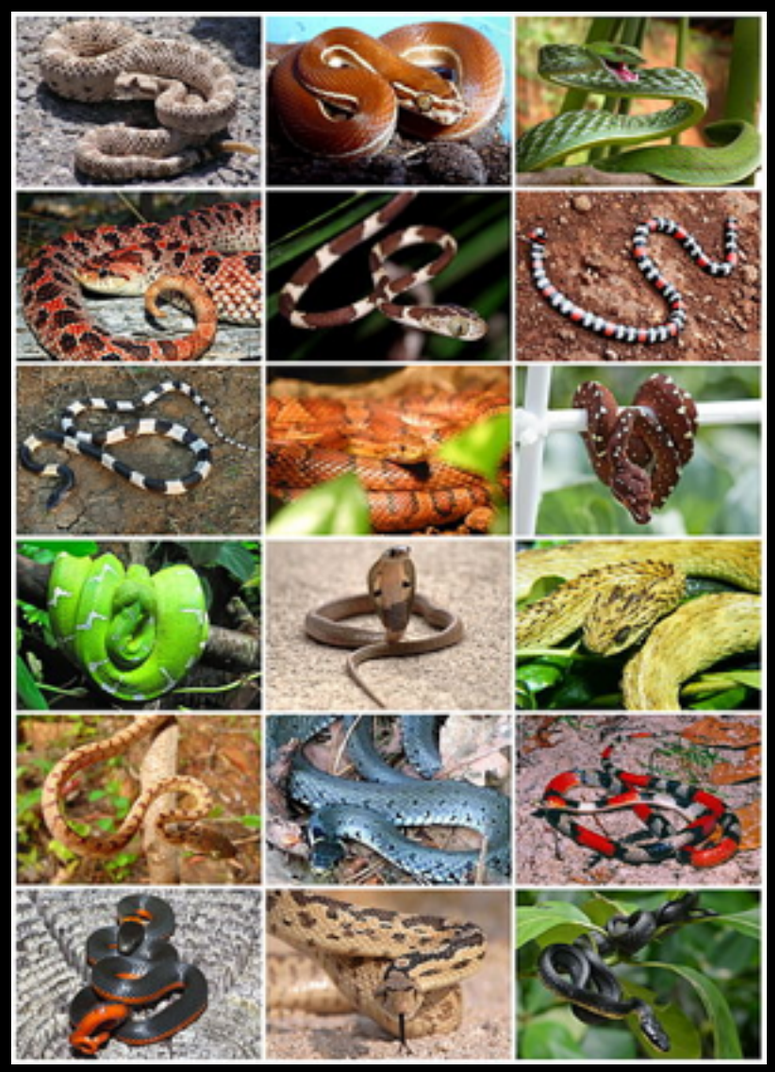
Diversification models

A FUNDAMENTAL PATTERN OF BIODIVERSITY: UNEVENNESS

> 9000 species



230 MY

TUATARA
Y U NO SPECIATE
imgflip.com

ECOLOGICAL ADAPTIVE RADIATION THEORY

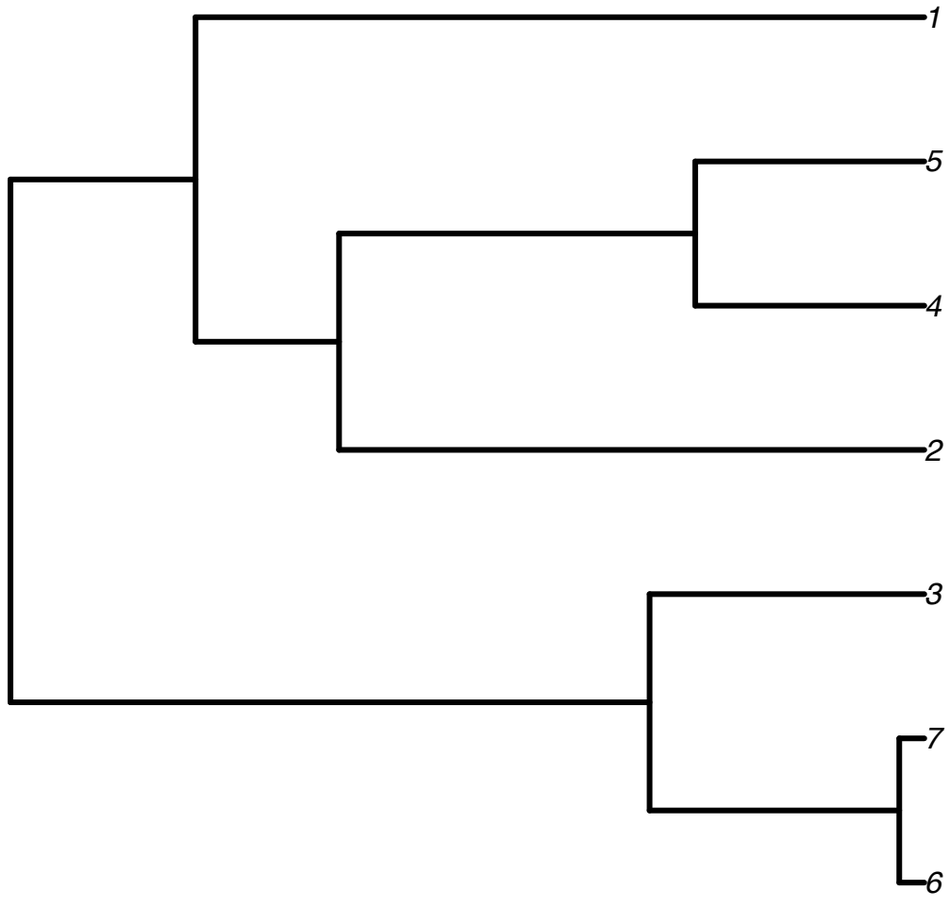
- innovation + ecological opportunity = rapid diversification

- ++speciation & ++disparification

Schluter 2000

Pure Birth model

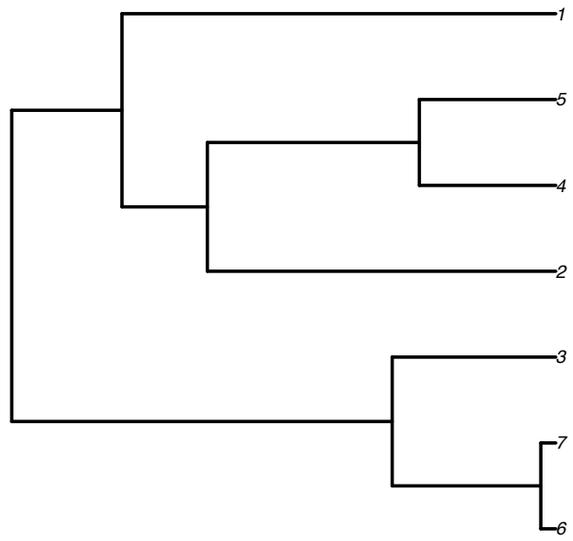
- each lineage has a constant probability of speciating, b
- b = birth rate
- $N(t) = N(0)e^{bt}$



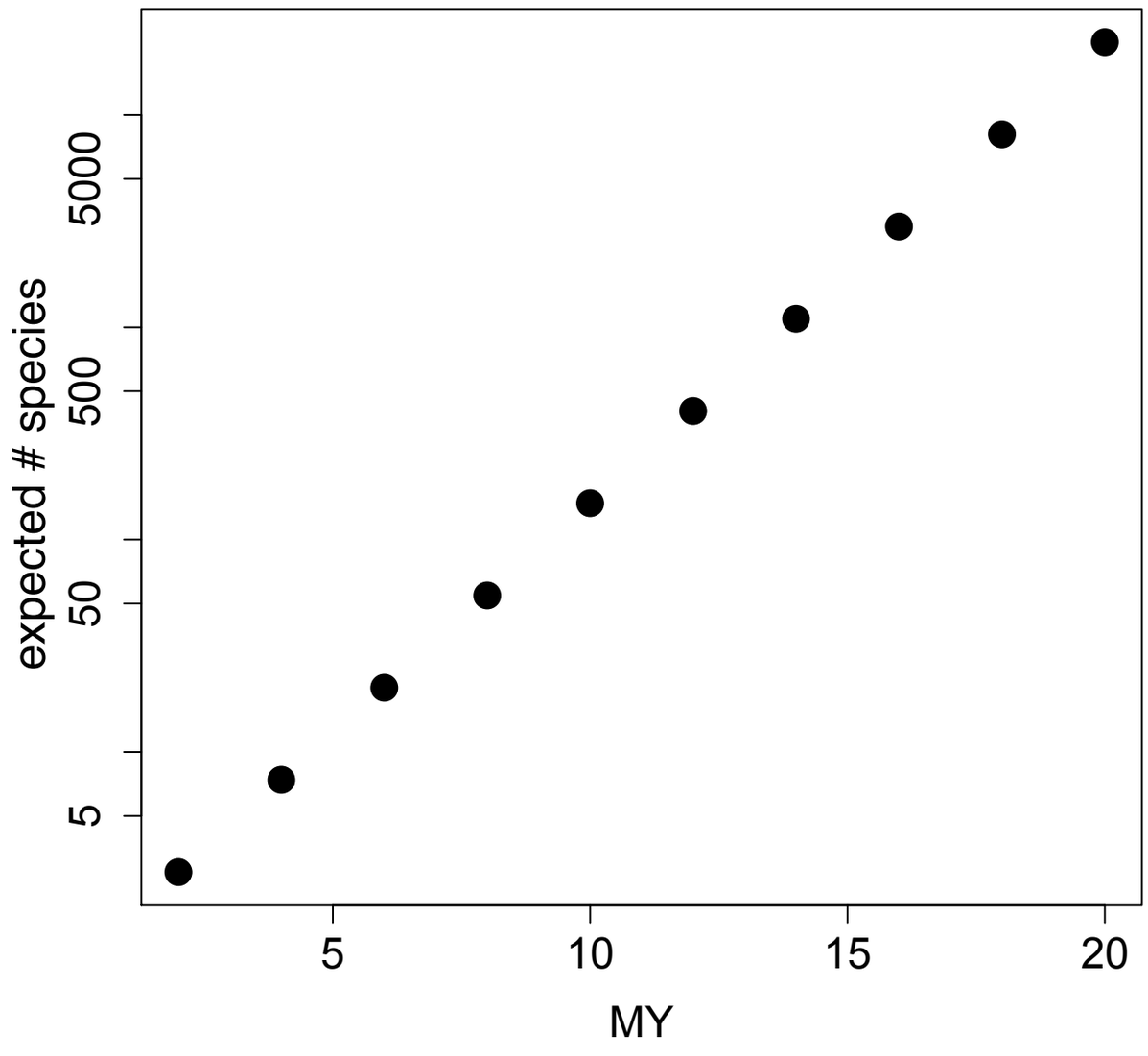
*Raup, D. 1975. Taxonomic survivorship curves and Van Valen's law. Paleobiology 1:82-96.

Stanley S.M. 1975. A theory of evolution above the species level. Proc. Natl. Acad. Sci. U. S. A. 72:646–650.

expectations



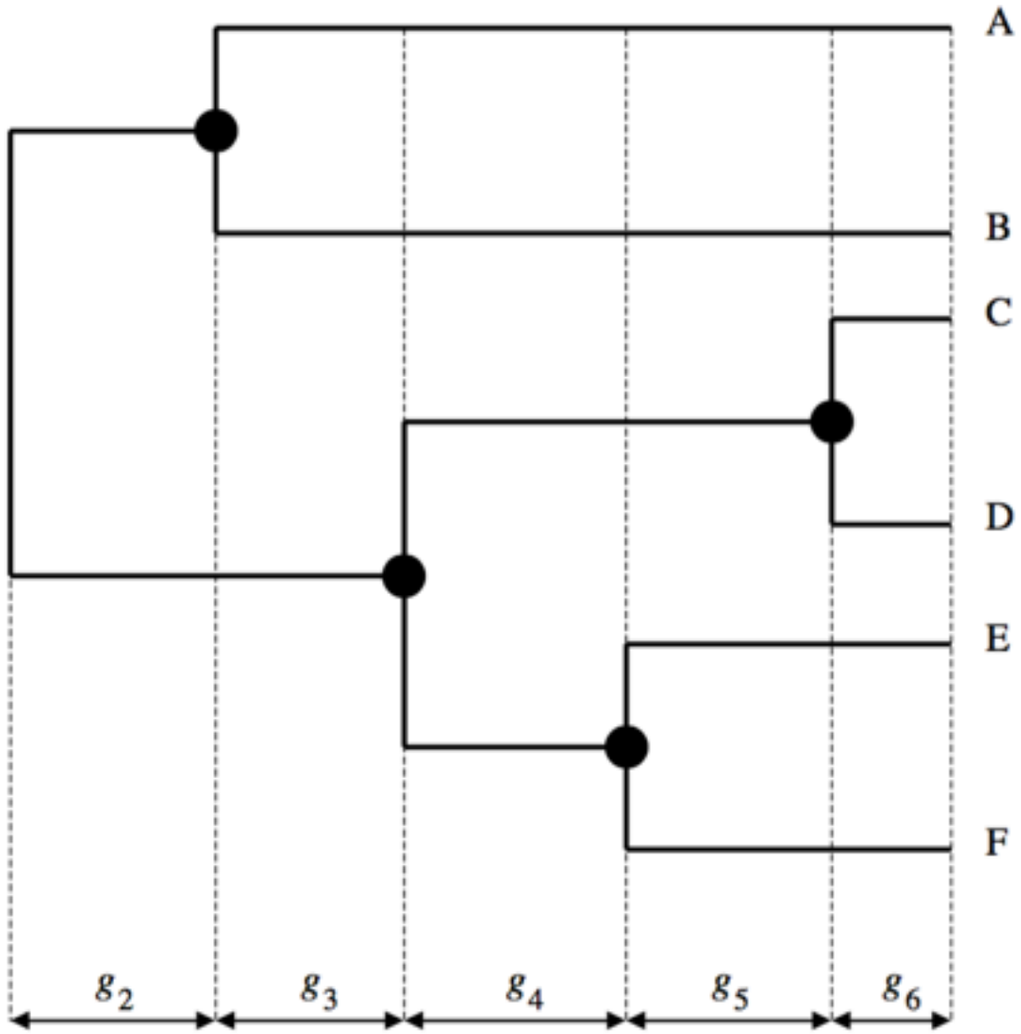
- $N(t) = N(0)e^{bt}$
- lineages accumulate exponentially (why?)



what can we do with the pure birth model?

- ask whether clades have diversified in a way that is consistent with pb
- estimate birth rate given a phylogeny

Testing for Rapid Radiation Over Clade History Using the Gamma Statistic



Pybus, O. G., and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Phil. Trans. Roy. Soc. (Lond.) B 267:2267-2272.

The Gamma Statistic

- describes relative depth of nodes in a tree

- function of *n* nodes and *g*, a vector of the internode distances or waiting times

- useful because gamma has a *standard normal distribution*

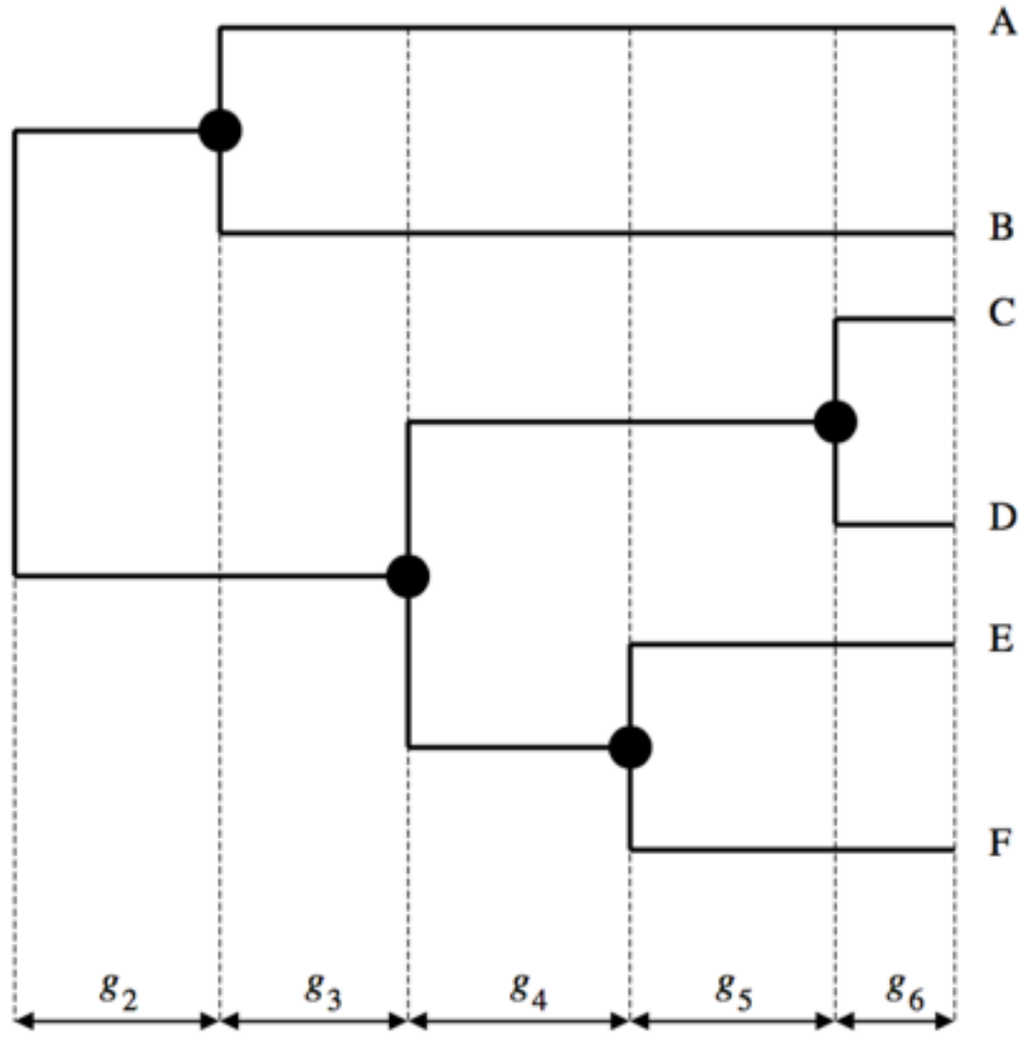
$$\gamma = \frac{\left(\frac{1}{n-2}\sum_{i=2}^{n-1}\sum_{k=2}^{i} k g_k\right) - \left(\frac{\tau}{2}\right)}{\tau\sqrt{\frac{1}{12(n-2)}}}, \tau = \left(\sum_{j=2}^{n} j g_j\right)$$

Pybus, O. G., and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Phil. Trans. Roy. Soc. (Lond.) B 267:2267-2272.

The Gamma Statistic

$$\gamma = \frac{\left(\frac{1}{n-2}\sum_{i=2}^{n-1}\sum_{k=2}^{i} k g_k\right) - \left(\frac{\tau}{2}\right)}{\tau\sqrt{\frac{1}{12(n-2)}}}, \tau = \left(\sum_{j=2}^{n} j g_j\right)$$

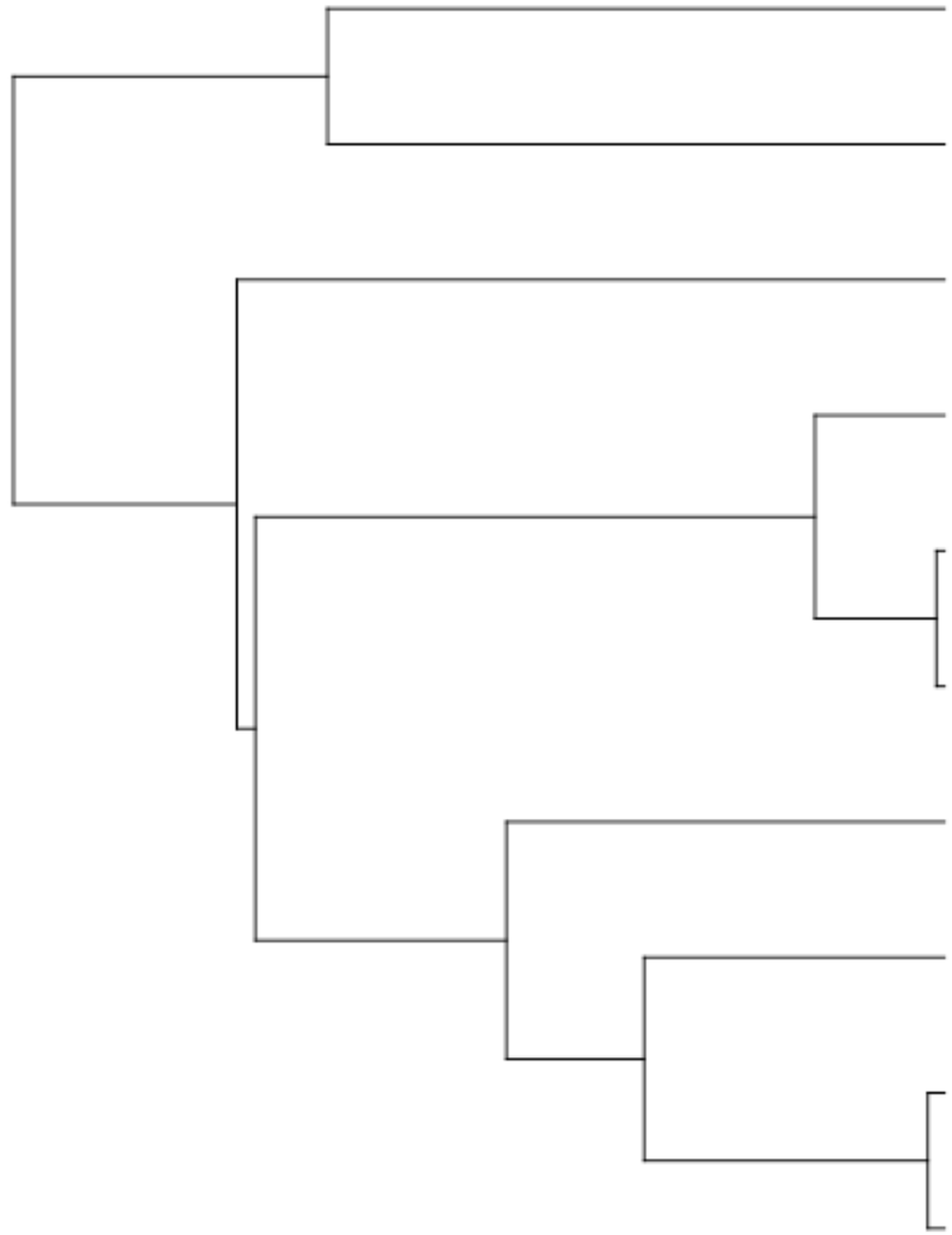
- measures the relative position of nodes within a phylogeny
- function of *n* nodes and *g* (summed branch lengths within a waiting time interval)
- useful because statistic has a *standard normal distribution*
- g is just a vector of the internode distances



Pybus, O. G., and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Phil. Trans. Roy. Soc. (Lond.) B 267:2267-2272.

