

An Introduction To MEDUSA

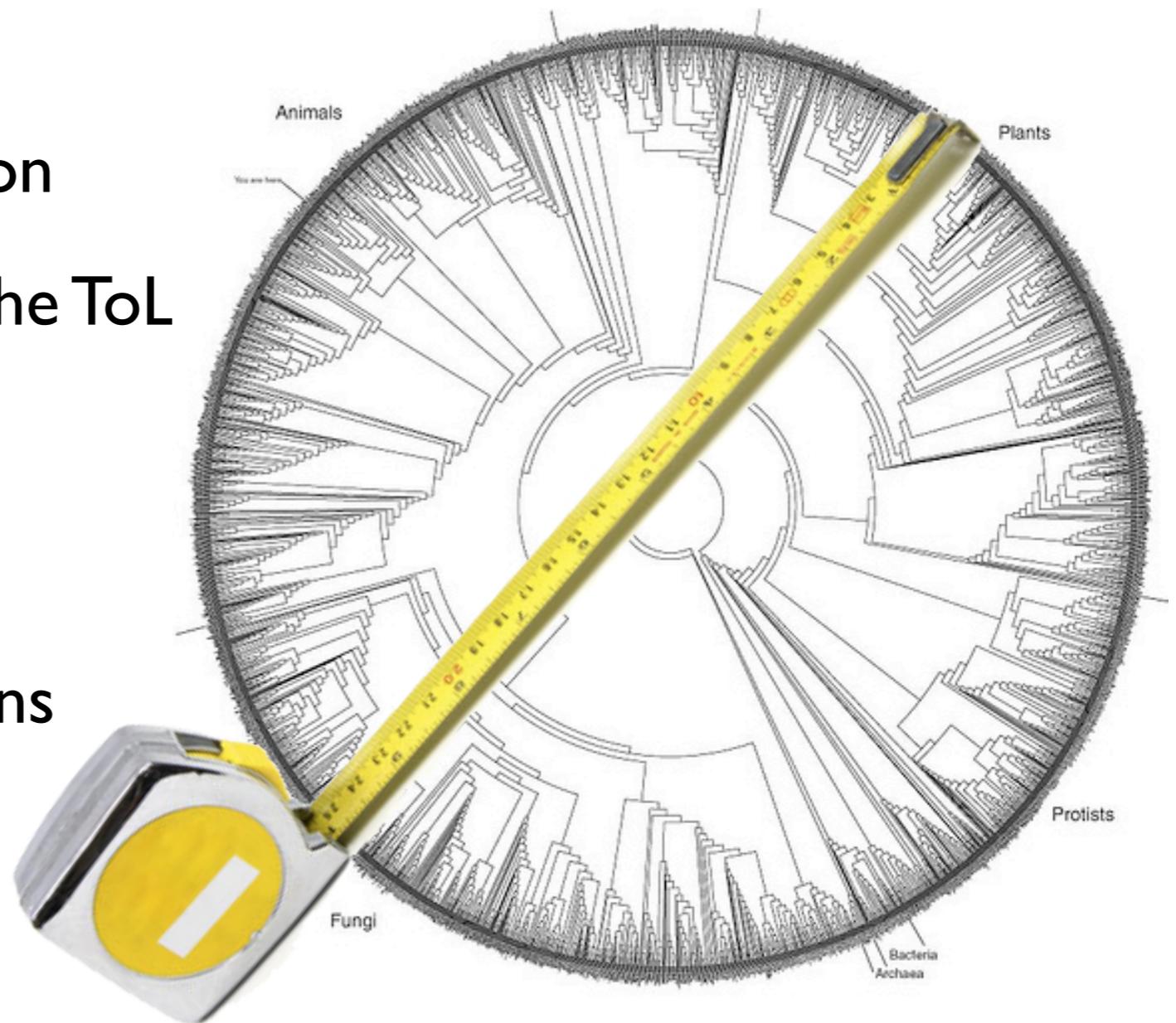


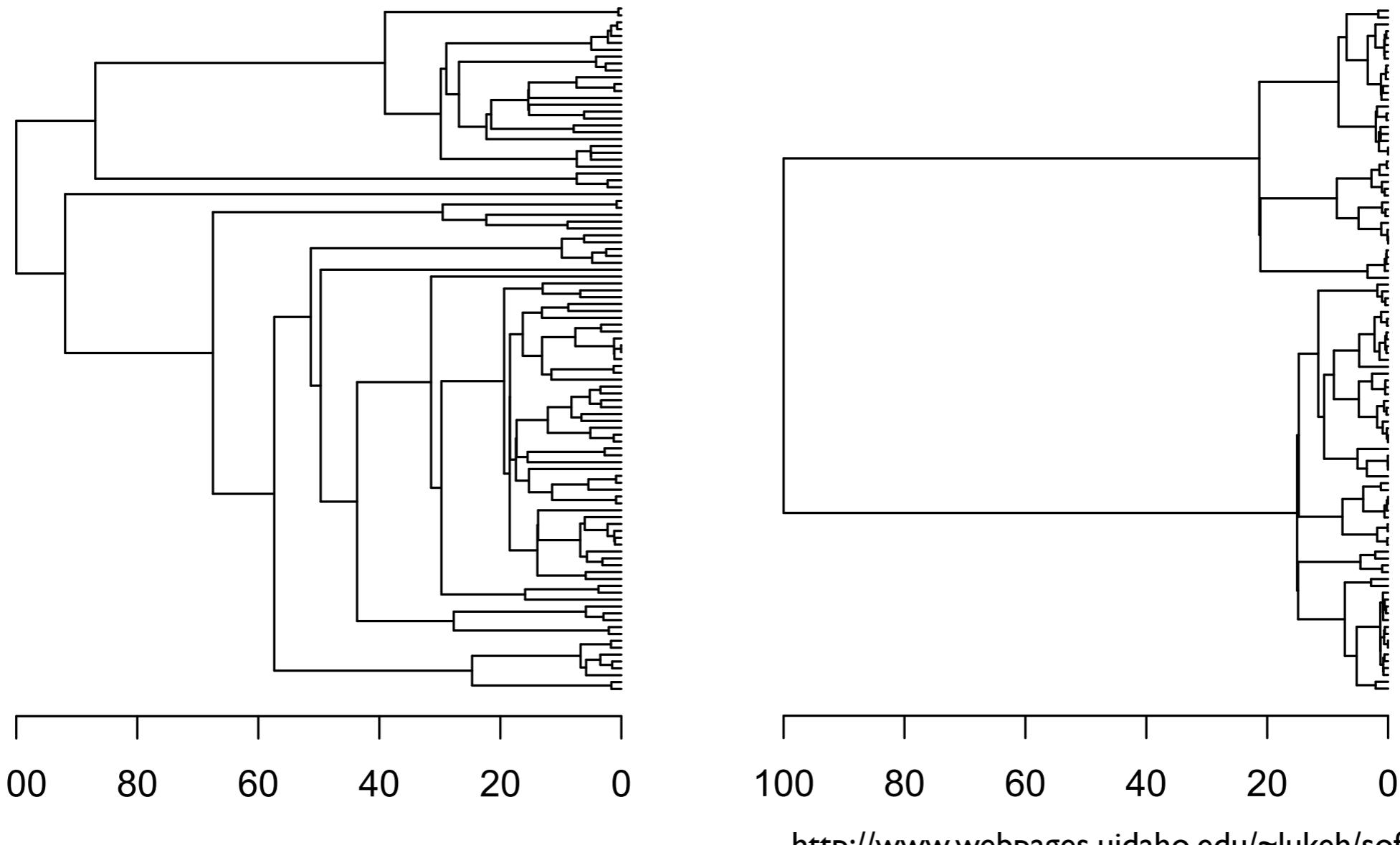
Joseph W. Brown, Richard G. FitzJohn, Graham J. Slater, Luke J. Harmon, and Michael E. Alfaro

josephwb@uidaho.edu

Talk Outline

- I. The tempo of macroevolution
- II. Clade size disparity across the ToL
- III. The problem of extinction
- IV. Cetaceans & Carnivores
- V. Conclusions/Future directions





<http://www.webpages.uidaho.edu/~lukeh/software/index.html>

An Introduction To MEDUSA - MacroInR 2012

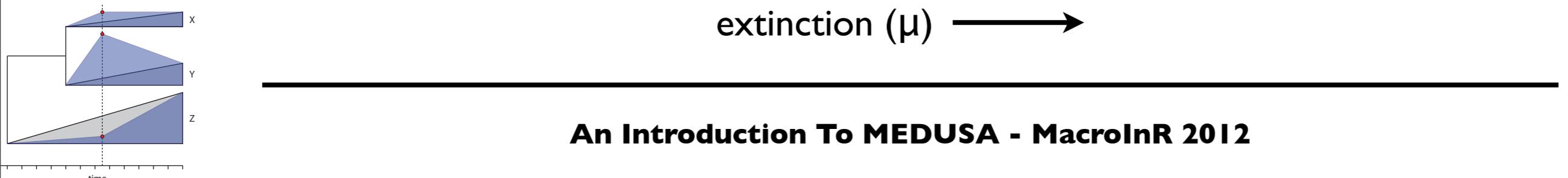
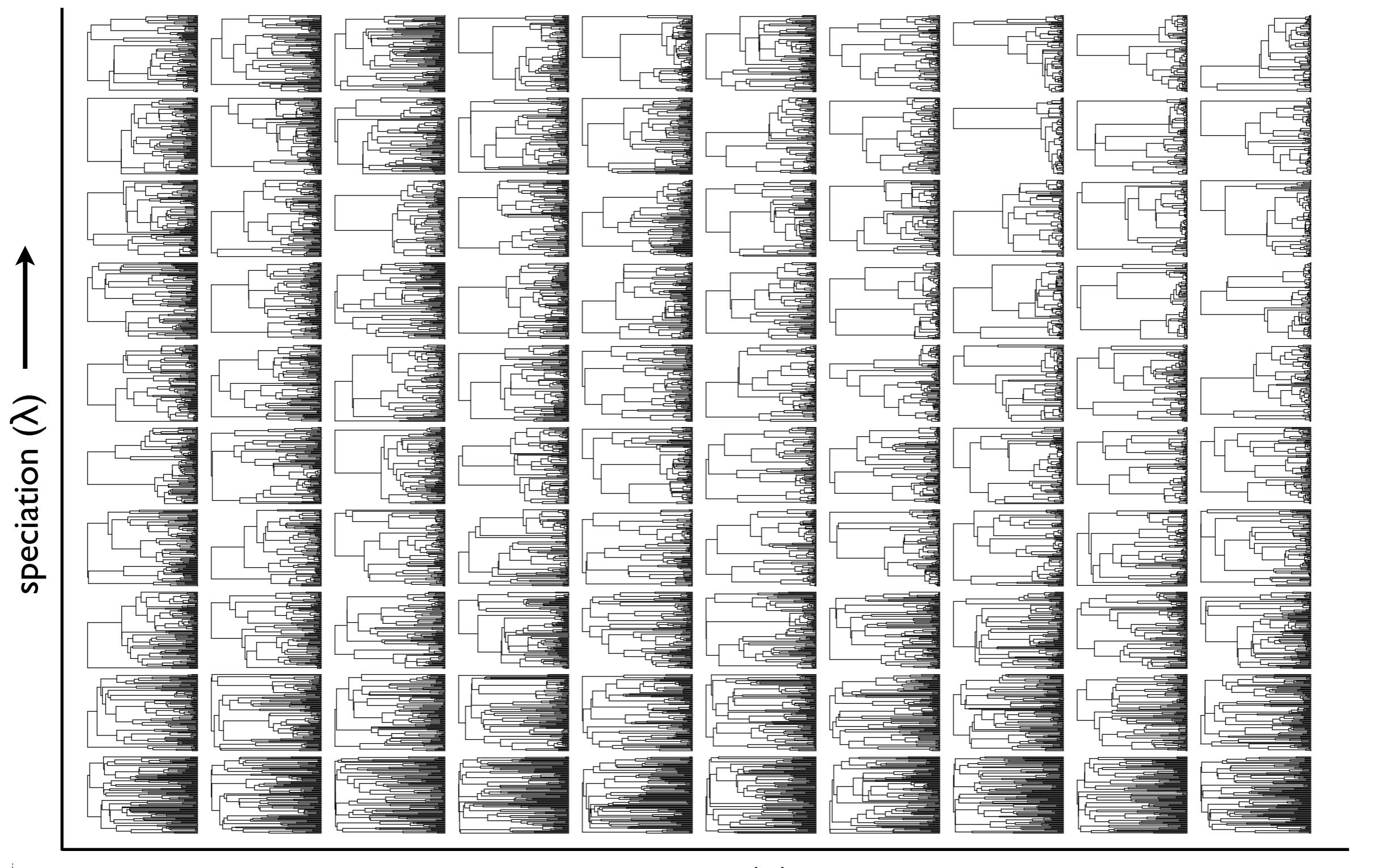
Measuring the Tree of Life



- Estimates of the *tempo* of evolution can guide us towards an understanding of the *mode* of evolution
- Allow contemplation of a number of questions
 - I. Why is the ToL shaped the way it is?
 2. Why are some lineages more diverse than others?
 3. How are rates of lineage diversification and trait evolution related?
 4. Do changes in lineage diversification rates follow mass extinctions?



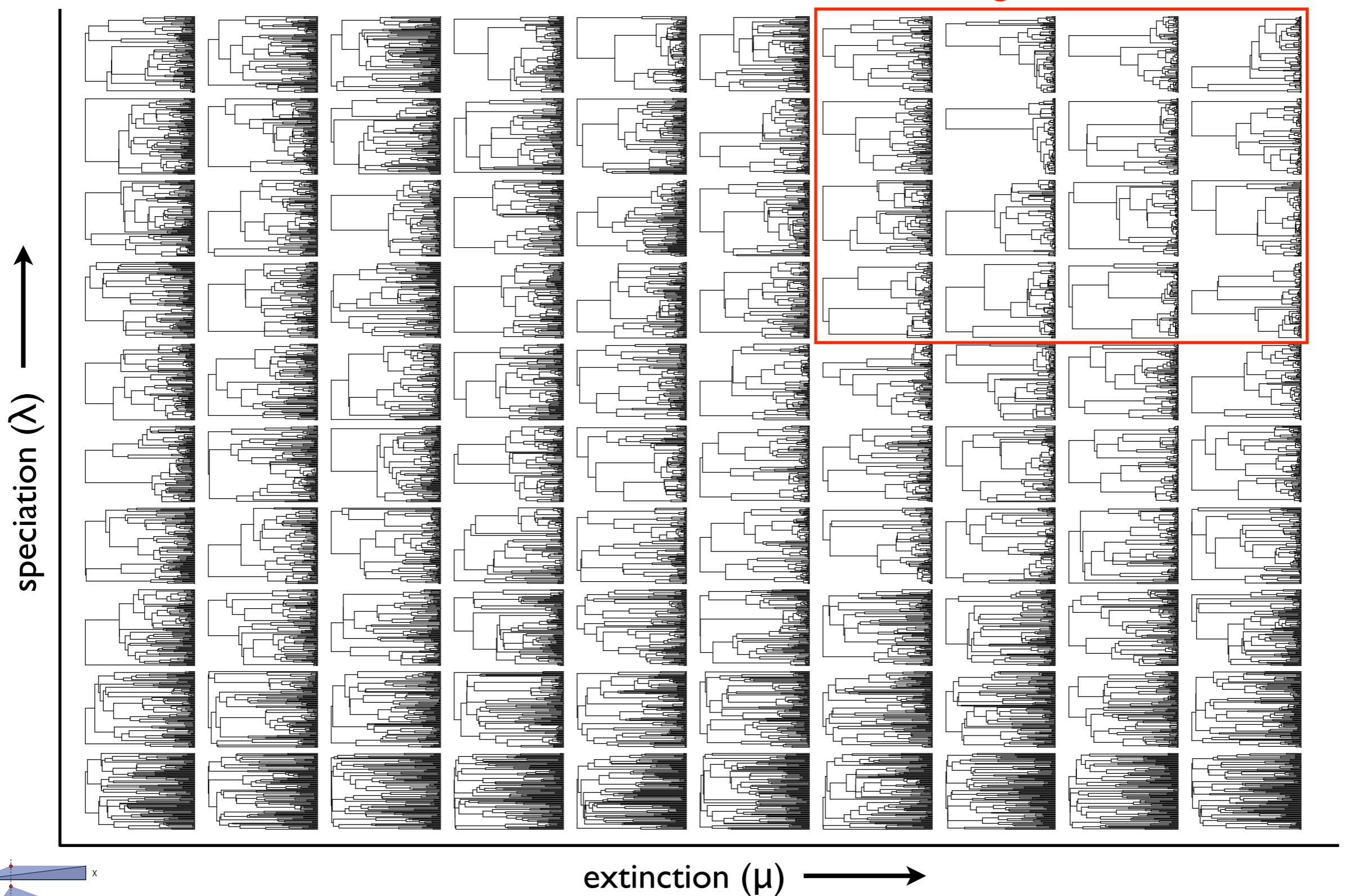
Neontology-based Methods



An Introduction To MEDUSA - MacroInR 2012

Neontology-based Methods

High turnover



An Introduction To MEDUSA - MacroInR 2012

Modelling Lineage Diversification

Birth-death Models

- Assume a time-homogeneous constant rate of speciation (λ) and extinction (μ)
 - units of lineage/lineage-unit-time
- Goal: to find the parameter values which best explain the observed edge lengths and speciation times.

The reconstructed evolutionary process

SEAN NEE, ROBERT M. MAY AND PAUL H. HARVEY

A.F.R.C. Unit of Ecology and Behaviour, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

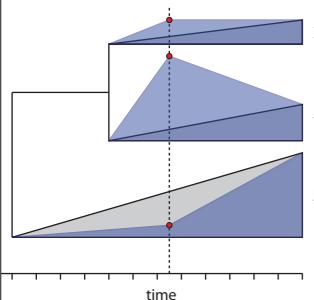
Extinction rates can be estimated from molecular phylogenies

SEAN NEE, EDWARD C. HOLMES, ROBERT M. MAY AND PAUL H. HARVEY

A.F.R.C. Unit of Ecology and Behaviour, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, U.K.

SUMMARY

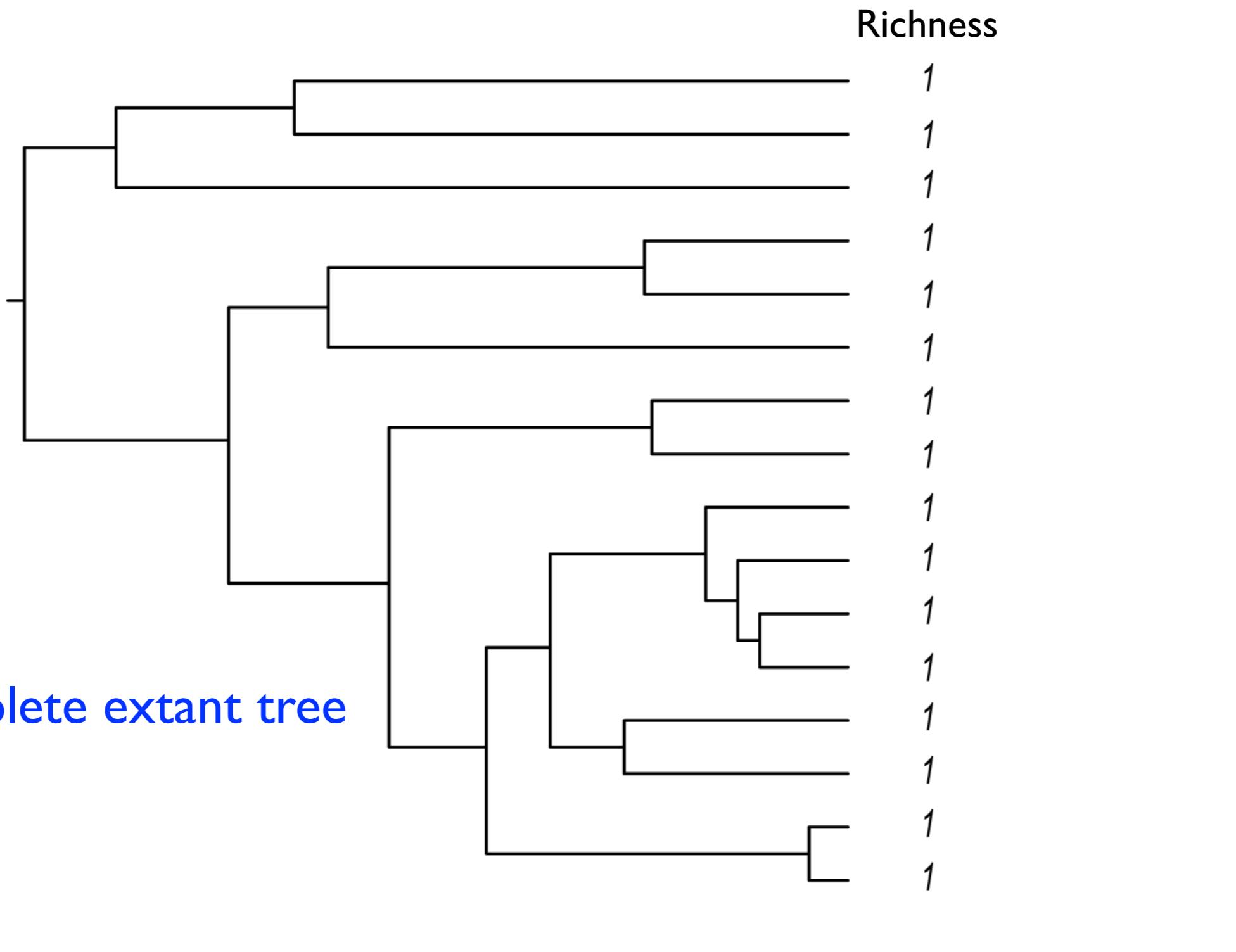
Molecular phylogenies can be used to reject null models of the way we think evolution occurred, including patterns of lineage extinction. They can also be used to provide maximum likelihood estimates of parameters associated with lineage birth and death rates. We illustrate: (i) how molecular phylogenies



Nee et al. 1994. Phil. Trans. Roy. Soc.

