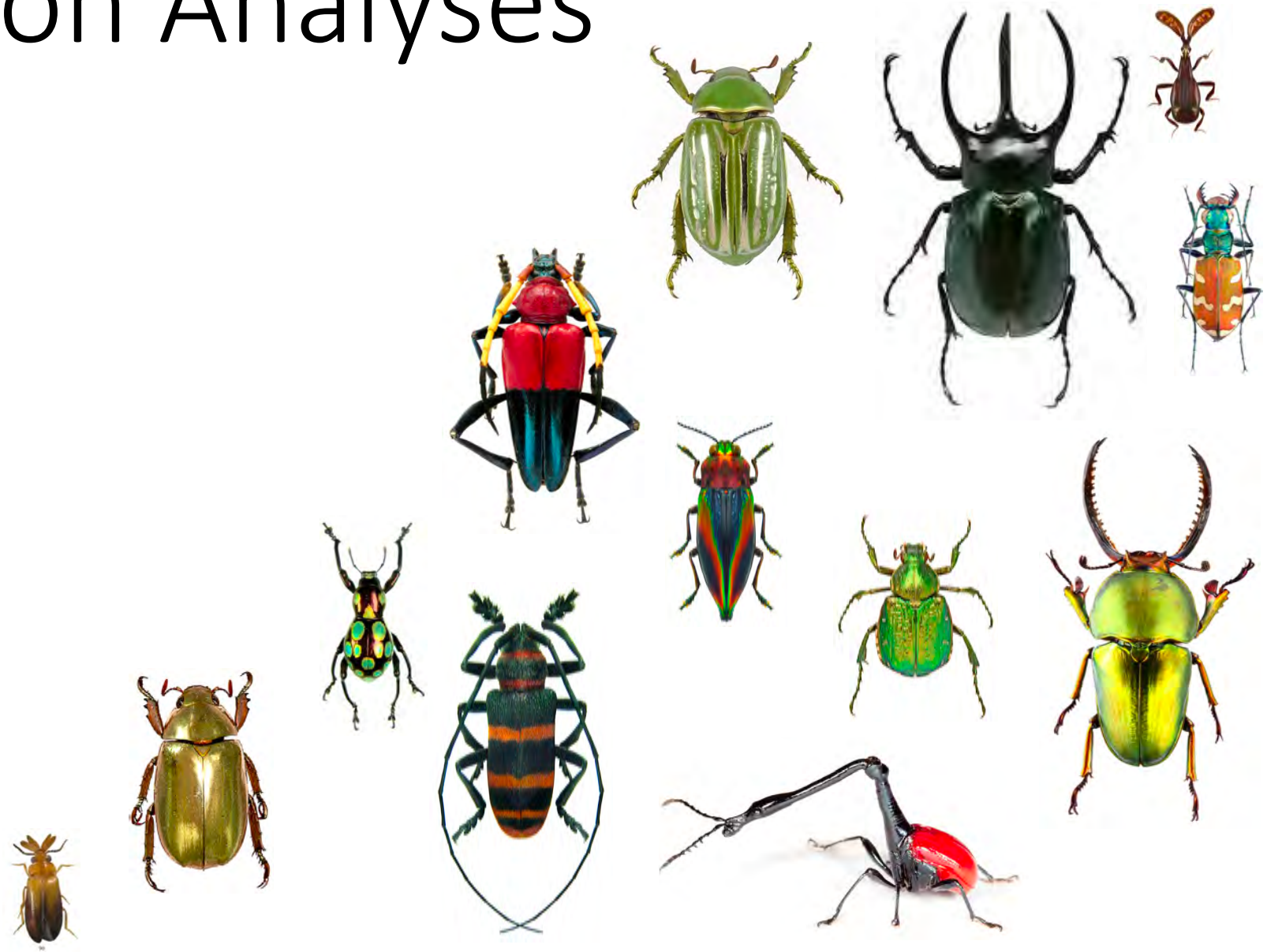
Diversification Analyses

EEBL 606

21 Feb 2018

Heath Blackmon

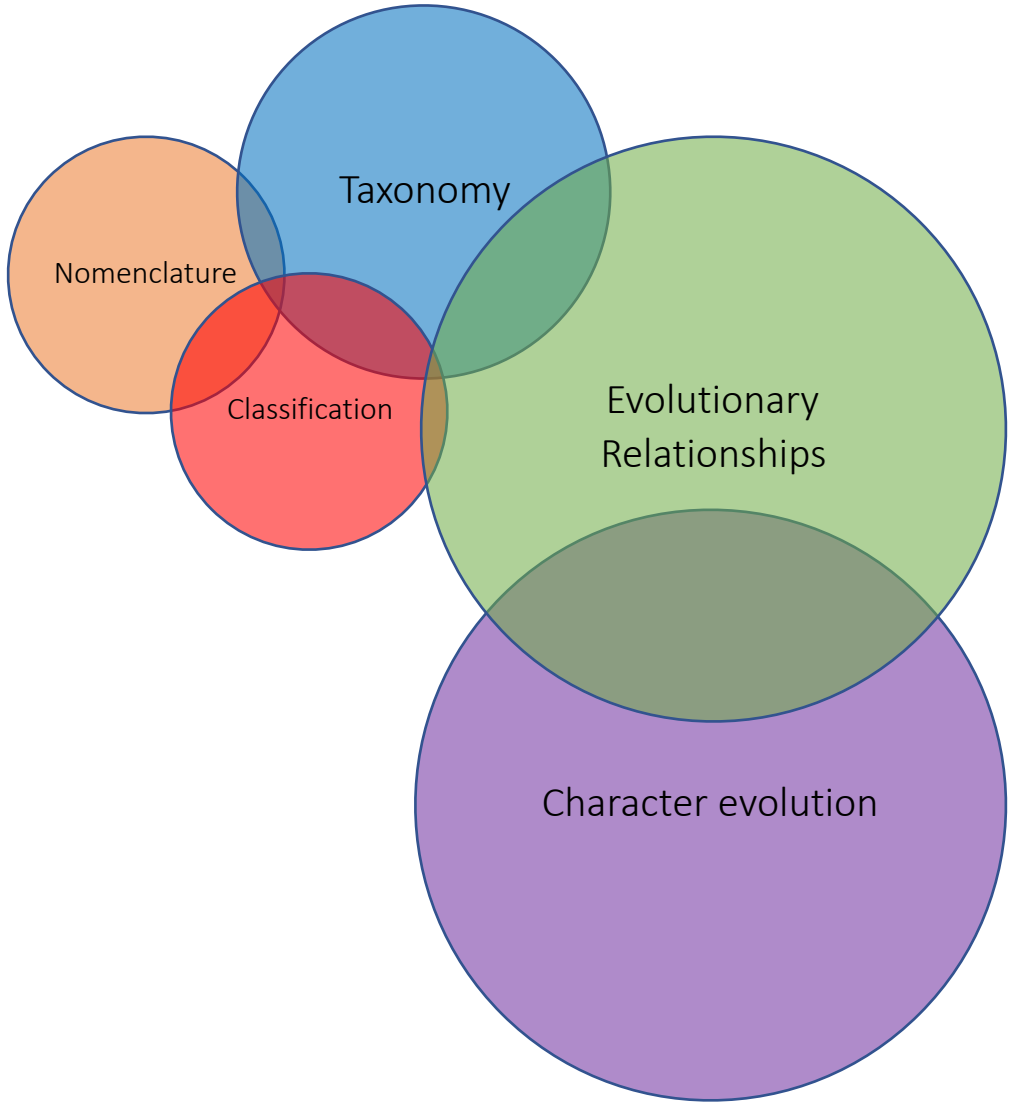


Previously

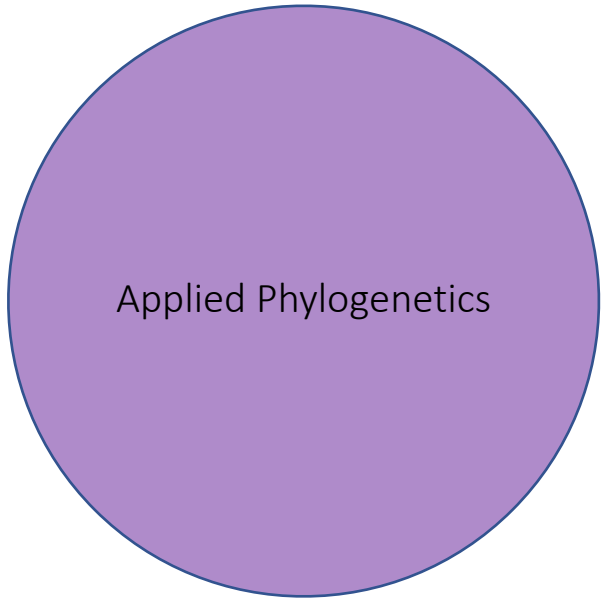
Dr. Conway: Why are we interested in phylogenies and what are the philosophical and logical premises that we base our work on.

Dr. Song: How do we go about making trees in the real world. What programs do we use what challenges do we face.

Systematics in 21st C?