The gamma statistic

- **γ of 0** indicates internal nodes have balanced distribution) compared to pb expectation)



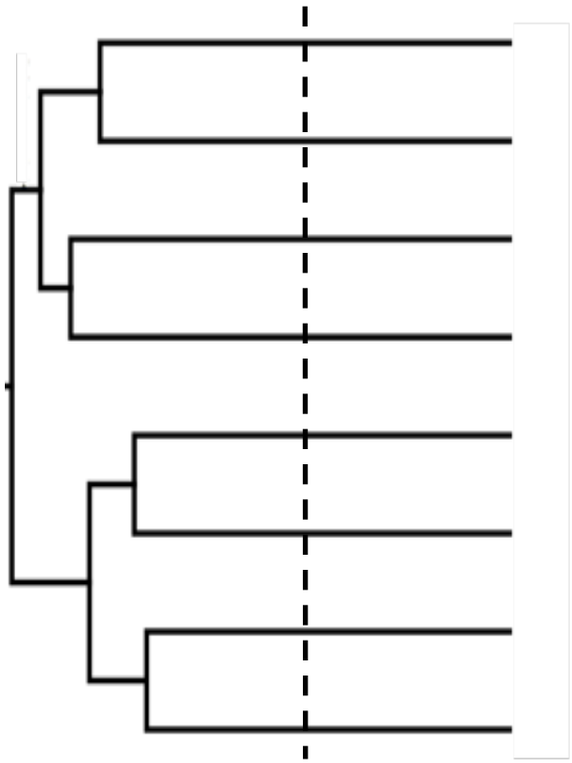
γ = - 0.03

negative gamma

- -γ means nodes are closer to root than expected

- interpretation: speciation rate was high initially and slowed through time

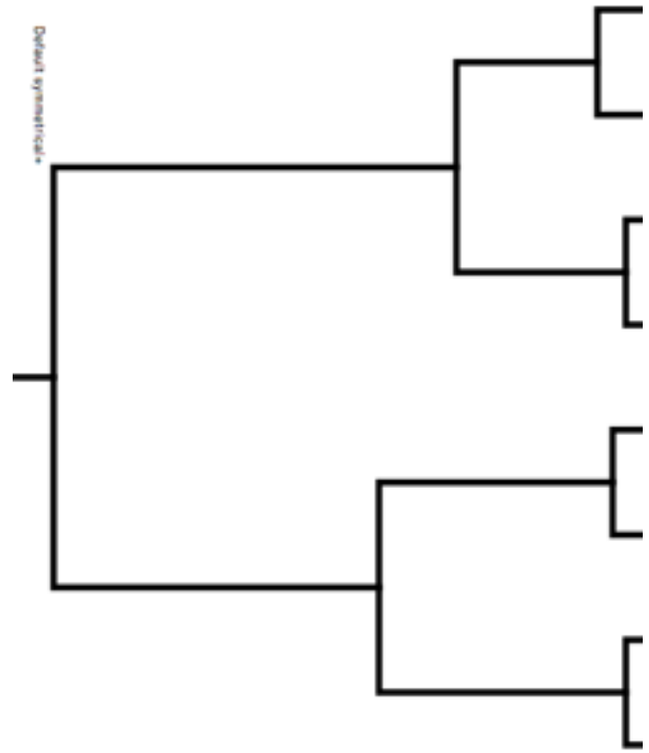
- evidence for adaptive radiation



positive gamma

- +γ means internal nodes closer to tips than expected

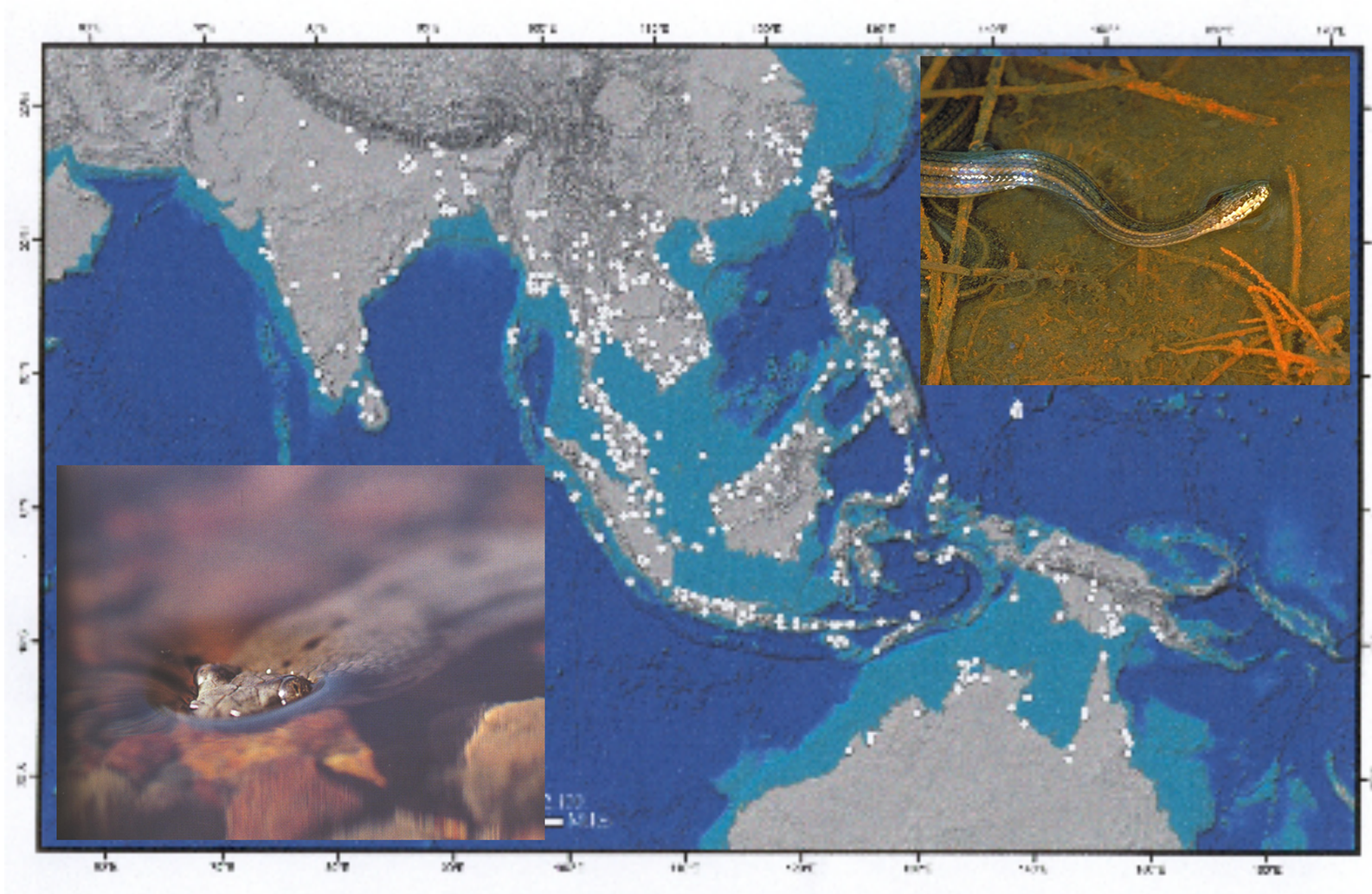
- interpretation: speciation rate low initially and *increased* through time *or* extinction decreased towards present



since γ is normally distributed we can test hypotheses about departures of gamma using a Z-test

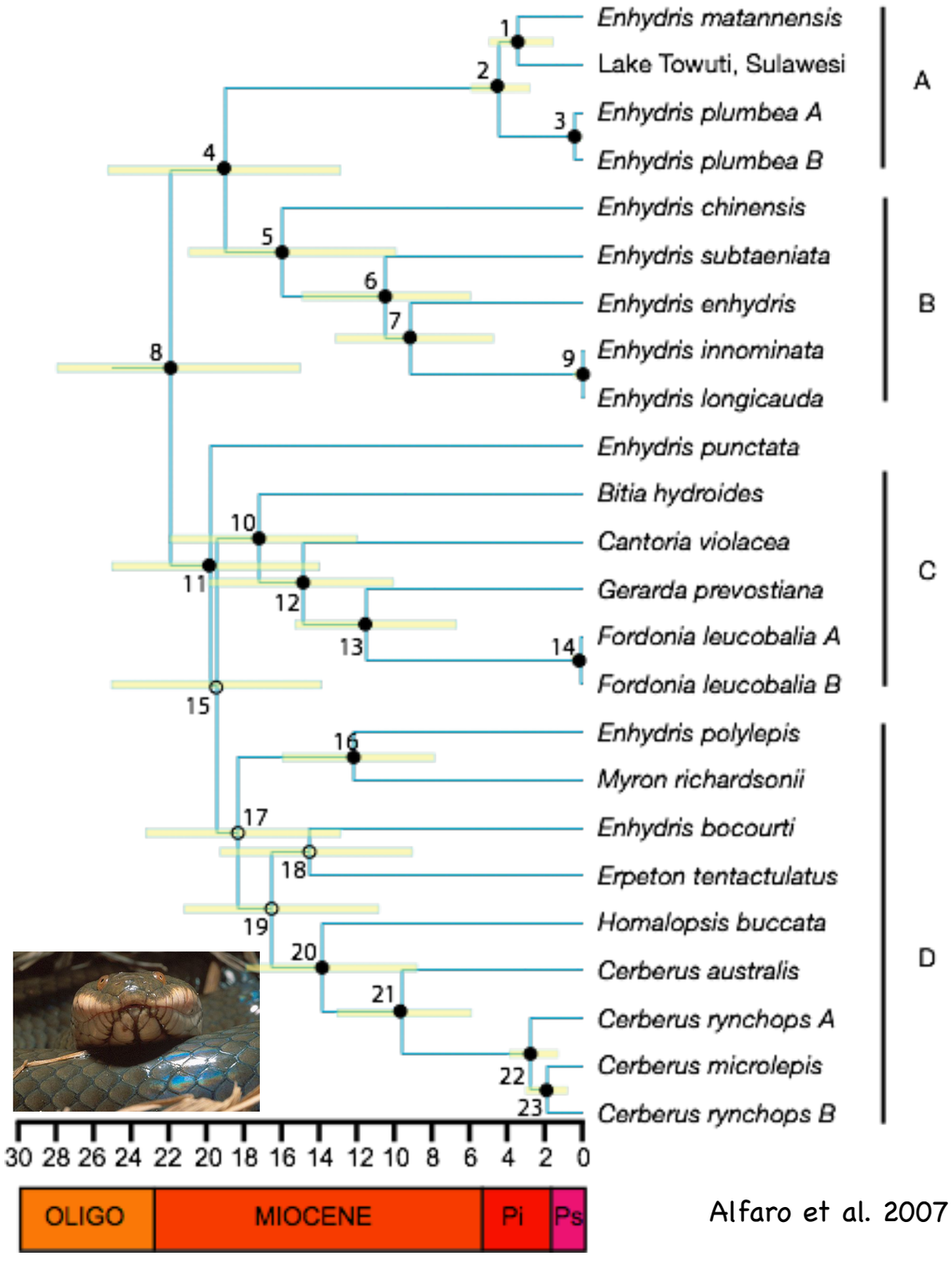
Example of Gamma Statistic: Diversification of Homalopsid Snakes

- 34 species distributed across S.E. Asia
- Aquatic to semi-aquatic



Have homalopsids adaptively radiated?

- 9 of 10 genera, 20 of 34 species
- CR (gamma) test suggests rapid initial diversification (P < 0.001)

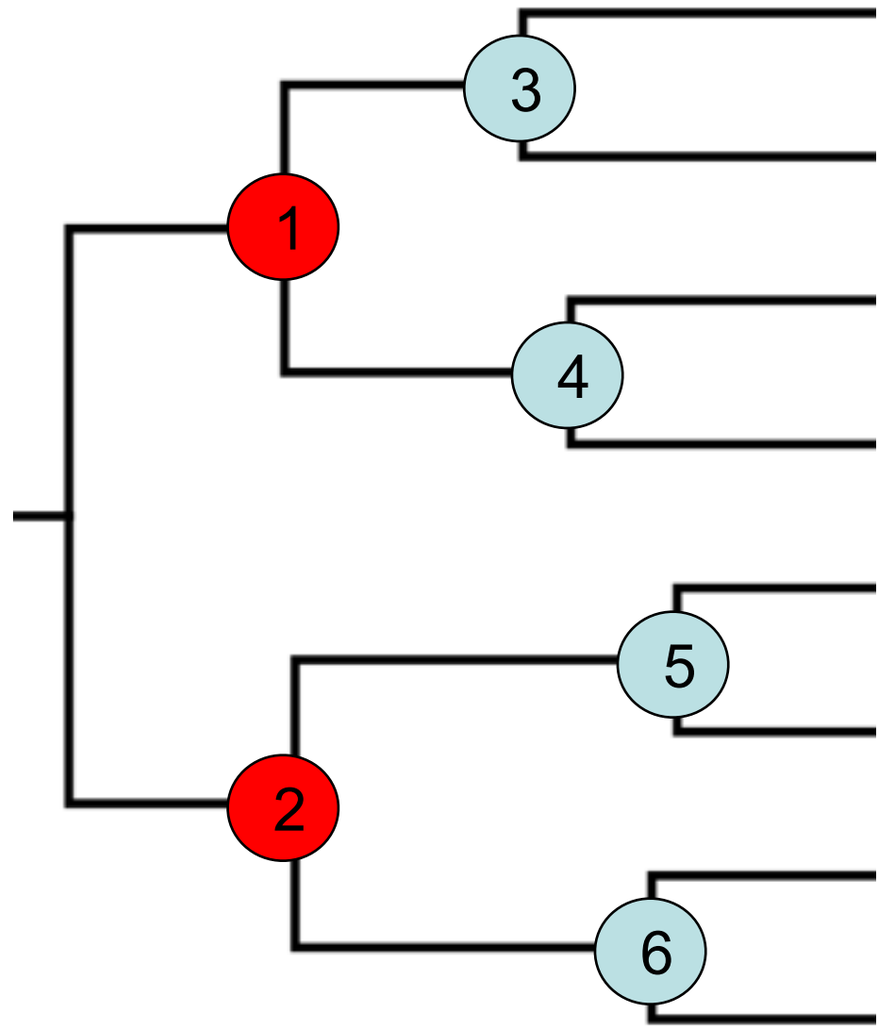


Alfaro et al. 2007

What about incomplete sampling?

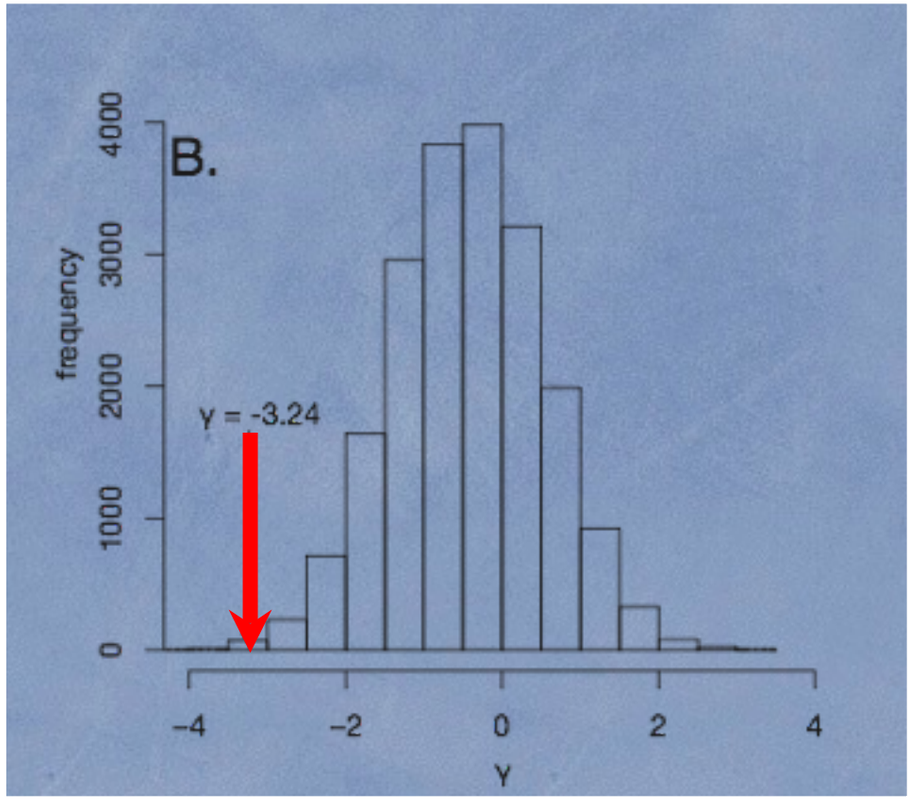
- with random sampling of four taxa, what nodes would be expected to miss?

- how would this affect gamma?



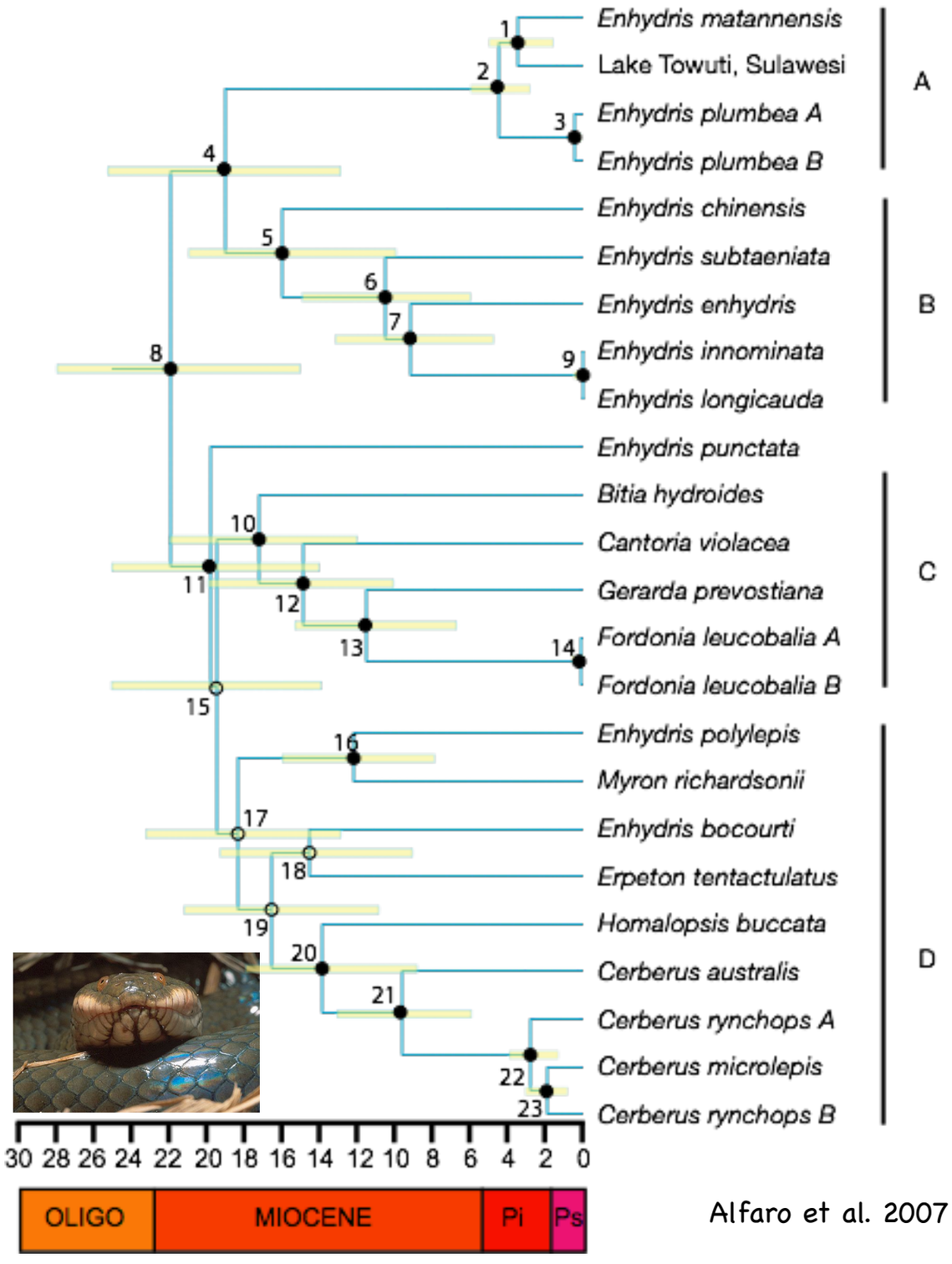
MCCR: The constant rates test for incompletely sampled trees*

- Complete sampling: $\gamma \leq -1.645$
- Incomplete sampling means that old nodes more likely to be recovered, biasing CR test
- Solution: MCCR test*
  - Simulate tree with n taxa
  - RANDOMLY prune missing taxa
  - Calculate the γ-statistic on pruned tree
  - Repeat
  - Calculate corrected p-value



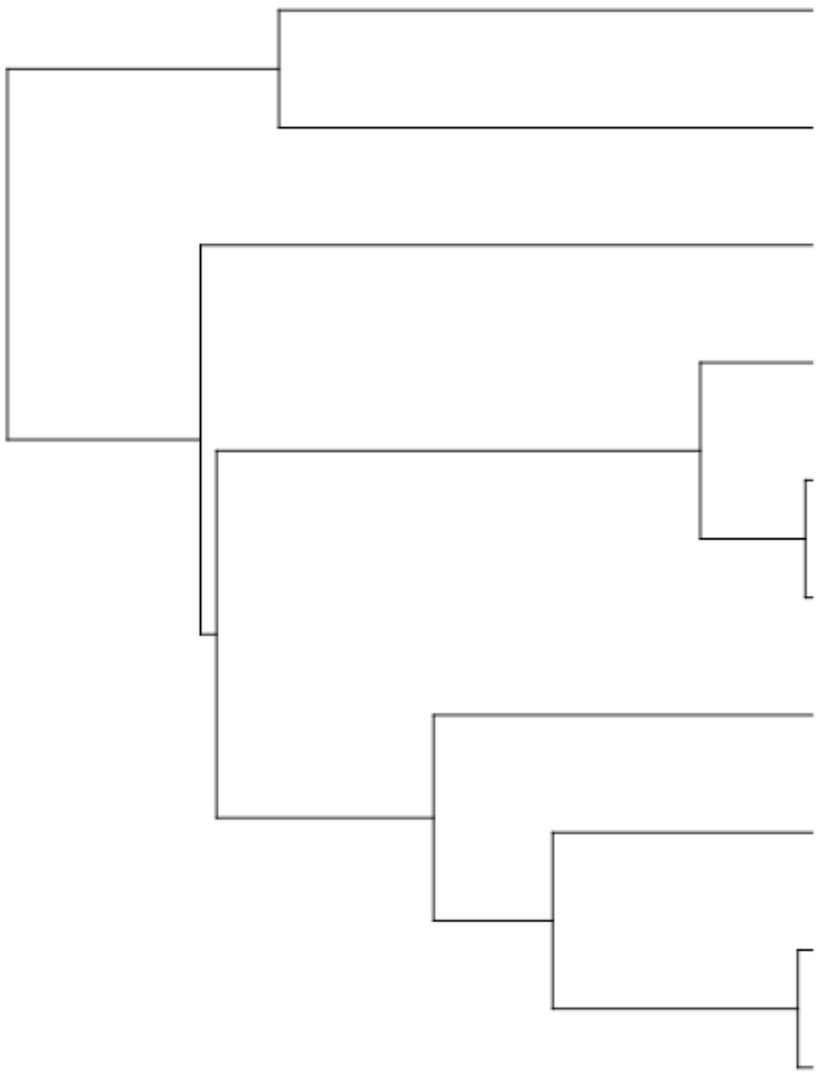
A timetree for homalopsid diversification

- Bayesian timetree using BEAST
- 4 genes (12S, 16S, cytB, C-Mos
- 9 of 10 genera, 20 of 34 species
- MCCR test suggests rapid initial diversification ($P < 0.001$)



Alfaro et al. 2007

fitting pure birth models



```
> bd.km(tt)
[1] 0.9748699
```

bd.ms {geiger} R Documentation

estimate net diversification rate

Description

estimating net diversification rate with confidence limits and testing diversities

