

## Assignment 4

Michelle

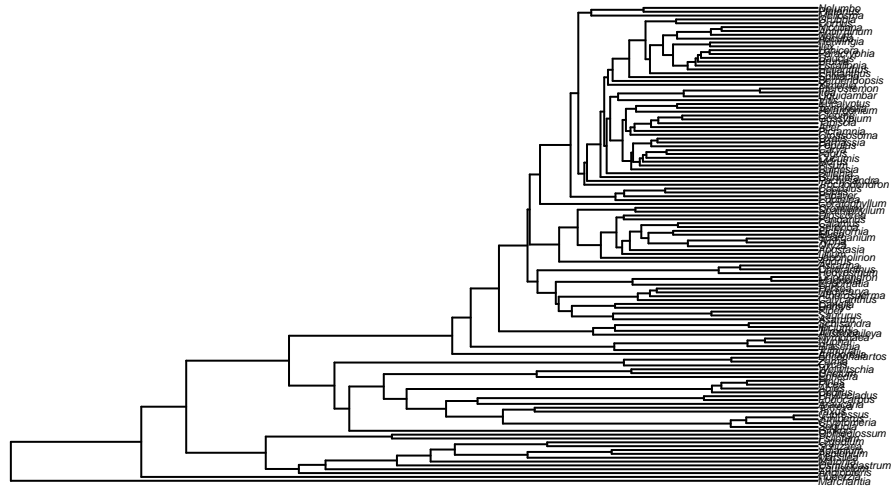
4/20/2020

```
devtools::install_github("bomeara/phybase")
```

#For real science, Liang's is the canonical one (and definitely cite his) but this will be easier for us to use for class.

#First, get a tree from Open Tree of Life. We'll get a recent plant tree from Beaulieu et al:

```
library(rotl)
library(ape)
phy <- get_study_tree("ot_485", "tree1")
plot(phy, cex=0.3)
```

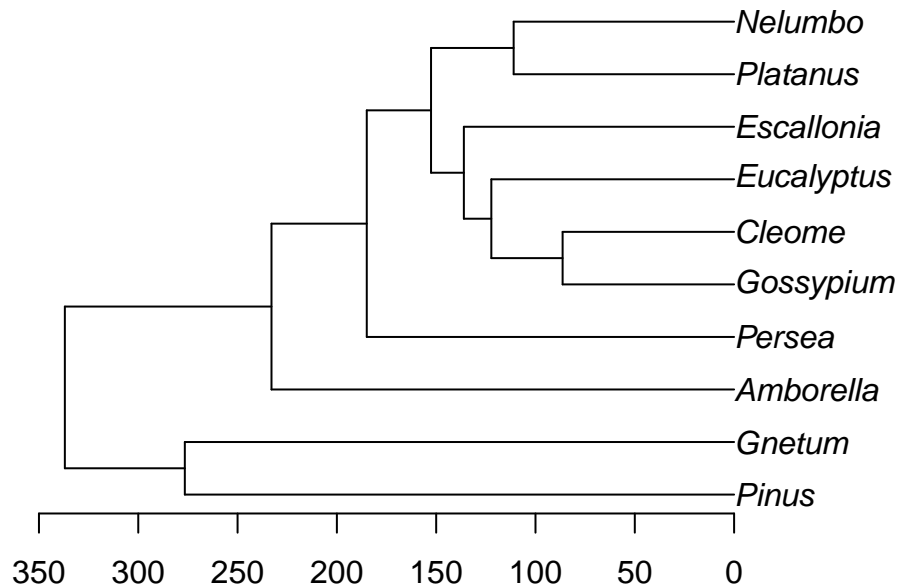


#Let's simplify by dropping some taxa

```
library(geiger)
```

```
## Registered S3 method overwritten by 'geiger':  
##   method      from  
## unique.multiPhylo ape
```

```
phy <- drop.random(phy, Ntip(phy) - 10)  
plot(phy)  
axisPhylo()
```



#We can simulate gene trees on this tree:

```
library(phybase)
```

```
## Loading required package: Matrix
```

```
##
```

