Fossils and phylogenies

Bethany Allen

Adapted from slides by Joëlle Barido-Sottani

Fossils

Age information

Location or geography

Environment

Anatomy -> **morphological characters**

Fossils represent a snapshot of life at a point in deep time

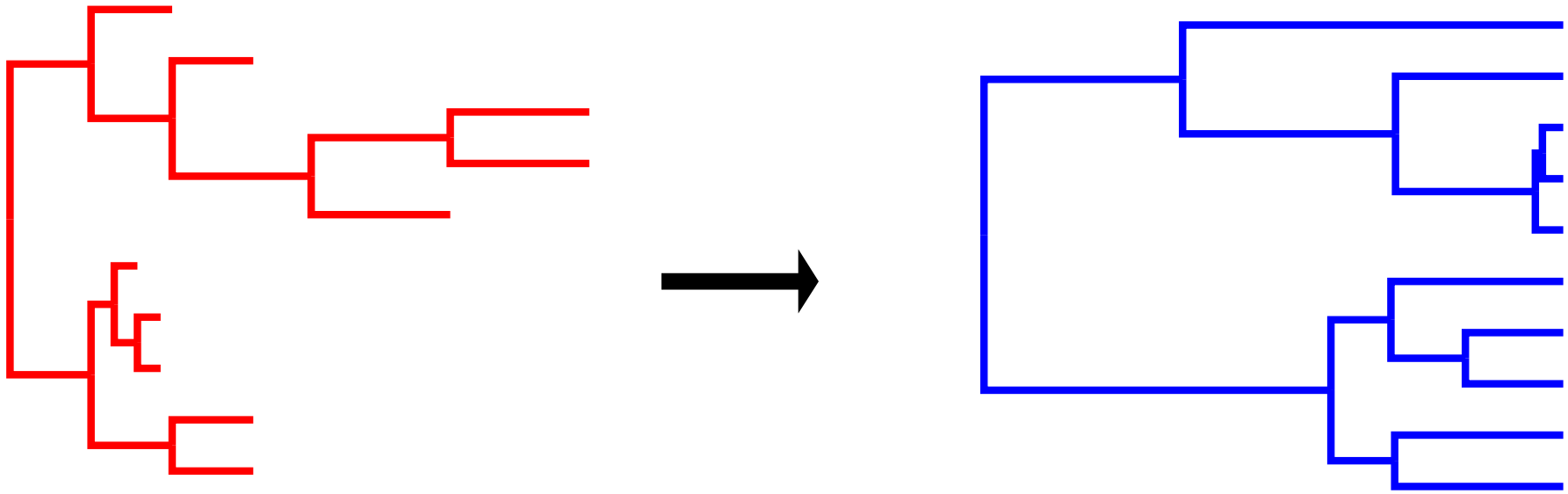
How can we use fossils in phylogenetics?

To timescale a phylogeny using **calibrated nodes**

To timescale a phylogeny using **fossils as age information**

To include **fossil taxa as tips** in a phylogeny

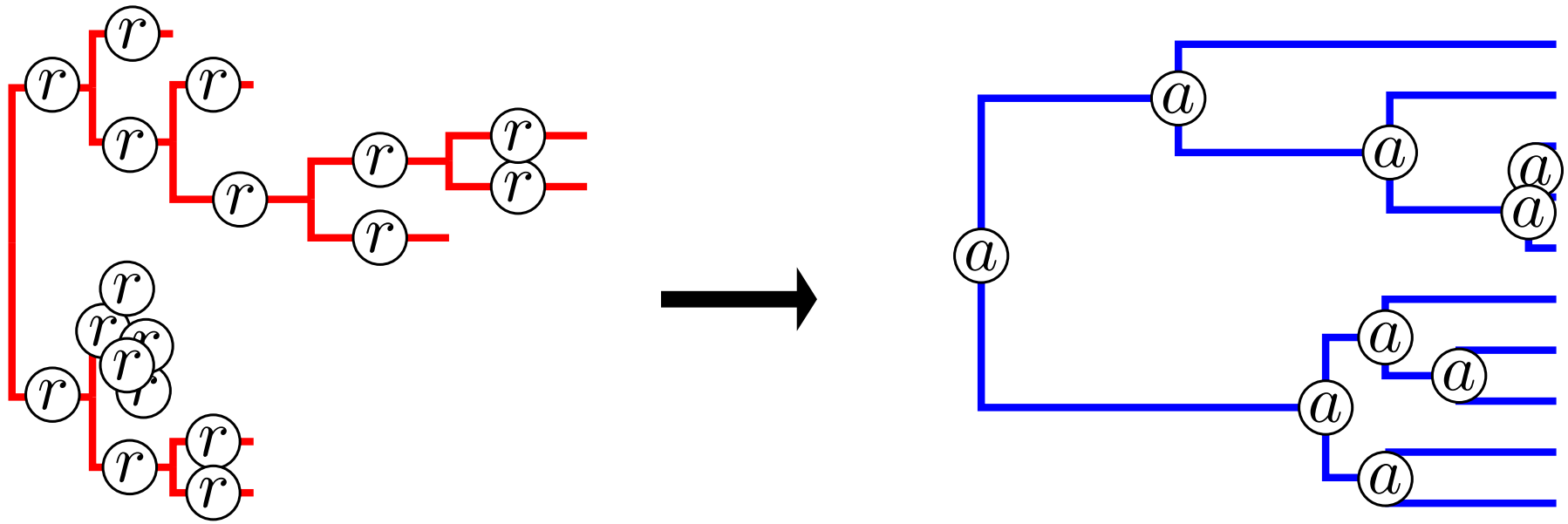
Divergence time estimation



Branch lengths represent
substitutions

Branch lengths represent
time

Divergence time estimation

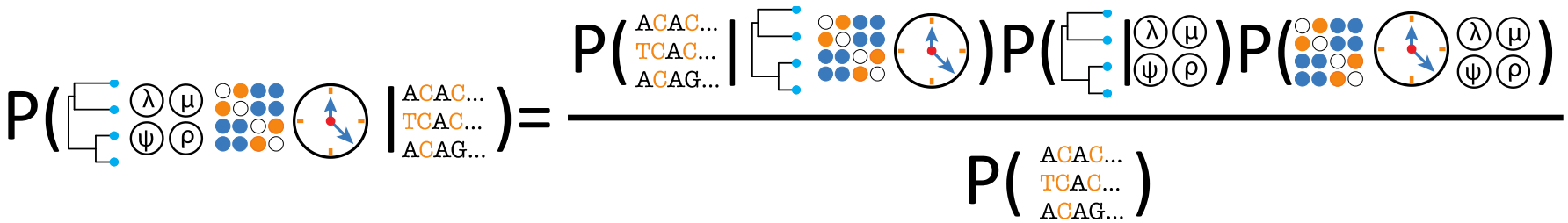


Branch lengths represent
substitutions

Branch lengths represent
time

$R = (r_1, r_2, r_3, \cdots, r_{2N-2})$
$A = (a_1, a_2, a_3, \cdots, a_{N-1})$
N = number of tips