Usage

```
bd.ms(phy=NULL, time, n, missing = 0, crown=TRUE, epsi
bd.km(phy=NULL, time, n, missing = 0, crown=TRUE)
crown.p(phy=NULL, time, n, r, epsilon)
stem.p(phy=NULL, time, n, r, epsilon)
crown.limits(time, r, epsilon, CI=0.95)
stem.limits(time, r, epsilon, CI=0.95)
```

Arguments

`phy`	a phylogenetic tree of class 'phylo' (see **Details**)
`time`	time interval (can be a vector)
`n`	number of extant species
`r`	net diversification rate, birth - death
`epsilon`	extinction rate as a fraction of speciation rate
`missing`	number of taxa missing from tree
`crown`	whether time is treated as crown age (or otherwise as stem age)
`CI`	confidence level for estimated parameters

estimating the birth rate

- you can estimate the net diversification rate, **r** from the present day **richness** of a clade and the **age** of the clade.

$$\hat{r}_{crown} = \frac{log(n) - log(2)}{time}$$

$$\hat{r}_{carnations} = \frac{log(200) - log(2)}{0.6}$$

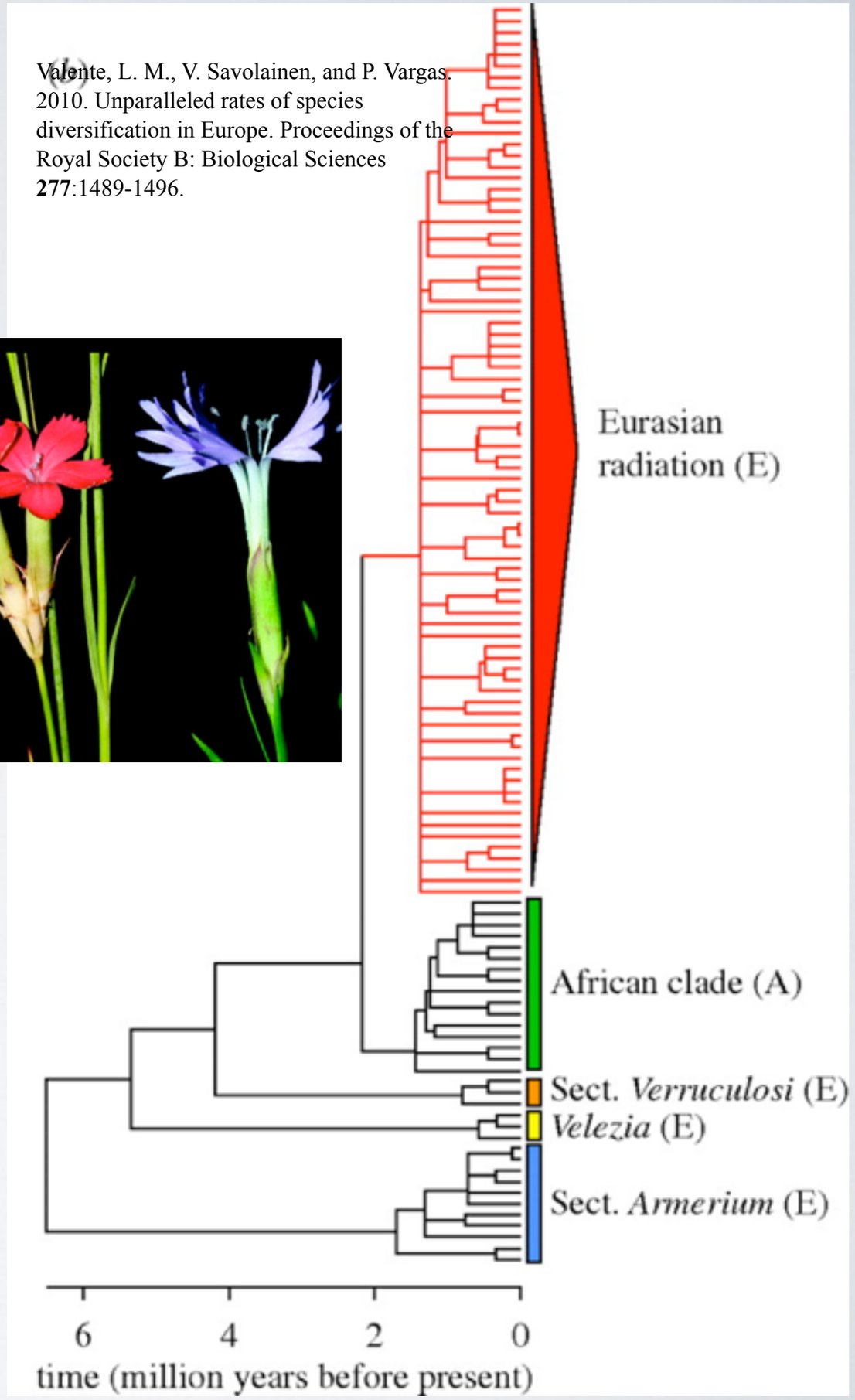
= 7.67

Magallón S., Sanderson M.J. 2001. Absolute diversification rates in angiosperm clades. Evolution. 55:1762–1780.

FUN WITH THE PURE BIRTH MODEL

$$N(t) = N(0)e^{bt}$$

- 200 carnations (*Dianthus*) today
- Carnation birth rate =b 7.6 species/my



Valente, L. M., V. Savolainen, and P. Vargas. 2010. Unparalleled rates of species diversification in Europe. Proceedings of the Royal Society B: Biological Sciences **277**:1489-1496.

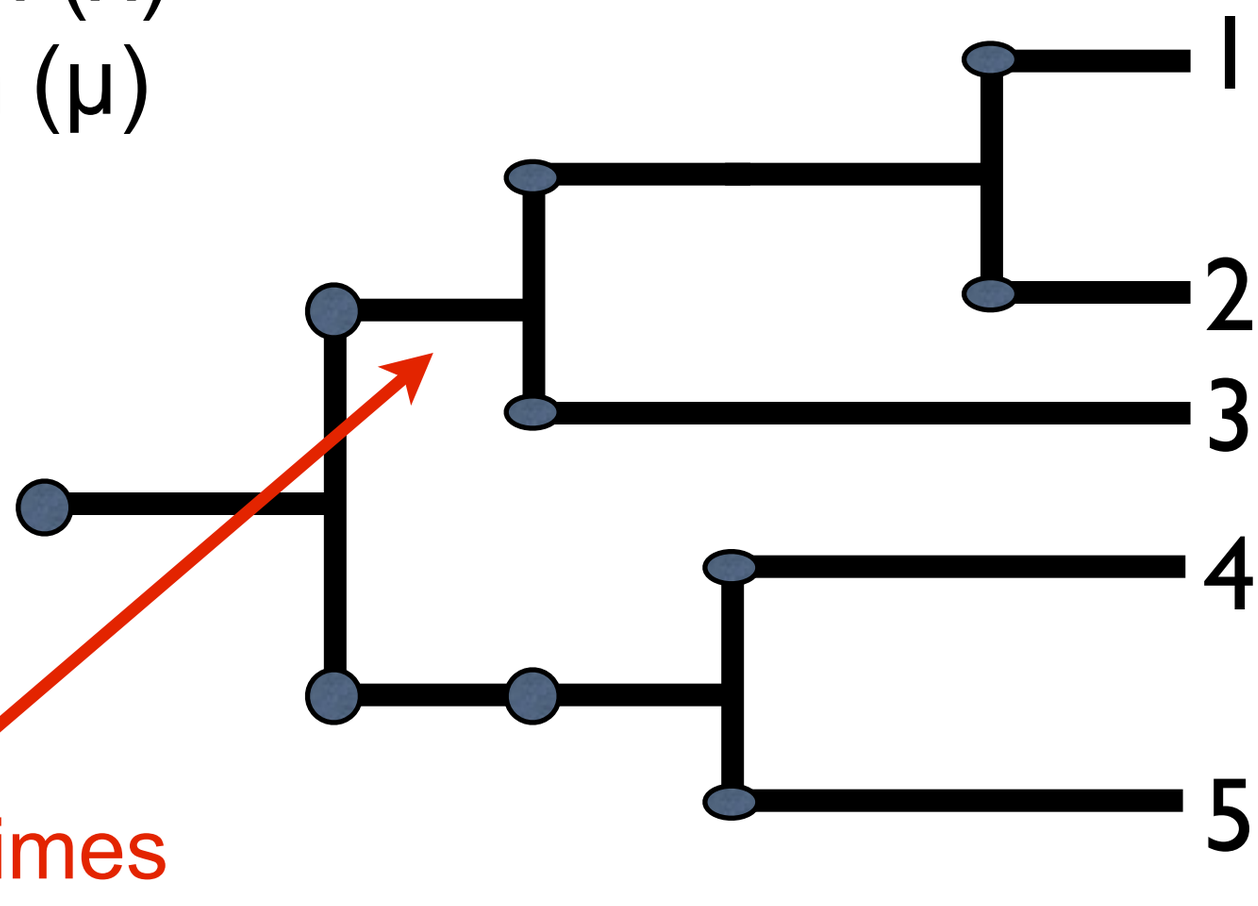
an improvement: the birth-death model

- Assumptions:
  - Speciation and extinction occur randomly
  - Each lineage has an equal and constant rate of speciation, *b*, and extinction, *d*

Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344:305–311.

Diversification Model

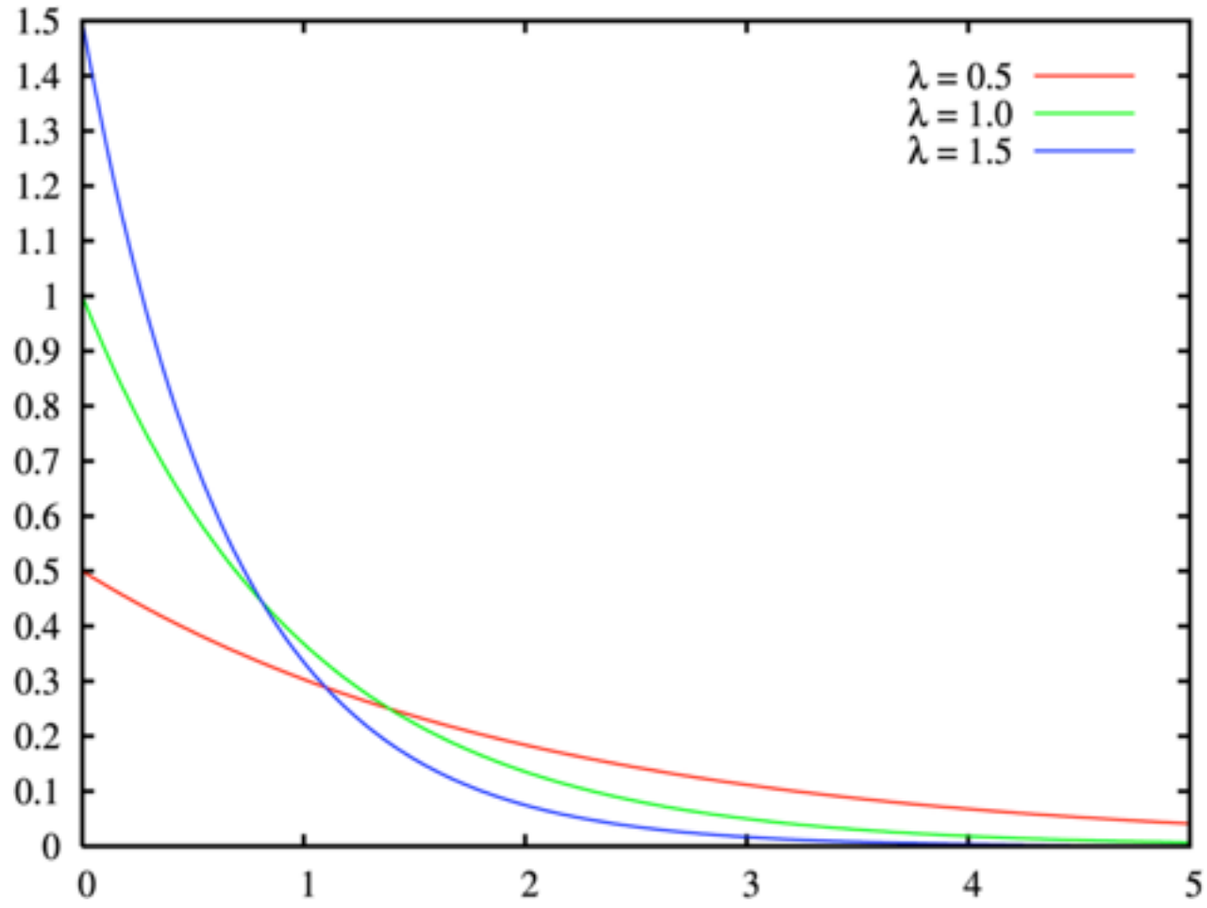
- Model: birth-death
- Every lineage has constant probability of:
  - –speciation (λ)
  - –extinction (μ)



Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344:305–311.

Waiting times

- Under a birth-death model, the time intervals between successive events are always drawn from exponential distributions



$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & , \ x \geq 0, \\ 0 & , \ x < 0. \end{cases}$$

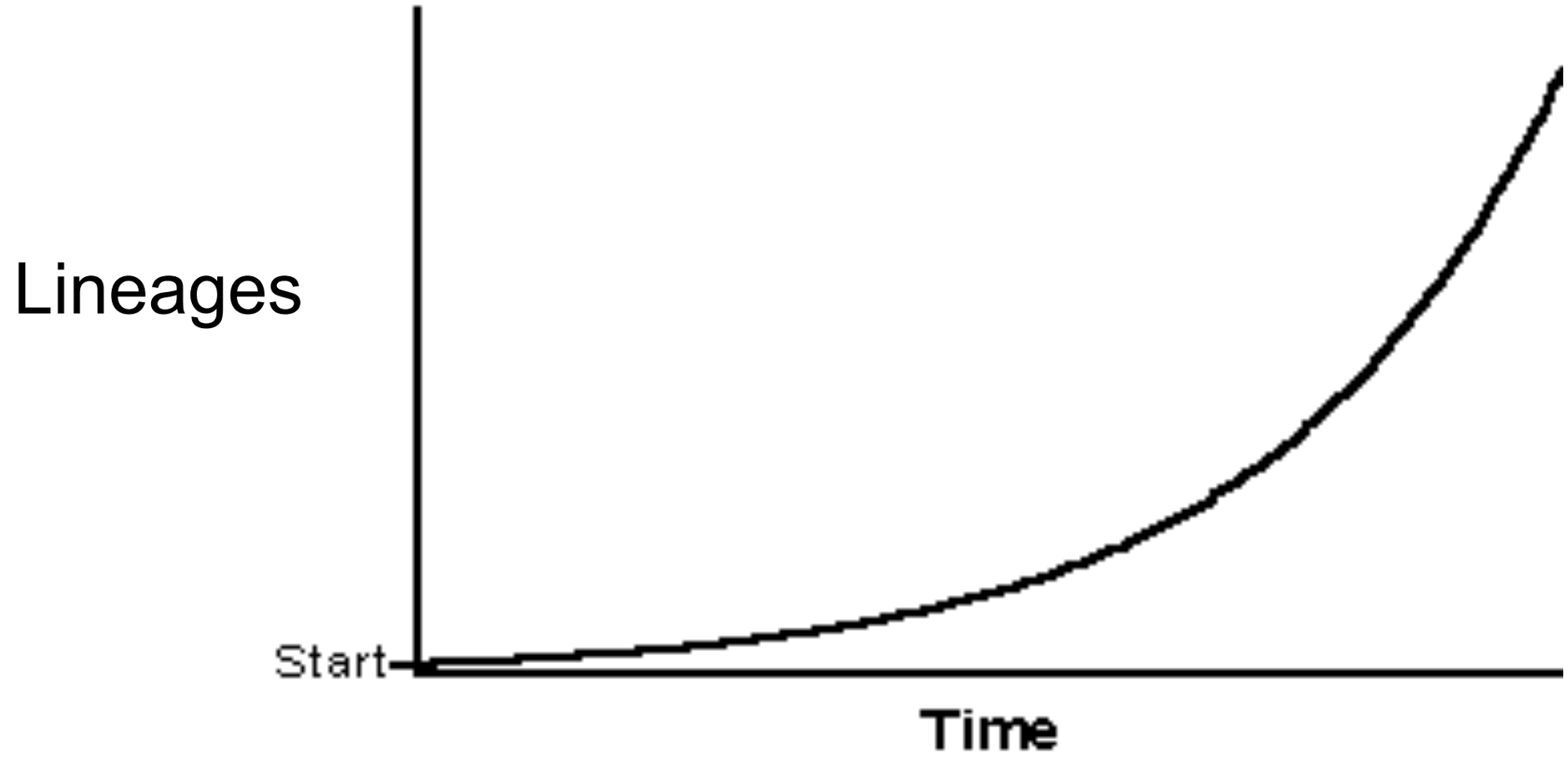
Expected species diversity under a birth-death model:

$$E[N_t] = N_o e^{(b-d)t}$$

N_t = species diversity after time t
N_o = starting species diversity
b = birth rate
d = death rate
t = time

Species Accumulation

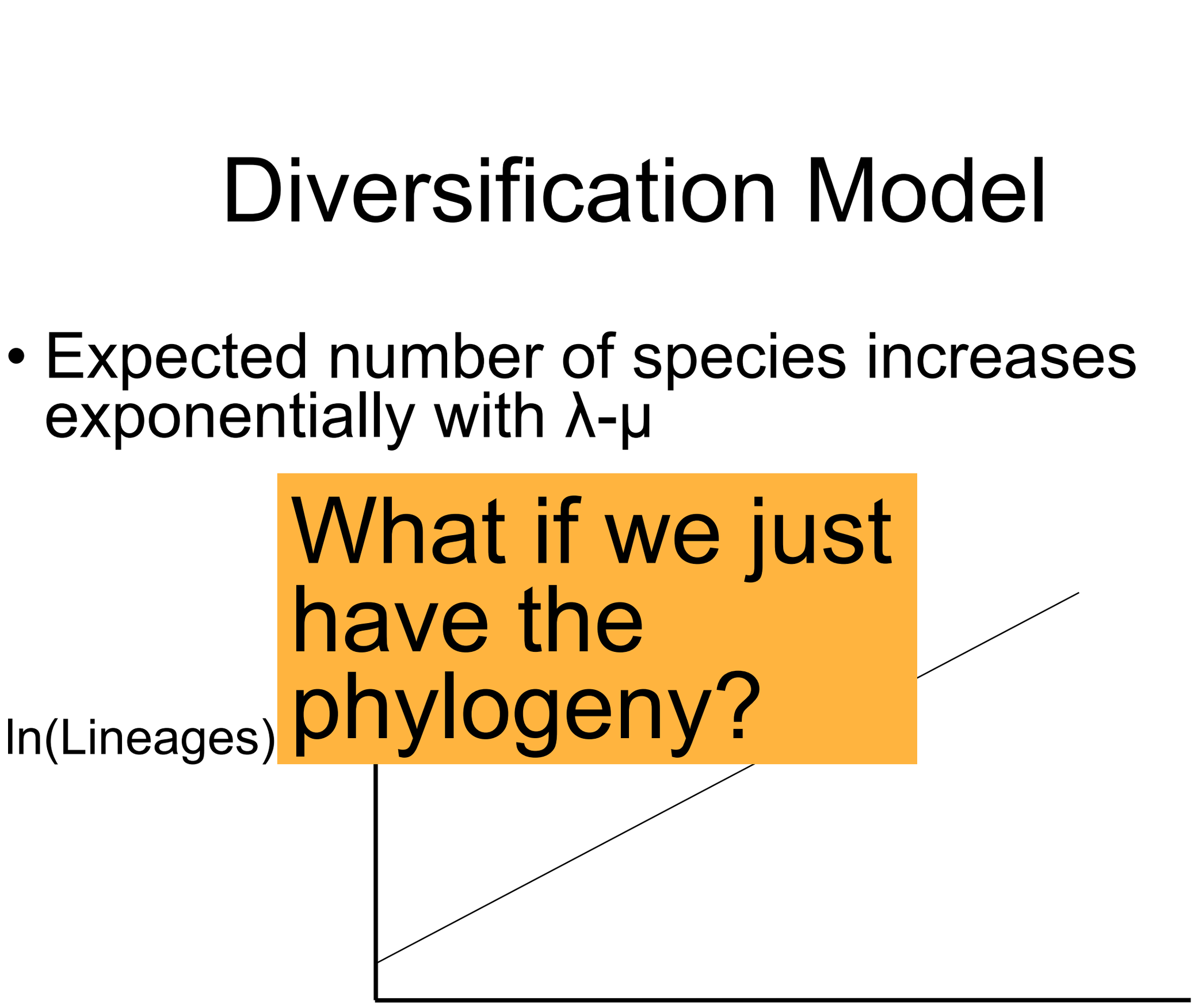
- Expected number of species increases exponentially with λ-μ



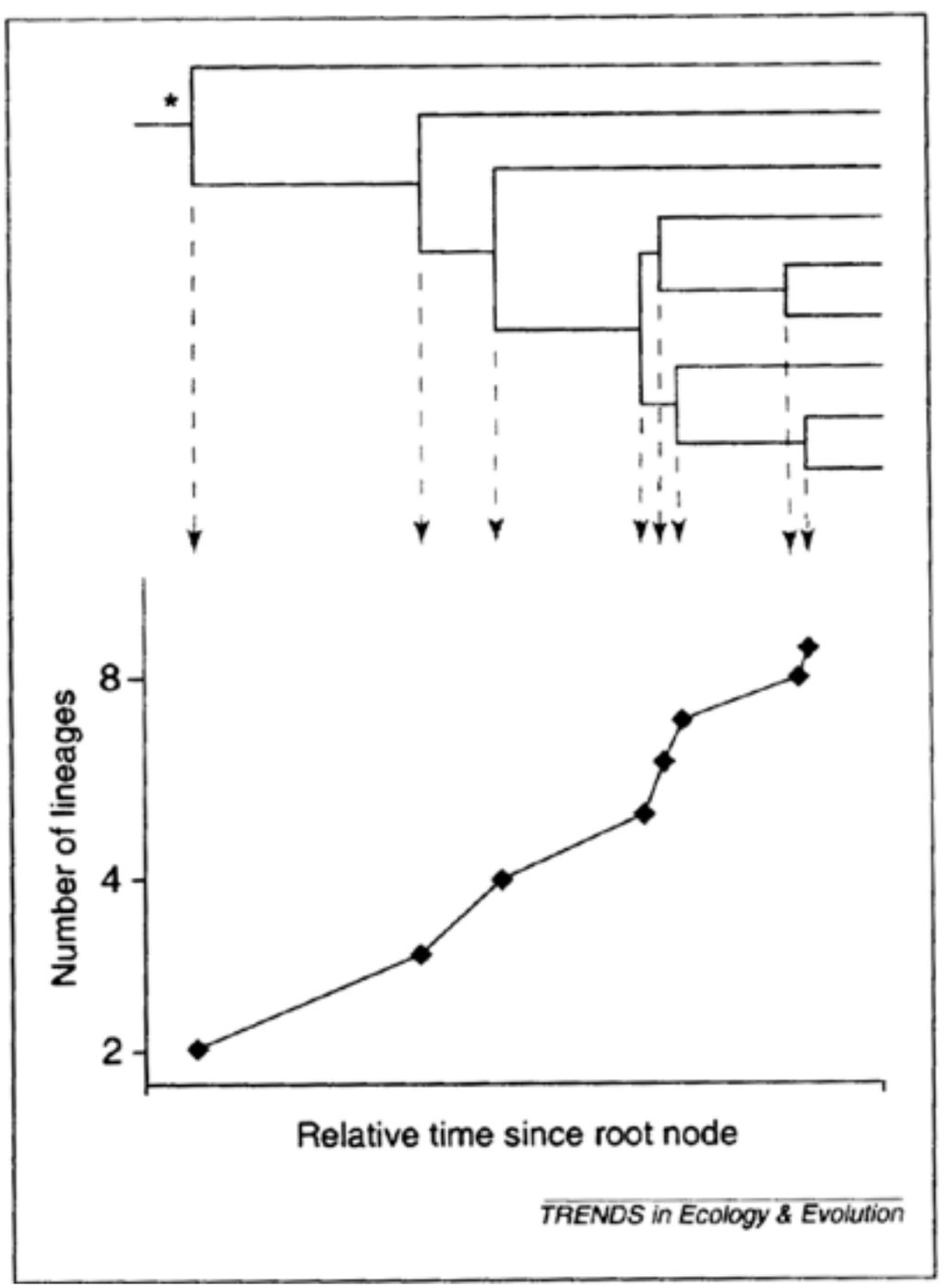
Diversification Model

- Expected number of species increases exponentially with λ-μ

What if we just have the phylogeny?

