Lab 1

TUTORIAL

Exercise 1

PACKAGES

```
library(phytools)
```

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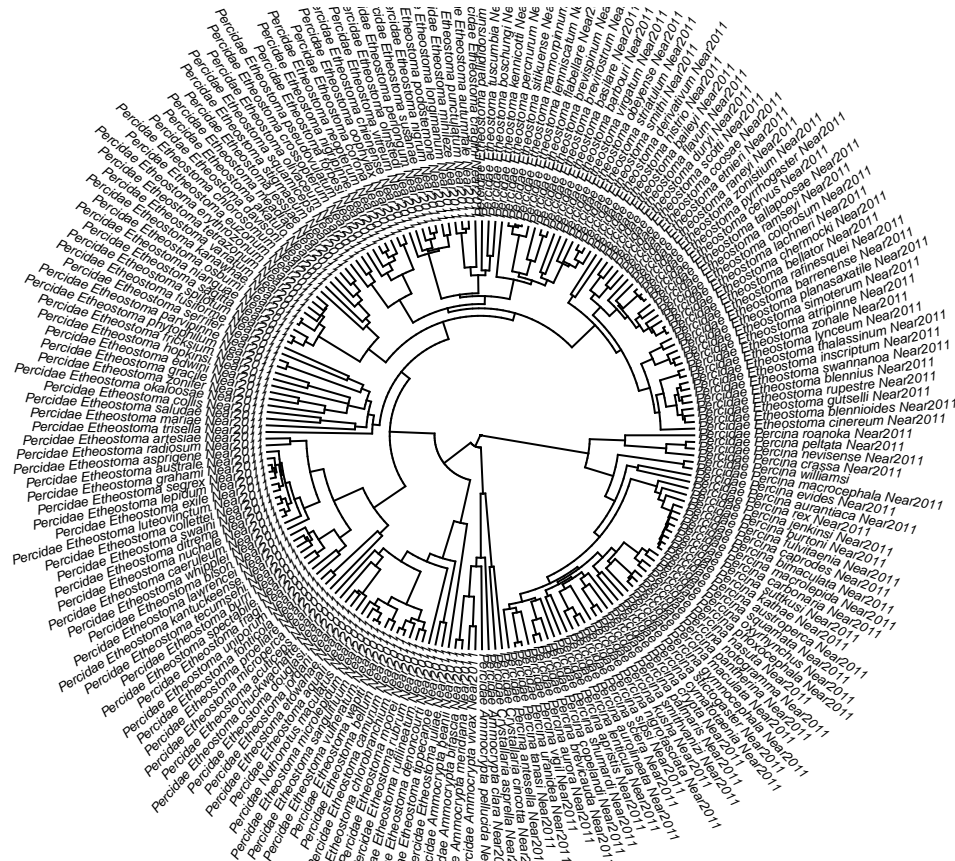
```
## Loading required package: maps
```