Bayesian divergence time estimation



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

Sequence data

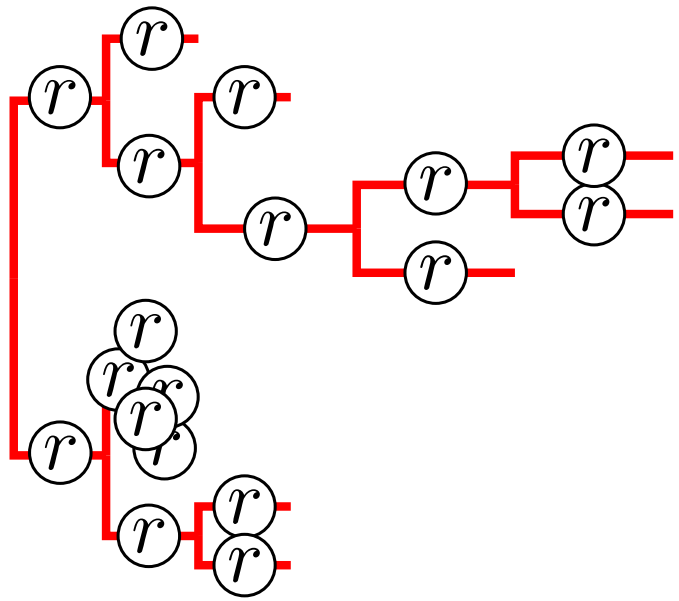
Substitution model

Clock model

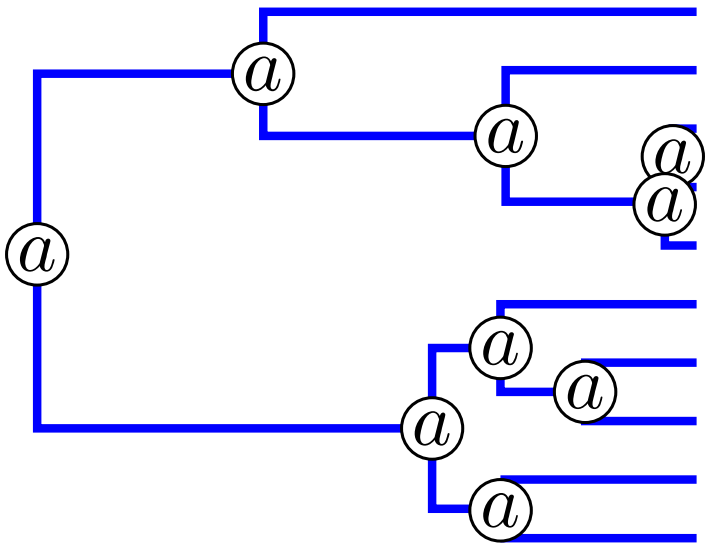
Time tree

Tree prior

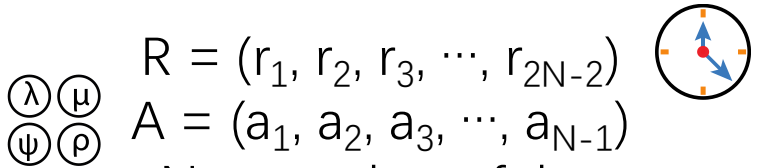
Divergence time estimation



Branch lengths represent
substitutions



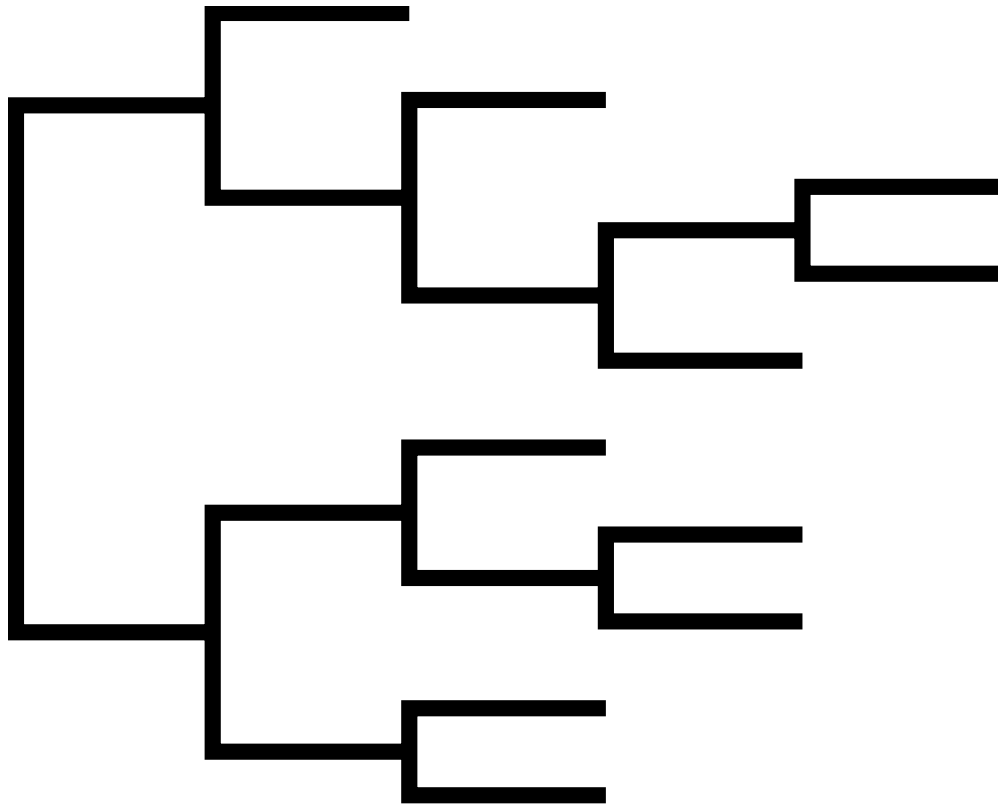
Branch lengths represent
time



$R = (r_1, r_2, r_3, \cdots, r_{2N-2})$
$A = (a_1, a_2, a_3, \cdots, a_{N-1})$
N = number of tips

Global molecular clock

All branches share the same clock rate

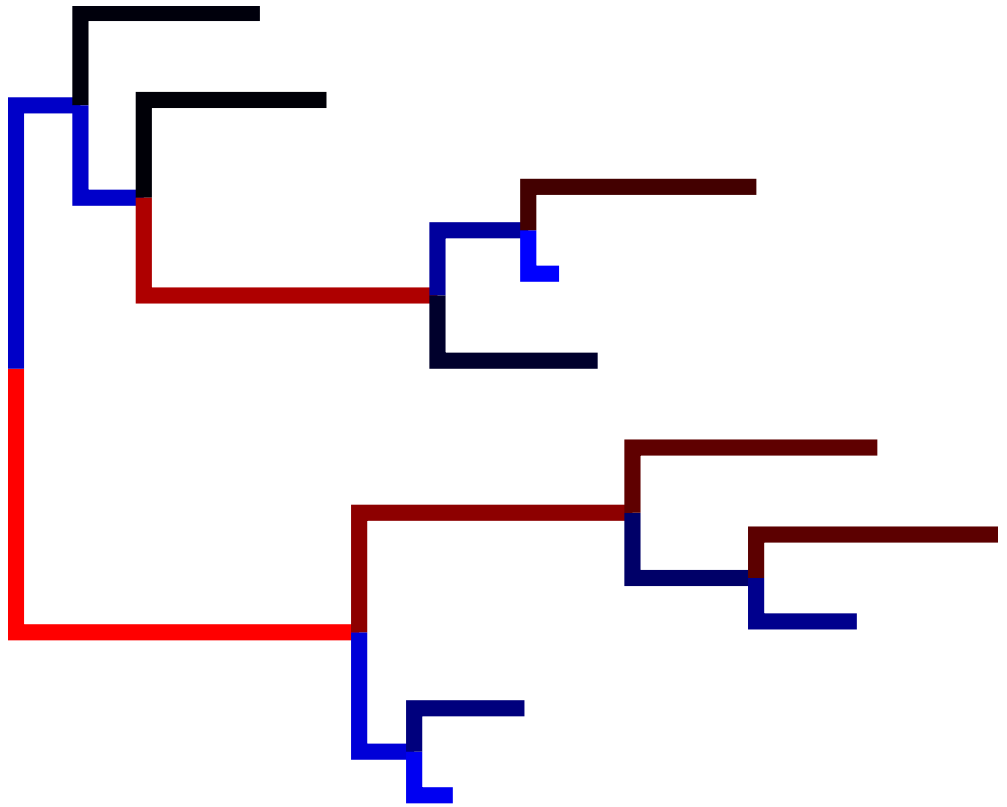


branch length = substitution rate

low high

Uncorrelated relaxed clock

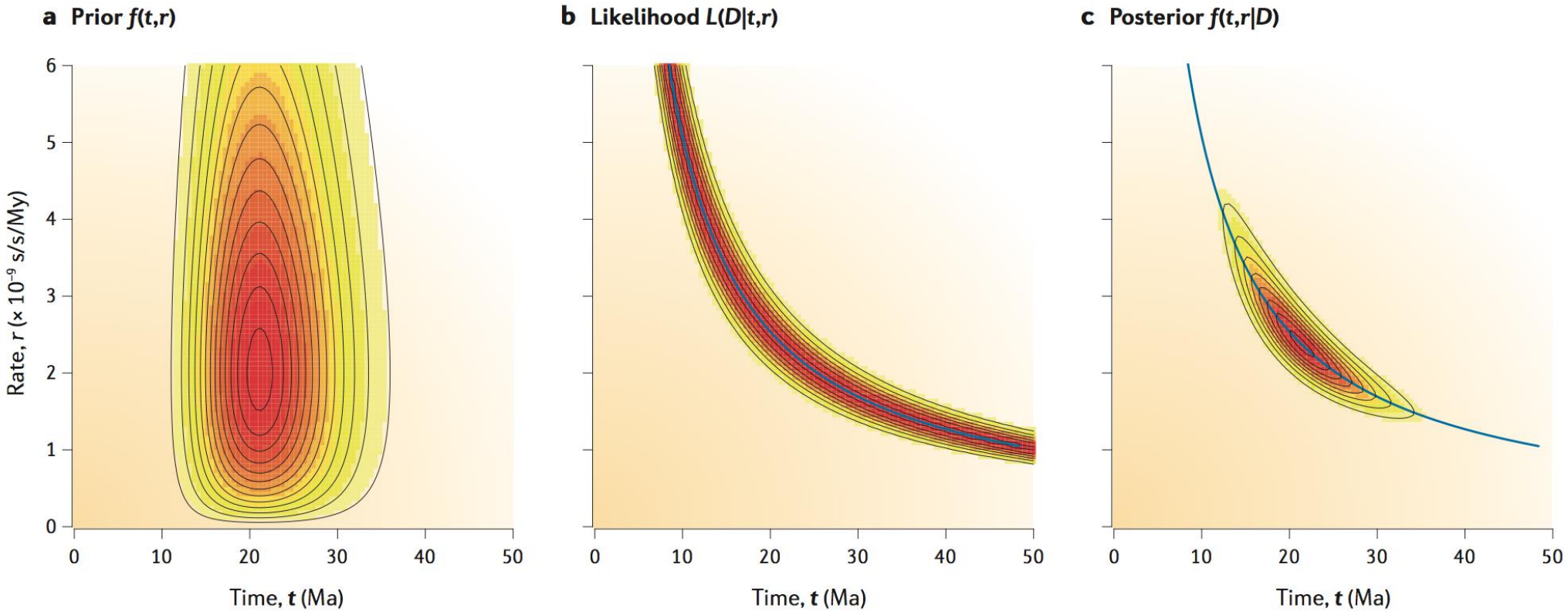
Each branch has its own clock rate, drawn from a global distribution (here, a lognormal distribution)