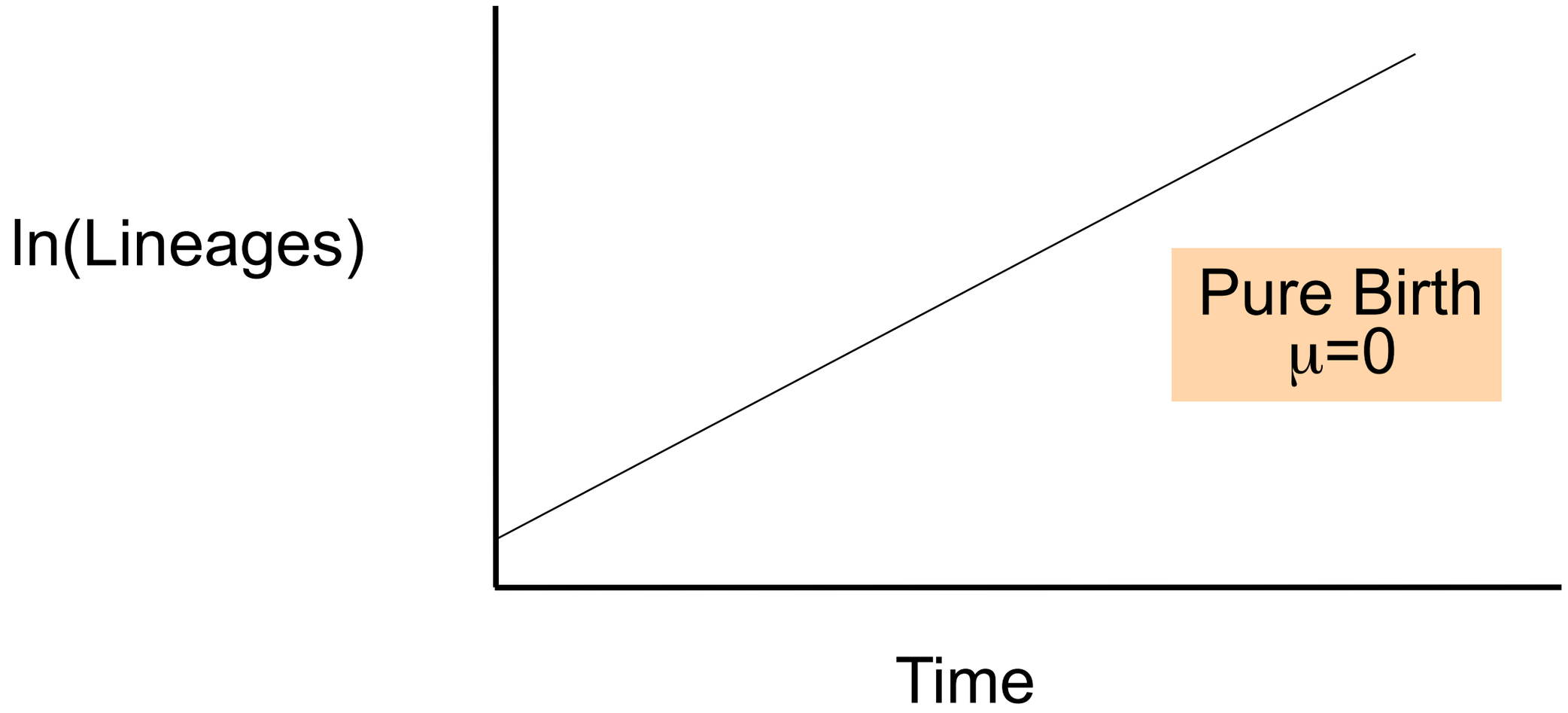


Lineage-through-time



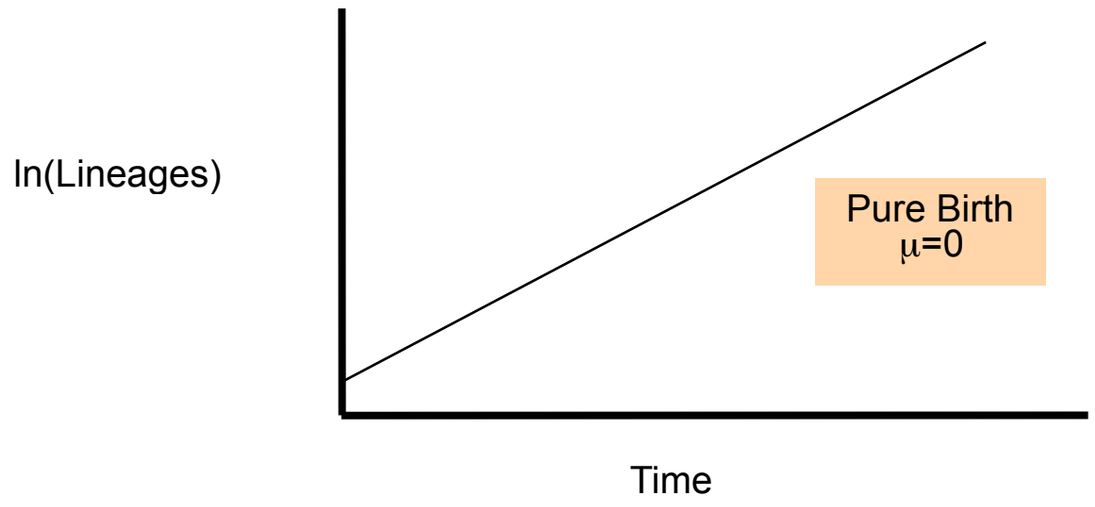
Barraclough T.G., Nee S. 2001. Phylogenetics and speciation. Trends Ecol. Evol. 16:391–399.

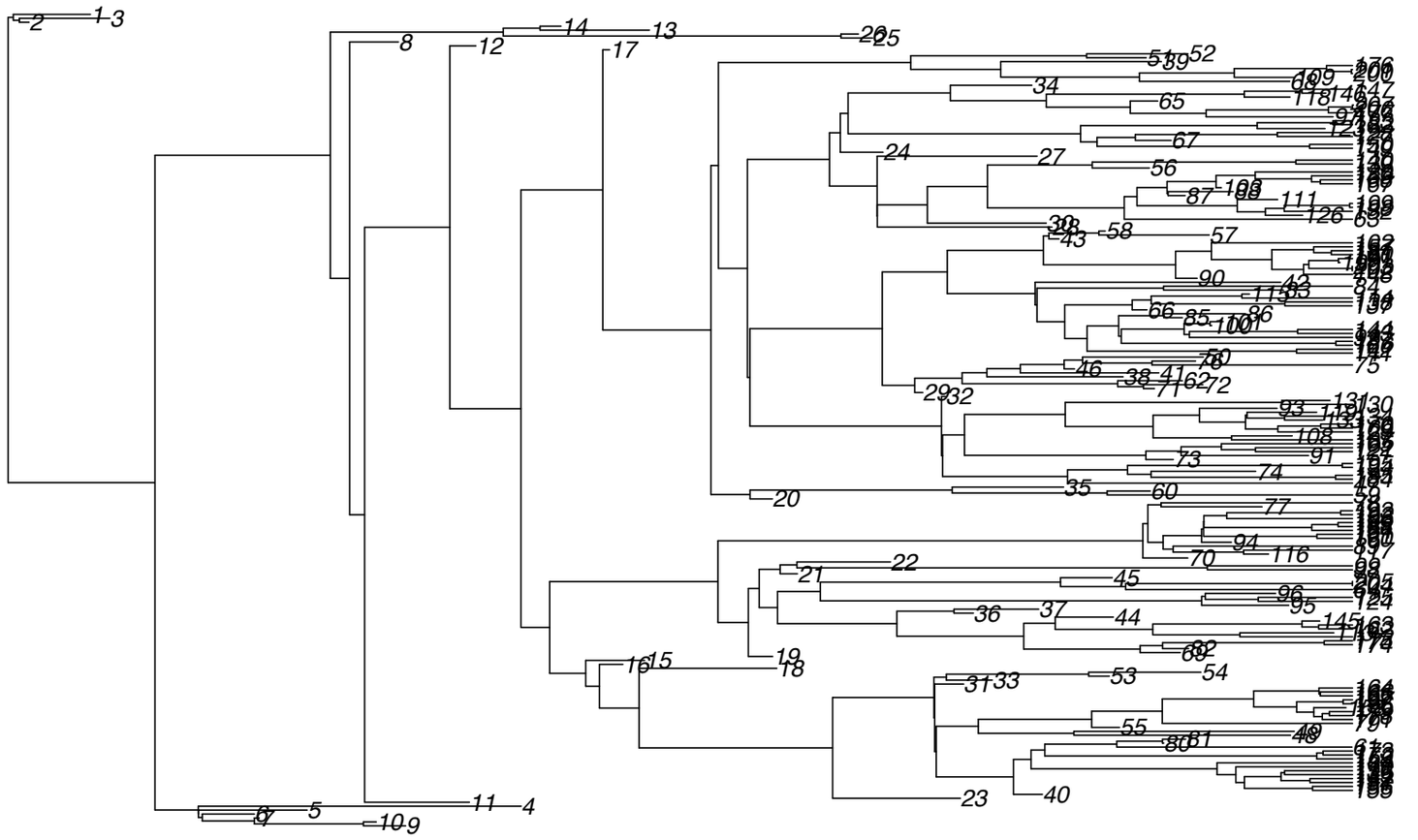
LTT Plot



Think about extinction

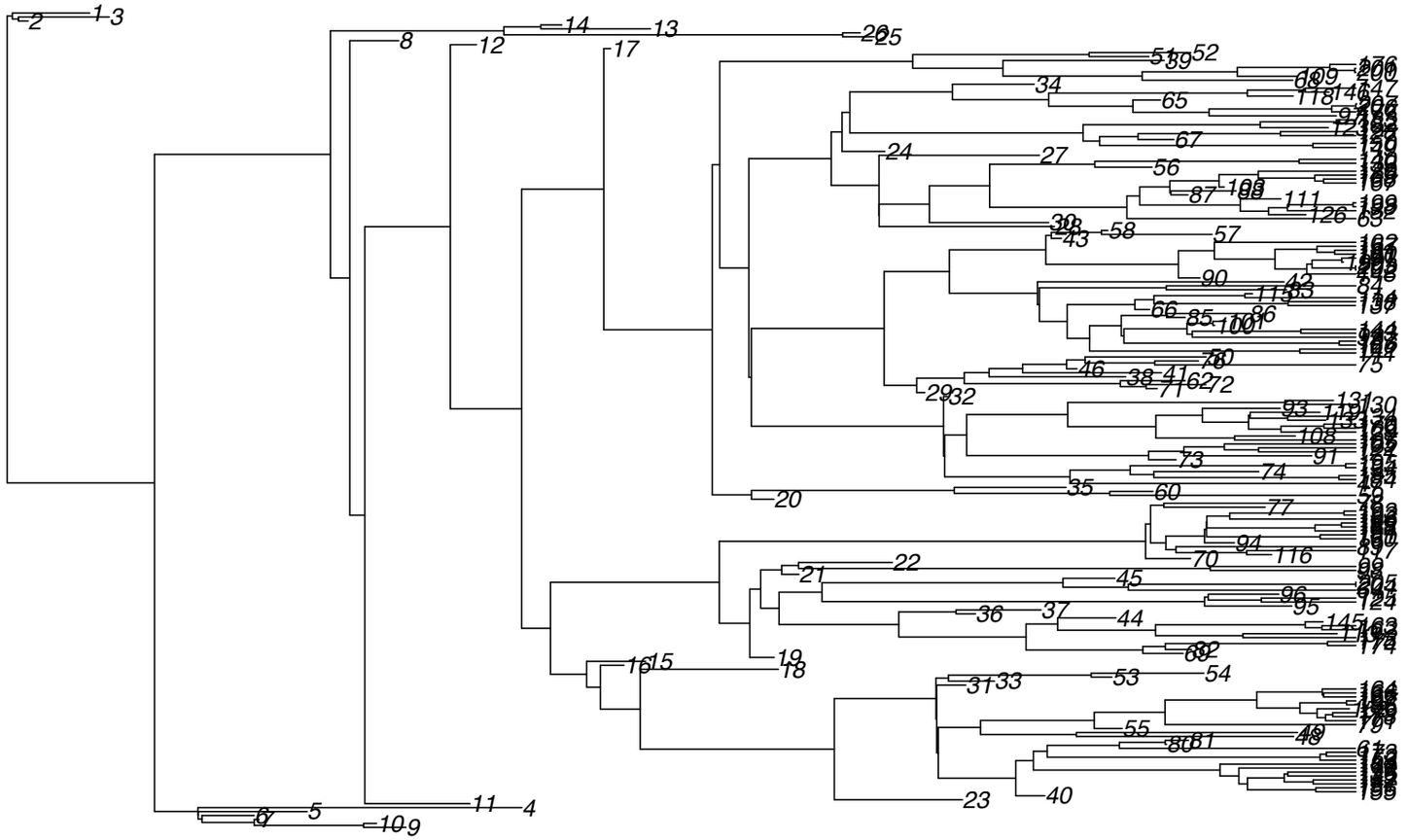
- how should extinction affect the rate of species accumulation in a phylogeny? why?
- can we infer extinction from phylogenetic data?

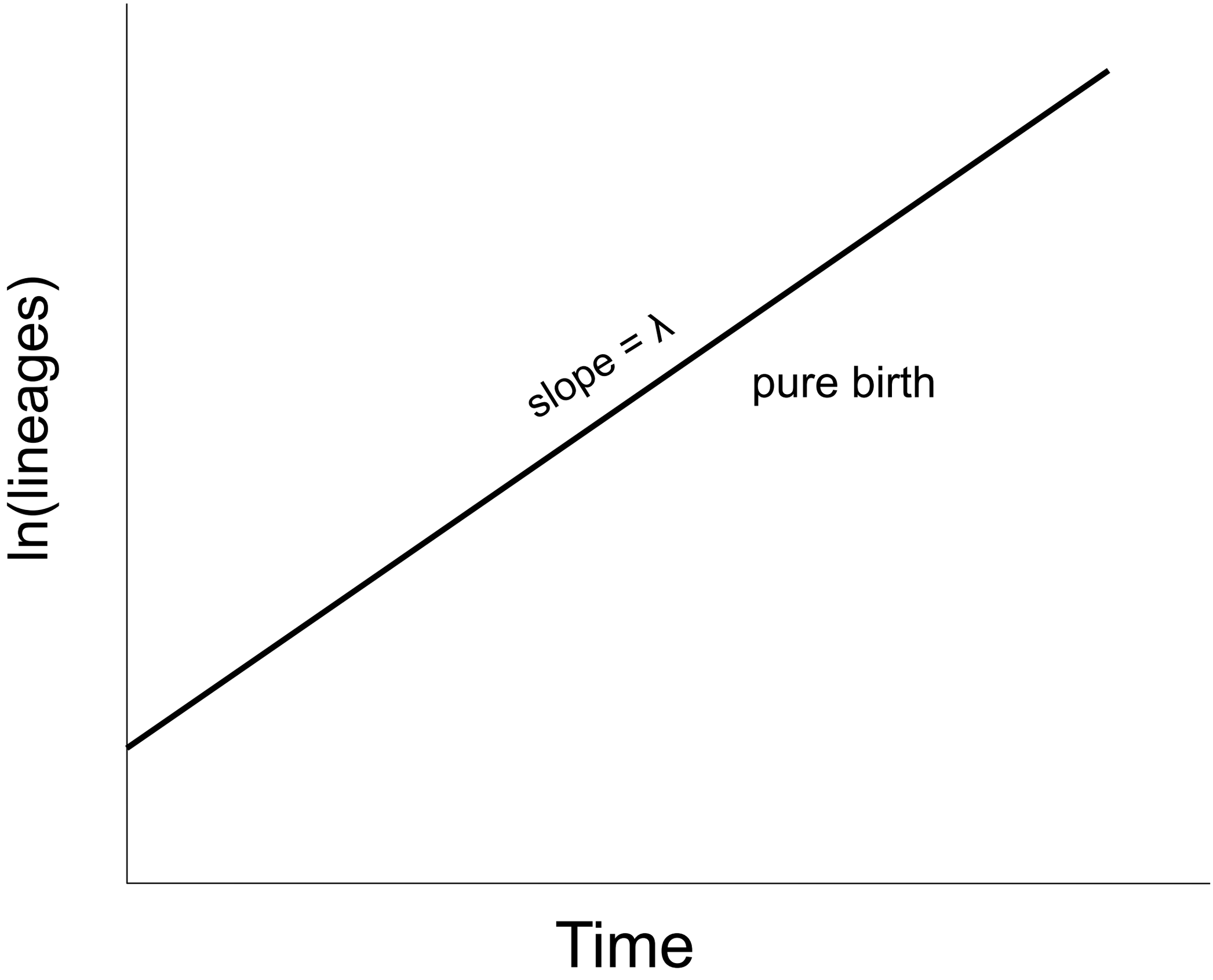


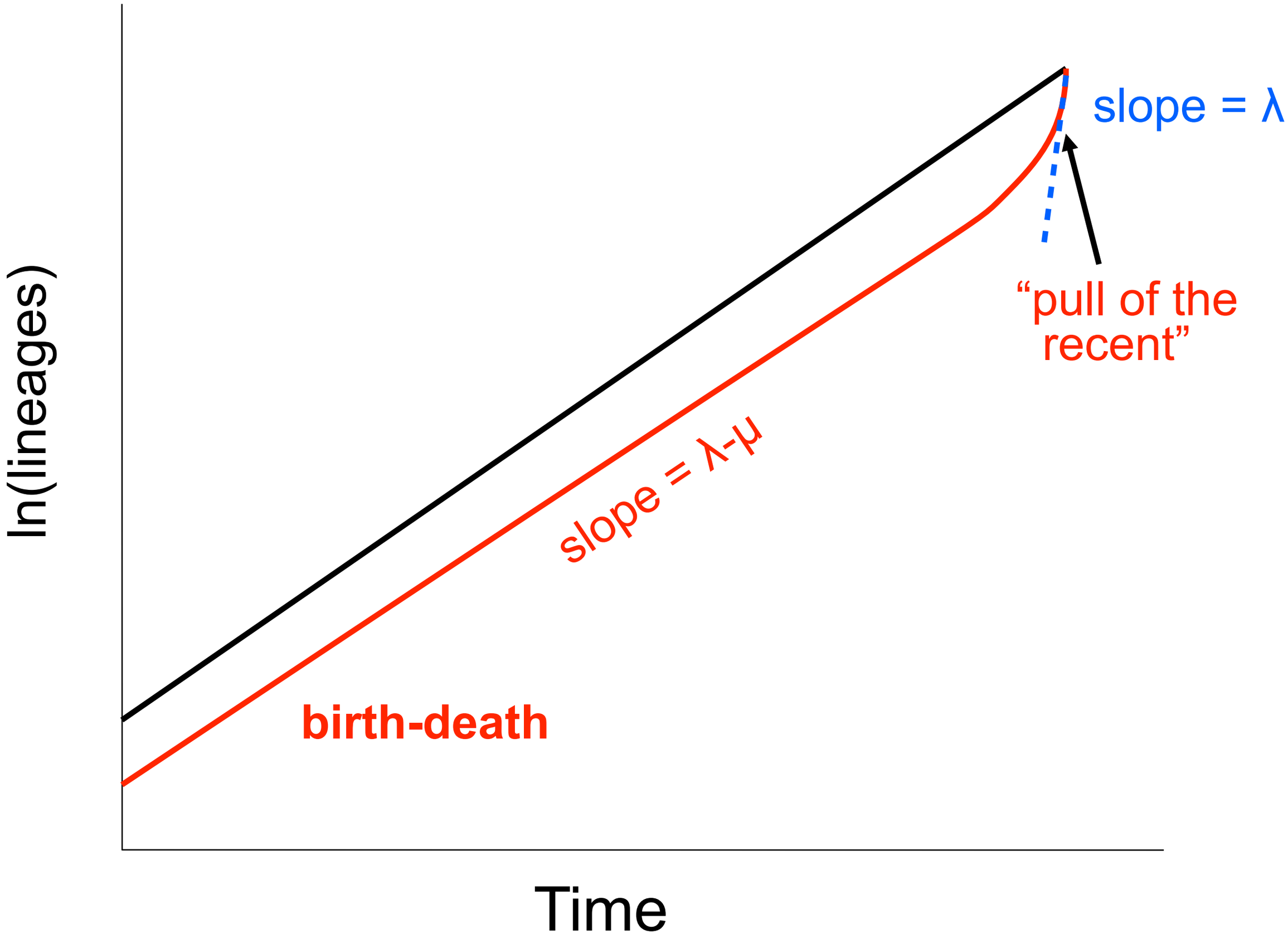


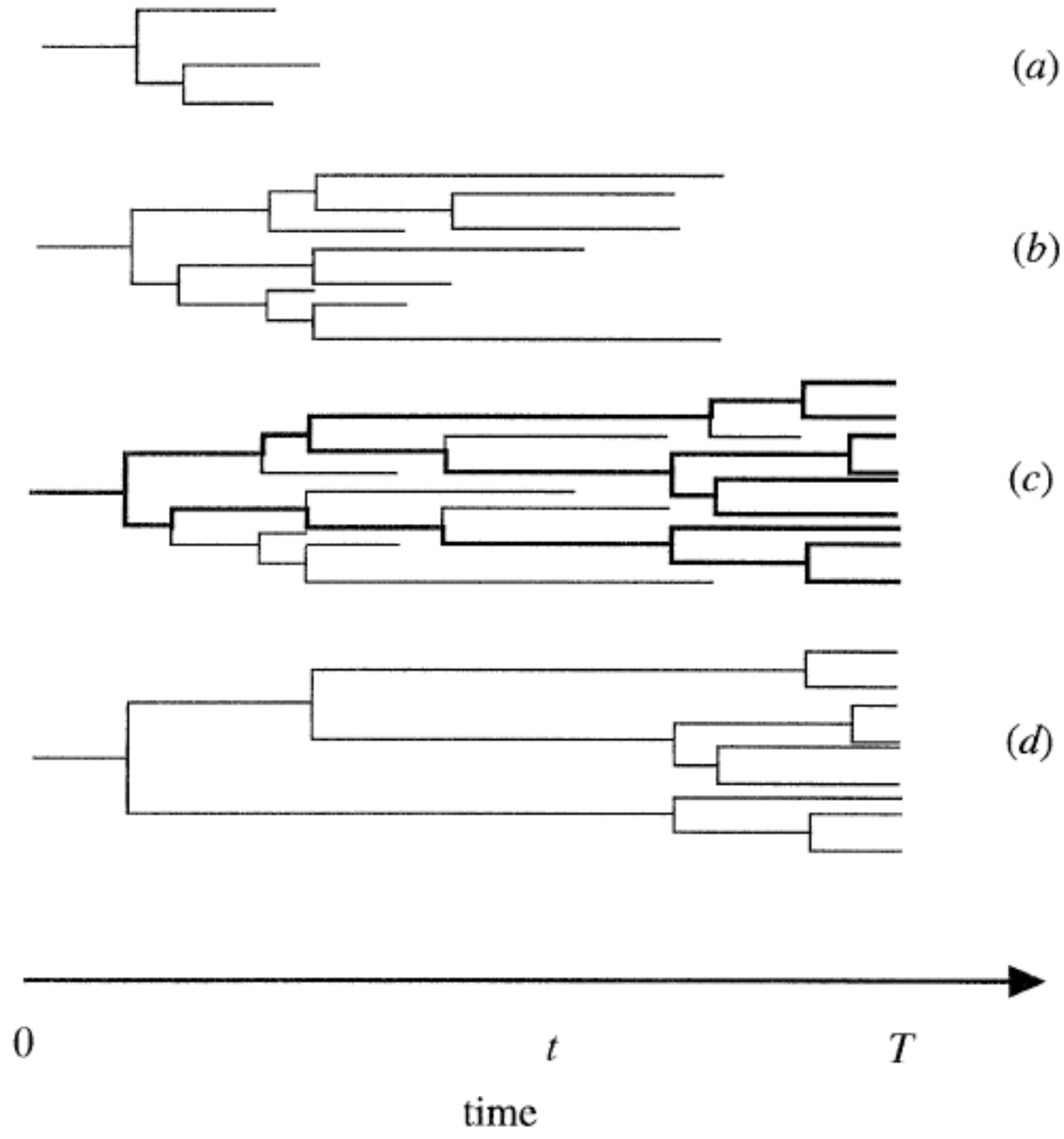
Extinction

- Can leave an imprint on present-day phylogenies
- Older lineages are more likely to have gone extinct than younger ones



