

# Phylogenetic Approaches for Exploring Differential Rates of Diversification\*

\*lineage diversification = (speciation – extinction)  
aka *real* diversification!

Brian R. Moore & Mike R. May

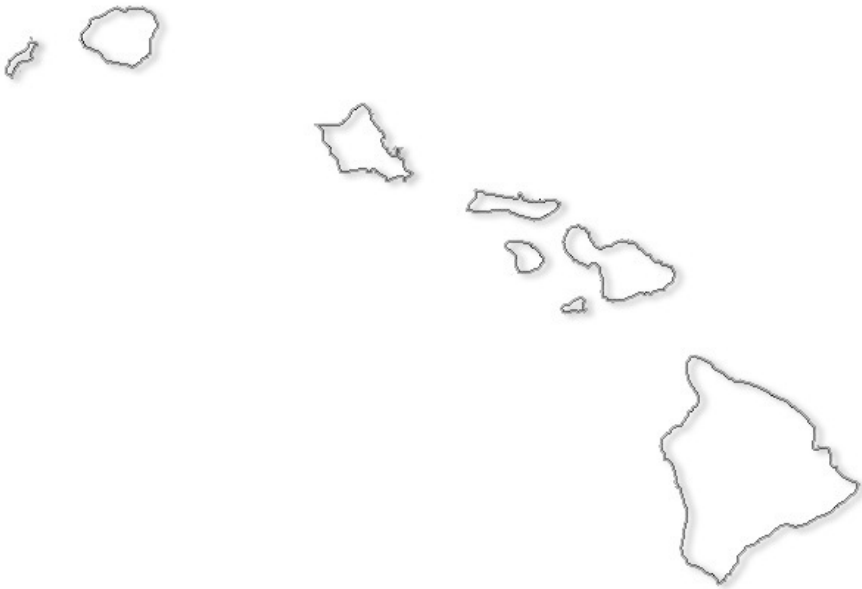
Department of Evolution & Ecology

UC, Davis

2014 Bodega Workshop

# Differential Rates of Diversification

adaptive radiation

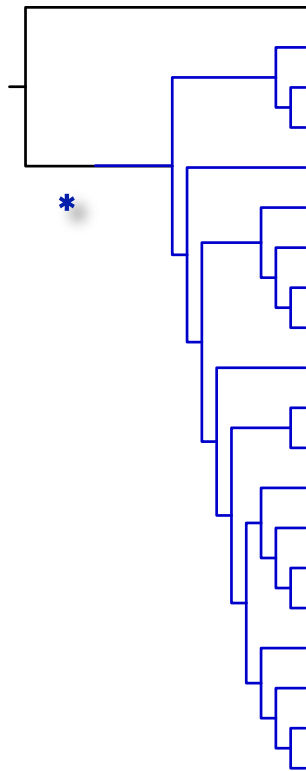


Hawaiian silverswords

# Differential Rates of Diversification

adaptive radiation

key innovations



nectar spurs in columbines

# Differential Rates of Diversification

adaptive radiation

key innovations

rapid cladogenesis



cichlids

# Outline & Organization

- I. Brief overview of general approach to the study of lineage diversification
- II. Brief overview of popular methods for studying lineage diversification  
A beginners guide to the types of methods available
- III. How to apply popular methods for studying lineage diversification  
Using R packages to address various questions about diversification rates

# Outline & Organization

**I. Brief overview of general approach to statistical study of lineage diversification**

**II. Brief overview of popular methods for studying lineage diversification**

A beginners guide to the types of methods available

Me

**III. How to apply popular methods for studying lineage diversification**

Using R packages to address various questions about diversification rates

Mike

# Statistical Phylogenetic Approaches for the Study of Diversification Rates

- I. Pose a substantive question
- II. Develop a stochastic model with parameters that, if known, would answer the question
- III. Collect observations that are informative about model parameters

# Statistical Phylogenetic Approaches for the Study of Diversification Rates



I. Pose a substantive question

II. Develop a stochastic model with parameters that, if known, would answer the question

III. Collect observations that are informative about model parameters

IV. Estimate the model parameters using some method conditioned on the data at hand

# The Five Fundamental Questions in the Study of Lineage Diversification Rates

## 1. Estimating parameters of the stochastic branching-process model

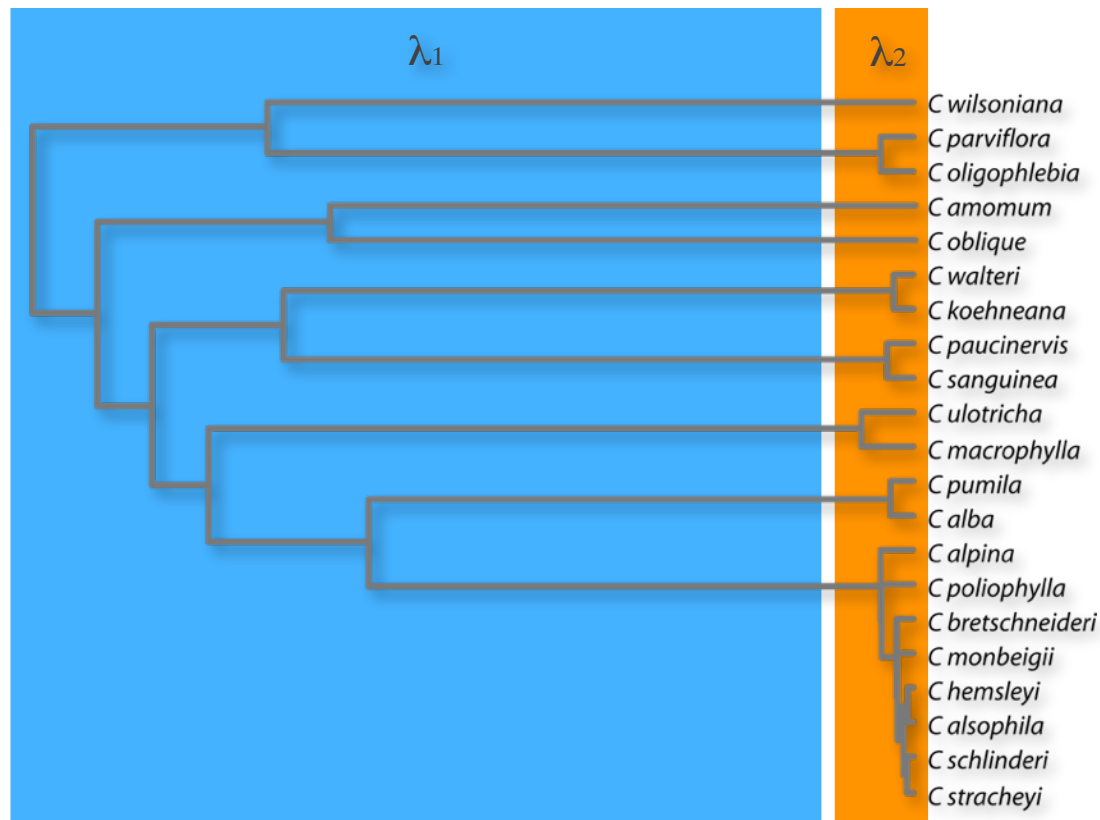
What are the absolute/relative rates of the branching process?

- speciation rate,  $\lambda$
- extinction rate,  $\mu$
- diversification rate,  $\partial = (\lambda - \mu)$
- relative extinction rate,  $\varepsilon = (\mu / \lambda)$

# The Five Fundamental Questions in the Study of Lineage Diversification Rates

1. Estimating parameters of the stochastic branching-process model
2. Identifying significant diversification rate shifts through time

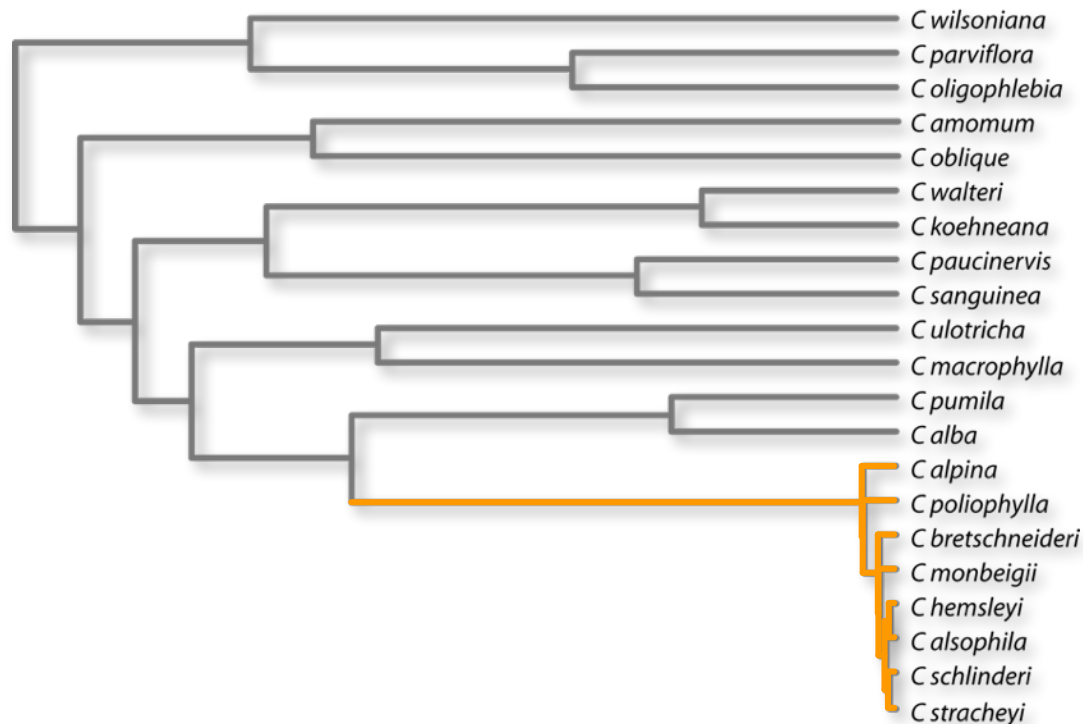
Have tree-wide rates of diversification changed through time?



