

Code used:

species list

```
$ perl -ne 'if(/>.*_(.*?)\\/) {print $1."\\n"}' RAPTORB_ENSEMBL_gene_tree.fa > sp_list.RAPTORB_ENSEMBL_gene_tree.fa
$ perl -ne 'if(/(Org_.*?)\\s/) {print $1."\\n"}' RAPTORB_phytozome_top100.fasta > sp_list.RAPTORB_phytozome_top100.fasta
$ gawk '{ if (match($0,/\\[(.*?)\\]/,m)) print m[0] }' RAPTORB_NCBI_BLAST_viridiplantae100hits.fasta >
sp_list.RAPTORB_NCBI_BLAST_viridiplantae100hits.fasta
```

Evaluate trees

```
$ raxml-ng --evaluate --msa ALIGNMENT_FILE --model LG+F --tree bestTree.result --prefix GENE_DATABASE_MODEL
```

Fixing tree branch names

```
$ grep ">" RENAMED_FASTA_FILE > DATABASE_old_label.txt
$ perl -ne 'if(/ref\\|(.*?)\\|.*\\[(.*?)\\]/) {print $1." ".$2."\\n"}' GENE_NCBI_HomoloGene.fasta > homolo_new_label.txt
$ perl -ne 'if(/>(.*?) .*\\[(.*?)\\]/) {print $1." ".$2."\\n"}' GENE_NCBI_BLAST_viridiplantae100hits.fasta >
blast_new_label.txt
$ perl -ne 'if(/Org_(.*?) .*peptide: (.*?) /) {print $1." ".$2."\\n"}' GENE_phytozome_top100.fasta > phyto_new_label.txt
```

Reformat the final tree

```
$ for FILE in *.fixed;do perl -ne 'if(/^.*?(\\(.*?;).*/) {print $1}' $FILE > $FILE.nwk; done
```

Get AIC/BIC and logL values

```
$ find -name *.raxml.log -exec grep -i 'AIC score' {} +
$ find -name *.raxml.log -exec grep 'Final LogLikelihood' {} +
```