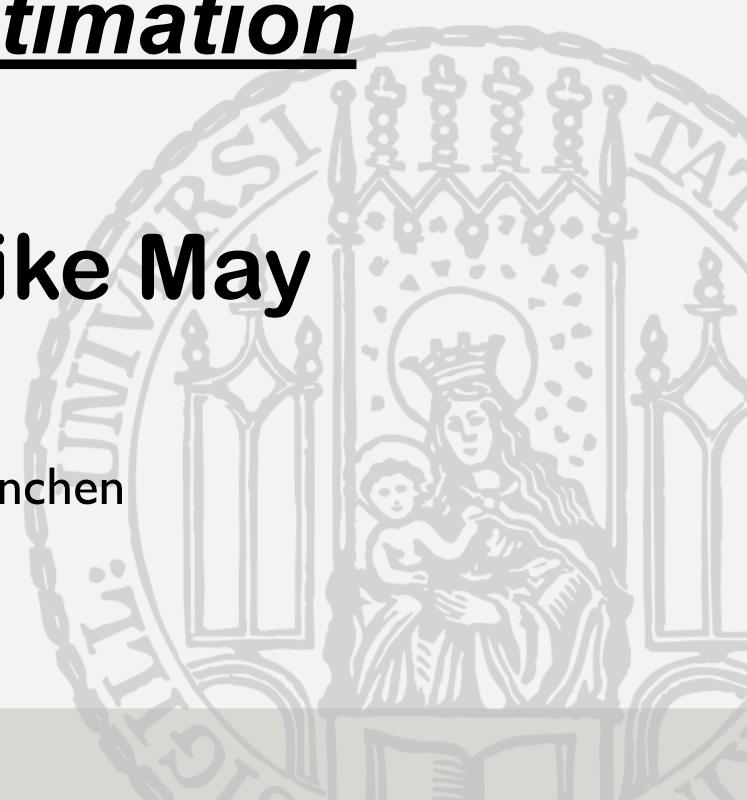


# Bayesian Phylogenetic Inference using RevBayes:

## Diversification Rate Estimation

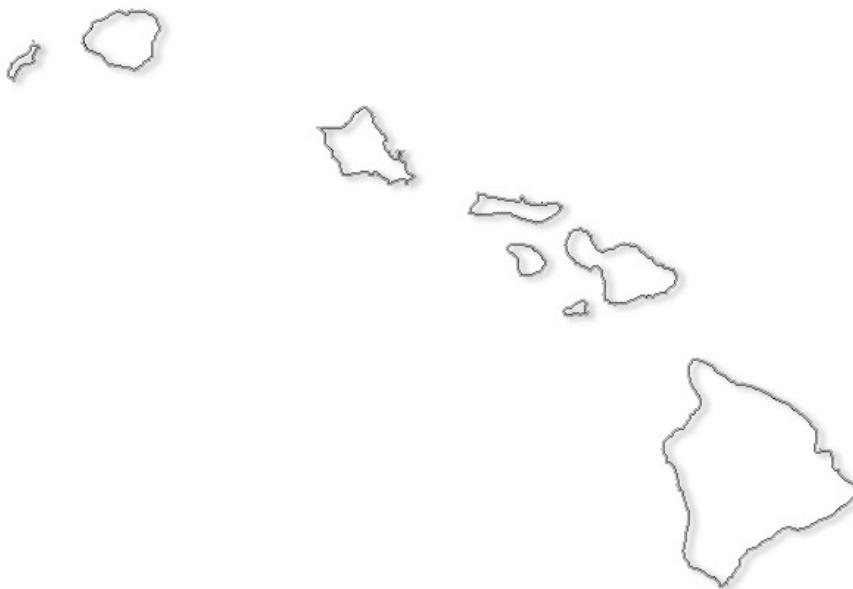
**Sebastian Höhna & Mike May**

GeoBio-Center  
Ludwig-Maximilians Universität, München



# Patterns of diversification

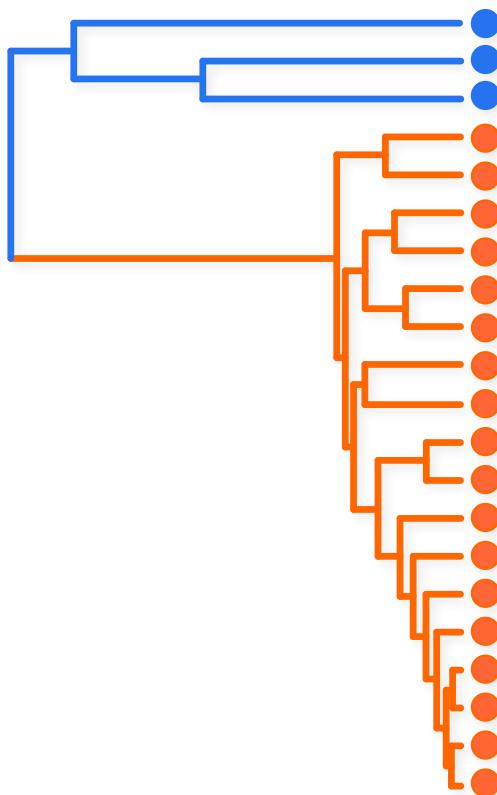
- adaptive radiation



Hawaiian silverswords

# Patterns of diversification

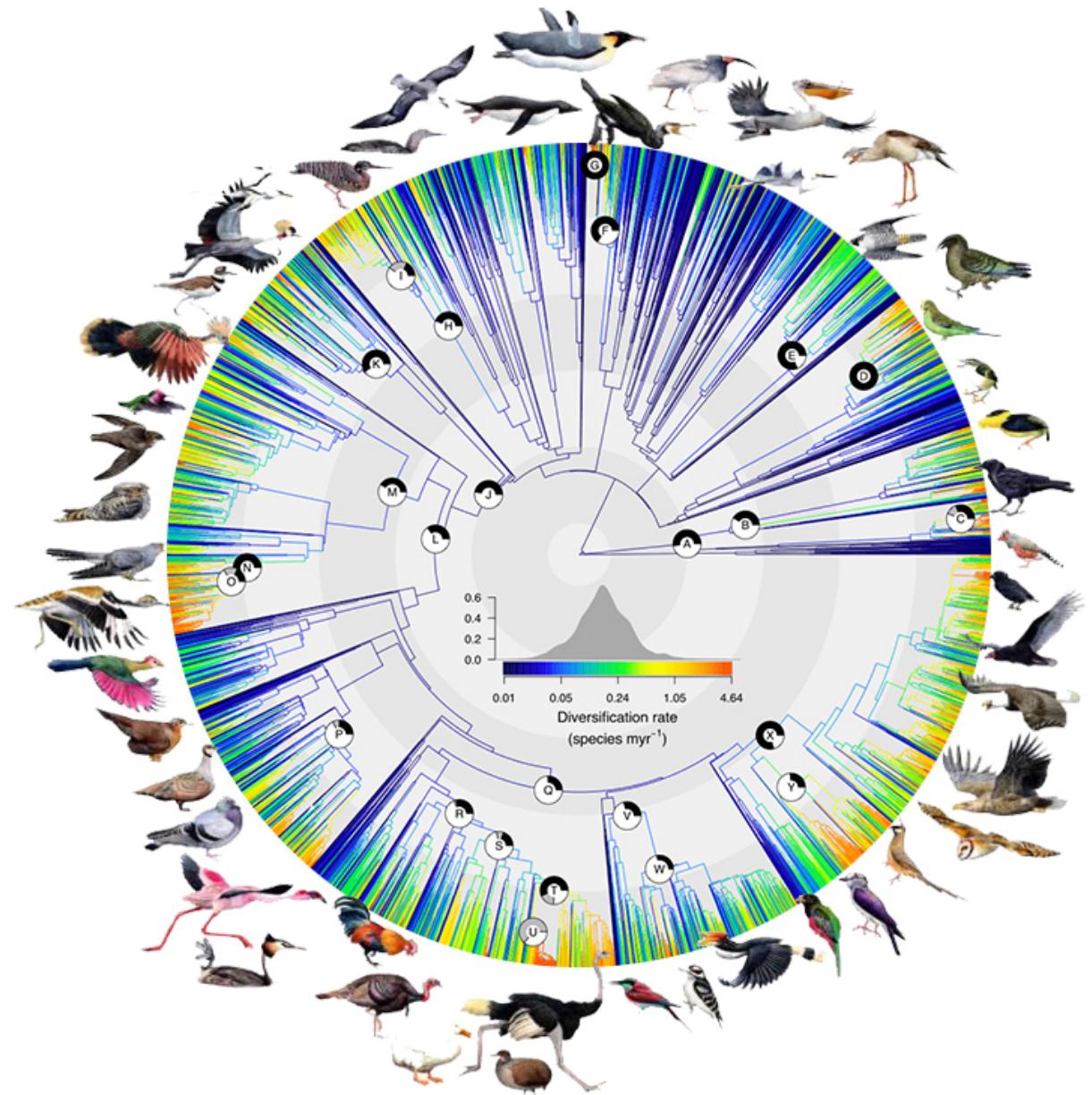
- adaptive radiation
- key innovation



Nectar spurs in columbines

# Patterns of diversification

- adaptive radiation
- key innovation
- diversity dependence



Global deceleration in birds

# Patterns of diversification

---

- adaptive radiation
- key innovation
- diversity dependence
- mass extinction



K-Pg bolide impact

## **1) Estimating constant rates of speciation and extinction**

- What are the speciation and extinction rates for my study group?

## **2) Estimating time-varying rates of speciation and extinction**

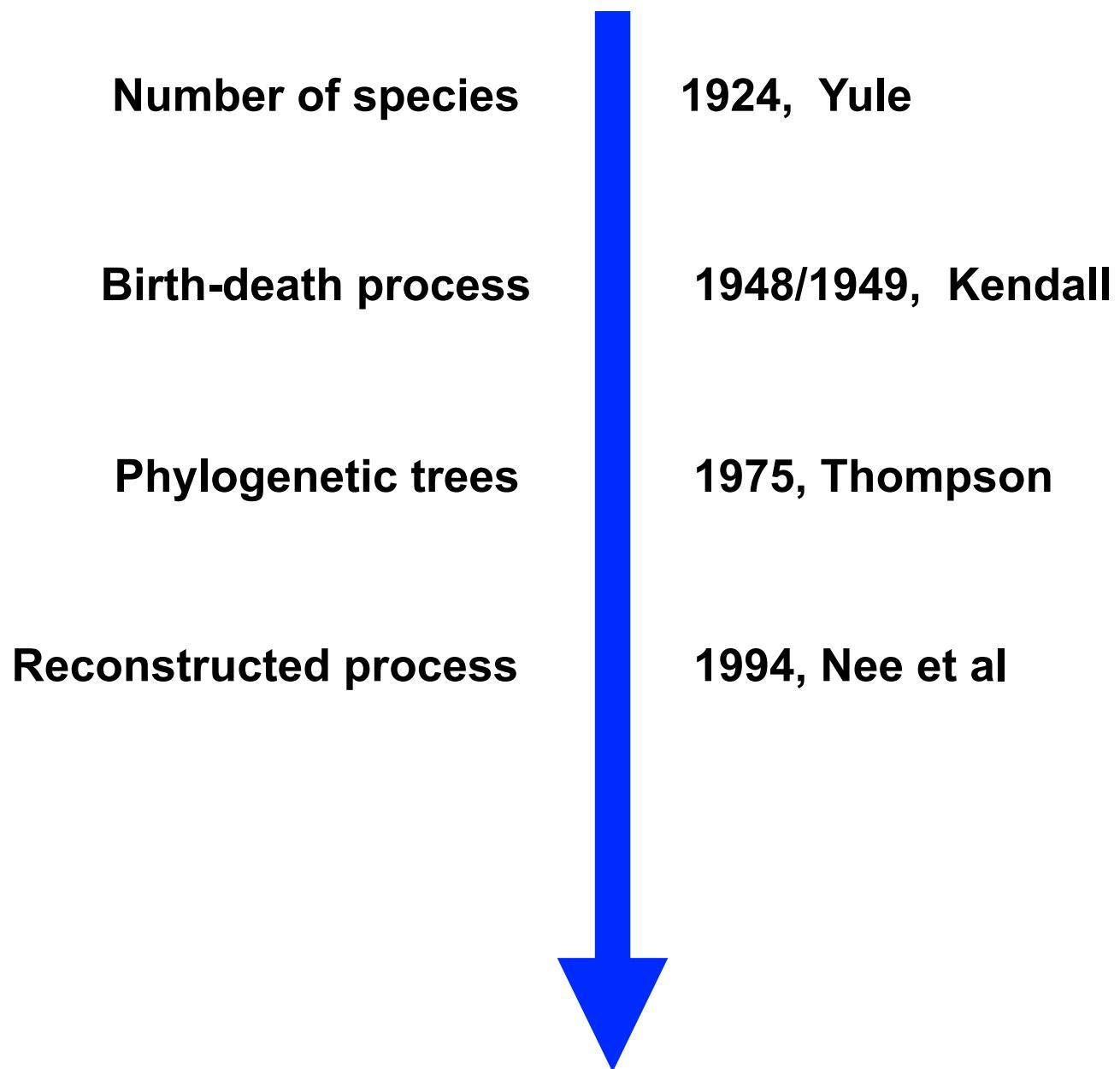
- Do speciation and extinction rates vary through time?
- Is this variation through time correlated with some abiotic factor?

## **3) Estimating branch-specific rates of speciation and extinction**

- Do speciation and extinction rates vary among lineages?
- Is this variation among lineages correlated with some biotic factor?

# A brief history

---



# Birth-Death Process

---

|      Origin (single species)