# The Five Fundamental Questions in the Study of Lineage Diversification Rates

1. Estimating parameters of the stochastic branching-process model
2. Identifying significant diversification rate shifts through time
3. Locating significant diversification rate shifts along branches

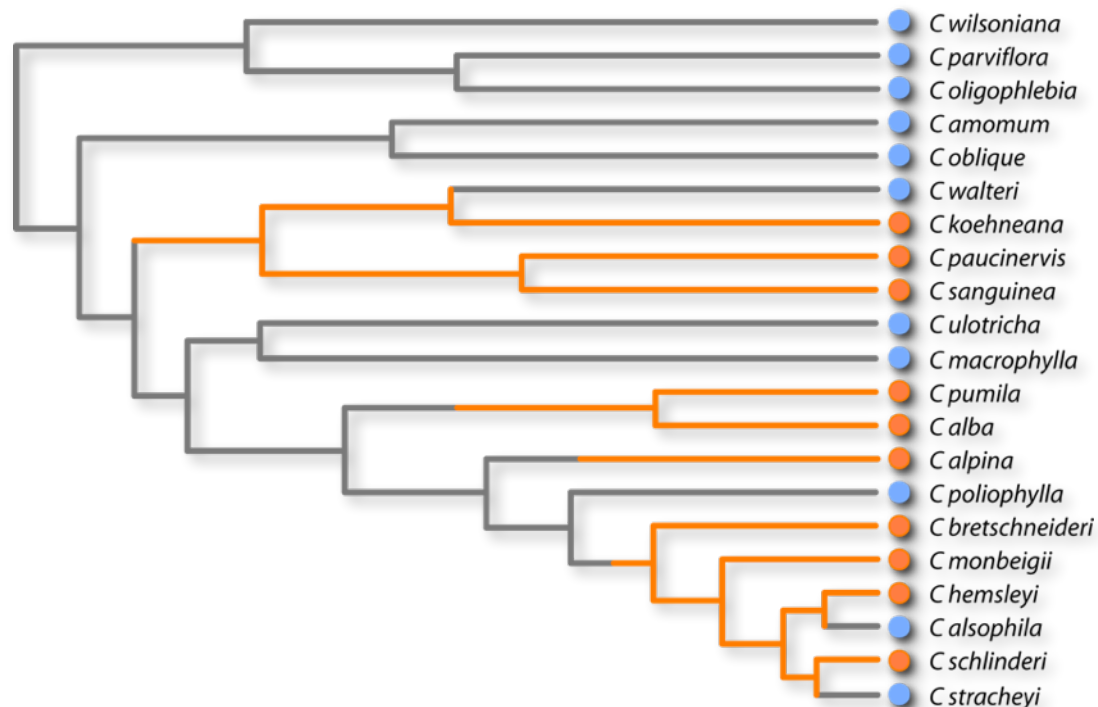
Along which branches have significant rate shifts occurred?



# The Five Fundamental Questions in the Study of Lineage Diversification Rates

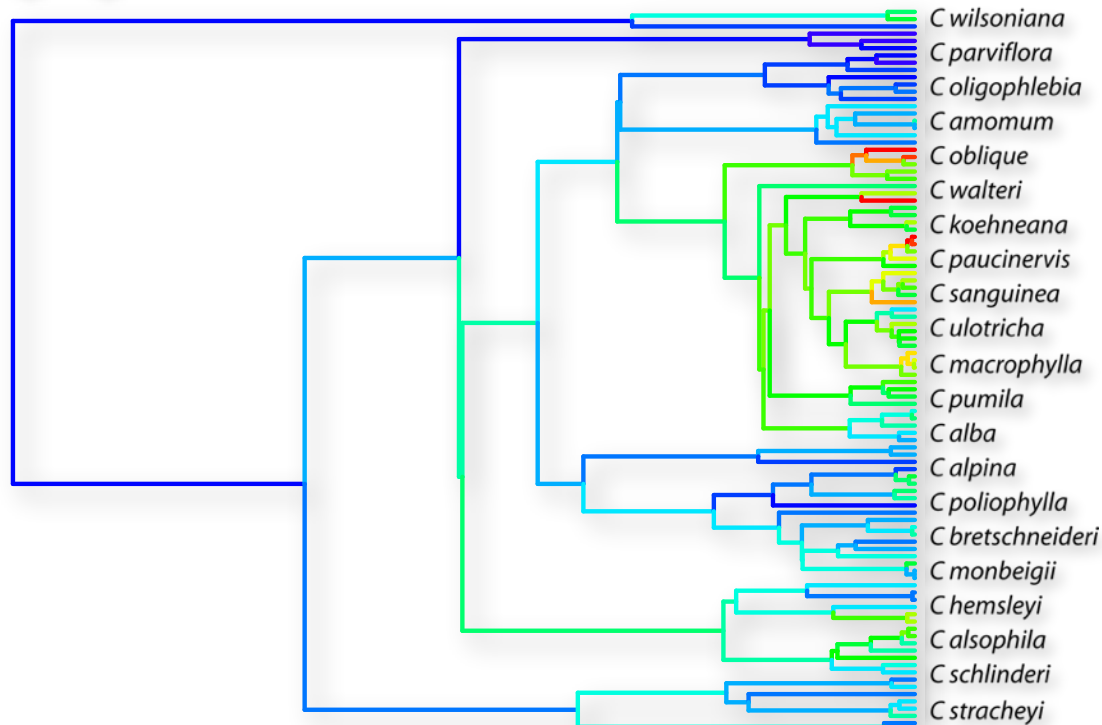
1. Estimating parameters of the stochastic branching-process model
2. Identifying significant diversification rate shifts through time
3. Locating significant diversification rate shifts along branches
4. Evaluating correlates of differential diversification rates

Are diversification rates correlated with other evolutionary variables?



# The Five Fundamental Questions in the Study of Lineage Diversification Rates

1. Estimating parameters of the stochastic branching-process model
2. Identifying significant diversification rate shifts through time
3. Locating significant diversification rate shifts along branches
4. Evaluating correlates of differential diversification rates
5. Detecting significant diversification rate variation across the tree



# Statistical Phylogenetic Approaches for the Study of Diversification Rates

I. Pose a substantive question

 II. Develop a stochastic model with parameters that, if known, would answer the question

III. Collect observations that are informative about model parameters

IV. Estimate the model parameters using some method conditioned on the data at hand

# Stochastic-Branching Process Models

*“All models are wrong, but some are useful”* Box (1976)

## Common Stochastic-Branching Process (SBP) Models

Generalized Birth-Death (GBD: Kendall, 1948)

- $\lambda(t)$  and  $\mu(t)$ ,  $\lambda > \mu$

Constant Birth-Death (CBD: Kendall, 1948)

- $\lambda$  and  $\mu$ ,  $\lambda > \mu$

Sampled Birth-Death (SBD: Gernhardt, 2008)

- $\lambda, \mu, \rho$ ,  $\lambda > \mu$

Generalized Pure Birth (GPB: Harris, 1964)

- $\lambda(t)$ ,  $\mu = 0$

Constant Pure Birth (CPB: Yule, 1924)

- $\lambda$ ,  $\mu = 0$

# Stochastic-Branching Process Models

*“All models are wrong, but some are useful”* Box (1976)

## General properties of SBP Markov models

Assume  $\lambda(t)$  and/or  $\mu(t)$  are:

- constant across all lineages at and instant
- independent across lineages at and instant
- events occur instantaneously

Assume  $\rho$  is:

- uniform/random with respect to phylogeny

# Statistical Phylogenetic Approaches for the Study of Diversification Rates

I. Pose a substantive question

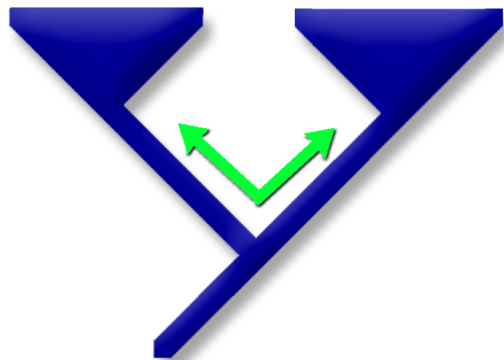
II. Develop a stochastic model with parameters that, if known, would answer the question

 III. Collect observations that are informative about model parameters

IV. Estimate the model parameters using some method conditioned on the data at hand

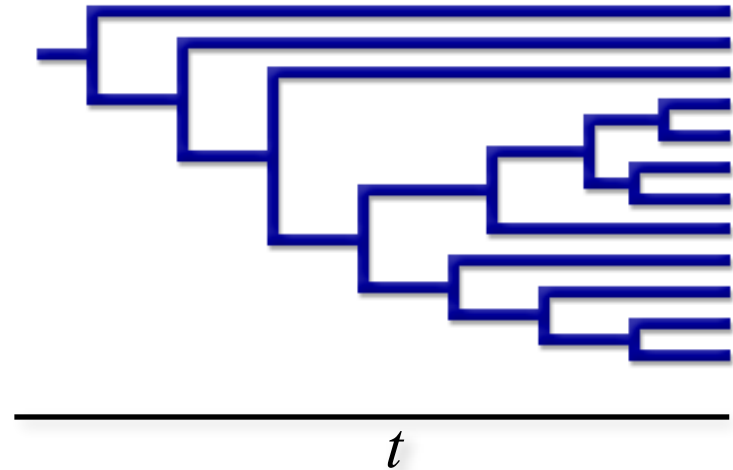
# What Are the Basic Phylogenetic Data for Studying Diversification Rates?

## Topological Information



Distribution of species diversity across the tree

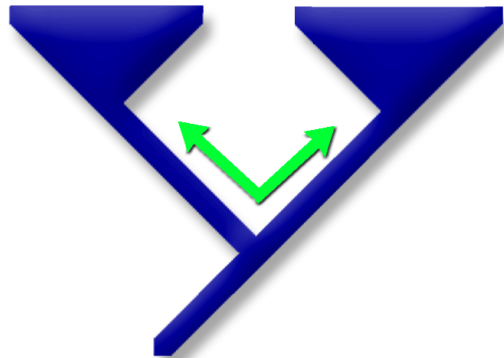
## Temporal Information



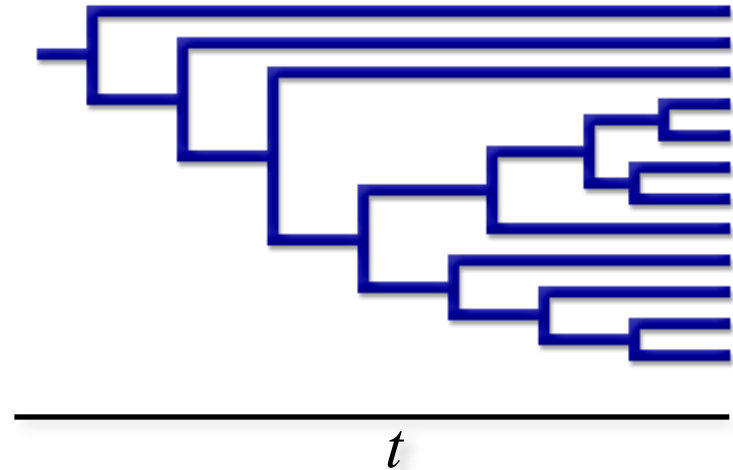
Distribution of speciation events through time

# What Are the Basic Phylogenetic Data for Studying Diversification Rates?

## Topological Information



## Temporal Information



The primary phylogenetic ‘observations’—both the topological distribution of species diversity and the temporal distribution of waiting times—arise via stochastic processes

These ‘observations’ are estimates from data (and therefore associated with uncertainty)

# Statistical Phylogenetic Approaches for the Study of Diversification Rates

I. Pose a substantive question

II. Develop a stochastic model with parameters that, if known, would answer the question

III. Collect observations that are informative about model parameters

 IV. Estimate the model parameters using some method conditioned on the data at hand

# Statistical Inference Under Stochastic-Branching Process Models

SBP models are employed in two qualitatively different ways

- Null-modeling approaches calculate a summary statistic from the phylogenetic ‘observations’, which is then compared to a statistical distribution generated under an appropriate stochastic-branching process model
- Model-fitting approaches estimate parameters of the SBP models from the phylogenetic ‘observations’; i.e., the topological distribution of species diversity and the temporal distribution of waiting times

