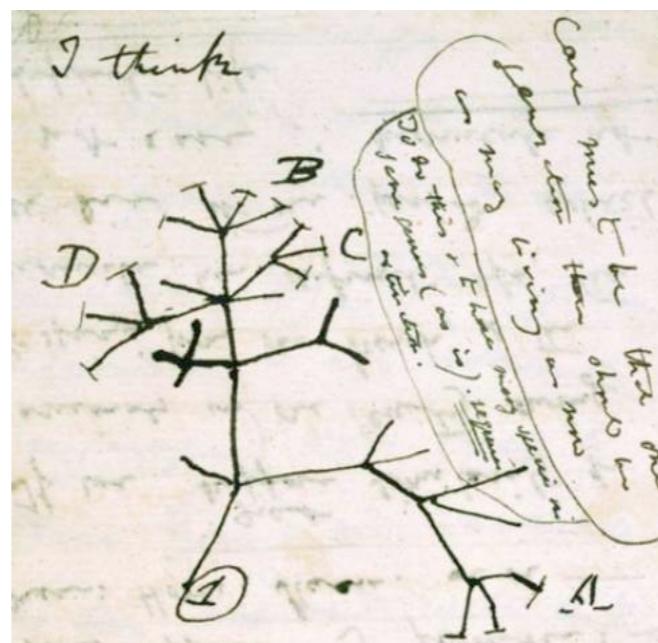
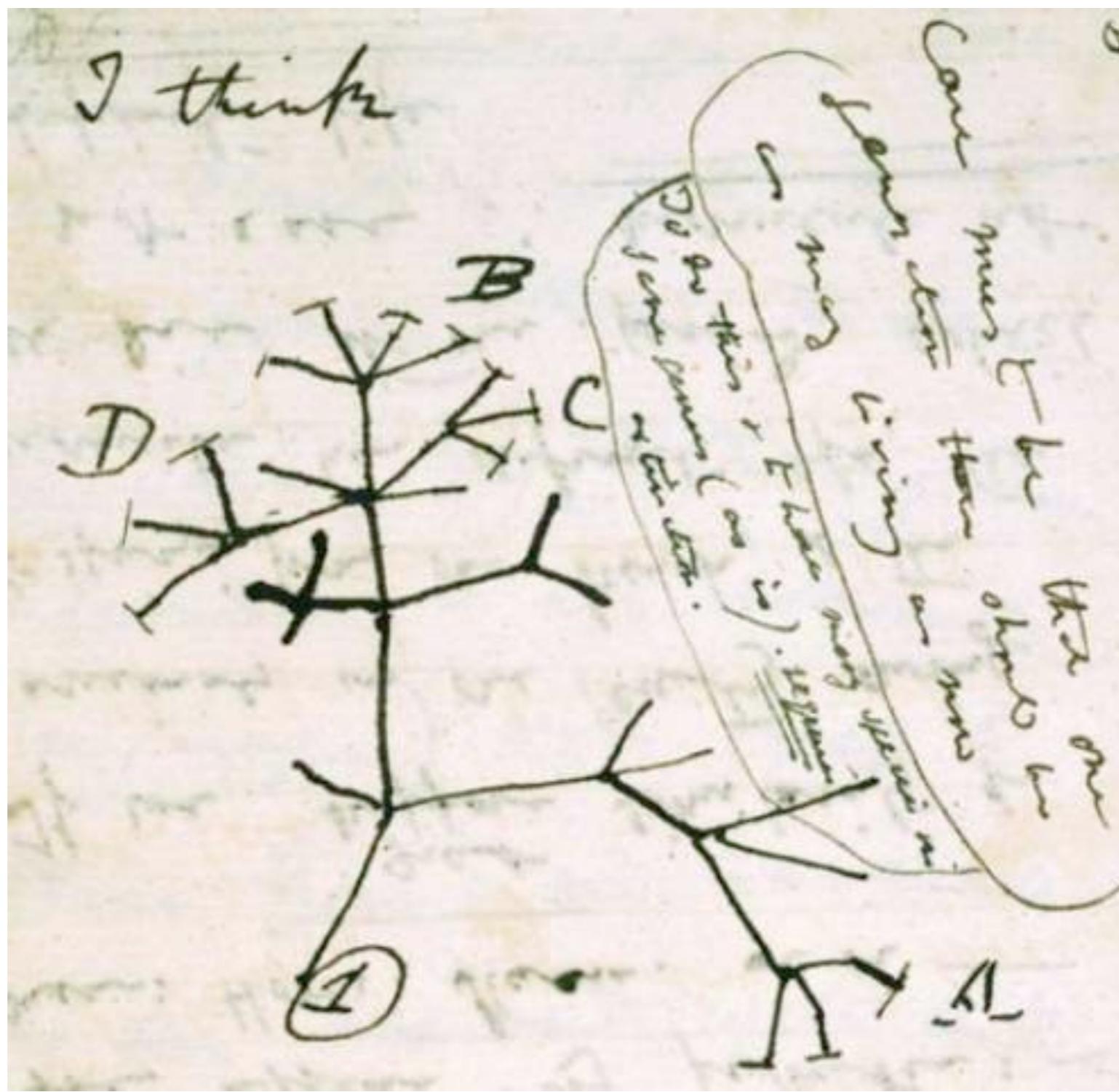


Phylogenetics

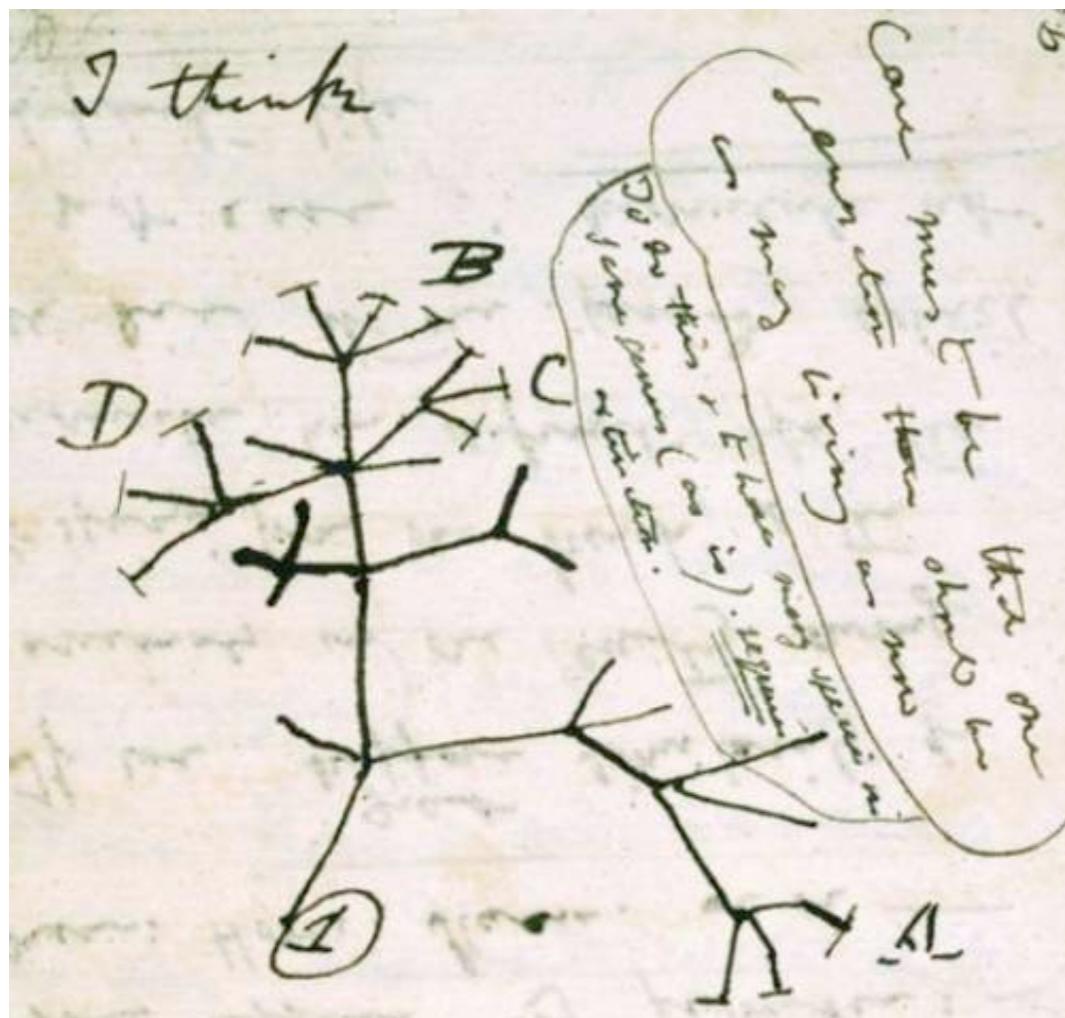
Uncovering
macro-evolutionary dynamics
based on genetic sequence data



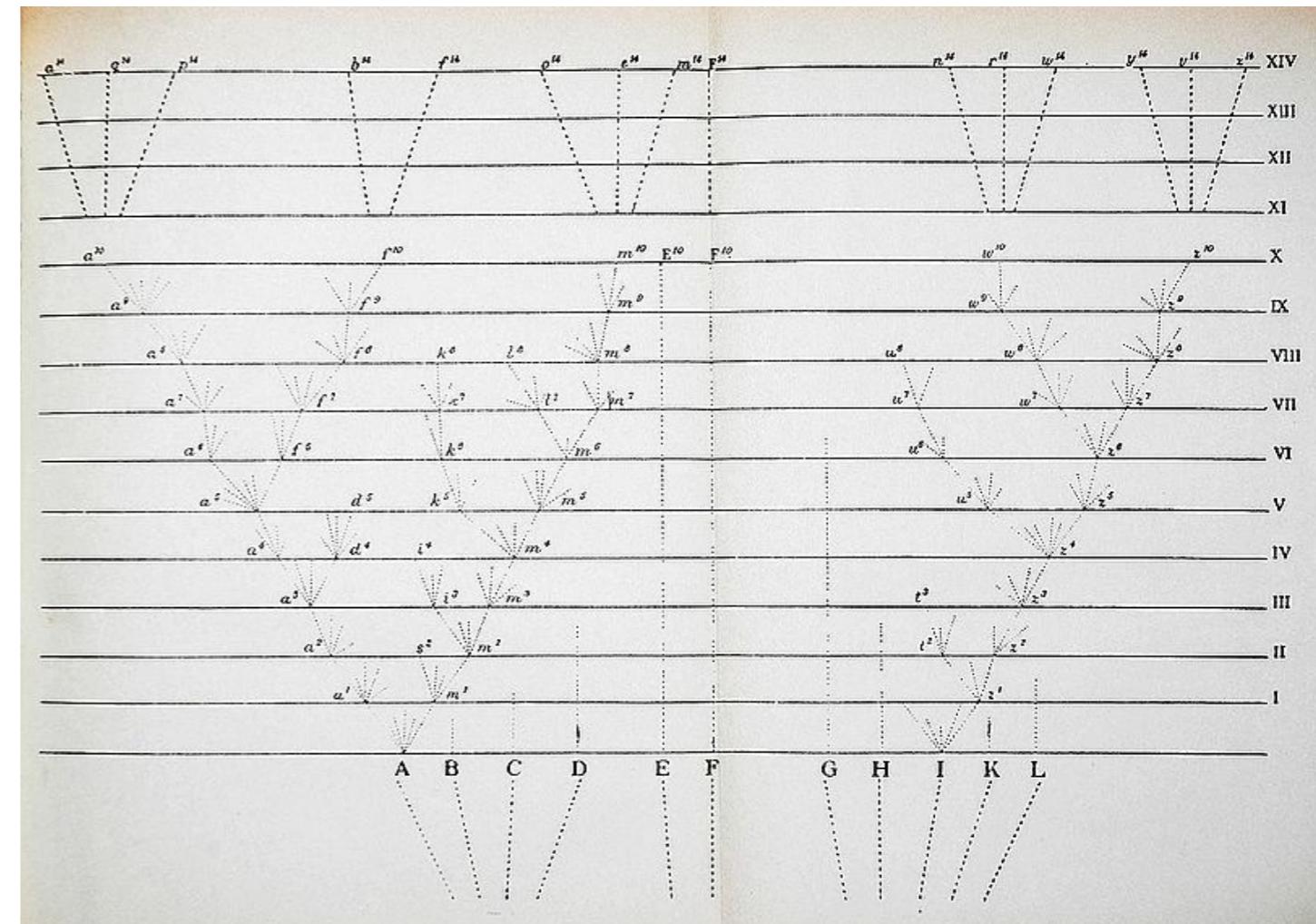
Darwin's first phylogenetic tree



Darwin's two types of trees



Notebook, 1837



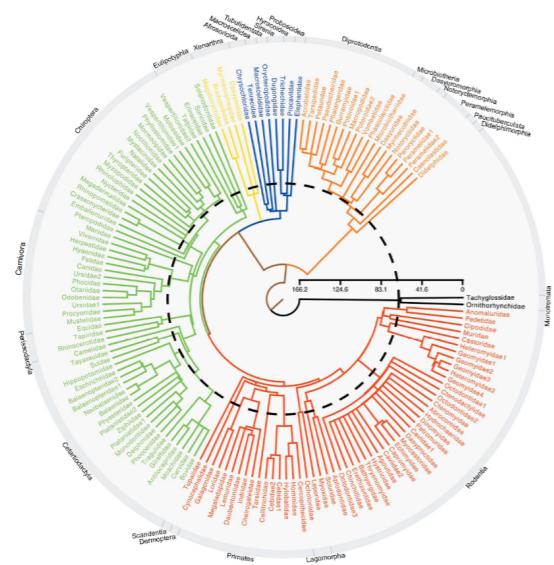
Origin of Species, 1859

Phylodynamic applications



Macro-evolution
Mammalian Evolution
Bininda-Emonds et al. (Nature, 2007)

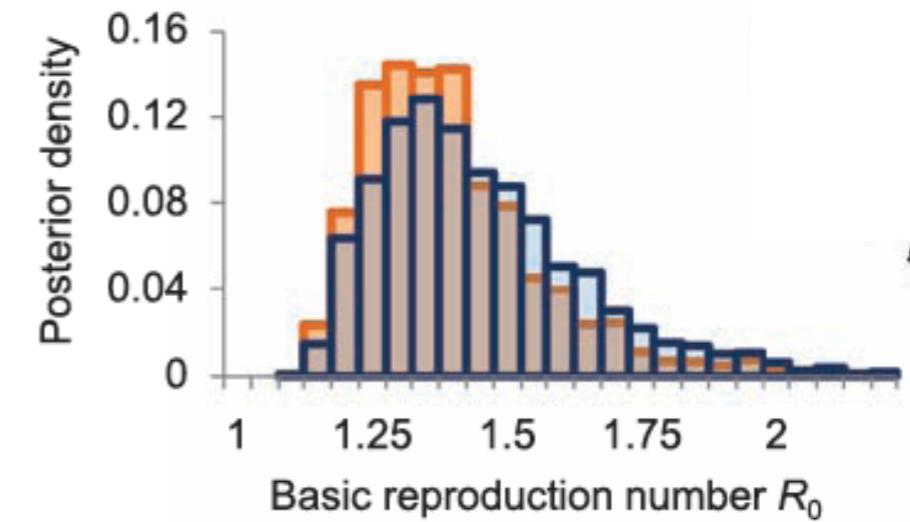
Phylodynamic applications



Macro-evolution

Mammalian Evolution

Bininda-Emonds et al. (Nature, 2007)

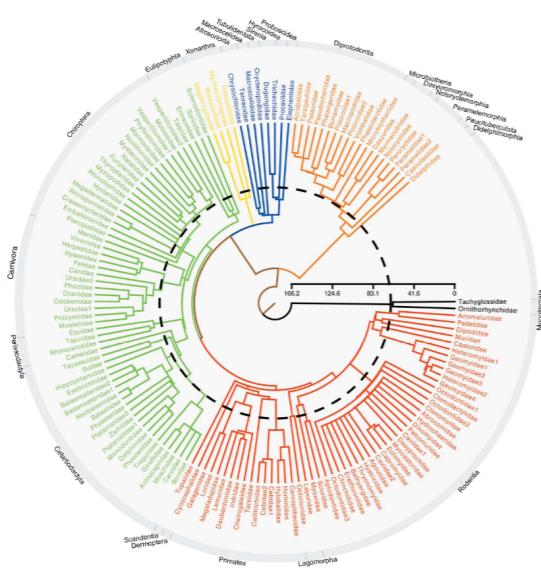


Epidemiology

Spread of an emerging pathogen (here H1N1)

Fraser et al. (Science, 2009)

