Lab Exercise 1: Introduction to estimating diversification rates

Tutorial

Lineage through time plots

```
library(phytools)
```

```
## Loading required package: ape
```

```
## Loading required package: maps
```
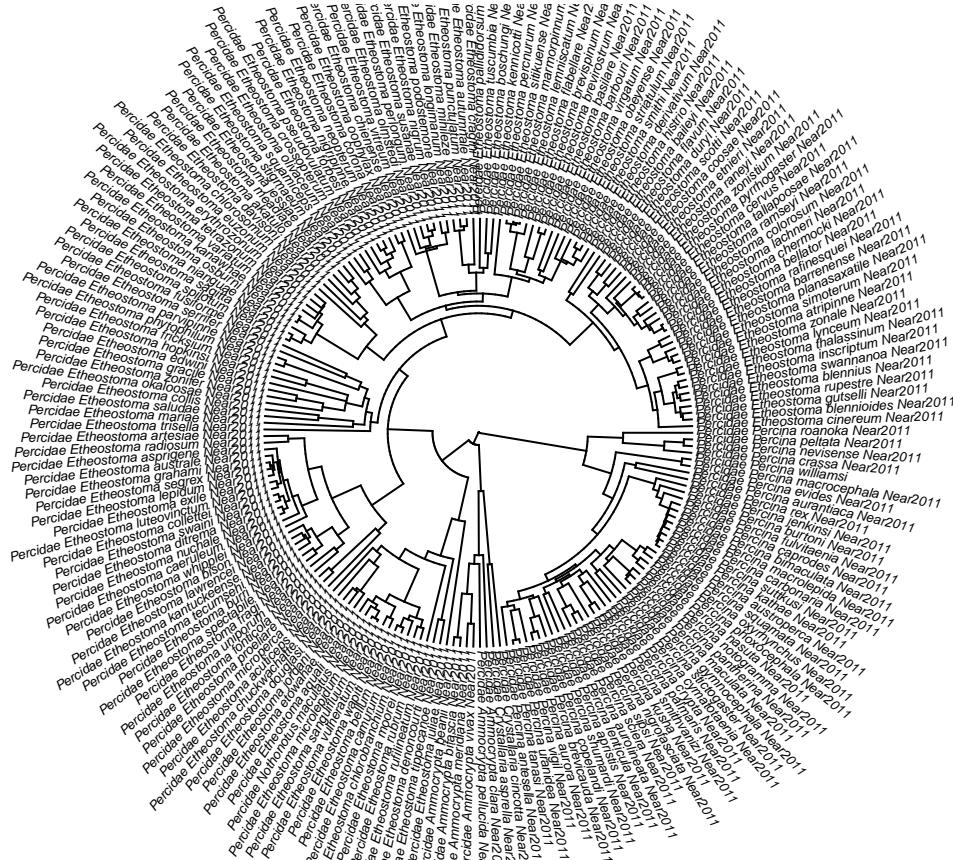
```
setwd("~/Dropbox/200A_Evolution/Labs/lab1")
# load tree
darter.tree <- read.tree("etheostoma_percina_chrono.tre")
# plot tree
plotTree(darter.tree,ftype="i",fsize=0.4,type="fan",lwd=1)
```



```
# generate LTT plot
obj<-ltt(darter.tree,log.lineages=FALSE)
```

```
# check tree
darter.tree
```

```
##
## Phylogenetic tree with 201 tips and 198 internal nodes.
##
## Tip labels:
##   Percidae_Etheostoma_cinereum_Near2011, Percidae_Etheostoma_blennioides_Near2011, Percidae_Etheostoma
##
## Rooted; includes branch lengths.
```

```
# a fully birfucating tree with N tips will have N-1 internal nodes
is.binary(darter.tree)
```

```
## [1] FALSE
```

```
# comes back false. we must fix this
darter.tree<-multi2di(darter.tree)
darter.tree
```

```
##
## Phylogenetic tree with 201 tips and 200 internal nodes.
##
## Tip labels:
##   Percidae_Etheostoma_cinereum_Near2011, Percidae_Etheostoma_blennioides_Near2011, Percidae_Etheostoma
##
## Rooted; includes branch lengths.
```

```
# now it has 201 tips and 200 nodes
is.binary(darter.tree)
```

[1] TRUE

```
we have resolved the inner nodes

# create a new object storing tree data
obj<-ltt(darter.tree,plot=FALSE); obj
```

```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 201 tips and 200 internal nodes.
##
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.
##
## (3) A value for Pybus & Harvey's "gamma" statistic of 0.2007, p-value = 0.841.
```

```
# make a new LTT plot
par(mfrow=c(1,1), mar=c(4, 4, 3, 2))
plot(obj,log.lineages=FALSE,main="LTT plot for darters")
# plot tree and LTT plot together
plot(obj,show.tree=TRUE,log.lineages=FALSE,main="LTT plot for darters")
```

