

```
python ../../phylogenomic_dataset_construction/cut_seq_ends.py all.fa all.rawblast
python ../../phylogenomic_dataset_construction/blast_to_mcl.py all.rawblast 0.4
mcl all.rawblast.hit-frac0.4.minusLogEvaluate --abc -te 5 -tf 'gq(5)' -l 1.4 -o hit-frac0.4_l1.4_e5
python ../../phylogenomic_dataset_construction/write_fasta_files_from_mcl.py all.fa.cut
hit-frac0.4_l1.4_e5 4 ../clusters/
```

Clusters larger than 2000 used fasttree

All others used fasta_to_tree.py

-----Folder Check

```
for x in Folder*;do echo $x; ls $x | grep -c "tre"; ls $x | grep -c "fa$";done
```

```
python ../../phylogenomic_dataset_construction/trim_tips.py Round1/ .tre 10 2
python ../../phylogenomic_dataset_construction/mask_tips_by_taxonID_transcripts.py Round1/
Round1/
python ../../phylogenomic_dataset_construction/cut_long_internal_branches.py Round1 .mm 2 4
Round1Cut/
python ../../phylogenomic_dataset_construction/write_fasta_files_from_trees.py Blast/all.fa.cut
Round1Cut/ .mm Round1Fastas/
python ../../phylogenomic_dataset_construction/write_fasta_files_from_trees.py Blast/all.fa.cut
Round1Cut/ .tre Round1Fastas/
```

Round 2

##Same as round 1

Move to separate folders for x in {0..7};do mv cluster\$x_rr.fa ../Folder0/;done

Move to separate folders for x in {0..7};do mv cluster\$x\rr.fa ../Folder0/;done

First 7 clusters use fasttree

```
python ../../phylogenomic_dataset_construction/mafft_wrapper.py Folder0/ .fa 1 aa && python
../../phylogenomic_dataset_construction/phyutility_wrapper.py Folder0/ 0.1 aa && python
../../phylogenomic_dataset_construction/fasttree_wrapper.py Folder0 aa
python ../../phylogenomic_dataset_construction/fasta_to_tree.py Folder1 4 aa n
```

```
python ../../phylogenomic_dataset_construction/trim_tips.py Round2/ .tre 10 2
python ../../phylogenomic_dataset_construction/mask_tips_by_taxonID_transcripts.py Round2/
Round2/
```

```
python ../../phylogenomic_dataset_construction/cut_long_internal_branches.py Round2 .mm 2 4  
Round2Cut/
```

-----Change the tre files created from cutting before to mm for easier processing downstream

```
for x in *.tre;do mv $x $x.mm;done
```

Prepare SH test

```
for x in *.tre;do cp $x $x.mm;done
```

Remove the fastas

```
python ../../phylogenomic_dataset_construction/write_fasta_files_from_trees.py Blast/all.fa.cut  
HomologTrees/ .mm Round2Fastas/
```

Move to folders

```
for x in {0..11}; do for file in $(ls -p | grep -v / | head -878);do mv $file Folder$x/;done; done
```

Align with Mafft

```
python ../../phylogenomic_dataset_construction/mafft_wrapper.py Folder0/ .fa 3 aa
```

Rename the trees for the SHtest

```
for x in *.fa; do raxml -T 9 -f J -p 12345 -m PROTCATWAG -s $x -t $x.tre -n $x;done
```

Species Tree Inference

```
python ../../phylogenomic_dataset_construction/filter_1to1_orthologs.py ../HomologTrees/ .tre  
13 1to1_NoMissing/
```

```
python ../../phylogenomic_dataset_construction/write_ortholog_fasta_files.py  
../Blast/all.fa.cut Trees/ Fastas/ 13
```

