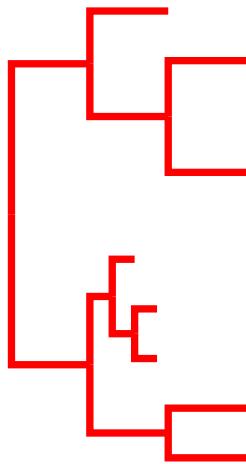


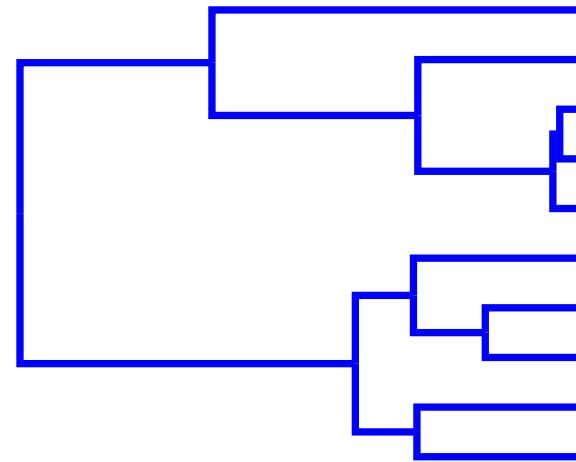
Divergence time estimation & the Fossilized Birth-Death process

Joëlle Barido-Sottani
Iowa State University

Divergence time estimation

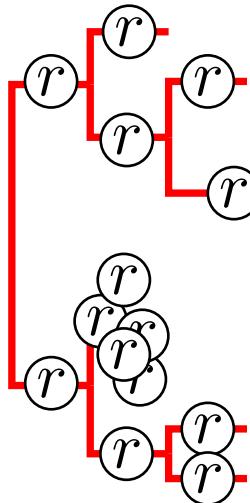


length = substitutions

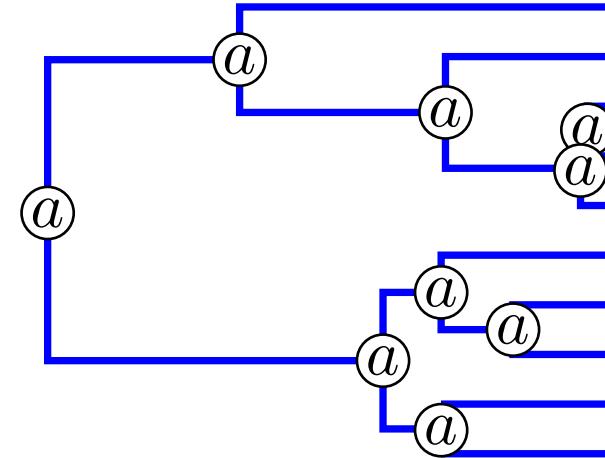


length = time

Divergence time estimation



length = substitutions



length = time

$$R = (r_1, r_2, r_3, \dots, r_{2N-2})$$

$$A = (a_1, a_2, a_3, \dots, a_{N-1})$$

N = number of tips

Bayesian divergence time estimation

$$P(\text{Sequence data} \mid \text{Substitution model}, \text{Clock model}, \text{Time tree}, \text{Tree prior}) = \frac{P(\text{Sequence data} \mid \text{Substitution model}, \text{Clock model}, \text{Time tree}, \text{Tree prior})}{P(\text{Sequence data})}$$

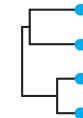
ACAC...
TCAC...
ACAG...



Sequence
data



Substitution
model



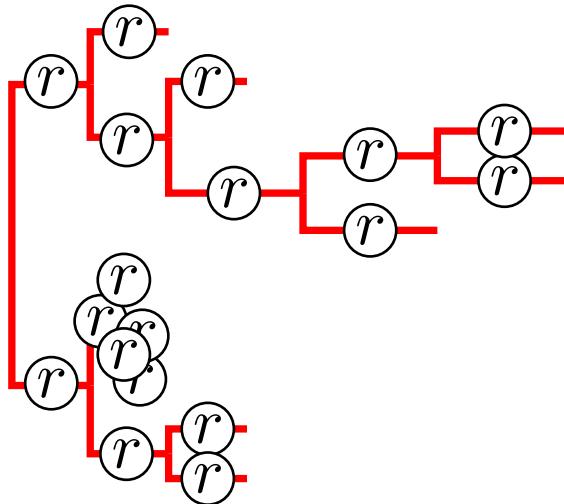
Time tree



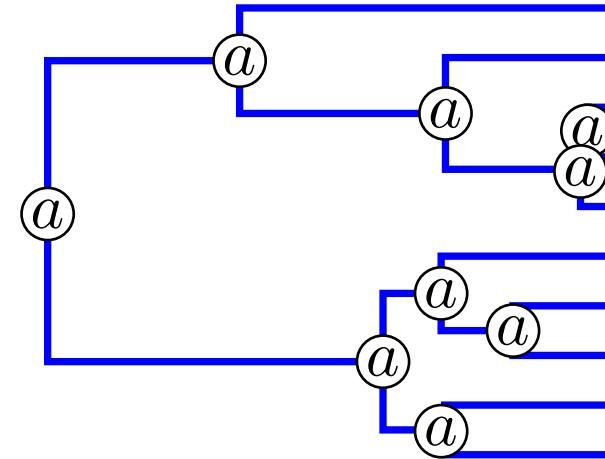
Tree prior

Adapted from du Plessis,
Taming the BEAST 2016

Divergence time estimation



length = substitutions



length = time

$$R = (r_1, r_2, r_3, \dots, r_{2N-2})$$

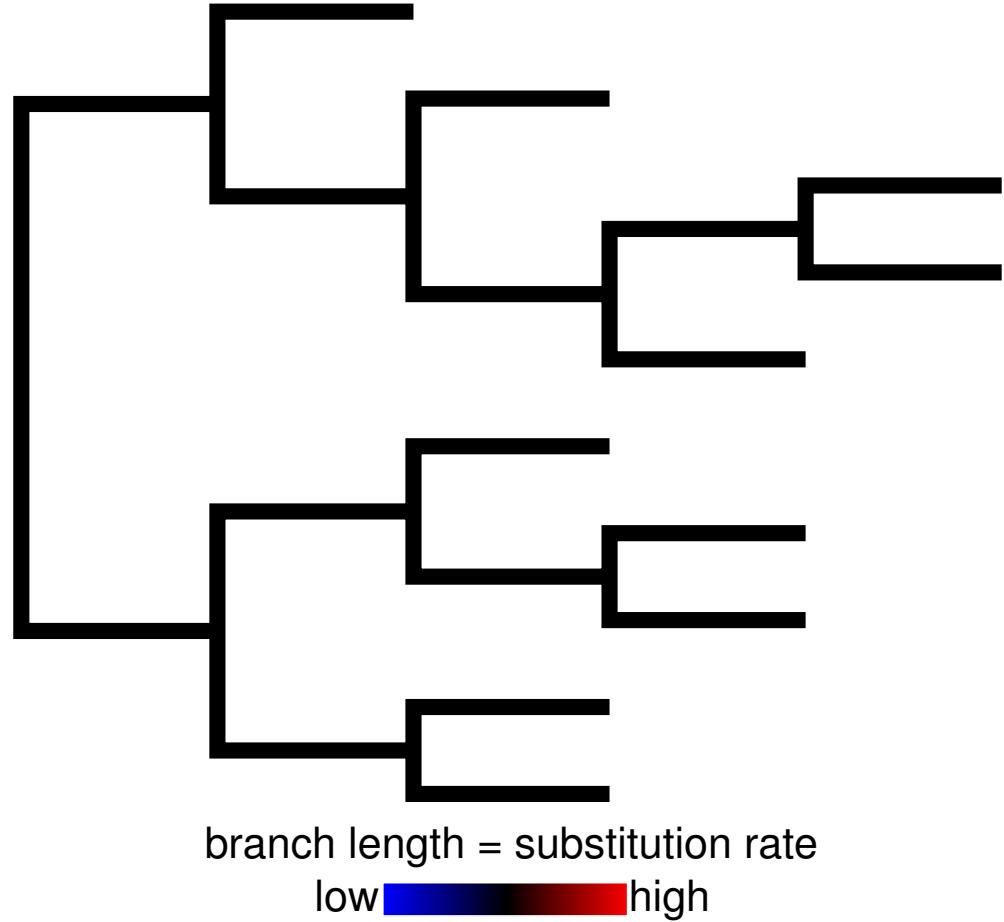

$$A = (a_1, a_2, a_3, \dots, a_{N-1})$$


N = number of tips



Global molecular clock

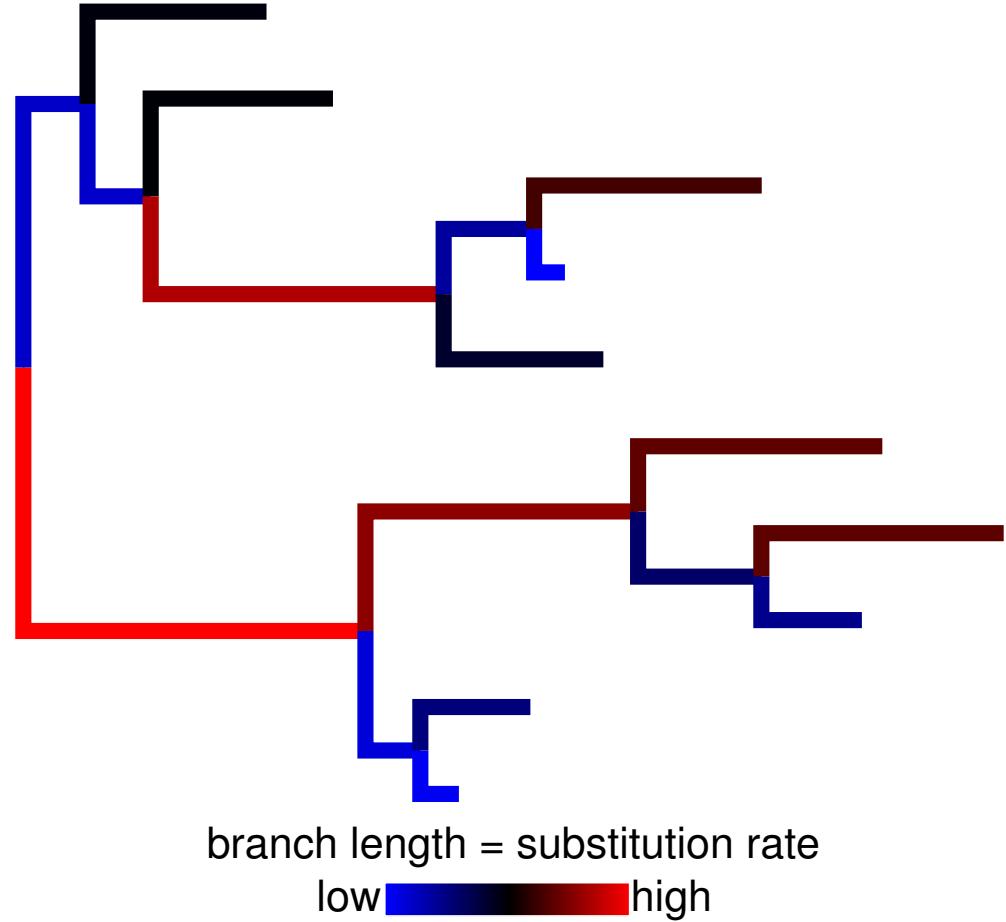
All branches
share the same
clock rate.





Uncorrelated relaxed clock

Each branch has its own clock rate, drawn from a global distribution (here, lognormal).



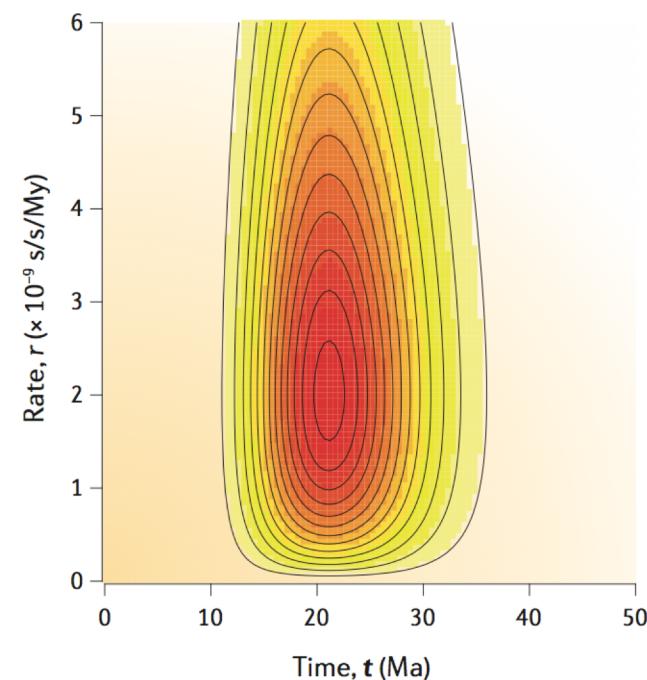


Clock models

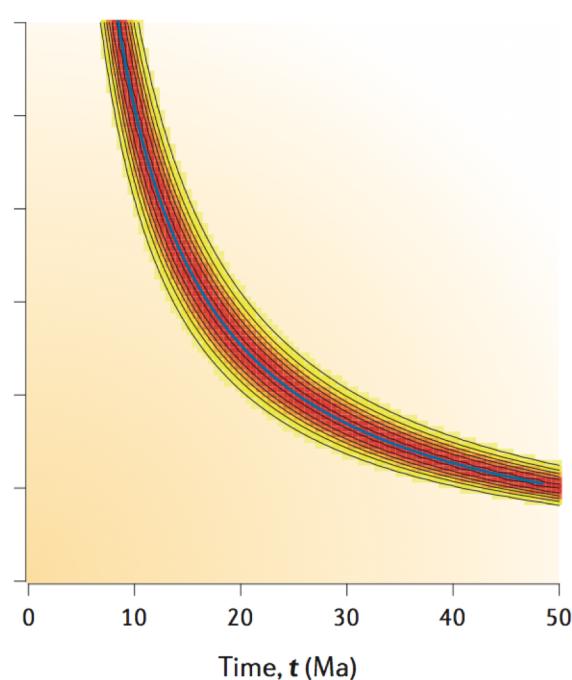
- Global clock (Zuckerkandl & Pauling, 1962)
- Uncorrelated/independent rates models (Drummond et al. 2006; Rannala & Yang 2007; Lepage et al. 2007)
- Log-normally distributed autocorrelated rates (Thorne, Kishino & Painter 1998; Kishino, Thorne & Bruno 2001; Thorne & Kishino 2002)
- Local clocks (Hasegawa, Kishino & Yano 1989; Kishino & Hasegawa 1990; Yoder & Yang 2000; Yang & Yoder 2003, Drummond and Suchard 2010)
- Mixture models on branch rates (Heath, Holder, Huelsenbeck 2012)
- Punctuated rate change model (Huelsenbeck, Larget and Swoford 2000)