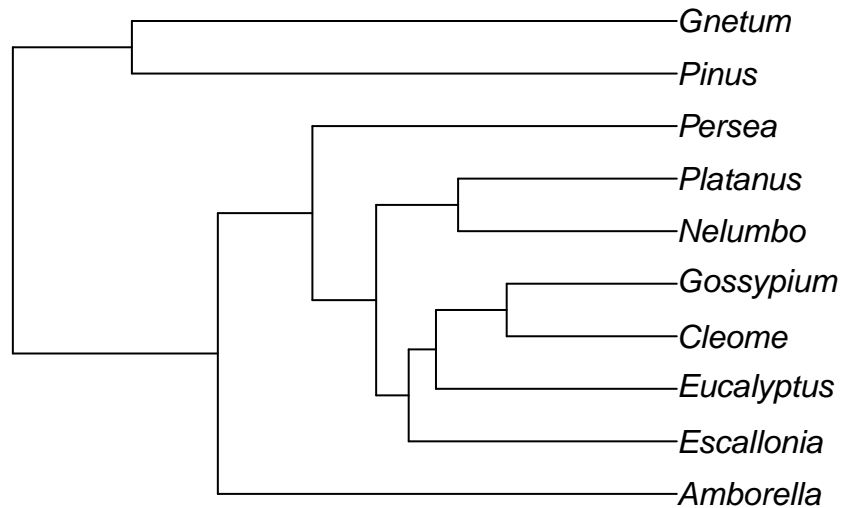
```
## Attaching package: 'phybase'
```

```
## The following objects are masked from 'package:ape':
```

```
##
```

```
##   node.height, write.dna
```

```
gene.tree <- phybase::sim.coaltree.phylo(phy, pop.size=1e-12)
plot(gene.tree)
```



#And it probably looks very similar to the initial tree:

```
library(phytools)
```

```
## Loading required package: maps
```

```
## Registered S3 method overwritten by 'phytools':
```

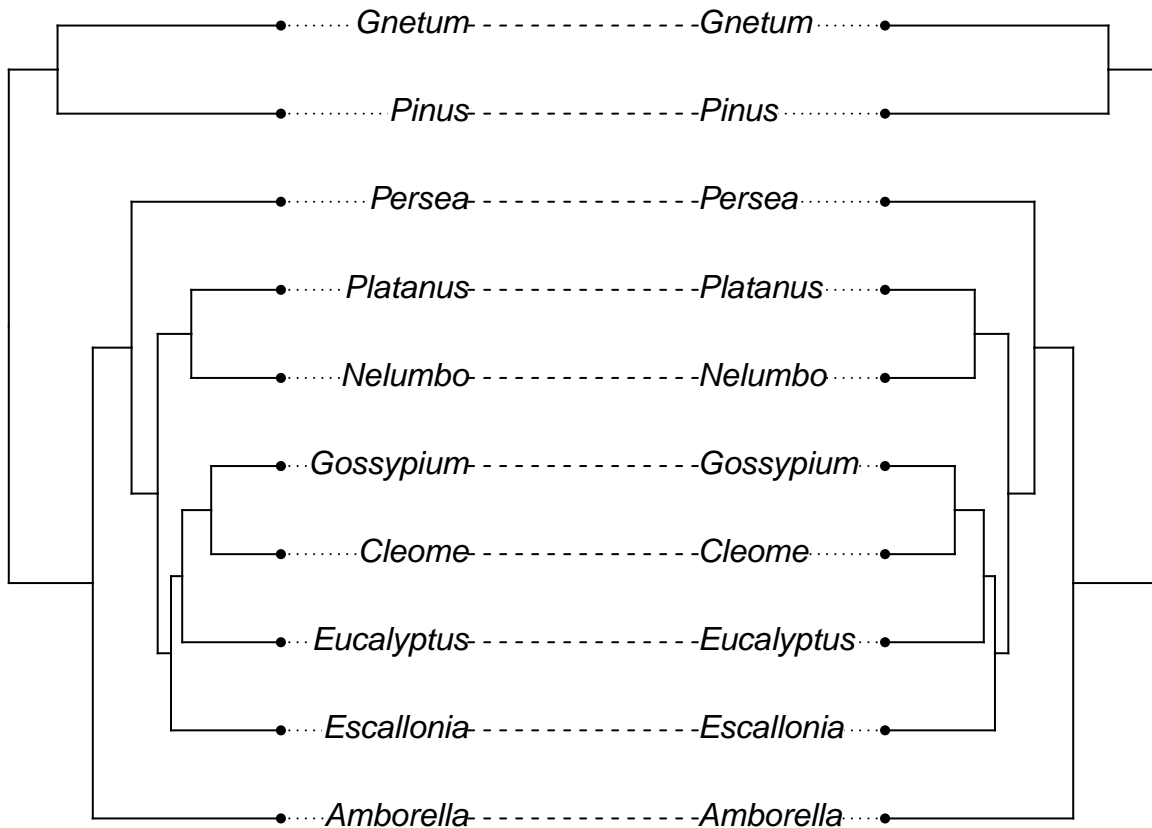
```
##   method      from
```

```
##   logLik.gfit geiger
```

```
plot(cophylo(phy, gene.tree, cbind(sort(phy$tip.label), sort(gene.tree$tip.label))))
```

```
## Rotating nodes to optimize matching...
```

```
## Done.
```