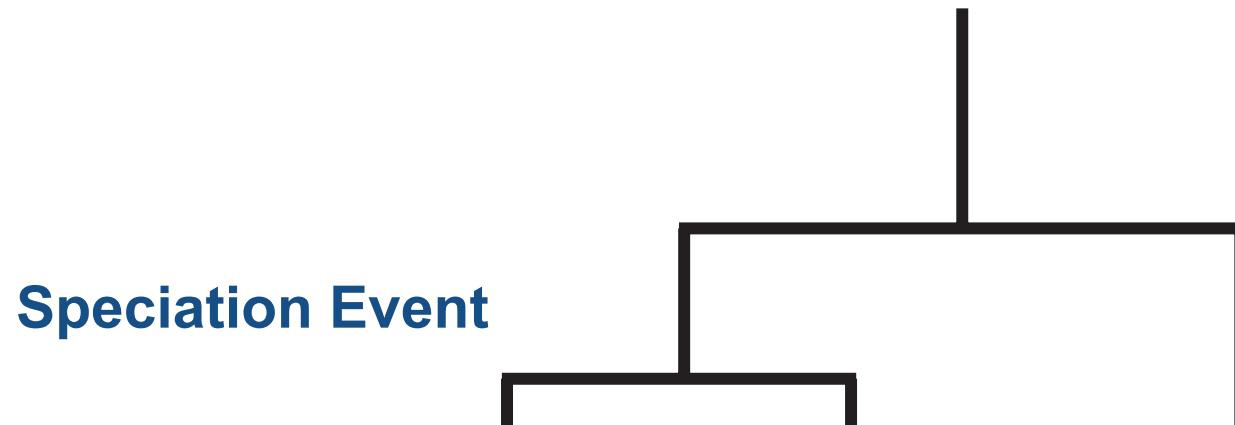
# Birth-Death Process

---



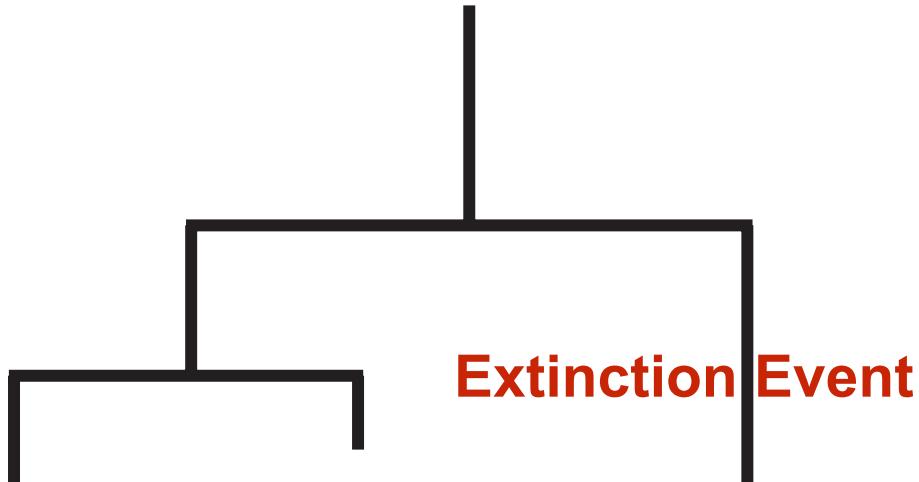
# Birth-Death Process

---



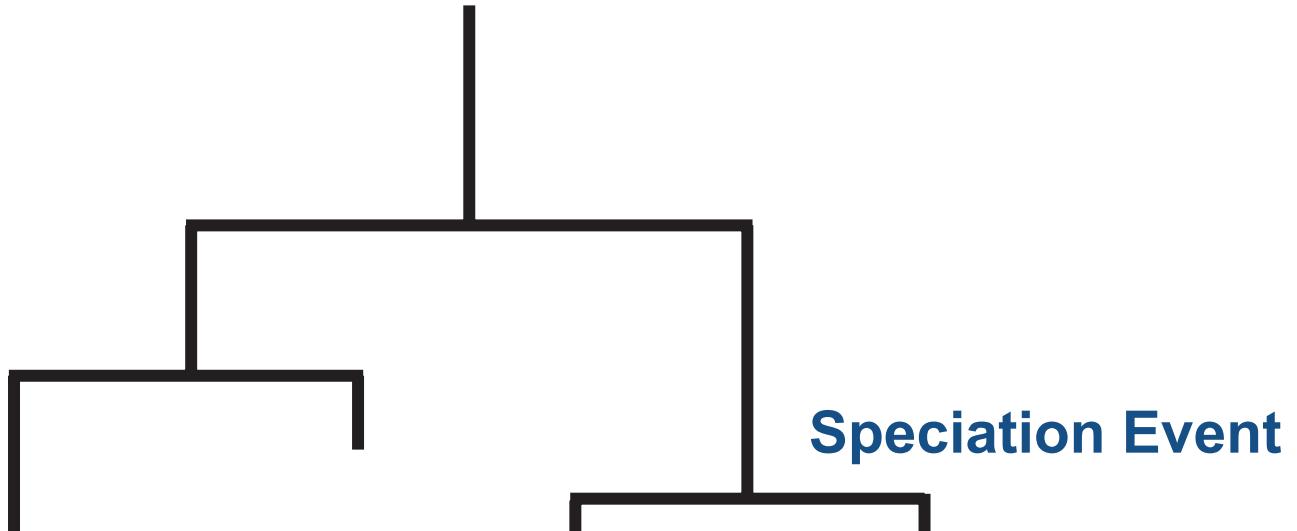
# Birth-Death Process

---



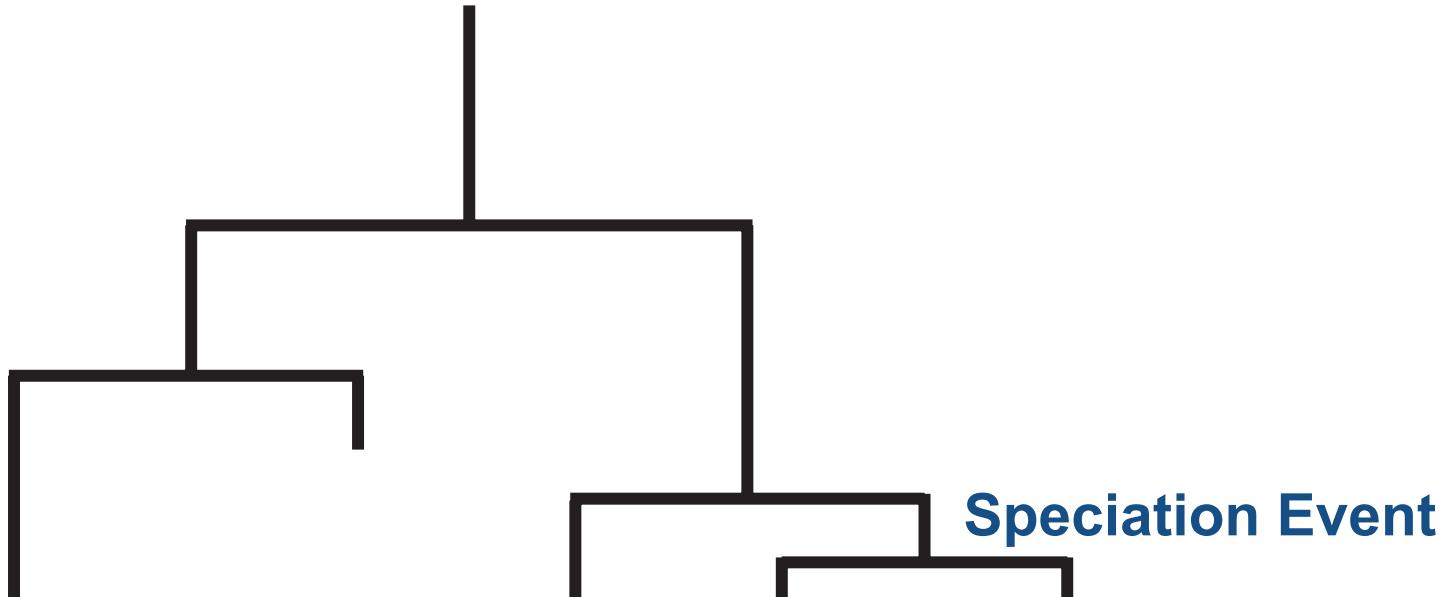
# Birth-Death Process

---



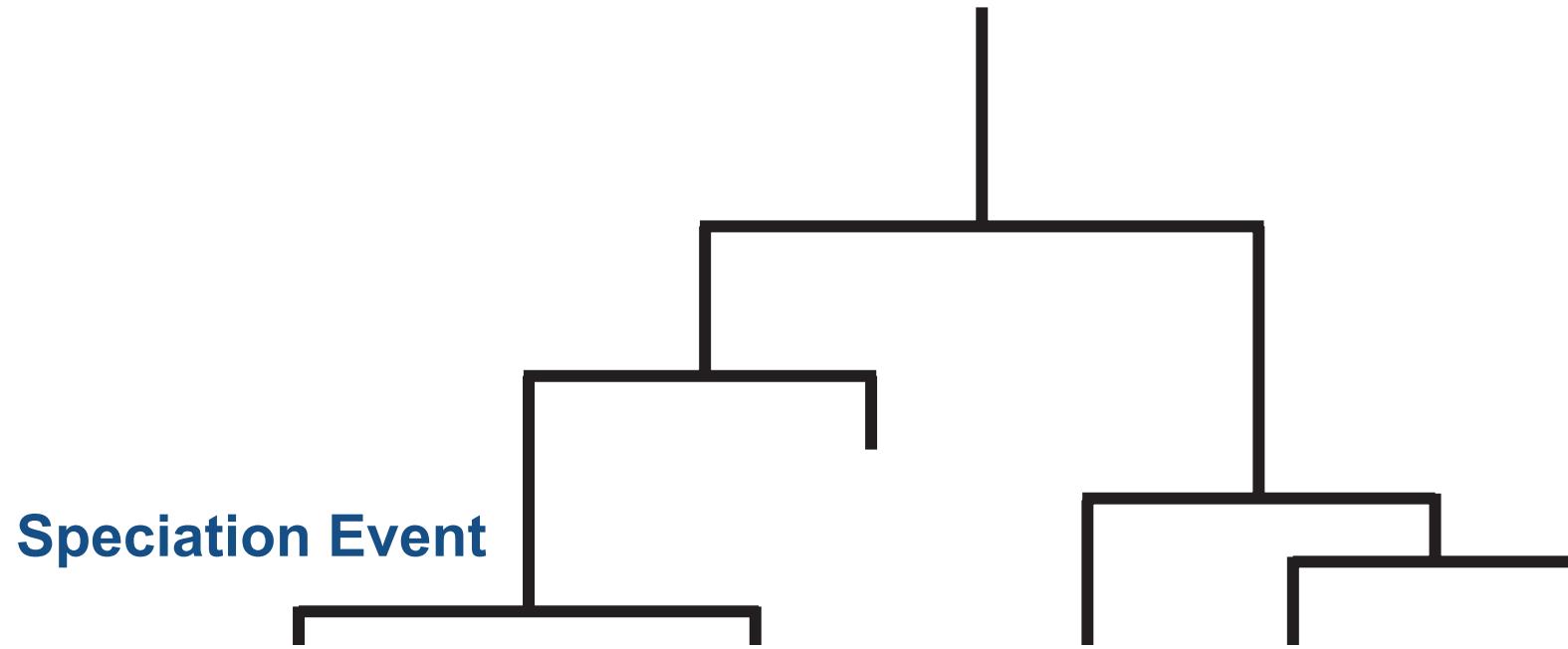
# Birth-Death Process

---



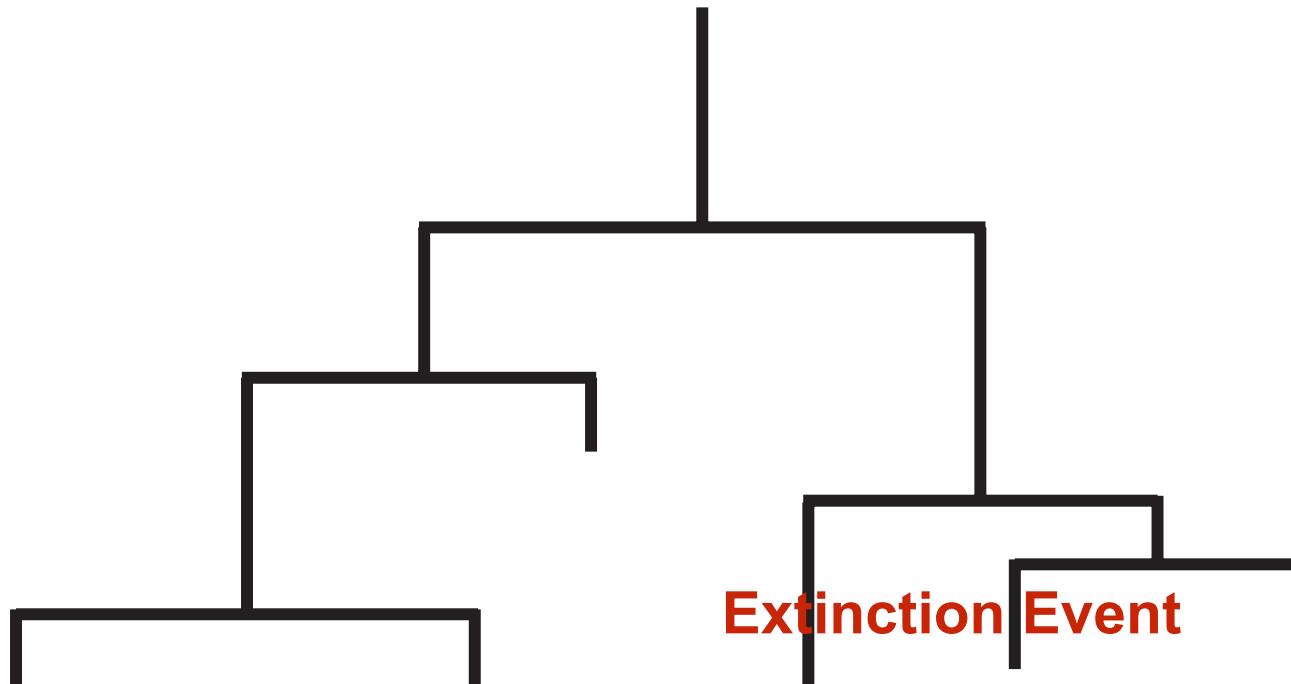
# Birth-Death Process

---



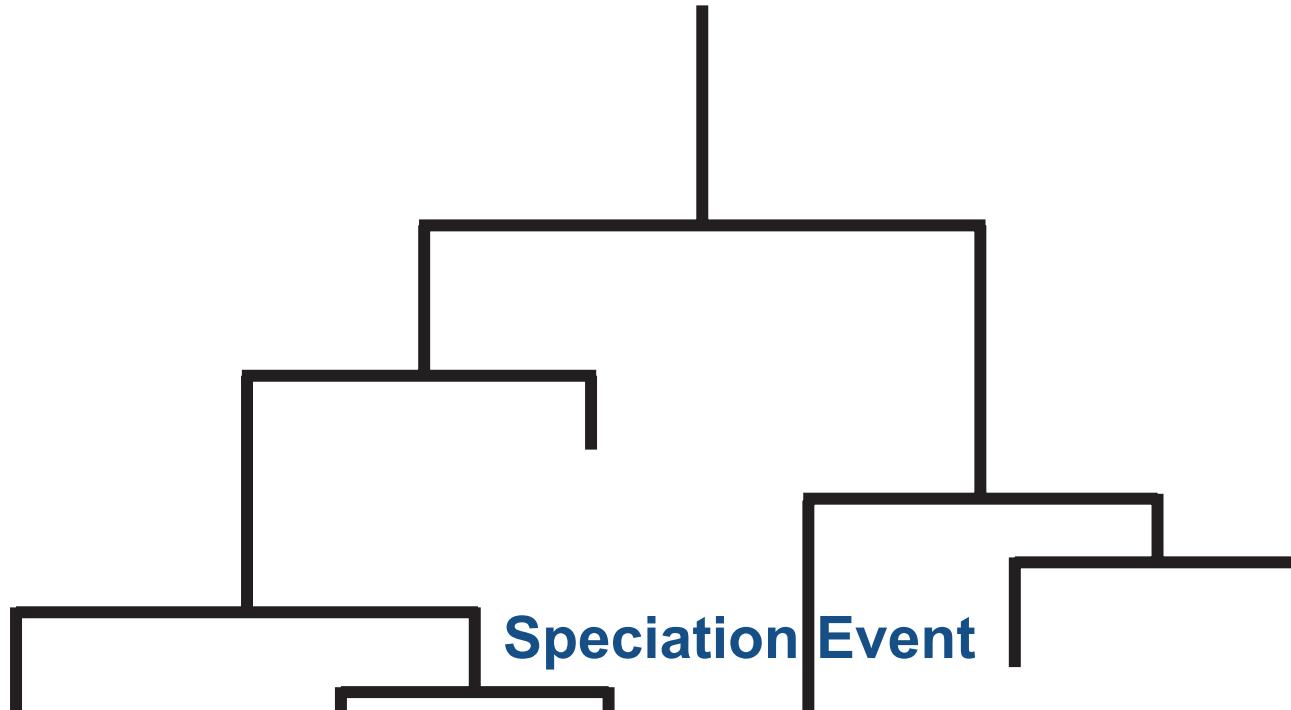
# Birth-Death Process

---



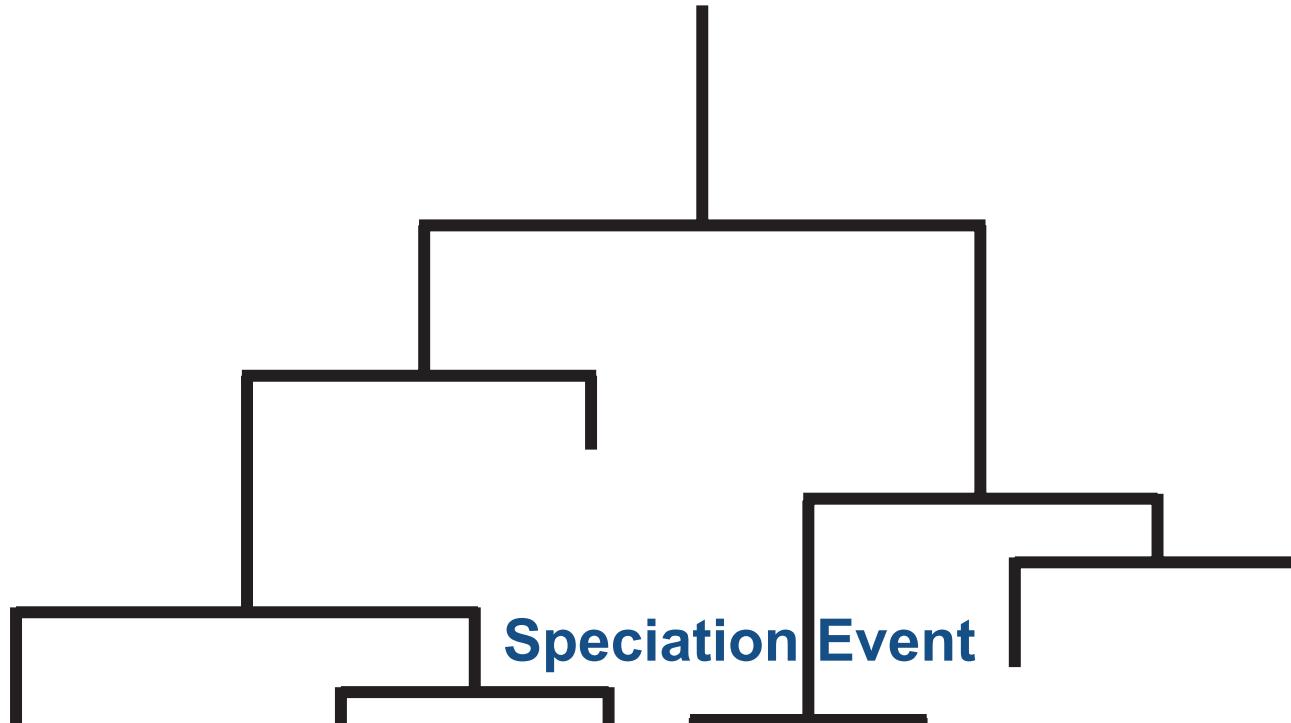
# Birth-Death Process

---



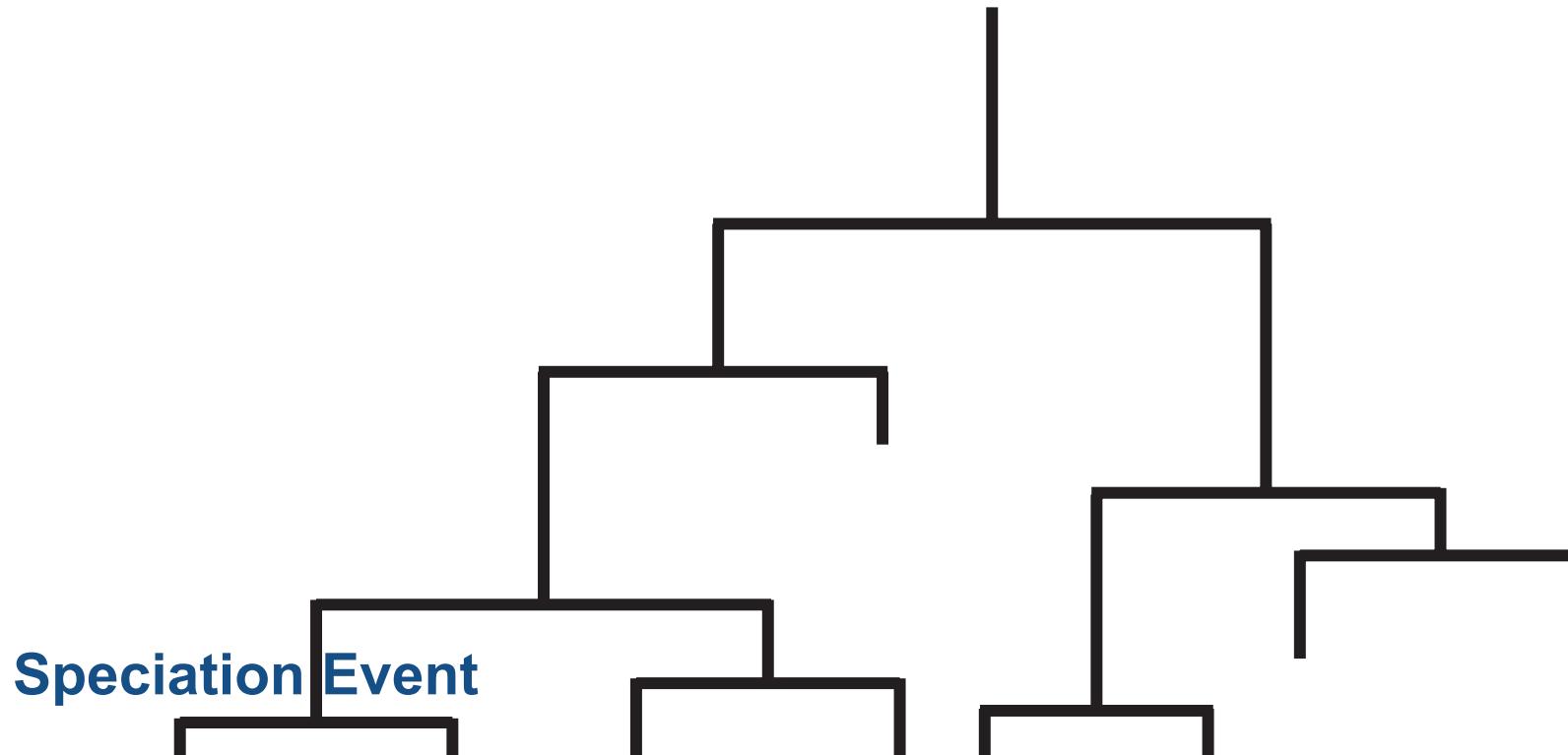
# Birth-Death Process

---



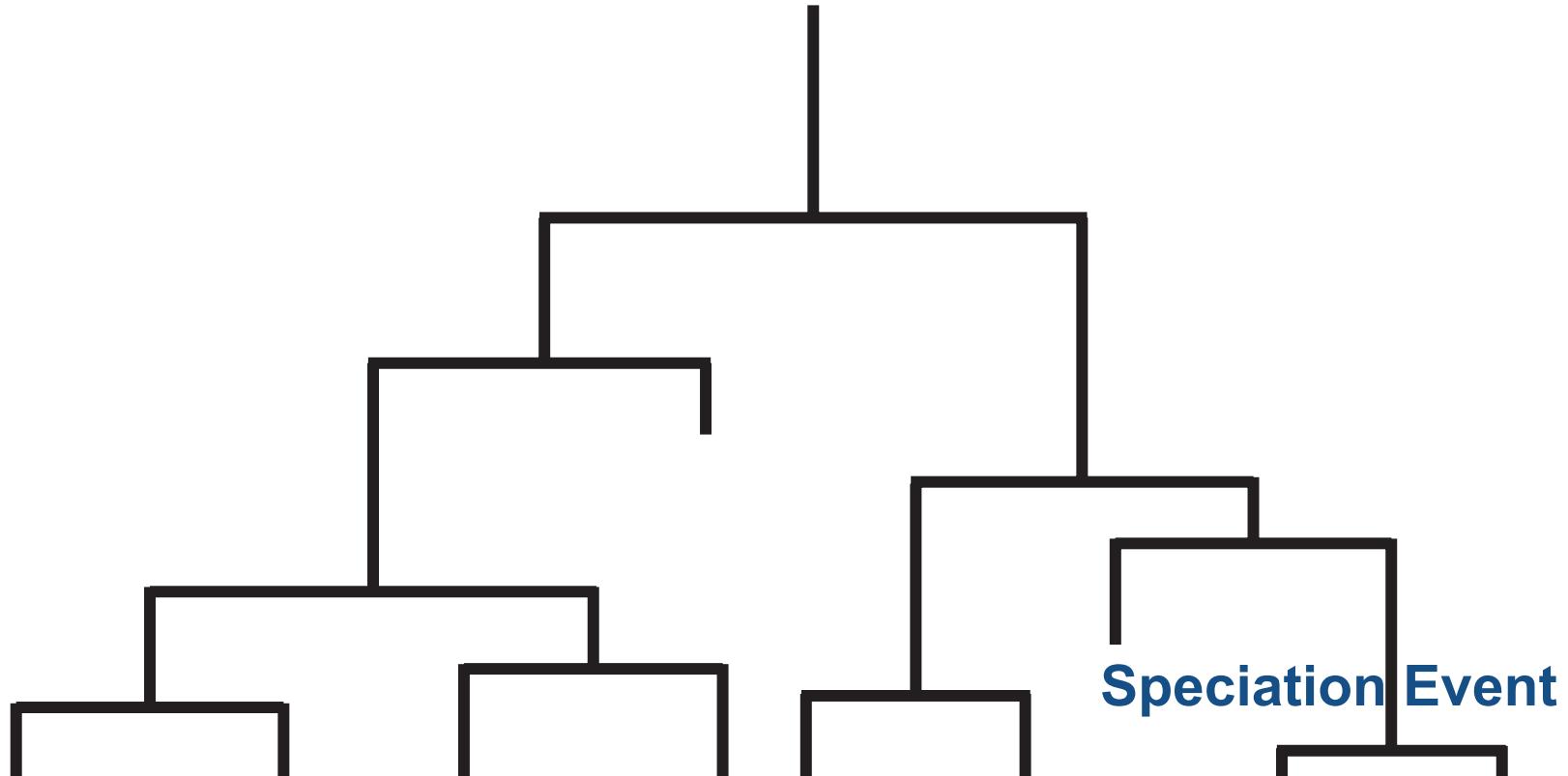
# Birth-Death Process

---



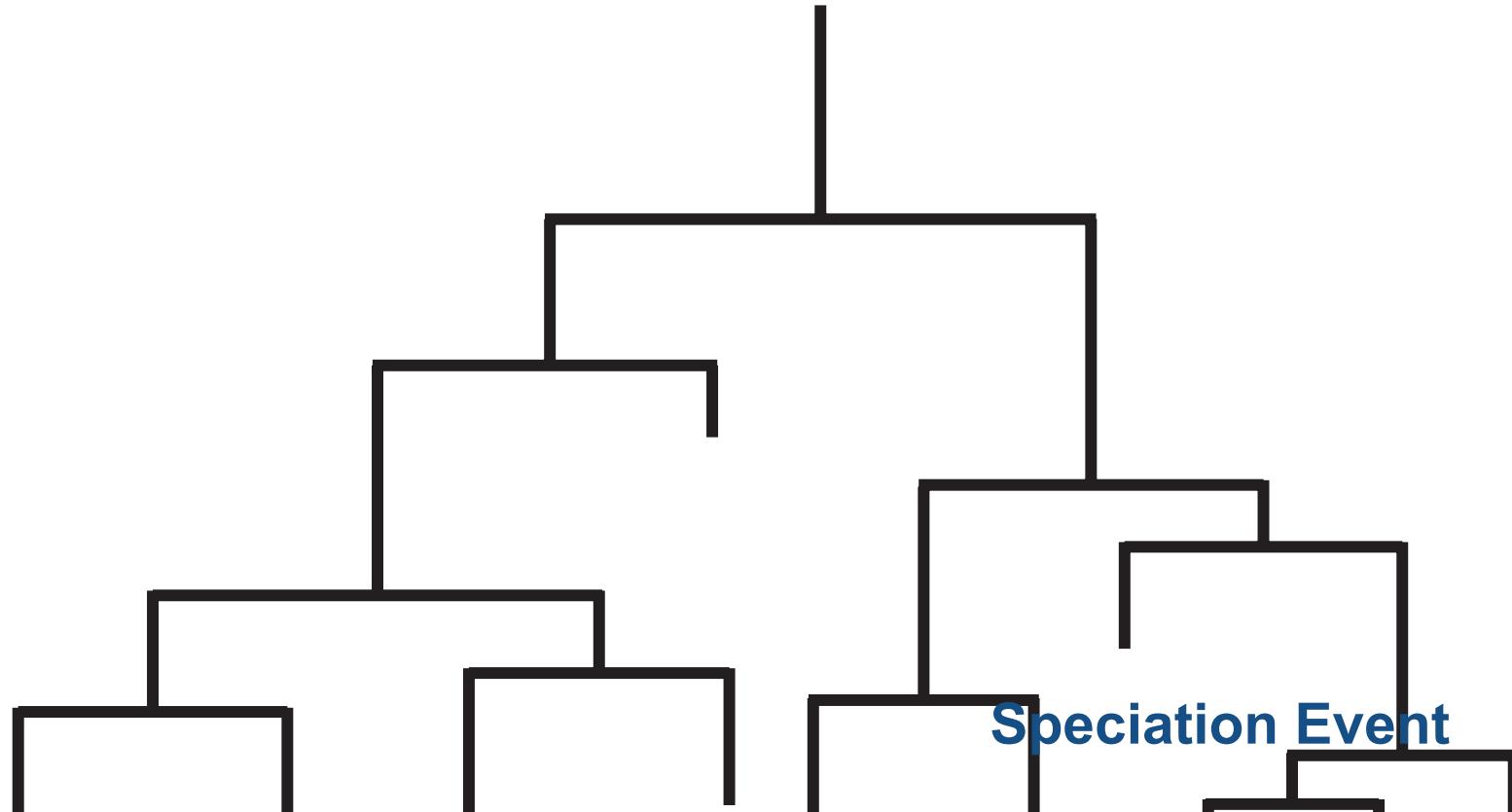
# Birth-Death Process

---



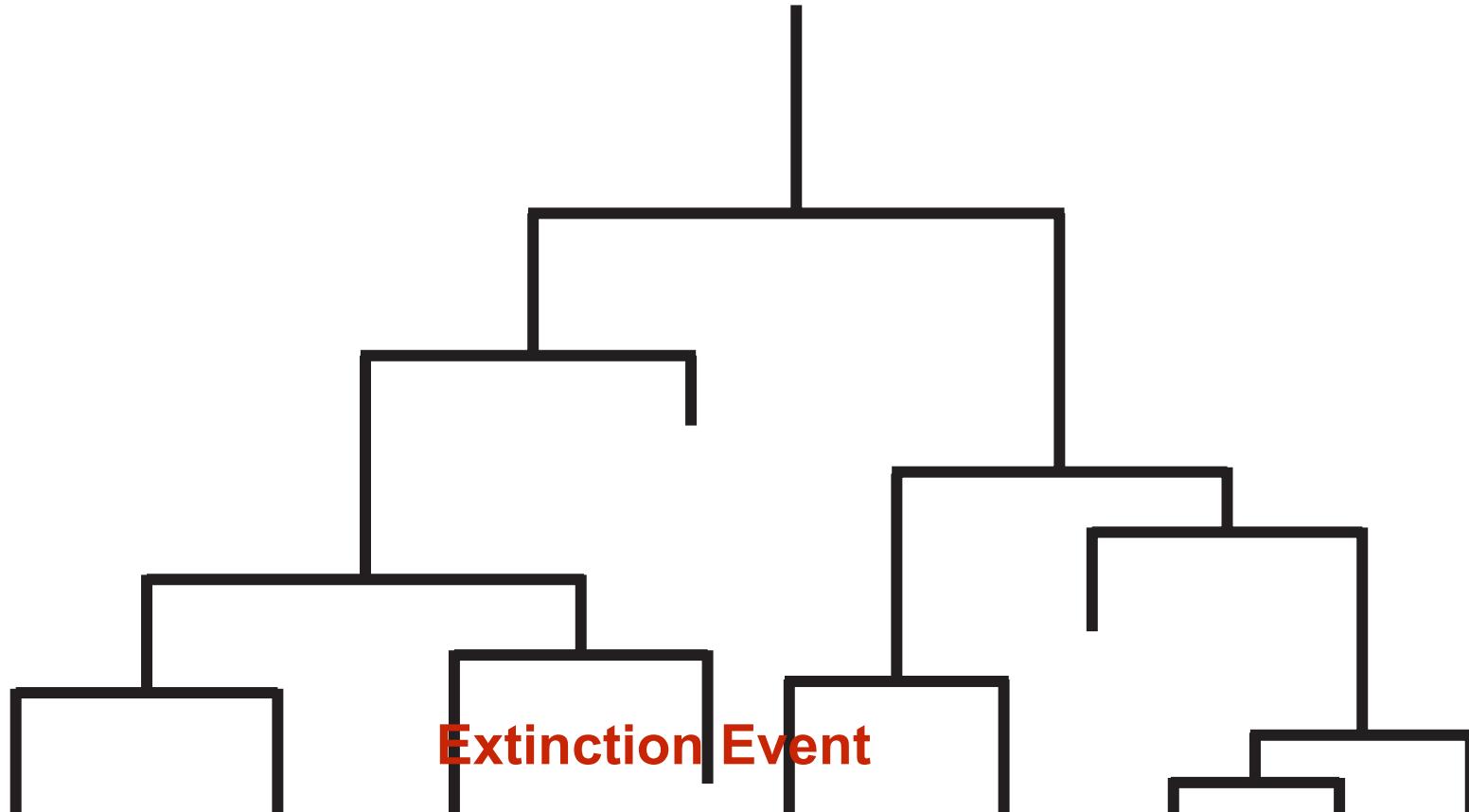
# Birth-Death Process

---



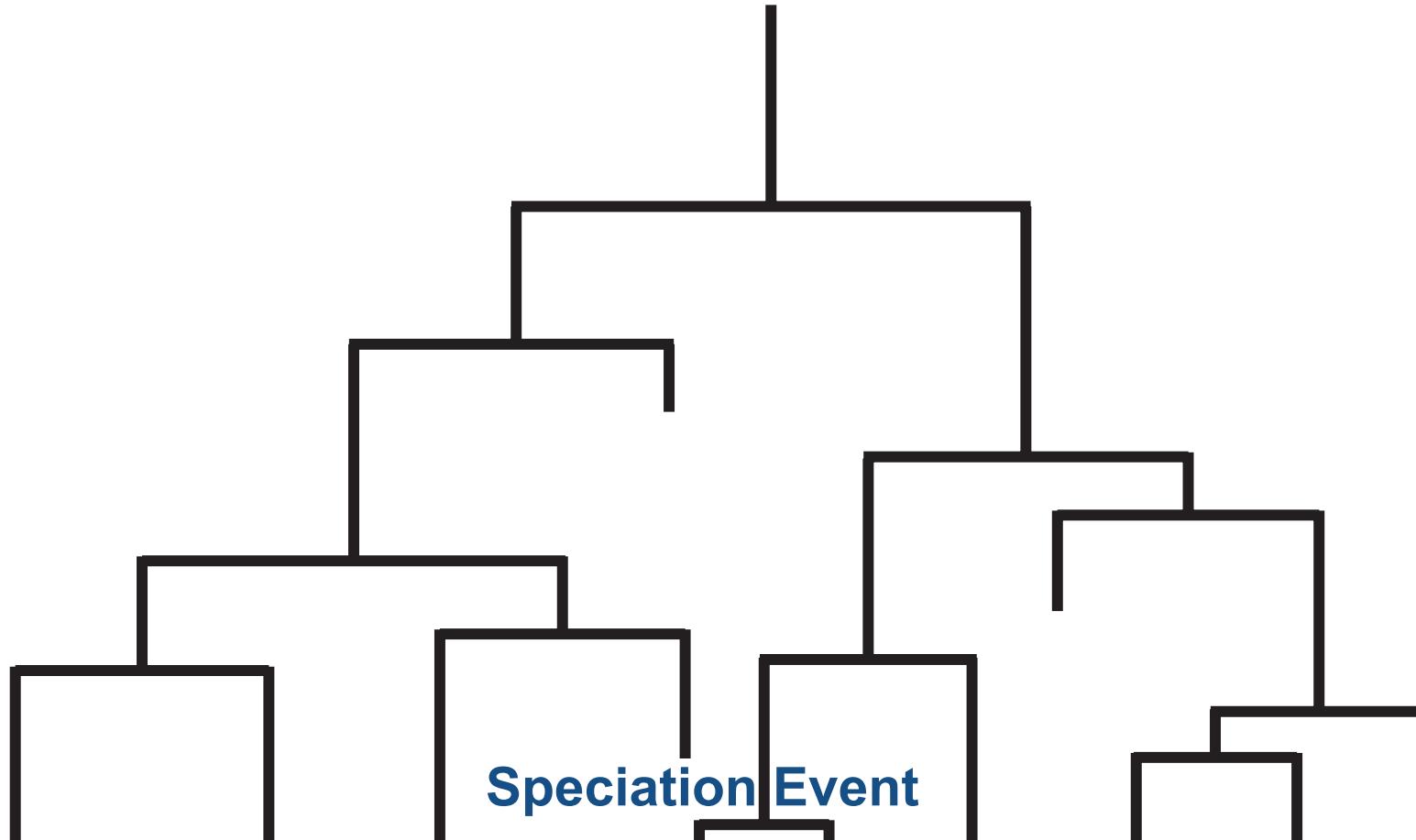
# Birth-Death Process

---



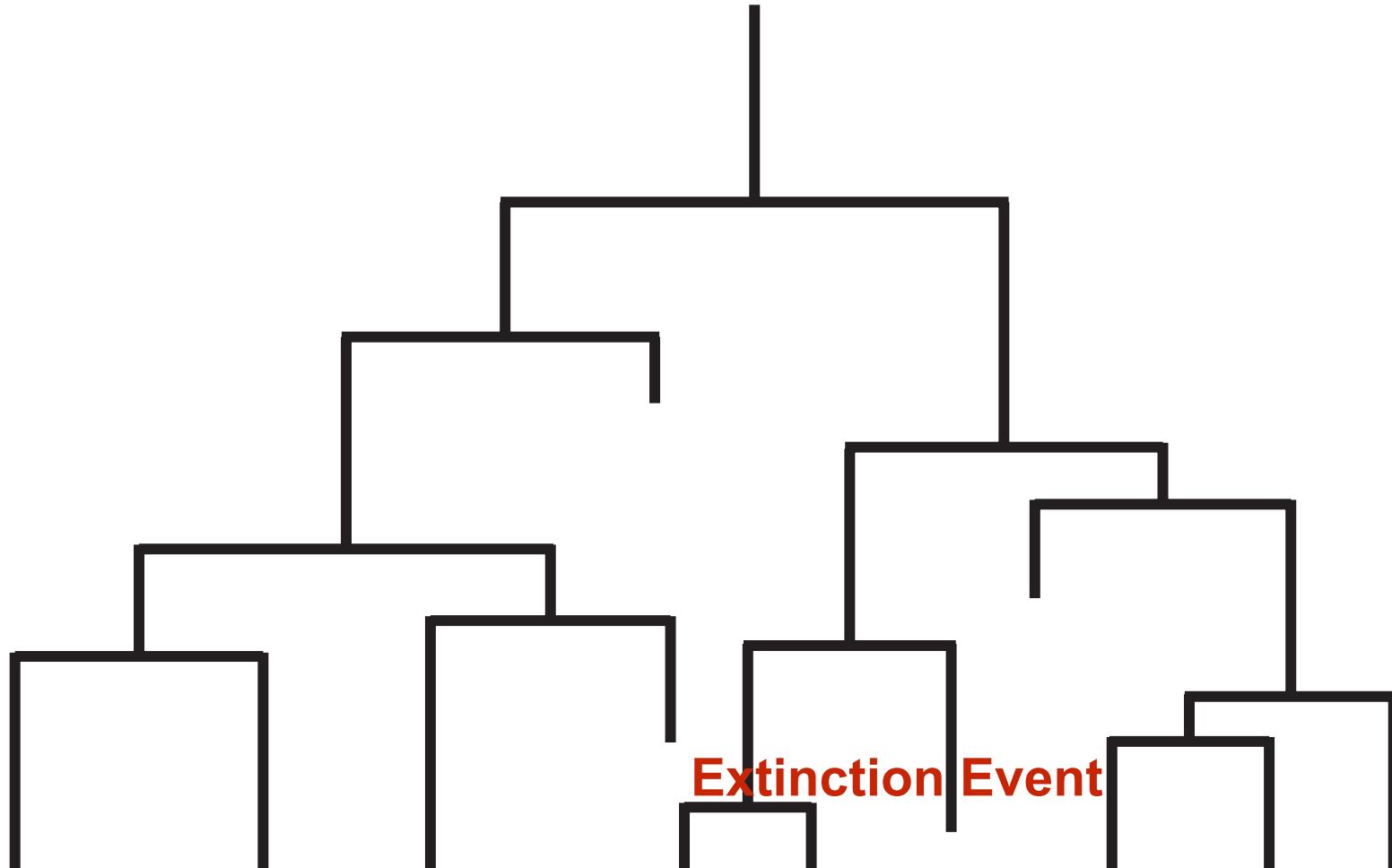
# Birth-Death Process

---



# Birth-Death Process

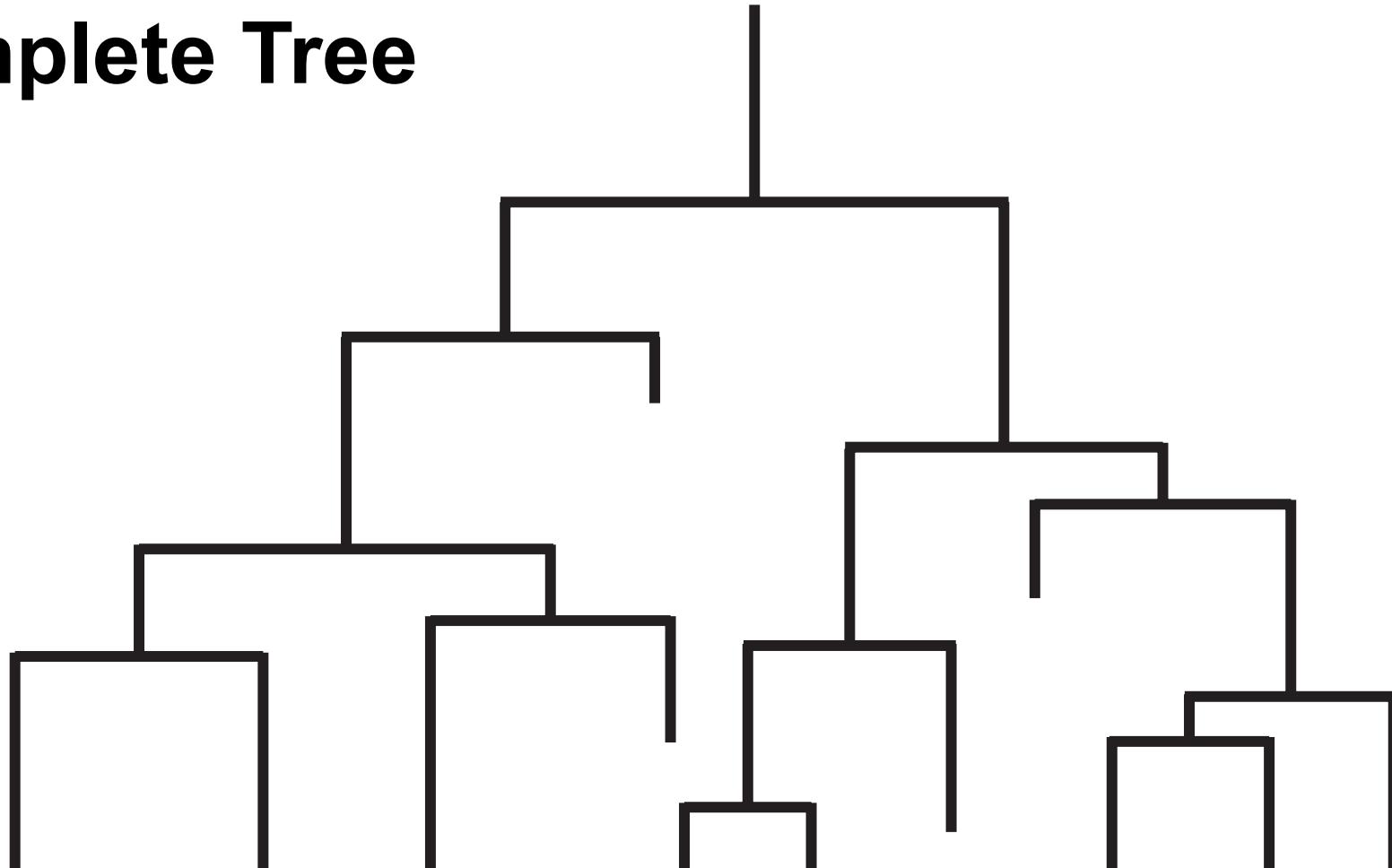
---



# Birth-Death Process

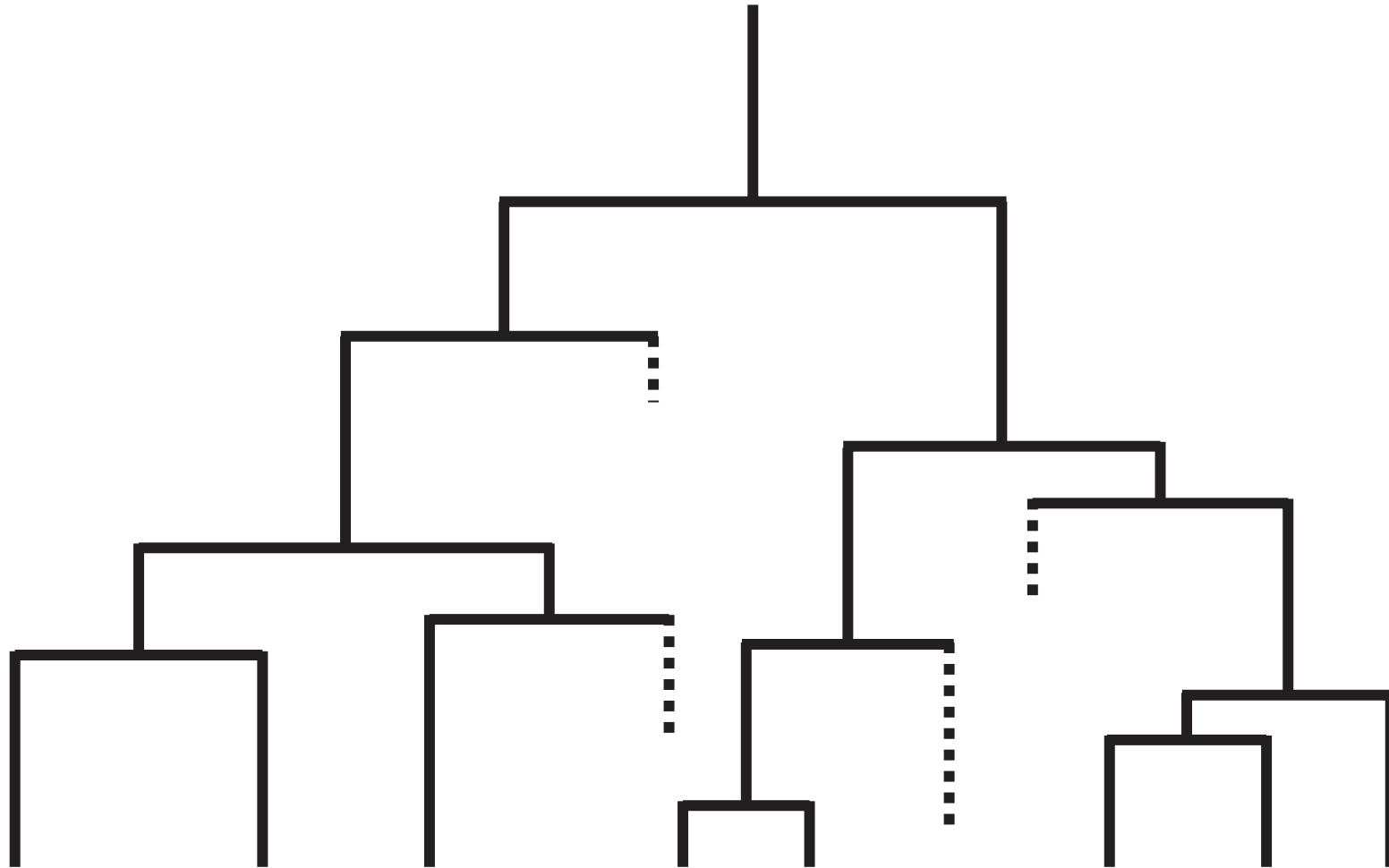
---

## Complete Tree



# Birth-Death Process

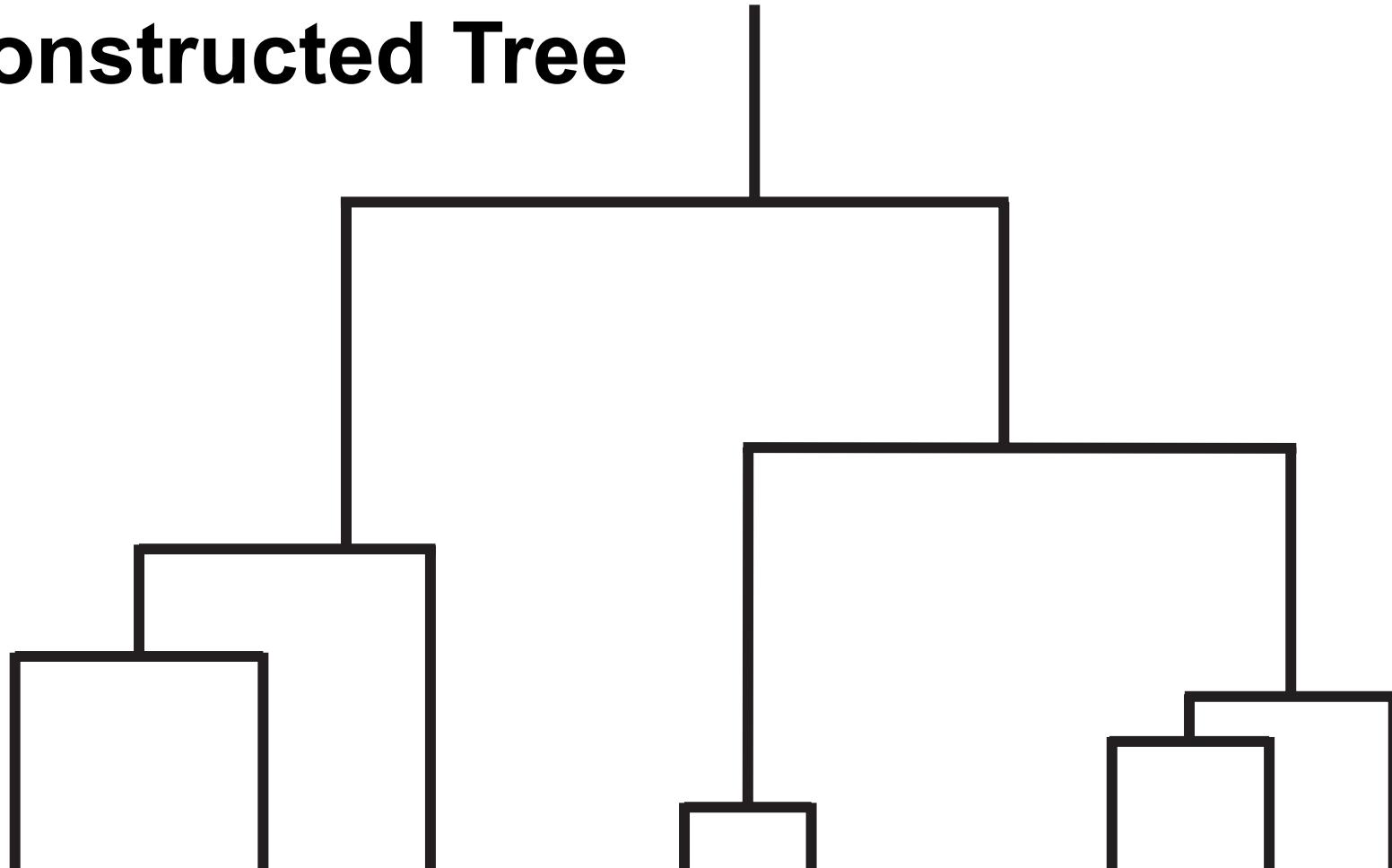
---



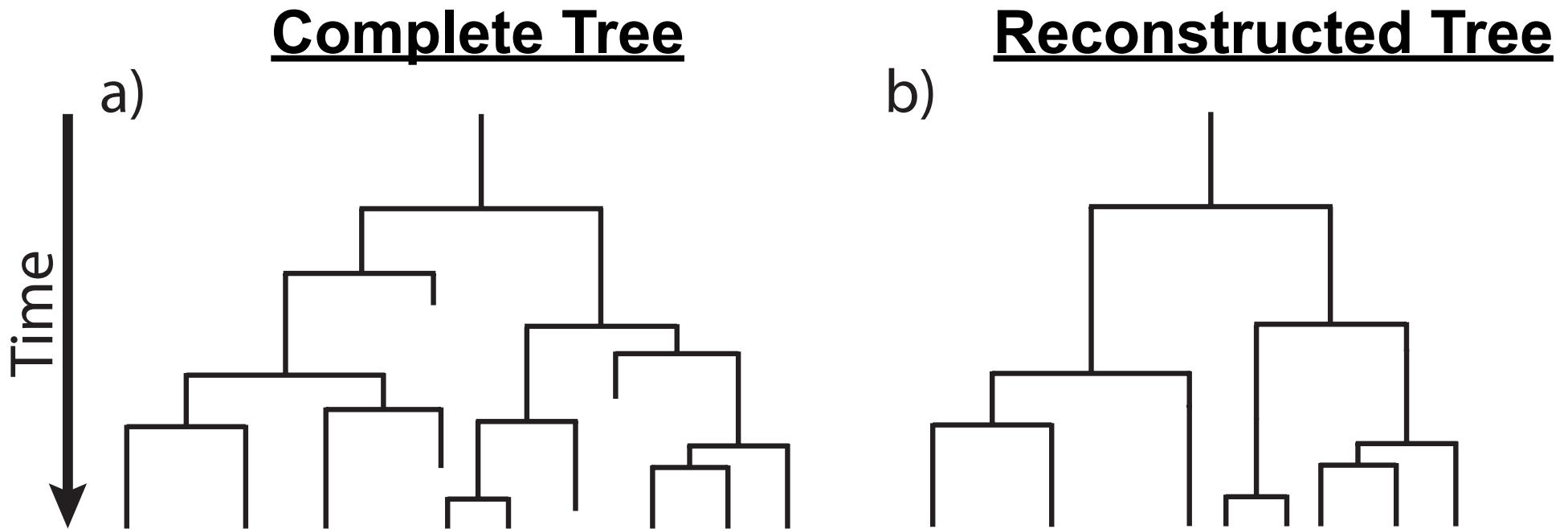
# Birth-Death Process

---

## Reconstructed Tree



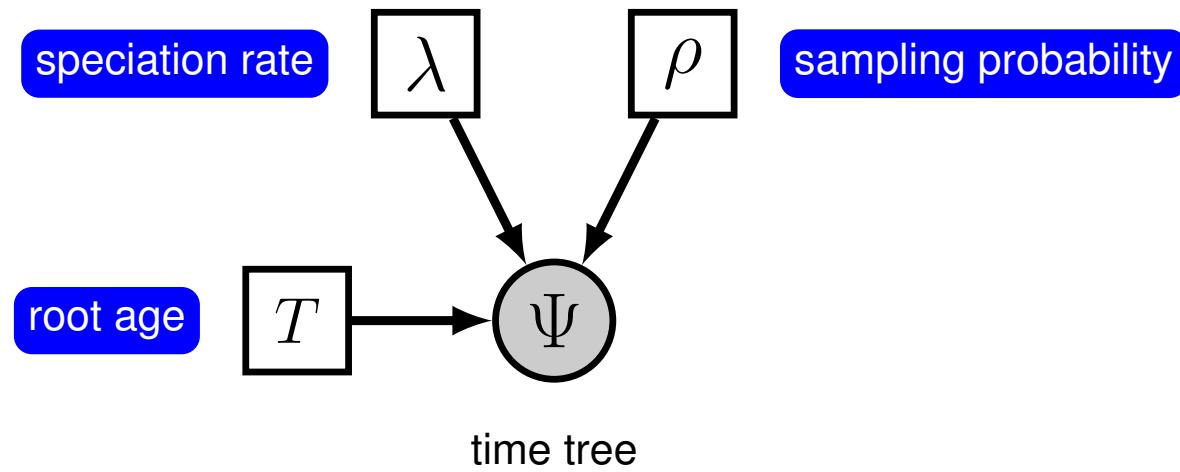
# Birth-Death Process



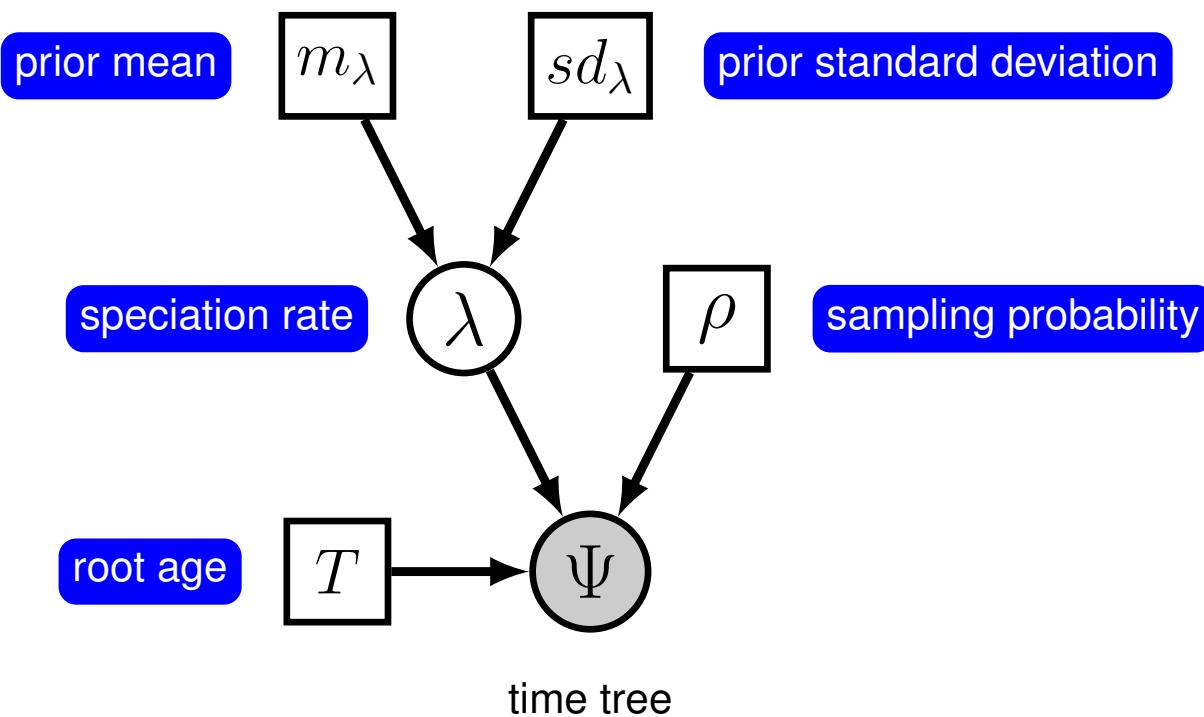
- A species gives birth to exactly one new species with rate  $\lambda(t)$
- A species dies with rate  $\mu(t)$
- Only extant species can be observed/sampled.

# Simple pure-birth (Yule) model

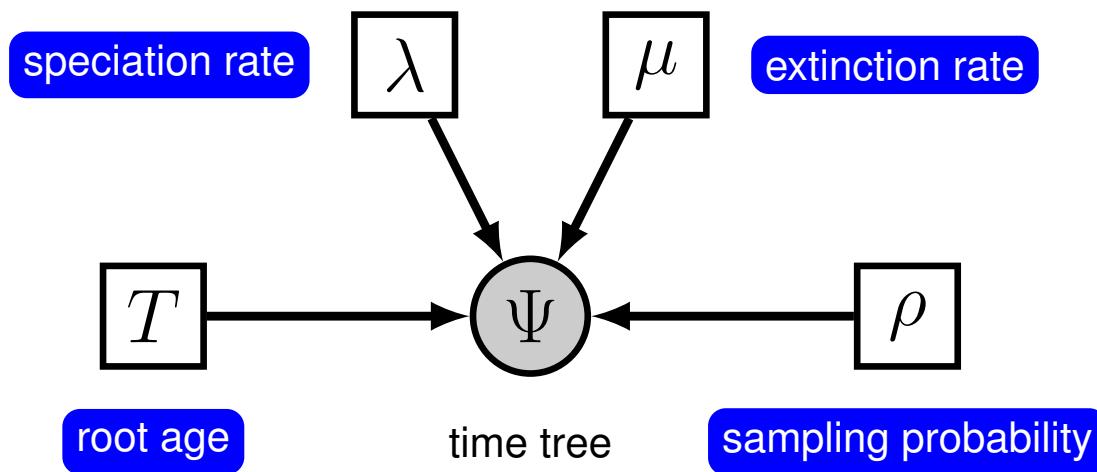
---



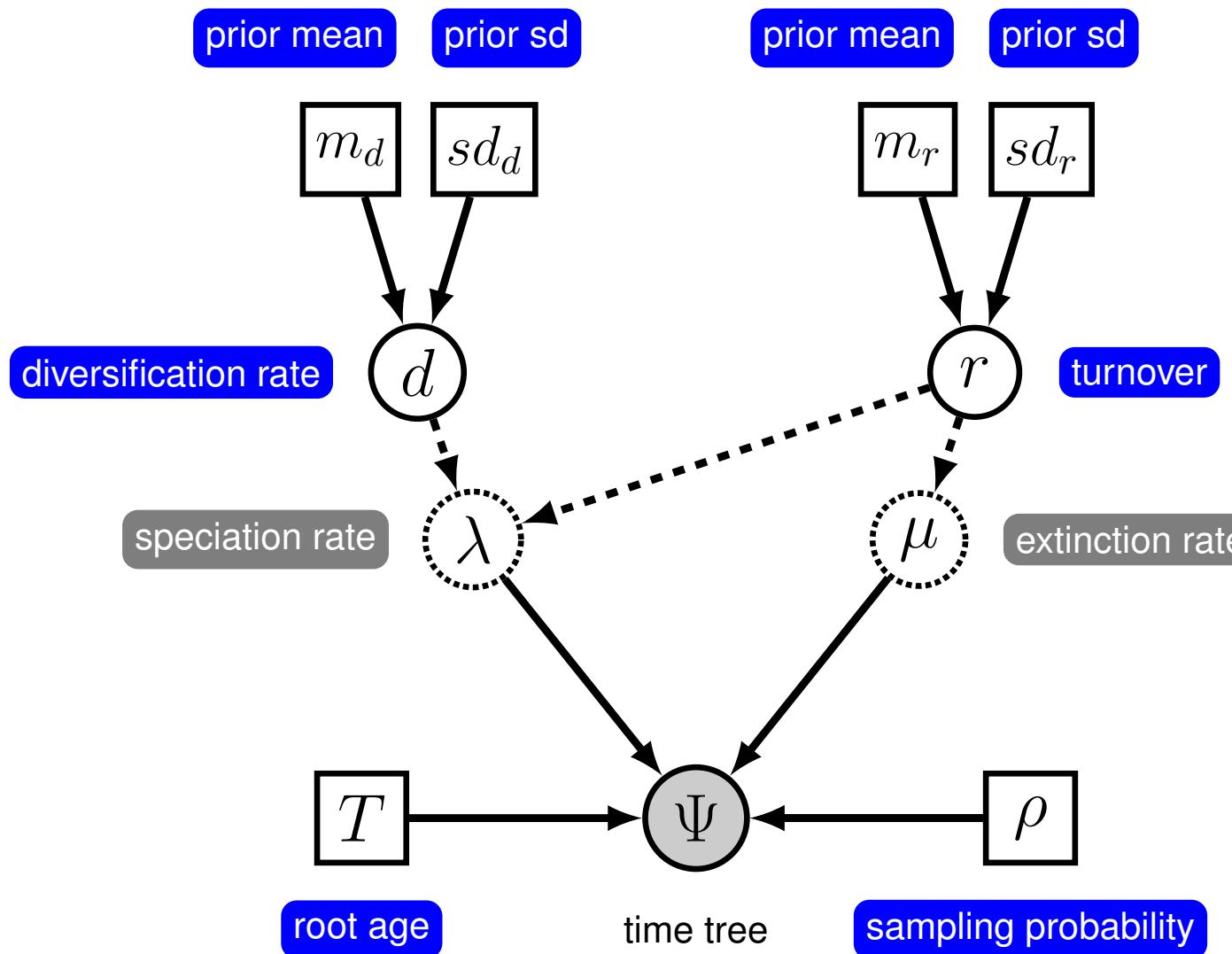
# Pure-birth (Yule) model with prior



# Simple Birth-Death Model



# Birth-Death Model with Prior



# Parameterization and interpretation

---

1)	Speciation rate	$\lambda$
	Extinction rate	$\mu$
2)	Net-diversification rate:	$\lambda - \mu$
	Turnover rate	$\mu$
3)	Net-diversification rate:	$\lambda - \mu$
	Relative extinction rate	$\frac{\mu}{\lambda}$
4)	Speciation rate	$\lambda$
	Relative extinction rate	$\frac{\mu}{\lambda}$

# Parameterization and interpretation

---

Speciation rate

$$\lambda$$

Extinction rate

$$\mu$$

- + Natural parameterization of the birth-death process
- Extinction rate might be larger than speciation rate
- Difficult to estimate parameters

# Parameterization and interpretation

---

Net-diversification rate:  $\lambda - \mu$

Turnover rate  $\mu$

+ Good prior information about net-diversification:

$$E[\lambda - \mu] = \ln\left(\frac{N}{2}\right) / T$$

+ Extinction rate can be enforced to be smaller than speciation

+ Biological interpretation of turnover rate

# Parameterization and interpretation

---

Net-diversification rate:  $\lambda - \mu$

Relative extinction rate  $\frac{\mu}{\lambda}$

+ Good prior information about net-diversification:

$$E[\lambda - \mu] = \ln\left(\frac{N}{2}\right) / T$$

+ Extinction rate can be enforced to be smaller than speciation

+ Simple prior on turnover rate (Beta distribution)

- Weird induced priors on parameters

# Parameterization and interpretation

---

Speciation rate  $\lambda$