The issue of identifiability

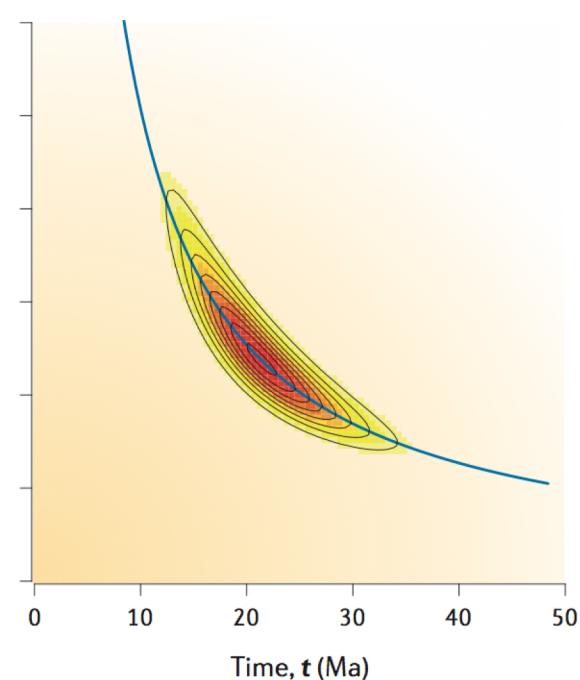
a Prior $f(t, r)$



b Likelihood $L(D|t, r)$



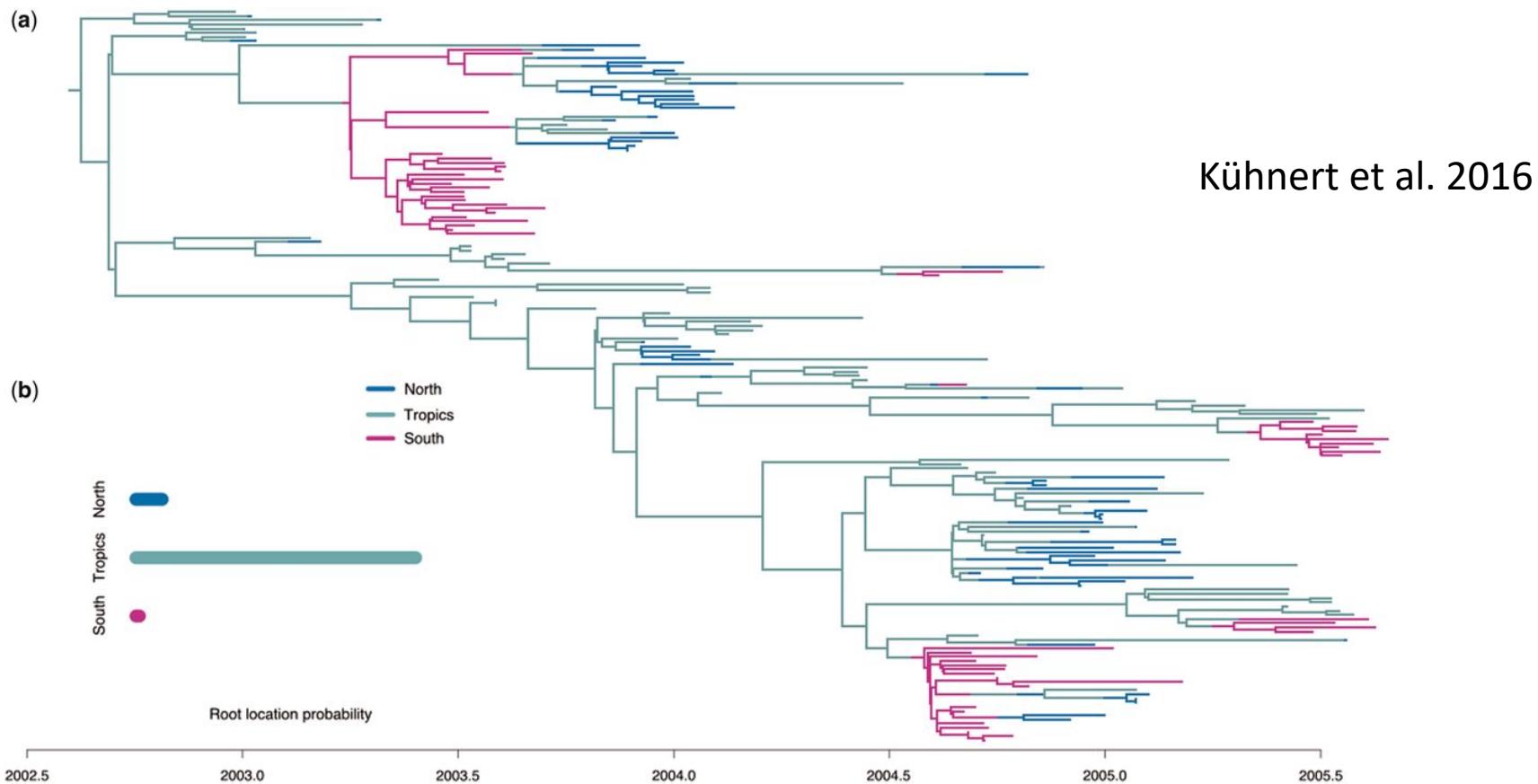
c Posterior $f(t, r|D)$



dos Reis et al. 2016. Nature Genetics Reviews

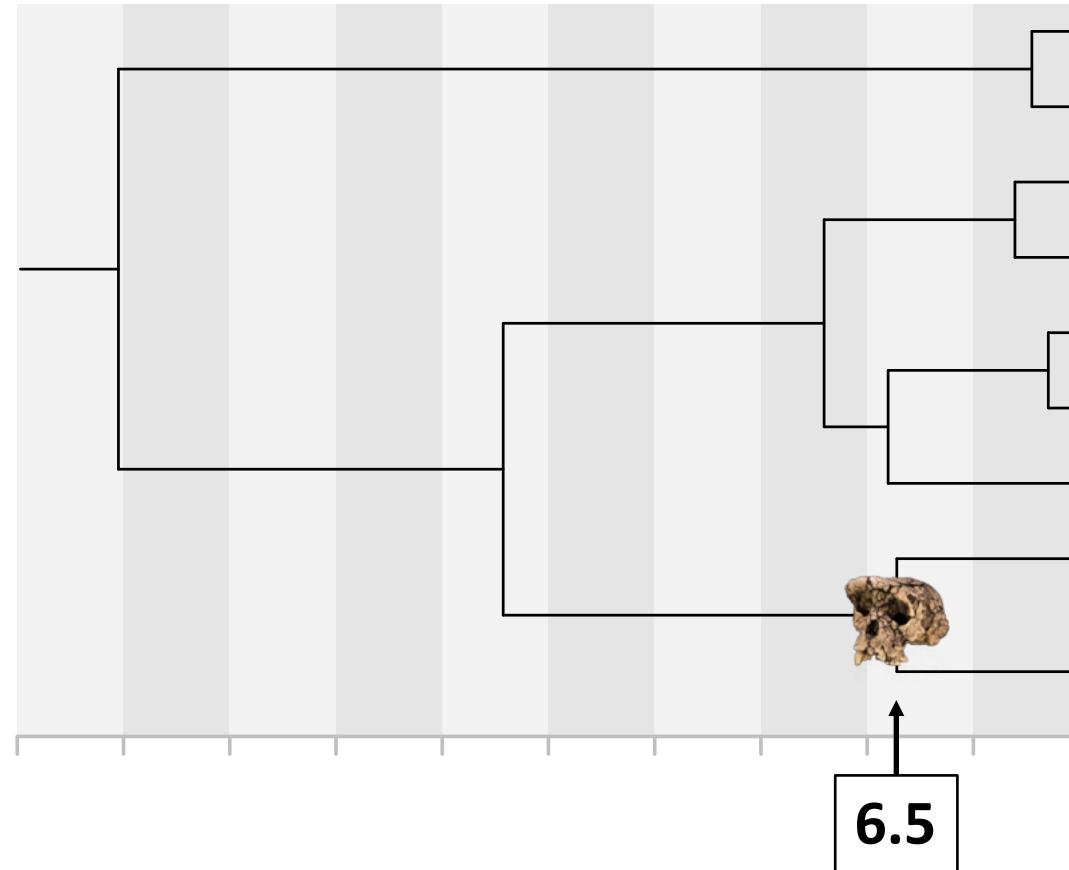
Calibrating the molecular clock

Sampling-through-time



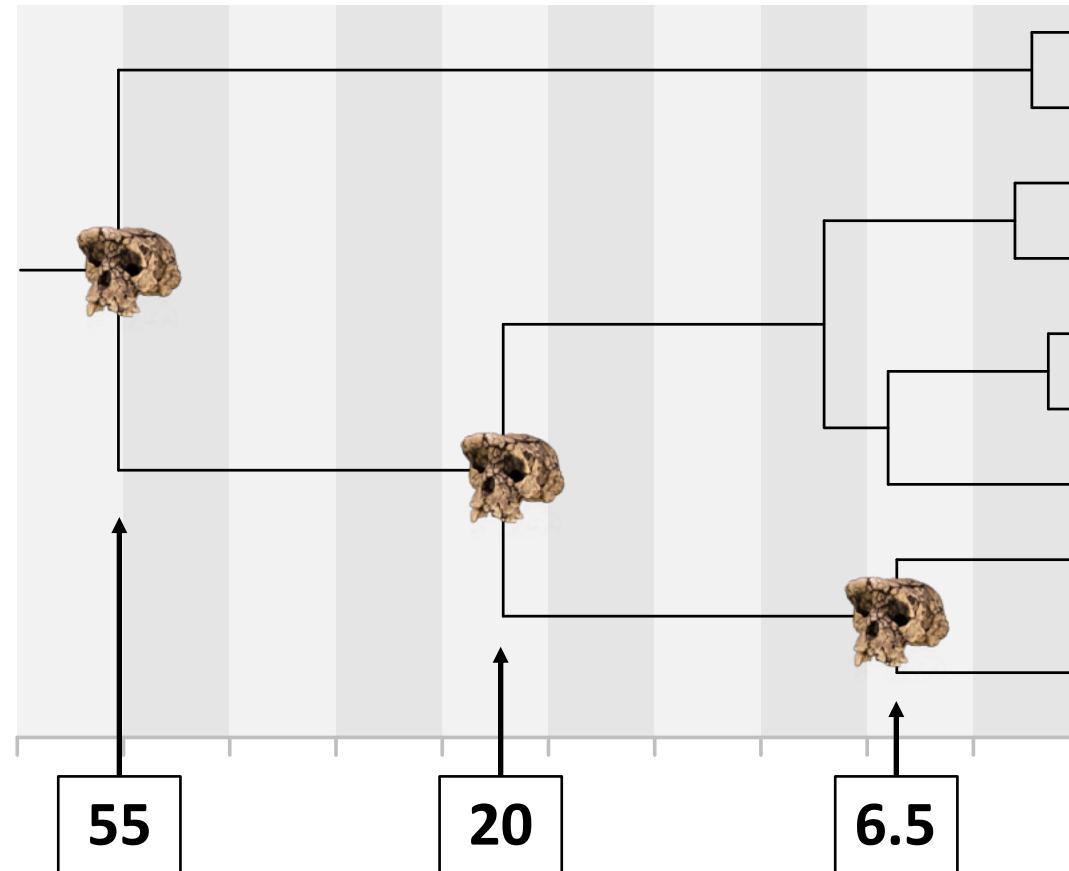
Calibrating the molecular clock

Fossil data

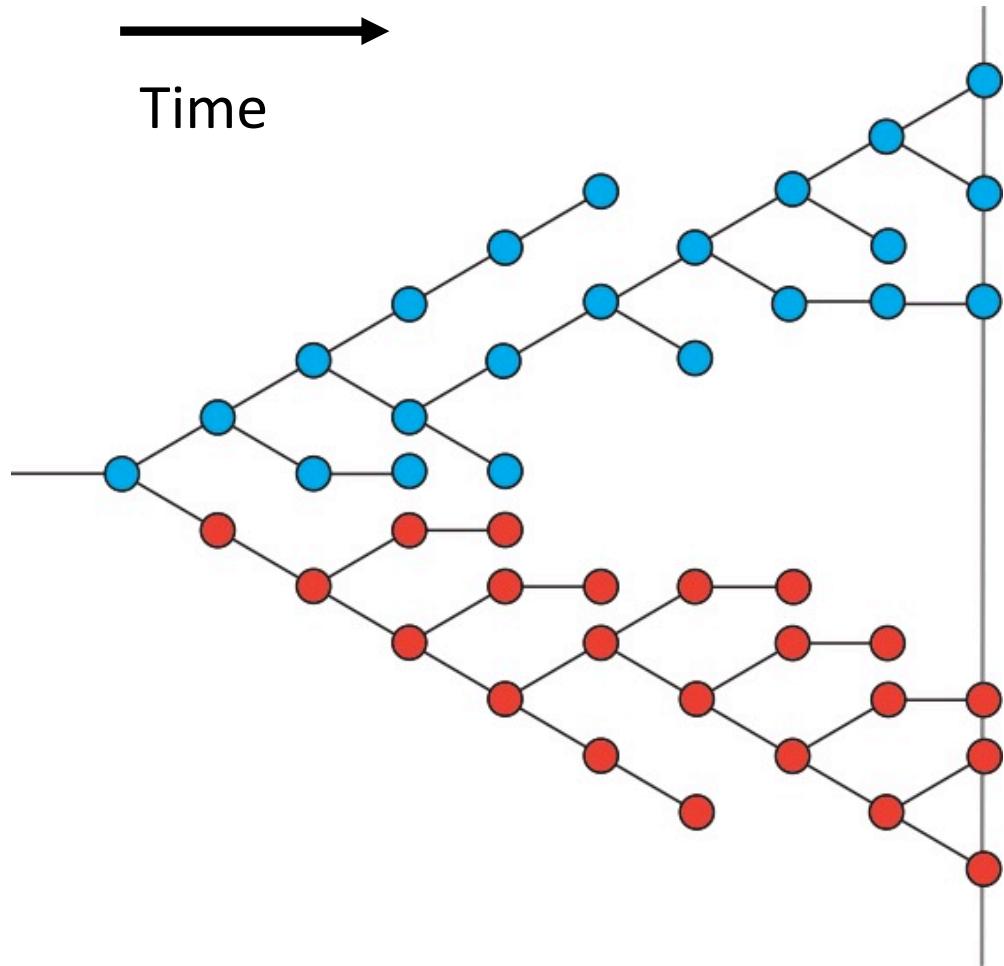


Calibrating the molecular clock