LOAD DATA

```
darter.tree<-read.tree("etheostoma_percina_chrono.tre")
```
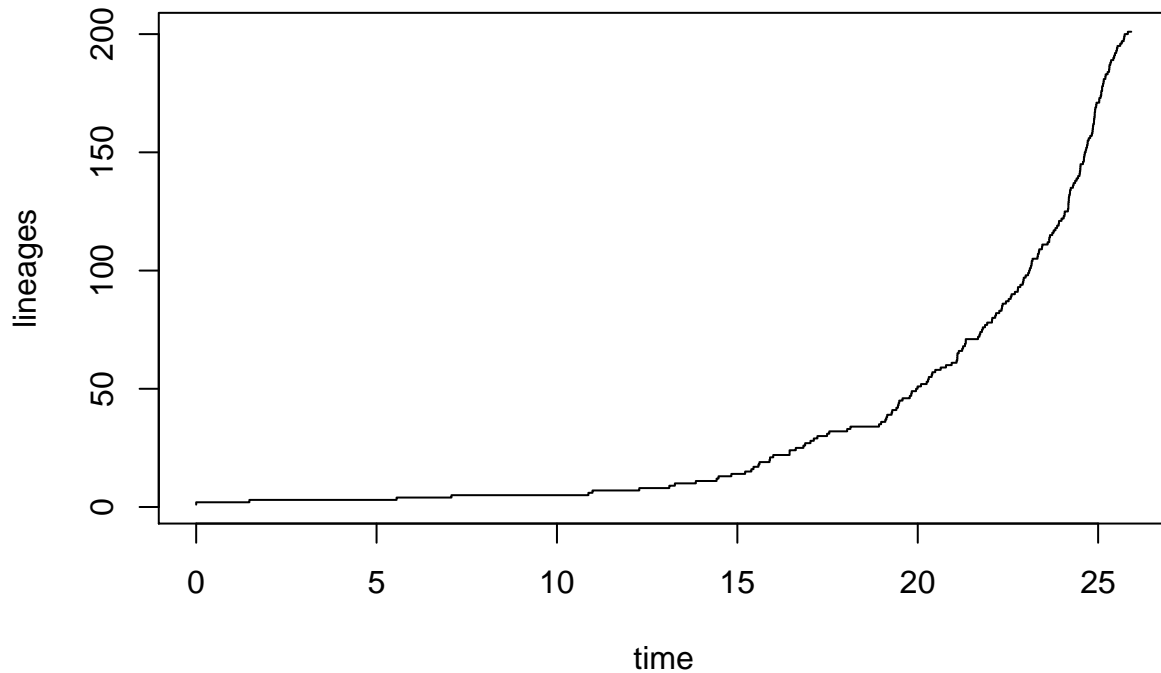
PLOT TREE

```
plotTree(darter.tree,ftype="i",fsize=0.4,type="fan",lwd=1)
```

LTT PLOT

```
obj<-ltt(darter.tree,log.lineages=FALSE) ; obj
```



```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 201 tips and 198 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 NA, p-value = NA.
```

EXAMINE DATA SET

```
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 bifurcated tree would have 200 internal nodes (n_tips-1). Another check of whether the tree is fully bifurcating...

```
is.binary(darter.tree)
```

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

FULLY BIFURCATE TREE

```
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.
```

```
is.binary(darter.tree)
```

```
## [1] TRUE
```

Have now forced multifurcating nodes to bifurcate

STORE LTT

```
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.
```

PLOT FULLY BIFURCATED DATA SET AND OVERLAY TREE ONTO LTT PLOT

