences from F3D1.

Selected 1499 sequences from F3D0.

Selected 1277 sequences from F3D9.

Selected 1058 sequences from F3D3.

Selected 1361 sequences from F3D6.

Selected 2705 sequences from F3D2.

It took 0 seconds to split the dataset by sample.

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Reducing processors to 10.

It took 1 secs to check 858 sequences from group F3D7.

It took 1 secs to check 892 sequences from group F3D5.

It took 1 secs to check 992 sequences from group F3D8.

It took 1 secs to check 1058 sequences from group F3D3.

It took 1 secs to check 1096 sequences from group F3D141.

It took 1 secs to check 1174 sequences from group F3D1.

It took 2 secs to check 1277 sequences from group F3D9.

It took 2 secs to check 1361 sequences from group F3D6.

It took 2 secs to check 1499 sequences from group F3D0.

It took 2 secs to check 2705 sequences from group F3D2.

It took 2 secs to check 12912 sequences.

It took 2 secs to check your sequences. 2124 chimeras were found.

The number of sequences checked may be larger than the number of unique sequences because some sequences are found in several samples.

Removing chimeras from your input files:

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Running command: remove.seqs(fasta=stability.trim.contigs.good.unique.fasta, accnos=stability.trim.contigs.good.unique.denovo.vsearch.accnos, count=stability.trim.contigs.good.count\_table)

Removed 2124 sequences from stability.trim.contigs.good.unique.fasta.

Removed 4616 sequences from stability.trim.contigs.good.count\_table.

Output File Names:

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

stability.trim.contigs.good.pick.count\_table

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Output File Names:

stability.trim.contigs.good.unique.denovo.vsearch.chimeras

stability.trim.contigs.good.unique.denovo.vsearch.accnos

stability.trim.contigs.good.denovo.vsearch.count\_table

stability.trim.contigs.good.unique.denovo.vsearch.fasta

#remove chimeric reads

mothur >

remove.seqs(fasta=stability.trim.contigs.good.unique.fasta, accnos=stability.trim.contigs.good.unique.denovo.vsearch.accnos, count=stability.trim.contigs.good.count\_table)

Removed 2124 sequences from stability.trim.contigs.good.unique.fasta.

Removed 4616 sequences from stability.trim.contigs.good.count\_table.

Output File Names:

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

stability.trim.contigs.good.pick.count\_table

# here you will train your Silva classifier before annotating your non-chimeric reads. This will save time and ensure better annotation

#start and end position are representative for V4 region of 16S rRNA it changes depending upon your target region during sequencing

#To know more about how to train your classifier according to your own primer <https://mothur.org/blog/2016/Customization-for-your-region/>

mothur >

pcr.seqs(fasta=silva.bacteria.fasta, start=11895, end=25318, keepdots=F)

Using 32 processors.

[NOTE]: no sequences were bad, removing silva.bacteria.bad.accnos

It took 2 secs to screen 14956 sequences.

Output File Names:

silva.bacteria.pcr.fasta

#rename your new classifier to silva.v4.fasta

mothur >

rename.file(input=silva.bacteria.pcr.fasta, new=silva.v4.fasta)

Current files saved by mothur:

accnos=stability.trim.contigs.good.unique.denovo.vsearch.accnos

fasta=silva.bacteria.pcr.fasta

contigsreport=stability.contigs\_report

count=stability.trim.contigs.good.pick.count\_table

processors=32

summary=stability.trim.contigs.summary

file=C:\Users\00113286\OneDrive - UWA\Desktop\mothur\stability.files

#this is where we annotate our non-chimeric reads

mothur >

classify.seqs(processors=12, fasta=stability.trim.contigs.good.unique.pick.fasta, count=stability.trim.contigs.good.pick.count\_table, reference=silva.v4.fasta, taxonomy=silva.bacteria.silva.tax)

Using 12 processors.

Generating search database... DONE.

It took 2 seconds generate search database.

Reading in the silva.bacteria.silva.tax taxonomy... DONE.

Calculating template taxonomy tree... DONE.

Calculating template probabilities... DONE.

It took 6 seconds get probabilities.

Classifying sequences from stability.trim.contigs.good.unique.pick.fasta ...

It took 2 secs to classify 6665 sequences.

It took 0 secs to create the summary file for 6665 sequences.