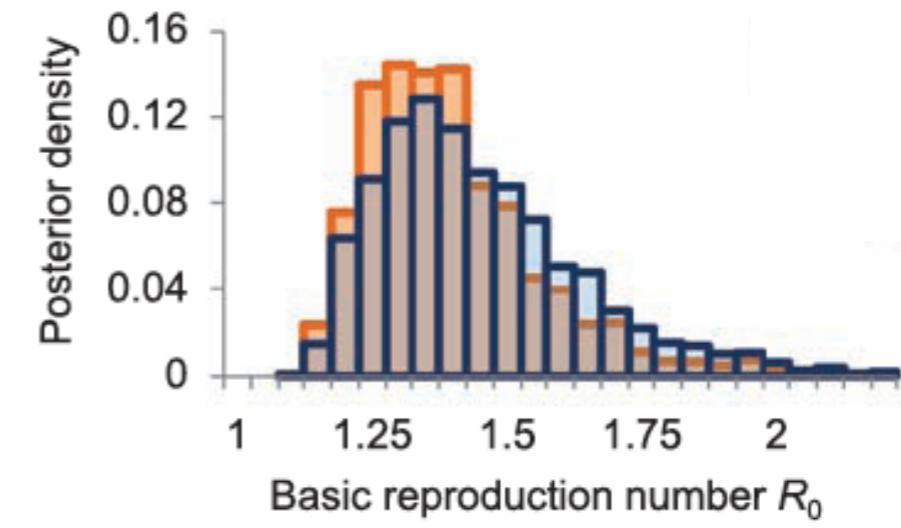
Phylogenetic applications



Macro-evolution

Mammalian Evolution

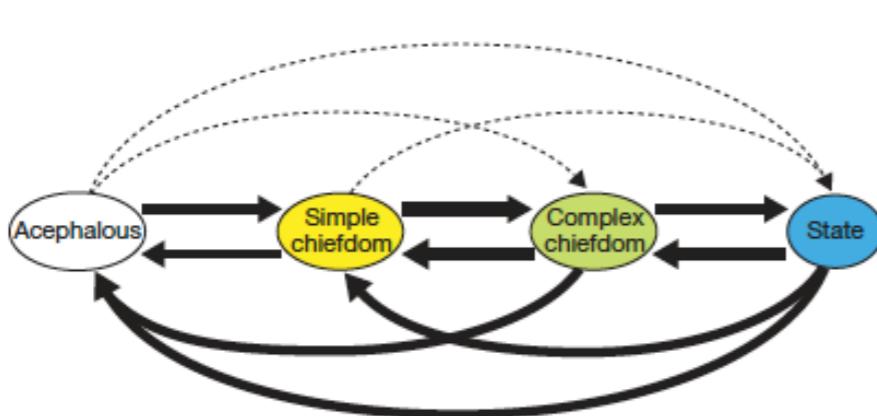
Bininda-Emonds et al. (Nature, 2007)



Epidemiology

Spread of an emerging pathogen (here H1N1)

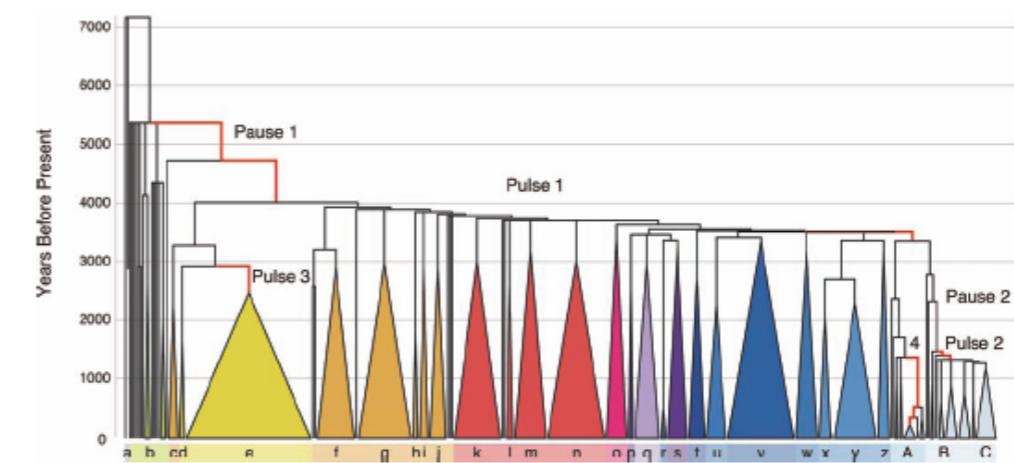
Fraser et al. (Science, 2009)



Sociology

Stability of political systems

Greenhill et al. (Nature, 2010)

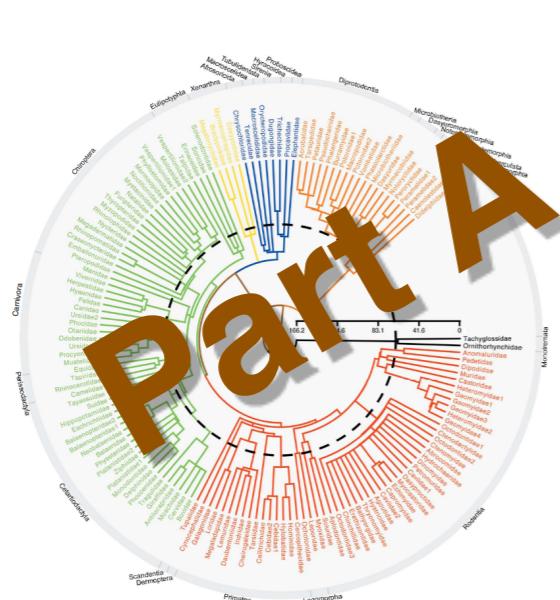


Linguistics

Language evolution

Gray et al. (Science, 2009)

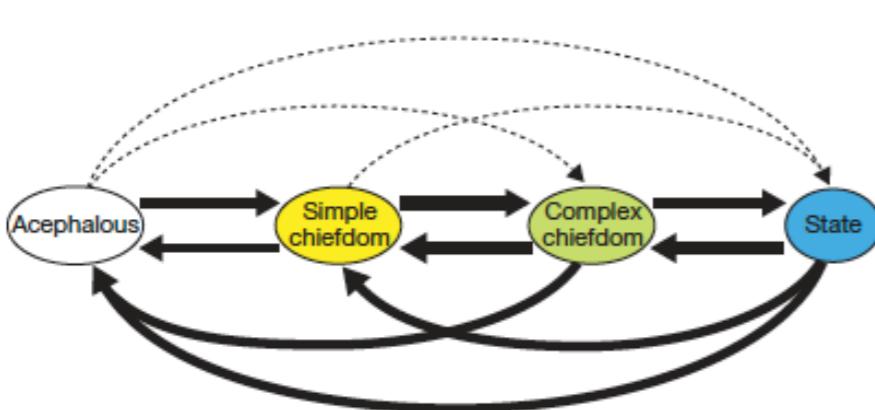
Phylodynamic applications



Macro-evolution

Mammalian Evolution

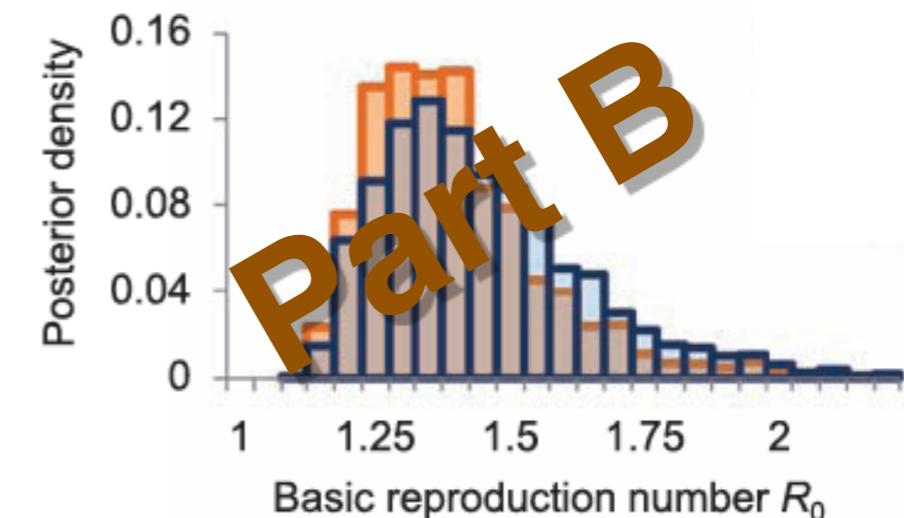
Bininda-Emonds et al. (Nature, 2007)



Sociology

Stability of political systems

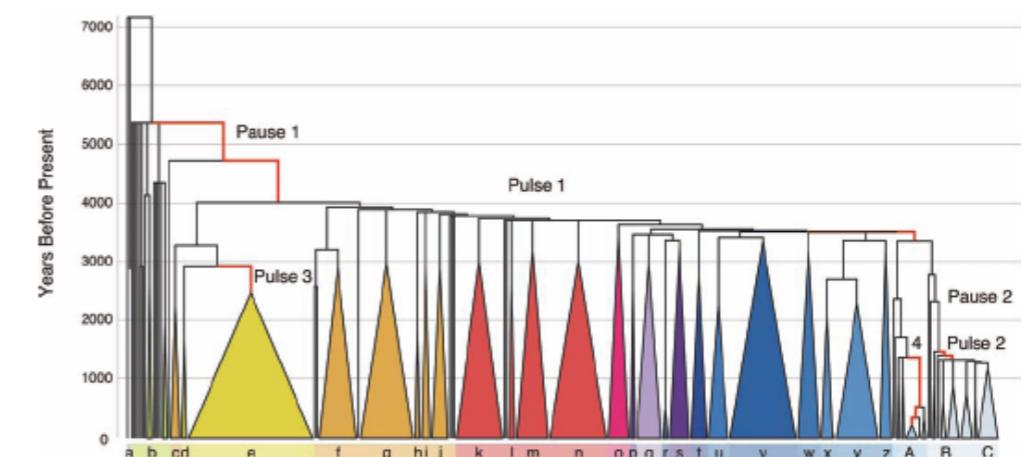
Greenhill et al. (Nature, 2010)



Epidemiology

Spread of an emerging pathogen (here H1N1)

Fraser et al. (Science, 2009)



Linguistics

Language evolution

Gray et al. (Science, 2009)

Part A:

Estimating macro-evolutionary processes



Major challenge in evolutionary biology,
conservation biology & paleontology:

- ▶ Identify main **factors governing macro-evolutionary dynamics**

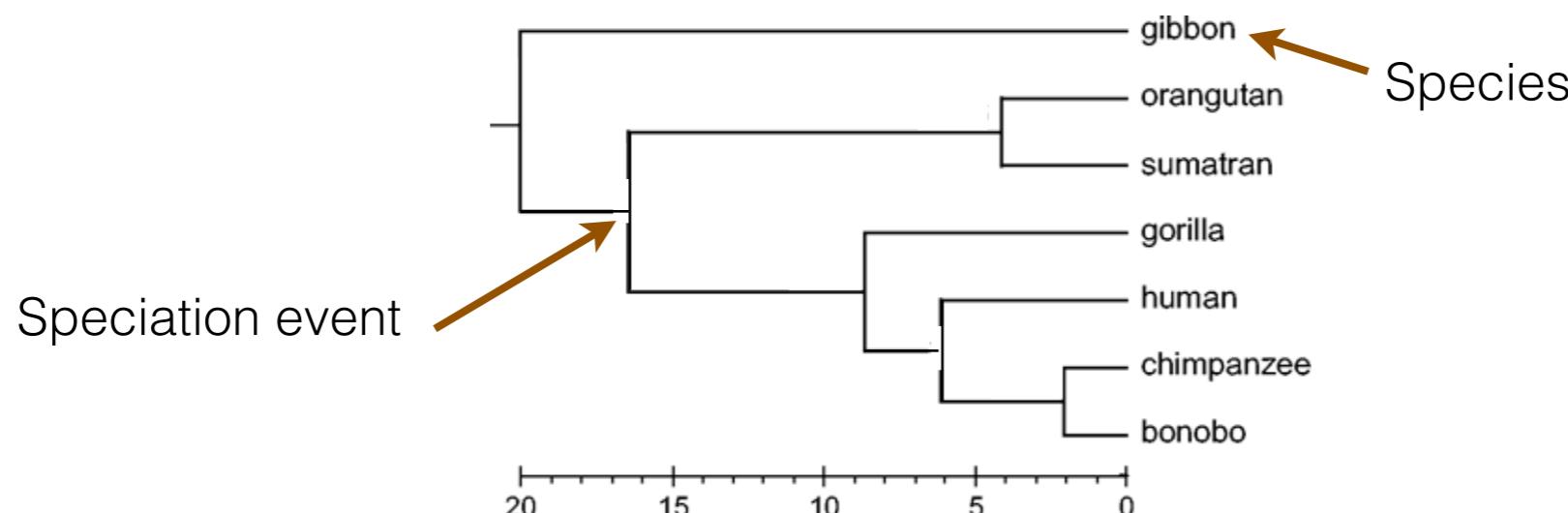
Part A:

Estimating macro-evolutionary processes



Major challenge in evolutionary biology, conservation biology & paleontology:

- ▶ Identify main **factors governing macro-evolutionary dynamics**



Phylogenetic tree of apes
(from Yang & Rannala, MBE, 2005)

Macro-evolutionary models

Null model

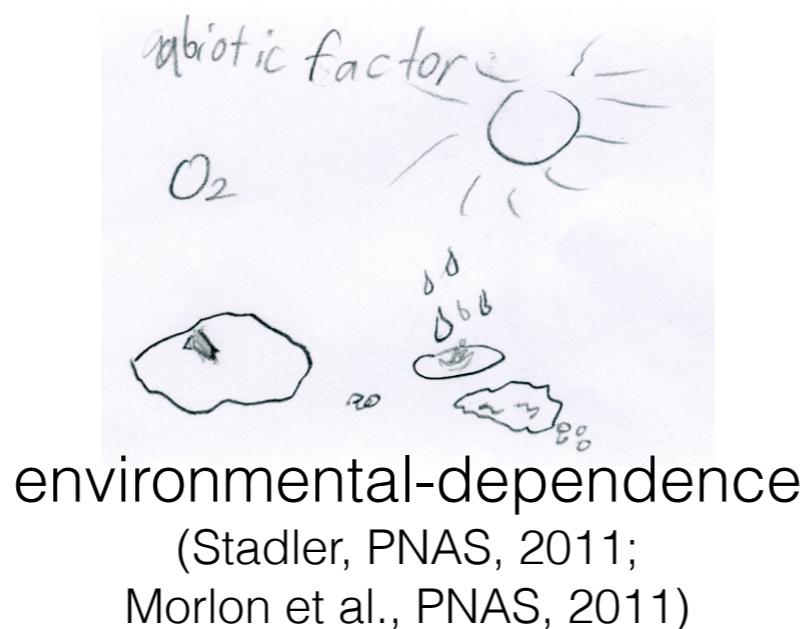
- ▶ Constant speciation and extinction rates
(Yule, Phil. Trans. Roy. Soc. B 1924; Nee et al., Phil. Trans. Roy. Soc. B, 1994)

Macro-evolutionary models

Null model

- Constant speciation and extinction rates
(Yule, Phil. Trans. Roy. Soc. B 1924; Nee et al., Phil. Trans. Roy. Soc. B, 1994)

Rate changes due to extrinsic factors

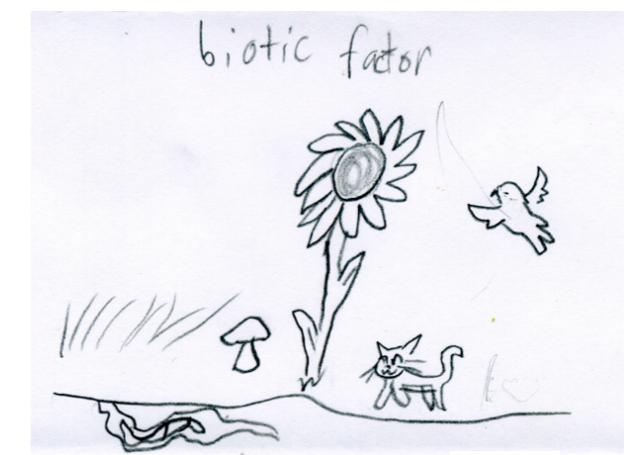
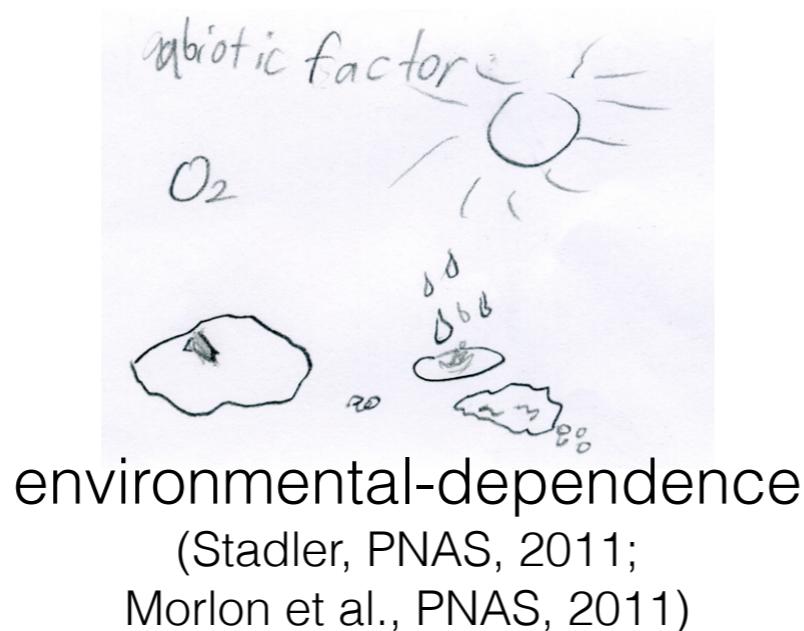


