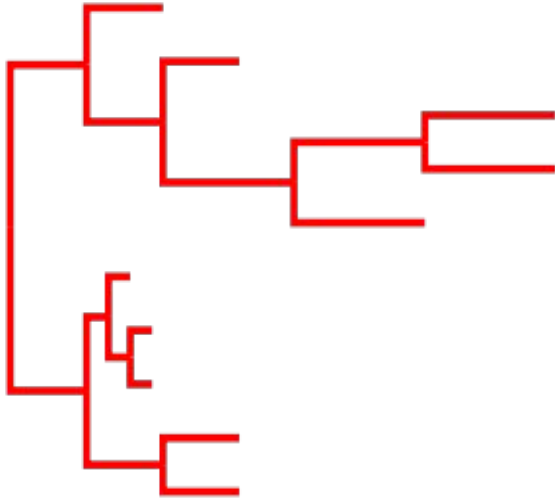


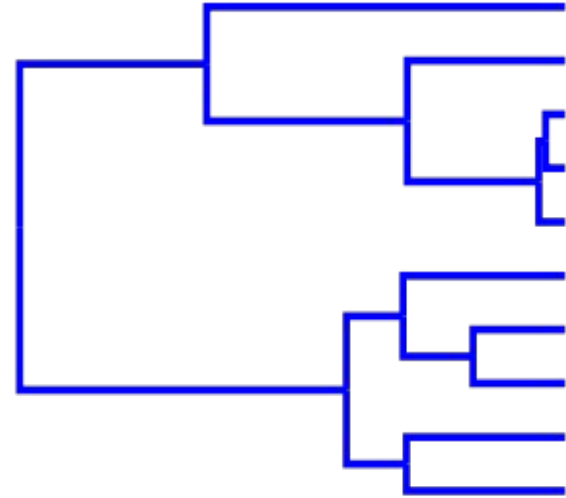
Sampled ancestors and divergence time estimation: the Fossilized Birth-Death process

Joëlle Barido-Sottani

Divergence time estimation

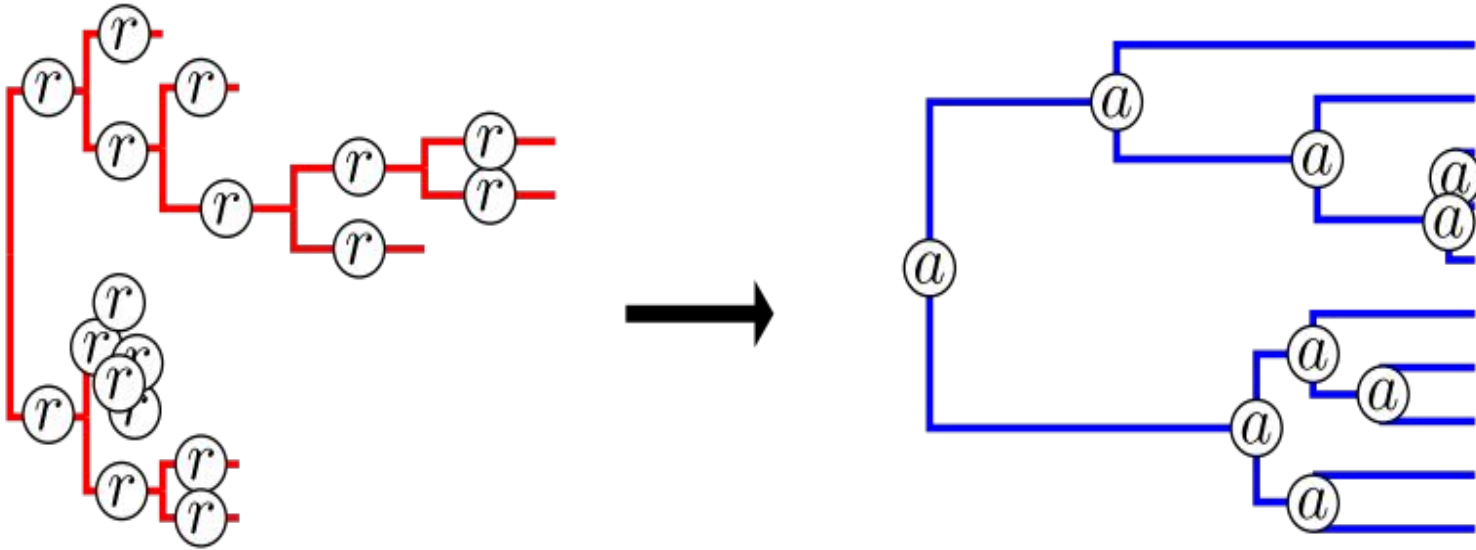


Branch lengths
represent
substitutions



Branch lengths
represent
time

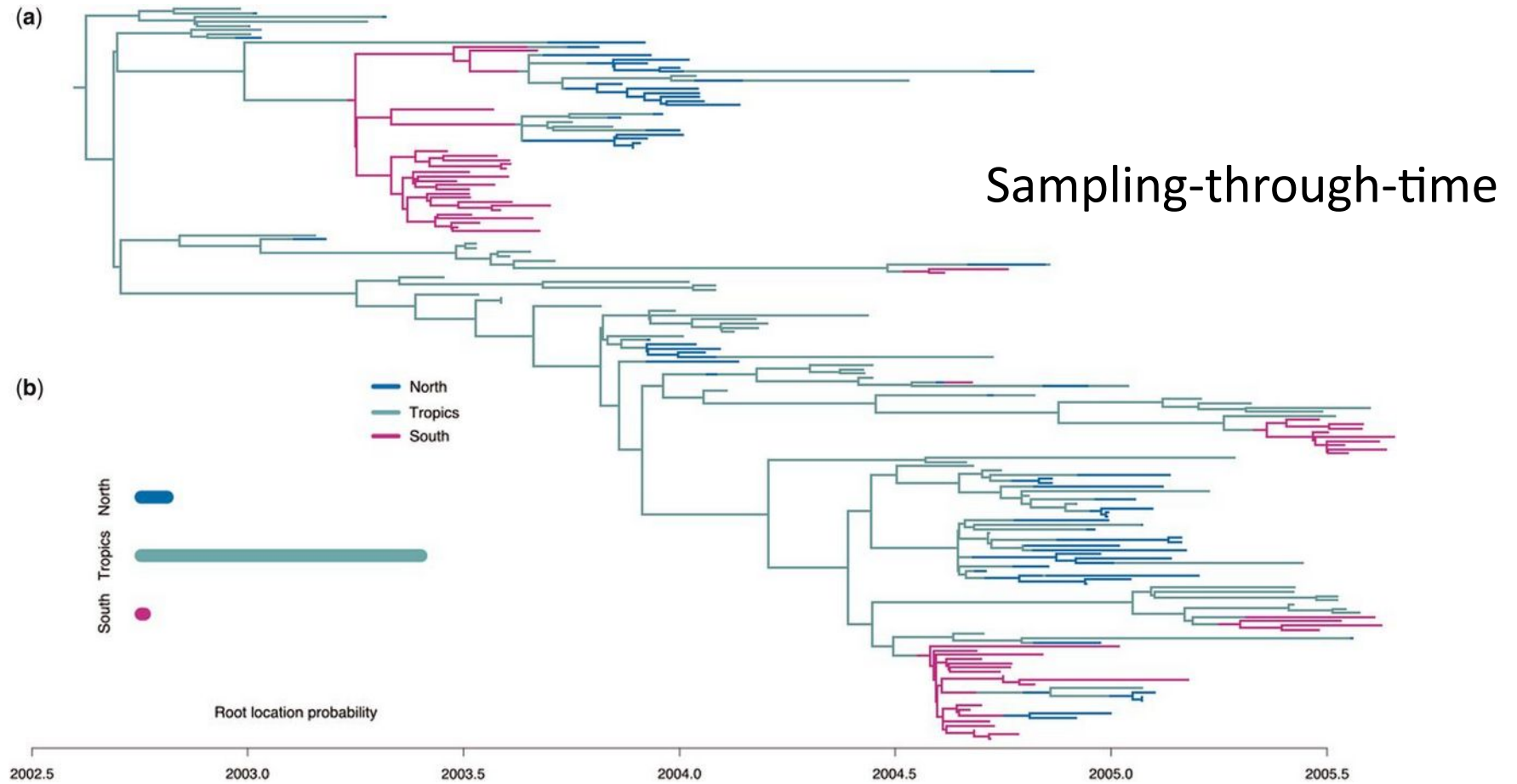
Divergence time estimation



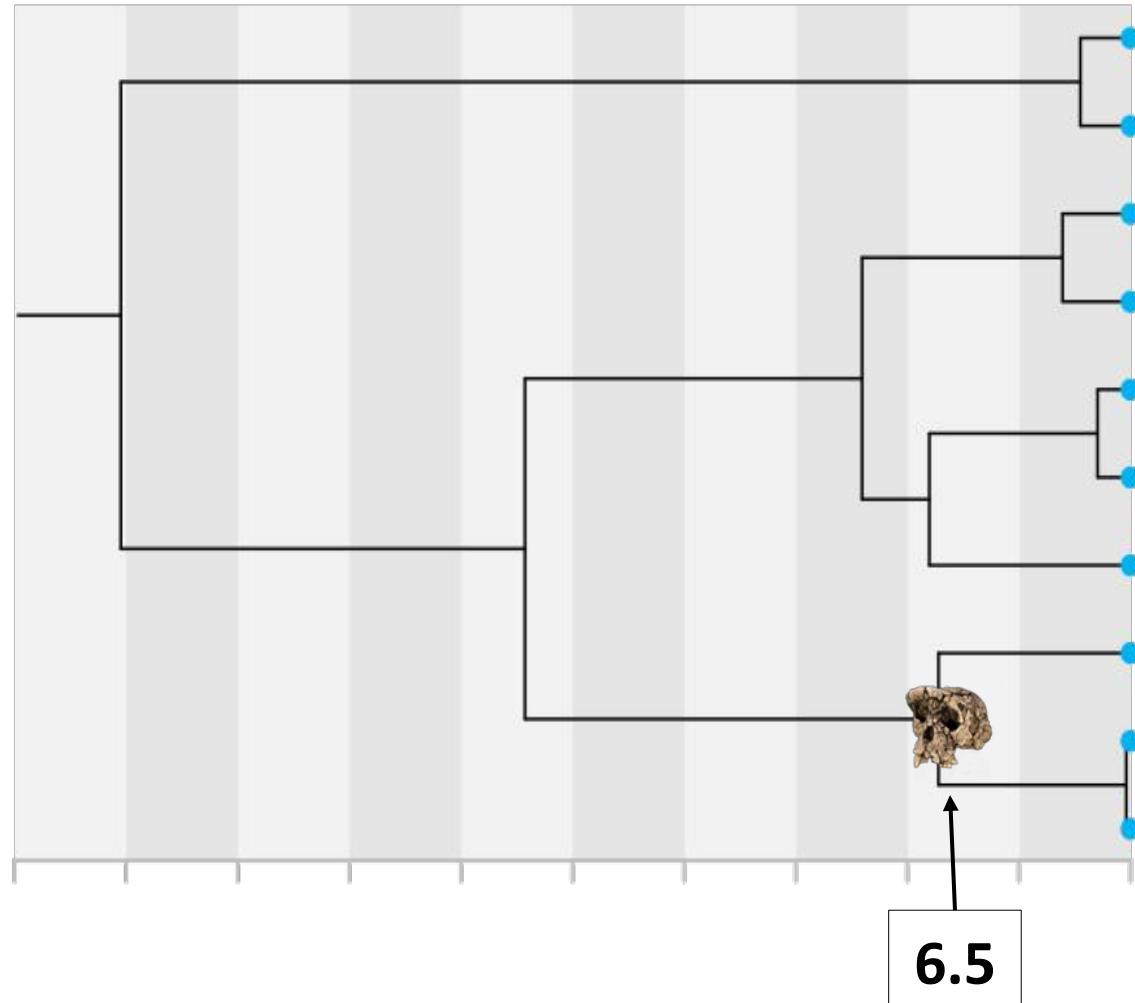
$$\text{Age} = r_{\text{clock}} \times N_{\text{substitutions}}$$