# Statistical Inference Under Stochastic-Branching Process Models

Inference under SBP models may adopt different statistical approaches

- Maximum Likelihood Estimation methods involve numerical optimization algorithms to identify the joint parameter estimates that collectively maximize the likelihood of the phylogenetic ‘observations’ under the SBP model
- Quasi-Bayesian Inference methods involve averaging ML estimates over a marginal posterior probability density of the phylogenetic ‘observations’
- Empirical Bayesian Inference methods involve numerical algorithms to approximate the joint posterior probability density of the SBP model parameters given point (maximum-likelihood) estimates of the phylogenetic ‘observations’
- Sequential-Bayesian Inference methods involve Bayesian inference of the SBP model parameters averaged over a previously estimated marginal posterior probability density of the phylogenetic ‘observations’
- Hierarchical-Bayesian Inference methods involve joint (simultaneous) estimation of the SBP and other phylogenetic model parameters

# Why Pursue These Questions in a Bayesian Statistical Framework?

Recent developments have largely been developed in Quasi-Bayesian ML framework

The study of diversification entails several sources of uncertainty

- phylogeny/topology
- branch lengths/durations
- rate parameters
- event histories

Bayesian framework provides a means for accommodating uncertainty

# Outline & Organization

## I. Brief overview of general approach to statistical study of lineage diversification

- I. Pose a substantive question
- II. Develop a stochastic model with parameters that can answer the question
- III. Collect observations that are informative about model parameters
- IV. Estimate the model parameters using some method conditioned on the data

## II. Brief overview of popular methods for studying lineage diversification

A beginners guide to the types of methods available

# Outline & Organization

I. Brief overview of general approach to statistical study of lineage diversification

 II. Brief overview of popular methods for studying lineage diversification  
A beginners guide to the types of methods available

# Five Fundamental Questions in the Study of Diversification Rates

- I. Estimating parameters of the branching process
- II. Identifying significant diversification rate shifts through time
- III. Locating significant diversification rate shifts along branches
- IV. Evaluating correlates of differential diversification rates
- V. Detecting significant diversification rate variation across the tree

# Estimating Diversification-Rate Parameters

## Lineage-Thru-Time Plots

(e.g., Nee et al., 1992, 1994a,b)

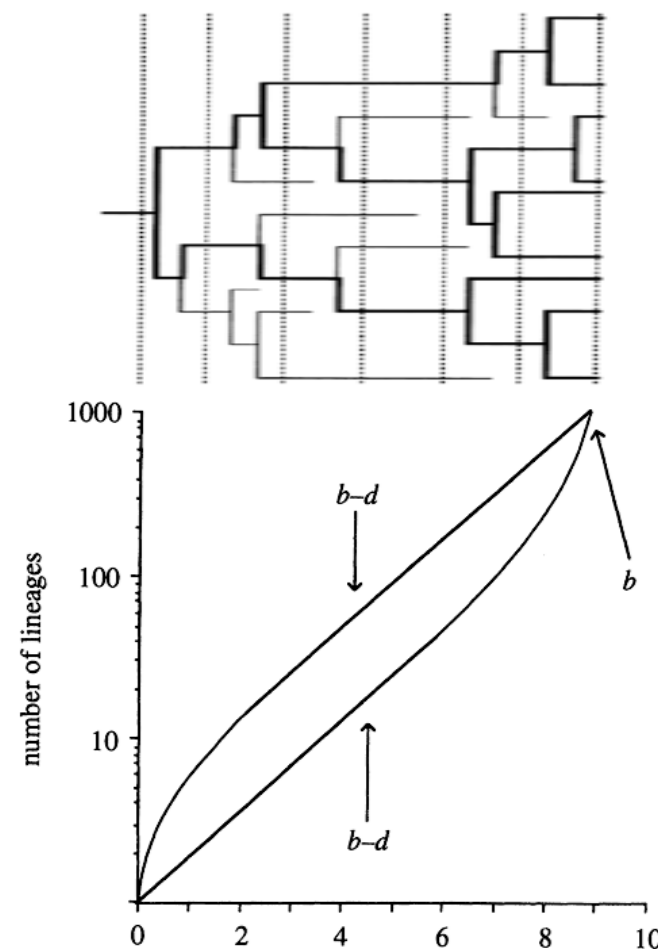
Semi-logarithmic accumulation of lineages

Primarily descriptive, characterized by:

- Push of the Past
- slope = diversification rate,  $\partial = (\lambda - \mu)$

If extinction rate,  $\mu = 0$ , also characterized by:

- Pull of the Present
- increased offset slope
- asymptotic behavior under random sub sampling



# Estimating Diversification-Rate Parameters

## ML Estimation from Lineage-Thru-Time Plots

(e.g., Nee , 2001, Magallon & Sanderson, 2001)

Maximum likelihood is used to estimate composite rate parameters:

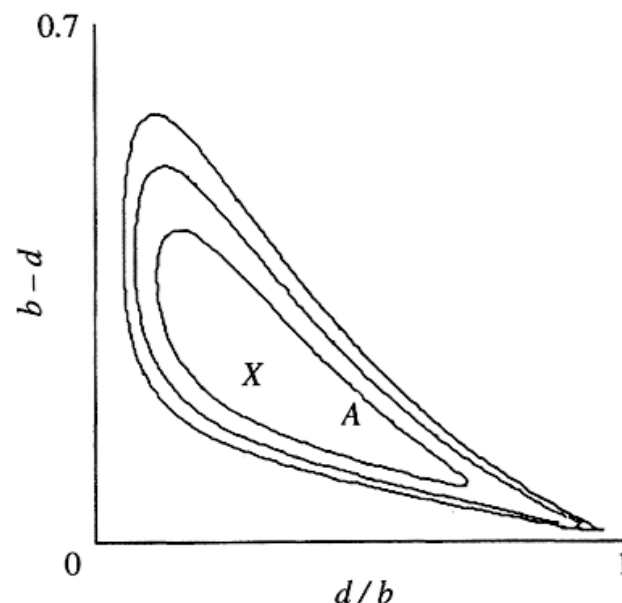
- diversification rate,  $r = (\lambda - \mu)$
- relative extinction rate,  $\varepsilon = (\mu / \lambda)$

It is not possible to estimate individual rate parameters (e.g., Kubo & Iwasa, 1995; Paradis, 2004)

Comparing absolute diversification rates across clades of different ages is largely invalid if  $\mu > 0$  (e.g., Kubo & Iwasa)

Reliable (valid) parameter estimation **REQUIRES**

- accommodating variance in divergence-time estimates
- no bias in divergence time estimates
- large phylogenies
- complete or phylogenetically unbiased species sampling
- demonstration that diversification has been constant
  - across lineages
  - through time



# Estimating Diversification-Rate Parameters

## Practical Demonstration 1

# Estimating Diversification-Rate Parameters

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.*

# Estimating Diversification-Rate Parameters

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.*

*Evolution*, 49(4), 1995, pp. 694–704

INFERRING THE RATES OF BRANCHING AND EXTINCTION FROM  
MOLECULAR PHYLOGENIES

TAKUYA KUBO AND YOH IWASA

*Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812-81, Japan*

*E-mail: tkuboscb@mbox.nc.kyushu-u.ac.jp; yiwasscb@mbox.nc.kyushu-u.ac.jp*

# Estimating Diversification-Rate Parameters

## 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.*

*Evolution*, 49(4), 1995, pp. 694–704

### INFERRING THE RATES OF BRANCHING AND EXTINCTION FROM MOLECULAR PHYLOGENIES

TAKI

*Department of Biology, Faculty  
E-mail: tkuboscb@mbox.nc*

Figure 3 clearly illustrates that, in general, different pairs of branching rates and extinction rates can give exactly the same pattern of ancestral phylogeny, expressed in terms of the number of ancestral lineages over various times. We have shown the equivalence relationship between the case with a time-dependent branching rate and a constant extinction rate and the second case with a constant branching rate and a time-dependent extinction rate. Generating the same ancestral phylogeny does not mean that these two processes are equivalent in evolution, because temporal change in the actual number of taxa, denoted by  $S(t)$ , may greatly differ between these processes. This shows a clear limitation of the method of inferring past events from detailed knowledge, such as of molecular phylogenies, of the extant species only.

# Five Fundamental Questions in the Study of Diversification Rates

- I. Estimating parameters of the branching process
- II. Identifying significant diversification rate shifts through time
- III. Locating significant diversification rate shifts along branches
- IV. Evaluating correlates of differential diversification rates
- V. Detecting significant diversification rate variation across the tree

# Locating Shifts in Diversification Rate Through Time

## Gamma Statistic (APE)

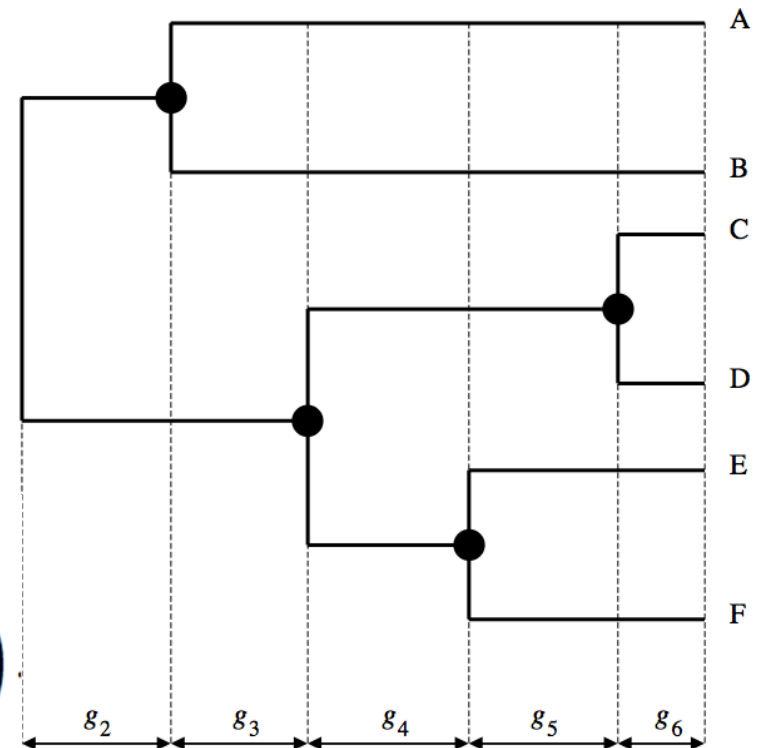
Under a Yule process, waiting times between speciation events are i.i.d. exponentially distributed random variables

If the process is constant, the sum of the waiting times will be a gamma distributed random variable

- the shape of the gamma is the speciation rate
- the scale of the gamma is  $(2N-2)$

The gamma statistic exploits the Yule process, where the waiting times are maximally concentrated near the root

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



Any tree with a greater concentration of waiting times near the root must be rate variable (with decreasing rate)