Macro-evolutionary models

Null model

- Constant speciation and extinction rates
(Yule, Phil. Trans. Roy. Soc. B 1924; Nee et al., Phil. Trans. Roy. Soc. B, 1994)

Rate changes due to extrinsic factors



Rate changes due to intrinsic factors



trait-dependence
(Maddison et al., Syst. Biol., 2007,
Fitzjohn et al., Syst. Biol, 2009)

Macro-evolutionary models

Null model

- Constant speciation and extinction rates
(Yule, Phil. Trans. Roy. Soc. B 1924; Nee et al., Phil. Trans. Roy. Soc. B, 1994)

Rate changes due to extrinsic factors



(Stadler, PNAS, 2011;
Morlon et al., PNAS, 2011)



(Etienne et al.,
Proc. Roy. Soc. B, 2011)

Rate changes due to intrinsic factors



trait-dependence
(Maddison et al., Syst. Biol., 2007,
Fitzjohn et al., Syst. Biol, 2009)

Likelihood approach for inferring macroevolutionary processes

Mathematical

- ▶ Define a macroevolutionary model
- ▶ Find efficient way to calculate likelihood function:
Probability of the phylogenetic tree given the parameters

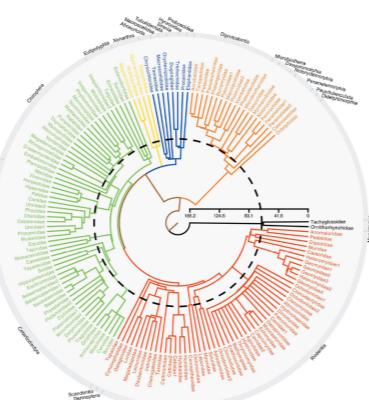


Computational

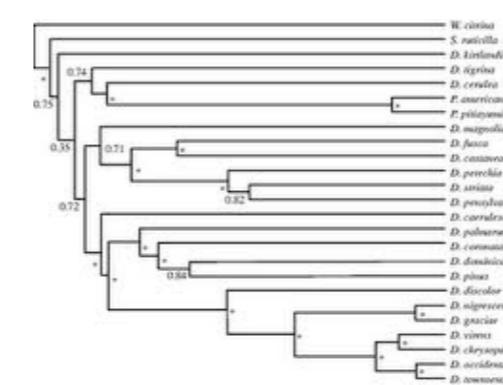
- ▶ Implement likelihood and find optimum
- ▶ Today: R packages TreePar & TreeSim



Empirical

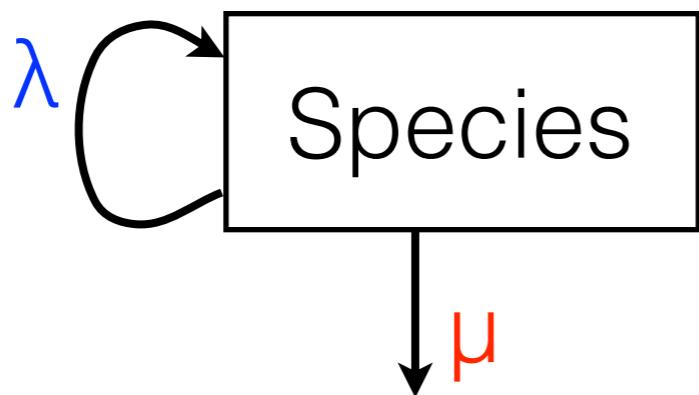
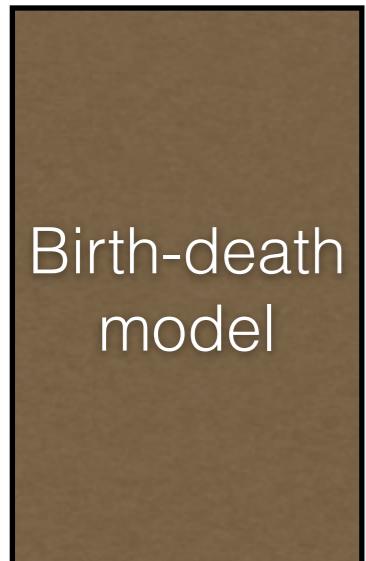


Mammals



Birds

The birth-death model as a model for speciation and extinction

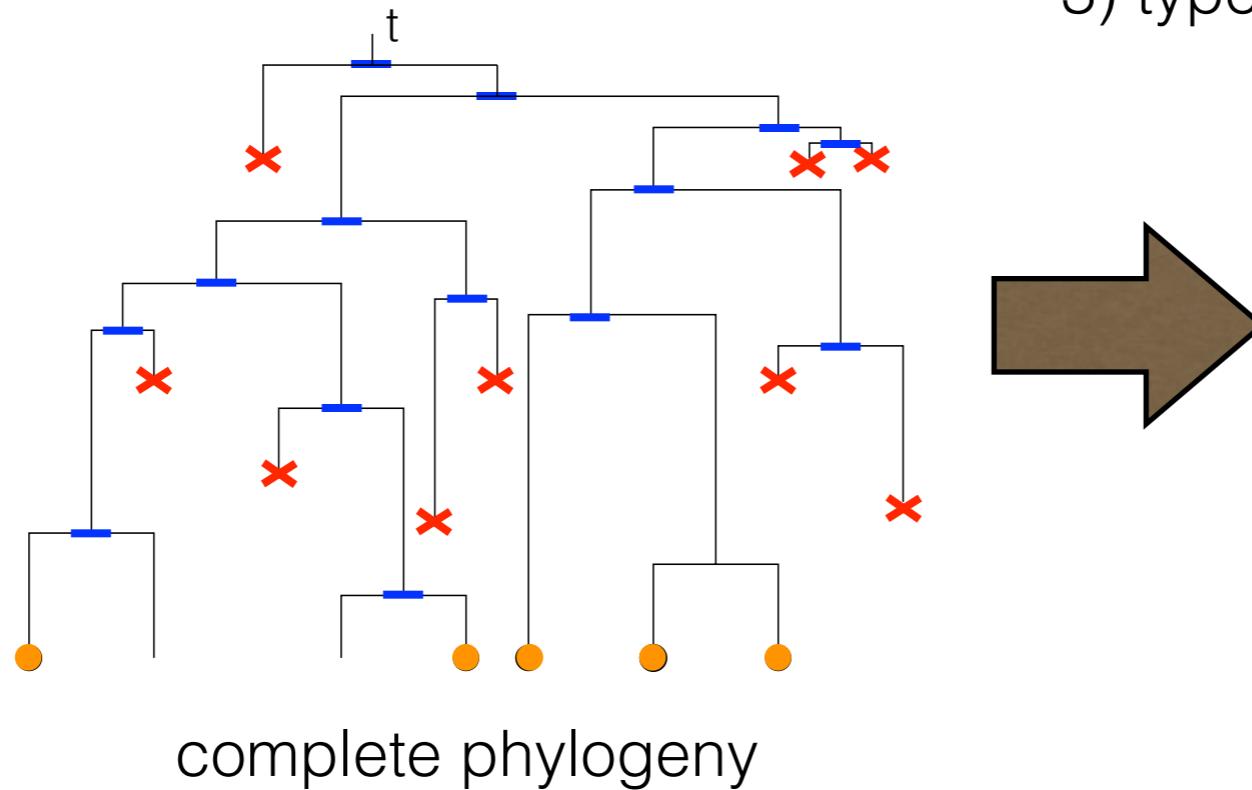
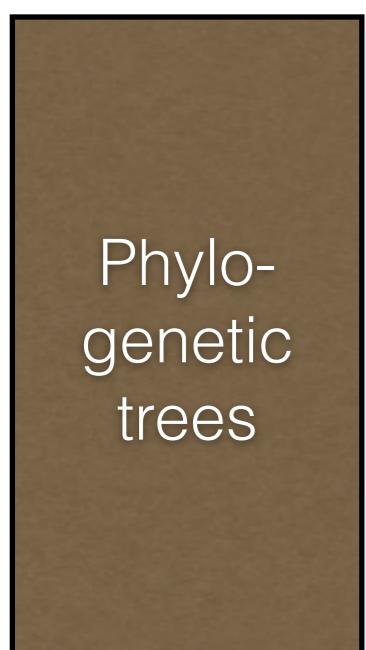
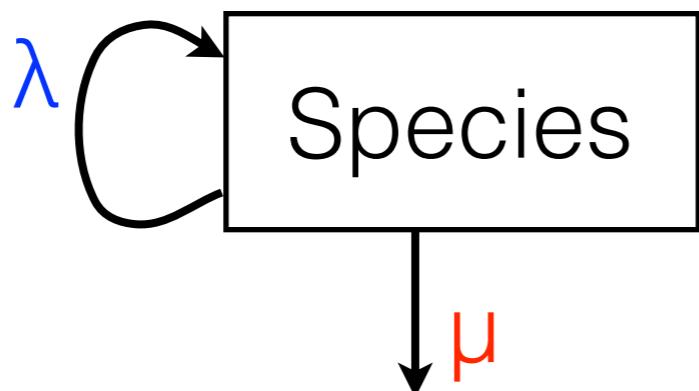
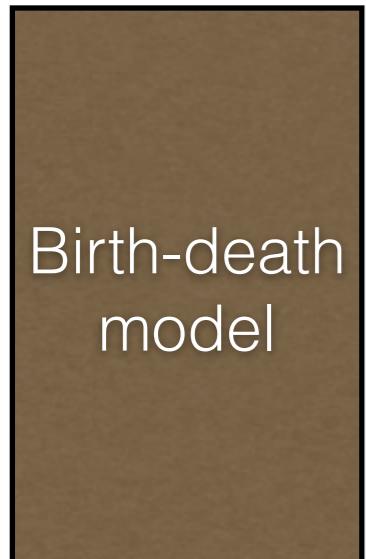


- stem age of a clade t
- speciation rate λ
- extinction rate μ
- sampling probability p

Rates may depend on:

- 0) constant
- 1) time (environmental-dependence)
- 2) # of species (diversity-dependence)
- 3) type of species

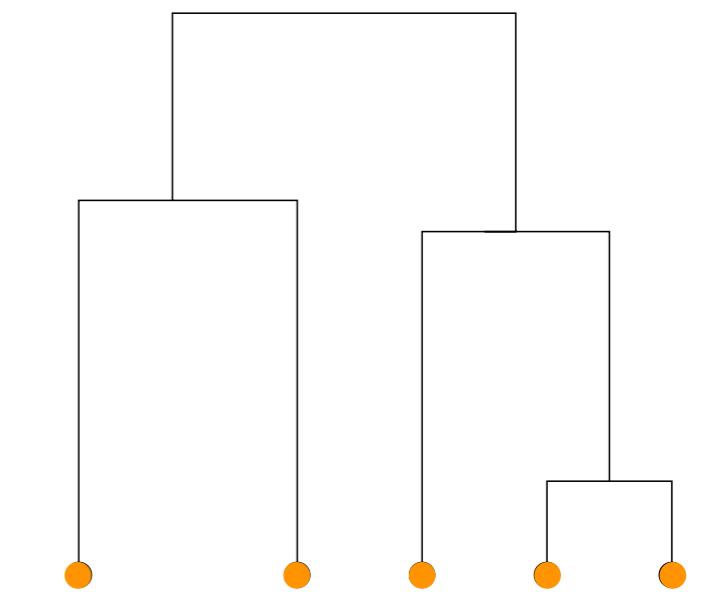
The birth-death model as a model for speciation and extinction



- stem age of a clade t
- speciation rate λ
- extinction rate μ
- sampling probability p

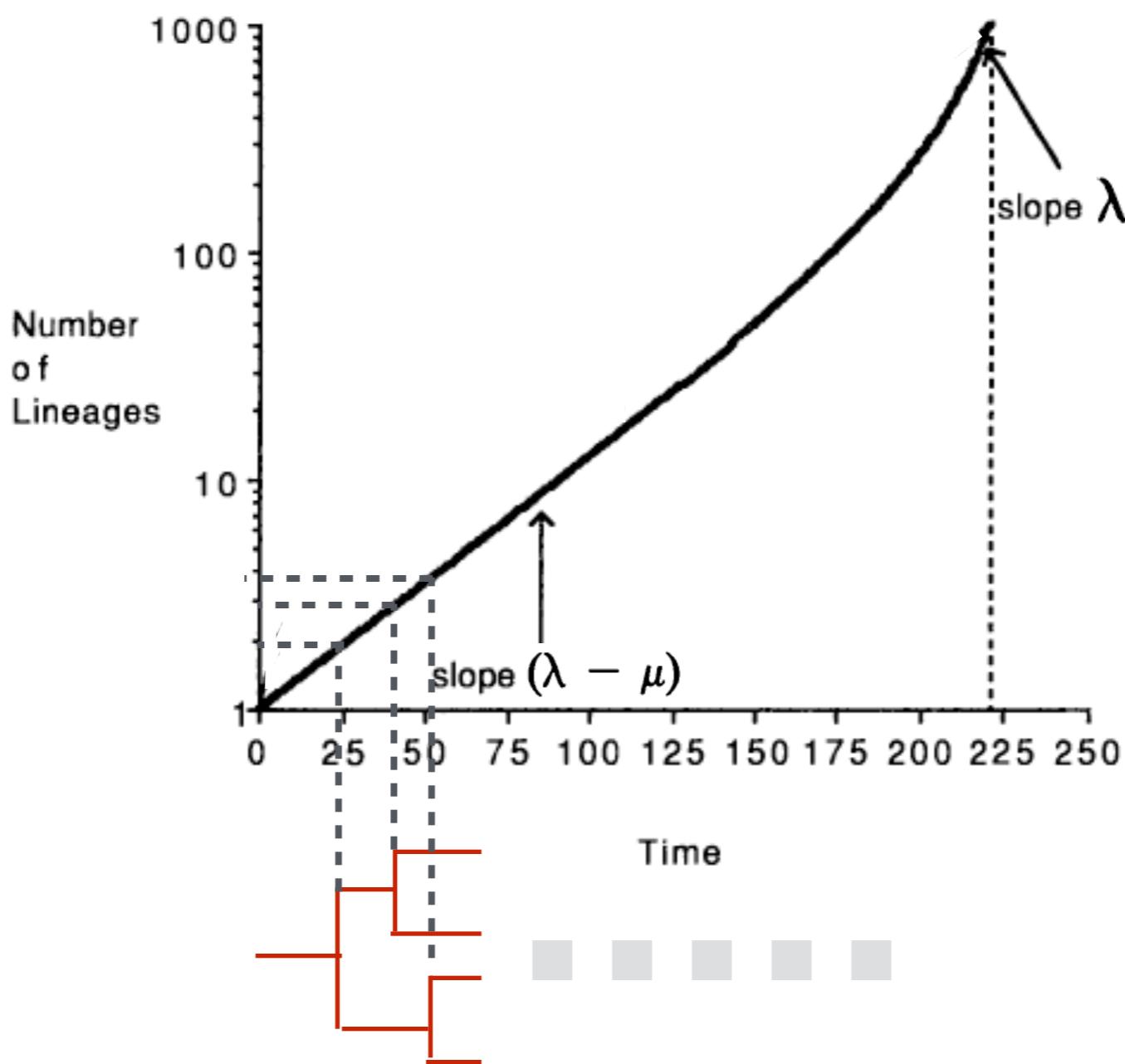
Rates may depend on:

- 0) constant
- 1) time (environmental-dependence)
- 2) # of species (diversity-dependence)
- 3) type of species



