

Phylogenetics

Introduction to
phyldynamics models
RL-V3 MPP

Rachel Warnock

18.06.2024



What did you learn about PCMs?

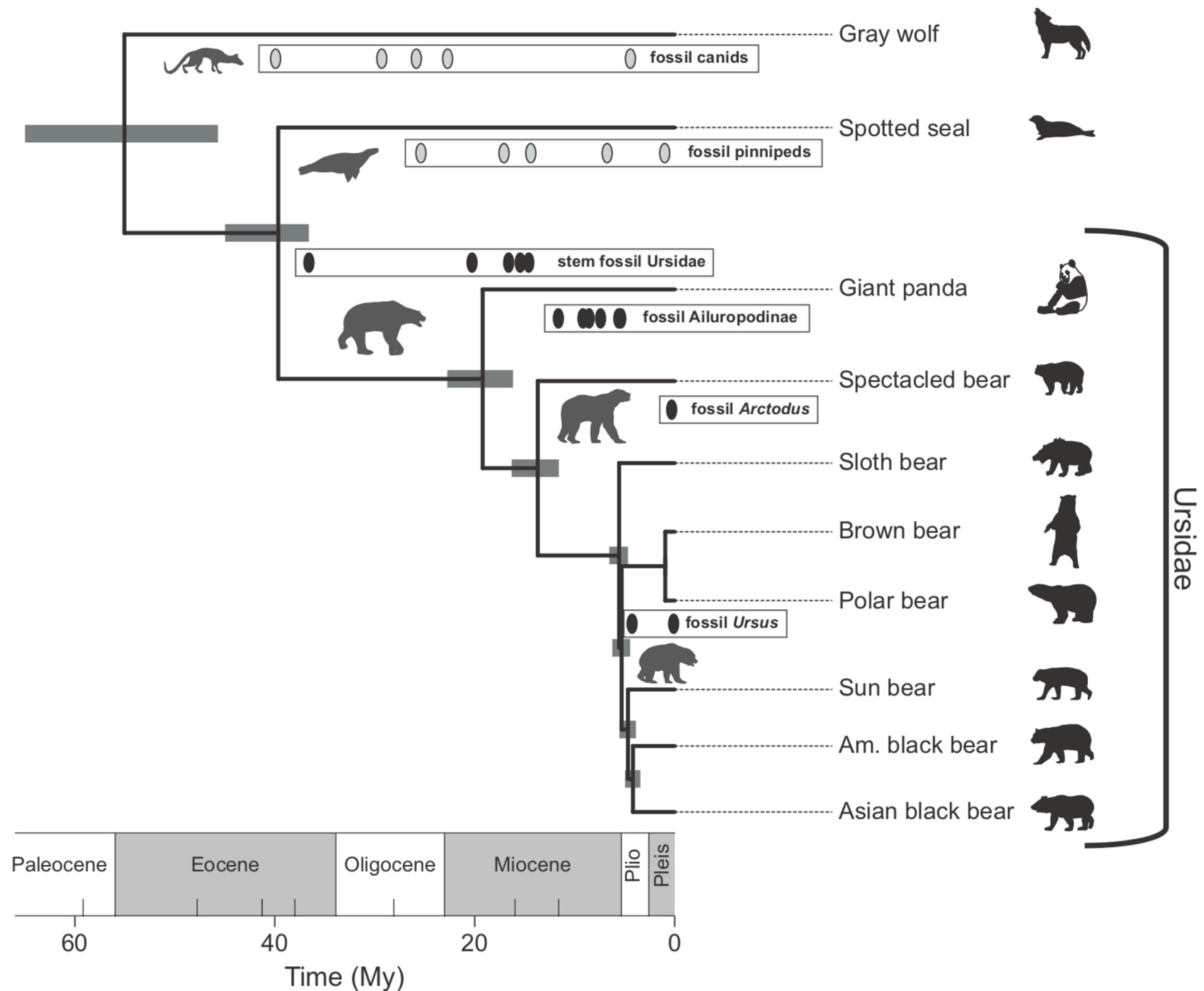
Emma's lecture

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
 - Phylodynamics



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

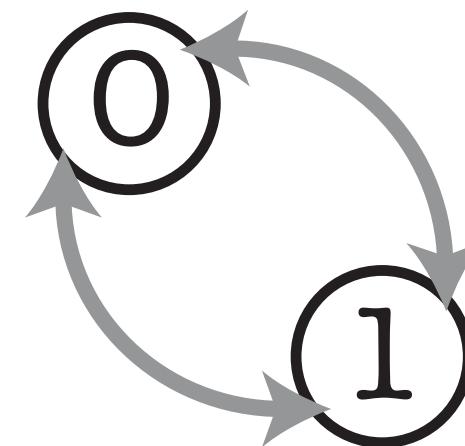
and / or
0101... ATTG...
1101... TTGC...
0100... ATTC...