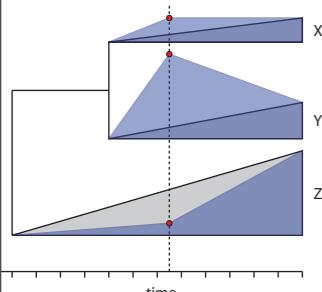
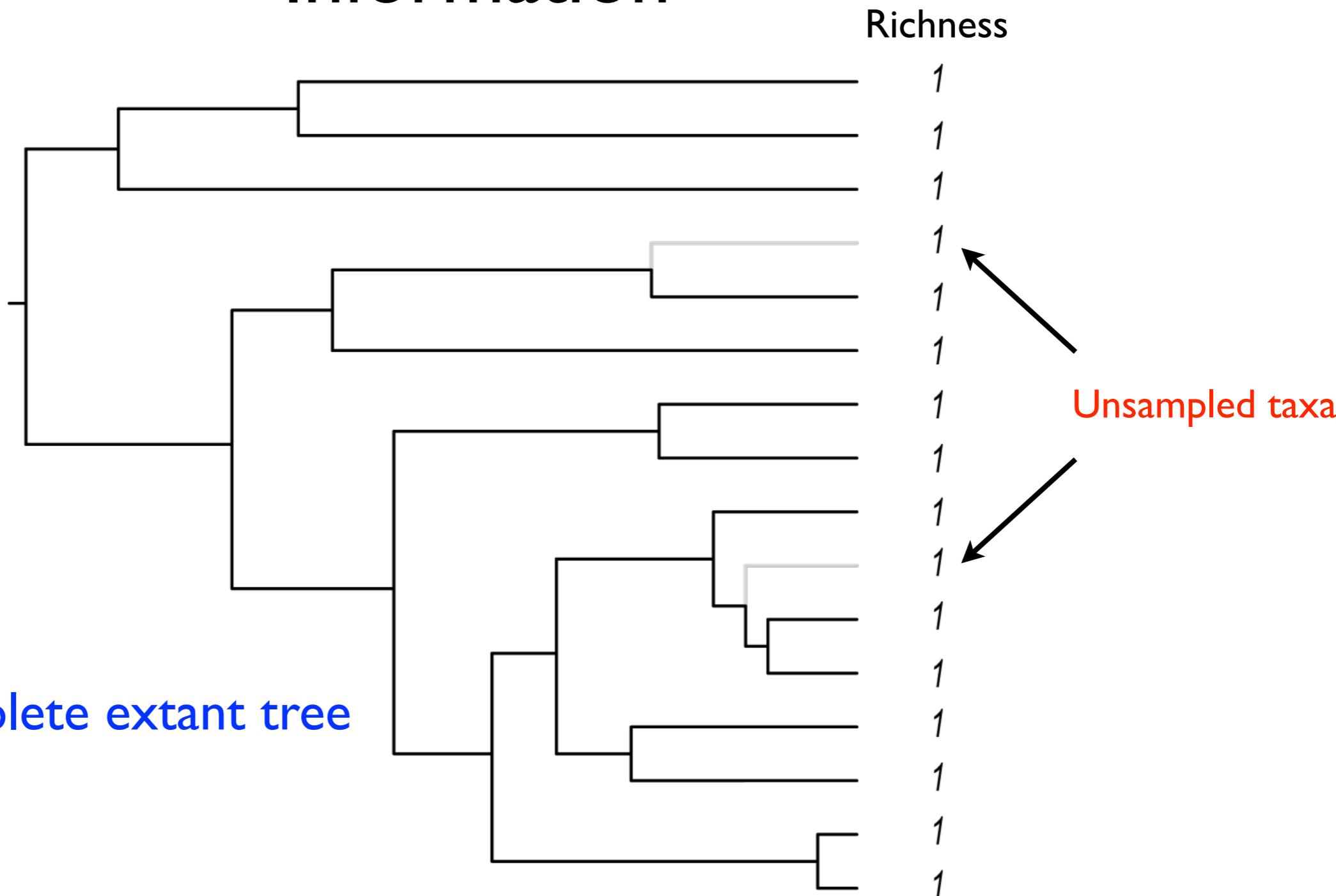
An Introduction To MEDUSA - MacroInR 2012

Fitting BD Models to Phylogenies



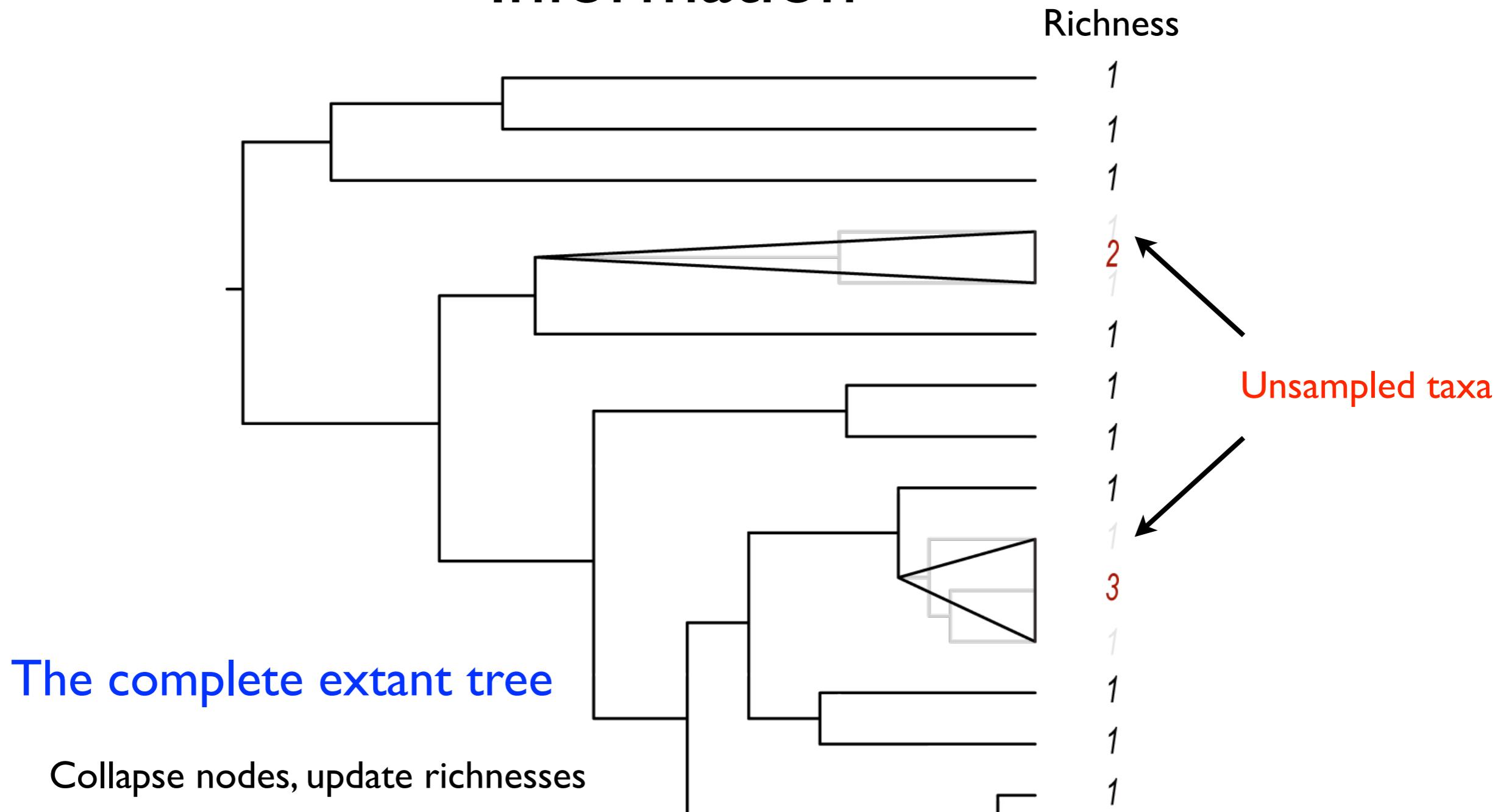
Merging Phylogenetic and Taxonomic Information

The complete extant tree

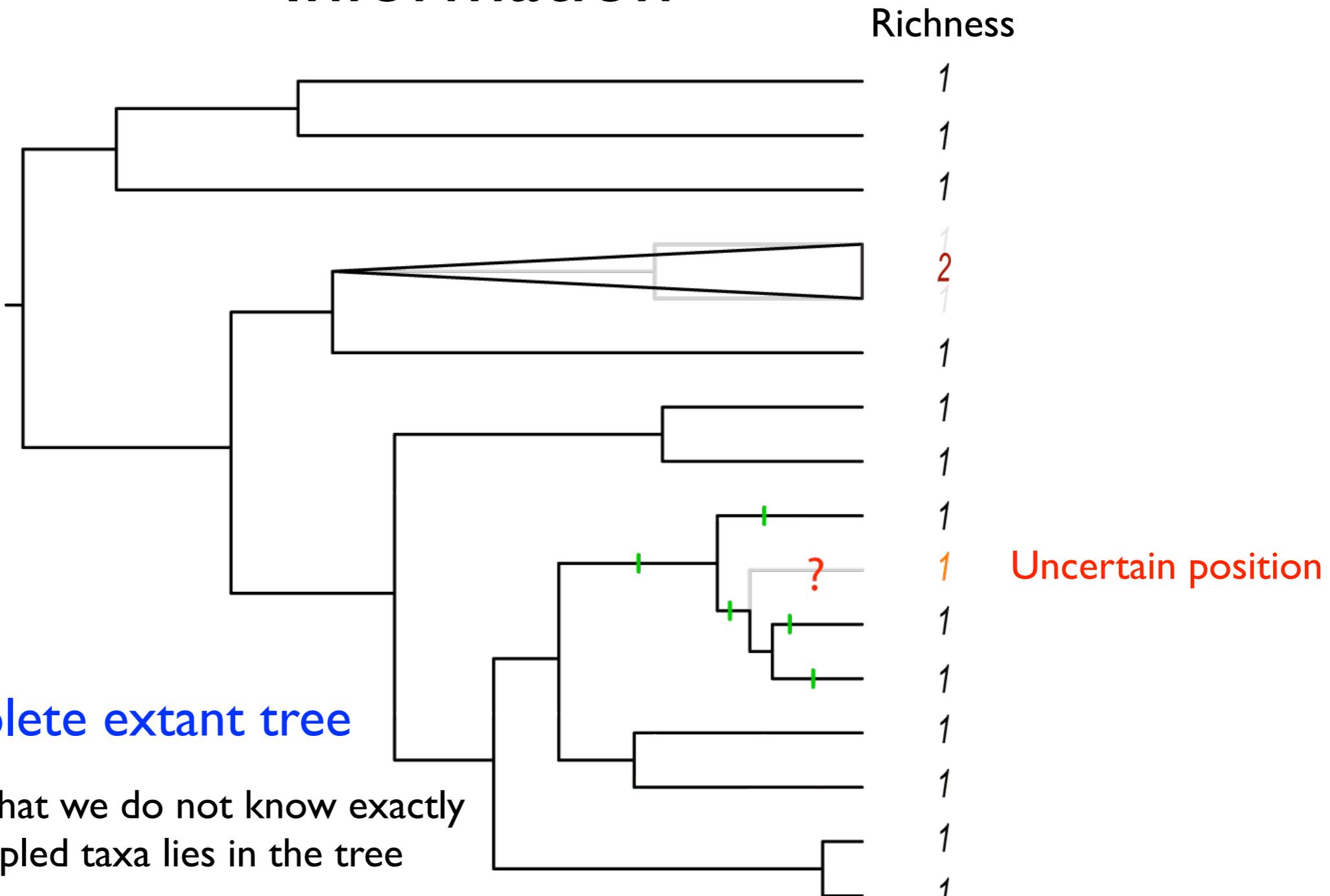


An Introduction To MEDUSA - MacroInR 2012

Merging Phylogenetic and Taxonomic Information



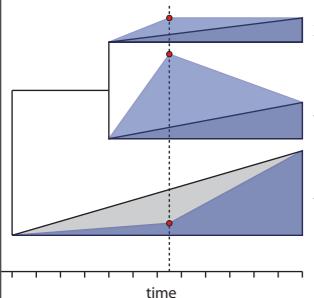
Merging Phylogenetic and Taxonomic Information



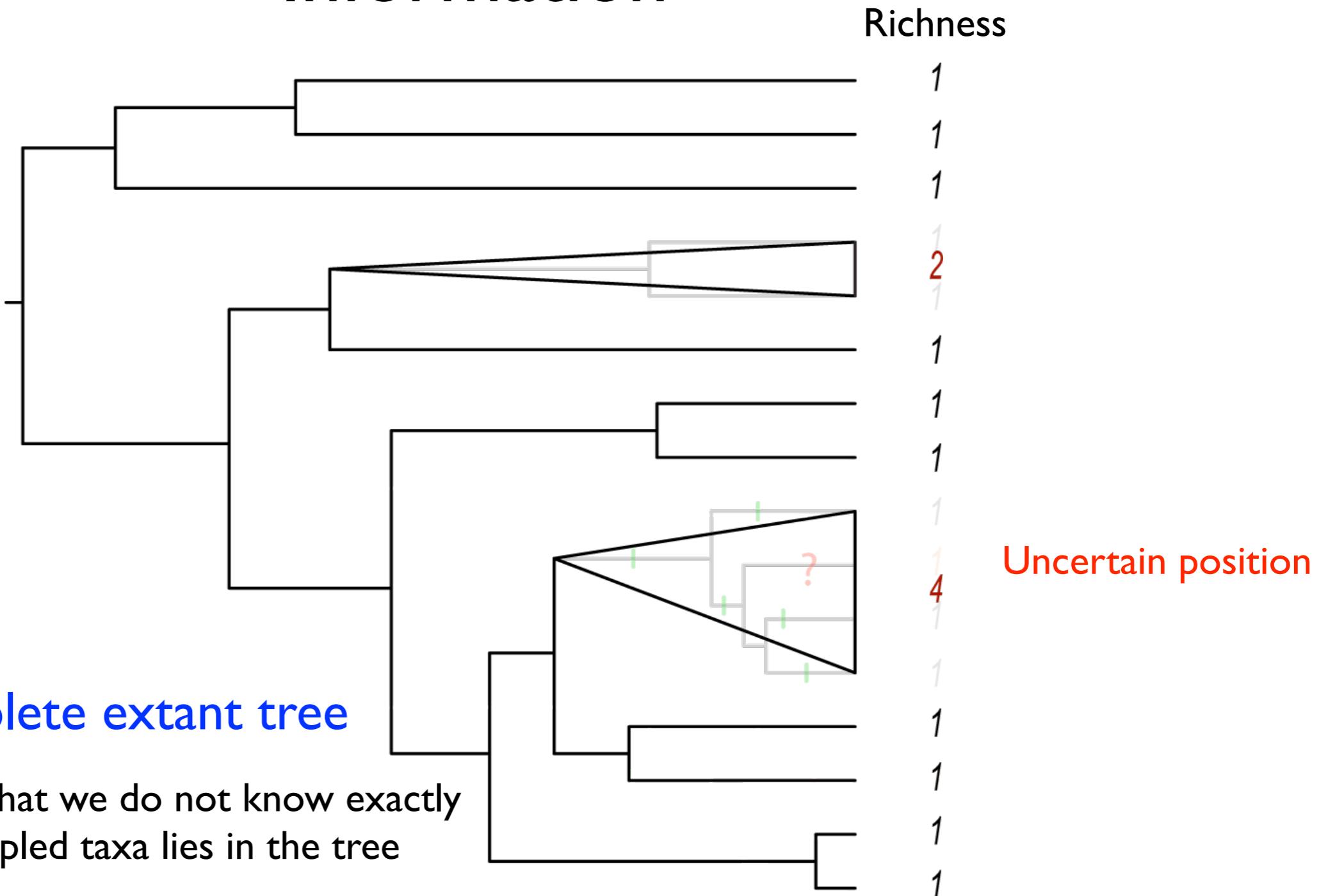
The complete extant tree

It is often the case that we do not know exactly where an unsampled taxa lies in the tree

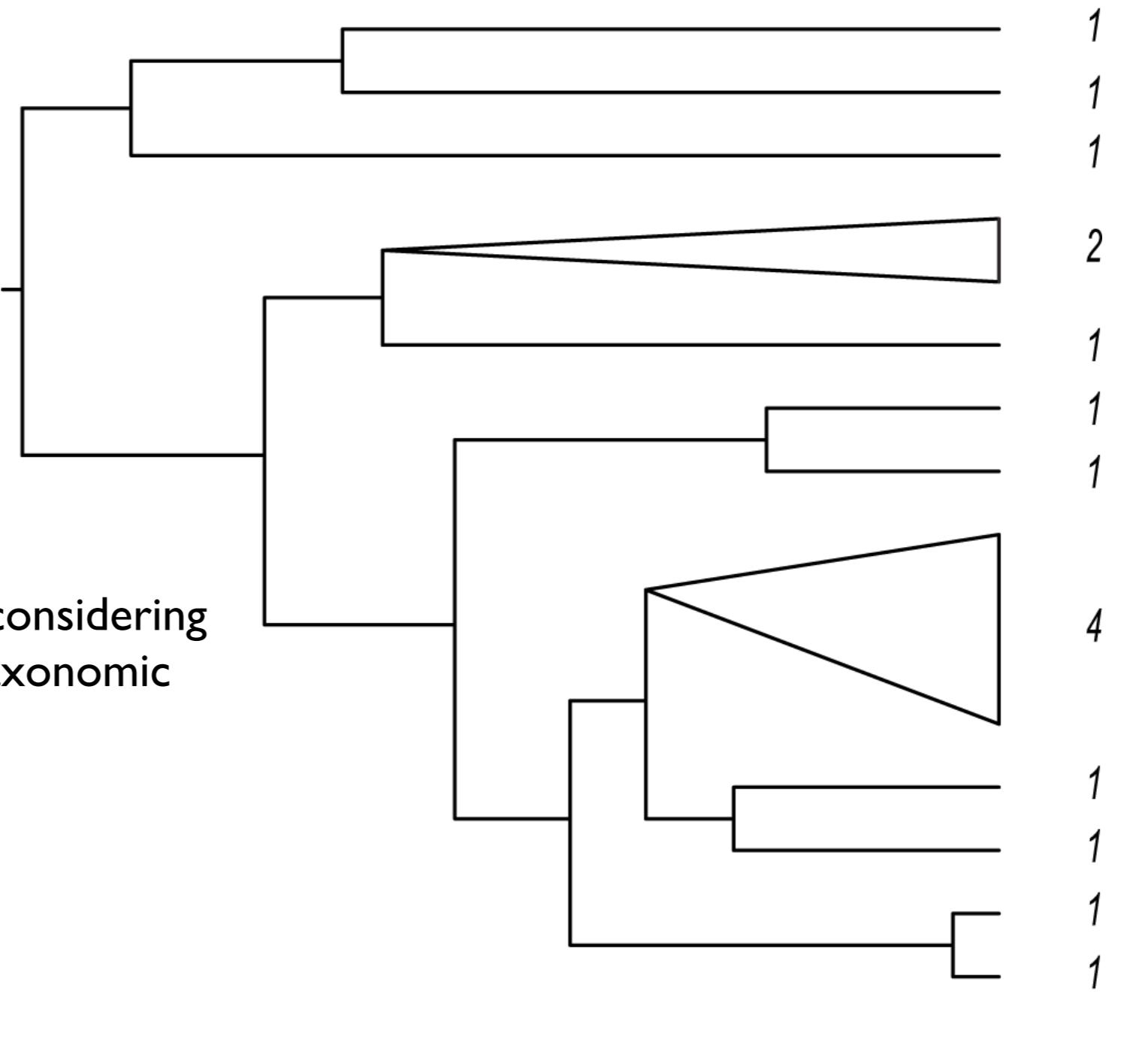
i.e. crown vs. stem



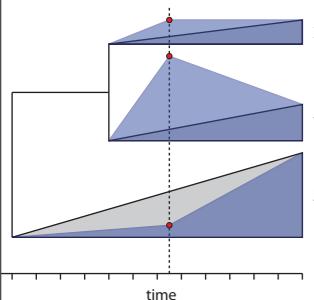
Merging Phylogenetic and Taxonomic Information



Fitting a Birth-Death Process to a Tree



Construct our likelihood model by considering both phylogenetic (internal) and taxonomic (pendant) information



Fitting a Birth-Death Process to a Tree

Internal Edges = Phylogenetic likelihood

$$\log L_P = N \cdot \log(r) - r \sum EL_i - \sum \log[1 - \varepsilon \cdot \exp(-r \cdot x_i)]$$

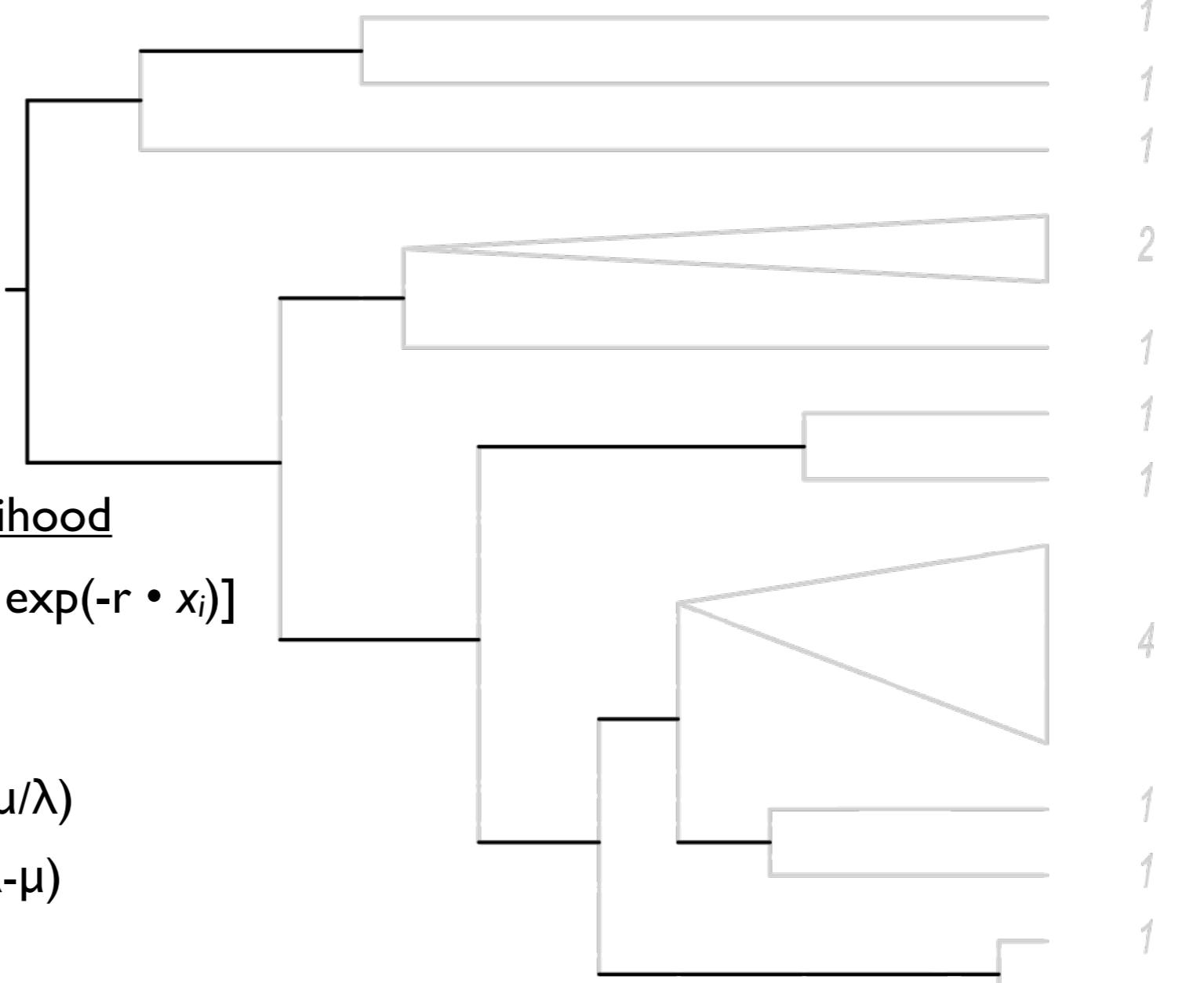
N = # internal edges

ε = extinction fraction (μ/λ)

r = net diversification ($\lambda-\mu$)