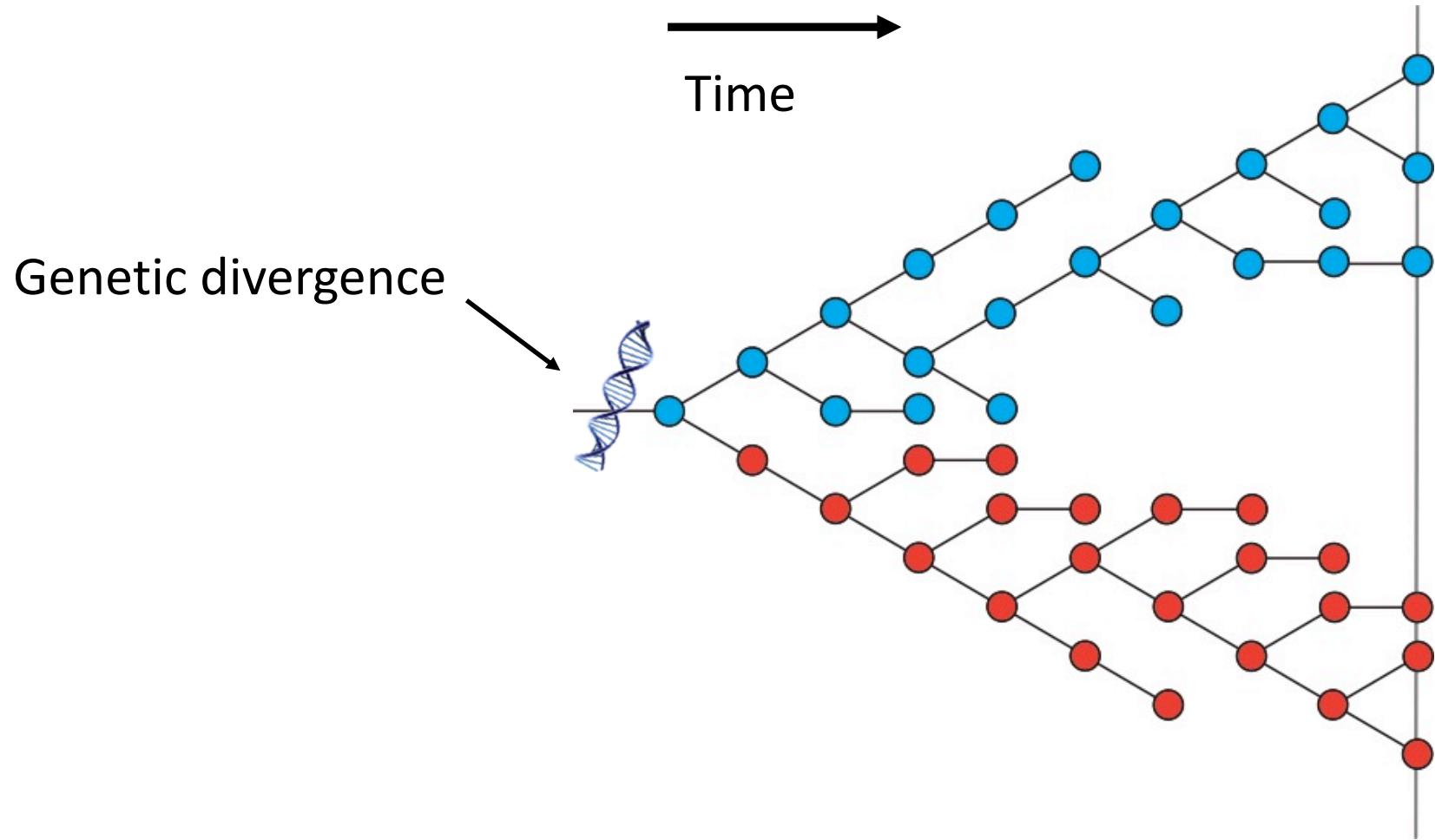
Fossil data



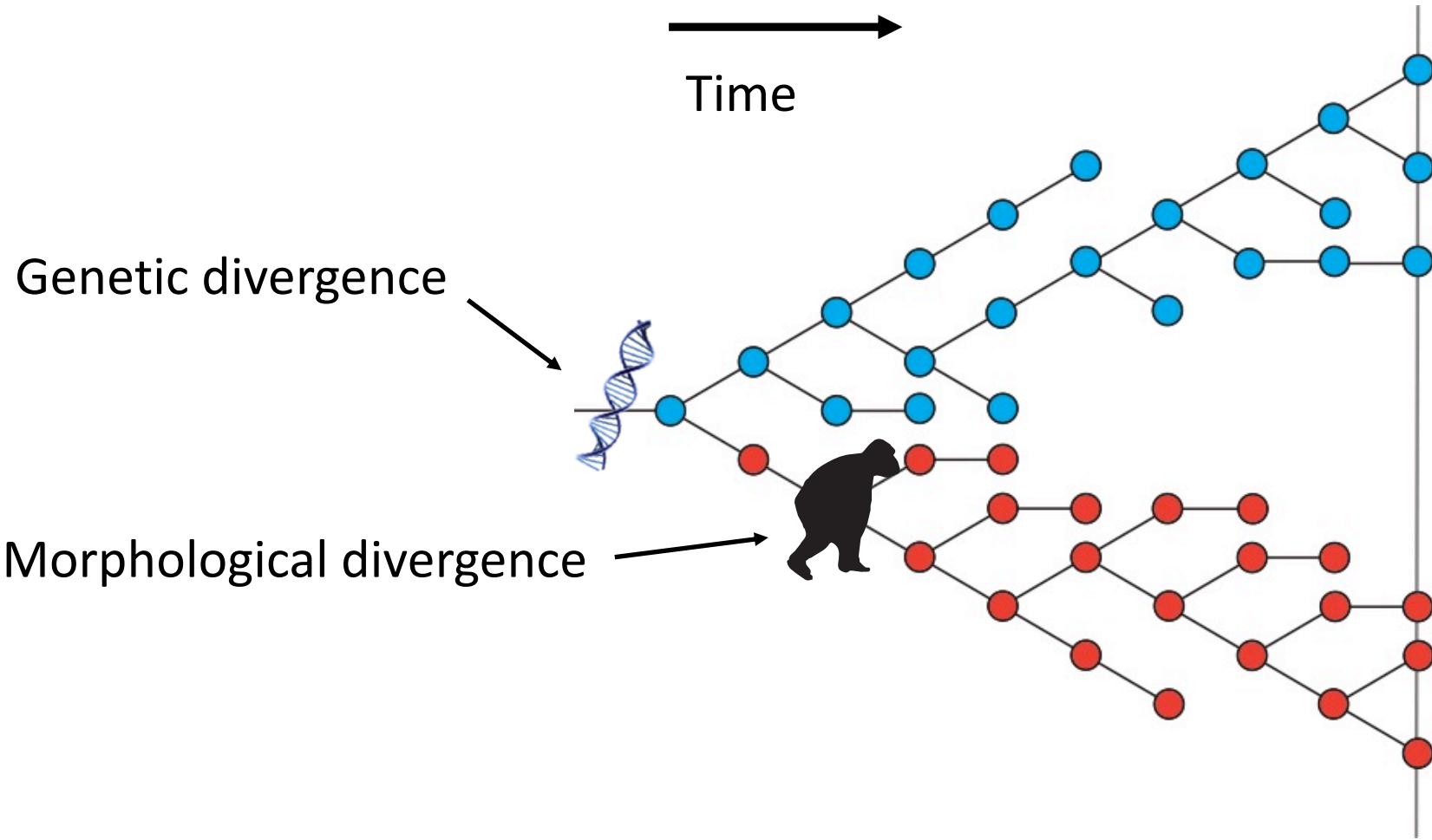
Fossils provide minimum ages only



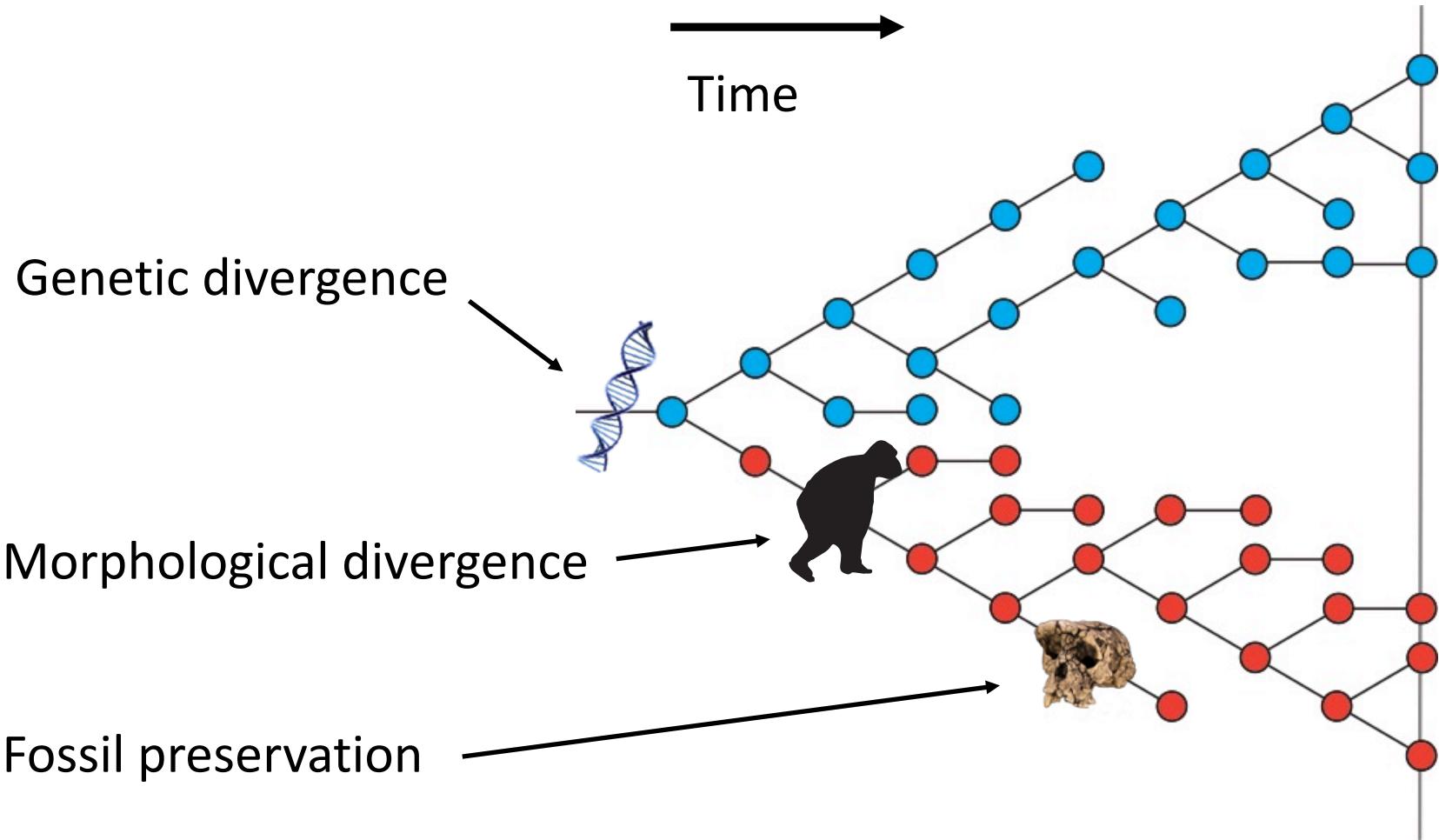
Fossils provide minimum ages only



Fossils provide minimum ages only

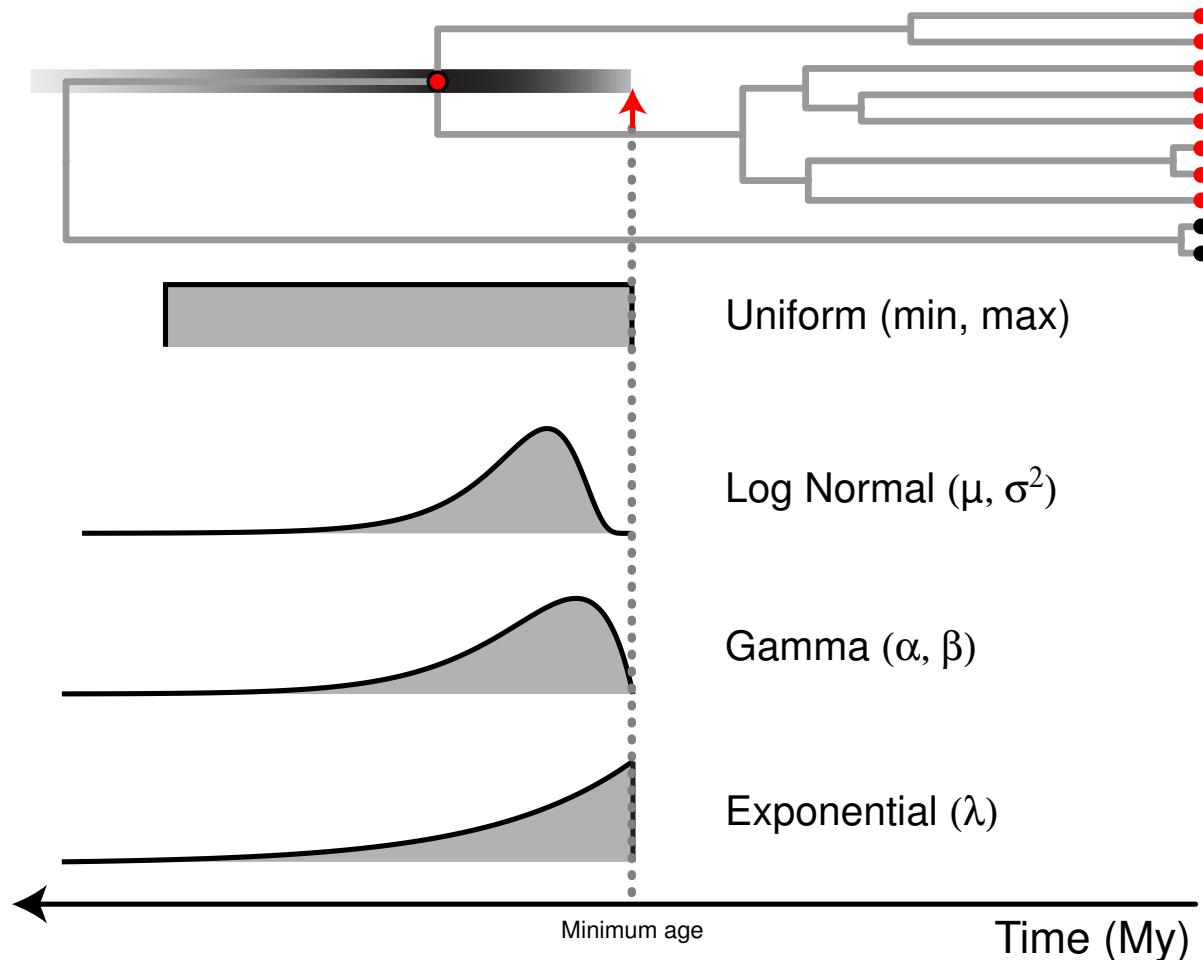


Fossils provide minimum ages only

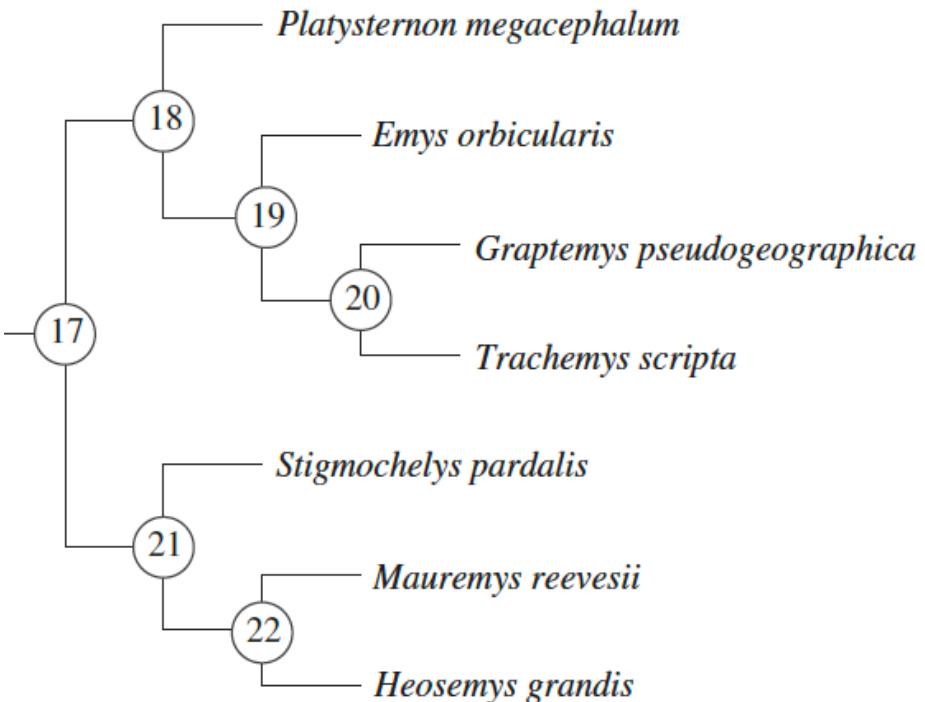
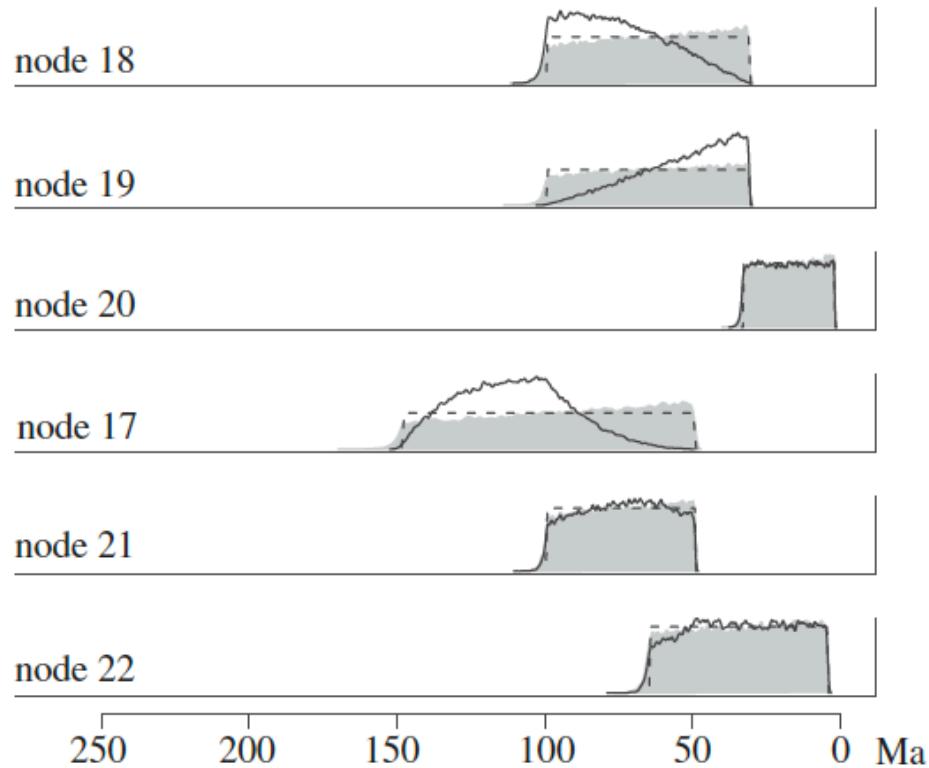


Calibration densities

Prior distributions
on the waiting
time between the
genetic divergence
and the sampled
fossil.