# Locating Shifts in Diversification Rate Through Time

## Gamma Statistic (APE)

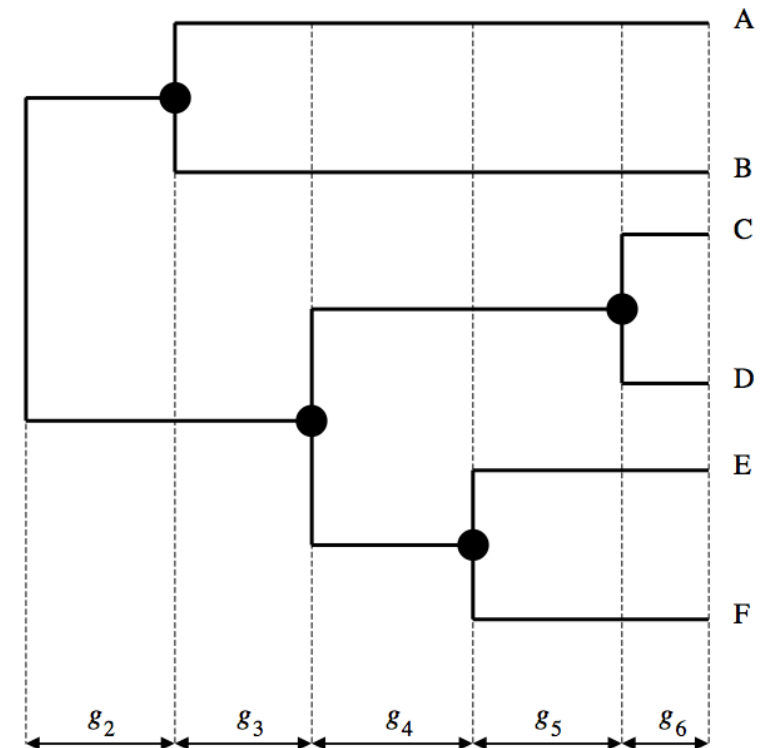
Incomplete species sampling can also create the illusion of temporally decreasing diversification rate

Accordingly, Monte Carlo simulation is required to generate a null distribution of the  $\gamma$ -statistic for incomplete trees

- assumes random species sampling
- departures from random sampling inflate FPR

Can detect diversification rate decreases, but not increases

Does not estimate parameters of the process:  
 $\lambda$ ,  $\mu$ , or the location of the temporal rate shift



# Locating Shifts in Diversification Rate Through Time

## Birth-Death Likelihood Method (LASER)

### Concept

Fits various stochastic-branching process models to (parts of) the tree  
Selects among diversification models using AIC

### Data

Vector of waiting times,  $\mathbf{x}$

### Models

constant Yule

- $(\lambda, \mu = 0)$

constant birth-death

- $(\lambda, \mu = 0, \lambda > \mu)$

Yule-shift

- $(\lambda_1, \lambda_2, \mu = 0, t_s)$

variable birth-death

- $(\lambda_1, \lambda_2 > 0; \mu_1, \mu_2 \geq 0; \epsilon_1 = \epsilon_2)$

# Locating Shifts in Diversification Rate Through Time

## Birth-Death Likelihood Method (LASER)

Likelihood function

$$\mathcal{L}(\mathbf{x}|\epsilon, r) = \prod_{n=2}^{N-1} nr \exp[-nr(x_n - x_{n+1})] \\ \times \frac{\{1 - \epsilon \exp[-r(x_{n+1})]\}^{n-1}}{\{1 - \epsilon \exp[-r(x_n)]\}^n}$$

Where  $r = \lambda - \mu$ ;  $\epsilon = \lambda/\mu$

# Locating Shifts in Diversification Rate Through Time

## Birth-Death Likelihood Method (LASER)

### Model selection

Diversification models are compared using the AIC model selection criterion:

$$AIC = 2p - 2 \log \mathcal{L}$$

↑                      ↑  
number of parameters      maximum likelihood

The fit of the data to the best rate variable and rate constant models is calculated:

$$\Delta AIC_{RC} = AIC_{RC} - AIC_{RV}$$

The significance of the  $\Delta AIC$  test statistic is assessed by Monte Carlo simulation under the null (Yule) stochastic-branching process model

# Locating Shifts in Diversification Rate Through Time

## Practical Demonstration 2

# Locating Shifts in Diversification Rate Through Time

## Methodological Concerns

Effects of branch-length estimation biases

- substitution-rate profile may lead to systematic estimation bias of deeper/shallower nodes (e.g., Revell et al., 2005)

Departure from random species sampling

- more realistic taxon sampling may produce spurious decrease in rate (e.g., Cusimano & Renner, 2010)

Effects of among-lineage rate variation

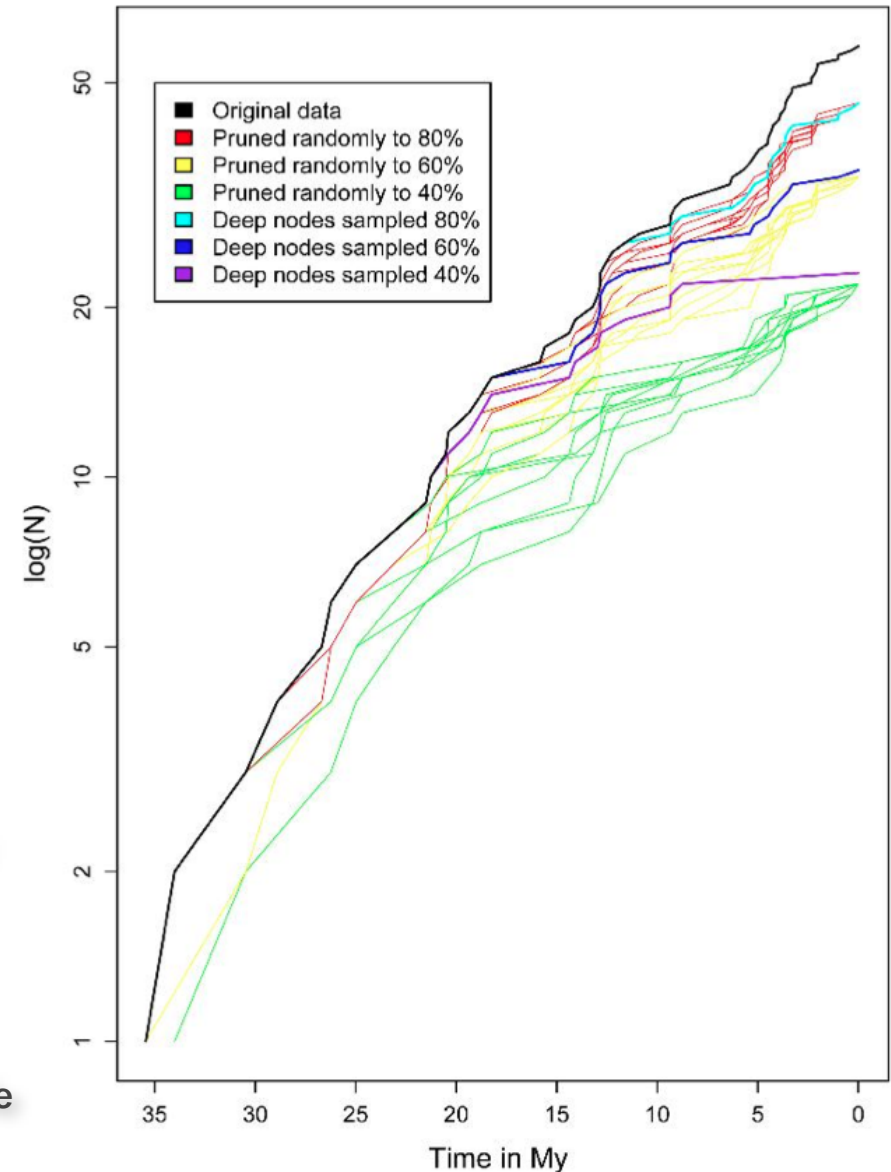
- violation of stochastically constant rates across lineages may produce spurious rate decreases

Effects of model-selection bias

- reliance on AIC may lead to inflated rejection of simpler rate constant models

Effects of phylogenetic uncertainty

- reliance on point estimates of phylogeny and divergence times may render inferences unreliable



# Locating Shifts in Diversification Rate Through Time

## (Unsolicited) Methodological Advice



**TreePar** (Stadler, 2011)

Maximum-likelihood estimation of temporal shifts in diversification rate



**TESS** (Hohna, 2013)

Bayesian inference of temporal shifts in diversification rate

# Five Fundamental Questions in the Study of Diversification Rates

- I. Estimating parameters of the branching process
- II. Identifying significant diversification rate shifts through time
- III. Locating significant diversification rate shifts along branches
- IV. Evaluating correlates of differential diversification rates
- V. Detecting significant diversification rate variation across the tree

# Locating Shifts in Diversification Rate Along Branches

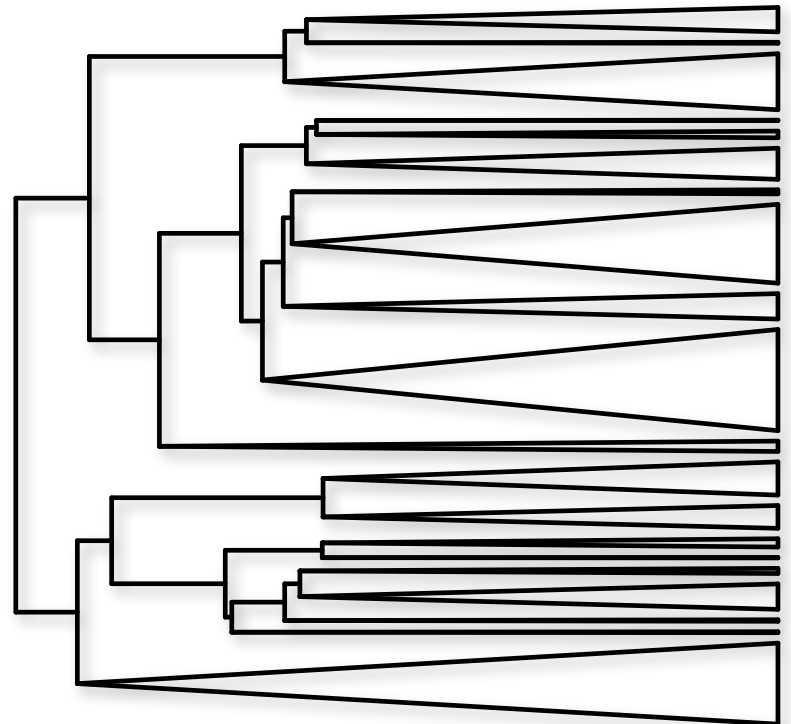
## MEDUSA: Sequential AIC Model Fitting

### Concept

Fits increasingly complex birth-death models to (parts of) the tree  
Selects among diversification models using stepwise AIC

### Data

A dated phylogenetic tree  $T$ , with  $s$  terminal lineages  
Terminal lineage  $i$  represents  $n_i$  species.



