

# Phylogenetics

Introduction to  
phyldynamics models  
RL-V3 MPP

Rachel Warnock

18.06.2024



# What did you learn about PCMs?

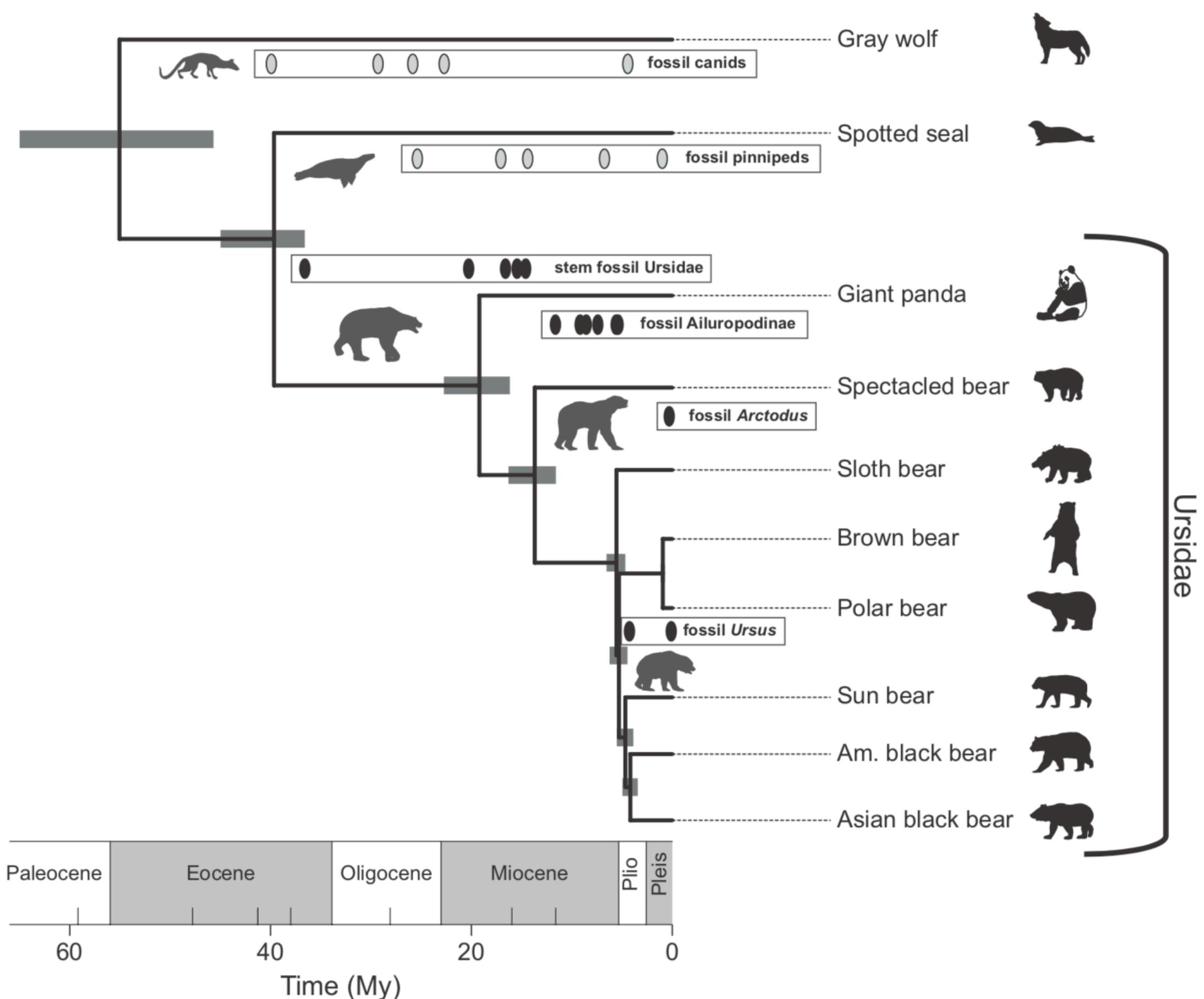
Emma's lecture

What did you learn about models of  
morphological evolution?

Laura's lecture

# Today's objectives

- Recap
  - Tripartite framework
  - The fossilised birth-death process
  - Total-evidence dating
  - Phylogenetics



# Bayesian divergence time estimation

Recap

# We use a Bayesian framework

$$P(\text{ model } | \text{ data }) = \frac{P(\text{ data } | \text{ model }) P(\text{ model })}{P(\text{ data })}$$

likelihood

priors

posterior

marginal probability of the data

# Bayesian divergence time estimation

## The data

and / or

0101... ATTG...

1101... TTGC...

0100... ATTC...



phylogenetics  
characters

sample  
ages

# Bayesian divergence time estimation

## The data

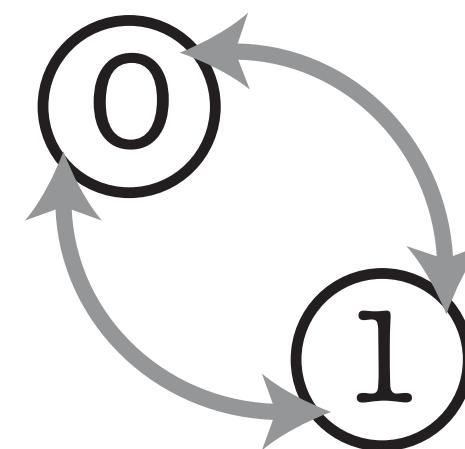
and / or  
0101... ATTG...  
1101... TTGC...  
0100... ATTC...



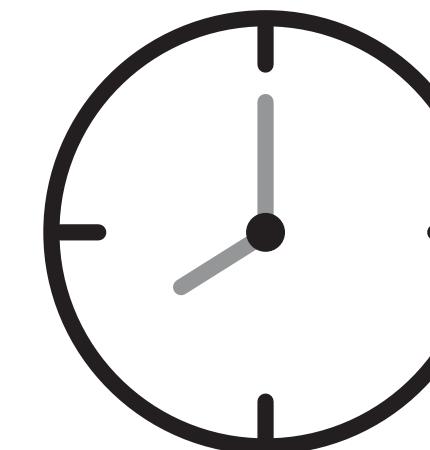
phylogenetics  
characters

sample  
ages

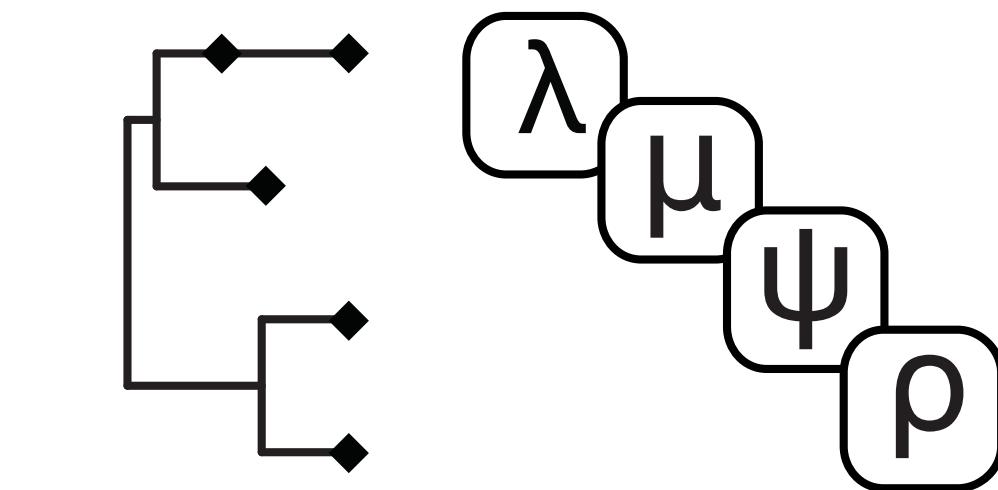
## 3 model components



substitution  
model

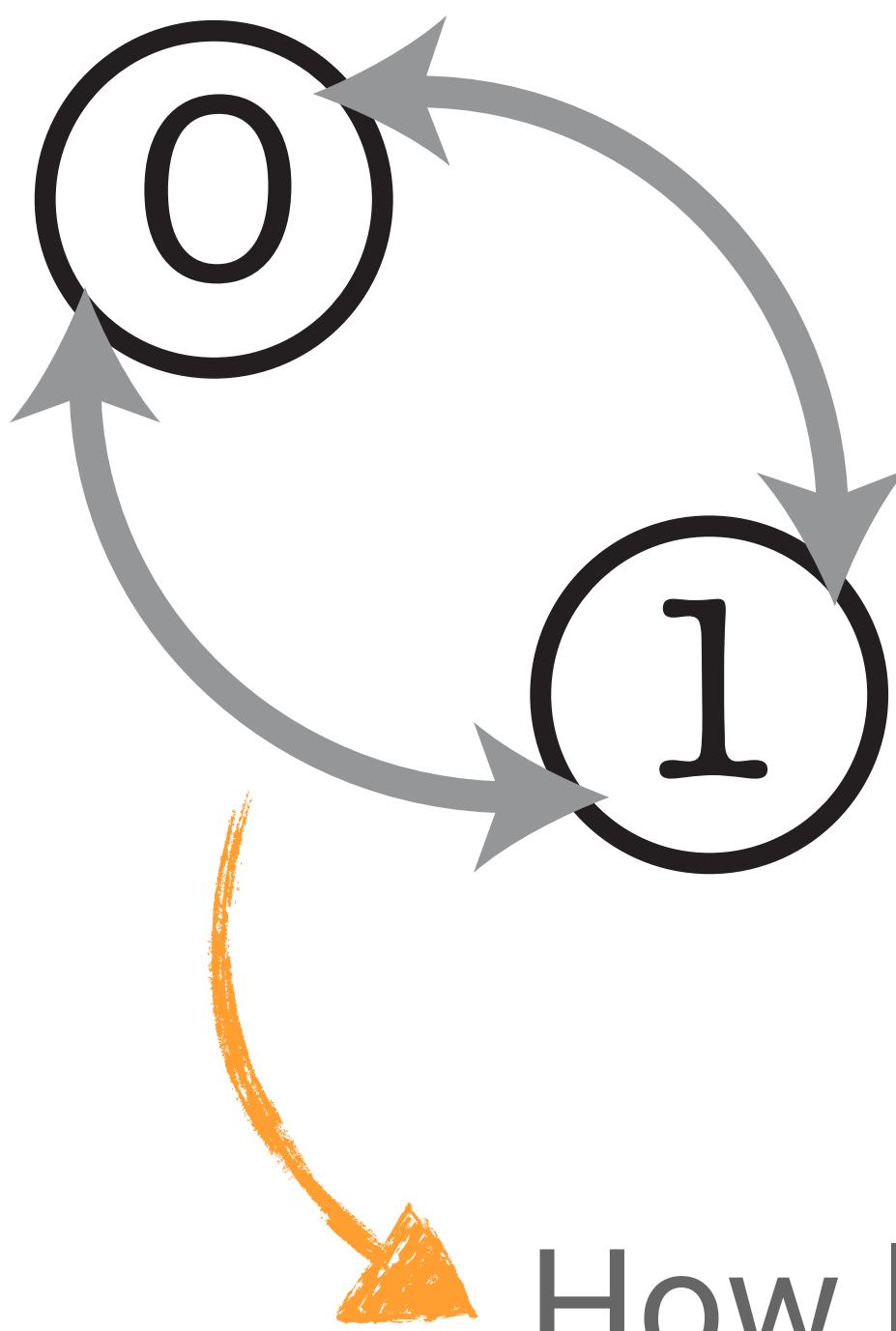


clock  
model

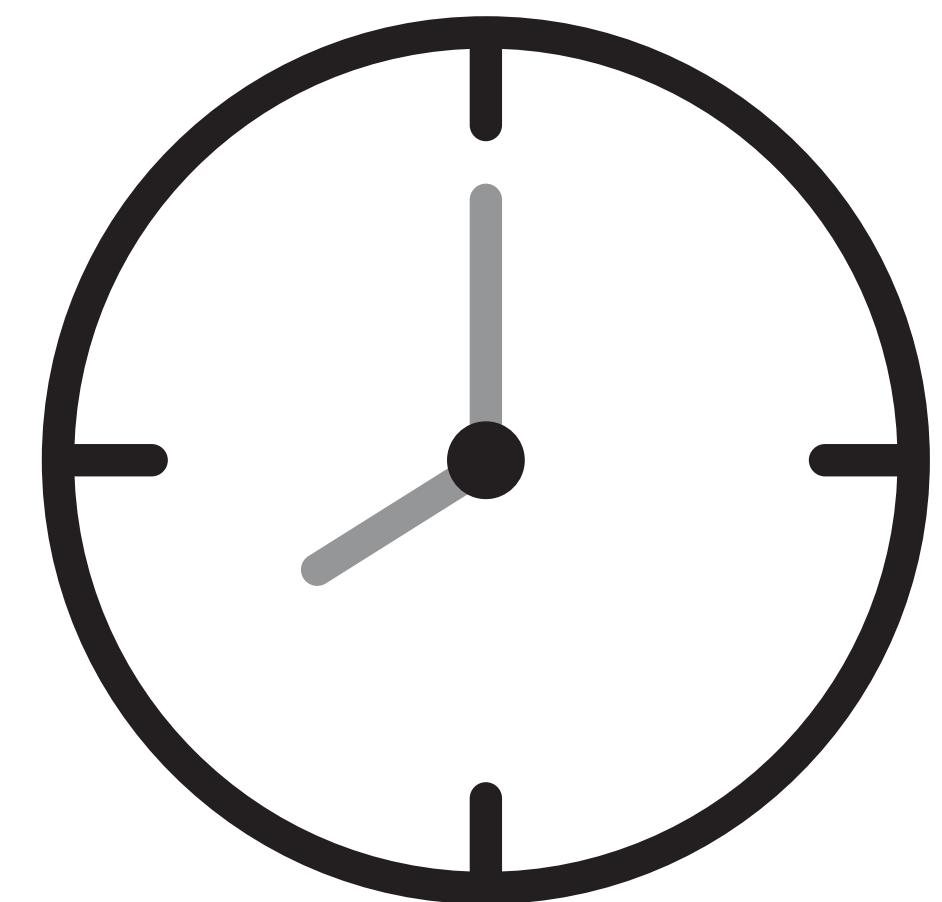


tree and tree  
model

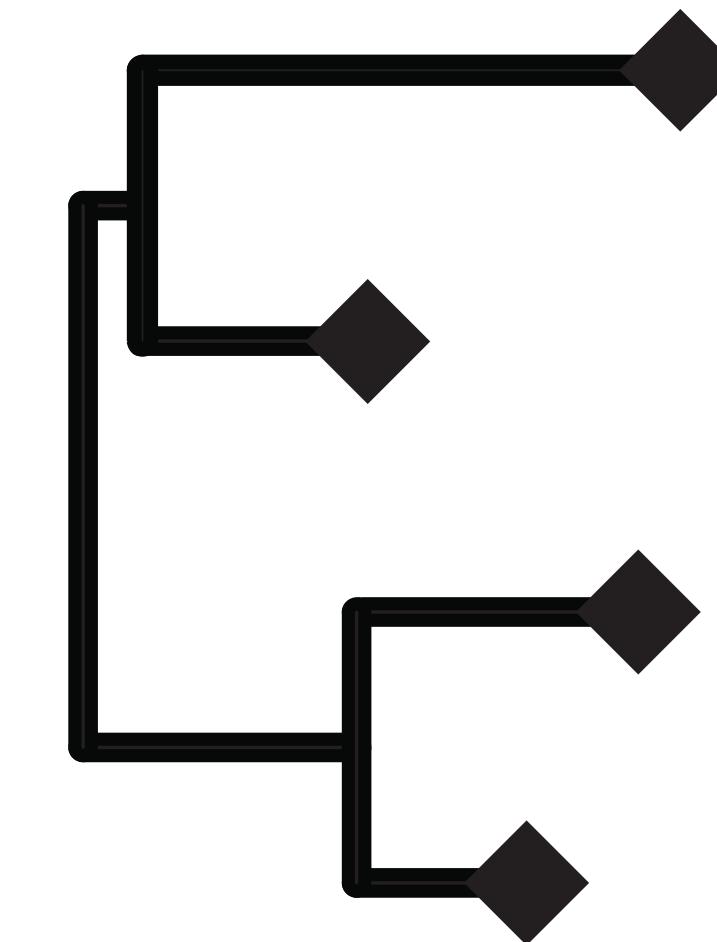
substitution model



clock model

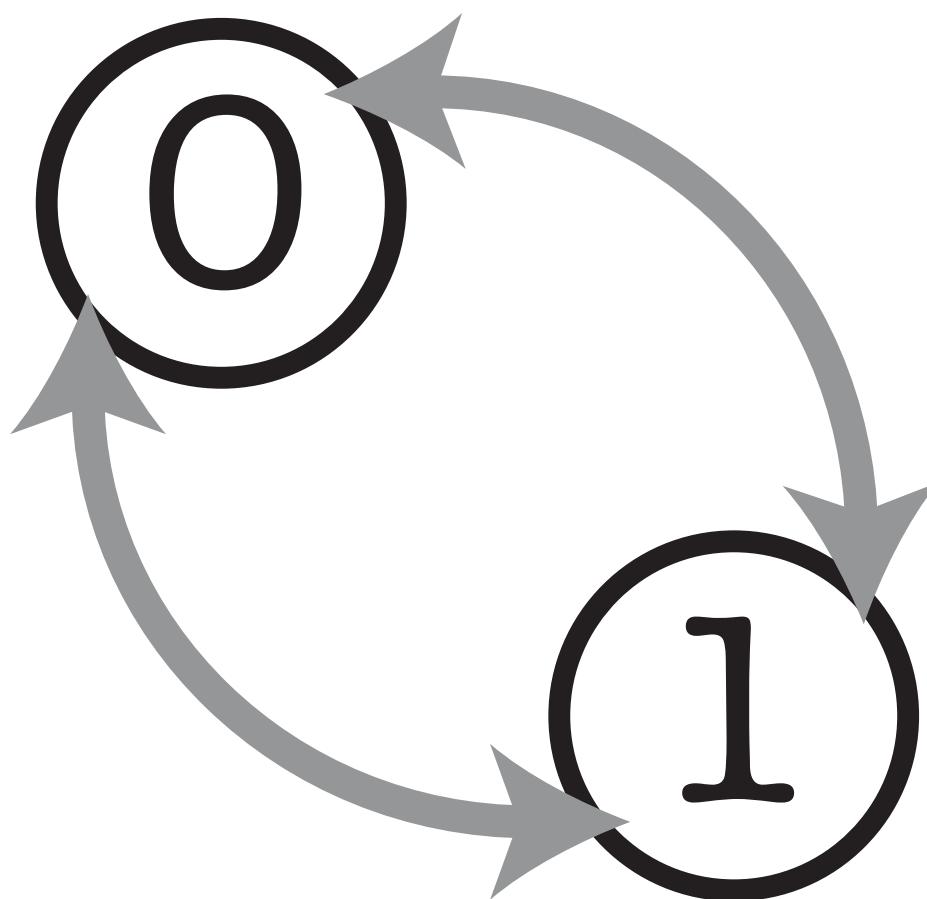


tree model

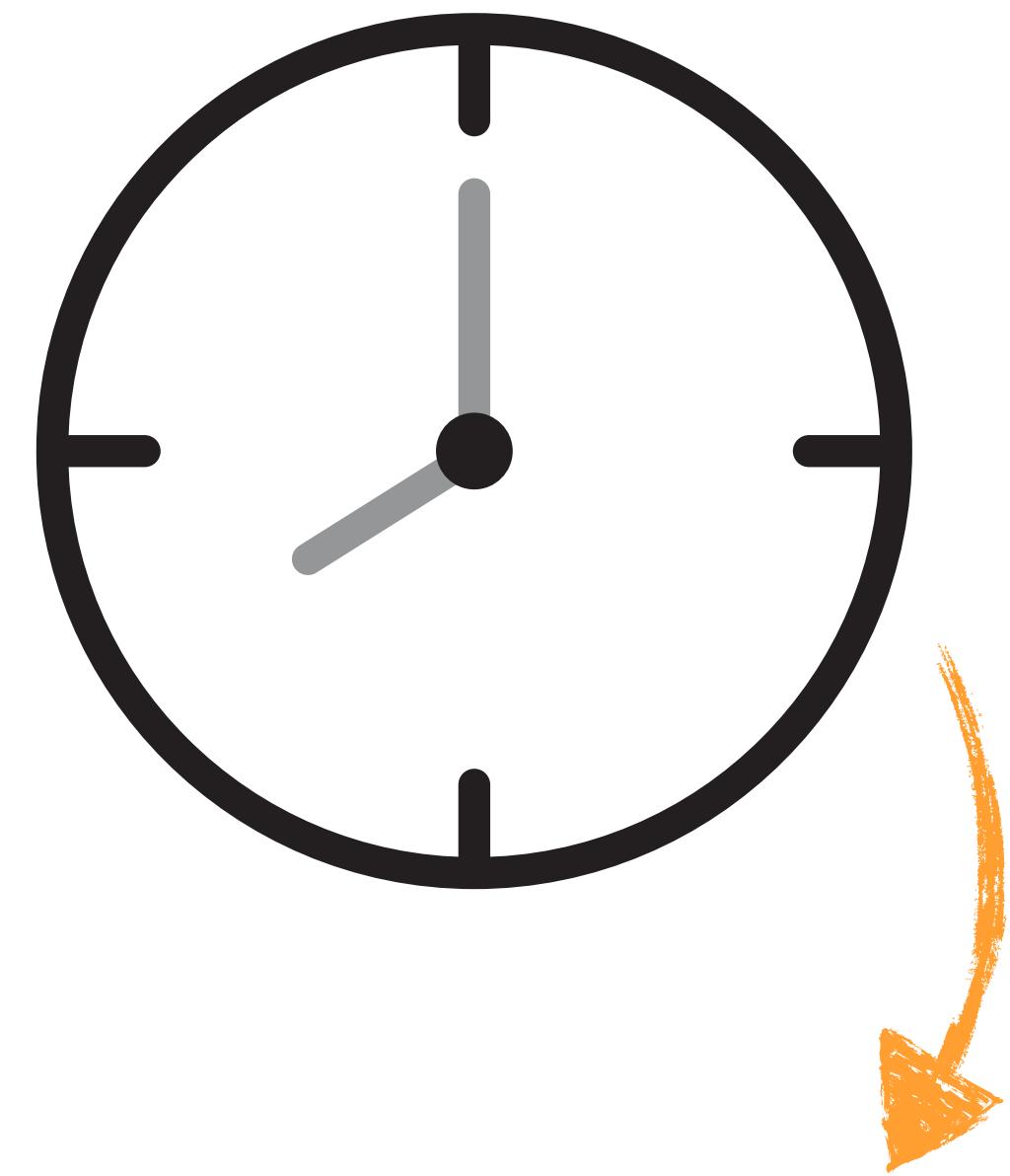


How likely are we to observe a change  
between character states? e.g.,  $A \rightarrow T$