#Rotating nodes to optimize matching... # Done.

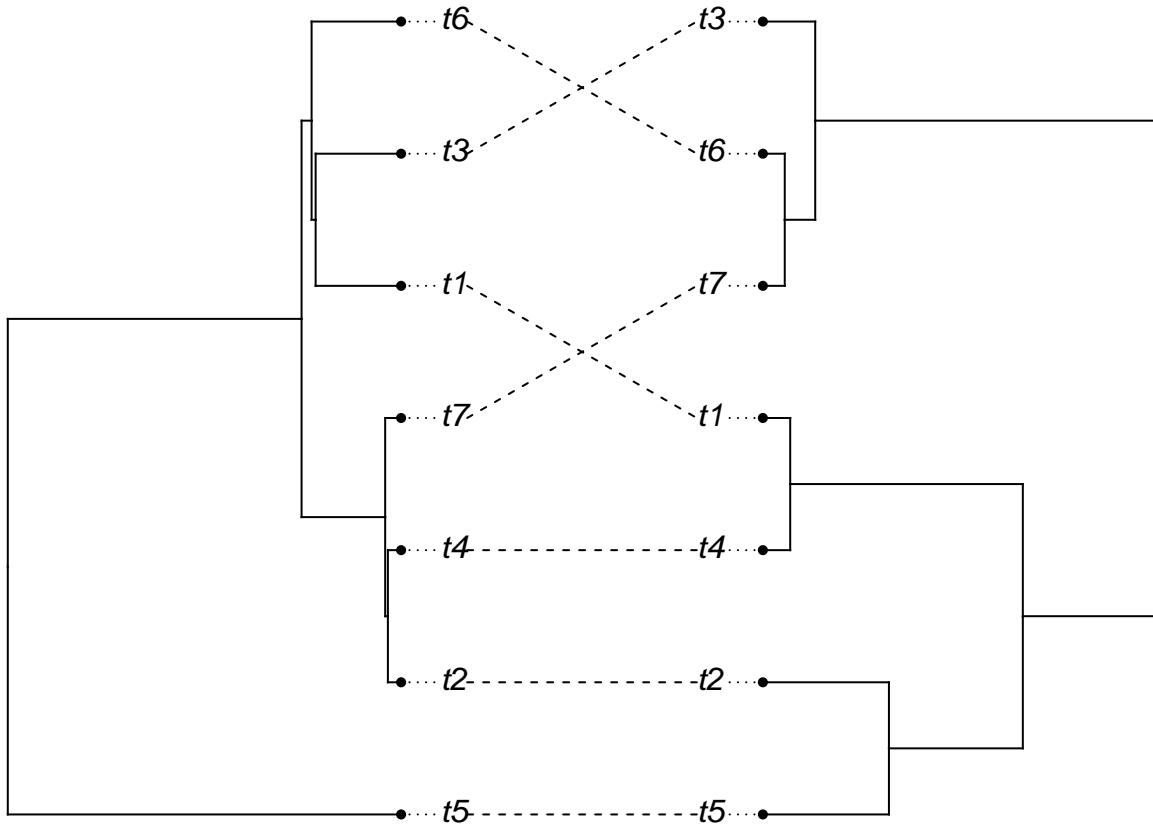
#So, does this mean gene tree species tree issues aren't a problem?

