

Phylogenetics

Ancestral state
reconstruction

RL-V3 MPP

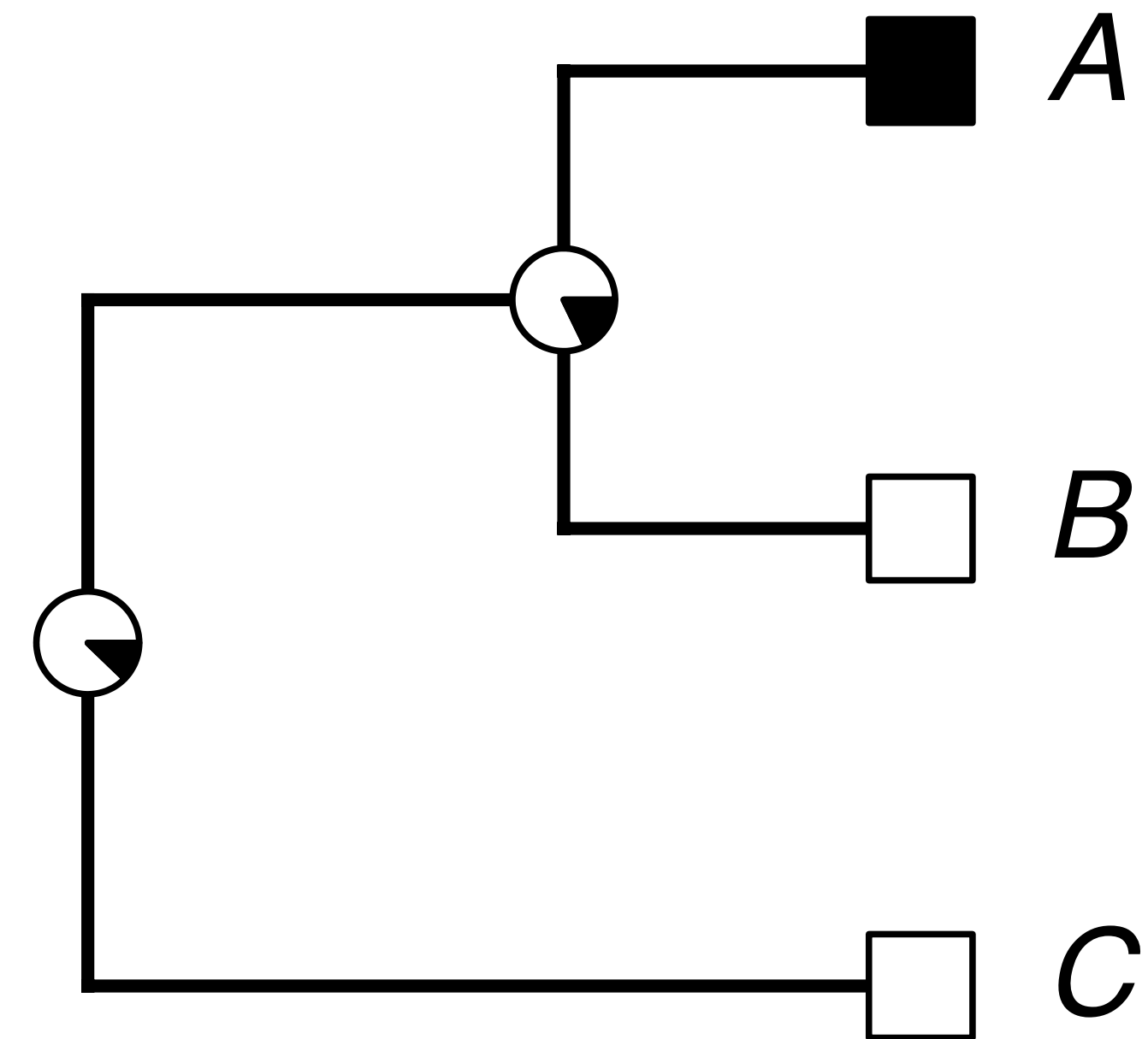
Rachel Warnock

30.06.26



Objectives

- Recap: models of morphological evolution
- Ancestral state reconstruction
- A very brief intro to biogeography



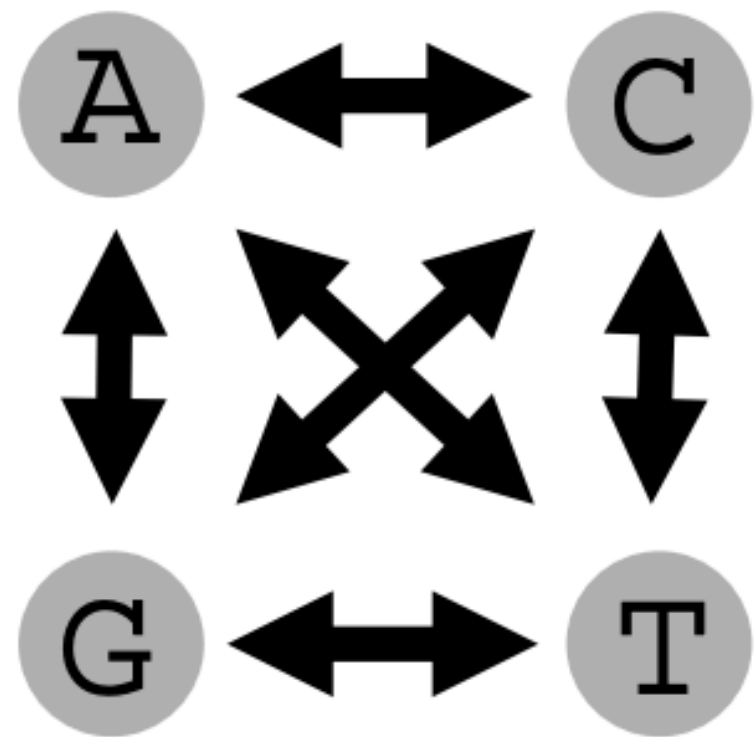
Recap

Models of morphological evolution

Slides adapted from Laura Mulvey (see week 11 materials)

Substitution models in molecular data

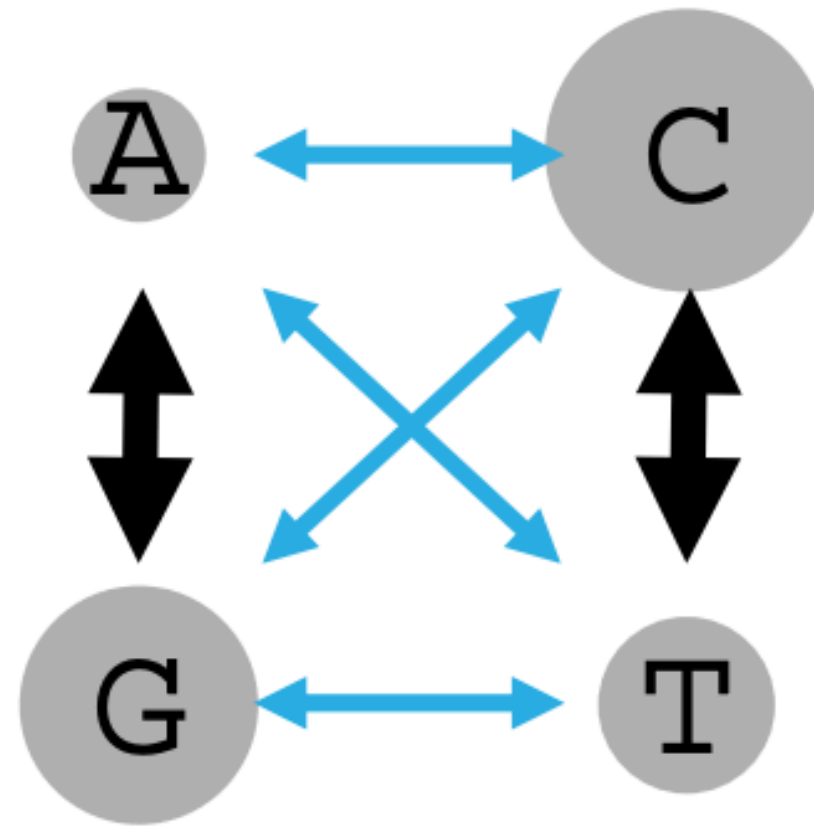
JC



All changes equal
Equal base frequencies

$$Q = \begin{pmatrix} * & \frac{\mu}{4} & \frac{\mu}{4} & \frac{\mu}{4} \\ \frac{\mu}{4} & * & \frac{\mu}{4} & \frac{\mu}{4} \\ \frac{\mu}{4} & \frac{\mu}{4} & * & \frac{\mu}{4} \\ \frac{\mu}{4} & \frac{\mu}{4} & \frac{\mu}{4} & * \end{pmatrix}$$

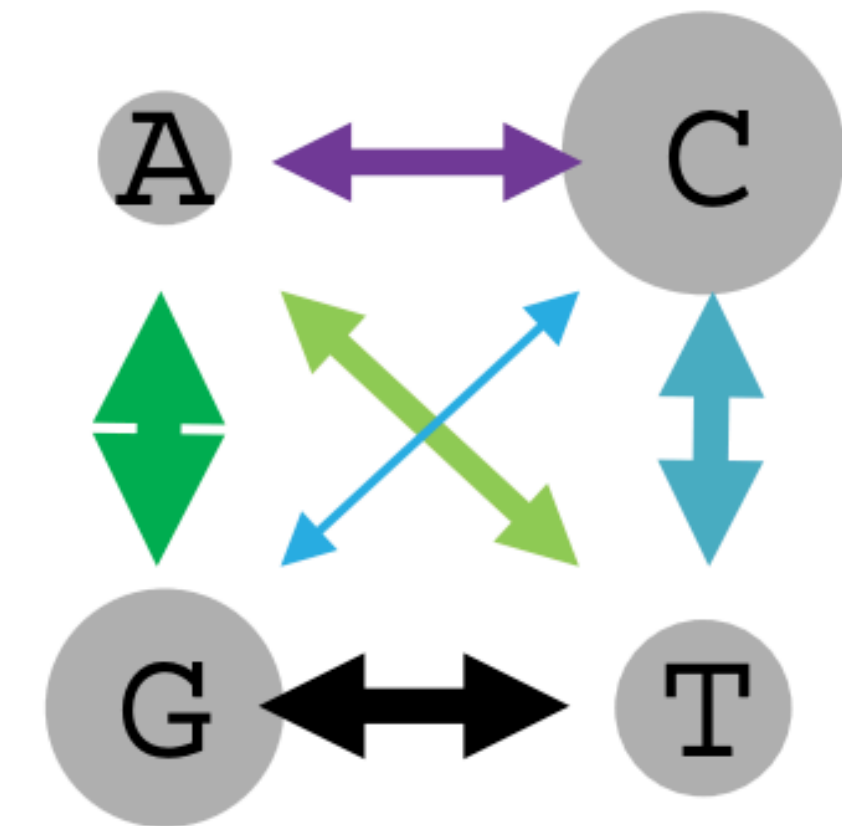
HKY



Transitions \neq Transversions
Unequal base frequencies

$$Q = \begin{pmatrix} * & \kappa\pi_G & \pi_C & \pi_T \\ \kappa\pi_A & * & \pi_C & \pi_T \\ \pi_A & \pi_G & * & \kappa\pi_T \\ \pi_A & \pi_G & \kappa\pi_C & * \end{pmatrix}$$