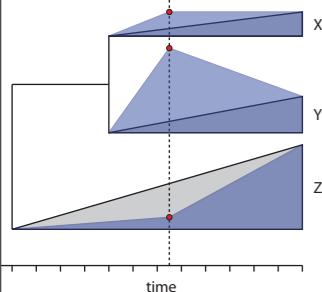
x_i = birth time of edge i

EL_i = length of edge i

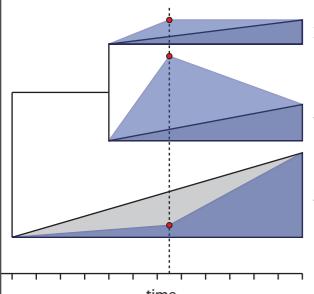
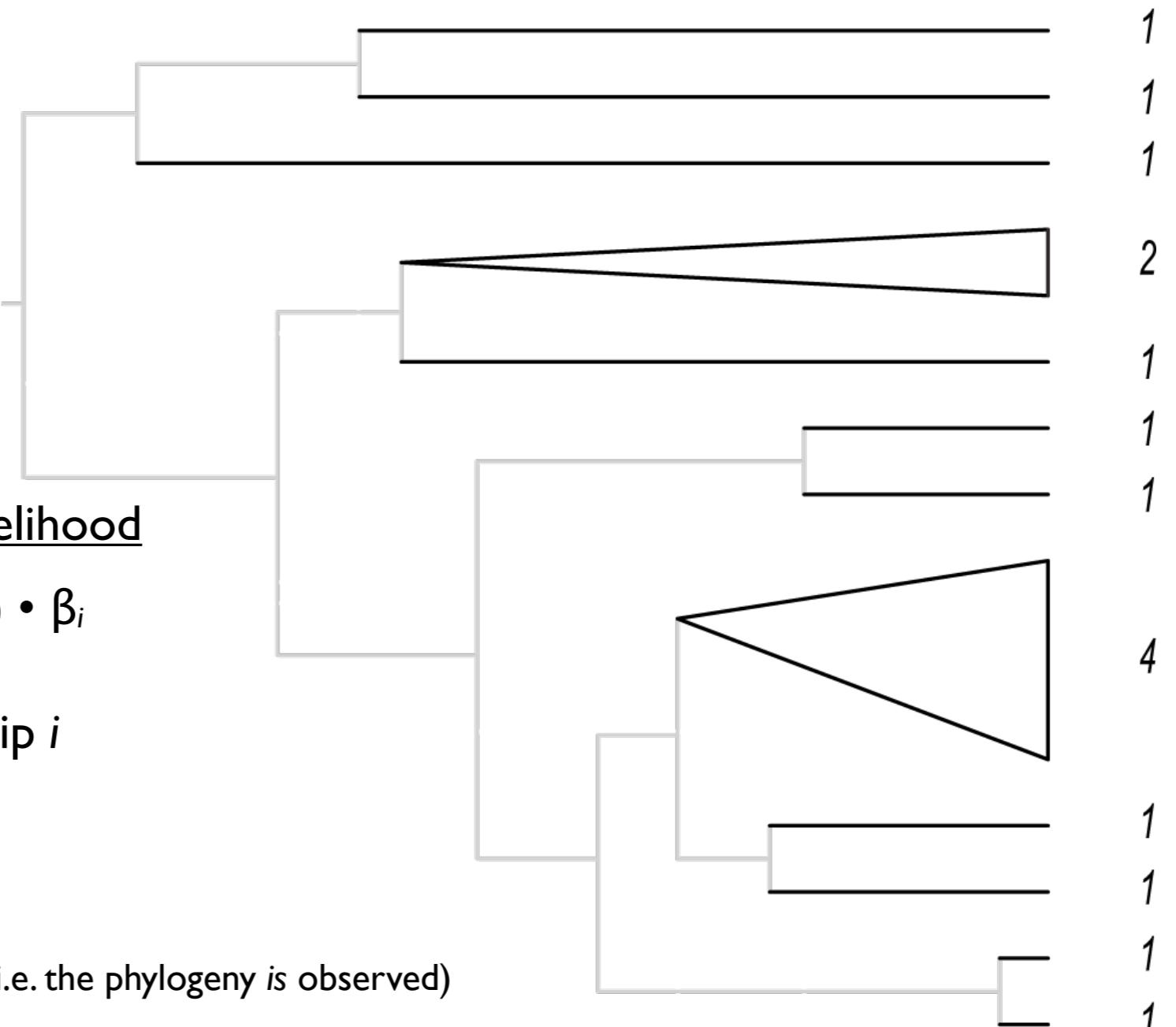


Rabosky et al. 2007. Proc. Roy. Soc.

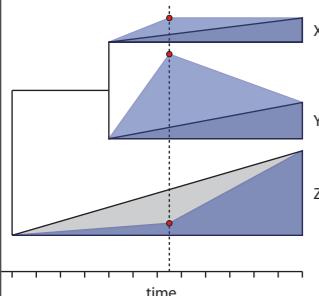
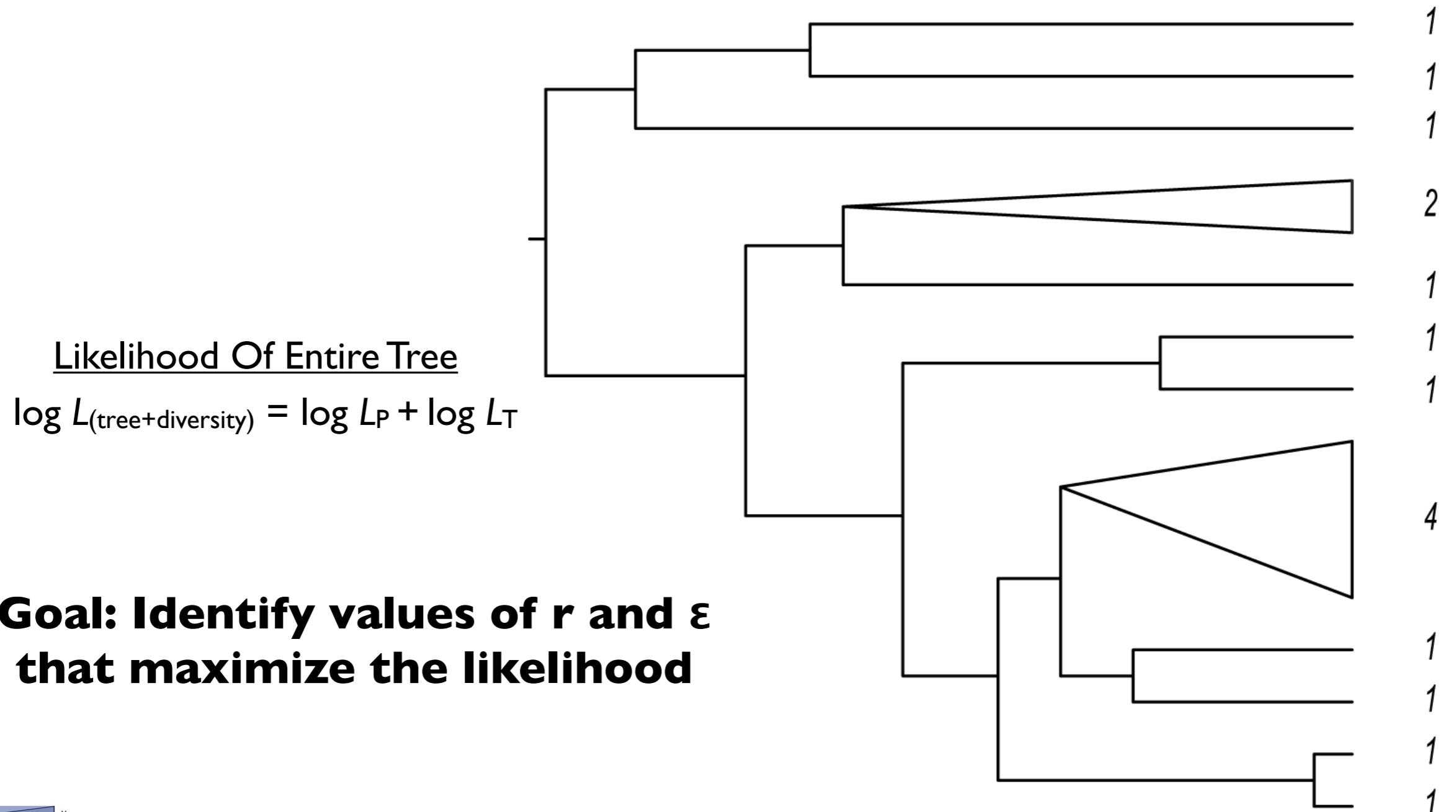
An Introduction To MEDUSA - MacroInR 2012



Fitting a Birth-Death Process to a Tree



Fitting a Birth-Death Process to a Tree

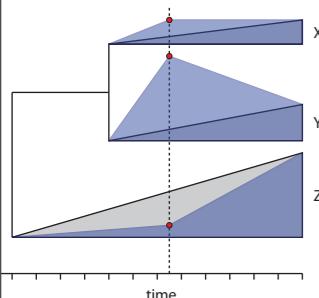


An Introduction To MEDUSA - MacroInR 2012

Clade Disparity Across the ToL

Problem

- The ToL is decidedly unbalanced
 - vertebrates, arthropods, plants, etc.
- Birth-death models predict trees more balanced than observed
- Suggests a single diversification model may be inadequate
 - there is a general problem of any model-based inference: as data grow larger, it becomes increasingly less likely that the data conform to (or are best explained by) a single generating model



Alfaro et al. 2009 PNAS

An Introduction To MEDUSA - MacroInR 2012

Stepwise AIC Approach (MEDUSA)

Algorithm

For (x in 1:(num.nodes)):

 1. Split move: Fit individual
 (piecewise) BD models to:

 1. The clade defined by x , plus
 the edge leading to x

 2. The remaining edges

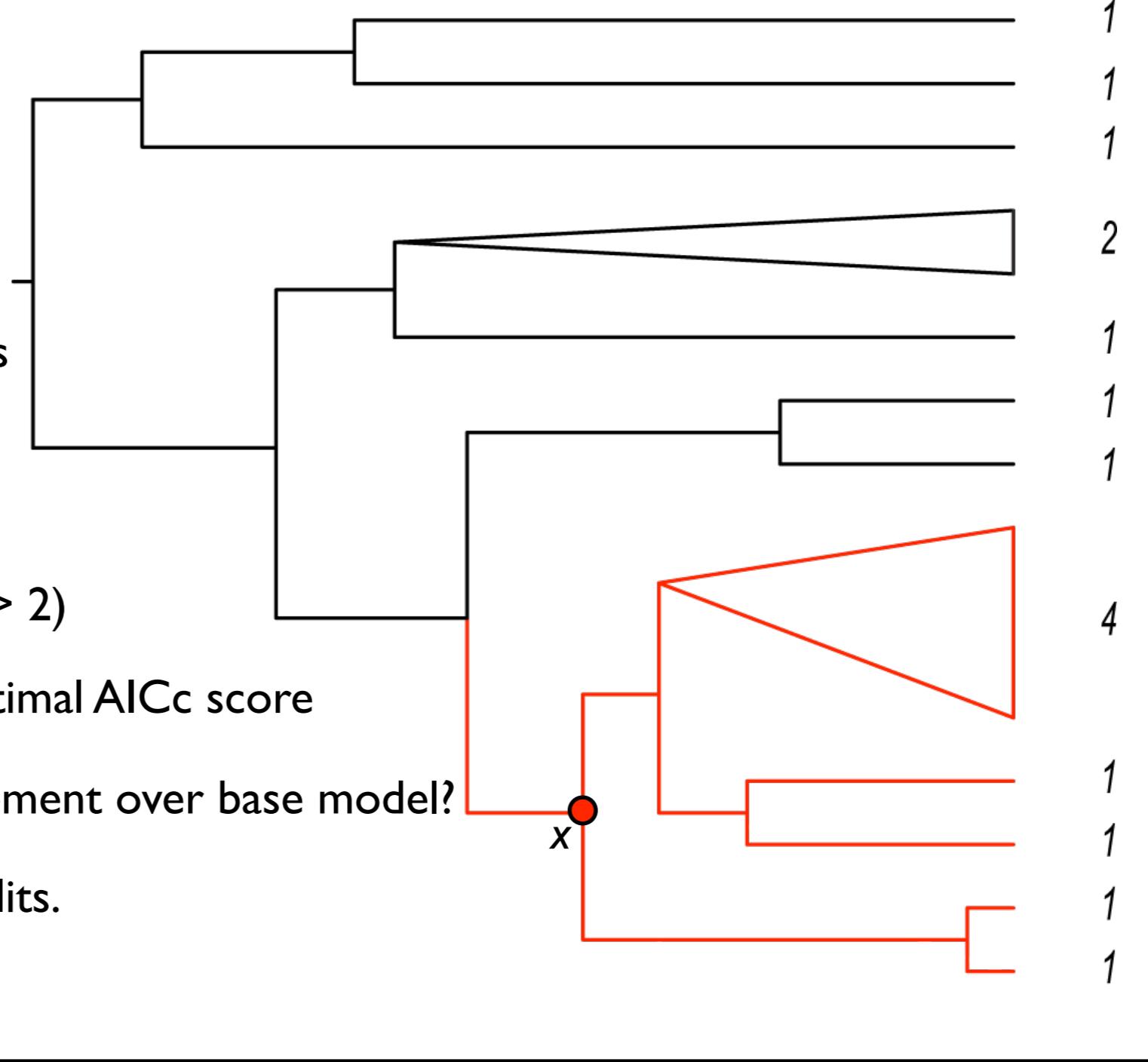
 2. Merge move (if num(models) > 2)

 3. Find piecewise model with optimal AICc score

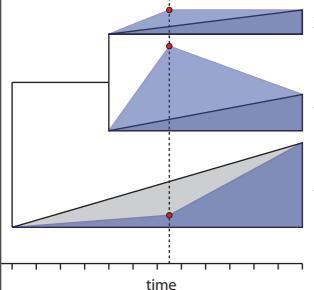
 • Is this a significant improvement over base model?

 - Yes: Test subsequent splits.

 - No: You're done.

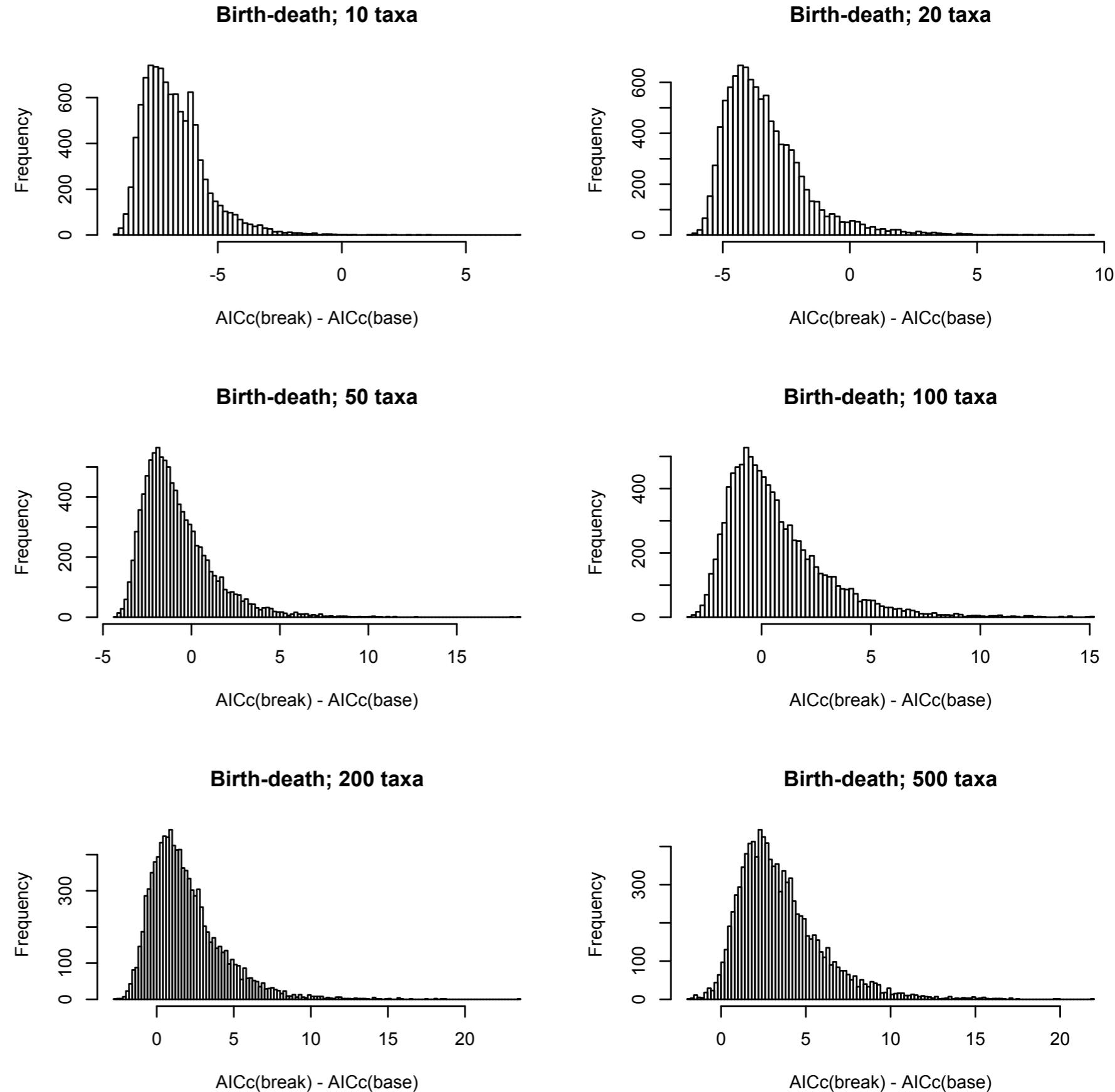


An Introduction To MEDUSA - MacroInR 2012

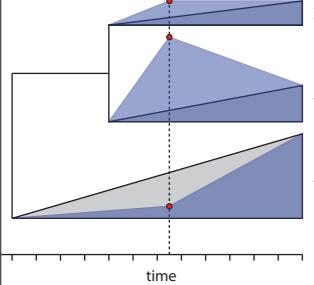


Data Dredging

- A drawback of fitting so many models is that spurious ‘significant’ results can arise

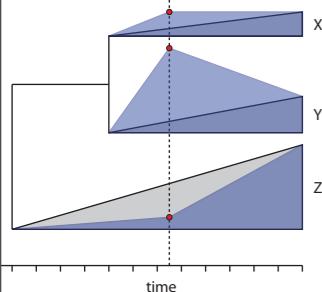
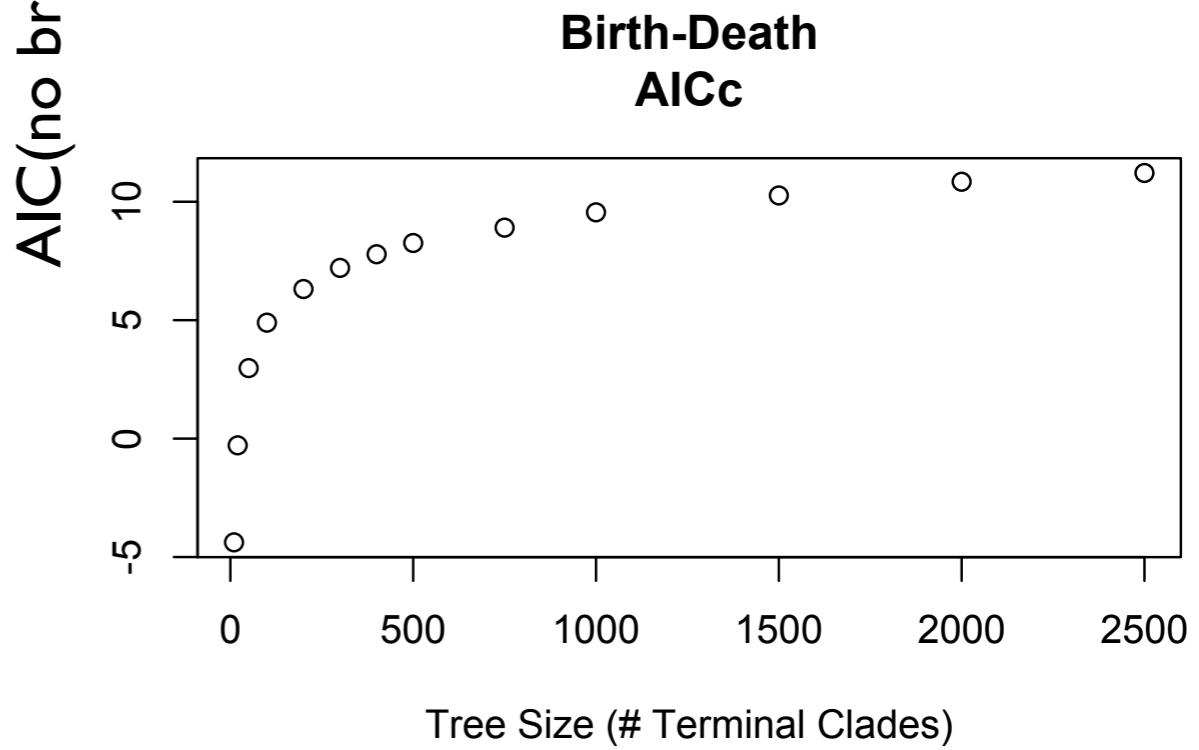
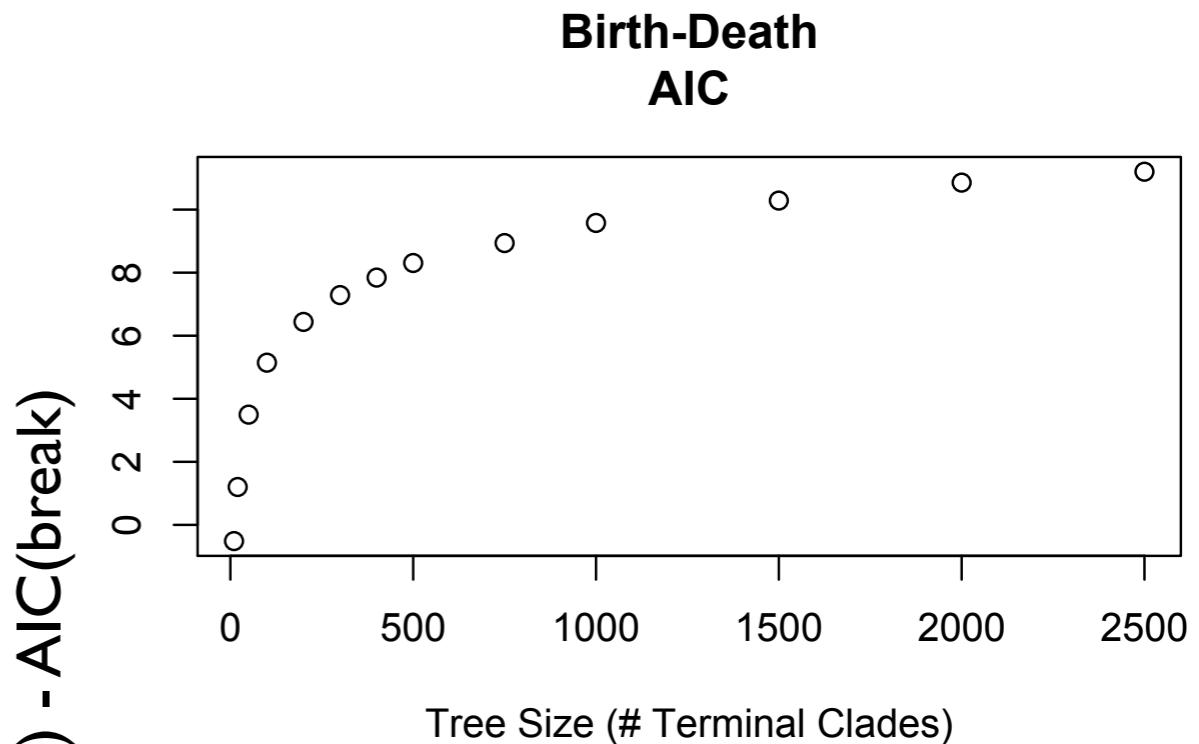


An Introduction To MEDUSA - MacroInR 2012



Data Dredging

- However, the relationship between tree size and Type I error is straightforward
- We can use a threshold value of AICc to determine significant improvements to the model
- **NOTE:** this renders AIC-weights as meaningless (being an artefact of the size of the tree).