LTT plot for darters
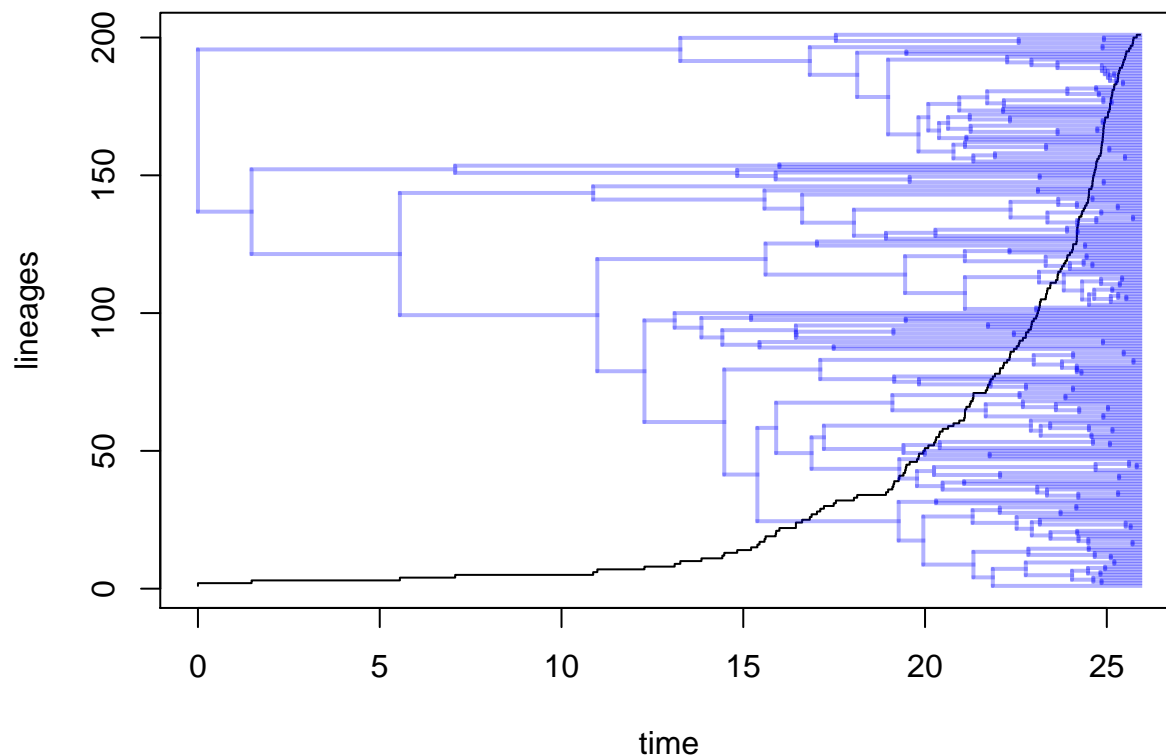

```
# number of linages rises exponentially through time with the pure birth model
# we want to show a linear relationship
# plot our pure-birth trees on a semi-log scale
# vertical axis is on a log-scale and horizontal axis is on a linear scale.
plot(obj,log.lineages=FALSE,log="y",main="LTT plot for darters",
```
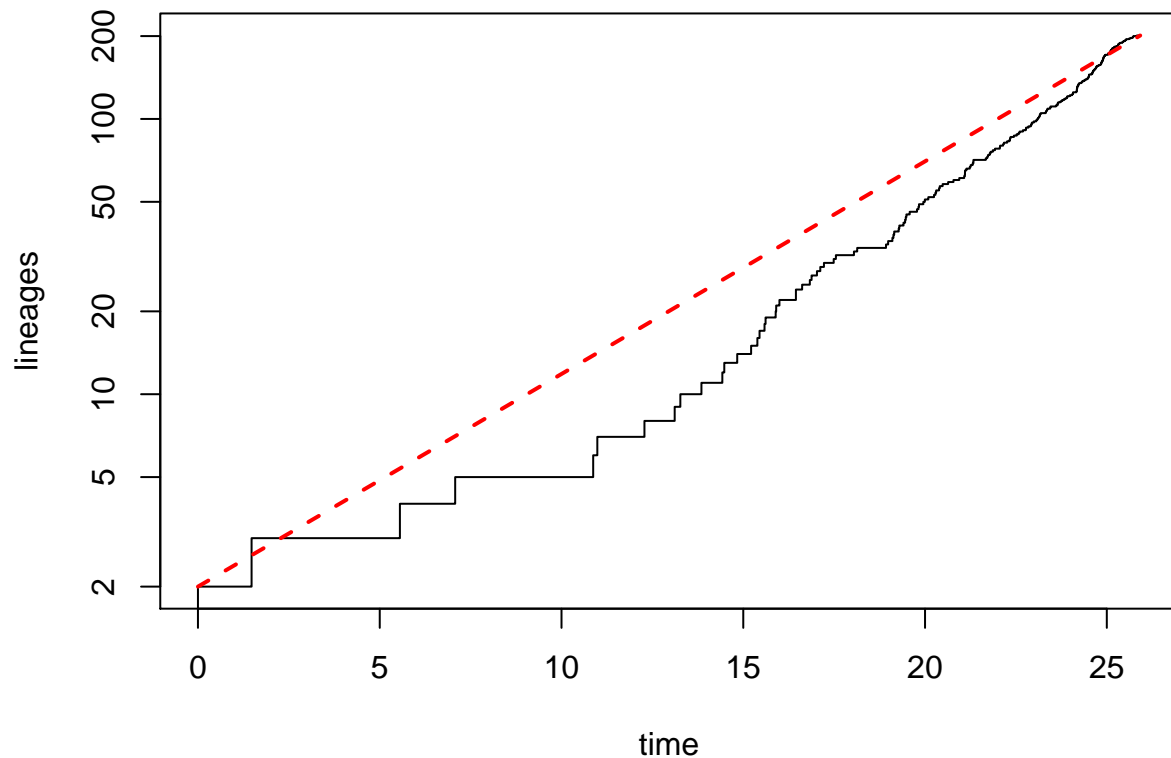
```
    ylim=c(2,Ntip(darter.tree)))
## we can overlay the pure-birth prediction:
h<-max(nodeHeights(darter.tree))
x<-seq(0,h,by=h/100)
b<-(log(Ntip(darter.tree))-log(2))/h
lines(x,2*exp(b*x),col="red",lty="dashed",lwd=2)
```