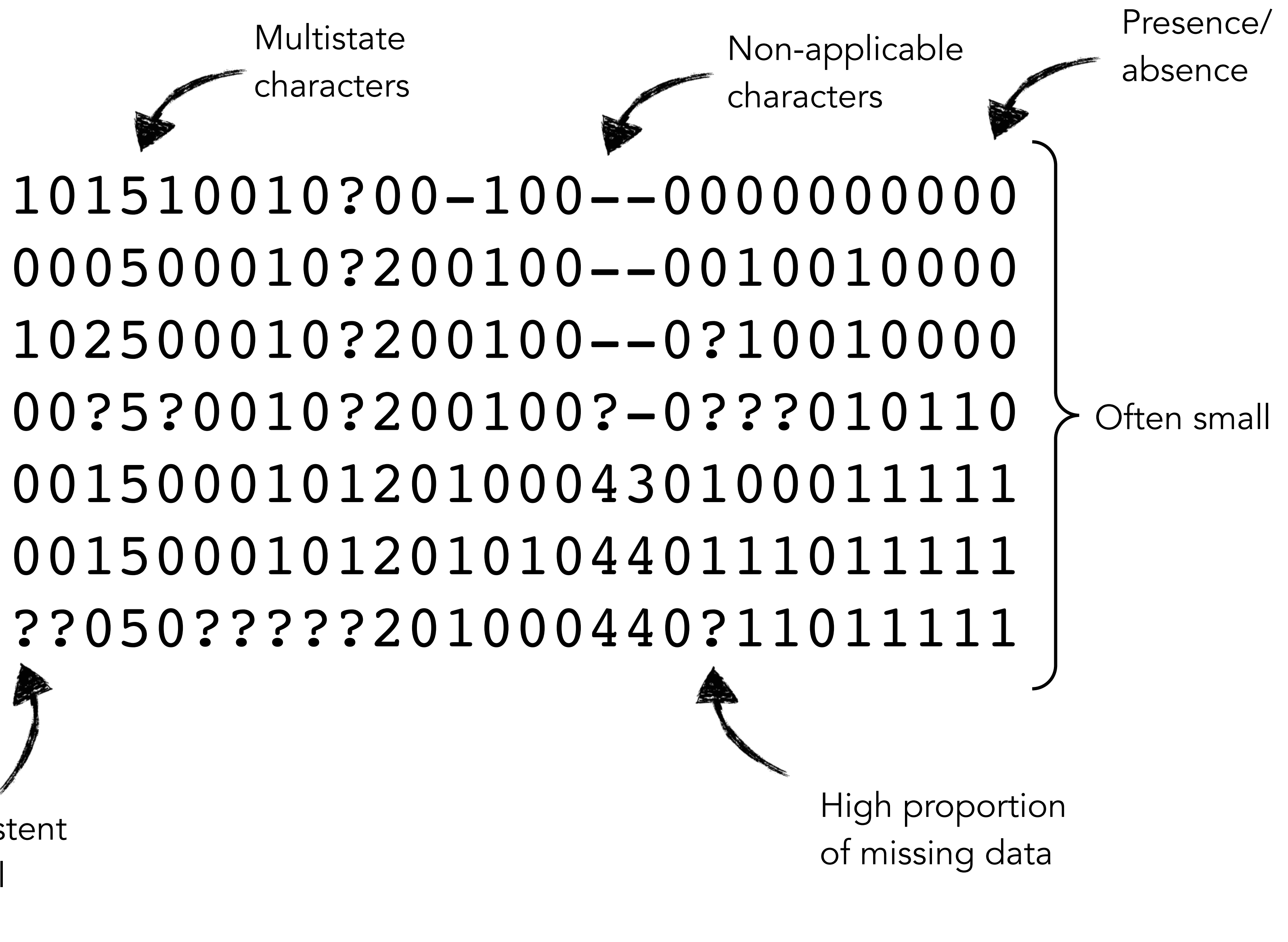
GTR



Separate rates for all sub. types
Unequal base frequencies

$$Q = \begin{pmatrix} * & \kappa_1\pi_G & \pi_C & \pi_T \\ \kappa_1\pi_A & * & \pi_C & \pi_T \\ \pi_A & \pi_G & * & \kappa_2\pi_T \\ \pi_A & \pi_G & \kappa_2\pi_C & * \end{pmatrix}$$

Making general assumptions about evolution from this data is extremely difficult



Mk model

Assumes **equal transition probabilities** between states

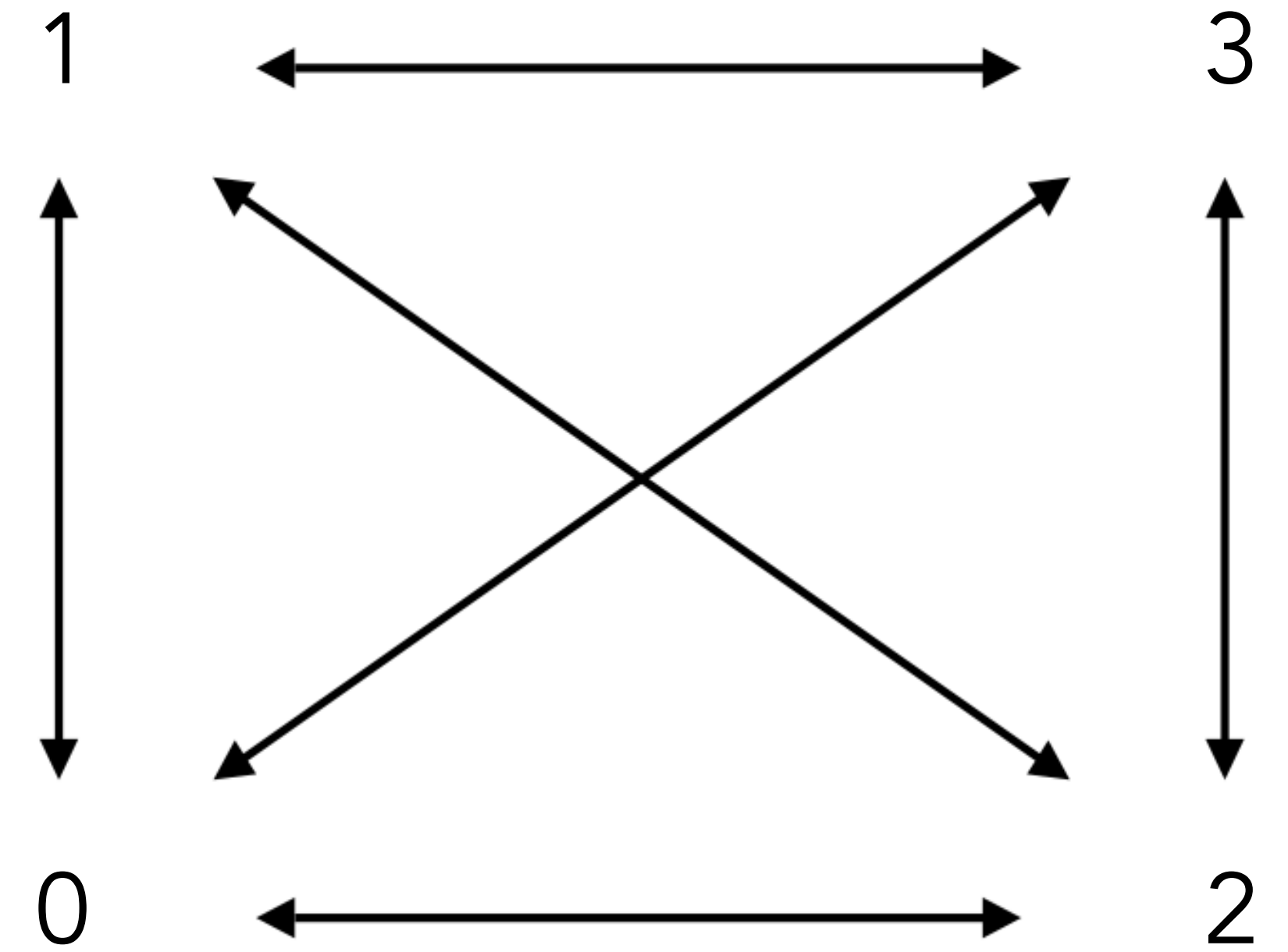


$$Q = \begin{pmatrix} -\mu_0 & \mu_{01} \\ \mu_{10} & -\mu_1 \end{pmatrix}$$

Mk model

K can be any number of states

$$Q = \begin{pmatrix} -\mu_0 & \mu_{01} & \mu_{02} & \mu_{03} \\ \mu_{10} & -\mu_1 & \mu_{12} & \mu_{13} \\ \mu_{20} & \mu_{21} & -\mu_2 & \mu_{23} \\ \mu_{30} & \mu_{31} & \mu_{32} & -\mu_3 \end{pmatrix},$$



*4 state here as an example, can be any number from 2!

Mk extensions

+V accounts for the fact that (usually) all characters vary

+G gamma model to account for rate variation

+P partitions the data into different rate matrices

You can use any combination of the above!

Tree inference using discrete morphology

Tutorials

RevBayes - [Standard tree inference](#) (more flexible than BEAST)

- [Total-evidence tutorial](#)

BEAST2 - [Total-evidence tutorial](#)

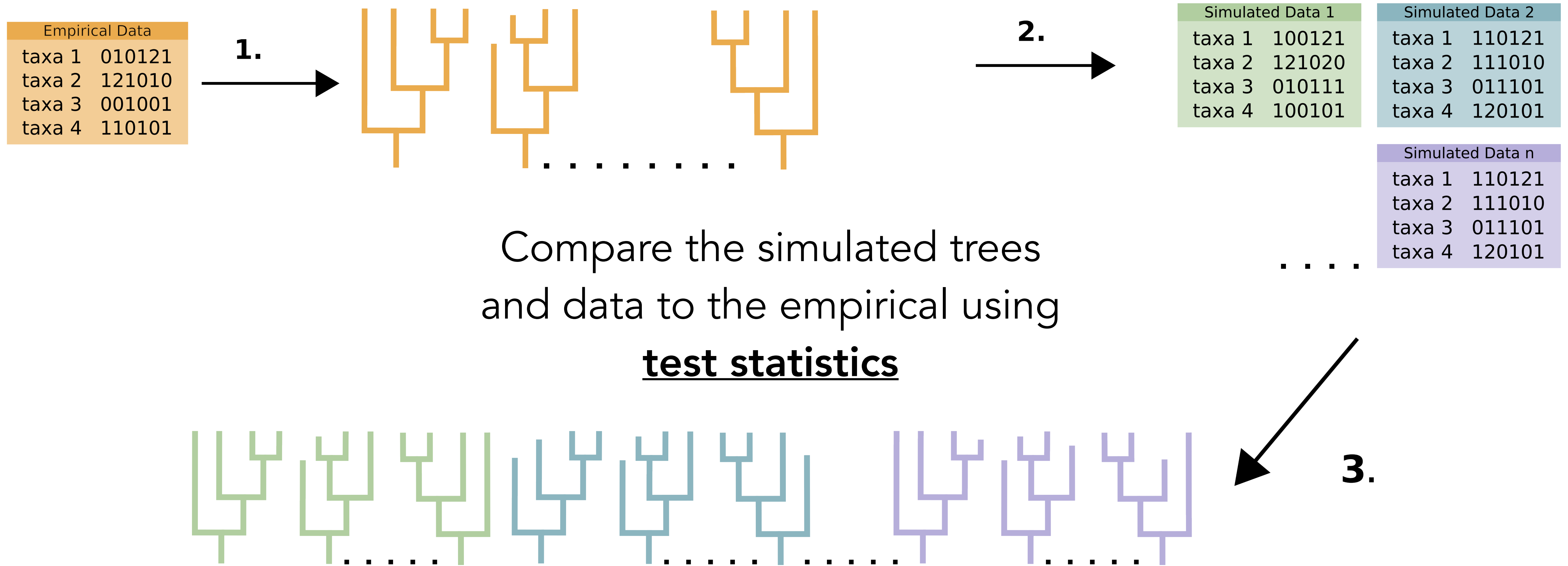
MrBayes - see software [manual](#)