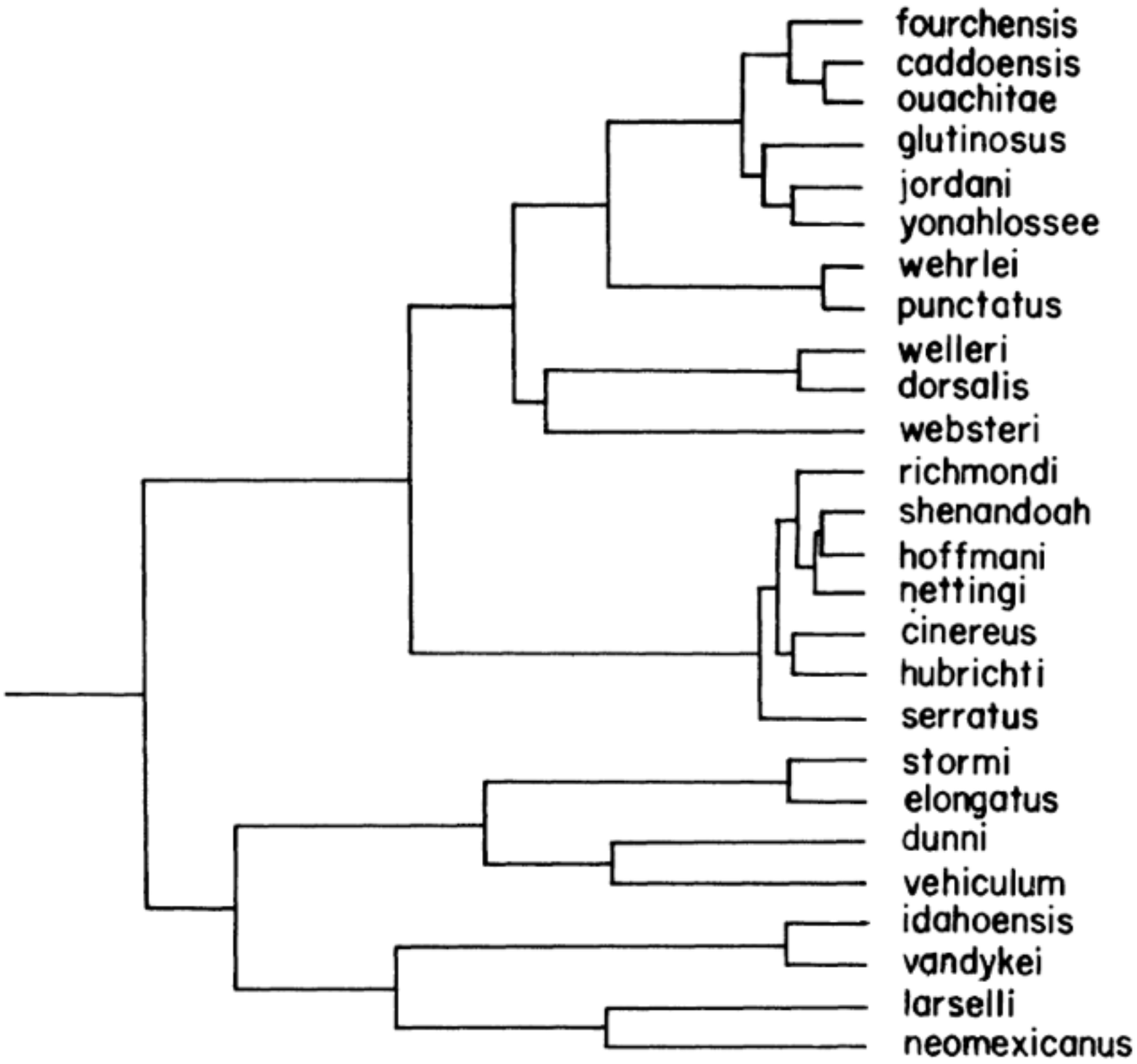

ln(lineages)
slope = λ
pure birth
Time


ln(lineages)
slope = λ
“pull of the recent”
slope = λ-μ
birth-death
Time



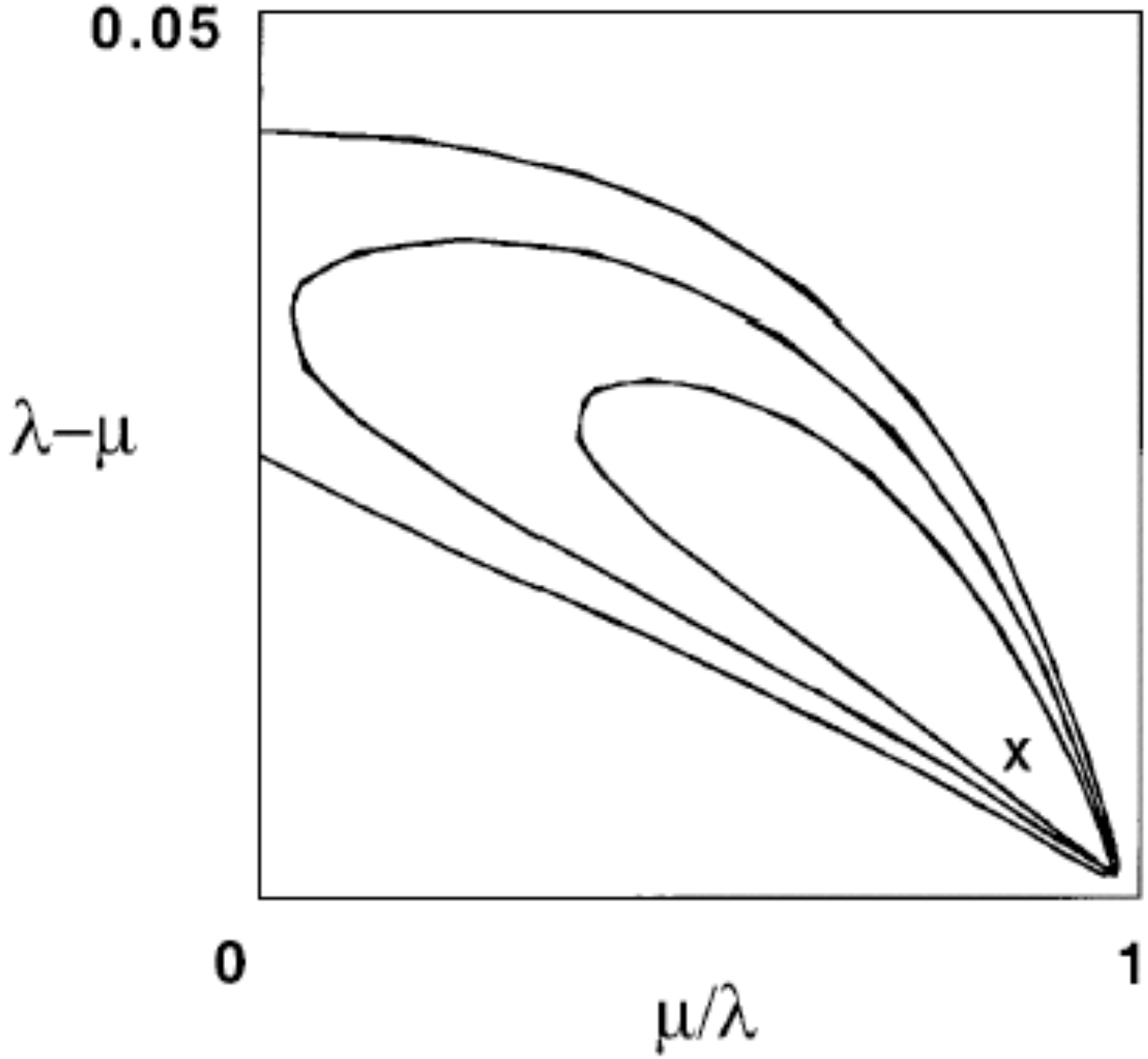
Nee S., May R.M., Harvey P.H. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B Biol. Sci. 344:305–311.

Plethodontid salamanders



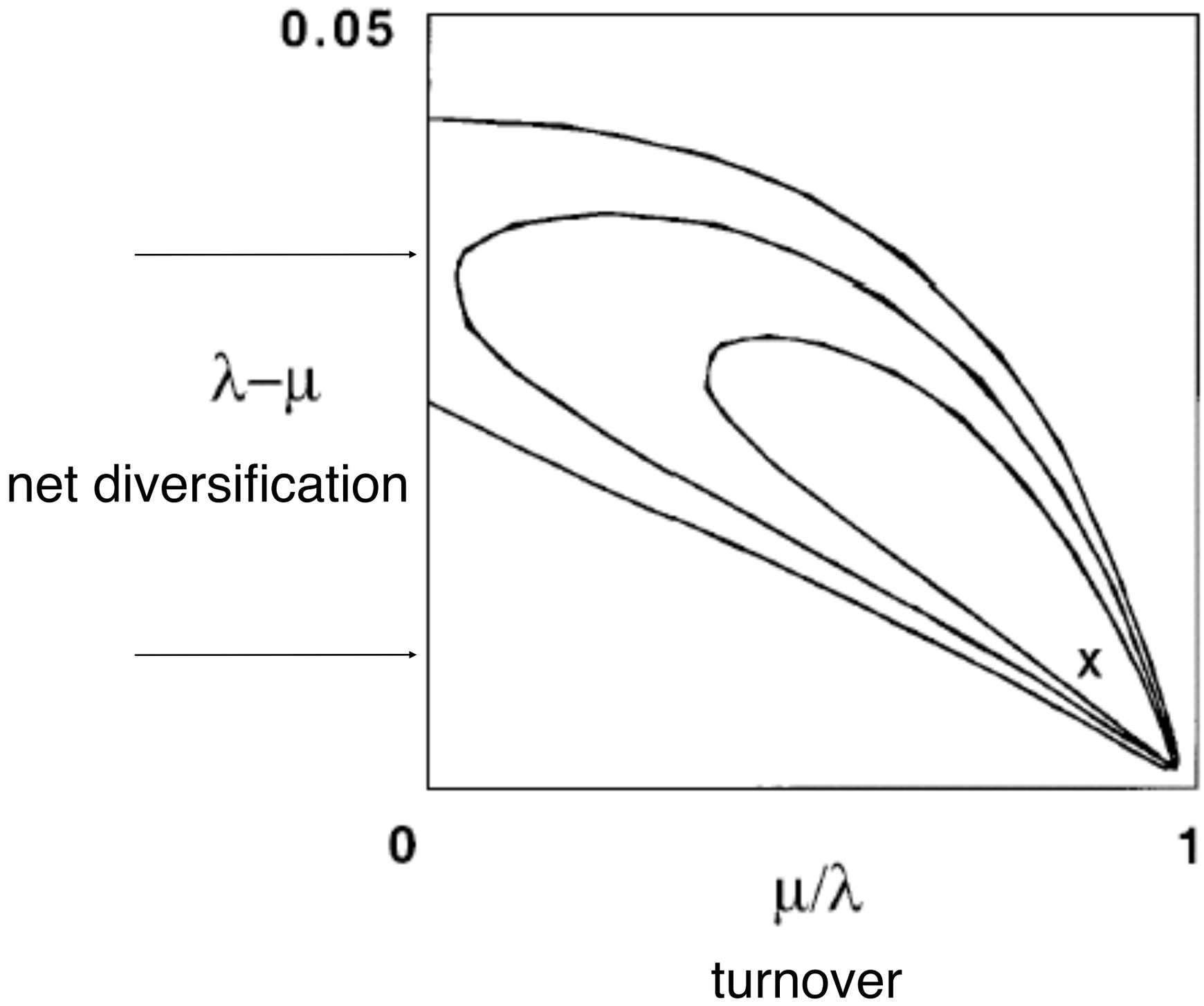
(Highton and Larson 1979)

Plethodontid salamanders



Nee S. 2001. Inferring speciation rates from phylogenies. Evolution. 55:661–668.

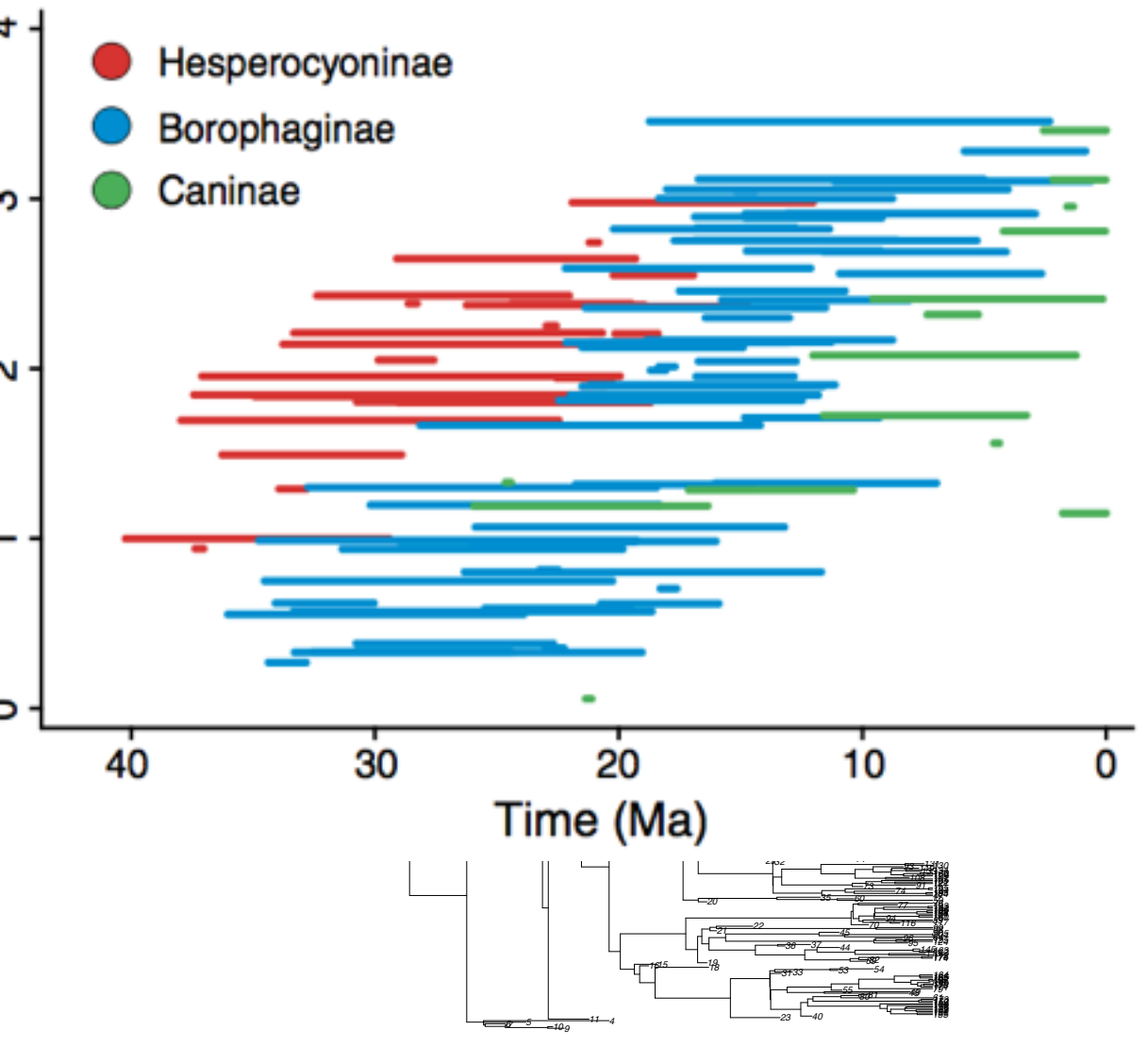
Plethodontid salamanders

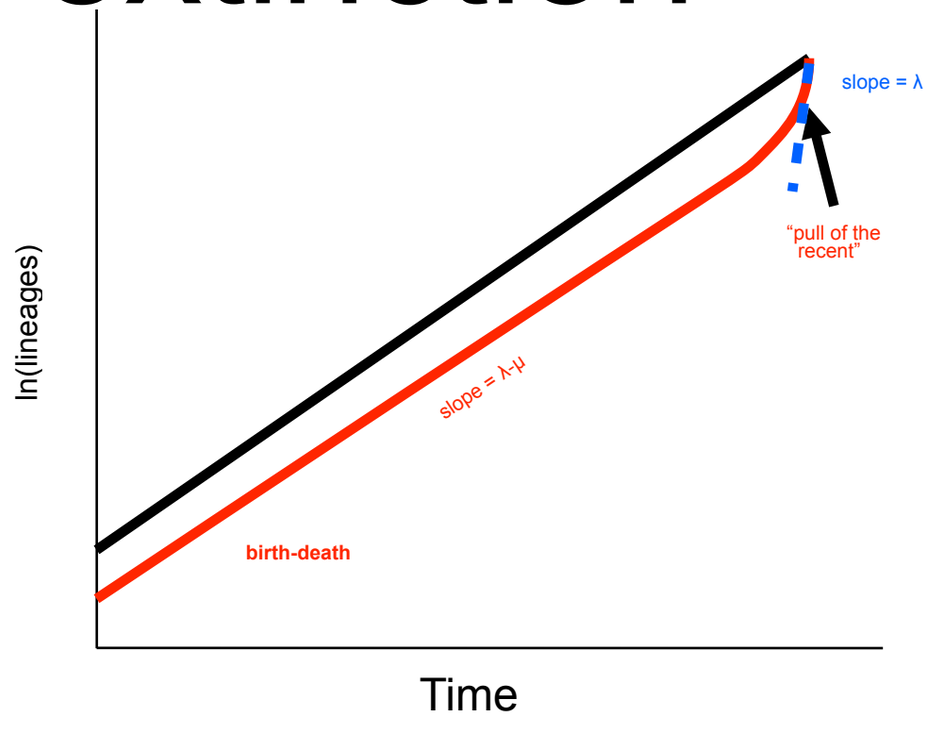


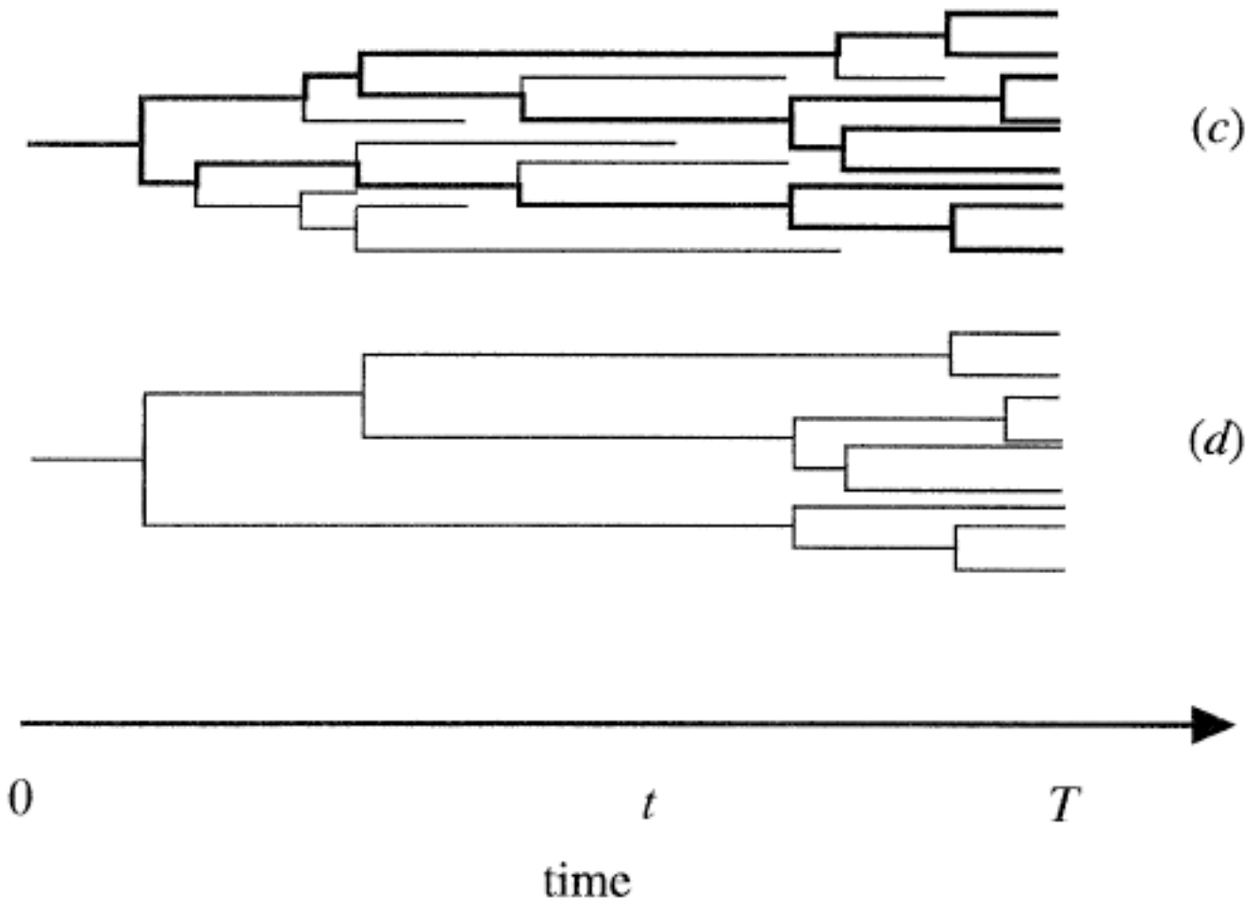
Nee S. 2001. Inferring speciation rates from phylogenies. Evolution. 55:661–668.