LTT plot for darters


```
# compare observed LTT to simulated LTTs assuming a pure-birth process of the
# same duration & resulting in the same total number of species
# use phytools function pbtree
trees<-pbtree(b=b,n=Ntip(darter.tree),t=h,nsim=100,method="direct",
            quiet=TRUE)
obj<-ltt(trees,plot=FALSE)
plot(obj,col="grey",main="LTT of darters compared to simulated LTTs")
lines(c(0,h),log(c(2,Ntip(darter.tree))),lty="dashed",lwd=2,col="red")
## now let's overlay our original tree
ltt(darter.tree,add=TRUE,lwd=2)
```
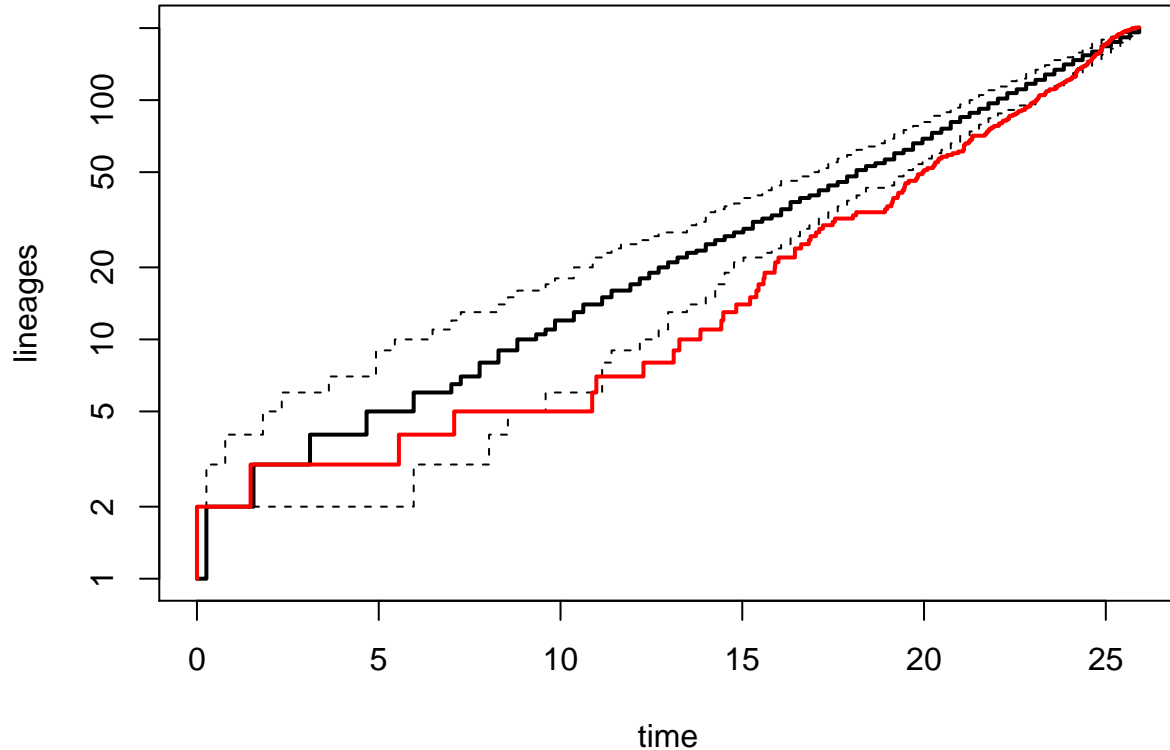
LTT of darters compared to simulated LTTs
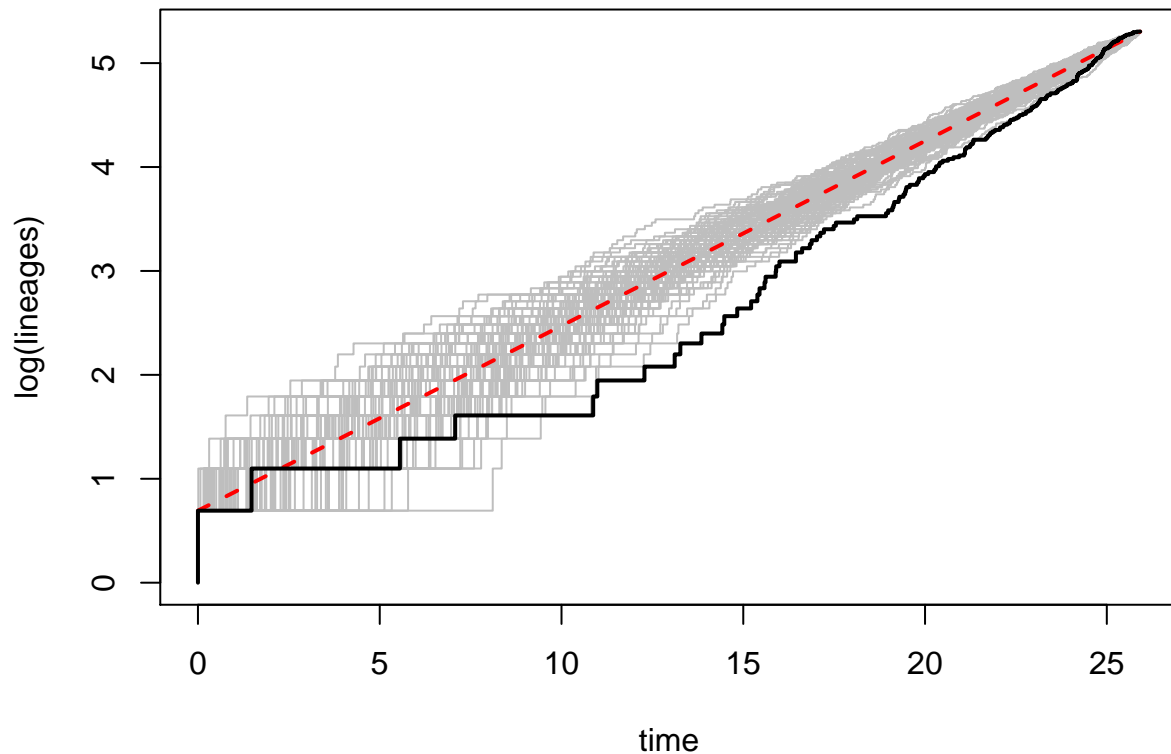

```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 201 tips and 200 internal nodes.
##
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.
##
## (3) A value for Pybus & Harvey's "gamma" statistic of 0.2007, p-value = 0.841.
```

```r
# ltt95 gives a (1-alpha)% CI for the LTT based on a set of trees
ltt95(trees,log=TRUE)
title(main="LTT of darters compared to simulated LTTs")
ltt(darter.tree,add=TRUE,log.lineages=FALSE,col="red",lwd=2)
```

LTT of darters compared to simulated LTTs


```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 201 tips and 200 internal nodes.
##
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.
##
## (3) A value for Pybus & Harvey's "gamma" statistic of 0.2007, p-value = 0.841.
```

Fitting pure birth and birth-death models to trees

```
library(phytools)
fitbd <- birthdeath(darter.tree)
fitbd
```

```
##
## Estimation of Speciation and Extinction Rates
##            with Birth-Death Models
##
##      Phylogenetic tree: darter.tree
##         Number of tips: 201
##               Deviance: -741.7514
##         Log-likelihood: 370.8757
##     Parameter estimates:
##       d / b = 0.06330651   StdErr = 0.1640933
##       b - d = 0.2212409   StdErr = 0.02342254
##    (b: speciation rate, d: extinction rate)
##    Profile likelihood 95% confidence intervals:
##       d / b: [-0.1838519, 0.2583648]
##       b - d: [0.1914747, 0.2539717]
```

```
# use a phytools function bd() to get b and d rate separately
bd(fitbd)
```

```
##          b          d
## 0.23619344 0.01495258
```

The gamma-statistic

```
# gamma-statistic has astandard normal distribution for trees generated under a pure-birth speciation m
# test this with our 100 simulated pure-birth phylogenies
g<-sapply(trees,function(x) ltt(x,plot=FALSE)$gamma)
hist(g,main=expression(paste("Distribution of ",gamma," from simulation")))
```