branch length = substitution rate

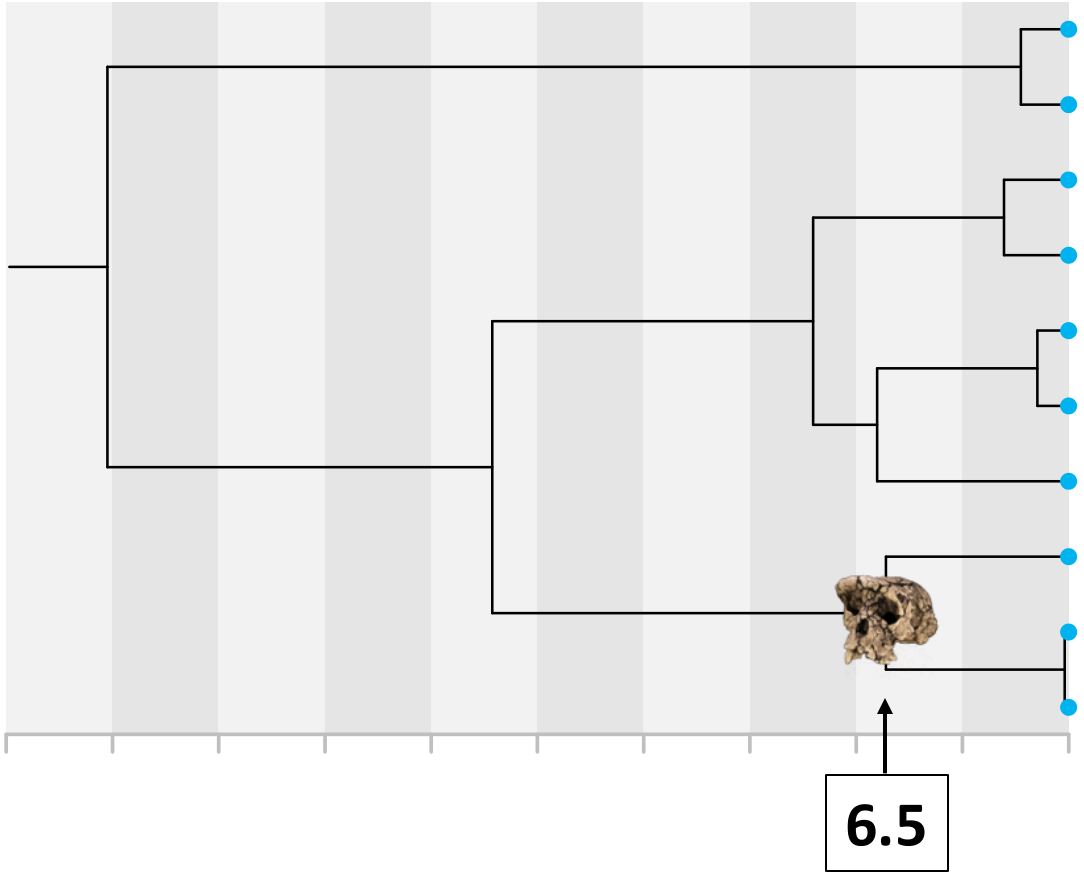
low high

Identifiability

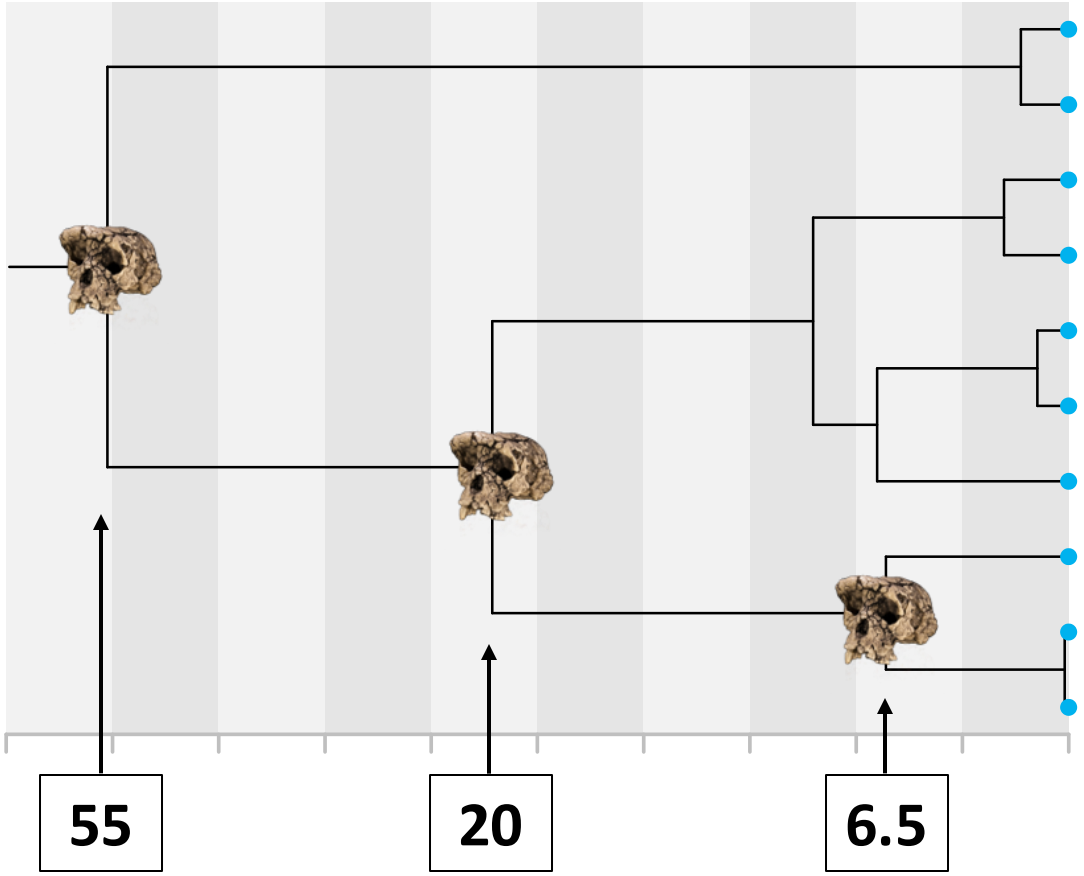


dos Reis et al. (2016)