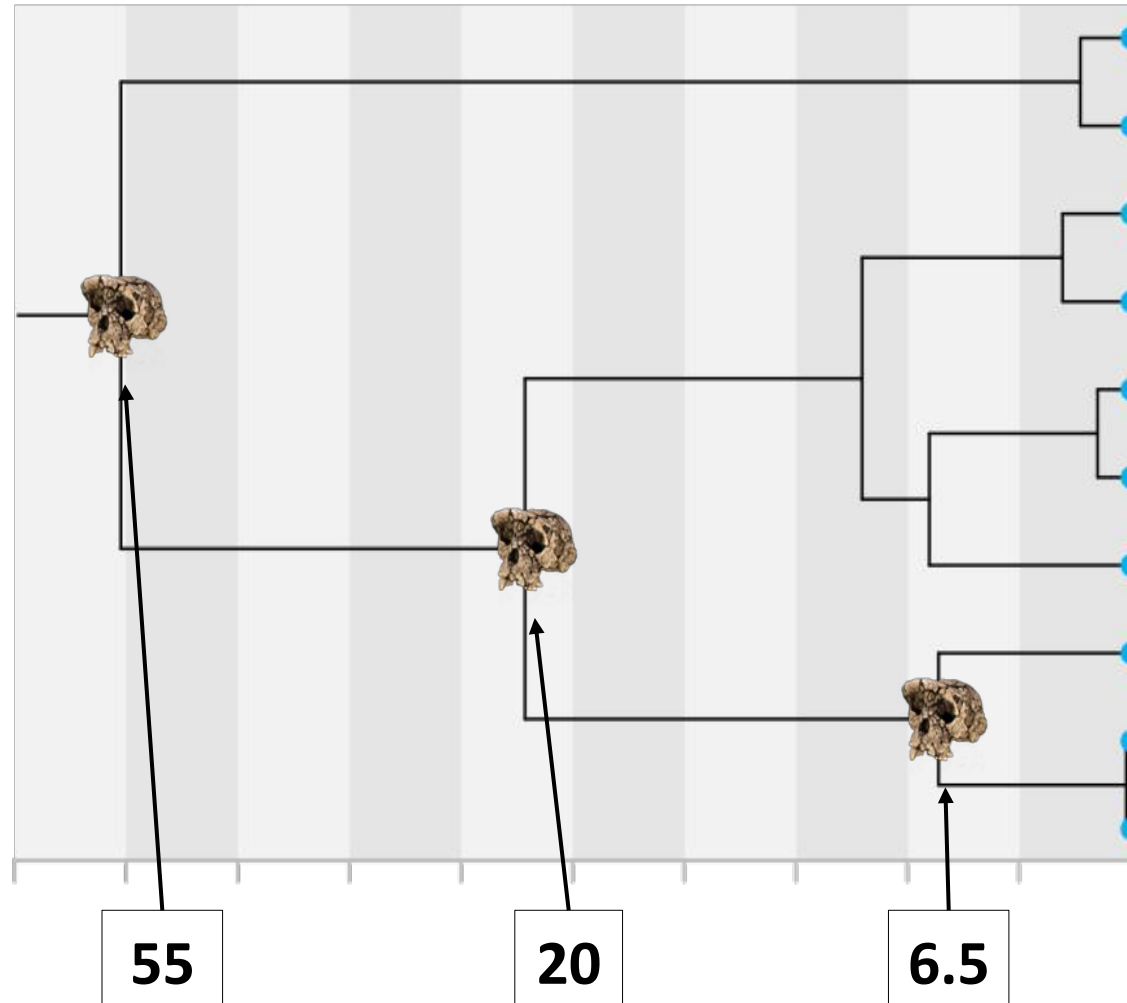
Calibrating the molecular clock



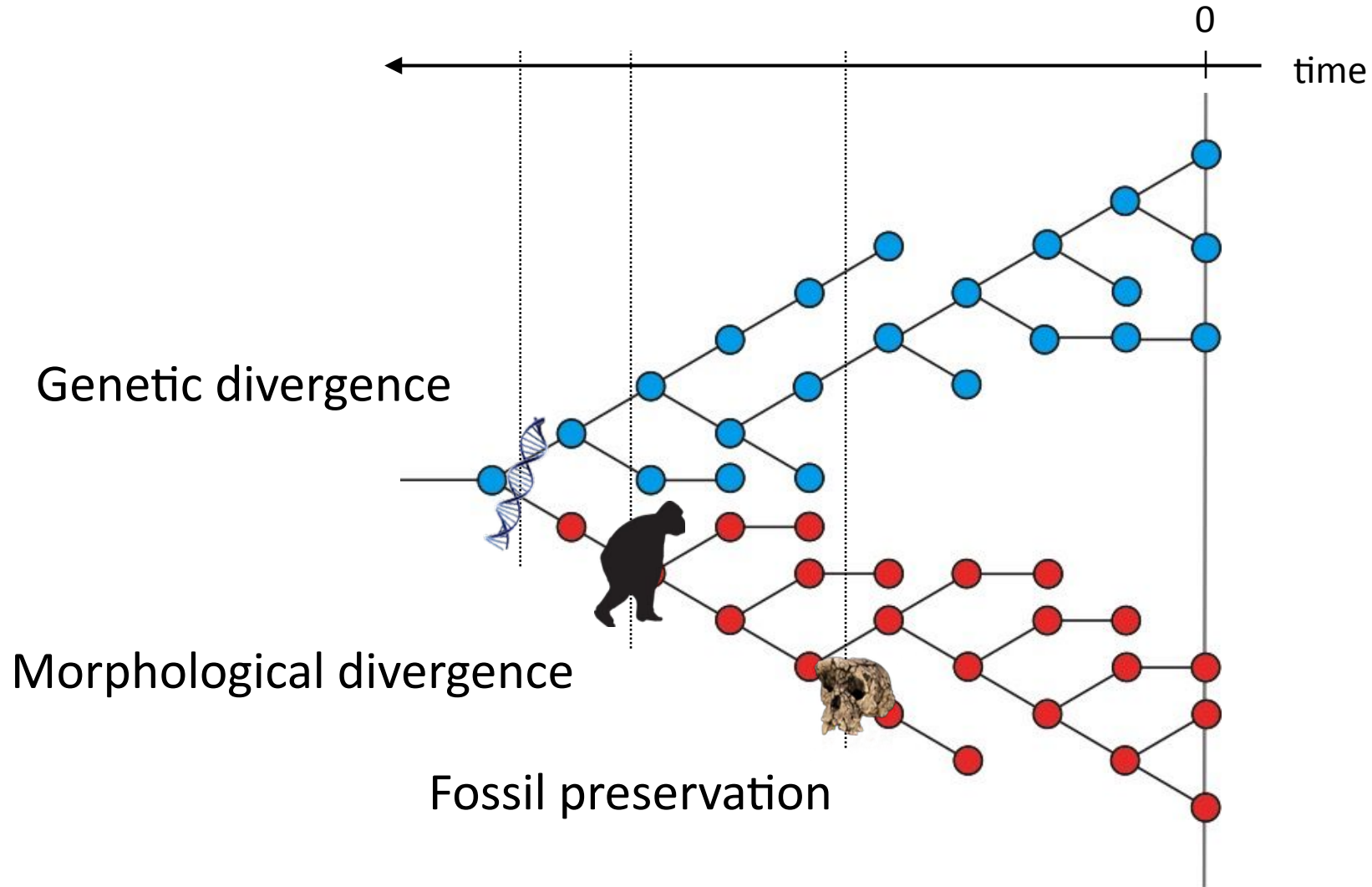
Calibrating the molecular clock



Calibrating the molecular clock

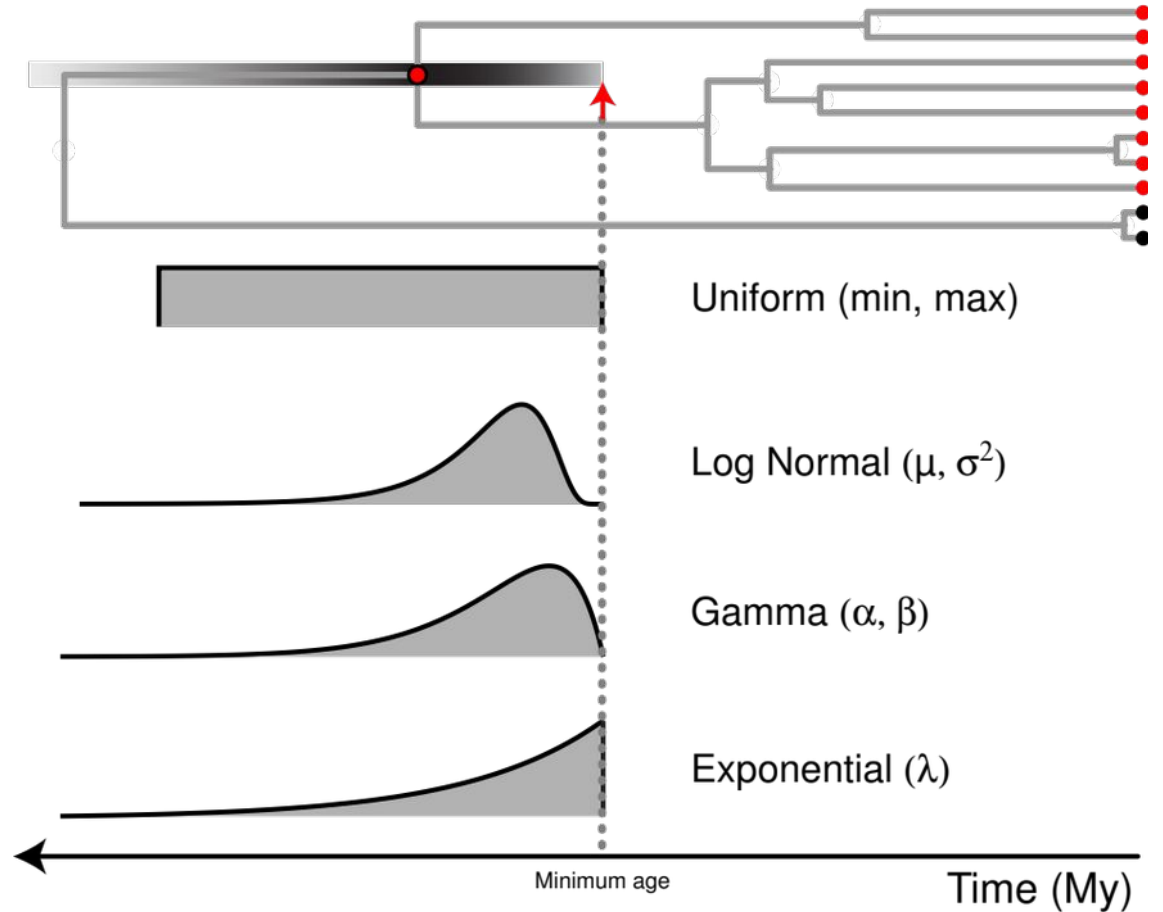


How do fossils constrain age?

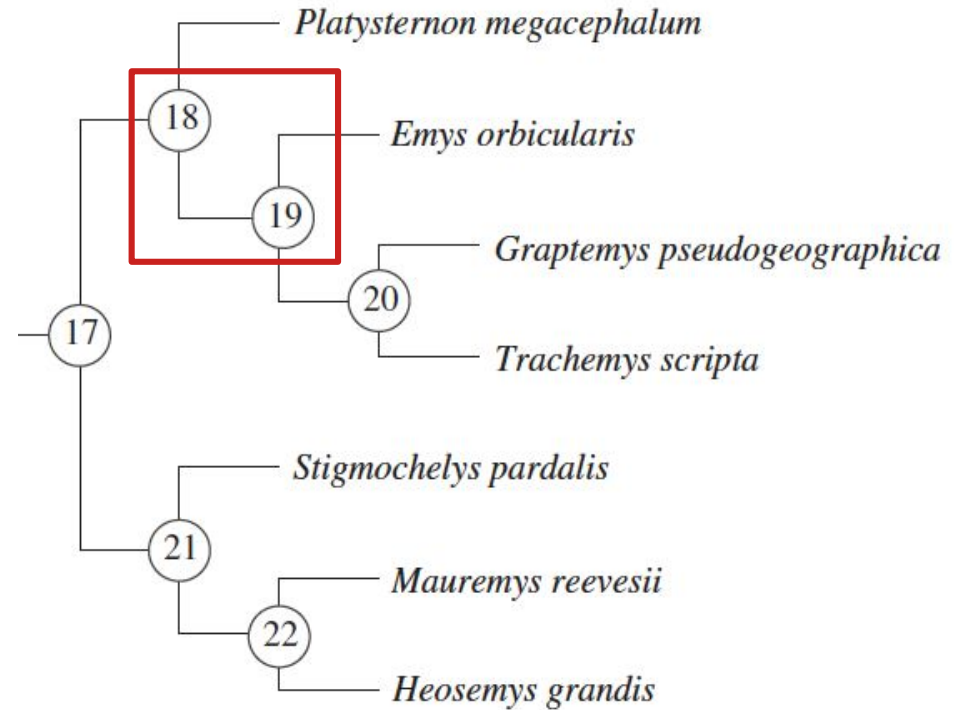
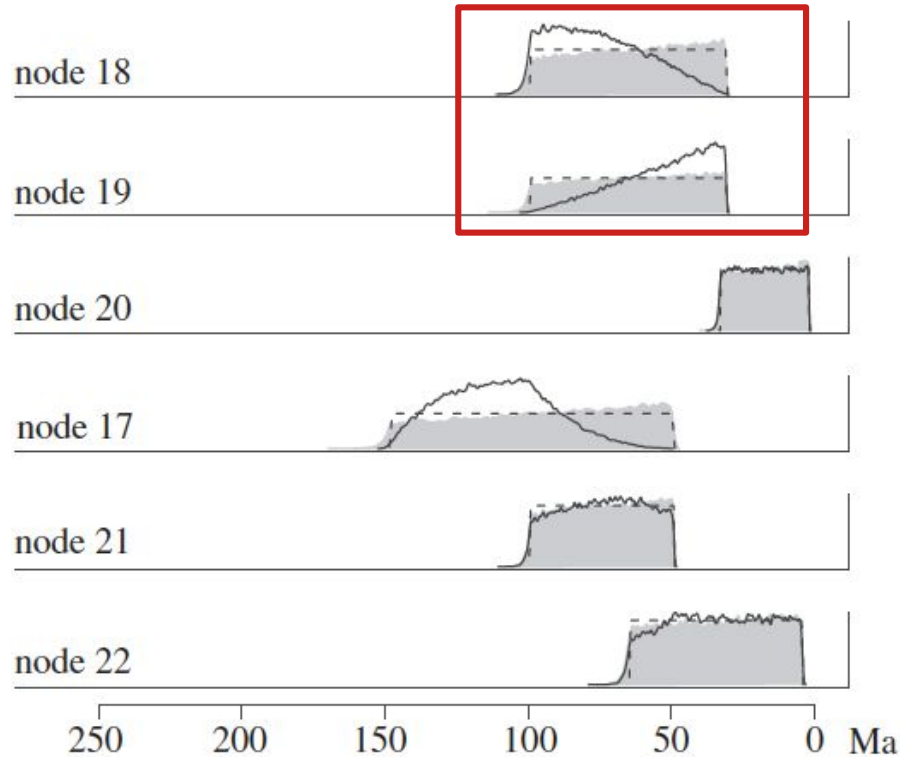


Using calibration densities

Prior distribution on
the difference
between the genetic
divergence and the
sampled fossil