```
plot(obj,log.lineages=FALSE,main="LTT plot for darters")
plot(obj,show.tree=TRUE,log.lineages=FALSE,main="LTT plot for darters")
```
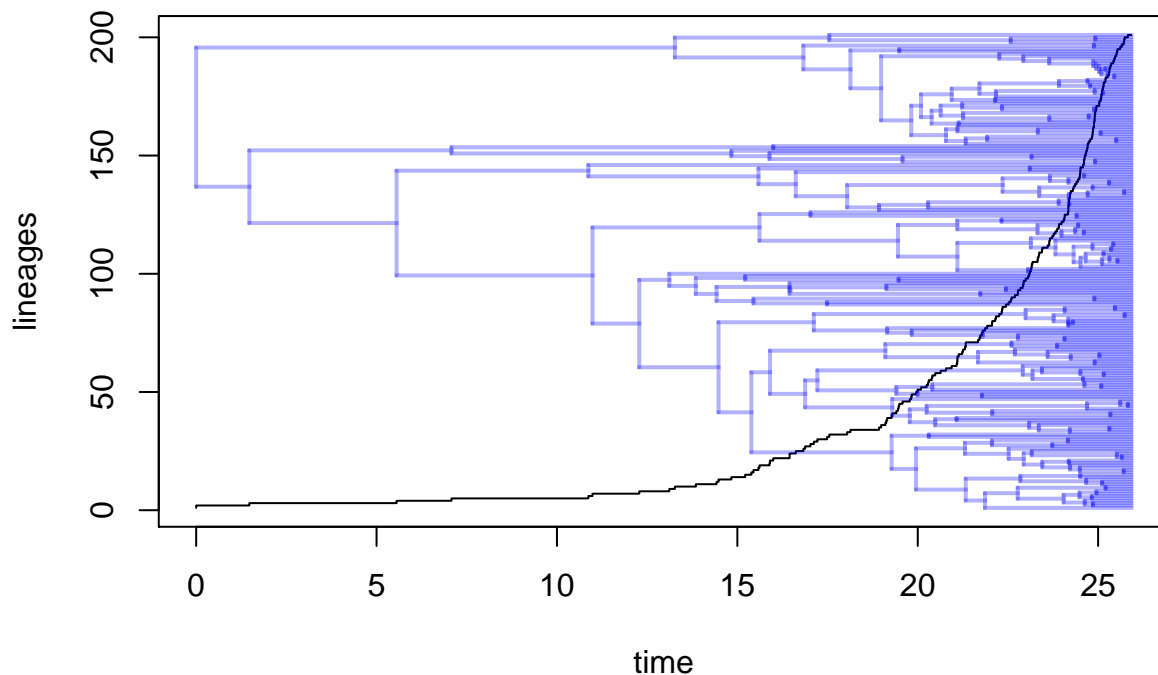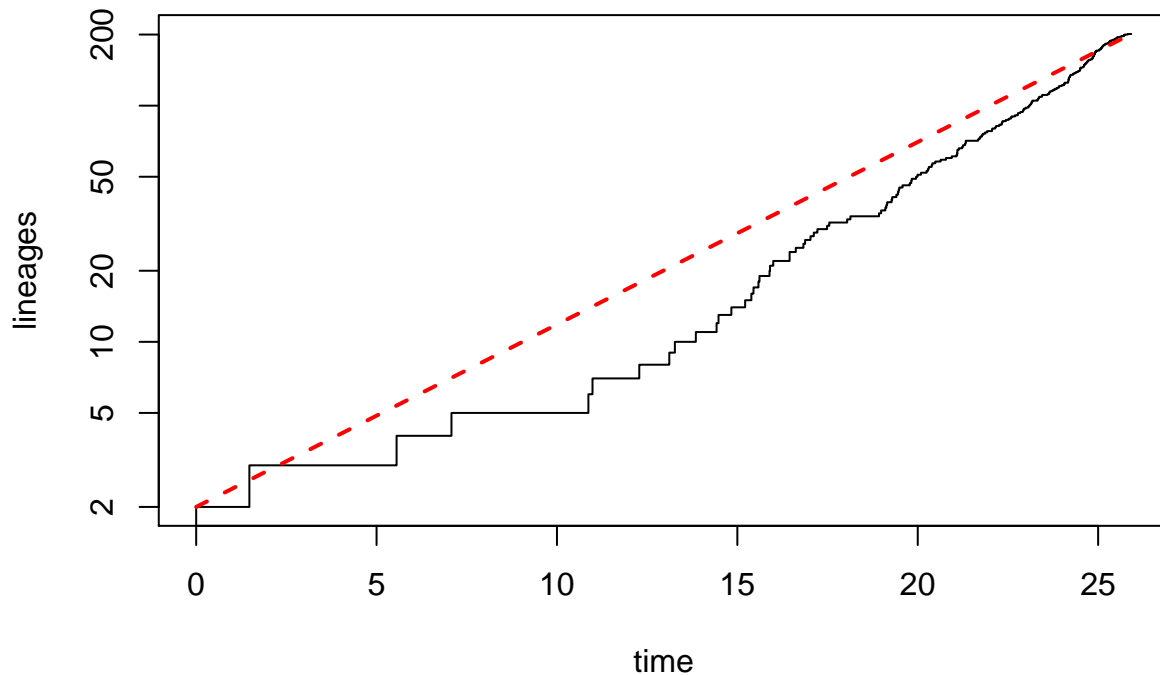
PLOT LOG OF LINEAGE NUMBER AGAINST TIME AND ADD LINE SHOWING PREDICTION UNDER PURE BIRTH MODEL

```
plot(obj,log.lineages=FALSE,log="y",main="LTT plot for darters",
     ylim=c(2,Ntip(darter.tree)))
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