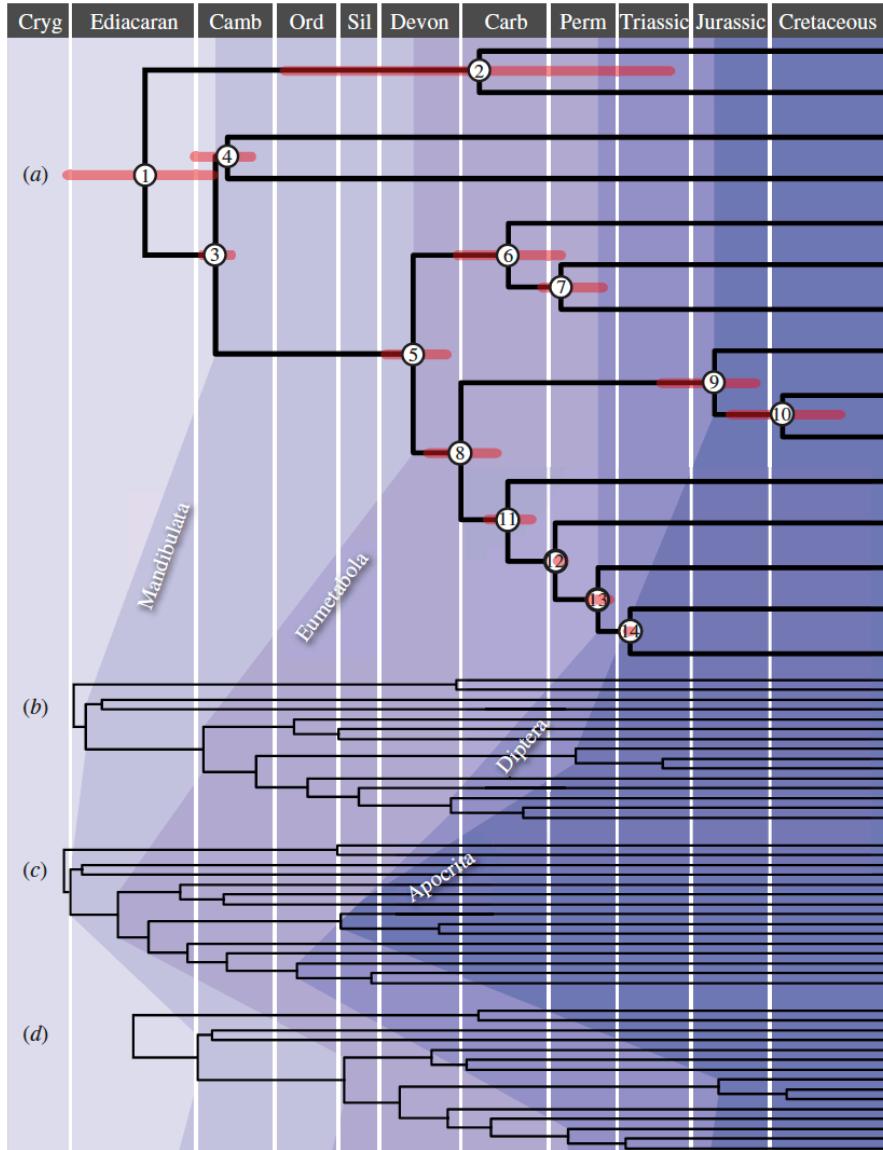
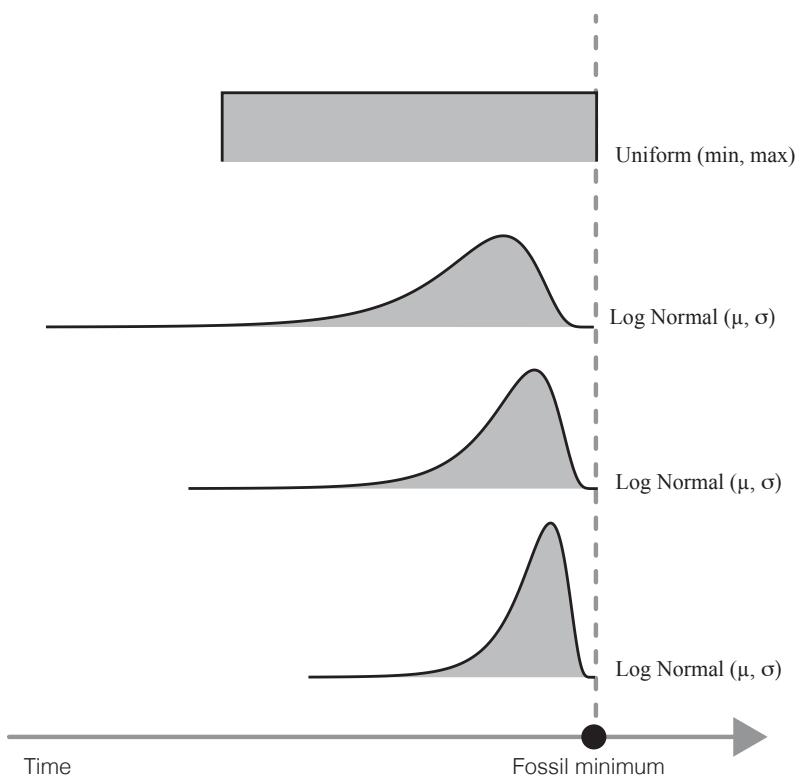


Effective priors



Warnock et al. 2014

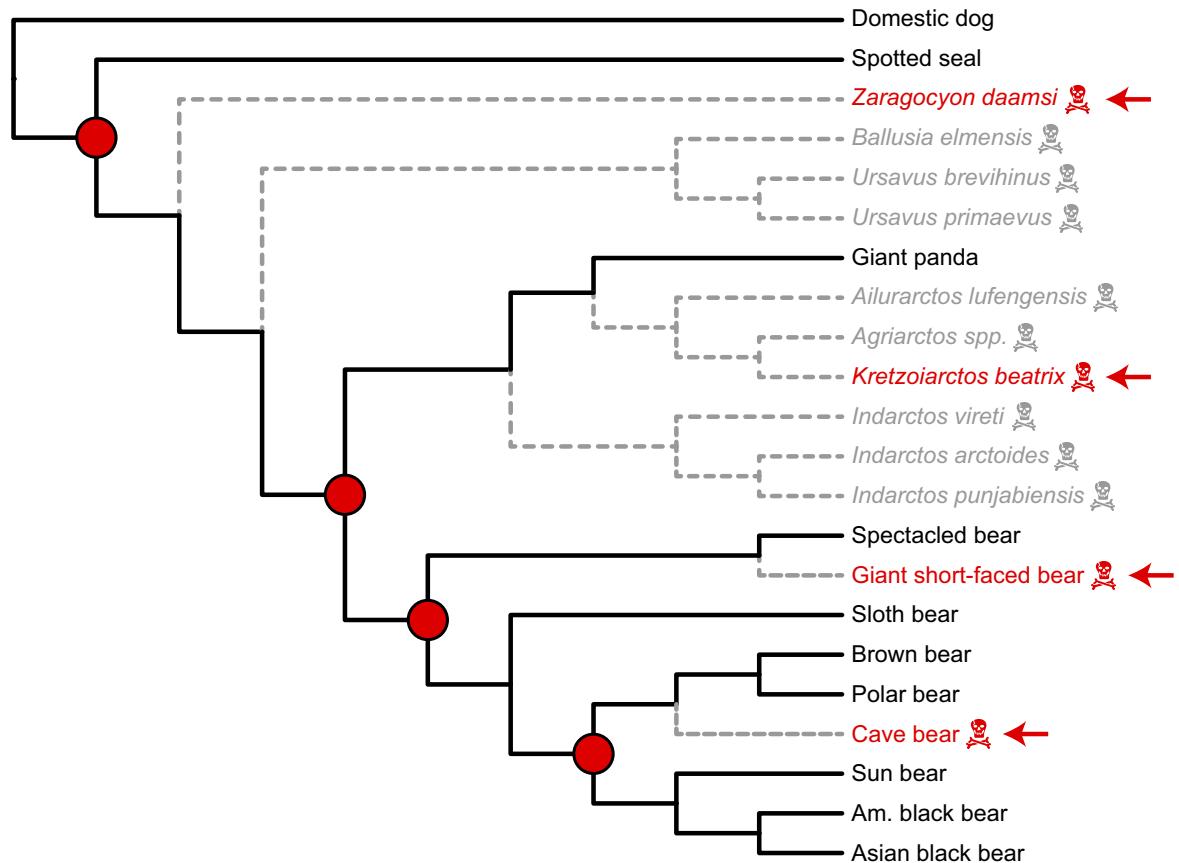
Impact of different calibration priors



Warnock et al. 2012

Calibration densities remove data

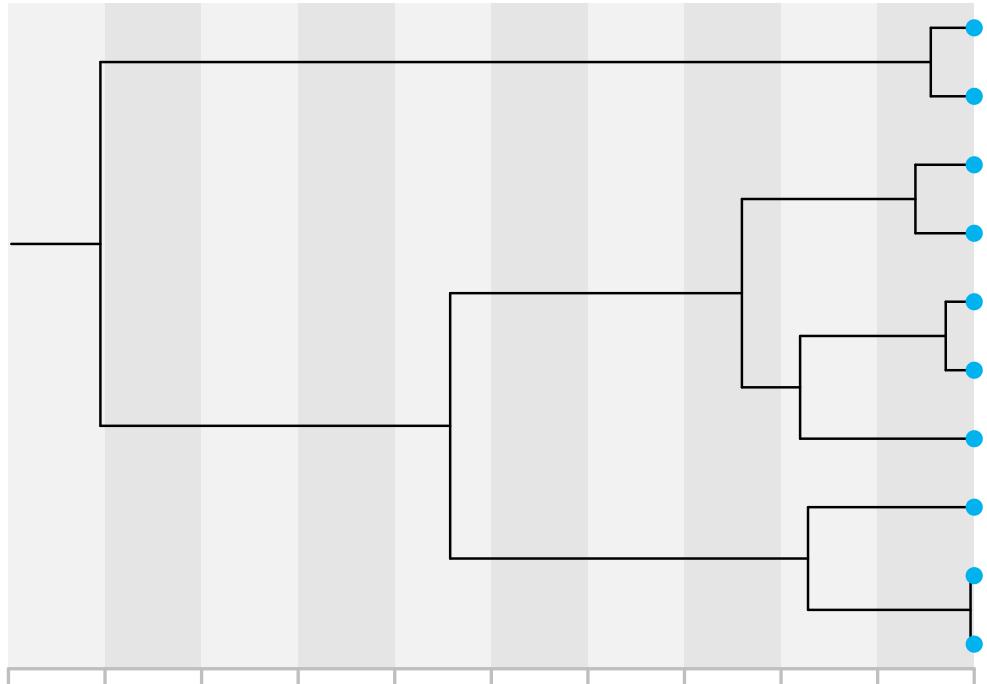
With calibration methods, only the oldest fossil for each node can be used.



Krause et al. BMC Evol. Biol. 2008; Abella et al. PLoS ONE 2012

The Birth-Death process

(λ) (μ)
(ψ) (ρ)

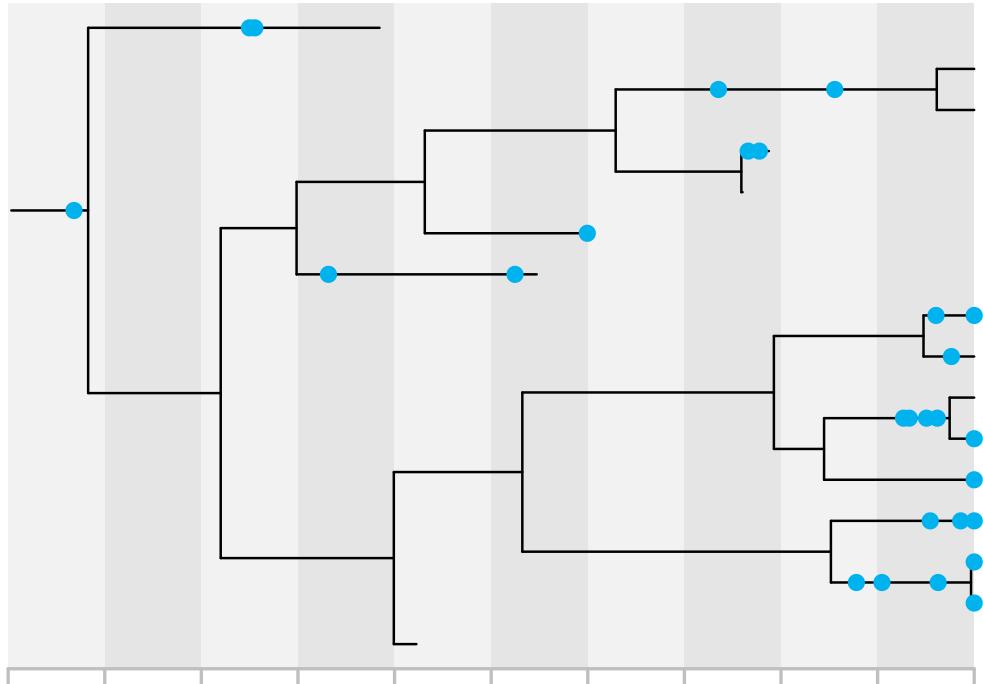


λ — speciation rate

μ — extinction rate

ρ — extant species
sampling probability

The Fossilized Birth-Death process



λ — speciation rate

μ — extinction rate

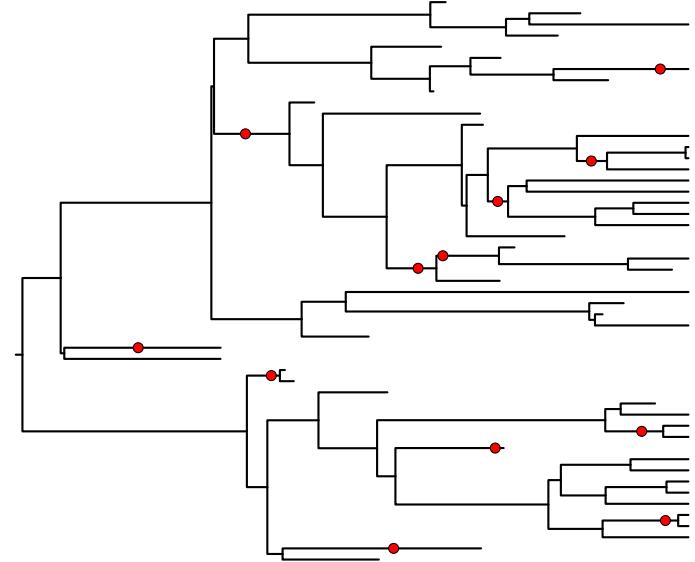
Ψ — **fossil sampling rate**

ρ — extant species
sampling probability

Stadler, 2010

Sampled ancestors

Sampled lineages with sampled descendants

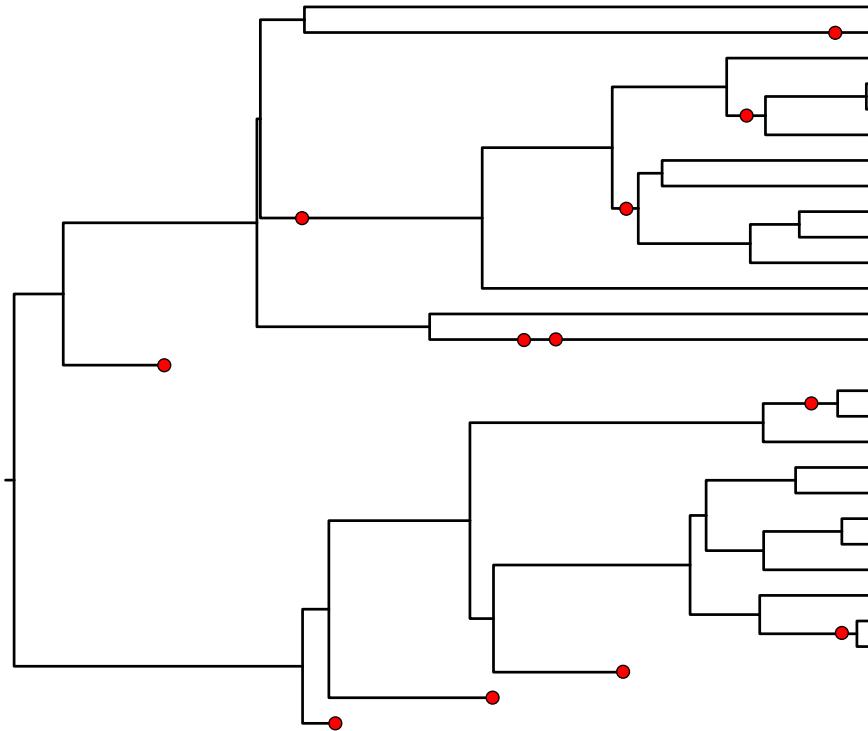


Paleobiology, 22(2), 1996, pp. 141–151

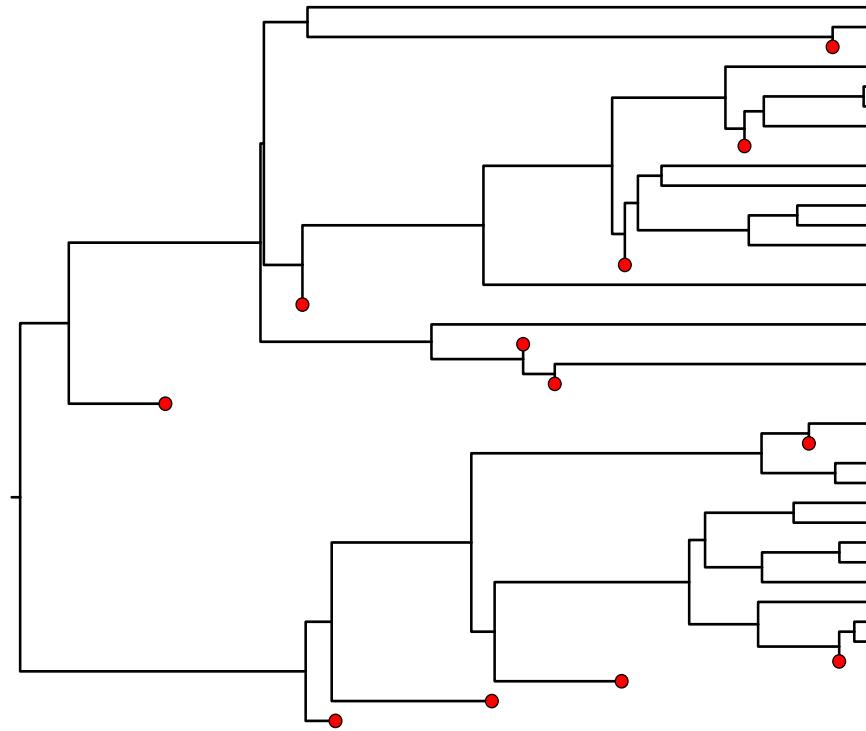
On the probability of ancestors in the fossil record