Recap

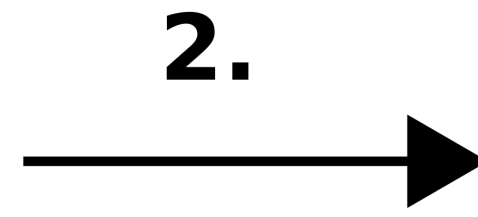
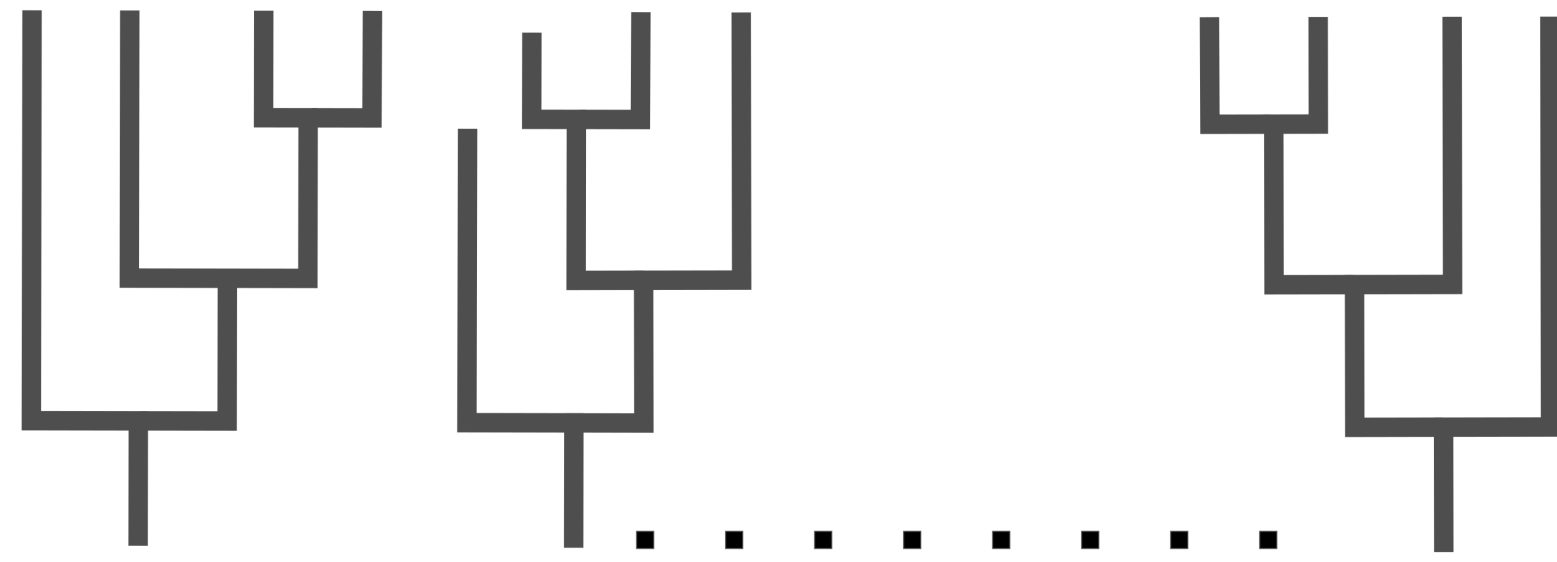
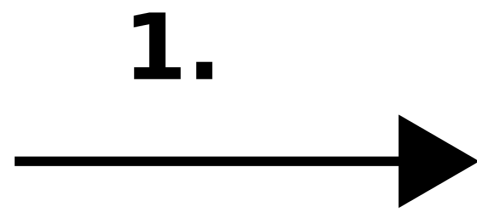
Model adequacy

Slides adapted from Laura Mulvey (see week 11 materials)

Posterior Predictive Simulations



Empirical Data 1					
taxa 1	0	0	1	2	1
taxa 2	1	2	2	0	2
taxa 3	0	1	0	1	0
taxa 4	0	0	0	1	1



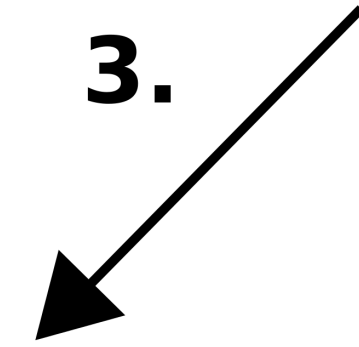
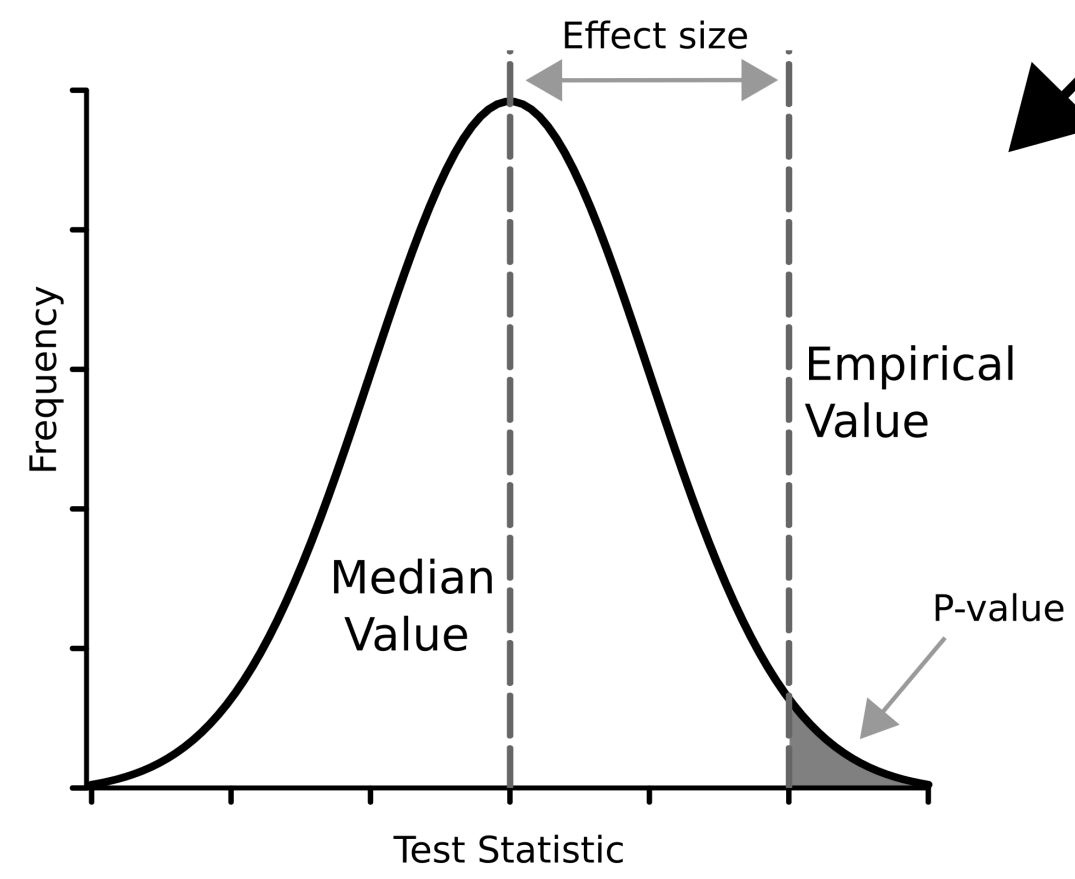
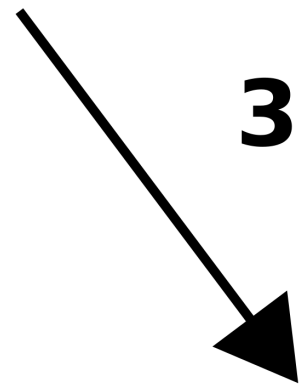
Simulated Data 1					
taxa 1	1	0	0	1	2
taxa 2	1	2	1	0	2
taxa 3	0	1	0	1	1
taxa 4	1	0	0	1	0

Simulated Data 2					
taxa 1	1	0	0	1	2
taxa 2	1	2	1	0	2
taxa 3	0	1	0	1	1
taxa 4	1	0	0	1	0

Simulated Data n					
taxa 1	1	0	0	1	2
taxa 2	1	2	1	0	2
taxa 3	0	1	0	1	1
taxa 4	1	0	0	1	0

Consistency Index
Retention Index

Consistency Index
Retention Index



Model adequacy exercise

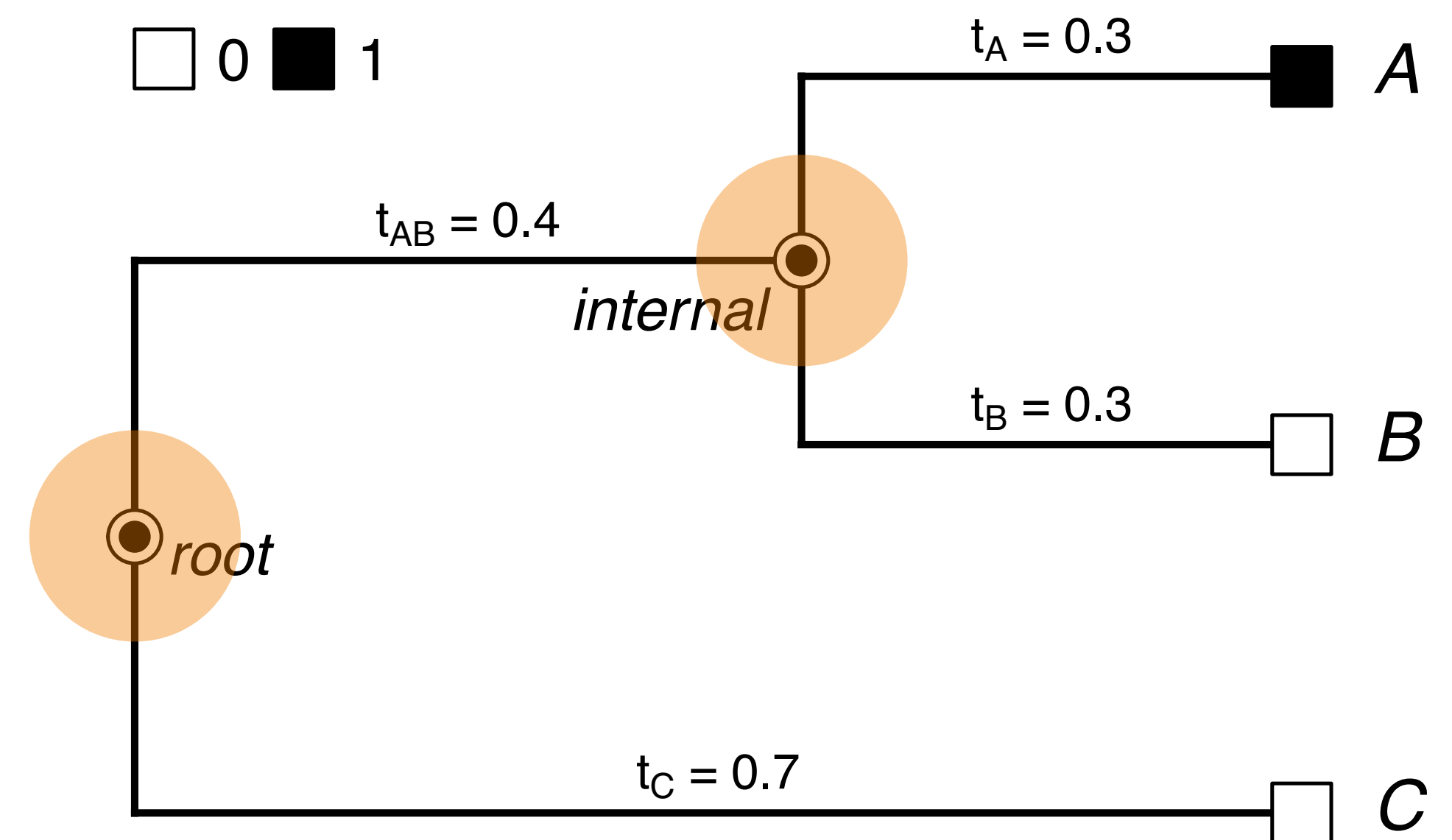
Ancestral state reconstruction

See also week 4 materials, intro to PCMs, ancestral state reconstruction using continuous trait models

Ancestral state reconstruction

The objective is to reconstruct the ancestral states for parts of the tree that are **not directly observable** using data observed at the tips of the tree

The input is usually a **dated tree**, with branch lengths in units of time - we usually do *ASR after* we have inferred the tree