Align with Prank

```
for x in {0..8}; do for file in $(ls -p | grep -v / | head -200);do mv $file Folder$x/;done; done
```

```
python ../../phylogenomic_dataset_construction/prank_wrapper.py Folder0/ Folder0/ .fa aa
```

```
python ../../phylogenomic_dataset_construction/phyutility_wrapper.py Fastas/ 0.3 aa
```

```
python ../../../../phylogenomic_dataset_construction/concatenate_matrices.py Fastas/ 150 13 aa  
SuperMatrix
```

```
raxml -T 9 -f a -x 112233 -# 200 -p 12345 -m PROTCATWAG -q SuperMatrix.model -s  
SuperMatrix.phy -n 1to1_NoMiss
```

```
raxml -T 9 -f J -p 12345 -m PROTCATWAG -s SuperMatrix.phy -t  
RAxML_bestTree.1to1_NoMiss -q SuperMatrix.model -n SHtest
```

For Phyparts Analysis

Relabeled from RAxML output using regex to remove everything before the first period

```
python ../../yangya-phylogenomic_dataset_construction/filter_1to1_orthologs.py HomologTrees/  
.labeled 13 1to1/
```

```
python ../RT/scripts/prune_paralogs_RT.py HomologTrees/ .labeled RT/ 11 INOUT.txt
```

```
for x in *.renamed; do pxrr -t $x -g Beta,Spol -o $x.rr;done
```

```
for x in *.rr;do perl ../../Programs/AddNameToNewick.pl ../../Programs/taxon_table $x >  
$x.full_name;done
```

```
java -jar  
../../WithWOBD/Phyparts_analyses/phyparts/target/phyparts-0.0.1-SNAPSHOT-jar-with-depend  
encies.jar -a 1 -d 1to1/Full_Name/ -m BootstrapTree.tre -s 80 -v
```

```
head out.concon.tre -n 1 > Concordant.tre
```

```
head out.concon.tre -n 2 | tail -n 1 > Conflict.tre
```

Astral Analysis

```
java -jar ../../../../Programs/ASTRAL/Astral/astral.4.10.8.jar -i Astral_in.tre -o Astral_out.tre
```

CDS analysis

```
cat *.fa > all.fa
```

```
python ../../../../phylogenomic_dataset_construction/write_ortholog_fasta_files.py Data/all.fa  
../../../../TreeBuilding_NoWOBD/SpeciesTreeInference/1to1_NoMissing/Trees/ Fastas/ 13
```

```
cp ../../../../TreeBuilding_NoWOBD/SpeciesTreeInference/1to1_NoMissing/Fastas/*.aln .
```

```
for x in *.fa;do ../../../../Phyx/phyx/src/pxaatocdn -a $x.aln -n $x -o $x.codon;done
```

```
for x in *.codon;do mv $x $x.aln;done
```

```
python ../../../../phylogenomic_dataset_construction/phyutility_wrapper.py CodonAligned/ 0.3 dna
```

```
python ../../../../phylogenomic_dataset_construction/concatenate_matrices.py CodonAligned/ 150  
13 dna Supermatrix
```

```
raxml -T 9 -f a -x 112233 -# 200 -p 12345 -m GTRCAT -q Supermatrix.model -s  
Supermatrix.phy -n 1to1_CDS
```

```
raxml -T 9 -f J -p 12345 -m GTRCAT -s Supermatrix.phy -t RAXML_bestTree.1to1_CDS -q  
Supermatrix.model -n SHtest
```