Mike Foote

Impact of sampled ancestors

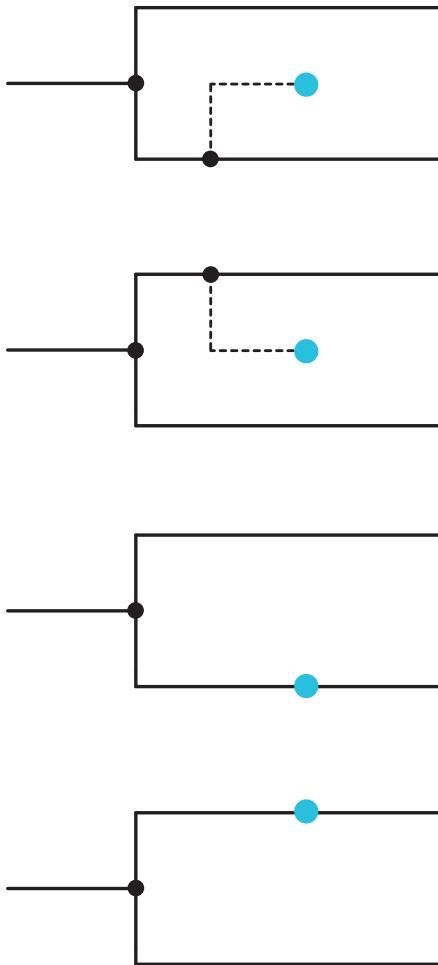


With sampled ancestors



Without sampled ancestors

MCMC implementation

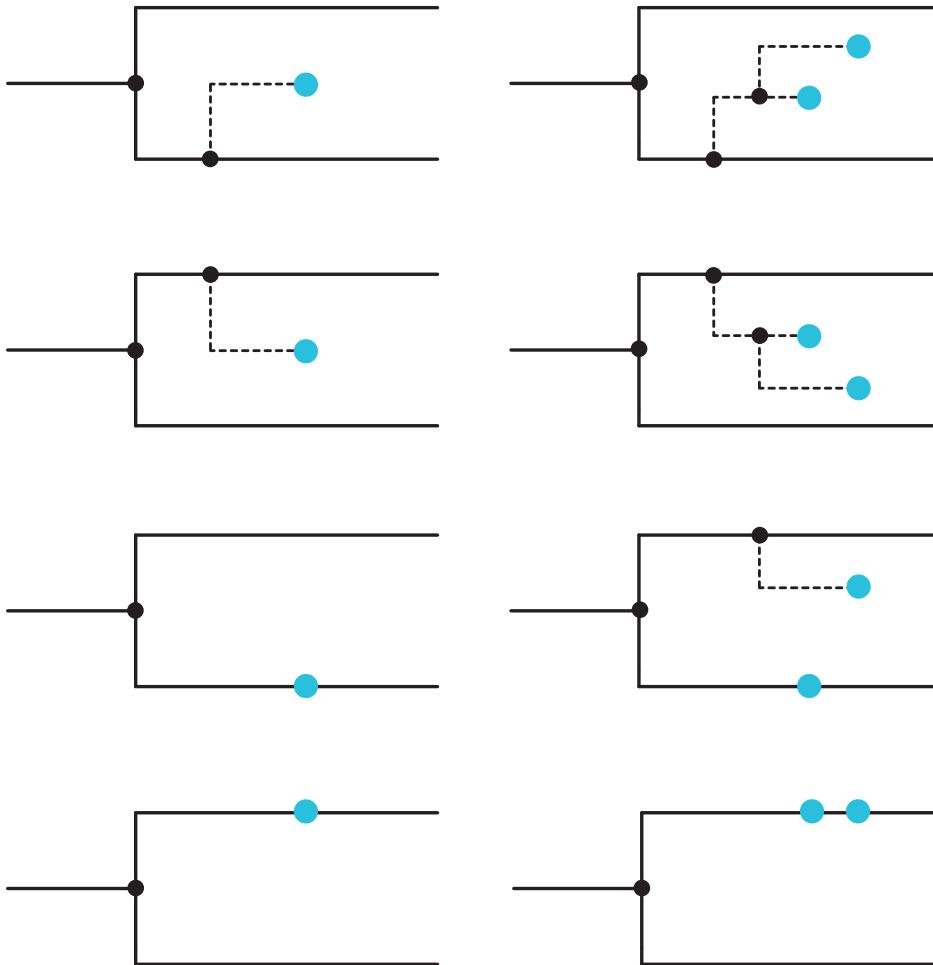


A fossil sample can be placed at the end of an unobserved branch or as a sampled ancestor on an observed branch.

- Fossil occurrence
- Speciation event

Figure courtesy of R. Warnock

MCMC implementation

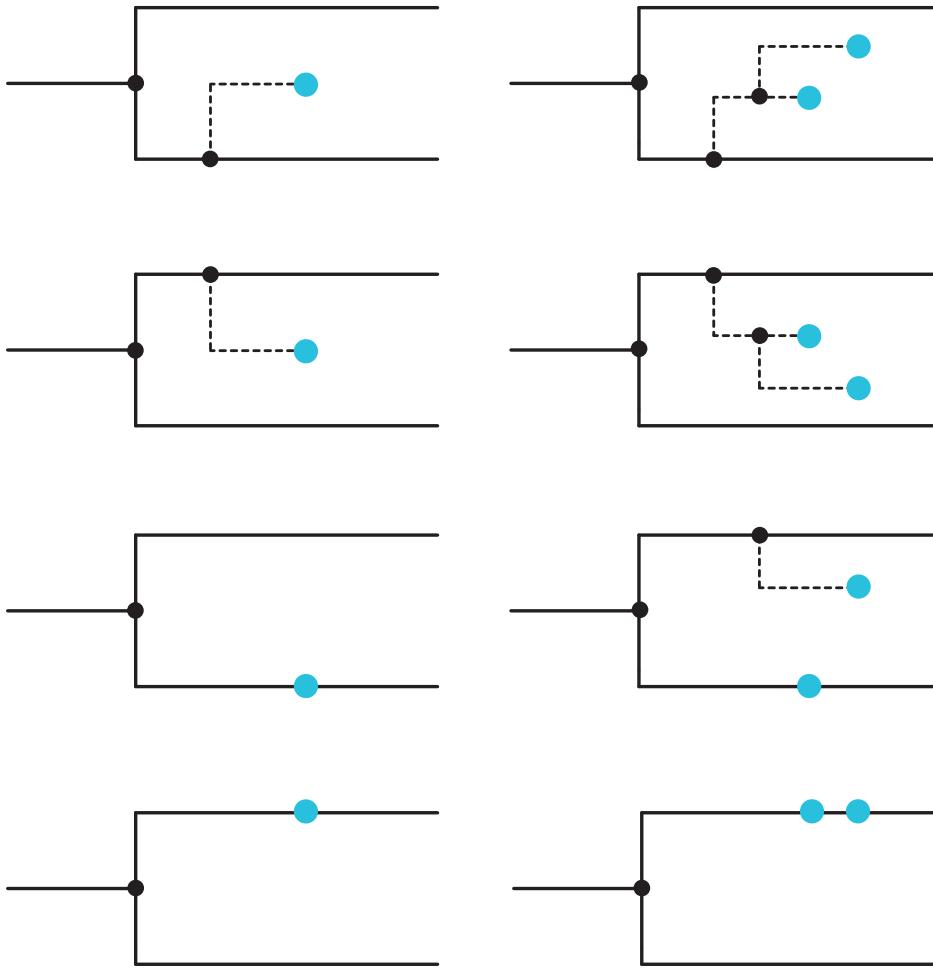


Each fossil can be attached anywhere on the tree, including on unobserved branches.

- Fossil occurrence
- Speciation event

Figure courtesy of R. Warnock

MCMC implementation

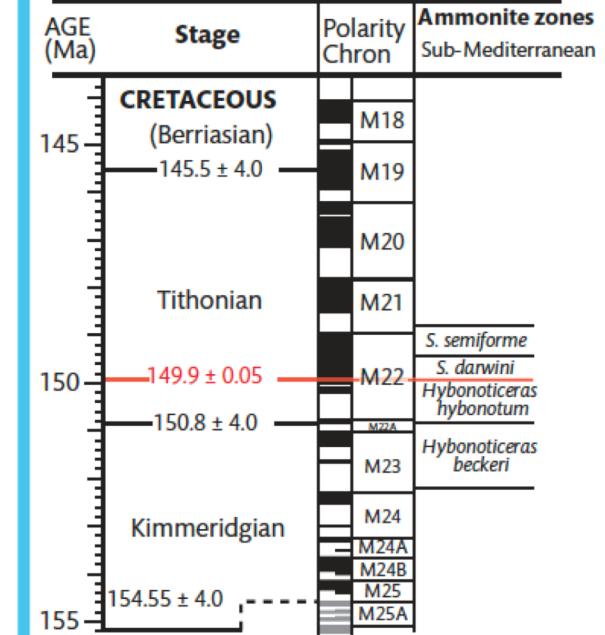
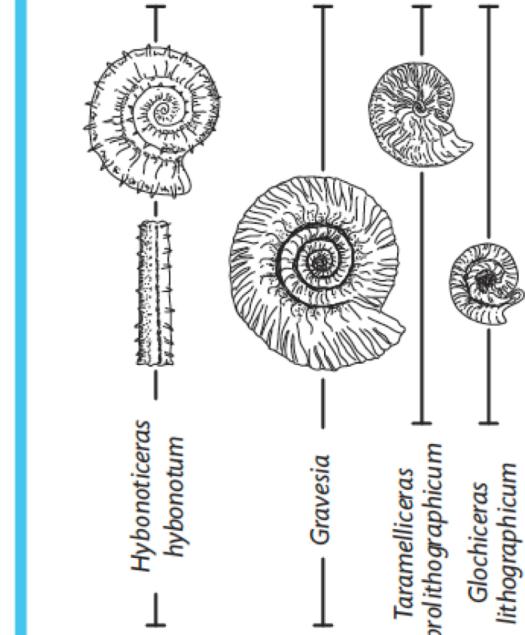
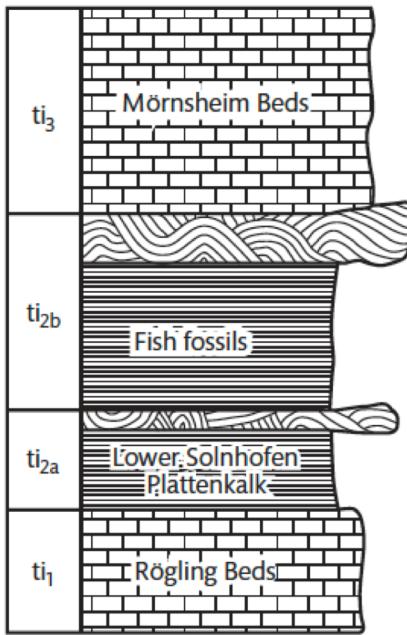


The FBD process gives the probability of each configuration conditioned on the model parameters.