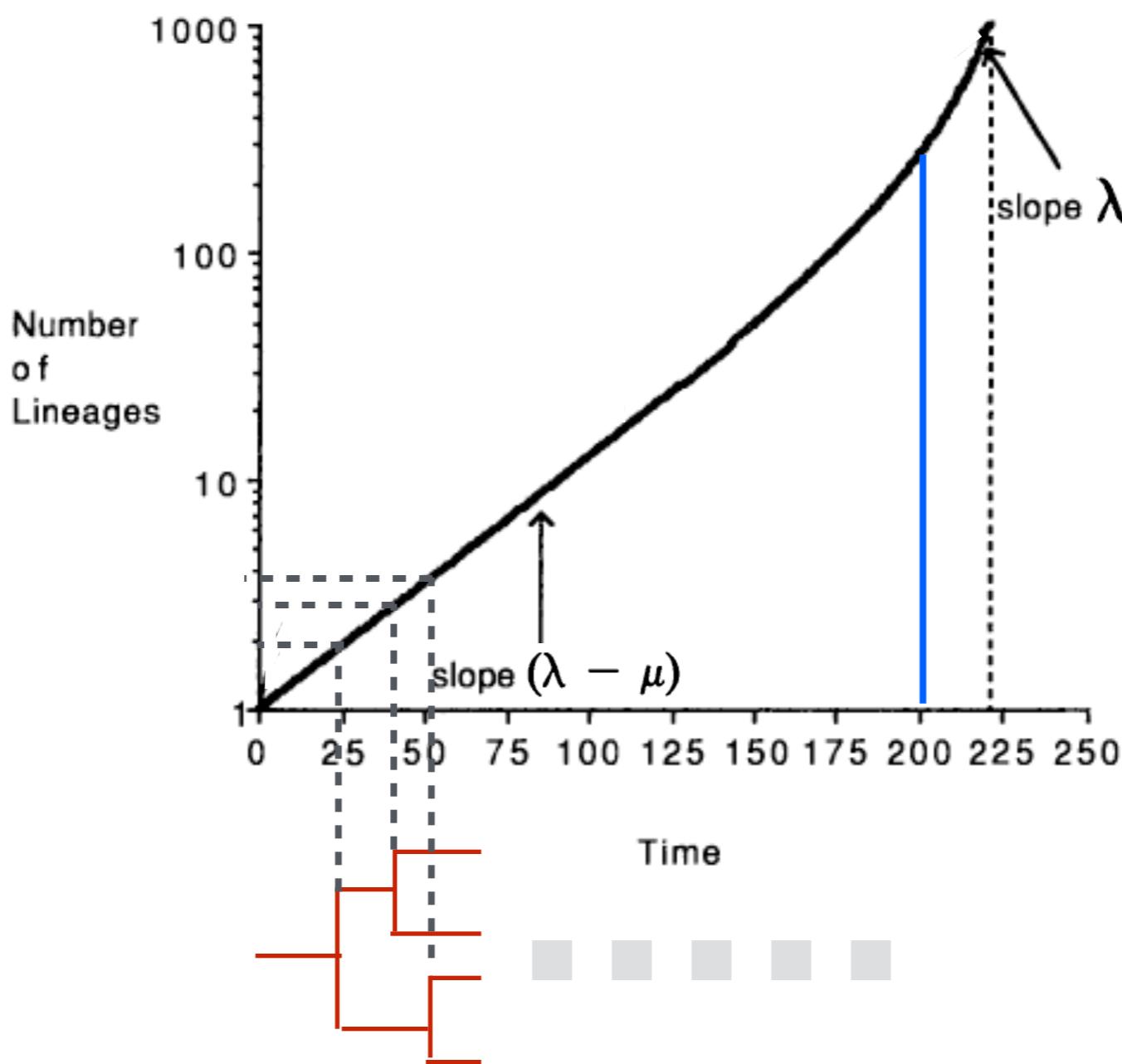
reconstructed phylogeny

Lineages-through-time (LTT) plots



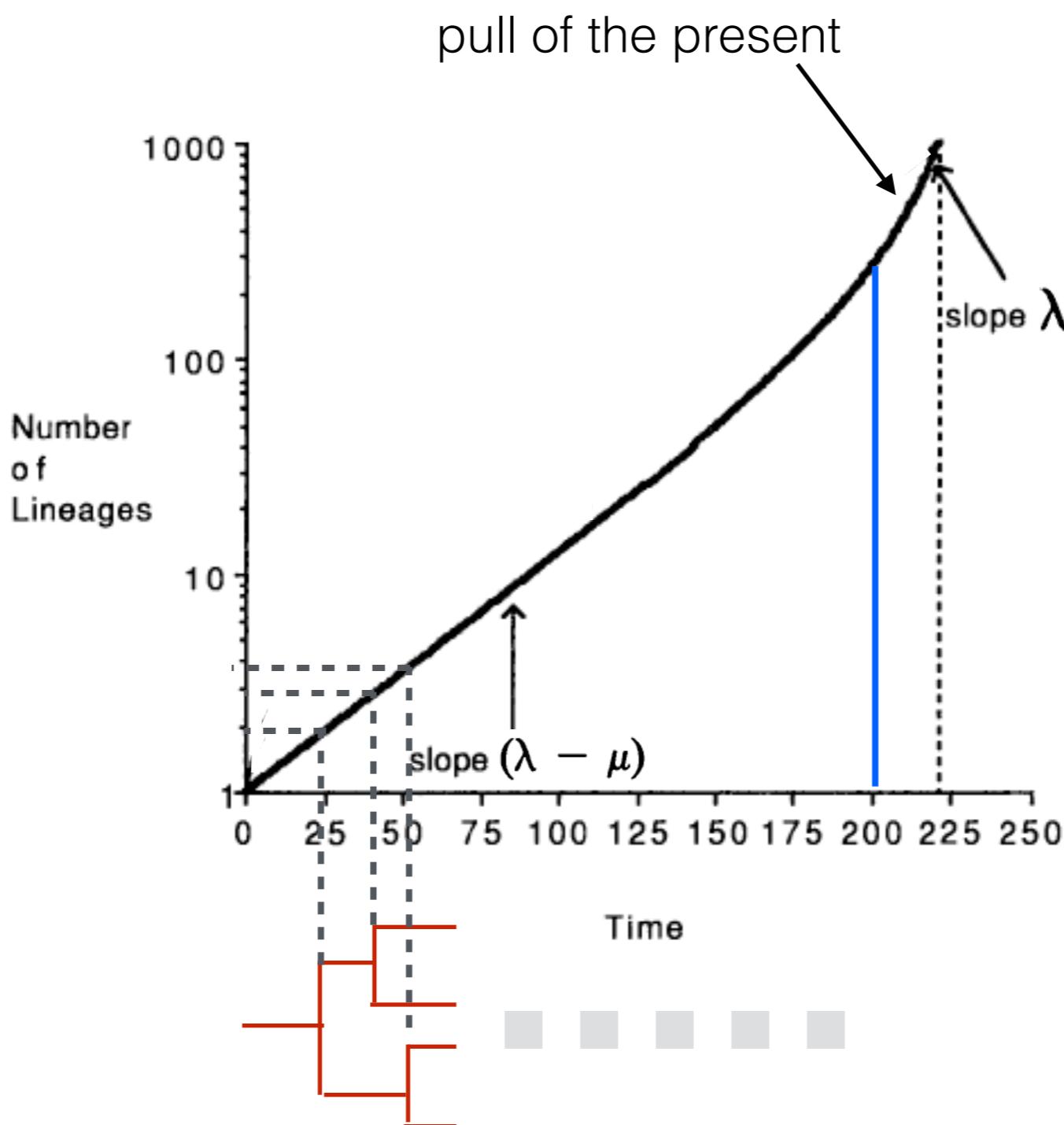
adapted from Harvey et al.
(1994, Evolution)

Lineages-through-time (LTT) plots



adapted from Harvey et al.
(1994, Evolution)

Lineages-through-time (LTT) plots

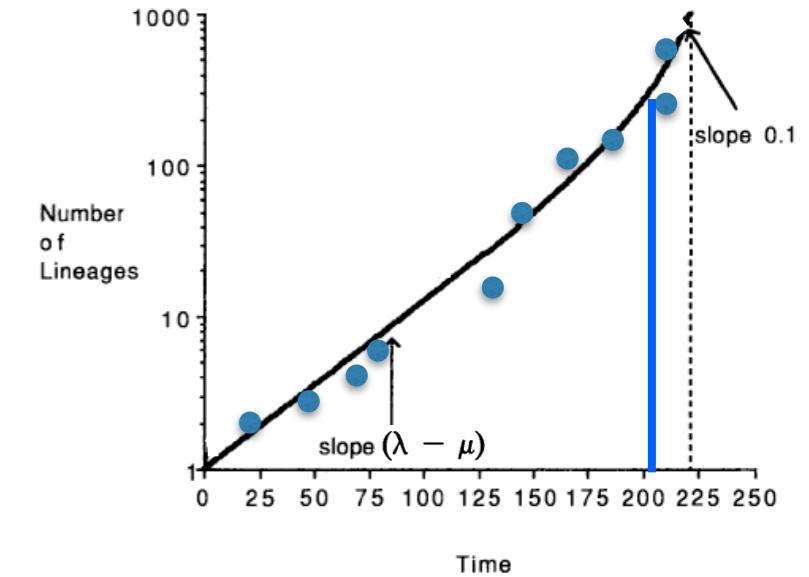


adapted from Harvey et al.
(1994, Evolution)

Lineages-through-time plots to estimate macroevolutionary rates

Naive

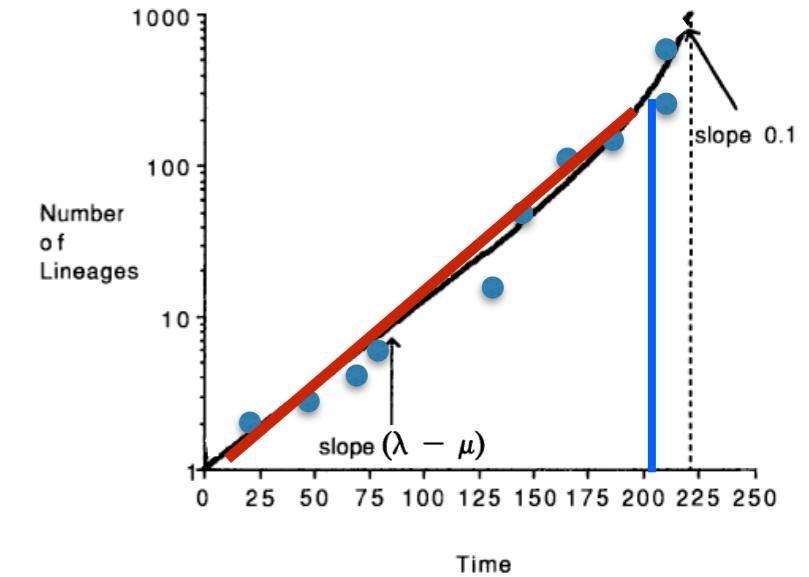
- ▶ Fit linear models to empirical data
- ▶ Problems:
 - ▶ Weights of the data points not clear
 - ▶ Confuses rate change and “pull of the present”



Lineages-through-time plots to estimate macroevolutionary rates

Naive

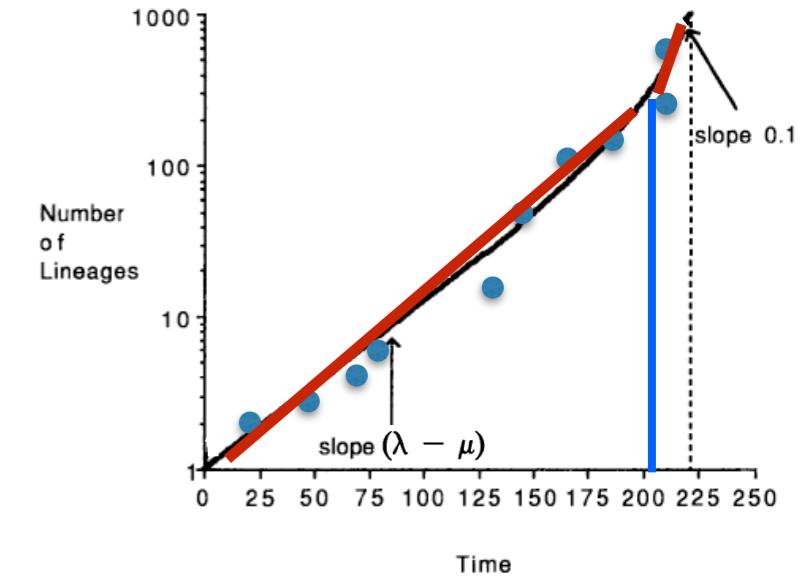
- ▶ Fit linear models to empirical data
- ▶ Problems:
 - ▶ Weights of the data points not clear
 - ▶ Confuses rate change and “pull of the present”



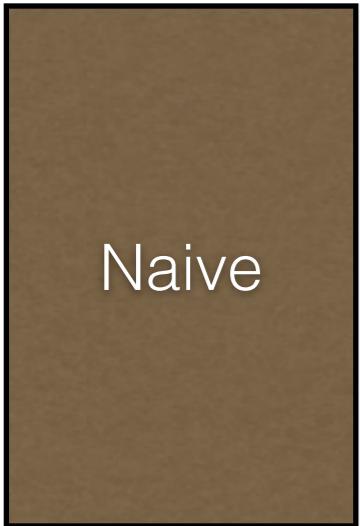
Lineages-through-time plots to estimate macroevolutionary rates

Naive

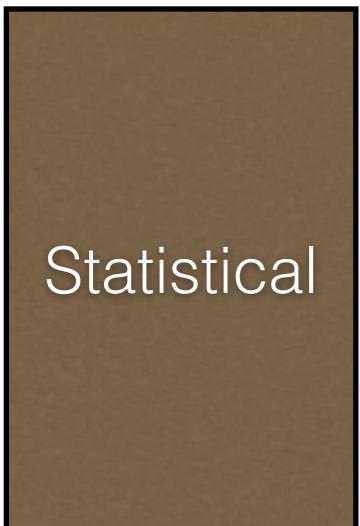
- ▶ Fit linear models to empirical data
- ▶ Problems:
 - ▶ Weights of the data points not clear
 - ▶ Confuses rate change and “pull of the present”



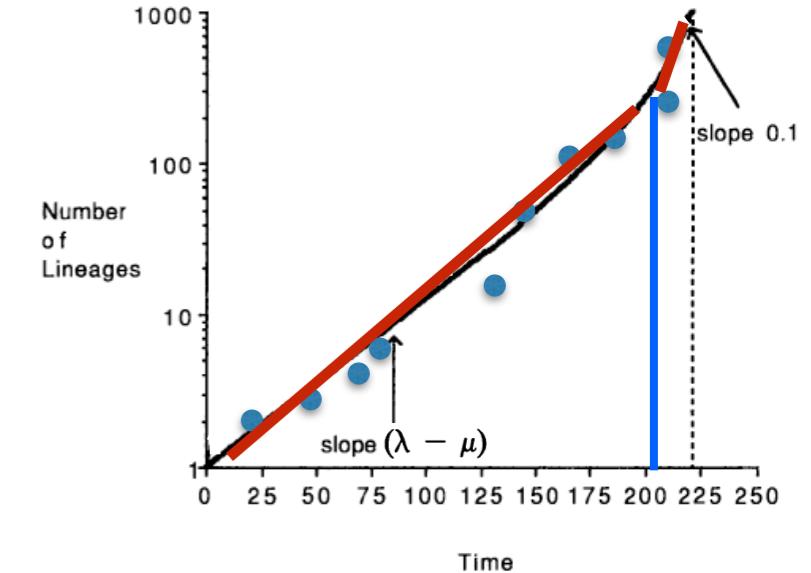
Lineages-through-time plots to estimate macroevolutionary rates