Output File Names:

stability.trim.contigs.good.unique.pick.silva.wang.taxonomy# this is your taxonomy file

stability.trim.contigs.good.unique.pick.silva.wang.tax.summary

# This step removes all of the unwanted annotations, remove everything apart from bacteria

mothur >

remove.lineage(fasta=stability.trim.contigs.good.unique.pick.fasta,count=stability.trim.contigs.good.pick.count\_table,taxonomy=stability.trim.contigs.good.unique.pick.silva.wang.taxonomy, taxon=Chloroplast-Mitochondria-unknown-Archaea-Eukaryota)

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Running command: remove.seqs(accnos=stability.trim.contigs.good.unique.pick.silva.wang.accnos, count=stability.trim.contigs.good.pick.count\_table, fasta=stability.trim.contigs.good.unique.pick.fasta)

Removed 13 sequences from stability.trim.contigs.good.unique.pick.fasta.

Removed 38 sequences from stability.trim.contigs.good.pick.count\_table.

Output File Names:

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

stability.trim.contigs.good.pick.pick.count\_table

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

Output File Names:

stability.trim.contigs.good.unique.pick.silva.wang.pick.taxonomy#

stability.trim.contigs.good.unique.pick.silva.wang.accnos

stability.trim.contigs.good.pick.pick.count\_table

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

#Assorting Amplicon Sequencing Variants: ASVs(For OTU: Operational Taxonomic Unit different method use count file and label=0.03( for more info go to MiSeq SOP protocol in mothur)

#Always better to work with ASVs as they are more accurate and very handy while comparing multiple datasets

#For further reading regarding OTU vs ASVs <https://www.nature.com/articles/ismej2017119>

mothur >

make.shared(count=current)

Using stability.trim.contigs.good.pick.pick.count\_table as input file for the count parameter.

ASV

Output File Names:

stability.trim.contigs.good.pick.pick.asv.list

stability.trim.contigs.good.pick.pick.asv.shared

# here taxonomic information for individual ASVs is generated from main taxonomic annotation we generated earlier

mothur >

classify.otu(list=current, count=current, taxonomy=current, label=ASV)

Using stability.trim.contigs.good.pick.pick.count\_table as input file for the count parameter.

Using stability.trim.contigs.good.pick.pick.asv.list as input file for the list parameter.

Using stability.trim.contigs.good.unique.pick.silva.wang.pick.taxonomy as input file for the taxonomy parameter.

ASV

Output File Names:

stability.trim.contigs.good.pick.pick.asv.ASV.cons.taxonomy : this is your main taxonomy file which you will use for stats

stability.trim.contigs.good.pick.pick.asv.ASV.cons.tax.summary

# This gives information regarding how many reads are present in each of samples (Mothur calls them groups)

mothur >

count.groups(shared=current)

Using stability.trim.contigs.good.pick.pick.asv.shared as input file for the shared parameter.

F3D0 contains 6050.

F3D1 contains 4479.

F3D141 contains 4562.

F3D2 contains 15475.

F3D3 contains 5134.

F3D5 contains 3405.

F3D6 contains 6350.

F3D7 contains 4008.

F3D8 contains 4216.

F3D9 contains 5658.

Size of smallest group: 3405.

Total seqs: 59337.

Output File Names:

stability.trim.contigs.good.pick.pick.asv.count.summary

# This step gives representative sequences for every ASV that you have generated. Comes in handly and can be used RefSeq for further analysis using Phyloseq

mothur >

get.oturep(fasta=current, count=current, list=current, sorted=bin, method=abundance)

Using stability.trim.contigs.good.pick.pick.count\_table as input file for the count parameter.

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

Using stability.trim.contigs.good.pick.pick.asv.list as input file for the list parameter.

You did not provide a label, using ASV.

ASV 6652

Output File Names:

stability.trim.contigs.good.pick.pick.asv.ASV.rep.count\_table

stability.trim.contigs.good.pick.pick.asv.ASV.rep.fasta

#Exporting to R for Phyloseq package

#Export stability.trim.contigs.good.pick.pick.asv.shared and rename to 16S\_Shared.shared

#Export stability.trim.contigs.good.pick.pick.asv.ASV.rep.fasta and rename to 16S.fasta

#Export stability.trim.contigs.good.pick.pick.asv.ASV.cons.taxonomy and rename to 16S\_Tax.taxonomy

#Create a metadata file in .csv format (Comma Separated Value) and export it to phyloseq as 16S\_Metadata.csv