MEDUSA

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

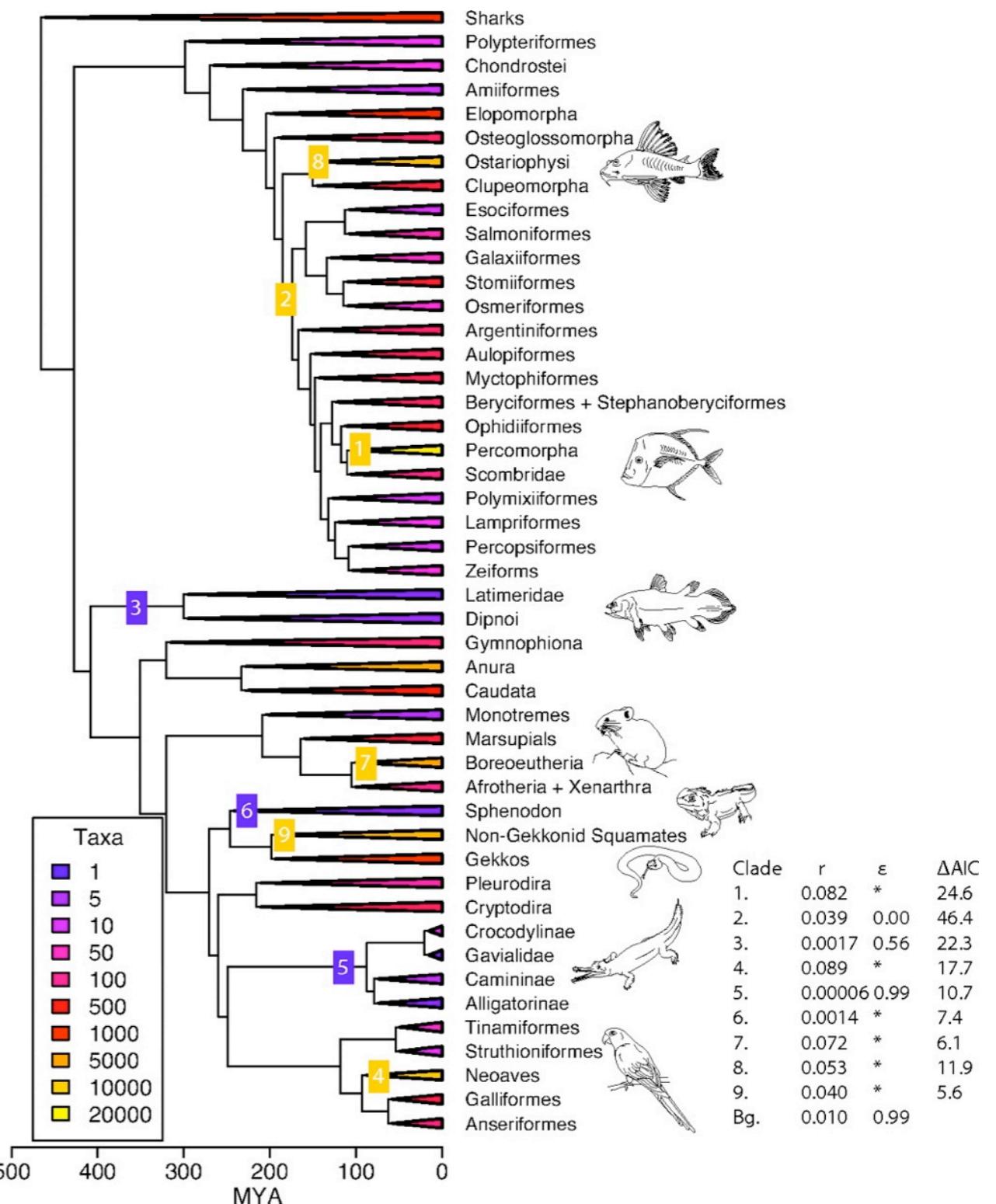
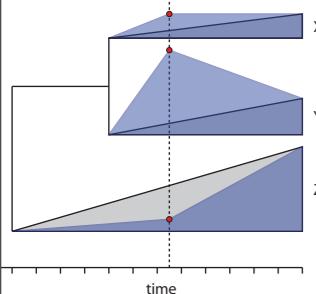
- Where have rates changed?
- By what magnitude?
- Why do clades differ in size?

r = net diversification rate

$$= \lambda - \mu$$

ϵ = extinction fraction

$$= \mu / \lambda$$

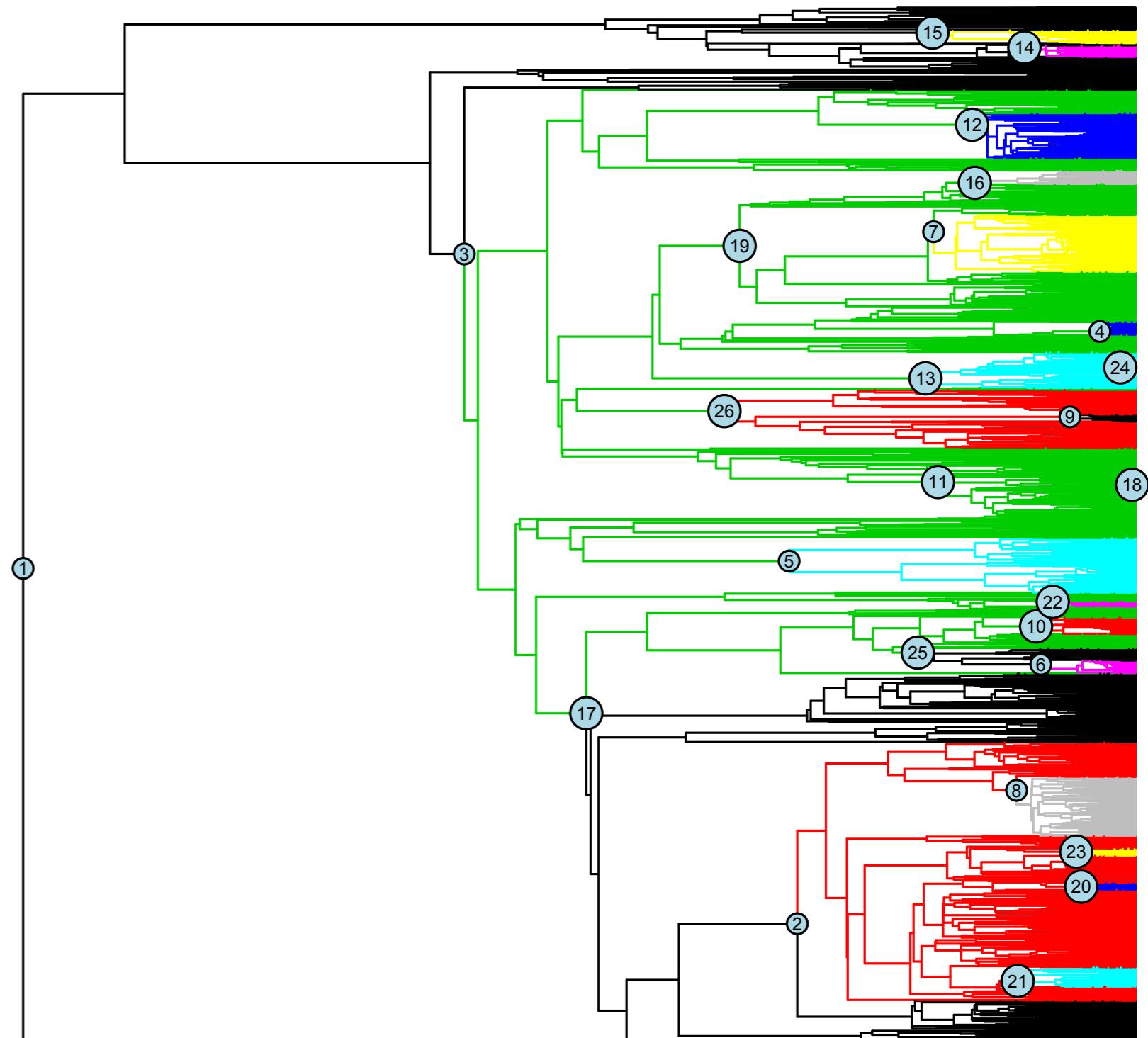


Alfaro et al. 2009 PNAS

An Introduction To MEDUSA - MacroInR 2012

Mammals

- Best described by a 26-piece diversification model



Fritz et al. 2009. Ecol. Lett.

150

100

50

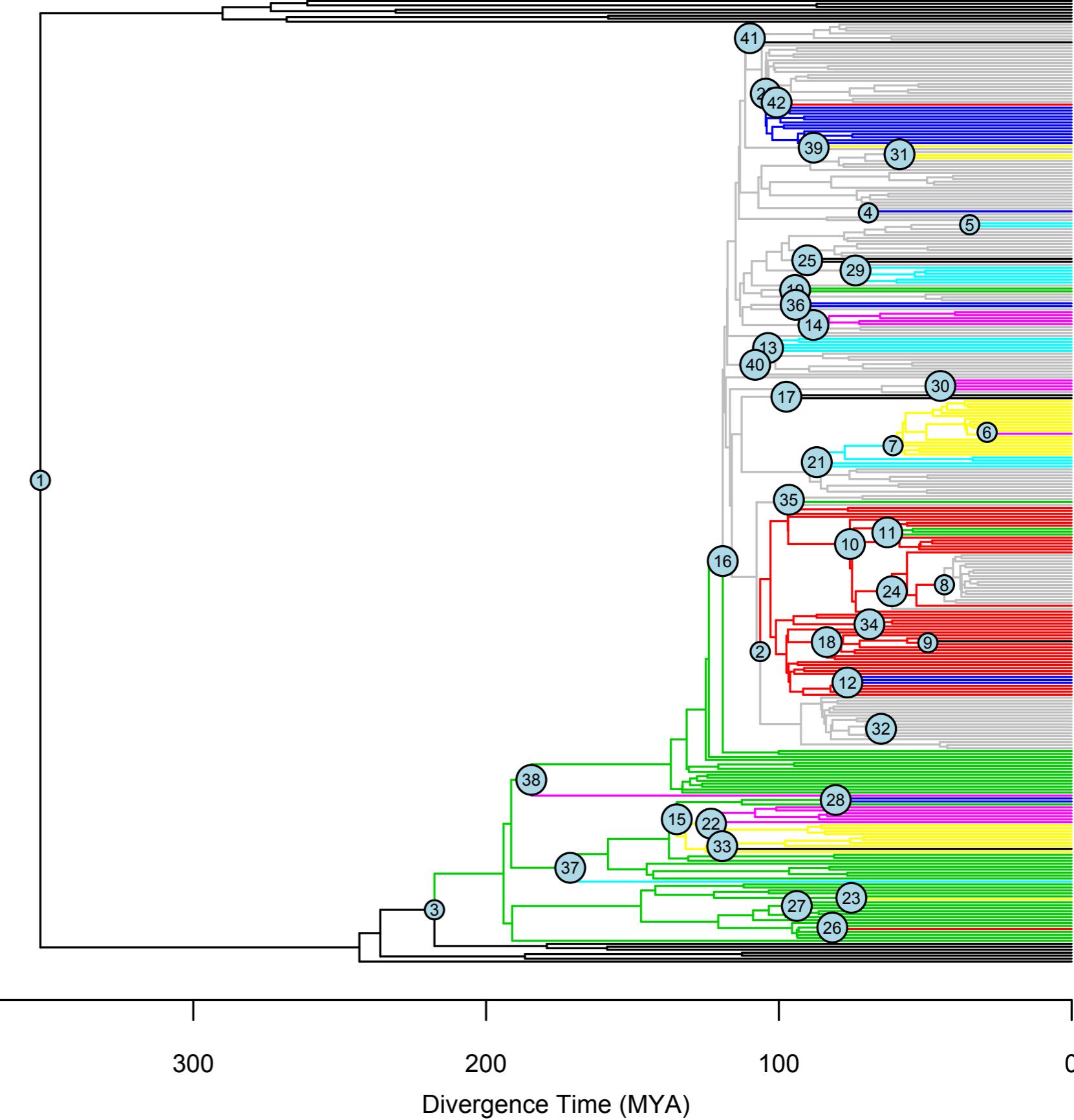
0

Divergence Time (MYA)

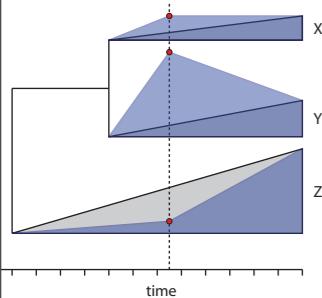
An Introduction To MEDUSA - MacroInR 2012

Angiosperms

- Best described by a 42-piece diversification model

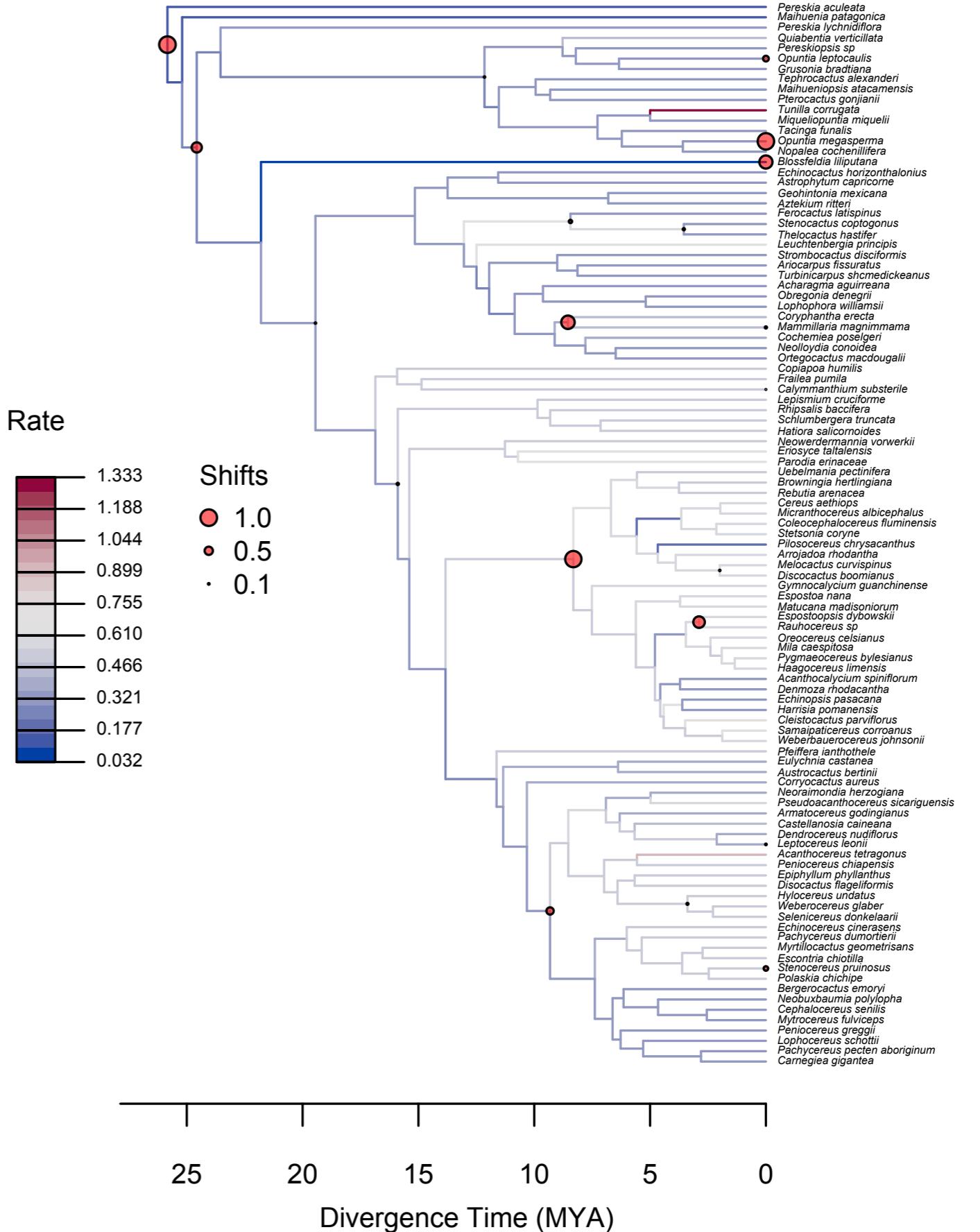


An Introduction To MEDUSA - MacroInR 2012

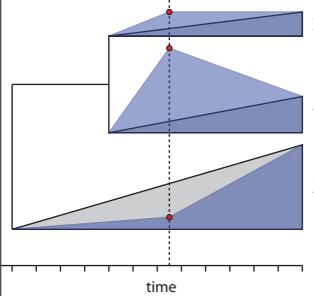


Multiple Trees

- A distribution of trees can be passed to MEDUSA
- Each tree is individually fit
- The results are plotted on some consensus or optimal tree



An Introduction To MEDUSA - MacroInR 2012



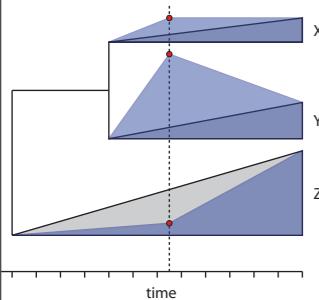
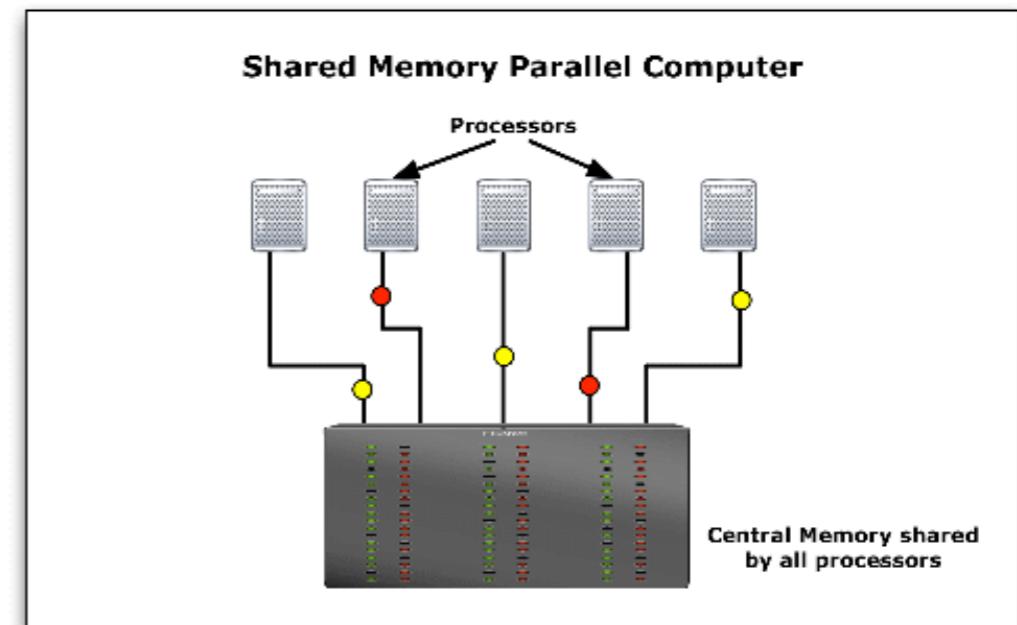
Hypothesis Testing With MEDUSA

- Sometimes we may not be interested in dredging for the best model, but instead some specific hypothesis
 - this may involve fixing specific parameter values (say, with extinction results from the fossil record)
 - comparison of a focal clade against the larger tree
 - comparisons across sister clades