Today: Applied phylogenetics



Character evolution
- Morphology
- Genes/Networks

Diversification analysis

Biogeographic analysis

Common Questions

1. Is group A diversifying more quickly than group B?

2. Is state 1 of a trait associated with faster diversification than state 2?

3. Is clade A an example of an adaptive radiation?

Terminology

1. Diversification: increase in the number of lineages through time - richness

2. Disparification: increase in the range of trait values in a clade through time - variance

3. Adaptive Radiation: diversification facilitated by new adaptive traits

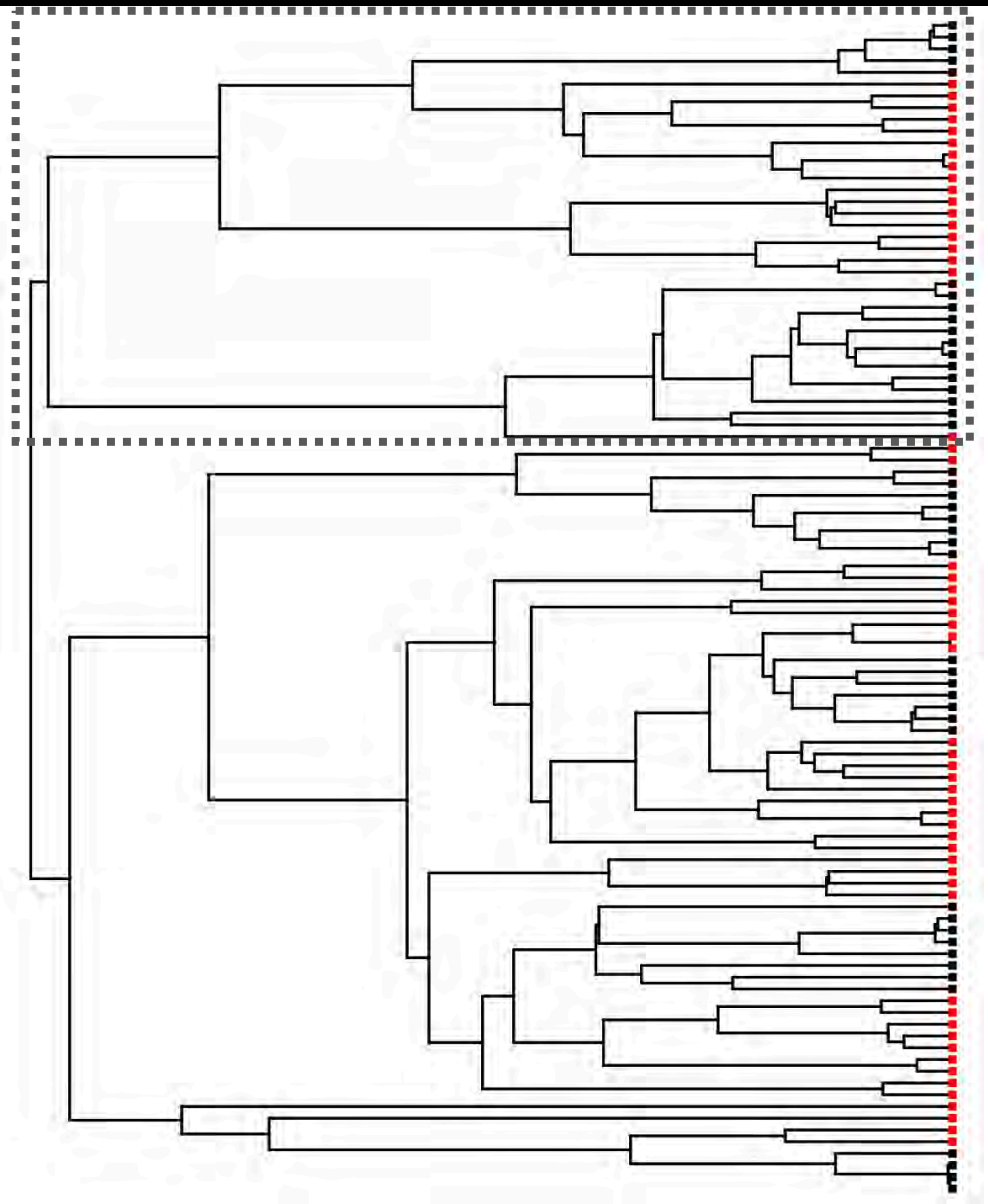
Methods

1. Sister clade contrasts

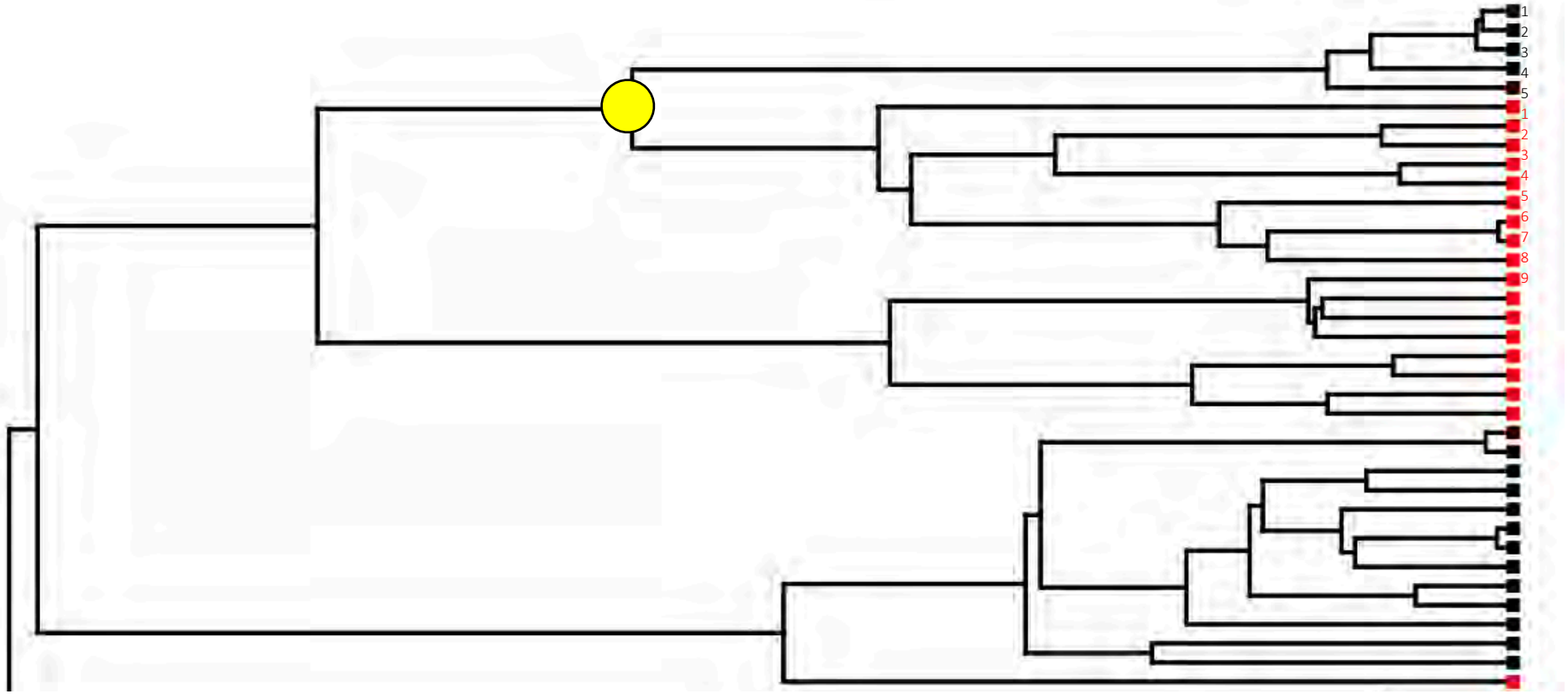
2. Explicit model based approaches (BiSSE, BAMM, etc.)
   - The controversy over BiSSE/BAMM