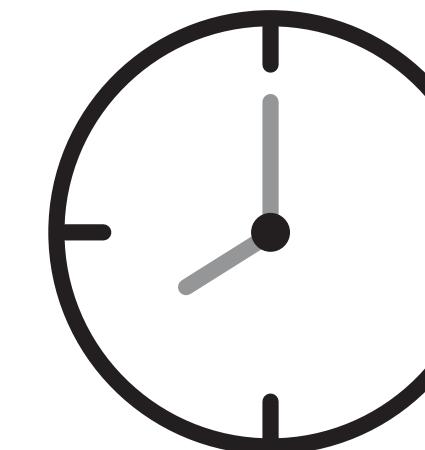
phylogenetics
characters

sample
ages

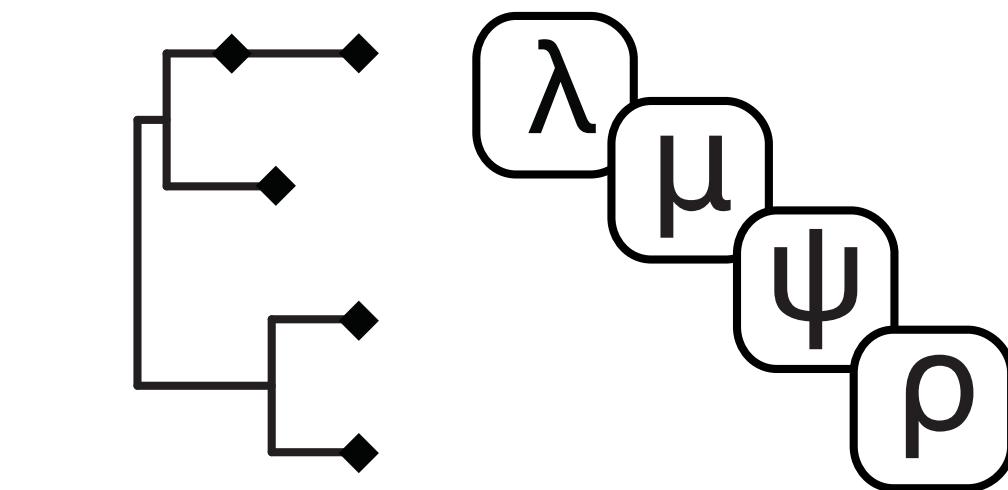
3 model components



substitution
model

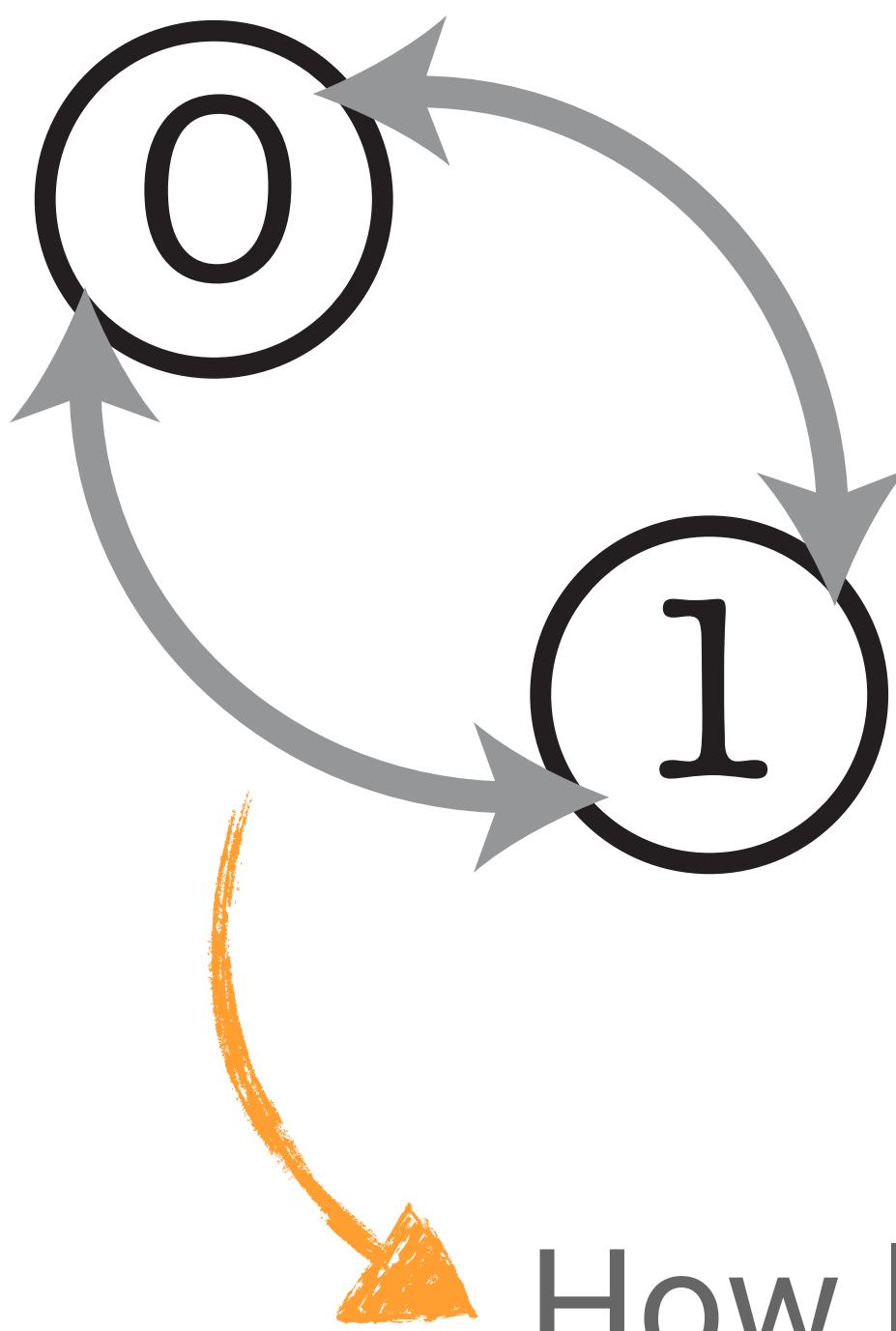


clock
model

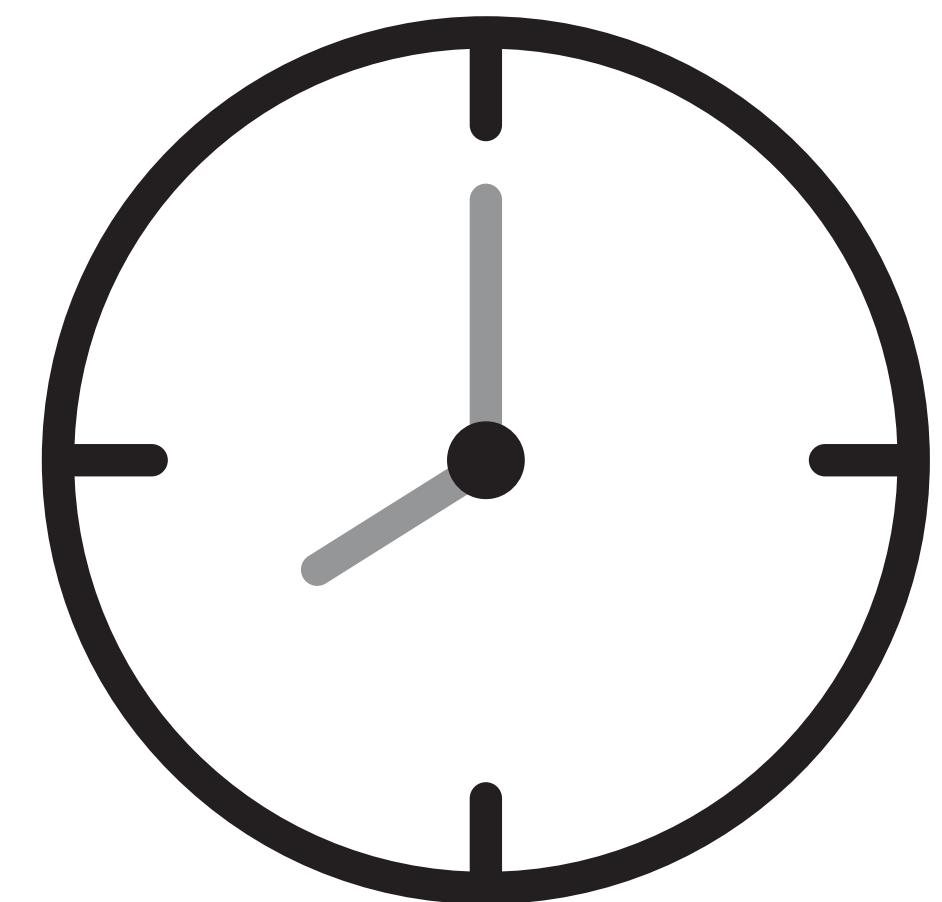


tree and tree
model

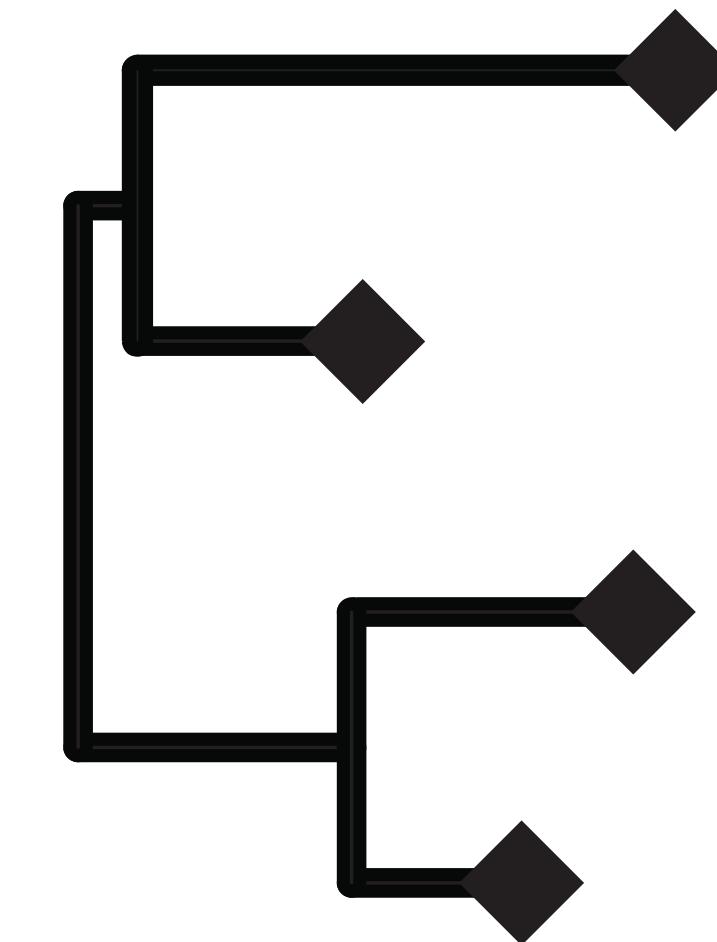
substitution model



clock model

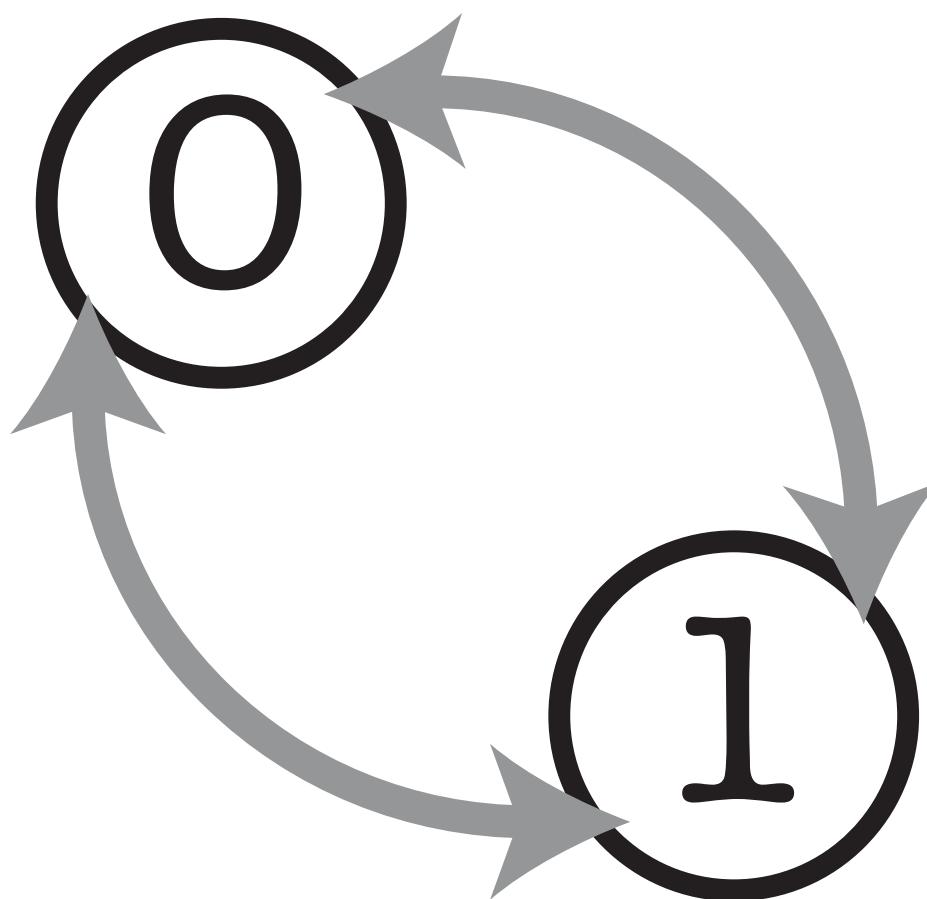


tree model

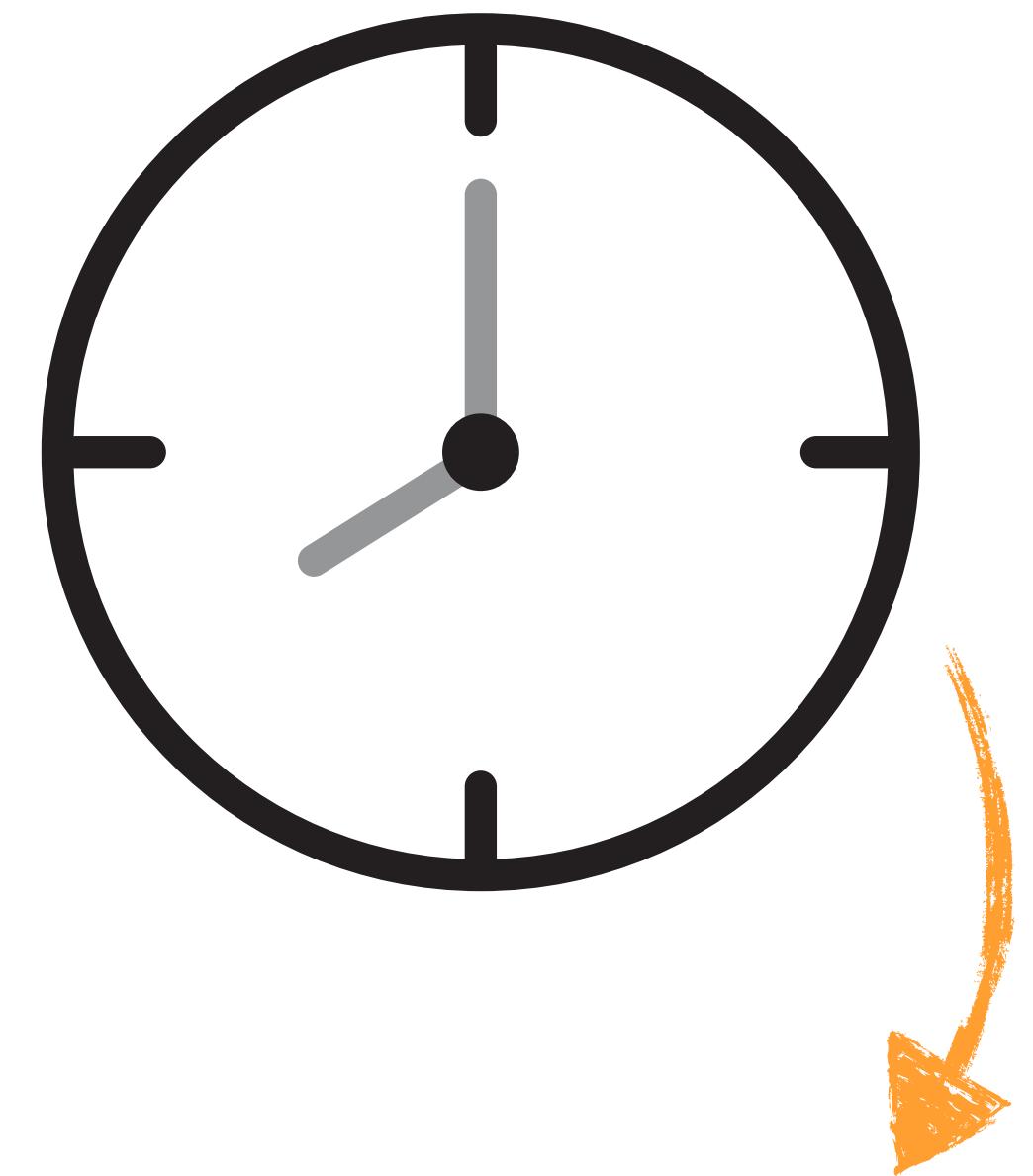


How likely are we to observe a change
between character states? e.g., $A \rightarrow T$