Large Problems

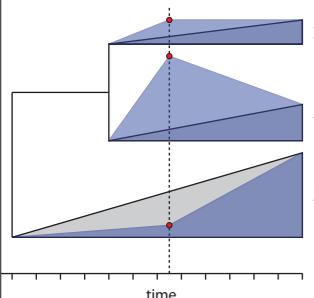
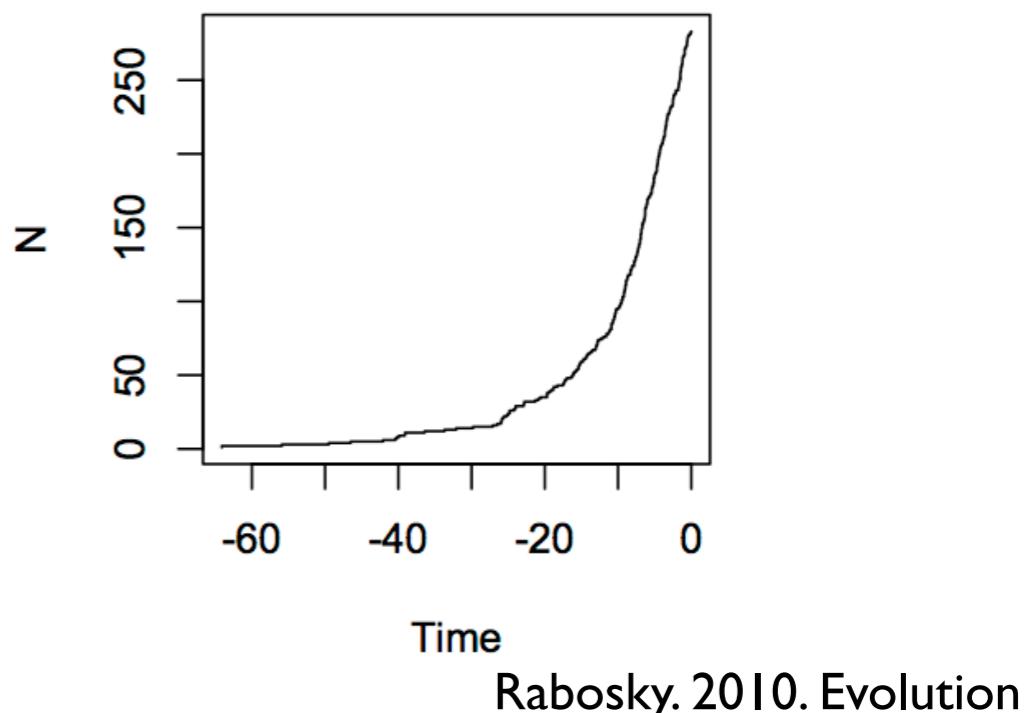
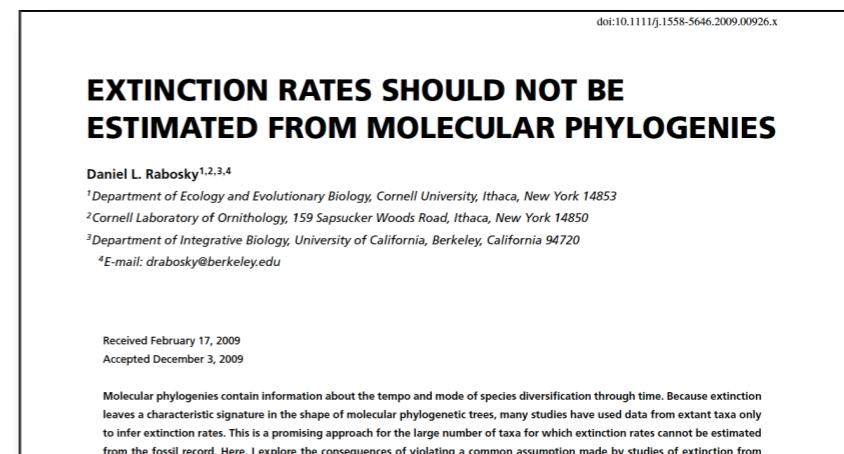
- For large and/or multiple trees, it is efficient to utilize multiple computing cores
- MEDUSA can make use of multiple cores, but:
 1. only non-Windows
 2. only non-GUI



The Problem of Extinction

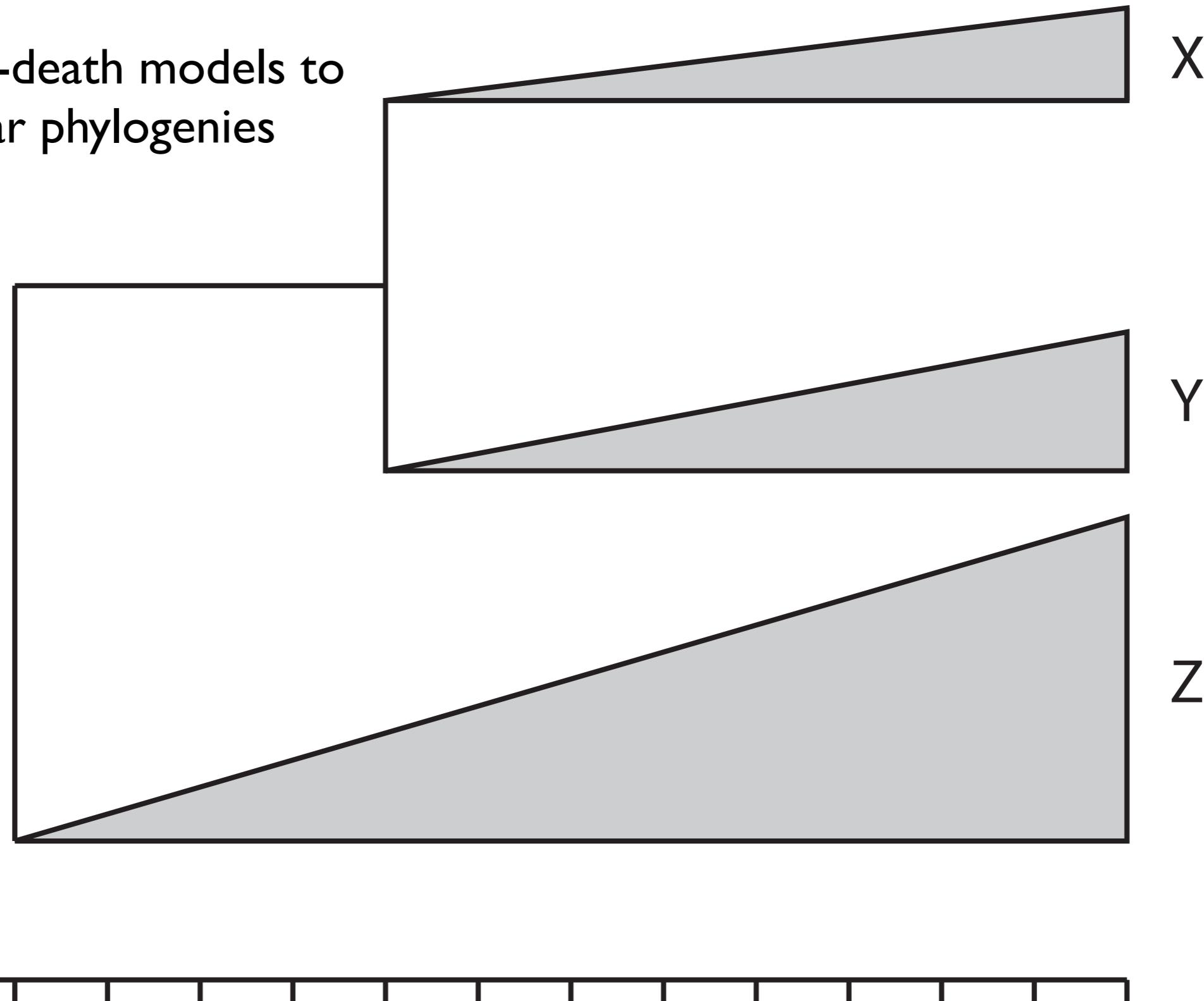
I. Extinction rates are poorly estimated if birth-death model assumptions are violated

- E.g. rates are not time-homogeneous

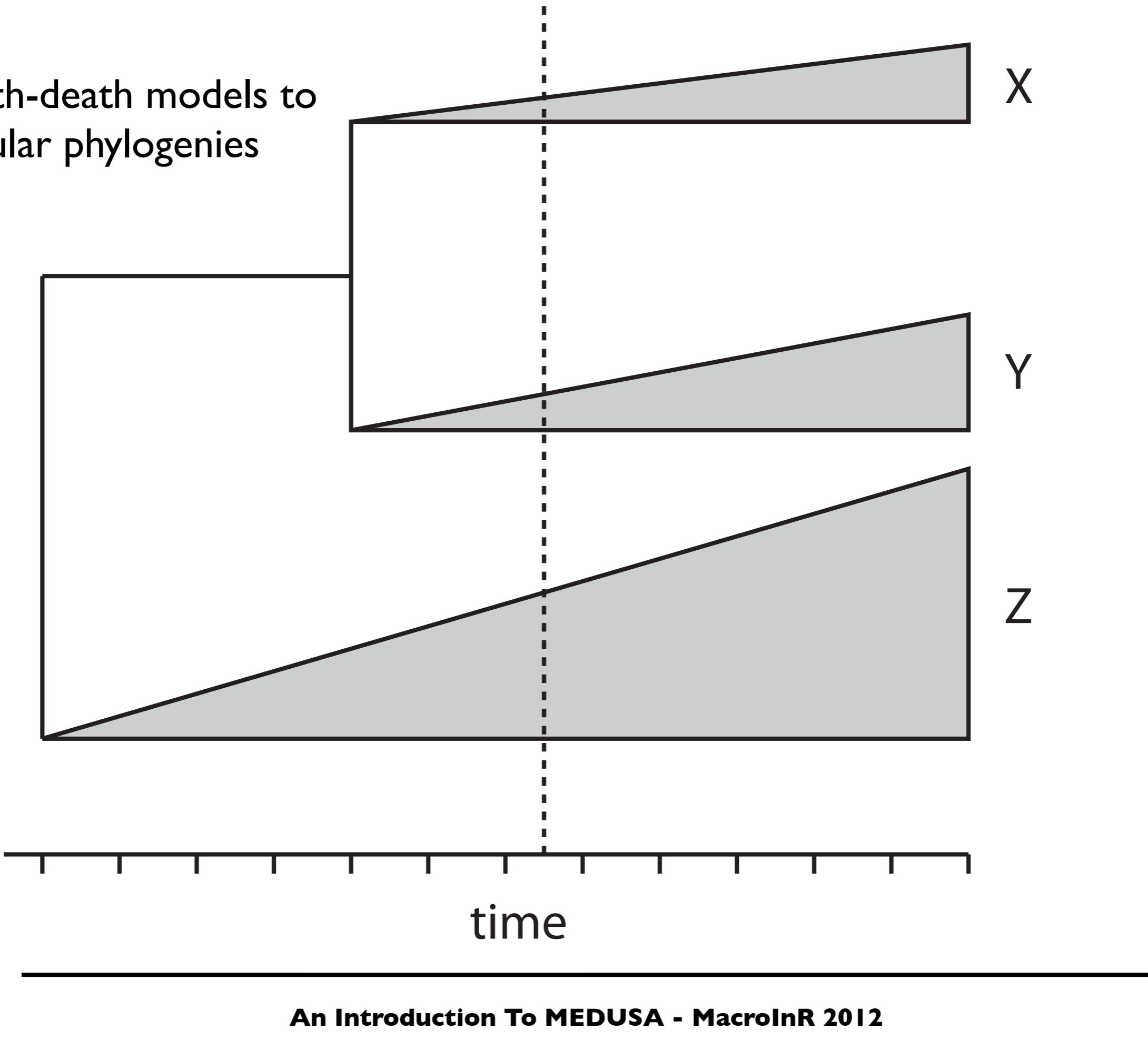


An Introduction To MEDUSA - MacroInR 2012

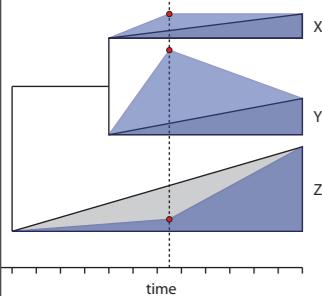
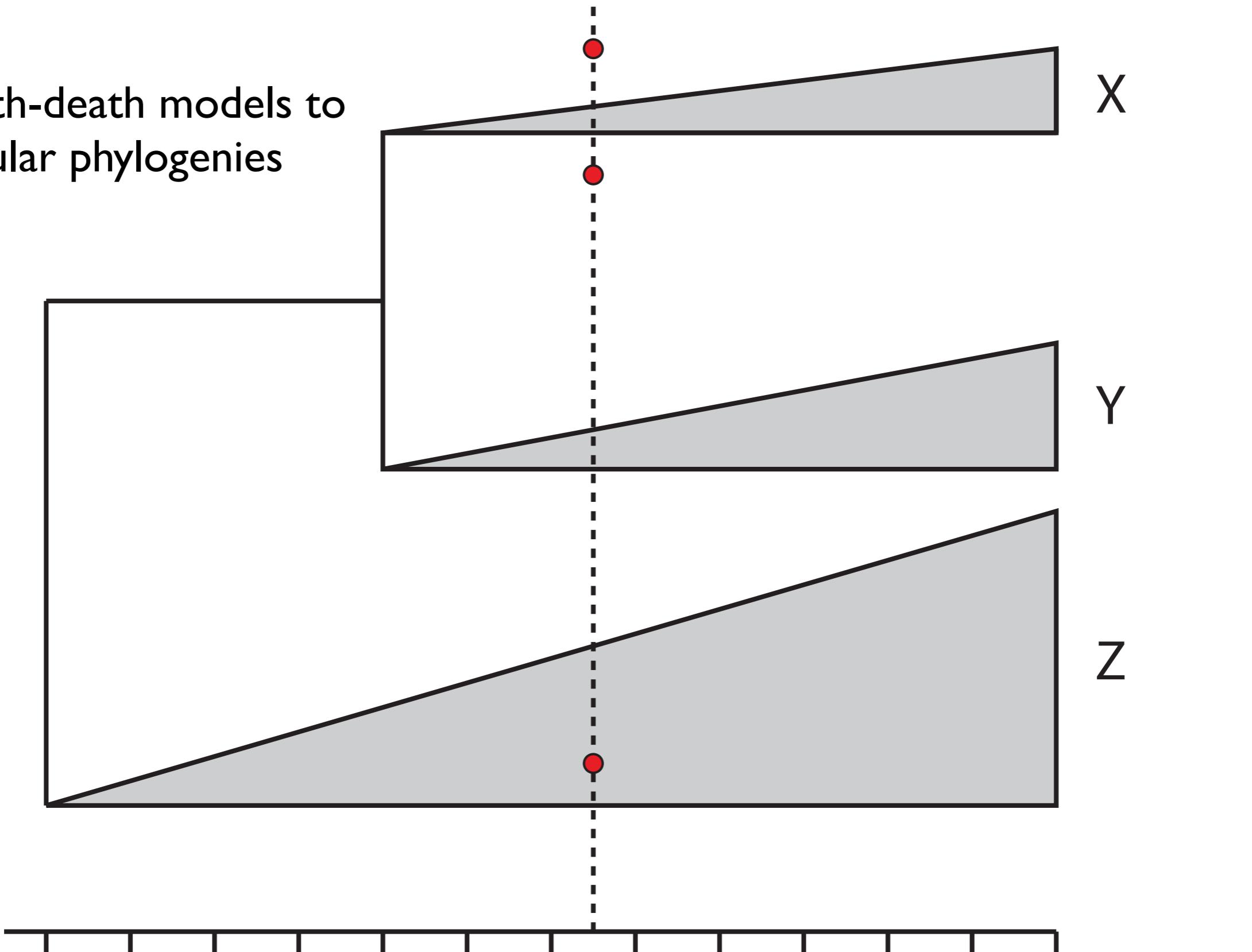
Fitting birth-death models to molecular phylogenies



Fitting birth-death models to molecular phylogenies

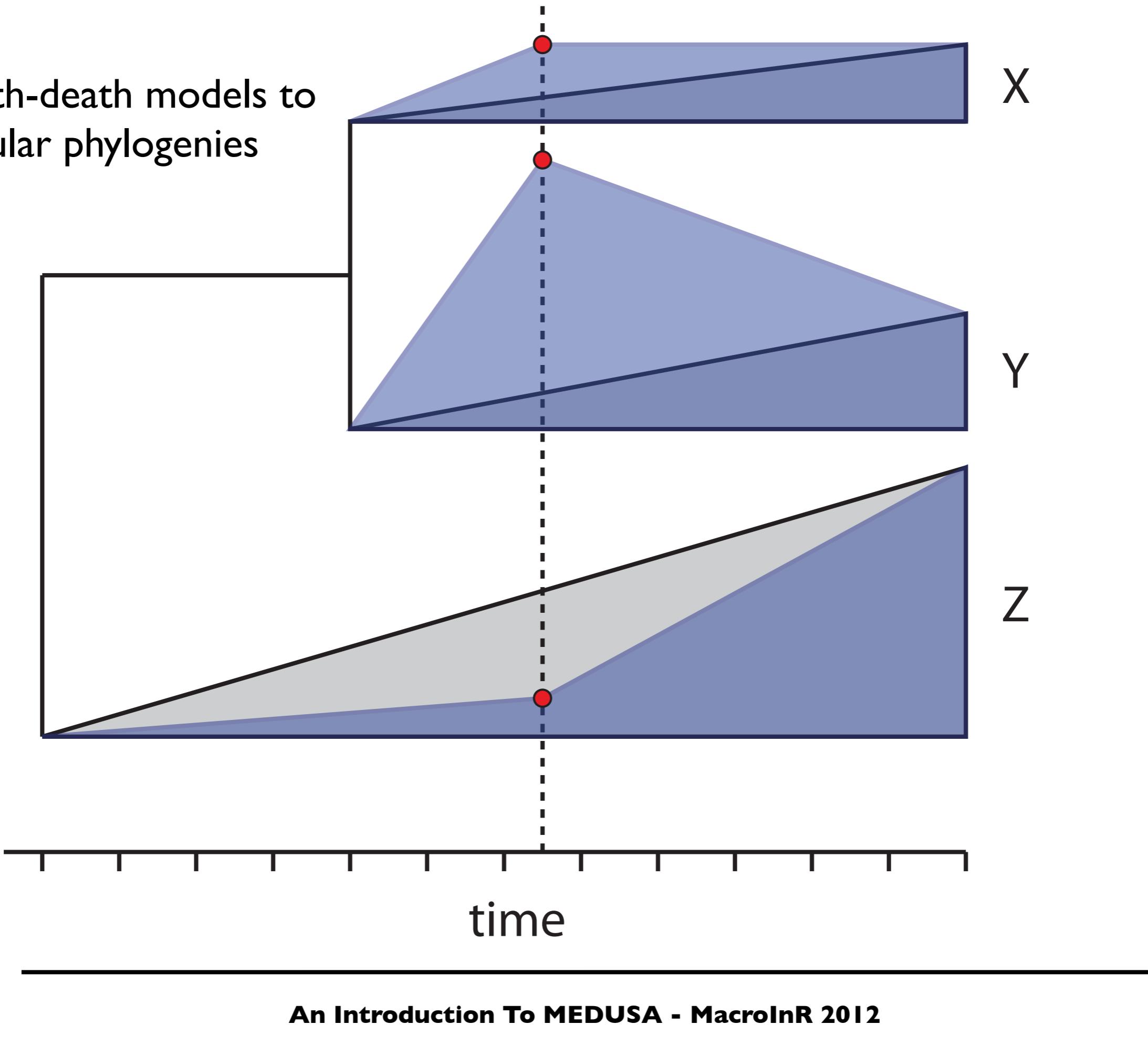


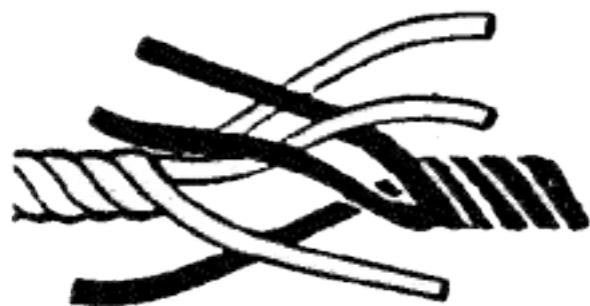
Fitting birth-death models to molecular phylogenies



An Introduction To MEDUSA - MacroInR 2012

Fitting birth-death models to molecular phylogenies





Existing Maths



- The paleontological literature is replete with maths describing changes in diversity (both increases and decreases)
- These paleontological likelihoods can be spliced with phylogenetic likelihoods to more fully explain macroevolutionary dynamics

Paleobiology, 11(1), 1985, pp. 42–52

Mathematical models of cladogenesis

David M. Raup

Abstract.—The evolutionary pattern of speciation and extinction in any biologic group may be described by a variety of mathematical models. These models provide a framework for describing the history of taxonomic diversity (clade shape) and other aspects of larger evolutionary patterns. The simplest model assumes time homogeneity; that is, speciation and extinction probabilities are constant through time and

REPORTS

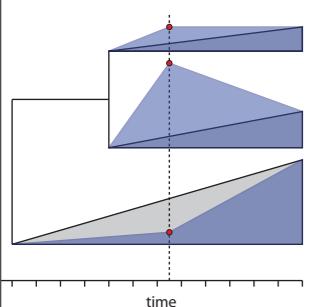
Evolutionary and Preservational Constraints on Origins of Biologic Groups: Divergence Times of Eutherian Mammals

Mike Foote,^{1*} John P. Hunter,² Christine M. Janis,³ J. John Sepkoski Jr.¹

Some molecular clock estimates of divergence times of taxonomic groups undergoing evolutionary radiation are much older than the groups' first observed fossil record. Mathematical models of branching evolution are used to estimate the maximal rate of fossil preservation consistent with a postulated missing history, given the sum of species durations implied by early origins under a range of species origination and extinction rates. The plausibility of postulated divergence times depends on origination, extinction, and preservation rates estimated from the fossil record. For eutherian mammals, this approach suggests that it is unlikely that many modern orders arose much earlier than their oldest fossil records.

preservation and preservation at least once. We consider a hypothesis of missing diversity plausible if the probability of complete nonpreservation of the group is at least 0.5. This is a conservative value.