3. LTT

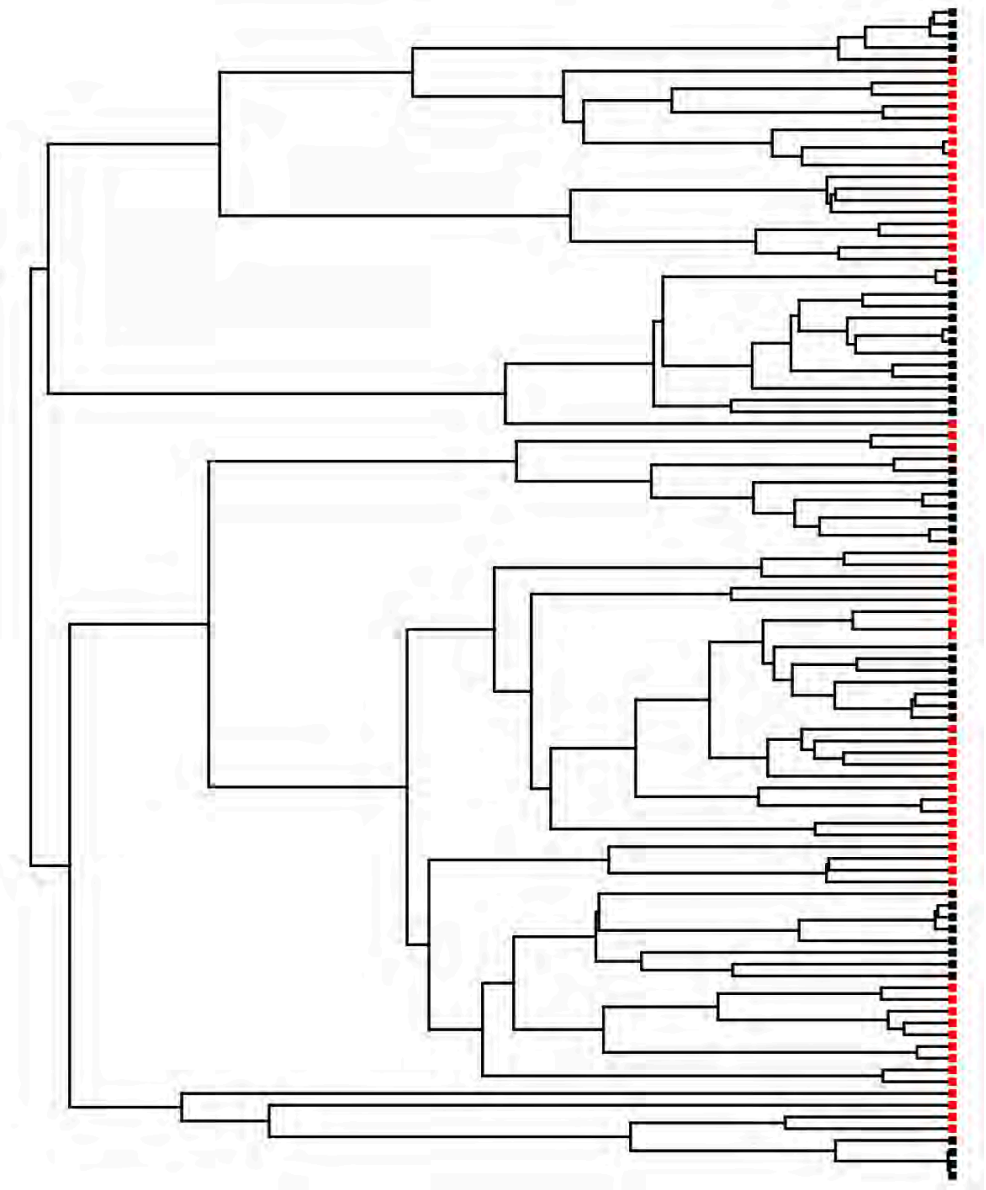
Sister clade contrasts



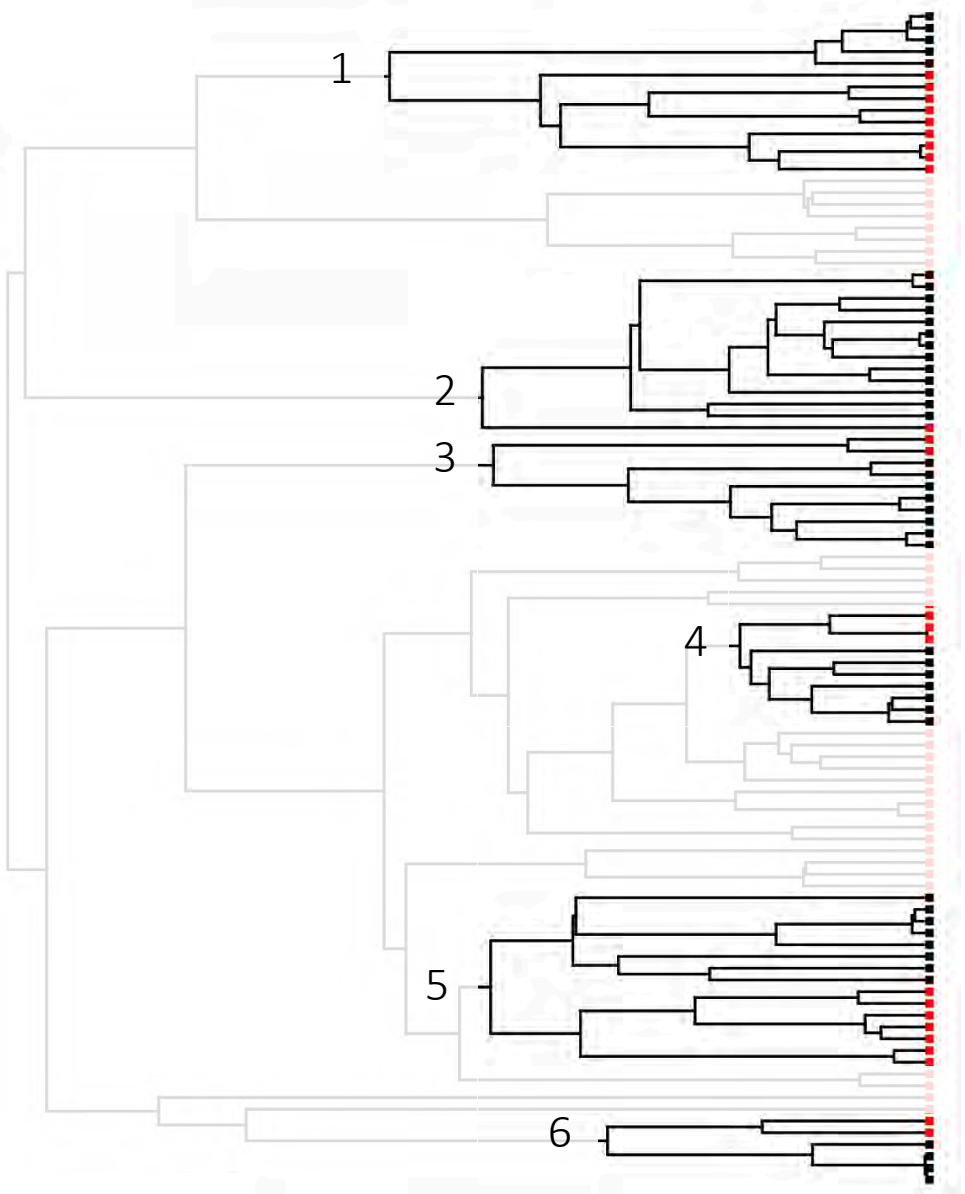
Sister clade contrasts



Sister clade contrasts



Sister clade contrasts



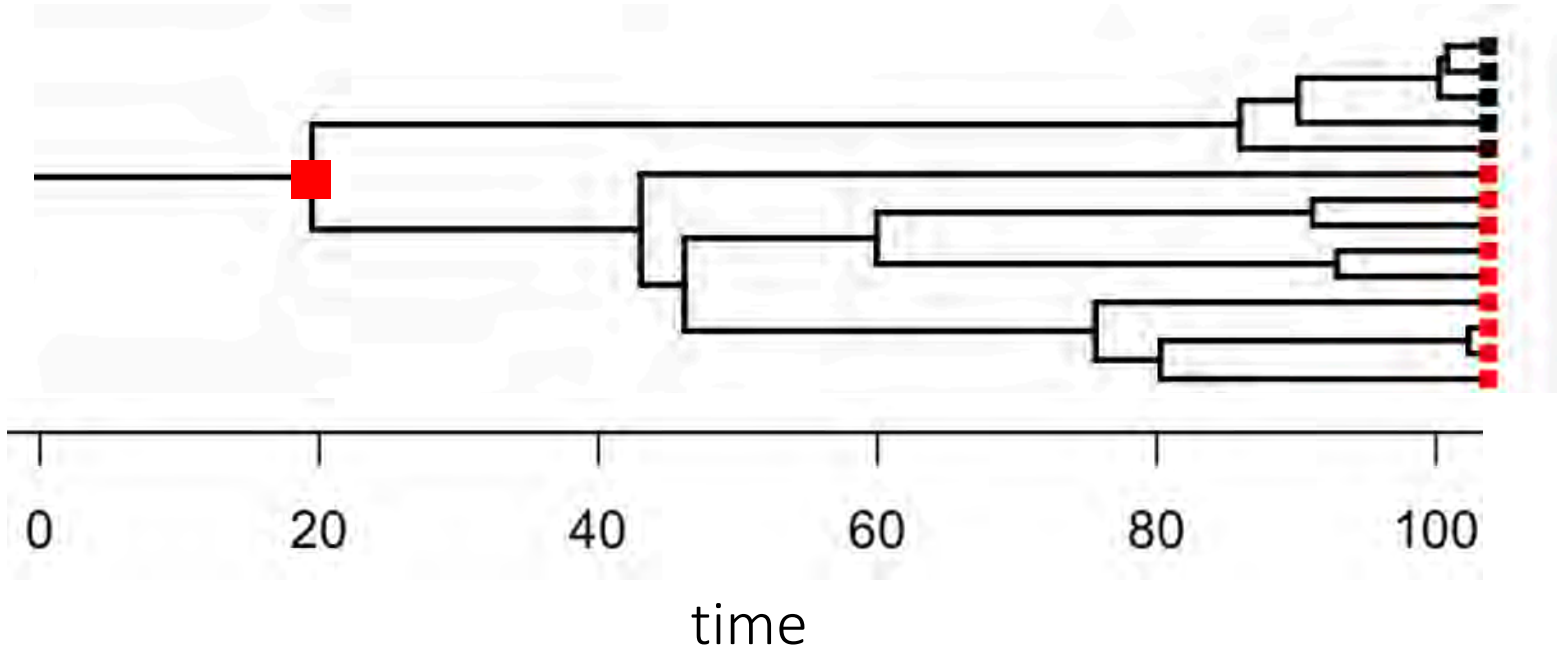
node	Red species	Black species
117	9	5
238	1	13
164	2	8
271	3	7
105	7	8
129	2	4