CDS clustering

```
for x in *.fa;do ../../../../apps/TransDecoder/transdecoder_plugins/cdhit/cd-hit-est -i $x -o  
$x.cdhitest -c 0.99 -n 10 -r 0 -T 9;done  
../../../../apps/ncbi-blast-2.3.0+/bin/makeblastdb -in all.fa -parse_seqids -dbtype nucl -out all.fa
```

```
../../../../apps/ncbi-blast-2.3.0+/bin/blastn -db all.fa -query all.fa -evaluate 10 -num_threads 9  
-max_target_seqs 1000 -out all.rawblast -outfmt '6 qseqid qlen sseqid slen frames pident nident  
length mismatch gapopen qstart qend sstart send eval evalue bitscore'
```

```
python ../../../../phylogenomic_dataset_construction/blast_to_mcl.py all.rawblast 0.4
```

```
mcl all.rawblast.hit-frac0.4.minusLogEvalue --abc -te 5 -tf 'gq(5)' -l 1.4 -o hit-frac0.4_l1.4_e5
```

```
python ../../phylogenomic_dataset_construction/write_fasta_files_from_mcl.py all.fa.cut  
hit-frac0.4_l1.4_e5 4 ../clusters/
```

```
python ../../phylogenomic_dataset_construction/fasta_to_tree.py Folder0/ 4 dna n
```

```
python ../../phylogenomic_dataset_construction/trim_tips.py Round1/ .tre 10 2
```

```
python ../../phylogenomic_dataset_construction/mask_tips_by_taxonID_transcripts.py Round1/  
Round1
```

```
python ../../phylogenomic_dataset_construction/cut_long_internal_branches.py Round1 .mm 2 4  
Round1Cut/
```

```
python ../../phylogenomic_dataset_construction/write_fasta_files_from_trees.py Blast/all.fa.cut  
Round1Cut/ .mm Round1Fastas/
```

```
python ../../phylogenomic_dataset_construction/write_fasta_files_from_trees.py Blast/all.fa.cut  
Round1Cut/ .tre Round1Fastas/
```

```
mv Folder3/cluster8545rr.fa* .
```

Rerun above ^

```
python ../../phylogenomic_dataset_construction/write_fasta_files_from_trees.py Blast/all.fa.cut  
HomologTrees/ .mm Round2Fastas/
```

Prepare SH test

```
for x in *.tre;do cp $x $x.mm;done
```

Remove the fastas

```
python ../../phylogenomic_dataset_construction/write_fasta_files_from_trees.py Blast/all.fa.cut  
HomologTrees/ .mm Round2Fastas/
```

Move to folders

```
for x in {0..11}; do for file in $(ls -p | grep -v / | head -898);do mv $file Folder$x/;done; done
```

Align with Mafft

```
python ../../../../phylogenomic_dataset_construction/mafft_wrapper.py Folder0/ .fa 3 dna
```

Rename the trees for the SHtest

```
for x in *.fa; do raxml -T 9 -f J -p 12345 -m GTRCAT -s $x -t $x.tre -n $x;done
```

Species Tree Inference

```
python ../../../../phylogenomic_dataset_construction/filter_1to1_orthologs.py ../HomologTrees/  
.mm 13 Trees/
```

```
python ../../../../phylogenomic_dataset_construction/write_ortholog_fasta_files.py ../Blast/all.fa.cut  
Trees/ Fastas/ 13
```

```
python ../../../../phylogenomic_dataset_construction/prank_wrapper.py Folder0/ Folder0/ .fa dna
```

```
python ../../../../phylogenomic_dataset_construction/phyutility_wrapper.py Fastas/ 0.3 dna
```

```
python ../../../../phylogenomic_dataset_construction/concatenate_matrices.py Fastas/ 150 13 dna  
Supermatrix
```

```
raxml -T 9 -f a -x 112233 -# 200 -p 12345 -m GTRCAT -q Supermatrix.model -s  
Supermatrix.phy -n 1to1_CDS
```

```
raxml -T 9 -p 12345 -m GTRCAT -q Supermatrix.model -s Supermatrix.phy -n JustML
```