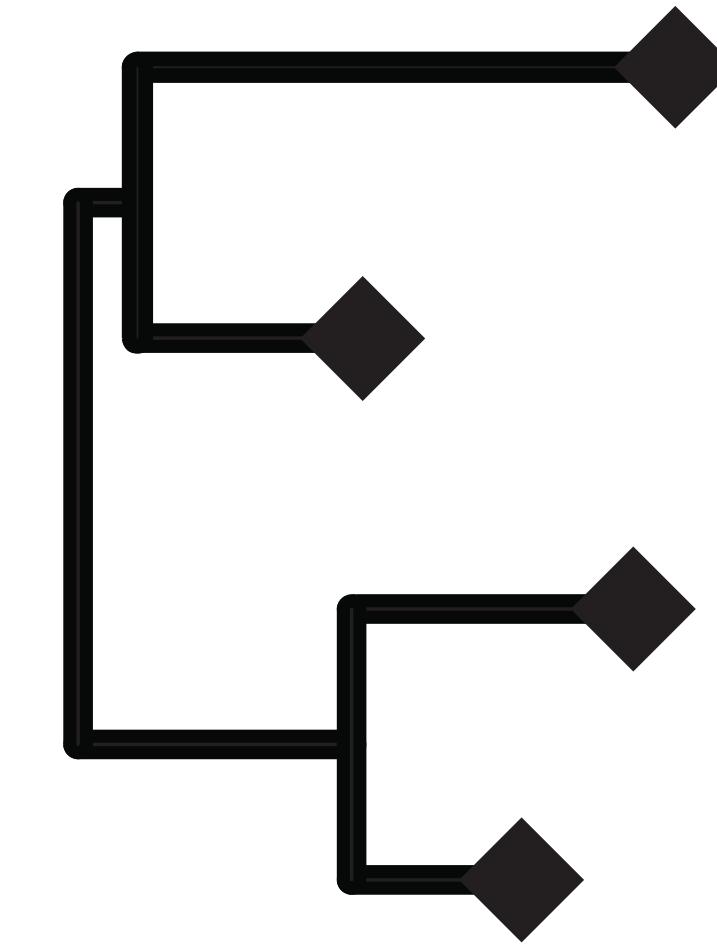
substitution model



clock model

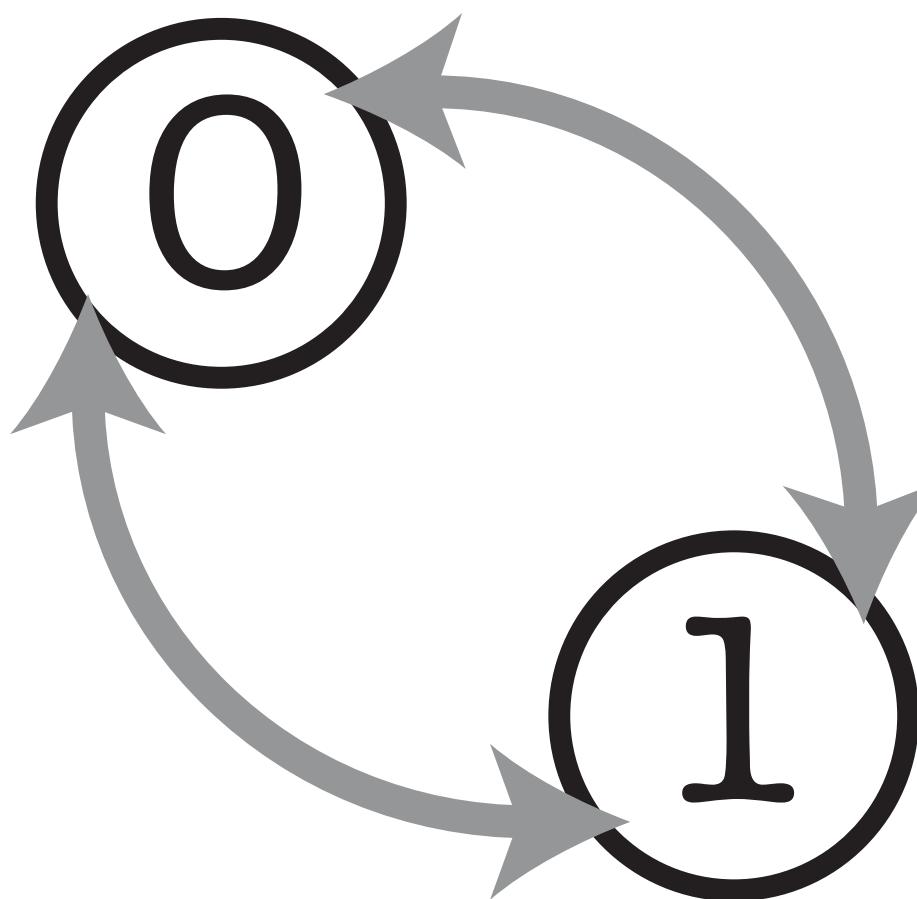


tree model

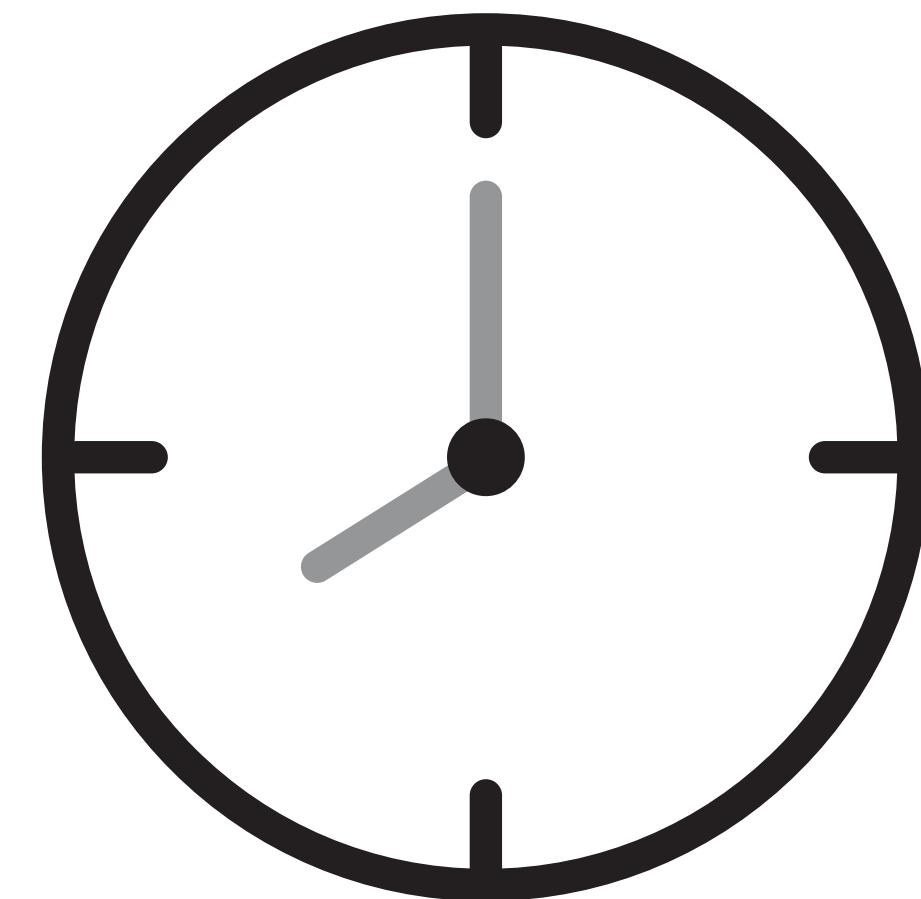


How have rates of evolution varied  
(or not) across the tree?

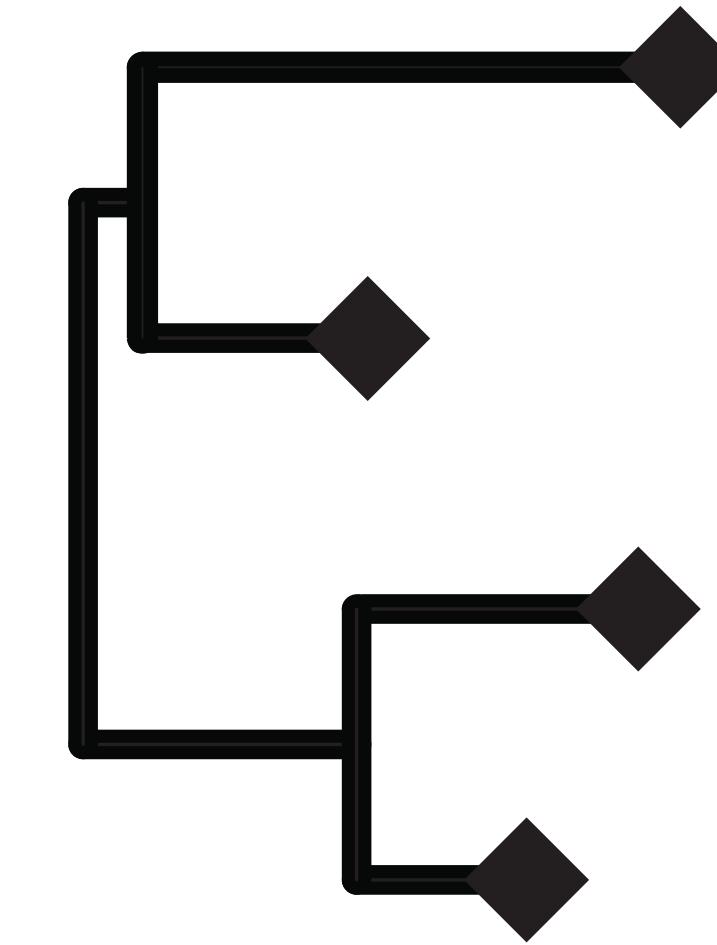
substitution model



clock model



tree model

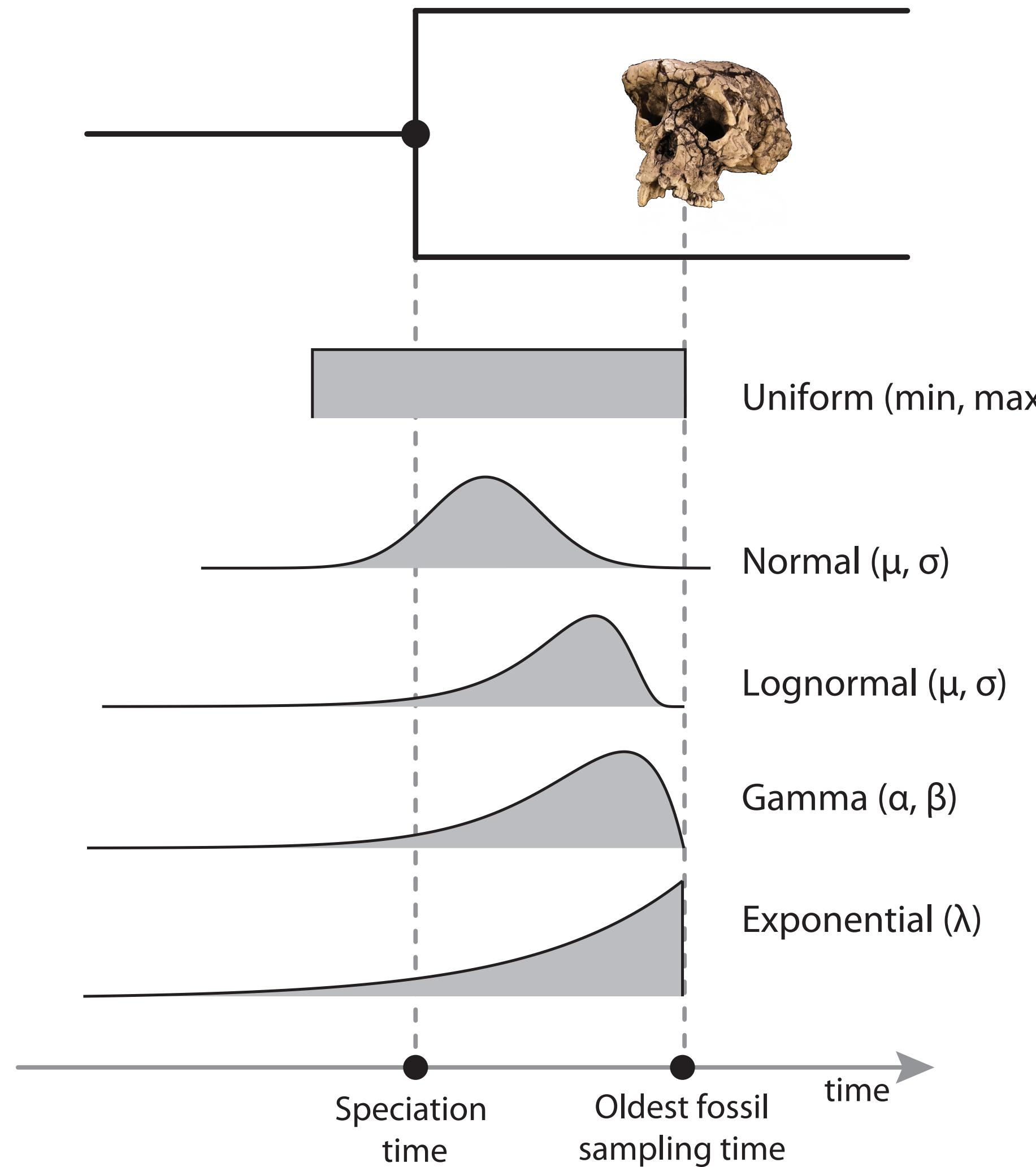


How have species originated, gone  
extinct and been sampled through time?

probability of the  
time tree

$$P(E | \mathcal{G}^{\lambda, \mu, \psi, p})$$

# Recap: Node dating



We can use a **calibration density** to constrain internal node ages

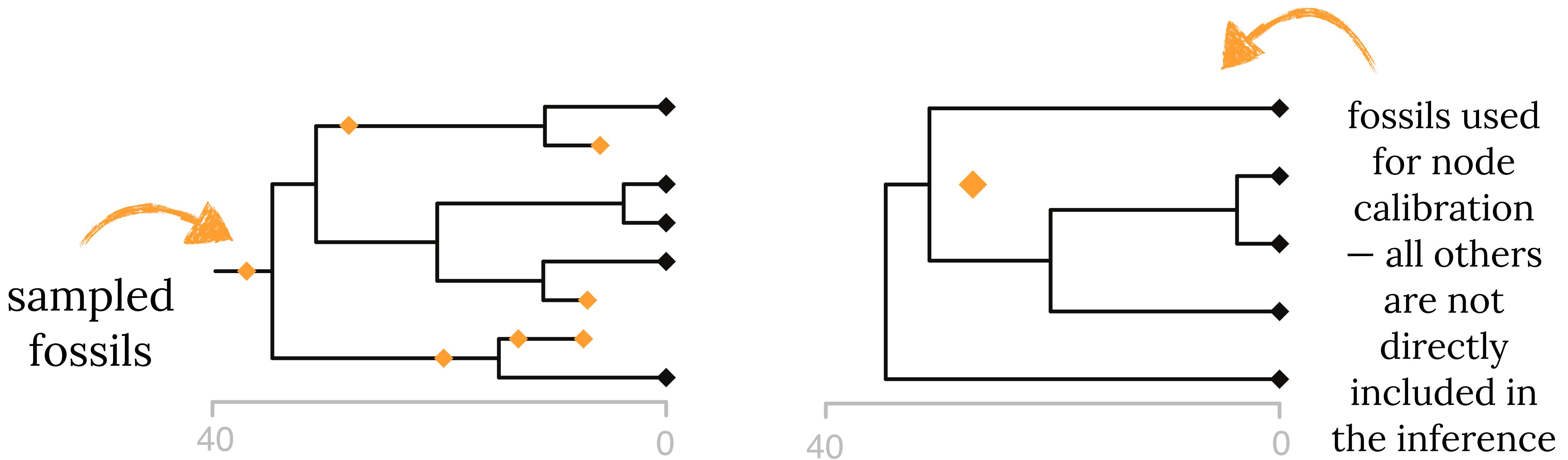
We typically use a **birth-death process** model to describe the tree generating process

Adapted from Heath 2012. Sys Bio

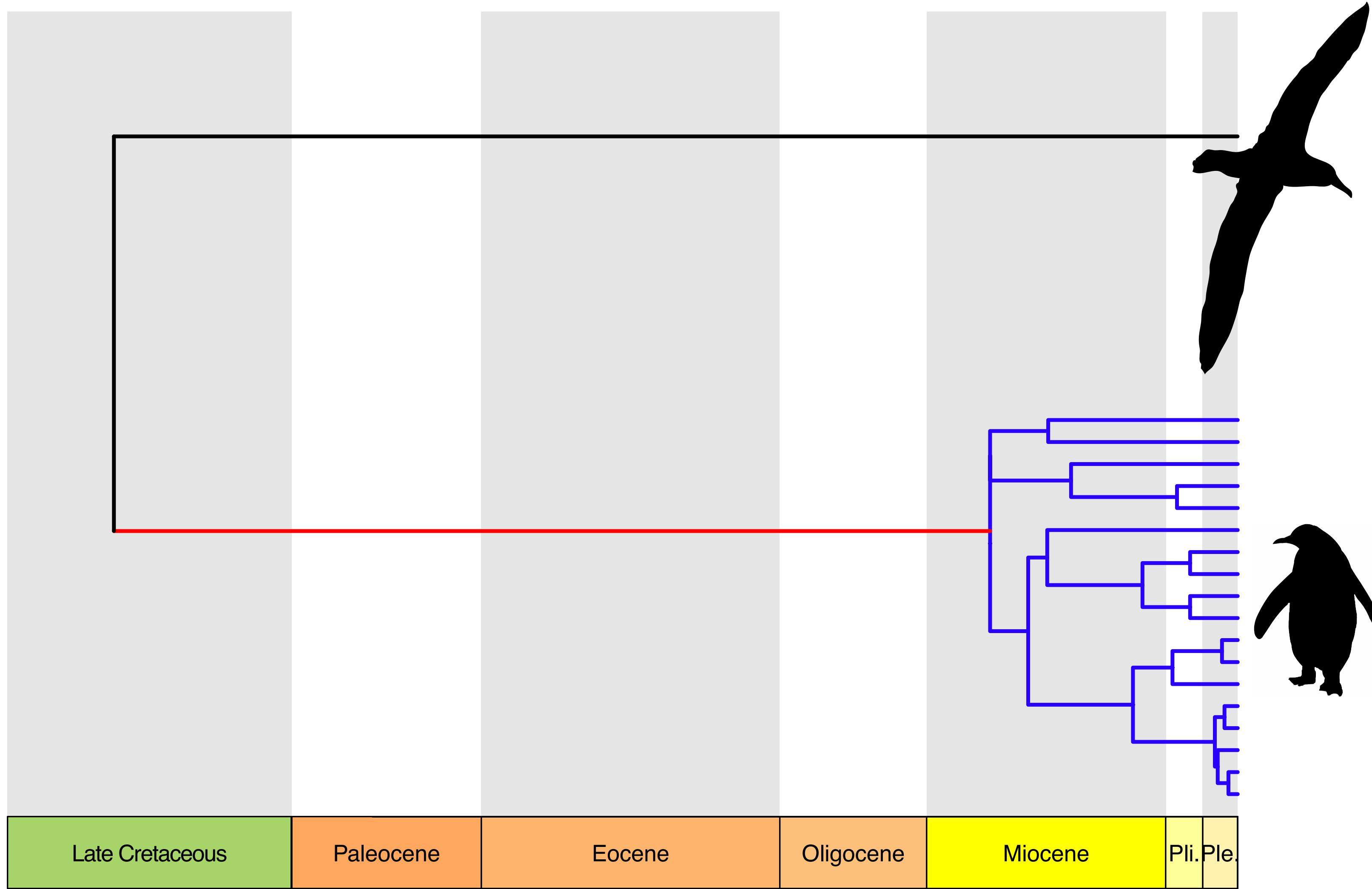
# Node dating: potential issues

# There are many!

A lot of information is excluded, since typically we assign one fossil per calibration node

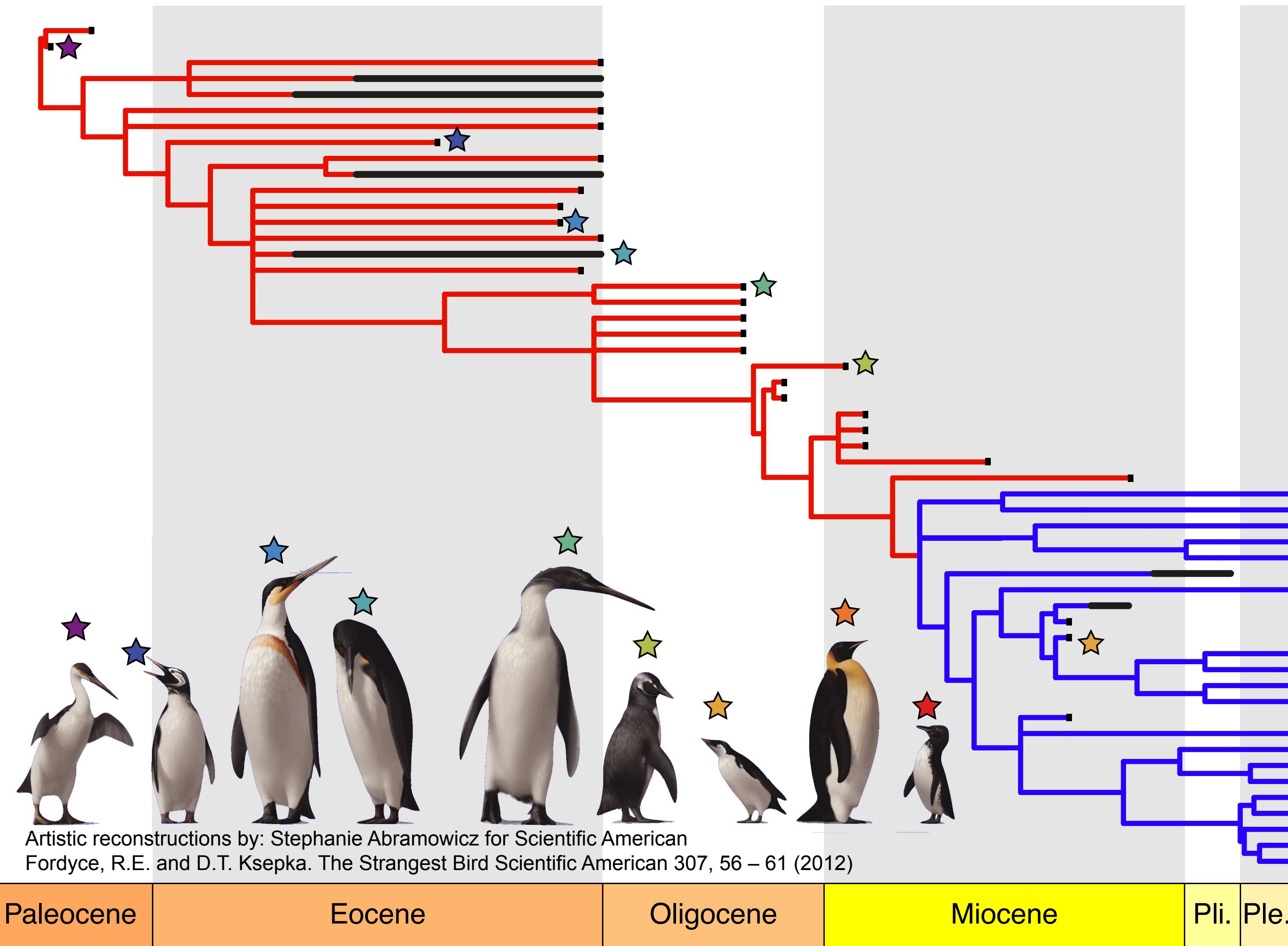


# Example: living penguins



Nearest living relative is the group containing falcons - separated by ~60 Ma

# Example: living penguins



But penguins  
have a rich  
fossil record!