# Locating Shifts in Diversification Rate Along Branches

## MEDUSA: Sequential AIC Model Fitting

### Likelihood function

Under constant diversification rates, the probability of observing a branch of length  $\tau$  that began at time  $t$ , conditional on the survival of its descendants is:

$$P(\tau \mid t, \lambda, \mu) = \frac{r \exp(-r\tau)}{1 - \epsilon \exp(-rt)}$$

and the probability of observing a subclade of stem age  $t$  with  $n$  descendants (conditional on  $n > 0$ ) is:

$$P(n \mid t, \lambda, \mu) = (1 - u) \times (u)^{n-1},$$
$$u = \frac{\exp(rt) - 1}{\exp(rt) - \epsilon}$$

# Locating Shifts in Diversification Rate Along Branches

## MEDUSA: Sequential AIC Model Fitting

### Likelihood function

To calculate the likelihood of the whole tree, we calculate the likelihood for each branch and each terminal unresolved subclade combine them as their product:

$$P(T, \mathbf{n} \mid \lambda, \mu) = \left( \prod_{\text{branches}} P(\tau_i \mid t_i, \lambda, \mu) \right) \times \left( \prod_{\text{triangles}} P(n_i \mid t_i, \lambda, \mu) \right)$$

# Locating Shifts in Diversification Rate Along Branches

## MEDUSA: Sequential AIC Model Fitting

### Likelihood function

Purportedly, this is valid even when diversification rates vary over branches and/or terminal unresolved subclades:

$$P(T, \mathbf{n} \mid \boldsymbol{\lambda}, \boldsymbol{\mu}) = \left( \prod_{\text{branches}} P(\tau_i \mid t_i, \lambda_i, \mu_i) \right) \times \left( \prod_{\text{triangles}} P(n_i \mid t_i, \lambda_i, \mu_i) \right)$$

# Locating Shifts in Diversification Rate Along Branches

## MEDUSA: Sequential AIC Model Fitting

### Model selection

Diversification models are compared using the AIC model selection criterion:

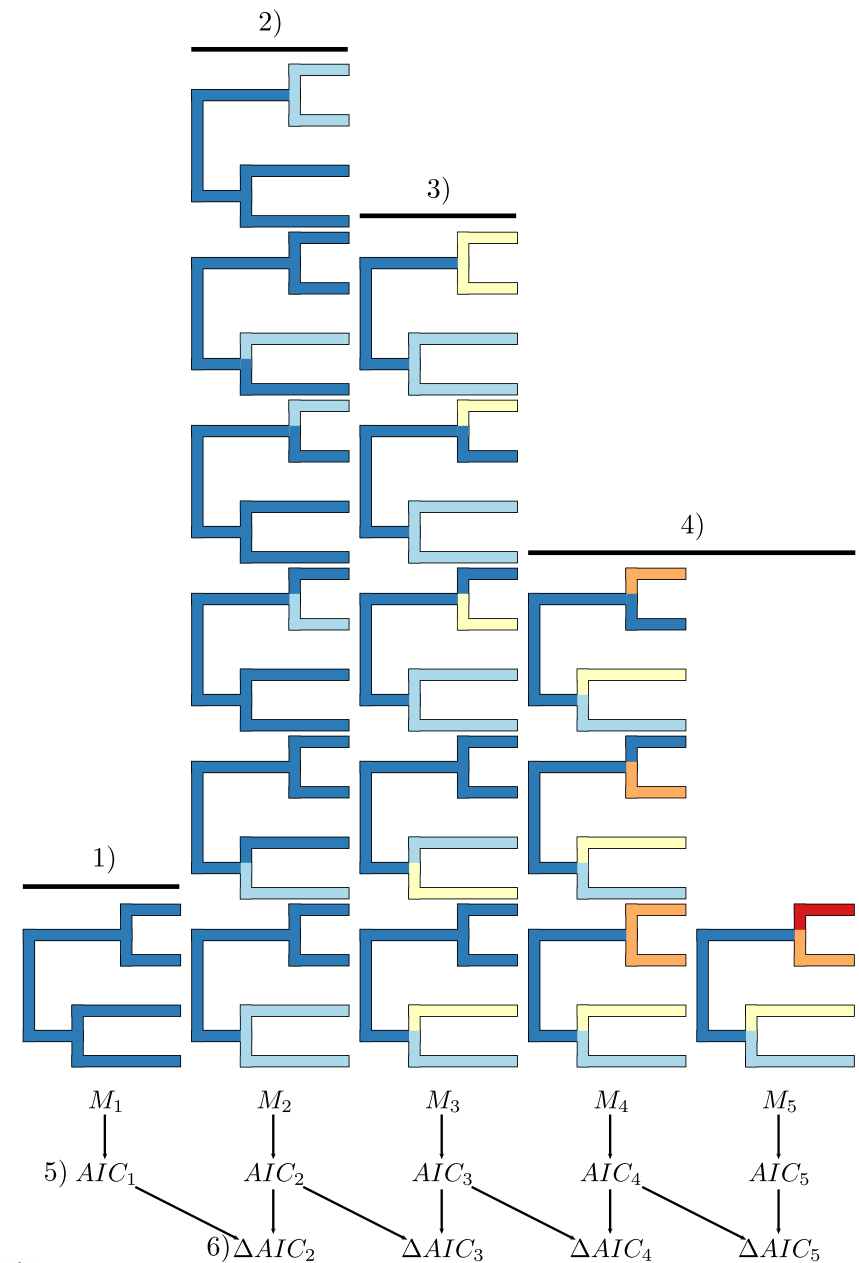
$$AIC = 2p - 2 \log \mathcal{L}$$

↑                      ↑  
number of parameters    maximum likelihood

The critical  $\Delta AIC$  for choosing among models is arbitrary

# Locating Shifts in Diversification Rate Along Branches

MEDUSA: Sequential AIC Model Fitting  
Algorithm



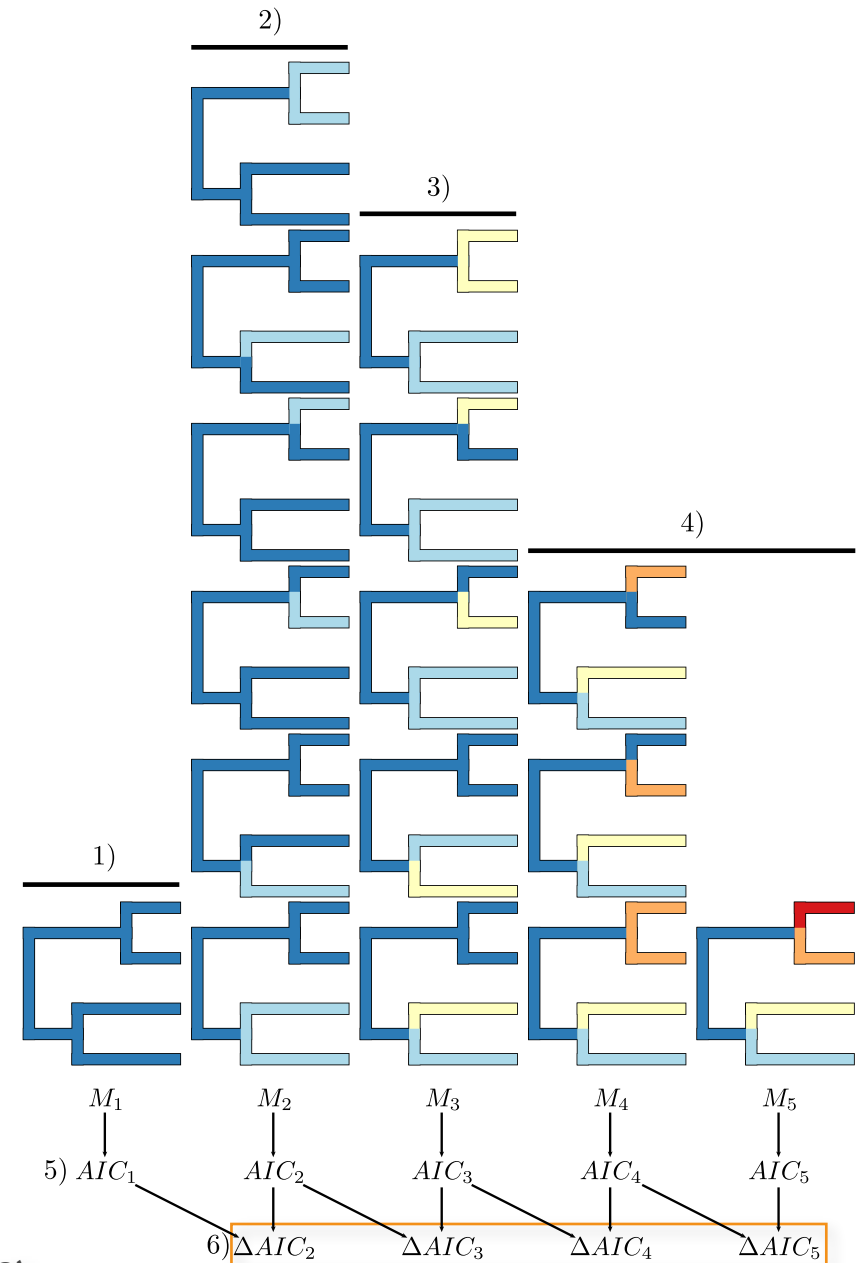
(Alfaro et al., 2009)

# Locating Shifts in Diversification Rate Along Branches

## MEDUSA: Sequential AIC Model Fitting

### Algorithm

- 1) Fit a one-rate model to the data.
- 2) Fit “every” two-rate model to the data.
- 3) Fit “every” three-rate model that contains the best two-rate model.
- 4) Keep fitting increasingly complex models.
- 5) Compute the AIC score for the best model in each level of model complexity.
- 6) Starting with the one-rate model, accept the next-most-complex model if the improvement in model fit is “good enough.”



(Alfaro et al., 2009)

# Locating Shifts in Diversification Rate Along Branches

## Practical Demonstration 3

# Five Fundamental Questions in the Study of Diversification Rates

- I. Estimating parameters of the branching process
- II. Identifying significant diversification rate shifts through time
- III. Locating significant diversification rate shifts along branches
- IV. Evaluating correlates of differential diversification rates
- V. Detecting significant diversification rate variation across the tree

# Evaluating Correlates of Differential Diversification Rates

Two fundamentally different types of questions:

Is there a general correlation between trait state and diversification rate?

e.g., Are nectar spurs correlated with increase rates of diversification?

Is there a correlation between a specific event and and diversification rate?

e.g., Was a particular origin of nectar spurs in columbines correlated with increase rates of diversification?

# Evaluating Correlates of Differential Diversification Rates

## Recent advances for evaluating diversification-rate correlates

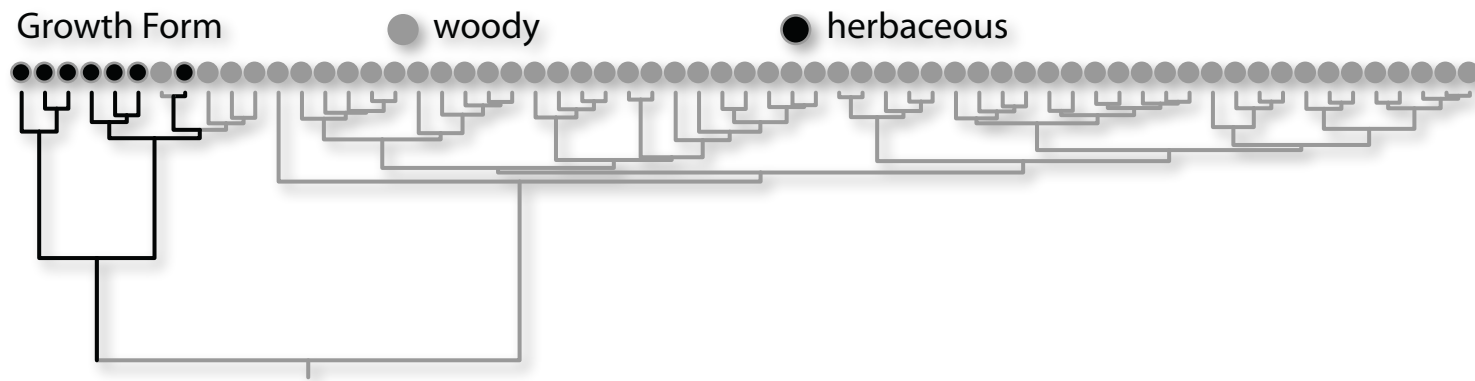
\*SSE model; DiversiTree (Maddison et al., 2007; Fitzjohn et al., 2009; Fitzjohn, 2010...)