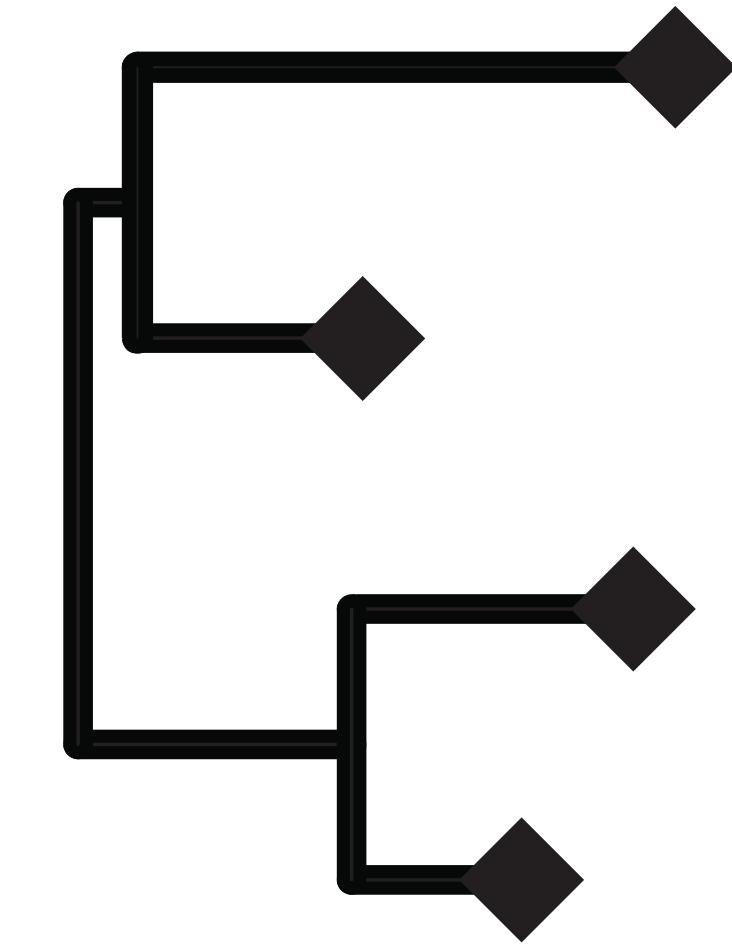
substitution model



clock model

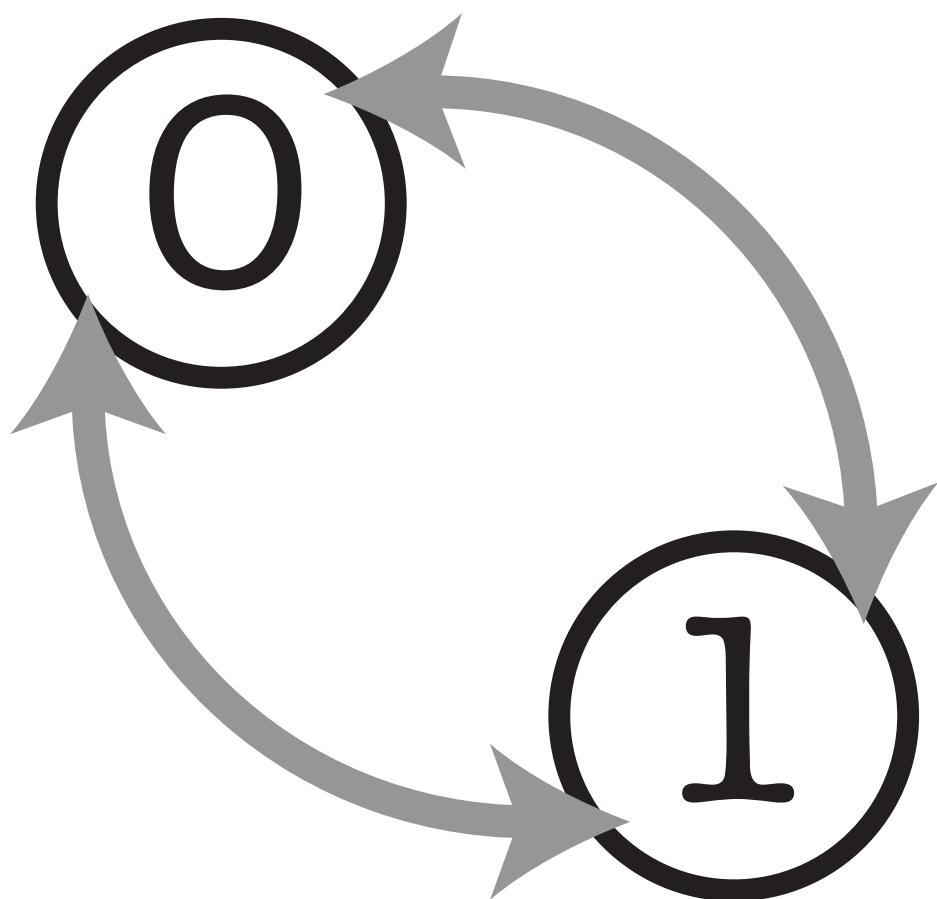


tree model

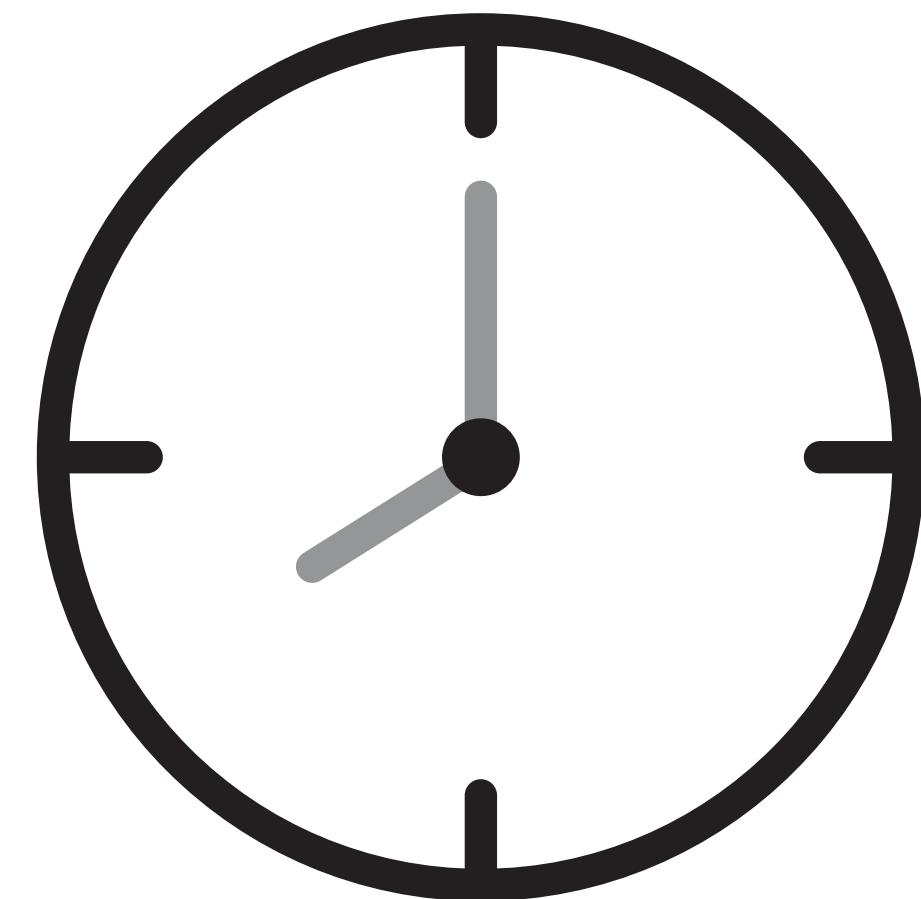


How have rates of evolution varied
(or not) across the tree?

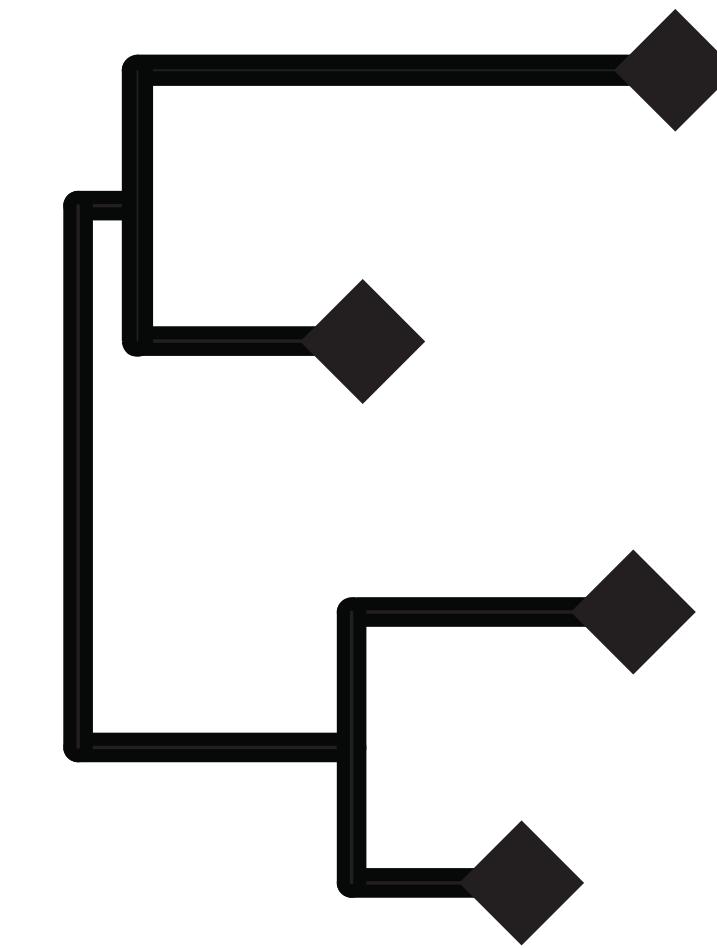
substitution model



clock model



tree model

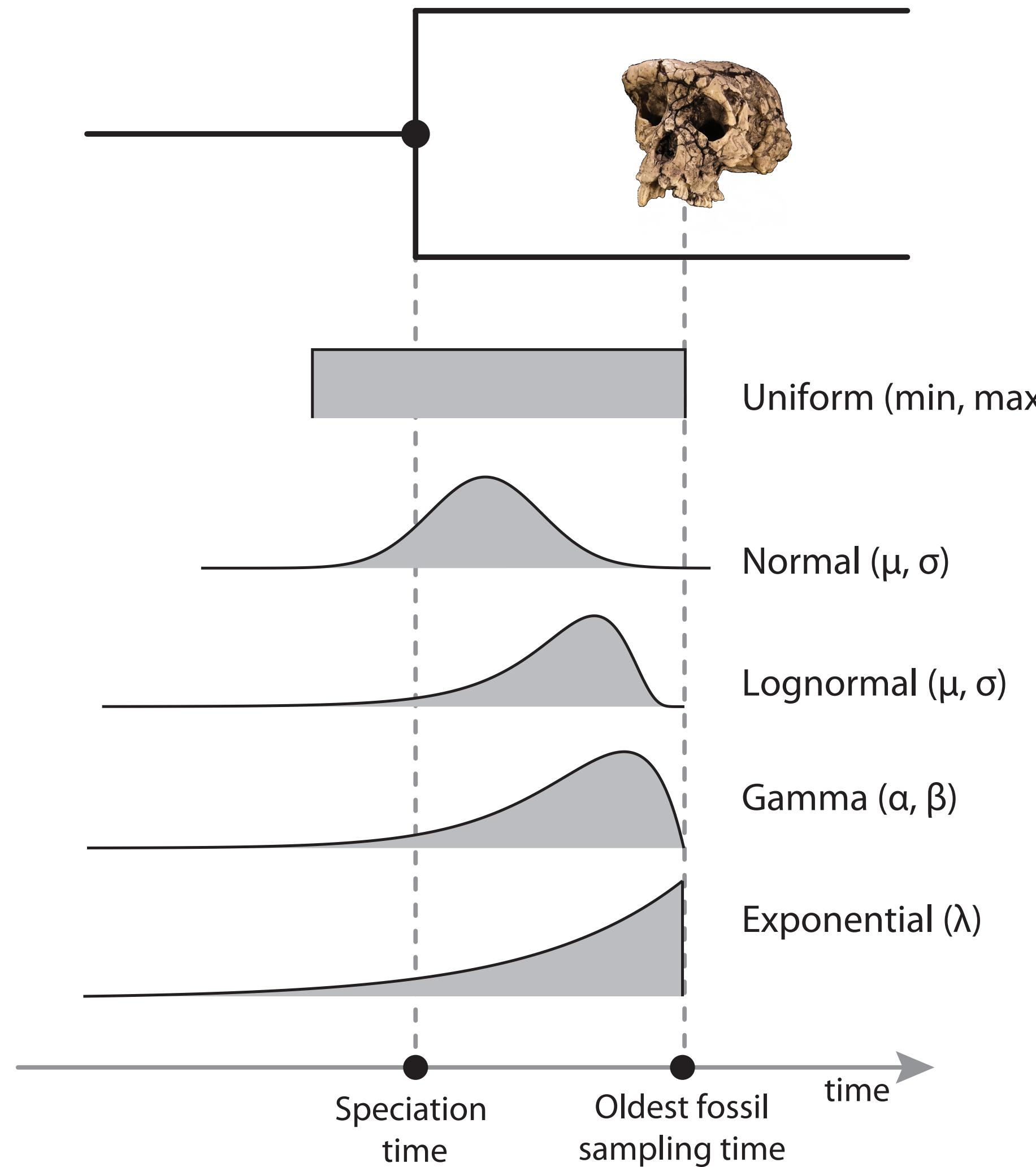


How have species originated, gone
extinct and been sampled through time?