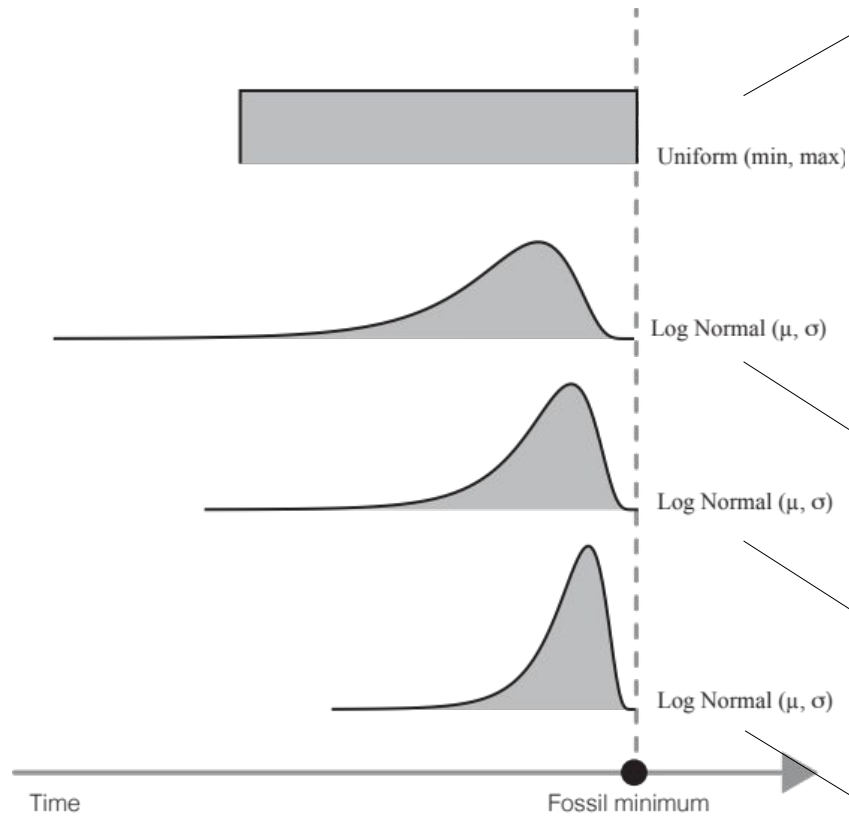


Issue #1: conflicting calibrations

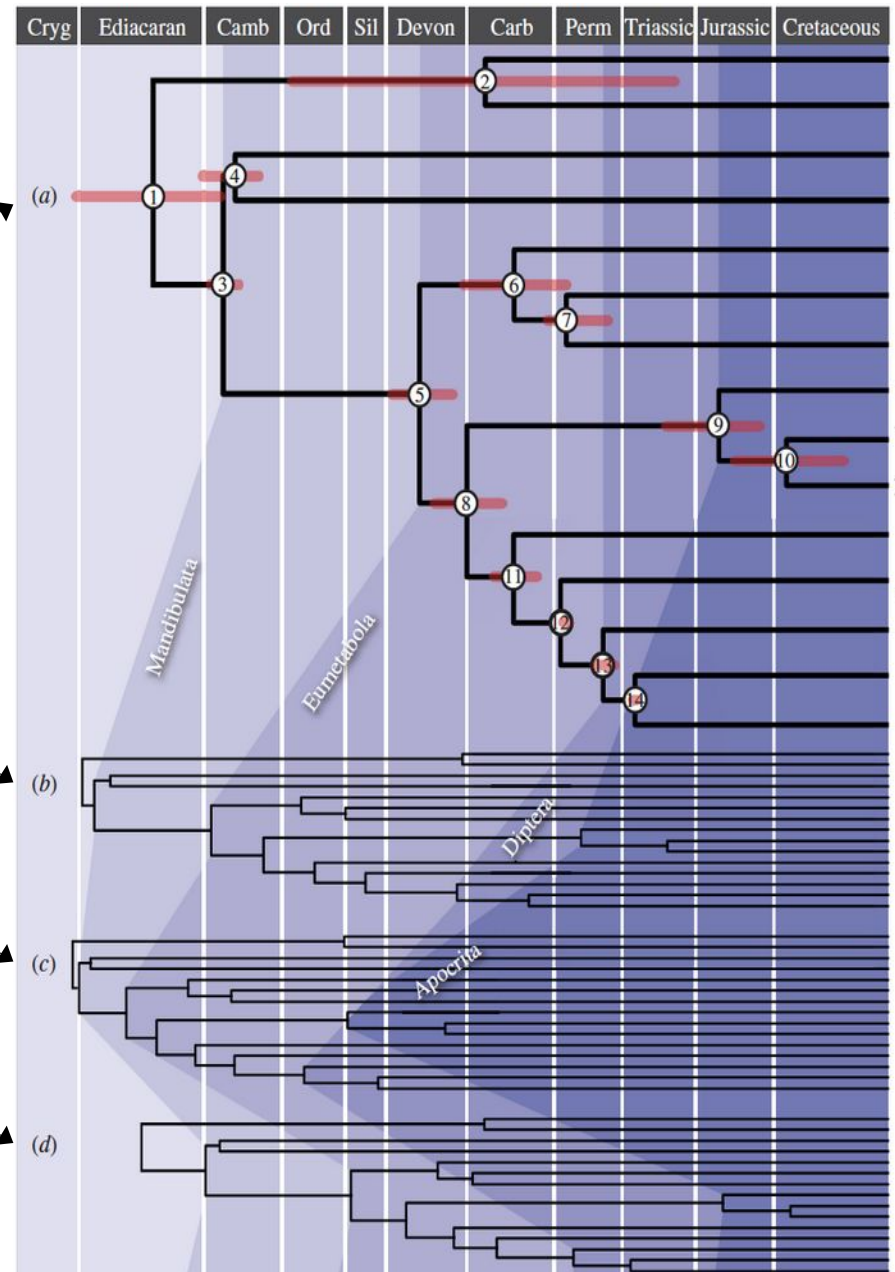


Effective priors may not match those specified!

Issue #2: prior shape



Warnock et al. (2012)

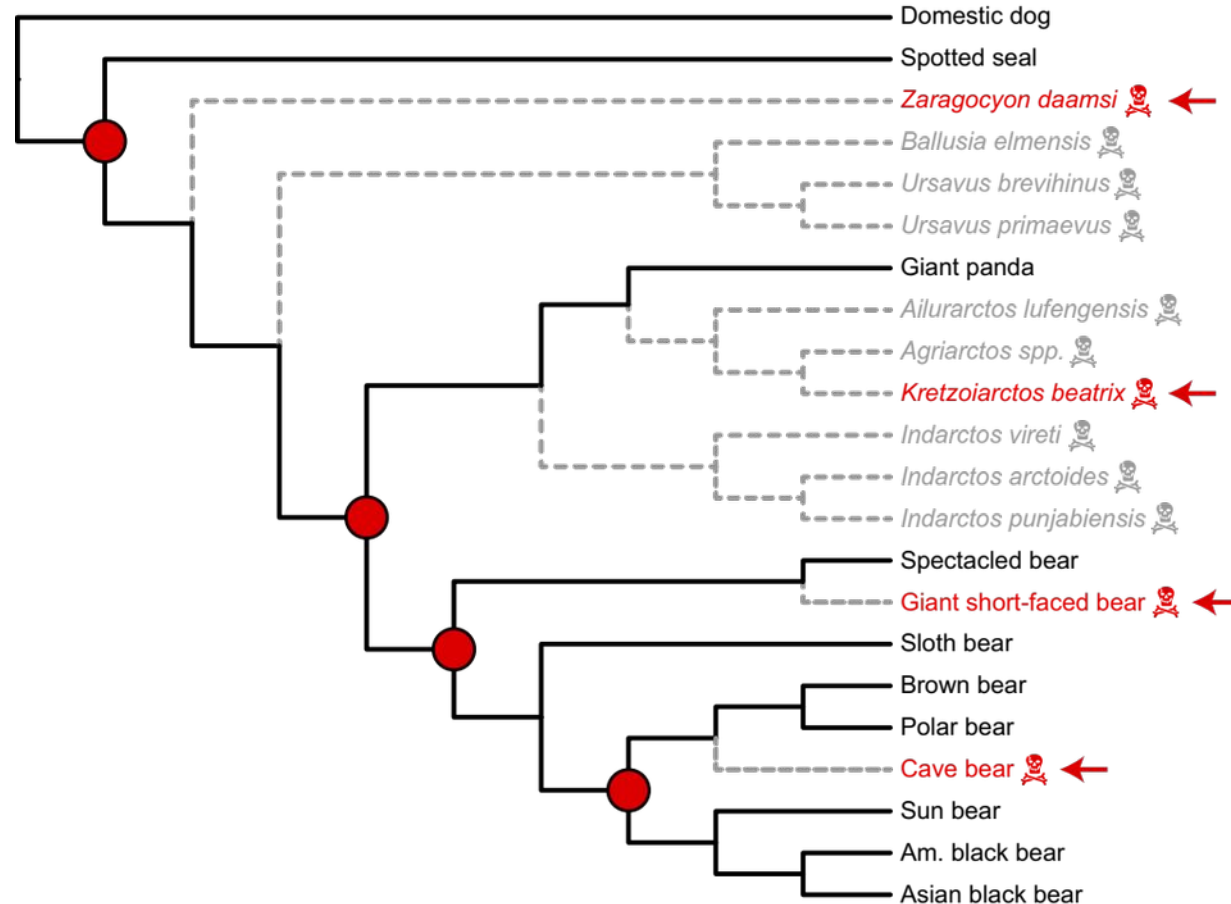


Issue #3: partial use of the fossil record

Node calibration only
uses the oldest fossil
occurrence for each
node

Node calibration is
reliant on correct fossil
placement

Node calibration
cannot use stem fossils

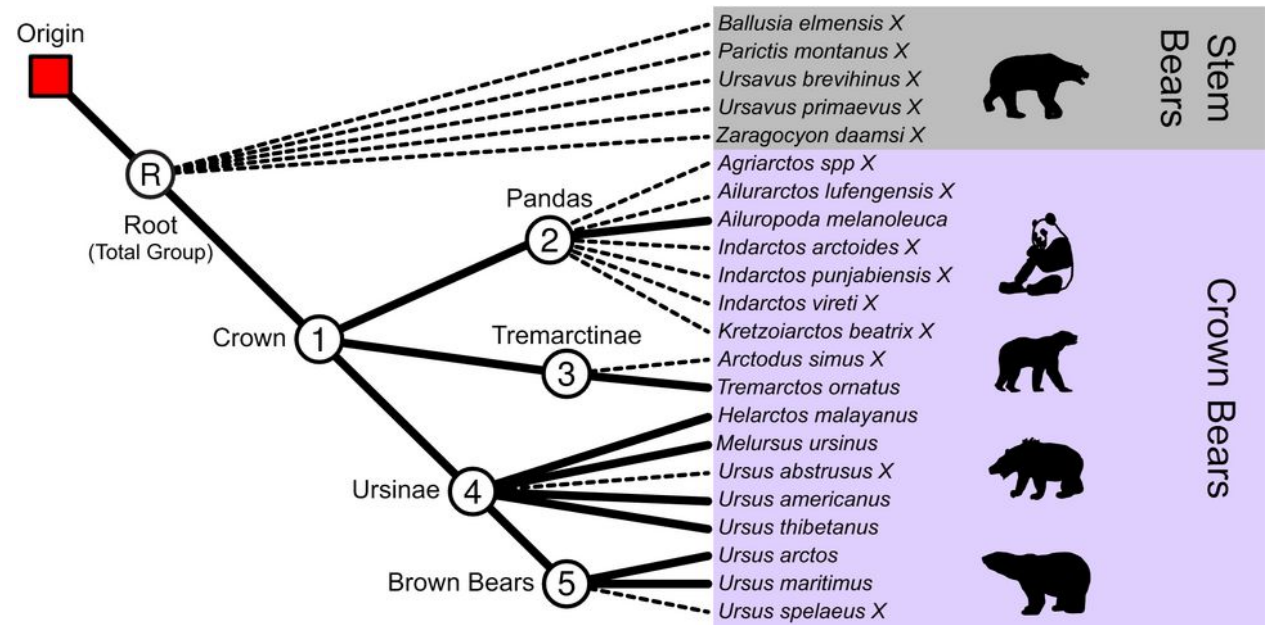


Krause et al. (2008); Abella et al. (2012)

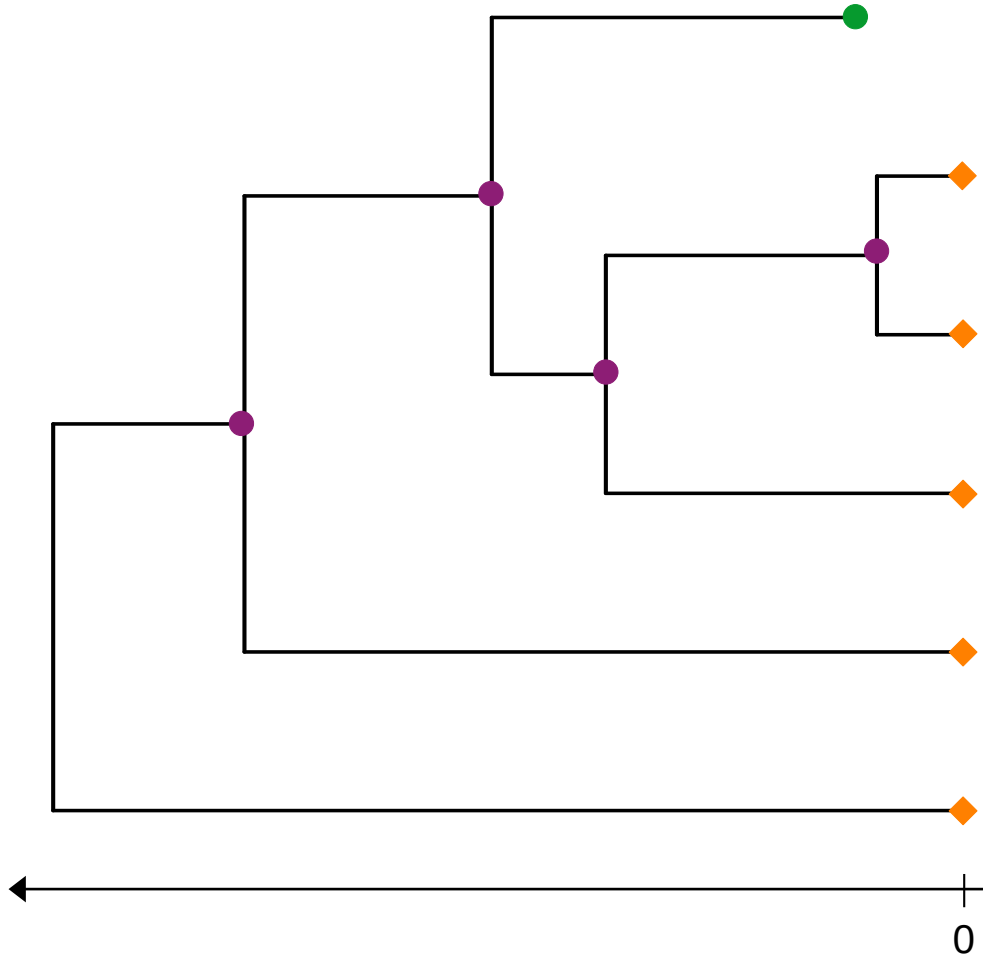
A better solution?

