

Assignment 1

ASSIGNMENT

Packages

```
library(phytools)
```

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

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

```
library(geiger)
```

Question 1

Load data and use both tests to check the tree is fully bifurcated

```
snake.tree<-read.tree("homalops.phy") ; snake.tree
```

```
##
```

```
## Phylogenetic tree with 21 tips and 20 internal nodes.
```

```
##
```

```
## Tip labels:
```

```
## Enhydris_plumbea_A, Enhydris_matannensis, Lake_Towuti, Enhydris_chinensis, Enhydris_subtaeniata, En
```

```
##
```

```
## Rooted; includes branch lengths.
```

```
is.binary(snake.tree)
```

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

Run ltt

```
snake.obj<-ltt(snake.tree,plot = FALSE) ; snake.obj
```

```
## 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.
```

```
snake.gamma <- snake.obj$gamma ; snake.gamma
```

```
## [1] -3.241081
```

```
GAMMA: -3.2411
```

```
p-val: 0.0012
```

Question 2

The gamma value is conclusively negative which means there are more nodes/divergences earlier in the tree. This would be reason for further study to determine if an adaptive radiation event has occurred to explain

the rise of more lineages farther back in time than in the more recent past.

Question 3

Determine if the observed gamma is due to incomplete sampling.

Create a set of simulated lineages and see how tree gamma varies when 15 lineages are excluded each time and plot a histogram of the resulting values.

```
age <- 22
richness <- 34
missing <- 13

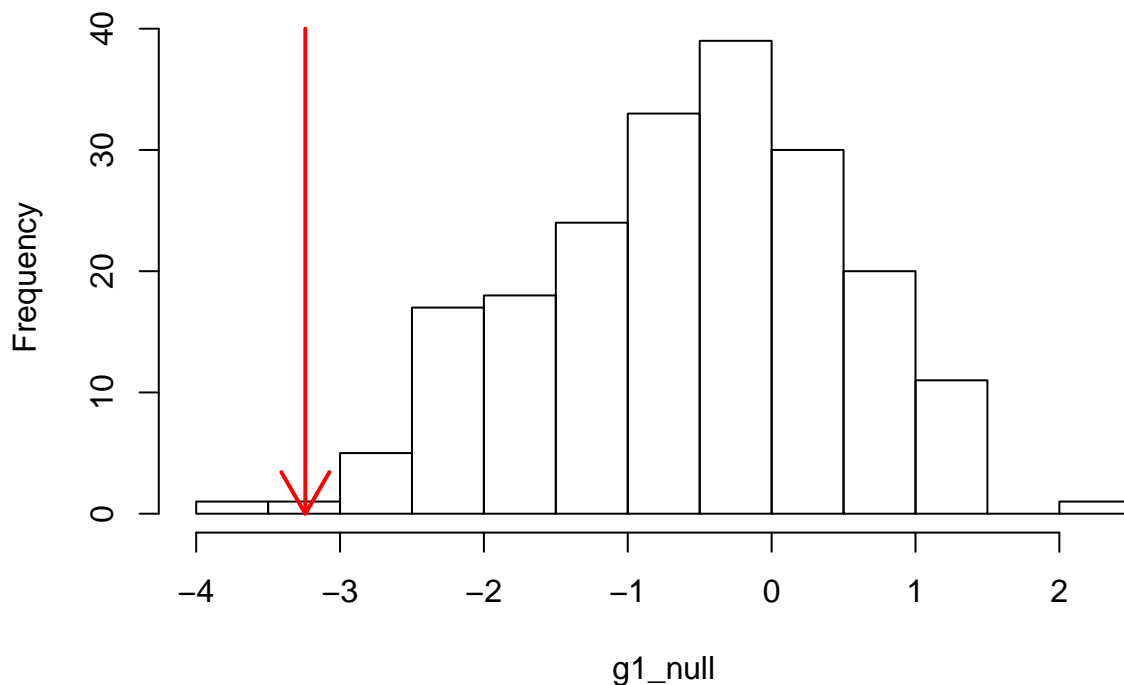
snakebirth = (log(richness) - log(2))/age ; snakebirth

## [1] 0.1287824
num_simulations<-200

# For loop to simulate populations
# COPIED FROM MATERIAL PROVIDED BY PROF. MIKE ALFARO
g1_null<-numeric(num_simulations) #g1_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)->g1_null[i]
}

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

Histogram of g1_null



The arrow indicating the observed gamma value seems to be well into the tails and therefore possibly significant. We will now determine the p value.

```
smallerNull<-g1_null<=snake.gamma
count<-sum(smallerNull)
mccr_pval<-(count+1)/(num_simulations+1) ; mccr_pval
```

```
## [1] 0.009950249
```

```
p-val: 0.00995
```

On the basis of the MCCR test, I conclude the rate of speciation slowing as time increases.

Question 4

Fit birth-death model and report rates:

```
fitbd <- birthdeath(snake.tree)
bd(fitbd)
```

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

Birth Rate: 0.0684 Death Rate: 0.0000 (The species are all still extant so none have gone extinct)

Question 5

(1) Clade description and source

Number of Tips: 50

Number of Species in Clades: 54

Mustelidae data set downloaded from:

<http://10ktrees.nunn-lab.org/Carnivora/downloadTrees.php>

Load tree and check it is fully bifurcated.