We can use ML or Bayesian methods to analyze data, estimate parameters, and carry out model selection

fossil versus molecular signature of extinction

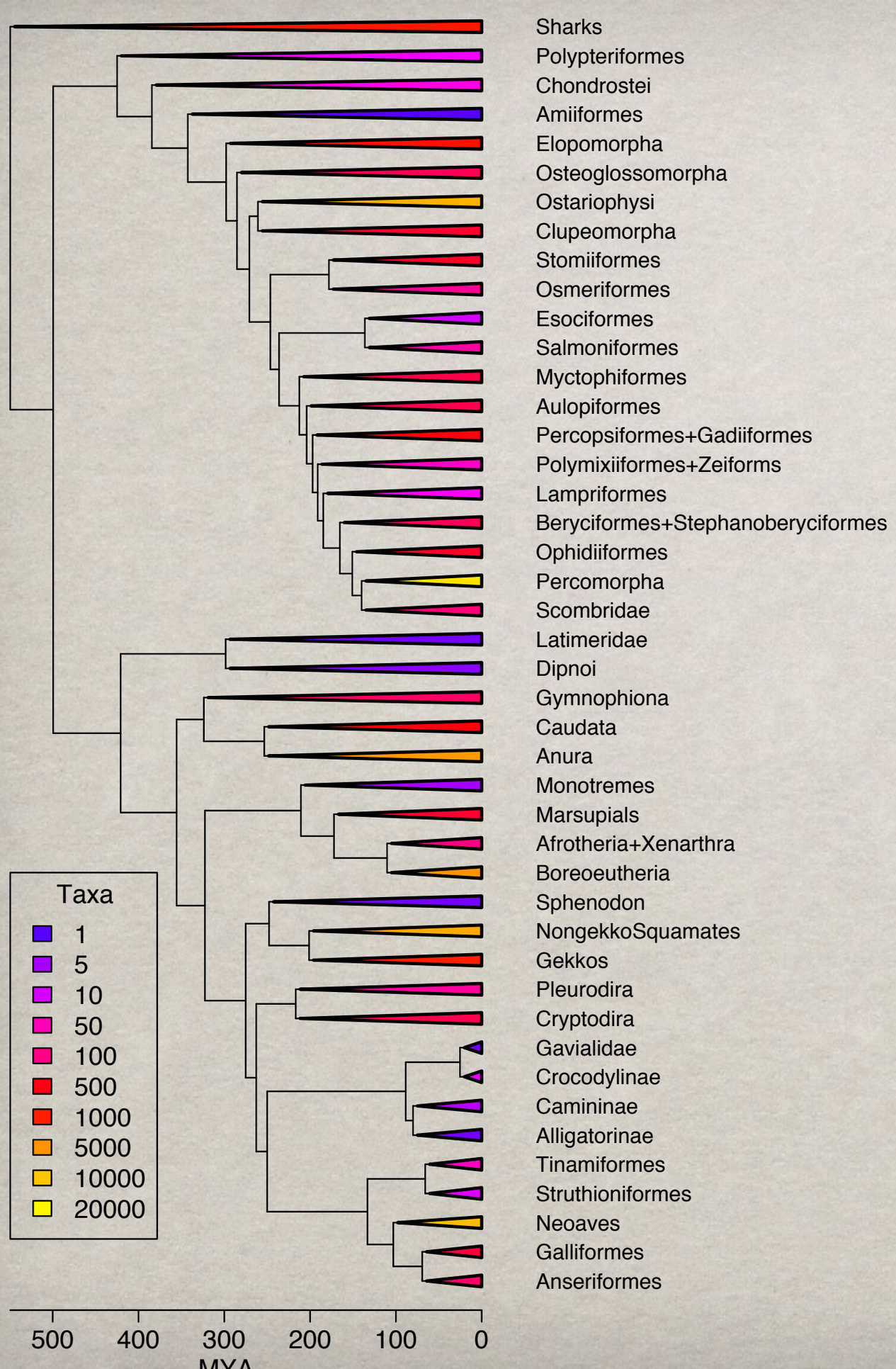






Silvestro D., Antonelli A., Salamin N., Quental T.B. 2015. The role of clade competition in the diversification of North American canids. Proc. Natl. Acad. Sci. U. S. A. 112:8684–8689.

Models have grown in sophistication since Nee 2001



How Fast Did Gnathostomes Diversify?

1. $\lambda_G = [\log(n) - \log 2]/t$

n = 57,859 species

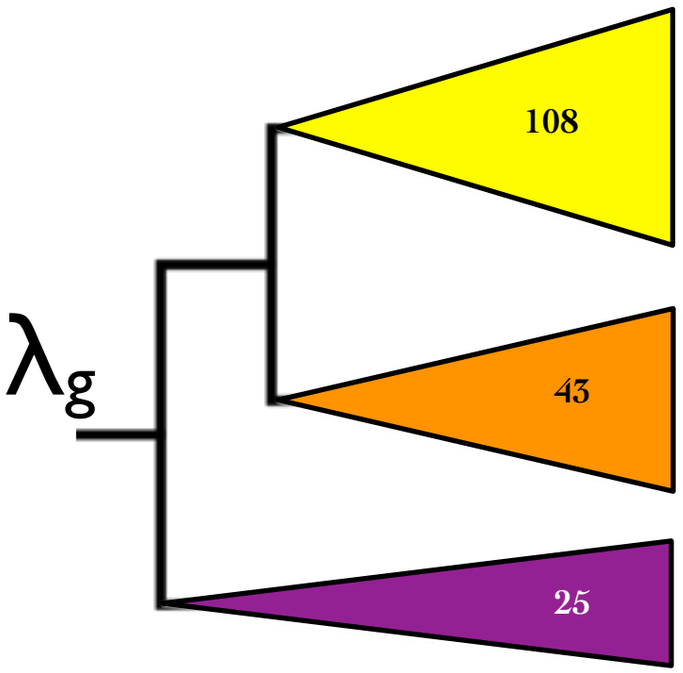
t = 549 mya

$$\lambda_G = \frac{\ln(n) - \ln(2)}{t}$$

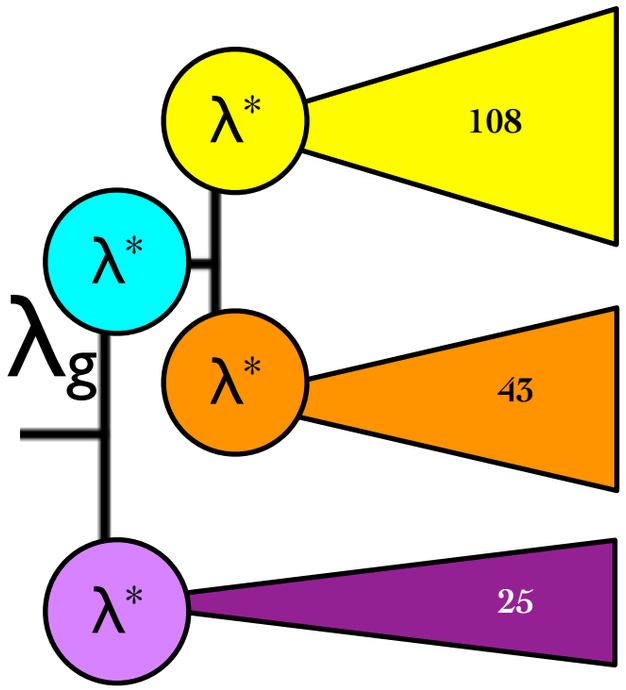
$$\lambda_G = \frac{\ln(57859) - \ln(2)}{549}$$

$$\lambda_G = 0.019$$

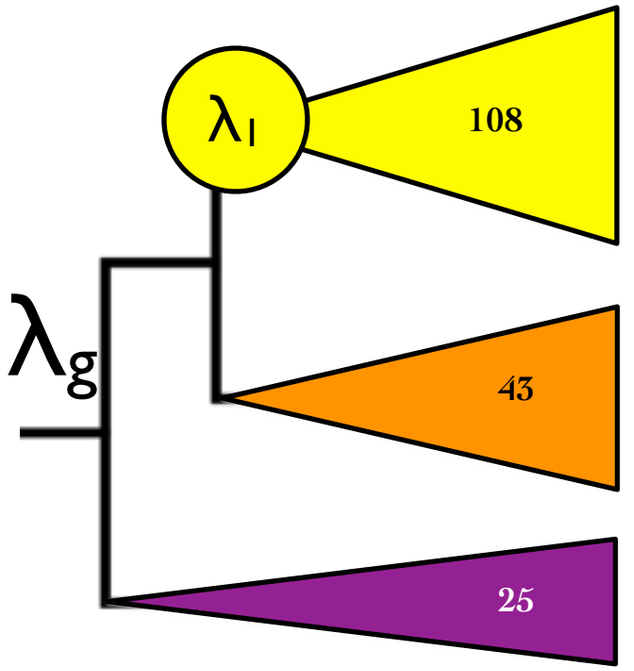
Pure Birth
High Turnover (ε=0.9985)
Percomorpha
Neoaves
crocodylians
Amia
Species Richness
10000
100
1
100
200
300
400
500
Stem Age (MY)



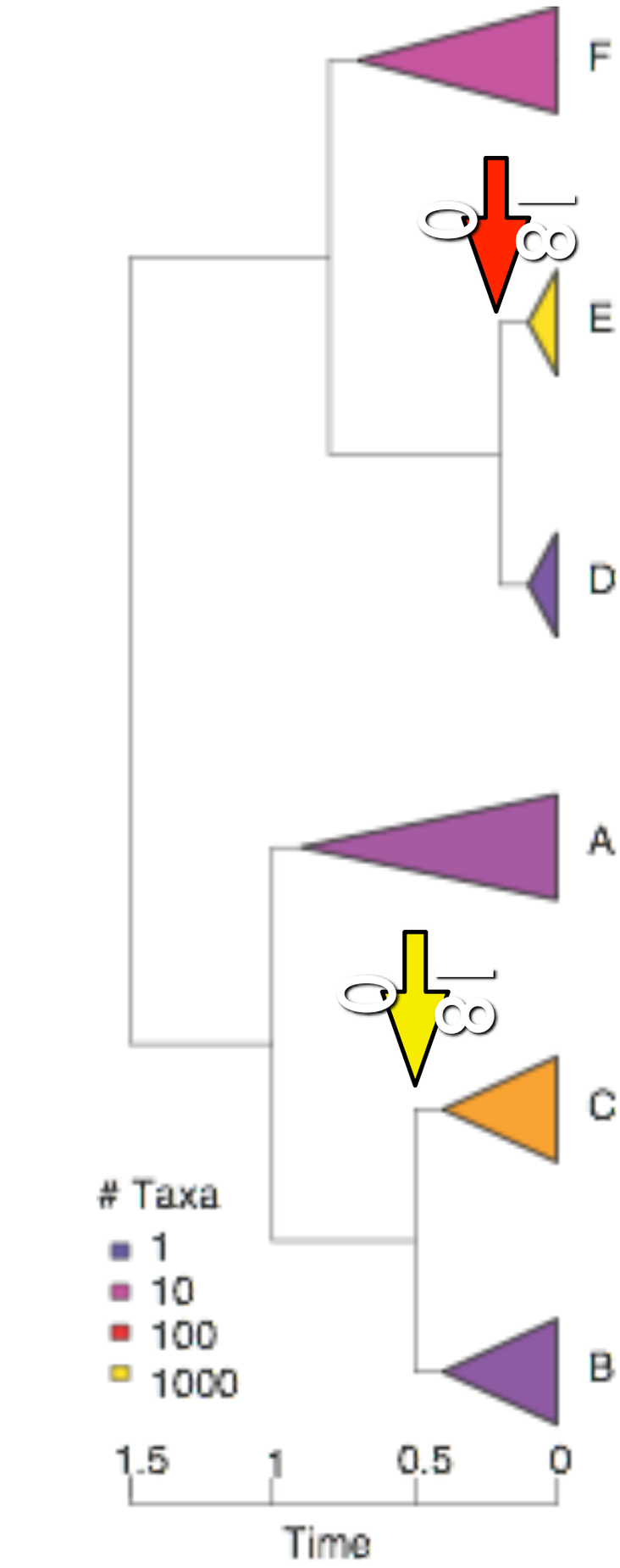
1. λ for whole tree

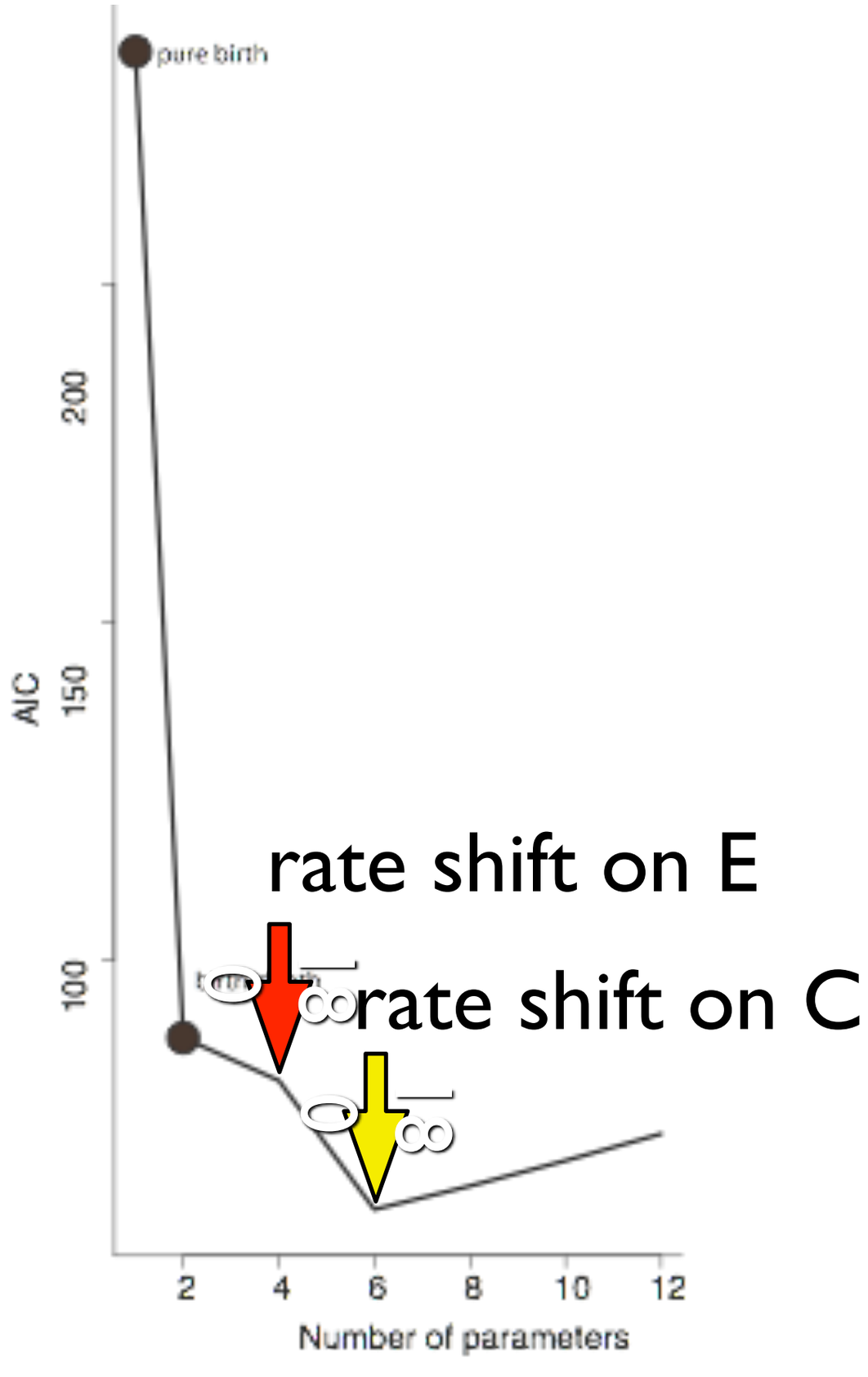


2. allow λ to change on each branch



3. if ΔAIC of best new λ and λ_g > 4, retain rate & repeat


F
E
D
A
C
B
Taxa
1
10
100
1000
1.5
1
0.5
0
Time


pure birth
200
150
100
AIC
rate shift on E
rate shift on C
2
4
6
8
10
12
Number of parameters

What Can MEDUSA Do For You?

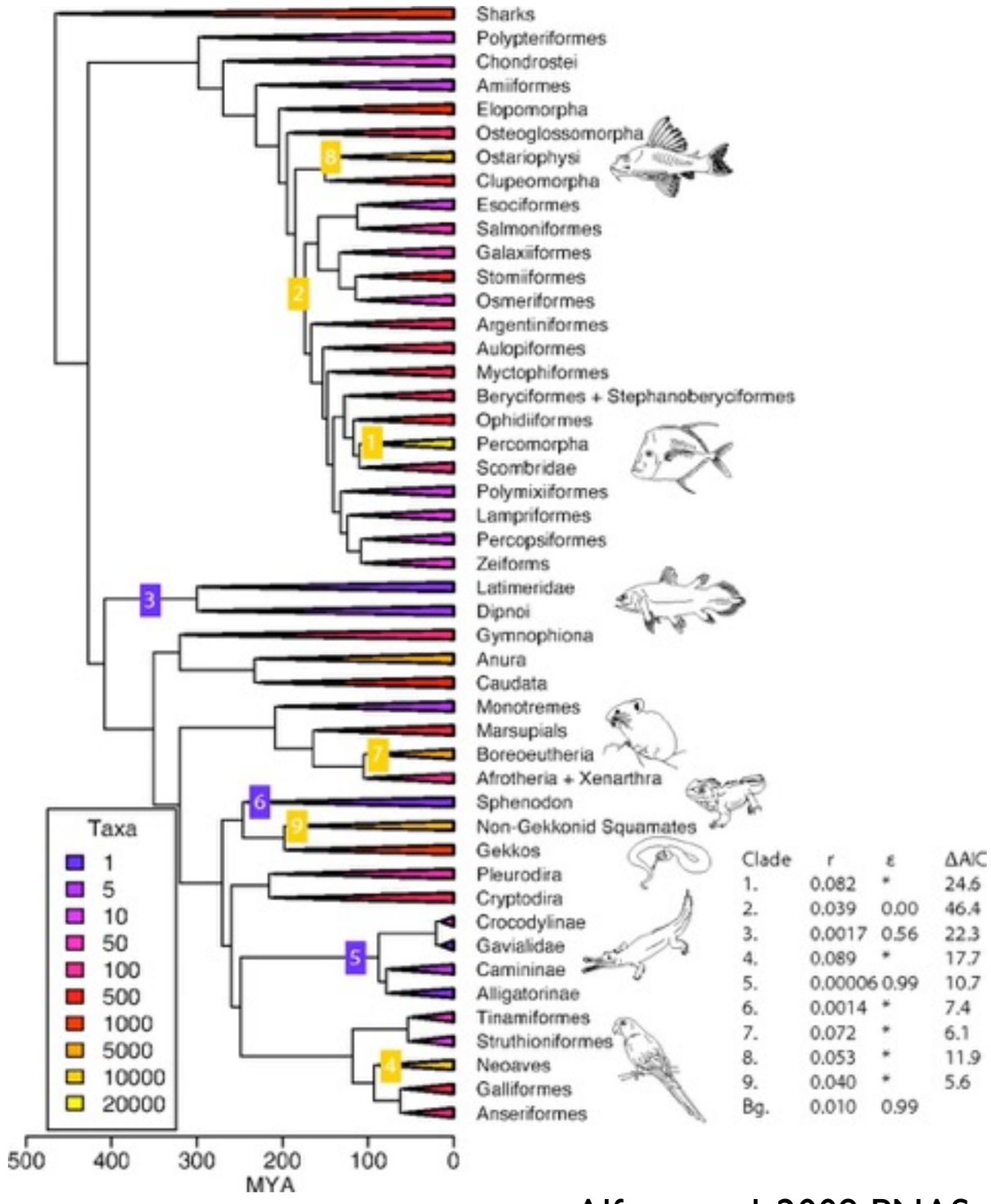
MEDUSA

- Learn about diversification dynamics through fitting <u>piecewise</u> models of birth (λ) and death (μ)

  - Where have rates changed?
  - By what magnitude?
  - Why do clades differ in size?