CVPPD; tRate (Moore & Donoghue, 2009)

# Evaluating Correlates of Differential Diversification Rates

## Binary-State Speciation Extinction (BiSSE) Model

Motivation: Imagine that we see a preponderance of species with a given state



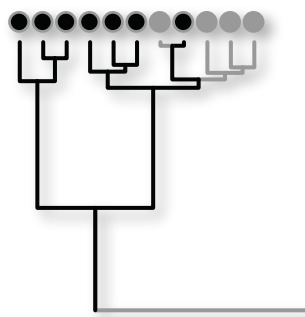
Explanations include:

- increased woody speciation rate  $\lambda_{\bullet}$
- decreased woody extinction rate  $\mu_{\bullet}$
- decreased herb speciation rate  $\lambda_{\circ}$
- increased herb extinction rate  $\mu_{\circ}$
- biased exchangeability rate  $q_{\bullet\circ}$

# Evaluating Correlates of Differential Diversification Rates

## Binary-State Speciation Extinction (BiSSE) Model

Solution: Specify a model that includes the parameters to tease these scenarios apart



$\lambda_0$  speciation rate for state 0

$\lambda_1$  speciation rate for state 1

$\mu_0$  extinction rate for state 0

$\mu_1$  extinction rate for state 1

$q_{01}$  rate of change to state 1

$q_{10}$  rate of change to state 0

Solve numerically with PDEs:

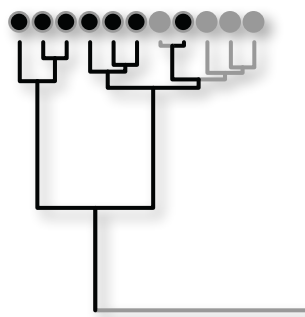
$$D_{N0}(t + \delta t) = (1 - \mu_0 \delta t) [(1 - q_{01} \delta t)(1 - \lambda_0 \delta t) D_{N0}(t) + (q_{01} \delta t)(1 - \lambda_0 \delta t) D_{N1}(t) + (1 - q_{01} \delta t)(\lambda_0 \delta t) E_0(t) D_{N0}(t) + (1 - q_{01} \delta t)(\lambda_0 \delta t) E_0(t) D_{N0}(t)] + (\mu_0 \delta t) 0$$

$$D_{N1}(t + \delta t) = (1 - \mu_1 \delta t) [(1 - q_{10} \delta t)(1 - \lambda_1 \delta t) D_{N1}(t) + (q_{10} \delta t)(1 - \lambda_1 \delta t) D_{N0}(t) + (1 - q_{10} \delta t)(\lambda_1 \delta t) E_1(t) D_{N1}(t) + (1 - q_{10} \delta t)(\lambda_1 \delta t) E_1(t) D_{N1}(t)] + (\mu_1 \delta t) 0$$

# Evaluating Correlates of Differential Diversification Rates

## Binary-State Speciation Extinction (BiSSE) Model

Solution: Specify a model that includes the parameters to tease these scenarios apart



$\lambda_0$  speciation rate for state 0

$\lambda_1$  speciation rate for state 1

$\mu_0$  extinction rate for state 0

$\mu_1$  extinction rate for state 1

$q_{01}$  rate of change to state 1

$q_{10}$  rate of change to state 0

Solve numerically with PDEs:

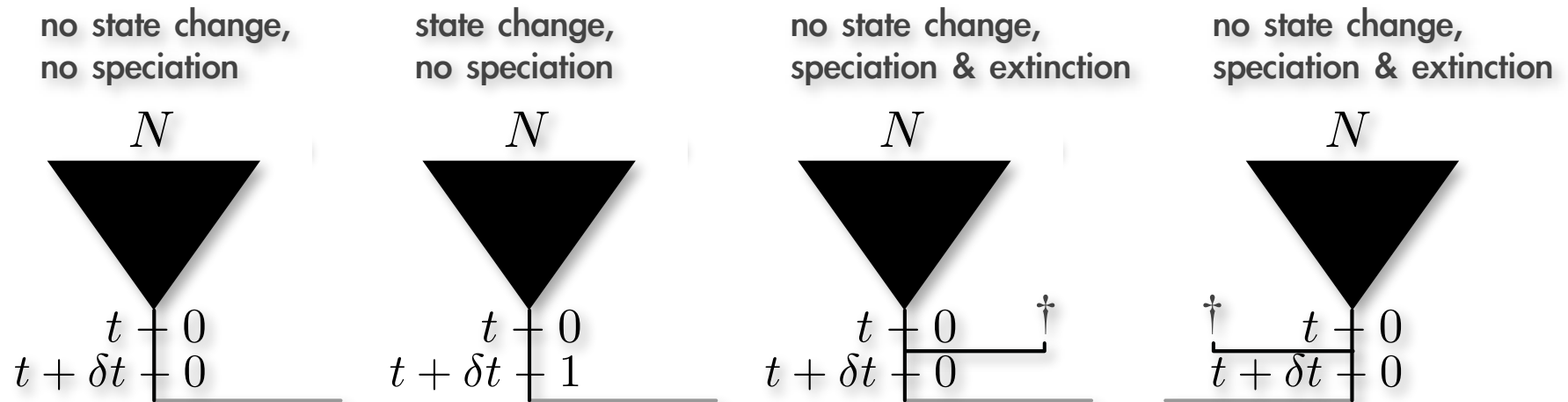
$$E_0(t + \delta t) = \mu_0 \delta t + (1 - \mu_0 \delta t)(1 - q_{01} \delta t)(1 - \lambda_0 \delta t)E_0(t) + (1 - \mu_0 \delta t)(q_{01} \delta t)(1 - \lambda_0 \delta t)E_1(t) + (1 - \mu_0 \delta t)(1 - q_{01} \delta t)(\lambda_0 \delta t)E_0(t)^2$$

$$E_1(t + \delta t) = \mu_1 \delta t + (1 - \mu_1 \delta t)(1 - q_{10} \delta t)(1 - \lambda_1 \delta t)E_1(t) + (1 - \mu_1 \delta t)(q_{10} \delta t)(1 - \lambda_1 \delta t)E_0(t) + (1 - \mu_1 \delta t)(1 - q_{10} \delta t)(\lambda_1 \delta t)E_1(t)^2$$

Maddison et al. (*Syst Biol*, 2007)

# Evaluating Correlates of Differential Diversification Rates

## Binary-State Speciation Extinction (BiSSE) Model



PDEs for branch probabilities:

$$D_{N0}(t + \delta t) = (1 - \mu_0 \delta t) [(1 - q_{01} \delta t)(1 - \lambda_0 \delta t) D_{N0}(t) + (q_{01} \delta t)(1 - \lambda_0 \delta t) D_{N1}(t) + (1 - q_{01} \delta t)(\lambda_0 \delta t) E_0(t) D_{N0}(t) + (1 - q_{01} \delta t)(\lambda_0 \delta t) E_0(t) D_{N0}(t)] + (\mu_0 \delta t) 0$$

no extinction in the interval

no state change, no speciation

state change, no speciation

no state change, speciation & extinction

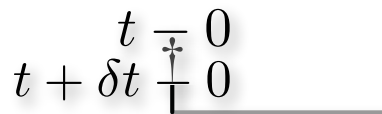
no state change, speciation & extinction

if extinct, zero probability of being observed

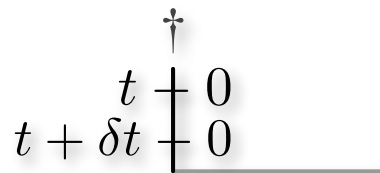
# Evaluating Correlates of Differential Diversification Rates

## Binary-State Speciation Extinction (BiSSE) Model

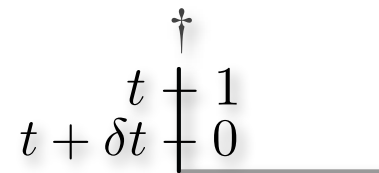
extinction in  
interval



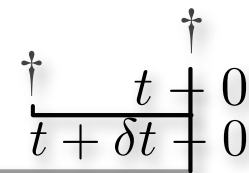
no state change,  
no speciation,  
extinction since  $t$



state change,  
no speciation,  
extinction since  $t$



no state change,  
speciation, extinction  
of both since  $t$



PDEs for branch probabilities:

$$E_0(t + \delta t) = \mu_0 \delta t$$

$$+ (1 - \mu_0 \delta t)(1 - q_{01} \delta t)(1 - \lambda_0 \delta t) E_0(t)$$

$$+ (1 - \mu_0 \delta t)(q_{01} \delta t)(1 - \lambda_0 \delta t) E_1(t)$$

$$+ (1 - \mu_0 \delta t)(1 - q_{01} \delta t)(\lambda_0 \delta t) E_0(t)^2$$

extinction in the interval

no state change, no speciation

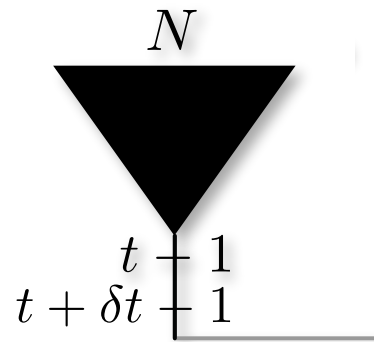
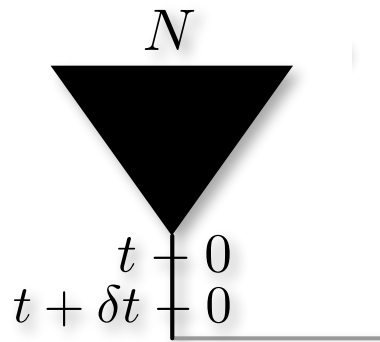
state change, no speciation

no state change, speciation

# Evaluating Correlates of Differential Diversification Rates

## Binary-State Speciation Extinction (BiSSE) Model

We start at the tips of the tree and move root-ward in small increments...



If  $N$  is a tip species with state 1:

$$D_{N0}(t_0) = 0$$

$$D_{N1}(t_0) = 1$$

If  $N$  is a tip species with state 0:

$$D_{N0}(t_0) = 1$$

$$D_{N1}(t_0) = 0$$

Maddison et al. (*Syst Biol*, 2007)

# Evaluating Correlates of Differential Diversification Rates

## Binary-State Speciation Extinction (BiSSE) Model

solving coupled differential equations to obtain conditional branch probabilities...



We take the derivative of the PDEs to shrink  $\delta t$ :

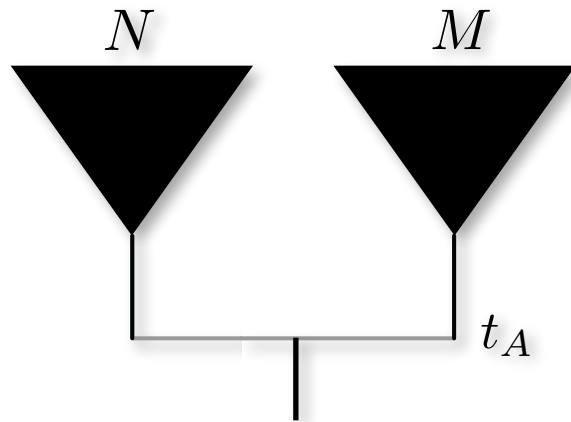
$$\frac{dD_{N0}}{dt} = -(\lambda_0 + \mu_0 + q_{01})D_{N0}(t) + q_{01}D_{N1}(t) + 2\lambda_0 E_0(t)D_{N0}(t)$$

$$\frac{dD_{N1}}{dt} = -(\lambda_1 + \mu_1 + q_{10})D_{N1}(t) + q_{10}D_{N0}(t) + 2\lambda_1 E_1(t)D_{N1}(t)$$

# Evaluating Correlates of Differential Diversification Rates

## Binary-State Speciation Extinction (BiSSE) Model

and then combine branch probabilities at internal nodes...