= cont.x

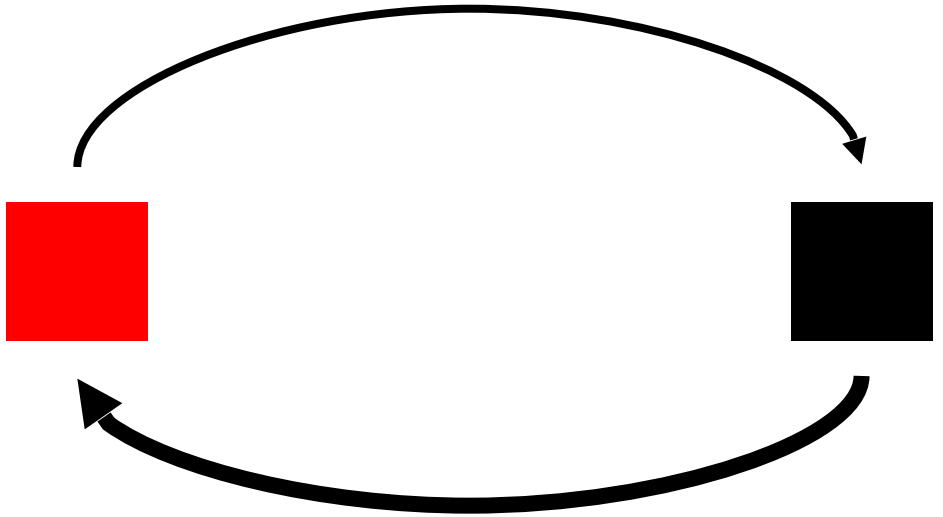
```
> diversity.contrast.test(x = cont.x,
+                         alternative = "less")
[1] 0.046875
```

- General rule need around 5 comparisons to get P < 0.05
- Use R package APE or beware that some old methods still implemented

Problems with sister clade contrasts



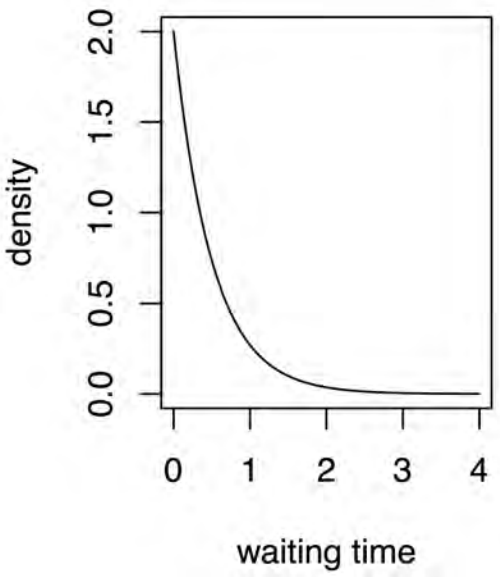
Problems with sister clade contrasts



Explicit model based approach

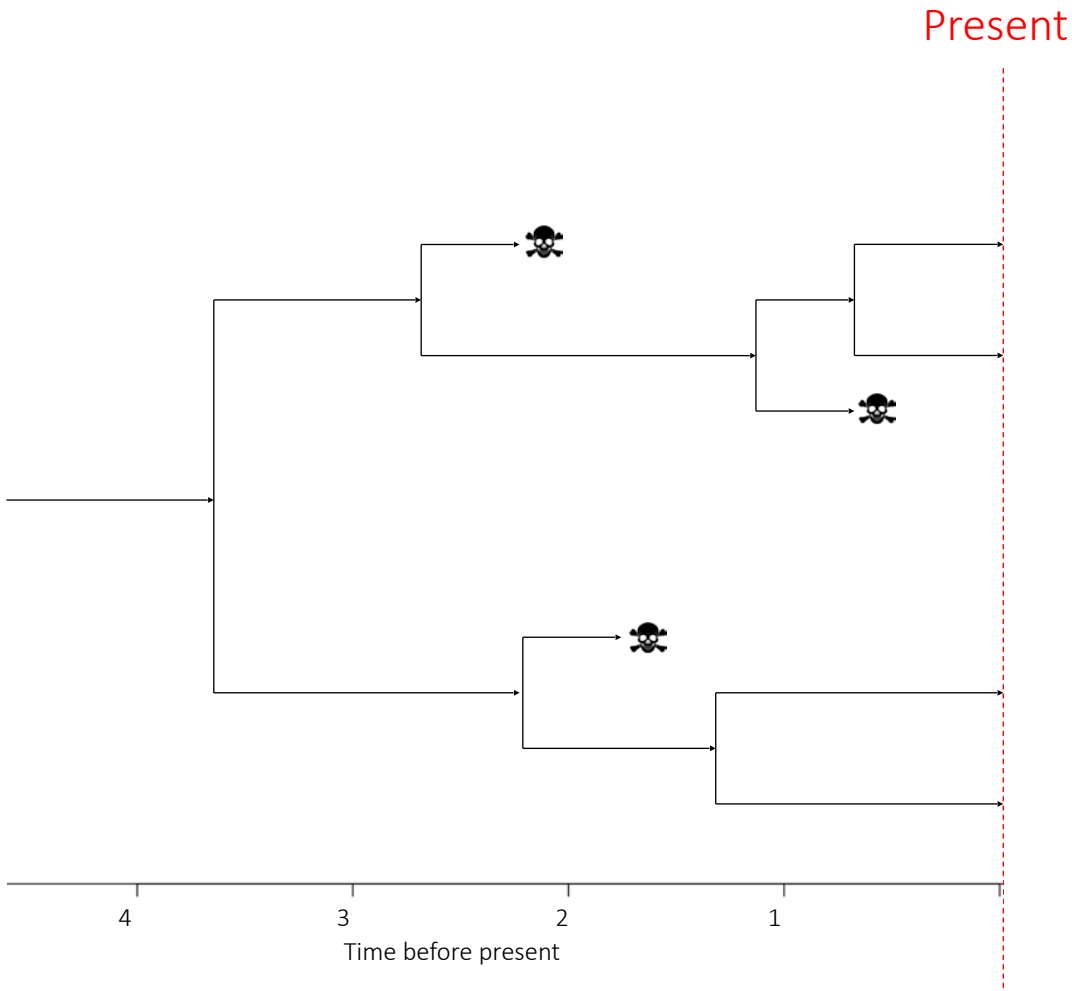
$\lambda = speciation$

$\mu = extinction$



Exponential distribution with rate parameter equal to $\lambda + \mu$

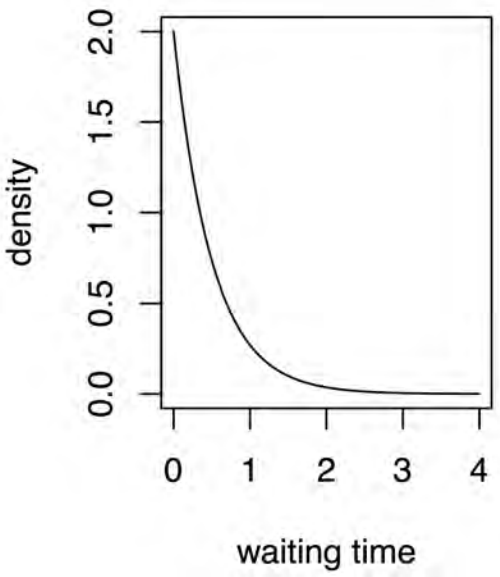
Probability that it is a speciation event is $\frac{\lambda}{\lambda+\mu}$ otherwise extinction