# Node dating: potential issues

The model doesn't describe the process that generated the fossil sampling times, meaning the model is **statistically incoherent**

The calibration priors are difficult to specify objectively and can have a massive impact on the divergence times. They can also interact with each other and / or the birth-death process prior in unintuitive ways

Some references on issues with specified vs effective priors

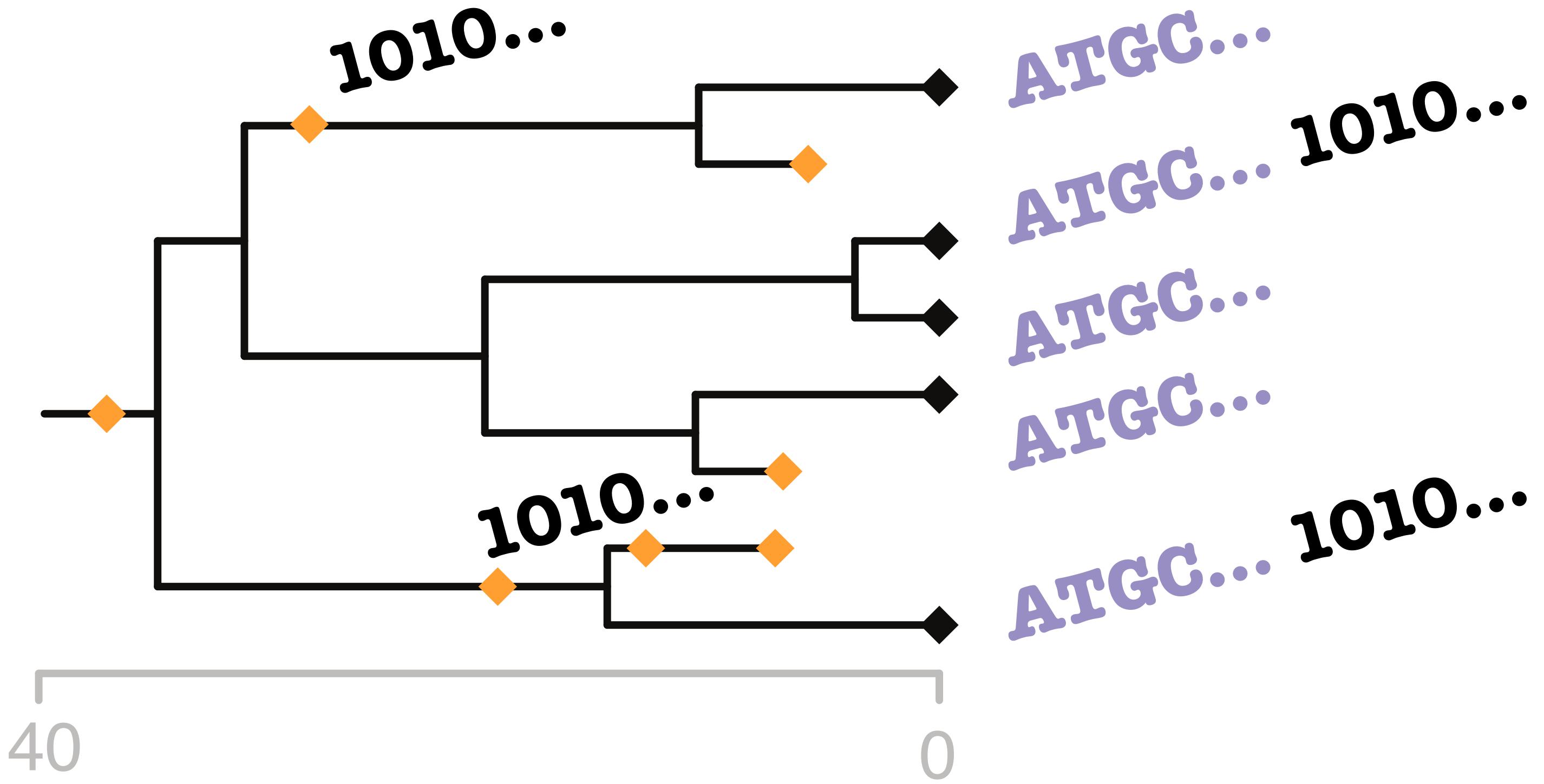
*Yang and Rannala. 2006. MBE*

*Heled and Drummond. 2012. Sys Bio*

*Warnock et al. 2012, 2015*

# Total-evidence dating

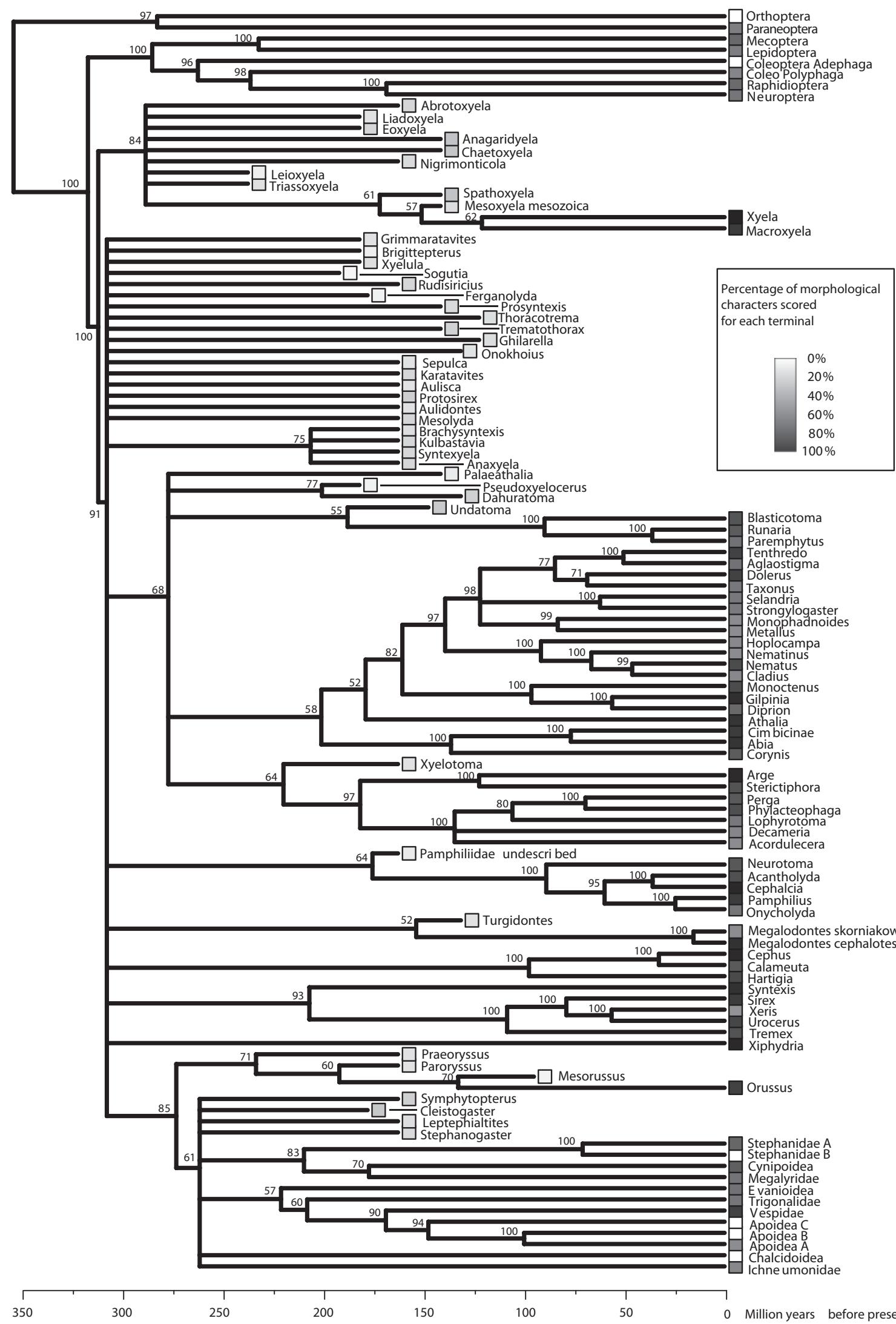
# Tip-dating or “total-evidence” dating



We have DNA for living species. We have morphology for living *and* fossil species

Fossils can be positioned on the basis of morphology  
→ accounts for uncertainty in fossil placement

# The uniform tree prior

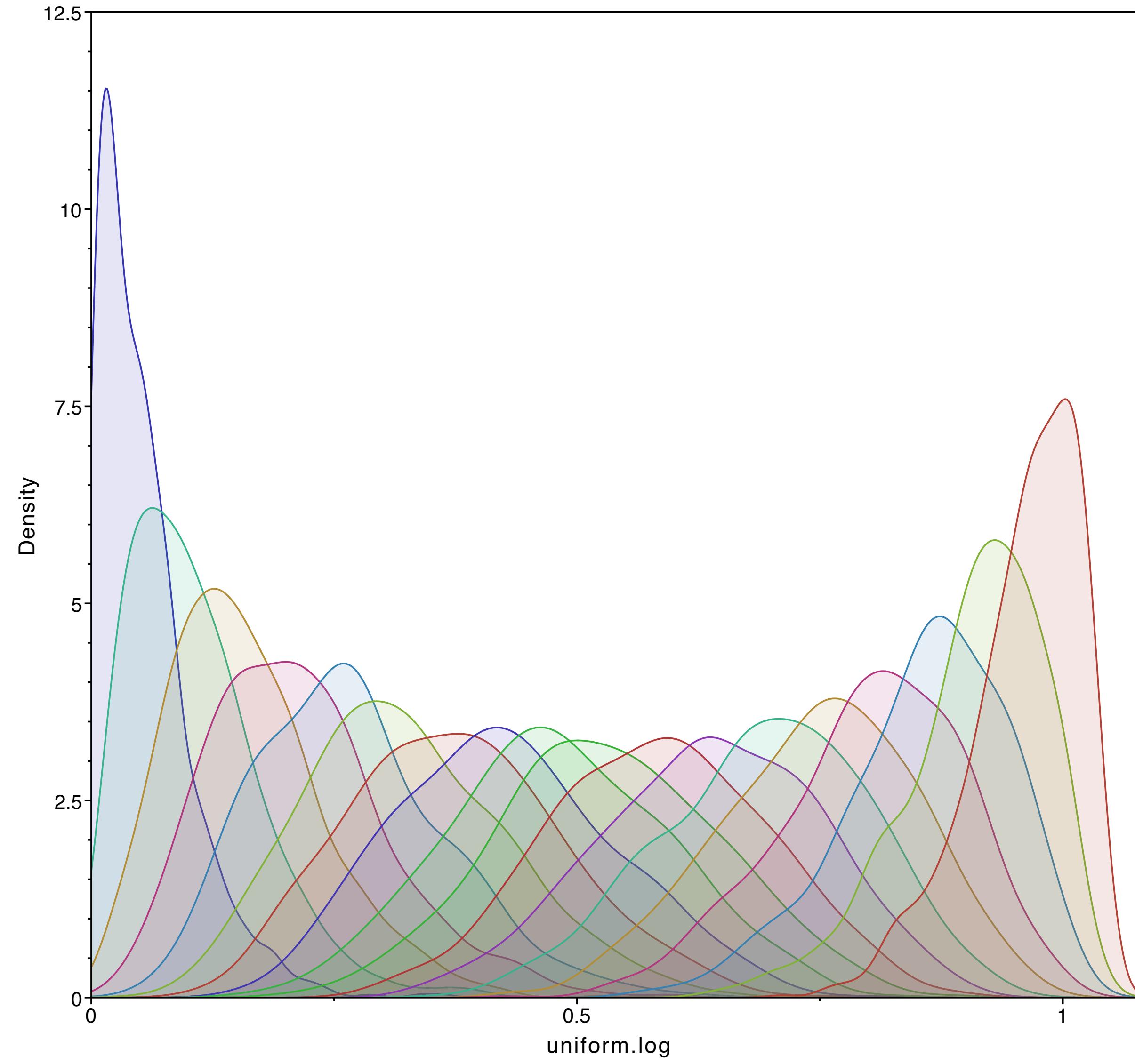


Dated tree of Hymenoptera



The uniform tree prior assumes all trees and branch lengths are equally likely within the bounds of the fossil ages (+ a max upper bound)

It does not explicitly account for the fossil sampling process

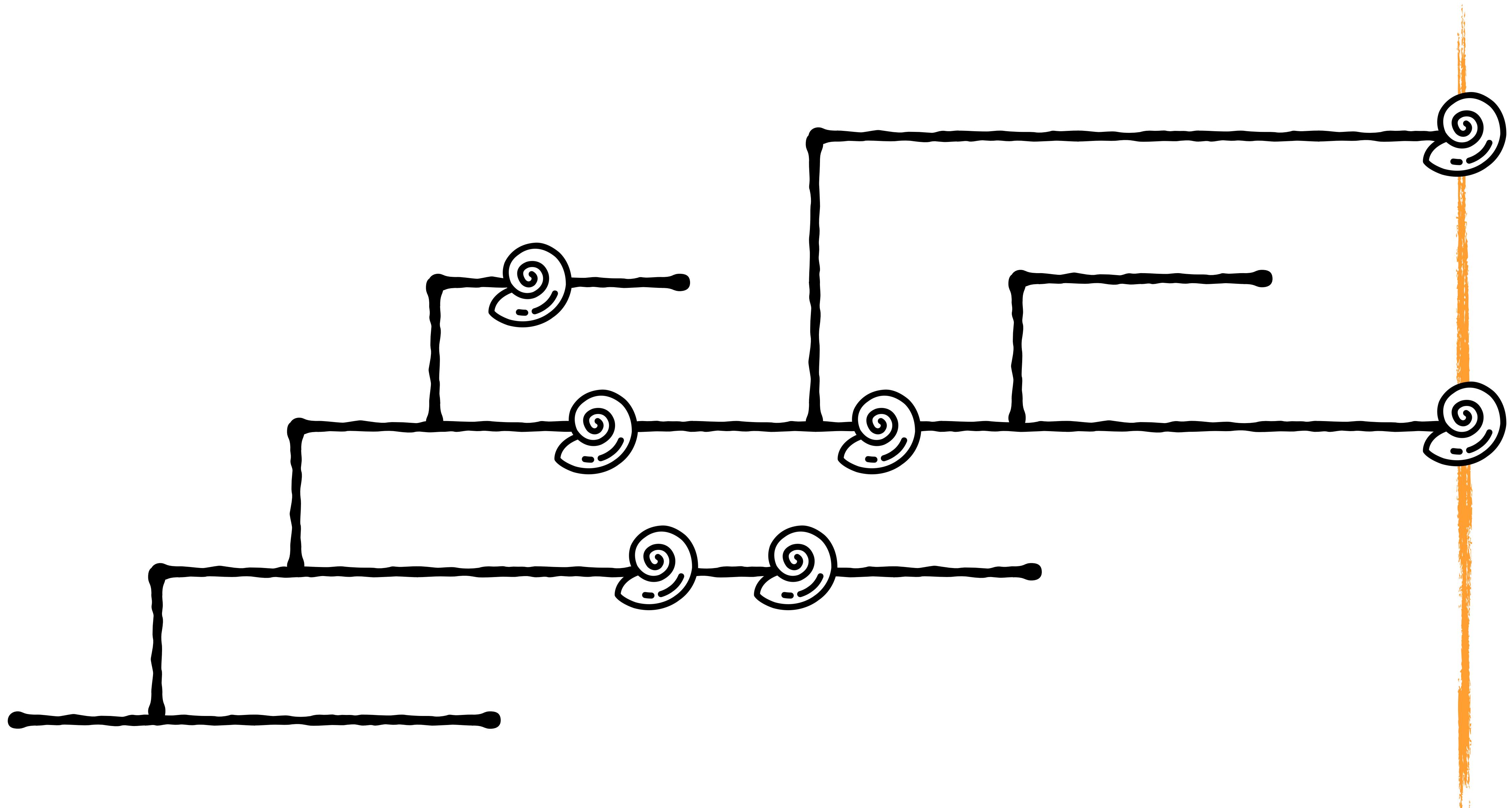


A uniform tree prior implies time till the next split is independent of how many lineages there are present

This is in contrast to birth-death processes, where more lineages mean a higher chance of observing a split in one of these lineages

What does a generating prior for the fossil record look like?