Distribution of γ from simulation
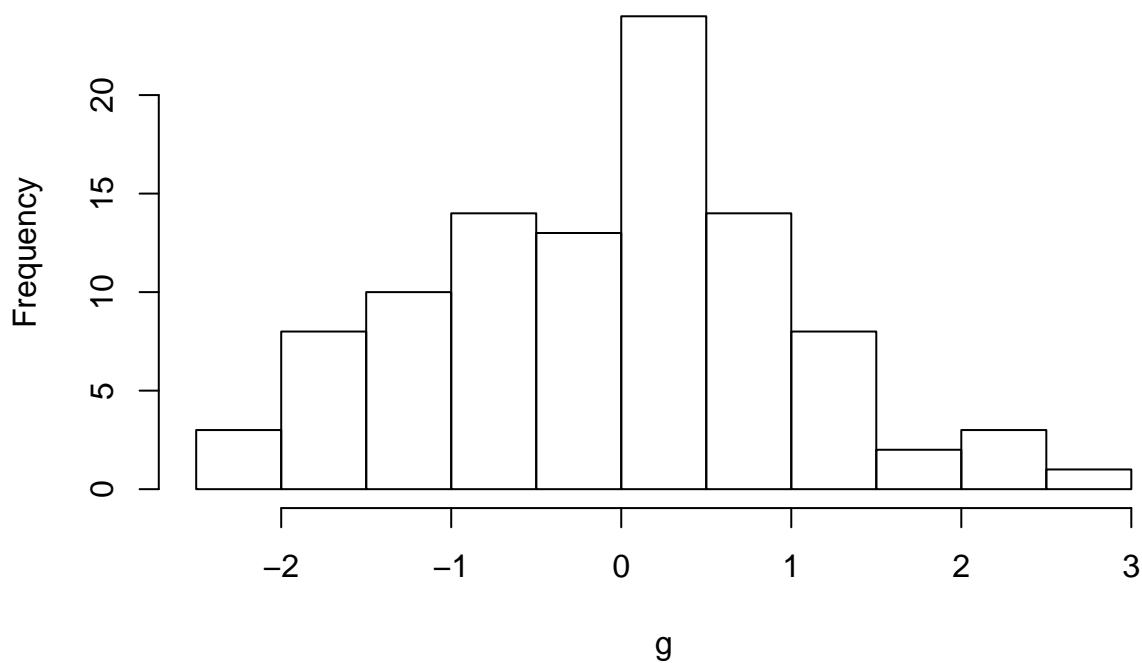

```
mean(g)
```

```
## [1] -0.07865308
```

```
var(g)
```

```
## [1] 1.102241
```

```
# test hypotheses about gamma automatically with ltt
obj<-ltt(darter.tree,plot=FALSE)
print(obj)
```

```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 201 tips and 200 internal nodes.
##
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.
##
## (3) A value for Pybus & Harvey's "gamma" statistic of 0.2007, p-value = 0.841.
```
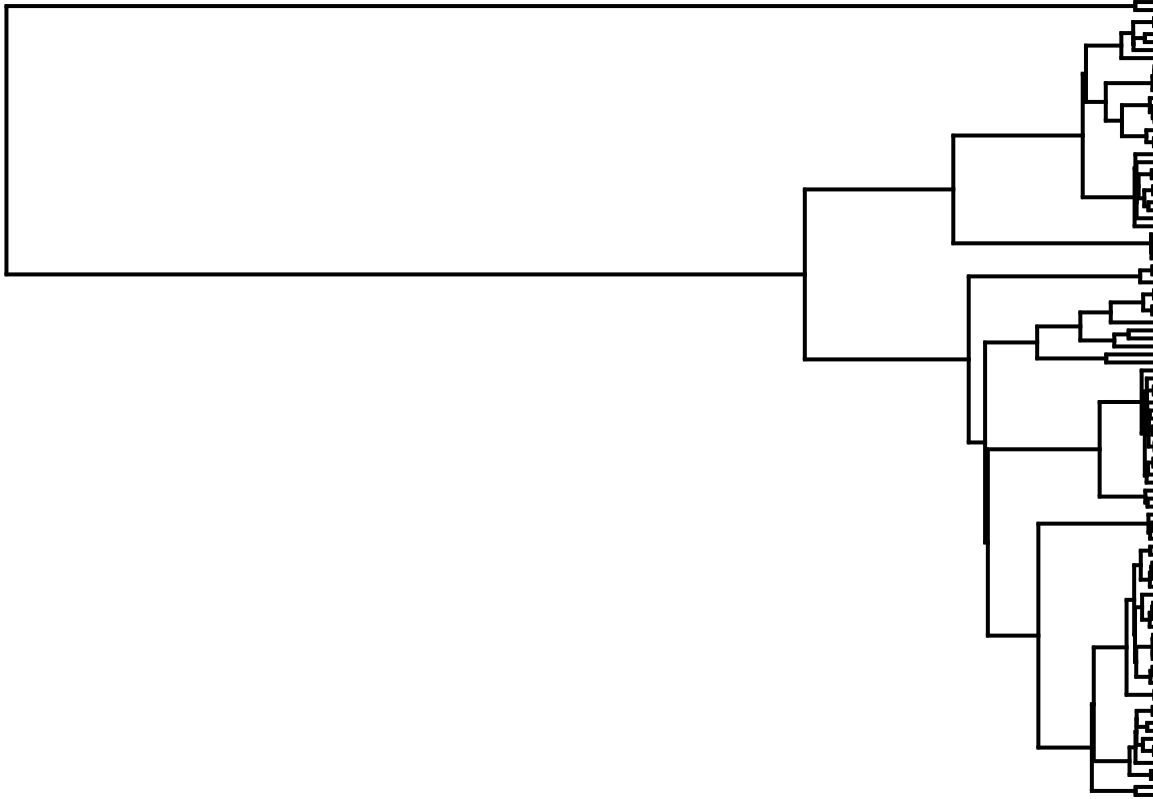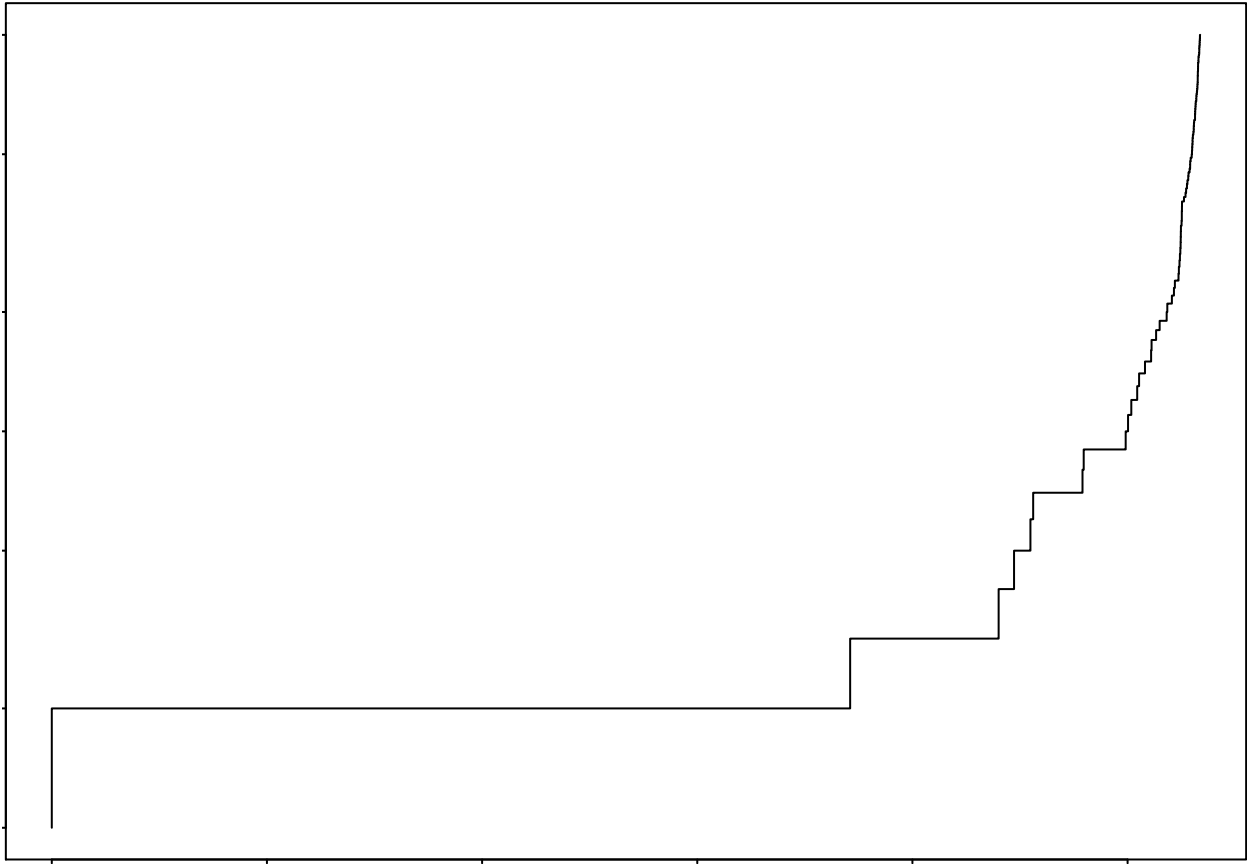
```
# simulate a tree under a different model of lineage accumulation – the coalescent –
# and see what the result is
# this should result in a significantly positive gamma
coal.tree<-rcoal(n=100)
plotTree(coal.tree,ftype="off")
```



```
# and it does
obj<-ltt(coal.tree,log.lineages=FALSE,log="y")
```

```
darter.gamma <- obj$gamma # ltt returns a gamma value as one of its elements

# compare the mean gamma value of 200 pb trees to gamma from the coalescent tree
par(mfrow=c(1,1), mar=c(4, 4, 3, 2))
trees<-pbtree(n=100,nsim=200,scale=max(nodeHeights(coal.tree)))
ltt95(trees,log=TRUE)
title(main="Simulated coalescent trees compared to pure-birth LTTs")
ltt(coal.tree,add=TRUE,log.lineages=FALSE,col="red",lwd=2,lty="dashed")
```