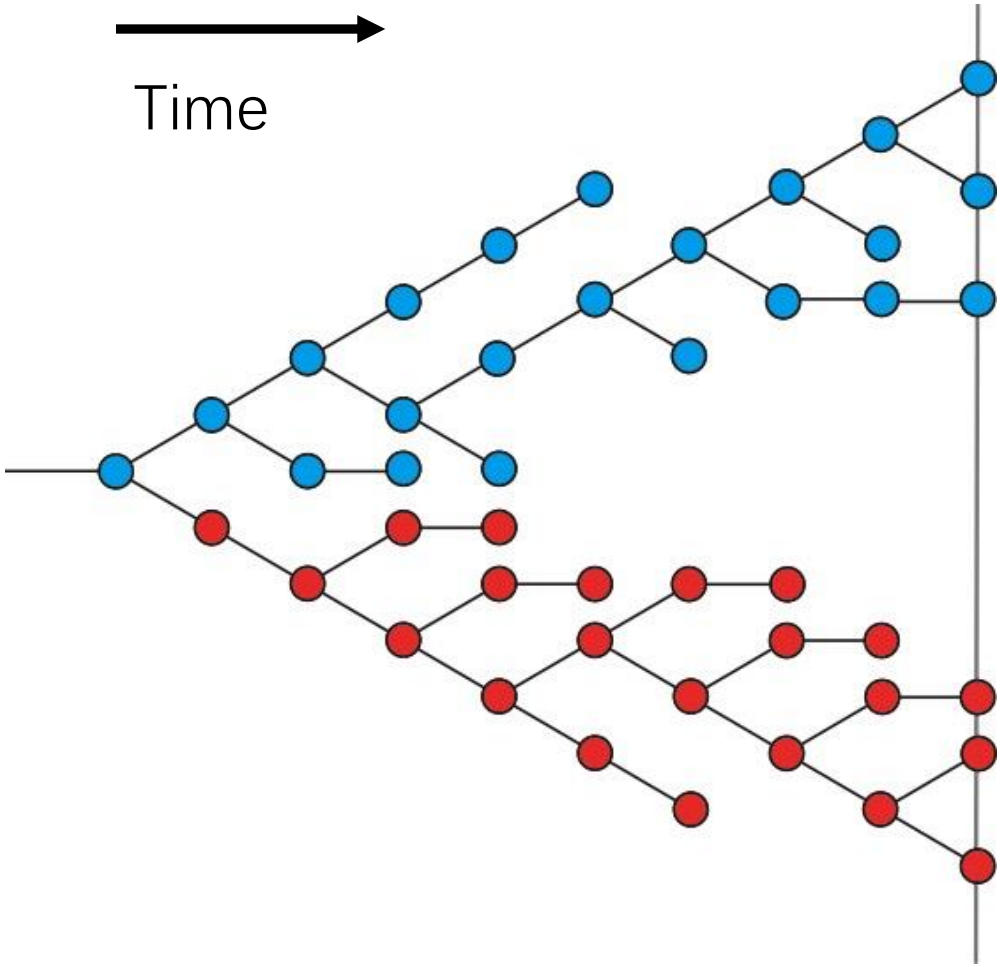
Fossil calibration



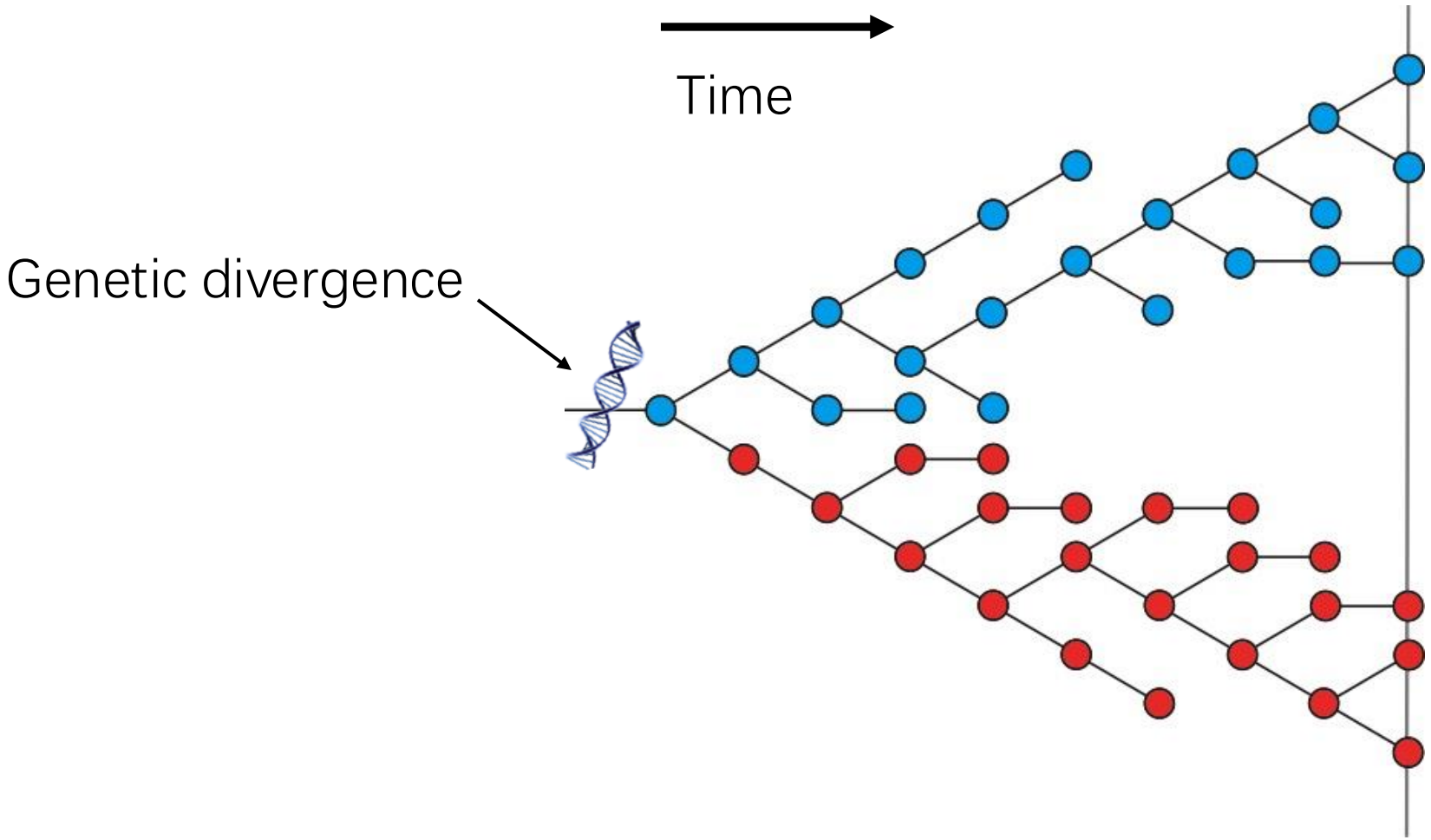
Fossil calibration



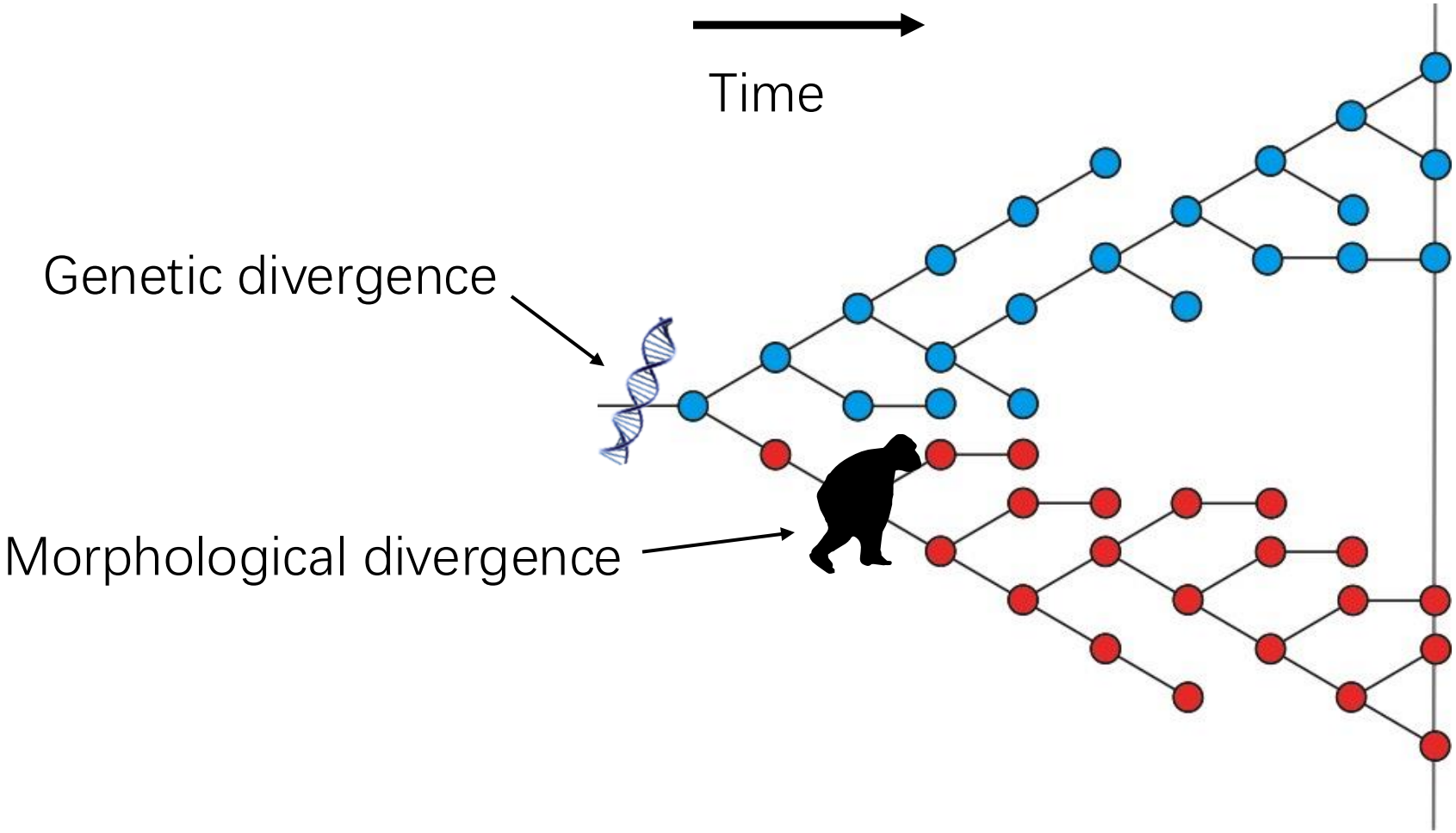
How do fossils constrain age?