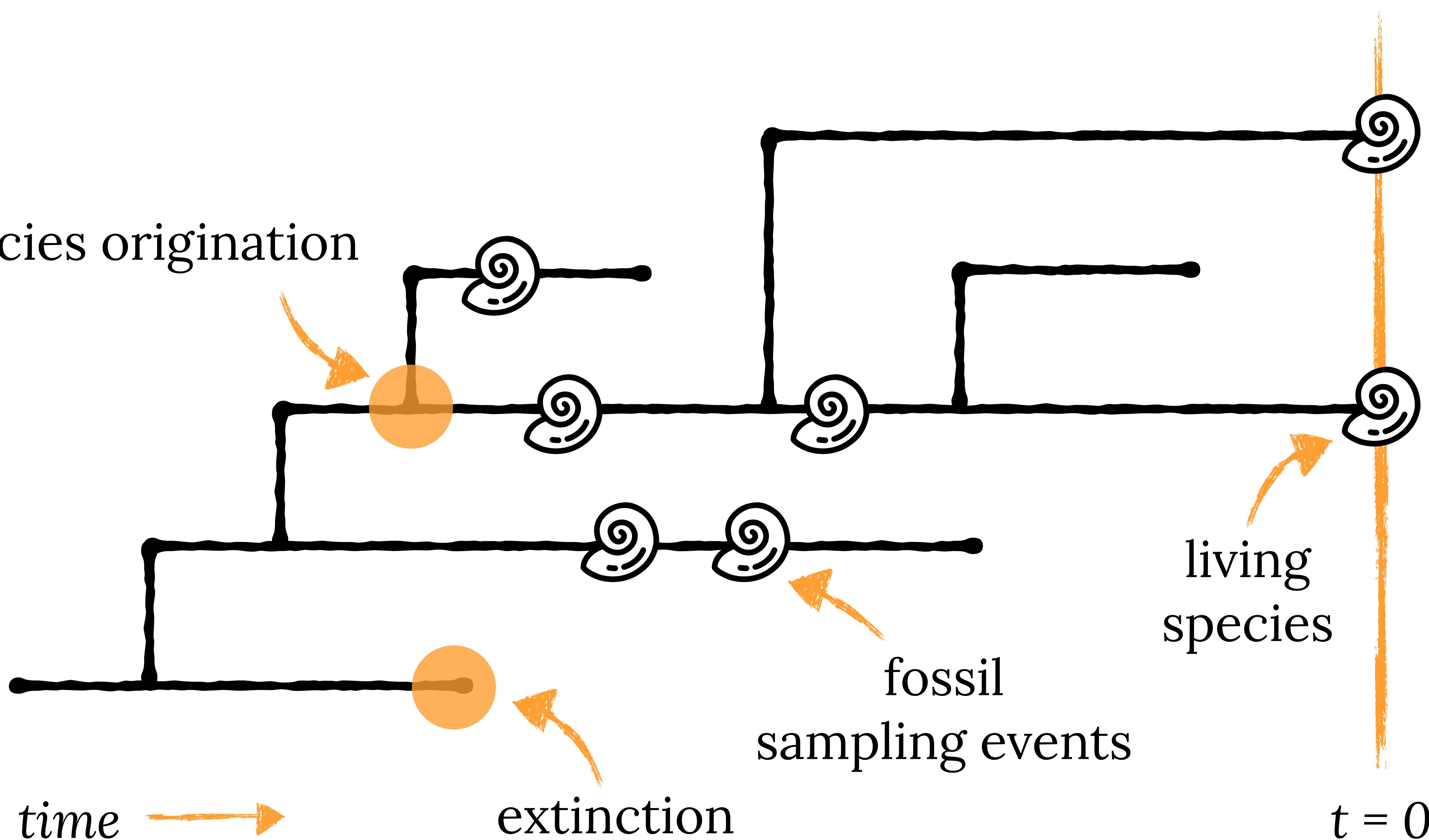
# The fossilised birth-death process



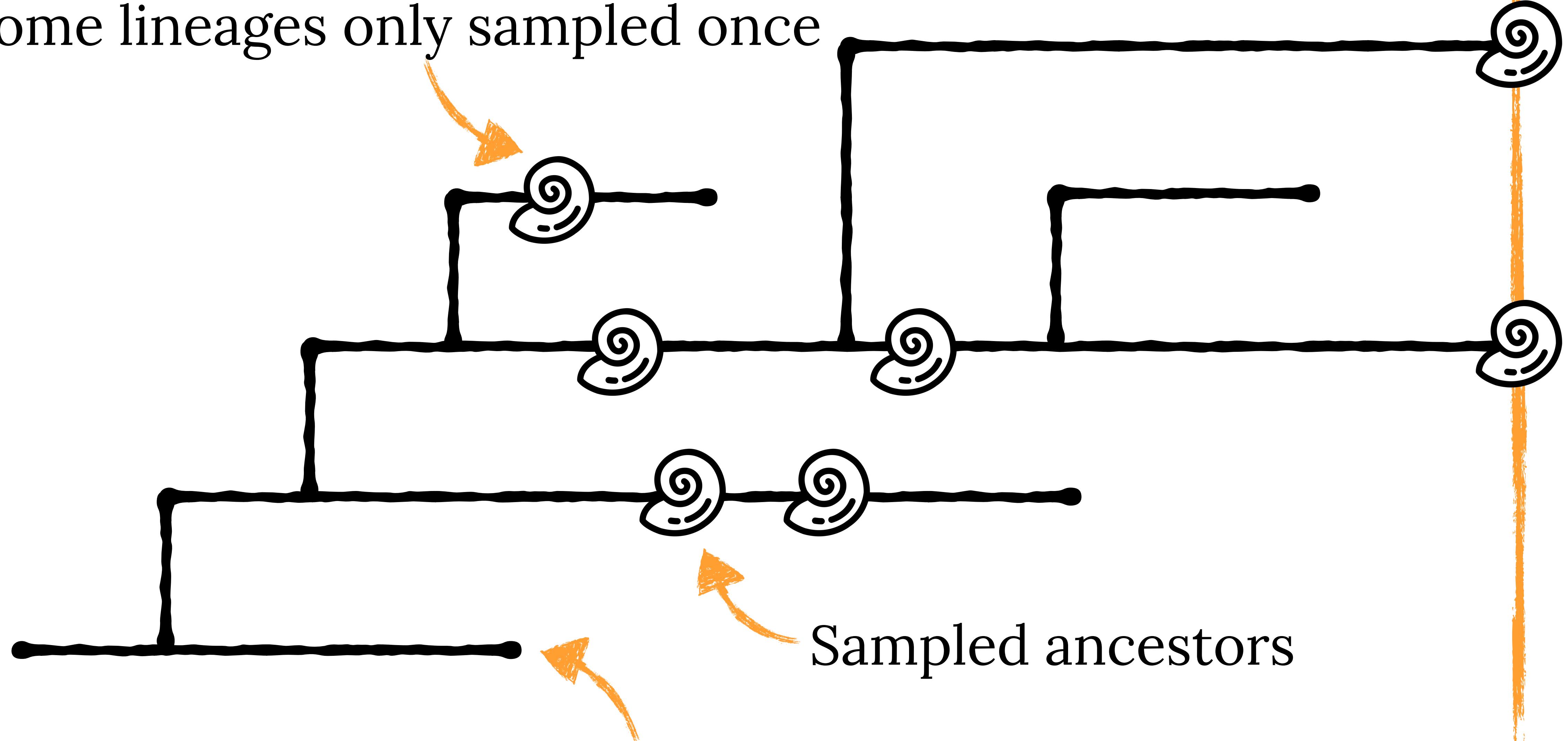
time

$t = 0$

species origination



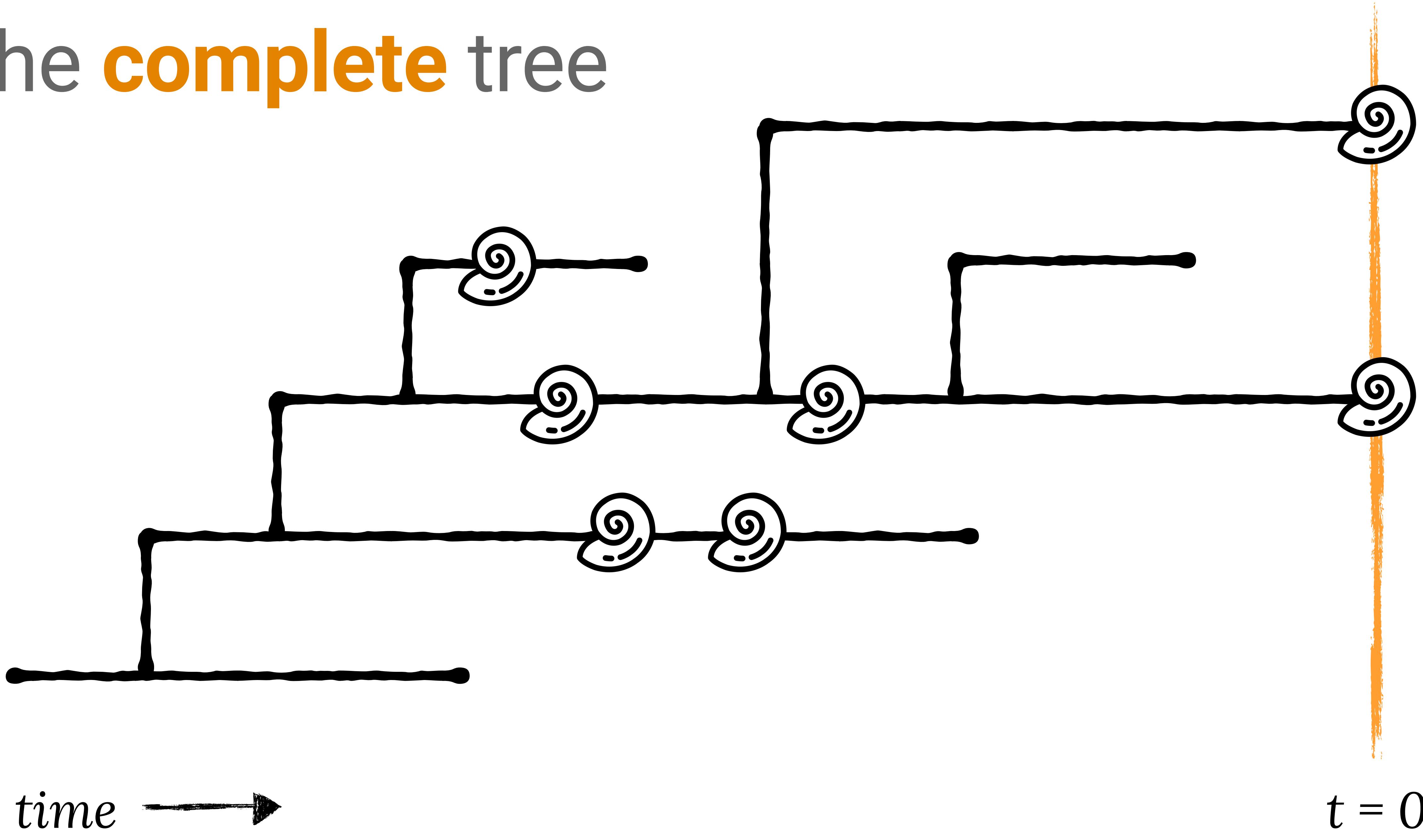
Some lineages only sampled once



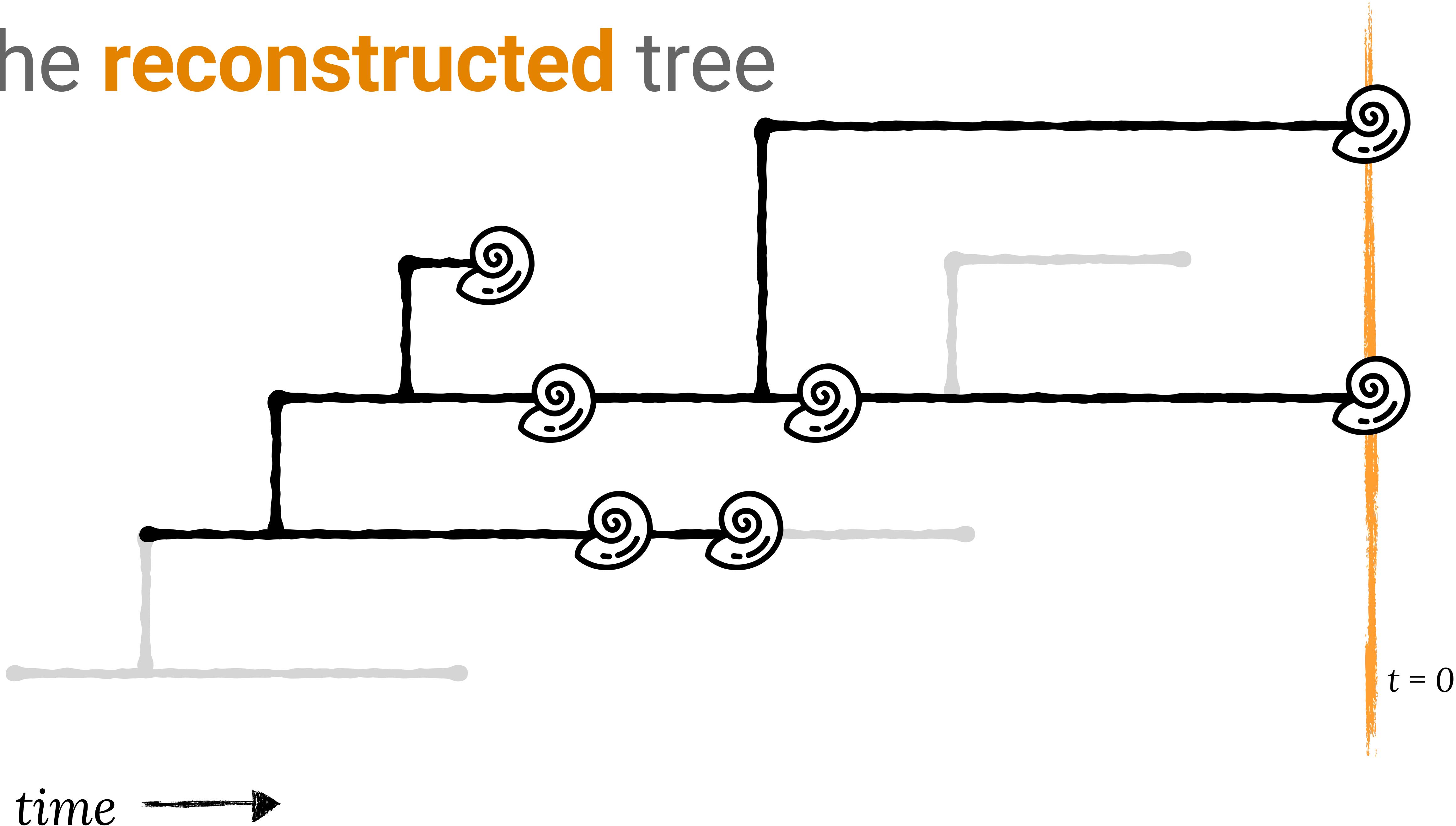
Some lineages go completely unsampled

$t = 0$

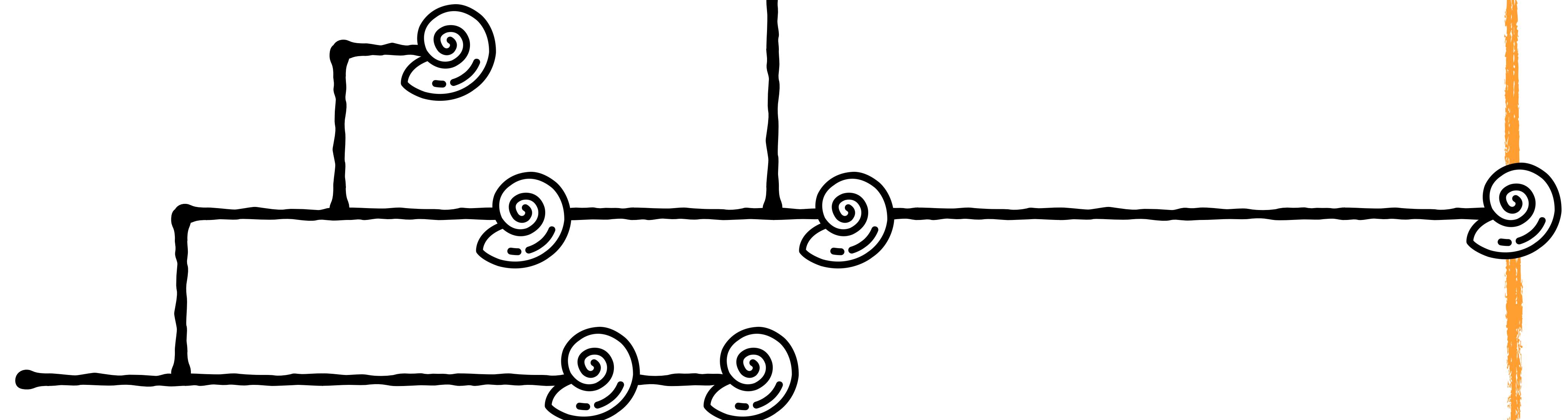
# The complete tree



# The reconstructed tree



The **fossilised birth-death (FBD) process** allows us to calculate the probability of observing the reconstructed tree

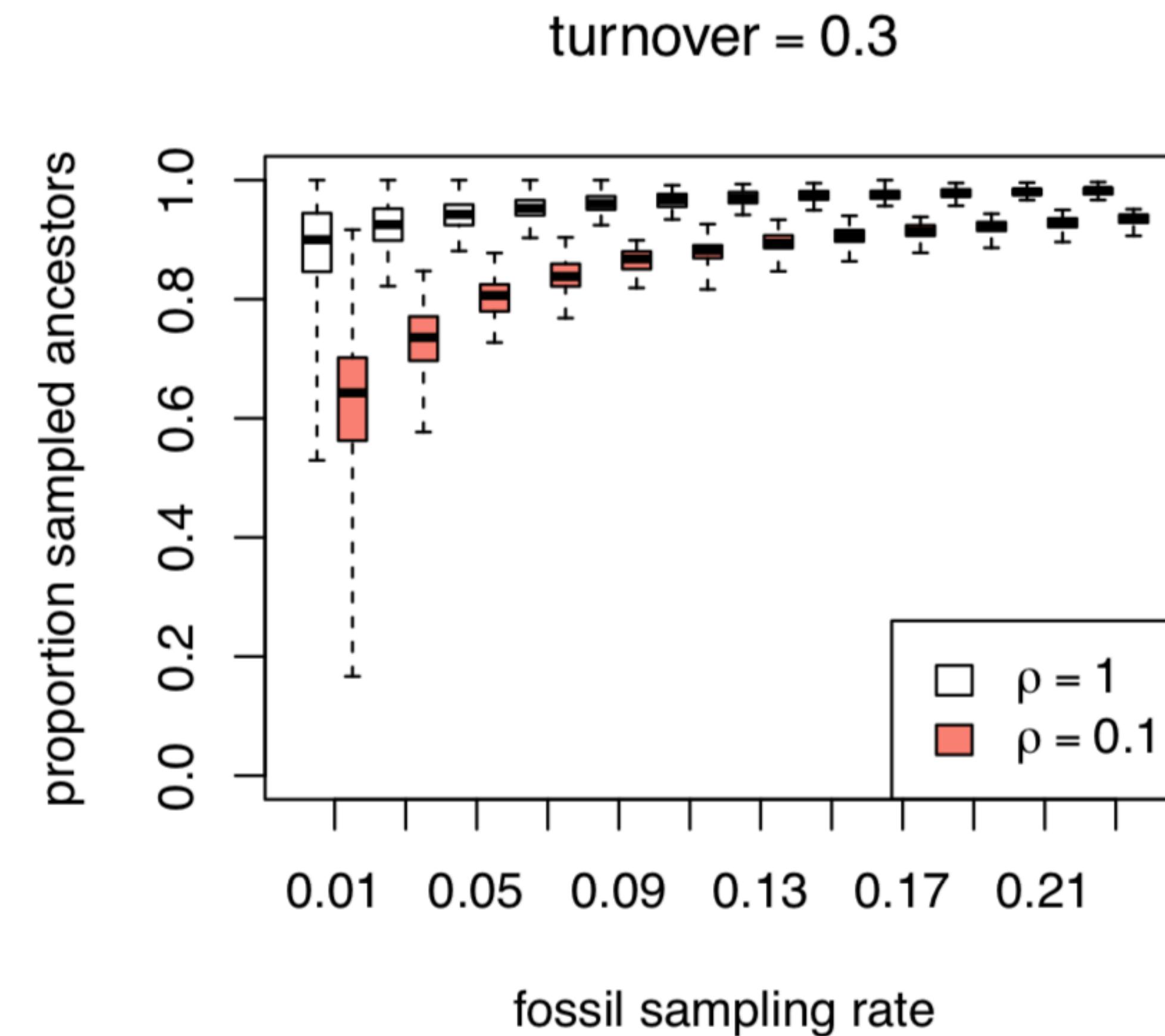
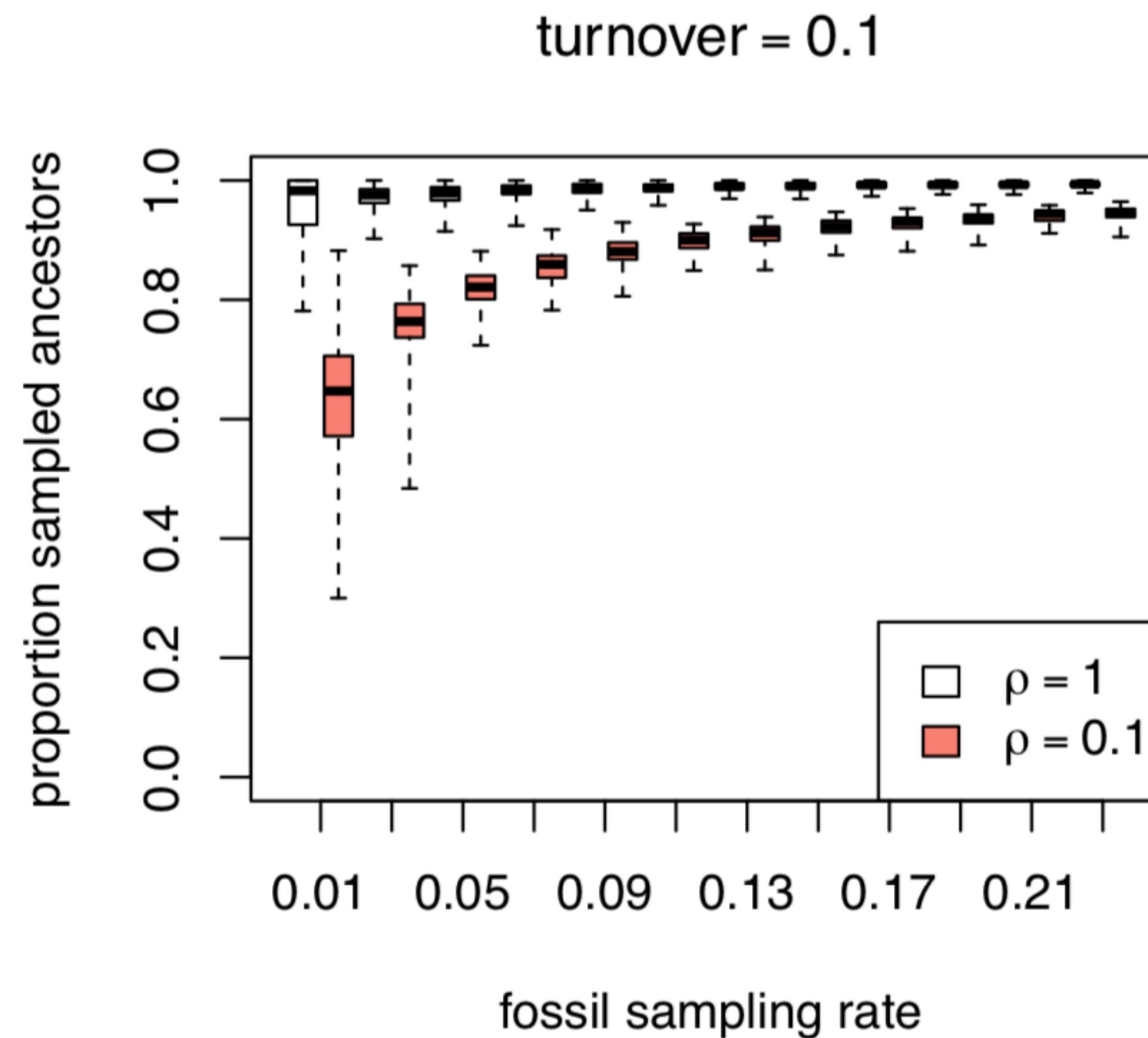


$$P(E | \text{snail}, \lambda, \mu, \psi, \rho)$$

Sampling-through-time in birth-death trees. Stadler. (2010)  
First implemented: Heath et al. (2014) and Gavryushkina et al. (2014)