Simulated coalescent trees compared to pure−birth LTTs
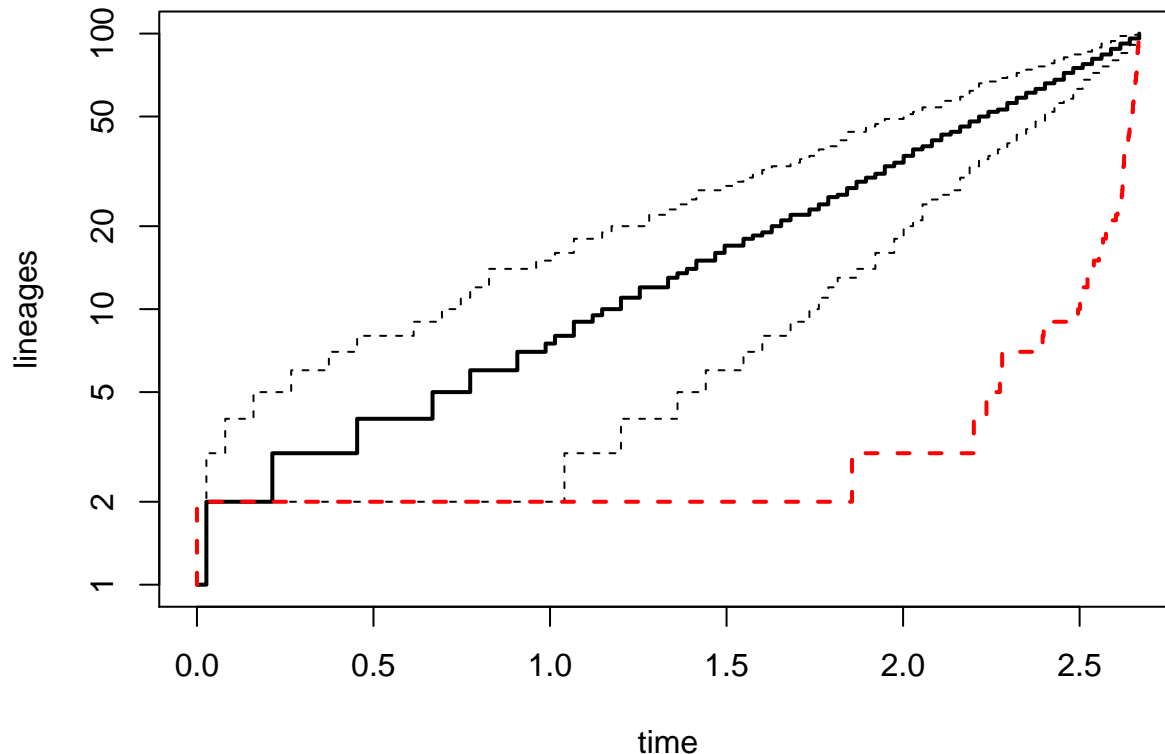
```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 100 tips and 99 internal nodes.
##
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.
##
## (3) A value for Pybus & Harvey's "gamma" statistic of 11.371, p-value = 0.
```