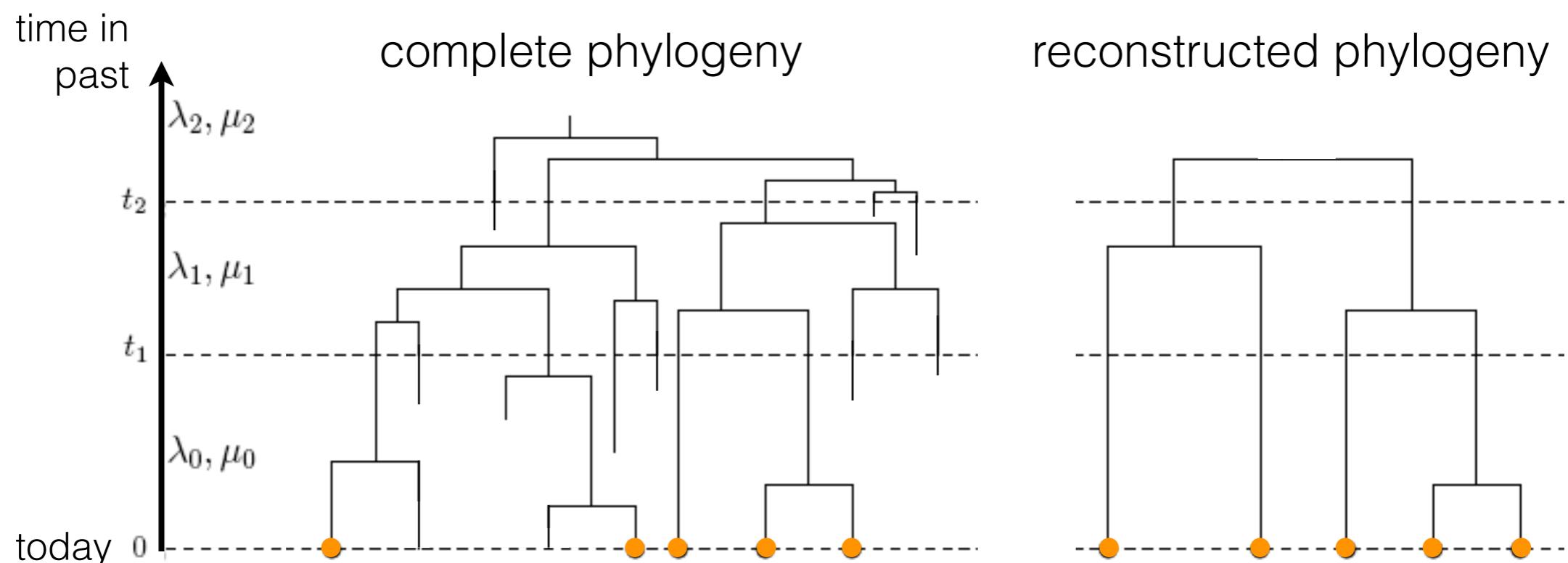
- ▶ Fit linear models to empirical data
- ▶ Problems:
 - ▶ Weights of the data points not clear
 - ▶ Confuses rate change and “pull of the present”



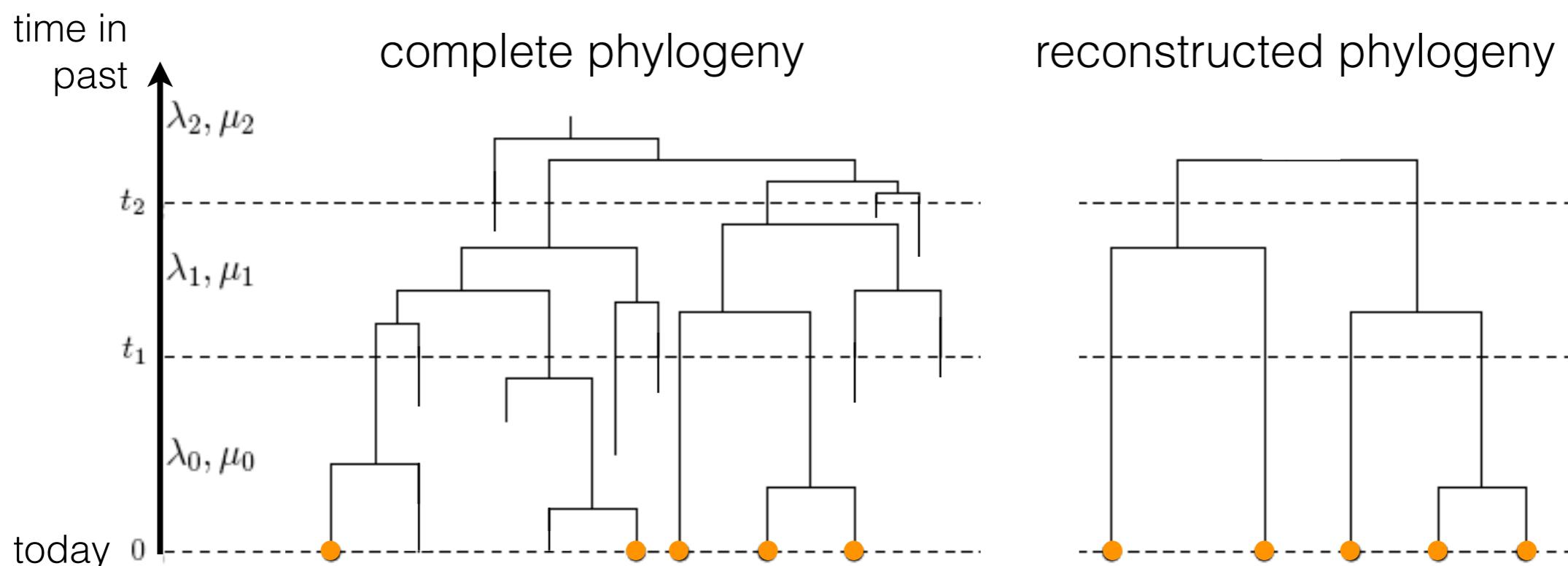
- ▶ Prob (LTT | λ, μ)
Nee et al., Phil. Trans. Roy. Soc. B, 1994
- ▶ Find parameters maximizing the probability:
 - ▶ maximum likelihood estimates



1) Modeling environmental effects



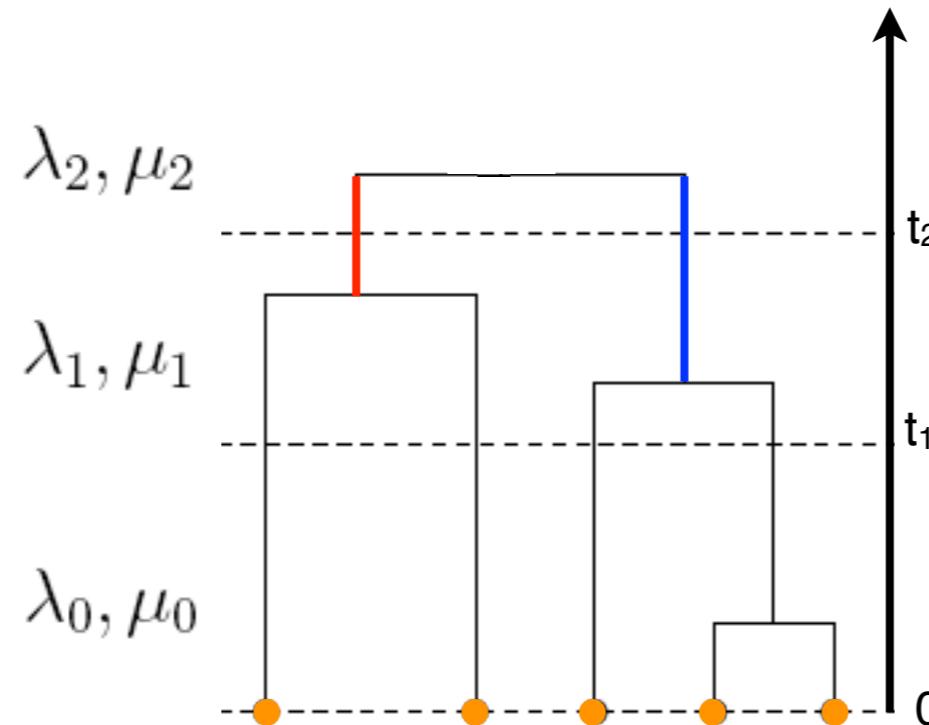
1) Modeling environmental effects



Calculate **likelihood** of reconstructed phylogeny by “integrating” over all complete phylogenies:

Efficiently done using differential equations and recursions

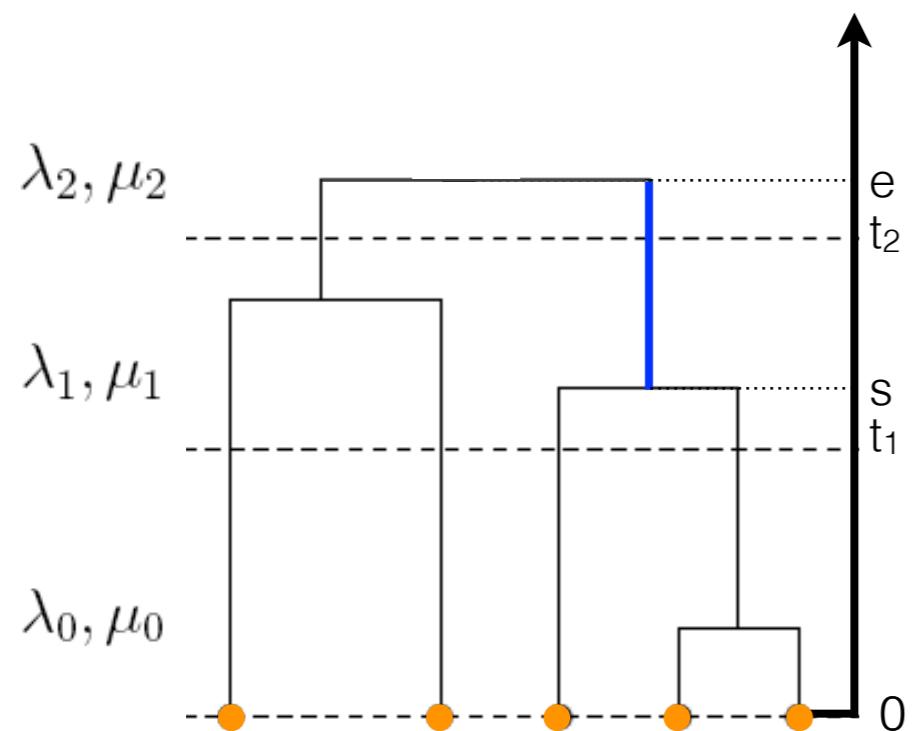
Calculating the likelihood: recursion



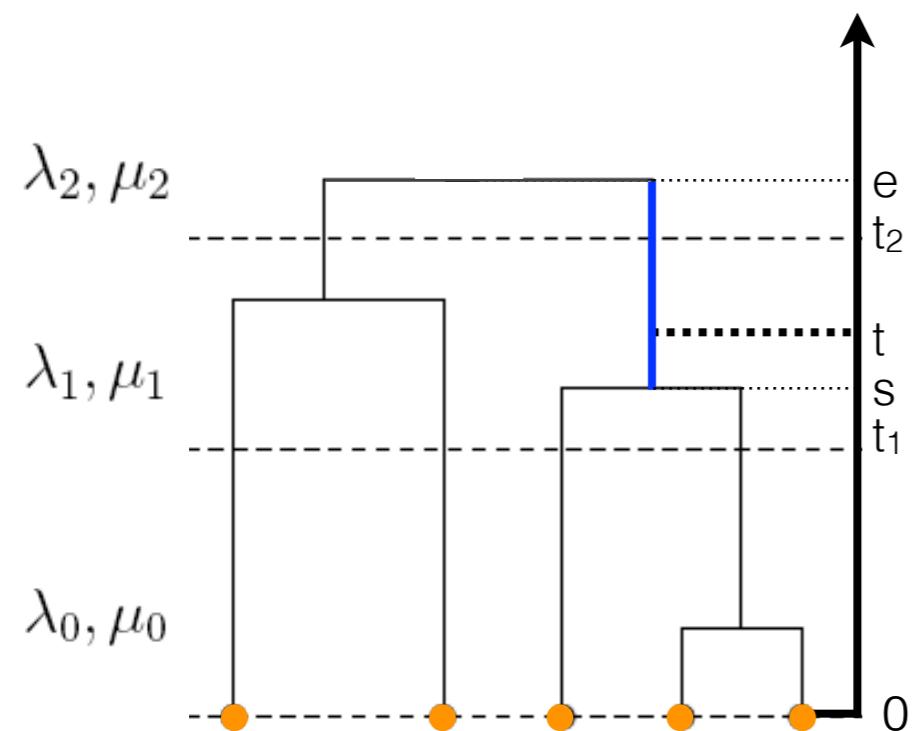
$$f(\text{[diagram of rectangles]}) = \lambda_2 \text{ Prob}(\text{[red bar]}) \text{ Prob}(\text{[blue bar]}) f(\text{[diagram of first two rectangles]}) f(\text{[diagram of last two rectangles]})$$

↑ ↑ ↑ ↑
Differential equation Recursive

Calculating the likelihood: differential equation

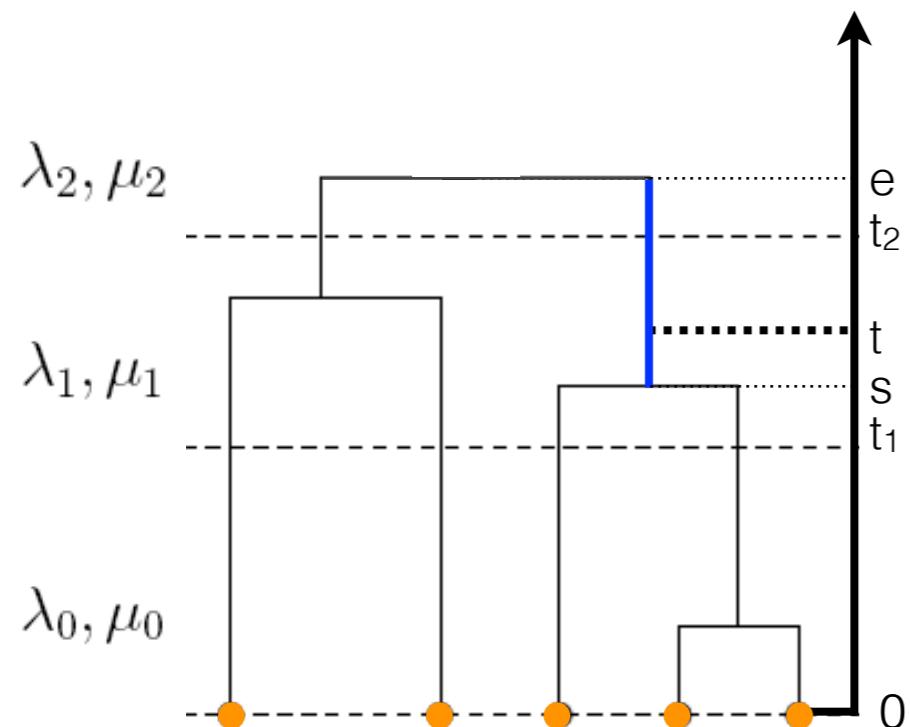


Calculating the likelihood: differential equation



$q(t)$: probability density that a given individual at time t produces the descending edge.

Calculating the likelihood: differential equation

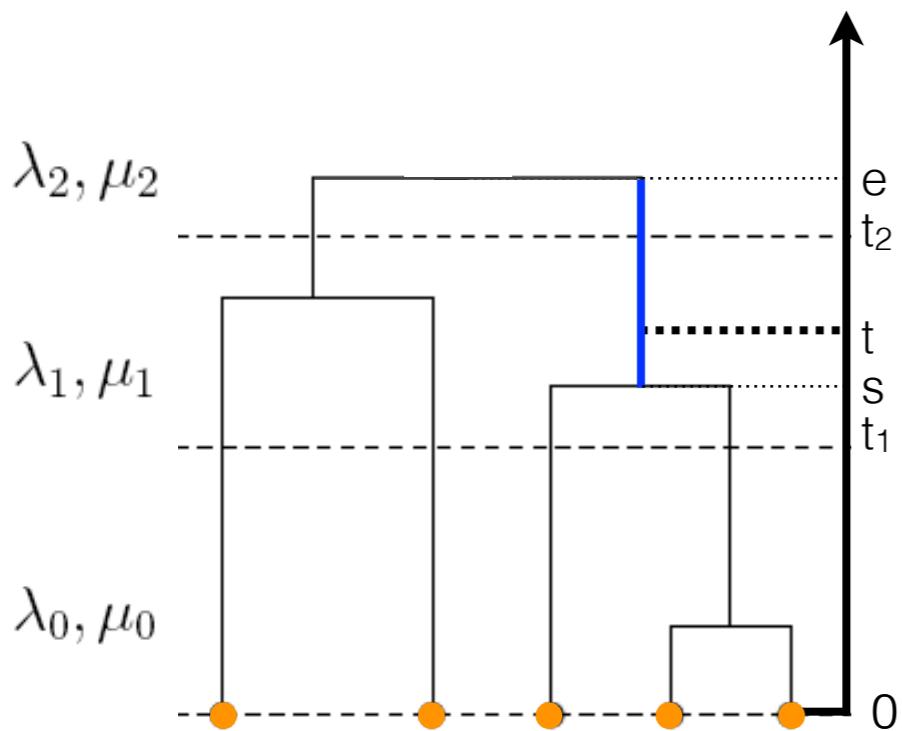


$q(t)$: probability density that a given individual at time t produces the descending edge.

$$q(e) = \text{Prob} (|)$$

$$q(s) = 1$$

Calculating the likelihood: differential equation



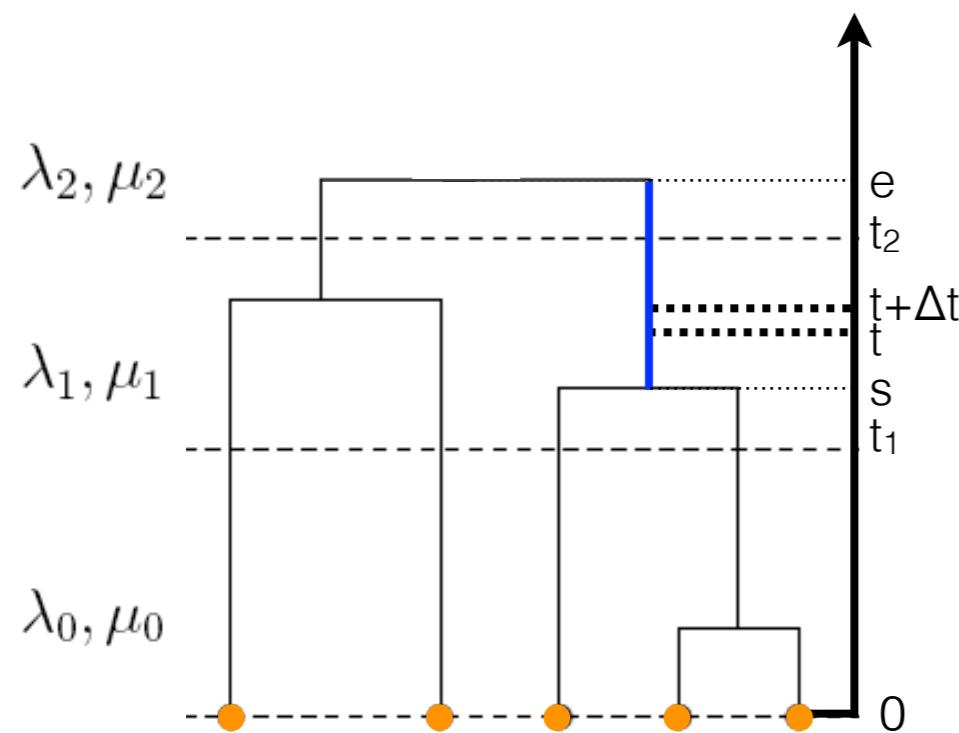
$q(t)$: probability density that a given individual at time t produces the descending edge.

$$q(e) = \text{Prob} (|)$$

$$q(s) = 1$$

$p(t)$: probability that a given individual at time t has no extant and sampled species offspring.

