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' {} +