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

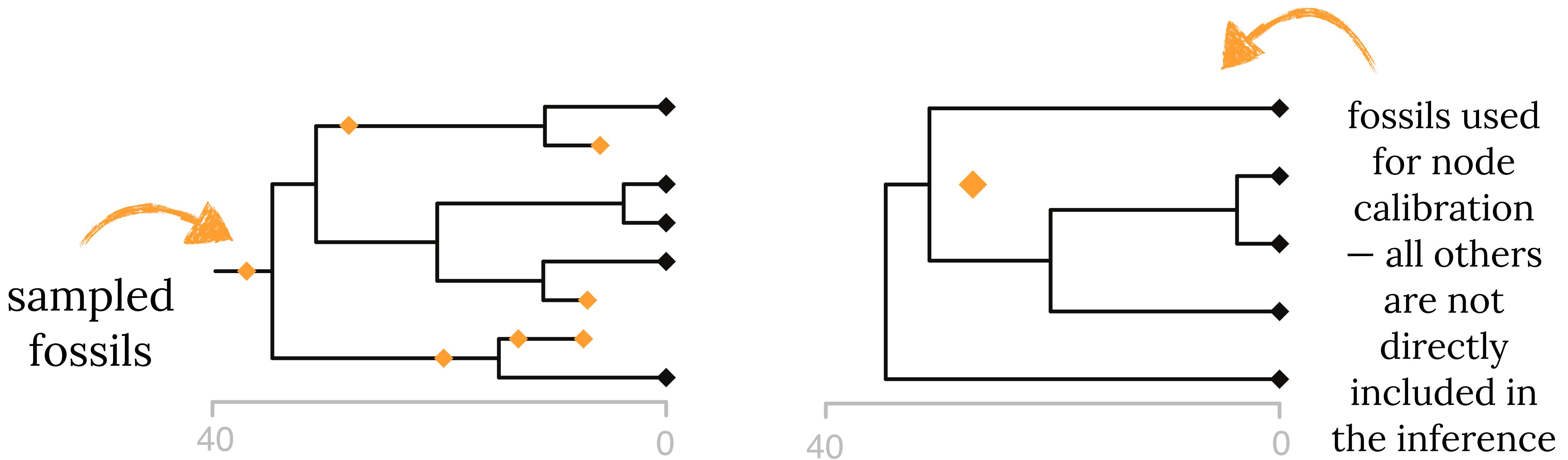
We typically use a **birth-death process** model to describe the tree generating process

Adapted from Heath 2012. Sys Bio

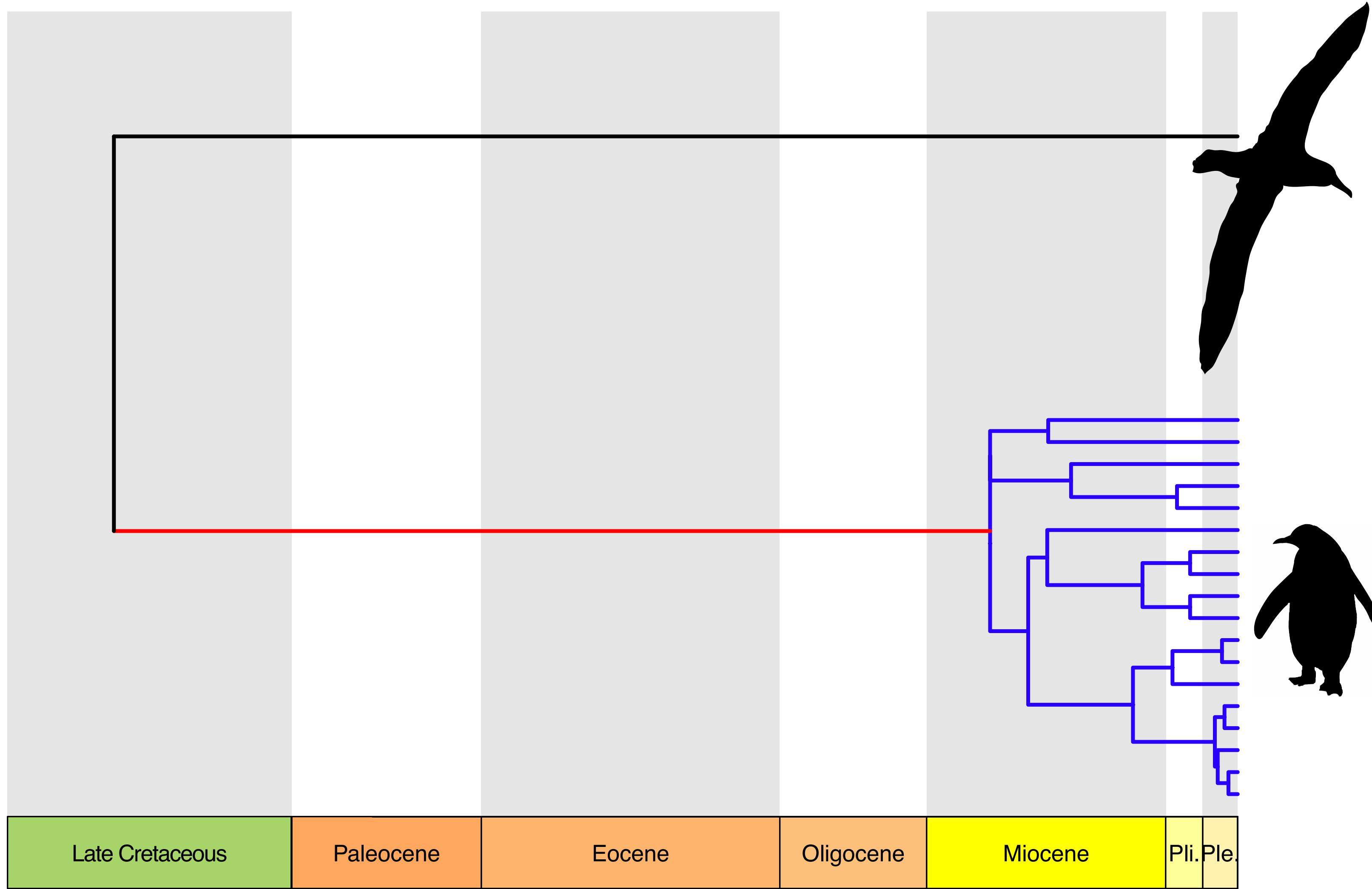
Node dating: potential issues

There are many!

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

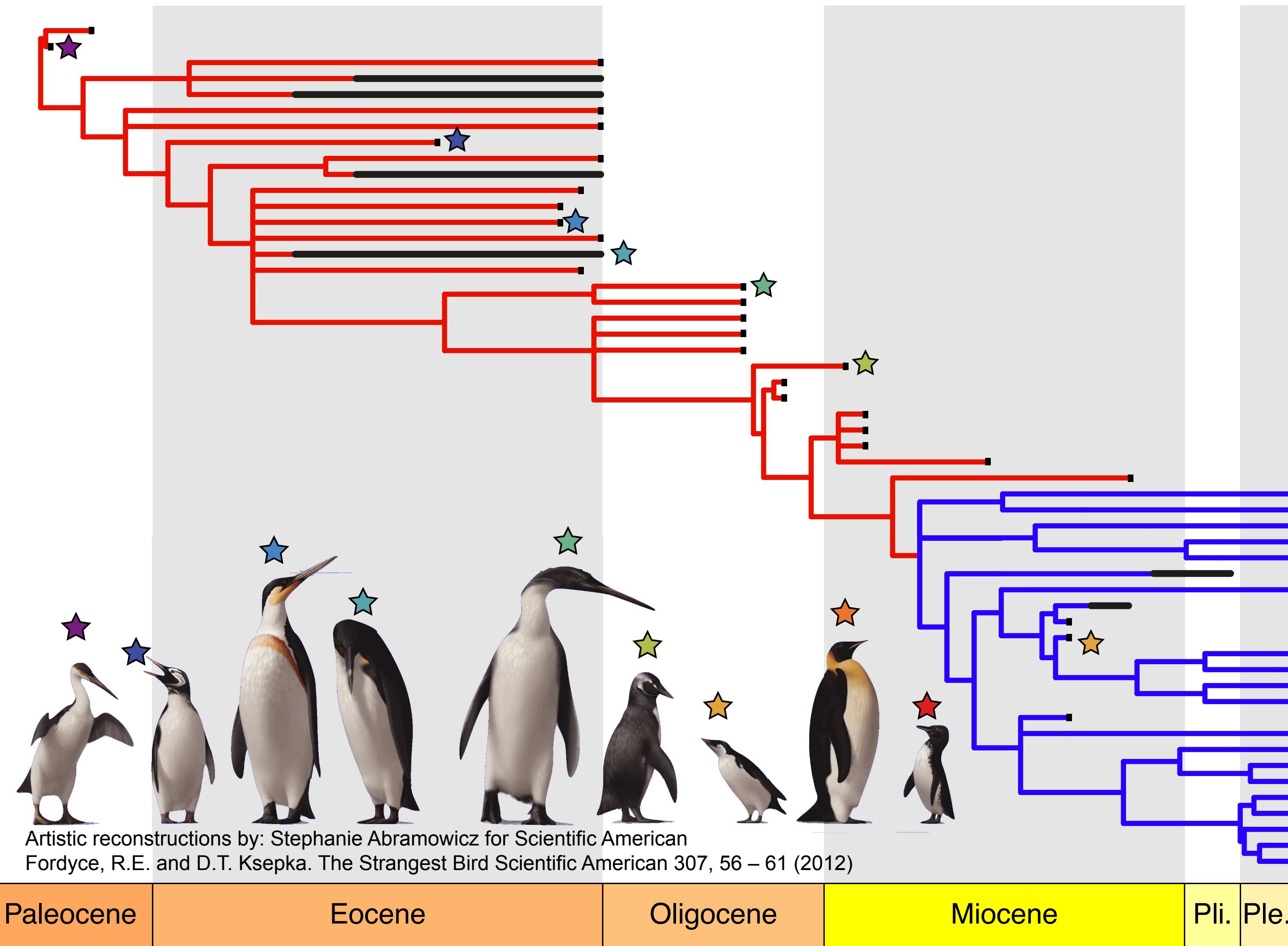


Example: living penguins



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

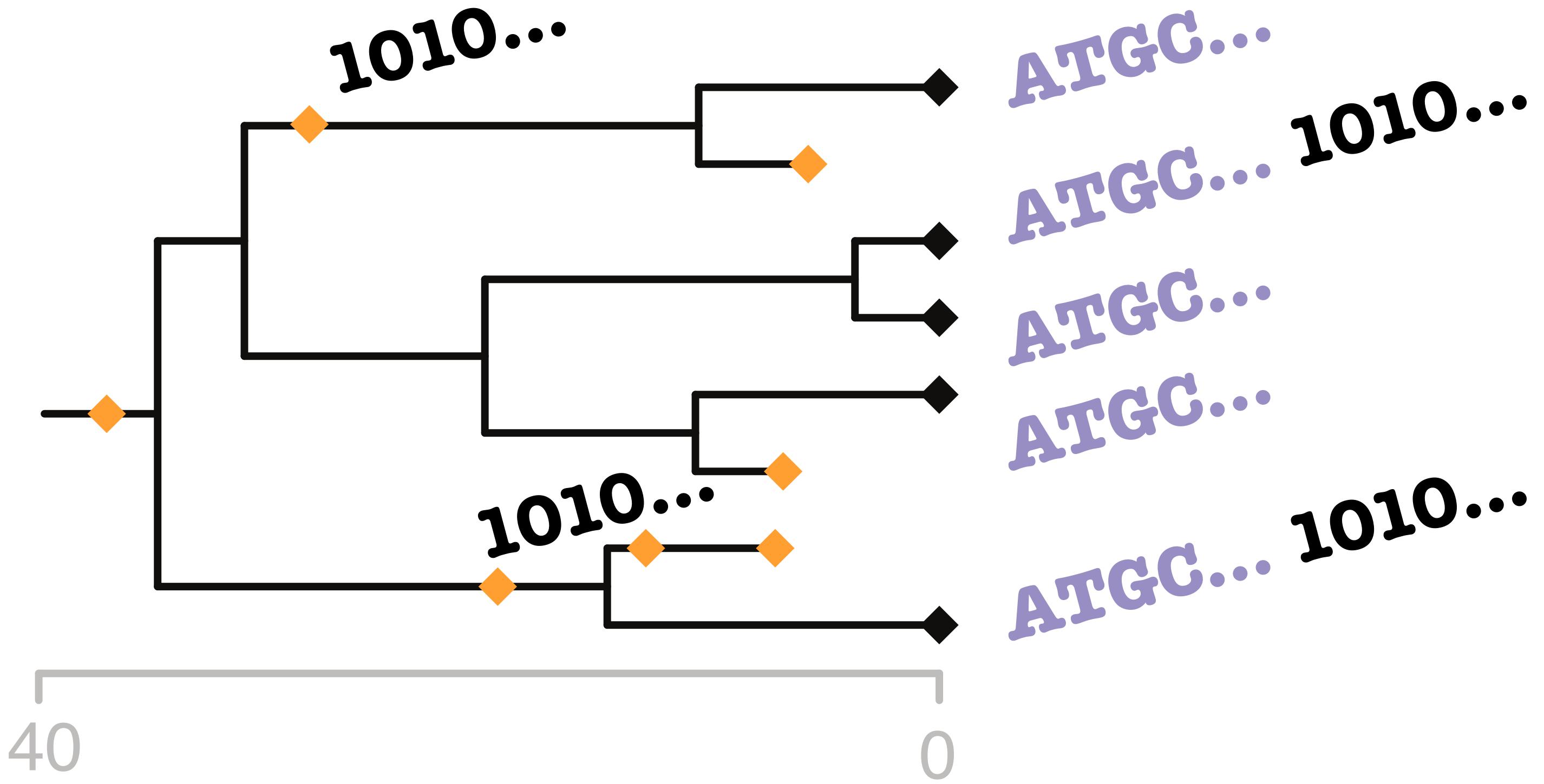
Yang and Rannala. [2006](#). MBE

Heled and Drummond. [2012](#). Sys Bio

Warnock et al. [2012, 2015](#)