Explicit model based approach

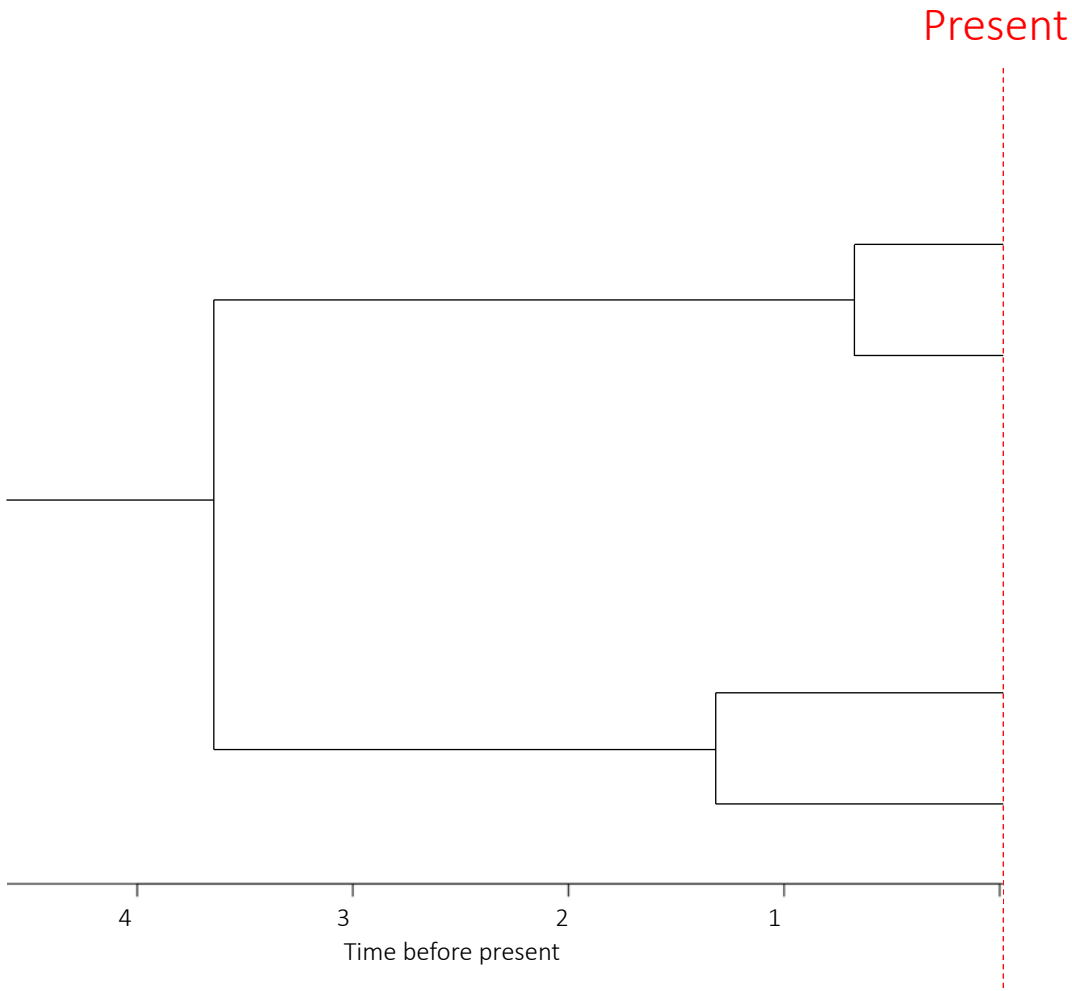
$\lambda = speciation$

$\mu = extinction$



Exponential distribution with rate parameter equal to $\lambda + \mu$

Probability that it is a speciation event is $\frac{\lambda}{\lambda+\mu}$ otherwise extinction

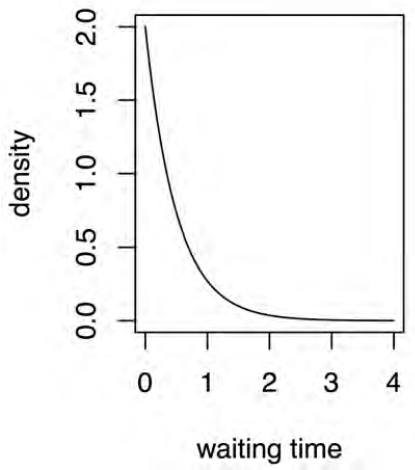


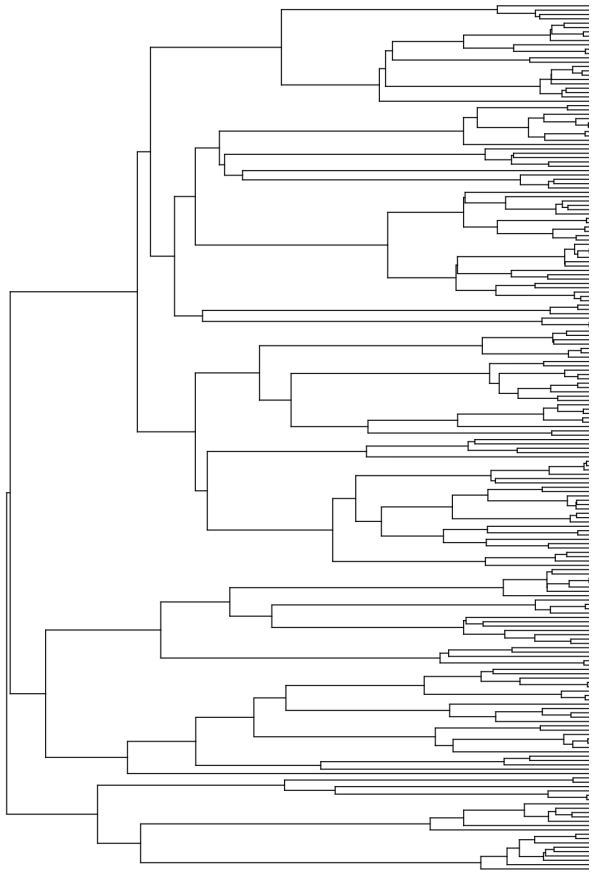
Explicit model based approach

$\lambda = speciation$

$\mu = extinction$

```
set.seed(1)
tree1 <- trees(pars=c(.2,.1), max.taxa=200, type="bd")[[1]]
```



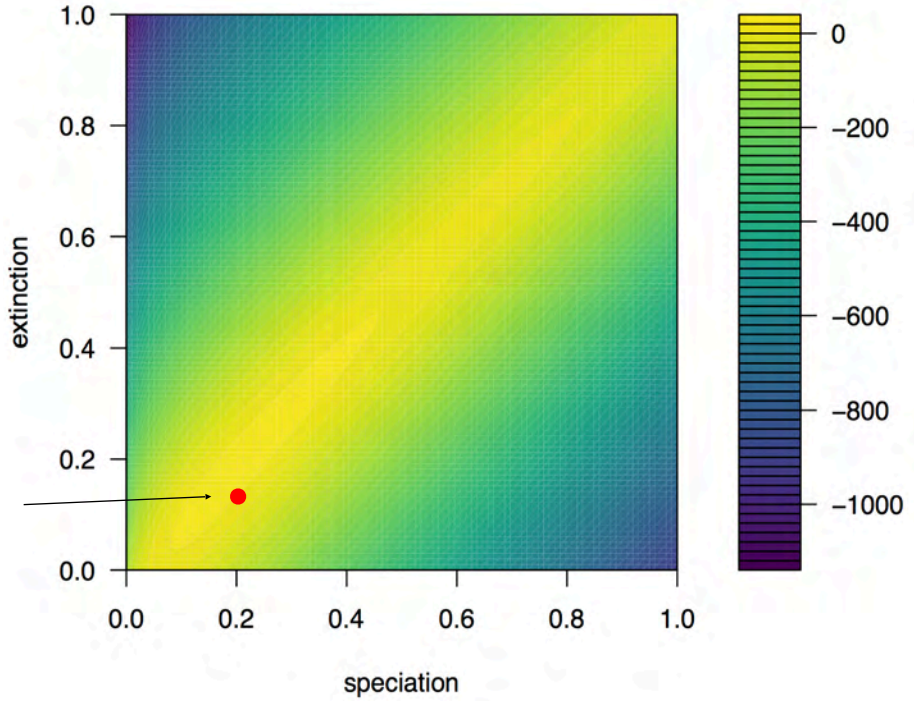


Exponential distribution with rate parameter equal to $\lambda + \mu$

Probability that it is a speciation event is $\frac{\lambda}{\lambda+\mu}$ otherwise extinction

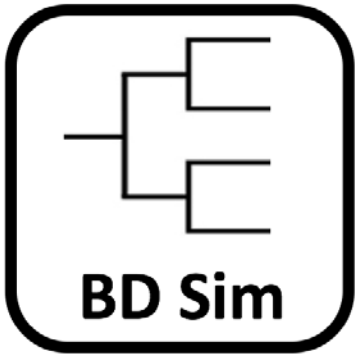
```
lik <- make.bd(tree1)
l <- seq(from = 0.01, to = 1, length.out=100)
m <- seq(from = 0.01, to = 1, length.out=100)
results <- matrix(,100,100)
for(i in 1:100){
  for(j in 1:100){
    results[i,j] <- lik(pars=c(l[i],m[j]))
  }
}
library(viridis)
vir.pal <- viridis(60)
filled.contour(results,
               ylab = "extinction",
               xlab = "speciation",
               nlevels = 50, col = vir.pal)
```

Pr(tree|model (λ, μ))