r = net diversification rate
  = λ - μ

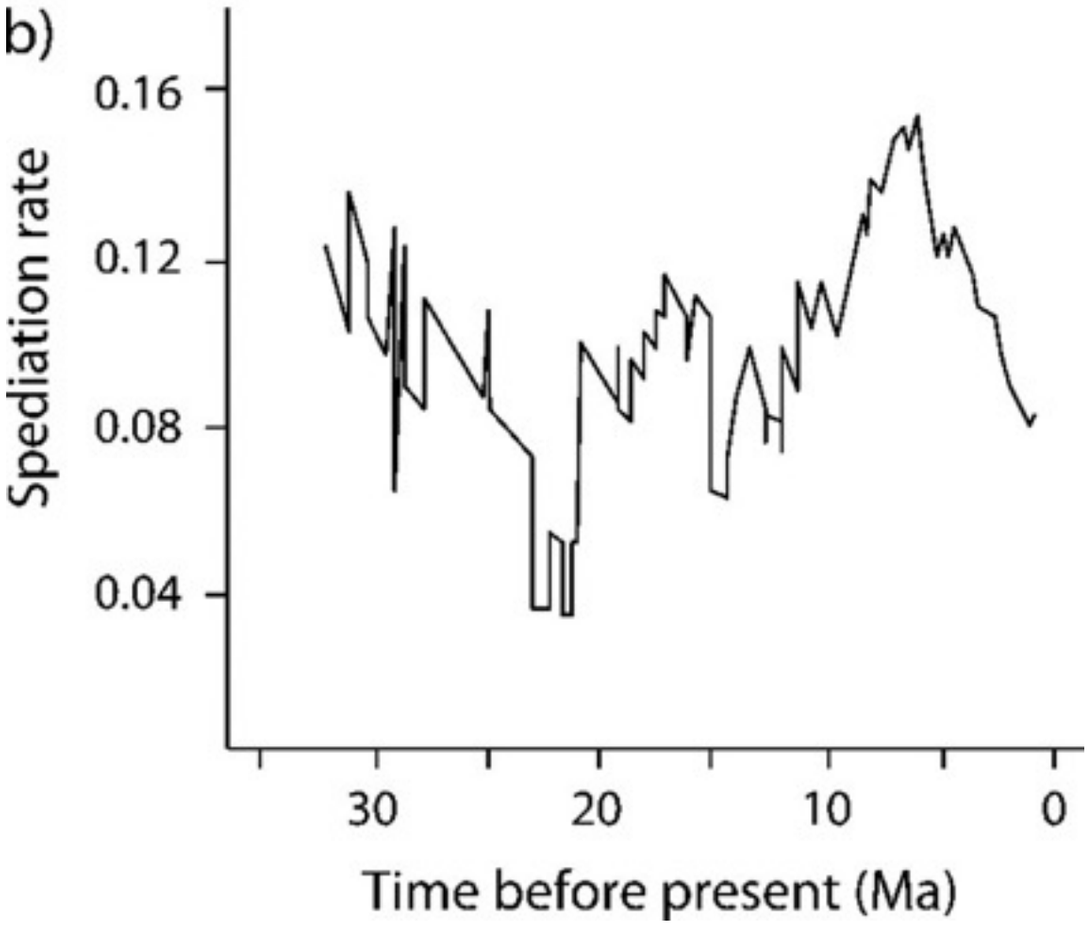
ε = extinction fraction
  = μ / λ



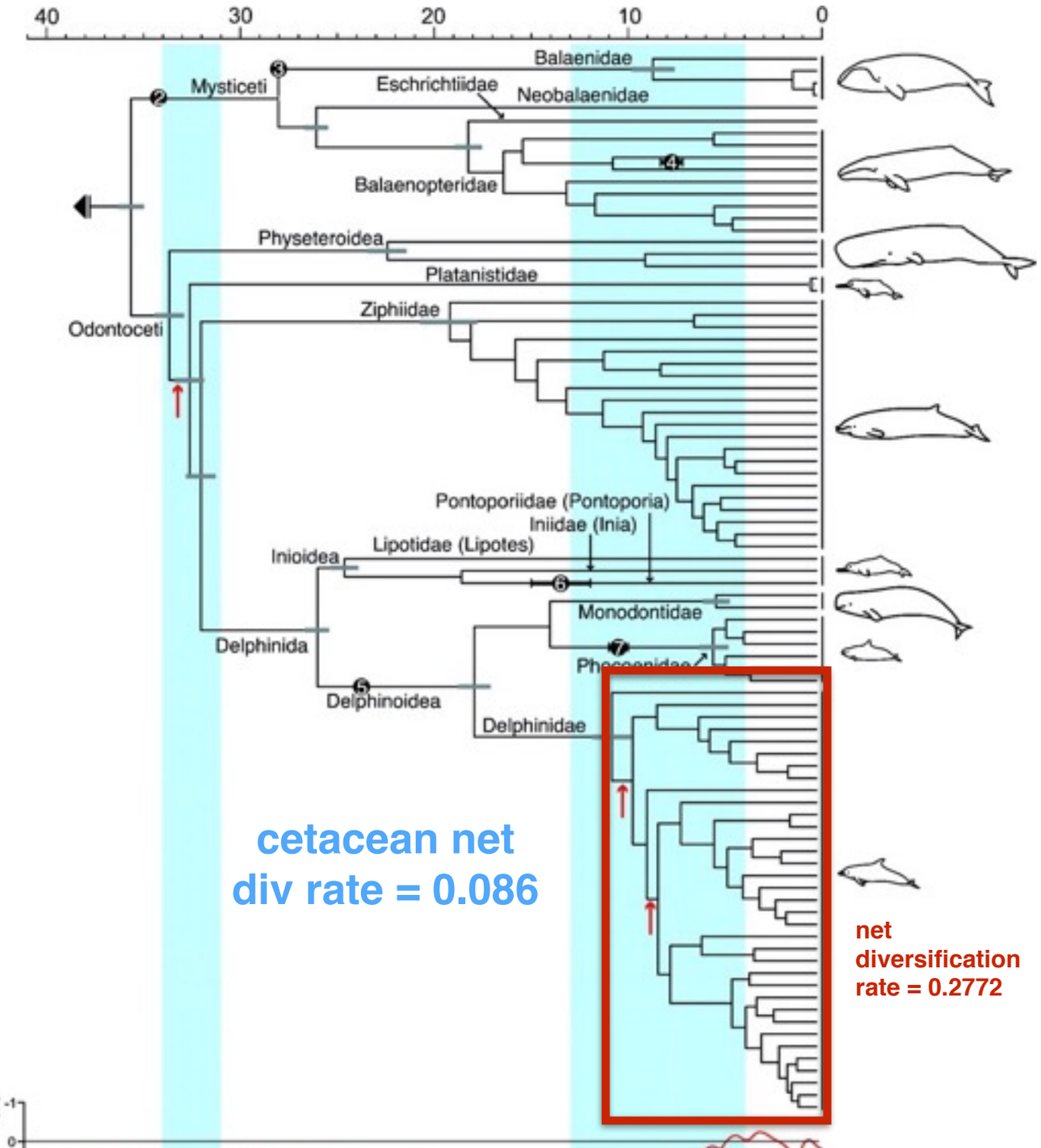
Alfaro et al. 2009 PNAS

More recent models allow for diversity dependence

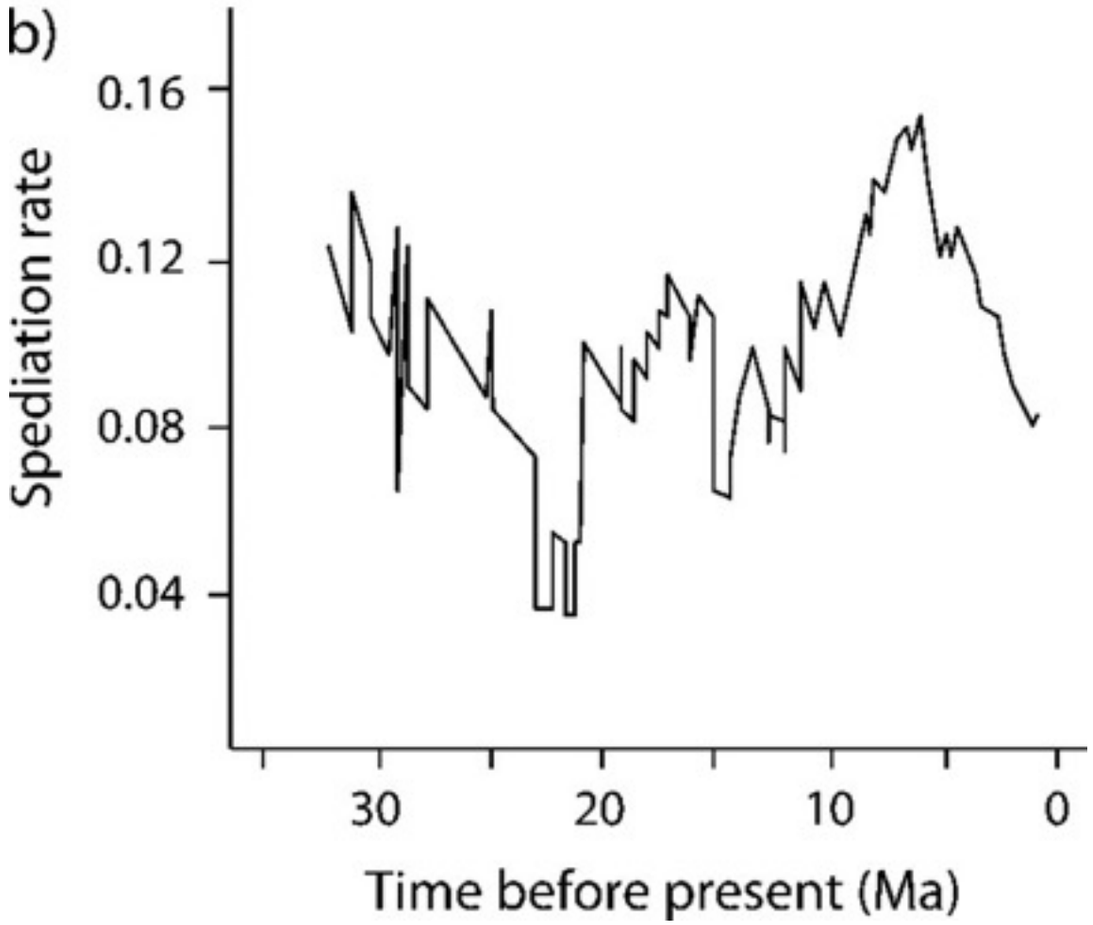
RPANDA and trait independent diversification

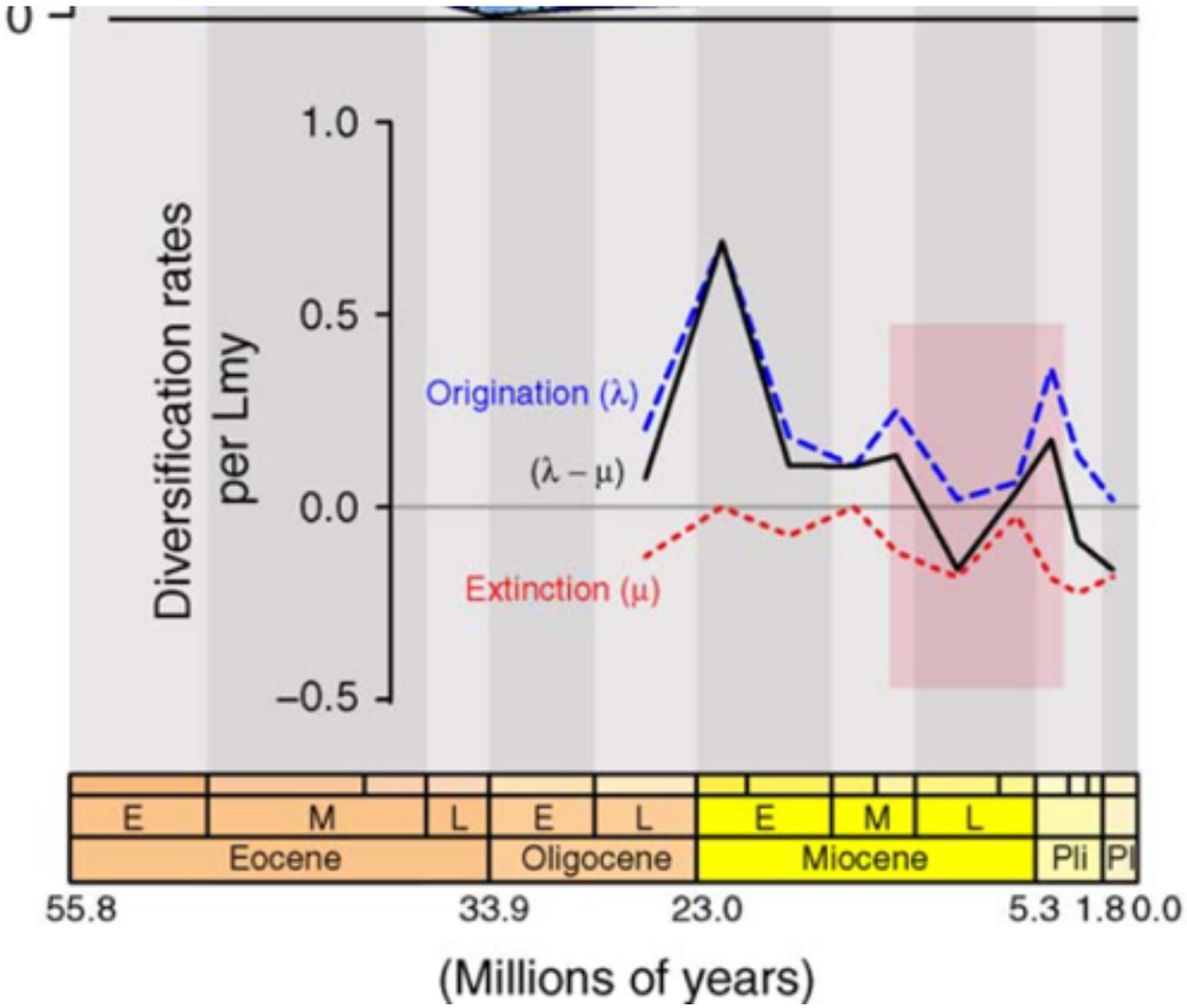


Steeman et al 2009

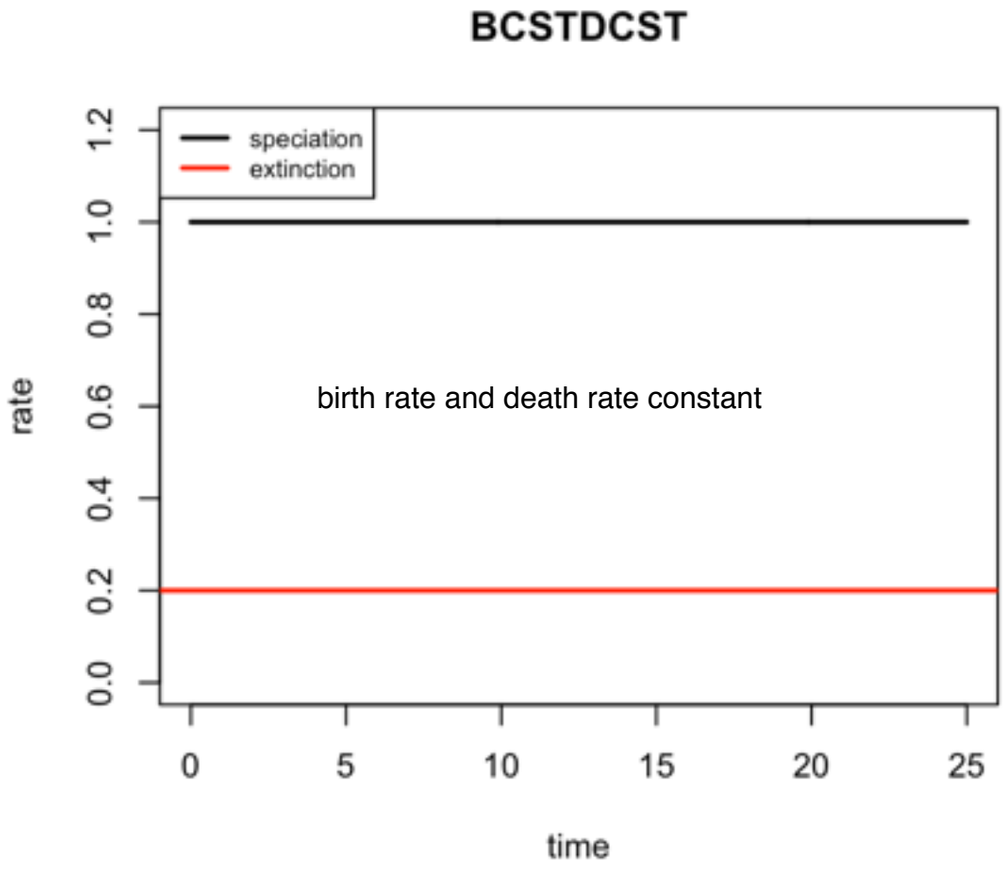


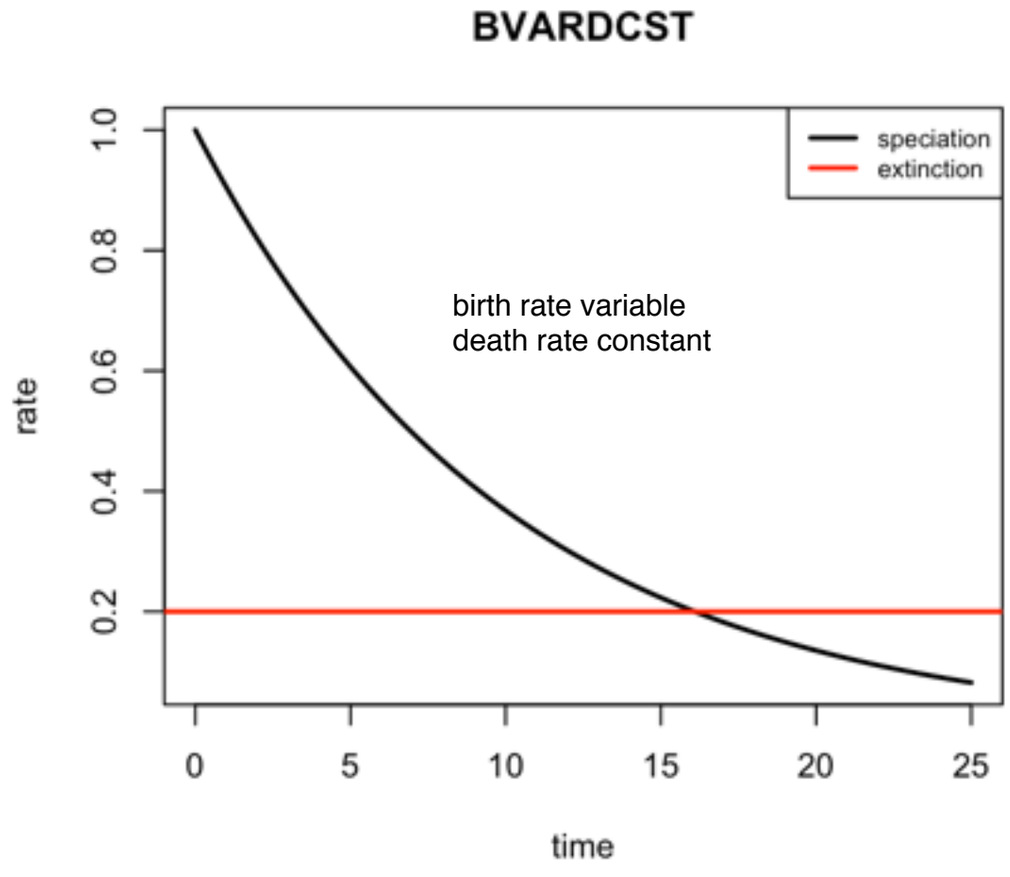
BAMM, RPANDA, and trait independent diversification

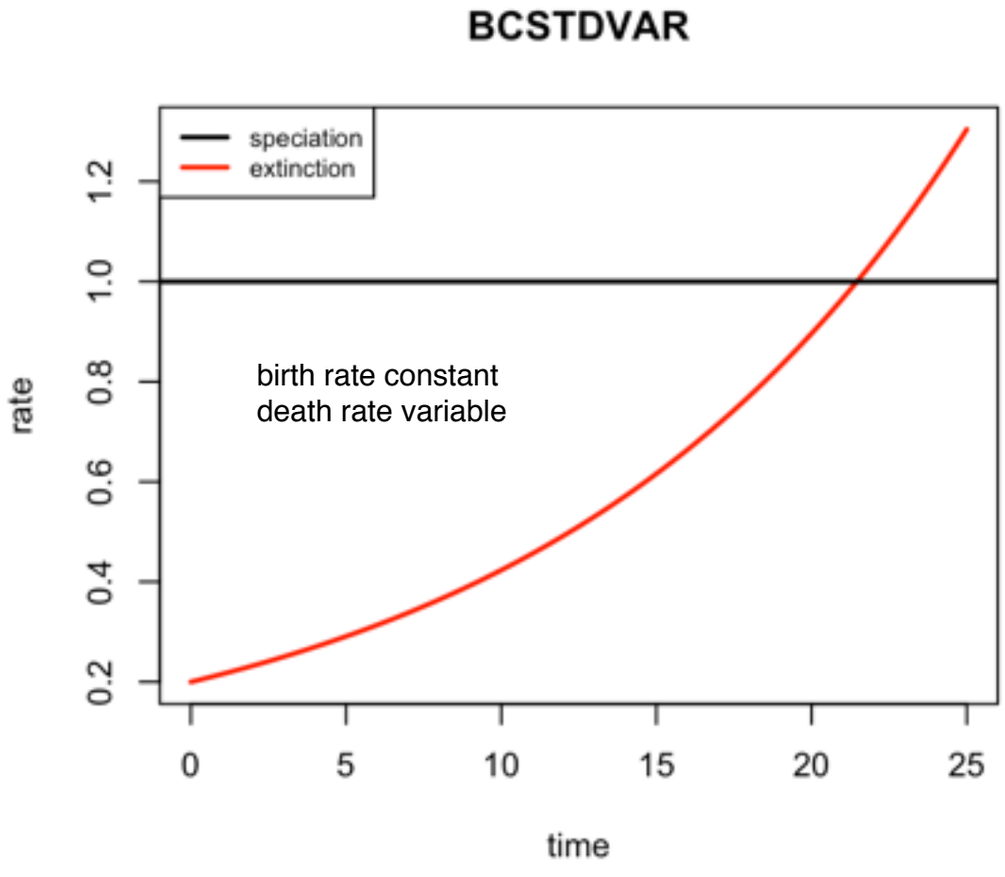


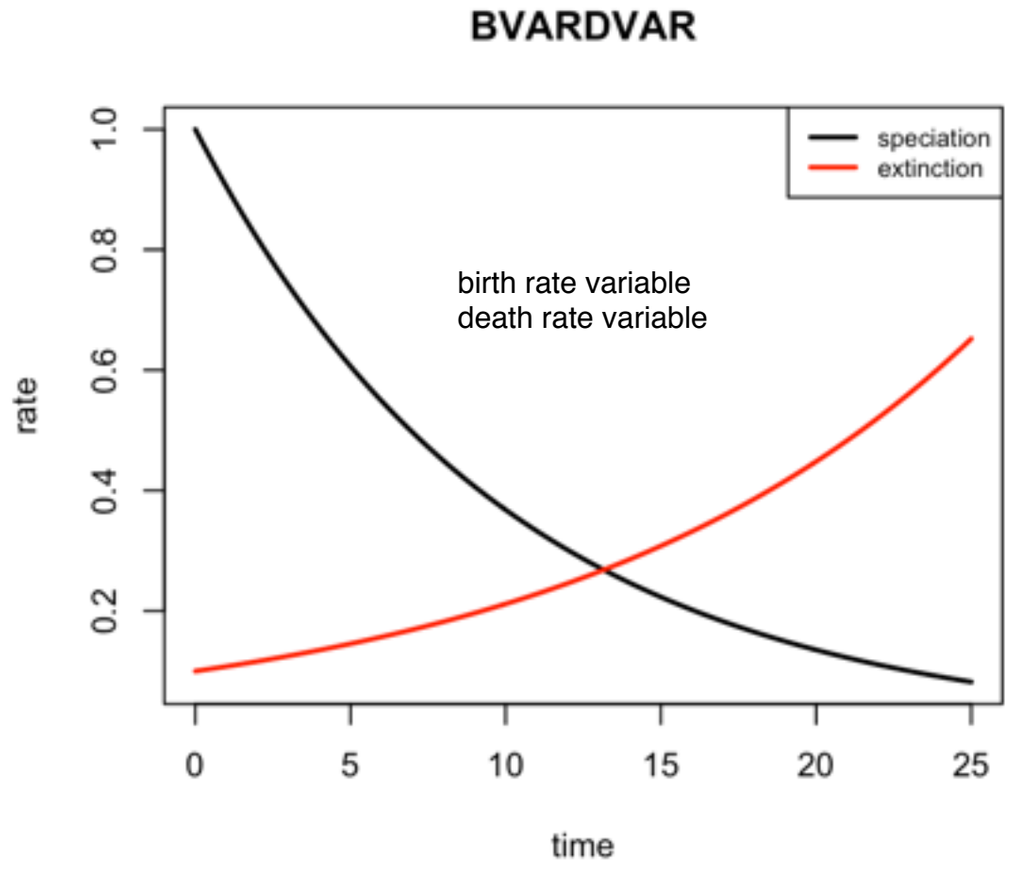


Steeman et al 2009


BCSTDCST
speciation
extinction
birth rate and death rate constant
rate
time
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
5
10
15
20
25


BVARDCST
speciation
extinction
birth rate variable
death rate constant
rate
time
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25

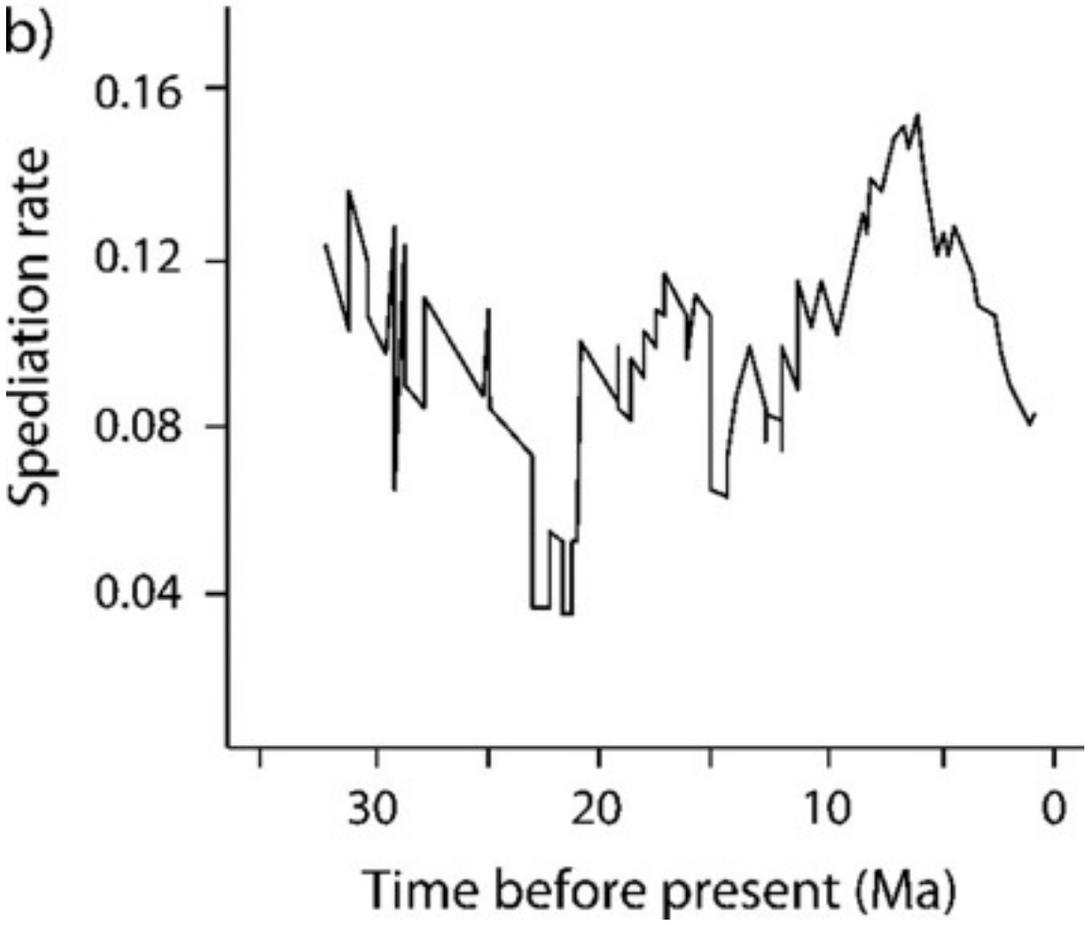

BCSTDVAR
speciation
extinction
birth rate constant
death rate variable
rate
time
0.2
0.4
0.6
0.8
1.0
1.2
0
5
10
15
20
25