Total-evidence dating

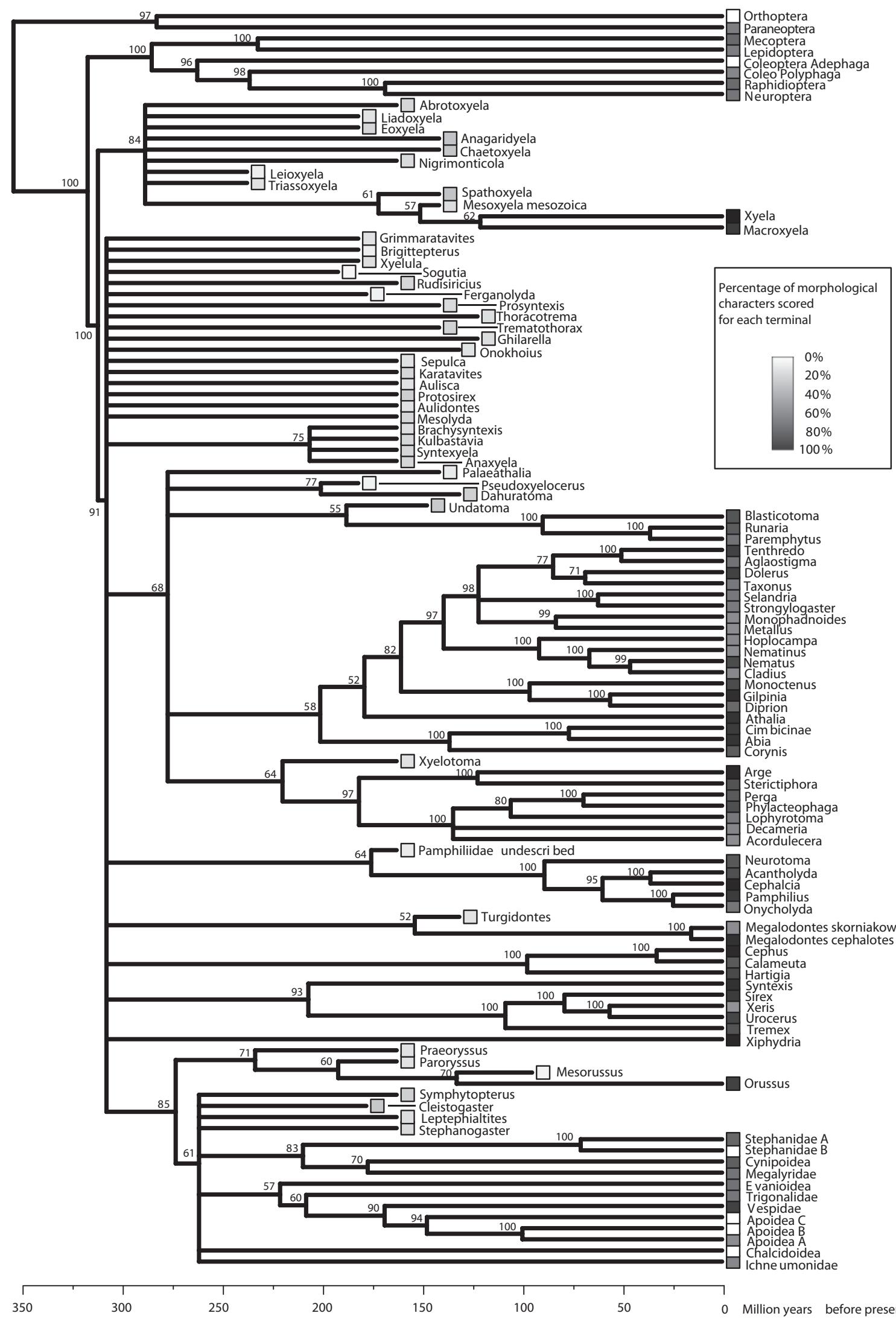
Tip-dating or “total-evidence” dating



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

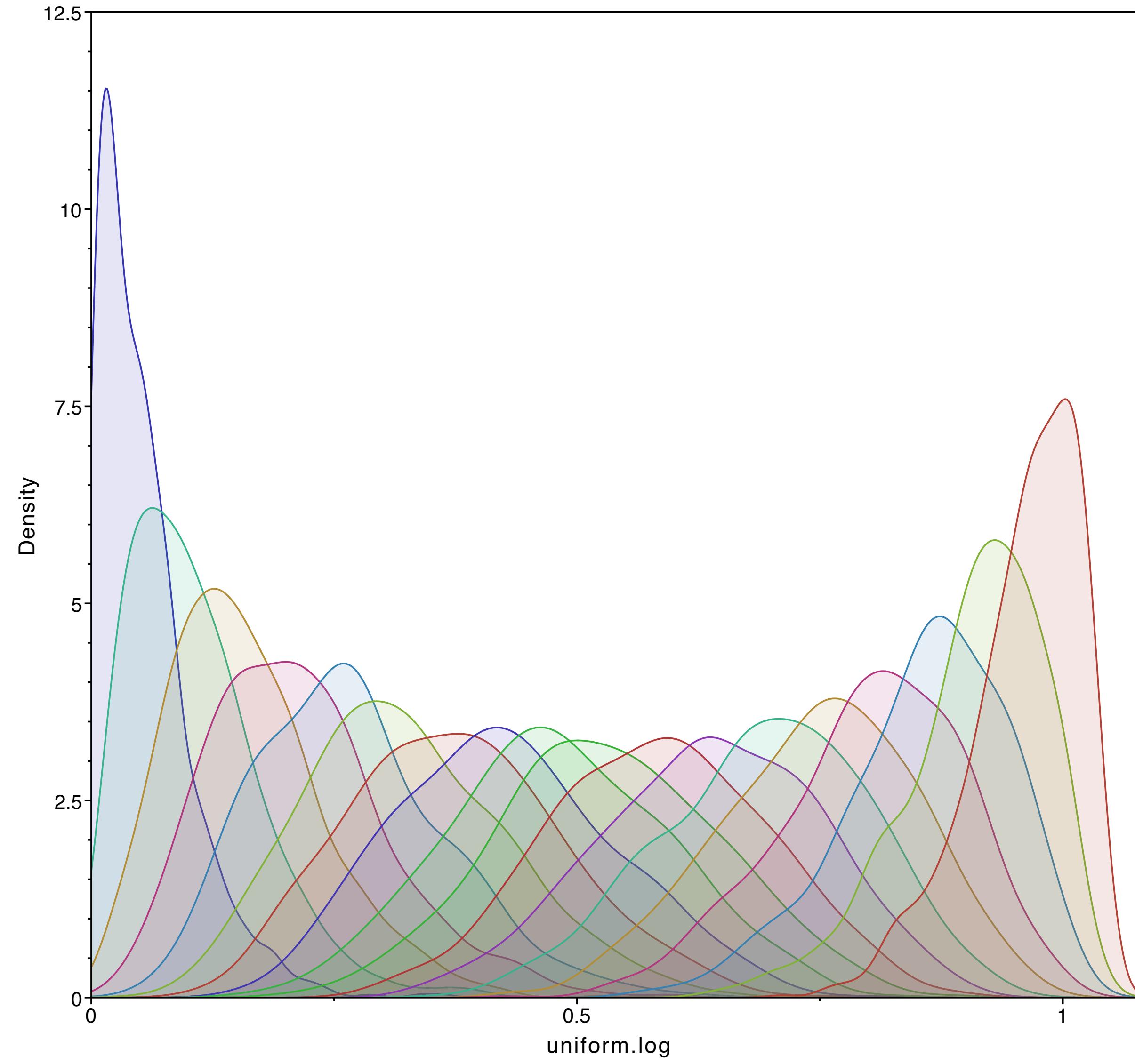
Fossils can be positioned on the basis of morphology
→ accounts for uncertainty in fossil placement

The uniform tree prior



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

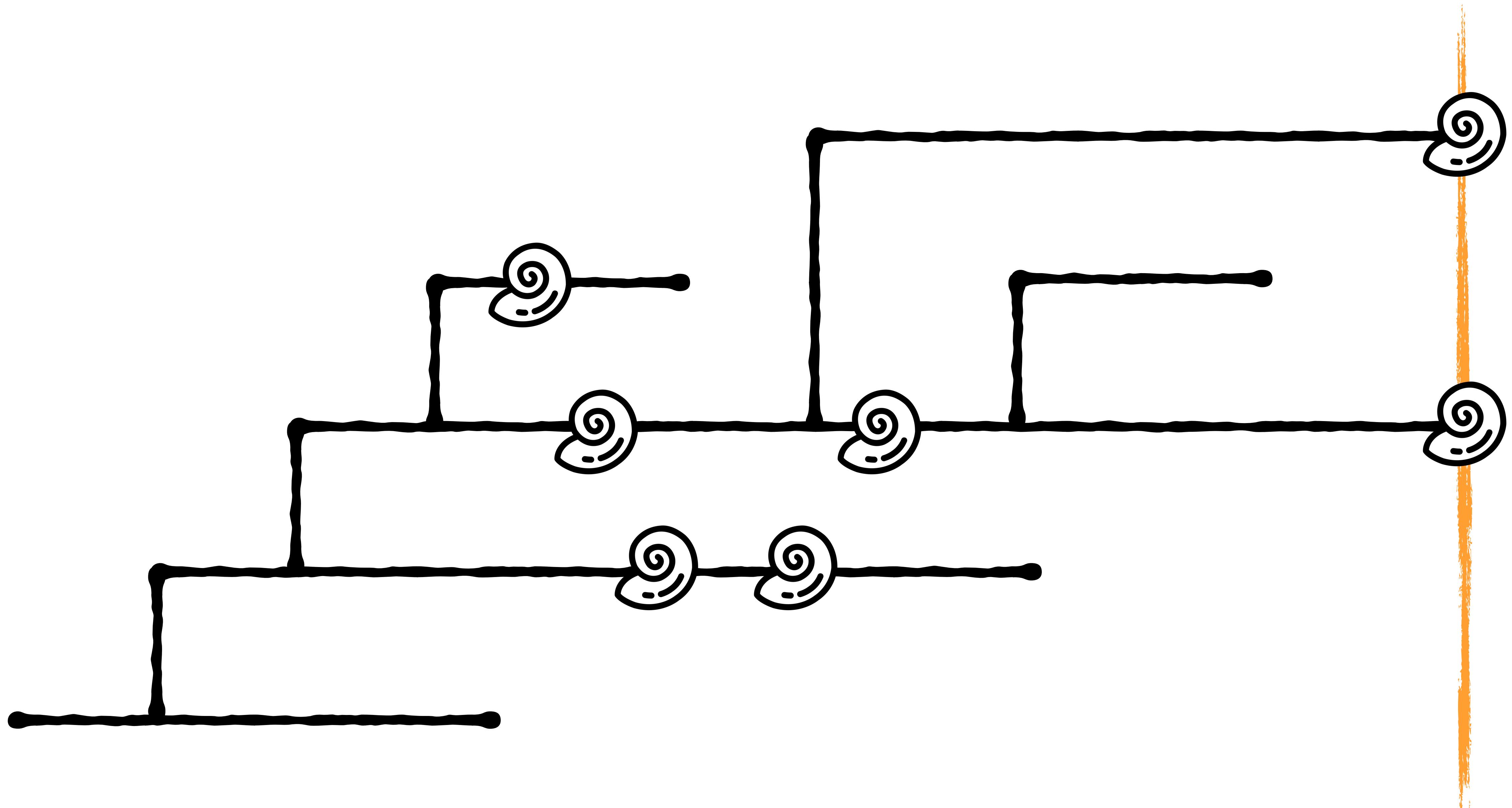


A uniform tree prior implies time till the next split is independent of how many lineages there are present

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?