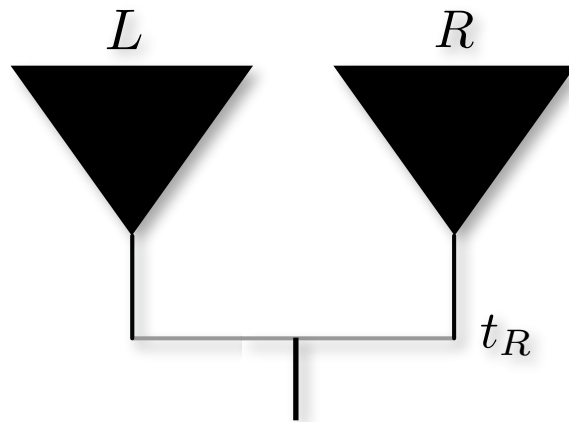
$$D_{A0}(t_A) = D_{N0}(t_A)D_{M0}(t_A)\lambda_0$$

$$D_{A1}(t_A) = D_{N1}(t_A)D_{M1}(t_A)\lambda_1$$

# Evaluating Correlates of Differential Diversification Rates

## Binary-State Speciation Extinction (BiSSE) Model

and at the root scale the conditional probabilities by the stationary frequencies



$$D_0 = D_{R0}(t_R)\pi_0$$

$$D_1 = D_{R1}(t_R)\pi_1$$

# Evaluating Correlates of Differential Diversification Rates

How well can we do inference under the BiSSE model?

Davis et al. *BMC Evolutionary Biology* 2013, **13**:38  
<http://www.biomedcentral.com/1471-2148/13/38>



**RESEARCH ARTICLE**

**Open Access**

## Exploring power and parameter estimation of the BiSSE method for analyzing species diversification

Matthew P Davis<sup>1\*</sup>, Peter E Midford<sup>2</sup> and Wayne Maddison<sup>3</sup>

# Evaluating Correlates of Differential Diversification Rates

How well can we do inference under the BiSSE model?

Set a:

$$\lambda_{\bullet} \neq \lambda_{\bullet}$$

$$\mu_{\bullet} = \mu_{\bullet}$$

$$q_{\bullet\bullet} = q_{\bullet\bullet}$$

Set b:

$$\lambda_{\bullet} = \lambda_{\bullet}$$

$$\mu_{\bullet} = \mu_{\bullet}$$

$$q_{\bullet\bullet} \neq q_{\bullet\bullet}$$

Set c:

$$\lambda_{\bullet} = \lambda_{\bullet}$$

$$\mu_{\bullet} \neq \mu_{\bullet}$$

$$q_{\bullet\bullet} = q_{\bullet\bullet}$$

\*\*\* Assuming no phylogenetic uncertainty

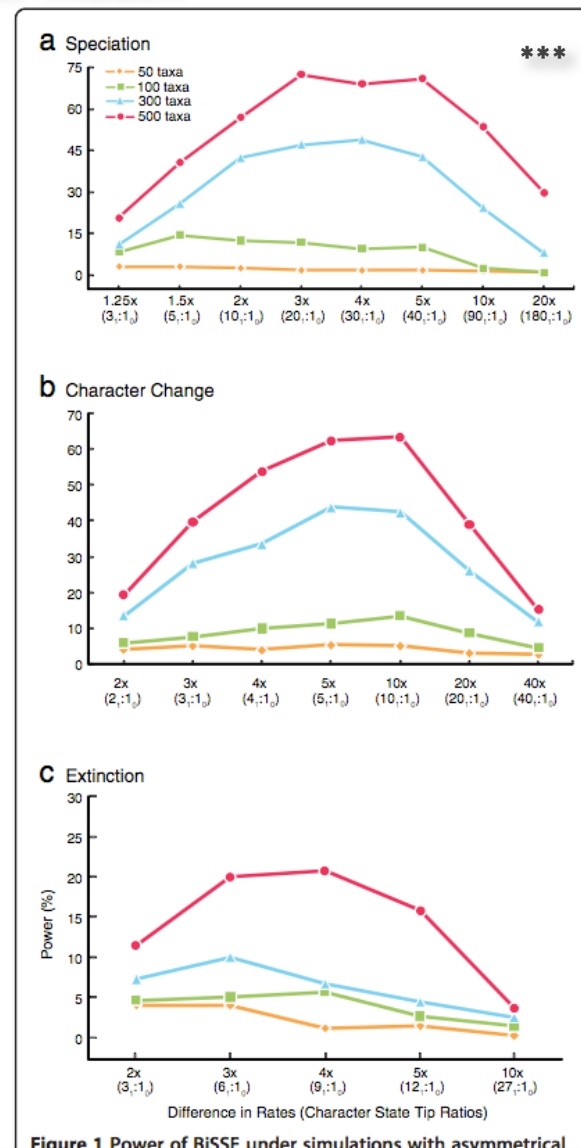


Figure 1 Power of BiSSE under simulations with asymmetrical

# Evaluating Correlates of Differential Diversification Rates

How well can we do inference under the BiSSE model?

## Conclusion

The power of the BiSSE likelihood method to test hypotheses of rate asymmetry is susceptible to both tree size and variation in parameter rates. Overall, power of the BiSSE method is low if the tree size is below 300 taxa, and investigators should take special care to investigate the power of their results if applying the BiSSE method to topologies with fewer than 300 tips. Power is increased when estimating fewer parameters, so utilizing a four parameter model to test hypotheses may be preferable if appropriate.

This study indicates that contrary to the hope expressed in Maddison [20], the problem of confounding effects can still occur while estimating process parameters simultaneously if there is low sample size and/or high tip ratio bias. Under scenar-

a Speciation

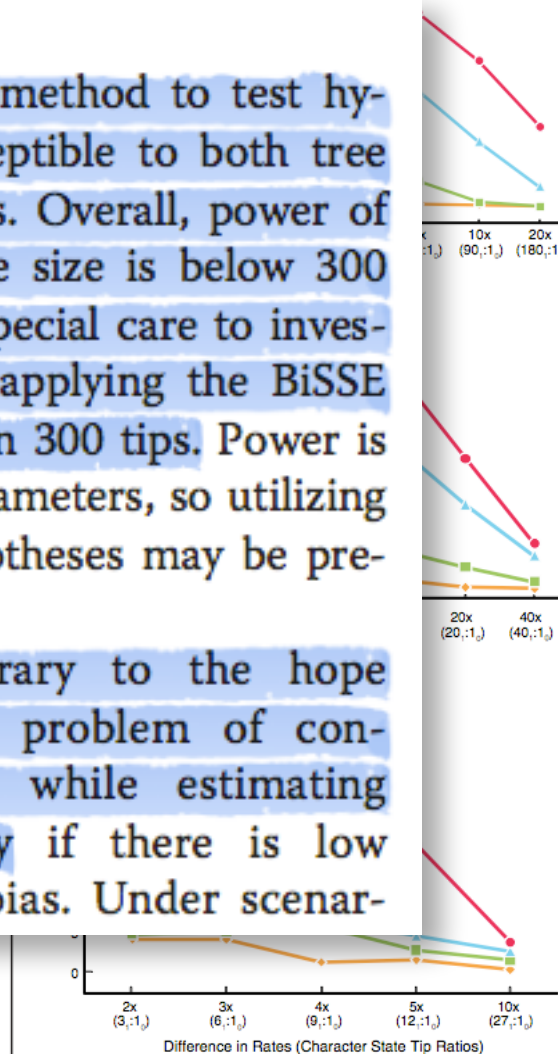


Figure 1 Power of BiSSE under simulations with asymmetrical

# Evaluating Correlates of Differential Diversification Rates

## Practical Demonstration 4

# Evaluating Correlates of Differential Diversification Rates

And now for some complicated models...meet the \*SSE family!

**BiSSE** (Binary-State Speciation Extinction): Maddison et al. (Syst Biol, 2007)

- Two states, 6 parameters

**MuSSE** (Multi-State Speciation Extinction): Fitzjohn et al. (Syst Biol, 2009)

- $(2^N - 2) + 2N$  parameters

**QuaSSE** (Quantitative-State Speciation Extinction): Fitzjohn (Syst Biol, 2010)

- lots and lots of parameters

**GeoSSE** (Geographic-State Speciation Extinction): Goldberg et al. (Syst Biol, 2011)

- even more parameters!

**BiSSEness** (BiSSE-Node Enhanced State Shift): Magnuson-Ford & Otto (Am Nat, 2012)

- two states, 10 parameters

# Evaluating Correlates of Differential Diversification Rates

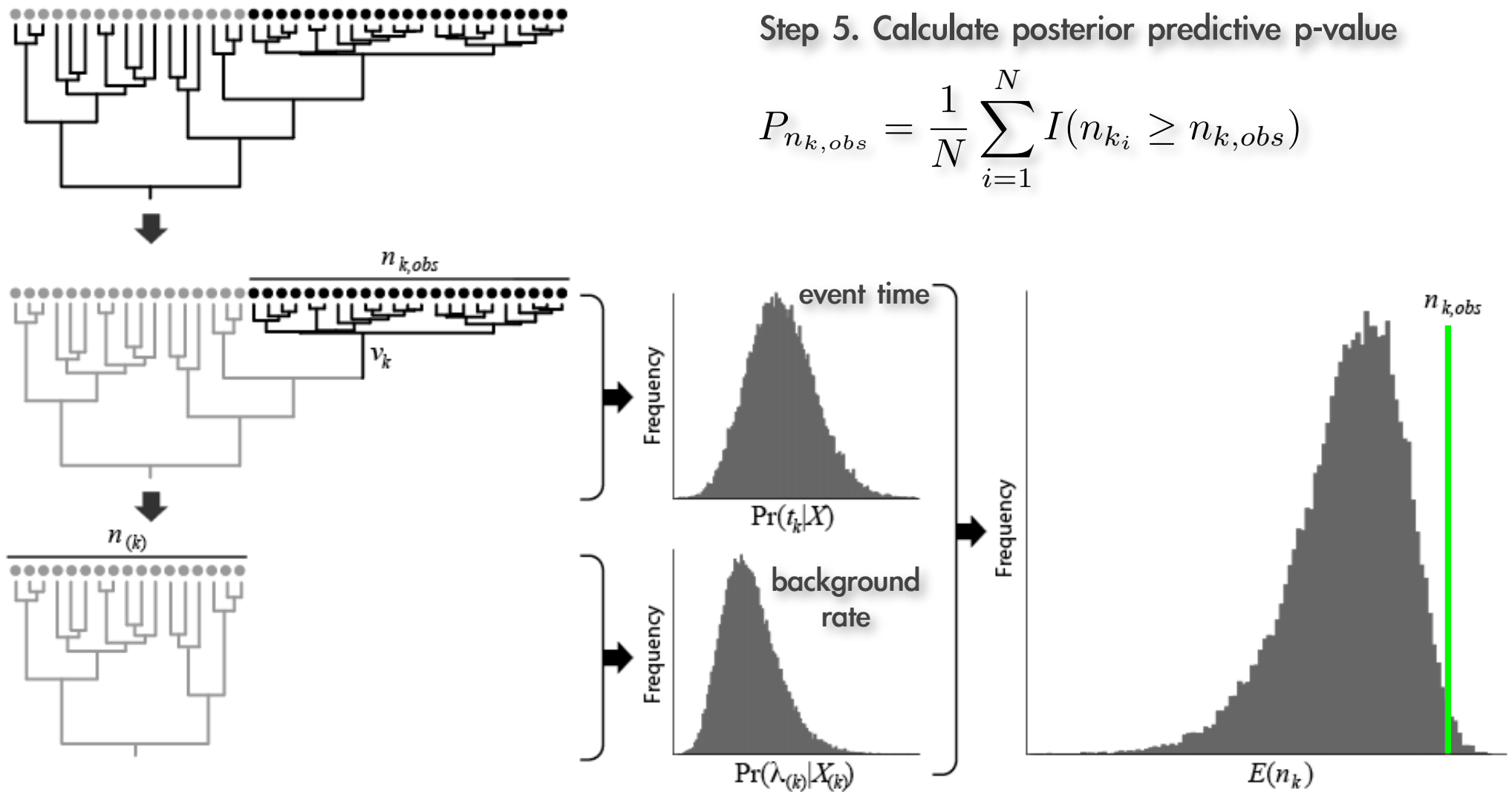
## Recent advances for evaluating diversification-rate correlates