Inserting fossil tips
into the phylogeny

This calibrates the
tree but incorporates
uncertainty in fossil
placement and age



The birth-death model



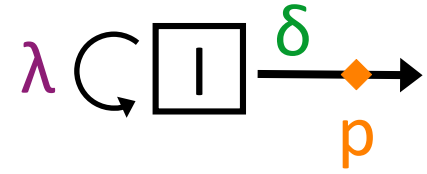
Parameters:

● λ — **birth** rate (= new lineage appearing)

● μ — **death** rate (= lineage disappearing)

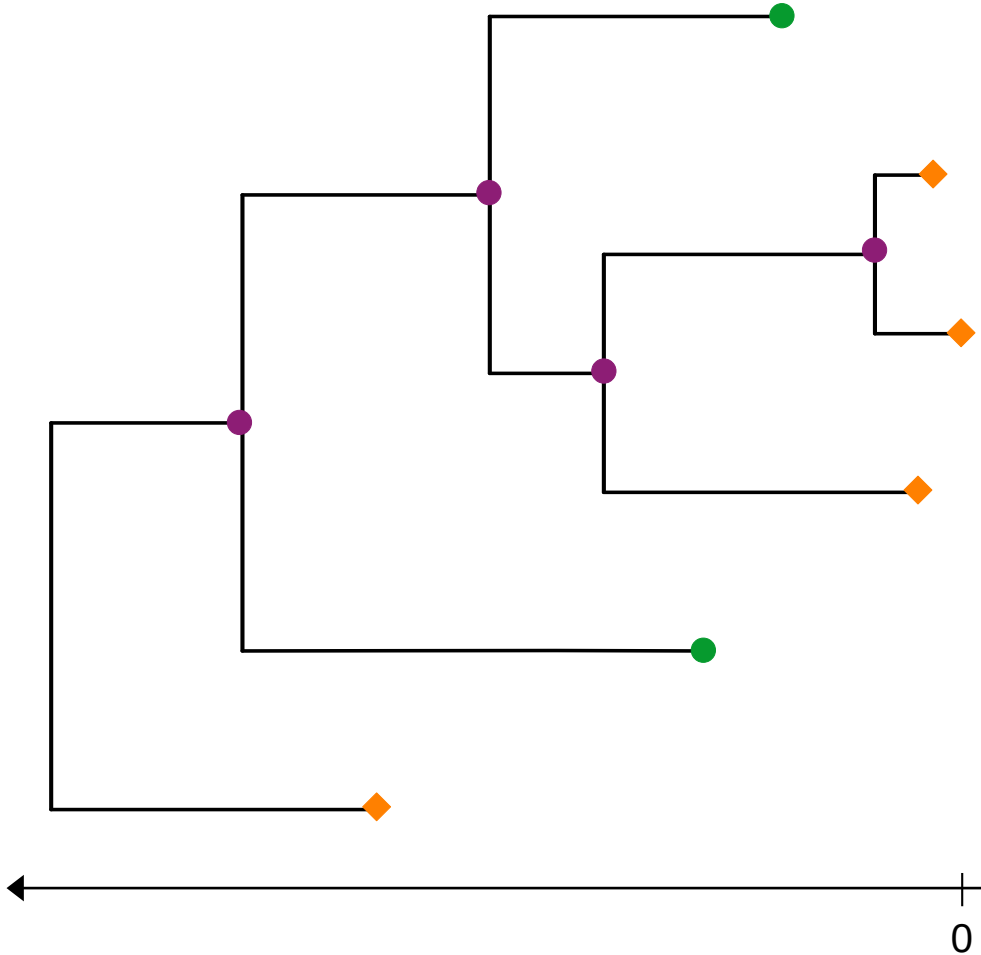
◆ ρ — extant species sampling probability

Birth-death for epidemiology

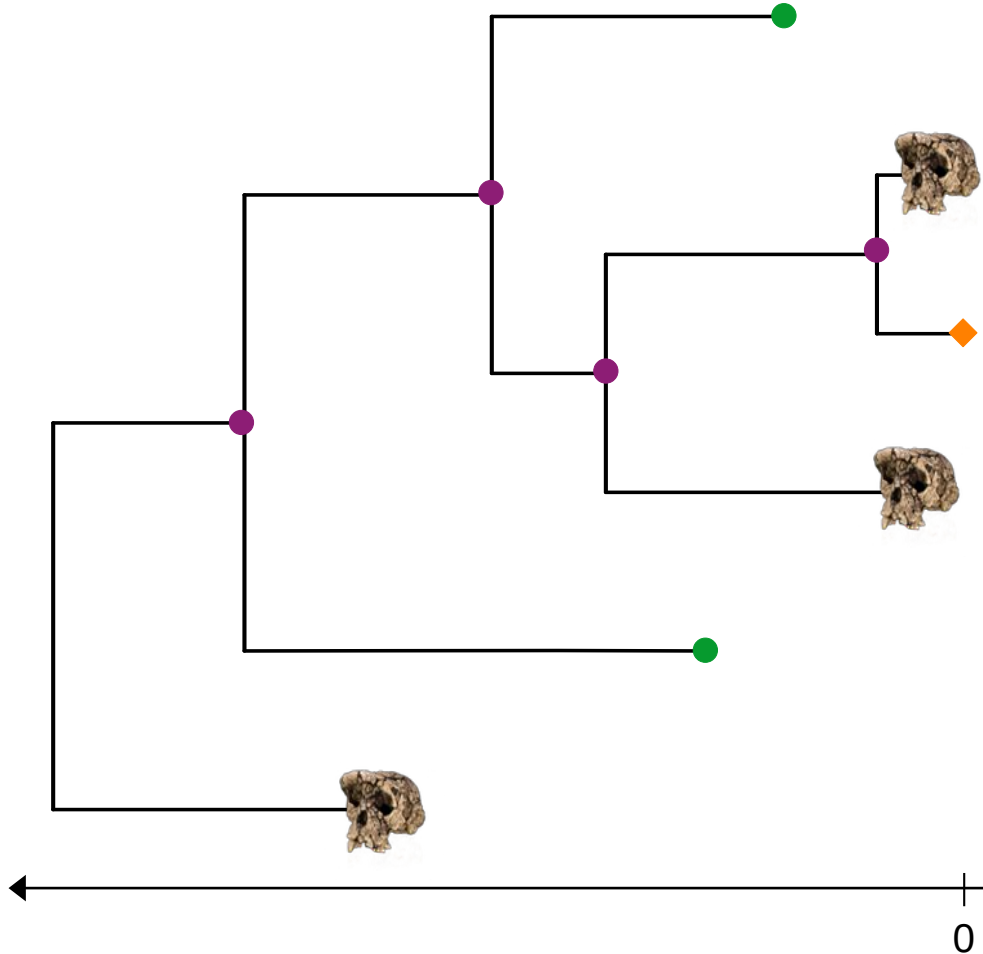


Processes:

- λ — transmission rate
- $\mu = \delta(1-p)$ — rate of removal without sampling
- ◆ $\psi = \delta p$ — rate of removal with sampling



Birth-death for fossils?



Processes:

● λ — birth rate

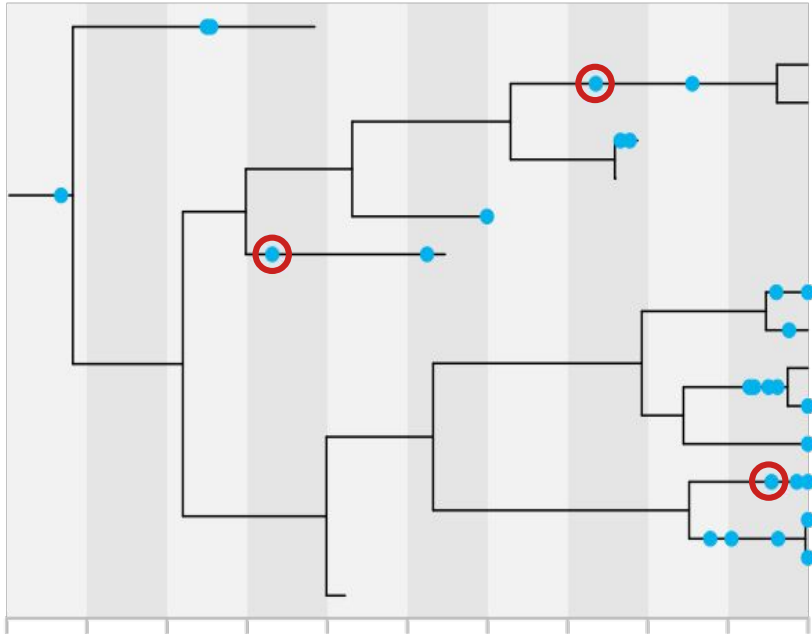
● $\mu = \delta(1-p)$ — rate of death without sampling

● $\psi = \delta p$ — rate of death with sampling

◆ ρ — extant species sampling probability

Sampled ancestors

Samples which are direct ancestors of other samples



Paleobiology, 22(2), 1996, pp. 141–151

On the probability of ancestors in the fossil record

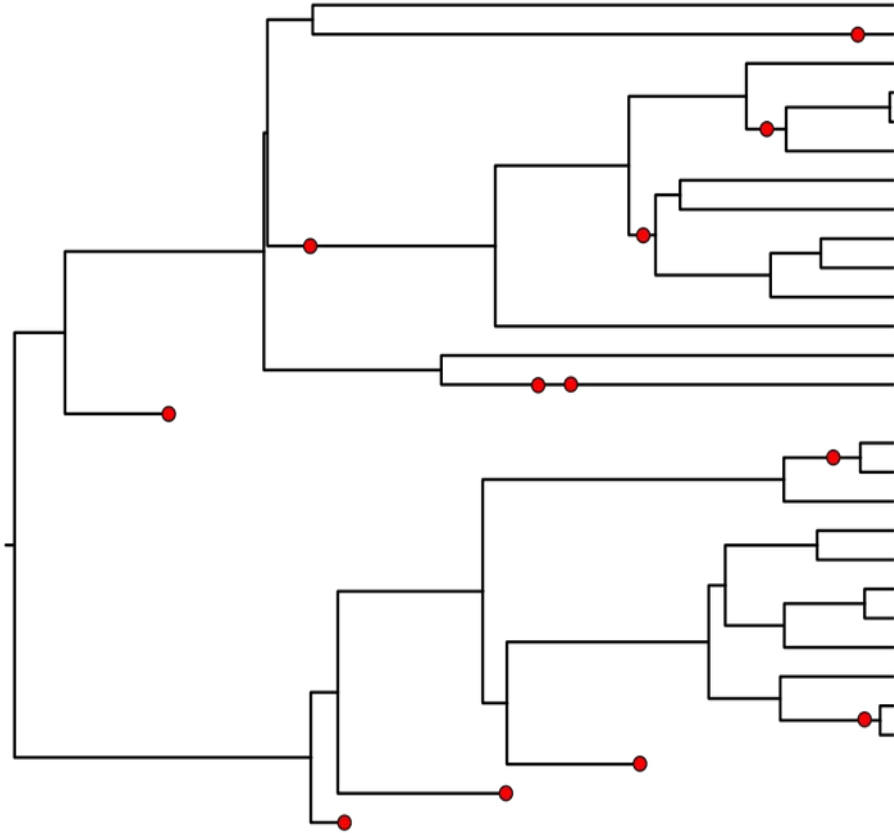
Mike Foote

HIV TRANSMISSIONS IN 2016		
% OF PEOPLE WITH HIV	STATUS OF CARE	ACCOUNTED FOR X% OF NEW TRANSMISSIONS*
15%	didn't know they had HIV	38%
23%	knew they had HIV but weren't in care	43%
11%	in care but not virally suppressed	20%
51%	taking HIV medicine and virally suppressed	0%

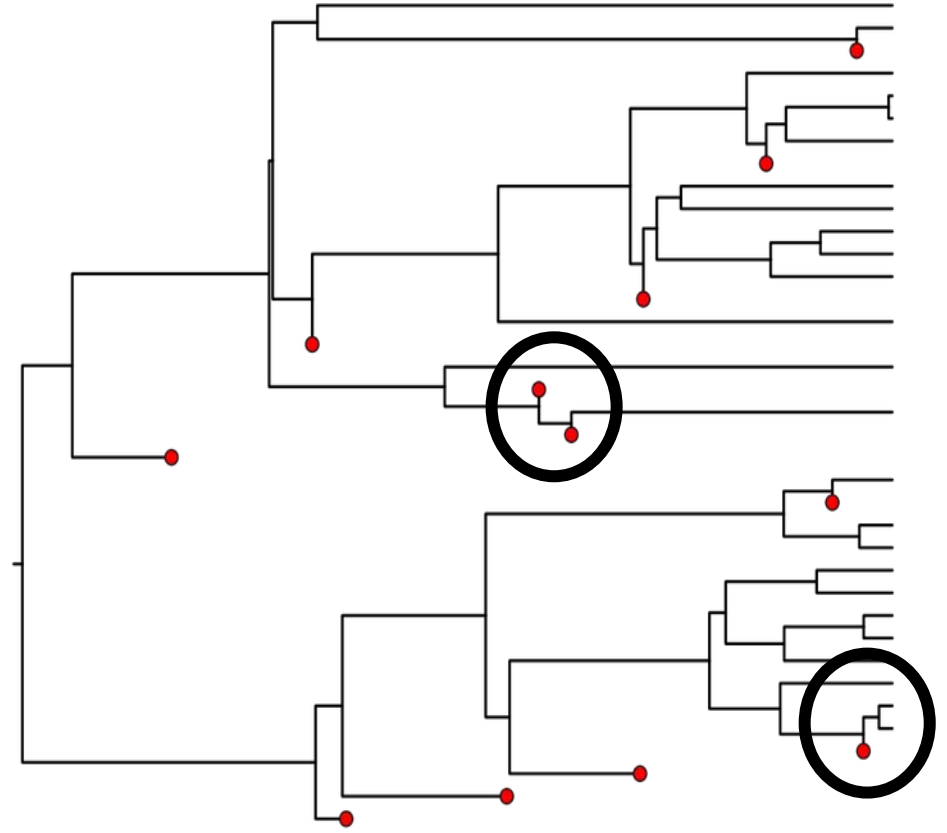
*Values do not equal 100% because of rounding