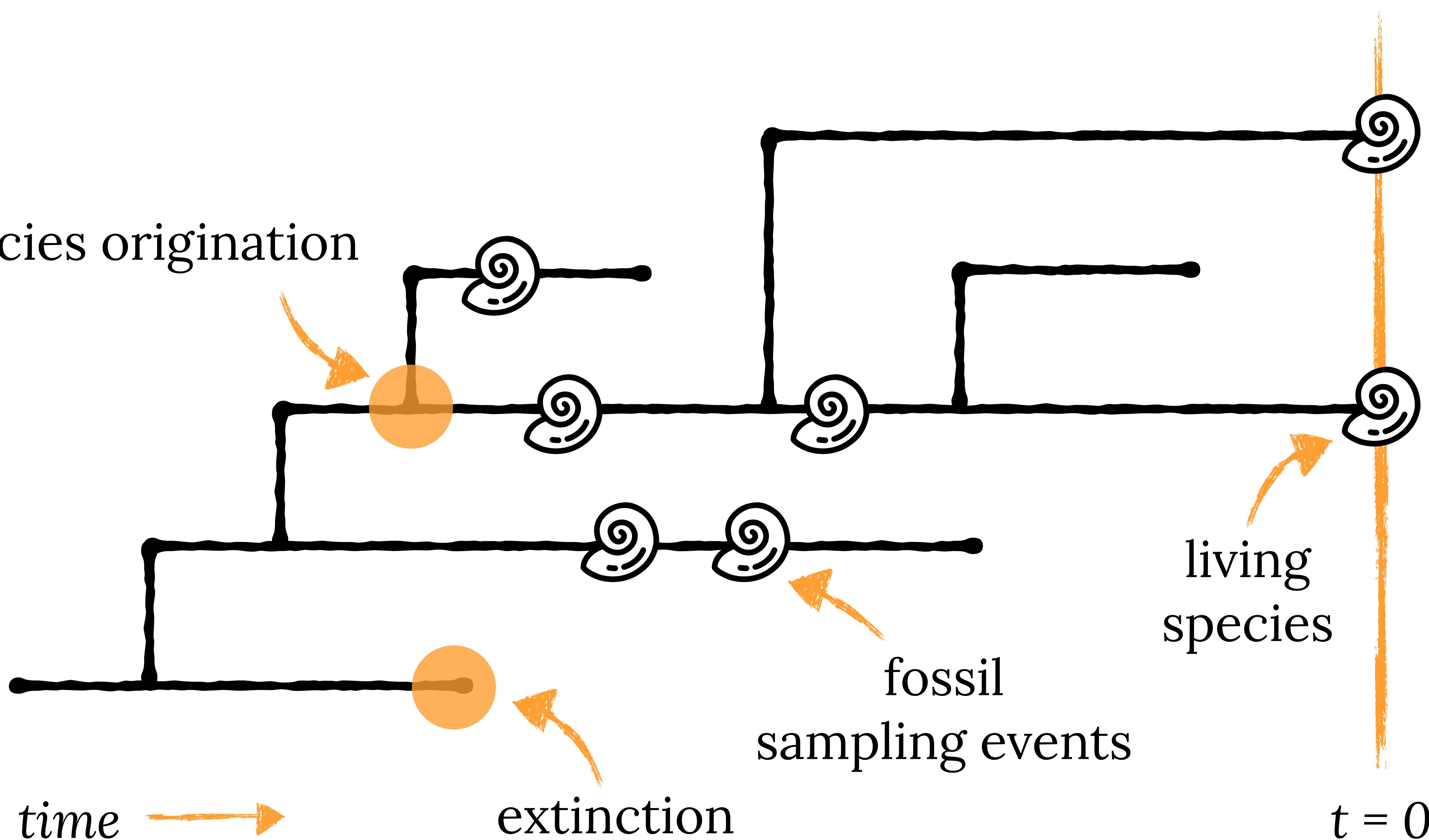
The fossilised birth-death process



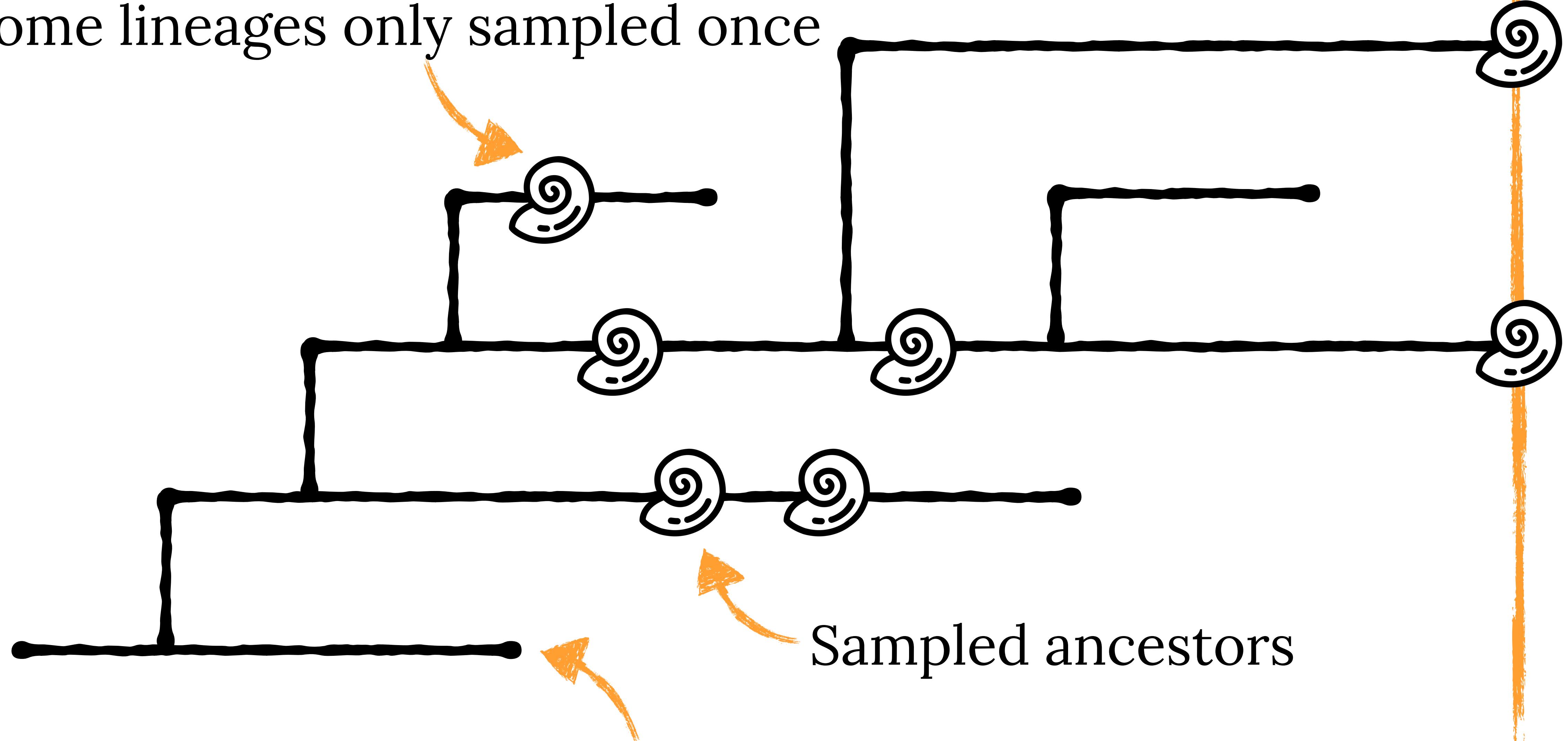
time →

$t = 0$

species origination



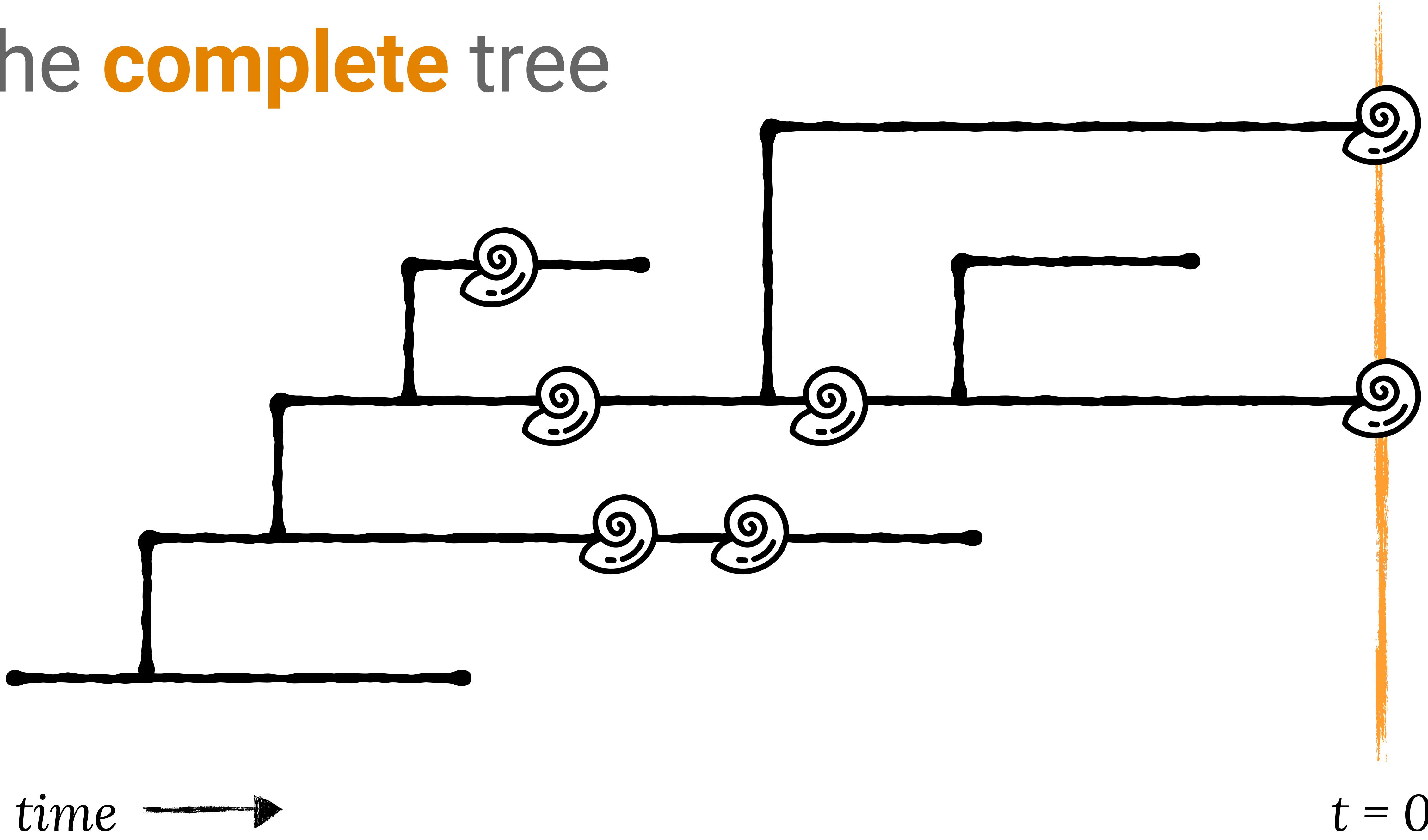
Some lineages only sampled once



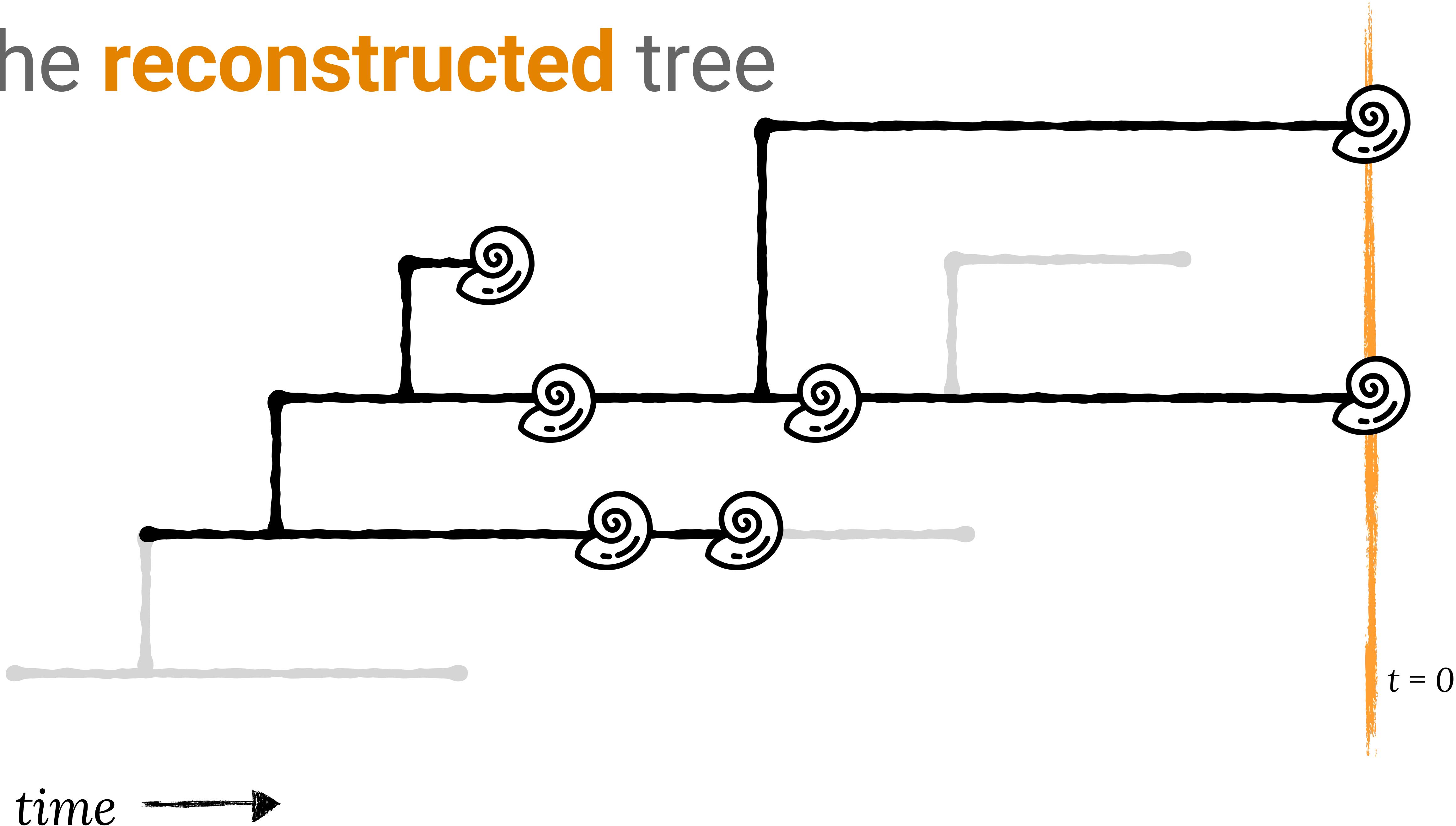