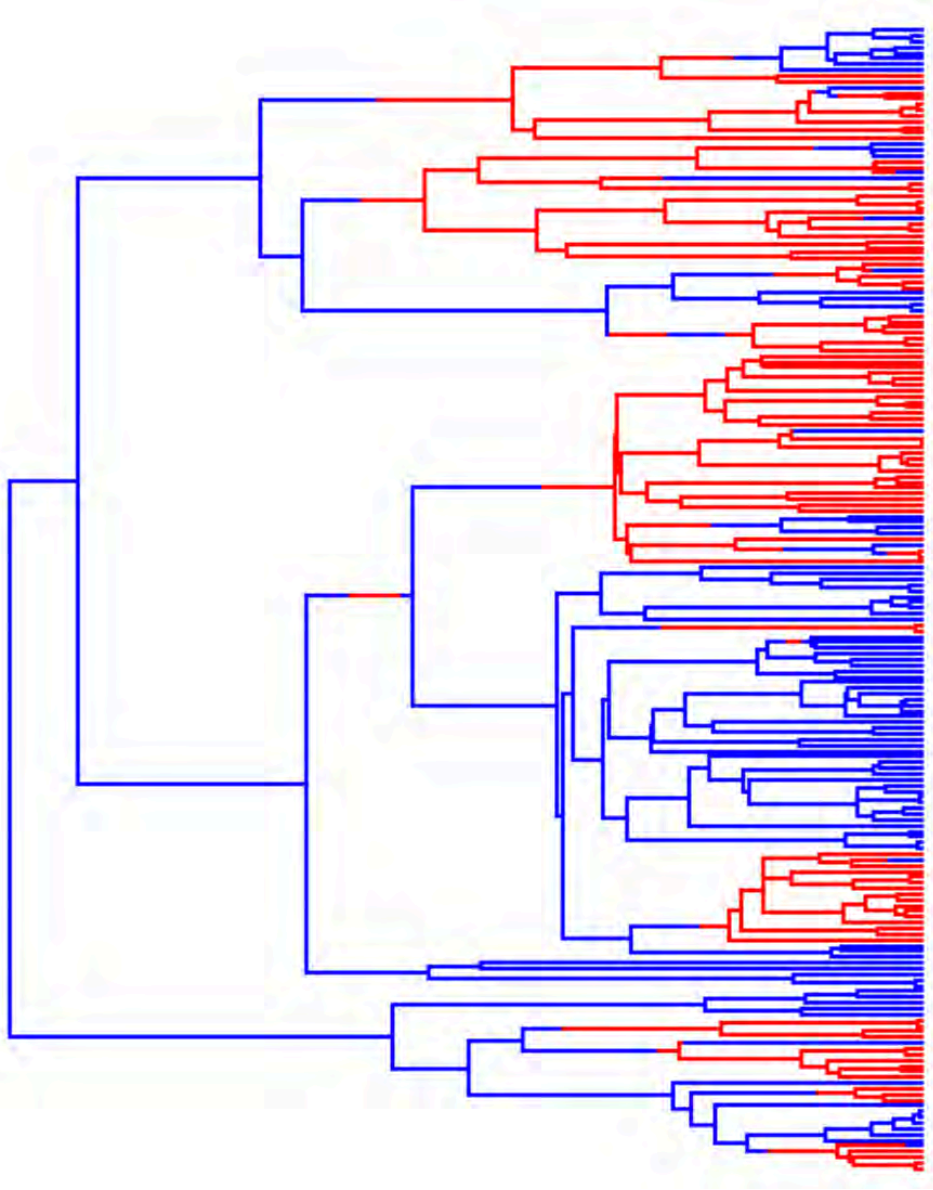


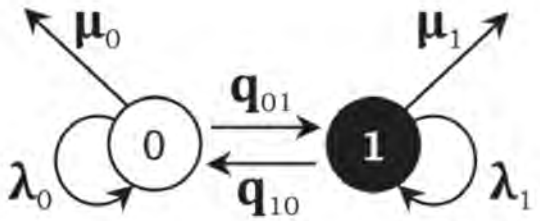
True value

High Variance in species under the BD model



BiSSE





SSE = state dependent
speciation and extinction

This method has gone on to spur the development of many extensions

BiSSE
MuSSE
QuaSSE
GeoSSE
HiSSE

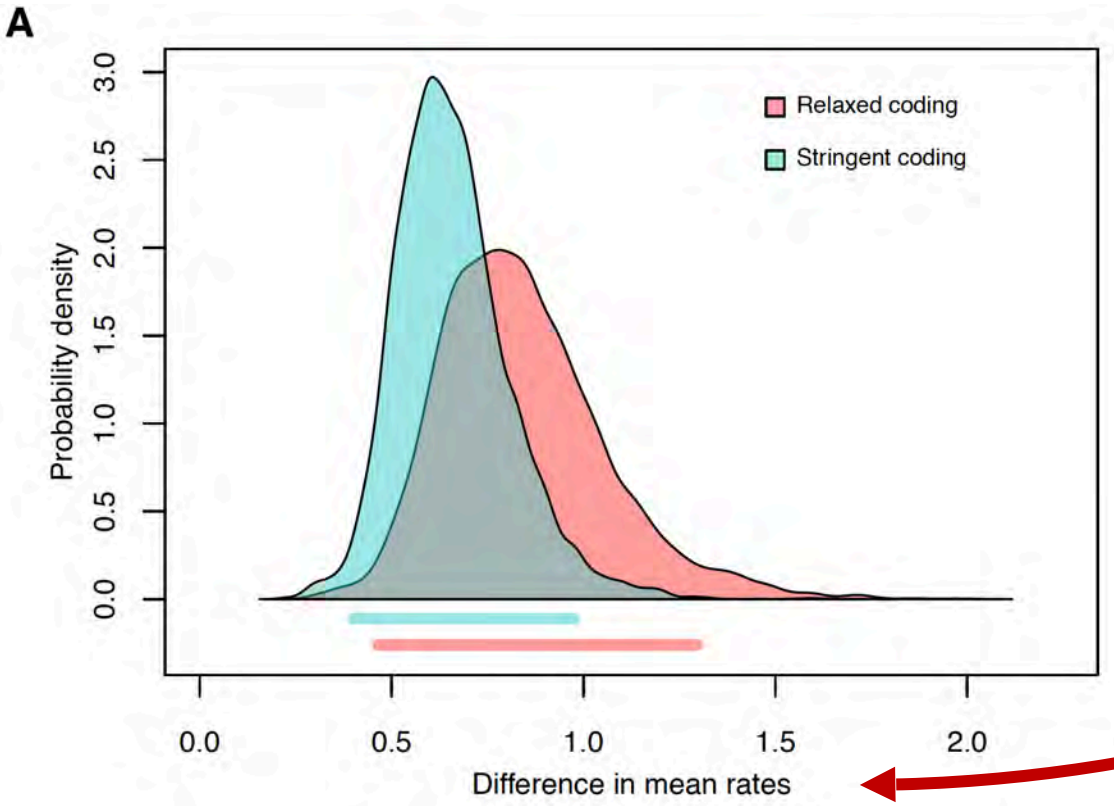
BiSSE

Despite the desire people have to use BiSSE(and extensions) it has some problems.

- pseudoreplication
- many/most empirical trees contain the signature of rate variation

Rabosky and Goldberg 2015

Bayesian refresher



$$Pr(A|B) = \frac{Pr(B|A) \times Pr(A)}{Pr(B)}$$

Bayesian refresher

Pr(B|A) this is just the likelihood of our data given our model and current parameters

Pr(A) is the prior this is a distribution that described the probabilities of different values of A

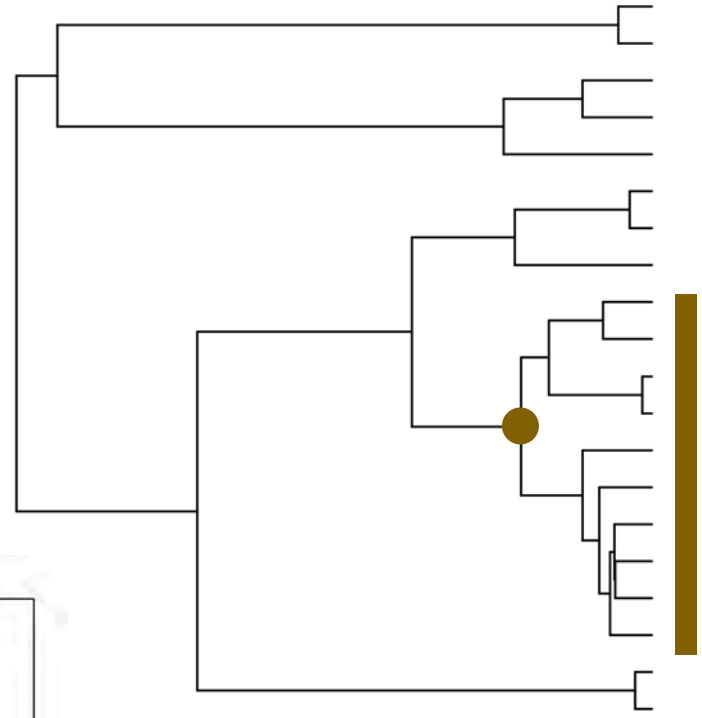
$$Pr(A|B) = \frac{Pr(B|A) \times Pr(A)}{Pr(B)}$$

A is the thing we are interested in (phylogeny, age of a node, number of rate shifts, extinction rate)

Pr(B) this is just the likelihood of our data we don't actually deal with this unless you are developing the software