SOURCE: Vital Signs, 2019

Sampled ancestors affect rate estimates

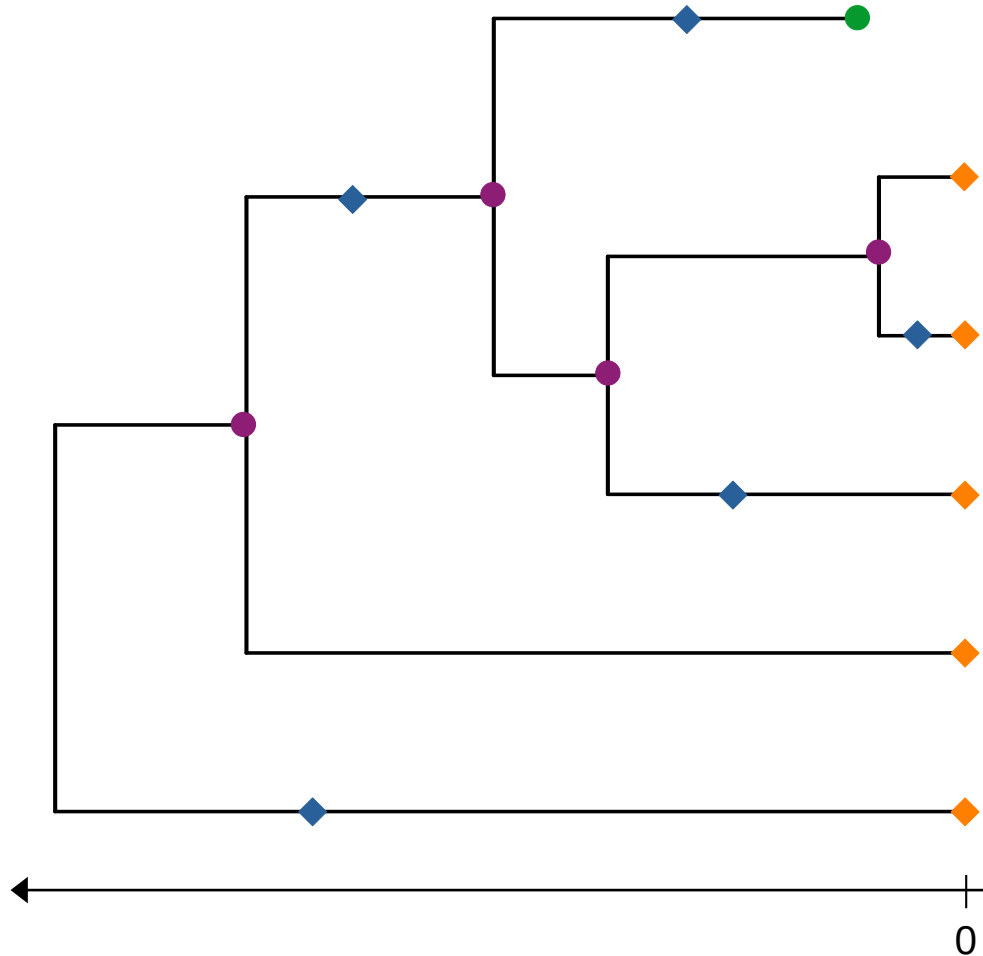


Inference with
sampled ancestors



Inference without
sampled ancestors

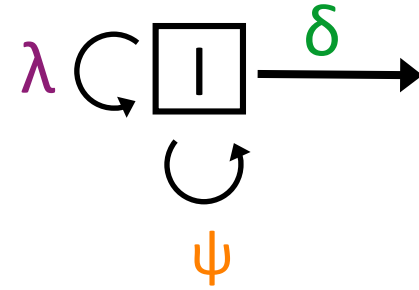
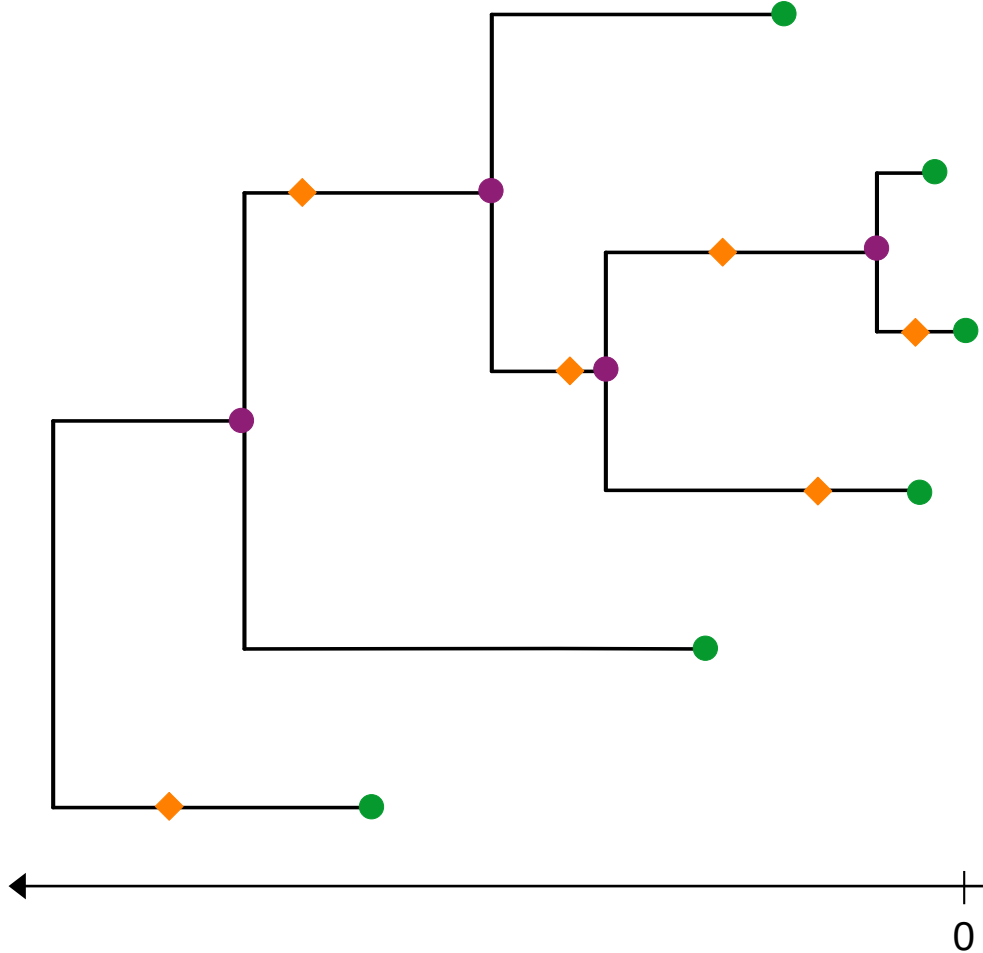
The fossilized birth-death (FBD) model



Parameters:

- λ — birth rate
- μ — death rate
- ◆ ψ — fossil sampling rate
- ◆ ρ — extant species sampling probability

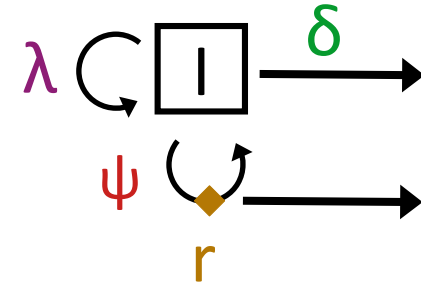
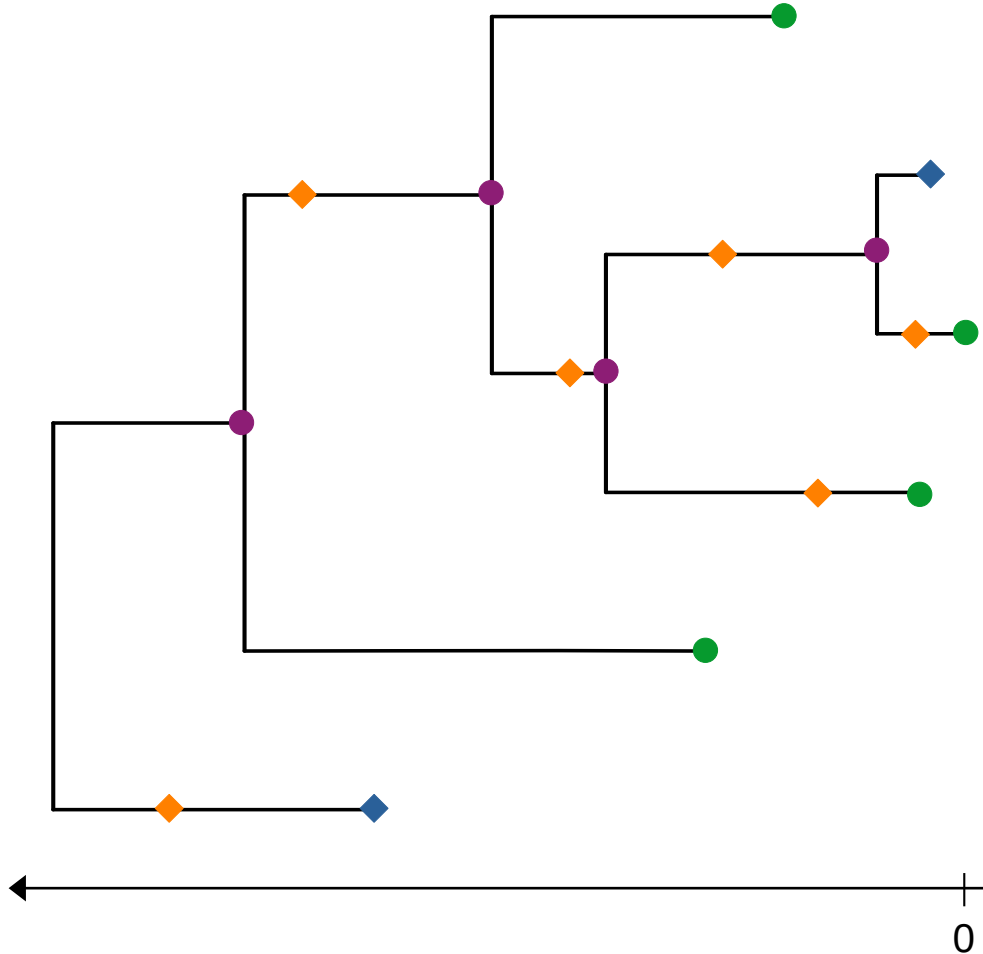
FBD/SA for epidemiology



Parameters:

- λ — transmission rate
- δ — rate of removal
- ◆ ψ — rate of sampling

FBD/SA for epidemiology



Processes:

- λ — transmission rate
- δ — rate of removal without sampling
- ◆ $\psi(1 - r)$ — rate of sampling without removal
- ◆ ψr — rate of sampling with removal

Fossils in Bayesian inference

$$P(\text{[Grid, Clock, Tree, Tree Model]} | \text{[Fossil, Alignment]}) =$$

Posterior

Likelihood

Probability of
the tree model

Priors

$$P(\text{[Alignment]} | \text{[Grid, Clock, Tree]})$$

$$P(\text{[Fossil]} | \text{[Tree, Tree Model]})$$

$$P(\text{[Grid, Clock, Tree, Tree Model]})$$

$$P(\text{[Fossil, Alignment]})$$



Fossil ages

ACAC...
TCAC...
ACAG...

Molecular alignment



Substitution model



Clock model

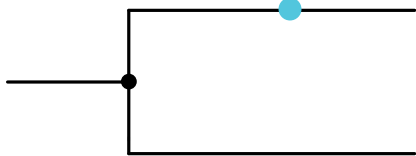
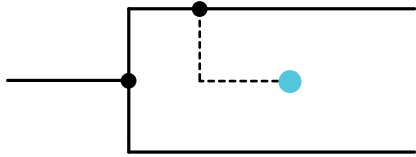
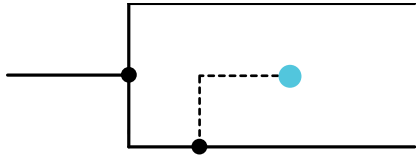


Time tree



Tree model

MCMC implementation



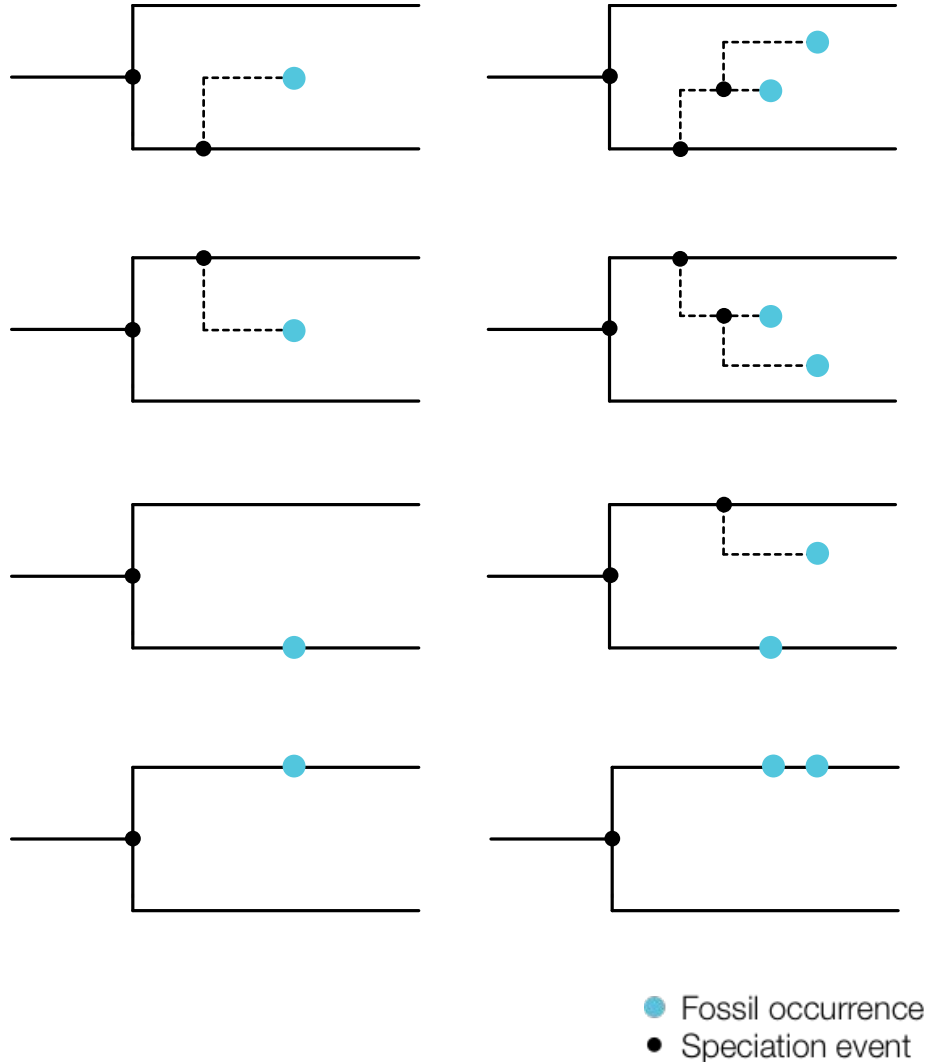
A fossil sample can be placed:

- at the end of an unobserved branch
- as a sampled ancestor on an observed branch