Some lineages go completely unsampled

$t = 0$

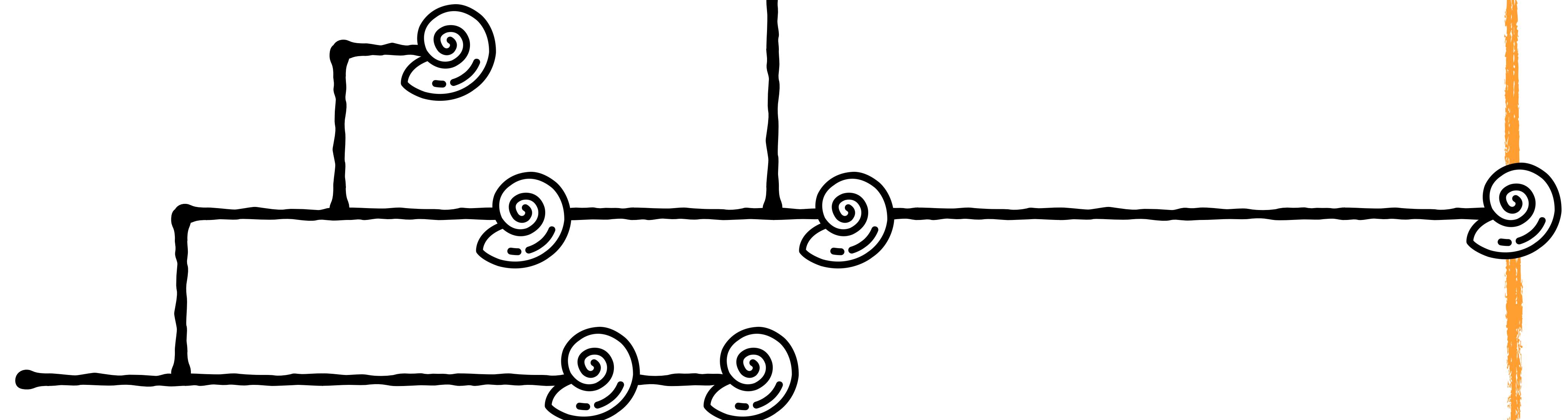
The complete tree



The reconstructed tree



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

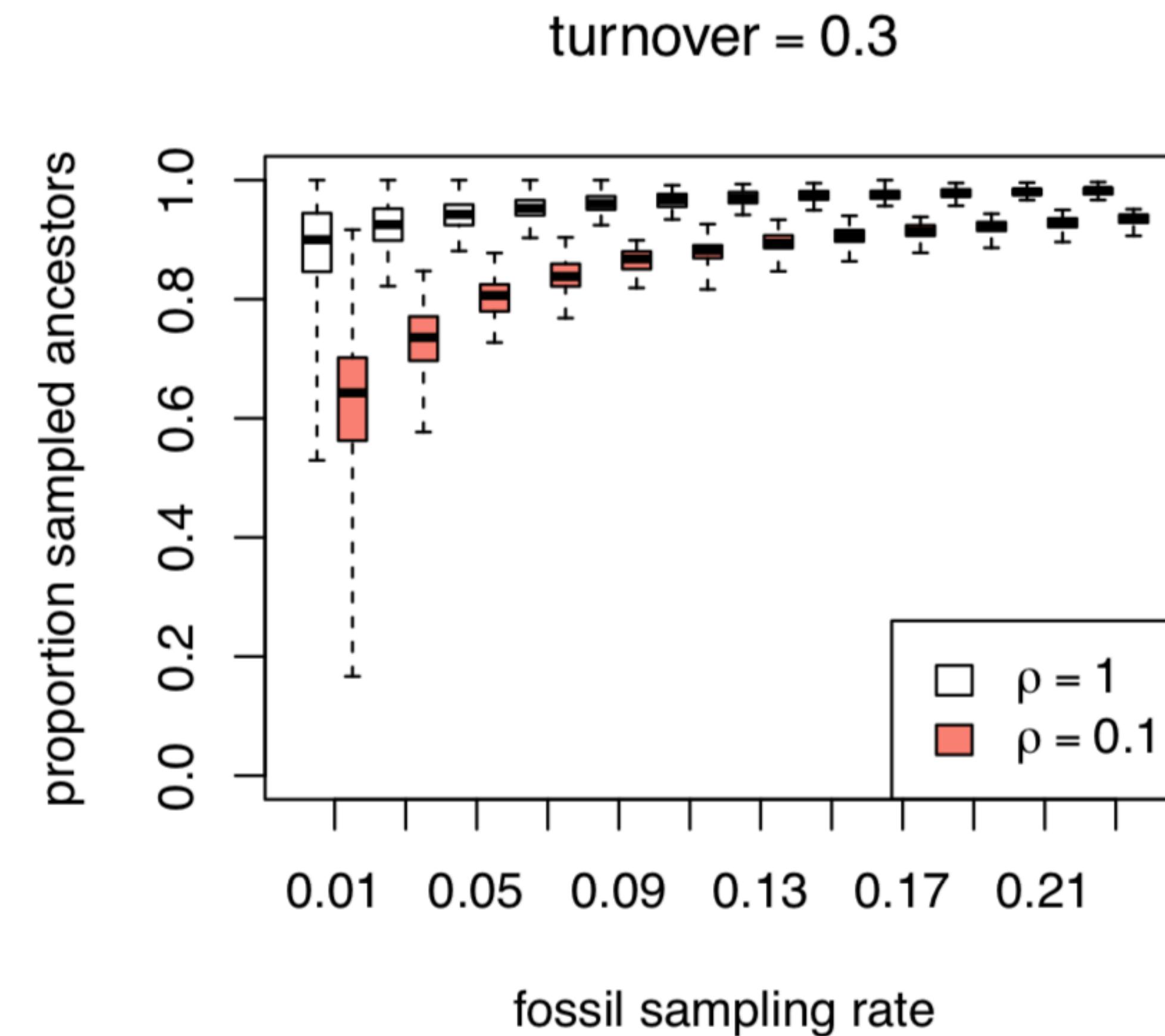
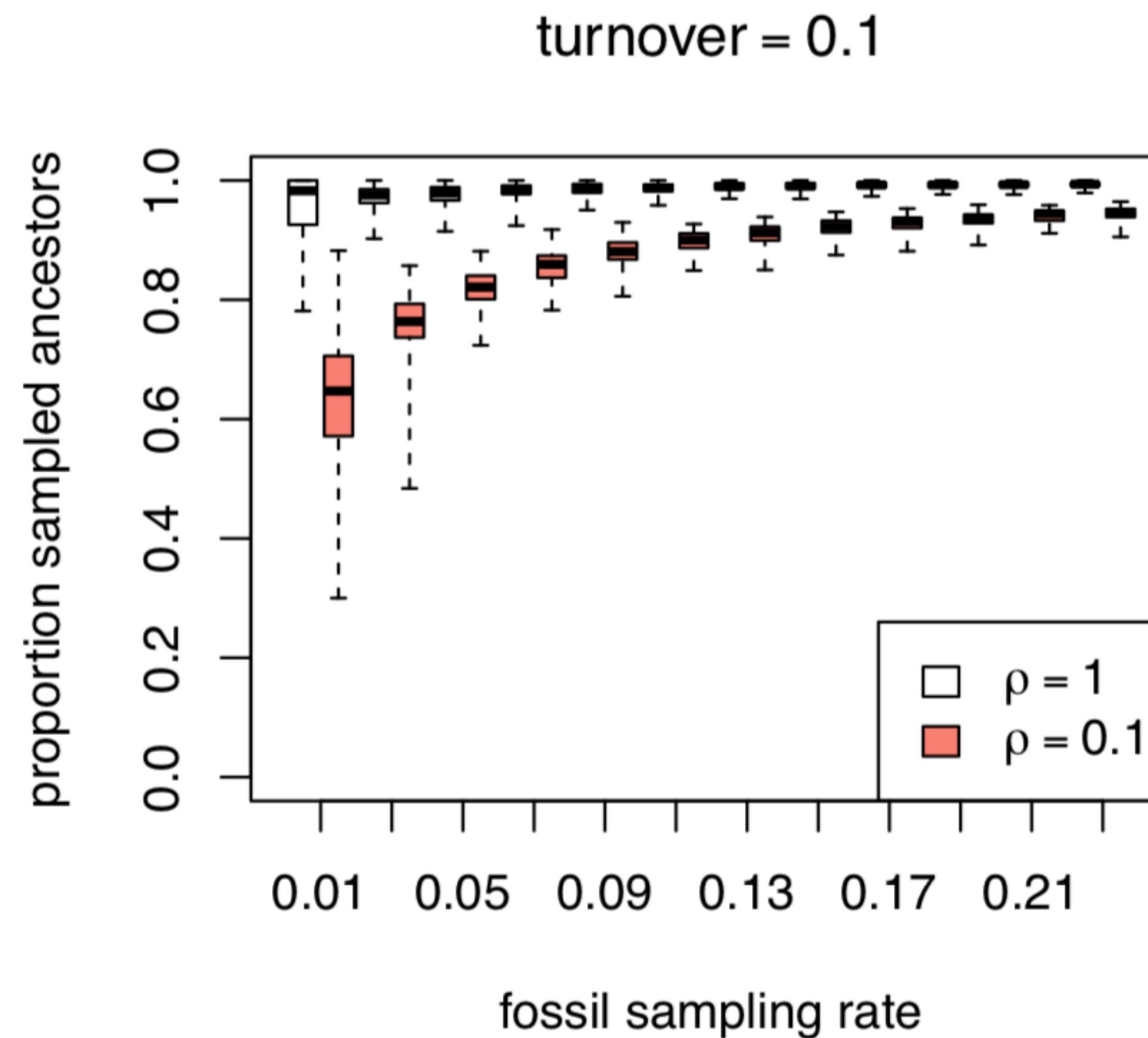