- Fossil occurrence
- Speciation event

Figure courtesy of R. Warnock

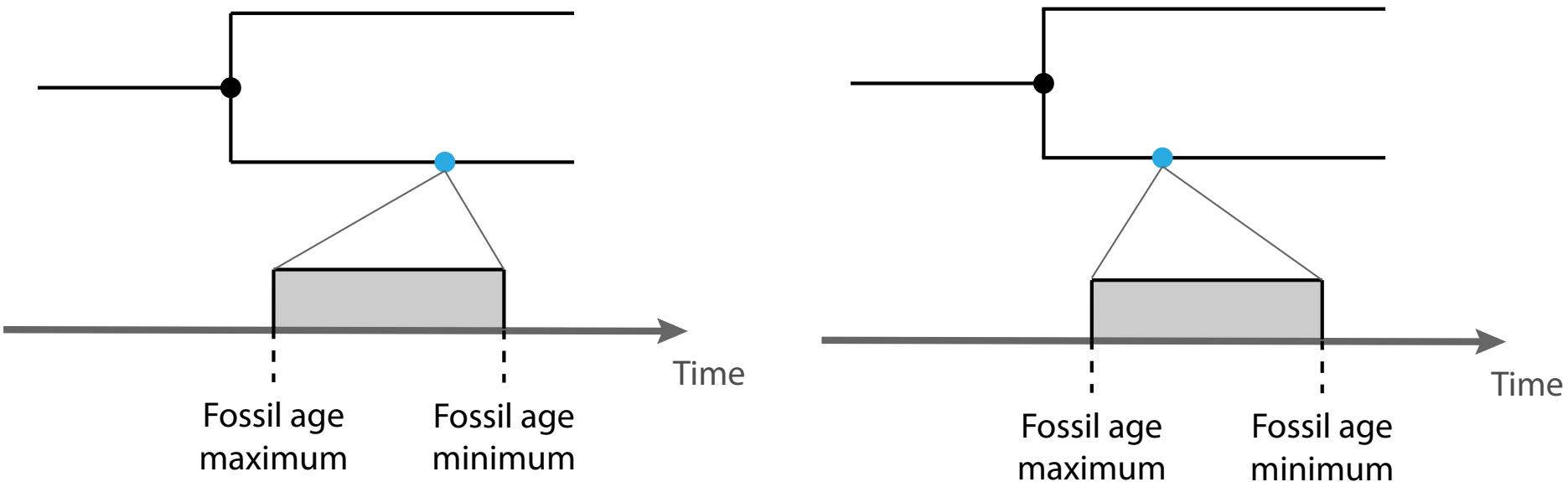
Fossil age uncertainty



Benton et al. 2009. TimeTree of Life

Integrating the uncertainty

Ages can be sampled as part of the MCMC



Integrating the uncertainty

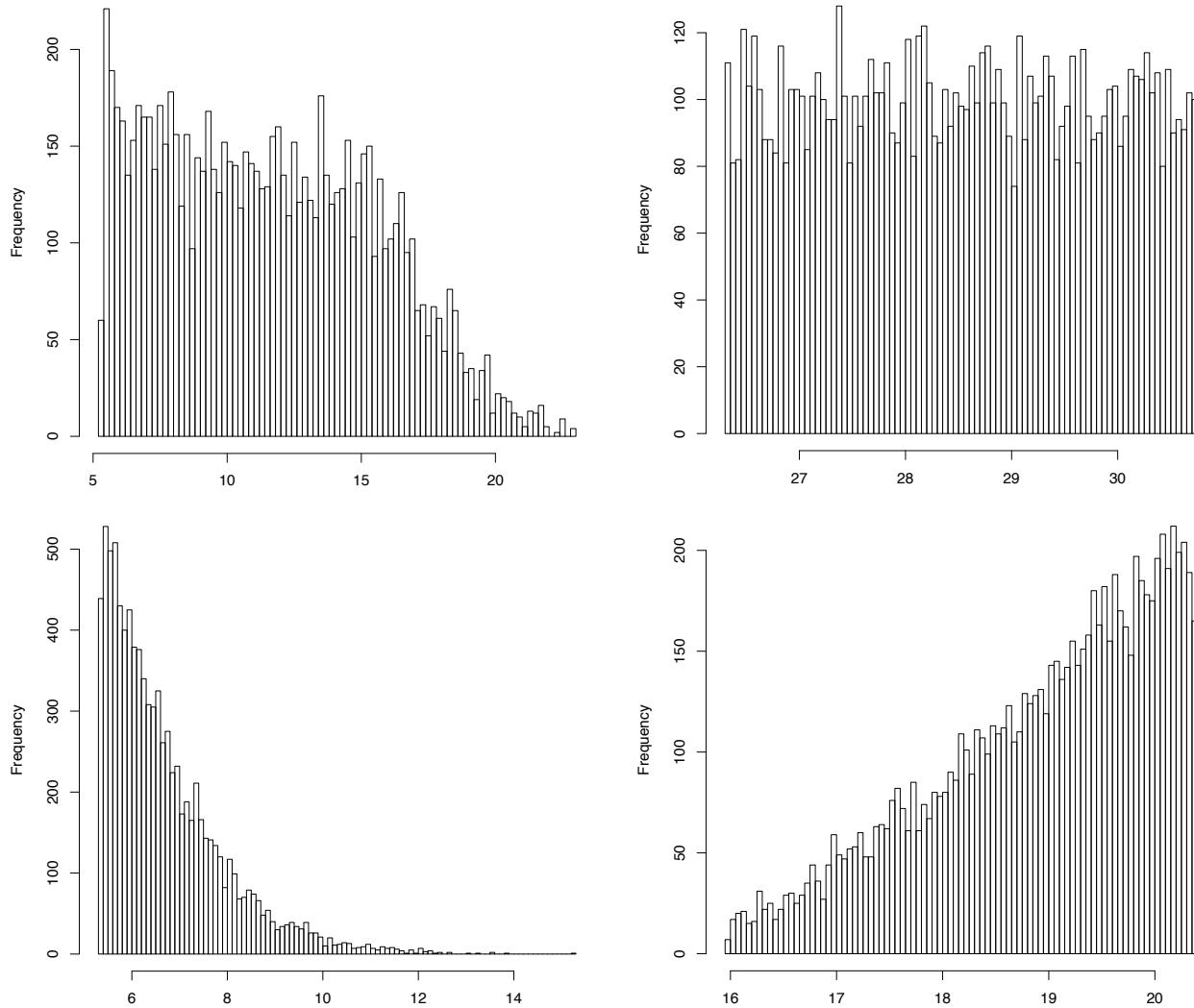
Ages **should** be sampled as part of the MCMC

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

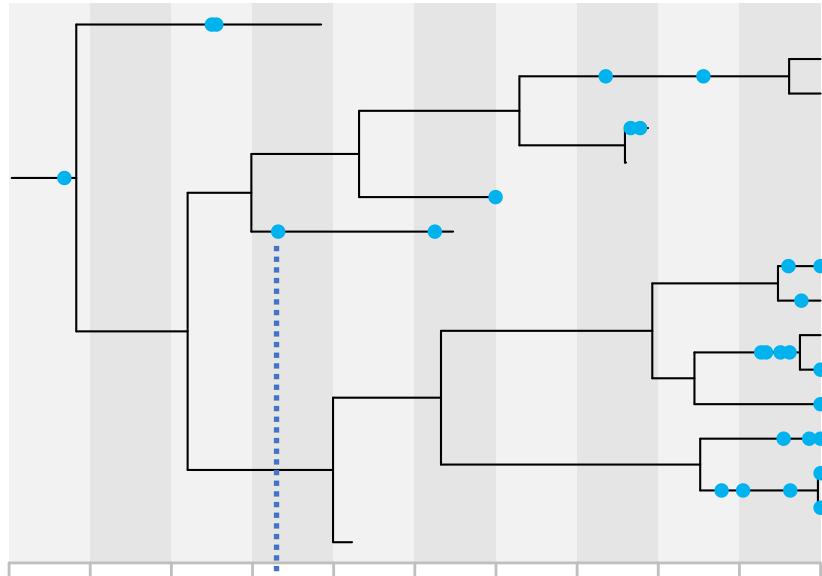
Joëlle Barido-Sottani^{1,2,3}, Gabriel Aguirre-Fernández⁴, Melanie Hopkins⁵,
Tanja Stadler^{1,2} and Rachel Warnock^{1,2,4}

Barido-Sottani et al. 2019. Proc. Roy. Soc. B

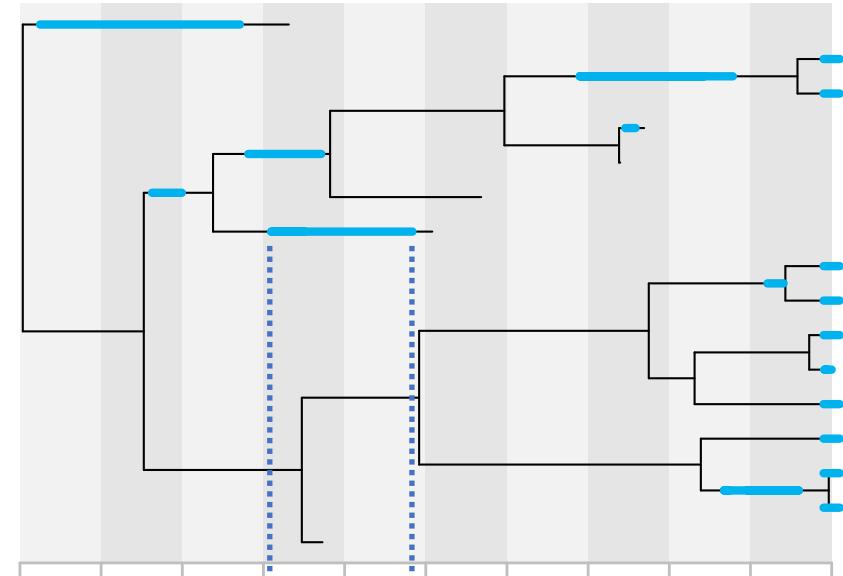
Effective priors



Specimen-level data vs species ranges



Fossil specimen
(one occurrence)



Last occurrence
First occurrence

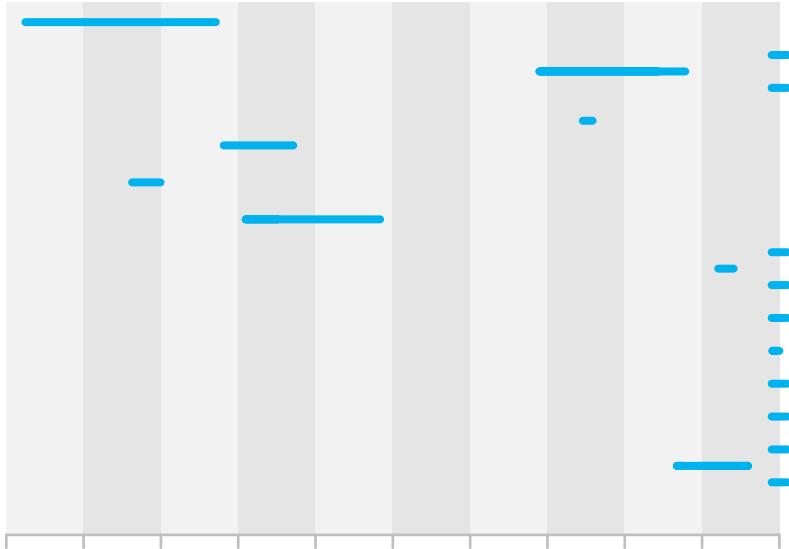
Figures courtesy of R. Warnock

The FBD range process

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

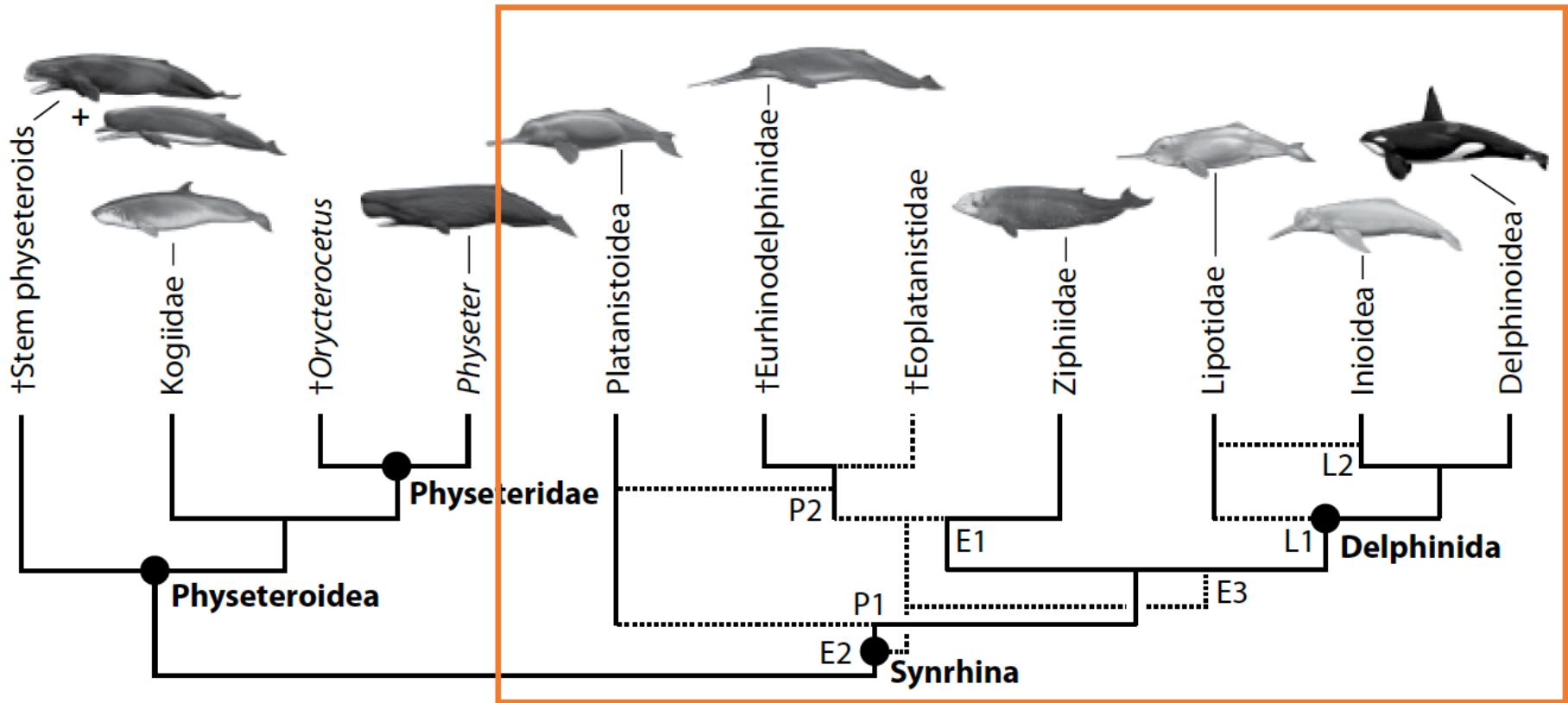
Journal of Theoretical Biology 447:41-55 (2018)



Can also be used without molecular data

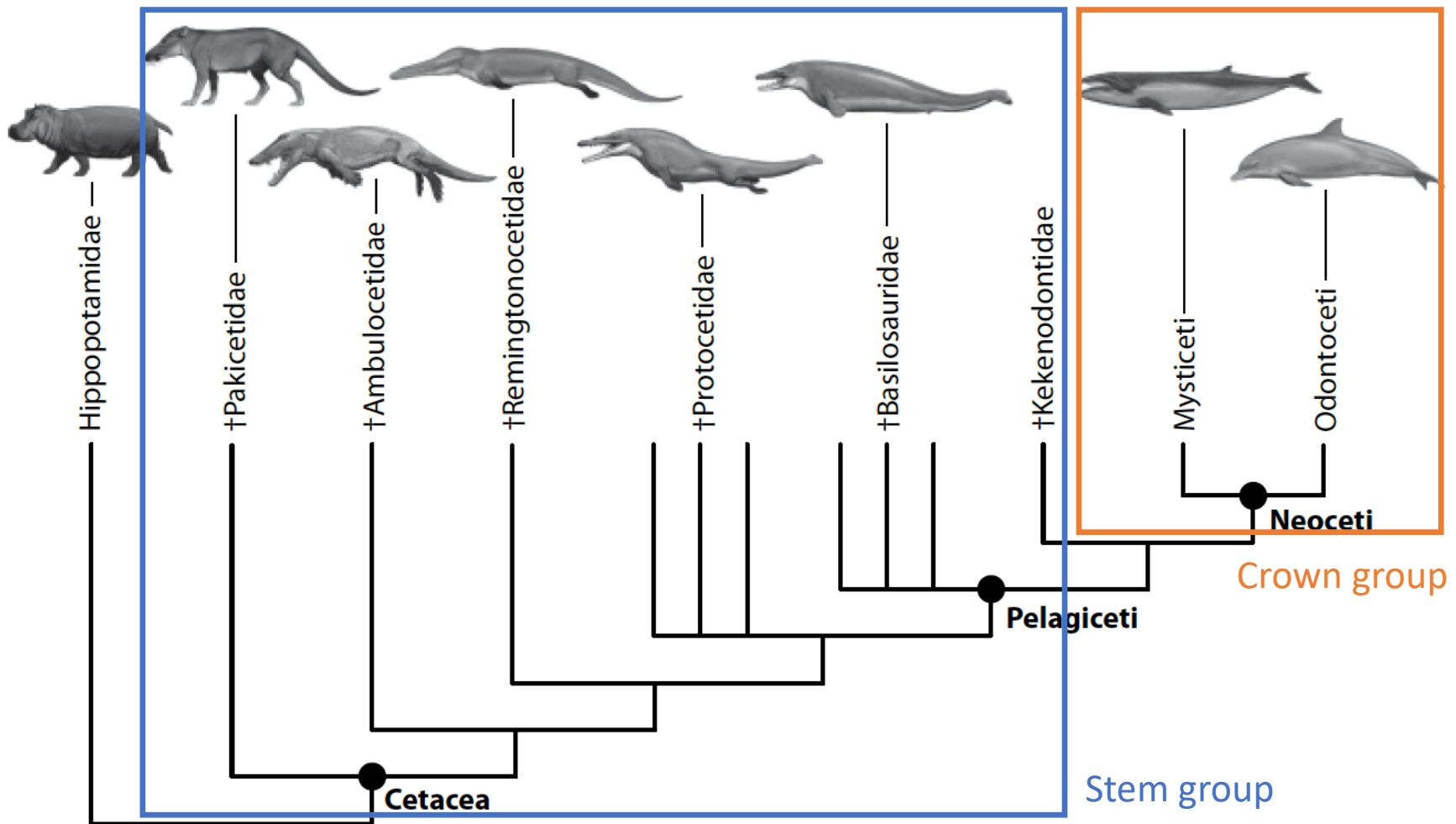
Currently only available in RevBayes, implementation in BEAST2 is planned

Taxonomic uncertainty



Marx et al. 2016. Cetacean Paleobiology

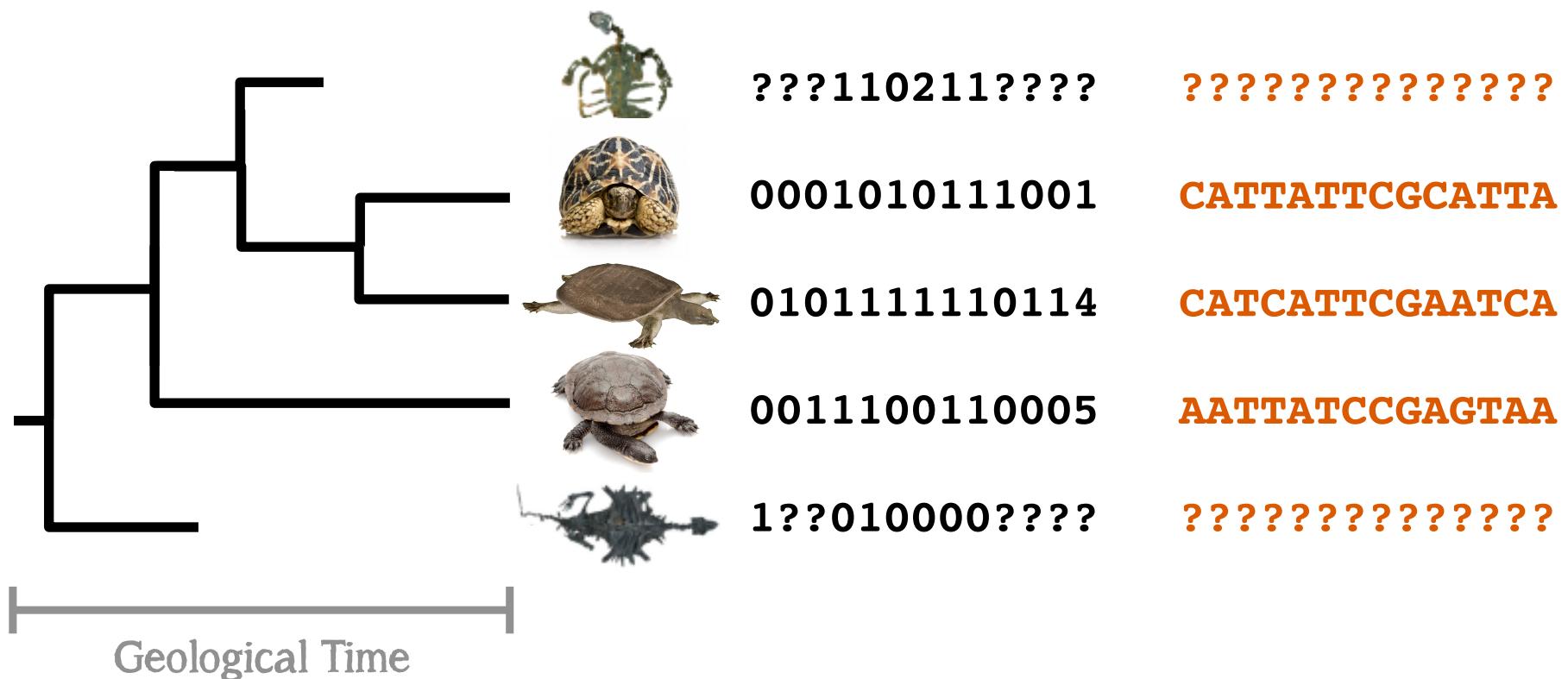
Taxonomic uncertainty



Marx et al. 2016. Cetacean Paleobiology

0101...
1101...
0100...