SIMULATED LINEAGE NUMBERS THROUGH TIME WITH ORIGINAL DATA

```
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")
ltt(darter.tree,add=TRUE,lwd=2)
```
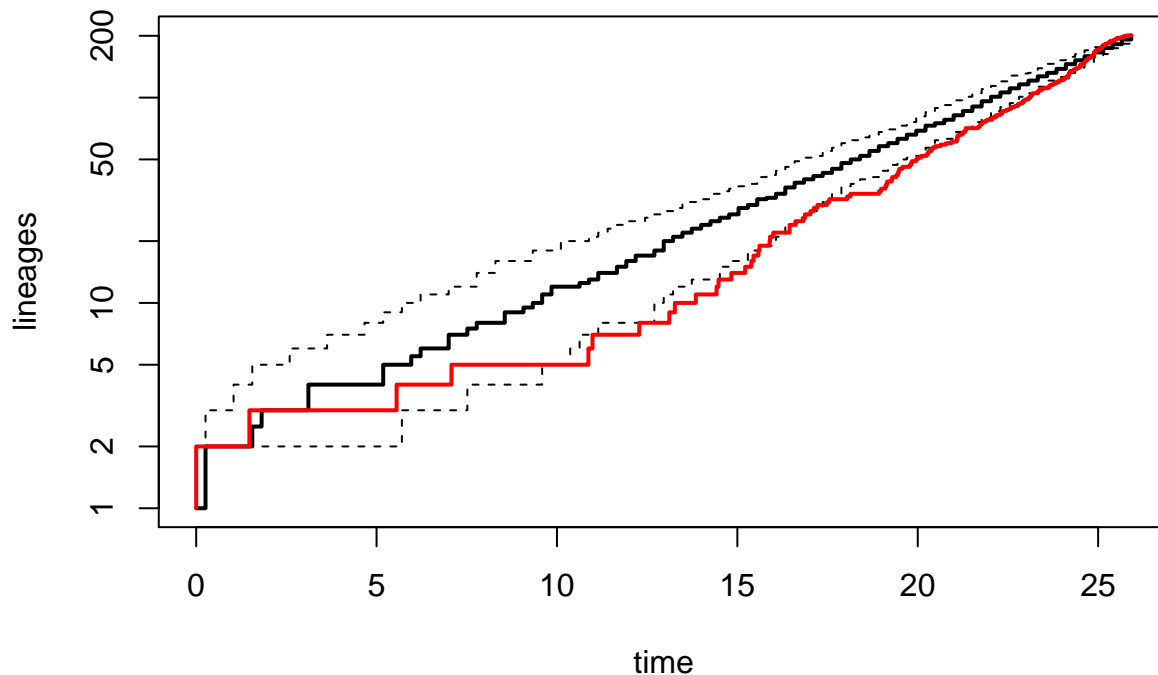
LTT of darters compared to simulated LTTs
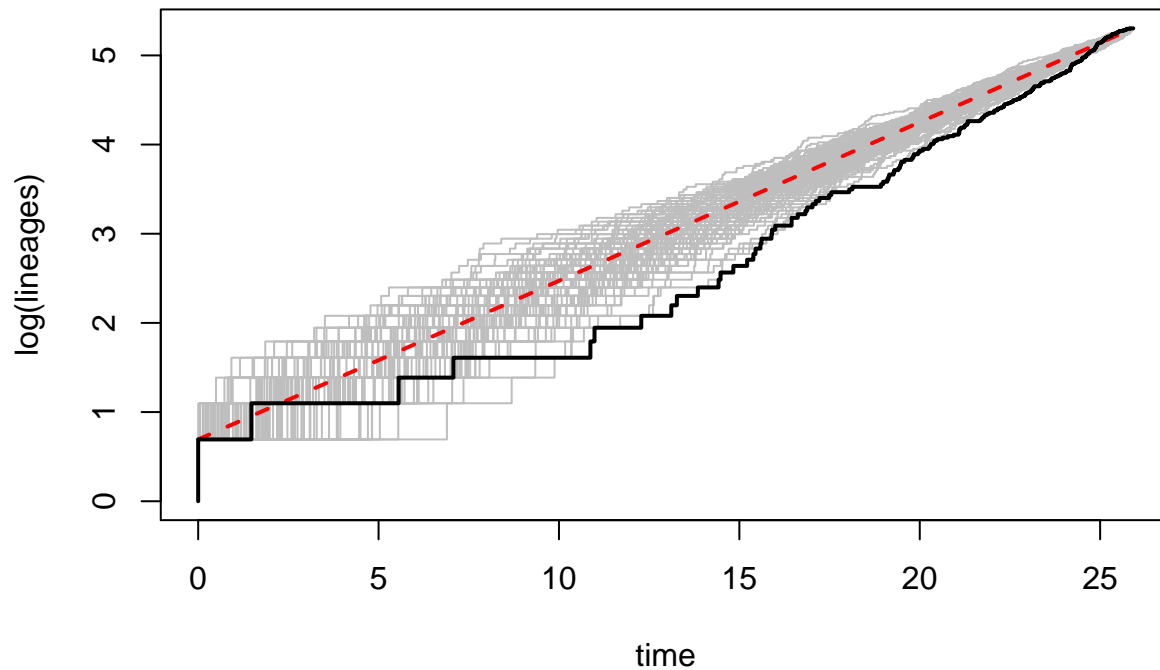

```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 201 tips and 200 internal nodes.
##
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.
##
## (3) A value for Pybus & Harvey's "gamma" statistic of 0.2007, p-value = 0.841.
```