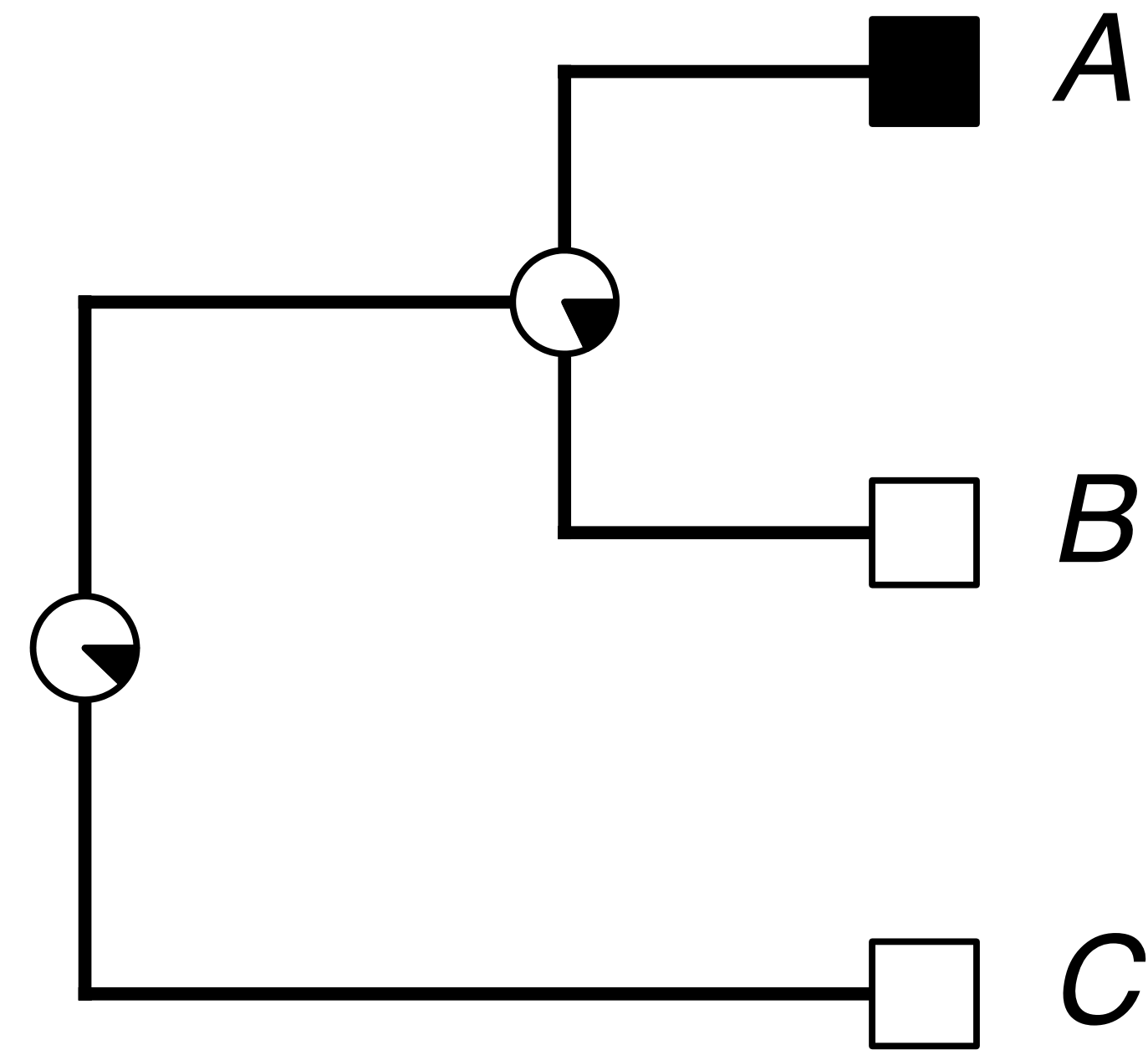


Ancestral state reconstruction

Discrete traits (see week 4 for models of continuous trait evolution)

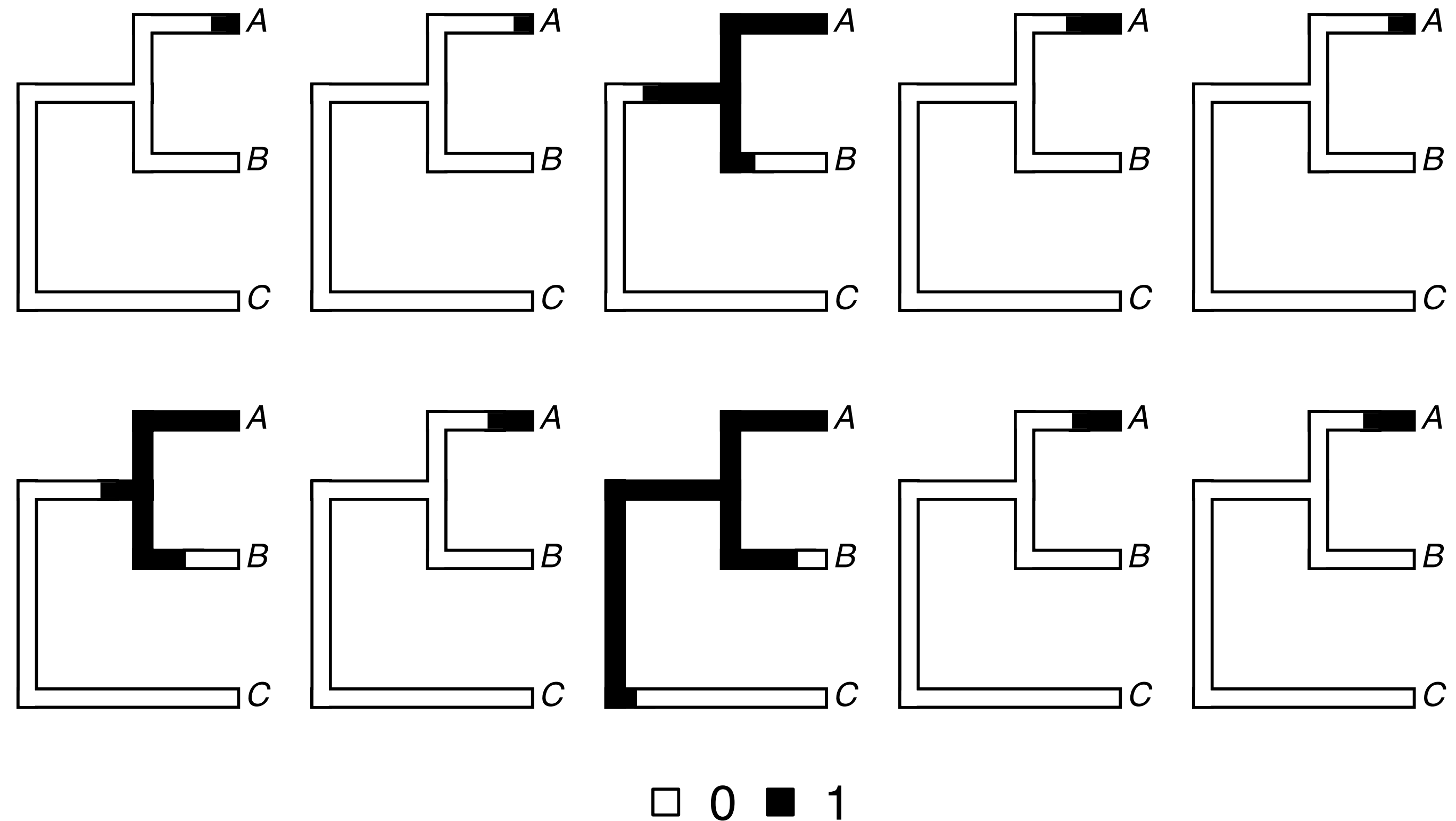
To estimate the states at internal nodes we can use standard models of character evolution, e.g., the Mk model (ML or Bayesian)

Most often we are interested in the most likely state at each individual node (the marginal reconstruction, shown right)

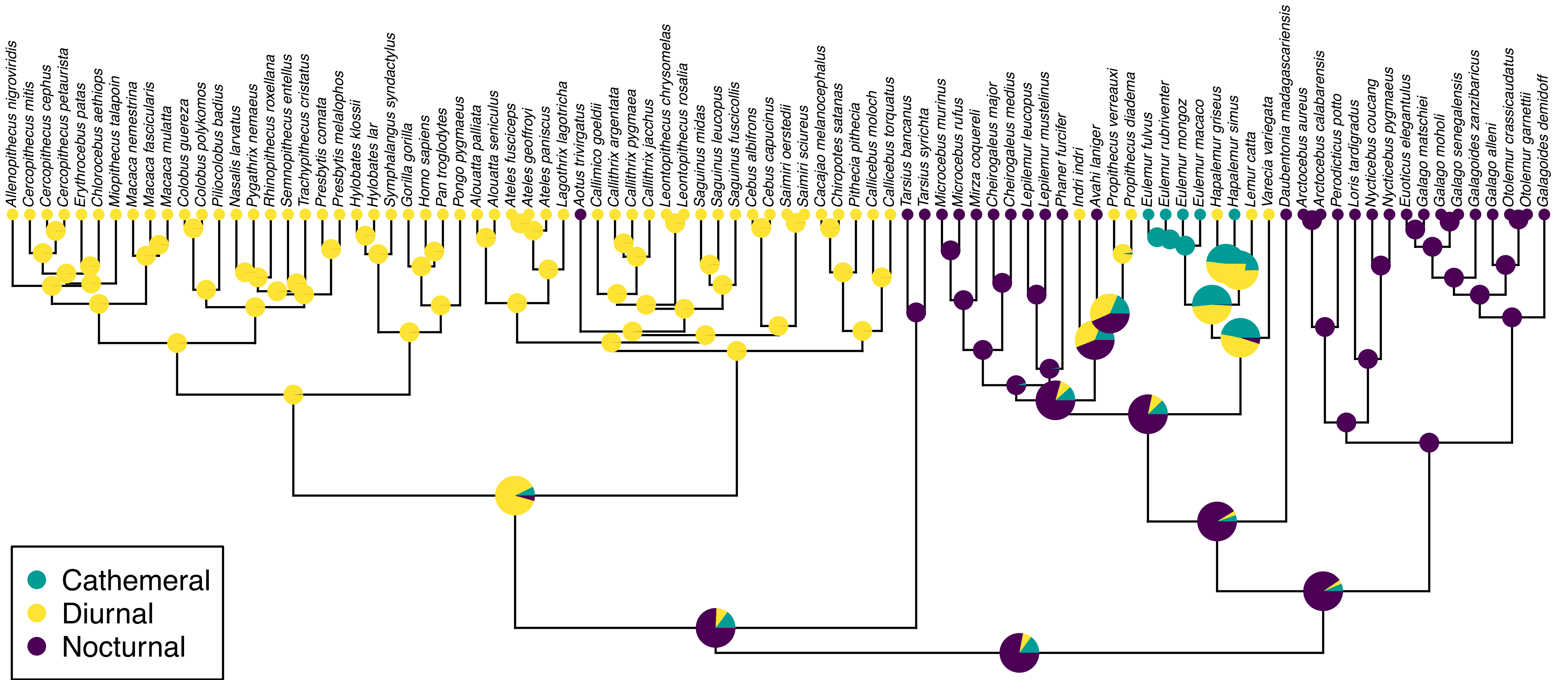


Ancestral state reconstruction

We need to account for all possible histories - we can do this using a variant of Felsenstein's pruning algorithm that incorporates information across the tree