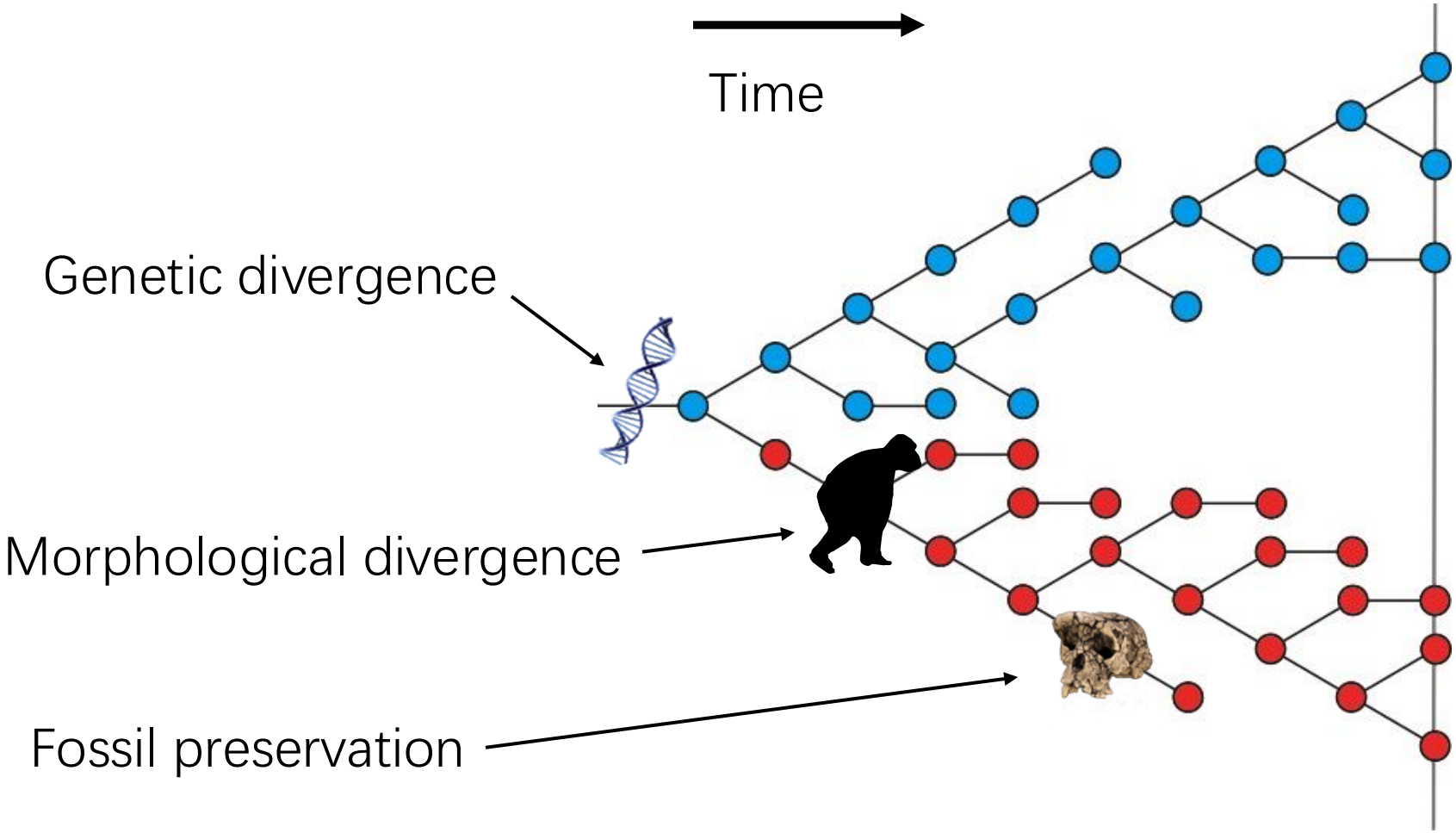
How do fossils constrain age?



How do fossils constrain age?

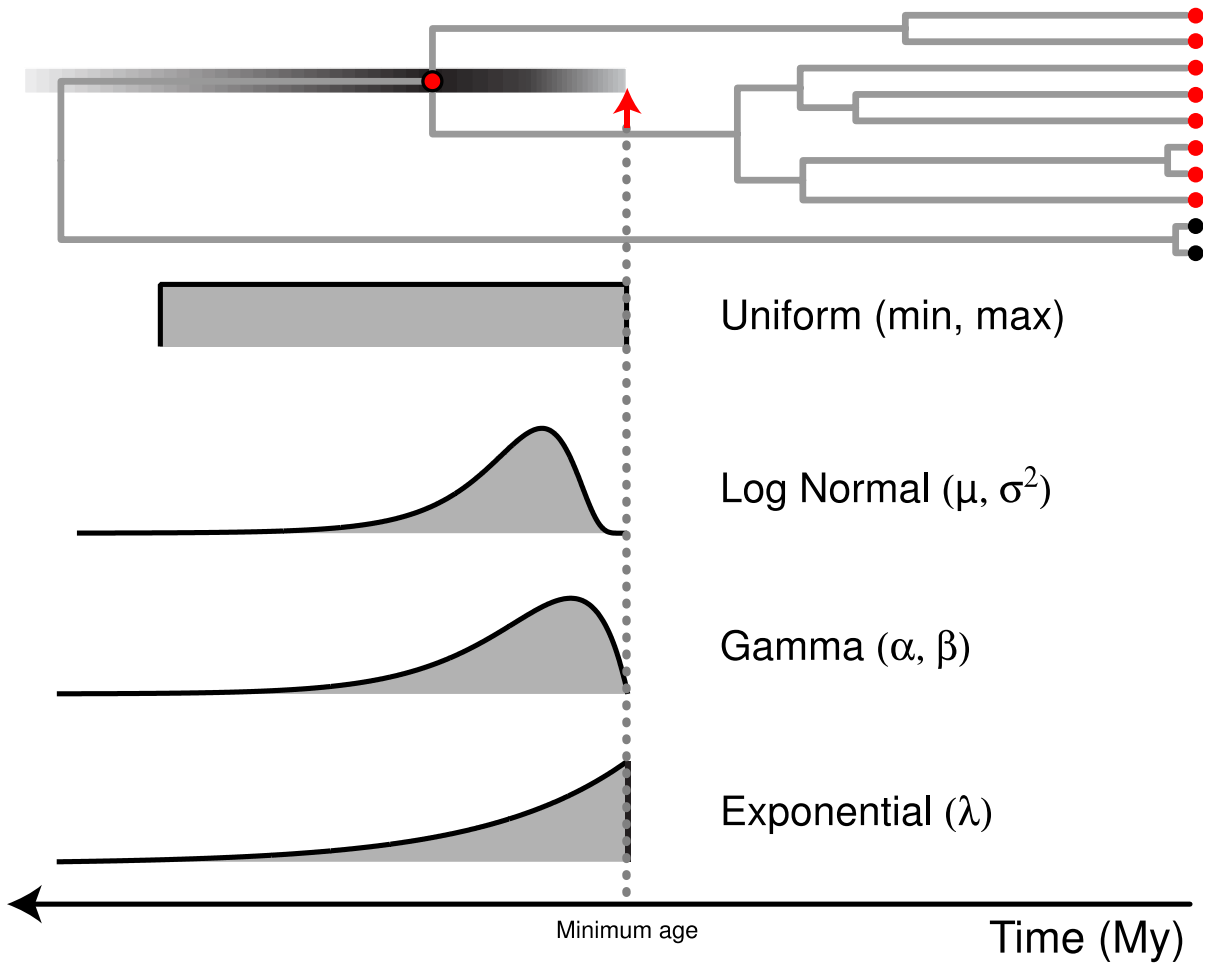


How do fossils constrain age?



