**Creation of Coral Ribosomal DNA database**

**Create files for clustering**

awk '/^>/ { p = ($0 ~ /Metazoa/)} p' ROD1.fasta > Metazoa.fasta **- 2677 (number of sequences)**

awk '/^>/ { p = ($0 ~ /Streptophyta/)} p' ROD1.fasta > Streptophyta.fasta -**1897**

awk '/^>/ { p = ($0 ~ /Ascomycota/)} p' ROD1.fasta > Ascomycota.fasta -**5670**

awk '/^>/ { p = ($0 ~ /Basidiomycota/)} p' ROD1.fasta > Basidiomycota.fasta **-772**

**Remove extracted sequences from main file**

awk -vRS=">" -vORS="" -vFS="\n" -vOFS="\n" ' NR>1 && $1!~/Metazoa/ {print ">"$0}' ROD1.fasta > nom.fasta **- 9190**

awk -vRS=">" -vORS="" -vFS="\n" -vOFS="\n" ' NR>1 && $1!~/Streptophyta/ {print ">"$0}' nom.fasta > nom2.fasta **- 7293**

awk -vRS=">" -vORS="" -vFS="\n" -vOFS="\n" ' NR>1 && $1!~/Ascomycota/ {print ">"$0}' nom2.fasta > nom3.fasta **- 1623**

awk -vRS=">" -vORS="" -vFS="\n" -vOFS="\n" ' NR>1 && $1!~/Basidiomycota/ {print ">"$0}' nom3.fasta > nom4.fasta **-851**

**Metazoa:**

Eurotatoria - 33 results:

awk '/^>/ { p = ($0 ~ /Eurotatoria/)} p' Metazoa.fasta > Eutoria.fasta  **33**

cat nom4.fasta Eutoria.fasta > nom5.fasta **884**

Cnidaria – 48 results

awk '/^>/ { p = ($0 ~ /Cnidaria/)} p' Metazoa.fasta > Cnidaria.fasta **48**

cat nom5.fasta Cnidaria.fasta> nom6.fasta **932**

Remove Eurotatoria and Cnidaria from Metazoa file:

awk -vRS=">" -vORS="" -vFS="\n" -vOFS="\n" ' NR>1 && $1!~/Eurotatoria/ {print ">"$0}' Metazoa.fasta > Metazoa1.fasta **2644**

awk -vRS=">" -vORS="" -vFS="\n" -vOFS="\n" ' NR>1 && $1!~/Cnidaria/ {print ">"$0}' Metazoa1.fasta > Metazoa2.fasta **2596**

Cluster the four fasta files

vsearch -sortbylength Metazoa2.fasta -output cluster/Metazoa.sorted.fasta -minseqlength 64

vsearch -sortbylength Ascomycota.fasta -output cluster/Ascomycota.sorted.fasta -minseqlength 64

vsearch -sortbylength Basidiomycota.fasta -output cluster/Basidiomycota.sorted.fasta -minseqlength 64

vsearch -sortbylength Streptophyta.fasta -output cluster/Streptophyta.sorted.fasta -minseqlength 64

**Cluster at 90%**

vsearch -cluster\_smallmem Metazoa.sorted.fasta -id 0.90 -centroids clustered/Metazoa.clustered.fasta -uc UC/Metazoa.clustered.uc

vsearch -cluster\_smallmem Ascomycota.sorted.fasta -id 0.90 -centroids clustered/Ascomycota.clustered.fasta -uc UC/Ascomycota.clustered.uc

vsearch -cluster\_smallmem Basidiomycota.sorted.fasta -id 0.90 -centroids clustered/Basidiomycota.clustered.fasta -uc UC/Basidiomycota.clustered.uc

vsearch -cluster\_smallmem Streptophyta.sorted.fasta -id 0.90 -centroids clustered/Streptophyta.clustered.fasta -uc UC/Streptophyta.clustered.uc

**COMBINE CLUSTERED FILES WITH MAIN FASTA FILE**

cat Ascomycota.clustered.fasta Metazoa.clustered.fasta Streptophyta.clustered.fasta Basidiomycota.clustered.fasta > clustered.fasta **1454**

cat nom6.fasta cluster/clustered/clustered.fasta > clustered.concatenated.fasta - **2386**

**align using maaft**

mafft --auto clustered.concatenated.fasta > clustered.aligned.fasta

**Create tree using fasttree**

fasttree -gtr -nt clustered.aligned.fa > clustered.fastree.tre

**Create tree using RAXML**

raxmlHPC-PTHREADS-SSE3 -s trimmed.fasta -n raxml\_trees -m GTRCAT -f a -N 100 -x 123 -p 256 -T 4