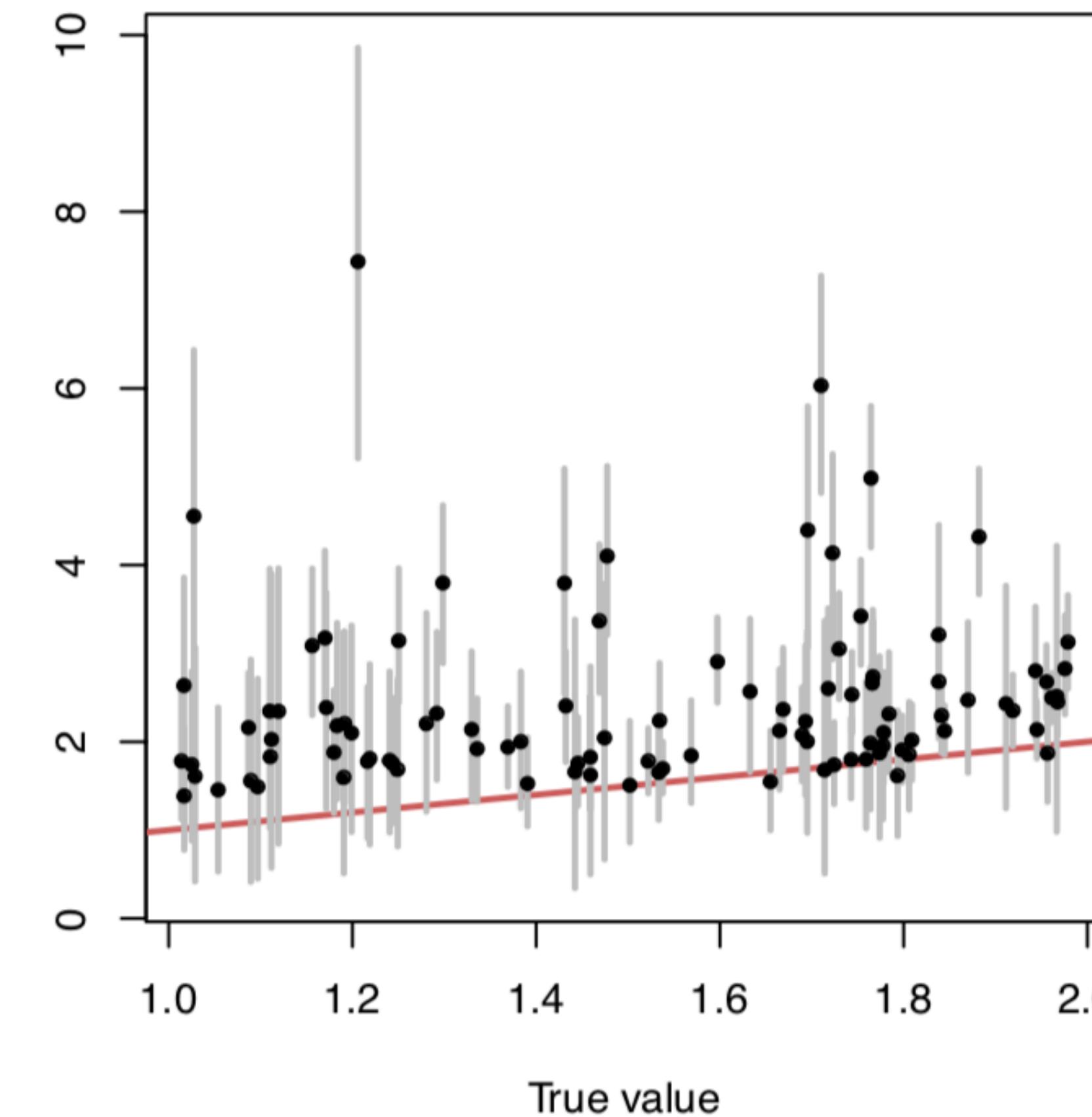
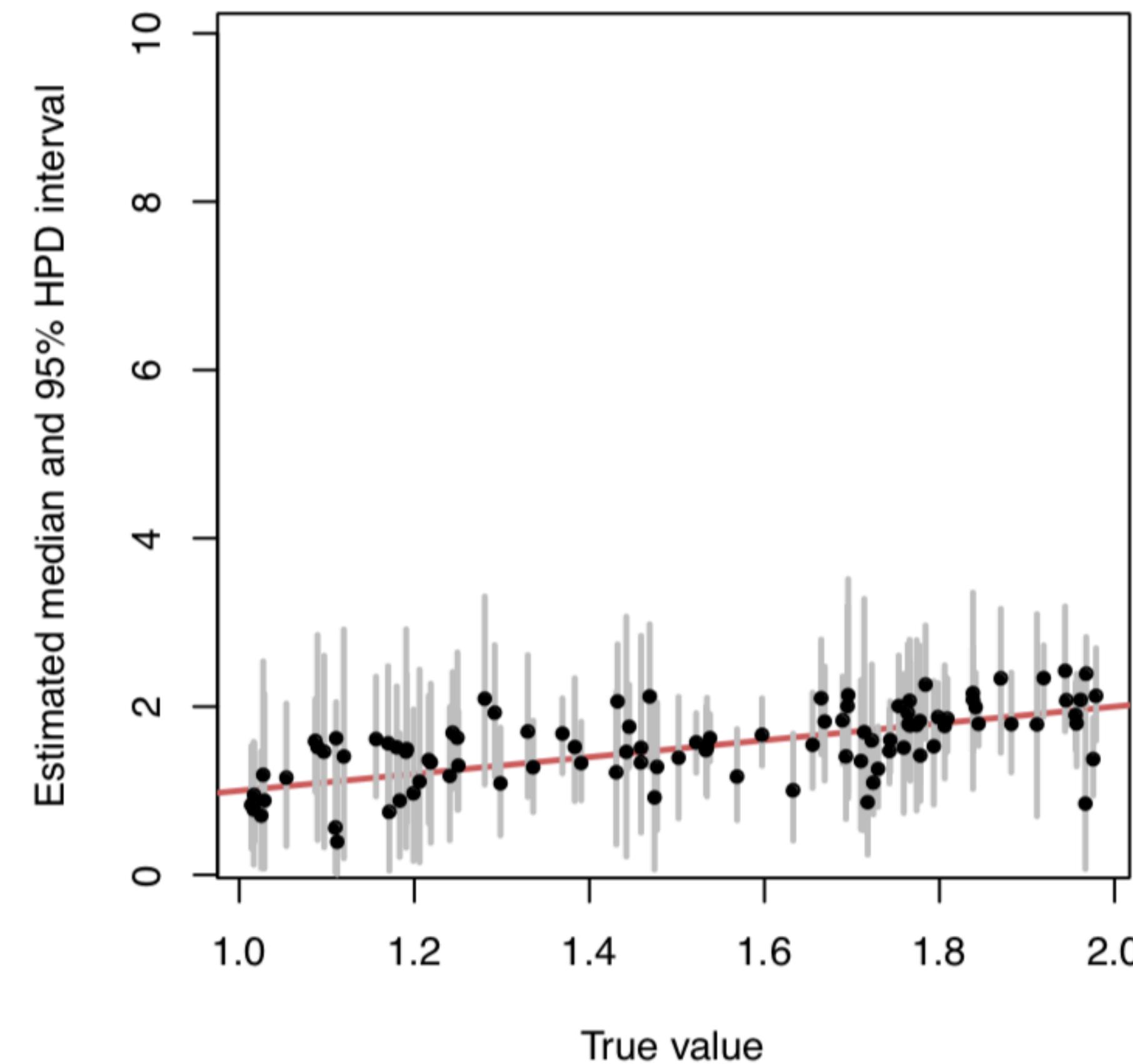
# Sampled ancestors

The proportion increases with higher turnover (birth - death) or higher sampling

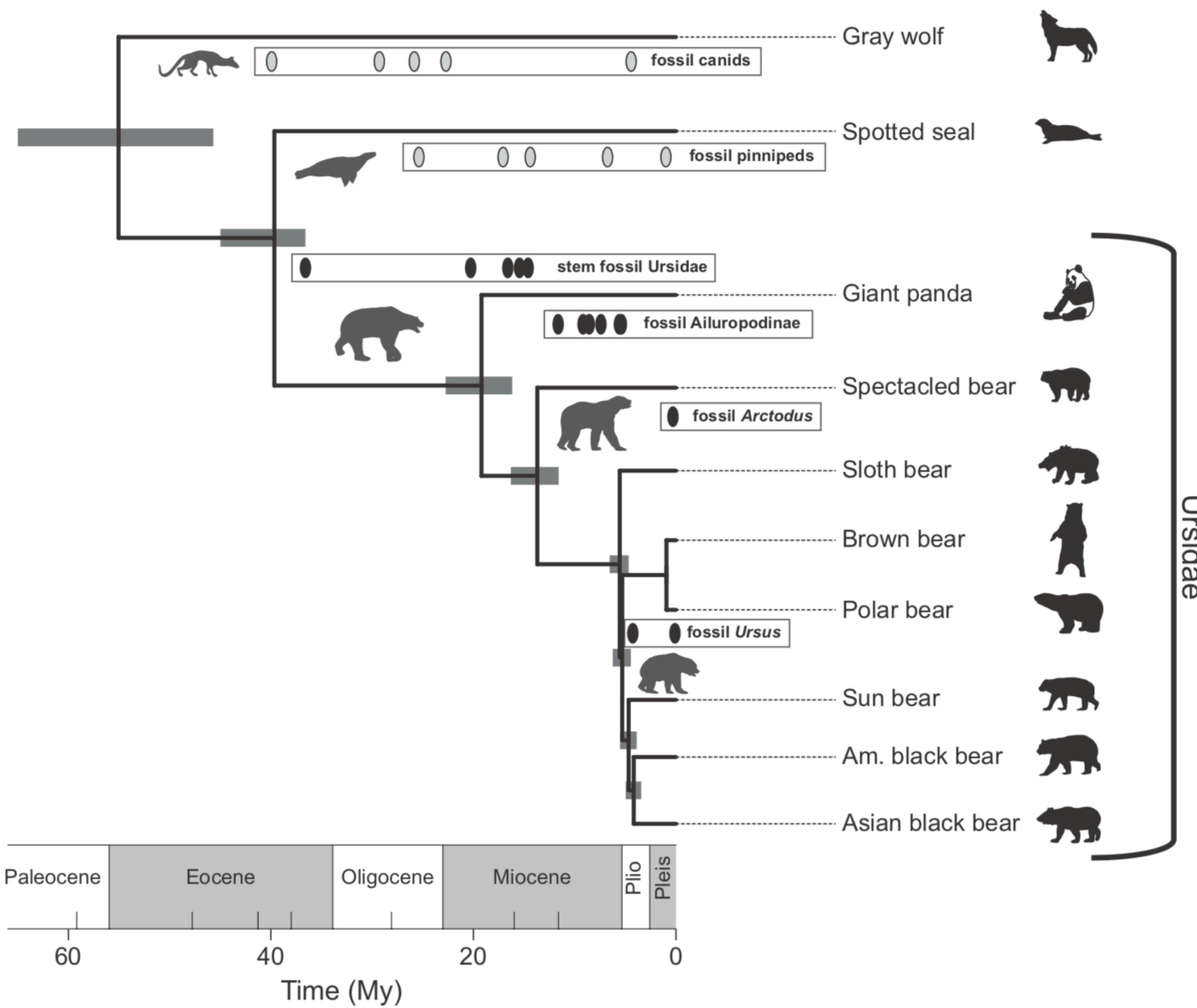


# Sampled ancestors

Ignoring sampled ancestors can lead to inaccurate parameter estimates



# Time calibrated tree of living and fossil bears



First application of the FBD model.

Fossils are incorporated via constraints, not character data. Their precise placement can not be inferred, but this uncertainty will be reflected in the posterior

# Exercise

# Fossils can be incorporated via **taxonomy** or **character data** (total-evidence)

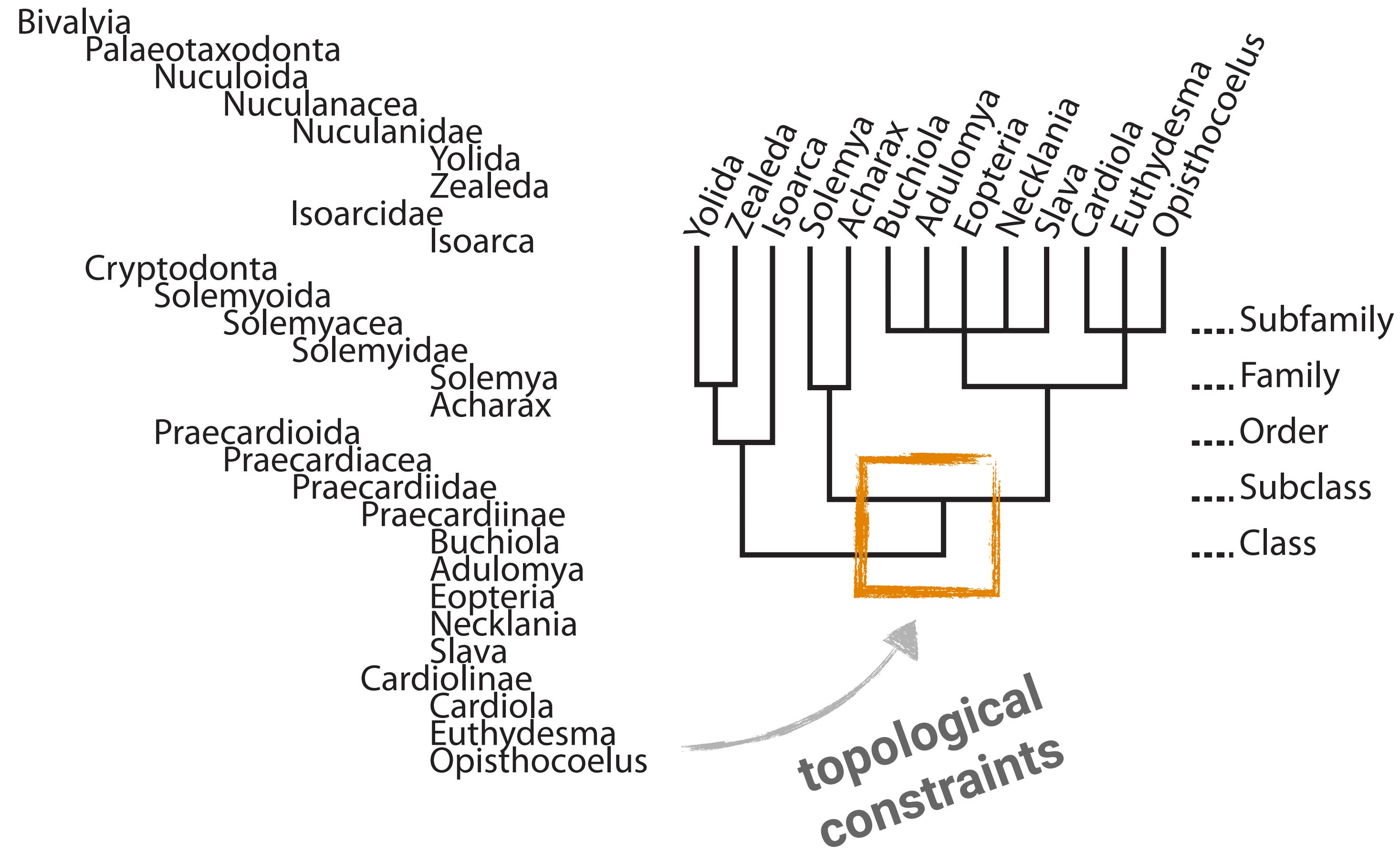
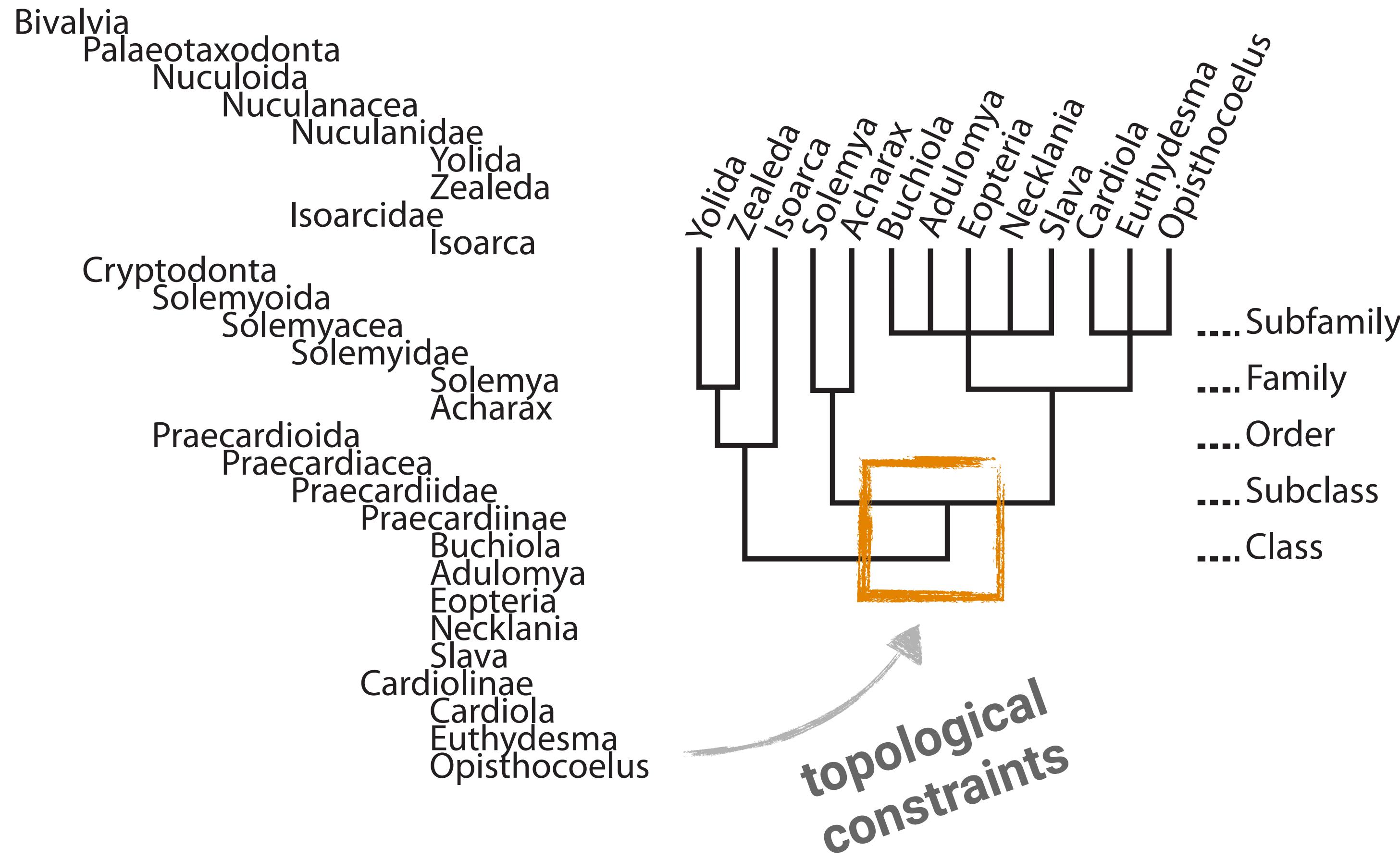


Image source Soul & Friedman (2015)

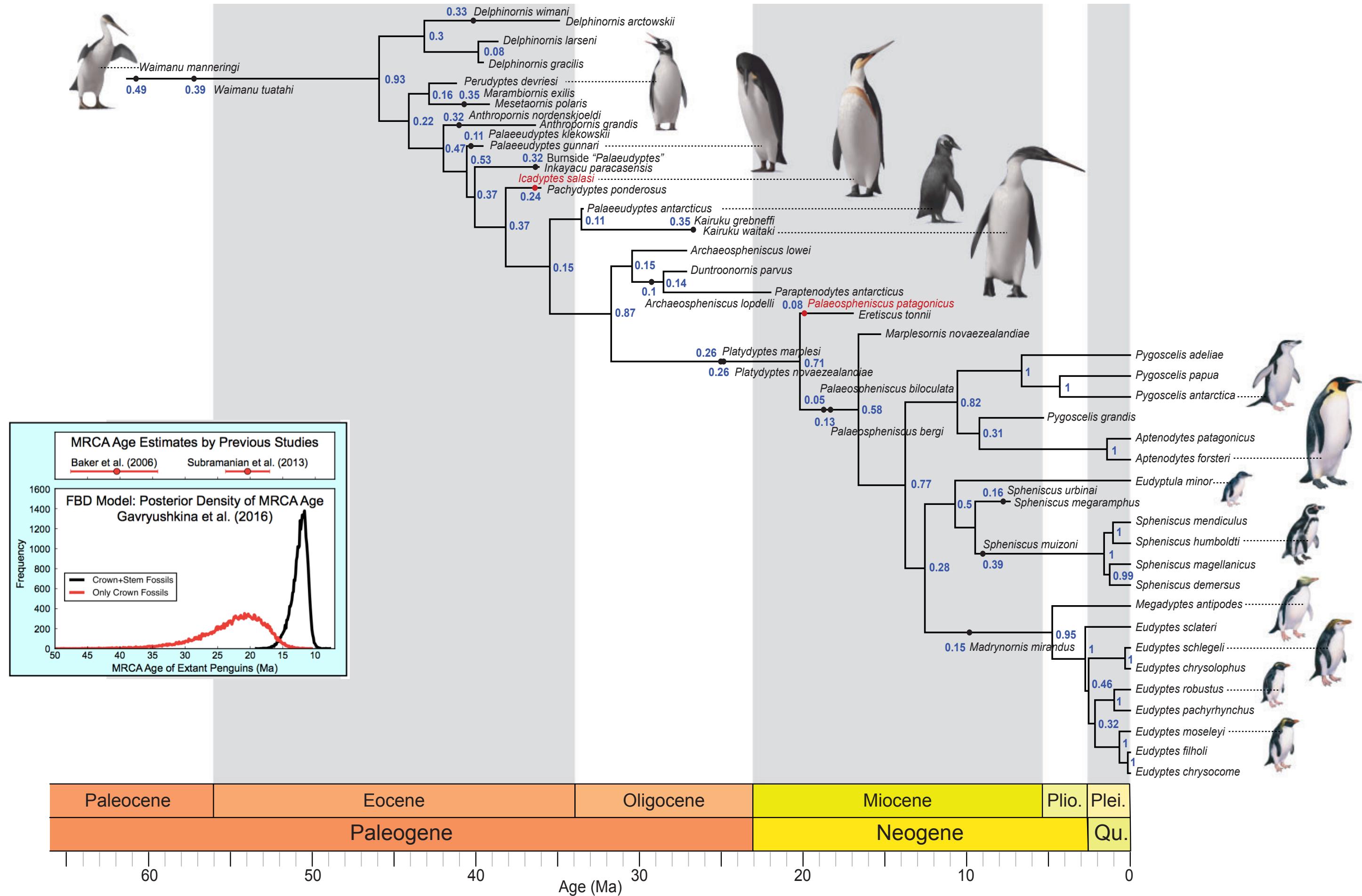
# Fossils can be incorporated via **taxonomy** or **character data** (total-evidence)



**ATAT...**  
**TCACT...**  
**?????...**  
**1001...**  
**1101...**  
**0100...**

Image source Soul & Friedman (2015)