● Fossil occurrence
● Speciation event

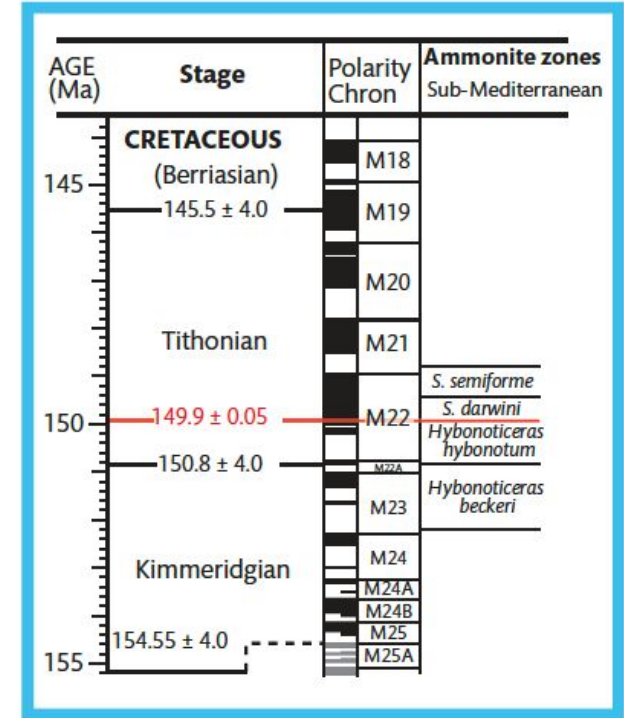
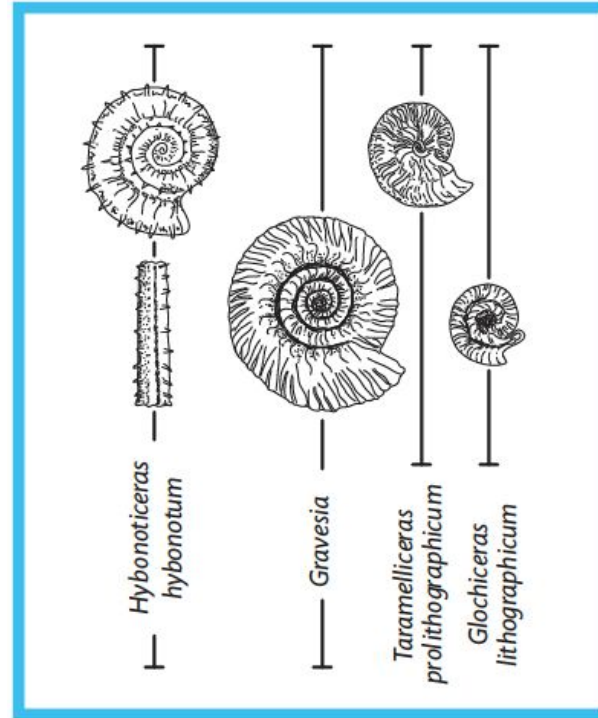
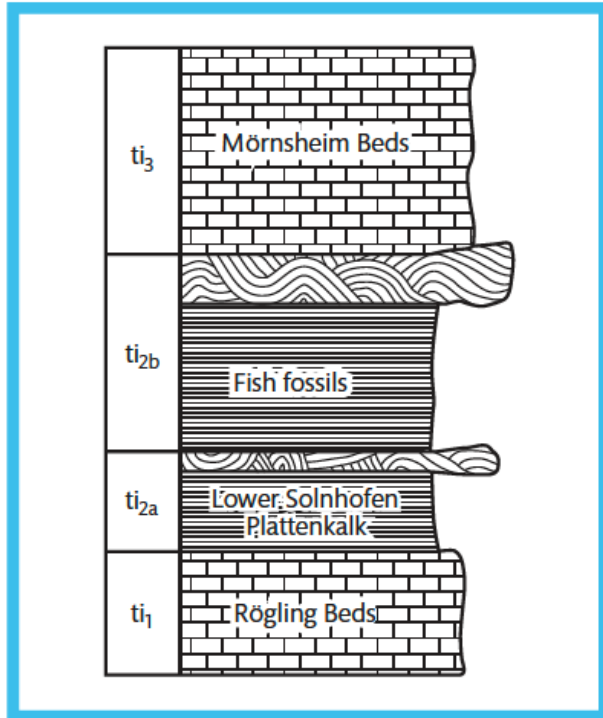
Gavryushkina et al. (2014)

MCMC implementation

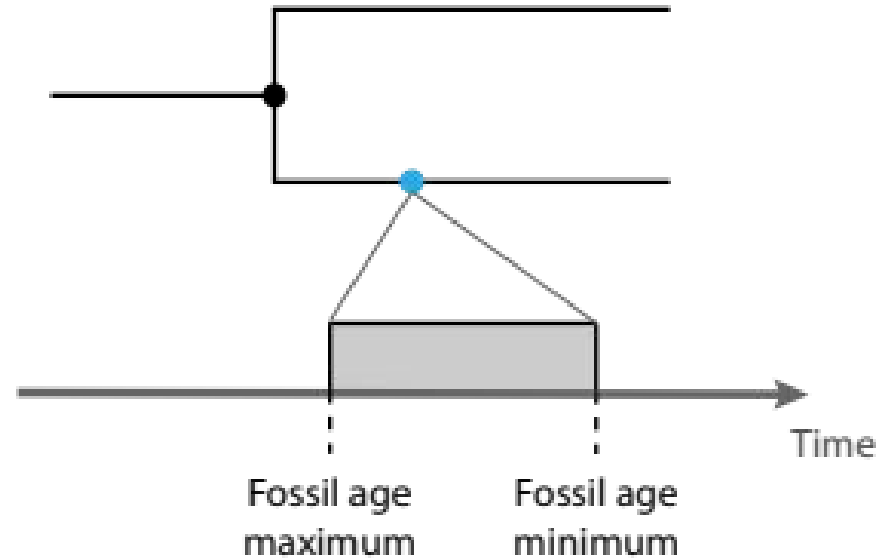
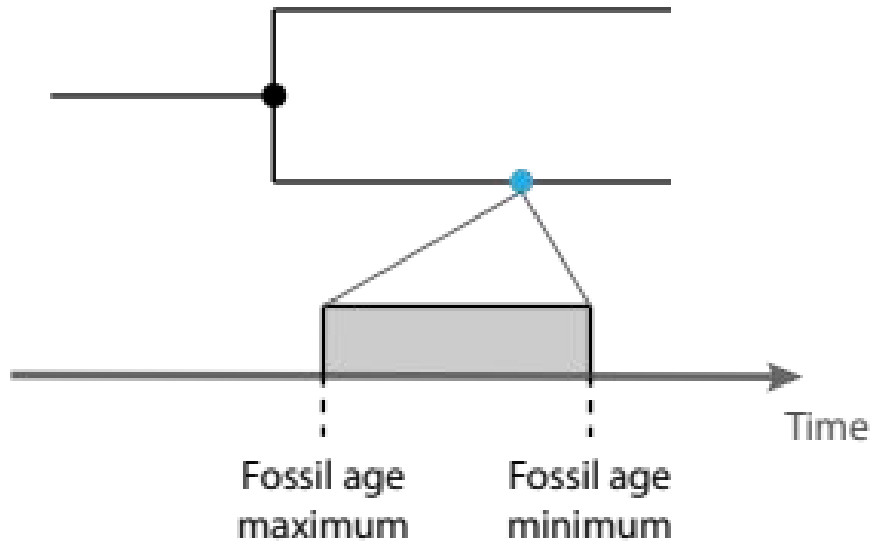


- each fossil can be attached anywhere on the tree, including on unobserved branches
- the FBD process gives the probability of each configuration conditioned on the model parameters

Fossil age uncertainty



Integrating the uncertainty



Fossil age uncertainty can be
sampled as part of the MCMC

Age uncertainty in Bayesian inference

$$P(\text{Genetic Data} \mid \text{Tree Model} \mid \text{Priors} \mid \text{Fossil Age Ranges}) =$$

Posterior

Likelihood

Probability of
the tree model

Priors

$$P(\text{ACAC... TCAC... ACAG...} \mid \text{Genetic Data} \mid \text{Tree Model} \mid \text{Priors}) \times P(\text{Tree Model} \mid \text{Fossil Age Ranges}) \times P(\text{Genetic Data} \mid \text{Tree Model} \mid \text{Priors})$$

$$P(\text{Fossil Age Ranges} \mid \text{ACAC... TCAC... ACAG...})$$



Fossil age ranges

Integrating the uncertainty

Ignoring stratigraphic age uncertainty
leads to erroneous estimates of species
divergence times under the fossilized
birth – death process

Joëlle Barido-Sottani^{1,2,3}, Gabriel Aguirre-Fernández⁴, Melanie Hopkins⁵,
Tanja Stadler^{1,2} and Rachel Warnock^{1,2,4}

Fossil age uncertainty **should** be
sampled as part of the MCMC

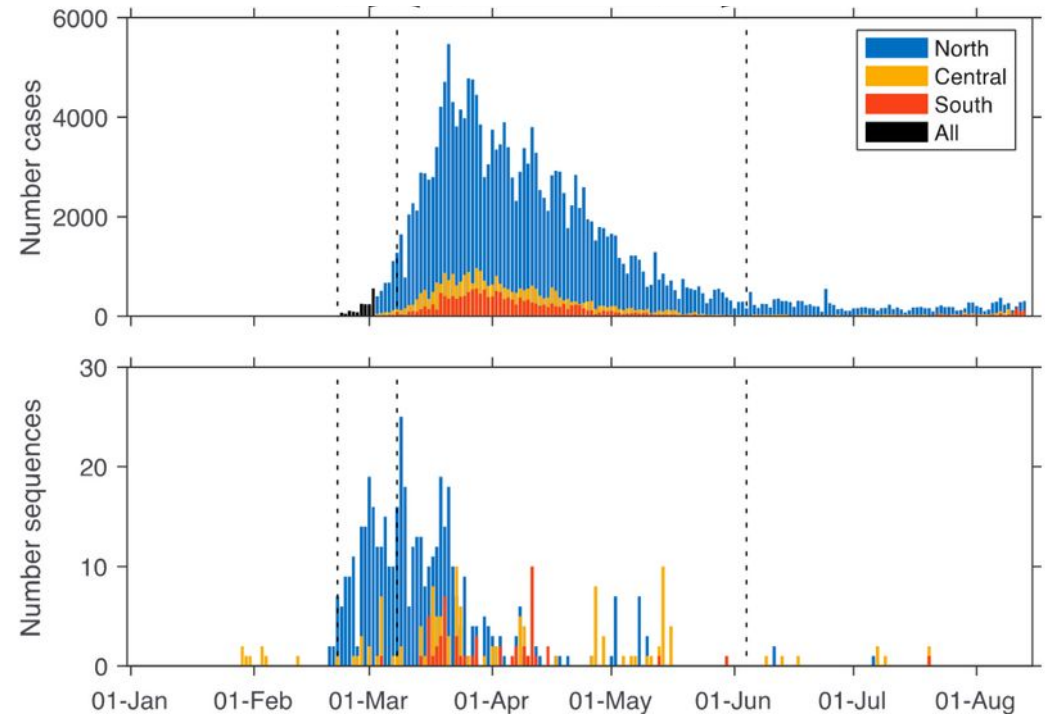
Integrating occurrences

Occurrence = sample without associated sequence information

- Database of fossil specimens
- Number of confirmed cases



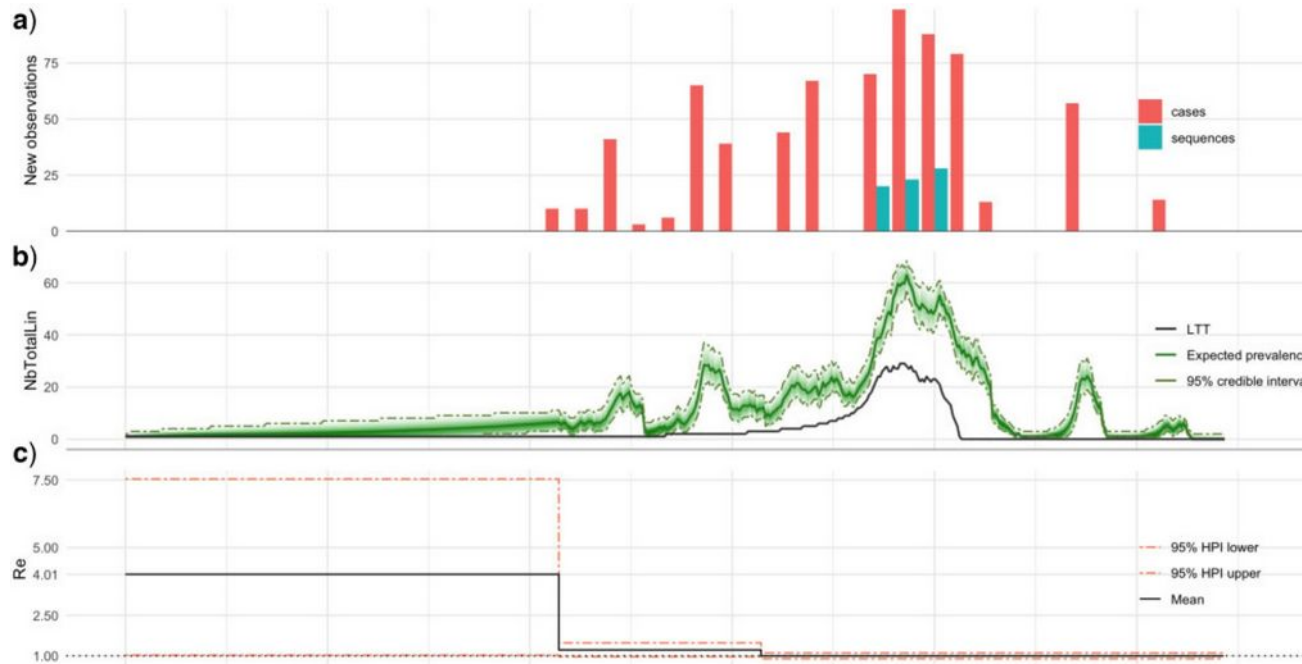
The Paleobiology Database
revealing the history of life



Di Giallonardo et al. (2020)

Occurrence birth-death process (OBDP)

Using occurrences to infer the total number of lineages through time and inform the population parameter estimates



TimTam

Gupta et al. (2020)

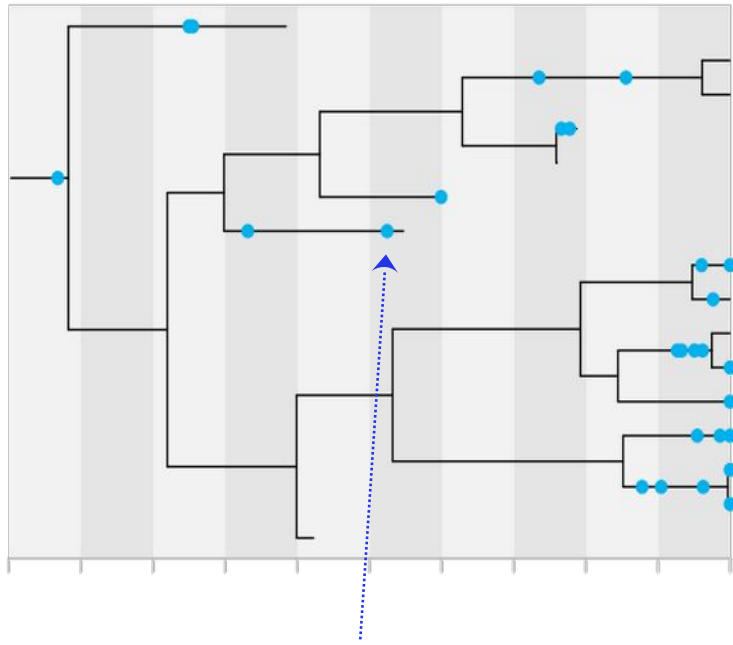
Manceau et al. (2021)