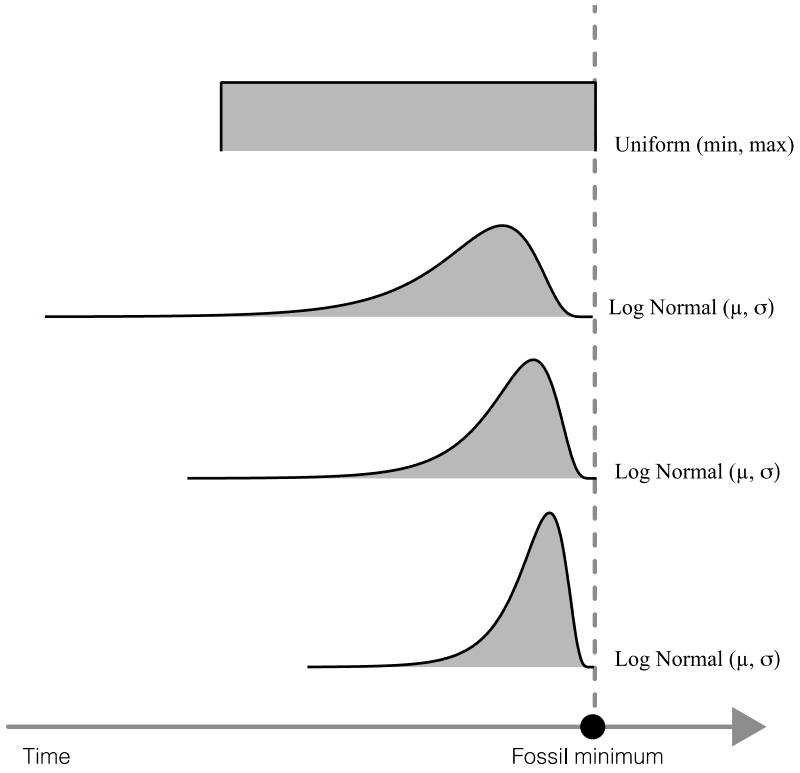
Calibration density

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

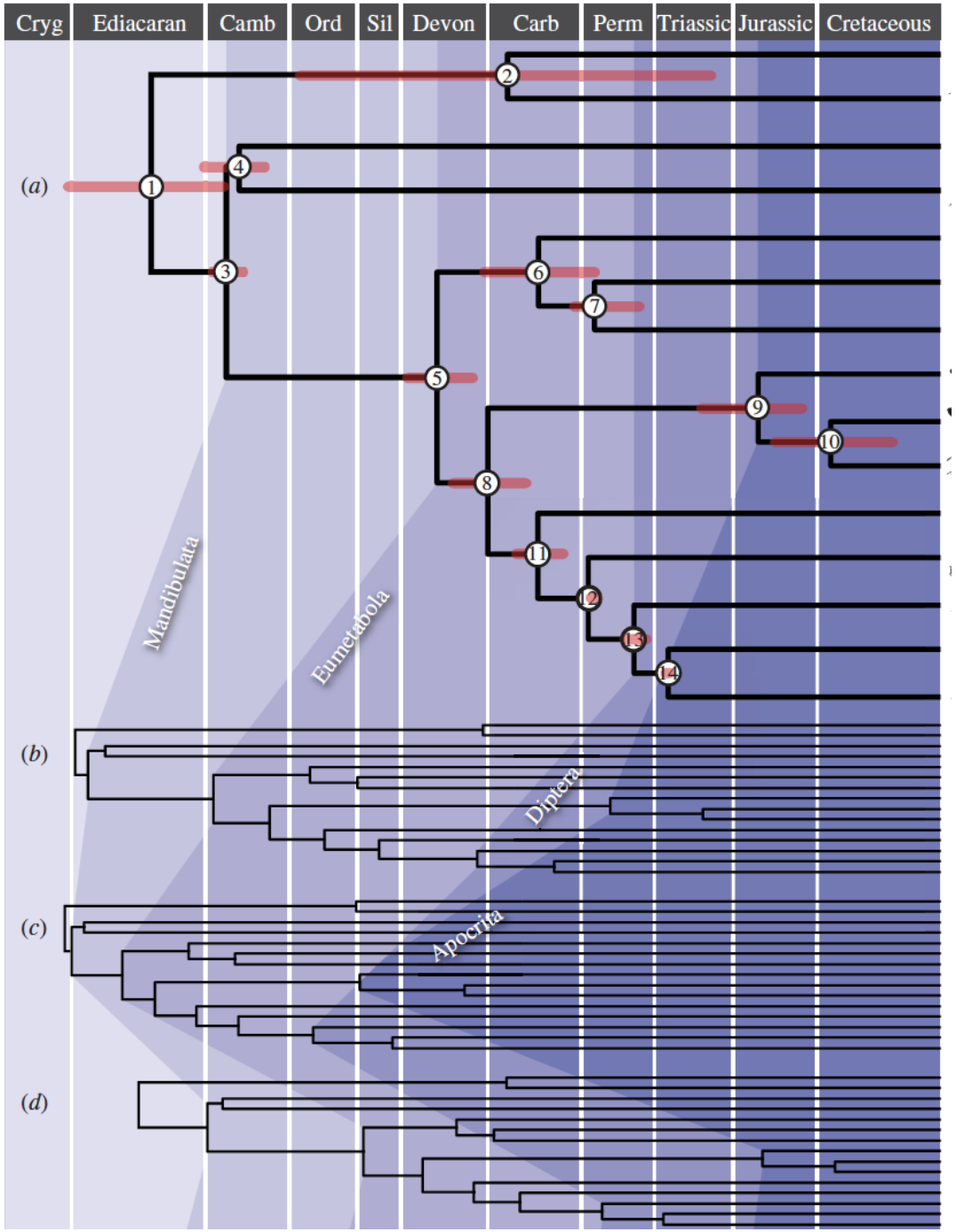


Heath (2012)

Prior shape matters



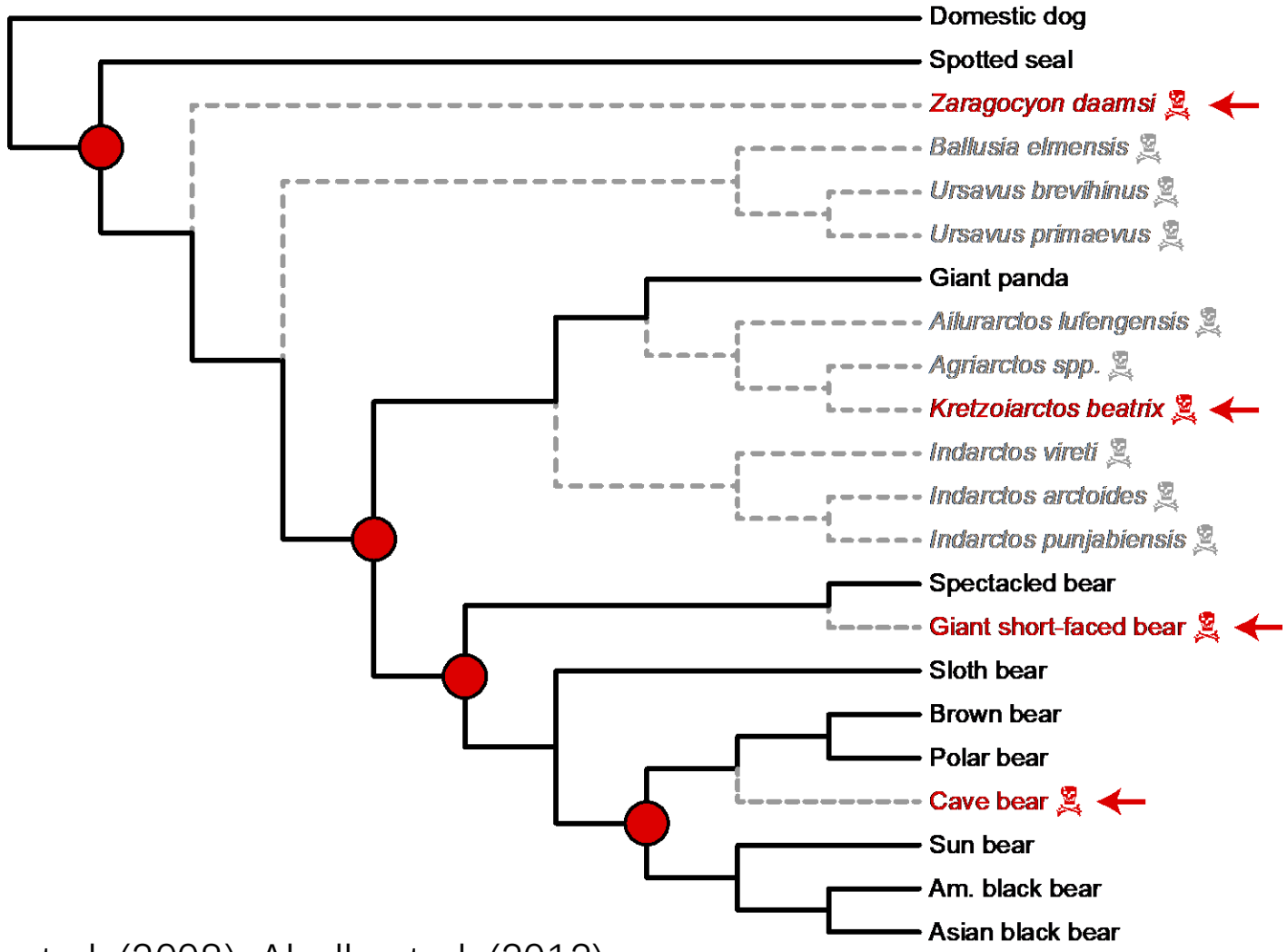
Warnock et al. (2012)



Why is calibration controversial?