# Time calibrated tree of living and fossil penguins

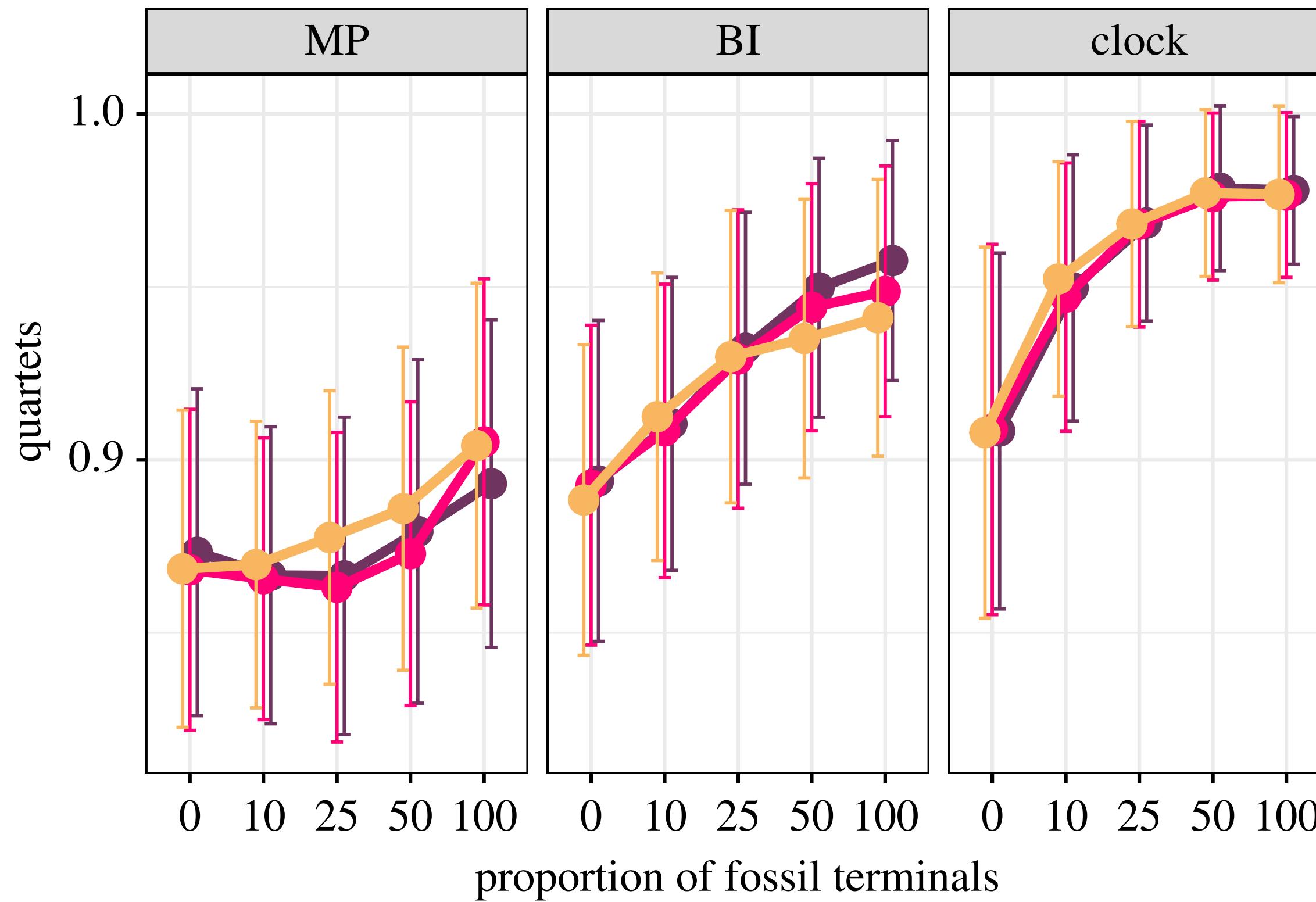


# First application of total evidence dating using the FBD model

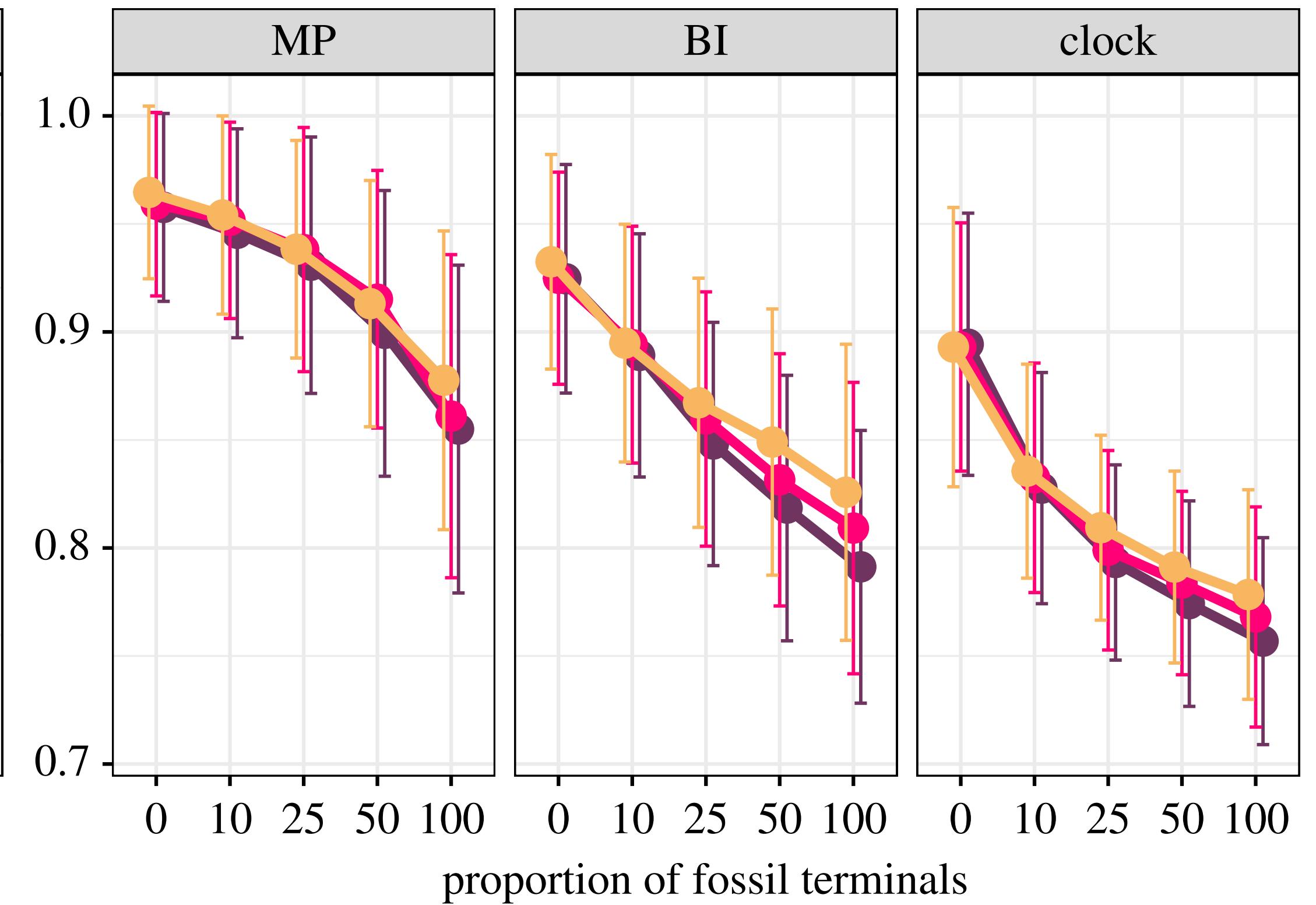
Fossils are incorporated  
using character data

Gavryushkina et al. (2016)

# accuracy



# precision



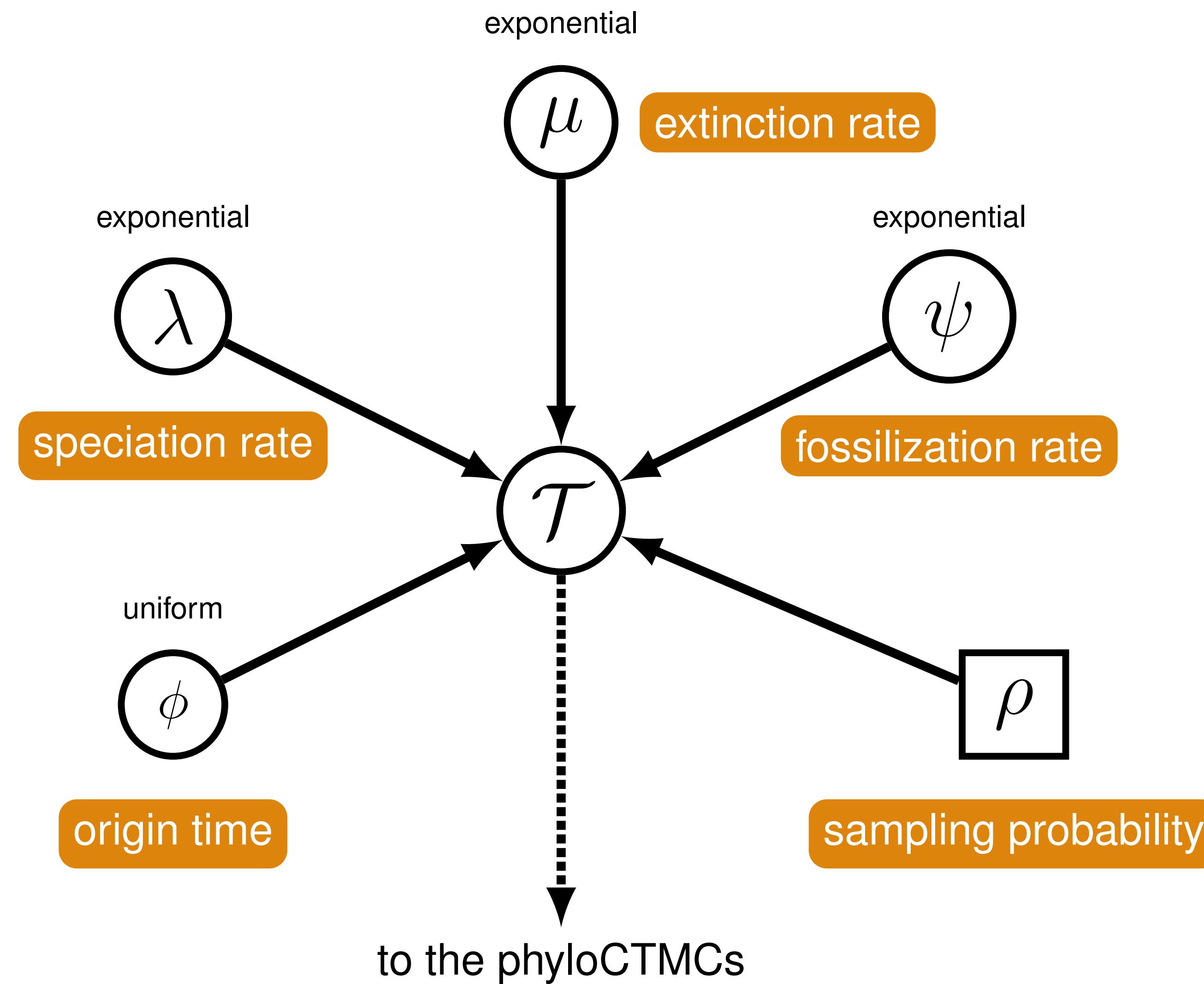
level of missing data      ● none    ● low    ● high

*Fossils improve phylogenetic analyses of morphological characters*  
Koch, Garwood, Parry. 2020. Proc B

# Some notes

- The topology of extant taxa is largely unaffected by how fossils are incorporated
- Fossils *and* age information help inform topology
- Divergence times are much more sensitive to errors in fossil placement and model misspecification
- Total-evidence dating is more robust to model misspecification

# Graphical model representation of the FBDP



# Exercise

# Phylogenetics

Diversification rate estimation

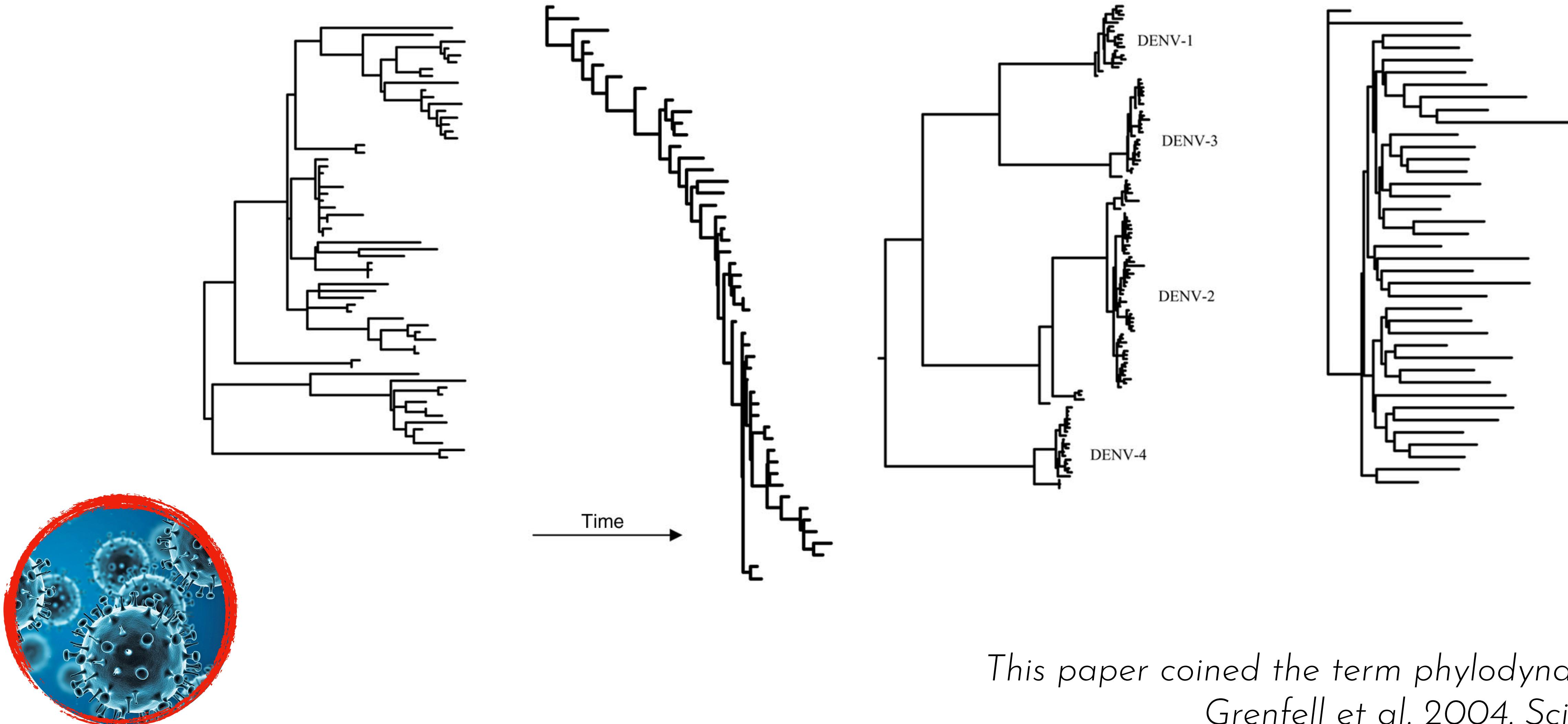
# Bayesian divergence time estimation

$$P(E \mid \lambda, \mu, \psi, p, O, t \mid 0101\dots, 1101\dots, 0100\dots, \text{snail}) =$$

probability of the  
time tree

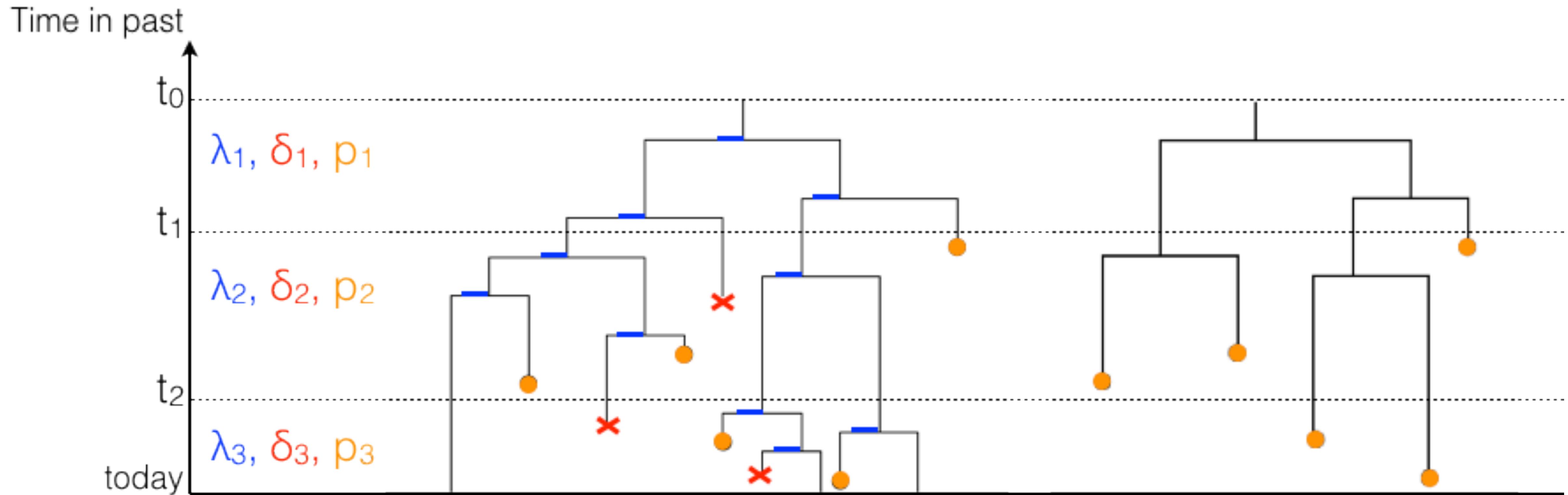
$$\frac{P(0101\dots, 1101\dots, 0100\dots \mid E) P(E \mid \lambda, \mu, \psi, p, O, t) P(\lambda, \mu, \psi, p) P(O) P(t)}{P(0101\dots, 1101\dots, 0100\dots, \text{snail})}$$