Phyparts analysis

```
for x in *.fa;do perl ../../../../Programs/RemoveRAxML.pl $x;done
```

```
python ../../../../yangya-phylogenomic_dataset_construction/filter_1to1_orthologs.py  
RAxMLRemoved/ .fa 13 1to1/
```

```
for x in *.tre;do python ../../../../Programs/move_notes_labels_trees_raxml.py $x >  
$x.moved;done
```

```
for x in *.moved;do perl ../../../../Programs/RemoveExtra.pl $x > $x.trim;done
```

```
for x in *.trim; do pxrr -t $x -g Beta,Spol -o $x.rr;done
```

```
for x in *.rr;do perl ../../../../Programs/AddNameToNewick.pl ../../../../Programs/taxon_table $x  
> $x.full_name;done
```

```
pxrr -t RAXML_bipartitions.1to1_CDS -g Beta,Spol -o Rerooted_tree.tre
```

```
perl ../../../../Programs/AddNameToNewick.pl ../../../../Programs/taxon_table Rerooted_tree.tre >  
Renamed_SpeciesTree.tre
```

```
java -jar  
../../../../WithWOBD/Phyparts_analyses/phyparts/target/phyparts-0.0.1-SNAPSHOT-jar-with-dep  
endencies.jar -a 1 -d FullName/ -m SpeciesTree/Renamed_SpeciesTree.tre -s 80 -v
```

```
head out.concon.tre -n 1 > Concordant.tre
```

```
head out.concon.tre -n 2 | tail -n 1 > Conflict.tre
```

Bug in pxrr redoing conflict analysis with bug fix

```
for x in *.trim;do pxrr -t $x -g Beta,Spol -s -o $x.rr;done
```

```
for x in *.rr;do perl ../../../../Programs/AddNameToNewick.pl ../../../../Programs/taxon_table $x  
> $x.full_name;done
```

```
java -jar  
../../../../WithWOBD/Phyparts_analyses/phyparts/target/phyparts-0.0.1-SNAPSHOT-jar-with-depen  
dencies.jar -a 1 -d Full_Name/ -m SpeciesTreeInference/Reroot_Renamed.tre -s 80 -v
```

Duplication Analysis

```
for x in *.fa;do python ../../../../Programs/move_notes_labels_trees_raxml.py $x >$x.labels;done
```

```
for x in *.labels;do perl ../../../../Programs/RemoveExtra.pl $x > $x.Renamed;done
```

```
java -jar
../../WithWOBD/Phyparts_analyses/phyparts/target/phyparts-0.0.1-SNAPSHOT-jar-with-dependencies.jar -a 2 -d HomologTreesForDuplication/ -m SpeciesTreeInference/Reroot.tre -s 80 -v
```

###Without mapping tree

```
java -jar ../../ModifiedPhyparts/phyparts-0.0.1-SNAPSHOT-jar-with-dependencies(2\).jar -a 2 -d HomologTreesForDuplication/ -s 80 > NoMapping.txt
```

```
perl GeneJoin.pl NoMapping.txt
```

```
java -jar
../../WithWOBD/Phyparts_analyses/phyparts/target/phyparts-0.0.1-SNAPSHOT-jar-with-dependencies.jar -a 2 -d ../Alltaxa/1to1/HomologTreesForDuplication/ -m GeneJoinStuff/Rerooted_PresenceAbsence.tre -s 80 -v
```

SiteLogLikelihoodTest

```
raxml -f g -T 9 -s SuperMatrix.phy.reduced -q SuperMatrix.model.reduced -m PROTGAMEAWAG -z 2trees.tre -n AA_SLL
```