Morphological evidence





The Mk Lewis model

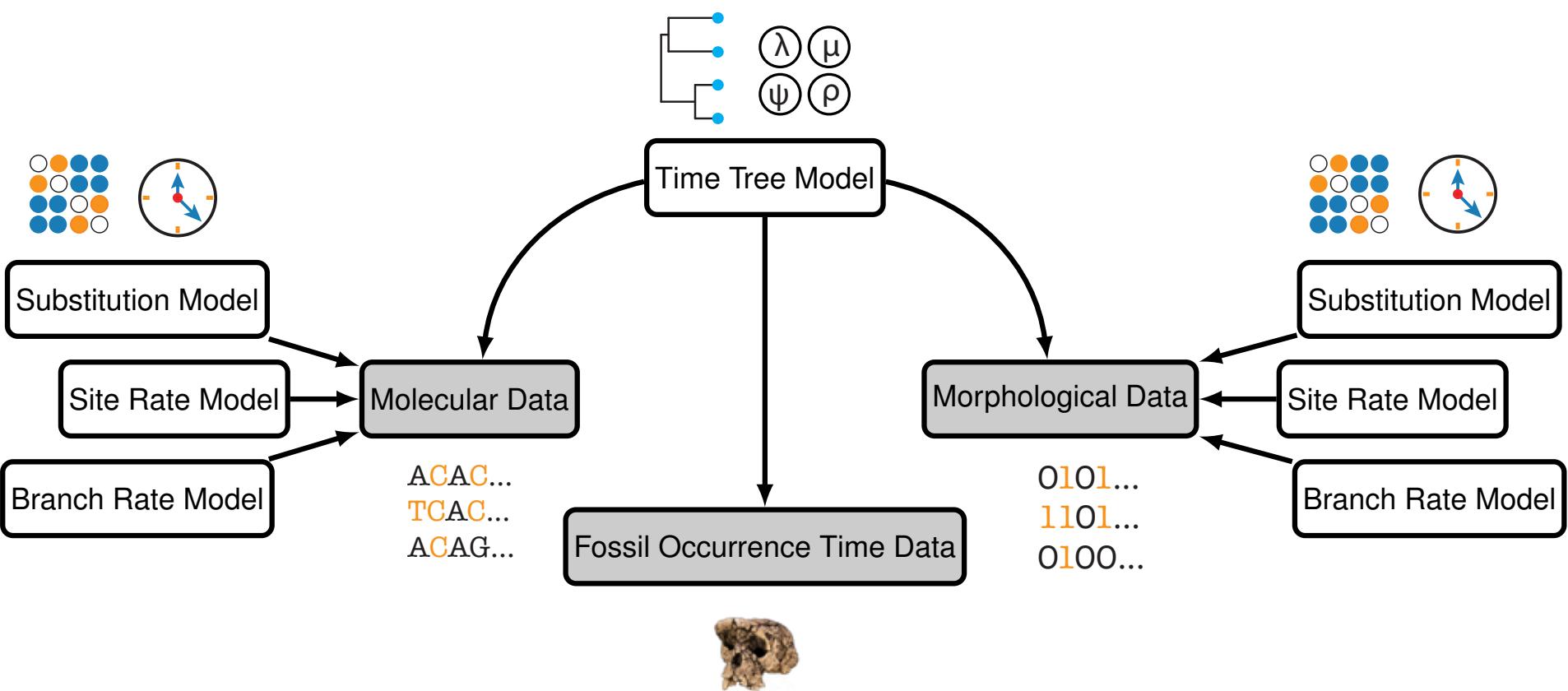
- k possible states for each character
- equal transition rates between all states