Calculating the likelihood: differential equation



$q(t)$: probability density that a given individual at time t produces the descending edge.

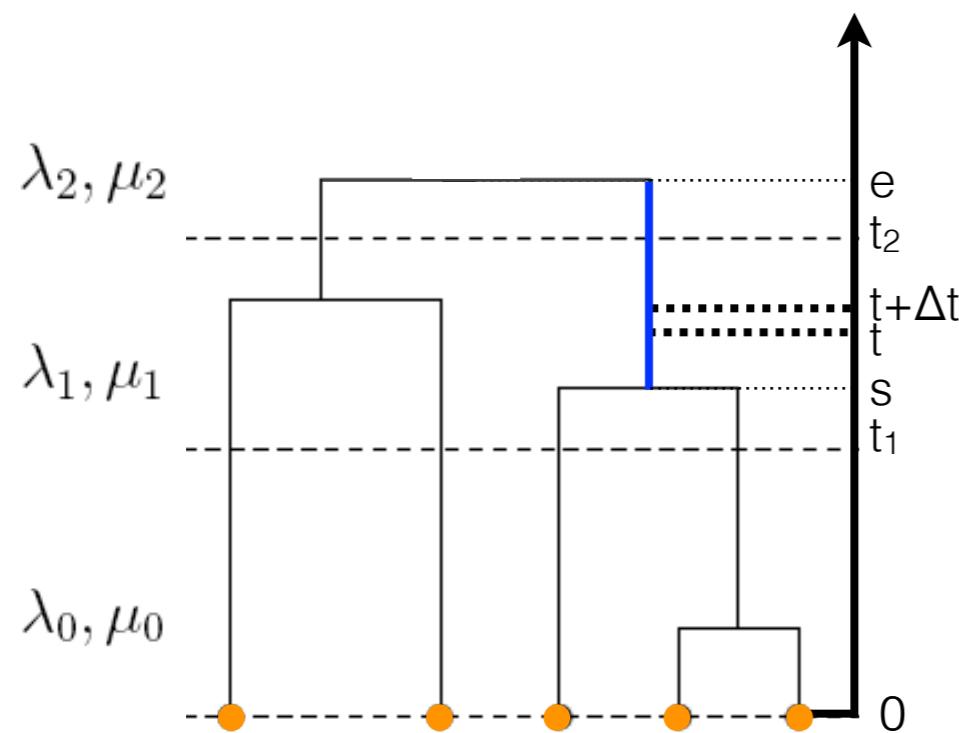
$$q(e) = \text{Prob} (|)$$

$$q(s) = 1$$

$p(t)$: probability that a given individual at time t has no extant and sampled species offspring.

$$q(t + \Delta t) = (1 - (\lambda_i + \mu_i)\Delta t - O(\Delta t^2))q(t) + \lambda_i \Delta t 2q(t)p(t) + O(\Delta t^2)$$

Calculating the likelihood: differential equation



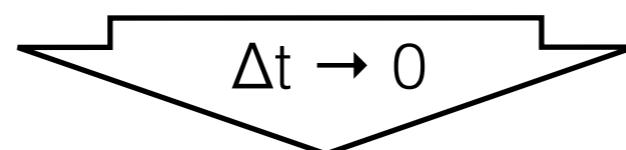
$q(t)$: probability density that a given individual at time t produces the descending edge.

$$q(e) = \text{Prob} (|)$$

$$q(s) = 1$$

$p(t)$: probability that a given individual at time t has no extant and sampled species offspring.

$$q(t + \Delta t) = (1 - (\lambda_i + \mu_i)\Delta t - O(\Delta t^2))q(t) + \lambda_i \Delta t 2q(t)p(t) + O(\Delta t^2)$$



$$\frac{d}{dt}q(t) = -(\lambda_i + \mu_i)q(t) + 2\lambda_i q(t)p(t)$$

Likelihood formula

$$\text{Lik} \left(\begin{array}{c} \text{---} \\ | \\ \text{---} \\ | \\ \text{---} \\ | \\ \text{---} \end{array} \right) = \rho^n (\lambda_0 - \mu_0)^{2n} \frac{q_{l(x_1)}(x_1)^2}{(1 - p_{l(x_1)}(x_1))^2} \prod_{i=2}^{n-1} \lambda_{l(x_i)} q_{l(x_i)}(x_i) \prod_{i=1}^m (\lambda_i - \mu_i)^{2n_i} q_{i-1}(t_i)^{n_i}$$

...with the x_i being the branching times and $p(x_i), q(x_i), l(x_i)$

being known functions depending on the parameters λ, μ, ρ, t

Likelihood formula

$$\text{Lik} \left(\begin{array}{c} \text{---} \\ | \\ \text{---} \\ | \\ \text{---} \\ | \\ \text{---} \end{array} \right) = \rho^n (\lambda_0 - \mu_0)^{2n} \frac{q_{l(x_1)}(x_1)^2}{(1 - p_{l(x_1)}(x_1))^2} \prod_{i=2}^{n-1} \lambda_{l(x_i)} q_{l(x_i)}(x_i) \prod_{i=1}^m (\lambda_i - \mu_i)^{2n_i} q_{i-1}(t_i)^{n_i}$$

Likelihood function used for maximum likelihood inference

...with the x_i being the branching times and $p(x_i)$, $q(x_i)$, $\lambda(x_i)$

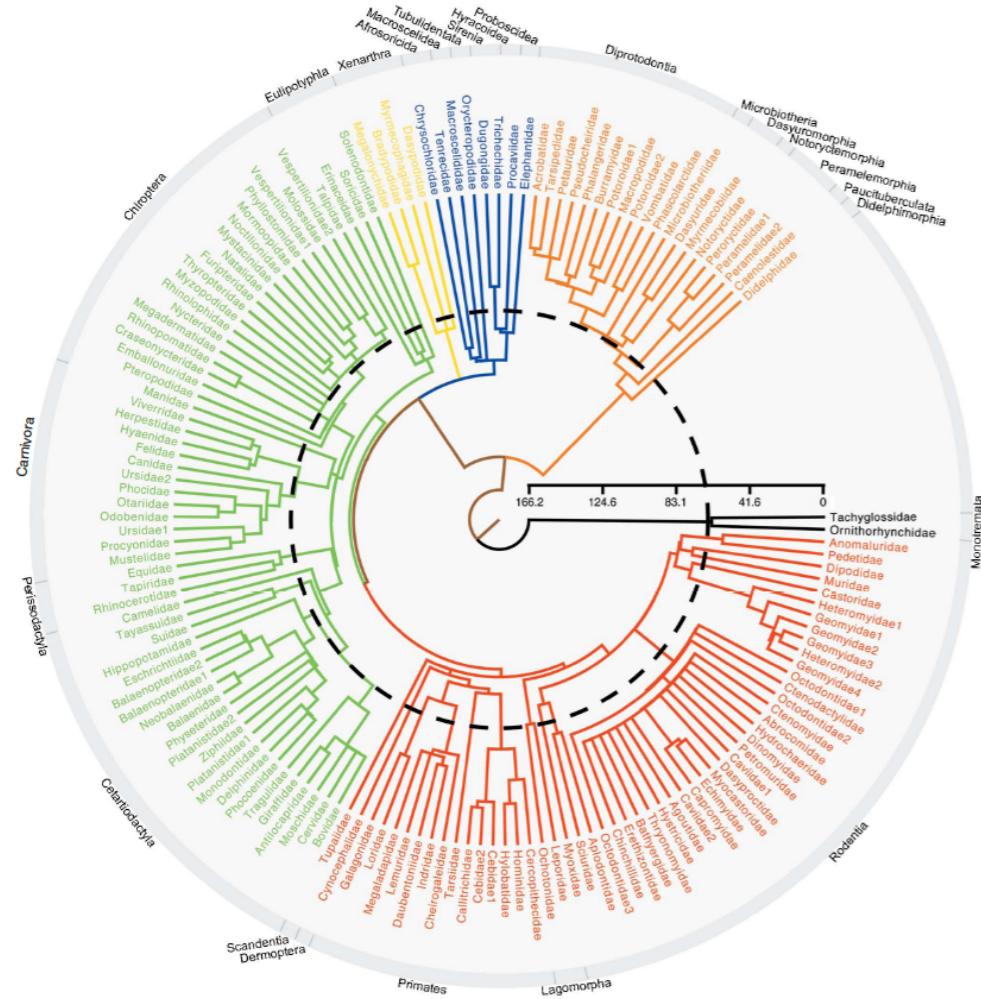
being known functions depending on the parameters λ, μ, ρ, t

Likelihood formula

$$f \left(\begin{array}{c} \text{Species Tree Diagram} \\ \text{(A-B-C-D-E)} \end{array} \right) = \rho^n (\lambda_0 - \mu_0)^{2n} \frac{q_{l(x_1)}(x_1)^2}{(1 - p_{l(x_1)}(x_1))^2} \prod_{i=2}^{n-1} \lambda_{l(x_i)} q_{l(x_i)}(x_i) \prod_{i=1}^m (\lambda_i - \mu_i)^{2n_i} q_{i-1}(t_i)^{n_i}$$

- ▶ The maximum likelihood parameters for a species tree are estimated for $m = 0, 1, 2, \dots$ rate shifts.
- ▶ The most likely number of shifts is determined using likelihood ratio tests.
- ▶ Method implemented in R library TreePar (available on CRAN).

Species phylogeny rejects increased mammalian diversification at KT-boundary

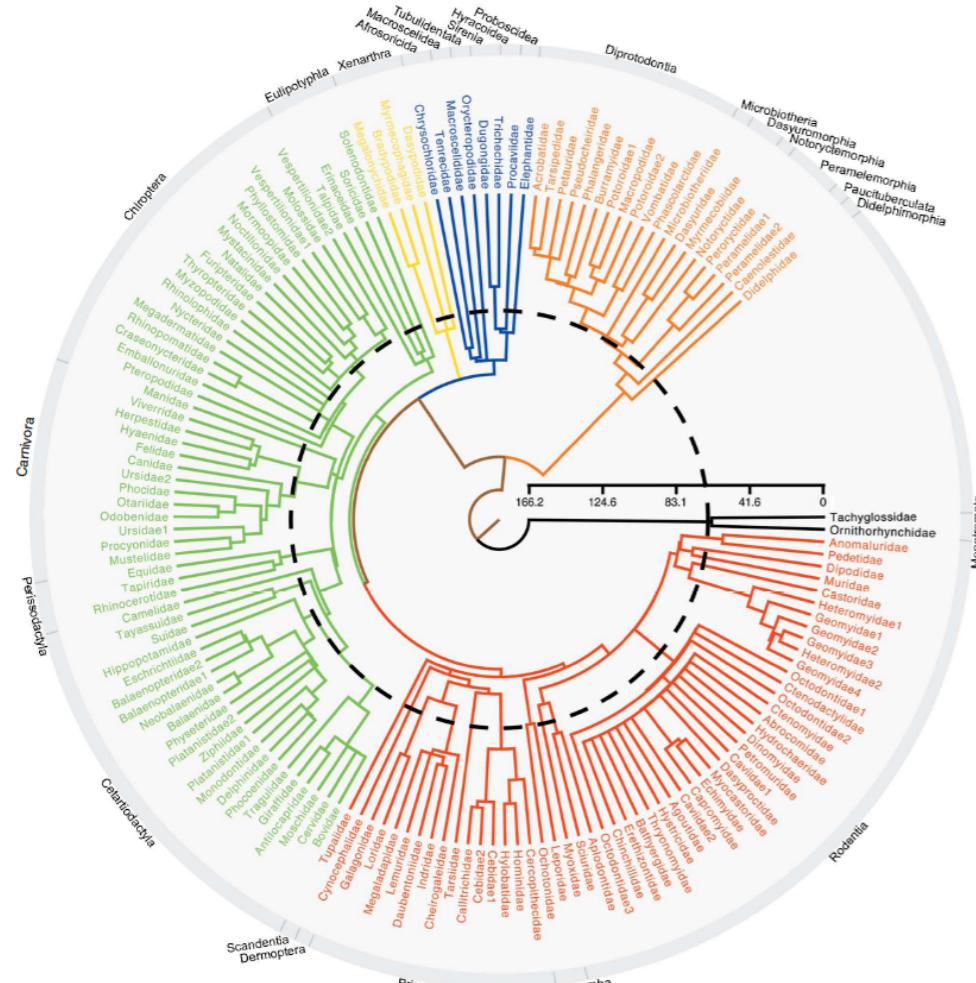


Mammalian phylogeny

(from Bininda et al., Nature, 2007)

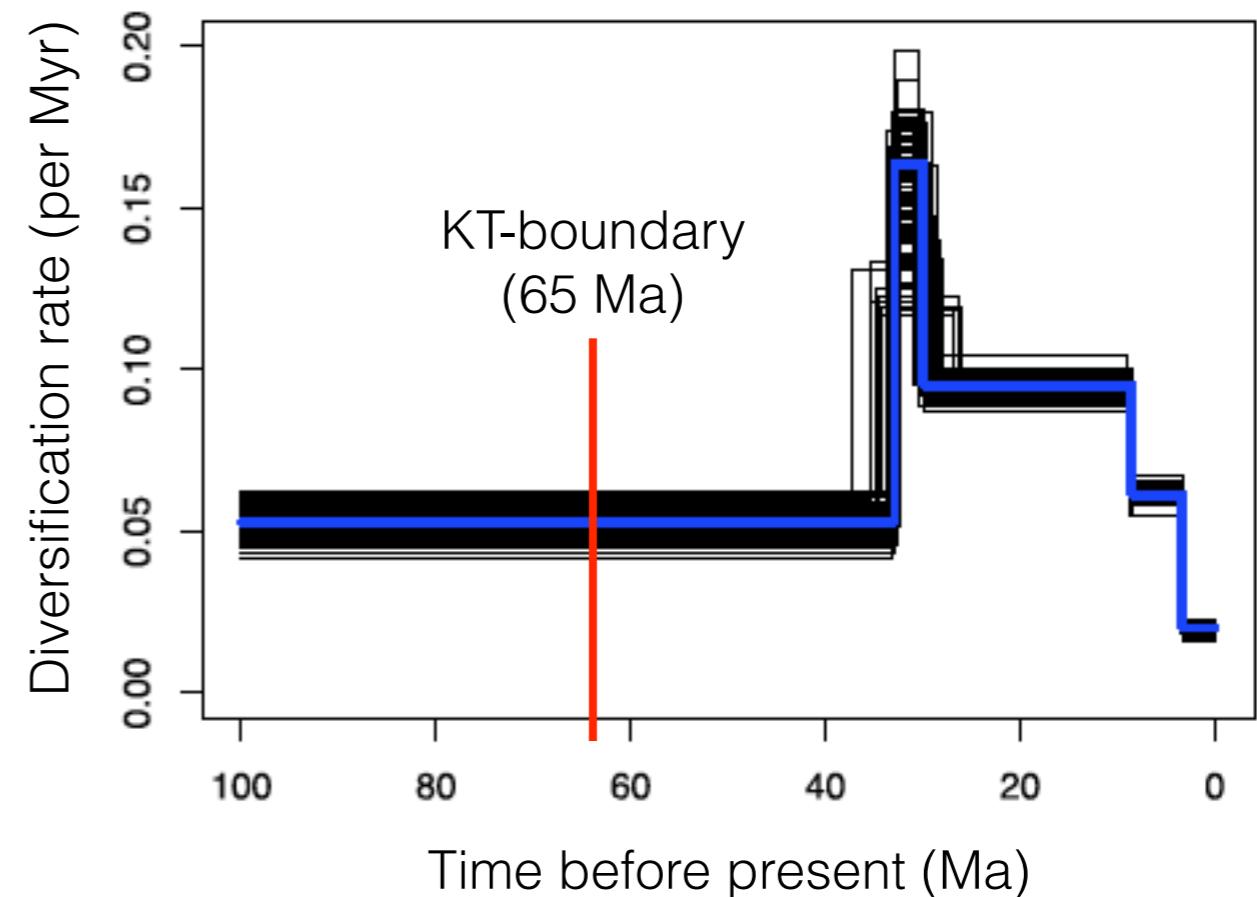
Species phylogeny rejects increased mammalian diversification at KT-boundary