See Part 1, Day 1 materials

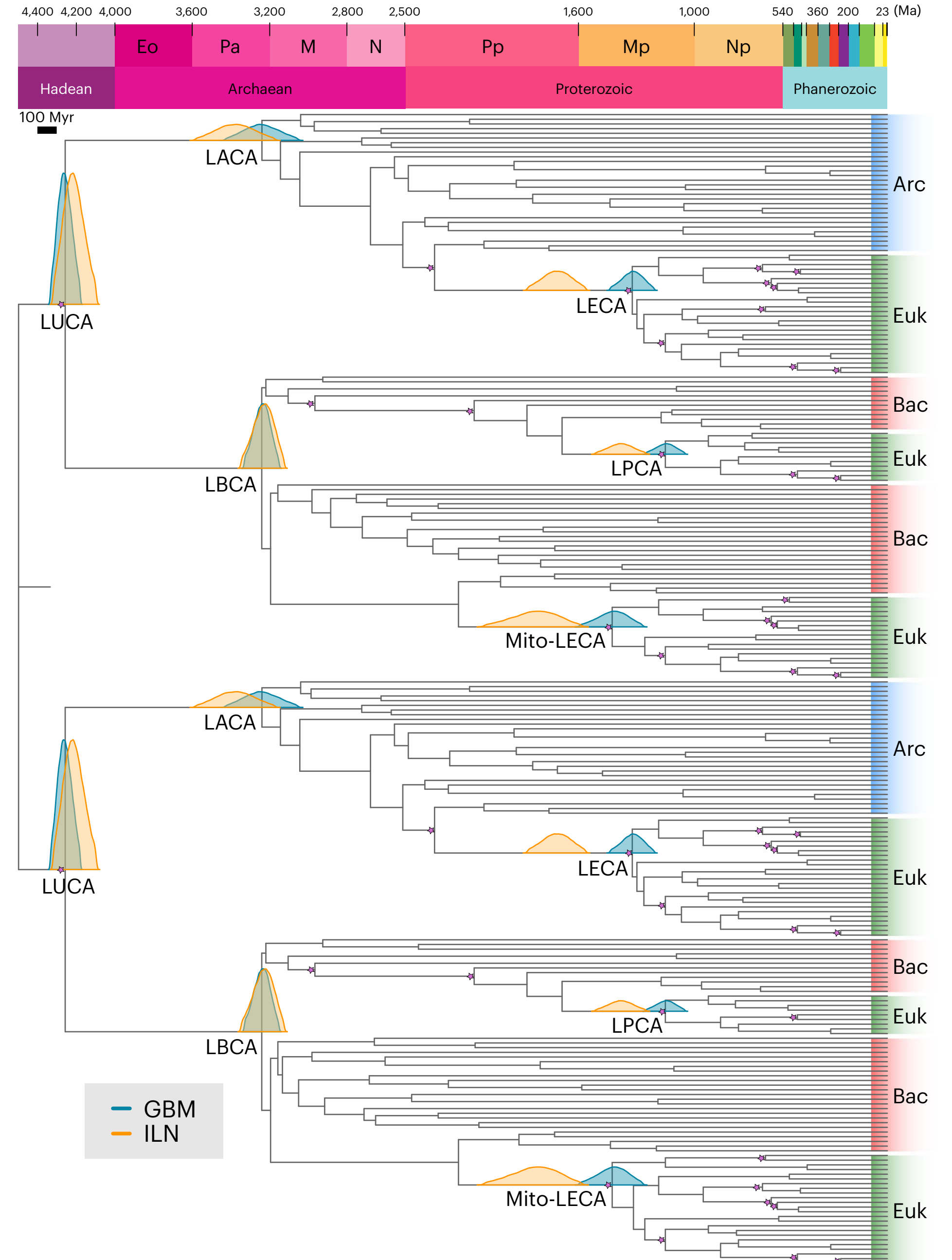
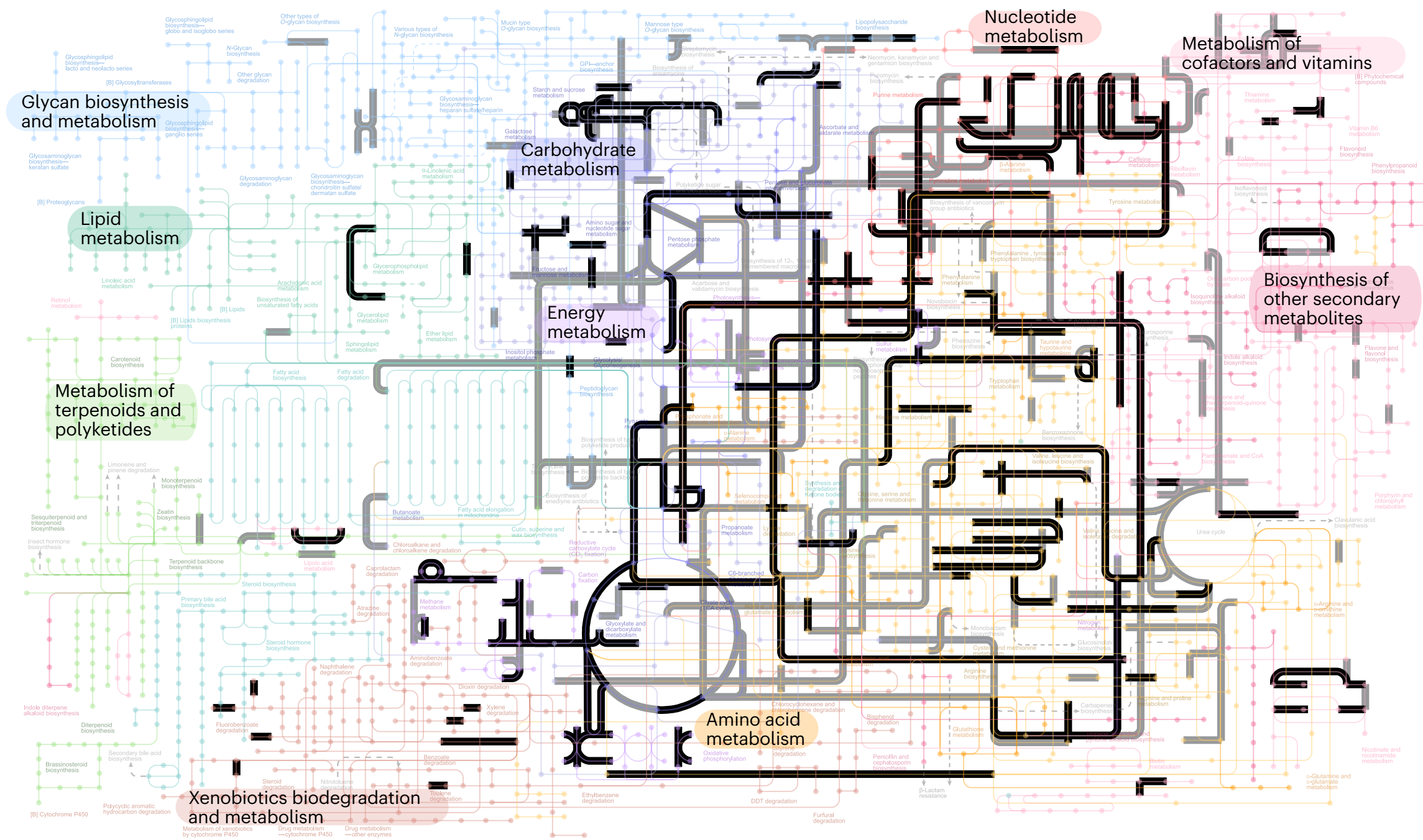




Article

<https://doi.org/10.1038/s41559-024-02461-1>

The nature of the last universal common ancestor and its impact on the early Earth system



Resources for ancestral state reconstruction



[Ancestral state reconstruction for discrete & continuous traits](#) *Liam Revell*



[Phylogenetic Comparative Methods: A Practical Introduction](#) *Natalie Cooper* ([Section 3.2](#))



[Ancestral State Reconstruction of Phenotypic Characters](#) *Liam Revell*

Tree inference using discrete morphology

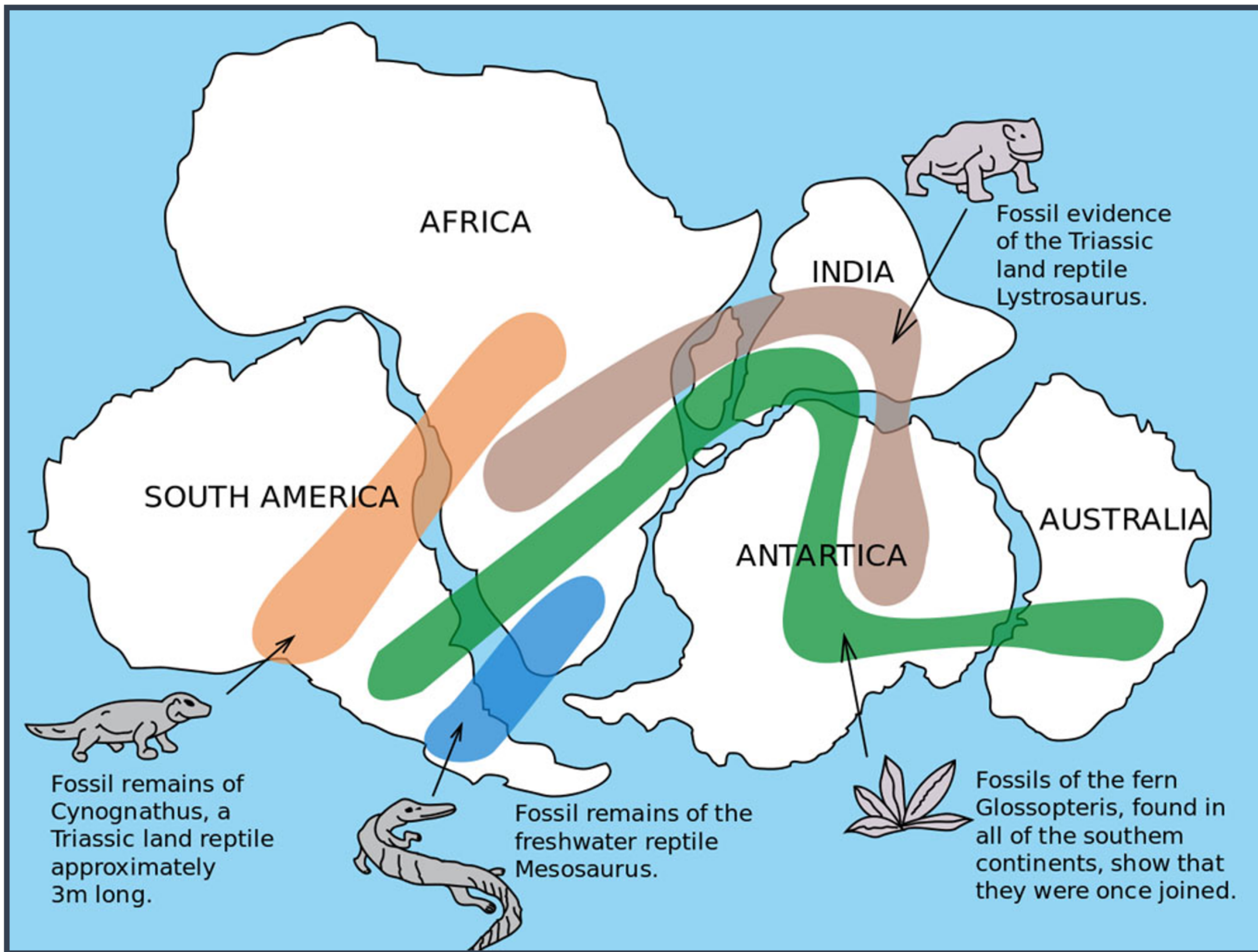
Tutorials

RevBayes - [Equal rates model](#) + many other options (see [main tutorial page](#))

BEAST2 - stochastic character mapping via [BDMMPrime](#) (more advanced)

R - geiger (see [exercise 11](#), week 7), phytools (see associated [blog](#))