# Tree shape is informative about underlying dynamics

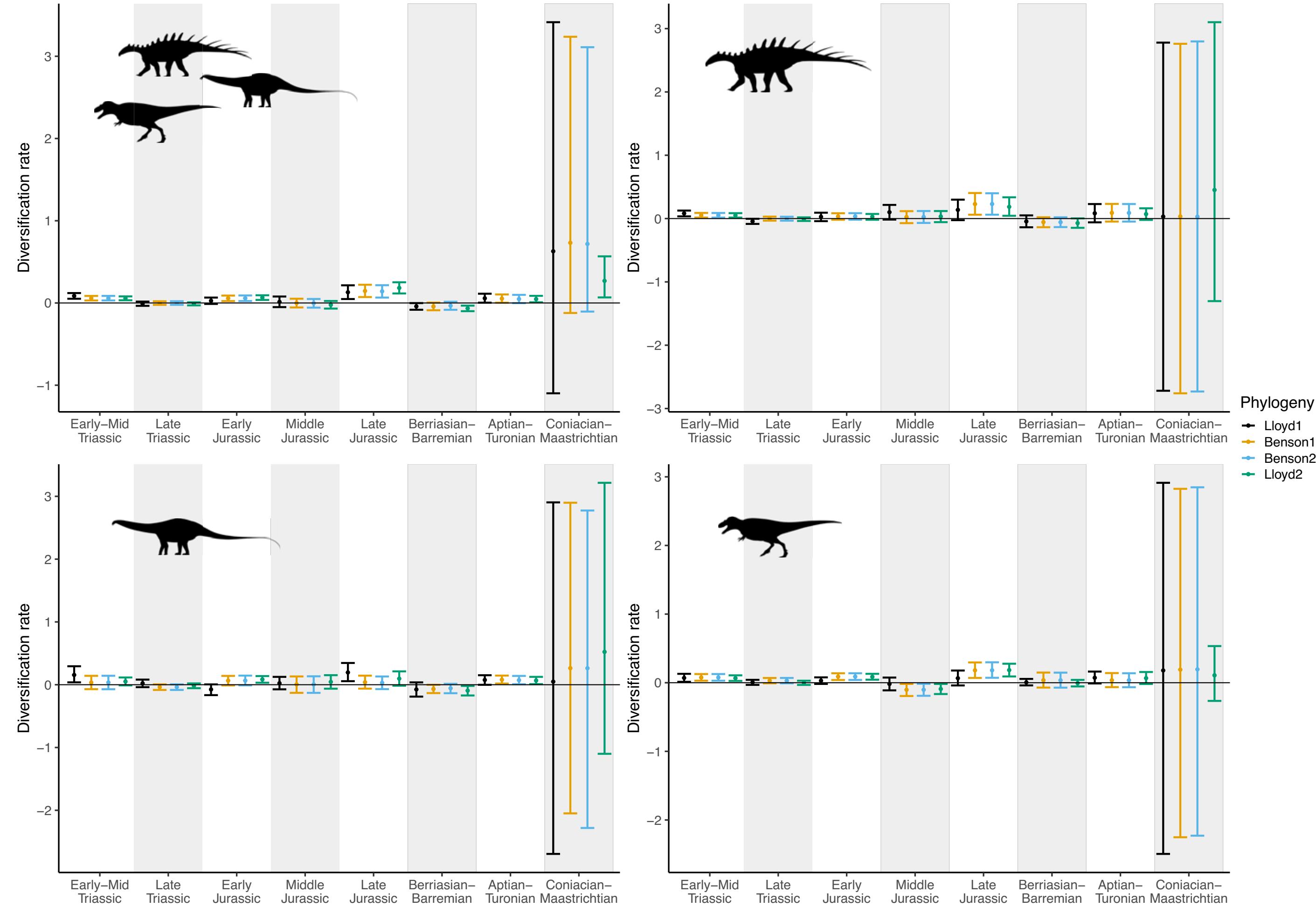


# The skyline birth-death process

First used for tracking the spread of infectious diseases



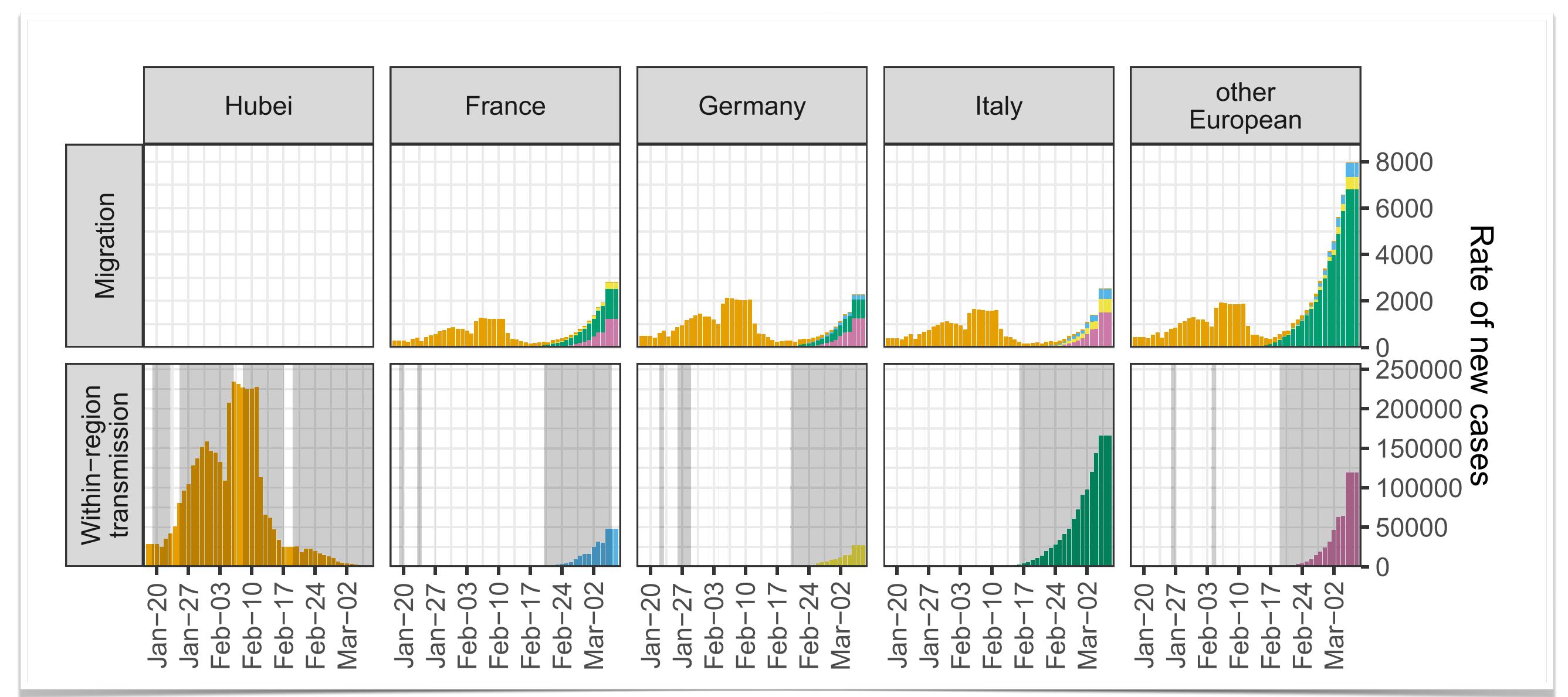
# Macroevolutionary case study



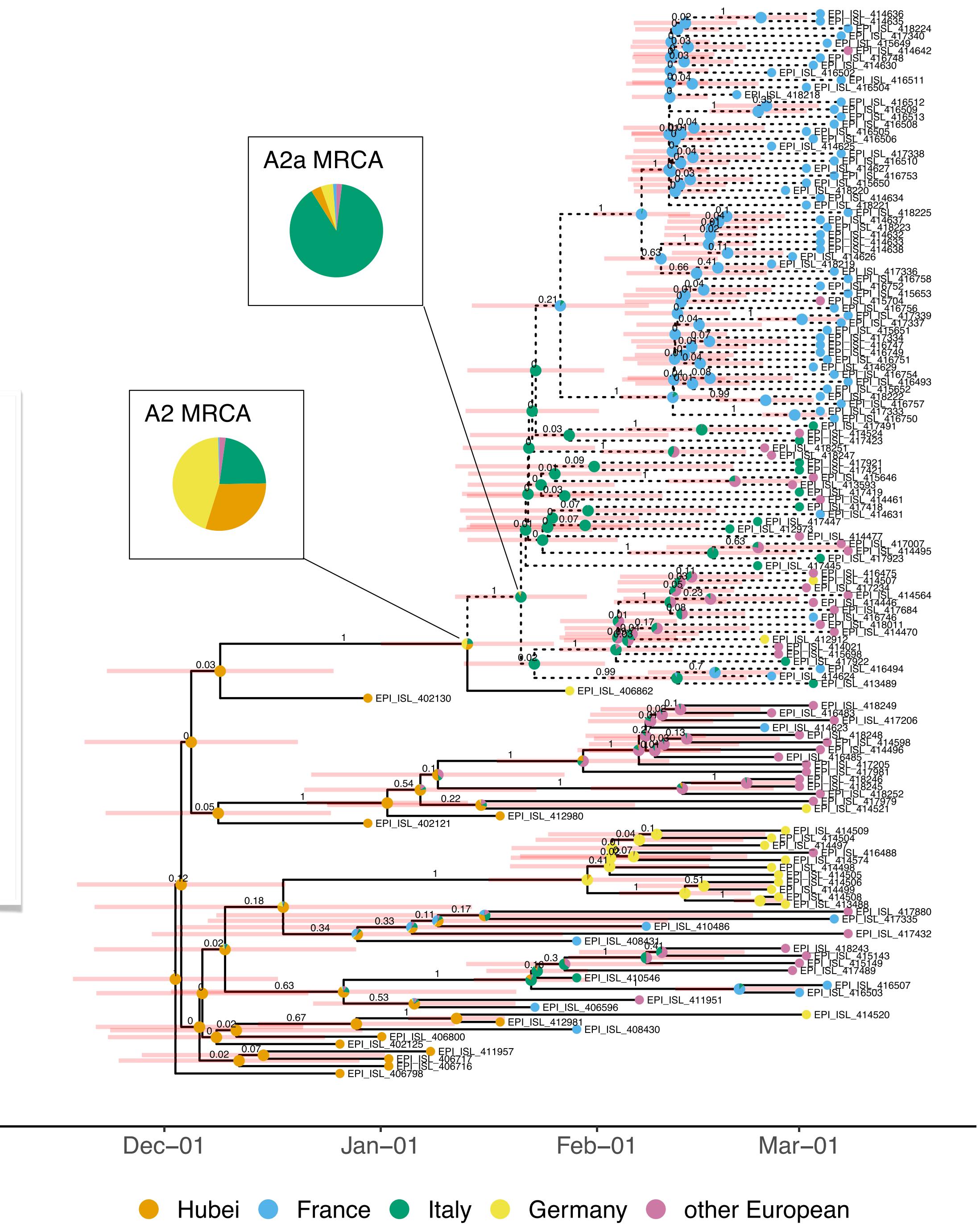
Phylogenies have been used to argue dinosaurs were incline prior to the KPg

FBD analyses suggest that we can not currently answer that question using phylogenies

# Models that include migration



The origin and early spread of SARS-CoV-2 in Europe  
Nadeau et al. 2021. PNAS



# Bayesian divergence time estimation

## The data

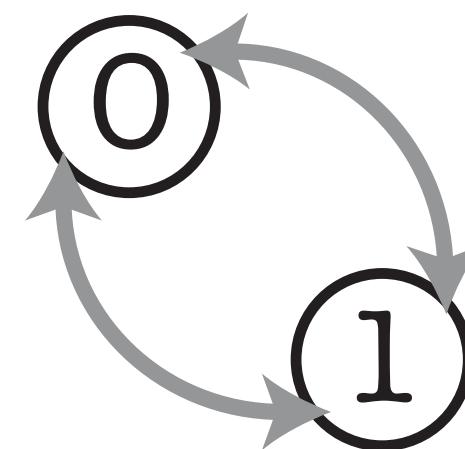
and / or  
0101... ATTG...  
1101... TTGC...  
0100... ATTC...



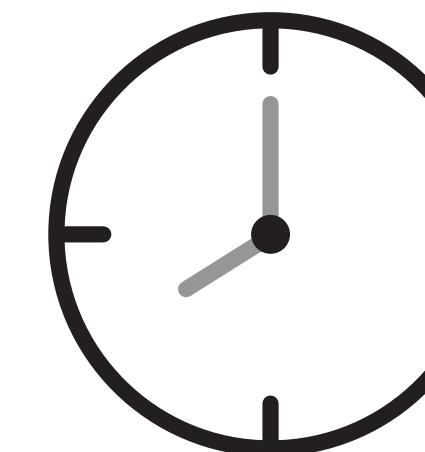
phylogenetics  
characters

sample  
ages

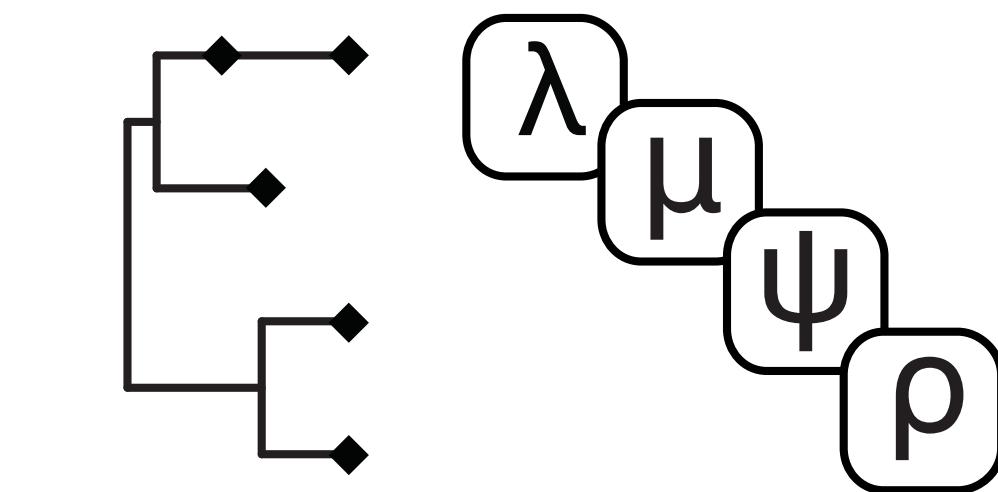
## 3 model components



substitution  
model



clock  
model



tree and tree  
model

# Using PCMs for dating

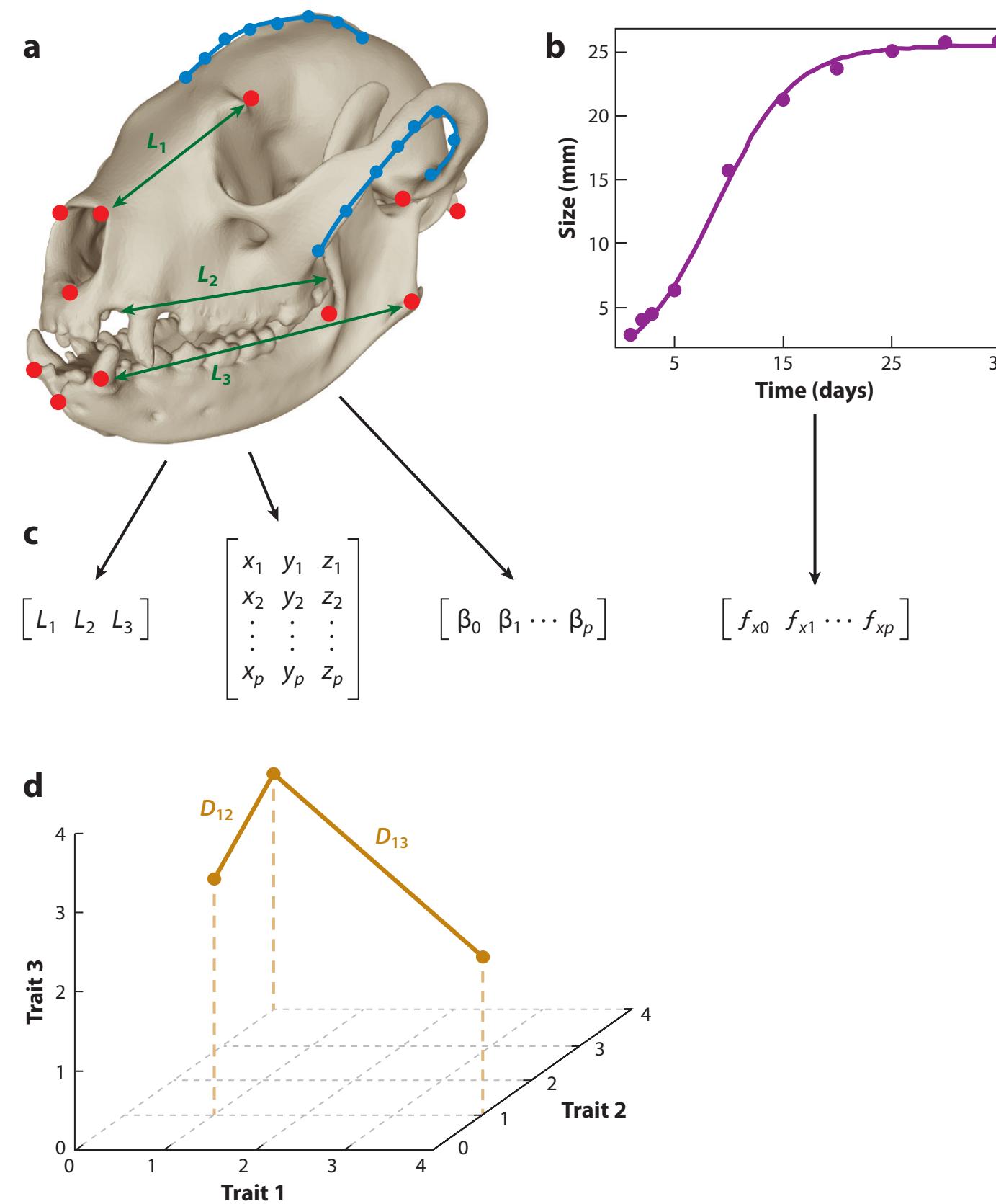
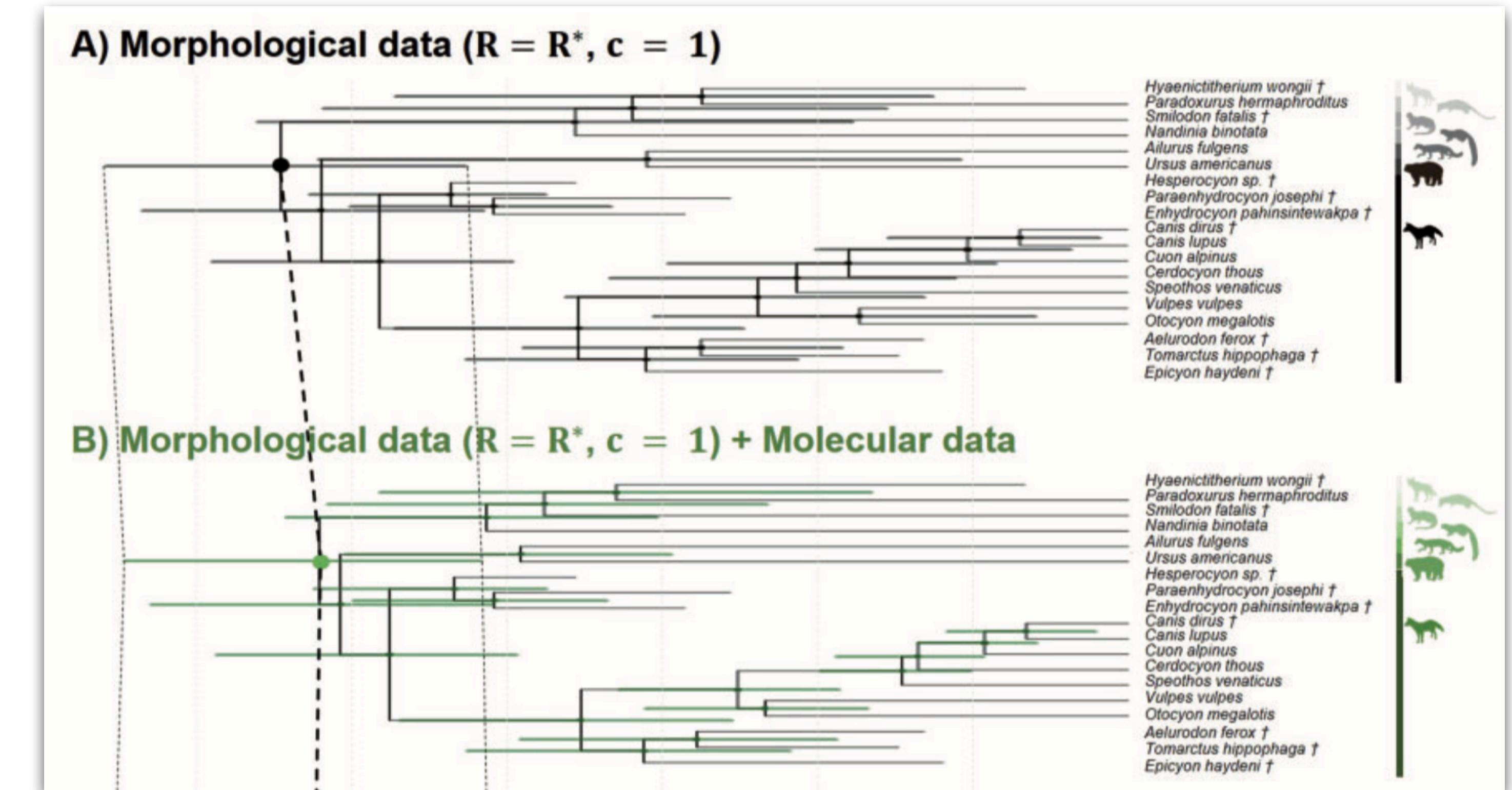
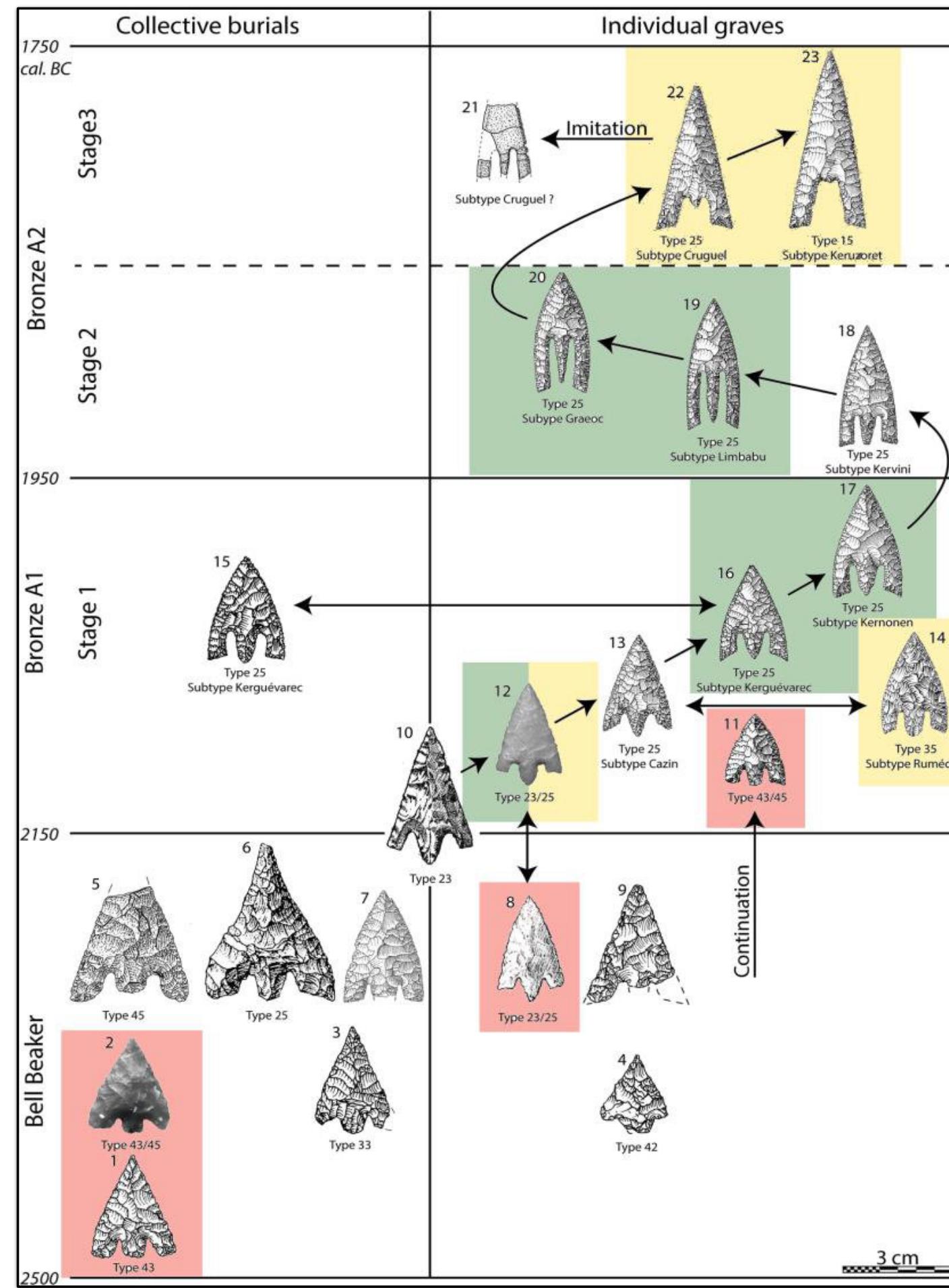


Image source Adams & Collyer ([2019](#))



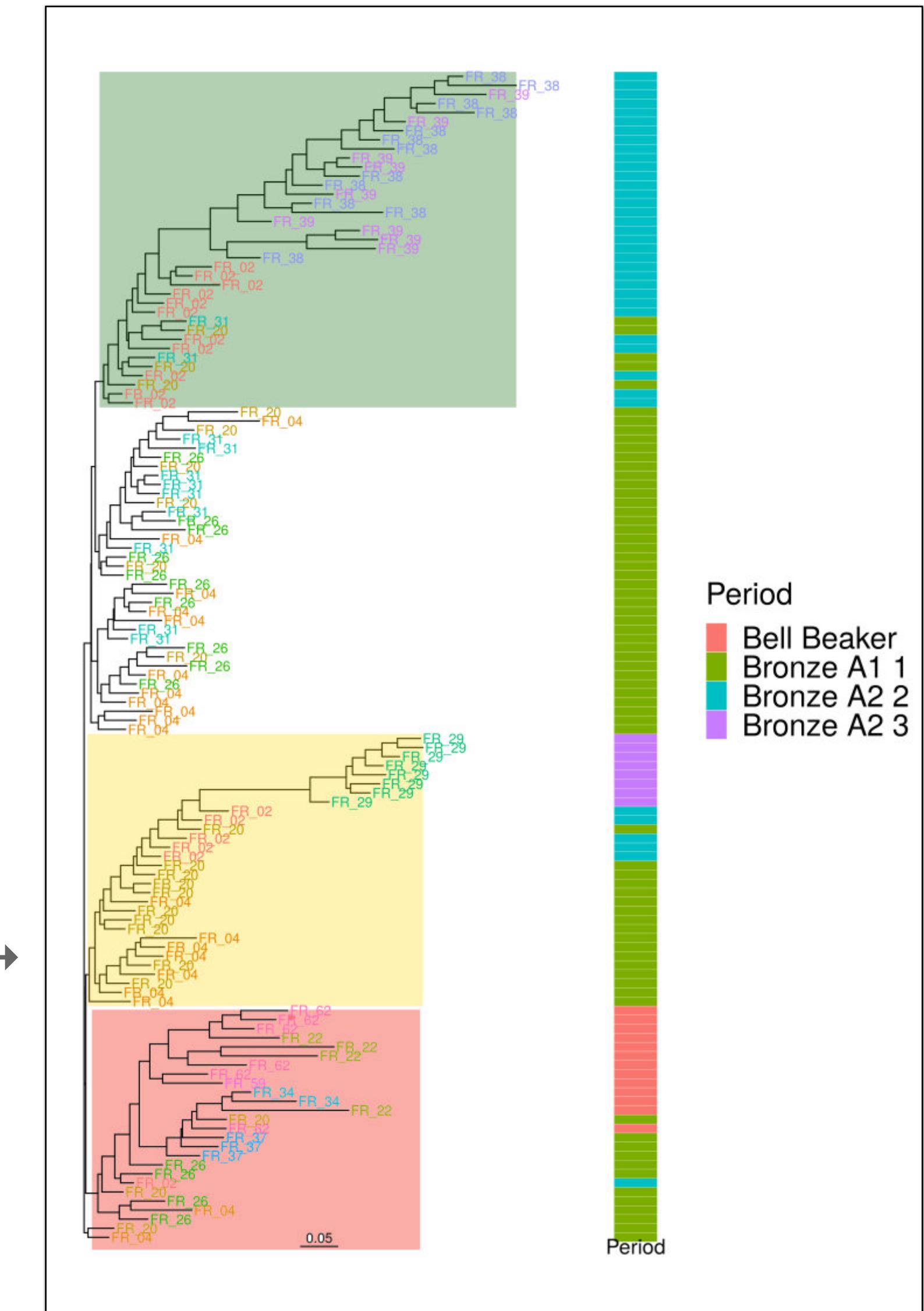
Álvarez-Carretero et al. ([2019](#)) Bayesian Estimation of Species Divergence Times Using Correlated Quantitative Characters