Incomplete sampling

```
# the authors only sampled 201 of 216 species
# run MCCR test
library(geiger)
age <- 25.91862
richness <- 216
darterbirth =  (log(richness) - log(2))/age
darterbirth
```

```
## [1] 0.1806474
```

```
richness <- 216
missing <- 15
#this simulates gamma values when trees are undersampled.
#we will grow trees with n=34 and prune them down to 13 taxa

num_simulations<-200 #number of simulations
g1_null<-numeric(num_simulations) #g1_null will hold the simulated gamma values
for(i in 1:num_simulations) {
  sim.bdtree(darterbirth, d=0, stop = "taxa", n=richness)->sim_tree
  drop.random(sim_tree, missing)->prune # prune down to the # of taxa in the phylogeny
```
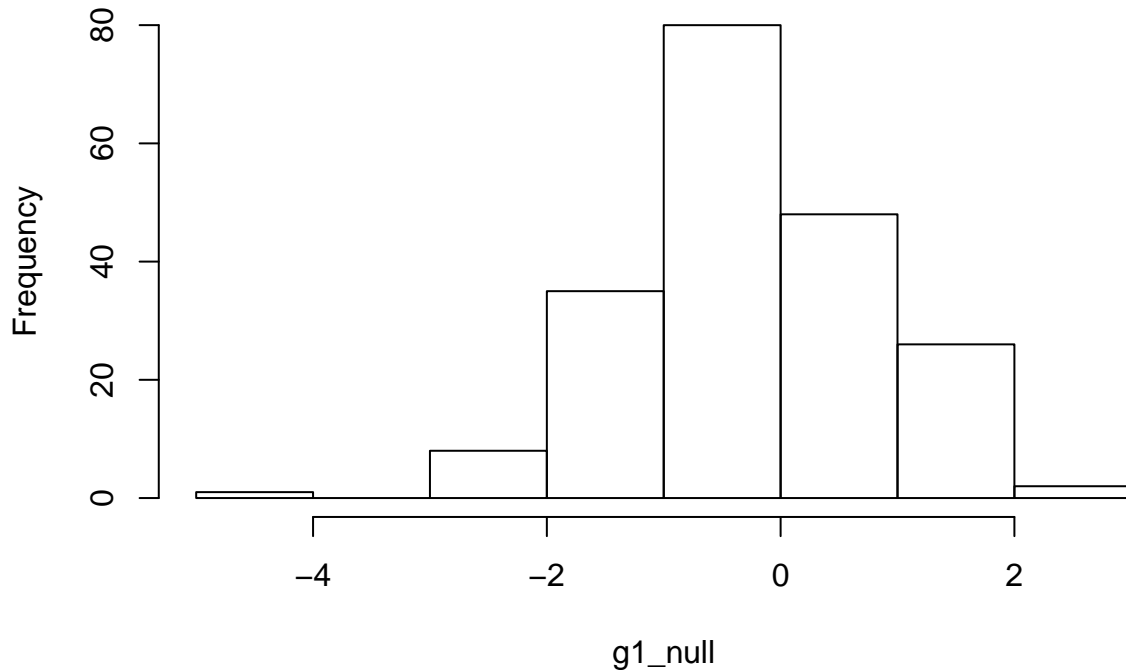
```
    gammaStat(prune)->g1_null[i]
}
# create a histogram of the null distribution
hist(g1_null)

#arrow indicates where the observed gamma falls in the null you just generated
arrows(darter.gamma, 40, darter.gamma, 0, col="red", lwd=2)
```

Histogram of g1_null

g1_null

```
# Which of the null values are smaller (more negative) than the data?
smallerNull<-g1_null<=darter.gamma
# How many TRUEs are there?
count<-sum(smallerNull); count
```

[1] 200

```
# finally, what is the p-value?
mccr_pval<-(count+1)/(num_simulations+1)
mccr_pval
```

[1] 1

Now that the tutorial is completed, it is time to move on to the lab exercise.

Exercise

1. Calculate the gamma statistic for this phylogeny of homalopsid snakes from Alfaro et al., 2008: snake.tre.

First we need to import the tree

```
library(phytools)
setwd("~/Dropbox/200A_Evolution/Labs/lab1")
snake_phy <- read.tree("homalops.phy")
plotTree(snake_phy,ftype="i",fsize=0.4,type="phylogram",lwd=1)
```
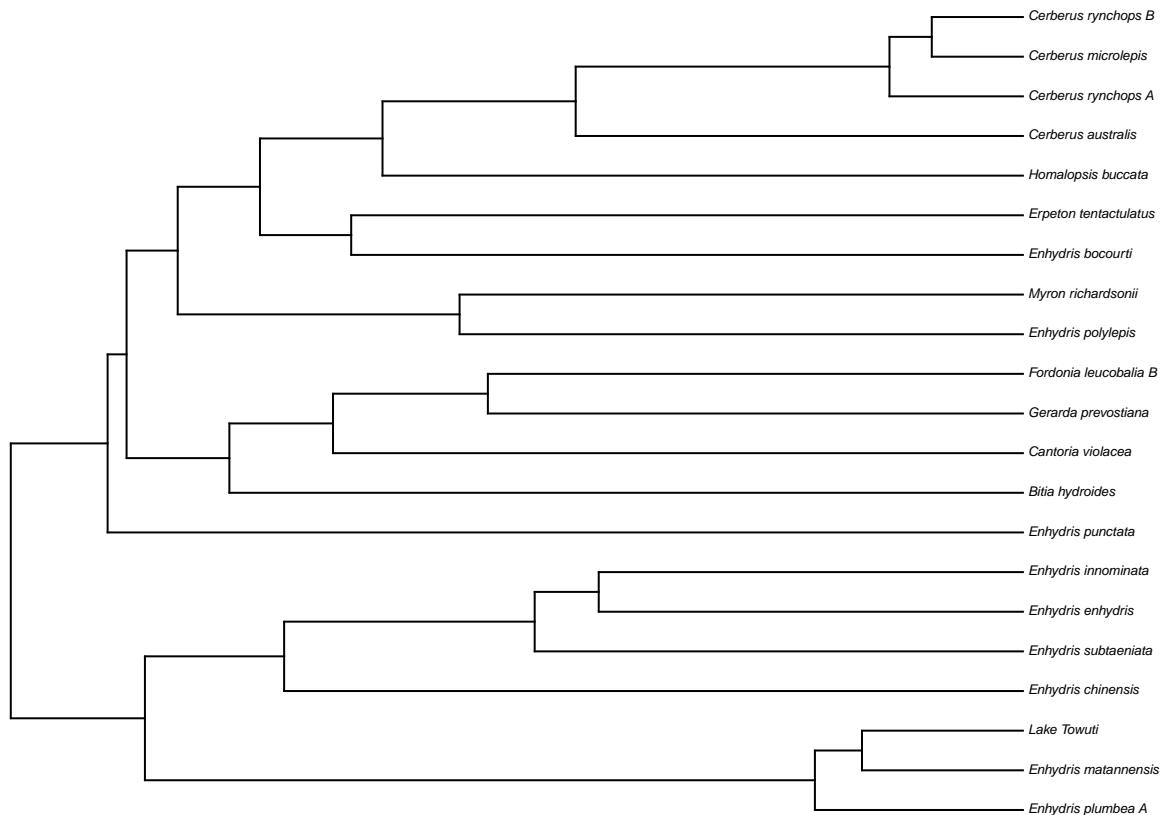

Now we need to generate an LTT plot

```
snakeLTT <- ltt(snake_phy, plot = FALSE); snakeLTT
```

```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 21 tips and 20 internal nodes.
##
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.
##
## (3) A value for Pybus & Harvey's "gamma" statistic of -3.2411, p-value = 0.0012.
```

This gives a gamma statistic of -3.2411.

2. Given this gamma value, what would you conclude about the tempo of speciation in this clade?

A negative gamma statistic suggests that the rate of speciation was more rapid earlier in the tree; the rate of speciation was historically higher and slowed down more recently.

3. Given that the crown age of the snake radiation is 22 MY and the total richness of the clade is 34 species, determine whether the observed gamma could be due to the amount of incomplete sampling in the empirical tree. On the basis of the MCCR test what can you conclude about the tempo of homalopsid snake diversification?