Relative extinction rate  $\frac{\mu}{\lambda}$

- + Extinction rate can be enforced to be smaller than speciation
- + Simple prior on turnover rate (Beta distribution)
- Difficult to specify prior on speciation rate

# **Exercise 1: Simple Diversification Rate Estimation**

## **1) Estimating constant rates of speciation and extinction**

- What are the speciation and extinction rates for my study group?

## **2) Estimating time-varying rates of speciation and extinction**

- Do speciation and extinction rates vary through time?
- Is this variation through time correlated with some abiotic factor?

## **3) Estimating branch-specific rates of speciation and extinction**

- Do speciation and extinction rates vary among lineages?
- Is this variation among lineages correlated with some biotic factor?

## **1) Estimating constant rates of speciation and extinction**

- What are the speciation and extinction rates for my study group?

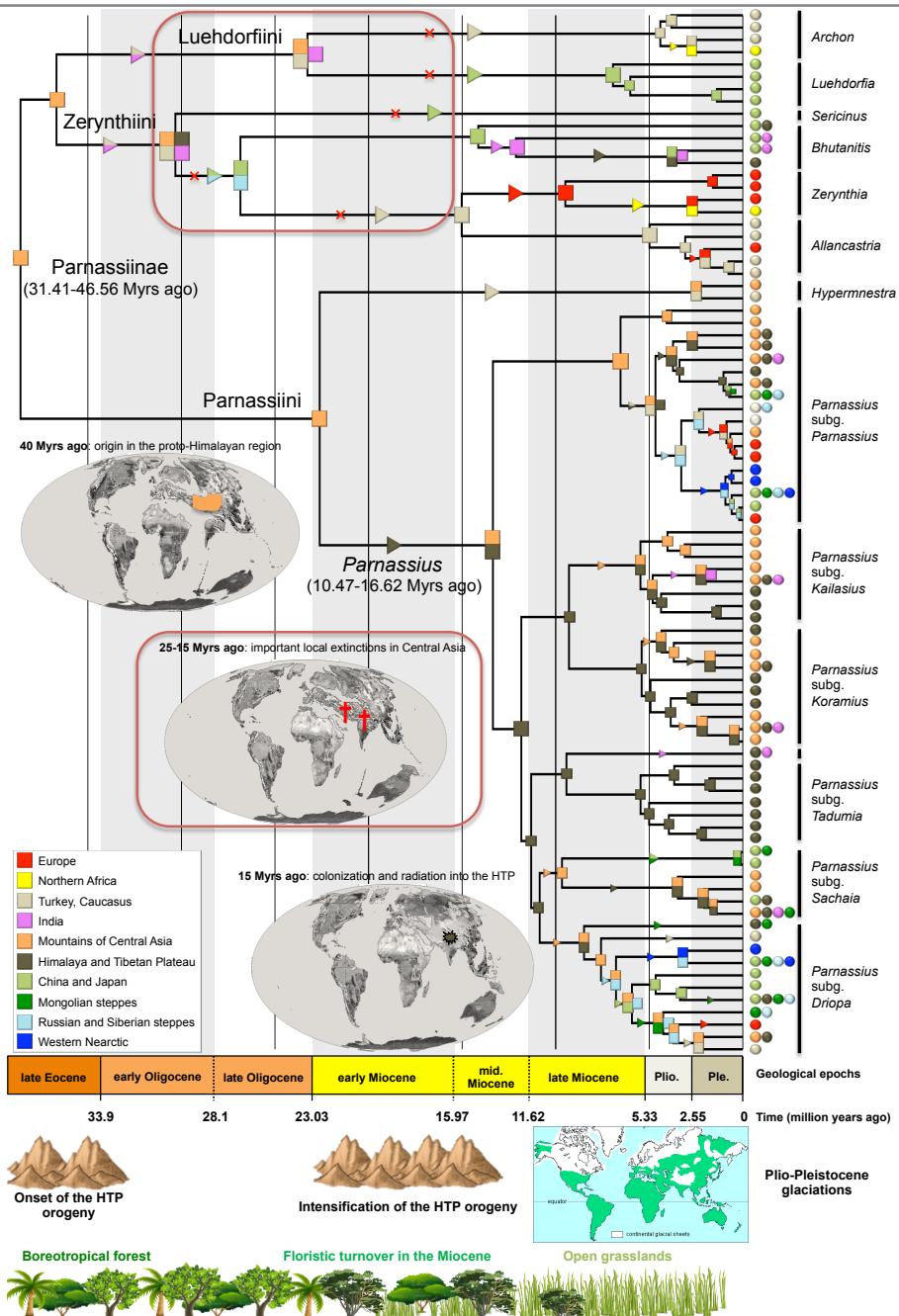
## **2) Estimating time-varying rates of speciation and extinction**

- Do speciation and extinction rates vary through time?
- Is this variation through time correlated with some abiotic factor?

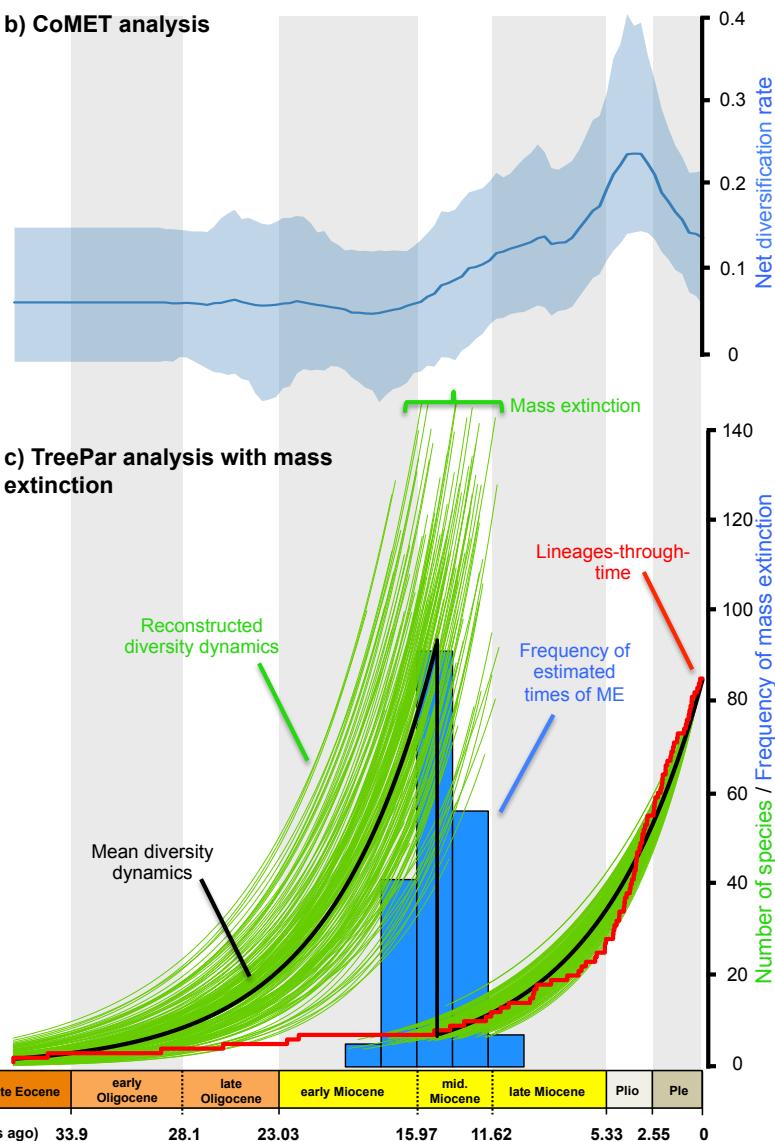
## **3) Estimating branch-specific rates of speciation and extinction**

- Do speciation and extinction rates vary among lineages?
- Is this variation among lineages correlated with some biotic factor?

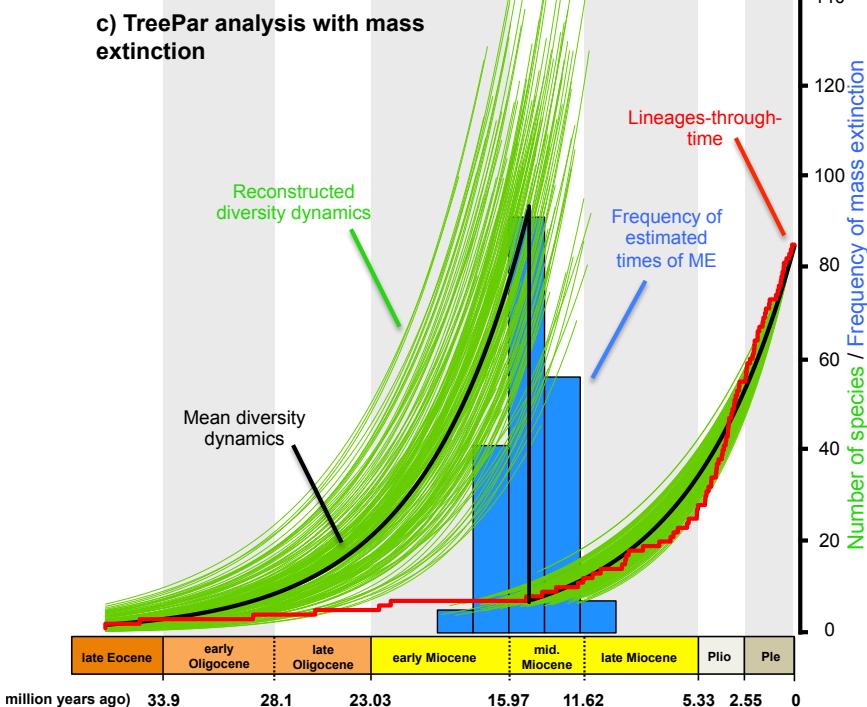
# Example: Diversification in Apollo Butterflies

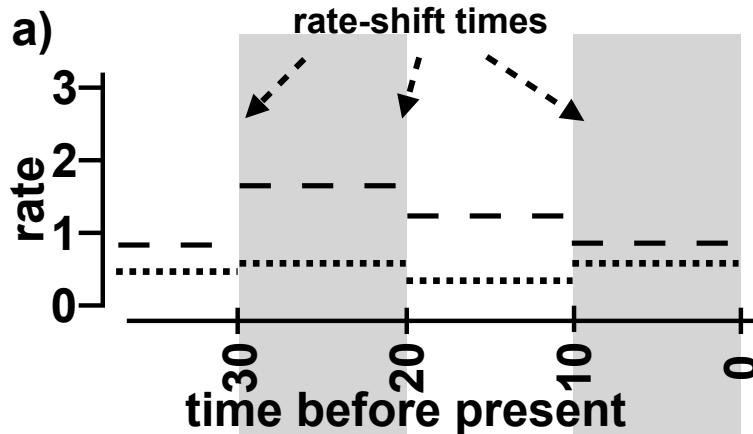


b) CoMET analysis



c) TreePar analysis with mass extinction



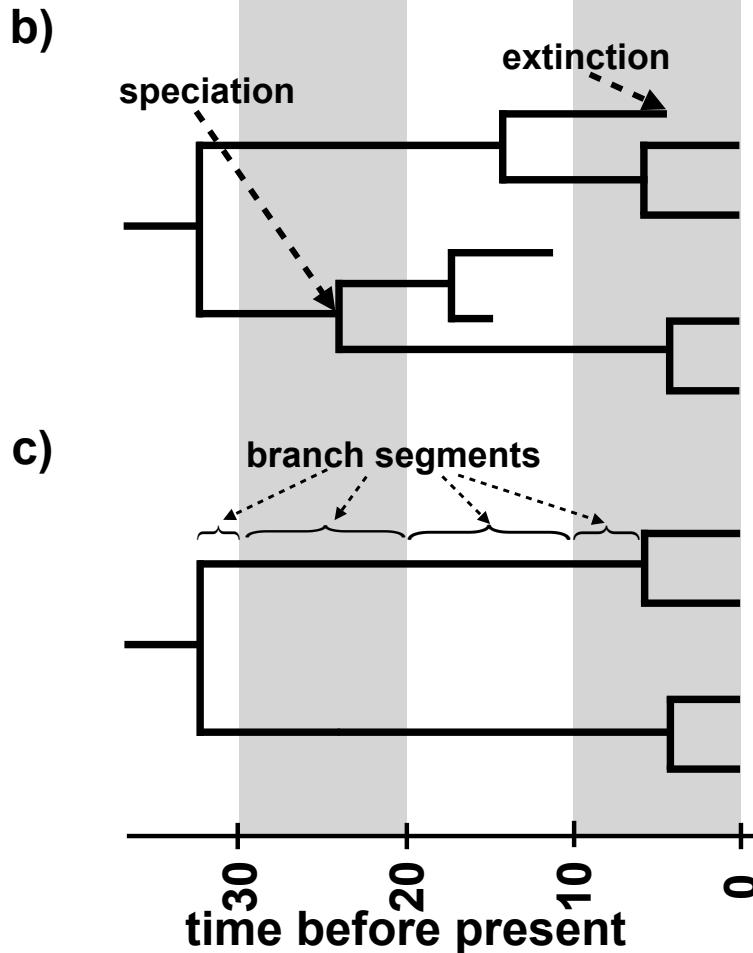


### Probability density of a reconstructed tree:

$$f(\Psi) = \frac{2^{N-1}}{N!} \times \begin{array}{l} \text{(i) the probability of the tree topology} \\ \rho^N \times \end{array}$$

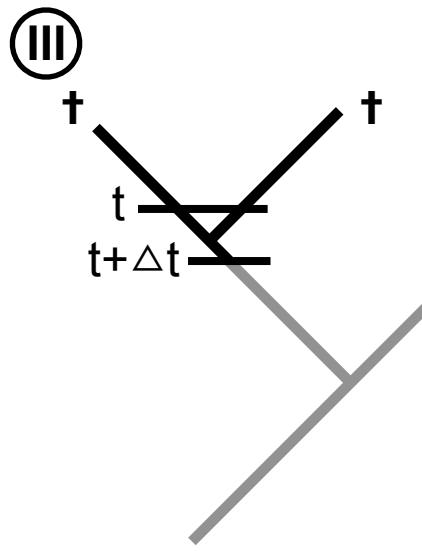
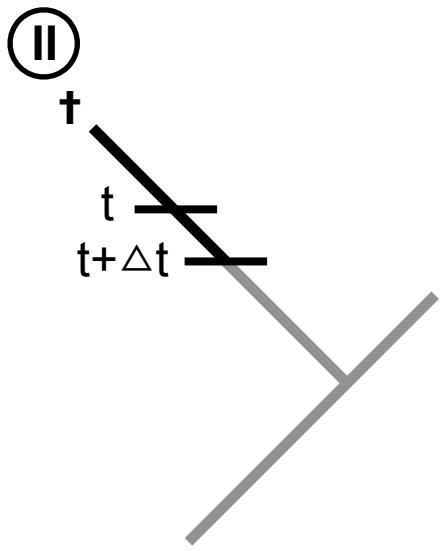
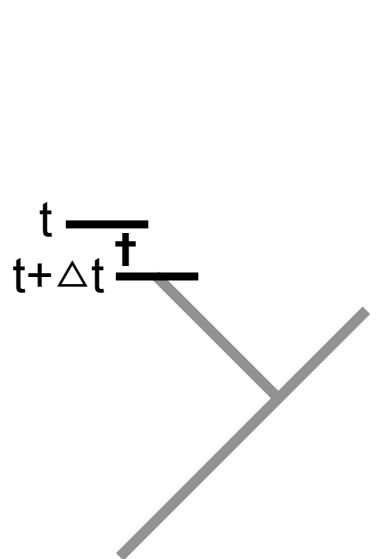
$$\prod_{t \in \mathcal{N}} [\lambda(t)] \times \begin{array}{l} \text{(ii) the probability of sampling the extant taxa} \\ \prod_{t \in \mathcal{B}} \left[ \frac{D(t_o)}{D(t_y)} \right] \end{array}$$

$$\begin{array}{l} \text{(iii) the probability of a speciation event} \\ \text{(iv) the probability of each branch-segment} \end{array}$$



# Theory: Computing the Probability of

I



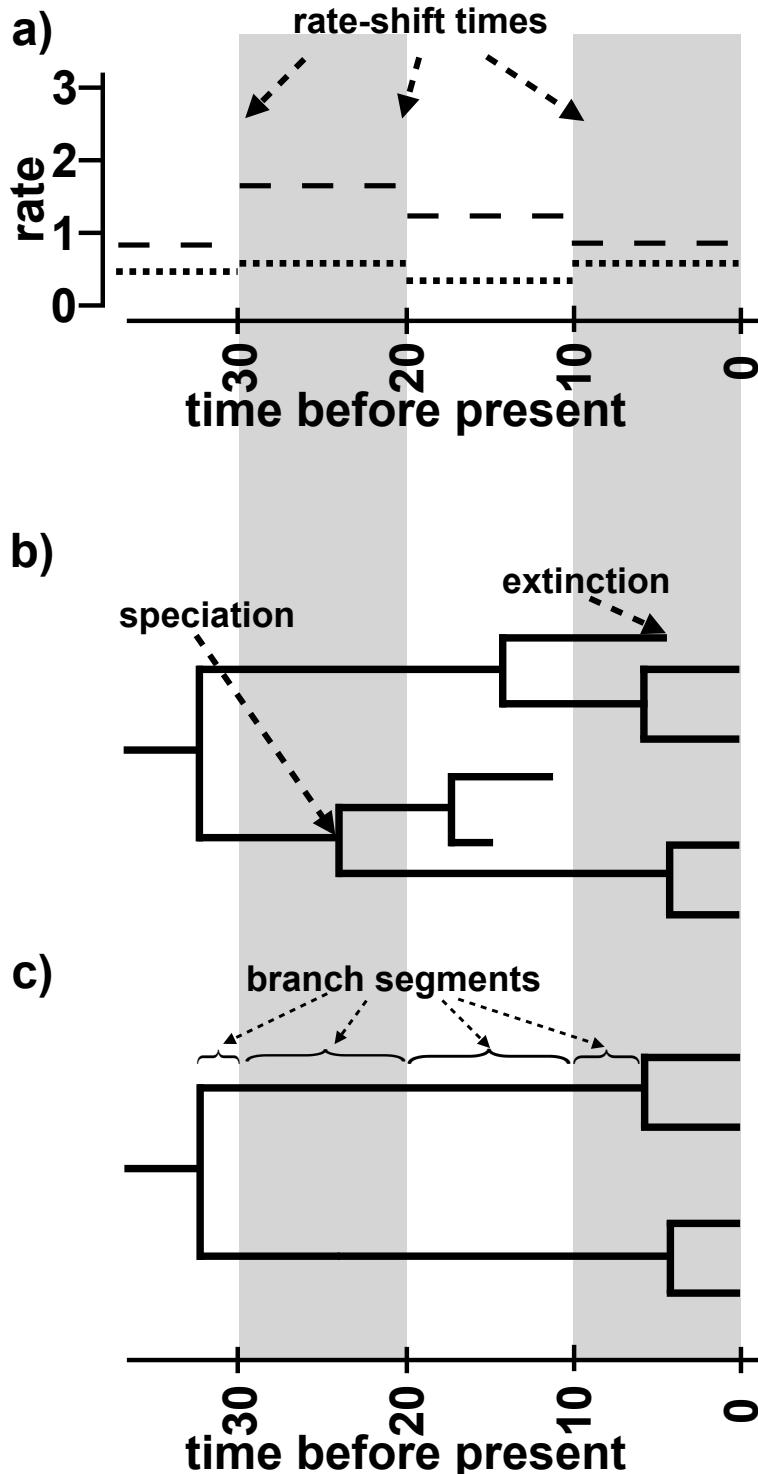
$$\begin{aligned} E(t + \Delta t) = & \mu(t)\Delta t \\ & + (1 - \mu(t)\Delta t - \lambda(t)\Delta t)E(t) \\ & + (1 - \mu(t)\Delta t)\lambda(t)\Delta t E(t)^2 \end{aligned}$$

(i) extinction event

(ii) no event

(iii) speciation event

$$\frac{dE}{dt} = \mu(t) - (\lambda(t) + \mu(t))E(t) + \lambda(t)E(t)^2$$



### Probability density of a reconstructed tree:

$$f(\Psi) = \frac{2^{N-1}}{N!} \times \begin{array}{l} \text{(i) the probability of the tree topology} \\ \rho^N \times \end{array}$$

$$\prod_{t \in \mathcal{N}} [\lambda(t)] \times \begin{array}{l} \text{(ii) the probability of sampling the extant taxa} \\ \prod_{t \in \mathcal{B}} \left[ \frac{D(t_o)}{D(t_y)} \right] \end{array}$$

$$\begin{array}{l} \text{(iii) the probability of a speciation event} \\ \text{(iv) the probability of each branch-segment} \end{array}$$