$$P(E | \text{snail}, \lambda, \mu, \psi, \rho)$$

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

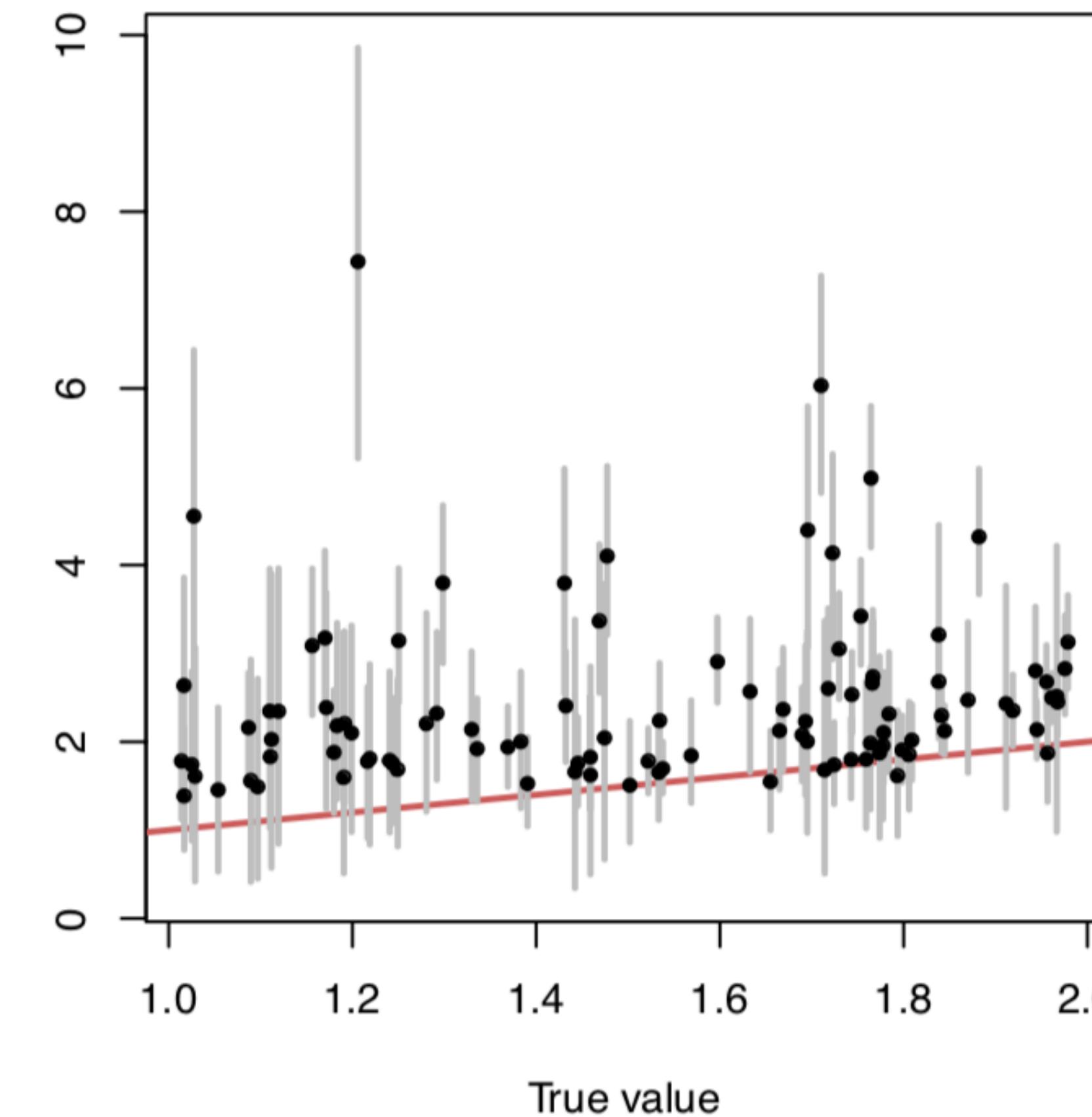
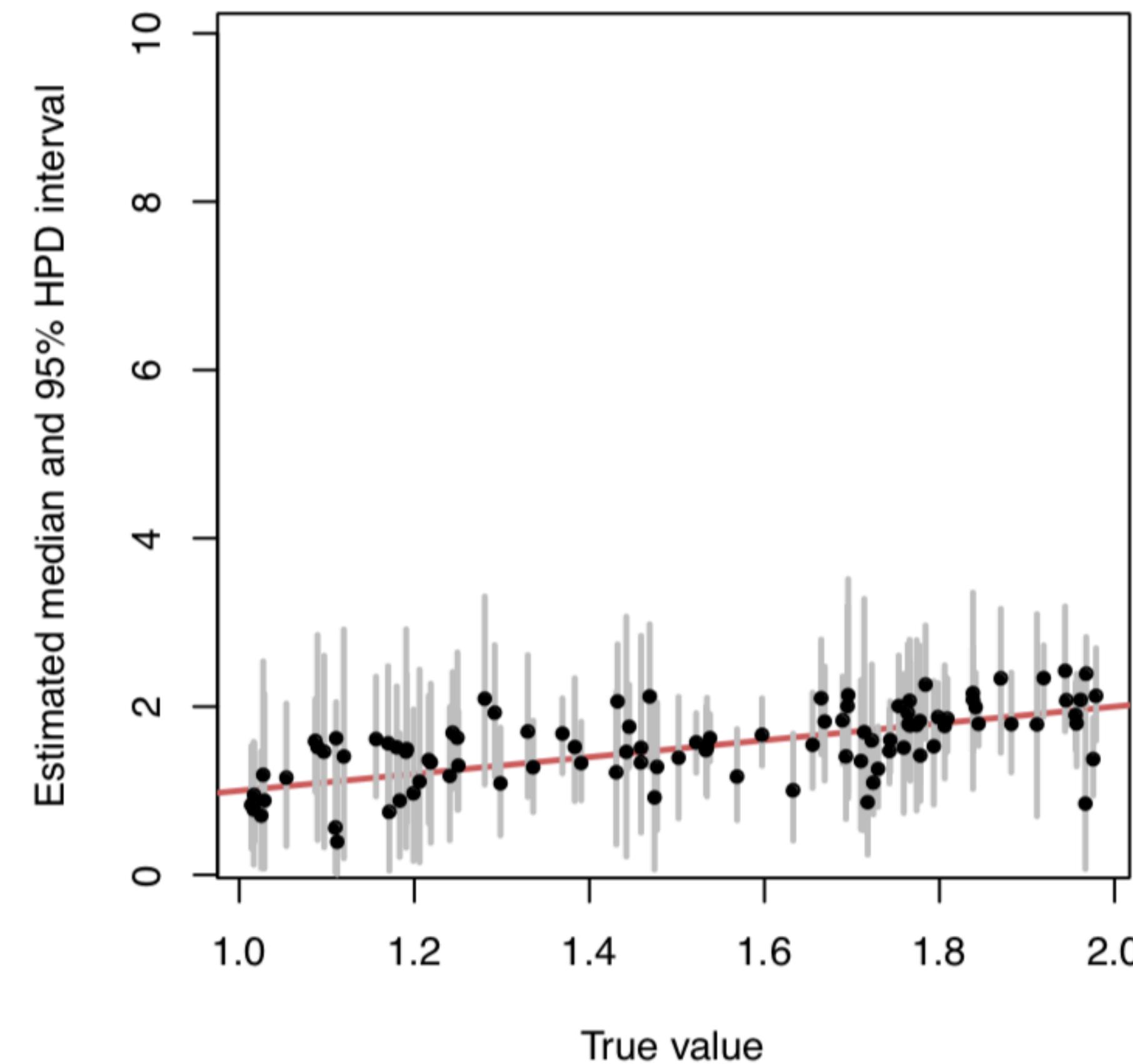
Sampled ancestors

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

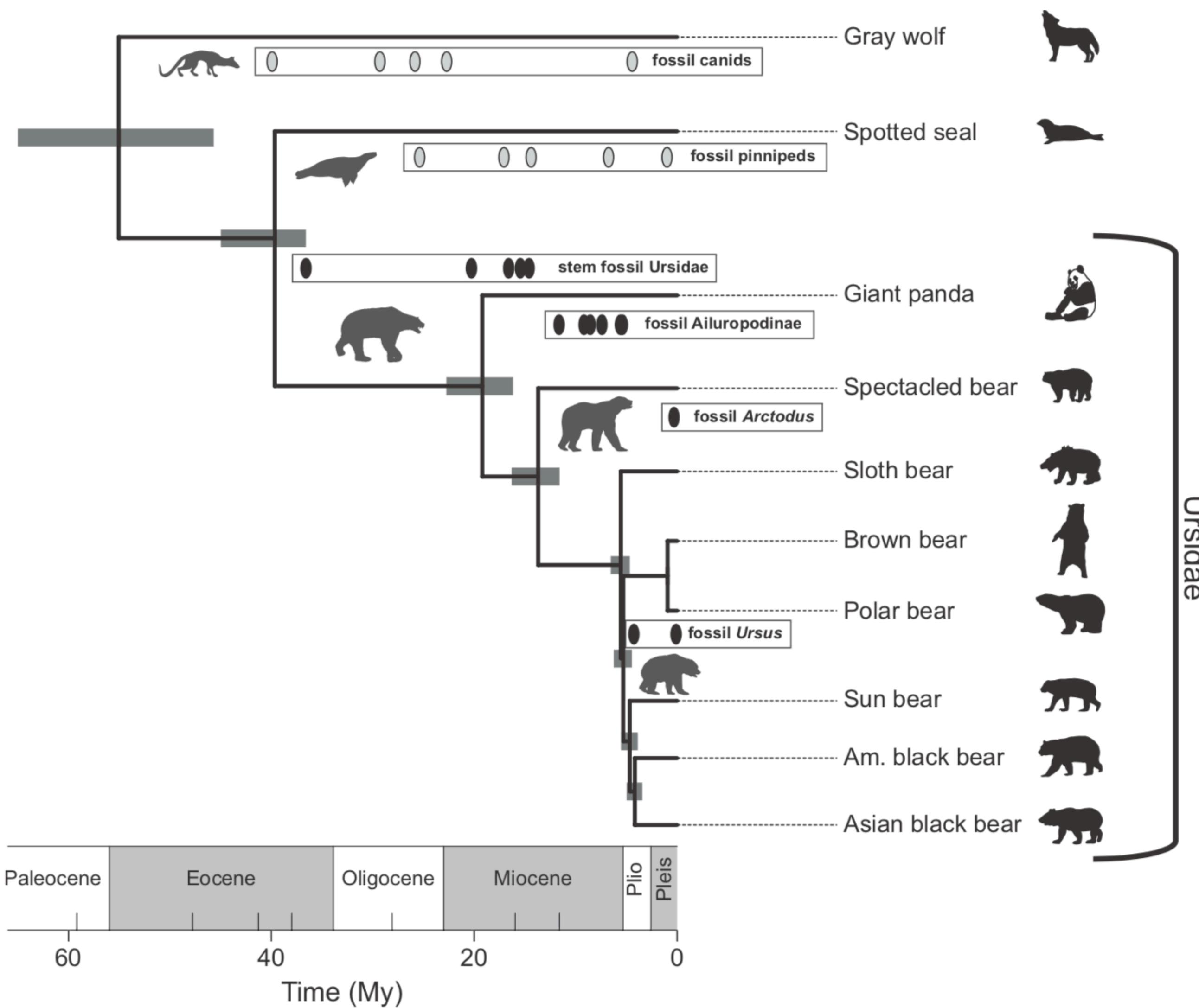


Sampled ancestors

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