We will run an MCCR test growing trees with n=34 and then pruning them down to 13 taxa.

```
library(geiger)
age <- 22
# use a purebirth estimate of the the tree based upon their total age and richness
richness <- 34
missing <- 13
snakebirth = (log(richness) - log(2))/age
snake_gamma <- snakeLTT$gamma


num_simulations<-200 #number of simulations
g_null<-numeric(num_simulations) #gamma_null will hold the simulated gamma values
for(i in 1:num_simulations) {
  sim.bdtree(snakebirth, d=0, stop = "taxa", n=richness)->sim_tree
  drop.random(sim_tree, missing)->prune # prune down to the # of taxa in the phylogeny
  gammaStat(prune)->g_null[i]
}
# create a histogram of the null distribution
hist(g_null)

#arrow indicates where the observed gamma falls in the null you just generated
arrows(snake_gamma, 40, snake_gamma, 0, col="red", lwd=2)
```

Histogram of g_null
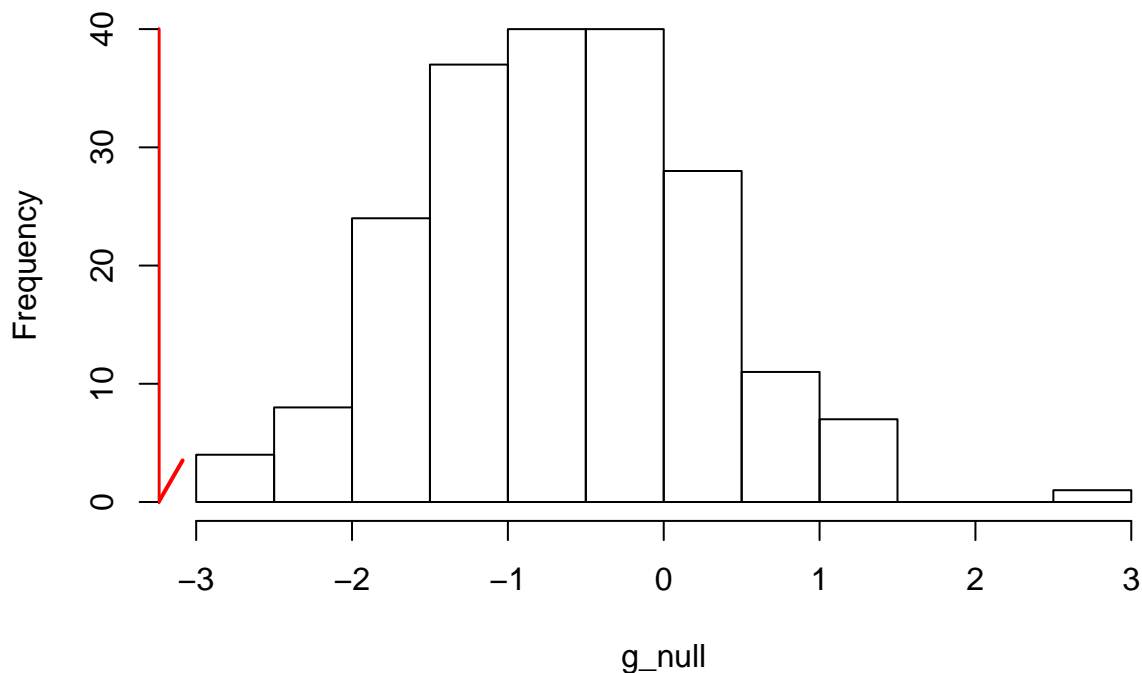
Create a vector with null values more negative than the data and count how many times that happens as the variable count

```
smallerNull<-g_null<=snake_gamma
count<-sum(smallerNull); count
```

[1] 0

Finally, get the p-value from the test.

```
mccr_pval<-(count+1)/(num_simulations+1); mccr_pval
```

[1] 0.004975124

The MCCR test yielded a very small p-value (< 0.01) with little to no values from the null simulated gamma distribution more negative than the gamma statistic from the original phylogeny. These results suggest that the observed gamma could be due to the amount of incomplete sampling in the given homalopsid tree. With the current level of sampling, the older nodes are more likely to be recovered and the negative gamma statistic favors more rapid diversification earlier in the tree. However, based on the results of the MCCR test, more recent rates of diversification could be more rapid than what is represented by this tree with its current level of sampling.

4. What is the birth rate and death rate of the homalopsid tree?

```
fitbd <- birthdeath(snake_phy)
bd(fitbd)
```

```
##          b          d
## 0.06839495 0.00000000
```

5. Find a time-calibrated phylogeny for a group that interests you (ideally with more than 30 tips and fewer than 200). Do the following: (1) Describe the clade (including a description of the number of tips in the tree and the total number of species in the clades) and provide a reference or citation to the source. (2) Fit a birthdeath model to this tree and report b and d. (3) Perform an MCCR test and describe whether the gamma value is extreme or not given the level of sampling in the tree.

(1) I have acquired a phylogeny of fish of the family Pomacentridae; the tree includes 208 of the 386 species that make up the family (Frédérich, B. et al. 2013). There are also some species in the tree that are not in the family that serve as outgroups. This tree has 216 tips and I pruned the outgroups for a tree with 208 tips (only a little bit over 200).

```
poma_tree <- read.nexus("damsels260miltime.tree.txt")
# remove the outgroups from tree
tip <- c("OUG_Centropyge_bicolor", "OUG_Cymatogaster_aggregata", "OUG_Cypho_purpurascens",
        "OUG_Dicentrarchus_labrax", "OUG_Embiotoca_jacksoni", "OUG_Ptychochromis_oligacanthus",
        "OUG_Semicossyphus_pulcher", "OUG_Thorichthys_meeki")
poma_only_tree <- drop.tip(poma_tree, tip)
# take a look at it
poma_only_tree
```

```
##
## Phylogenetic tree with 208 tips and 207 internal nodes.
```

```
##
## Tip labels:
##  Abudefduf_abdominalis, Abudefduf_bengalensis, Abudefduf_concolor, Abudefduf_declivifrons, Abudefduf_
##
## Rooted; includes branch lengths.
```