#Well, it depends on the details of the tree. One common misconception is that gene tree #species tree issues only relate to trees for recent events. This problem can happen any time #there are short, fat branches, where lack of coalescence of copies can occur.

```
species.tree <- rcoal(7)
species.tree$edge.length <- species.tree$edge.length / (10*max(branching.times(species.tree)))
gene.tree <- phybase::sim.coaltree.phylo(species.tree)
plot(cophylo(species.tree, gene.tree, cbind(sort(species.tree$tip.label), sort(gene.tree$tip.label))))
```

## Rotating nodes to optimize matching...

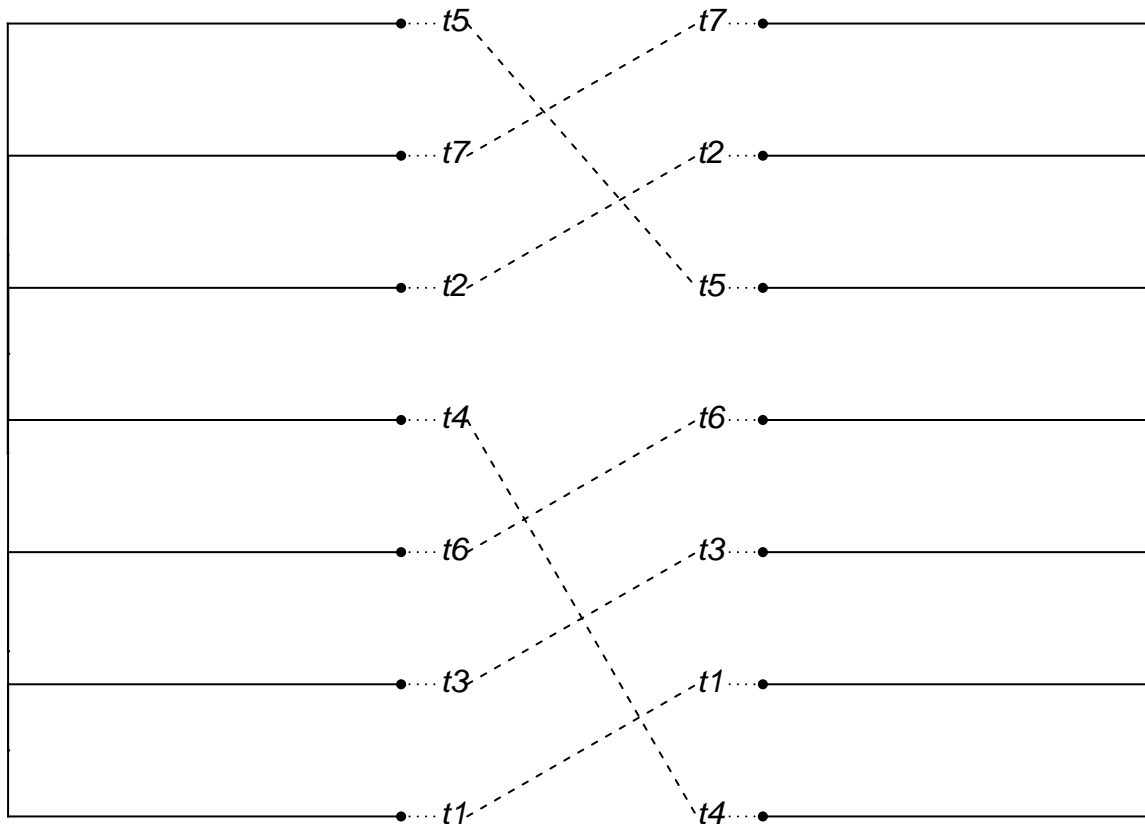
## Done.