# Cultural evolution



# ← Typo-Chronology of Palaeolithic stone tools

# Outline based NJ tree –

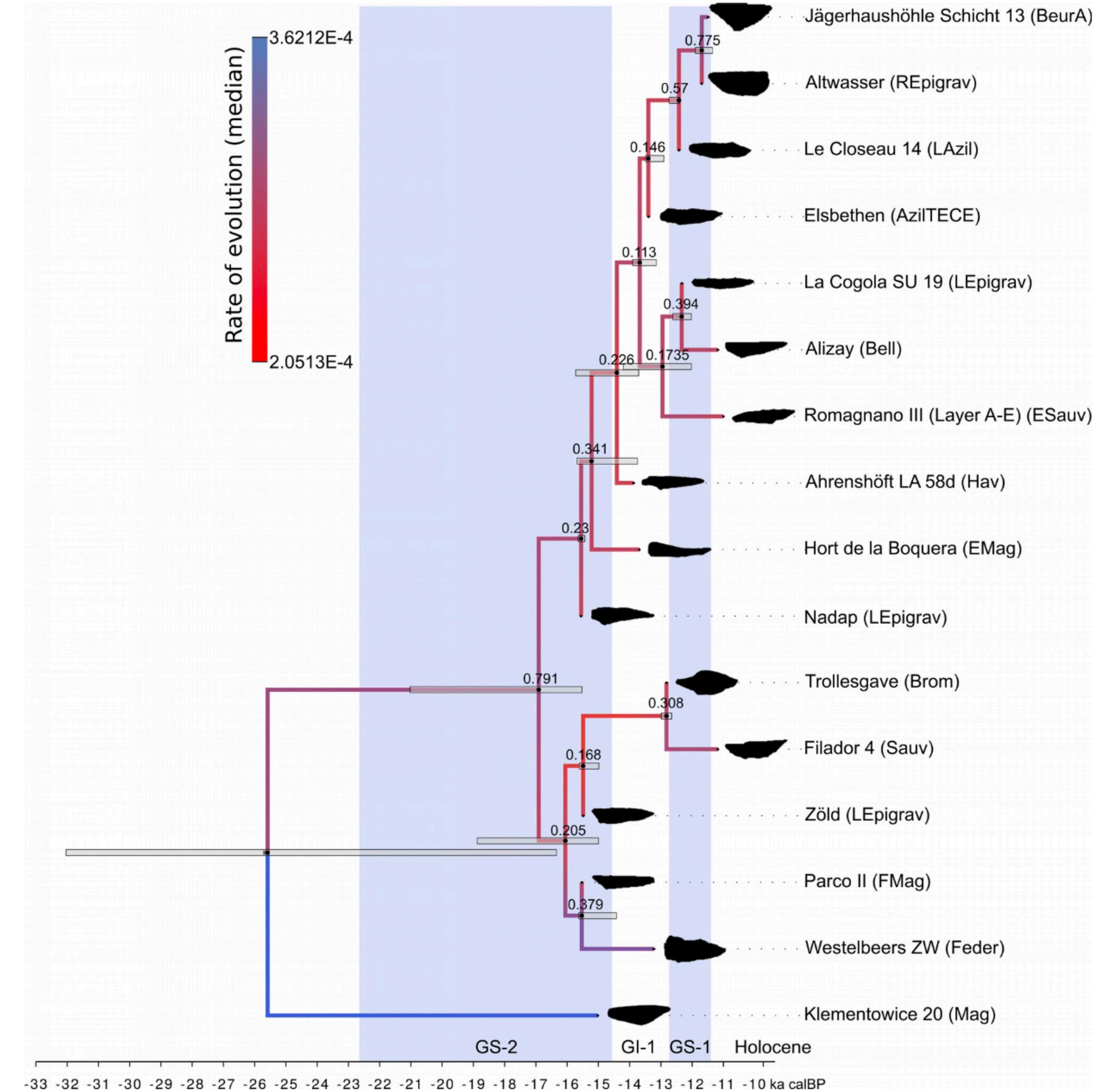


Matzig et al. 202

After Nicolas (2017)

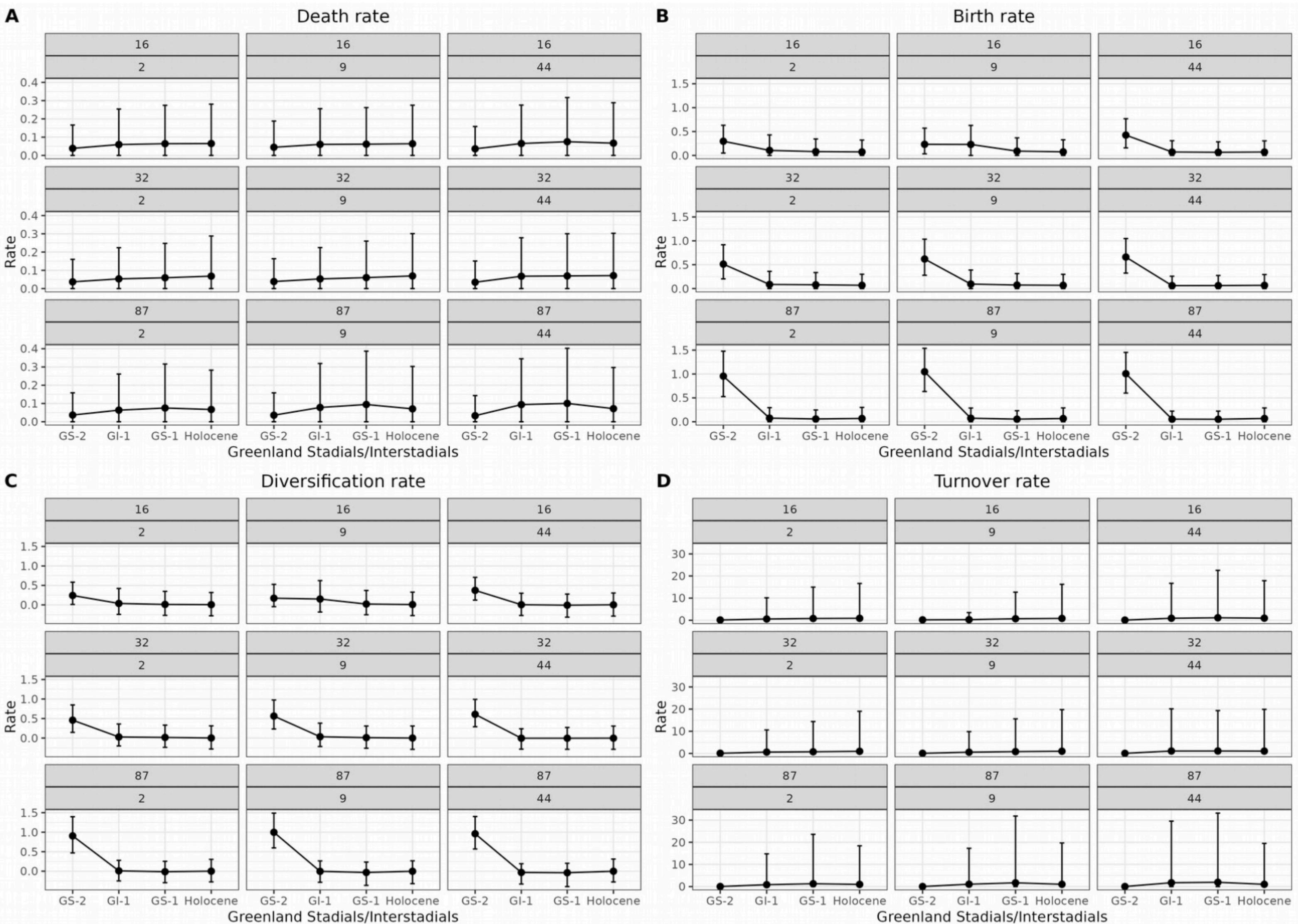
# The tree topology of stone tools exhibits a lot of uncertainty

Matzig et al. (in review) A macroevolutionary analysis of European Late Upper Palaeolithic stone tool shape using a Bayesian phylodynamic framework (preprint available)



# Sensitivity analyses

Birth, death, and sampling rates are impacted by trait and taxon sampling



# Joint phylogenetic estimation of geographic movements and biome shifts



# Global diversity of *Viburnum*

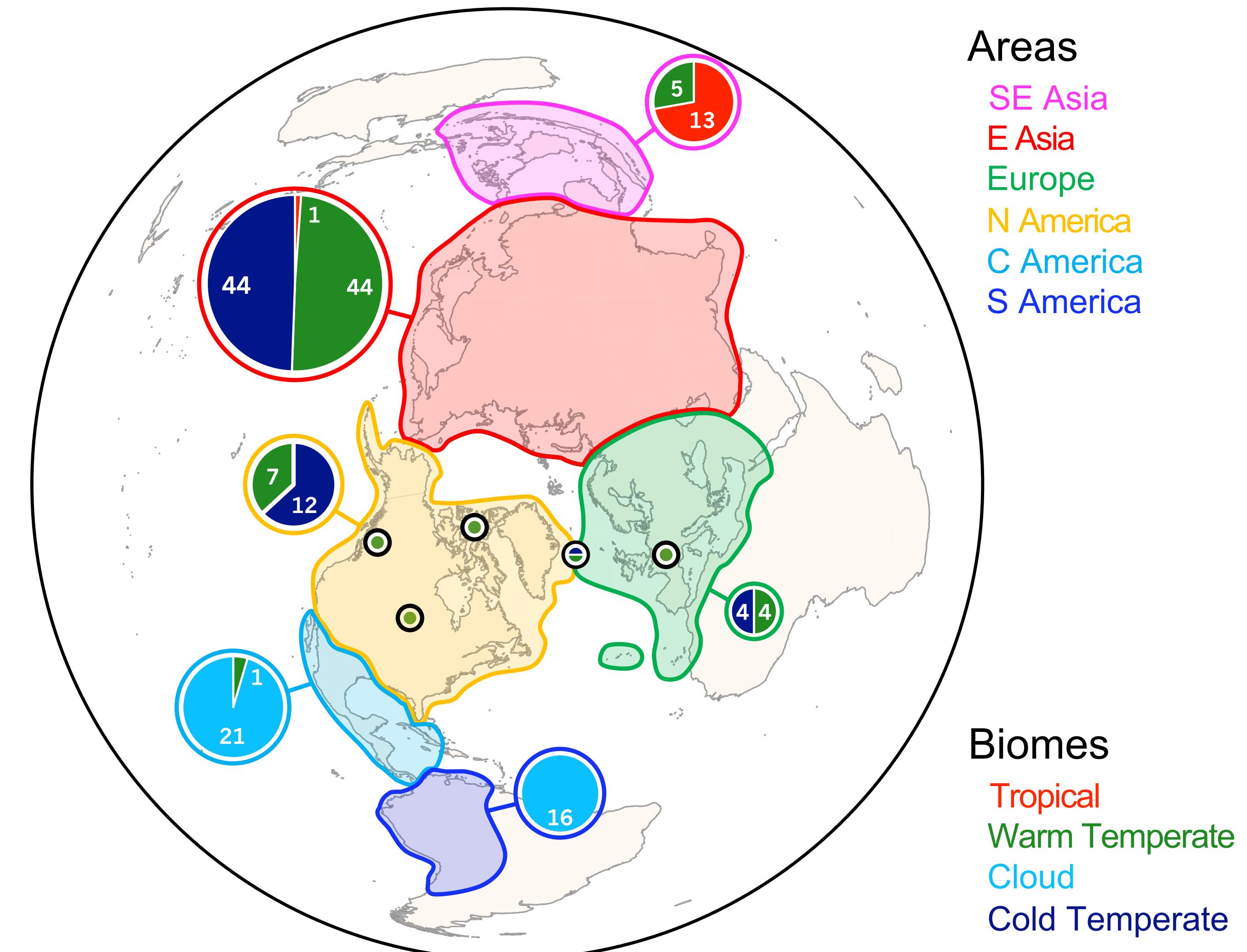
## Data

163 extant species (127 with DNA)

5 fossils (with taxonomic constraints)

6 geographic areas

4 biomes



Areas

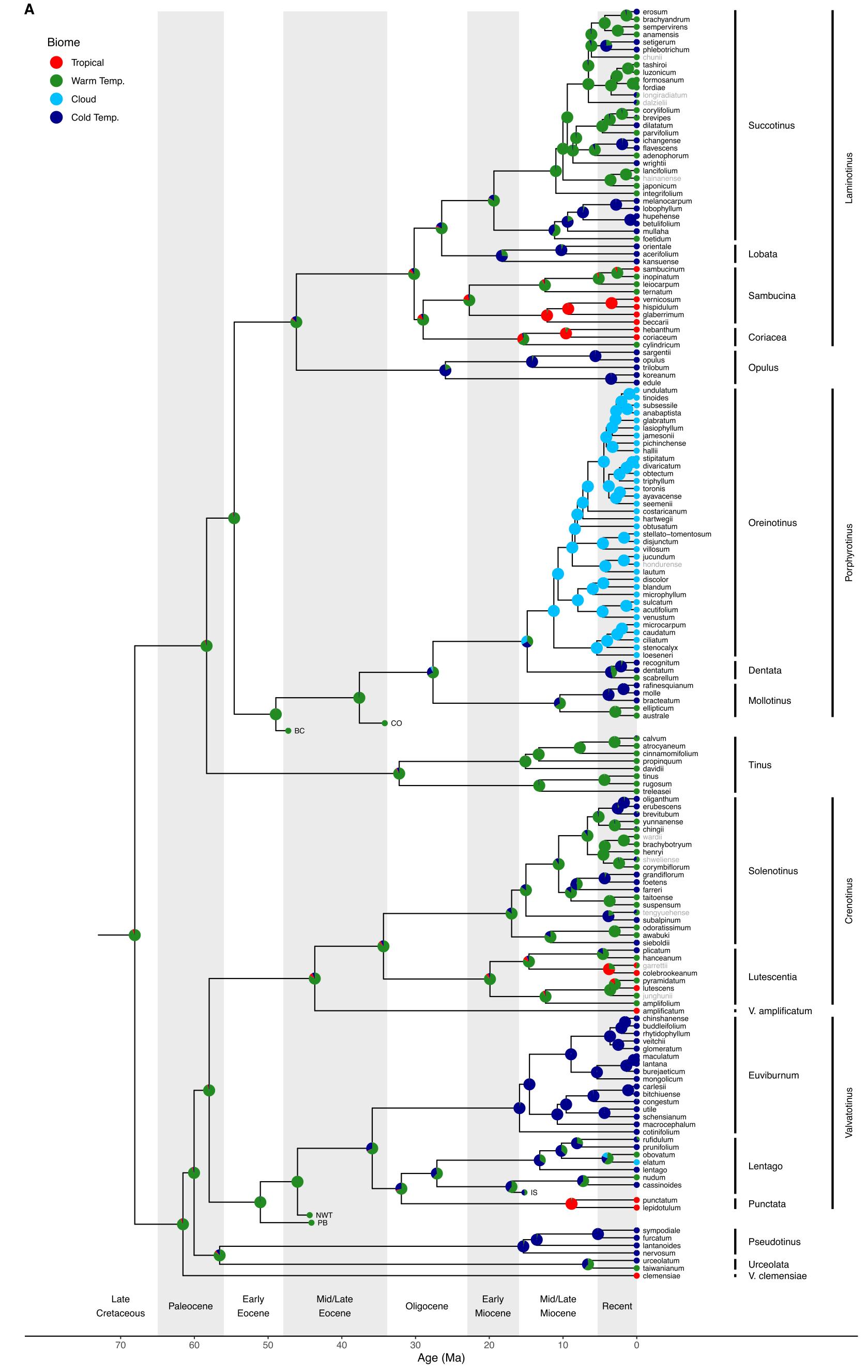
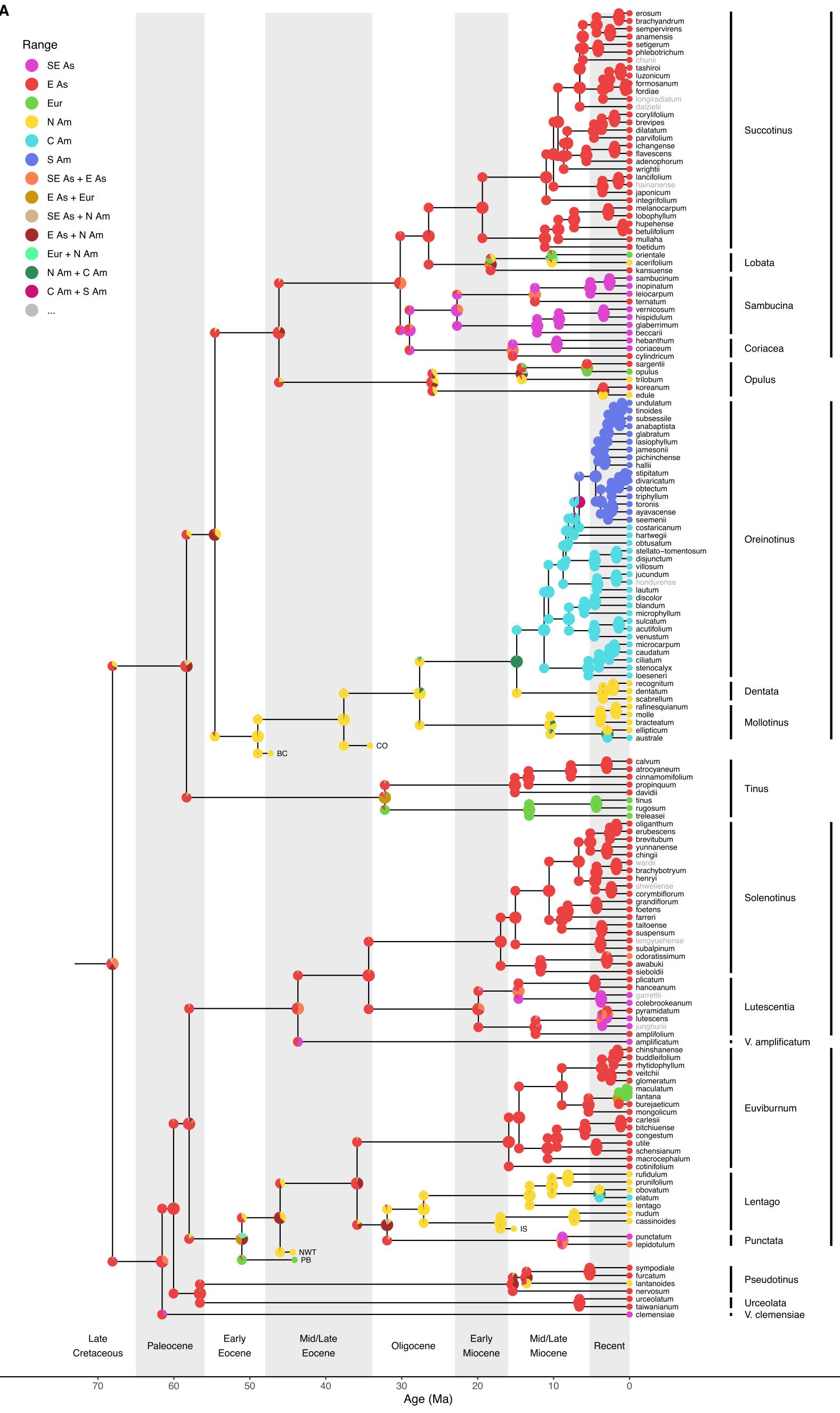
- SE Asia
- E Asia
- Europe
- N America
- C America
- S America

Biomes

- Tropical
- Warm Temperate
- Cloud
- Cold Temperate

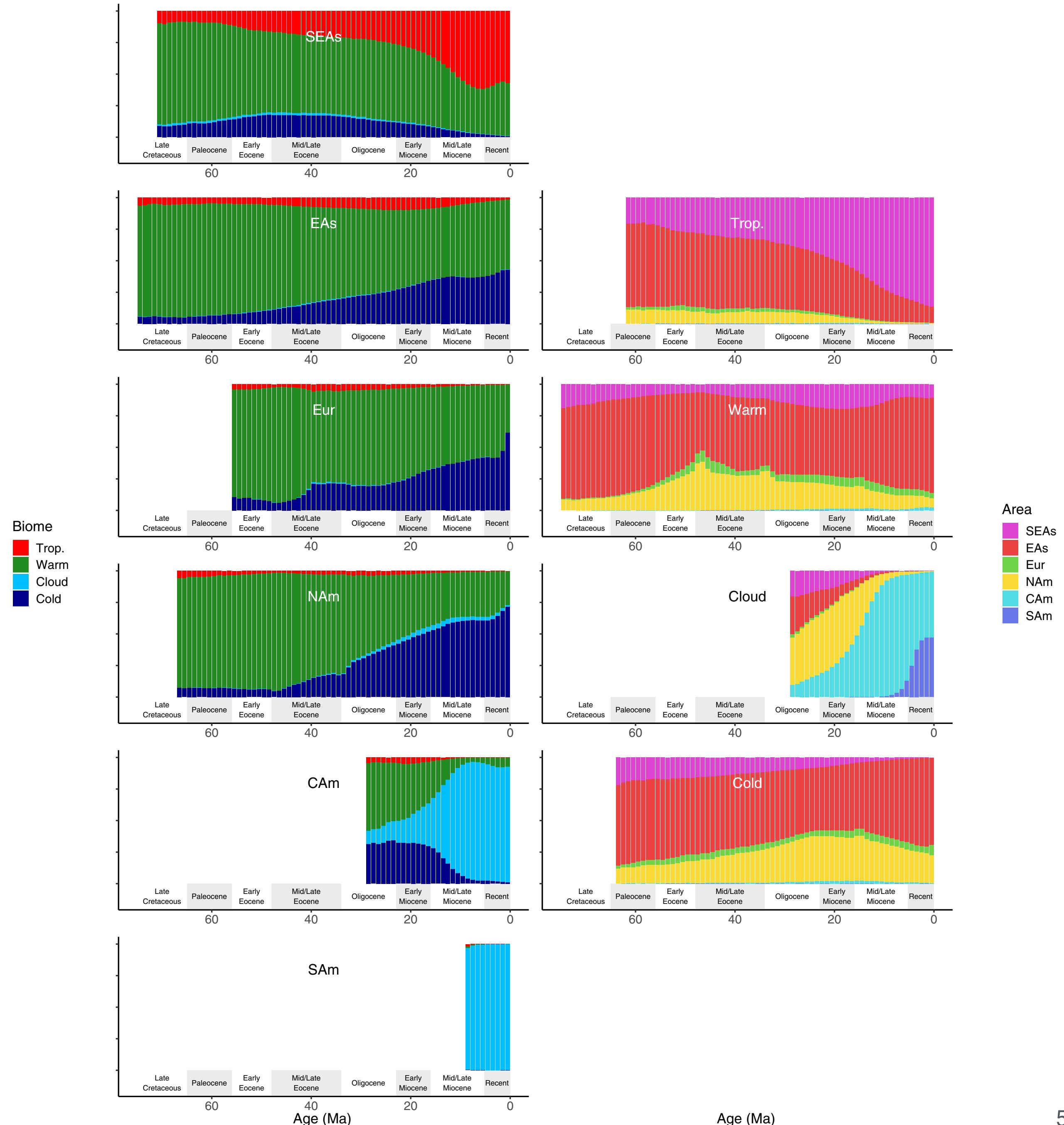
# Analysis

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



*“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.”*

Landis et al. (2023) *Systematic Biology*  
*Joint phylogenetic estimation of geographic movements and biome shifts*

No lecture next week but we can meet in  
the Old Library if you want to discuss  
projects ideas

# Exam registration