95% CONFIDENCE INTERVAL FOR LTT FROM A SET OF TREES

```r
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.
```

Exercise 2

FIT BIRTH-DEATH MODEL AND RETURN RATES

```
fitbd <- birthdeath(darter.tree)
bd(fitbd)
```
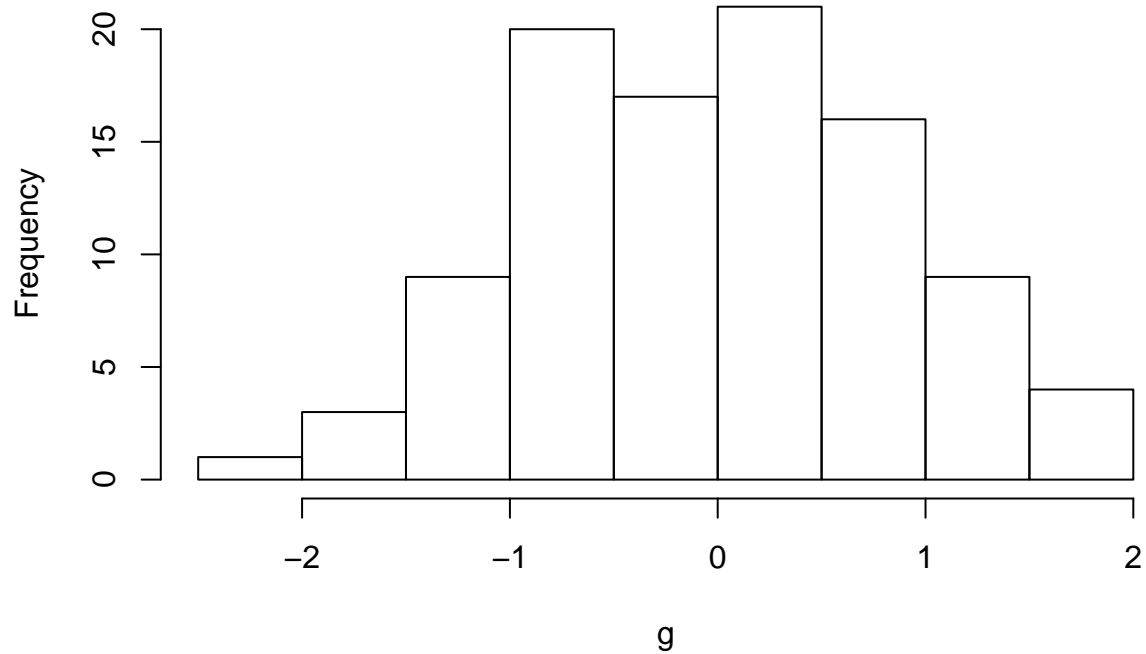
```
##          b          d
## 0.23619344 0.01495258
```

Exercise 3

PLOT HISTOGRAM OF SIMULATED GAMMA VALUES AND RETURN SUMMARY STATS

```
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