A very brief intro to biogeography



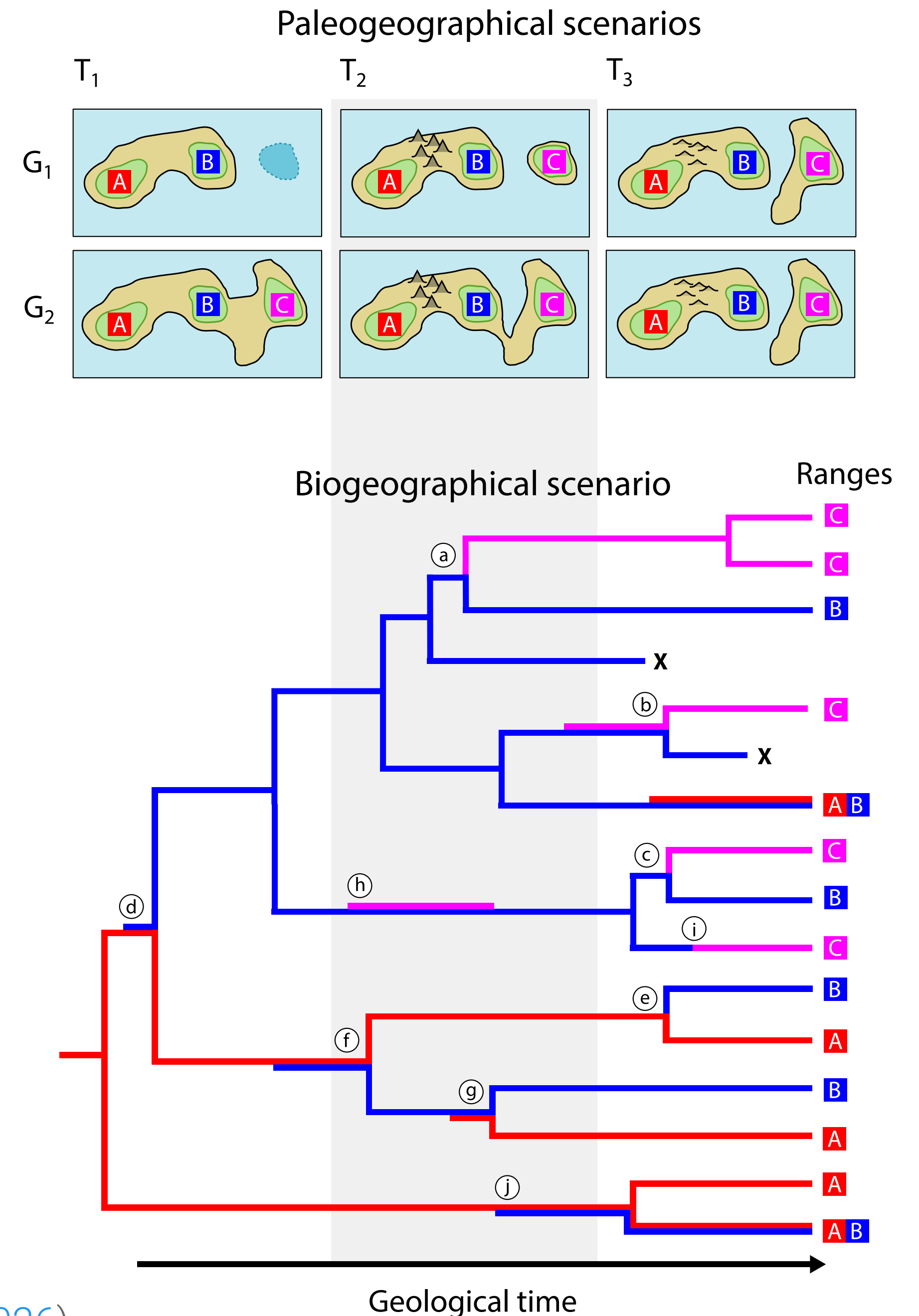
- How has biogeography shaped evolution?
- How has Earth history shaped biogeography?
- When and **where** do major lineages originate?
- How do biogeographic processes vary across lineages?
- How do dispersal (migration), speciation, and extinction shape biogeographic patterns

Phylogenetic biogeography

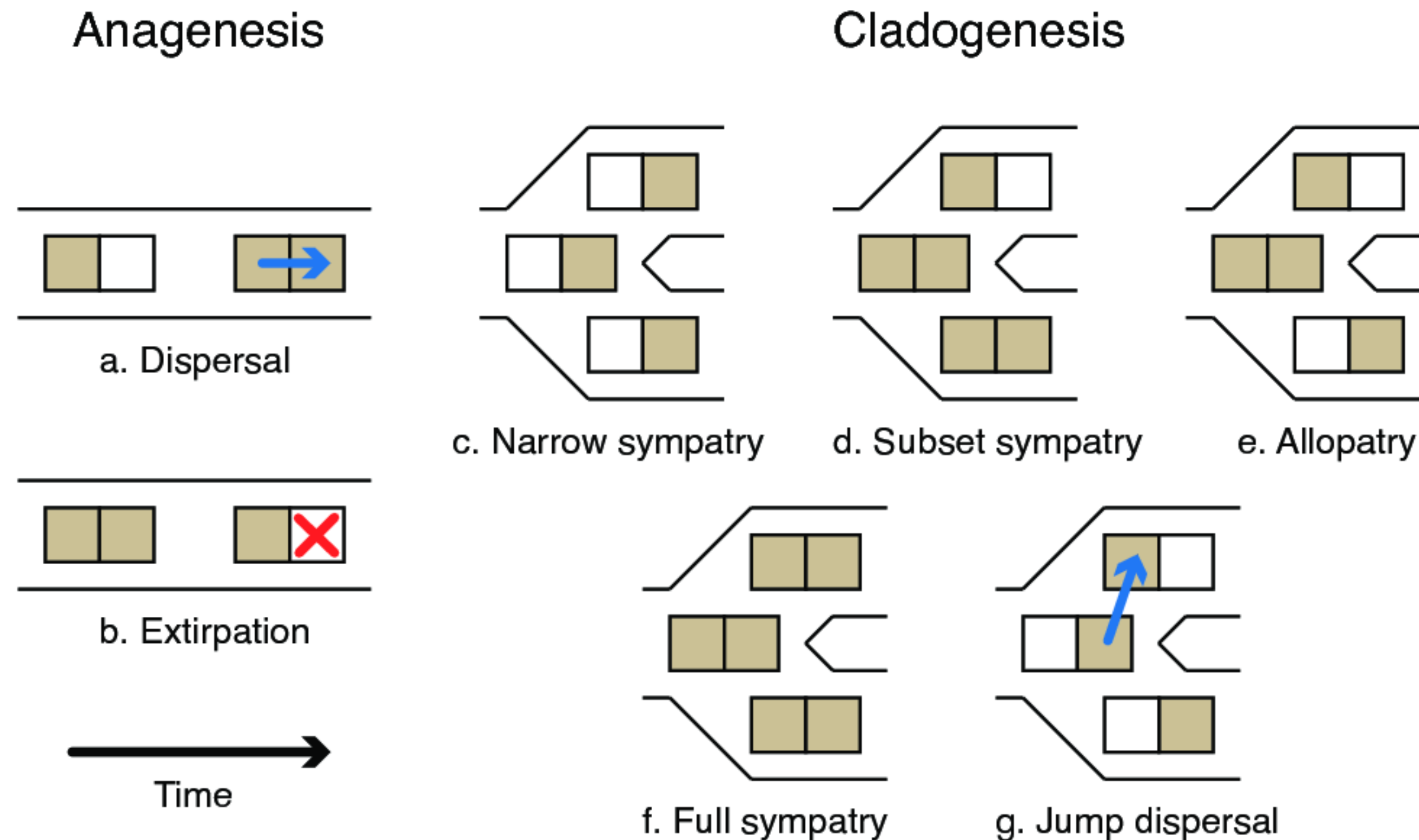
The aim is to reconstruct biogeographic history using data observed at the tips of the tree

Related objectives:

- Reconstructing biogeography and ancestral ranges
- Exploring the link between biogeography and diversification (λ , μ)
- Biogeographic dating



The Dispersal-Extinction-Cladogenesis (DEC) model



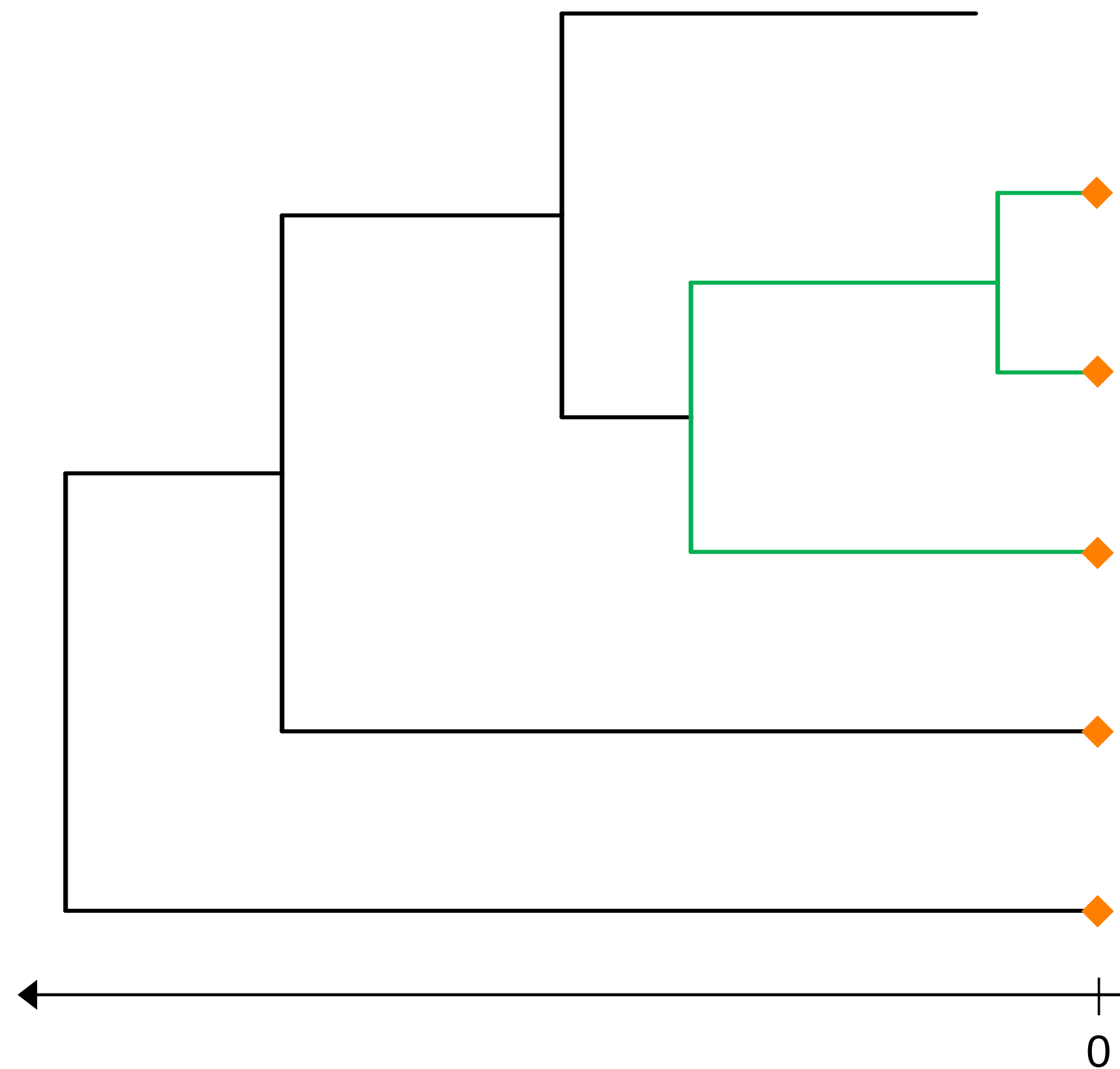
Each set of occupied areas represents a character state, e.g., 00, 01, 11, 10