Andréoletti et al. (2022)

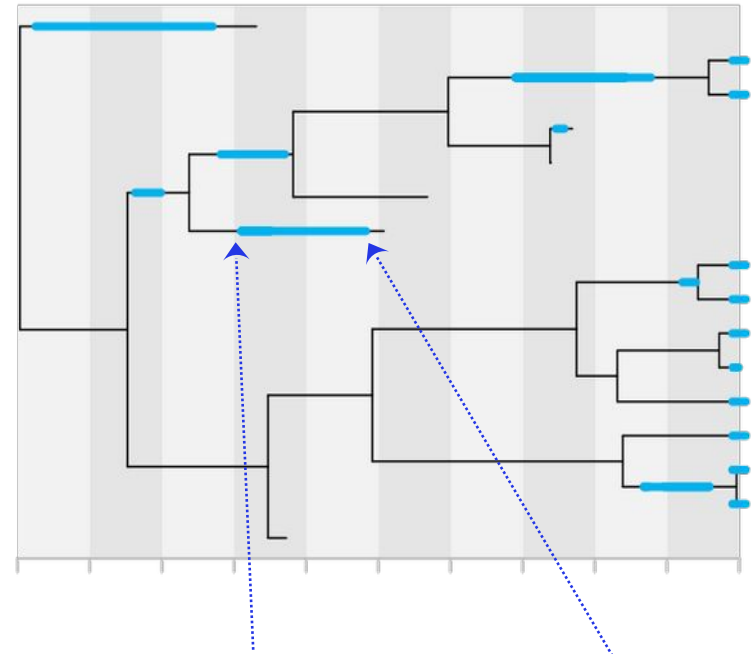
Zarebski et al. (2023)

Inferred LTT & R_e for the Diamond Princess outbreak
(Feb-Mar 2020)

Specimen-level data vs. range data



Specimen (one occurrence)



First occurrence

Last occurrence

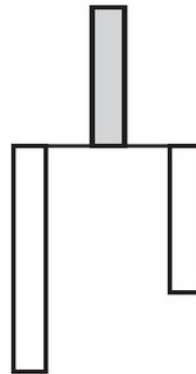
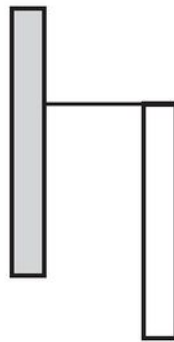
The FBD for stratigraphic ranges

The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes

Tanja Stadler^{a,b,*}, Alexandra Gavryushkina^{a,b}, Rachel C.M. Warnock^{a,b},
Alexei J. Drummond^c, Tracy A. Heath^d

Journal of Theoretical Biology 447:41-55 (2018)

(i) asymmetric speciation (ii) symmetric speciation (iii) anagenetic speciation



Available in RevBayes – Coming soon to BEAST2

The FBD Range process


Paleobiology, 2020, pp. 1–21
DOI: 10.1017/pab.2020.12

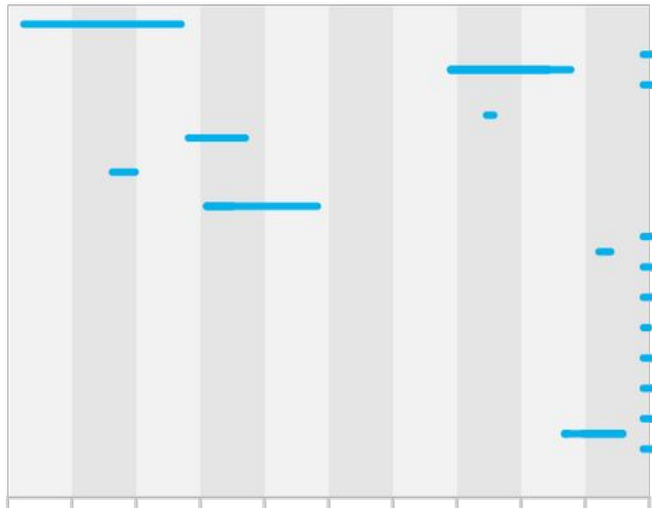
PALEOBIOLOGY
A PUBLICATION OF THE



Article

Assessing the impact of incomplete species sampling on estimates of speciation and extinction rates

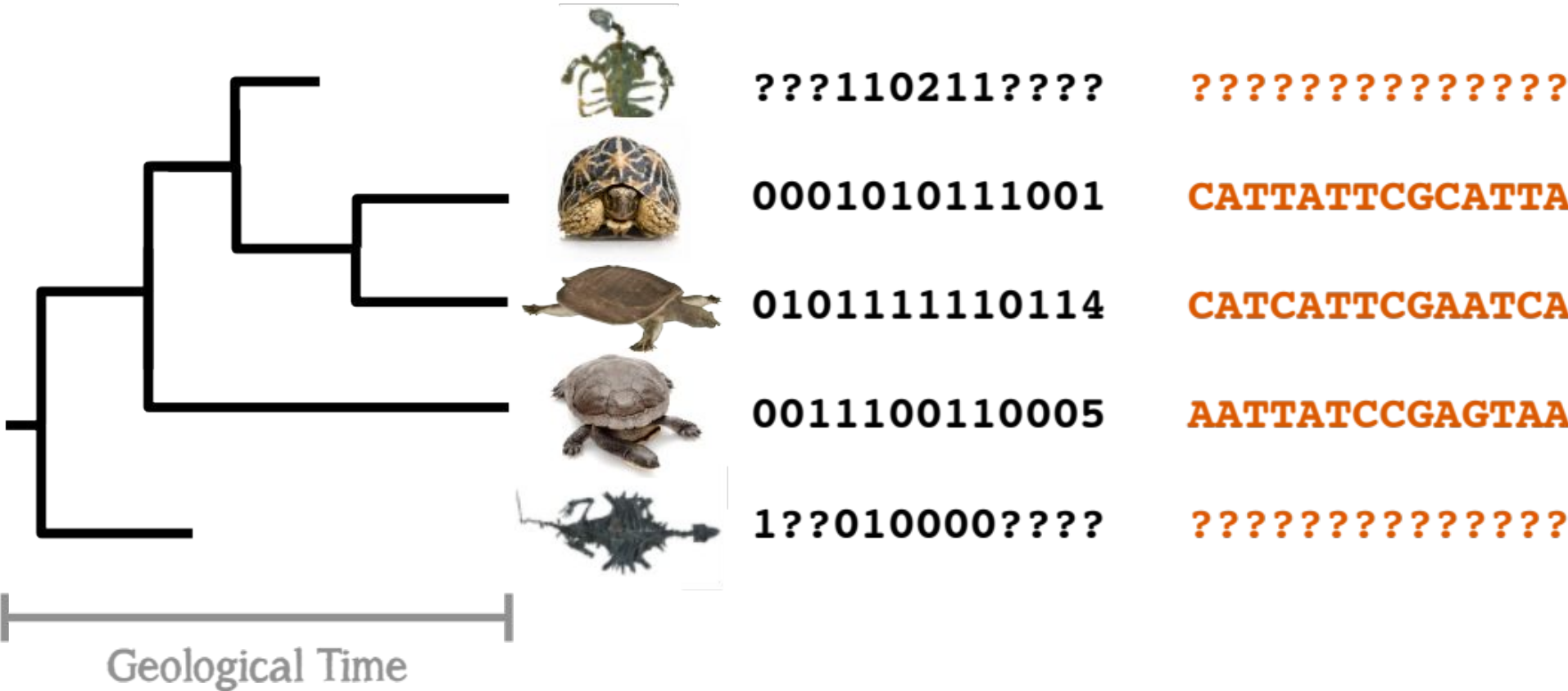
Rachel C. M. Warnock , Tracy A. Heath, and Tanja Stadler



No phylogeny: can be used
without molecular data

Only available in RevBayes
at the moment

Using morphological data



Total-evidence analysis

$$P(\text{Molecular data, Morphological data, Tree model} \mid \text{Skull}) =$$

Posterior

Molecular
likelihood

Morphological
likelihood

Probability of
the tree model

$$P(\text{ACAC... TCAC... ACAG...} \mid \text{Molecular data, Tree model})$$

$$P(\text{1001... 1101... 0100...} \mid \text{Morphological data, Tree model})$$

$$P(\text{Skull} \mid \text{Tree model})$$

$$P(\text{Molecular data, Morphological data} \mid \text{Tree model})$$

Priors

$$P(\text{Skull} \mid \text{ACAC... TCAC... ACAG...} \mid \text{1001... 1101... 0100...})$$

Mk Lewis substitution model

$$Q = \alpha \begin{bmatrix} 1-k & 1 & \dots & 1 \\ 1 & 1-k & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1-k \end{bmatrix}$$

Lewis 2001

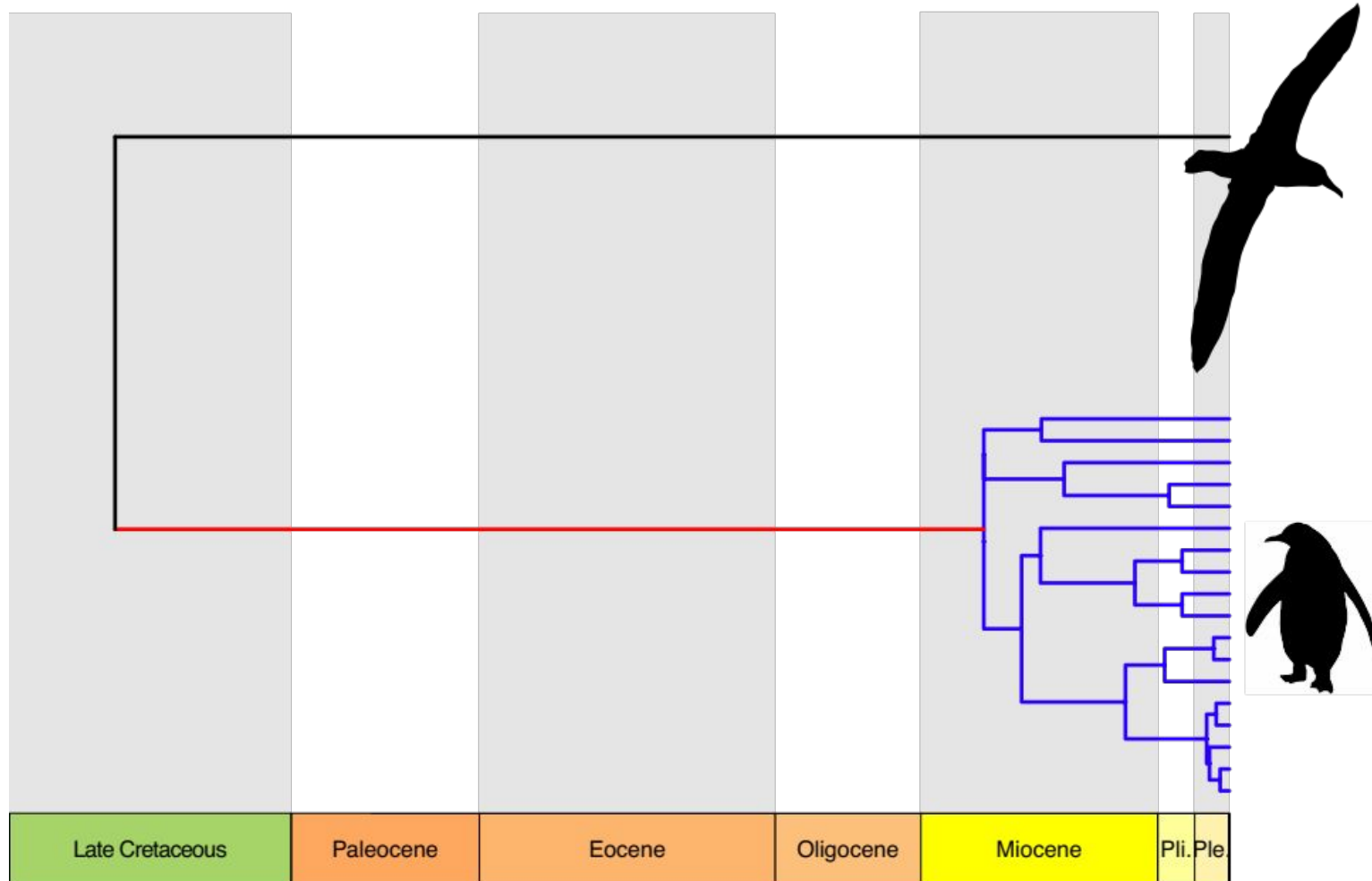
Adaptation of the Jukes-Cantor model

k possible states for each character

Equal transition rates between all states

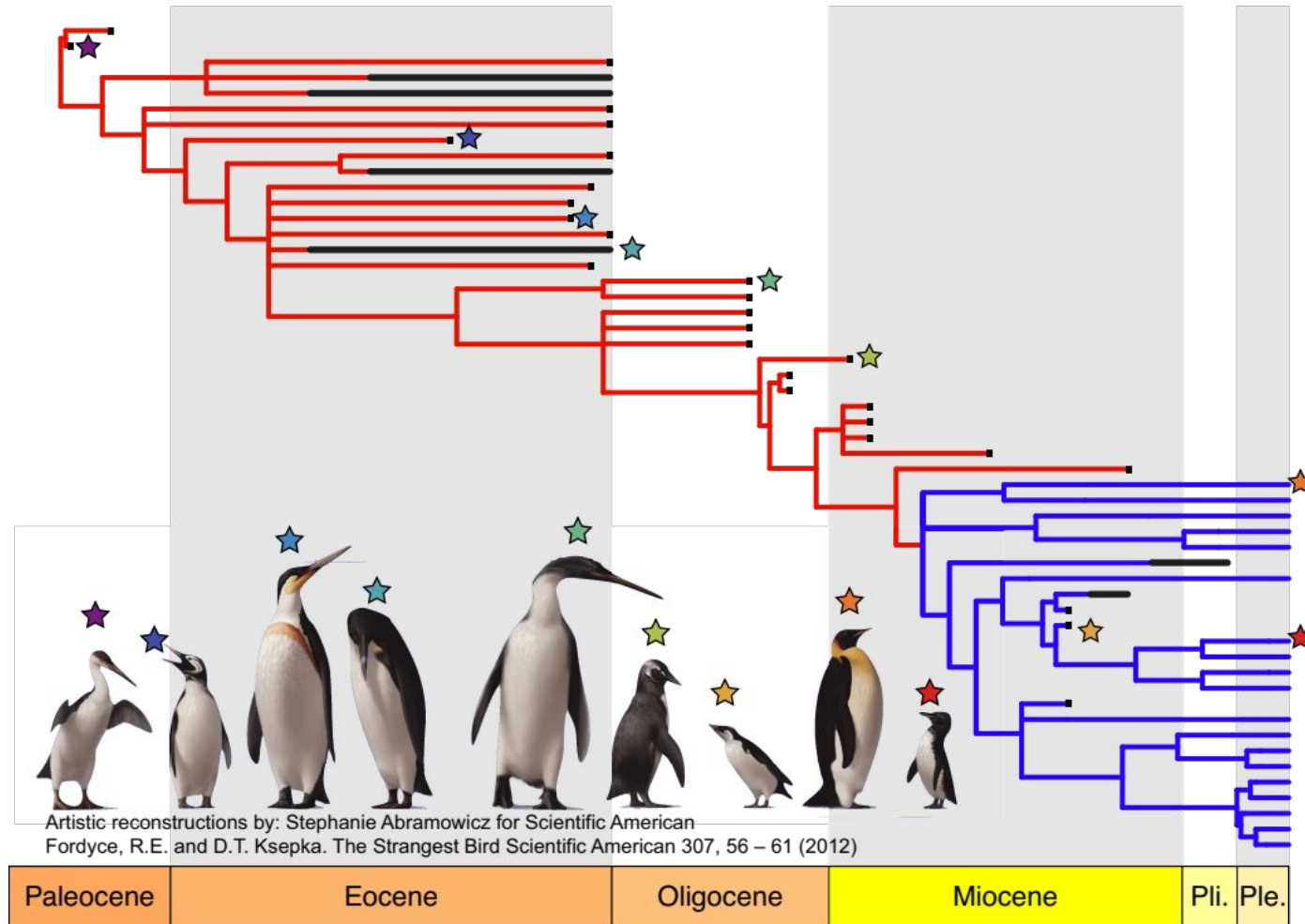
Usually assumes that all possible states are represented in the dataset