We estimate the sum of missing species durations implied by a hypothesized divergence time. This sum increases with (i) the length of missing history, (ii) the diversity at the end of this interval, and (iii) the extinction rate in most diversity models (Fig. 3) (J3, J8). Increases in all three parameters demand more extinct species evolving before the time that the group is first observed. Because the length of missing history and the minimal diversity at the group's first fossil appearance are given by the hypothesized time of origin and by observed fossils, the parameters that need to be constrained are extinction rate and preservation rate. For a group's summed species durations to be unobserved, the extinction rate, the preservation rate, or both must fall below some threshold (Fig. 3). We can thus place upper probabilistic bounds on the rates consistent with the hypothesis of



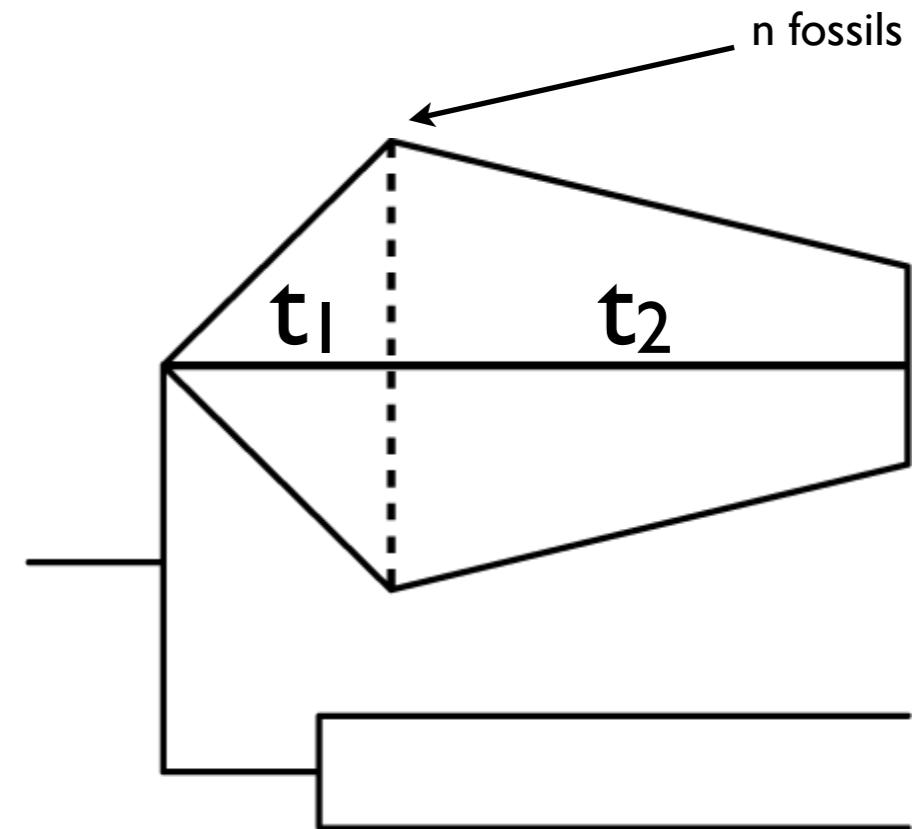
Raup. 1985. *Paleobiology*; Foote et al. 1999. *Science*

An Introduction To MEDUSA - MacroInR 2012

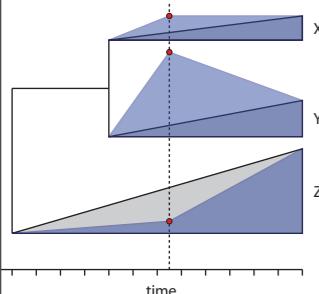
Fit Birth-Death Process to Tree, With Fossils

Divide edge into 2 (or more) sub-edges
For interval t_1 , calculate like pendant edge
For interval t_2 , assuming fossils are exact:

$$P(n,t,a) = \sum_{j=1}^{\min(a,n)} \binom{a}{j} \binom{n-1}{j-1} \cdot A^{a-j} [(1-A)(1-B)]^j B^{n-j}$$



Entire Tree $\log L_T = \log L_I + \log L_P + \log L_F$



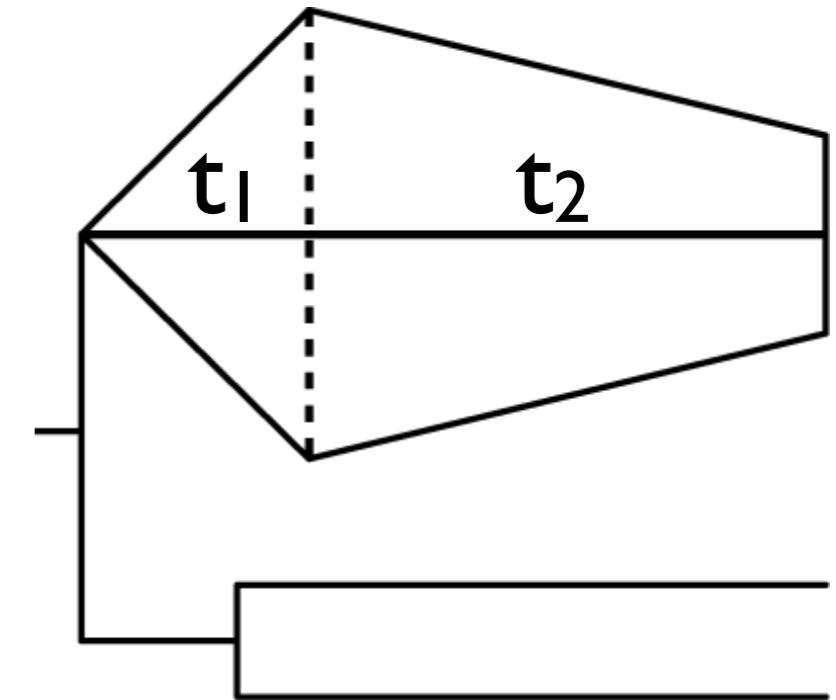
Raup. 1985. Paleobiology; Foote et al. 1999. Science

An Introduction To MEDUSA - MacroInR 2012

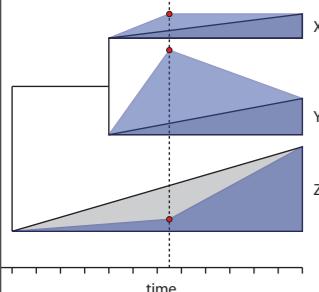
Fit Birth-Death Process to Tree, With Fossils

Treating a fossil as a minimum count:

$$P(\geq n, t, a) = 1 - P(0, t, a) - \sum_{m=1}^{n-1} P(m, t, a)$$



Entire Tree $\log L_T = \log L_I + \log L_P + \log L_F$



Raup. 1985. Paleobiology; Foote et al. 1999. Science

An Introduction To MEDUSA - MacroInR 2012

FOSSIL MEDUSA



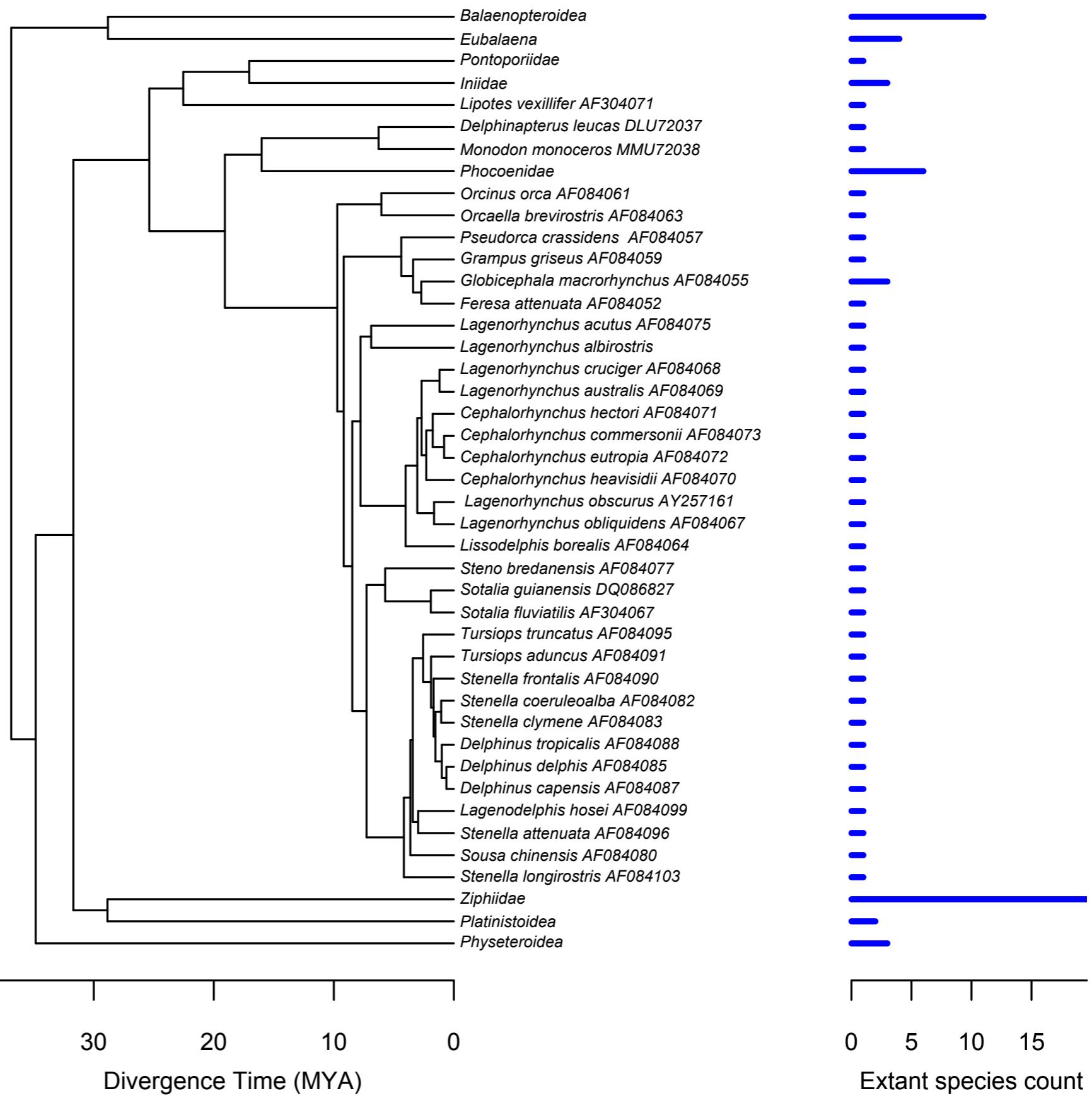
- Fossil richness information can come from:
 - I. Time-slice(s) through the tree
 - May be appropriate when considering periods of following mass extinction
 2. Unique time periods for each taxon
 - Appropriate when temporal quality of the fossil record differs across taxa, or if trying to incorporate the maximum amount of information from the fossil record



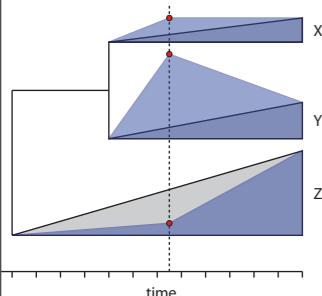
An Introduction To MEDUSA - MacroInR 2012

Cetaceans

- Cetaceans are of particular interest, as several lineages have previously enjoyed higher diversity in the past



Slater et al. 2010. Proc. Roy. Soc.



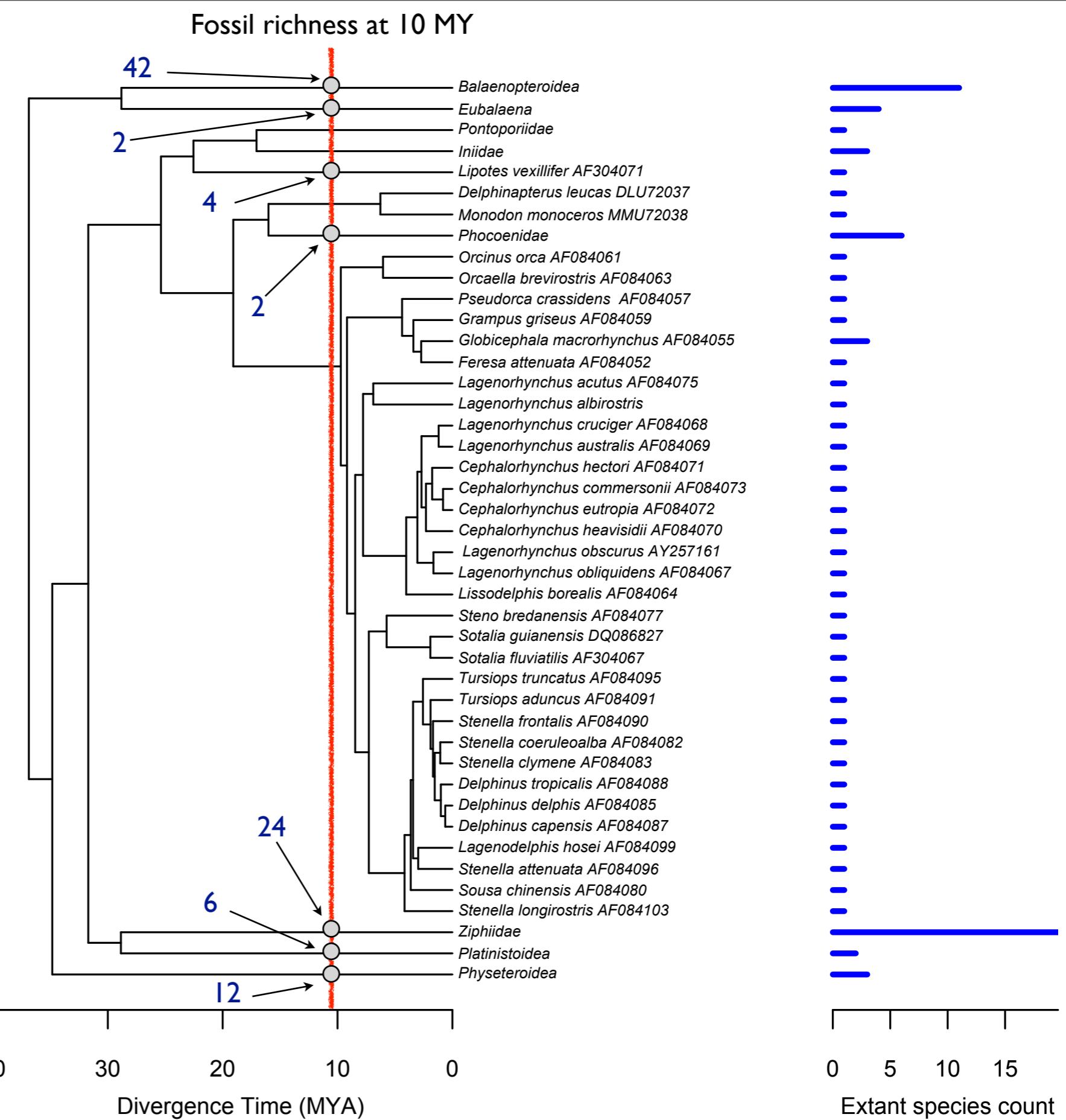
An Introduction To MEDUSA - MacroInR 2012

Cetaceans

- Cetaceans are of particular interest, as several lineages have previously enjoyed higher diversity in the past

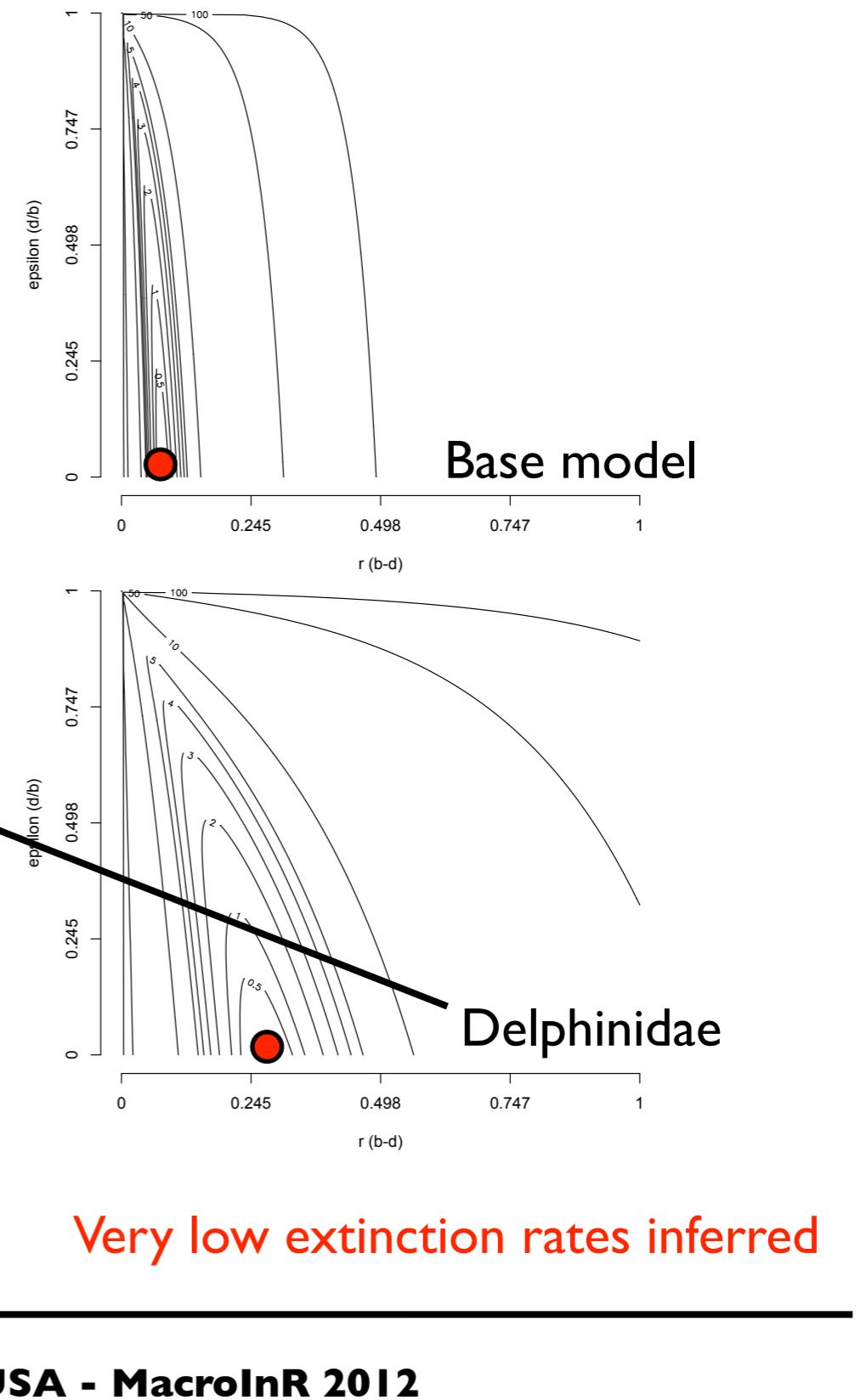
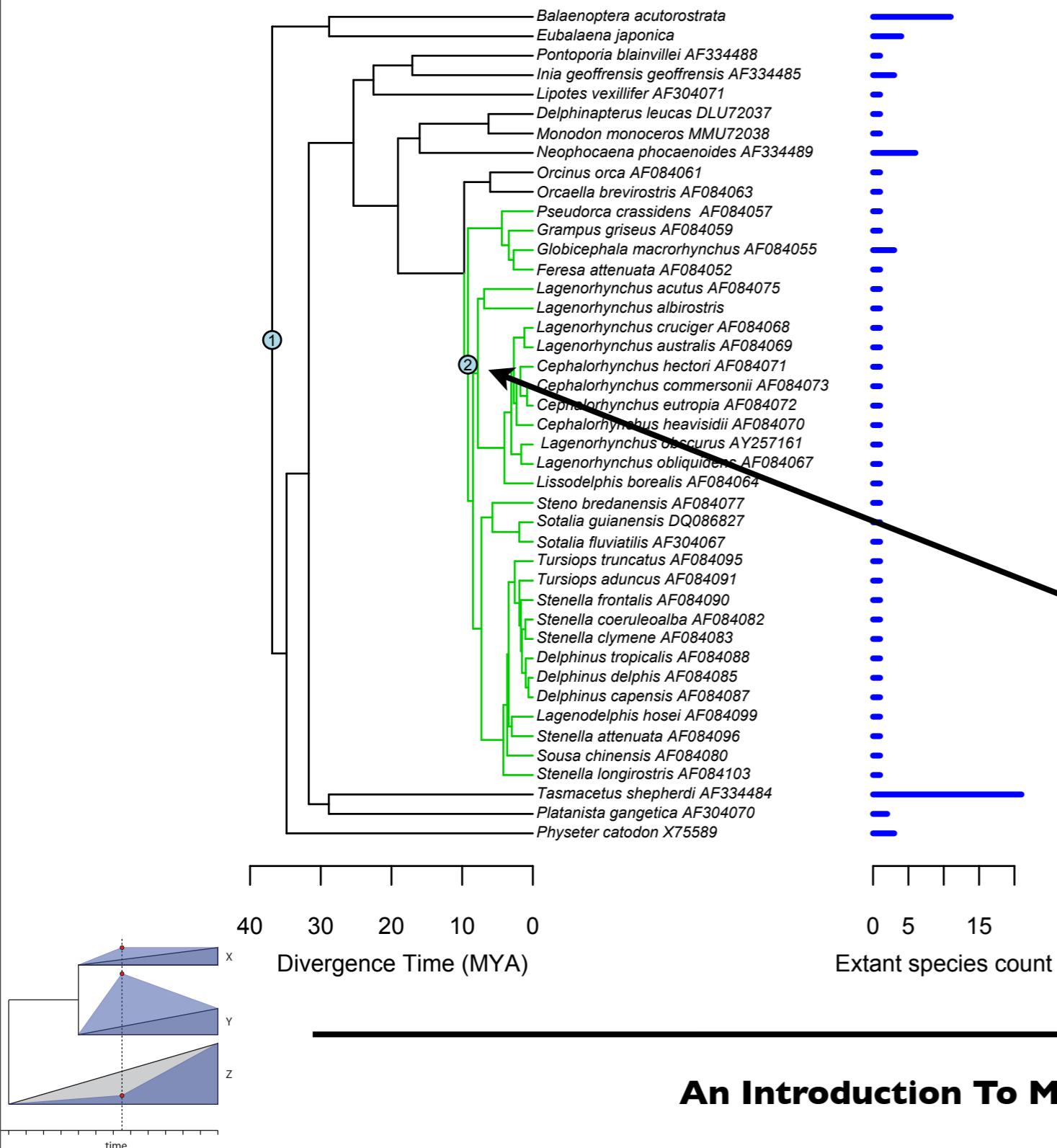
Data from G Slater via the Paleobiology Database (<http://paleodb.org>)

Slater et al. 2010. Proc. Roy. Soc.