### Differential equations:

$$\frac{dE}{dt} = \mu(t) - (\lambda(t) + \mu(t))E(t) + \lambda(t)E(t)^2$$

$$\frac{dD}{dt} = -D(t)(\lambda(t) + \mu(t)) + 2D(t)\lambda(t)E(t)$$

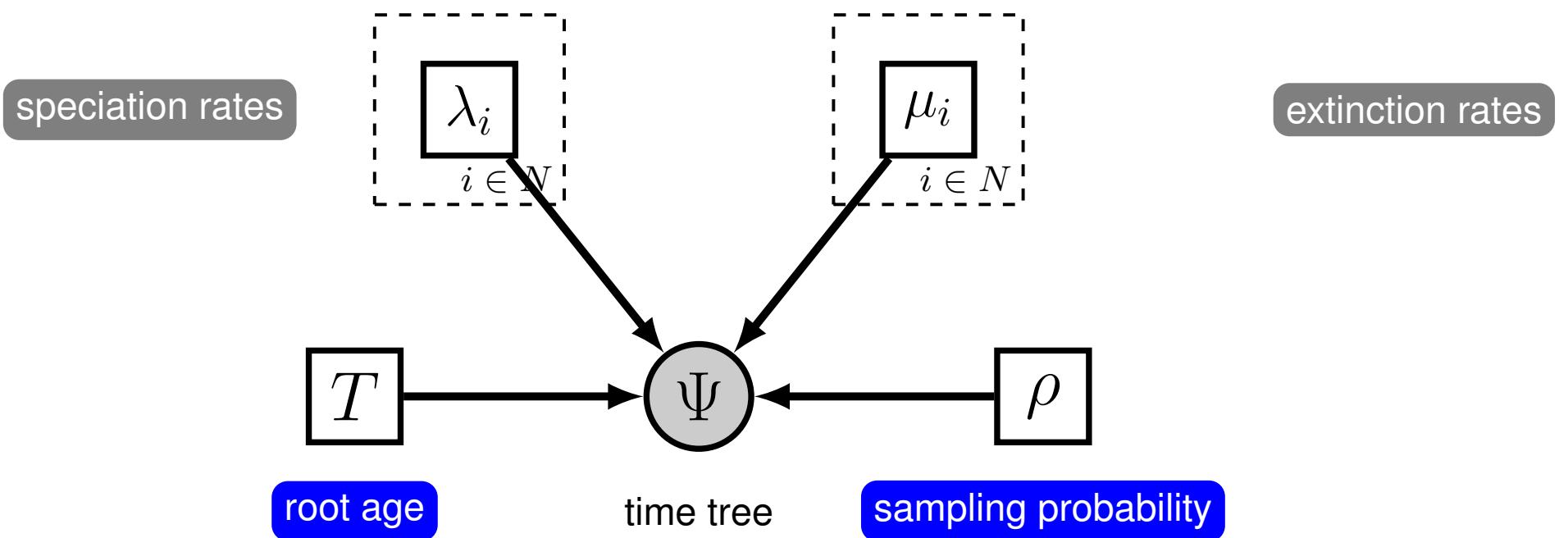
### Analytical solutions:

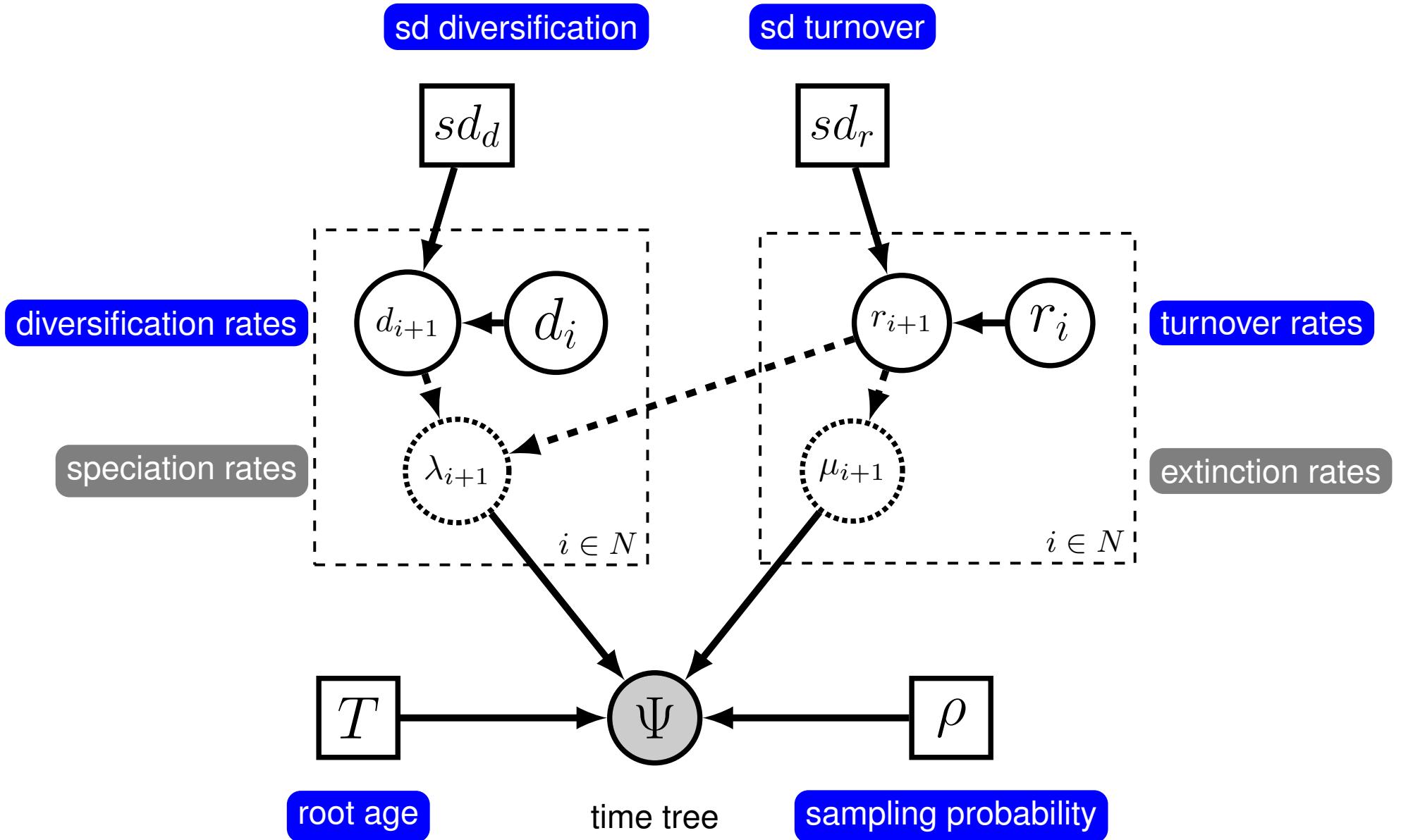
$$E_i(t) = \frac{\lambda_i + \mu_i - A_i \frac{1+B_i - e^{-A_i(t-s_i)}(1-B_i)}{1+B_i + e^{-A_i(t-s_i)}(1-B_i)}}{2\lambda_i}$$

$$D_i(t) = \frac{4e^{-A_i(t-s_i)}}{(1 + B_i + e^{-A_i(t-s_i)}(1 - B_i))^2}$$

$$A_i = \sqrt{(\lambda_i - \mu_i)^2 + (4 \times \lambda_i)}$$

$$B_i = \frac{(1 - 2E_{i-1}(s_i))\lambda_i + \mu_i}{A_i}$$





## **1) Estimating constant rates of speciation and extinction**

- What are the speciation and extinction rates for my study group?

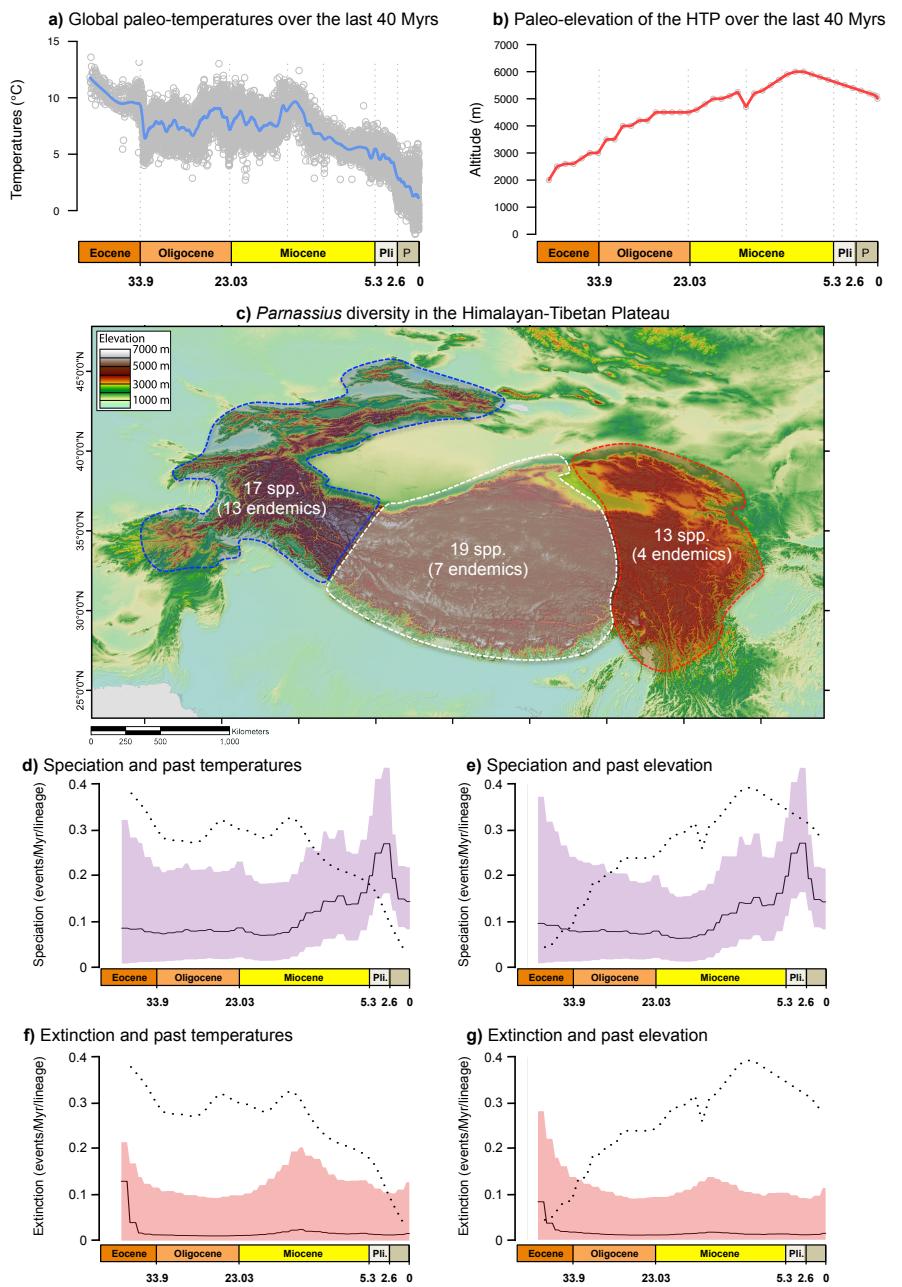
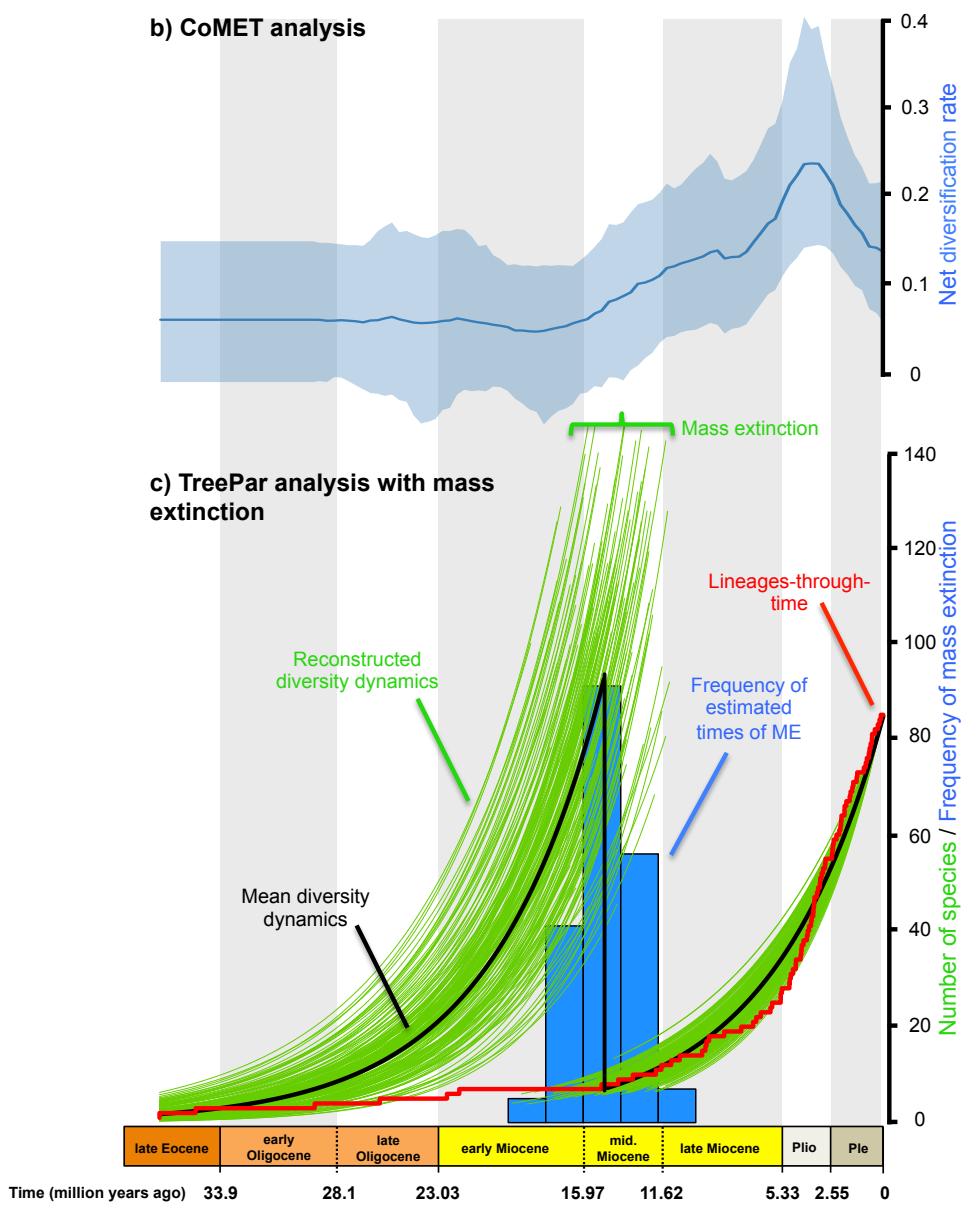
## **2) Estimating time-varying rates of speciation and extinction**

- Do speciation and extinction rates vary through time?
- Is this variation through time correlated with some abiotic factor?

## **3) Estimating branch-specific rates of speciation and extinction**

- Do speciation and extinction rates vary among lineages?
- Is this variation through time correlated with some biotic factor?

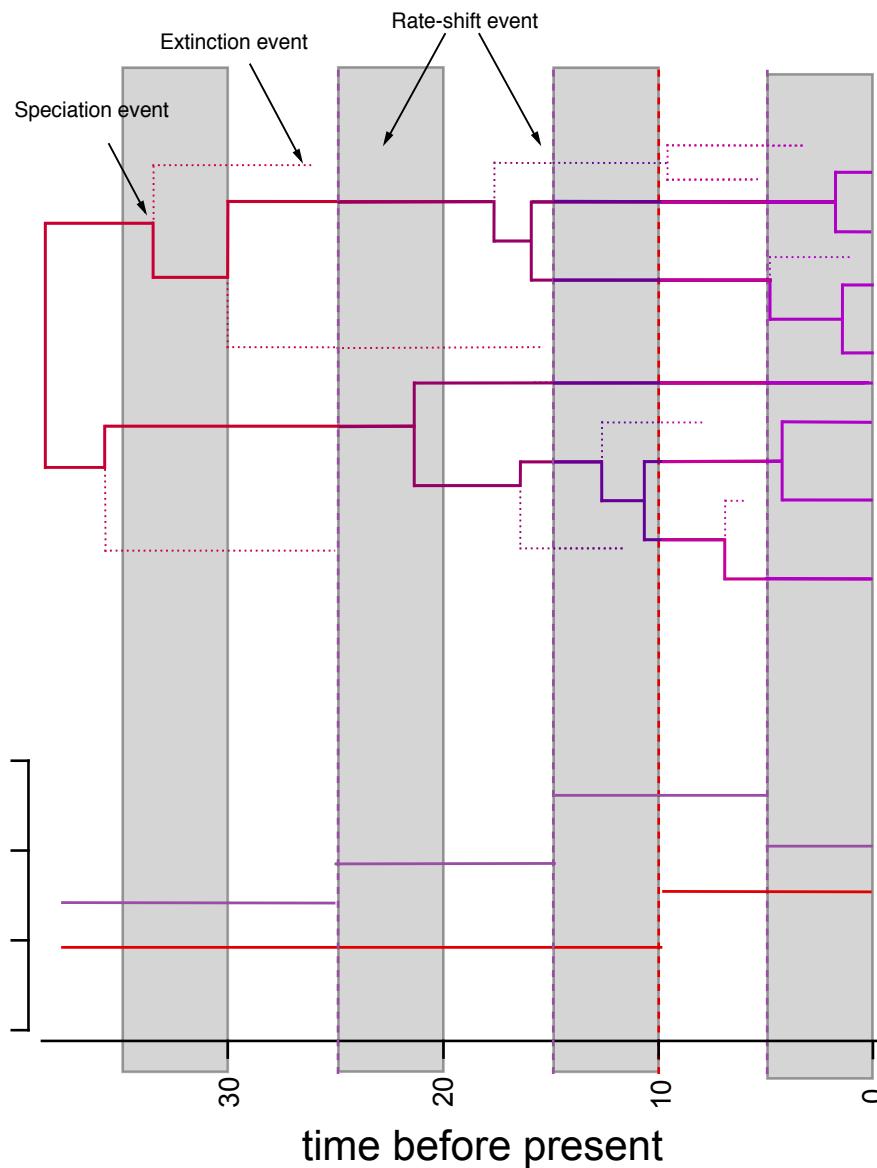
# Example: Diversification in Apollo Butterflies



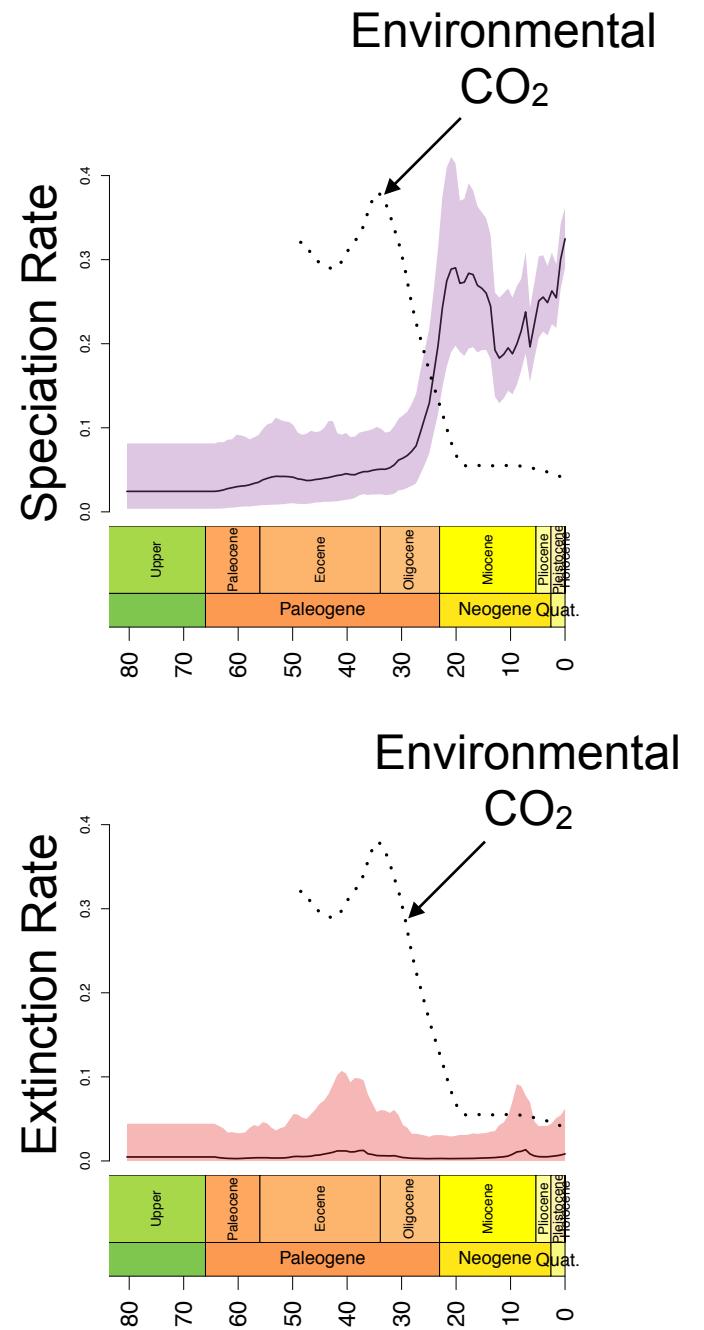
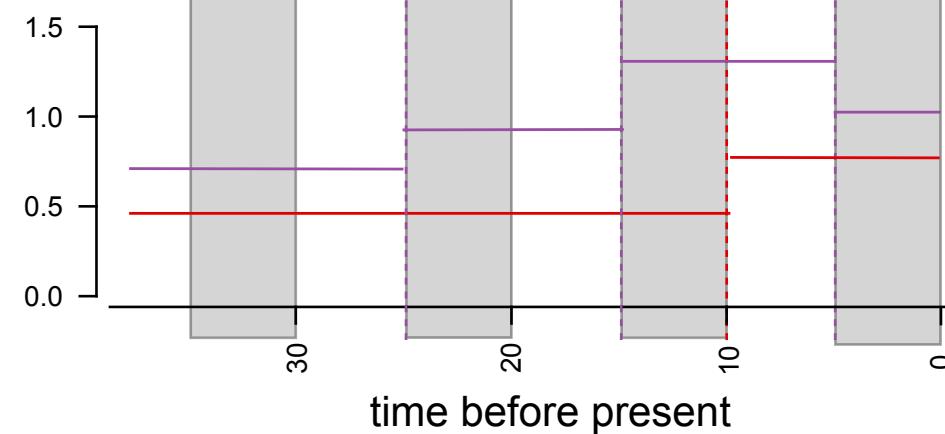
# Hypothesis: CO<sub>2</sub> is correlated with diversification rate

## Episodic birth-death process:

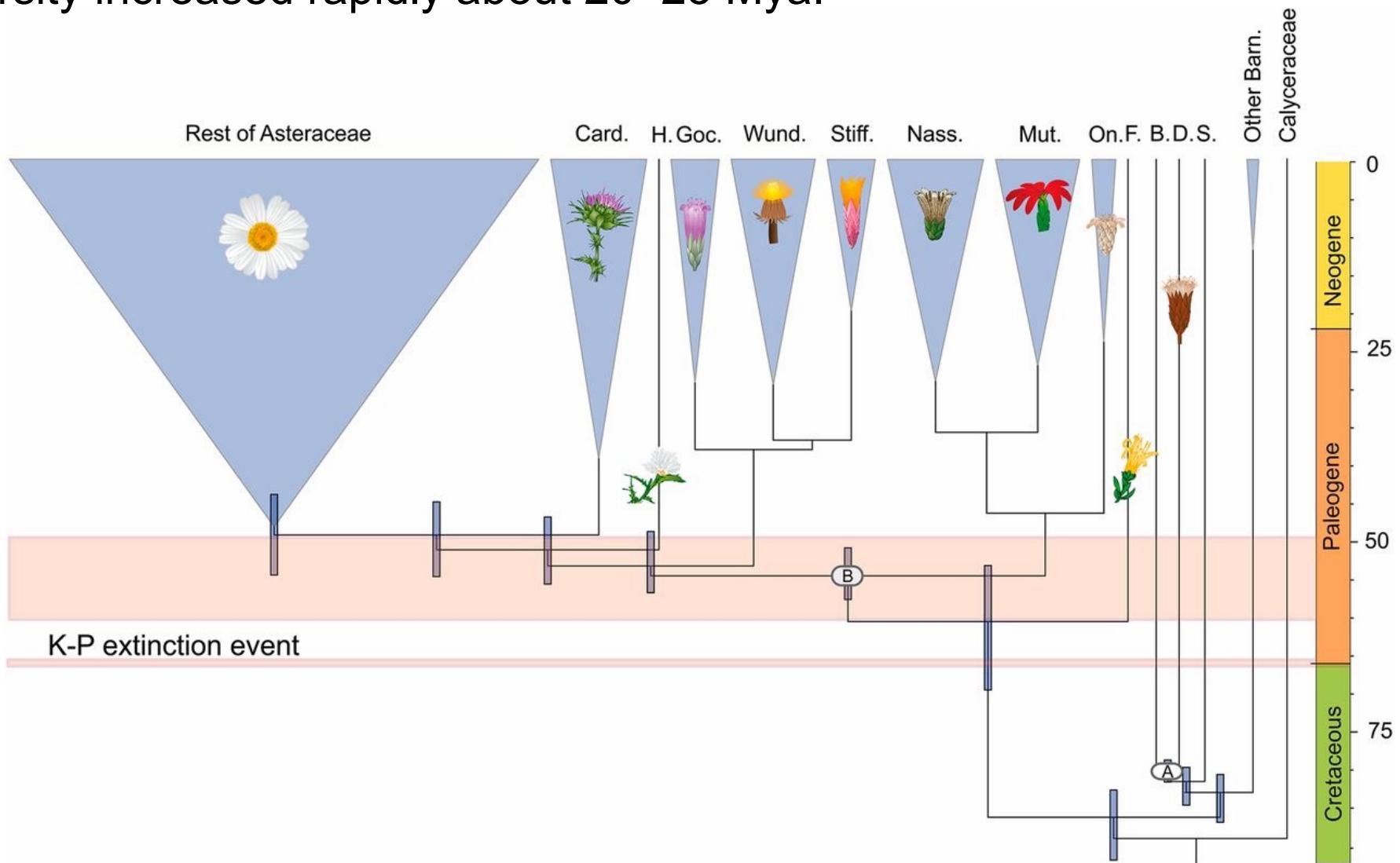
A



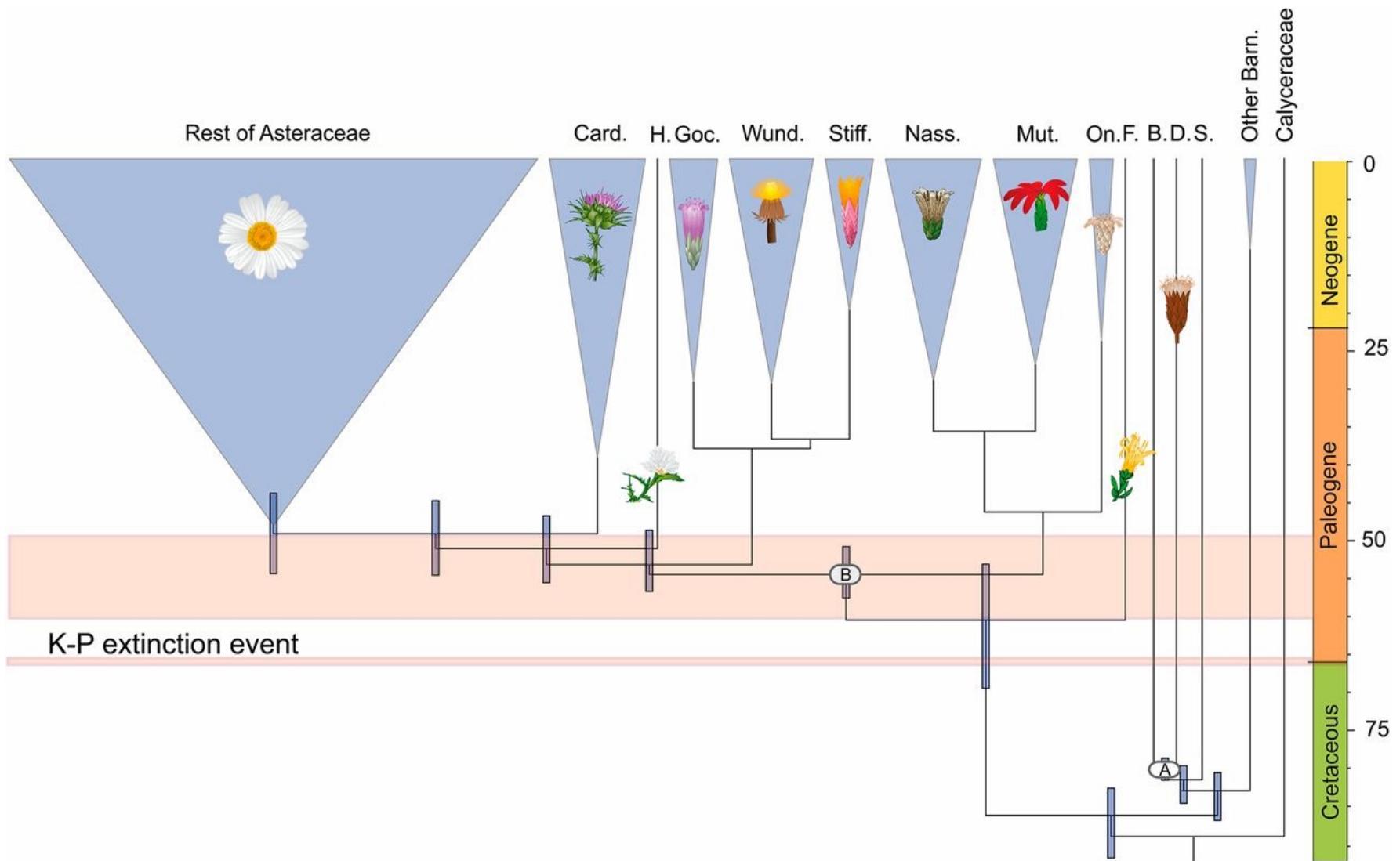
B



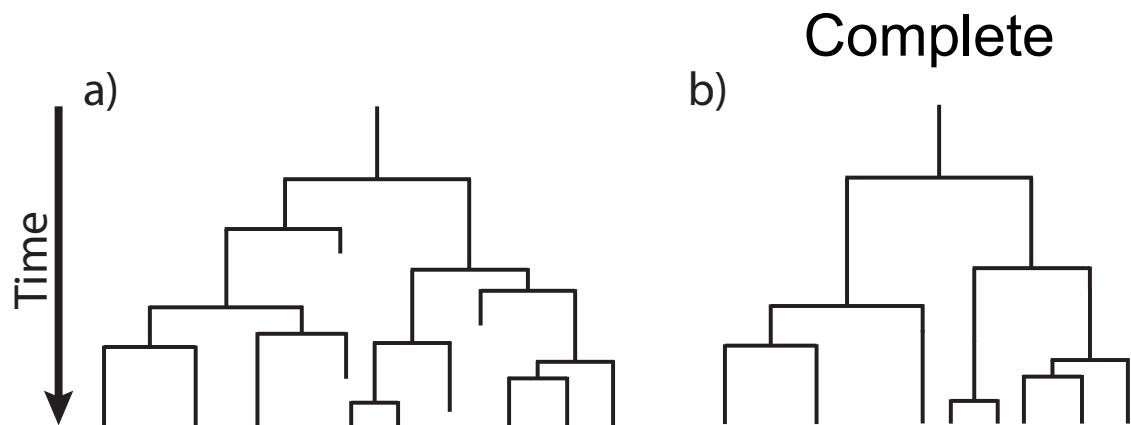
Asteraceae is one of the largest (plant) families: >23,000 species.  
Very few fossils from the Cretaceous known.  
Diversity increased rapidly about 20–25 Mya.



Estimated dated super tree with ~2200 species.  
Only 10% of taxa sampled.



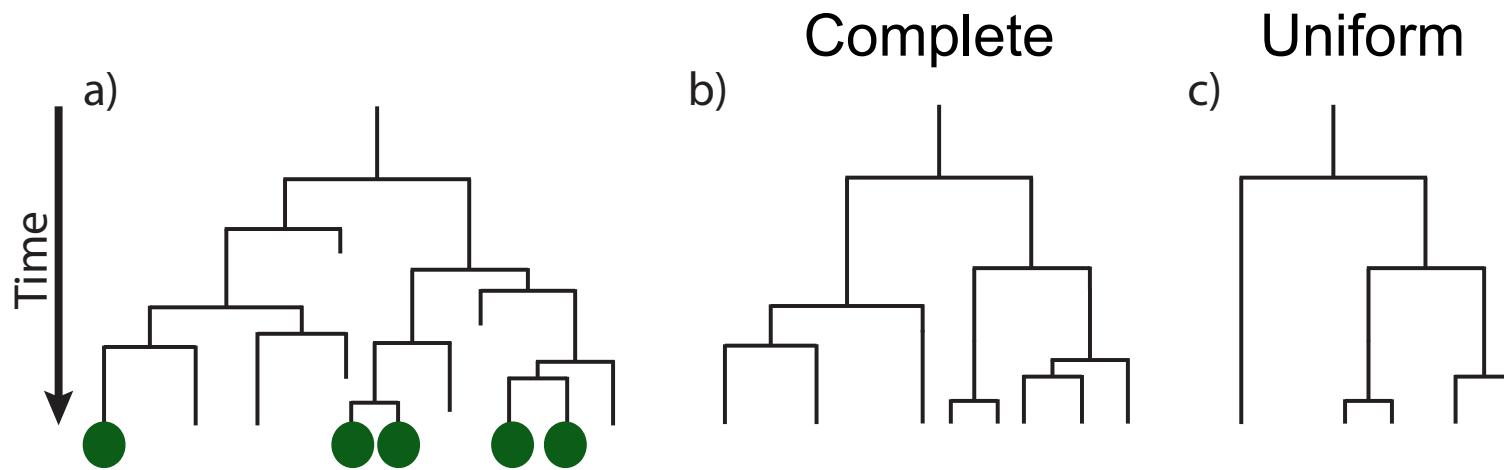
# Incomplete taxon sampling



Complete taxon sampling:

All extant taxa are included.