Example: dating extant penguins

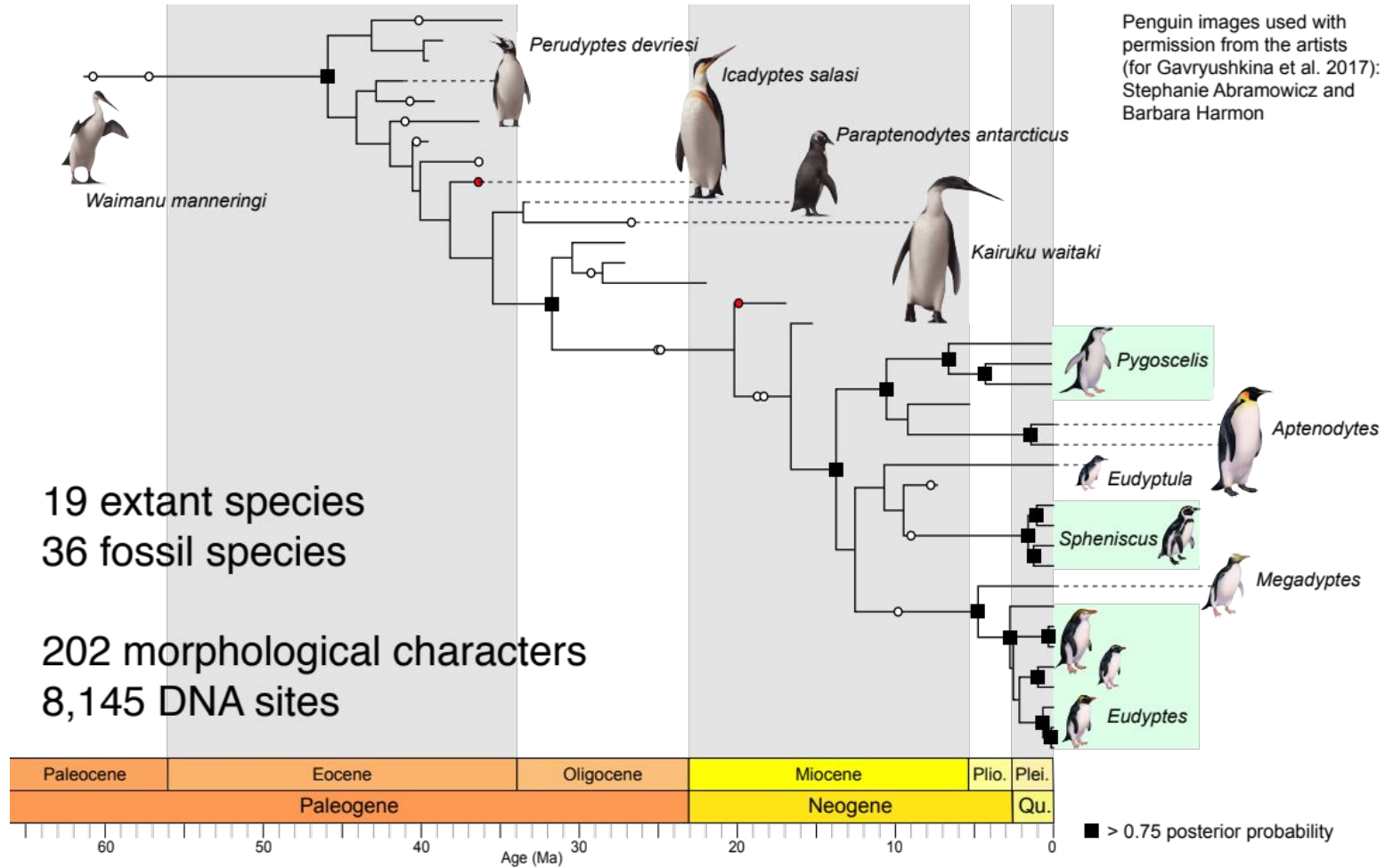


Gavryushkina et al. (2017)

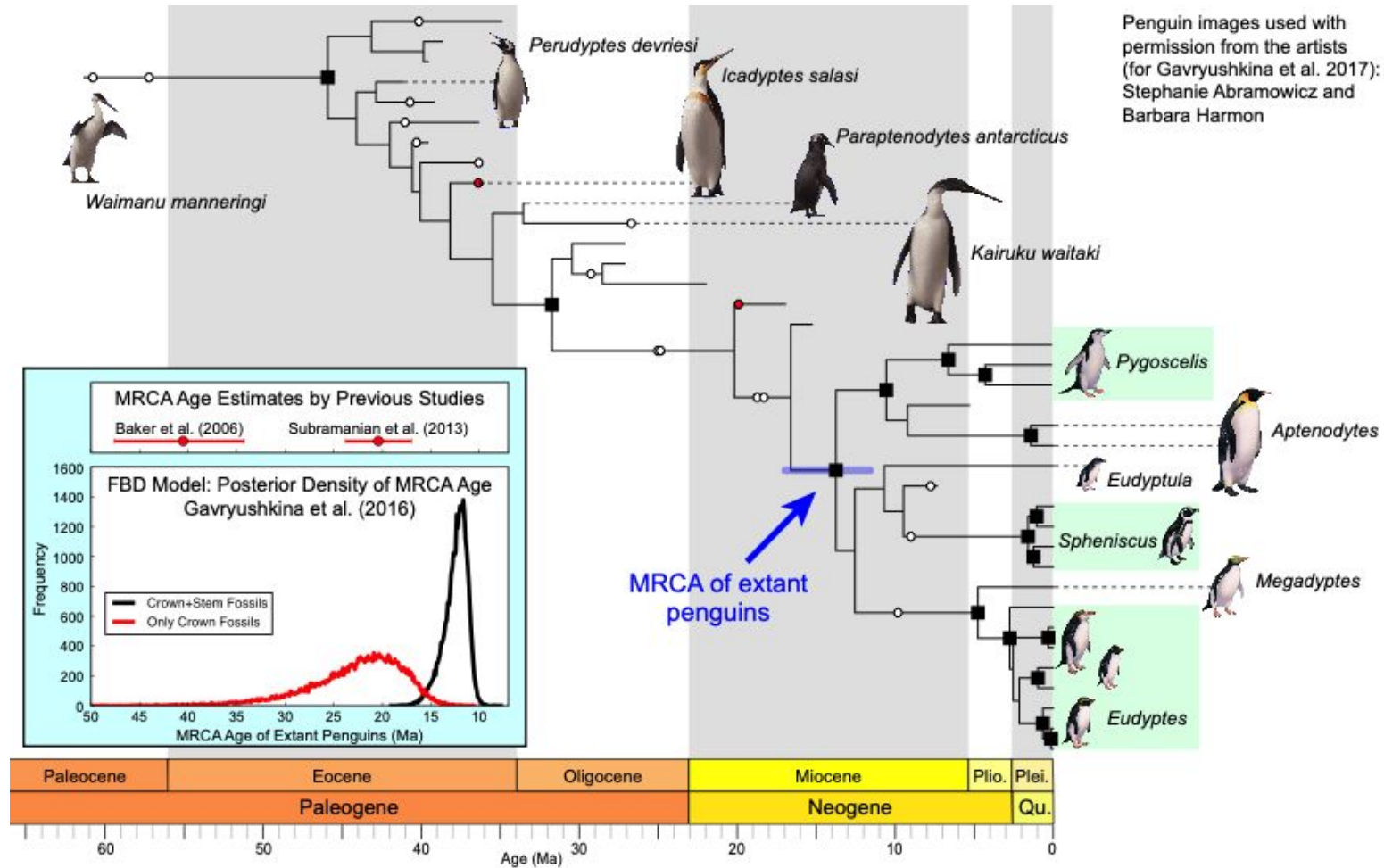
Example: dating extant penguins



Example: dating extant penguins



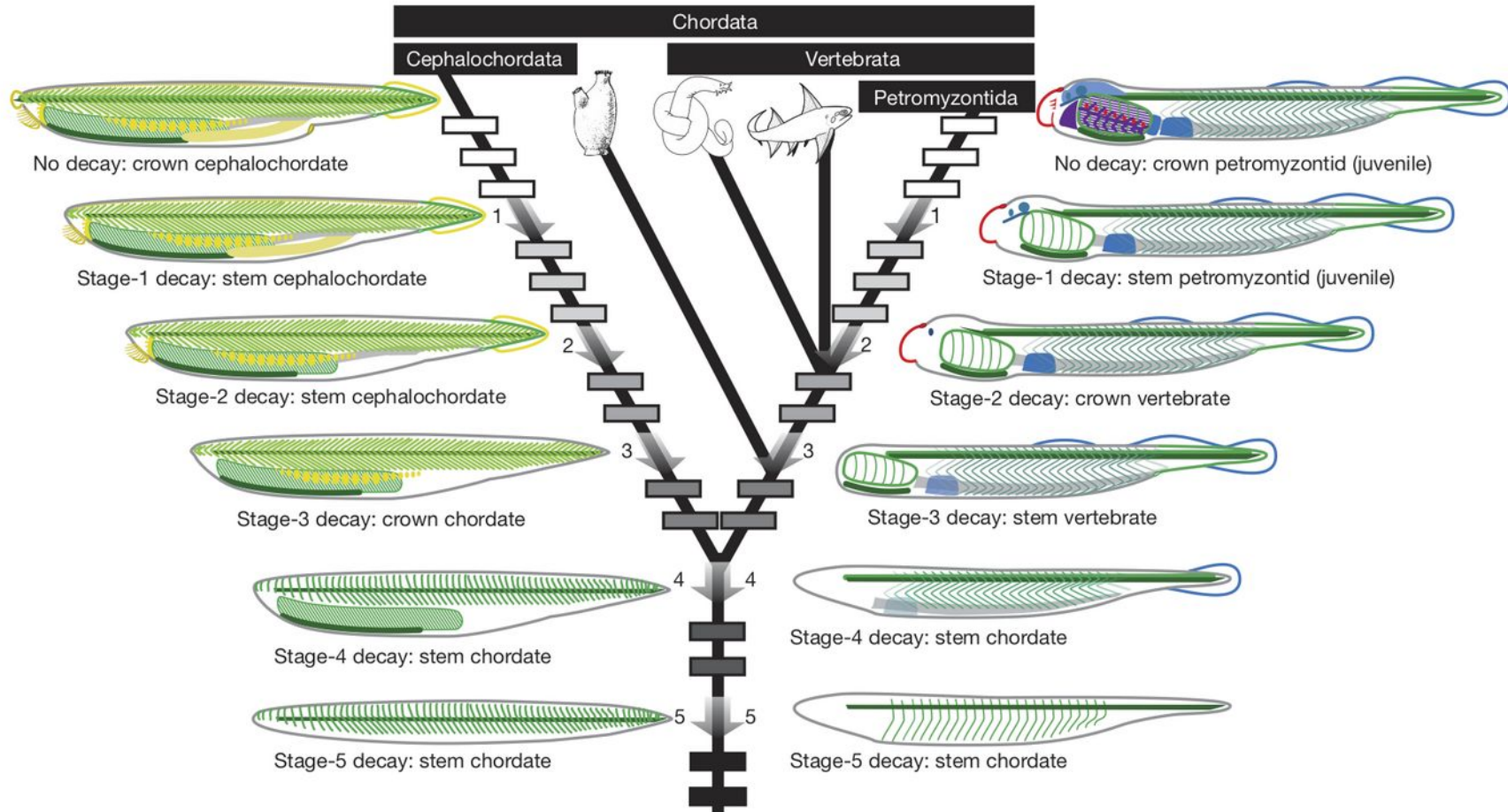
Example: dating extant penguins



Challenge #1: substitution model

- Partitioning by number of characters: automatic in BEAUti
- Ordered characters: available in BEAST2 through manual XML editing
- Partitioning by character similarity: R package EvoPhylo (Simões et al. 2023)
- Many more:
 - Adding frequencies to the model (Wright et al. 2016)
 - Partitioning based on anatomy (Casali et al. 2022)

Challenge #2: fossil decay



Sansom et al. (2010)

Tutorial time



Taming the BEAST [news](#) [workshops](#) [tutorials](#) [contribute](#) 

Divergence Time Estimation using BEAST v2.x



Dating Species Divergences with the Fossilized Birth-Death Process

by Tracy A. Heath

Tutorial

-  FBD-tutorial.pdf
-  Github repository
-  License
-  Statistics

Data

-  bears_irbp_fossils.nex
-  bears_cytb_fossils.nex

Central among the questions explored in biology are those that seek to understand the timing and rates of evolutionary processes. Accurate estimates of species divergence times are vital to understanding historical biogeography, estimating diversification rates, and identifying the causes of variation in rates of molecular evolution.

This tutorial will provide a general overview of divergence time estimation and fossil calibration using a stochastic branching process and relaxed-clock model in a Bayesian framework. The exercise will guide you through the steps necessary for estimating phylogenetic relationships and dating species divergences using the program BEAST v2.x.

<https://taming-the-beast.org/tutorials/FBD-tutorial/>