1. Define working directory

cd Desktop\Host plant on life history and behavior\Fungal community analysis\Sequences\LSU

1. Demultiplex the FASTA files

I got data already demultiplexed: for each sample I got 1 file corresponding to the forward read and 1 file corresponding to the reverse read

1. Merging

usearch -fastq\_mergepairs 01\_S1\_L001\_R1\_001.fastq 02\_S13\_L001\_R1\_001.fastq 03\_S25\_L001\_R1\_001.fastq 04\_S37\_L001\_R1\_001.fastq 05\_S49\_L001\_R1\_001.fastq 06\_S61\_L001\_R1\_001.fastq 07\_S73\_L001\_R1\_001.fastq 08\_S85\_L001\_R1\_001.fastq 09\_S2\_L001\_R1\_001.fastq 10\_S14\_L001\_R1\_001.fastq 11\_S26\_L001\_R1\_001.fastq 12\_S38\_L001\_R1\_001.fastq 13\_S50\_L001\_R1\_001.fastq 14\_S62\_L001\_R1\_001.fastq 23\_S75\_L001\_R1\_001.fastq 24\_S87\_L001\_R1\_001.fastq 45\_S54\_L001\_R1\_001.fastq 46\_S66\_L001\_R1\_001.fastq negC-LSU2\_S184\_L001\_R1\_001.fastq posContr-LSU2\_S173\_L001\_R1\_001.fastq negContr\_S24\_L001\_R1\_001.fastq posC\_S12\_L001\_R1\_001.fastq -fastq\_minmergelen 200 -fastq\_maxdiffs 20 -relabel @ -fastqout HostPlantLSUmerged

#Merging outpout

413691 Pairs (413.7k)

385583 Merged (385.6k, 93.21%)

70316 Alignments with zero diffs (17.00%)

27063 Too many diffs (> 20) (6.54%)

0 Fwd tails Q <= 2 trimmed (0.00%)

6 Rev tails Q <= 2 trimmed (0.00%)

1022 No alignment found (0.25%)

0 Alignment too short (< 16) (0.00%)

23 Merged too short (< 200)

35 Staggered pairs (0.01%) merged & trimmed

205.63 Mean alignment length

396.37 Mean merged length

0.70 Mean fwd expected errors

2.98 Mean rev expected errors

0.13 Mean merged expected errors

1. Trimming primers

usearch -fastx\_truncate HostPlantLSUmerged -stripleft 17 -stripright 21 -fastqout HostPlantLSUtrimmed

#Trimming output

100.0% Processing, 0 (0.0%) too short

1. Filtering

usearch -fastq\_filter HostPlantLSUtrimmed -fastq\_maxee 1 -fastaout HostPlantLSUecoldriv\_reads

#Filtering output

FASTQ base 33 for file LSUtrimmed

100.0% Filtering, 98.1% passed

385583 Reads (385.6k)

7438 Discarded reads with expected errs > 1.00

378145 Filtered reads (378.1k, 98.1%)

1. Dereplicating

usearch -fastx\_uniques HostPlantLSUecoldriv\_reads -sizeout -fastaout HostPlantLSUuniques

# Dereplicating output

100.0% Reading LSUecoldriv\_reads

100.0% DF

378145 seqs, 128898 uniques, 112544 singletons (87.3%)

Min size 1, median 1, max 38558, avg 2.93

100.0% Writing LSUuniques

1. Removing singletons

usearch -sortbysize HostPlantLSUuniques -minsize 2 -fastaout HostPlantLSUno\_singletons

#Removing singletons output

100.0% Reading LSUuniques

Getting sizes

Sorting 16354 sequences

100.0% Writing output

1. Denoising

usearch -unoise3 HostPlantLSUno\_singletons -zotus HostPlantLSUzotus -tabbedout HostPlantLSUunoise3

#Denoising output

100.0% Reading LSUno\_singletons

0.0% 0 amplicons, 0 bad (size >= 38558)

WARNING: Shifted sequences detected

100.0% 437 amplicons, 86279 bad (size >= 8)

100.0% 105 good, 332 chimeras

100.0% Writing zotus

1. Contructing zOTUs table

usearch -usearch\_global HostPlantLSUecoldriv\_reads -db HostPlantLSUzotus -strand both -id 0.97 -otutabout HostPlantLSUzotus\_table -mapout HostPlantLSUzmap

#Constructiong ZOTUs table output

100.0% Reading LSUzotus

100.0% Masking (fastnucleo)

100.0% Word stats

100.0% Alloc rows

100.0% Build index

100.0% Searching, 96.8% matched

366210 / 378145 mapped to OTUs (96.8%)

Writing LocalAdaptLSUzotus\_table

Writing LocalAdaptLSUzotus\_table ...done

1. Assign taxonomy

##Assign to JD’s database

usearch -usearch\_global HostPlantLSUzotus -db ref\_LSU\_Sequences.fasta -id 0.97 -strand both -otutabout HostPlantLSUmatchJD -fastapairs HostPlantLSUzotus\_JD -notmatched HostPlantLSUzotus\_JD\_notmatched

100.0% Reading ref\_LSU\_Sequences.fasta

100.0% Masking (fastnucleo)

100.0% Word stats

100.0% Alloc rows

100.0% Build index

100.0% Searching, 34.3% matched

36 / 105 mapped to OTUs (34.3%)

Writing LocalAdaptLSUmatchJD

Writing LocalAdaptLSUmatchJD ...done

##Assign to NCBI database 99% identity

usearch -usearch\_global HostPlantLSUzotus\_JD\_notmatched -db sequences\_NCBI\_tax.fa -id 0.99 -strand both -otutabout HostPlantLSUmatchNCBI -fastapairs HostPlantLSUzotus\_NCBI -notmatched HostPlantLSUzotus\_NCBI\_notmatched

#Output

100.0% Reading sequences\_NCBI\_tax.fa

100.0% Masking (fastnucleo)

100.0% Word stats

100.0% Alloc rows

100.0% Build index

100.0% Searching, 40.6% matched

28 / 69 mapped to OTUs (40.6%)

Writing LocalAdaptLSUmatchNCBI

Writing LocalAdaptLSUmatchNCBI ...done

##Assign taxonomy

usearch -sintax HostPlantLSUzotus -db ref\_LSU\_Sequences.fasta -strand both -sintax\_cutoff 0.8 -tabbedout HostPlantLSUtax\_JD

usearch -sintax HostPlantLSUzotus\_JD\_notmatched -db sequences\_NCBI\_tax.fa -strand both -sintax\_cutoff 0.8 -tabbedout HostPlantLSUtax\_endofline