Anagenetic events are modelled using a standard rate (Q) matrix

Cladogenetic events modelled using a probability (P) matrix

State-dependent speciation-extinction (SSE) models

Also known as multi-type birth death process models



2 types, type 1 & **type 2**

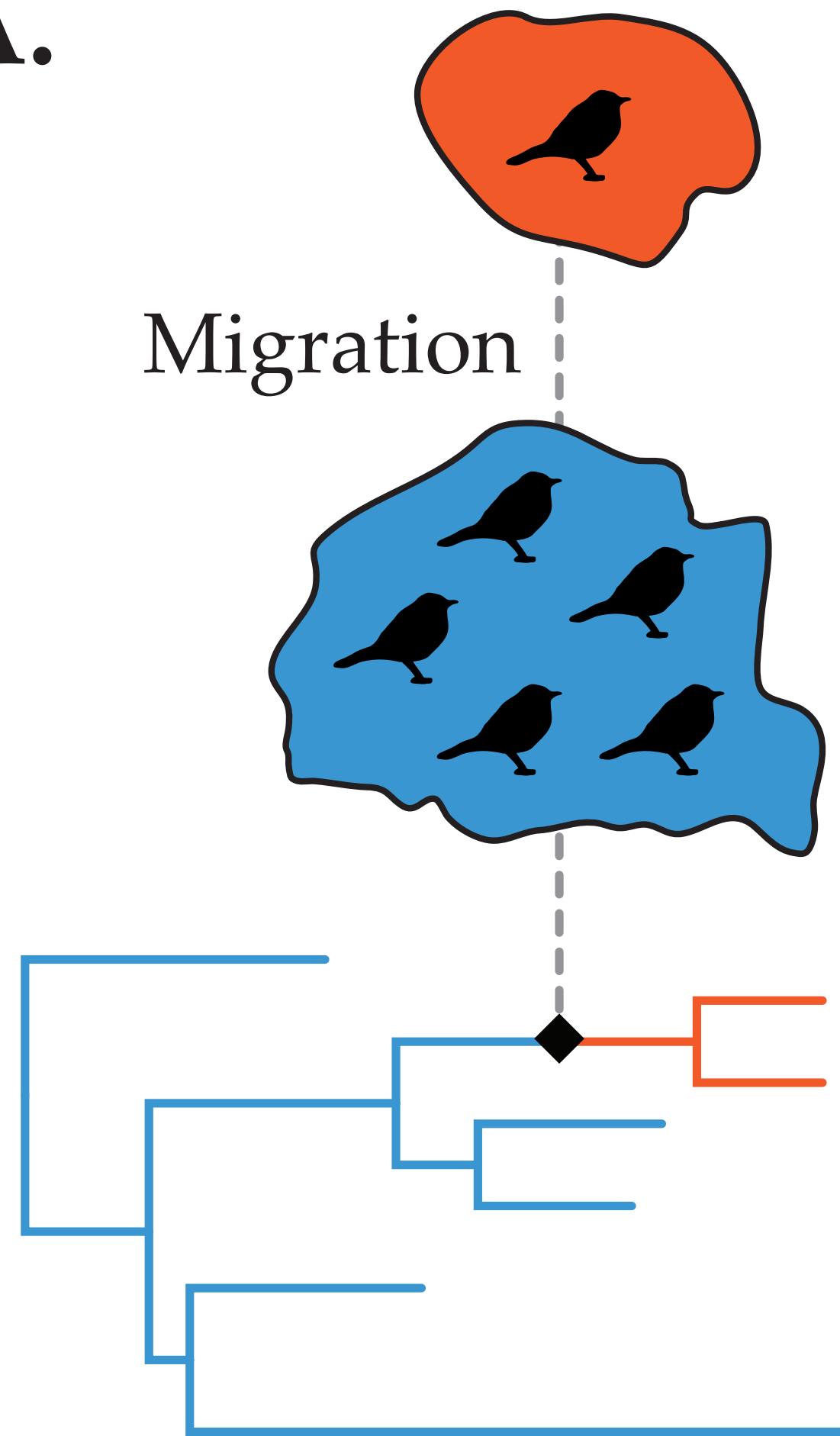
λ_1 & λ_2 – birth rates

μ_1 & μ_2 – death rates

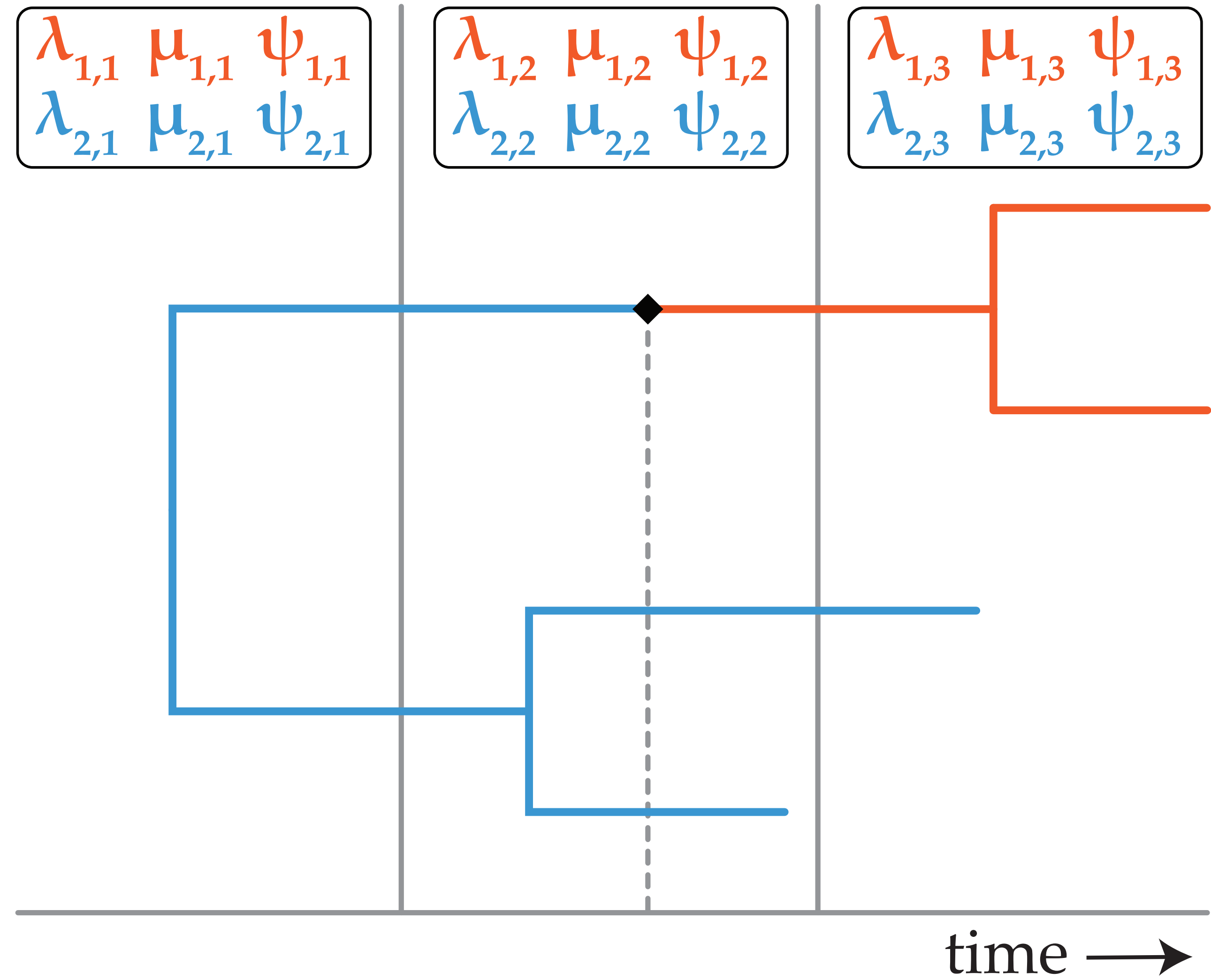
ρ – extant species sampling probability

See also week 4 materials, intro to PCMs, ancestral state reconstruction using continuous trait models

A.



B.



Joint phylogenetic estimation of geographic movements and biome shifts

Example



Global diversity of *Viburnum*

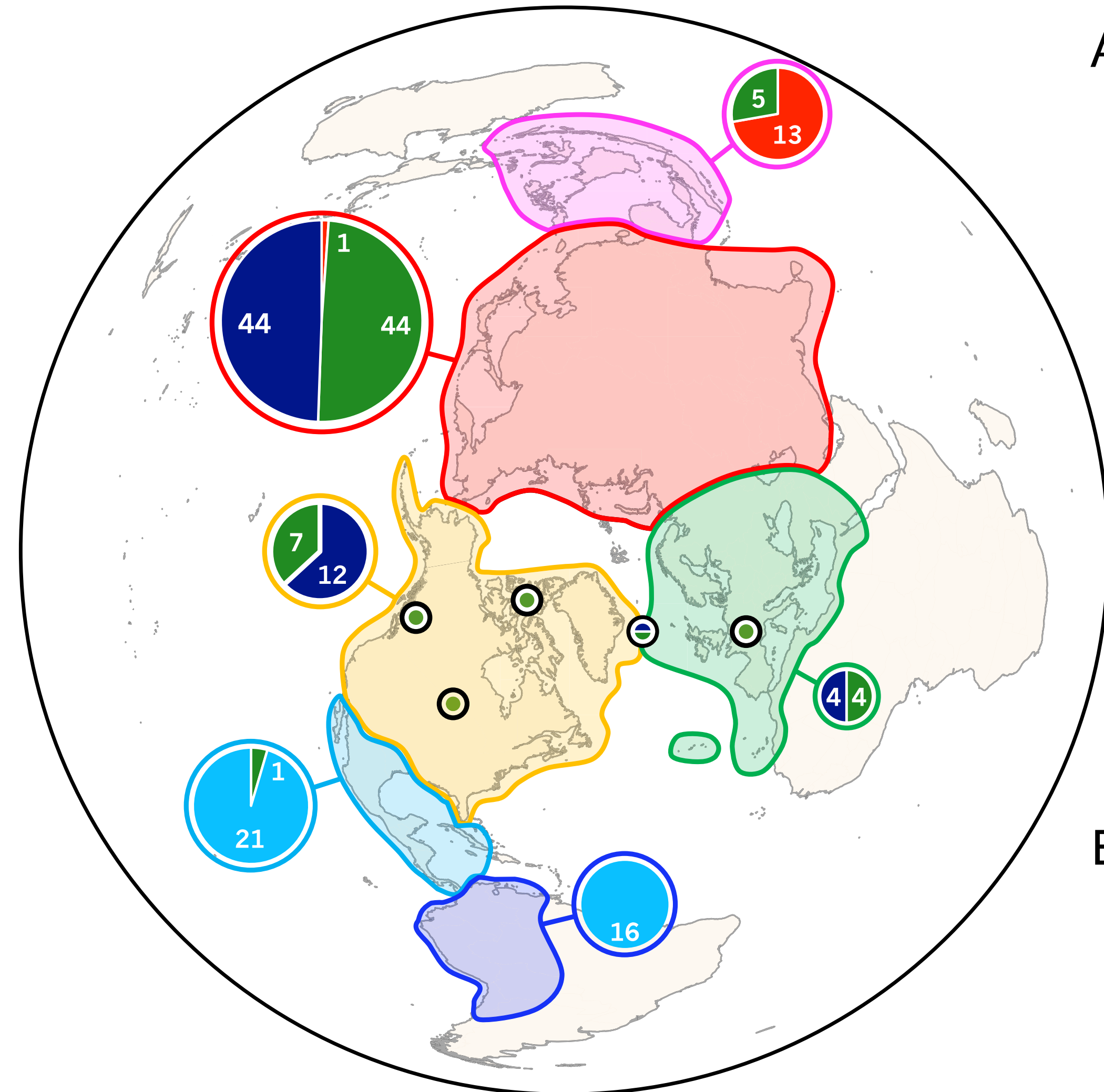
Data

163 extant species (127 with DNA)

5 fossils (with taxonomic constraints)