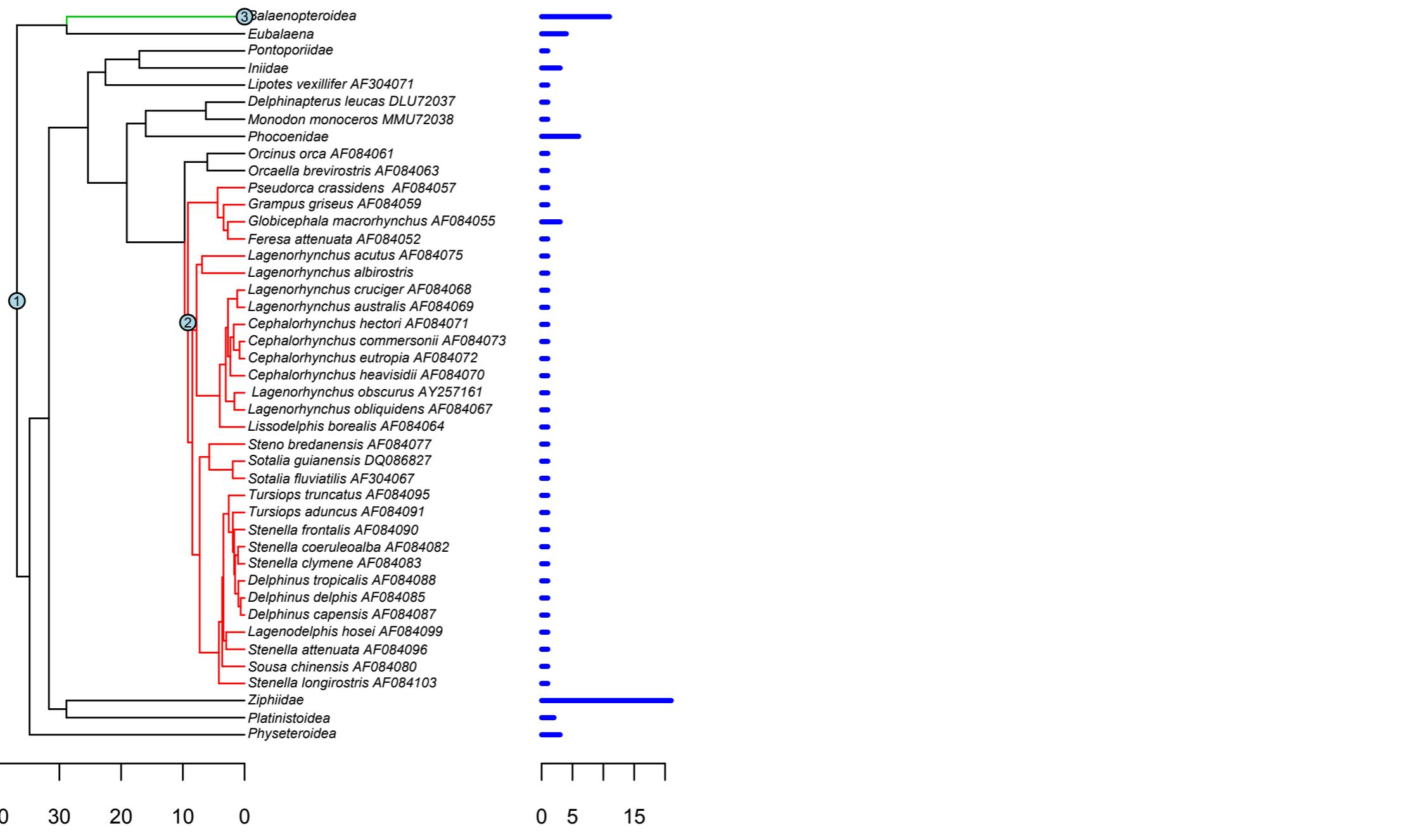
An Introduction To MEDUSA - MacroInR 2012

Without Fossil Information



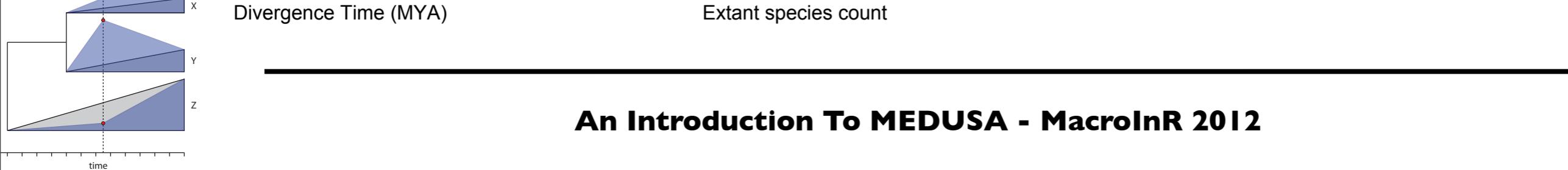
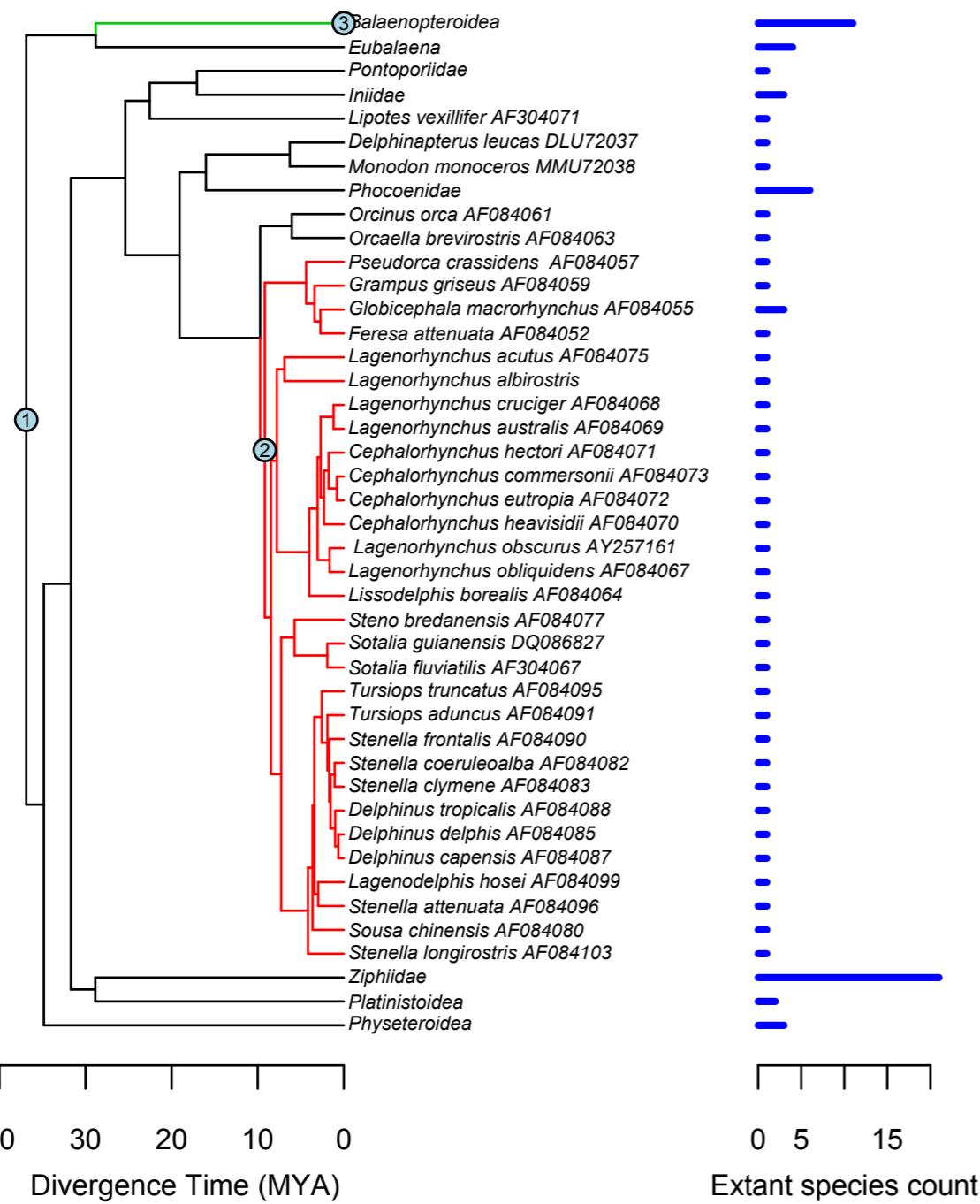
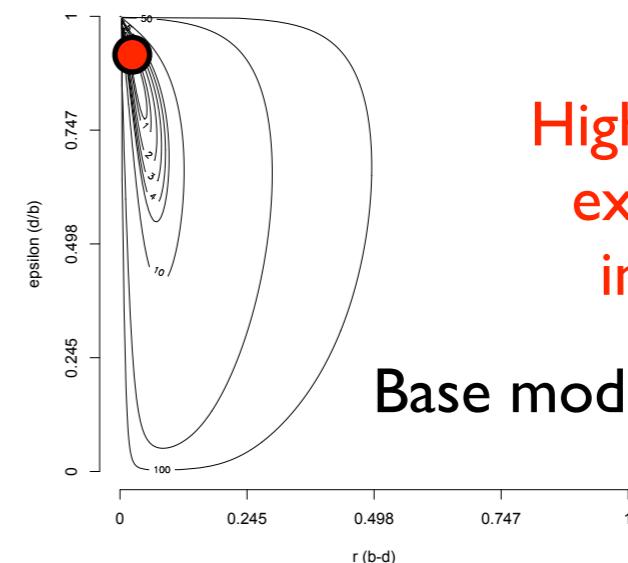
An Introduction To MEDUSA - MacroInR 2012

Incorporating Fossils

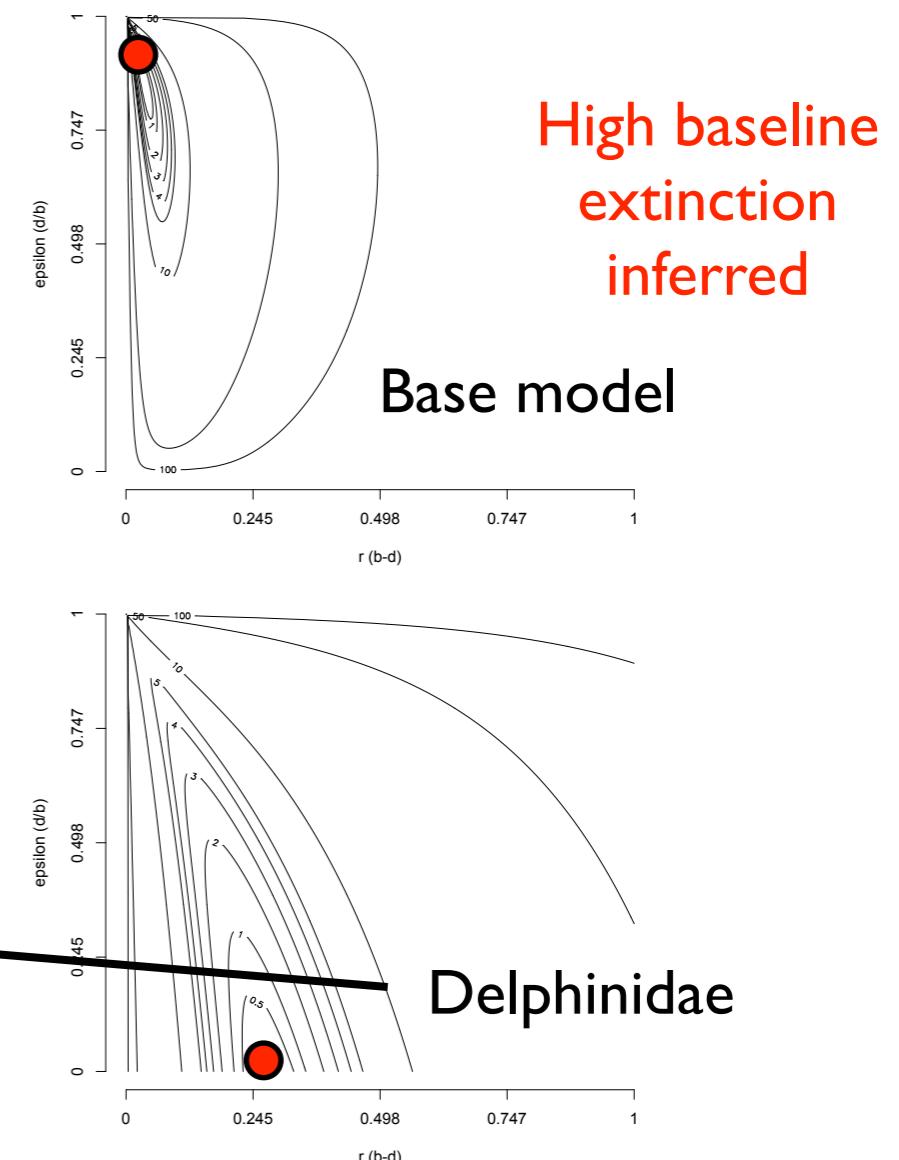
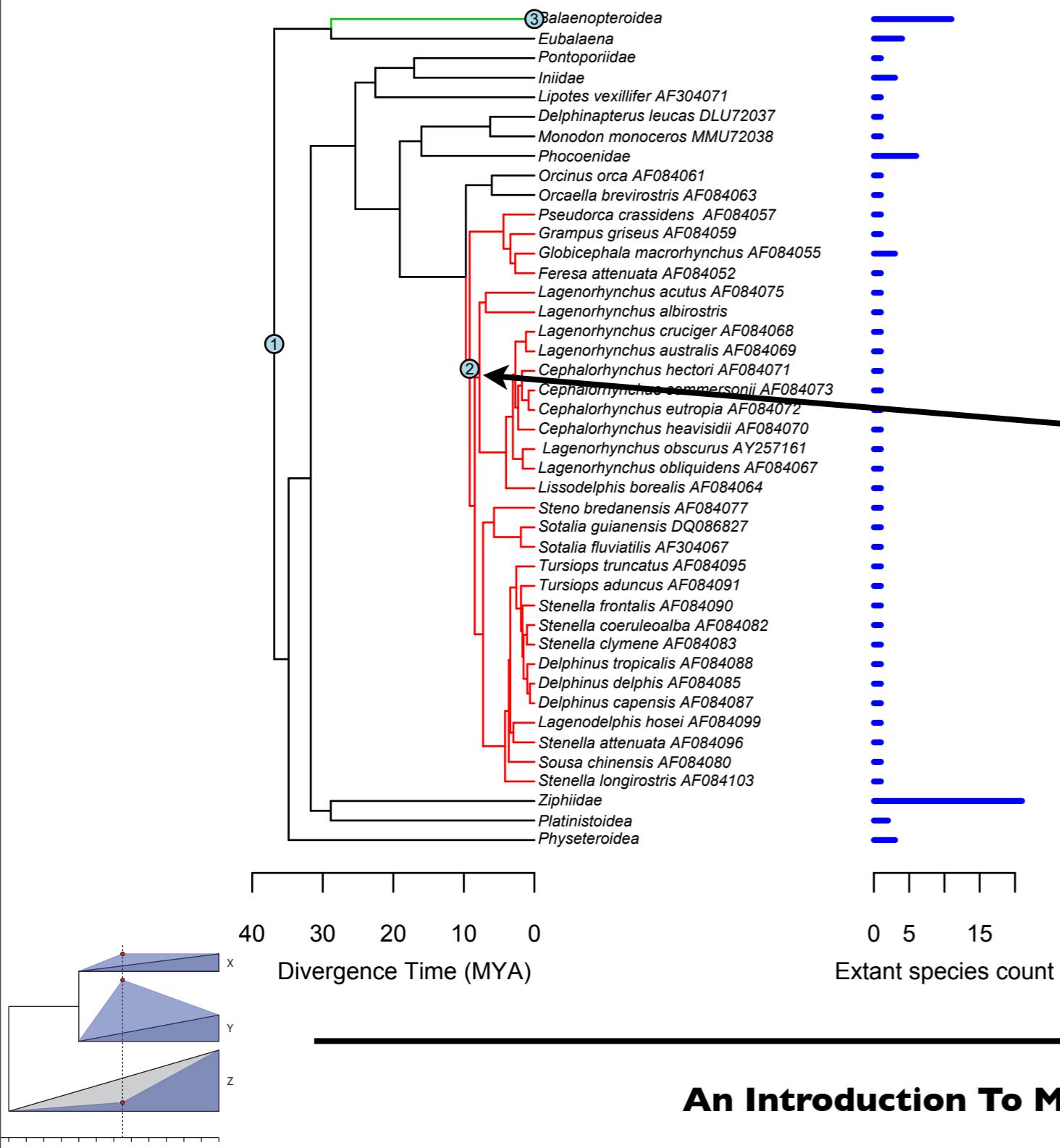


An Introduction To MEDUSA - MacroInR 2012

Incorporating Fossils

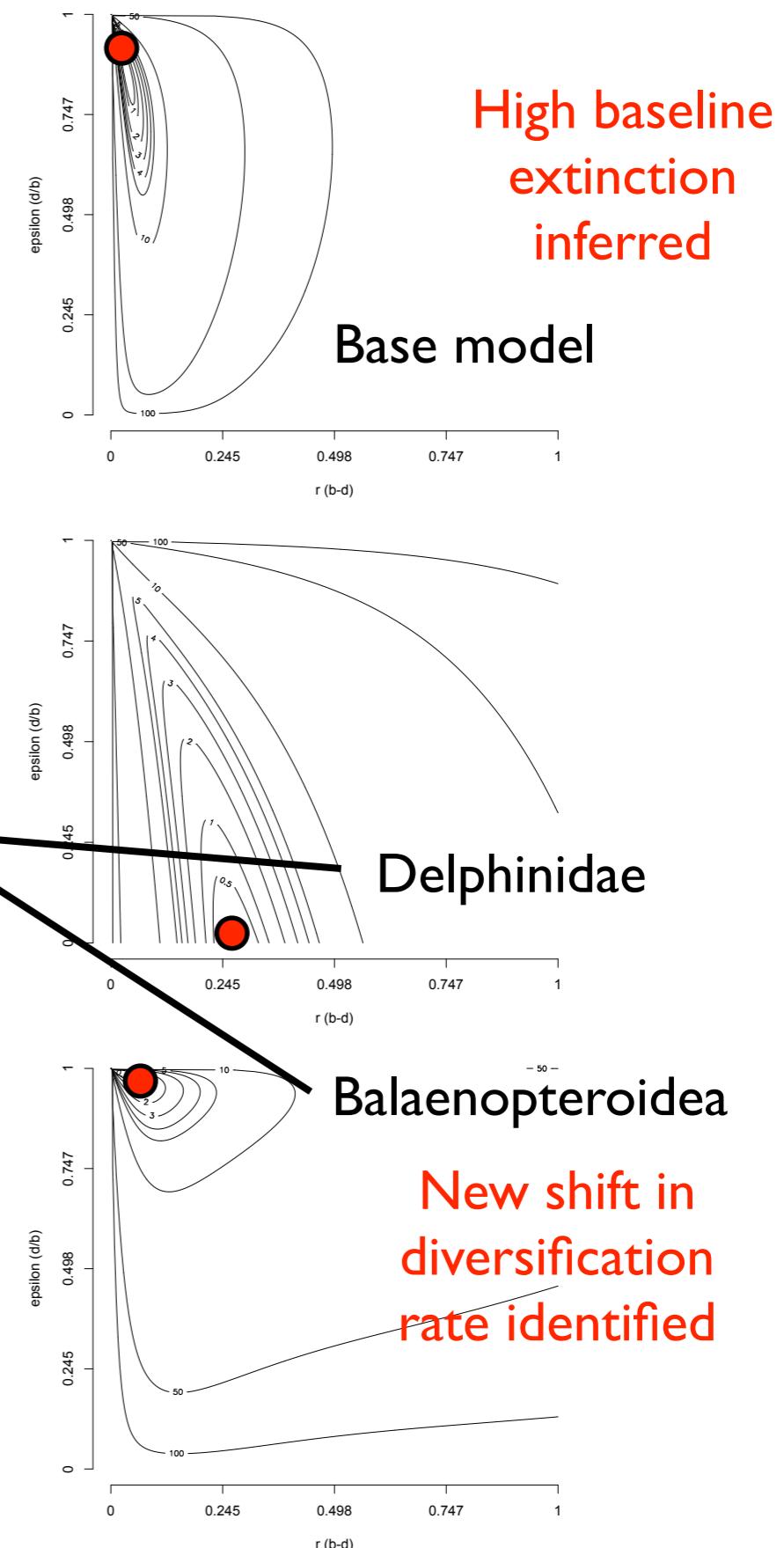
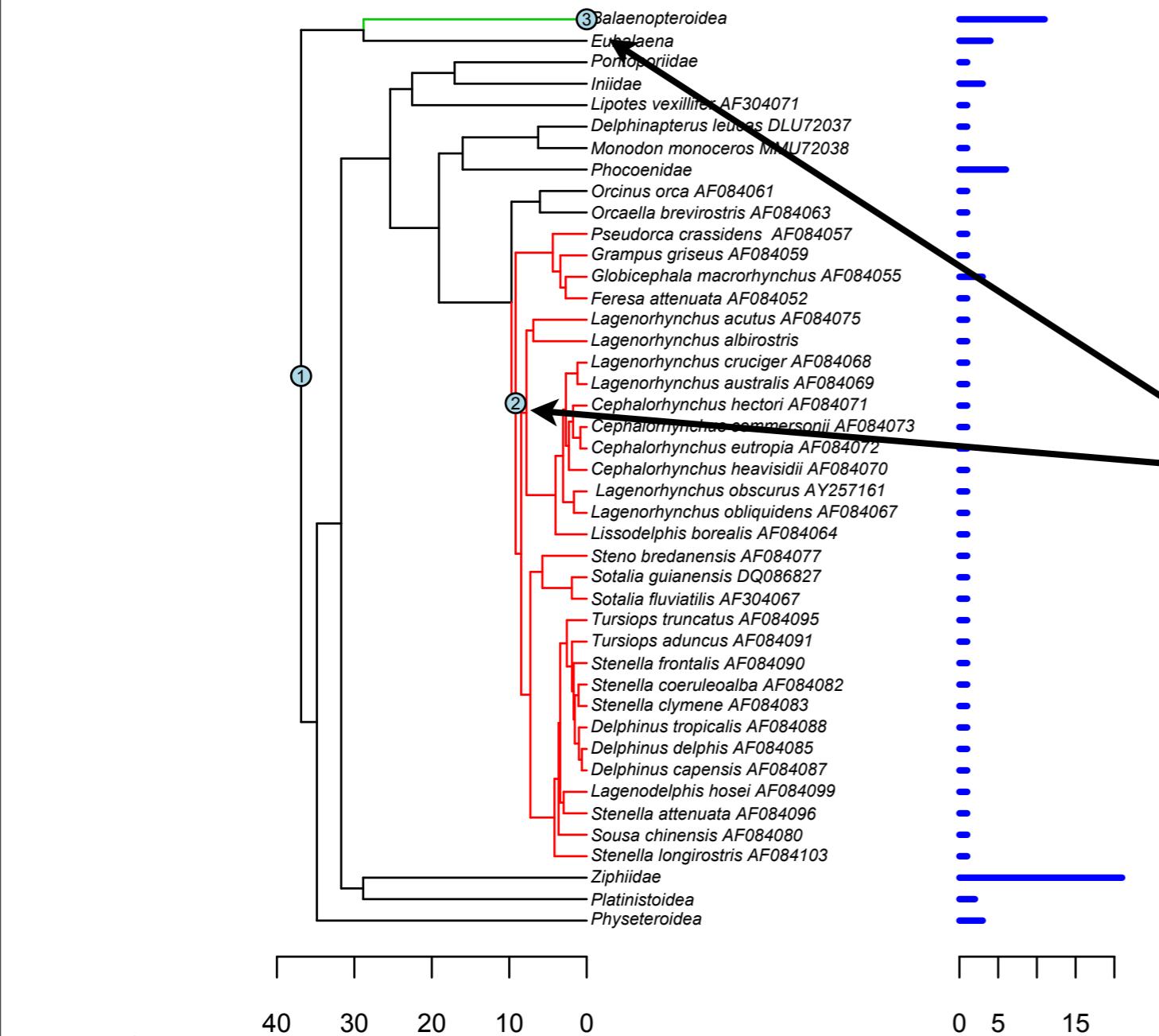