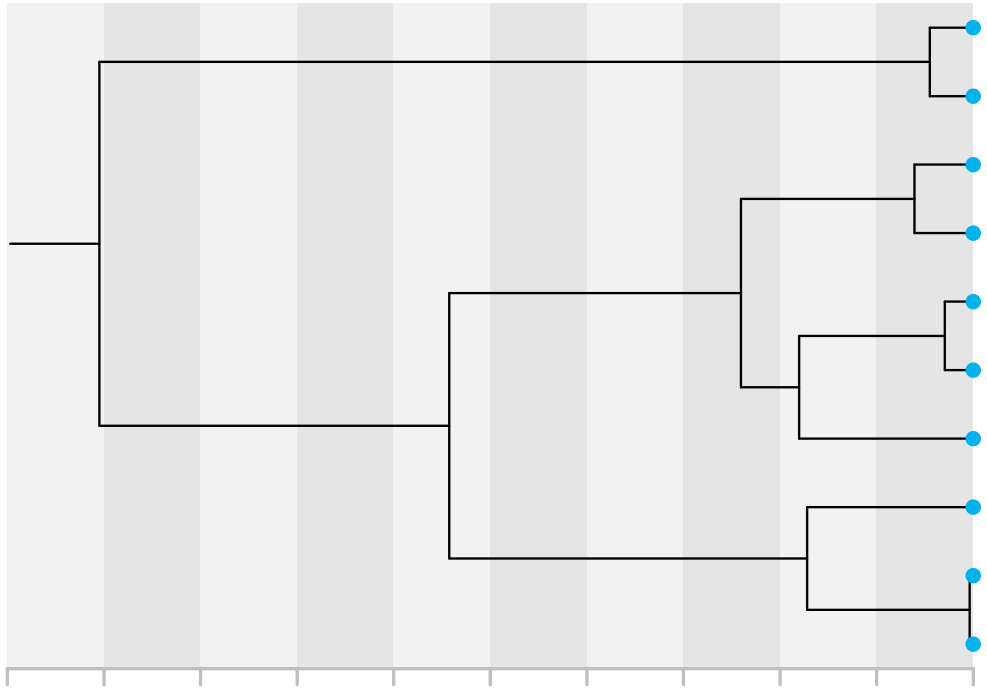
Node calibration only uses the oldest fossil occurrence for each node

Choosing the right fossil for a node is not easy!



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

The birth-death process

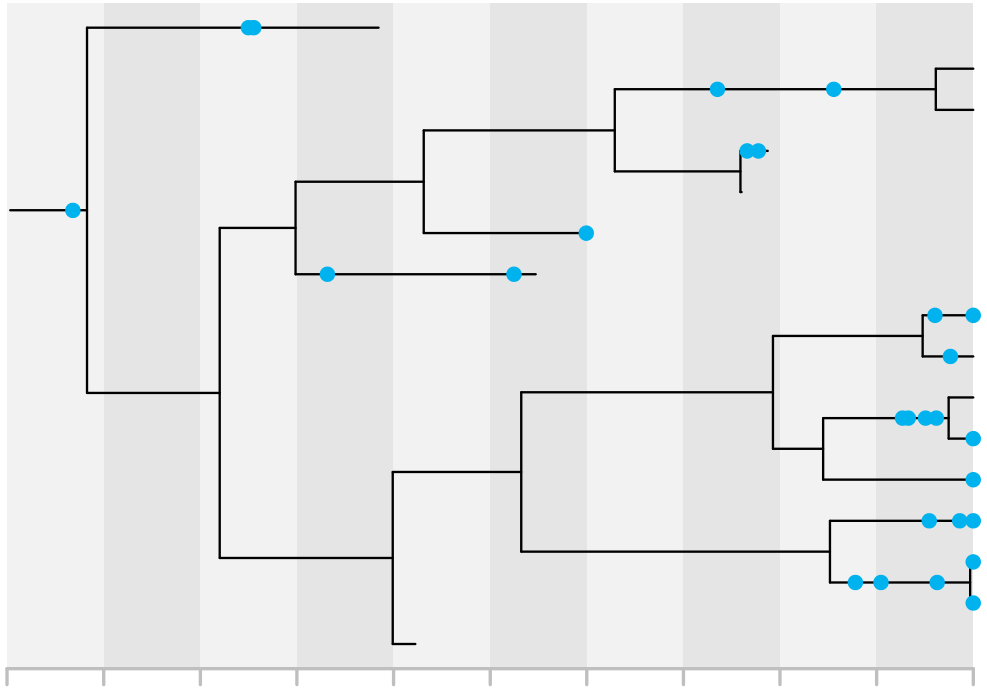


λ — speciation rate

μ — extinction rate

ρ — extant species
sampling probability

The fossilised birth-death process



λ — speciation rate

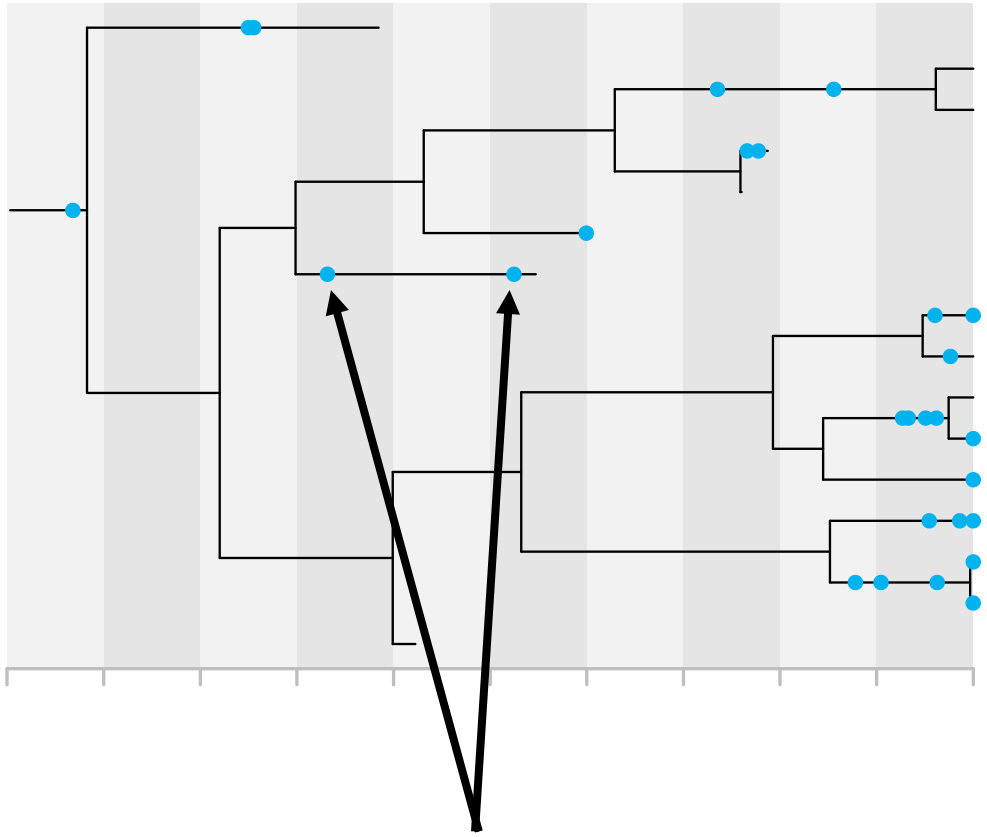
μ — extinction rate

ψ — fossil sampling rate

ρ — extant species sampling probability

Stadler (2010)