```r
mean(g)
```

```
## [1] -0.02340854
```

```r
var(g)
```

```
## [1] 0.7616468
```

TESTING HYPOTHESES

```r
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.
```
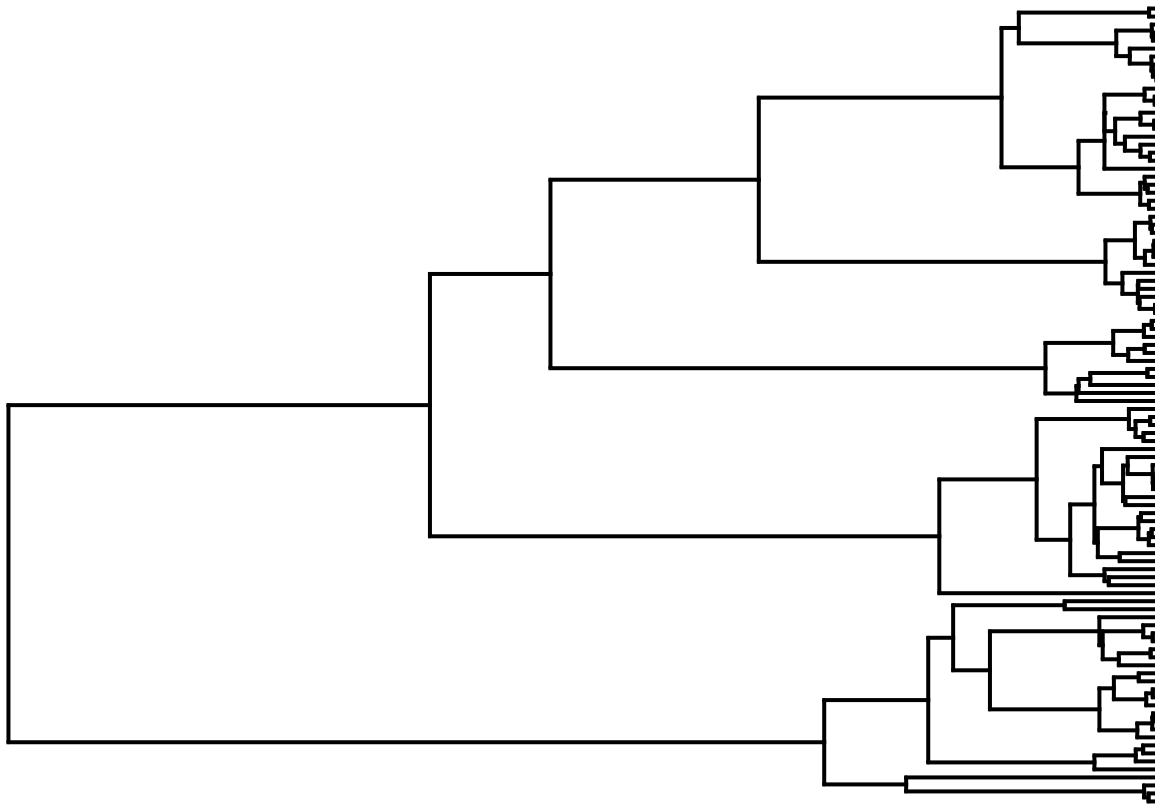
SIMULATING A COALESCENT TREE

```r
coal.tree<-rcoal(n=100)
plotTree(coal.tree,ftype="off")
```
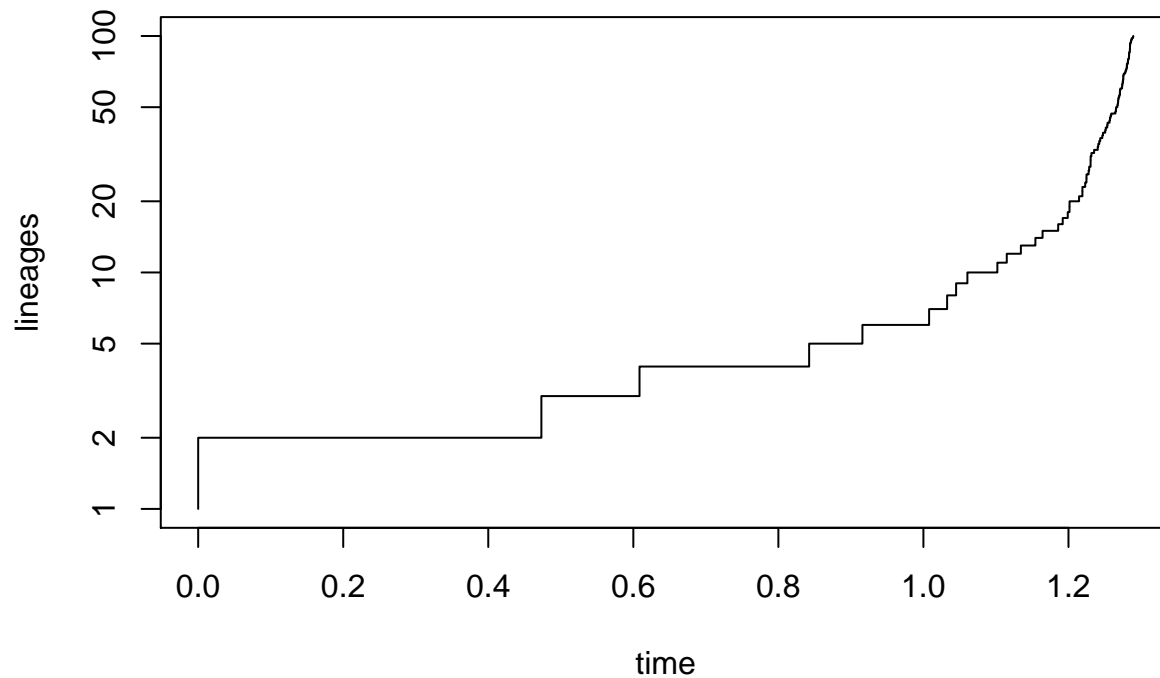
```
coal.obj<-ltt(coal.tree,plot=FALSE)
print(coal.obj)
```