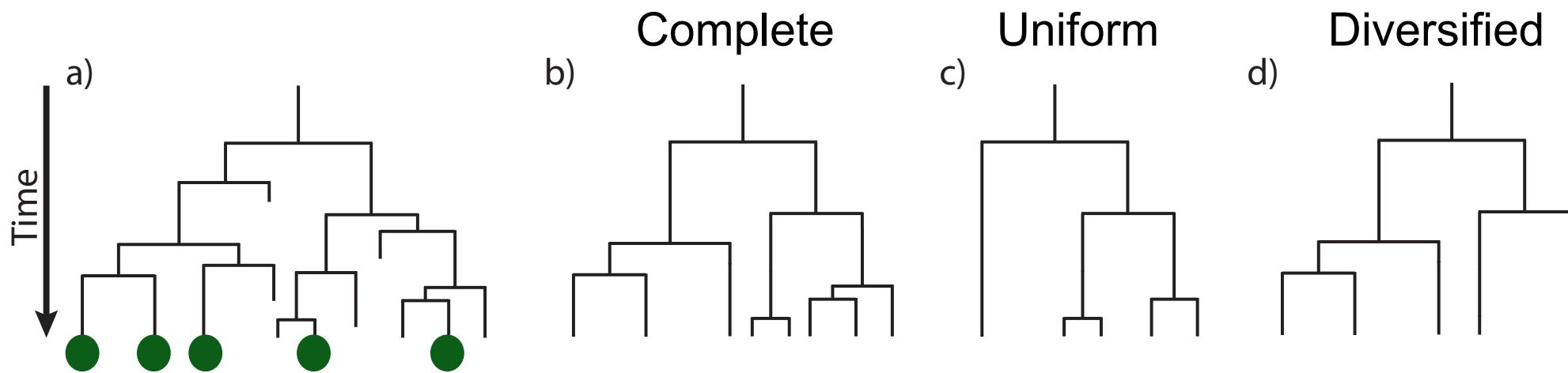
# Incomplete taxon sampling



Uniform taxon sampling:

Each extant taxon has the same probability to be included.

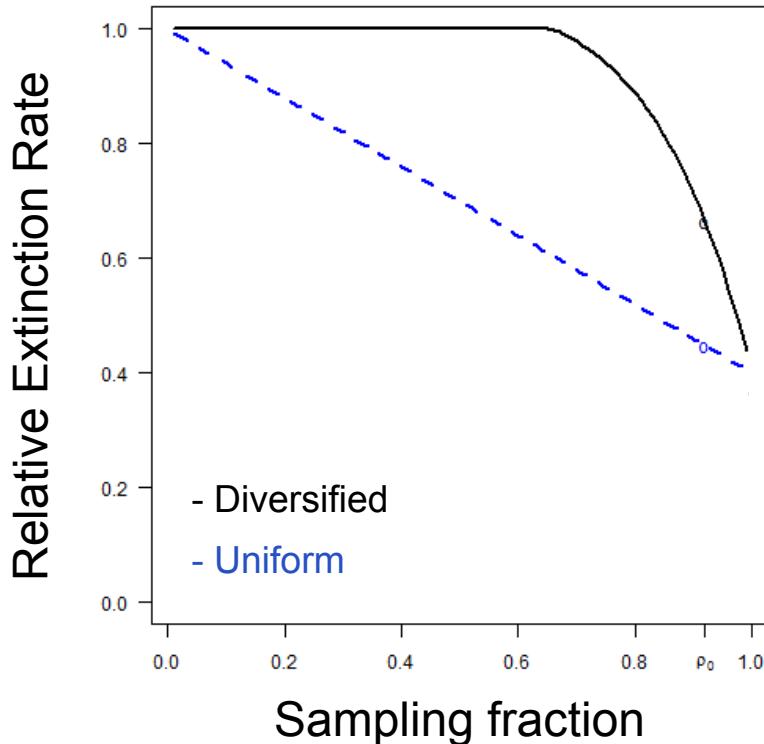
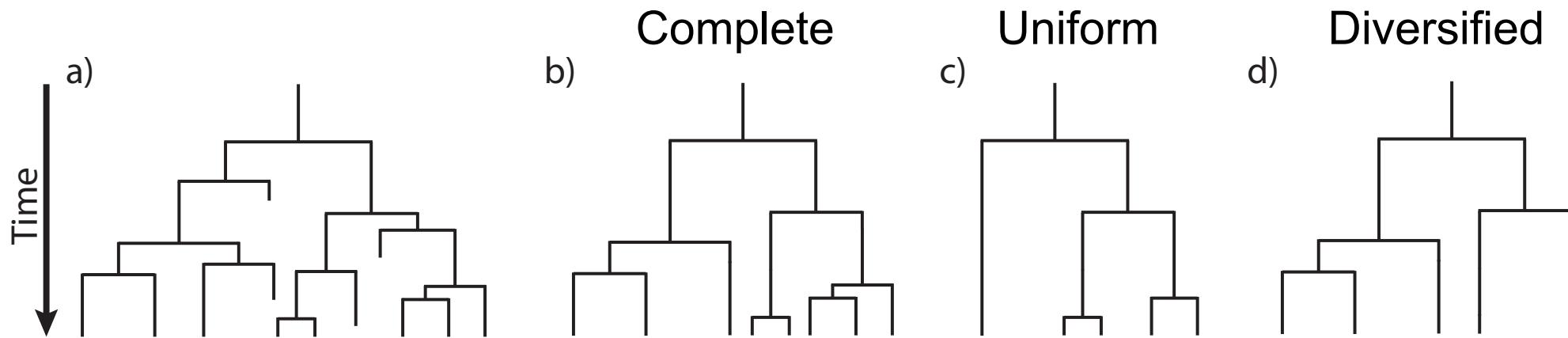
# Incomplete taxon sampling



Diversified taxon sampling:

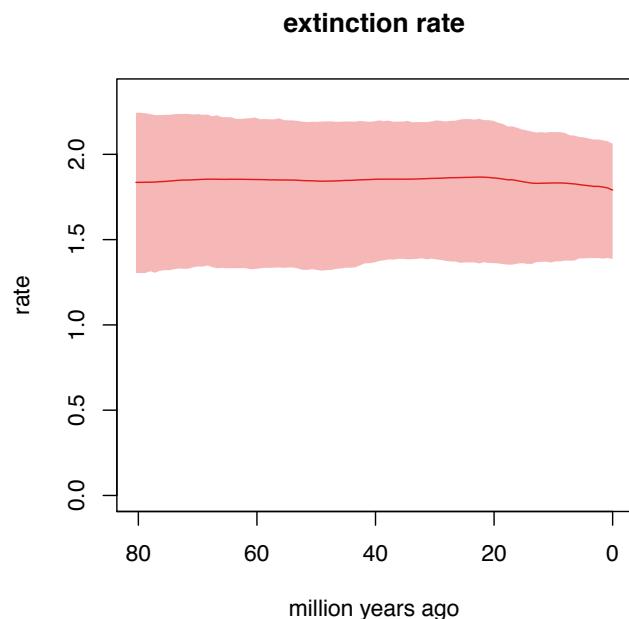
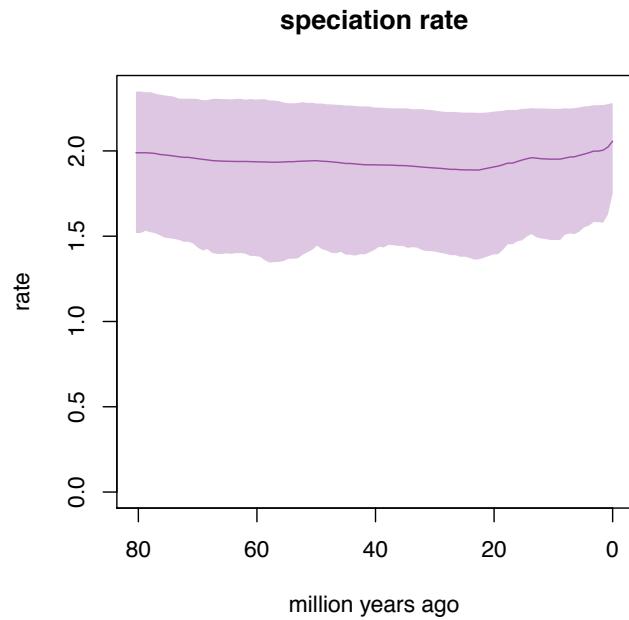
Only one taxon per group (e.g., genus) is included.

# Incomplete taxon sampling

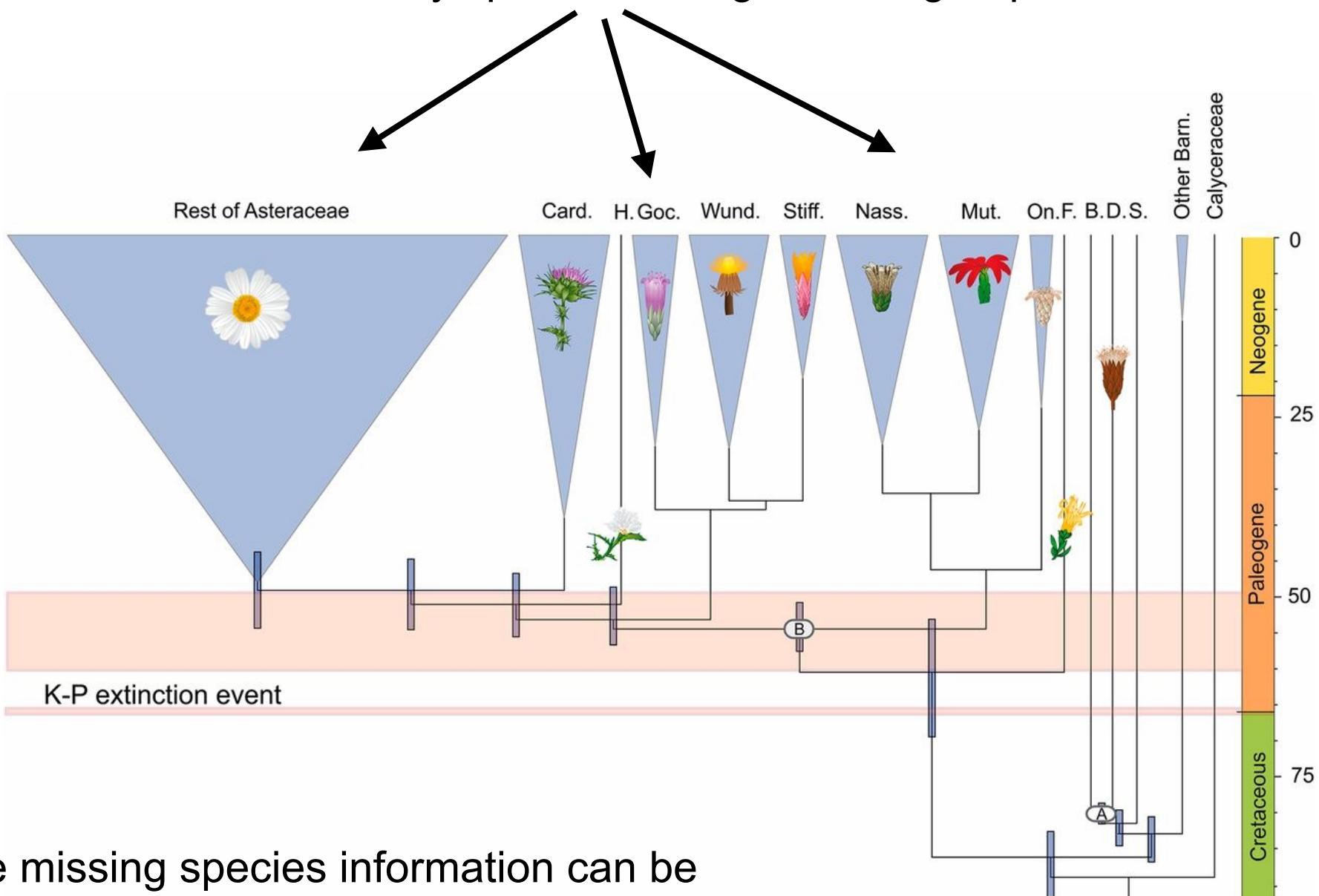


Incomplete taxon sampling  
biases diversification rate  
estimate!

# Uniform Sampling

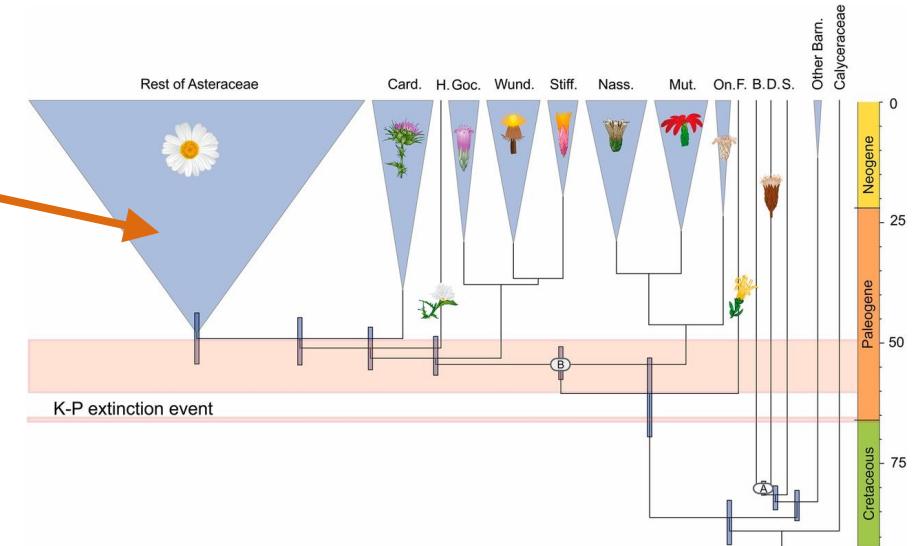
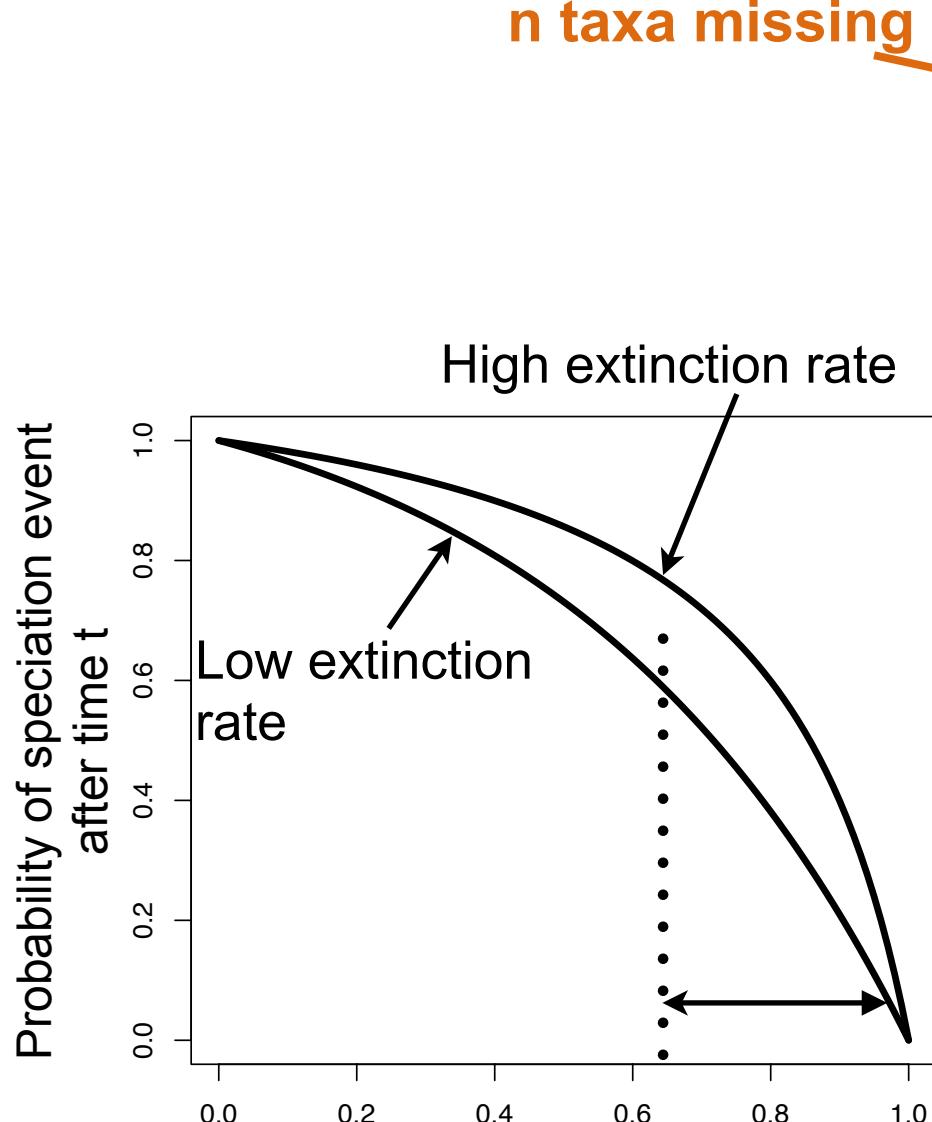


We know how many species belong to each group!



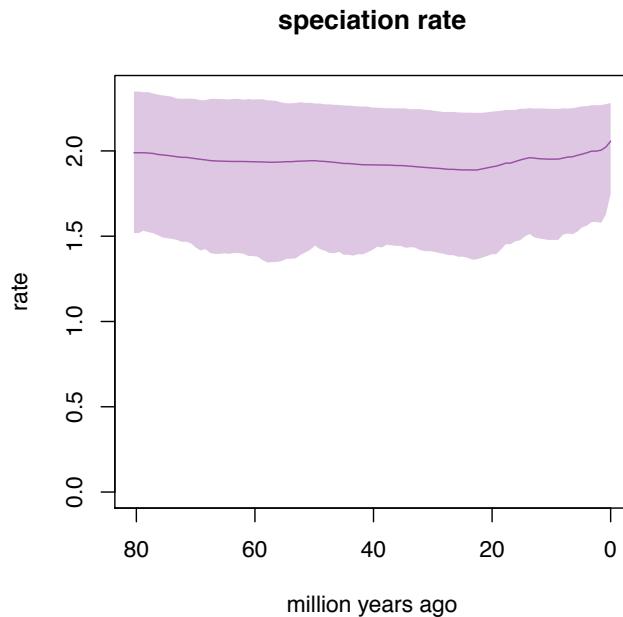
The missing species information can be used to inform diversification rate estimates.

# Empirical taxon sampling

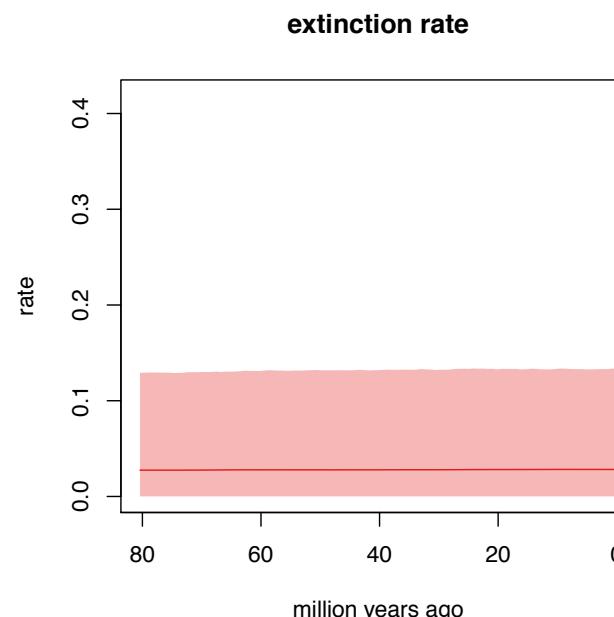
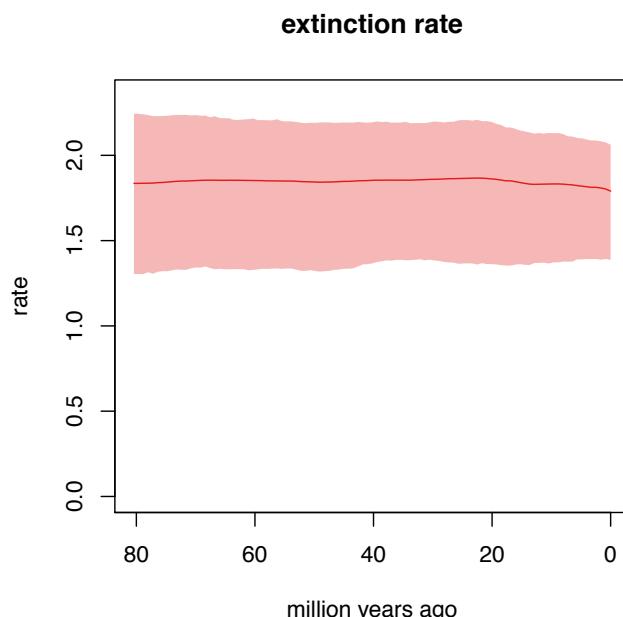
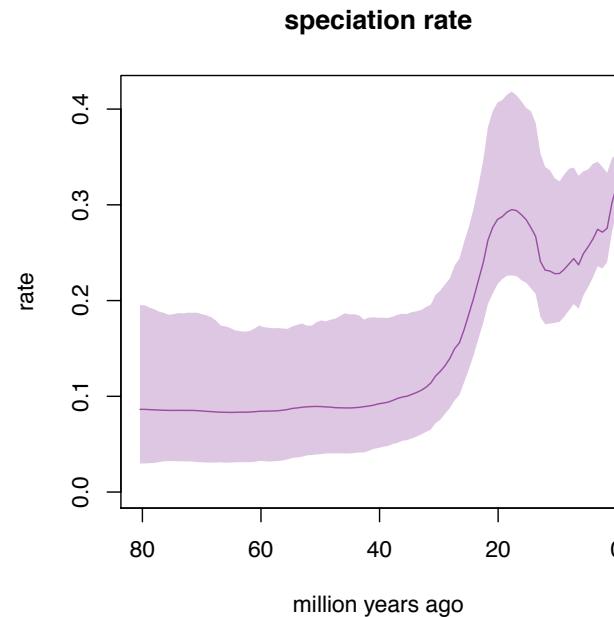


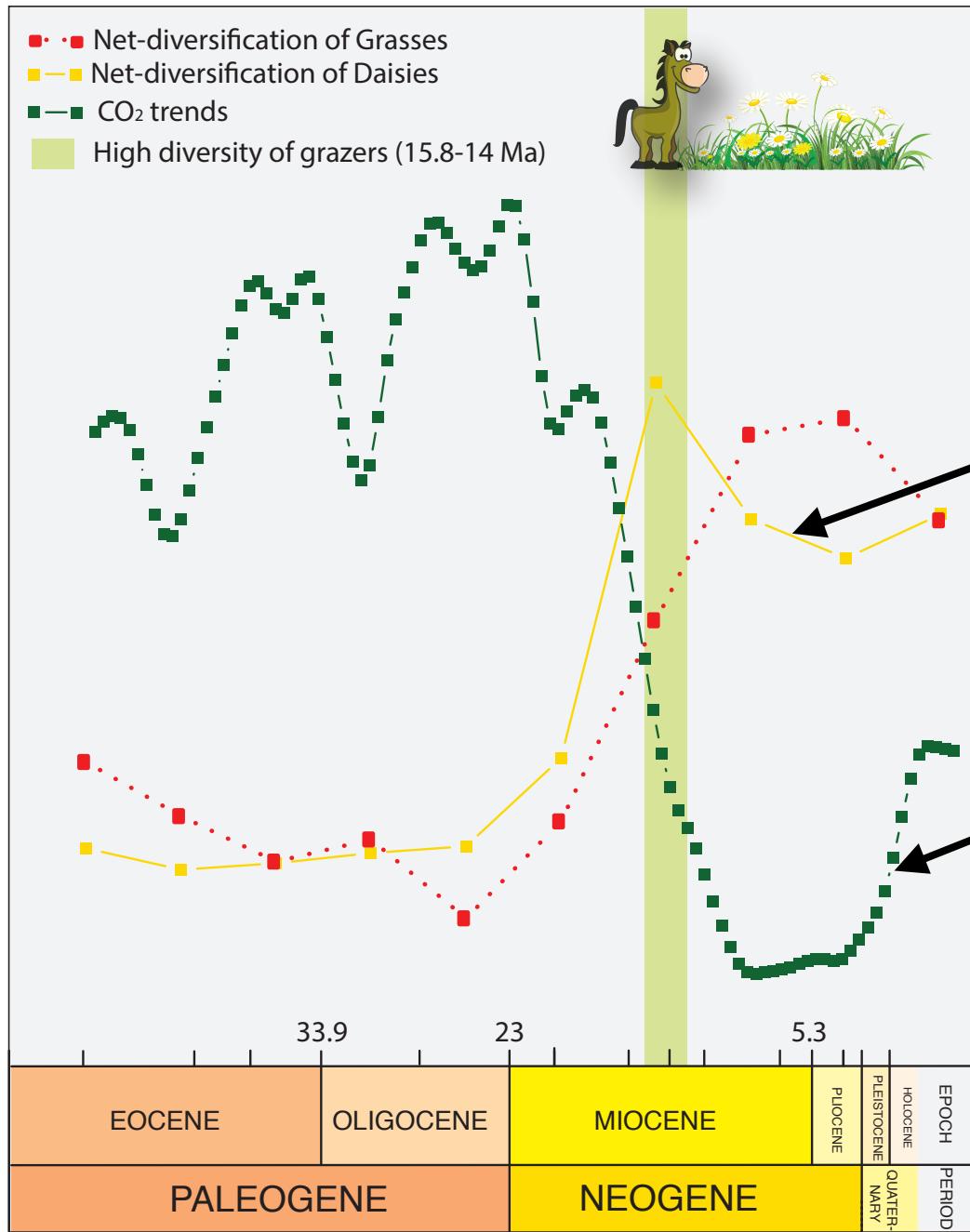
n events  
within interval

# Uniform Sampling



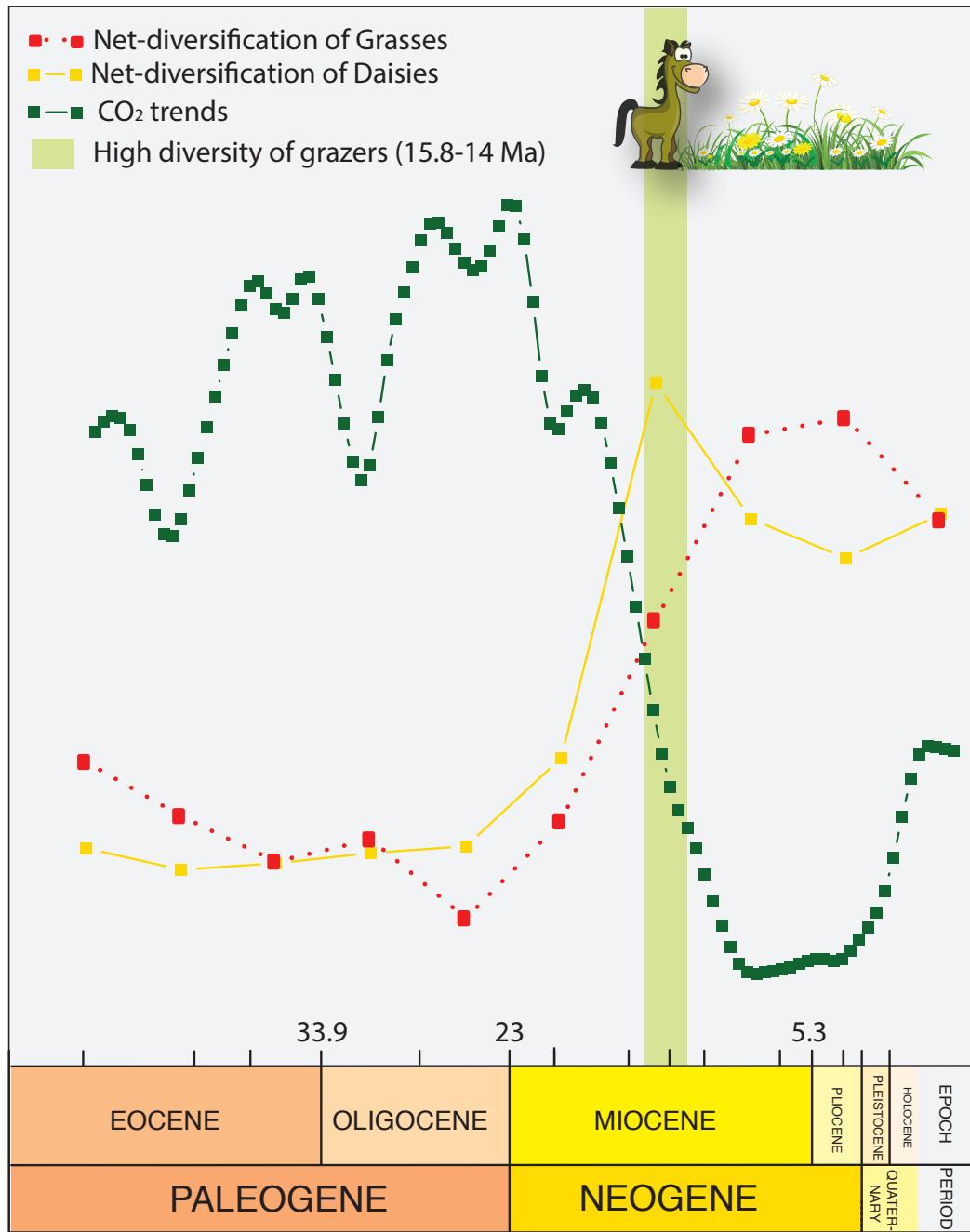
# Empirical Sampling





Inferred net-diversification  
(speciation-extinction)  
rate through time

Inferred CO<sub>2</sub>  
through time



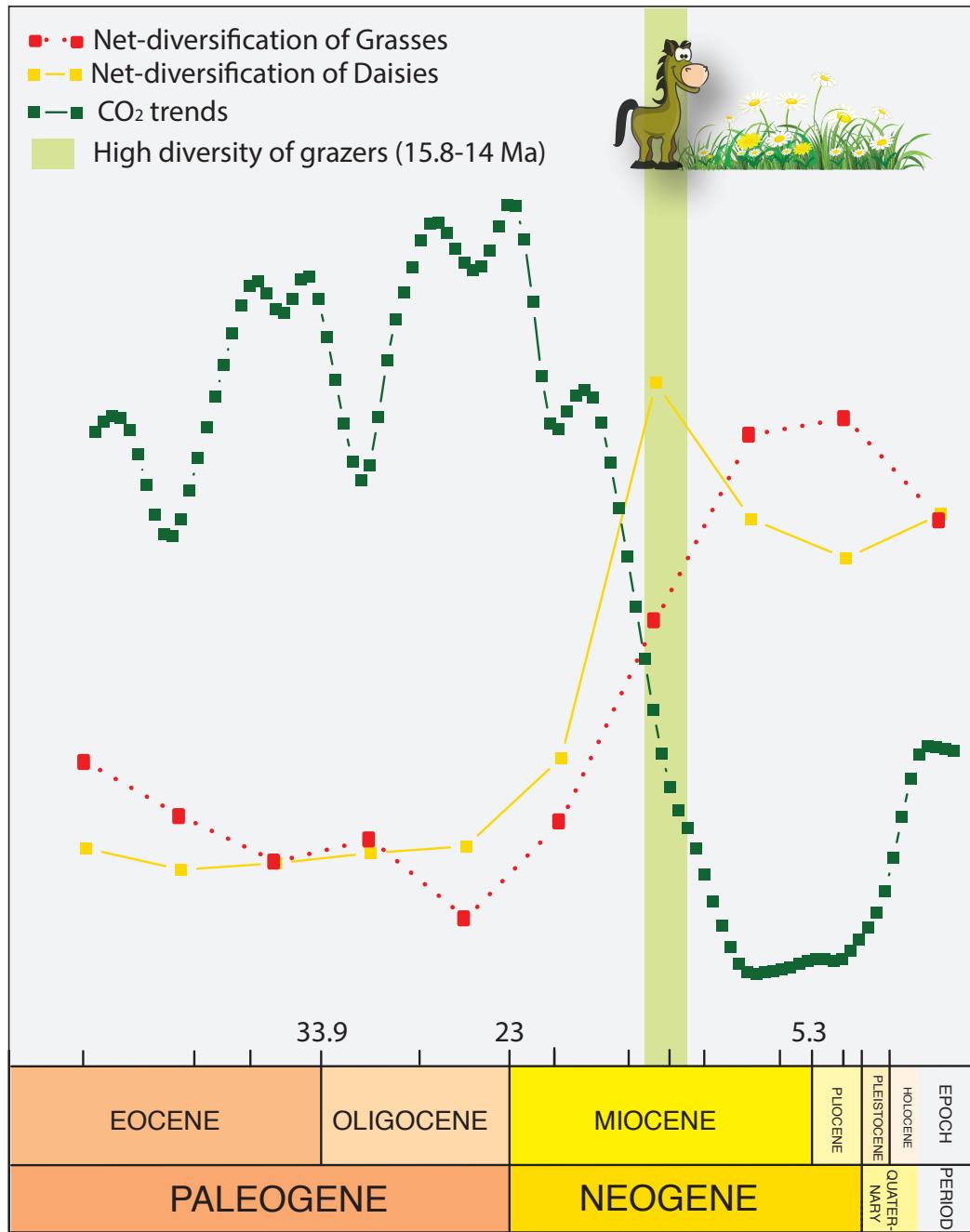
$$\lambda = \alpha_\lambda + \beta_\lambda * \text{CO}_2 + \epsilon_\lambda$$