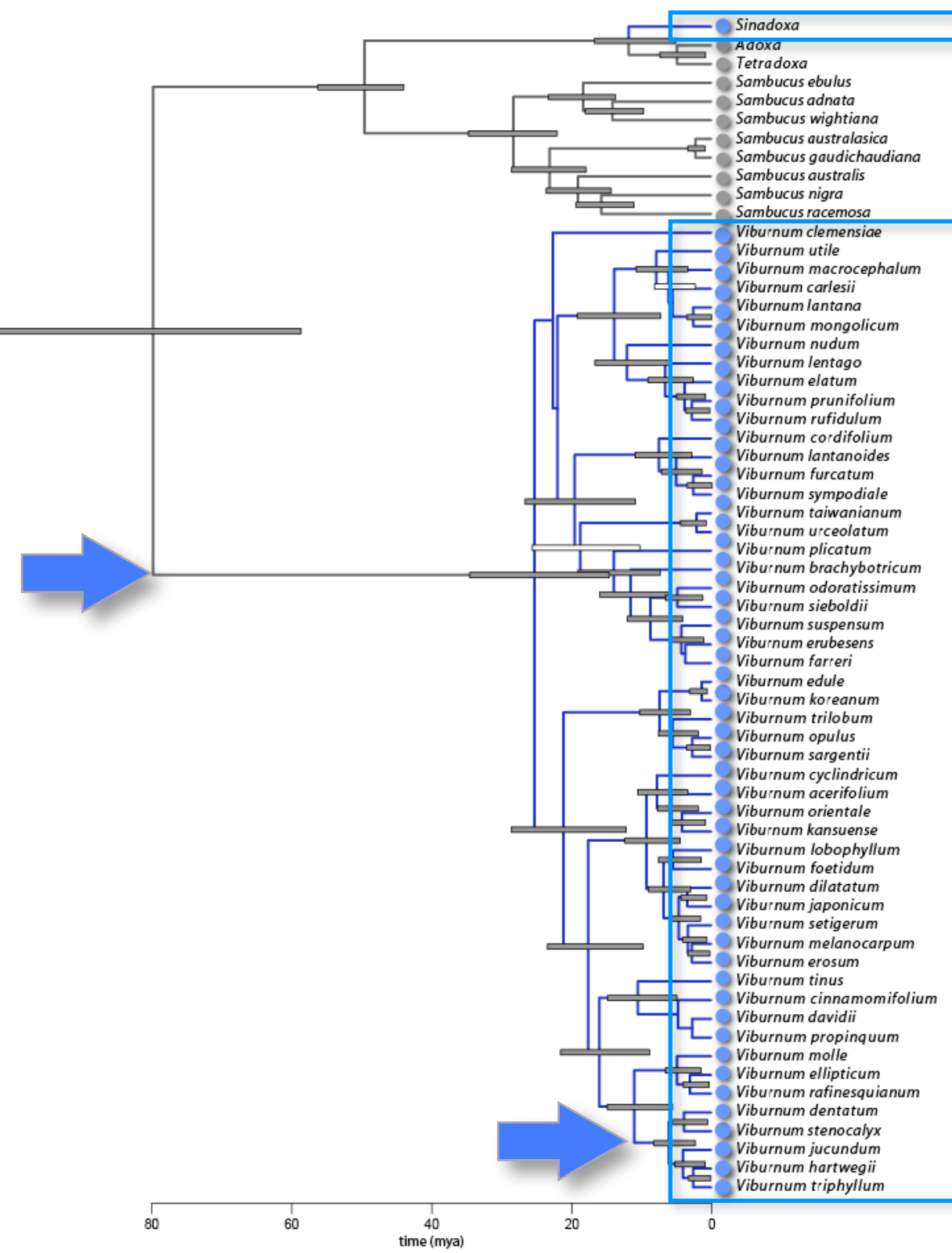
\*SSE model; DiversiTree (Maddison et al., 2007; Fitzjohn et al., 2009; Fitzjohn, 2010...)



CVPPD; tRate (Moore & Donoghue, 2009)

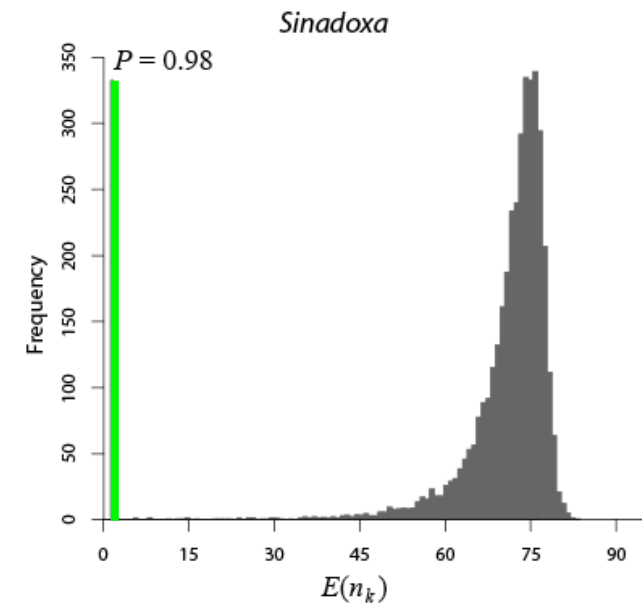
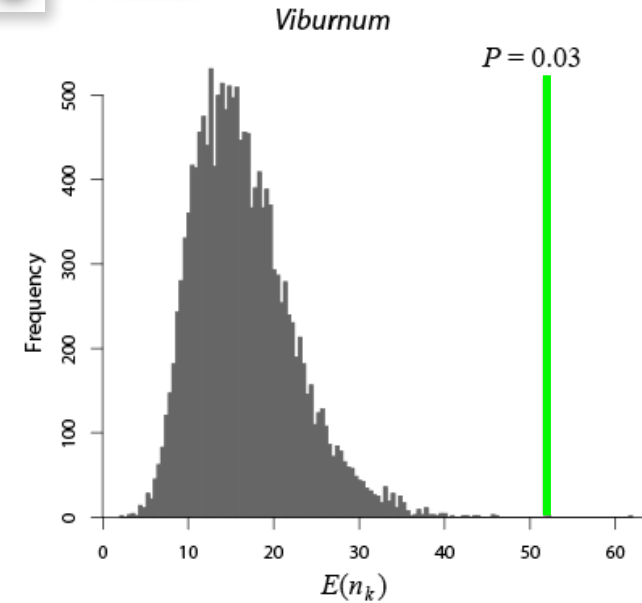
# Cross-Validation Predictive Densities





## Fruit Type

- 1 seed
- >1 seed



# Summary: Some General Advice for Exploring Diversification Rates

1. Expectations under SBP models are diffuse to a degree that defies intuition  
this makes it difficult to detect departures from stochastic expectations
2. When using methods that entail SBP models for estimation it's critical to:  
carefully assess model fit/adequacy  
carefully assess our ability to reliably estimate under the assumed model
3. It's important to accommodate various sources of phylogenetic uncertainty  
inferences based on point estimates are unlikely to be reliable
4. The statistical behavior of many recent methods is poorly characterized  
power analyses may be useful for assessing the ability of methods to make the desired inferences from your data
5. Make an effort to understand—and assess—the (implicit) assumptions  
the assumptions are often cryptic, critical and frequently violated in real data

# Five Fundamental Questions in the Study of Diversification Rates

- I. Estimating parameters of the branching process
- II. Identifying significant diversification rate shifts through time
- III. Locating significant diversification rate shifts along branches
- IV. Evaluating correlates of differential diversification rates
- V. Detecting significant diversification rate variation across the tree

# Detecting Diversification Rate Variation Across Branches

## Whole-Tree Model-Based Methods

(Chan & Moore, 2002; Moore, Chan & Donoghue, 2004)

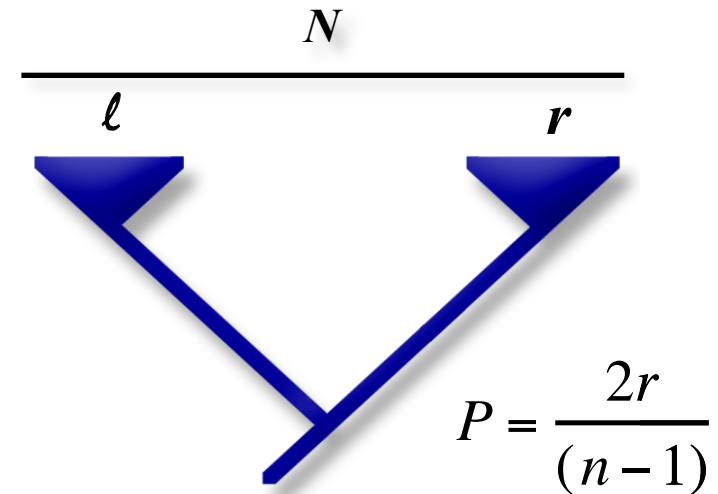
Topological approach

Combines individual ERM nodal probabilities over internal nodes as the product,  $M \prod$ , sum,  $M \Sigma$ , which may be weighted by the sample size of each node,  $M^* \prod$ , sum,  $M^* \Sigma$

Significance estimated by Monte Carlo simulation of the null distribution of the test statistic under an ERM stochastic branching process

Power: HIGH

- + statistically robust (accommodates phylogenetic uncertainty, polytomies)
- + can accommodate incomplete taxon sampling
- + incorporates more of the tree
- does not identify anomalously large/small groups



$$\Pi^* = \frac{\sum_{i=1}^{n-1} \ln(n_i) \ln(P_i)}{\sum_{i=1}^{n-1} \ln(n_i)} \quad \Sigma^* = \frac{\sum_{i=1}^{n-1} \ln(n_i) P_i}{\sum_{i=1}^{n-1} \ln(n_i)}$$