## Rotating nodes to optimize matching... ## Done. #You should see (in most iterations), the above code giving a mismatch between the gene tree #and the species tree (the species tree has little height). Now, let's lengthen the tips of #the species tree:

```
tip.rows <- which(species.tree$edge[,2]<=Ntip(species.tree))
species.tree2 <- species.tree
species.tree2$edge.length[tip.rows] <- 100 + species.tree2$edge.length[tip.rows]
gene.tree2 <- phybase::sim.coaltree.phylo(species.tree2)
plot(cophylo(species.tree2, gene.tree2, cbind(sort(species.tree2$tip.label), sort(gene.tree2$tip.label))
```

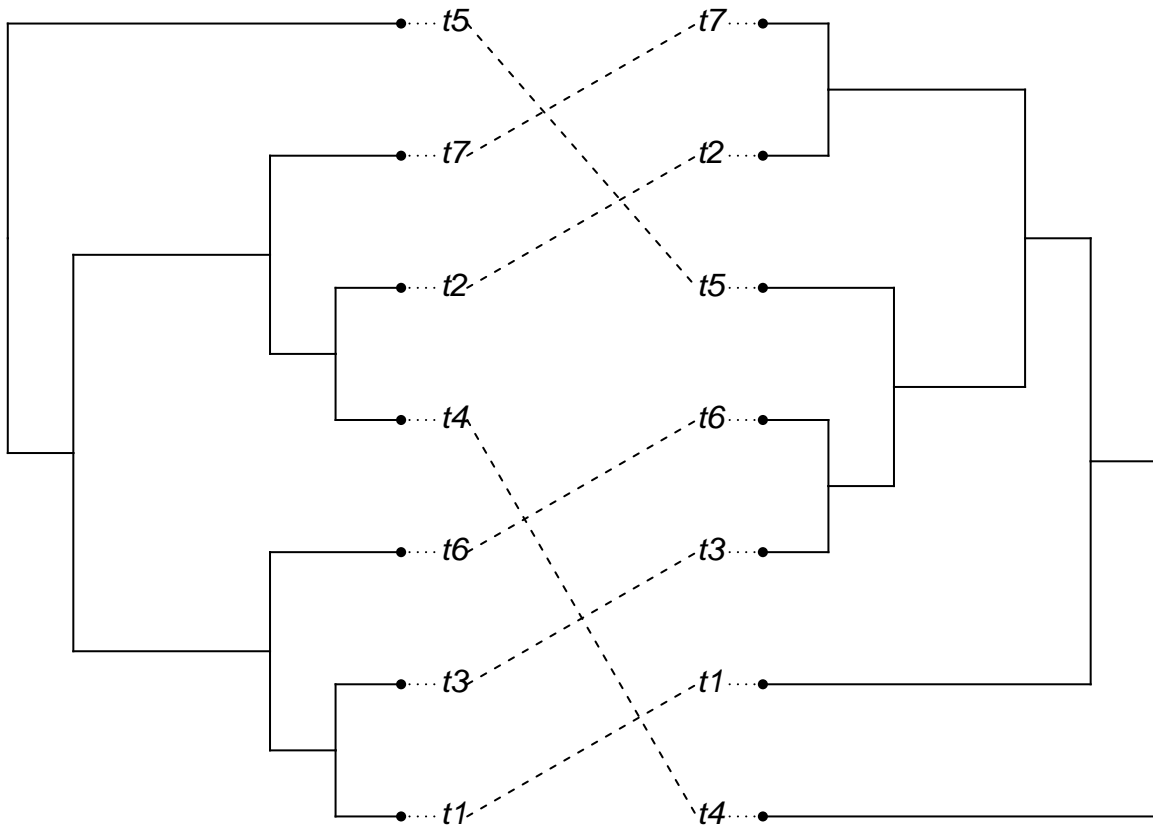
```
## Rotating nodes to optimize matching...
## Done.
```



## Rotating nodes to optimize matching... ## Done. #It looks like a mismatch, but it's hard to see, since the tips are so long. So plot the #cladogram instead [we need to manually change branch lengths to do this, though note we do #not resimulate the gene tree].

```
species.tree2.clado <- compute.brlen(species.tree2)
gene.tree2.clado <- compute.brlen(gene.tree2)
plot(cophylo(species.tree2.clado, gene.tree2.clado, cbind(sort(species.tree2.clado$tip.label),
sort(gene.tree2.clado$tip.label))))
```

## Rotating nodes to optimize matching...  
## Done.



#So we can see that even though the relevant divergences happened long ago, gene tree species tree issues are still a problem.