$$\mu = \alpha_\mu + \beta_\mu * \text{CO}_2 + \epsilon_\mu$$

$\alpha$  = background rate

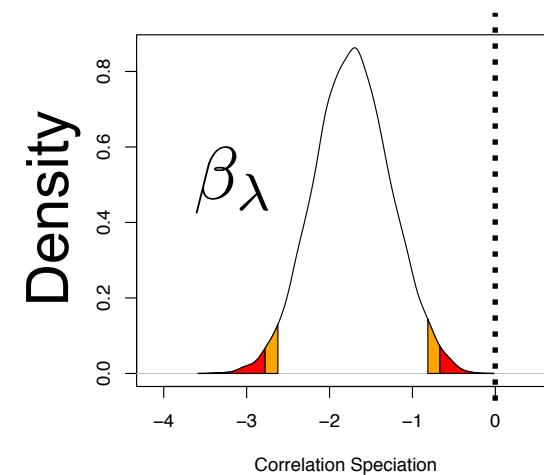
$\beta$  = regression slope

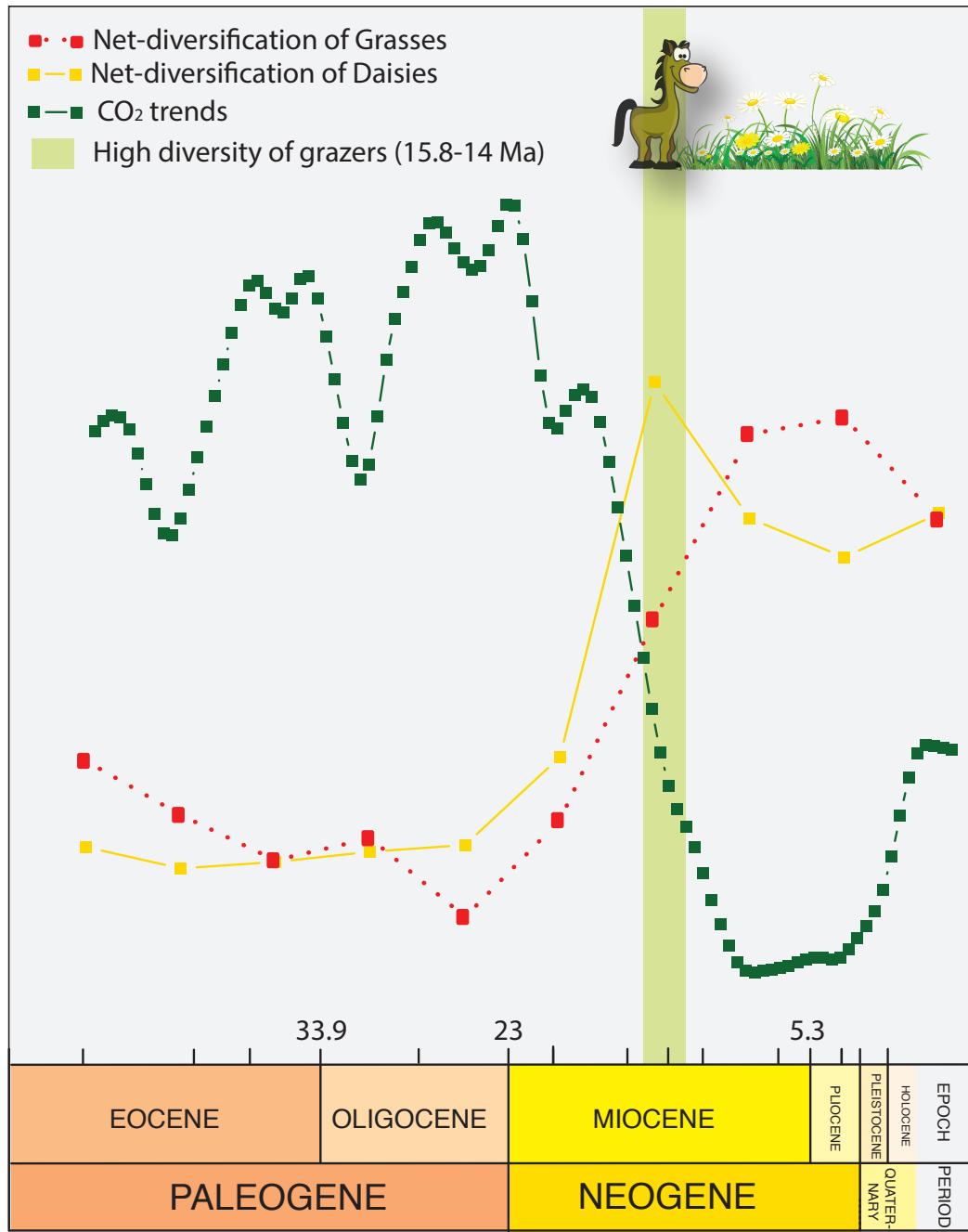
$\epsilon$  = residual



$$\lambda = \alpha_\lambda + \beta_\lambda * \text{CO}_2 + \epsilon_\lambda$$

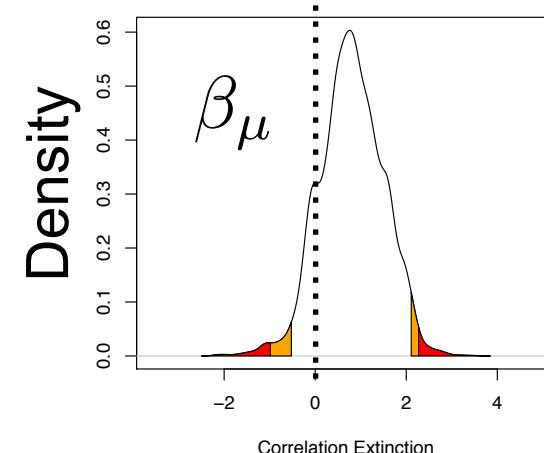
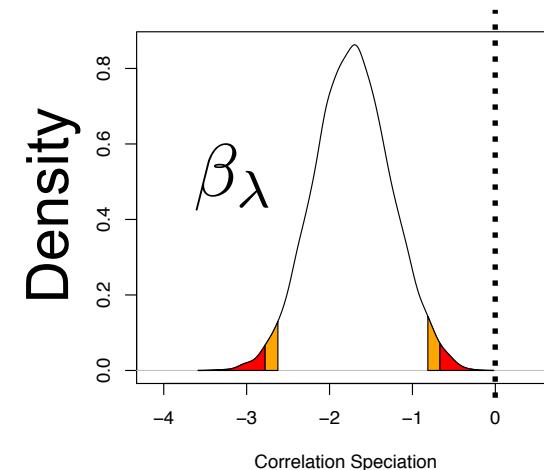
$$\mu = \alpha_\mu + \beta_\mu * \text{CO}_2 + \epsilon_\mu$$





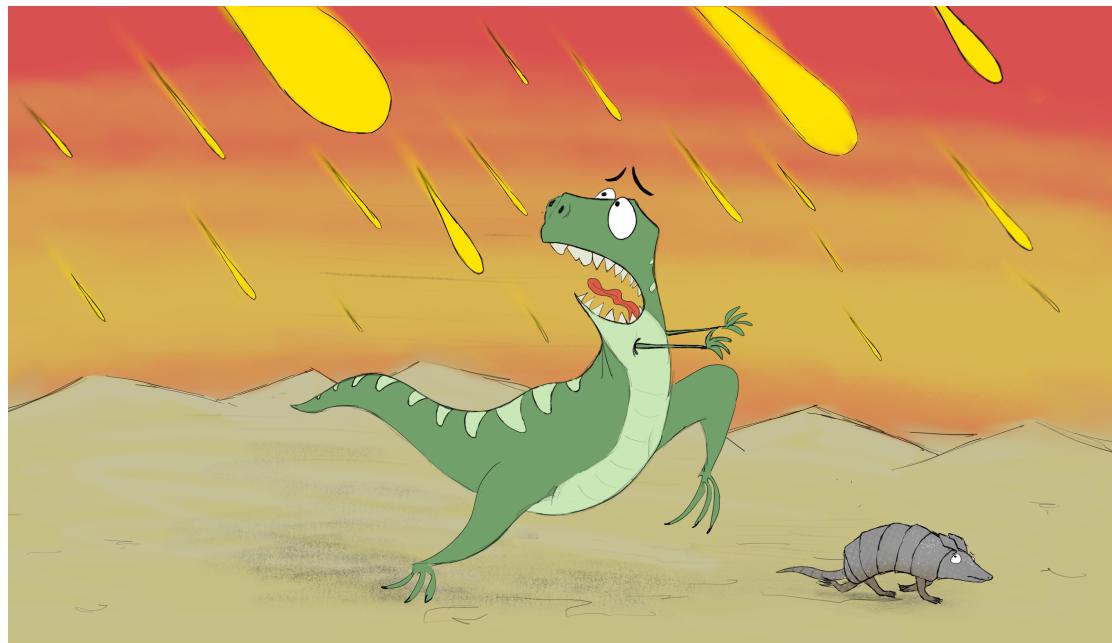
$$\lambda = \alpha_\lambda + \beta_\lambda * \text{CO}_2 + \epsilon_\lambda$$

$$\mu = \alpha_\mu + \beta_\mu * \text{CO}_2 + \epsilon_\mu$$



# Mass-extinction events

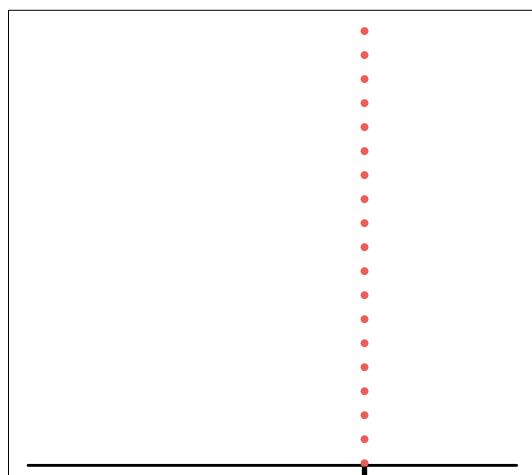
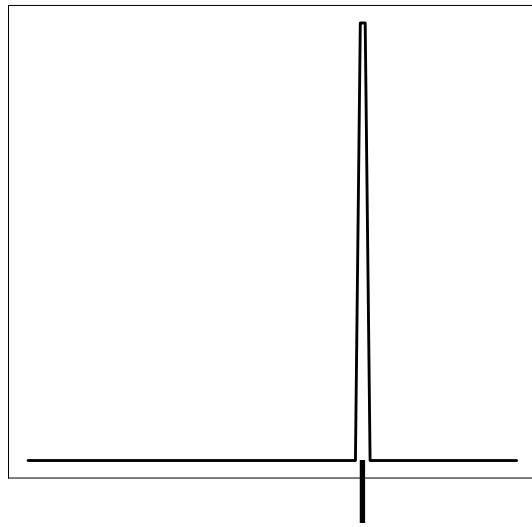
---



Mass-extinction events are:

- short periods of time with high extinction
- 50–75% of genera go extinct
- around 90% of species diversity goes extinction

# Modeling mass extinction



Mass-extinction  
event

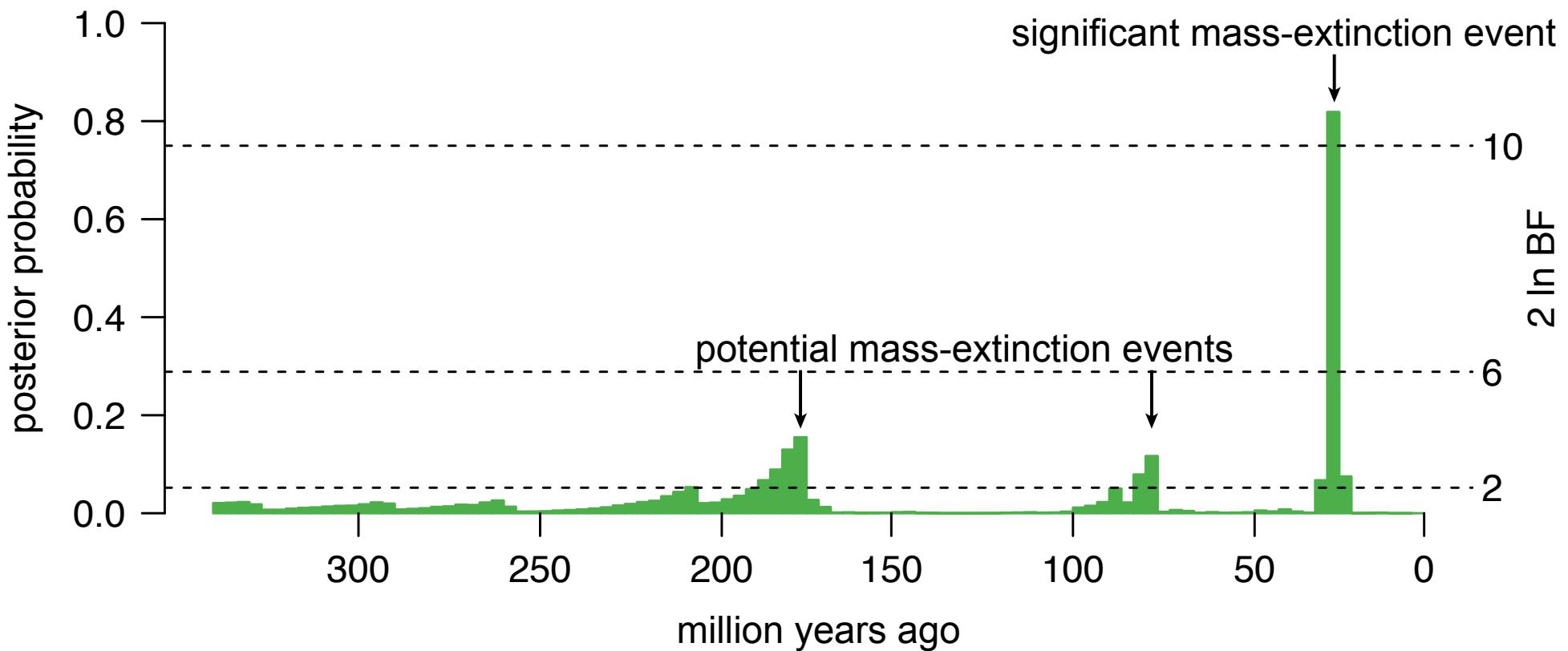
Mass-extinction events can be modeled by:

- a) a short period of time with high extinction
- b) an explicit mass-extinction event with a given survival probability

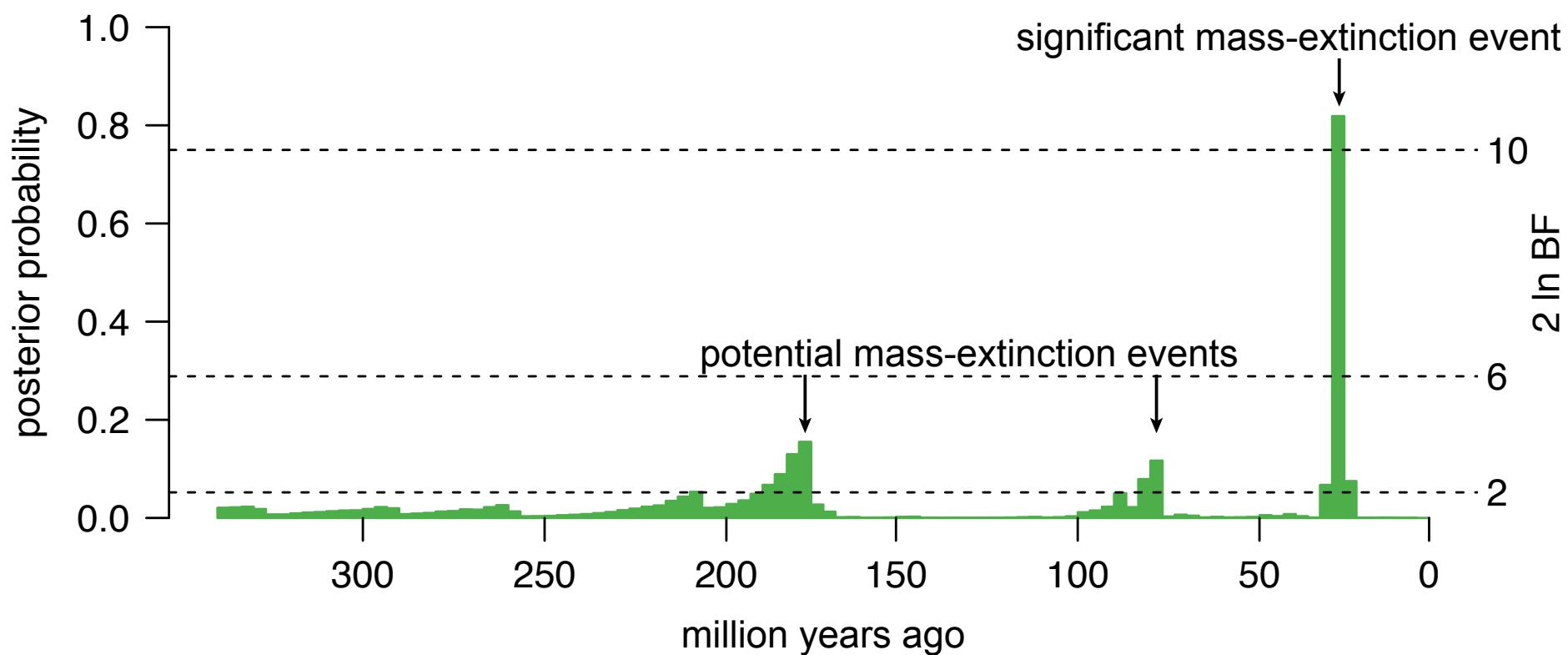
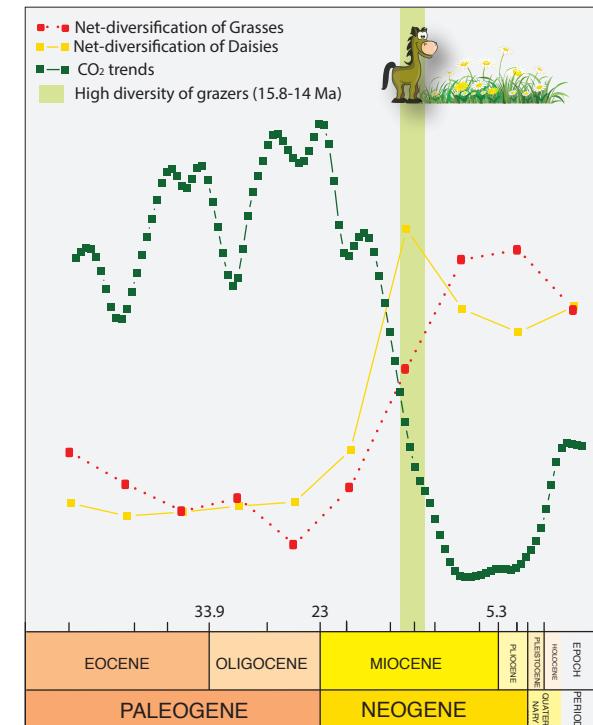
*using explicit mass extinction and biologically informed priors we can distinguish between background rate variation and mass extinctions*

# Example: Inference of mass-extinction events in

$$\text{Bayes factor: } \frac{\text{posterior mass extinction}}{\text{posterior no mass extinction}} \div \frac{\text{prior mass extinction}}{\text{prior no mass extinction}}$$



- Conifers were impacted by a mass extinction about 23 Mya
- Daisies and grasses increased in diversity about 20 Mya
- Diversification rates in daisies are linked to CO<sub>2</sub>



# **Exercise 2: Episodic Diversification Rate Estimation**

## **1) Estimating constant rates of speciation and extinction**

- What are the speciation and extinction rates for my study group?

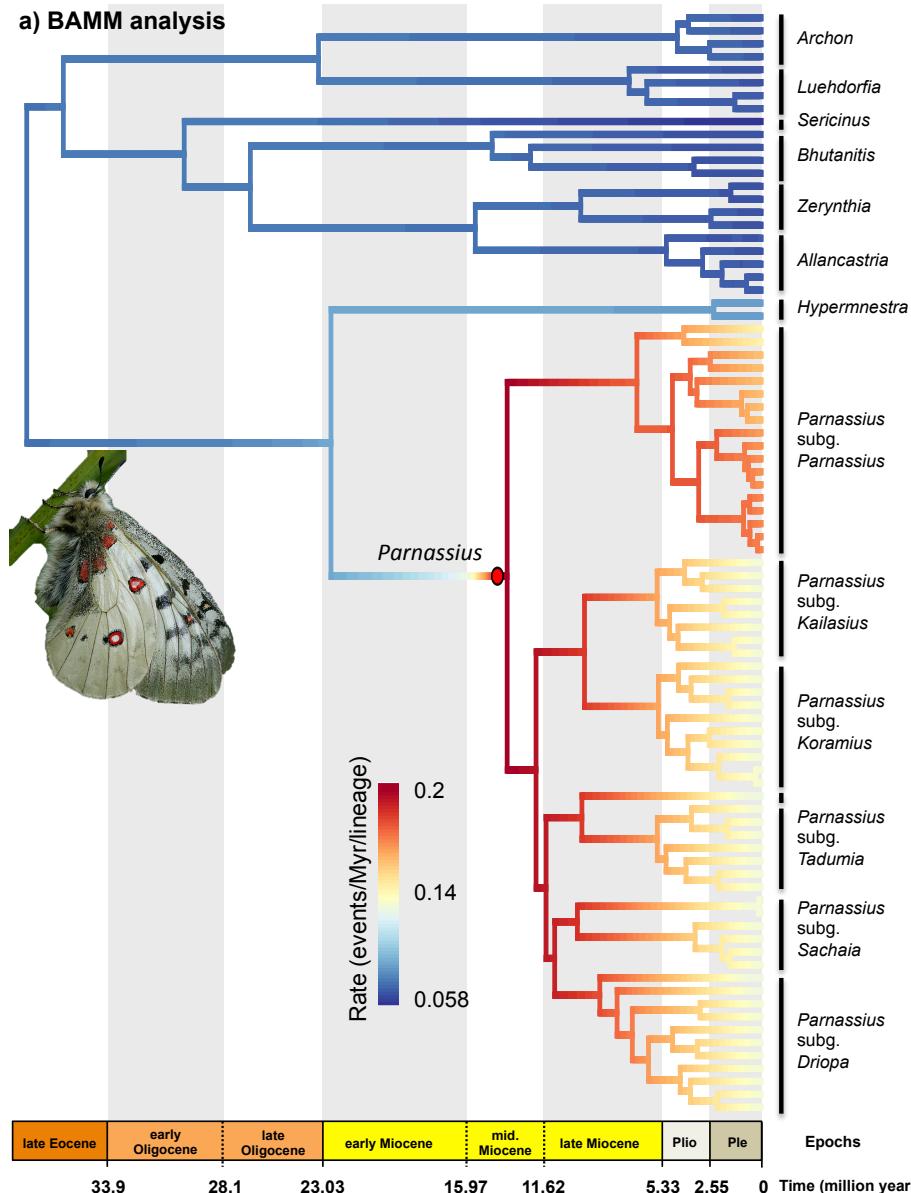
## **2) Estimating time-varying rates of speciation and extinction**

- Do speciation and extinction rates vary through time?
- Is this variation through time correlated with some abiotic factor?

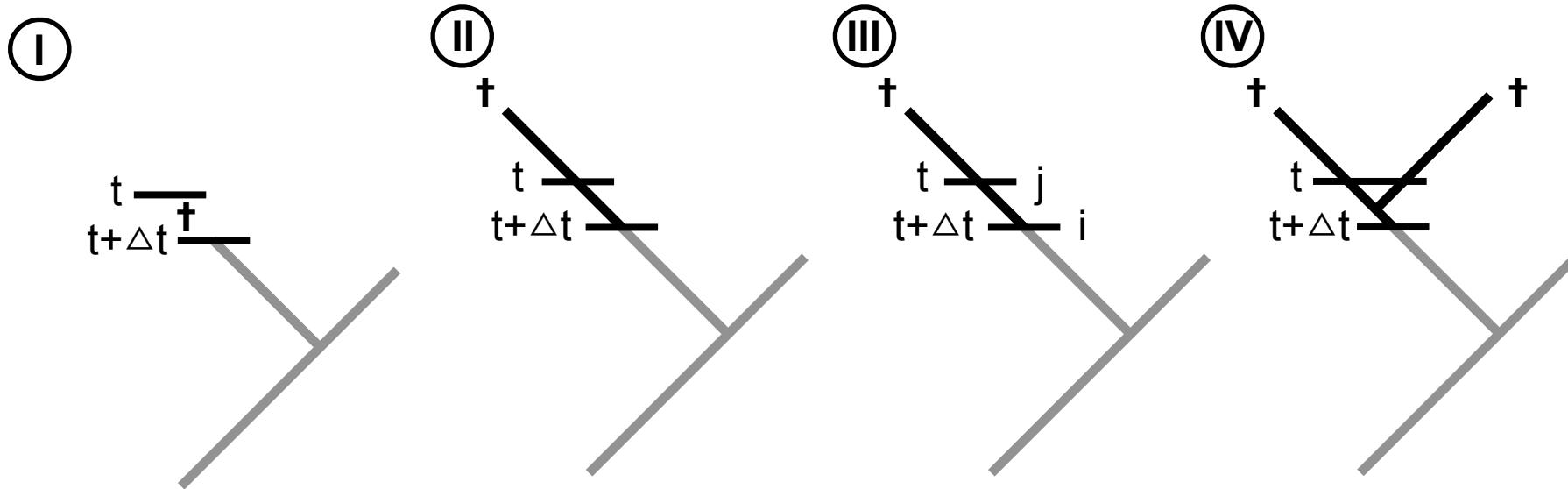
## **3) Estimating branch-specific rates of speciation and extinction**

- Do speciation and extinction rates vary among lineages?
- Is this variation among lineages correlated with some biotic factor?

# Example: Diversification in Apollo Butterflies



# Theory: Computing the Probability of



$$E_i(t + \Delta t) = \mu_i \Delta t$$

(i) extinction

$$+ \left[ 1 - \mu_i \Delta t - \lambda_i \Delta t - \sum_{i \neq j} (q_{ij} \Delta t) \right] E_i(t)$$

(ii) no event

$$+ \left[ 1 - \mu_i \Delta t - \lambda_i \Delta t \right] \times \sum_{i \neq j} (q_{ij} \Delta t E_j(t))$$

(iii) a state change event

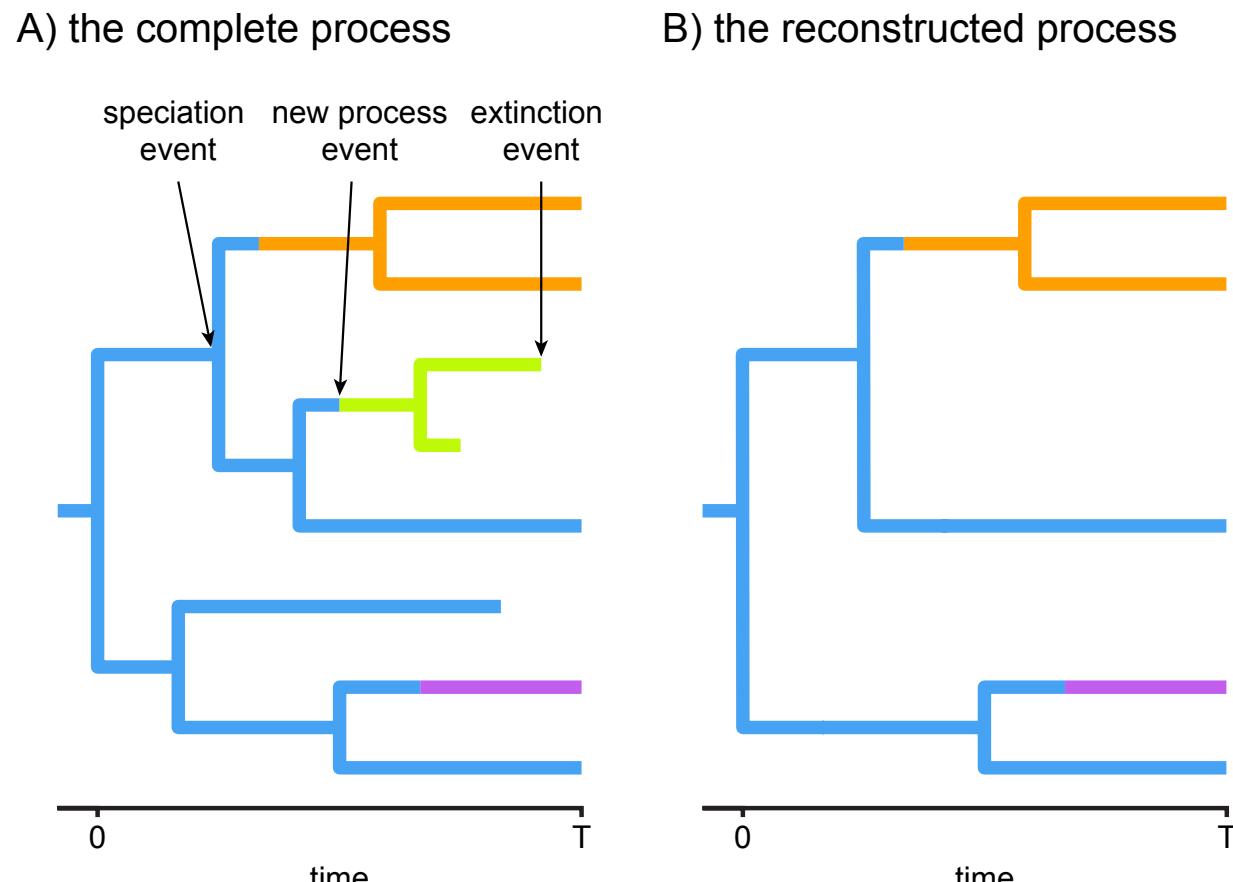
$$+ \left[ 1 - \mu_i \Delta t - \sum_{i \neq j} (q_{ij} \Delta t) \right] \lambda_i \Delta t E_i(t)^2$$

(iv) a speciation event

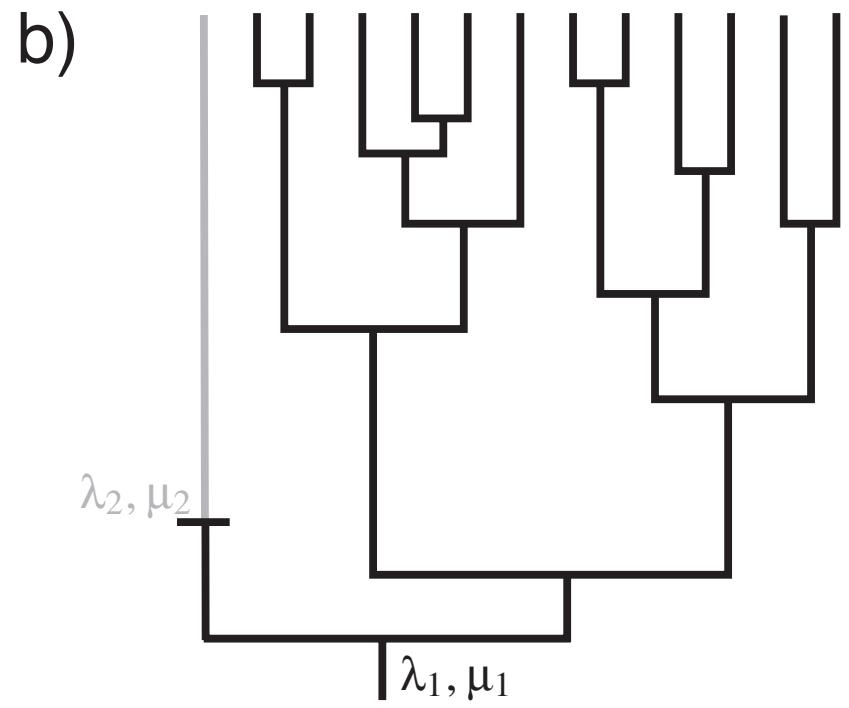
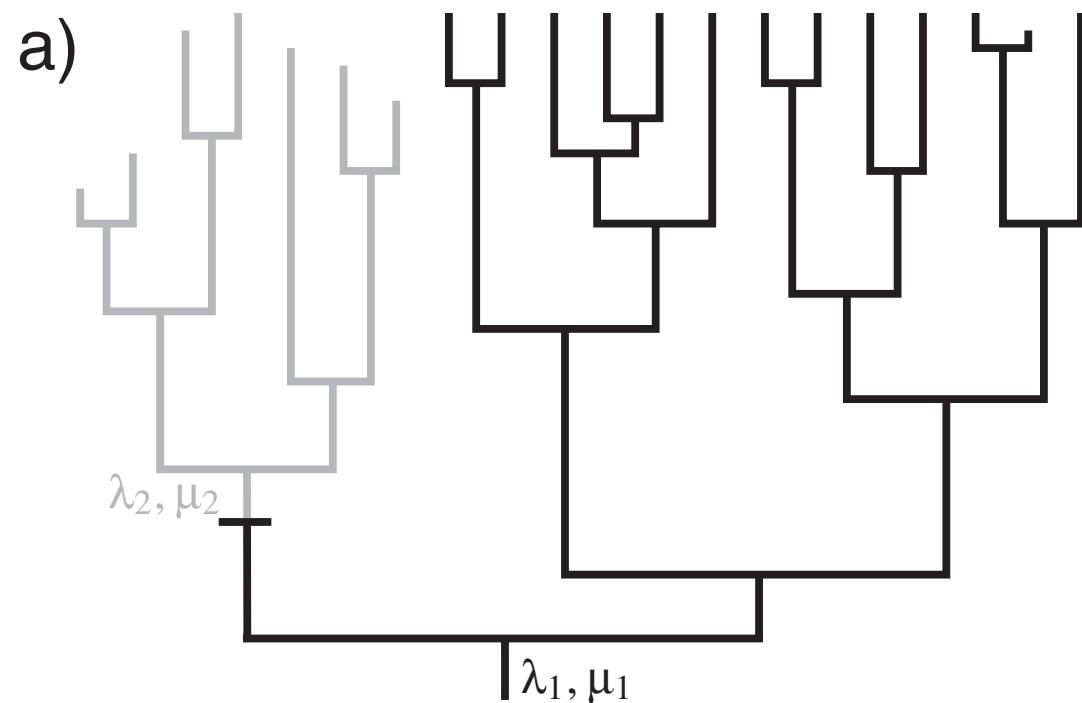
# Branch-Specific Diversification

Lineages / clades have heterogeneous diversification rates.

Locations of shifts are defined a priori (prior hypothesis) or estimated.