Incorporating Fossils



An Introduction To MEDUSA - MacroInR 2012

Incorporating Fossils



An Introduction To MEDUSA - MacroInR 2012

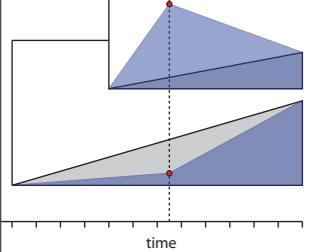
High baseline extinction inferred

Base model

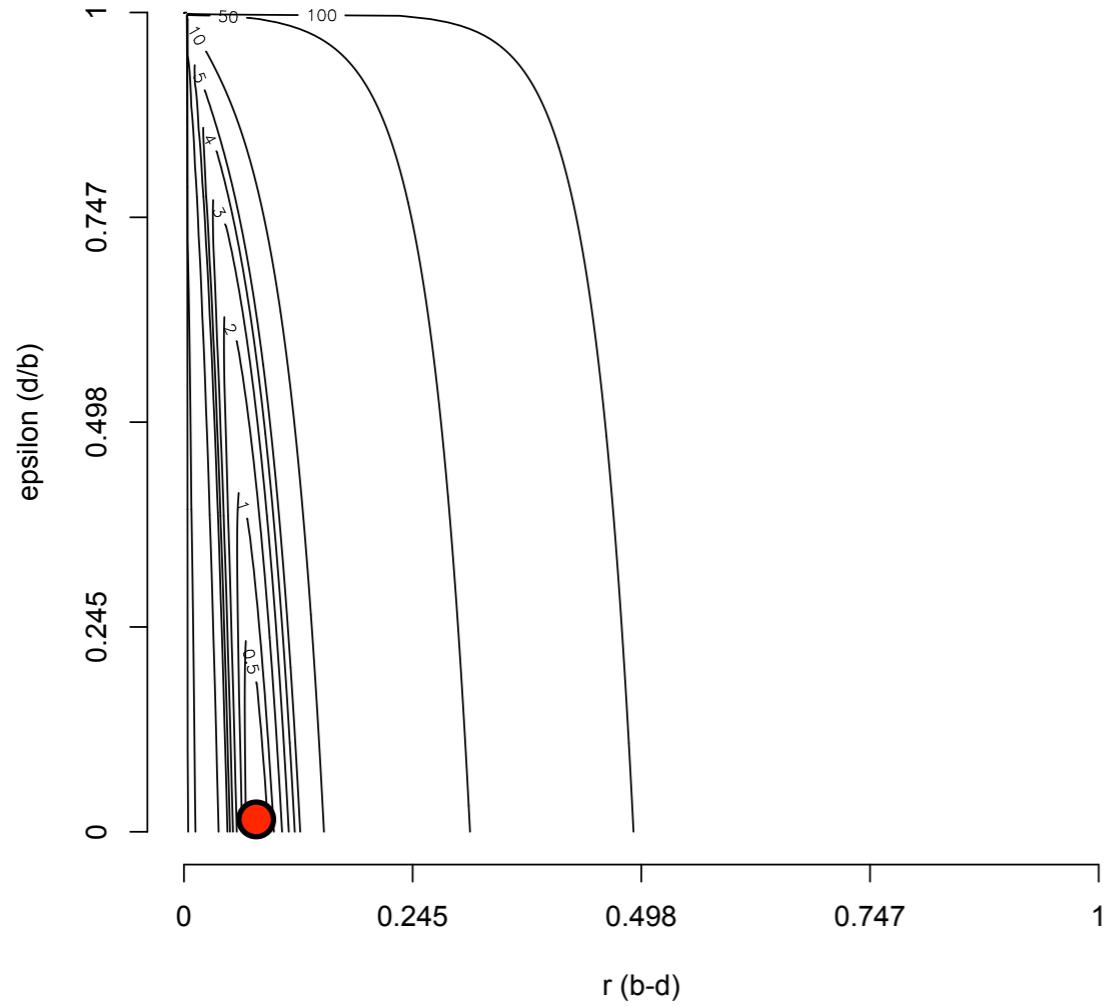
Delphinidae

Balaenopteroidea

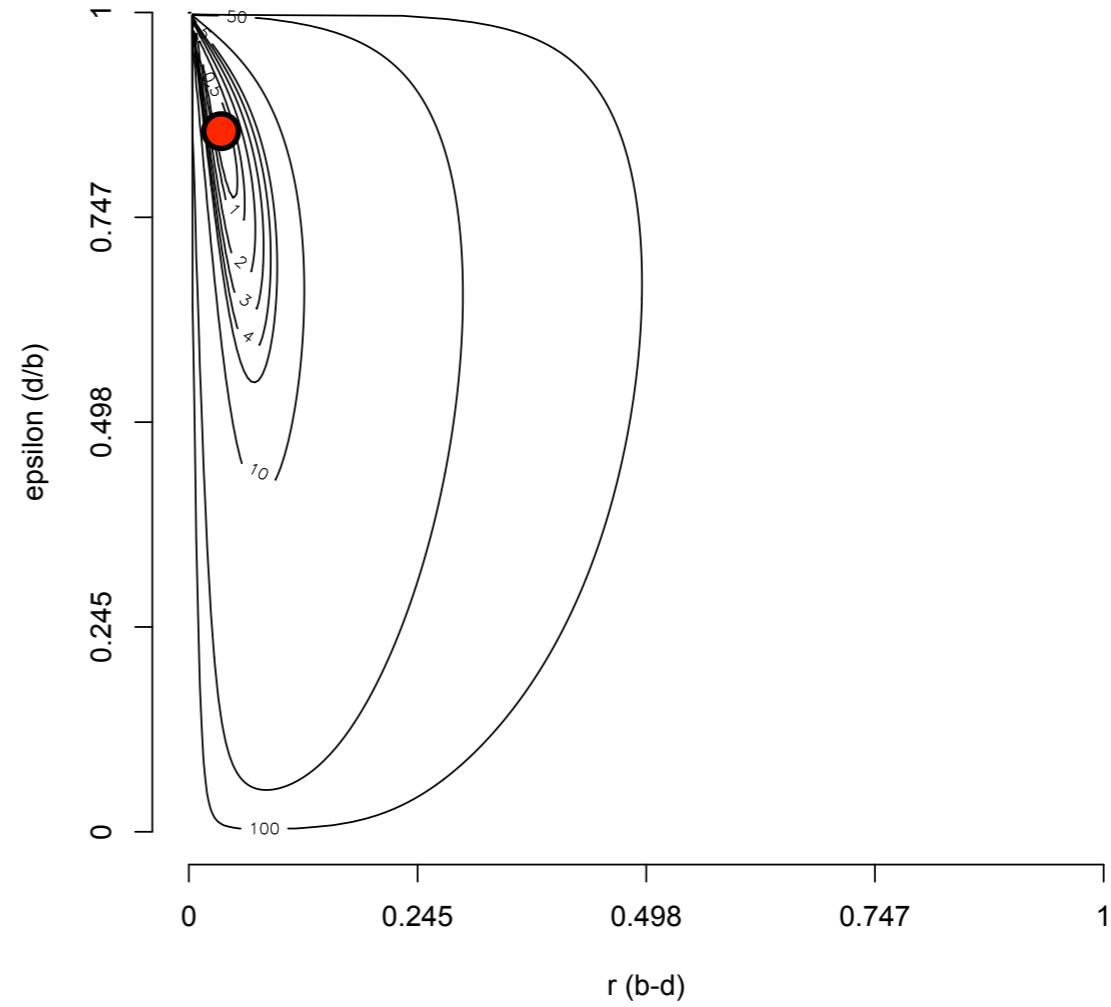
New shift in diversification rate identified



Cetaceans

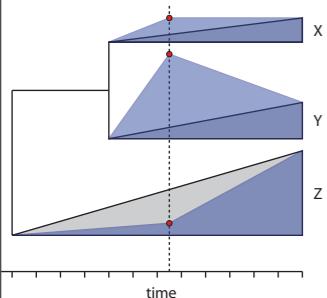


No fossils



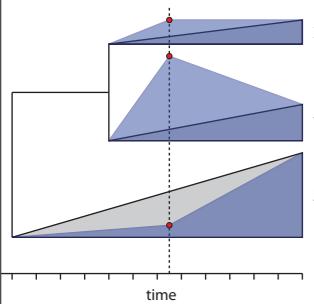
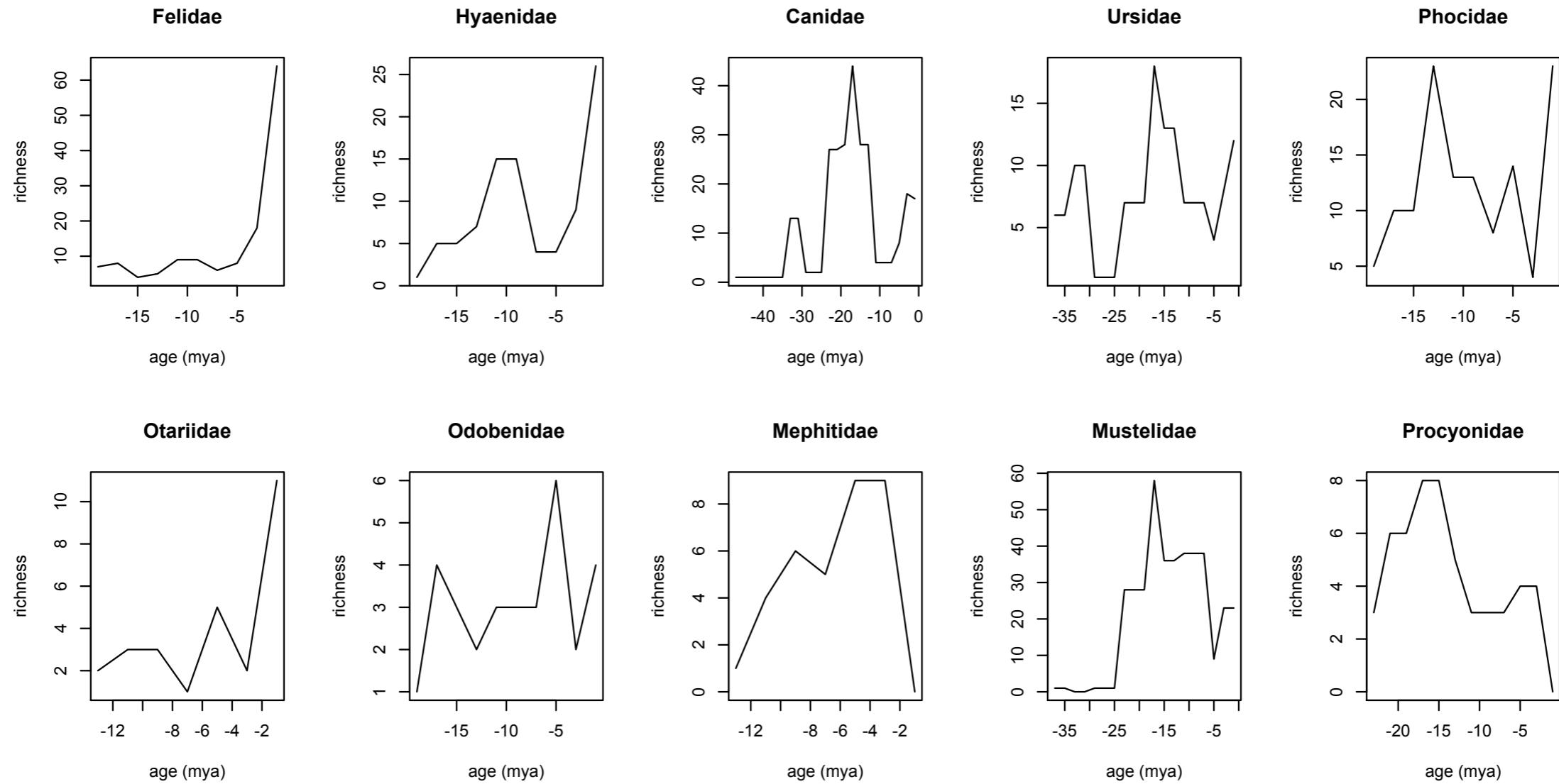
With fossils

The incorporation of fossil richness information dramatically changes our inference from little extinction to very high extinction



An Introduction To MEDUSA - MacroInR 2012

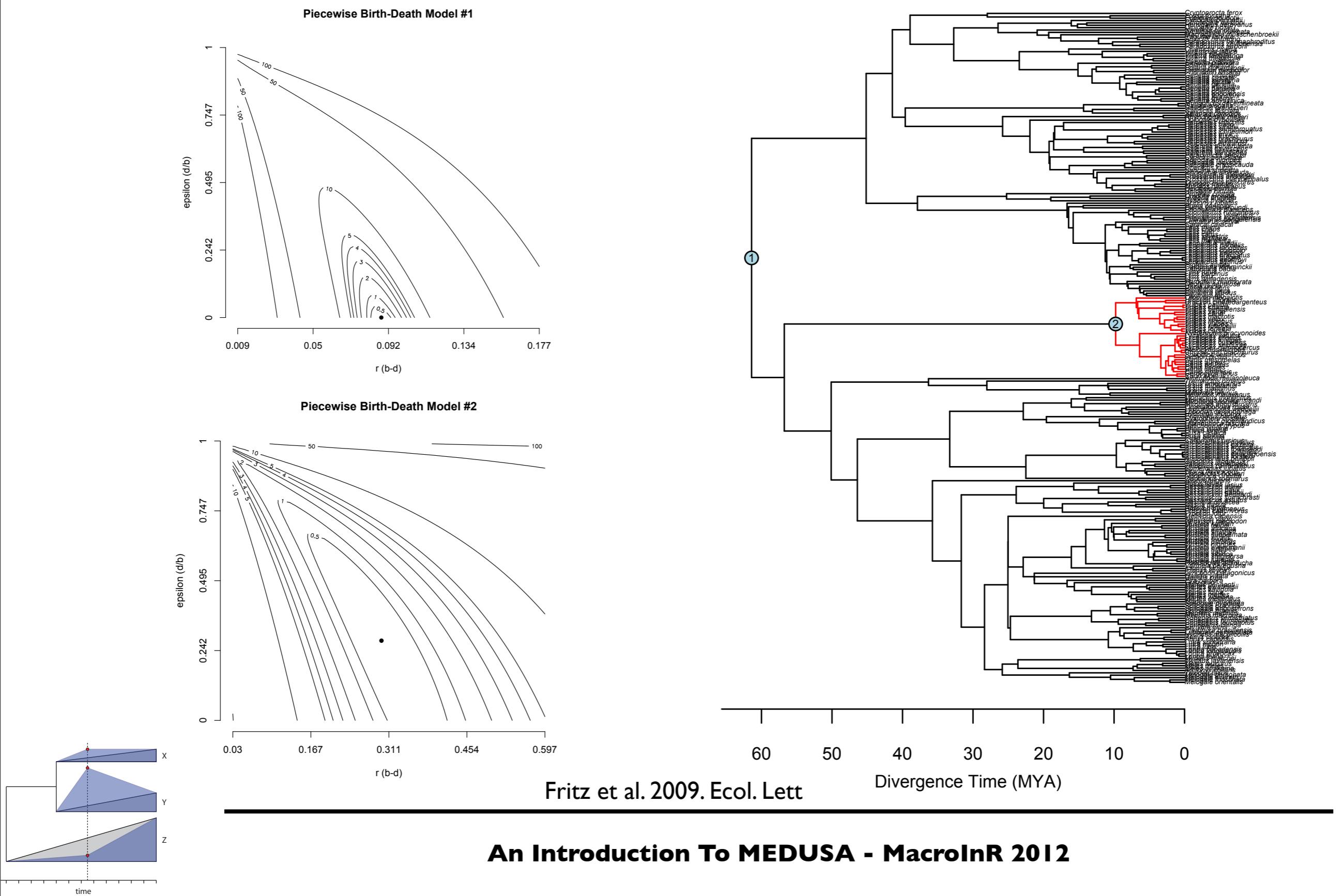
Carnivore Fossil Richness



G Slater via the Paleobiology Database (<http://paleodb.org>)

An Introduction To MEDUSA - MacroInR 2012

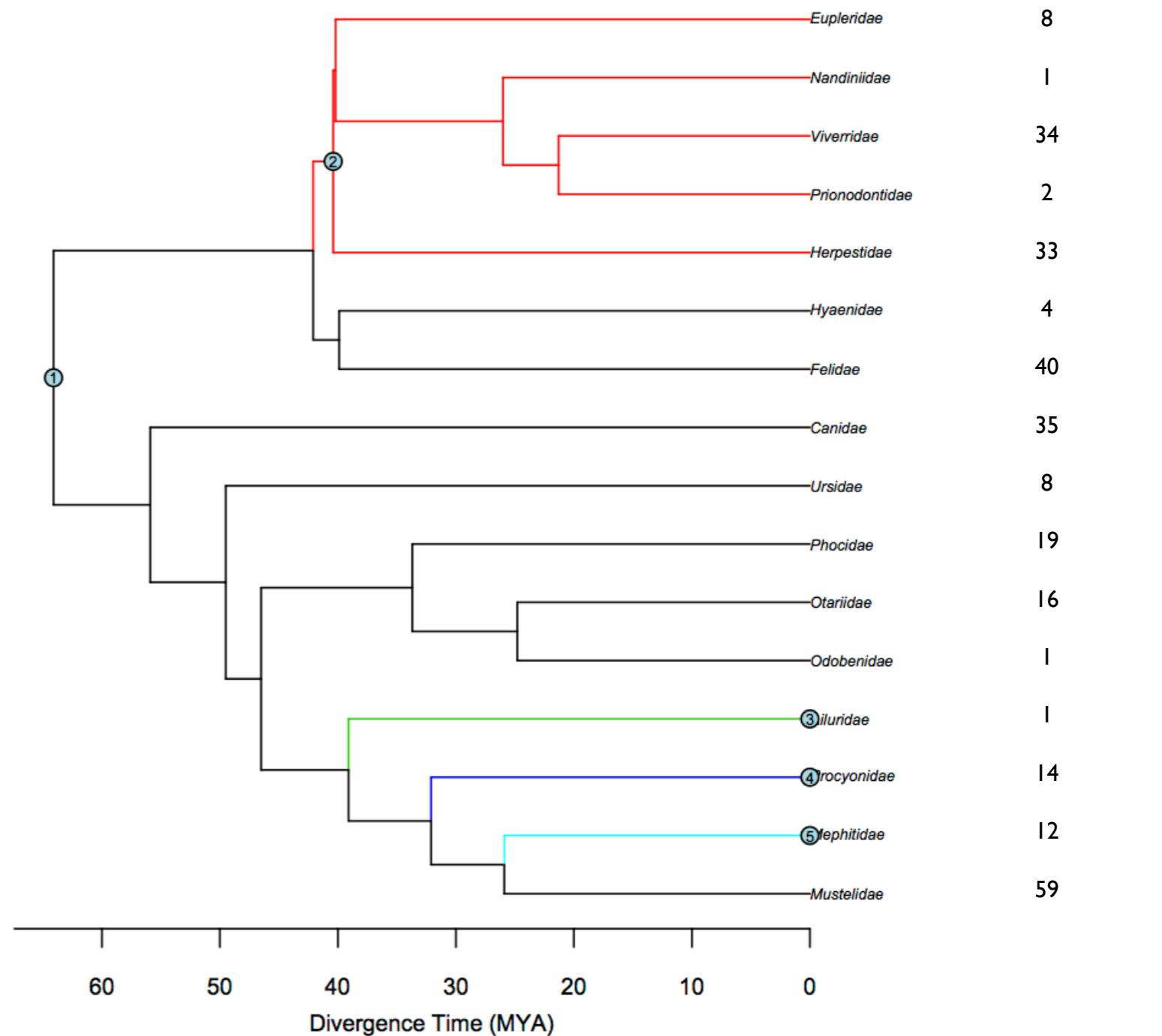
Without Fossil Information



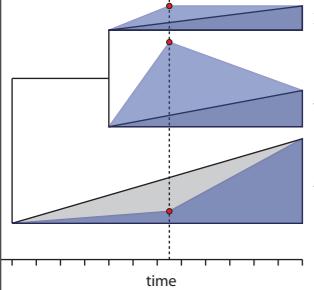
Incorporating Fossils

Current Richness

split	r	epsilon
1	0.039537	0.98054
2	0.081911	0.42352
3	4.79E-11	0.51603
4	0.075288	0.72951
5	0.085828	0.60761



An Introduction To MEDUSA - MacroInR 2012



Conclusions/Implications

I. Incorporating fossil richness information:

- gives reasonable estimates of extinction rate
 - refines/overturns our understanding of patterns of macroevolutionary diversification
2. However, methods are only applicable to lineages with a decent fossil record
- Can we accurately reconstruct past diversification dynamics from taxa lacking a fossil record?