This is what my pruned tree looks like:

```
plotTree(poma_only_tree,ftype="i",fsize=0.4,type="phylogram",lwd=1)
```
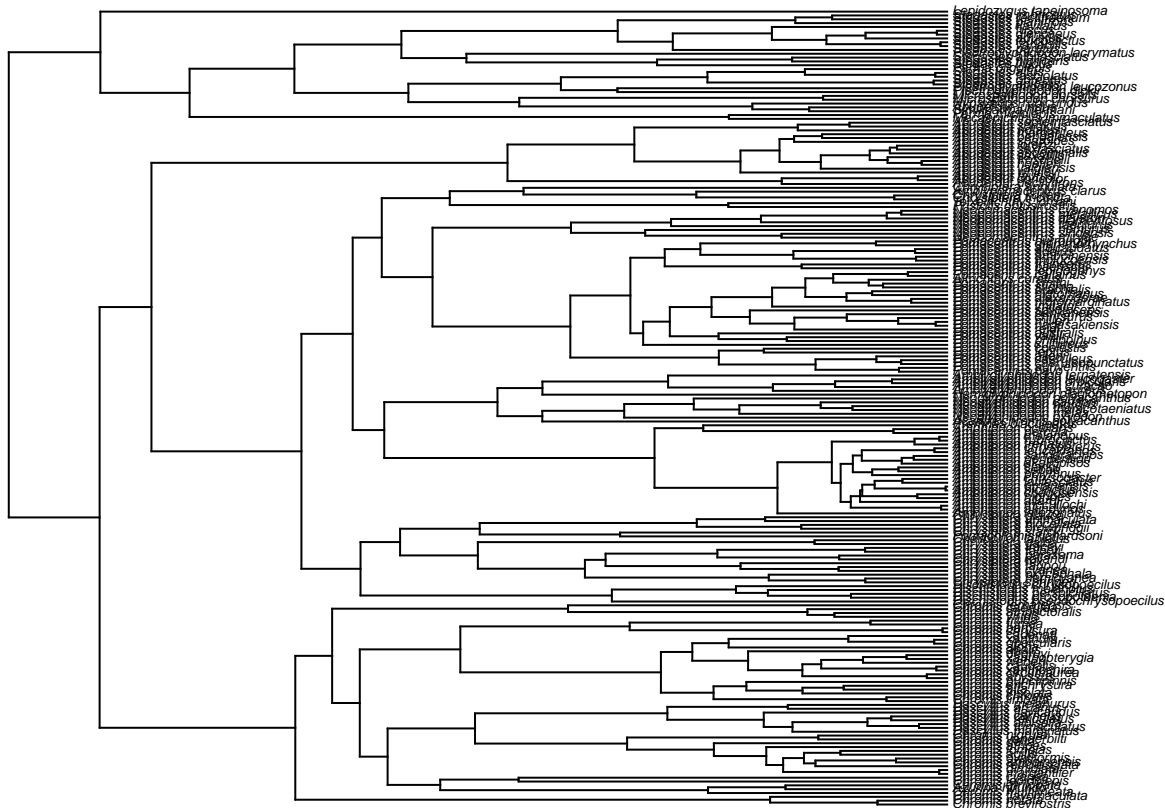

(2) Now fit a birth-death model and report b and d:

```
# fit a birth-death model
poma_bd <- birthdeath(poma_only_tree)
# report b and b
bd(poma_bd)
```

```
##         b         d
## 0.0823782 0.0000000
```

Now I will run an MCCR test starting with 386 taxa and pruning down to 178

```
# get an LTT plot
pomaLTT <- ltt(poma_only_tree, plot = FALSE); pomaLTT
```

```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 208 tips and 207 internal nodes.
##
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.
##
```

15

```
## (3) A value for Pybus & Harvey's "gamma" statistic of -1.7917, p-value = 0.0732.
# age of clade in MY
age <- 50
# use a purebirth estimate of the the tree based upon their total age and richness
richness <- 386
missing <- 178
poma_birth = (log(richness) - log(2))/age
poma_gamma <- pomaLTT$gamma



num_simulations<-200 #number of simulations
gamma_null<-numeric(num_simulations)
for(i in 1:num_simulations) {
  sim.bdtree(poma_birth, d=0, stop = "taxa", n=richness)->sim_tree
  drop.random(sim_tree, missing)->prune # prune down to the # of taxa in the phylogeny
  gammaStat(prune)->gamma_null[i]
}
# create a histogram of the null distribution
hist(gamma_null)

#arrow indicates where the observed gamma falls in the null you just generated
arrows(poma_gamma, 40, poma_gamma, 0, col="red", lwd=2)
```

Histogram of gamma_null
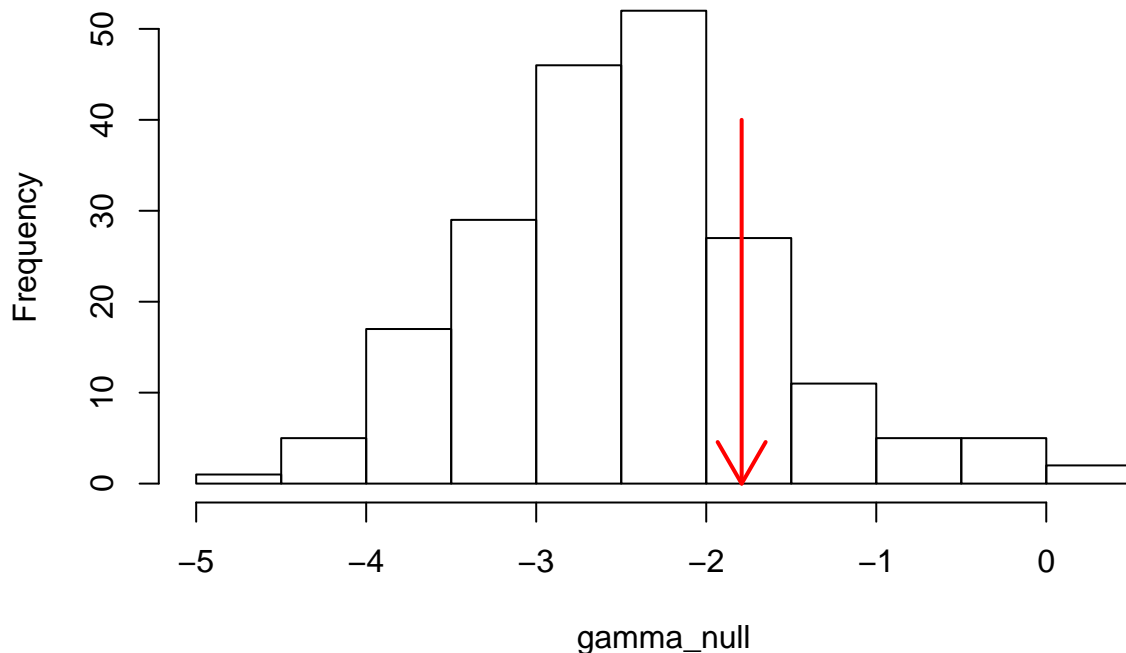
Create a vector with null values more negative than the data and count how many times that happens as the variable count

```
smallerNull<-gamma_null<=poma_gamma
count<-sum(smallerNull); count
```

```
## [1] 165
```

Get the p-value from the test

```
mccr_pval<-(count+1)/(num_simulations+1); mccr_pval
```

```
## [1] 0.8258706
```

(3) The MCCR test yielded a large p-value with a large number values from the null simulated gamma distribution more negative than the gamma statistic from the original Pomacentridae phylogeny. These results suggest that the observed gamma value is not extreme given the level of sampling in the tree.

Reference:

Frédérich, B. et al., 2013. Iterative ecological radiation and convergence during the evolutionary history of damselfishes (Pomacentridae). The American naturalist, 181(1), pp.94–113. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23234848.