```
## 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 8.851, p-value = 0.
```

PRINT COAL LTT PLOT

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

COMPARING GAMMA BETWEEN PURE BIRTH AND COALESCENT SIMULATED TREES

```
darter.gamma <- obj$gamma
darter.gamma
```

```
## [1] 8.851037
```

```
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
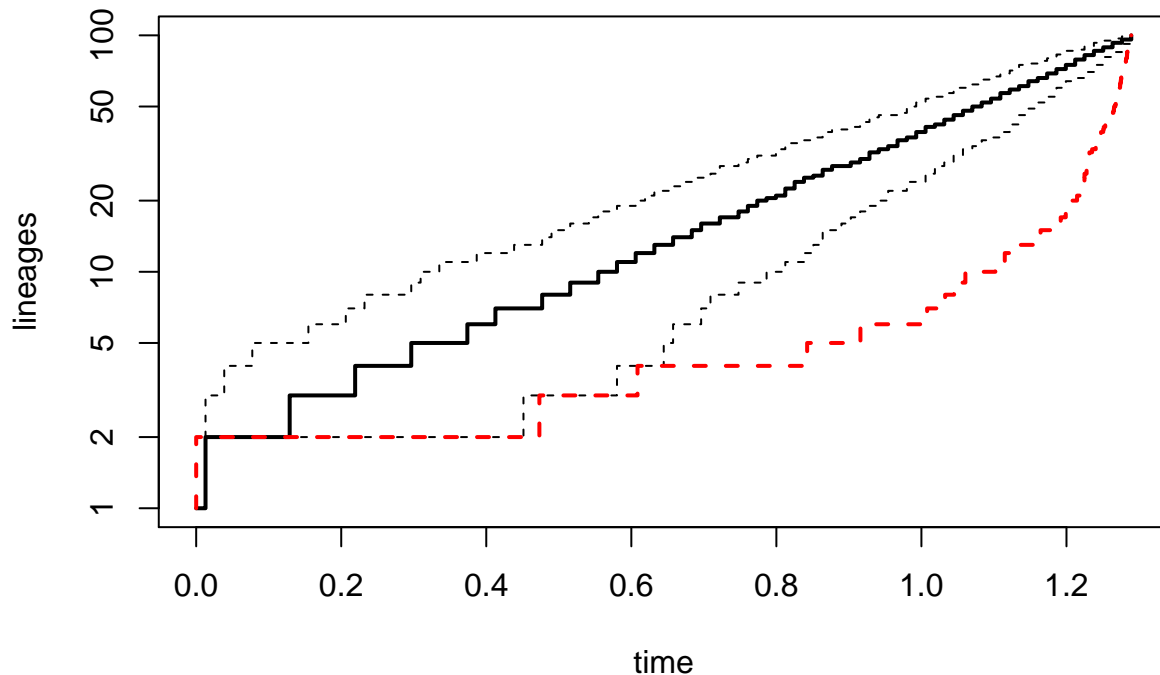

```
## Object of class "ltt" containing:
##
## (1) A phylogenetic tree with 100 tips and 99 internal nodes.
##
## (2) Vectors containing the number of lineages (ltt) and branching times (times) on the tree.
##
## (3) A value for Pybus & Harvey's "gamma" statistic of 8.851, p-value = 0.
```