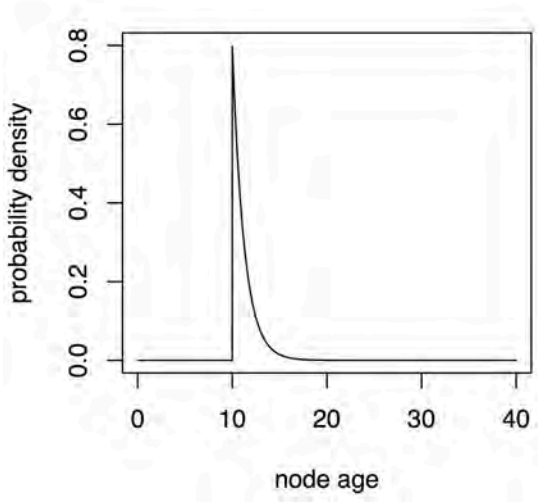
Bayesian refresher – age of a node

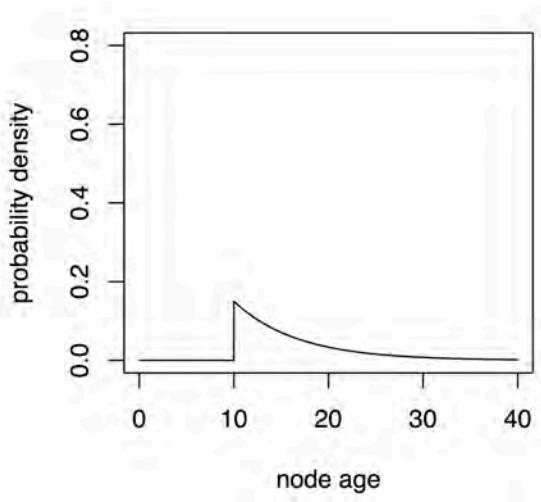
$$Pr(A|B) = \frac{Pr(B|A) \times Pr(A)}{Pr(B)}$$



Age estimate
10 MYA

Two possible priors

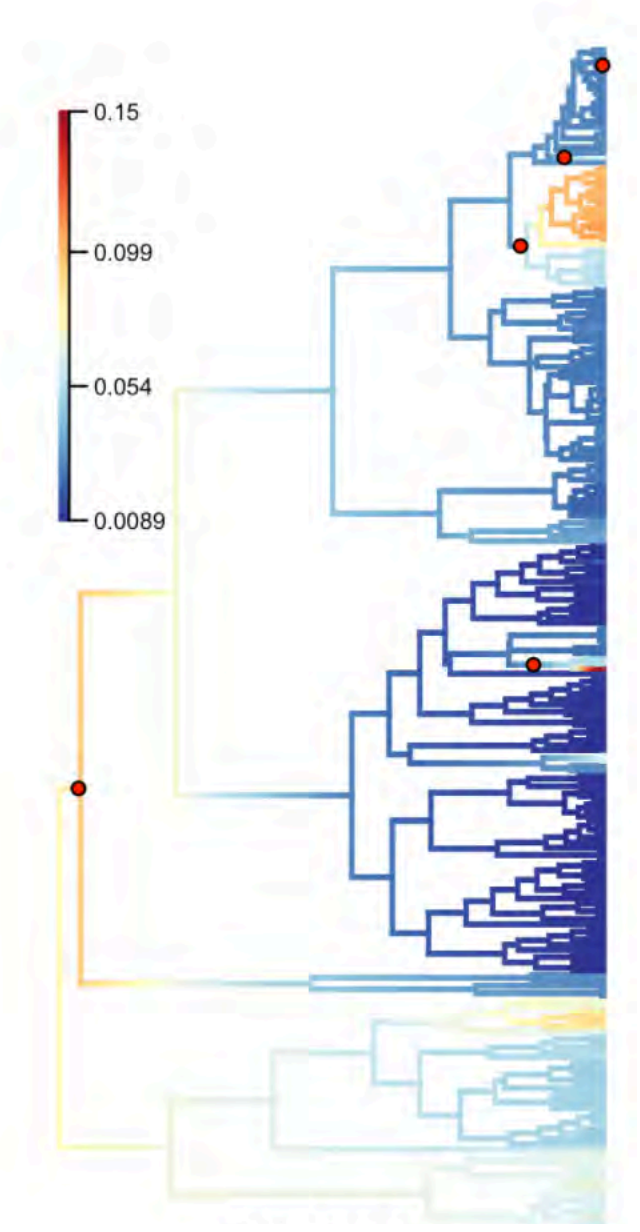




Prior might drive the result if the data are not very informative

BAMM

Dan Rabosky



BAMM (Bayesian analysis of macroevolutionary mixtures)

Infers

1) Number of shifts among diversification regimes
2) Characteristics of each regime

The idea is powerful you can give it a phylogeny it will find the interesting points where the mode of evolution has changed.

Became popular almost immediately.

351 citations for the 2014 paper that includes initial description.

BAMM controversy

Brian Moore

1) The compound Poisson process distribution that describes how we expect rate shifts to be observed is statically hard to identify making analyses very sensitive to the prior specification.
2) A problem with how we account for rate shifts in unobserved lineages leads to a bias in the estimates of extinction probabilities. Making the likelihoods calculated incorrect – suggest a computationally costly solution.
3) Take together these lead to unreliable hypothesis testing – based on simulation study.

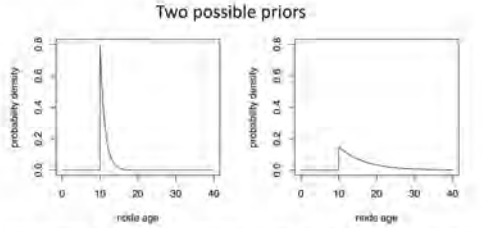
Dan Rabosky

1) Can't replicate with version 2.5 or later.
2) Disagree about the problem – and claim mathematical error in possible solution.
3) Can't replicate problems even when making simulations more difficult

[molecular ecologist blog]

[treethinkers blog]

$$Pr(A|B) = \frac{Pr(B|A) \times Pr(A)}{Pr(B)}$$



Prior might drive the result if the data are not very informative

My Advice

If you are interested in diversification.

1) Use more than one method
2) Play with priors and compare your results with very different priors

If you use BAMM and find the same rate shifts regardless of the prior that you use some reviewers will accept your results.

If you can also show that these results are consistent with other approaches to analyzing your data more reviewers will accept your results.

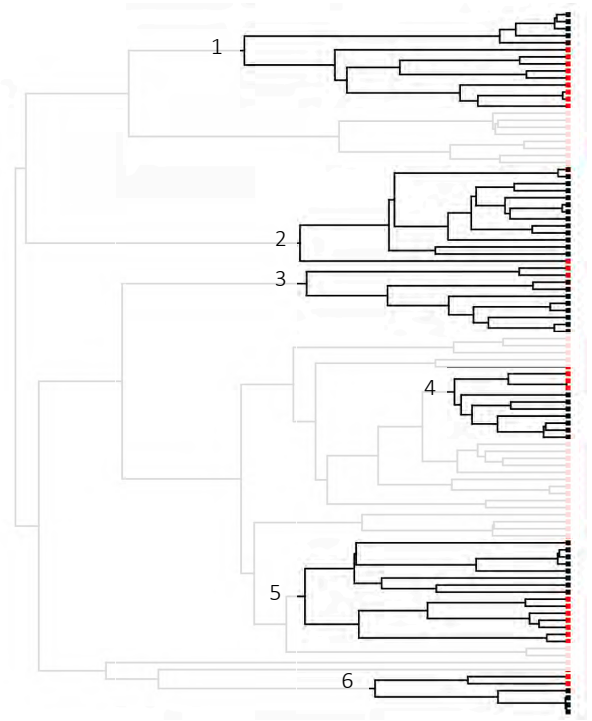
My Advice

If you are interested in diversification.

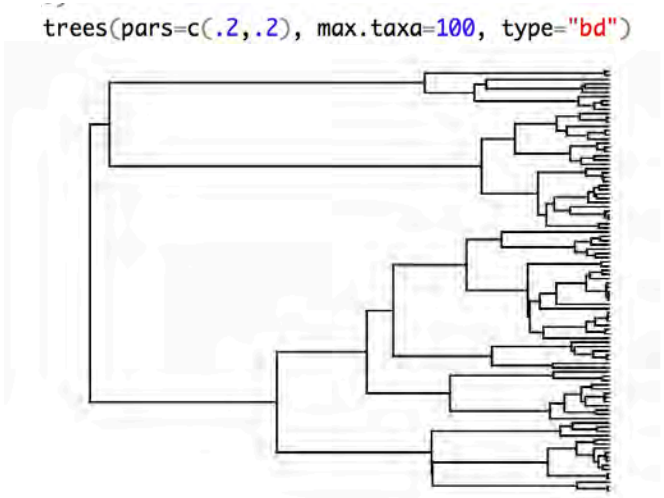
1) Use more than one method
2) Play with priors and compare your results with very different priors

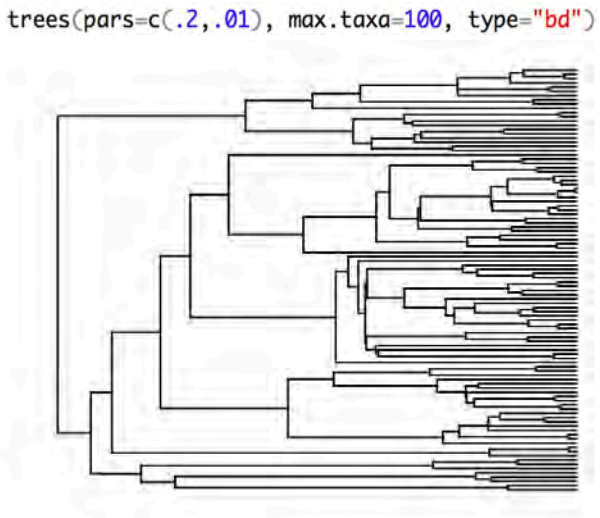
If you use BiSSE and find a significant difference in diversification between your two character states.
(our reading discusses why reviewers might not trust this)

Consider running BiSSE on subclades within your phylogeny that differ in the trait of interest.