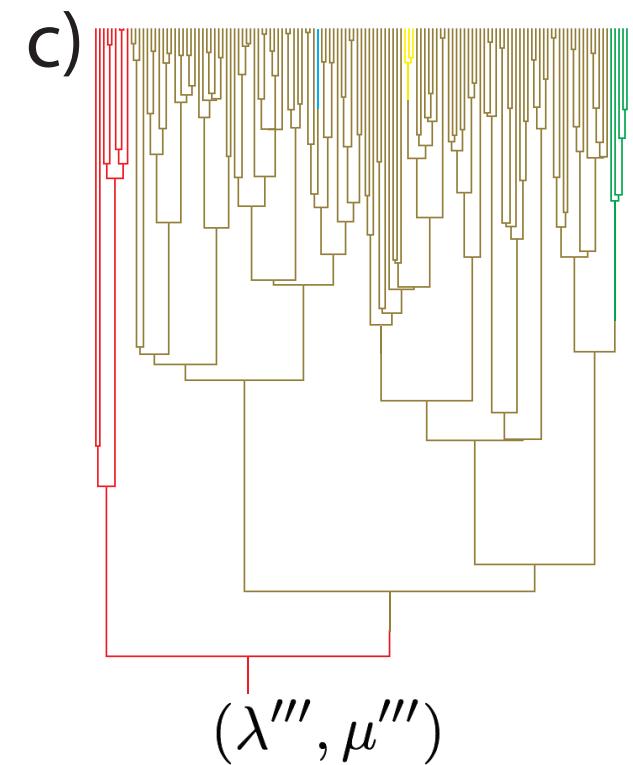
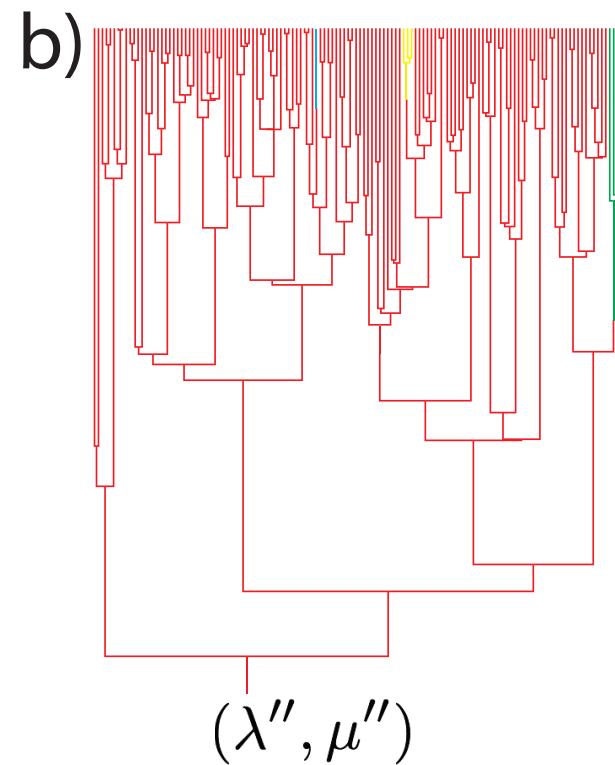
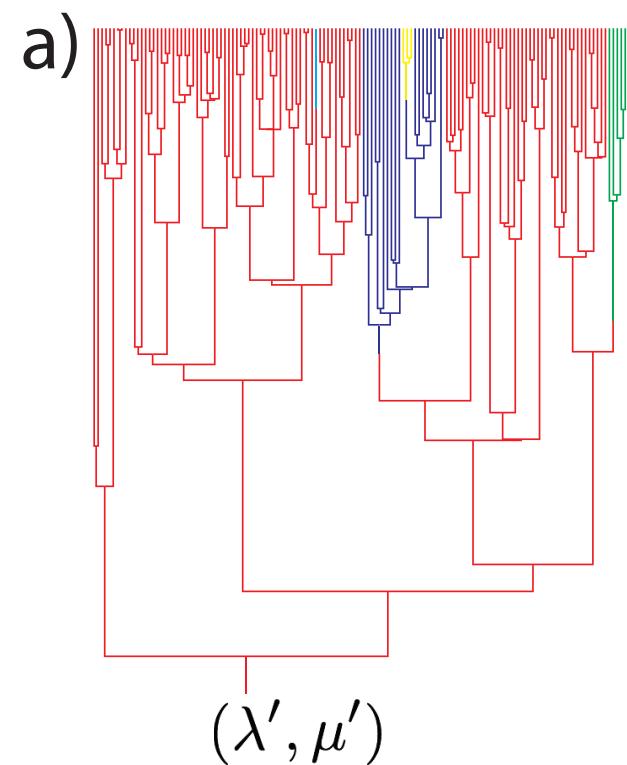


# Full tree vs reconstructed tree



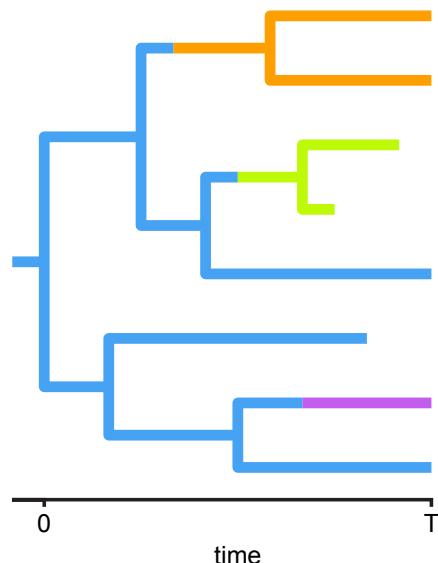
# Data Augmentation

---

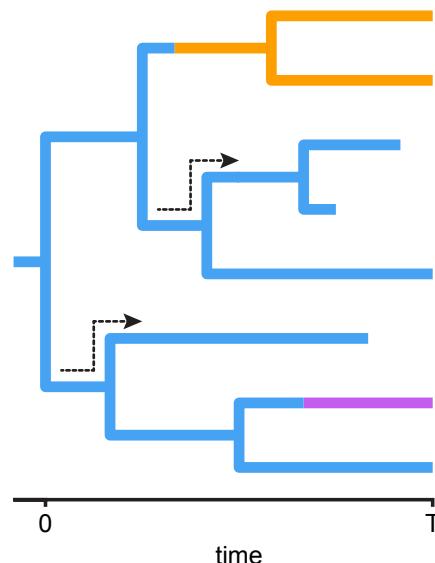


# Modeling Challenges

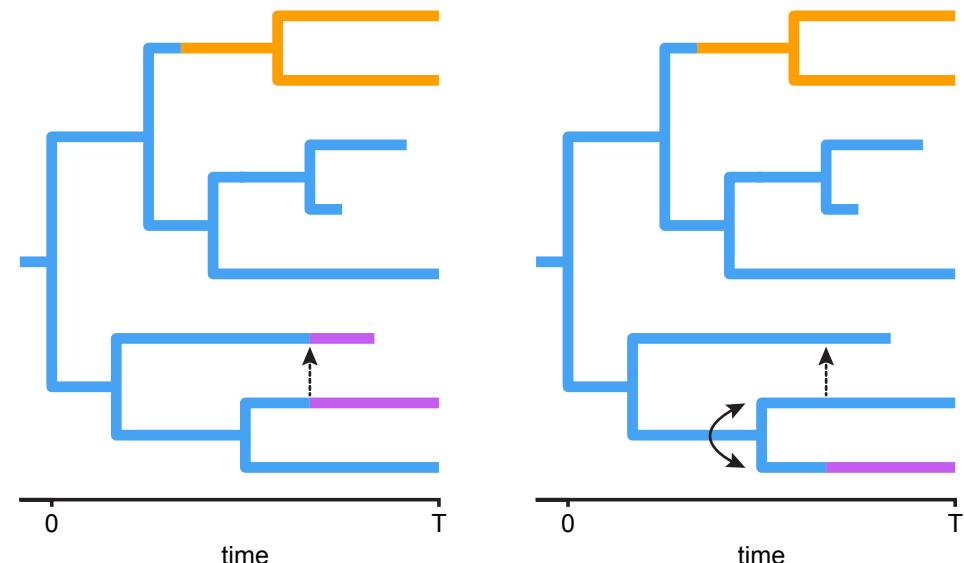
A) actual process  
(process may vary on extinct lineages)



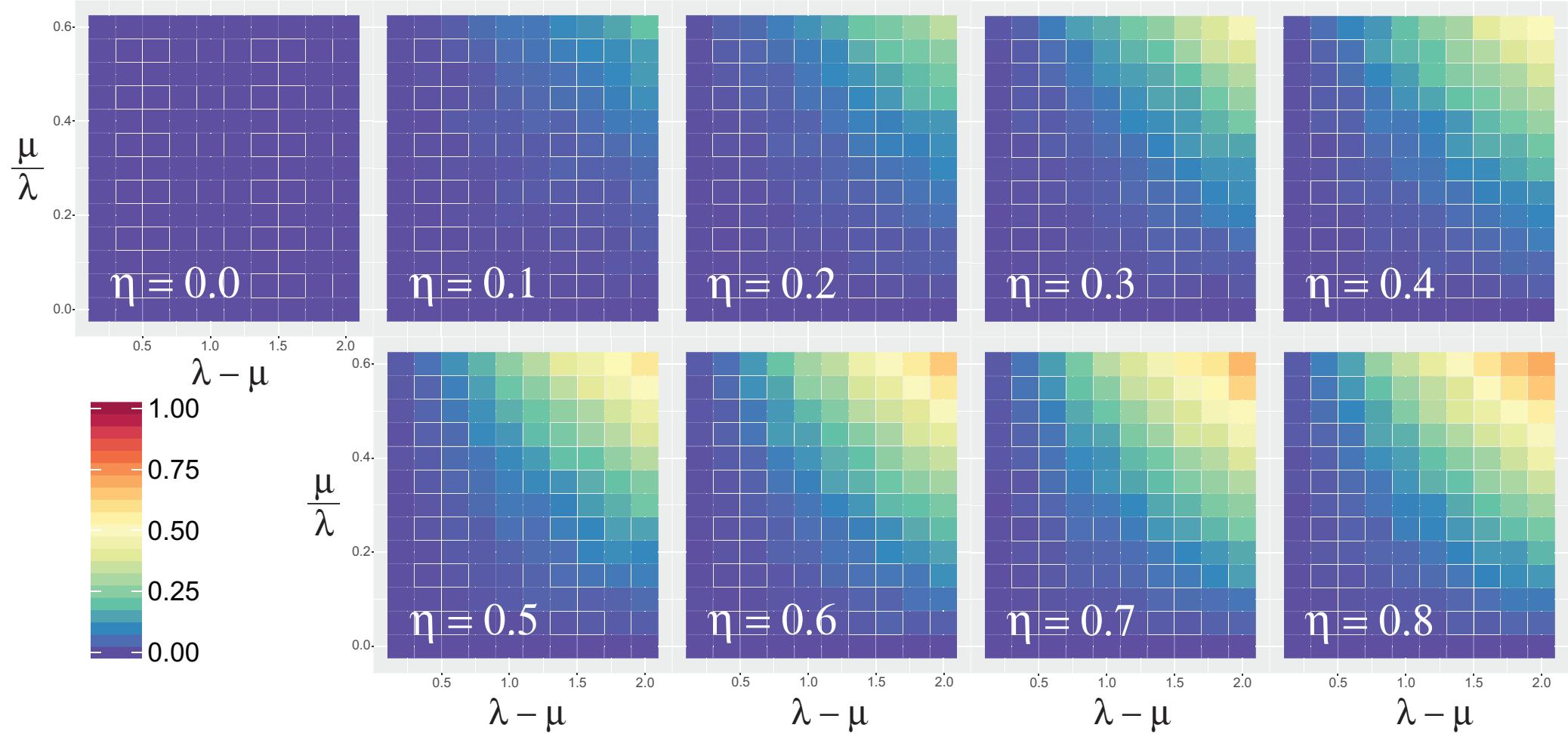
B) described process  
(extinct lineages inherit ancestral process)



C) implemented process  
(extinct lineages laterally inherit the process of the left but not the right observed branch)

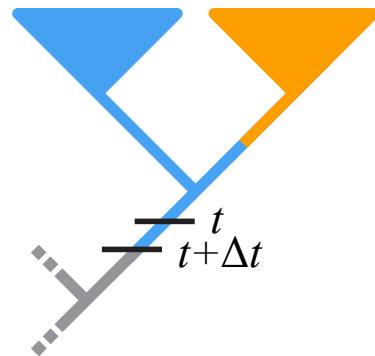


# Percentage of trees with rate shifts at extinct

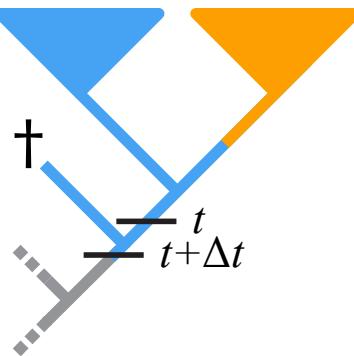


# Branch-specific likelihood in RevBayes

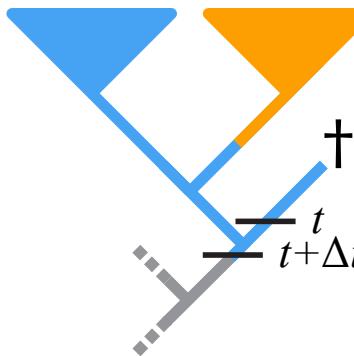
A) Scenarios for observed lineages



i. no speciation

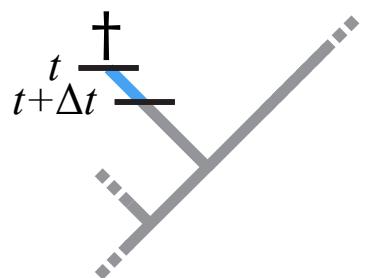


ii. speciation left branch,  
subsequent extinction

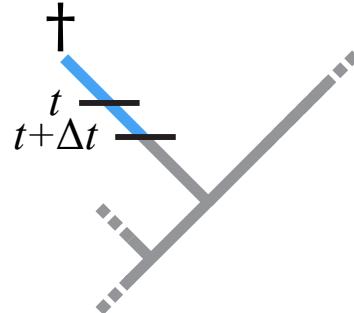


iii. speciation right branch,  
subsequent extinction

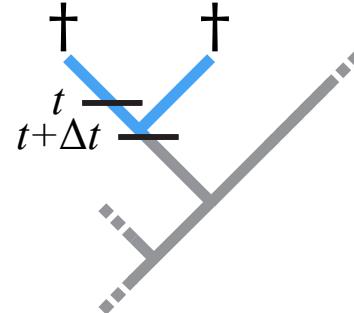
B) Scenarios for unobserved lineages



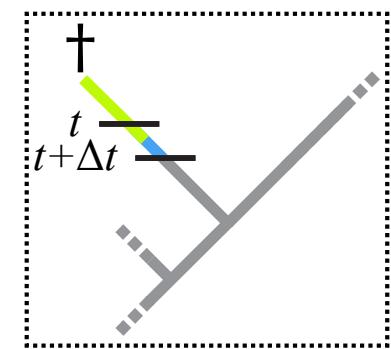
i. no rate shift,  
no speciation,  
extinction



ii. no rate shift,  
no speciation,  
subsequent extinction



iii. no rate shift,  
speciation,  
subsequent extinctions



iv. rate shift,  
no speciation,  
subsequent extinction

# The Finite Rate Category Model

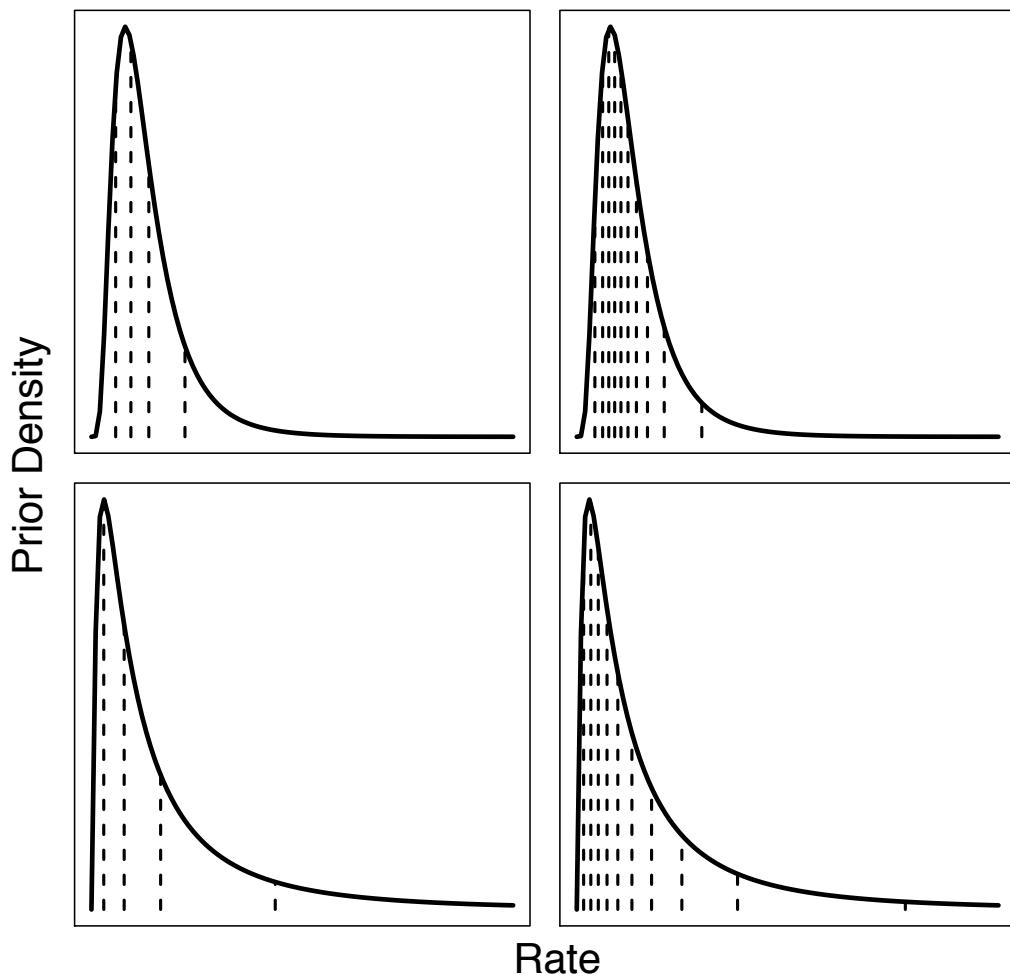
---

We assume an underlying birth-death-shift model:

1. Species  $i$  speciates with a given speciation rate  $\lambda_i$ . Both descendant species inherit the speciation and extinction rates.
2. Species  $i$  dies with a given extinction rate  $\mu_i$ .
3. Species  $i$  changes to new speciation and extinction rates which are drawn from the discrete prior distribution.

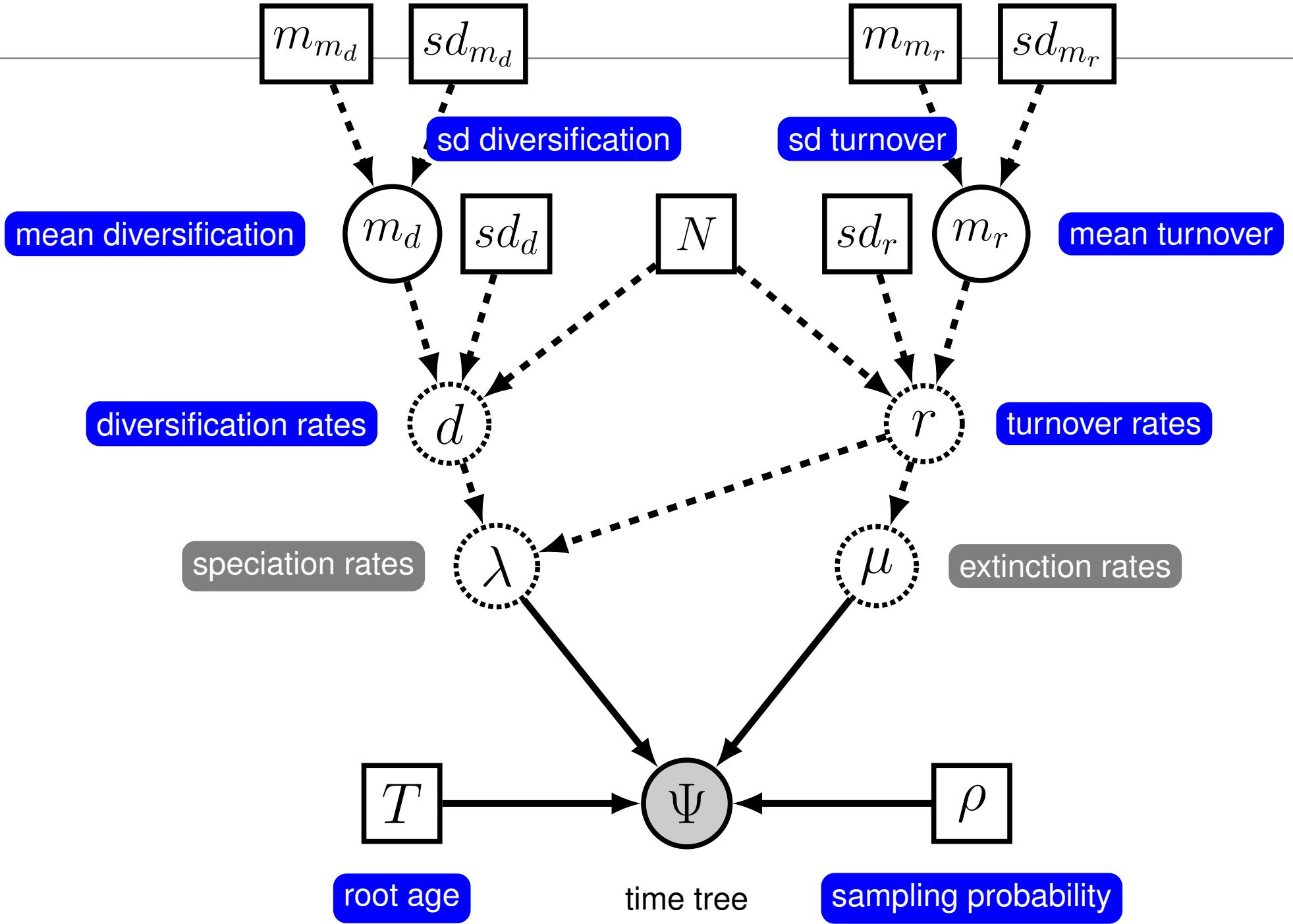
- **We allow for extinction events to occur on extinct lineages too!**
- We can calculate the probability of the tree numerically (slow).
- This is a corrected version of the BAMM model.

# The Finite Rate Category Model



Assume there are only  $N$  discrete rate categories.

Approximate continuous distribution by  $N$  quantiles of distribution (discretization).



# **Exercise 3: Branch-specific Diversification Rate Estimation**

## **1) Estimating constant rates of speciation and extinction**

- What are the speciation and extinction rates for my study group?

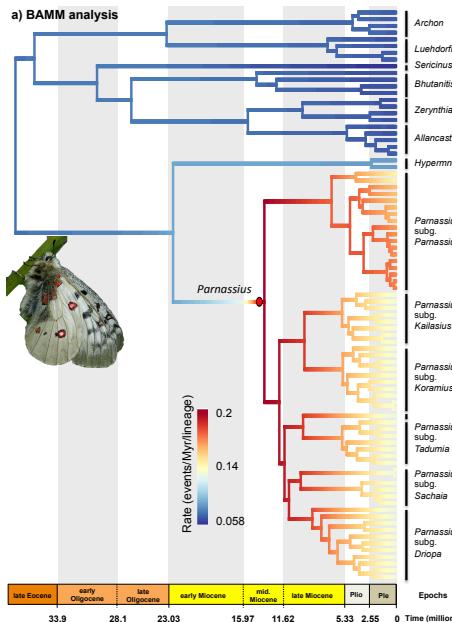
## **2) Estimating time-varying rates of speciation and extinction**

- Do speciation and extinction rates vary through time?
- Is this variation through time correlated with some abiotic factor?

## **3) Estimating branch-specific rates of speciation and extinction**

- Do speciation and extinction rates vary among lineages?
- Is this variation among lineages correlated with some biotic factor?

# Example: Diversification in Apollo Butterflies

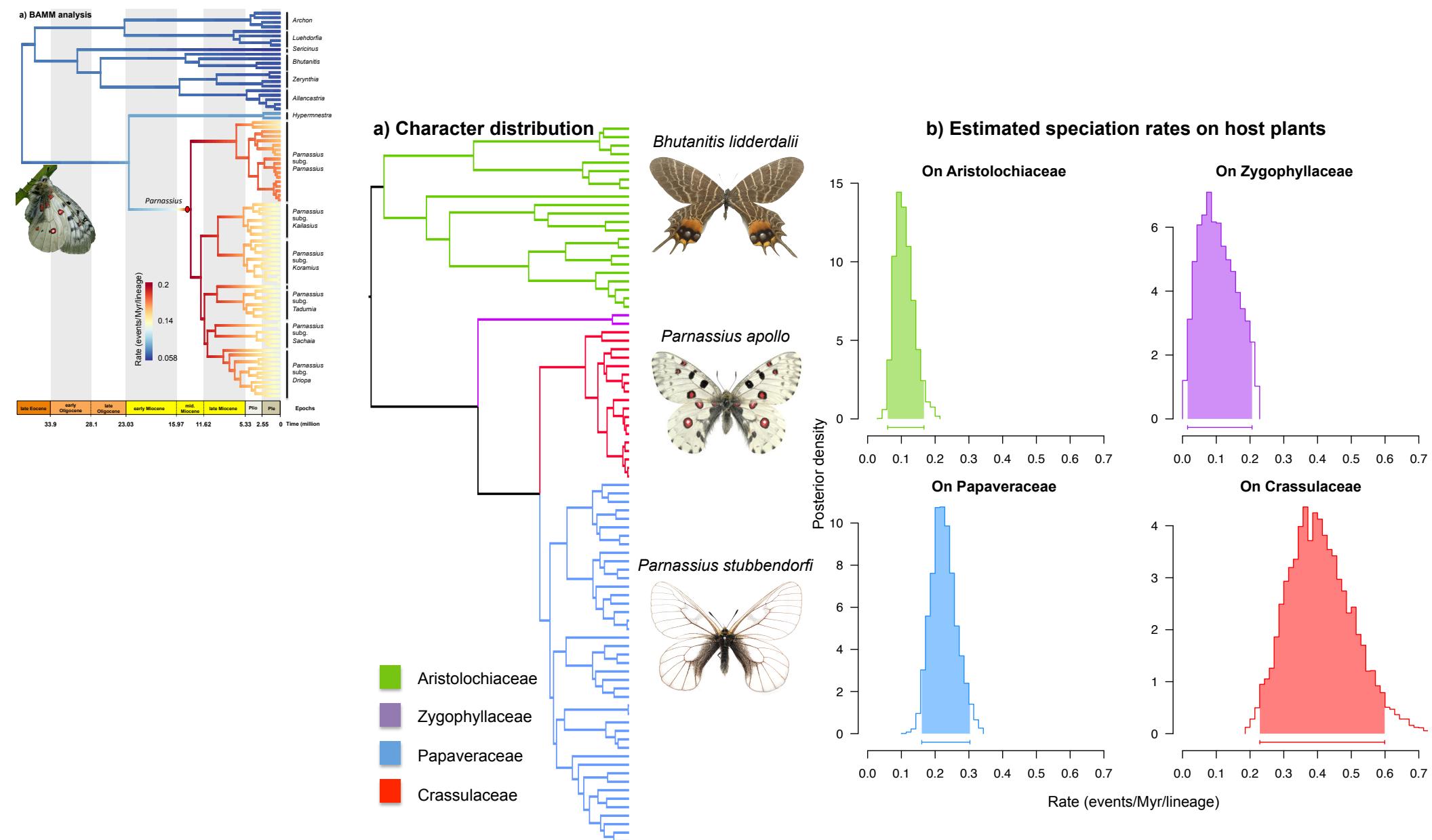


Speciation/extinction rates are dependent on character state.

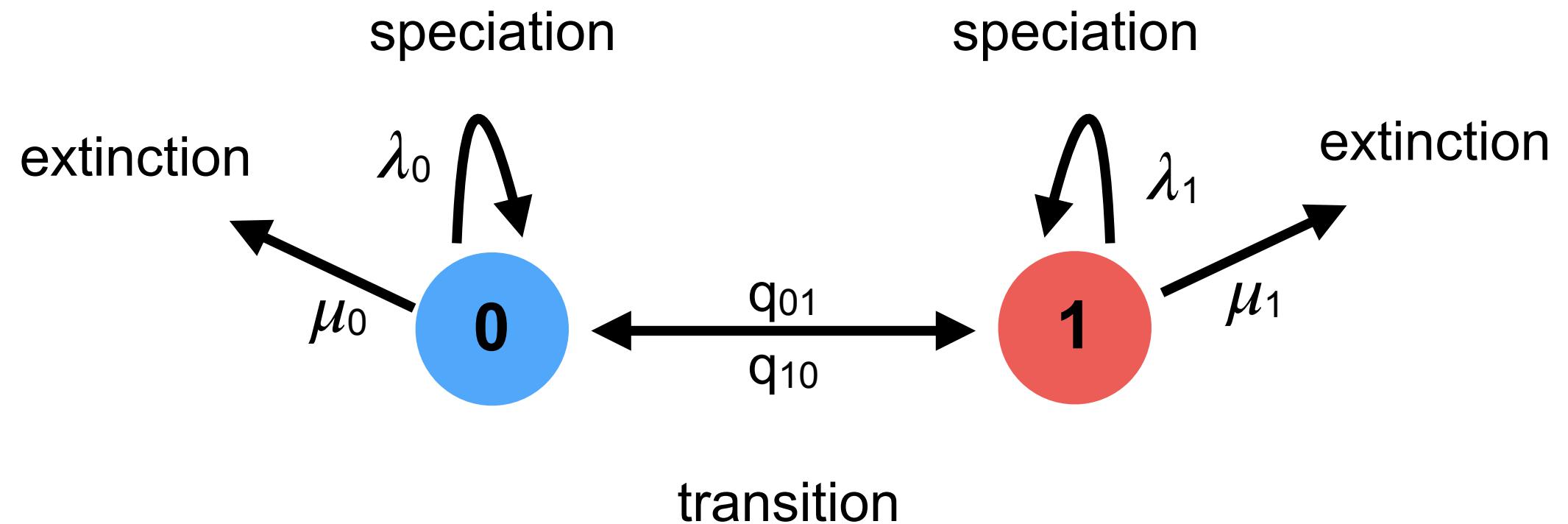
Character states are only observed at tips.

Joint model of diversification and character evolution.