The coalescent tree shows an exponential growth even when logged

Incomplete Sampling

Authors were missing 15 species Going to see what affect that has by generating 216 species and subsampling down to 201 and examining each gamma. If ours is extreme in comparison, we can conclude it's not caused by incomplete sampling

SIMULATE PURE BIRTH TREE

```
library(geiger)
age <- 25.91862
richness <- 216
darterbirth =  (log(richness) - log(2))/age ; darterbirth
```

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

SIMULATE INCOMPLETE SAMPLING

```
richness <- 216
missing <- 15
```

```
num_simulations<-200
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]
}
```
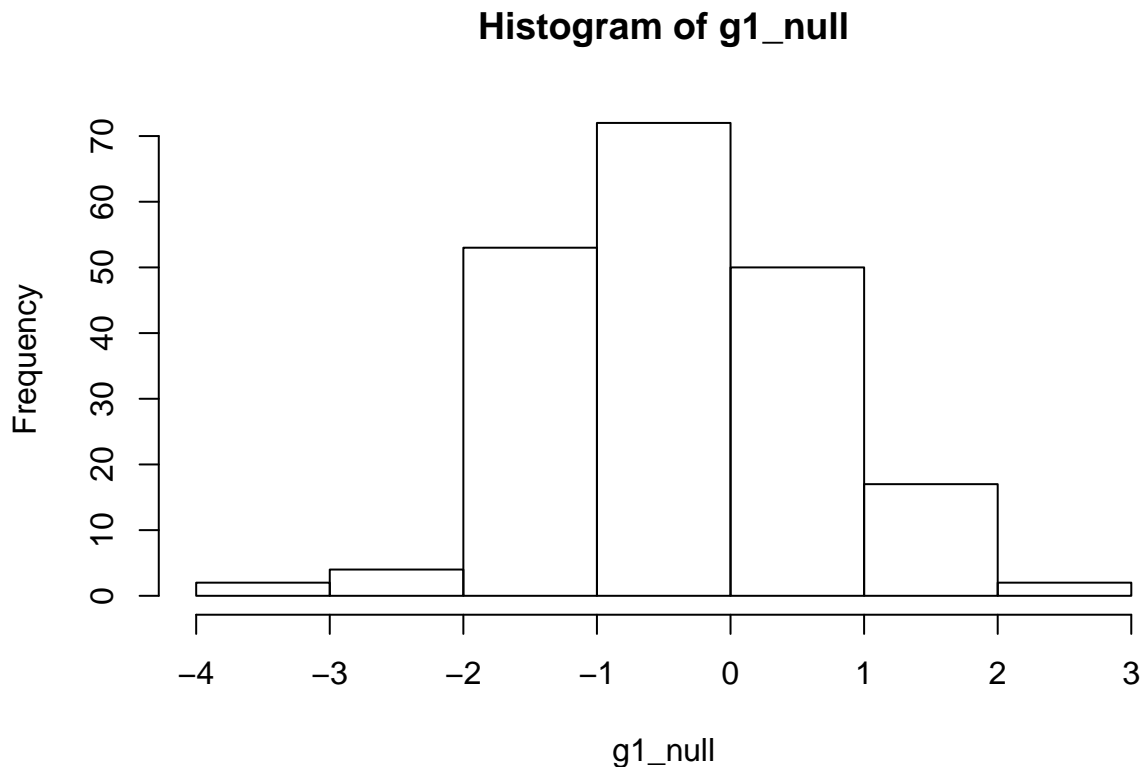
PLOT HISTOGRAM OF NULL DISTRIBUTION AND ADD ARROW TO SHOW OB-SERVED GAMMA

```
hist(g1_null)
arrows(darter.gamma, 40, darter.gamma, 0, col="red", lwd=2)
```


Histogram of g1_null

COMPARE OBSERVED GAMMA TO SIMULATED VALUES

```
smallerNull<-g1_null<=darter.gamma
count<-sum(smallerNull)
```

USE TO GENERATE P VALUE

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

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

The gamma value is definitely just due to subsampling