6 geographic areas

4 biomes



Areas

- SE Asia
- E Asia
- Europe
- N America
- C America
- S America

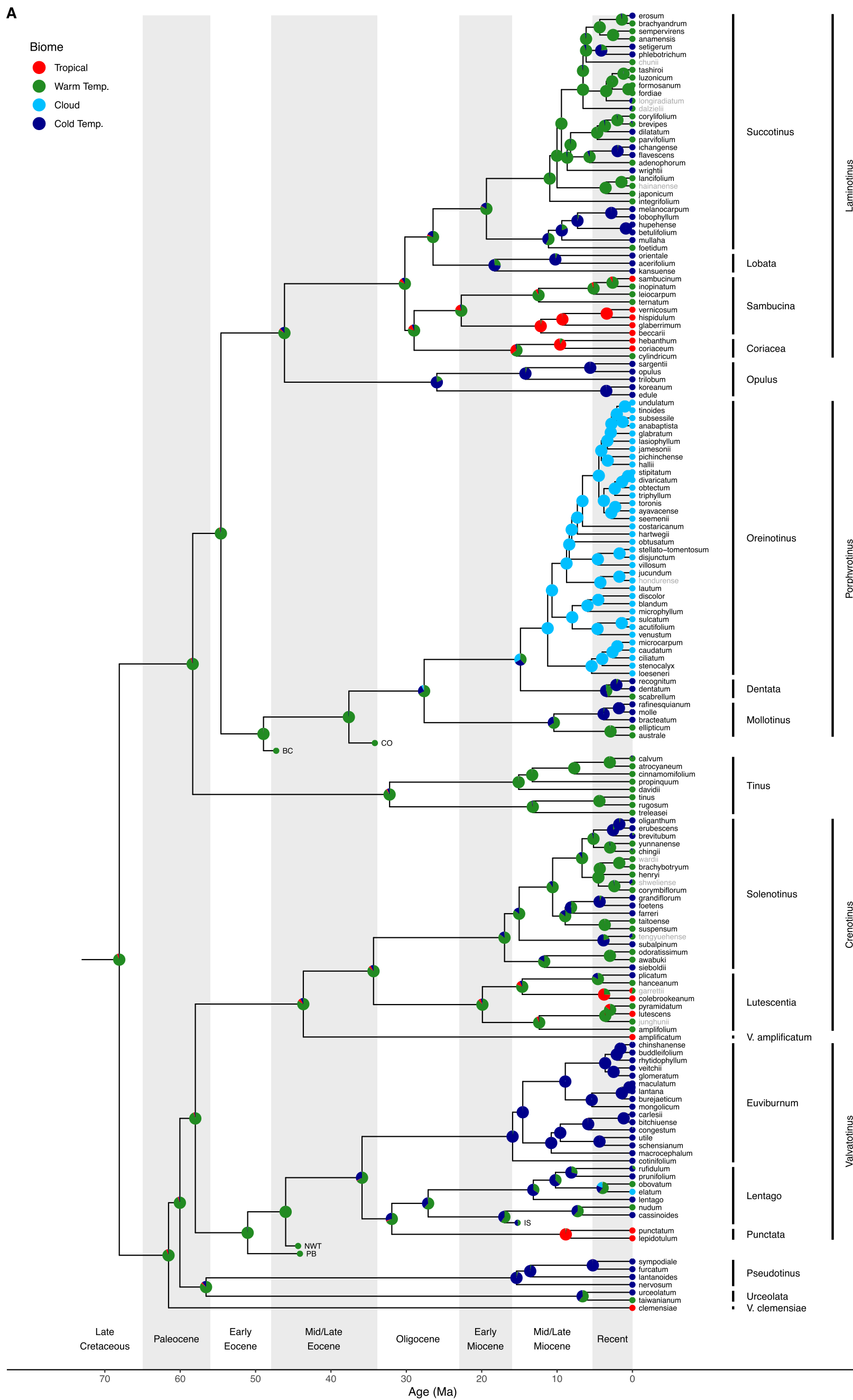
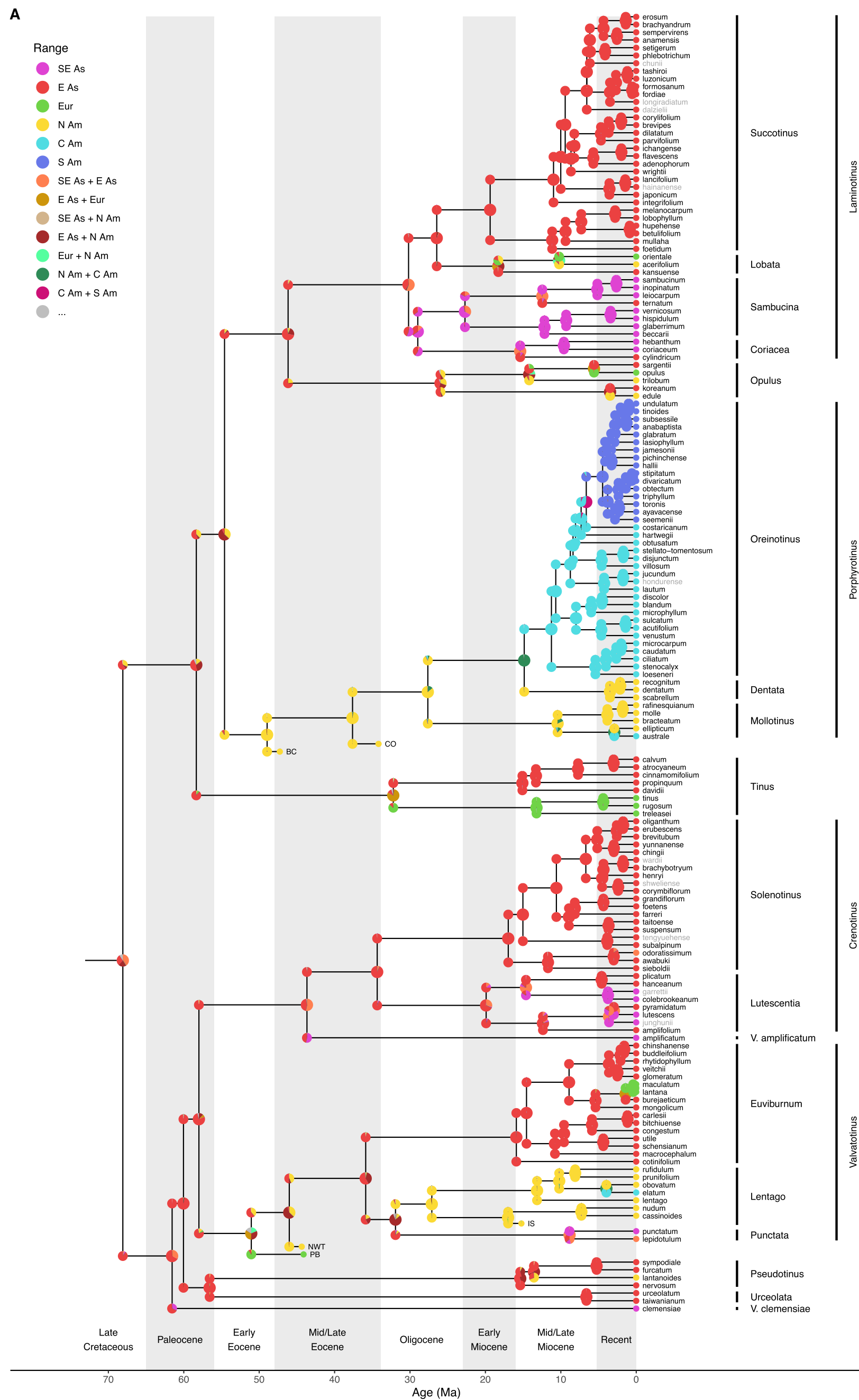
Biomes

- Tropical
- Warm Temperate
- Cloud
- Cold Temperate

- Can we integrate **geographic range** and **biome data** into analysis using the FBD model?
- What can we learn about the **diversification history** of the *Viburnum*?

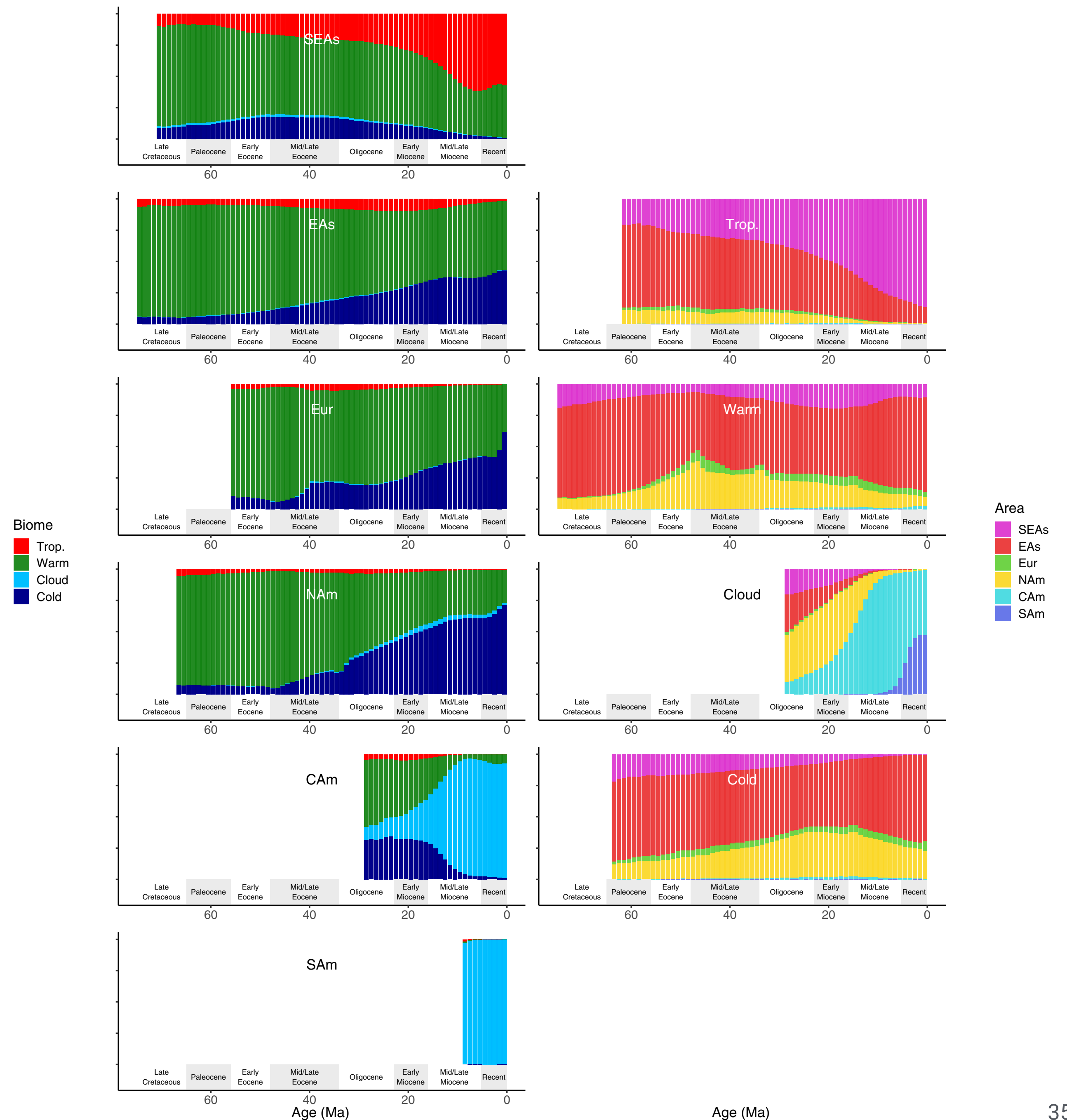
Analysis

1. Estimate the extant topology using maximum likelihood
2. Joint inference divergence times, biogeographic and biome history (normally we *first* infer a dated tree, and *then* separately infer biogeographic history)
3. Ancestral state reconstruction
4. + various sensitivity analyses



Results summary

- Joint inference using the FBD and biogeographic models allows us to estimate a rich diversification history
- Major lineages of *Viburnum* likely originated in warm / temperate regions and later adapted to the cold
- Fossils can change the results



Further reading



[Phylogenetic models of historical biogeography](#) *Michael Landis*

Funk Biogeography Seminar Series



[An Evolving View of Phylogenetic Biogeography](#) *Landis et al. (2026)*

Phylogenetic biogeography

Tutorials

RevBayes - [DEC model](#) + many other options (see [main tutorial page](#))

R - [BioGeoBears](#)

*“It is, it must be admitted, a **humbling** task to infer ancient events, and the results in many cases are tenuous at best. Given the obvious limitations of working with extant species and few, if any, fossils, **it is necessary to integrate all of the available sources of evidence** if we hope to produce assuring answers.”*

Course evaluation