# Example: Diversification in Apollo Butterflies

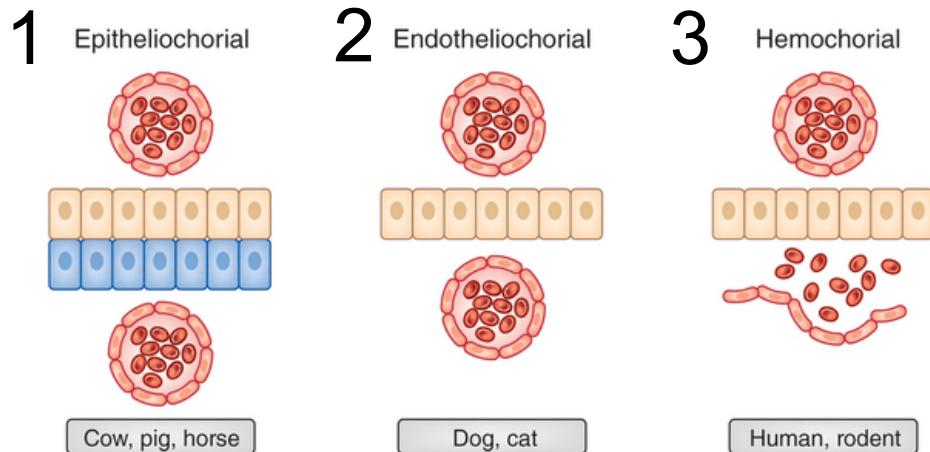


# Binary State Speciation & Extinction (BiSSE)

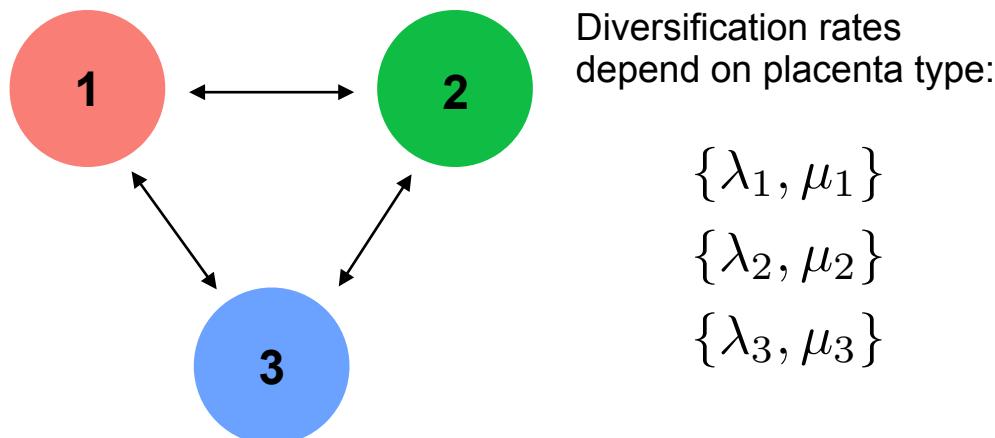


# Hypothesis: Invasiveness of placenta is correlated with speciation rate

## Placenta Types:

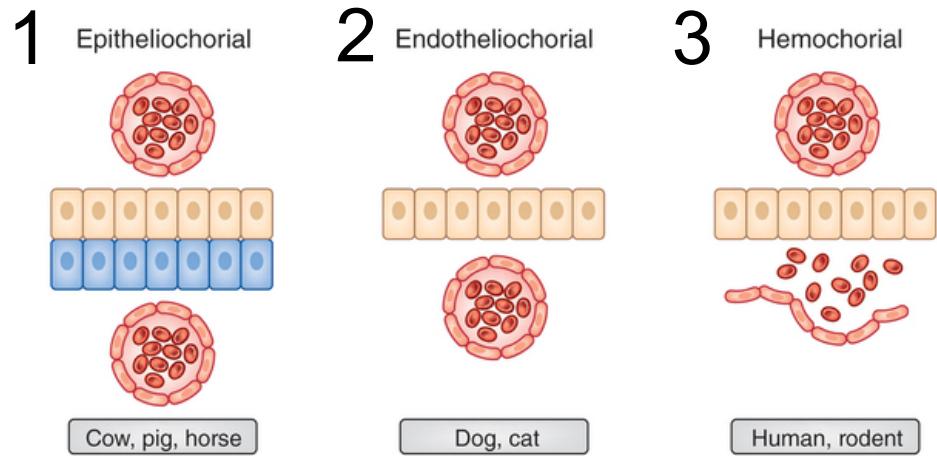


## Model:

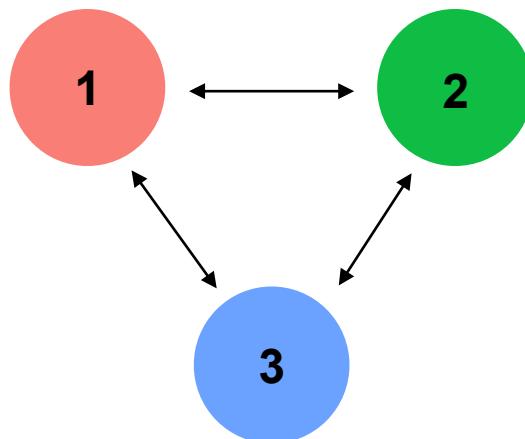


# Hypothesis: Invasiveness of placenta is correlated with speciation rate

## Placenta Types:



## Model:

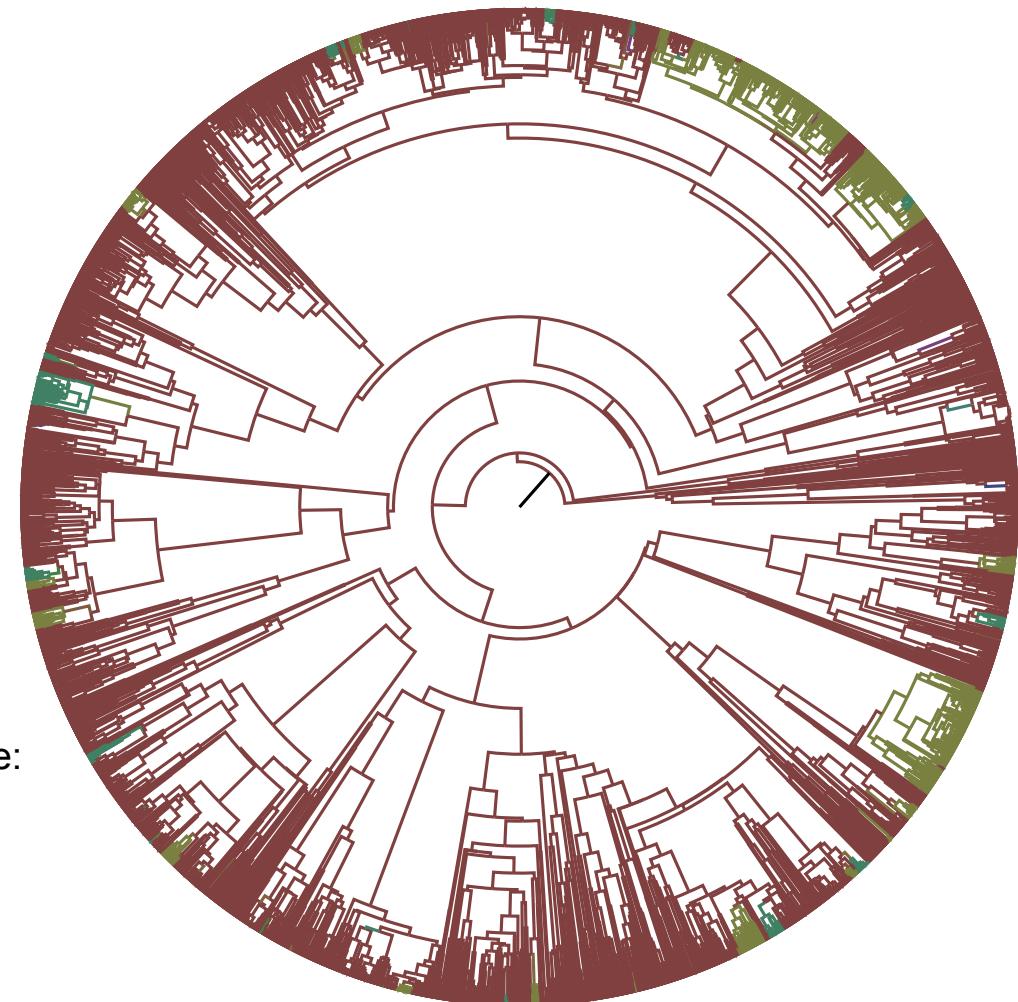


Diversification rates  
depend on placenta type:

$$\{\lambda_1, \mu_1\}$$

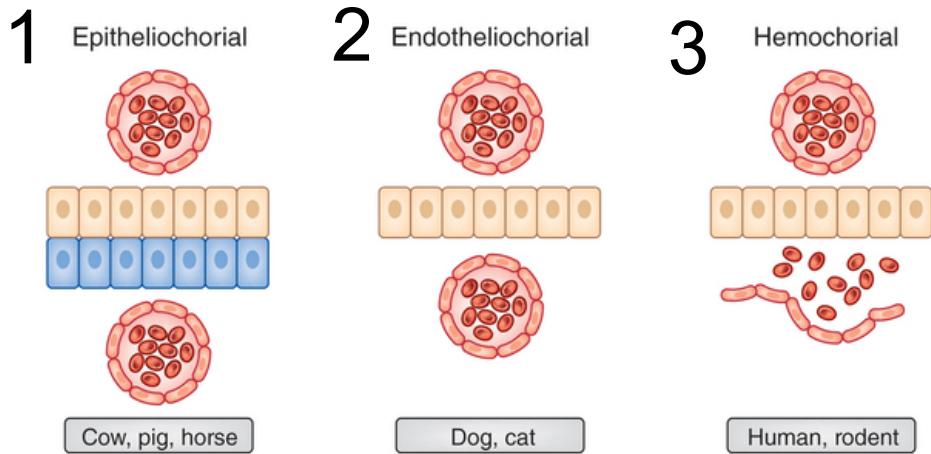
$$\{\lambda_2, \mu_2\}$$

$$\{\lambda_3, \mu_3\}$$

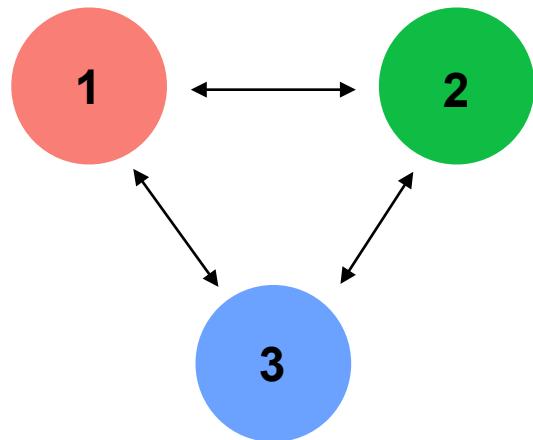


# Ancestral State Estimation

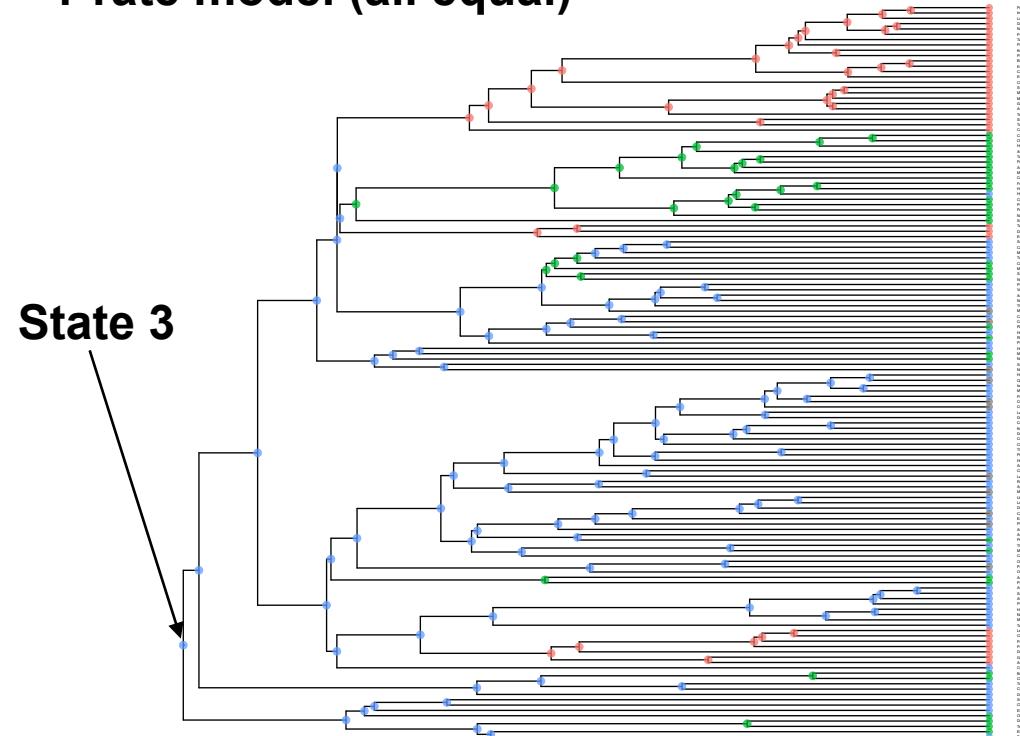
## Placenta Types:



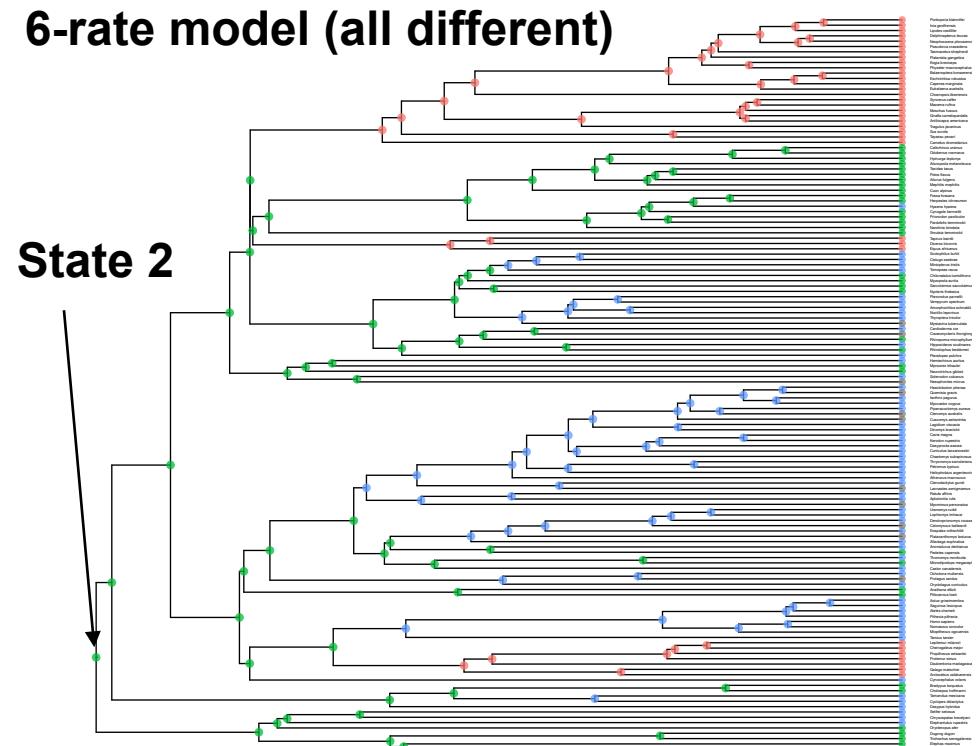
## Model:



## 1-rate model (all equal)



## 6-rate model (all different)

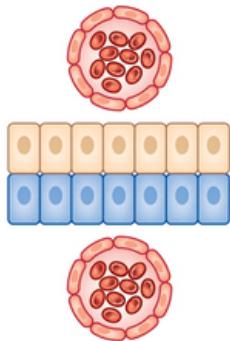


Furness, Freyman, Springer & Höhna (in prep)

# Diversification Rate Estimation

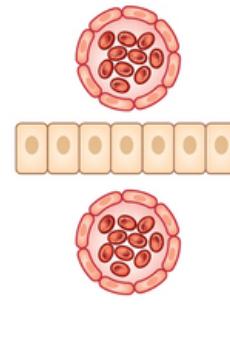
## Placenta Types:

1 Epitheliochorial



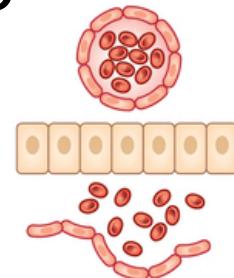
Cow, pig, horse

2 Endotheliochorial



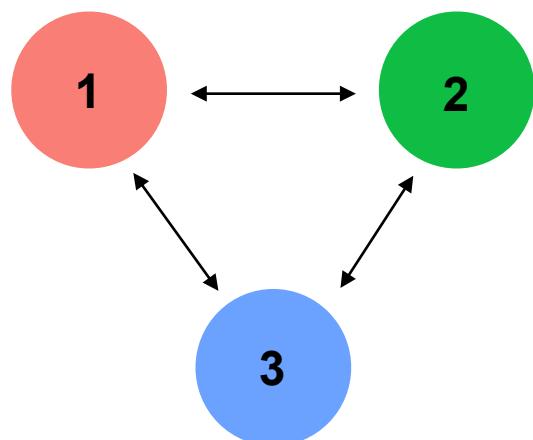
Dog, cat

3 Hemochorial

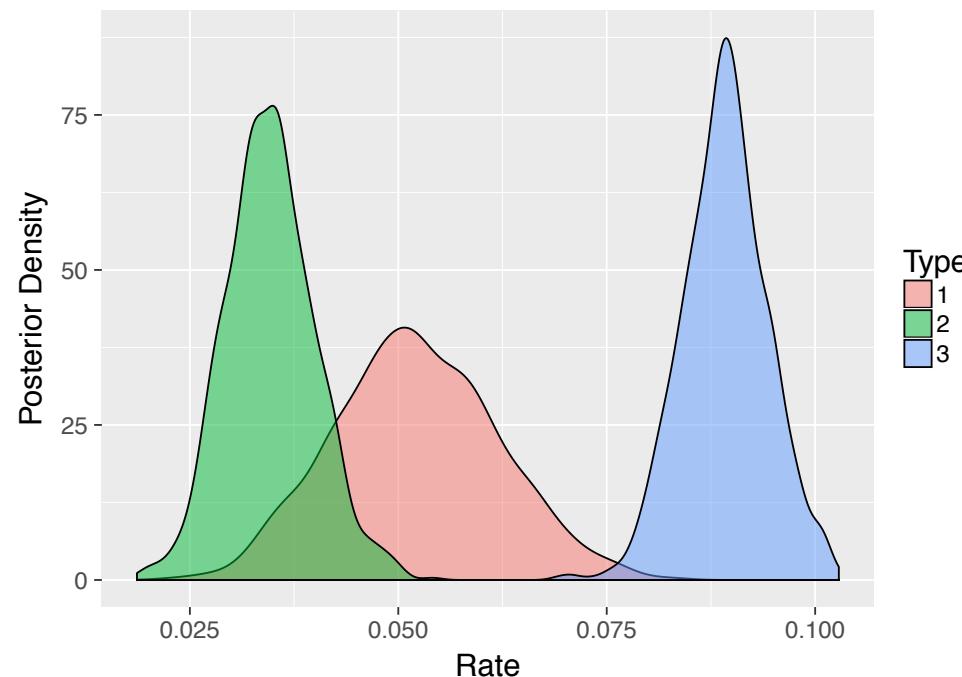


Human, rodent

## Model:



## Net-Diversification



# Problems with the BiSSE/MuSSE models

---

## **Problems:**

1. Speciation and extinction rate changes ALWAYS coincide with discrete character (morphology) changes.
2. What if there are rate changes that are uncorrelated with the observed character?
3. Leads to high false-positive rate (Rabosky & Goldberg, 2015).

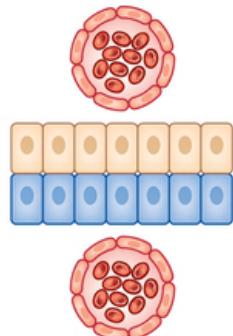
## **Solution:**

1. Add an unobserved (hidden) character (Beaulieu & O'Meara, 2016).
2. Rate shifts might be correlated with the observed character, the unobserved character, or both.

# Diversification Rate Estimation

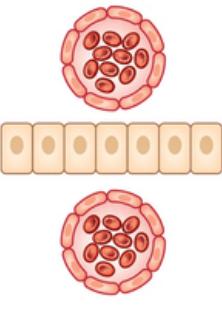
## Placenta Types:

1 Epitheliochorial



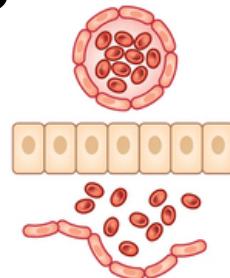
Cow, pig, horse

2 Endotheliochorial



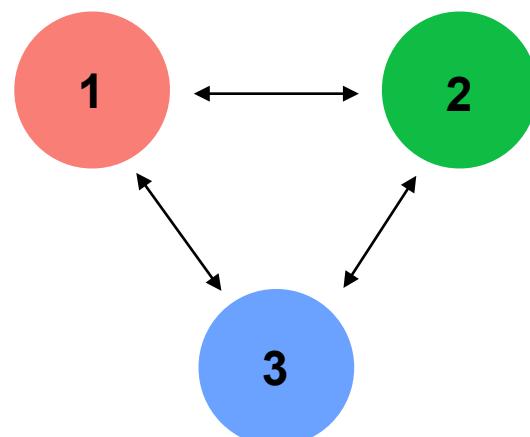
Dog, cat

3 Hemochorial

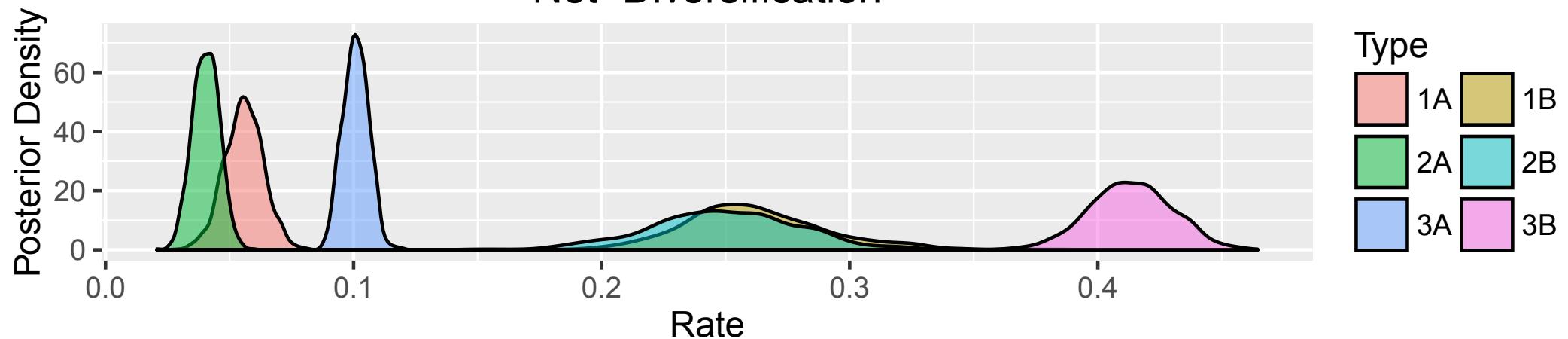


Human, rodent

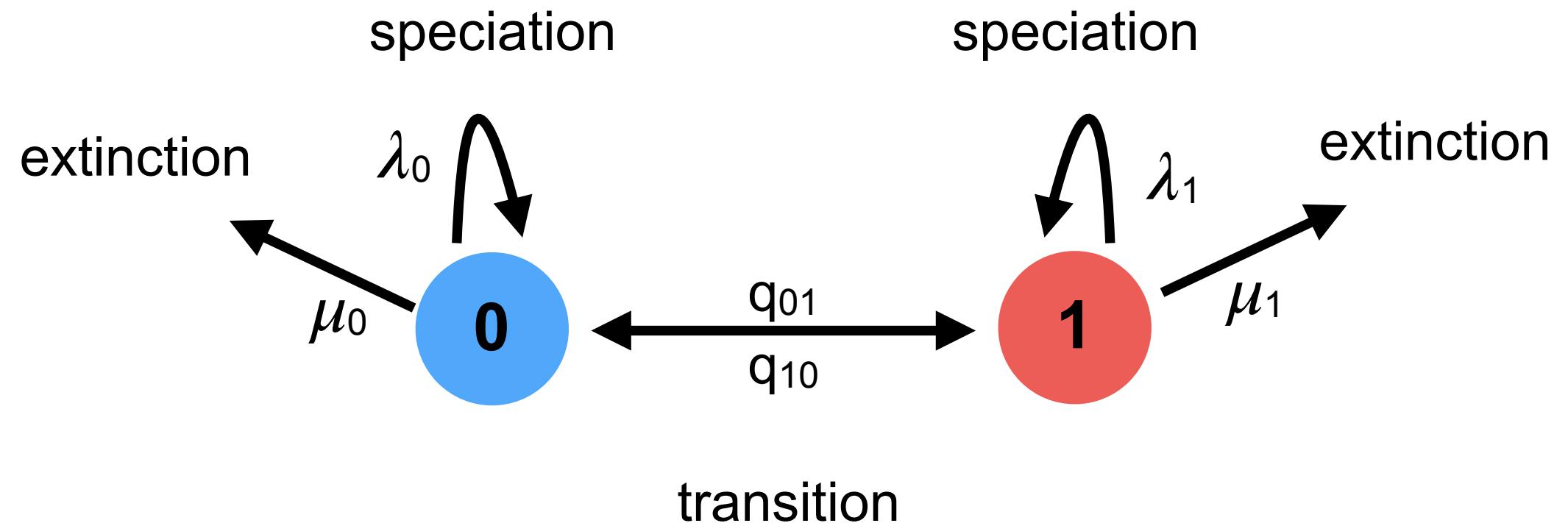
## Model:



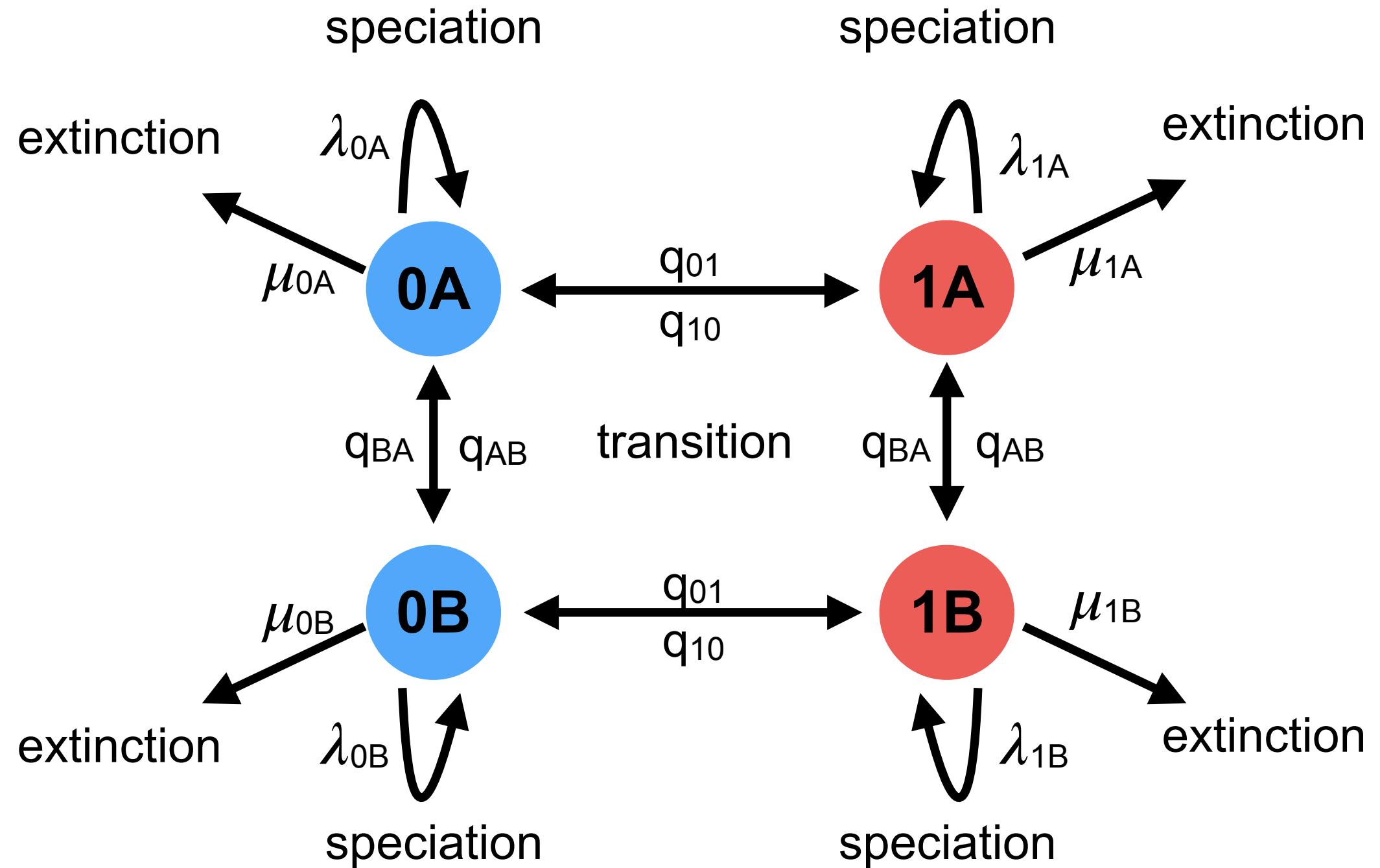
## Net-Diversification



# Binary State Speciation & Extinction (BiSSE)



# Hidden State Speciation & Extinction (HiSSE)



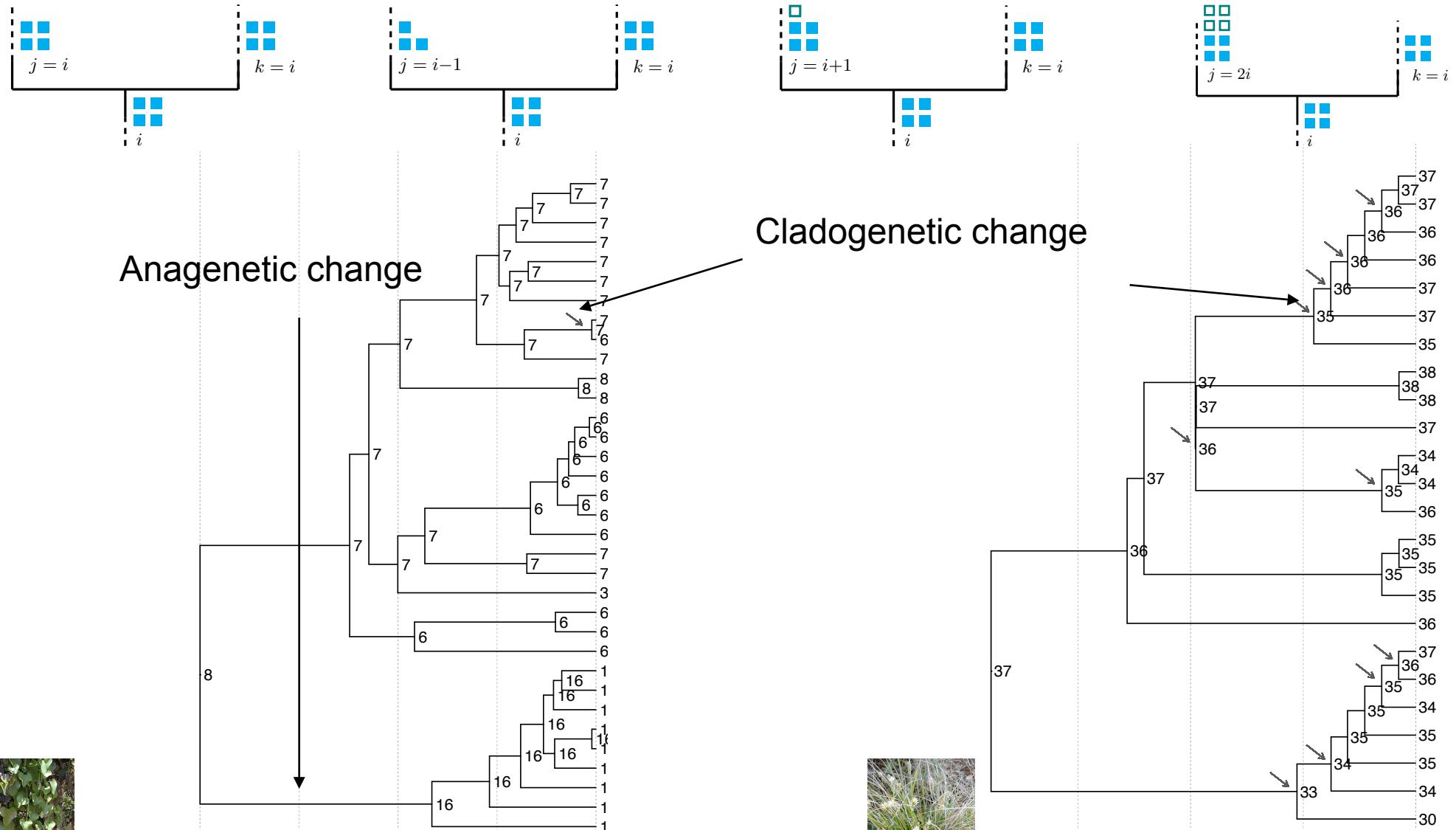
# Extensions: ChromoSSE

No Change

Loss

Gain

Duplication



Aristolochia (birthwort & pipevine)

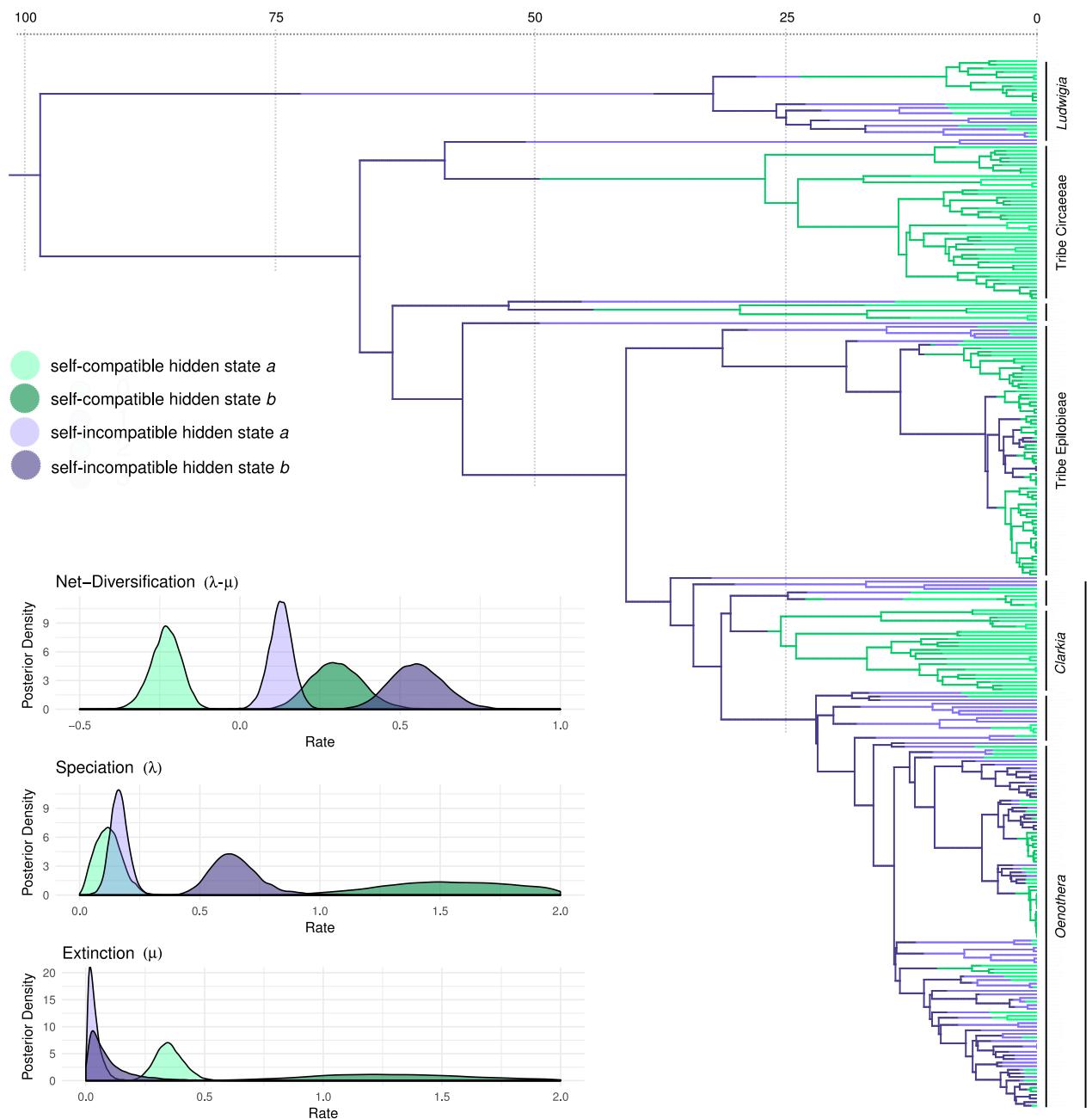
Carex (Sedges)

Freyman & Höhna (Systematic Biology, 2018)

# Extensions: Stochastic Character Mapping

Is self-compatibility an evolutionary dead-end?

Stochastic Character Mapping reveals a short-term evolutionary advantage followed by a decline (negative net-diversification).



# **Exercise 4: State-dependent Diversification Rate Estimation**