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

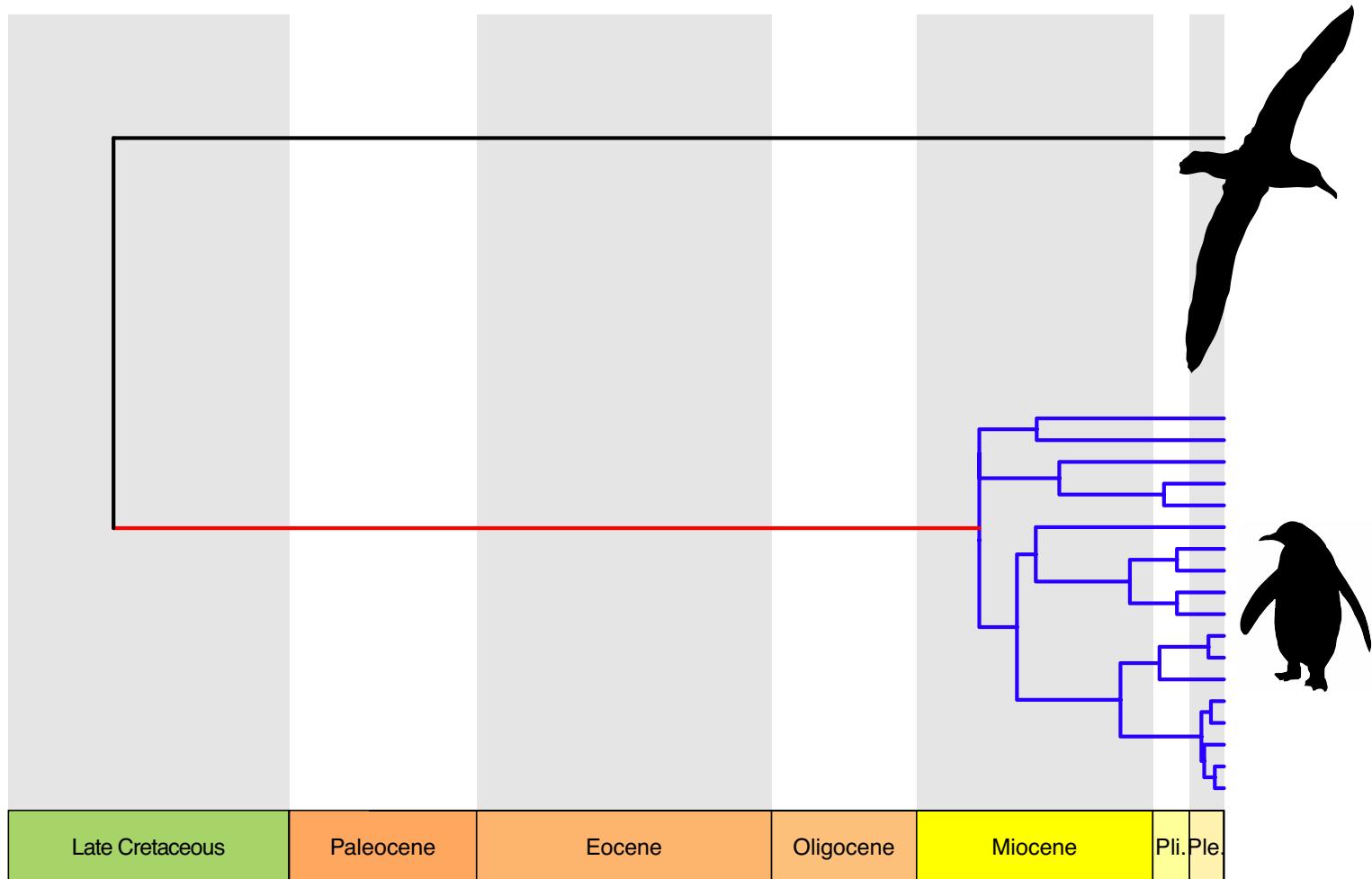
T1	0
T2	0
T3	1
T4	2
T5	2
T6	1
T7	1

Lewis 2001. Systematic Biology

Total-evidence approach

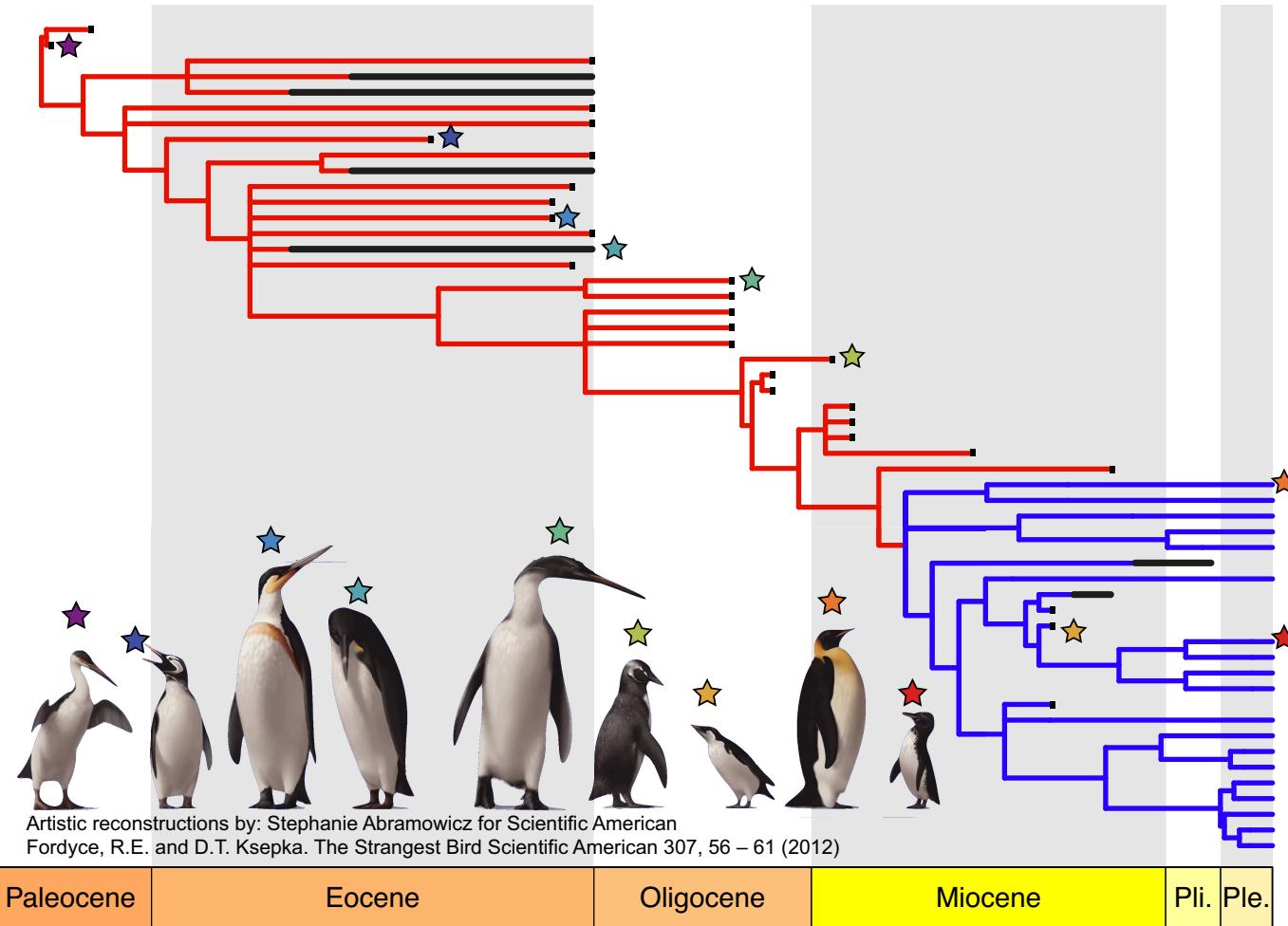


Dating extant penguins



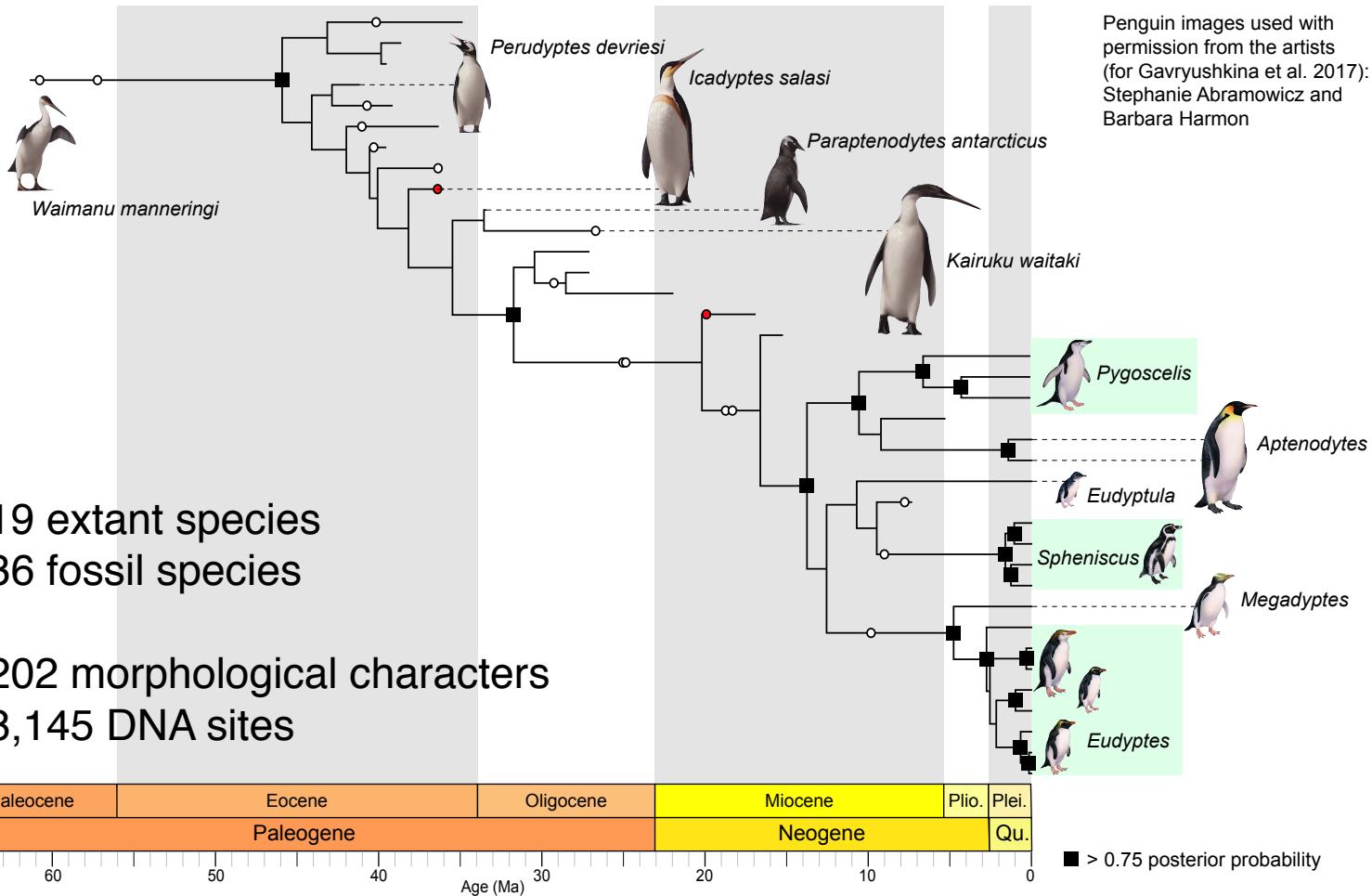
Gavryushkina et al. 2017. *Sys Bio*

Dating extant penguins



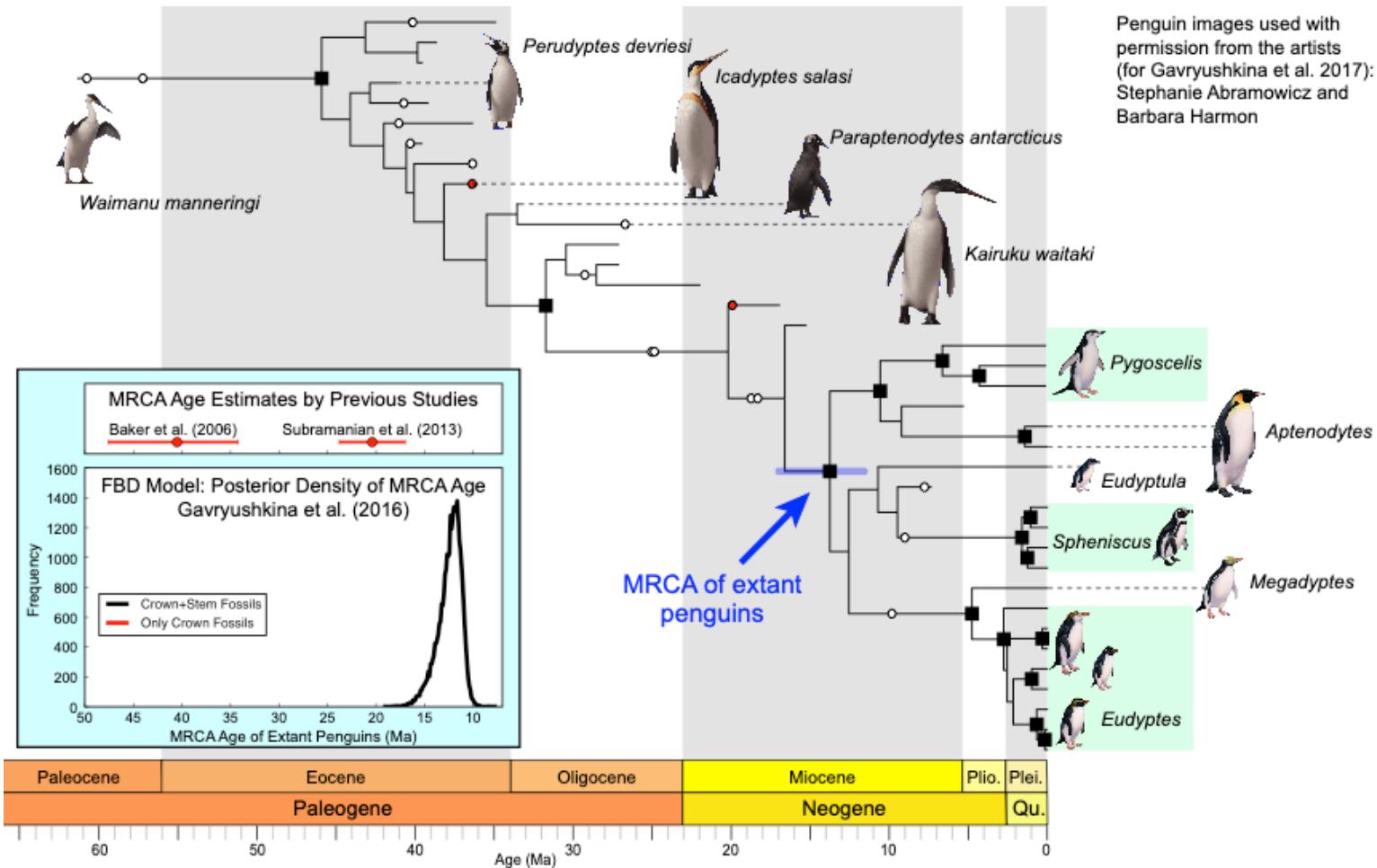
Gavryushkina et al. 2017. *Sys Bio*

Dating extant penguins



Gavryushkina et al. 2017. *Sys Bio*

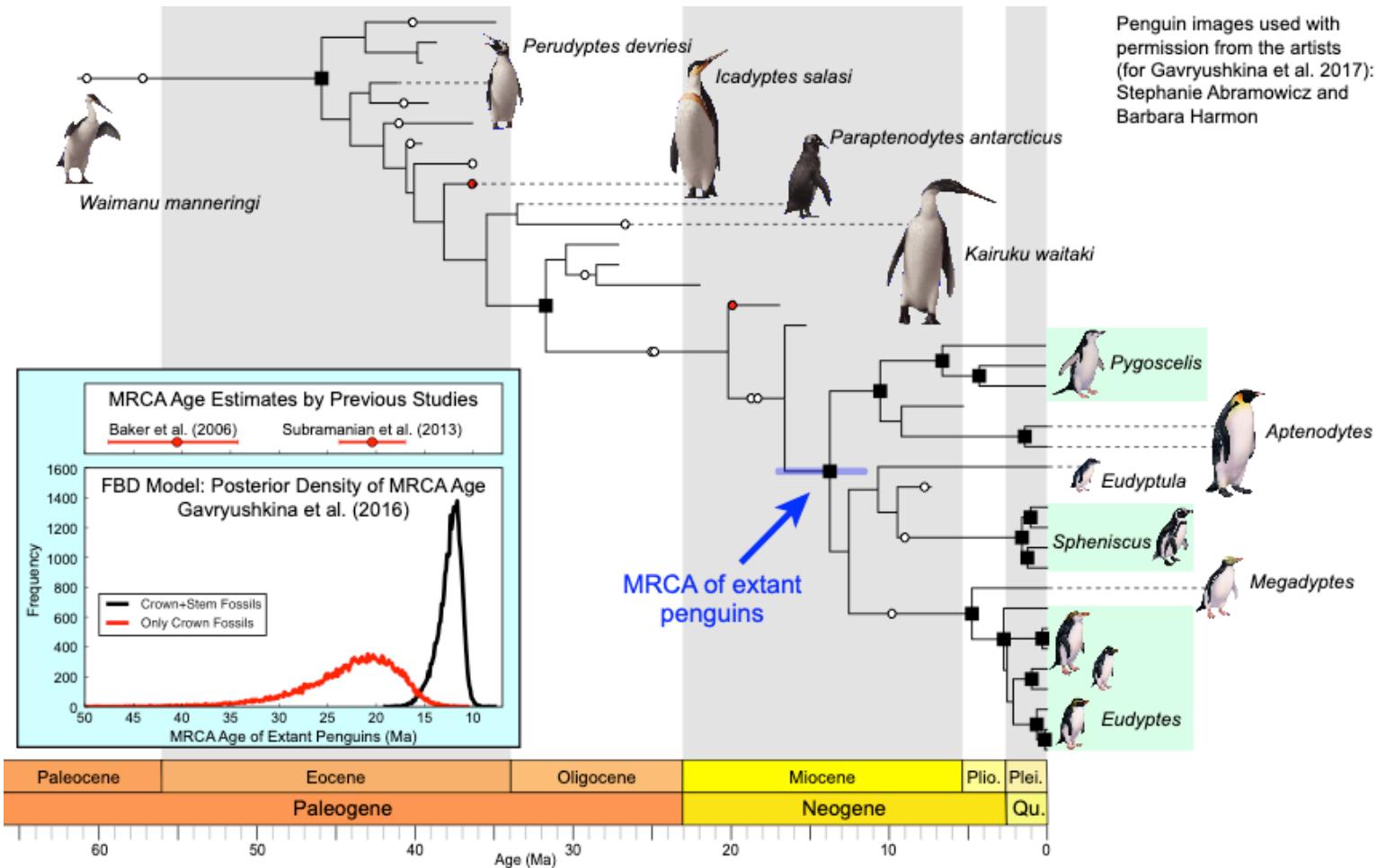
Dating extant penguins



Penguin images used with permission from the artists (for Gavryushkina et al. 2017):
Stephanie Abramowicz and Barbara Harmon

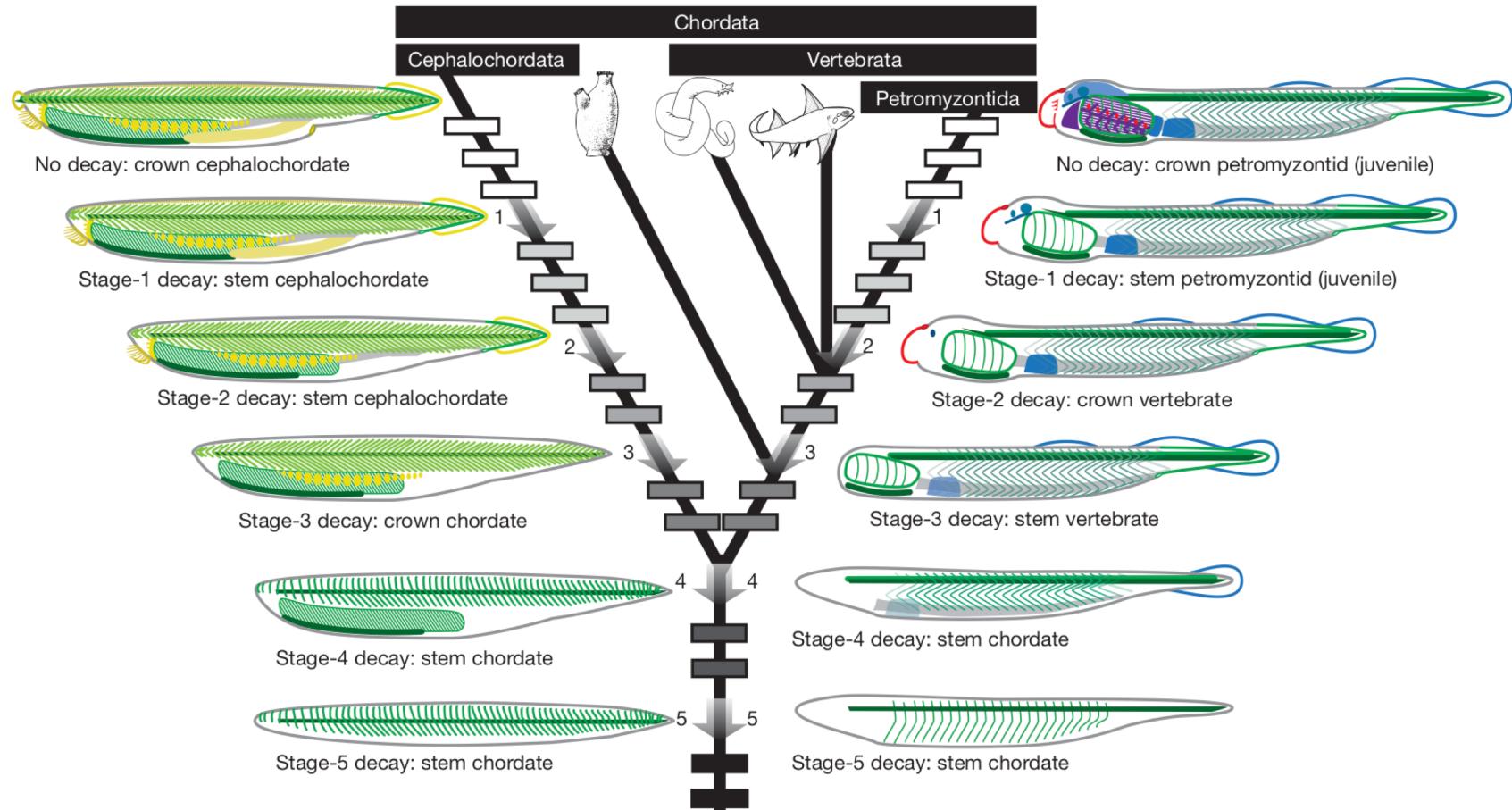
Gavryushkina et al. 2017. *Sys Bio*

Dating extant penguins



Gavryushkina et al. 2017. Sys Bio

Morphological uncertainty



Sansom et al. 2010. Nature

In summary (fossils)

The FBD process allows you to:

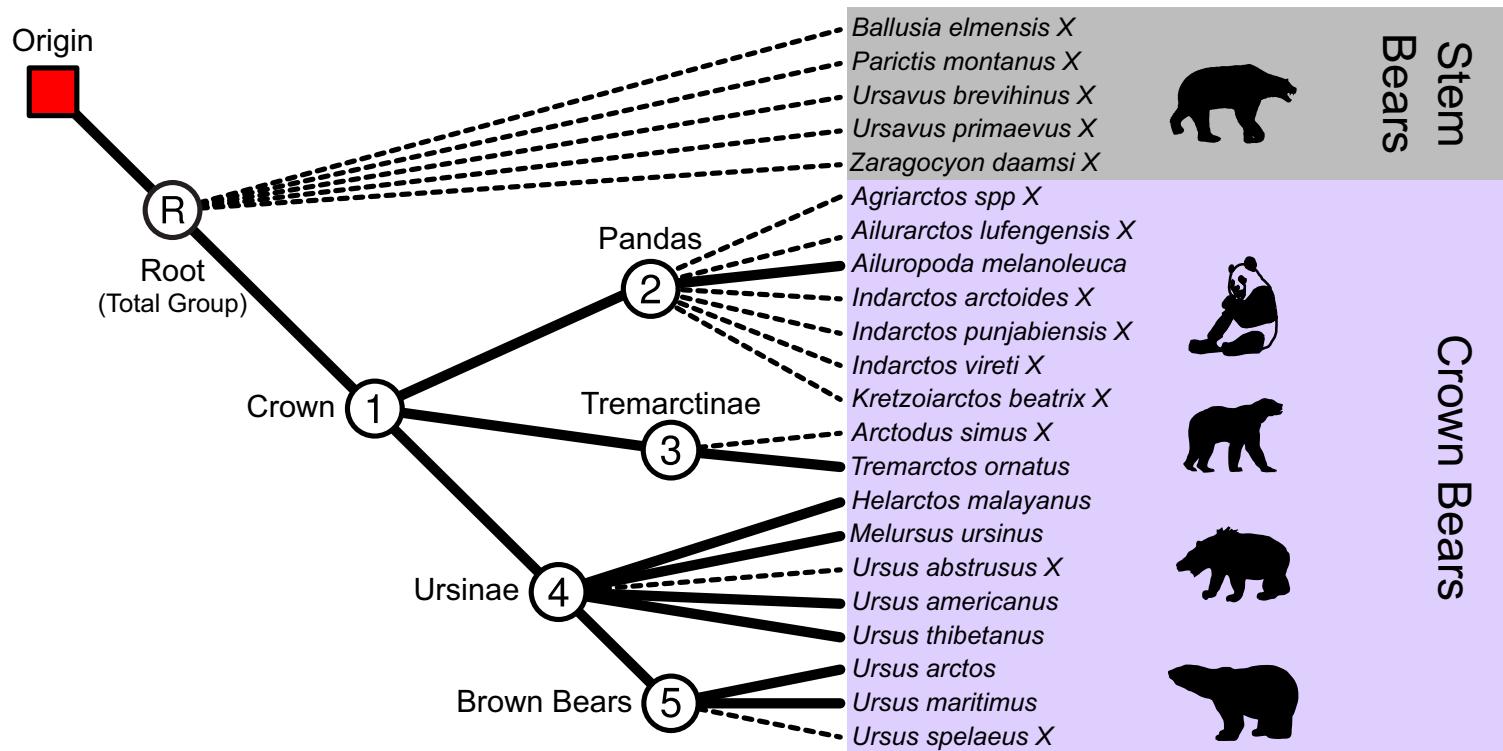
- integrate fossil information into the tree prior
- use all the available fossil information, including stem fossils
- integrate the uncertainty around fossil ages
- infer the placement of fossils in the topology, using morphological data

In summary (epidemics)

The FBD process allows you to:

- analyze datasets where sampling is not linked to removal from the infectious population
- integrate the uncertainty around sample ages

Next: FBD tutorial in BEAST2



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