```
prac.tree<-read.nexus("consensusTree_10kTrees_Carnivora_Version1.nex") ; prac.tree
```

```
##
## Phylogenetic tree with 46 tips and 45 internal nodes.
##
## Tip labels:
## Aonyx_capensis, Aonyx_cinerea, Arctonyx_collaris, Eira_barbara, Enhydra_lutris, Galictis_cuja, ...
##
## Rooted; includes branch lengths.
```

```
is.binary(prac.tree)
```

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

(2) Fit birth death model and return b and d

```
fitbd <- birthdeath(prac.tree)
bd(fitbd)
```

```
##           b           d
## 0.07690858 0.00000000
```

Birth Rate: 0.0769
Death Rate: 0.0000
Again, all species are extant

(3) Perform MCCR test and analyse

Run ltt and get observed gamma

```
prac.obj<-ltt(prac.tree,plot = FALSE) ; prac.obj
```

```
## Object of class "ltt" containing:
```

```
##
```

```
## (1) A phylogenetic tree with 46 tips and 45 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 -1.8281, p-value = 0.0675.
```

```
prac.gamma <- prac.obj$gamma ; prac.gamma
```

```
## [1] -1.828136
```

Simulate incomplete sampling

```
age <- 15
```

```
richness <- 54
```

```
pracbirth = (log(richness) - log(2))/age ; pracbirth
```

```
## [1] 0.2197225
```

```
missing <- 4
```

```
num_simulations<-200
```

```
# For loop to simulate populations
```

```
# COPIED FROM MATERIAL PROVIDED BY PROF. MIKE ALFARO
```

```
g1_null<-numeric(num_simulations) #g1_null will hold the simulated gamma values
```

```
for(i in 1:num_simulations) {
```

```
  sim.bdtree(pracbirth, d=0, stop = "taxa", n=richness)->sim_tree
```

```
  drop.random(sim_tree, missing)->prune # prune down to the # of taxa in the phylogeny
```

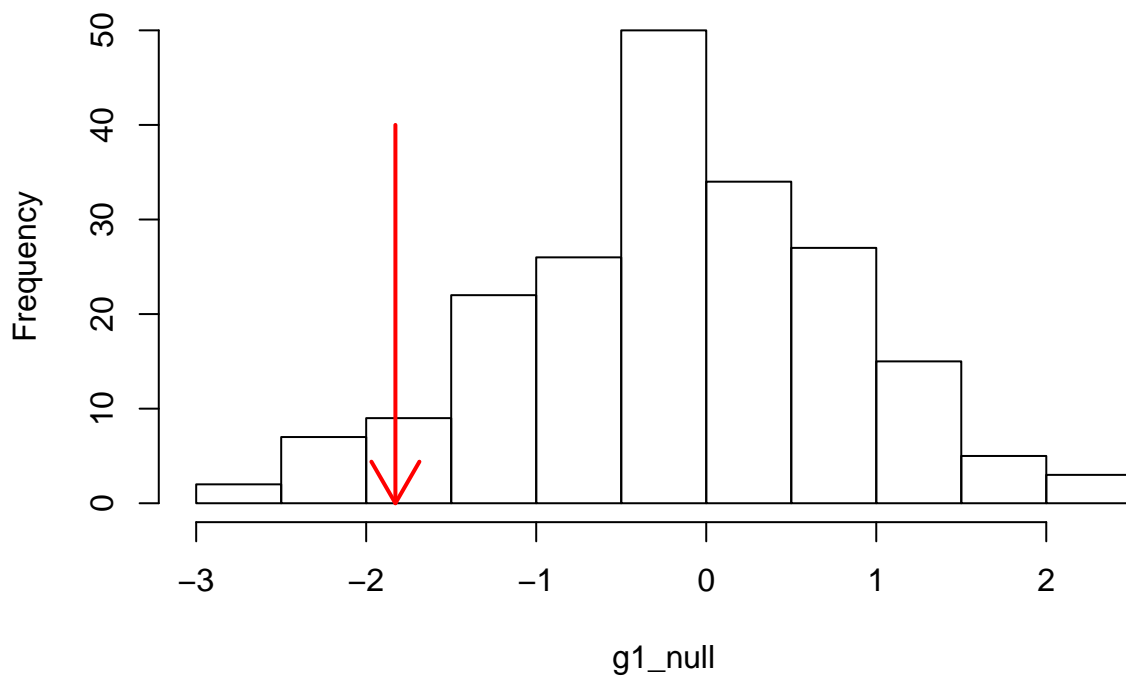
```
  gammaStat(prune)->g1_null[i]
```

```
}
```

```
hist(g1_null)
```

```
arrows(prac.gamma, 40, prac.gamma, 0, col="red", lwd=2)
```

Histogram of g1_null



Seems to be somewhat into the tails. Looks possibly significant.

Generate p-value of observed gamma

```
smallerNull<-g1_null<=prac.gamma  
count<-sum(smallerNull)  
mccr_pval<-(count+1)/(num_simulations+1) ; mccr_pval
```

```
## [1] 0.05970149
```

p-val: 0.0448

Interpretation:

The gamma value does seem to be significant compared to simulations suggesting a slight skew of the nodes towards the beginning of the tree which also suggests the rate of speciation declined over time.