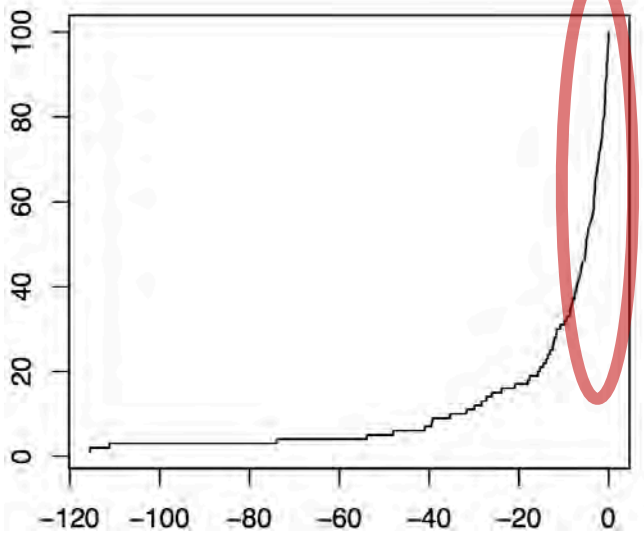


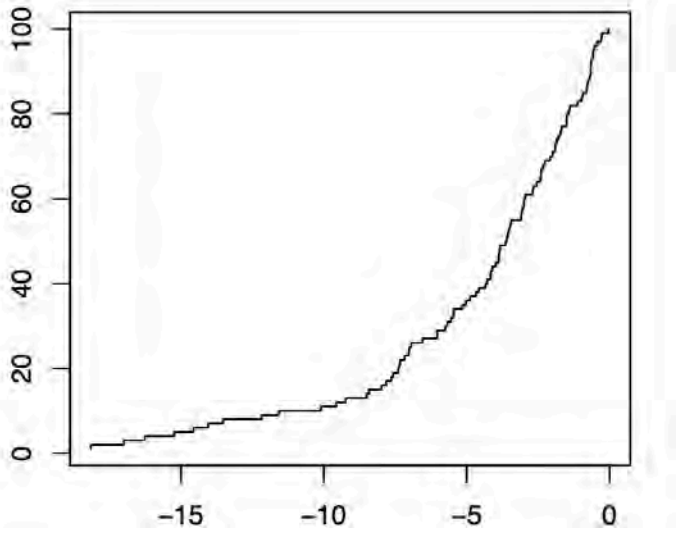
LTT Plots





Pull of the present

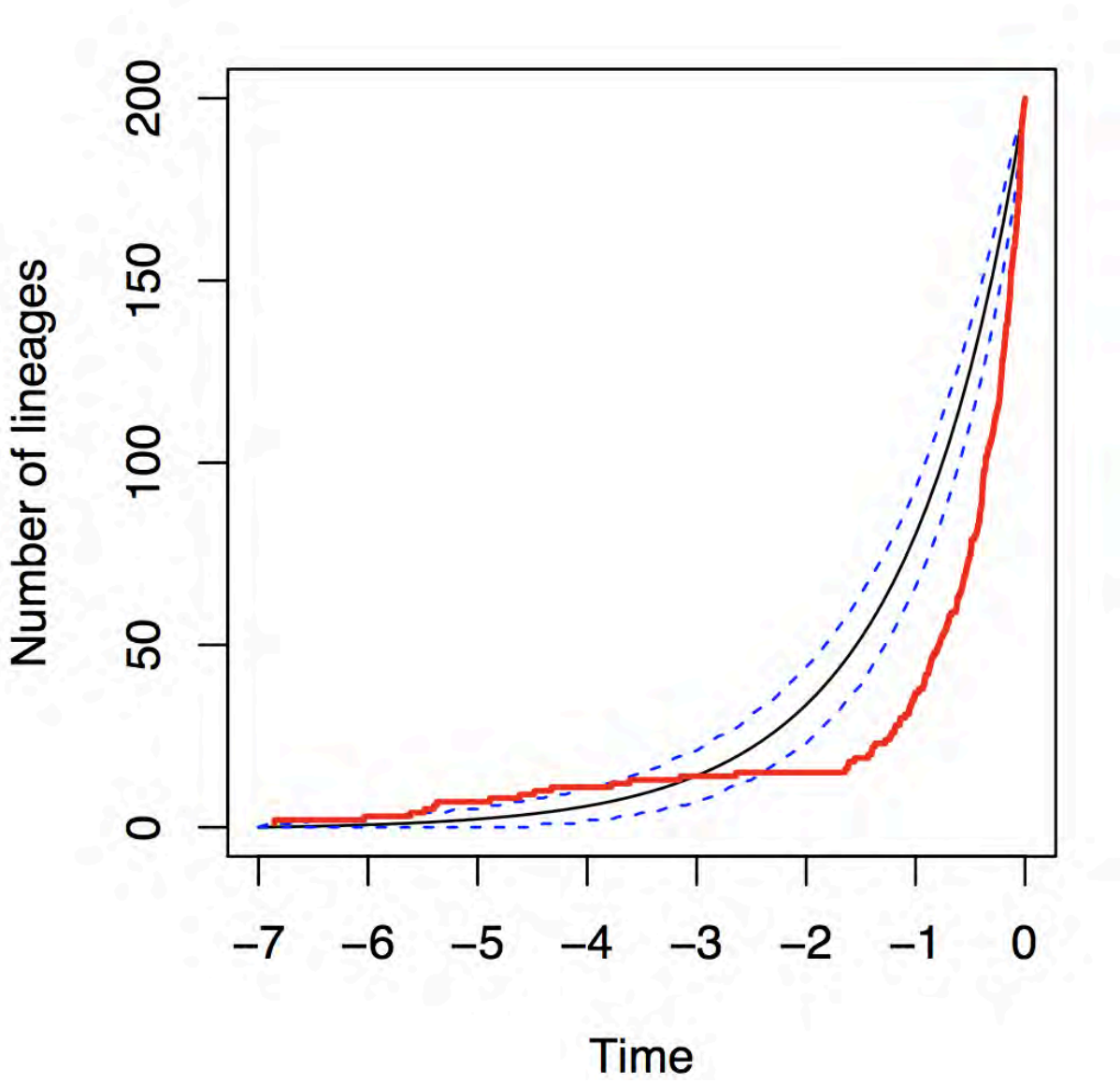




If we assume a birth death model the large uptick in an LTT plot as we reach the present day is caused by sampling lineages that are likely to go extinct in the future.

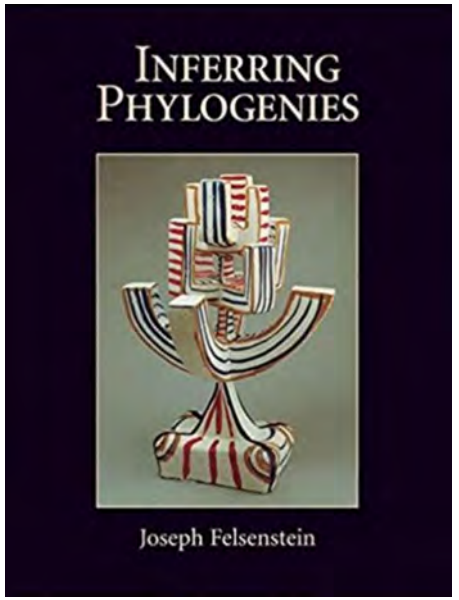
LTT Plots are put to many uses

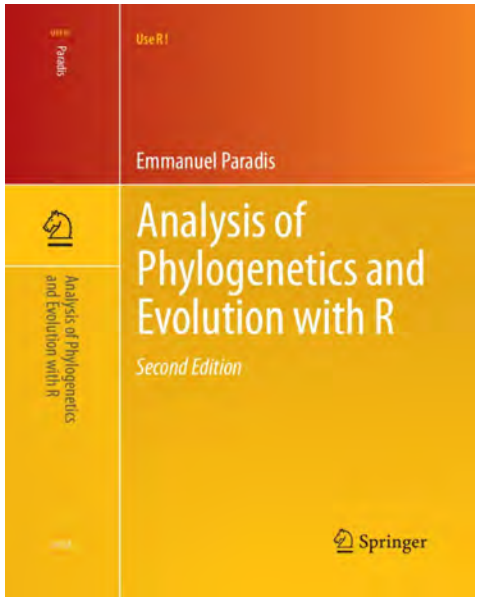
```
library(ape)
phy <- read.tree("my.tree.new")
est.rates <- bd.time(phy, 0, 0)$par
LTT(birth = est.rates[1],
    death = est.rates[2],
    N = 200, PI = 95,
    Tmax = 7, scaled=F)
ltt.lines(phy, col="red", lwd=2)
ltt.plot(phy)
```



Take homes

- Differential diversification analyses are hard!
- Easy to get a false positive
- Easy to use the wrong approach
- You can still do differential diversification
- My door is always open





CRAN Taskview

Workshops
Bodega – Bayesian/cheap
MBL – More molecular/very expensive
Nimbios – RevBays
MN course – coming soon!