(diversification = speciation λ - extinction μ)



Mammalian phylogeny

(from Bininda et al., Nature, 2007)

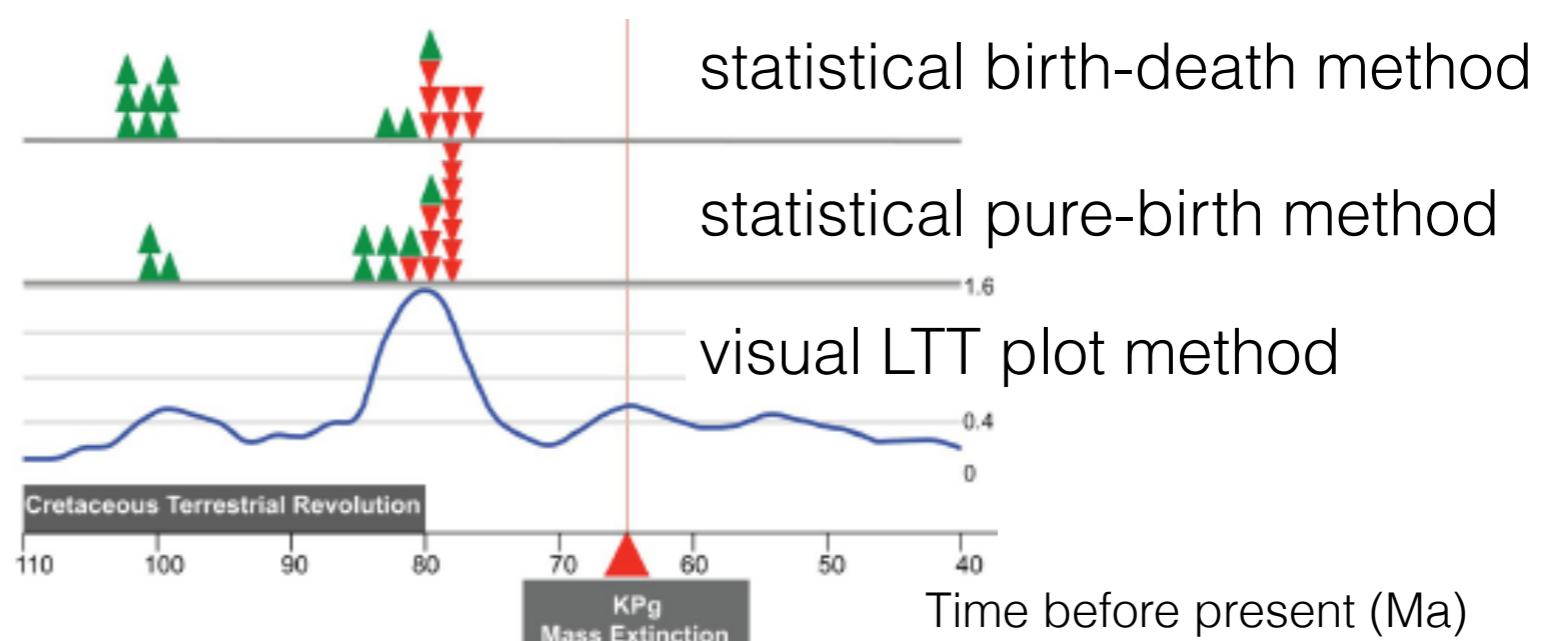
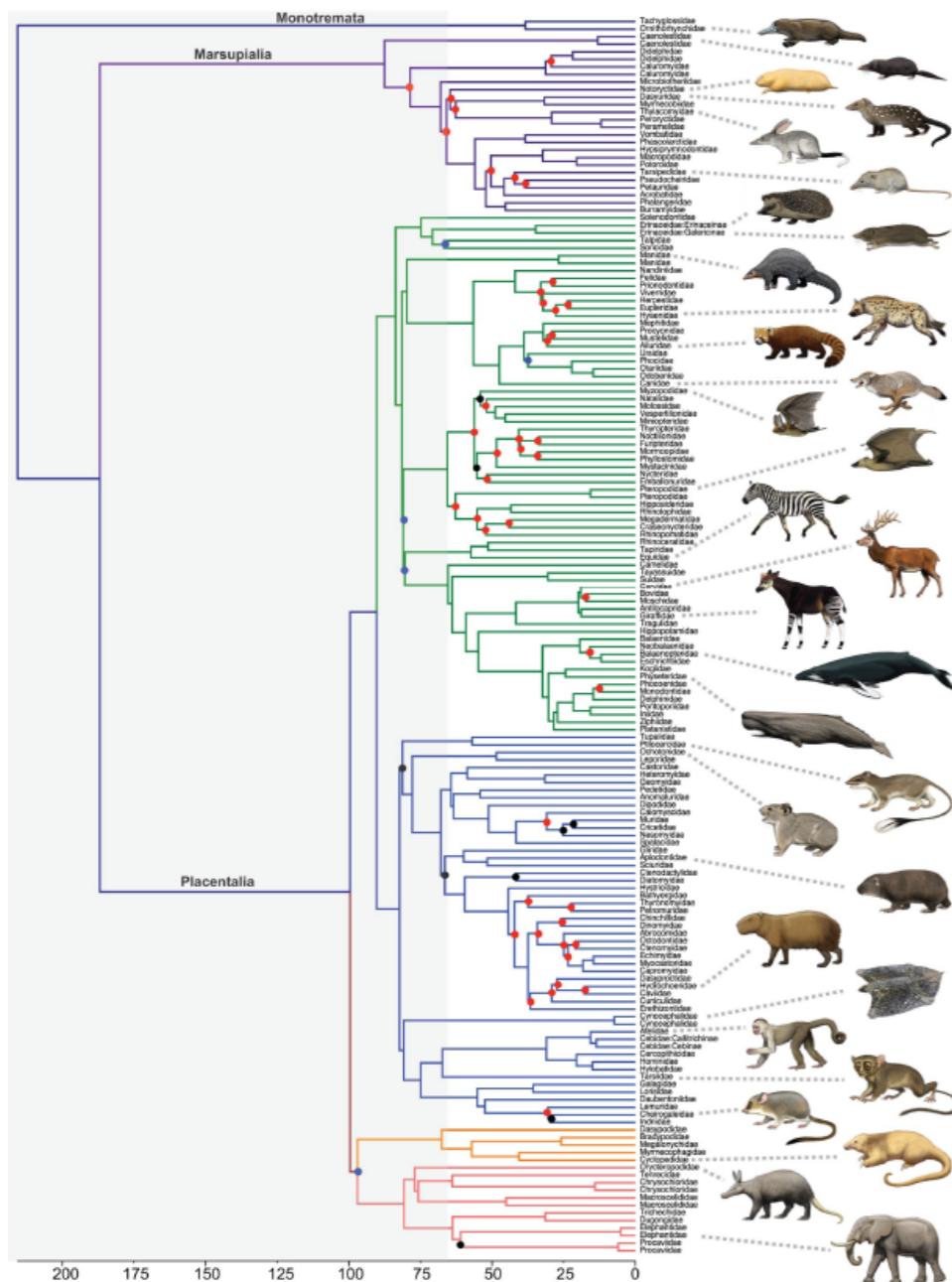


Null model: constant rates bad fit

Environmental model: 4 shifts ($p=1.0$)

Simulation results

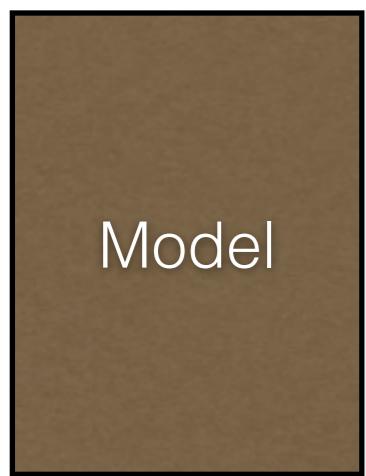
Family phylogeny rejects increased mammalian diversification at KT-boundary



Green arrow: increase in diversification rate

Red arrow: decrease in diversification rate

2) Modeling diversity-dependent effects

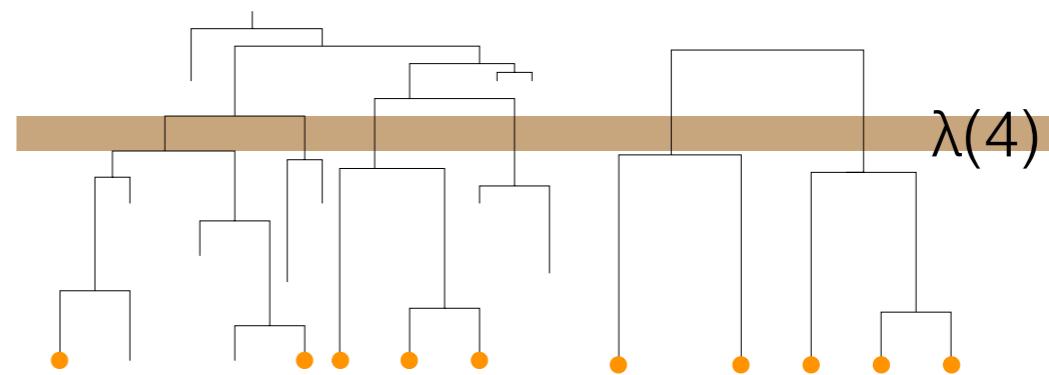


N: # of species

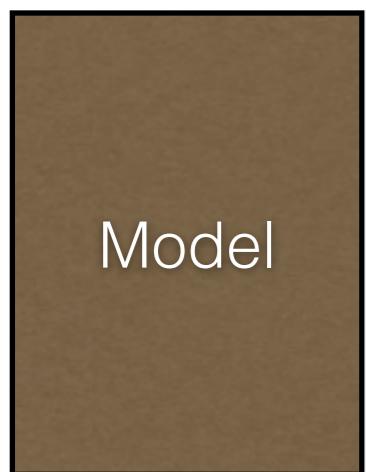
Model

K: carrying capacity

$$\lambda(N) = \lambda_{\text{const}} (1-N/K)$$



2) Modeling diversity-dependent effects

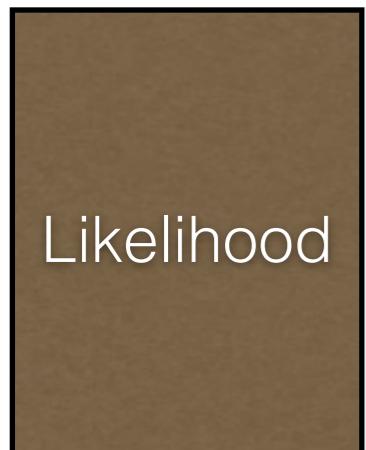
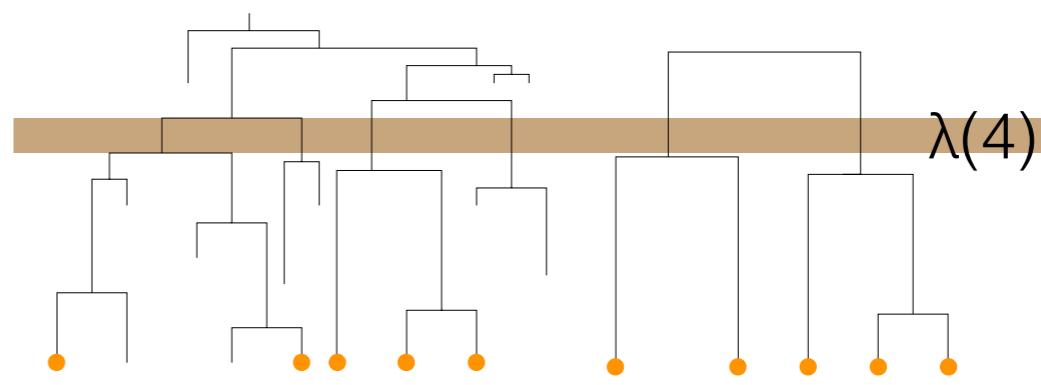


Model

N: # of species

K: carrying capacity

$$\lambda(N) = \lambda_{\text{const}} (1 - N/K)$$



Likelihood

$$\text{Lik} \left(\begin{array}{c} \text{phylogenetic tree} \\ \text{with orange dots at tips} \end{array} \right) = \prod_{i=2}^n \left(L e^{M_i(x_{i-1} - x_i)} \right) e_n$$

with L and M_i matrices of dimension K, e_n vector.

Computational challenges calculating the likelihood

Likelihood

$$\text{Lik} \left(\begin{array}{c} \text{---} \\ | \\ \text{---} \\ | \\ \text{---} \\ | \\ \text{---} \end{array} \right) = \prod_{i=2}^n \left(L e^{M_i(x_{i-1} - x_i)} \right) e_n$$

with L and M_i matrices of dimension K , e_n vector.

Numerical
method

Evaluating vectors applied to matrix exponentials

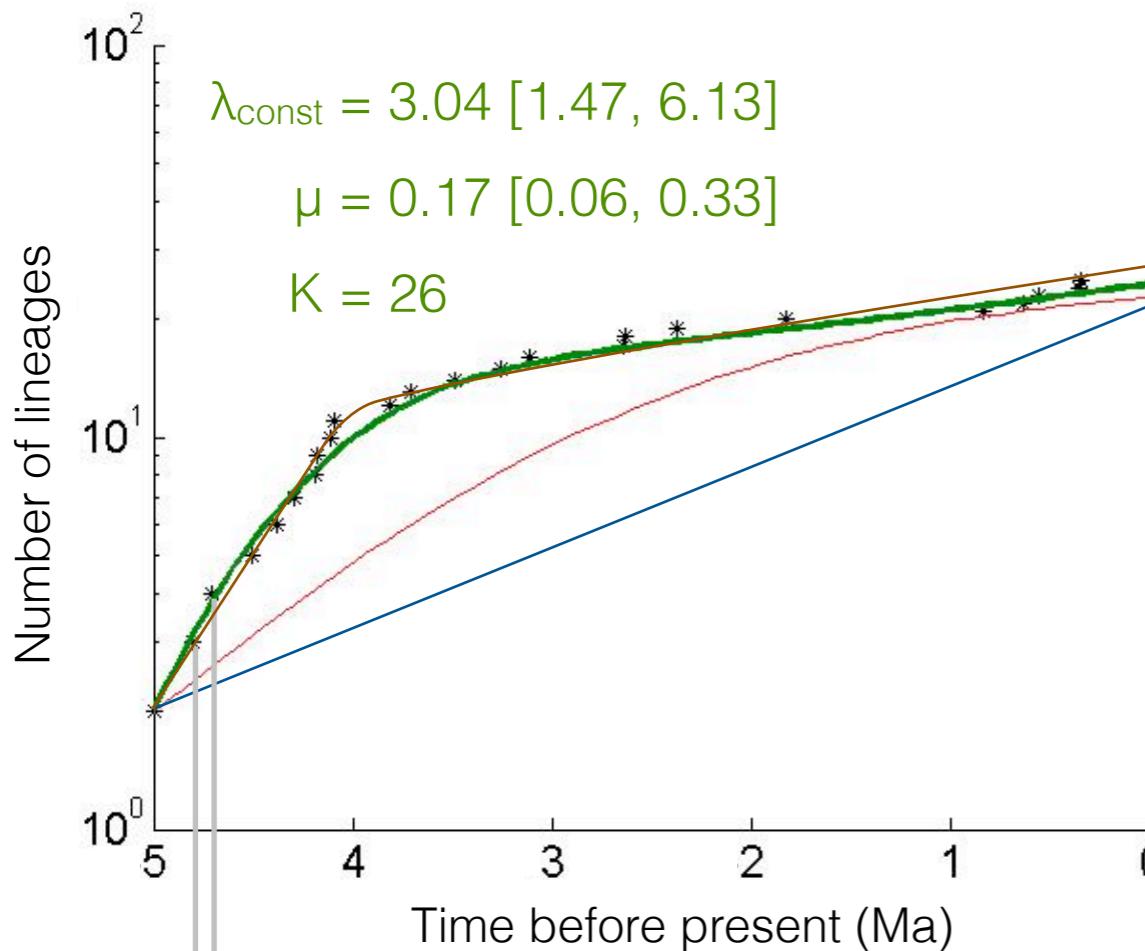
Standard methods:

all inaccurate or too slow

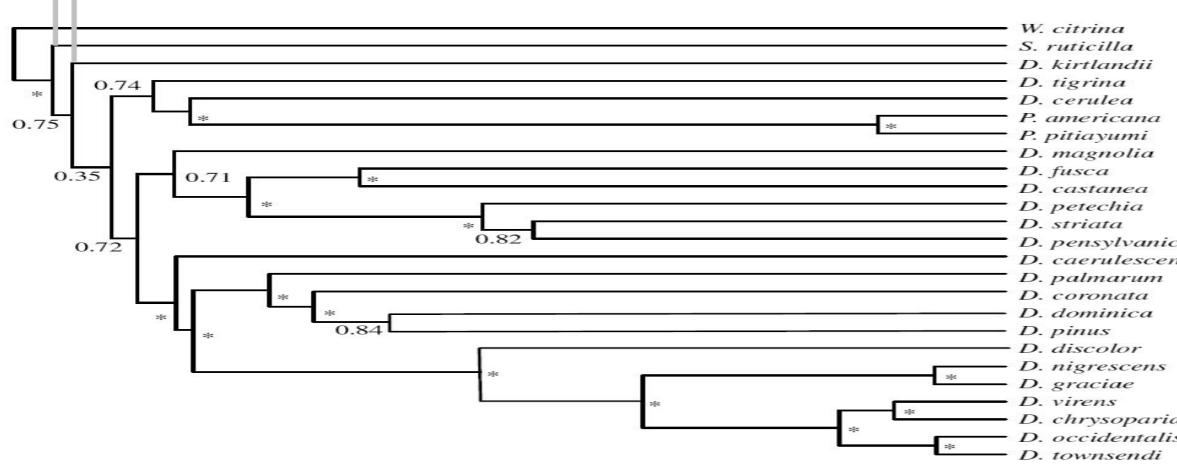
Al-Mohy & Higham (2011):

our method of choice

Dendroica warbler phylogeny



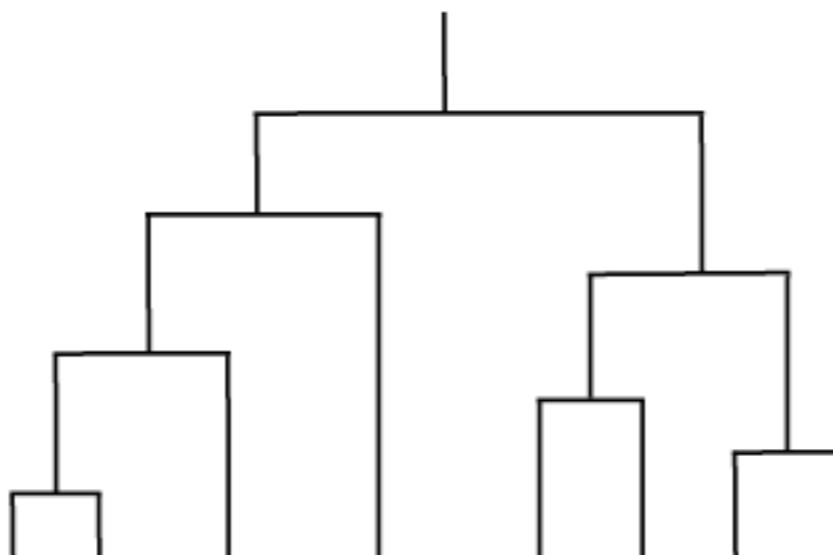
* empirical data
our diversity-dependence model
our environmental model ($\Delta \text{AIC} = 5.93$)
density-dependence w/o extinction ($p = 0.97$)
constant rates model ($p = 1.00$)



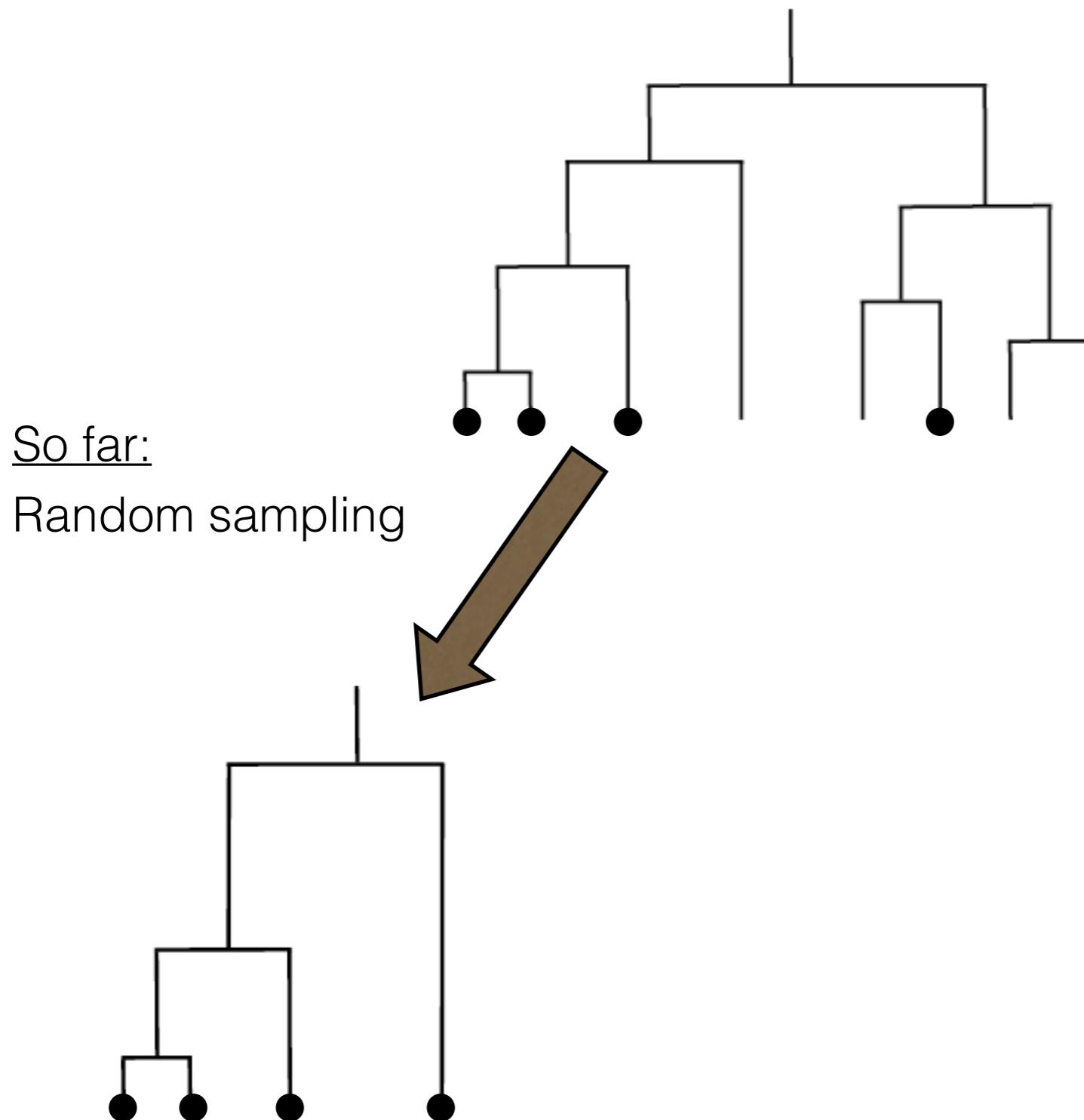
Warbler phylogeny

(from Rabosky et al. 2008, Proc. Roy. Soc. B)

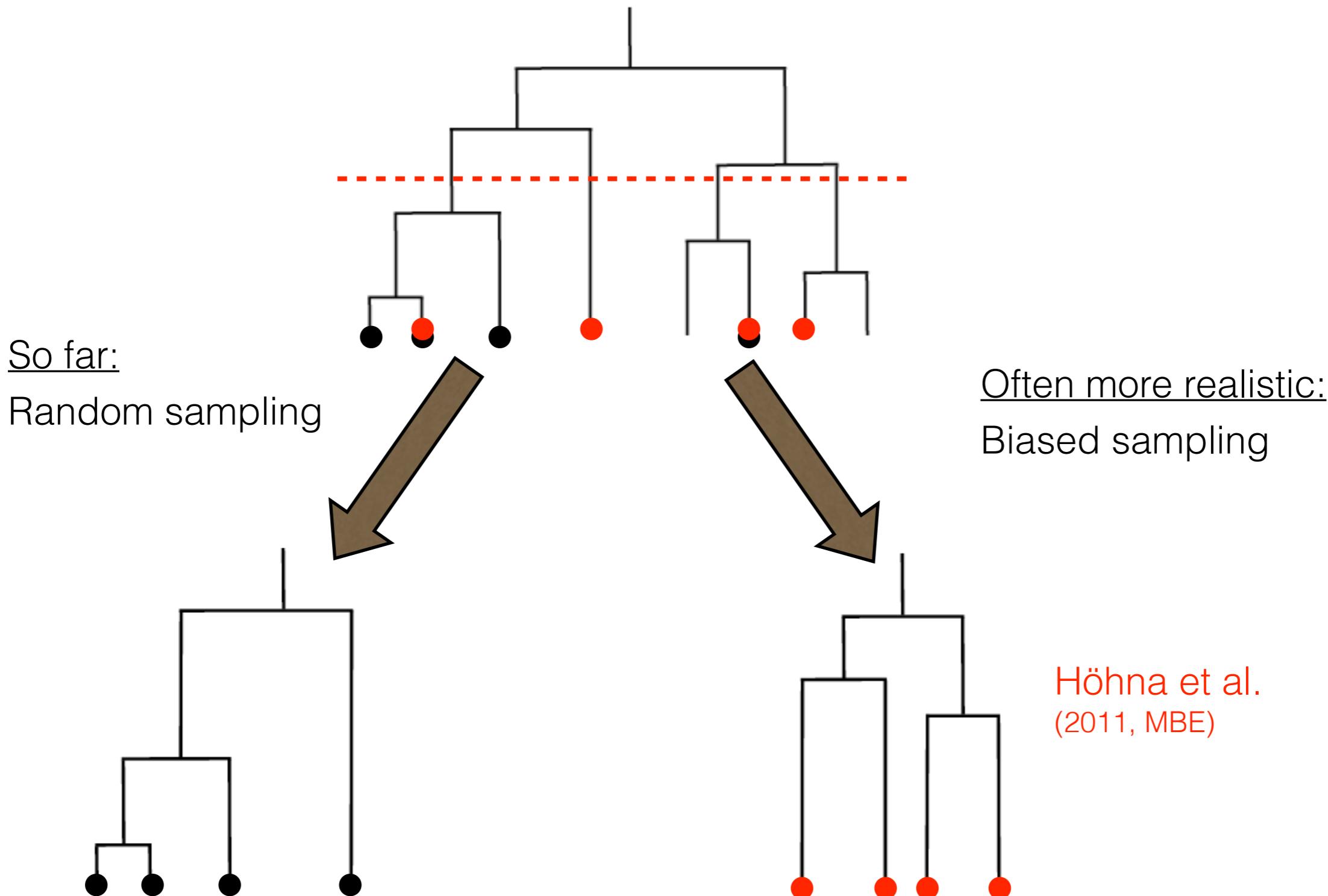
Accounting for missing species



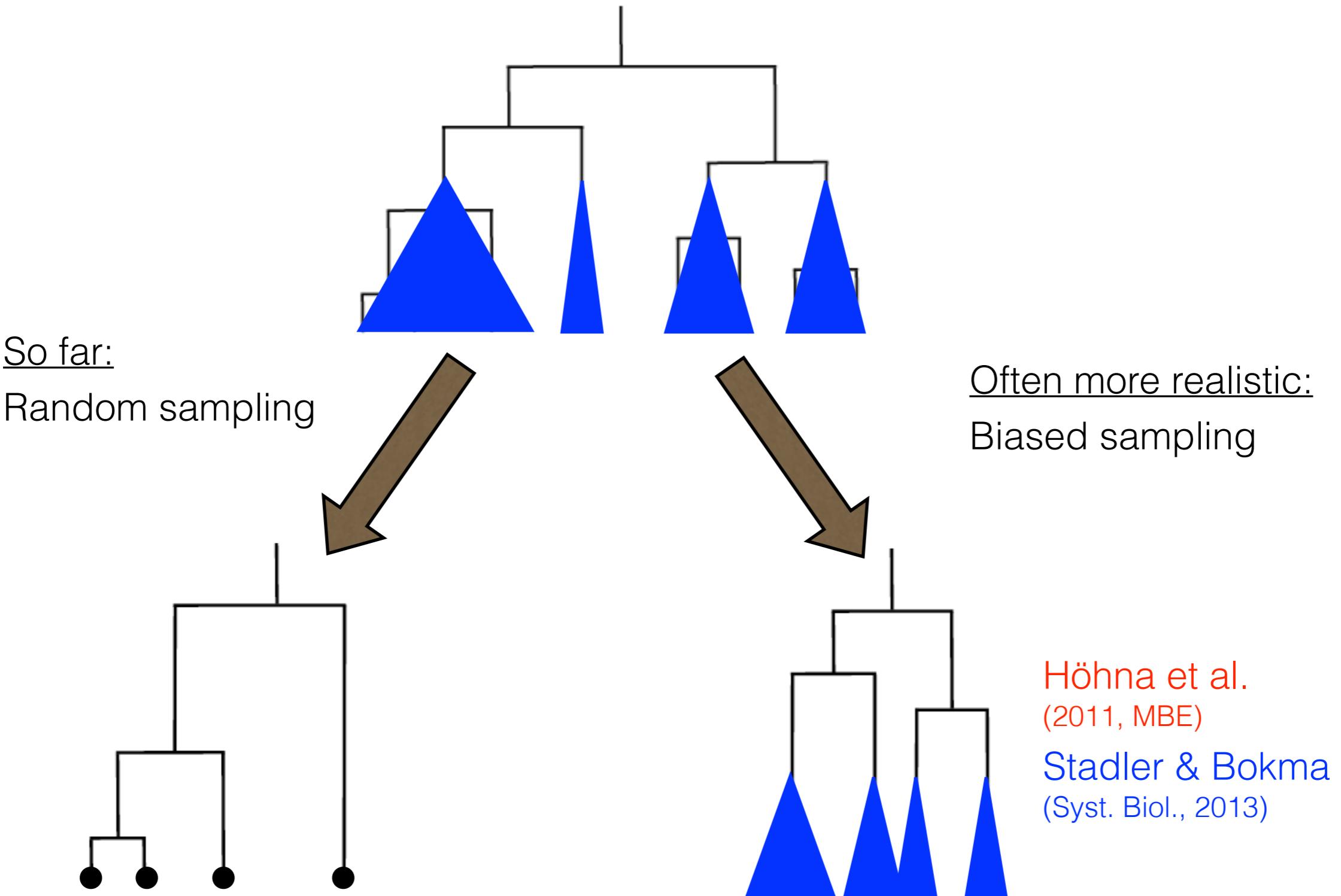
Accounting for missing species



Accounting for missing species



Accounting for missing species



Phylogenetics in R

Libraries

- ▶ **ape** - Manipulating phylogenetic trees (*E. Paradis*)
- ▶ **TreeSim & TreeSimGM** - Simulating phylogenetic trees (*T. Stadler, O. Hagen*)

- ▶ **Laser, BAMM** - Clade-dependent diversification (*D. Rabosky*)
- ▶ **DDD** - Diversity-dependent diversification (*R. Etienne*)
- ▶ **TreePar** - Time & trait dependent diversification (*T. Stadler*)
- ▶ **diversitree** - Trait-dependent diversification (*R. Fitzjohn*)
- ▶ **Geiger** - Trait evolution (*L. Harmon, M. Pennell*)

Theoretical
Overview

Review: *T. Stadler, JEB, 2013*

Recovering speciation and extinction dynamics based on phylogenies.

Comparison of approaches: *Stadler, Syst Biol, 2013*

How can we improve accuracy of macroevolutionary rate estimates?