Tree 0 is drosophyllum early

Tree 1 is sister to Nepenthes

MultiSpecies Ks

```
cdhit -i NepSFB.pep.fa -o NepSFB.pep.fa.cdhit -c 0.99 -n 5 -T 2
```

```
cdhit -i MJM2940.pep.fa -o "MJM2940.pep.fa.cdhit" -c 0.99 -n 5 -T 2
```

```
makeblastdb -in MJM2940.pep.fa.cdhit -out MJM2940.db -dbtype="prot"
```



```
blastp -db MJM2940.db -query NepSFB.pep.fa.cdhit -evaluate 10 -num_threads 15  
-max_target_seqs 20 -out NepSFB.rawblastp -outfmt '6 qseqid qlen sseqid slen frames pident  
nident length mismatch gapopen qstart qend sstart send evaluate bitscore'
```

```
cat MJM2940.pep.fa.cdhit NepSFB.pep.fa.cdhit > For_Program.pep.fa  
cat MJM2940.cds.fa NepSFB.cds.fa > For_Program.cds.fa
```

```
perl remove_top_blast.pl NepSFB.rawblastp For_Program.pep.fa For_Program.cds.fa
```

```
blastp -db DrobinSFB.db -query NepSFB.pep.fa.cdhit -evaluate 10 -num_threads 15  
-max_target_seqs 20 -out NepSFB_Drobin.rawblastp -outfmt '6 qseqid qlen sseqid slen frames  
pident nident length mismatch gapopen qstart qend sstart send evaluate bitscore' && blastp -db  
DrobinSFB.db -query MJM2940.pep.fa.cdhit -evaluate 10 -num_threads 15 -max_target_seqs 20  
-out MJM2940_Drobin.rawblastp -outfmt '6 qseqid qlen sseqid slen frames pident nident length  
mismatch gapopen qstart qend sstart send evaluate bitscore' && blastp -db MJM3360.db -query  
RuprSFB.pep.fa.cdhit -evaluate 10 -num_threads 15 -max_target_seqs 20 -out  
Rupr_MJM3360.rawblastp -outfmt '6 qseqid qlen sseqid slen frames pident nident length  
mismatch gapopen qstart qend sstart send evaluate bitscore'
```

```
perl remove_top_blast.pl NepSFB.rawblastp For_Program.pep.fa For_Program.cds.fa
```

Ks_DrosPosition

Conflict rerun with AA on CDS tree

```
for x in ../Trees/*;do echo $x > out.txt; pxbp -t $x > out.txt;done
```

Wrote Program GeneHybridSplitter

Conditions of sister these bipartitions must exist

Ancistrocladus_robertsonianum,Drosophyllaceae_Drosophyllum_lusitanicum,Nepenthes_ampul
laria,Nepenthaceae_Nepenthes_alata_SFB

Ancistrocladus_robertsonianum,Drosophyllaceae_Drosophyllum_lusitanicum

Droseraceae_Aldrovanda_vesiculosa,Droseraceae_Drosera_binata,Dionaea_muscipula,Ancistr
ocladus_robertsonianum,Drosophyllaceae_Drosophyllum_lusitanicum,Nepenthes_ampullaria,N
epenthaceae_Nepenthes_alata_SFB

Conditions of Basal these bipartitions must exist

Droseraceae_Aldrovanda_vesiculosa,Droseraceae_Drosera_binata,Dionaea_muscipula,Nepenthes_ampullaria,Nepenthaceae_Nepenthes_alata_SFB
Ancistrocladus_robertsonianum,Drosophyllaceae_Drosophyllum_lusitanicum
Droseraceae_Aldrovanda_vesiculosa,Droseraceae_Drosera_binata,Dionaea_muscipula,Ancistrocladus_robertsonianum,Drosophyllaceae_Drosophyllum_lusitanicum,Nepenthes_ampullaria,Nepenthaceae_Nepenthes_alata_SFB

Wrote files from trees with those conditions

```
python ../yangya-phylogenomic_dataset_construction/write_ortholog_fasta_files.py Protein.fa  
TreesBasal/ FastaBasal/ 13
```

Change names for forloop

```
for x in *.fa;do cp $x $x.cds.pep;done
```

Take out pairs pep

```
for x in *.pep;do perl RemovePair.pl $x >$x.pairs;done
```

Take out pairs cds, I know this is a dumb way

```
for x in *.cds;do perl RemovePair.pl $x >$x.pep;done
```

Align All

```
for x in *.pairs;do mafft --auto --genafpair --maxiterate 1000 --amino $x > $x.aln;done
```

Codon Align

```
for x in *.pep;do pxaatocdn -a $x.pairs.aln -n $x -o $x.codon;done
```

Move the Codon to a folder called Codon

Get Ks

```
for x in Codon/*;do perl Ks_test.pl $x;done
```