BVARDVAR
speciation
extinction
birth rate variable
death rate variable
rate
time
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25

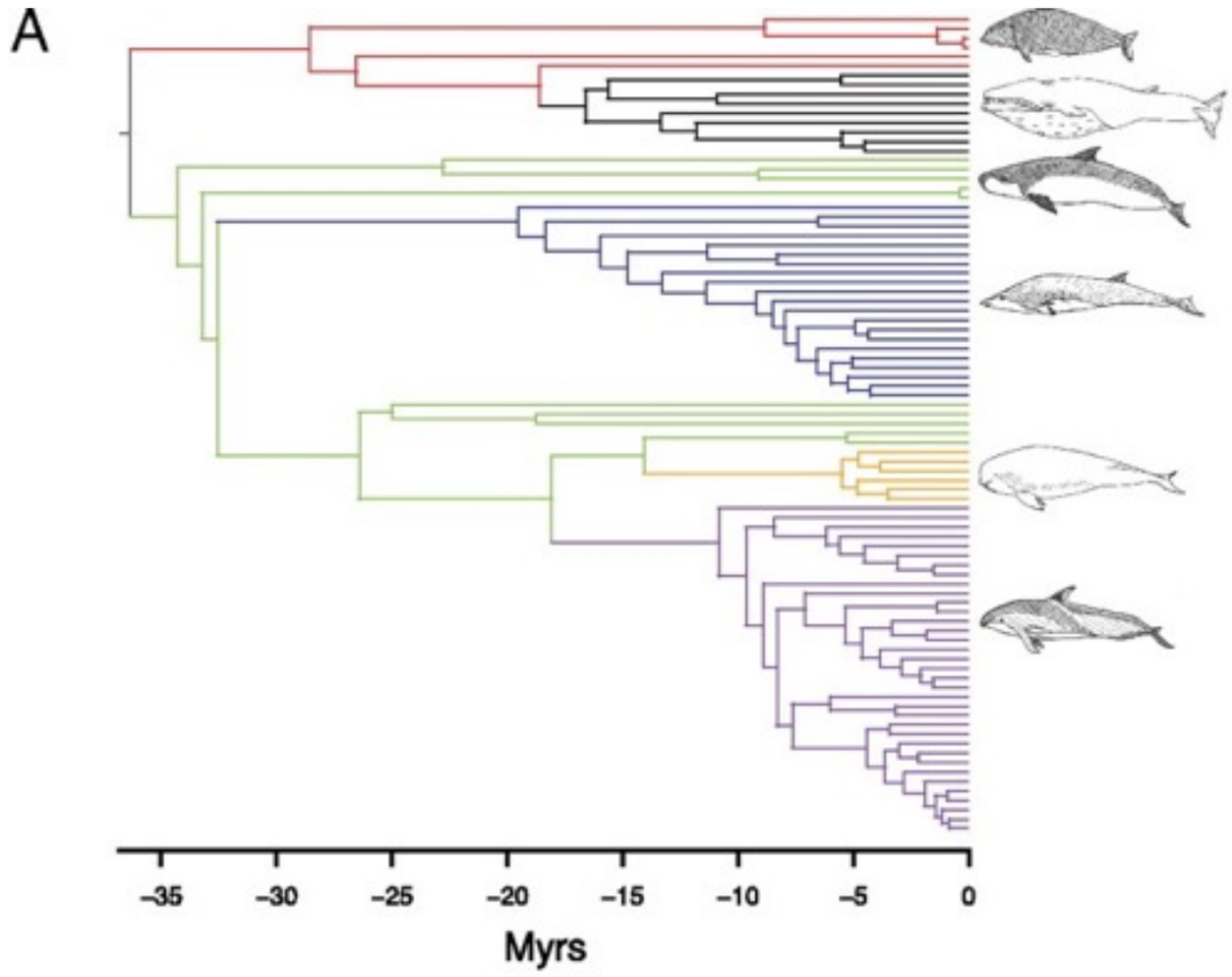
these models are more flexible than constant rates models

- allow rates to change with time over a tree
- avoid the “carnation problem” by allowing extinction to be higher than speciation
- implemented in the RPANDA package in R
- requires subclades to be identified a-priori
- give better match of molecular phylogenetic and fossil inferred diversification dynamics
- *still do not integrate the two data types*

BAMM, RPANDA, and trait independent diversification

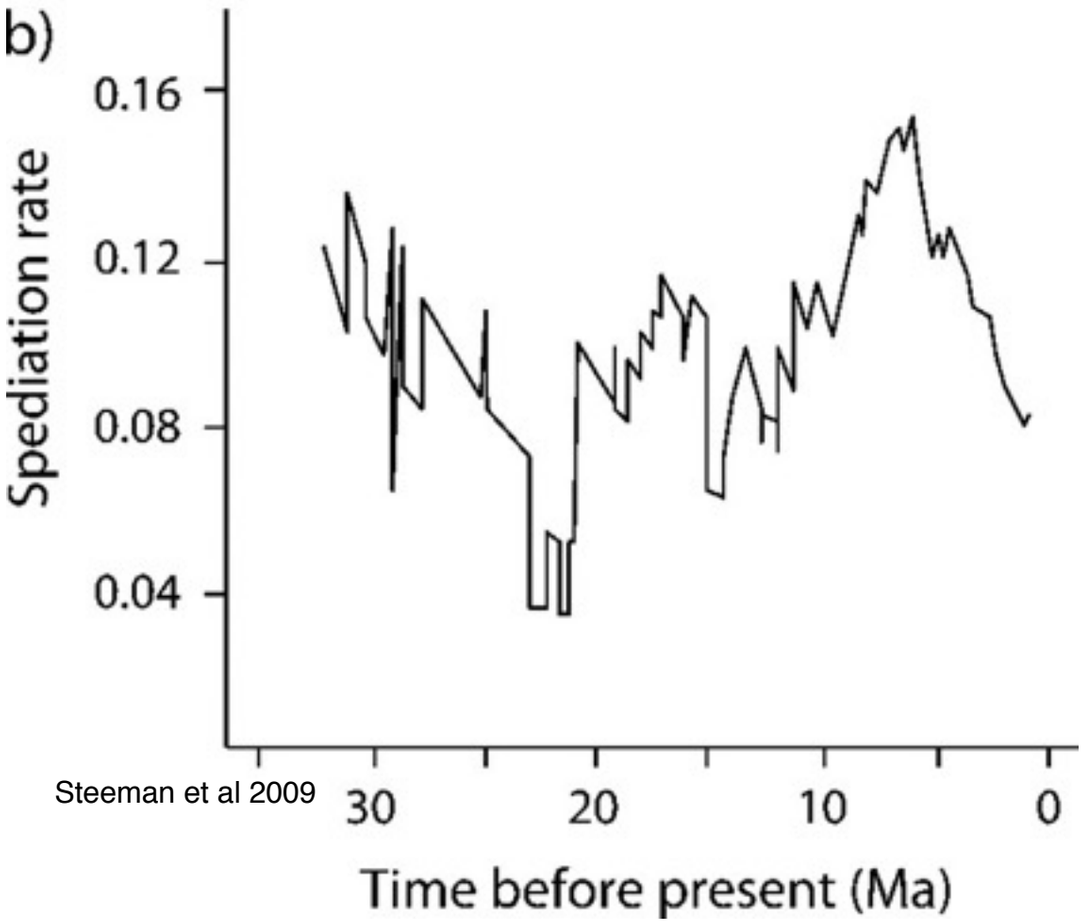


Steeman et al 2009

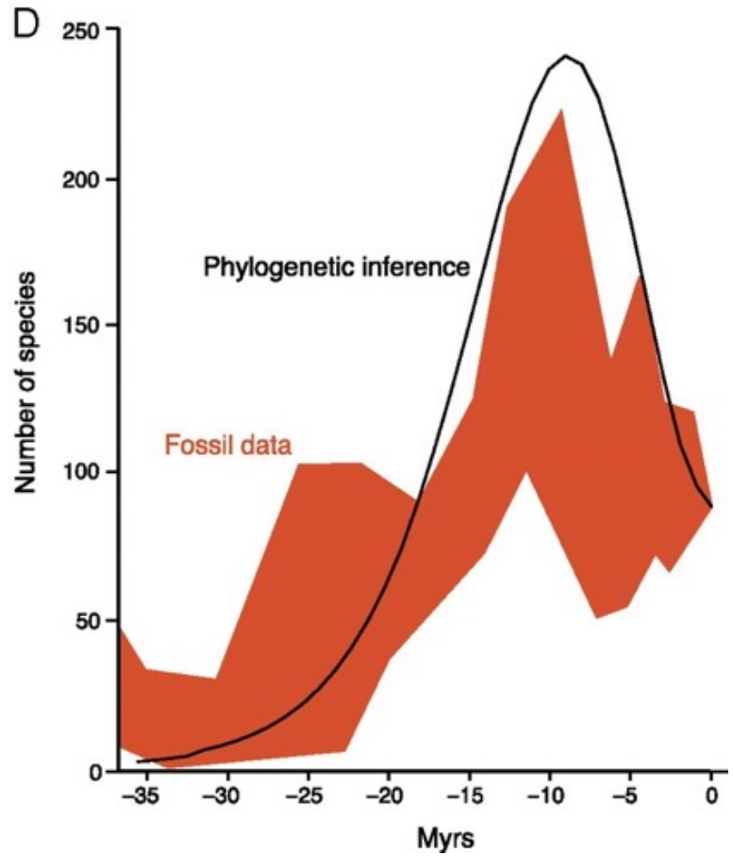


Morlon et al 2010

BAMM, RPANDA, and trait independent diversification



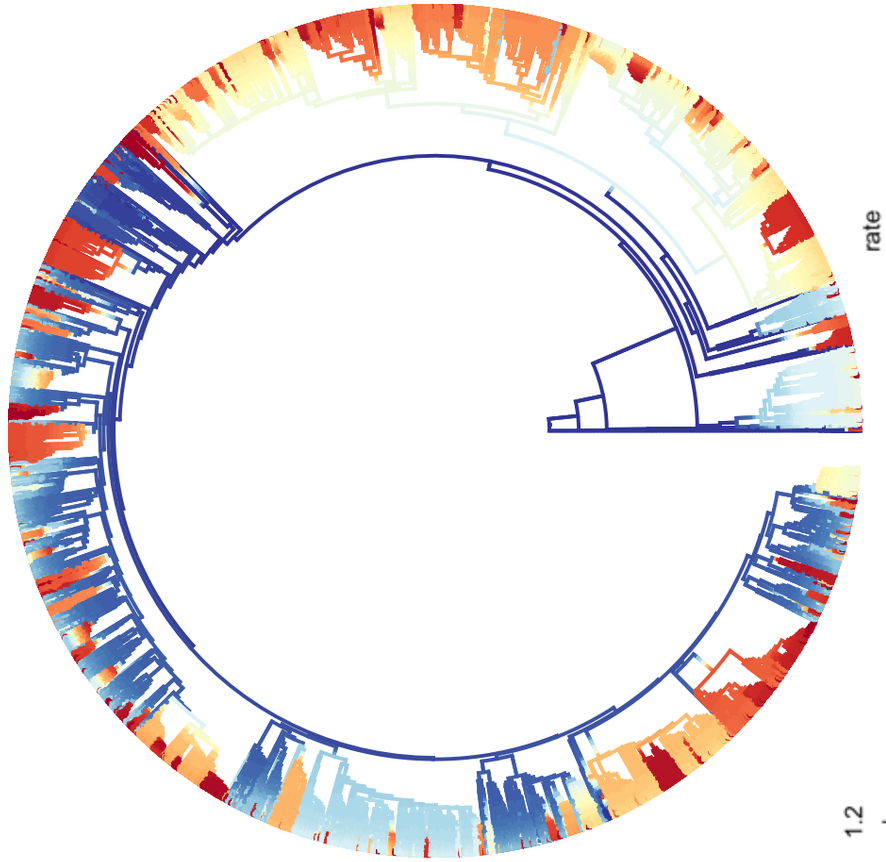
Steeman et al 2009

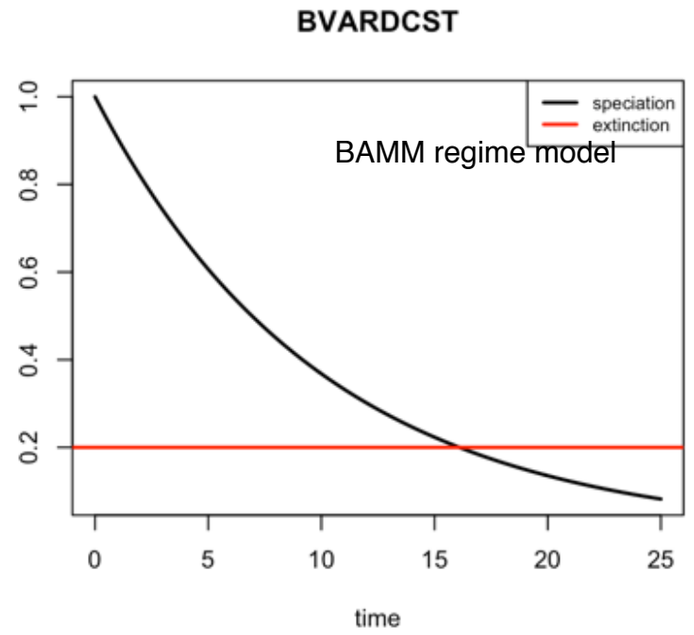


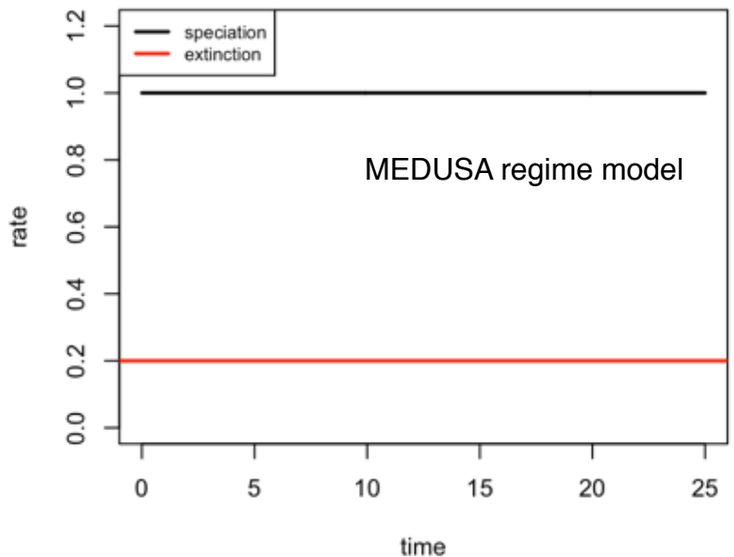
Morlon et al 2010

BAMM a further extension

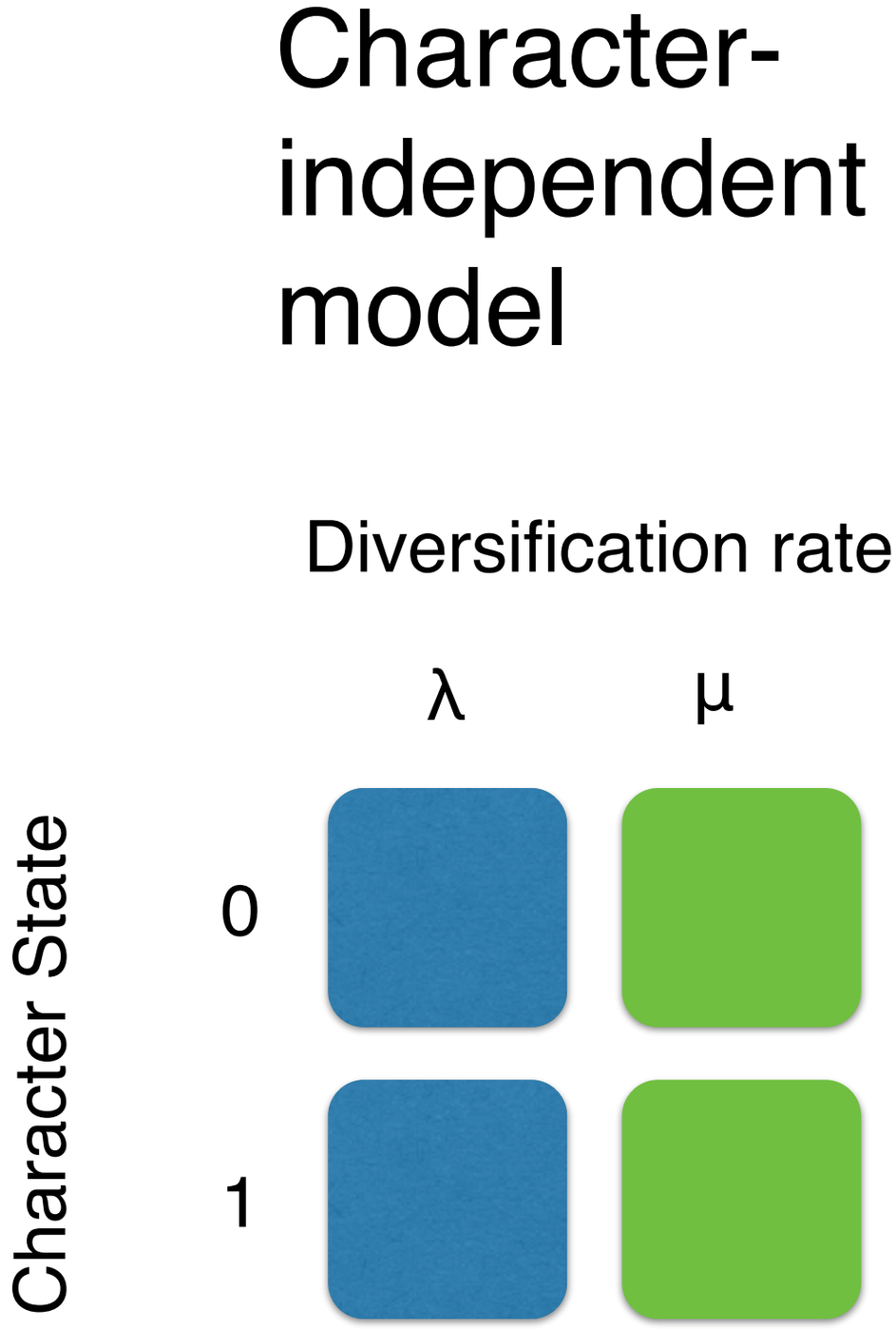
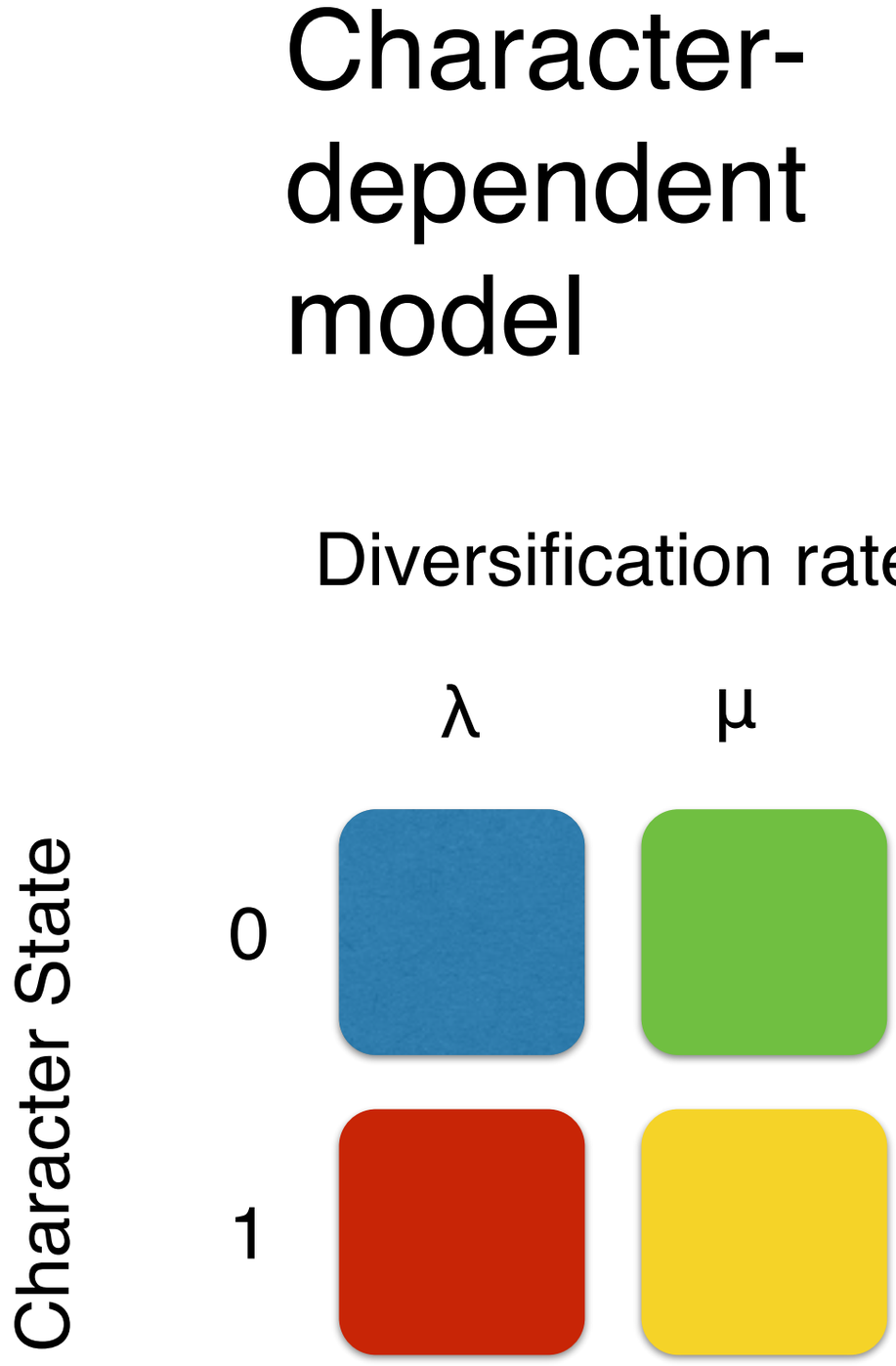
- general model: speciation variable, extinction constant
- allows tree to have as many diversification regimes as the data require (similar to MEDUSA but with a more flexible model)
- integrates over all reasonable fits of this multi regime model to any particular tree







Character-independent model

vs.

Character-dependent model