Sampled ancestors



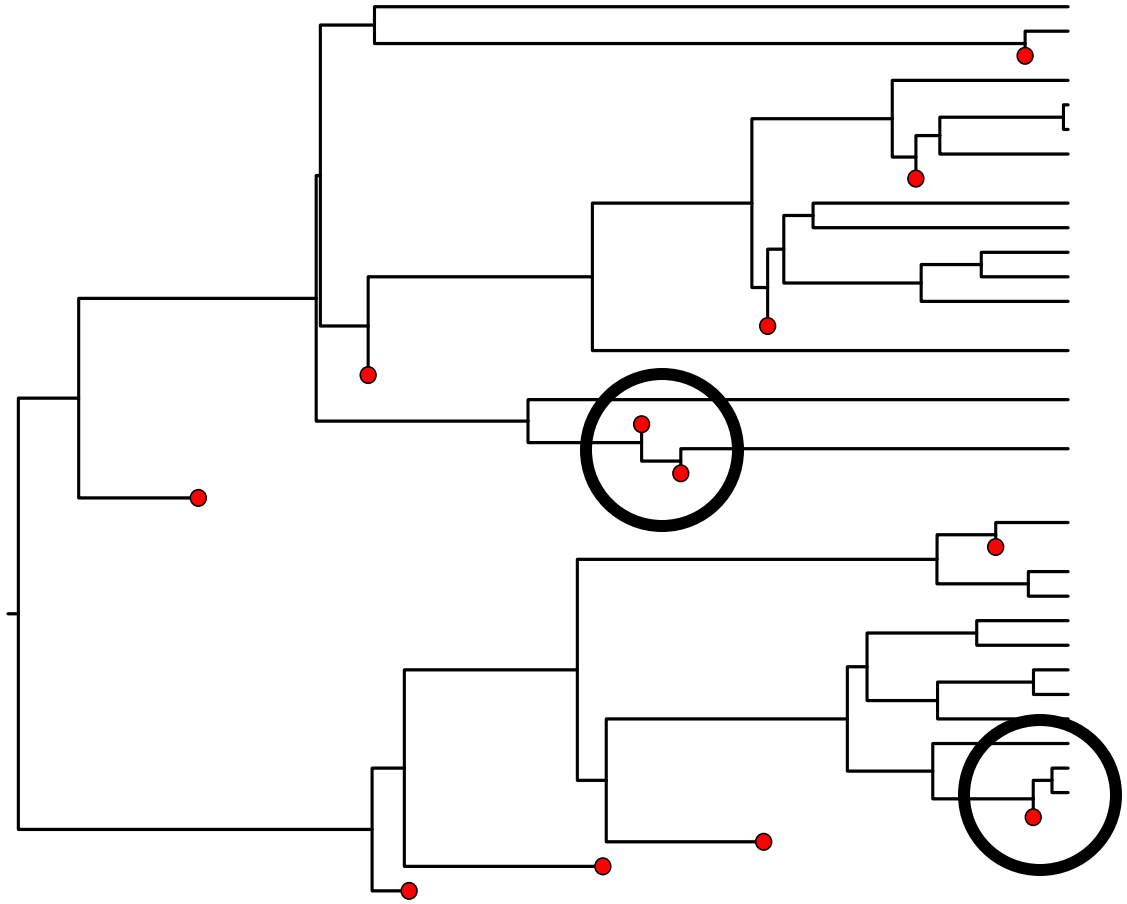
Lineages with **more than one sample**

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

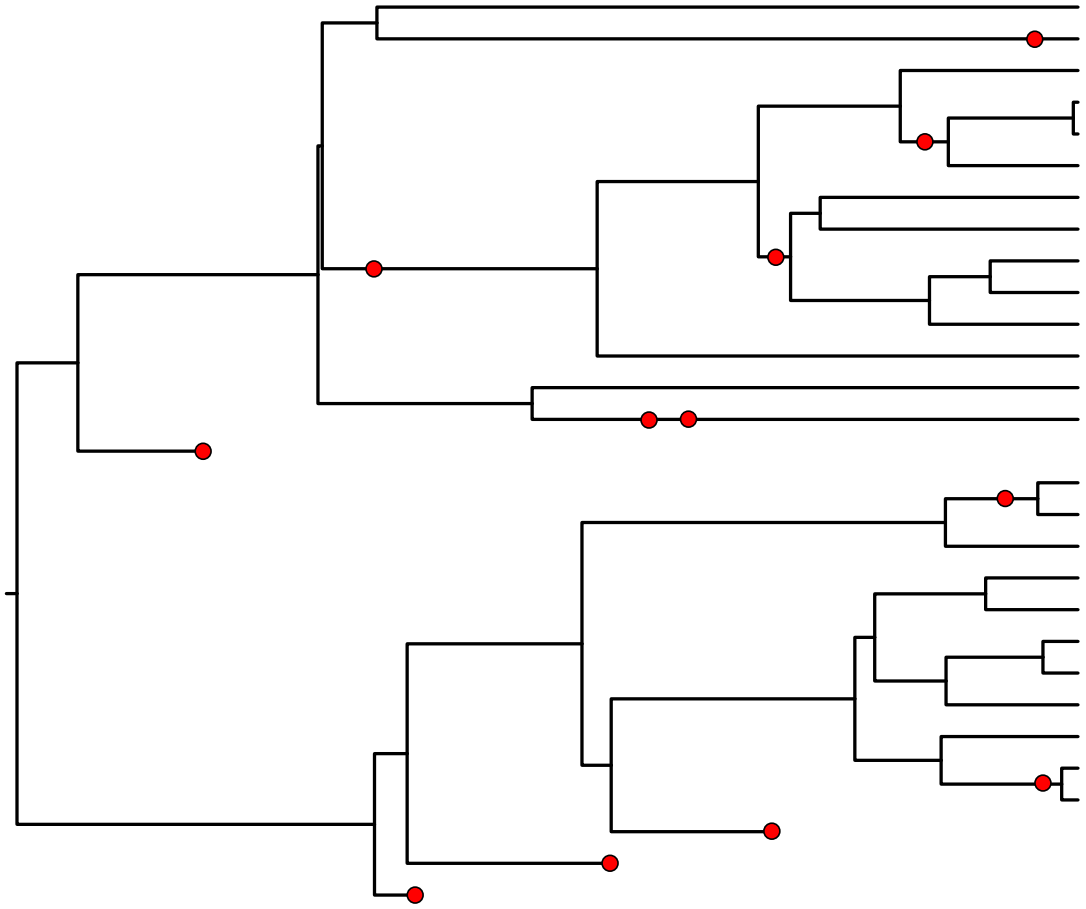
On the probability of ancestors in the fossil record

Mike Foote

Sampled ancestors

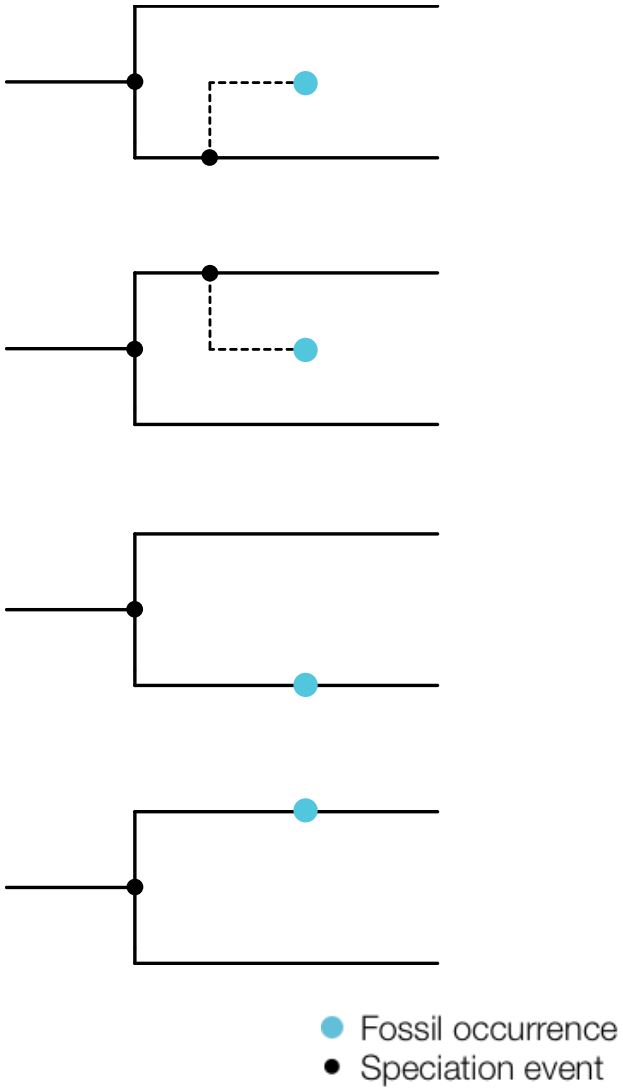


Without sampled ancestors



With sampled ancestors

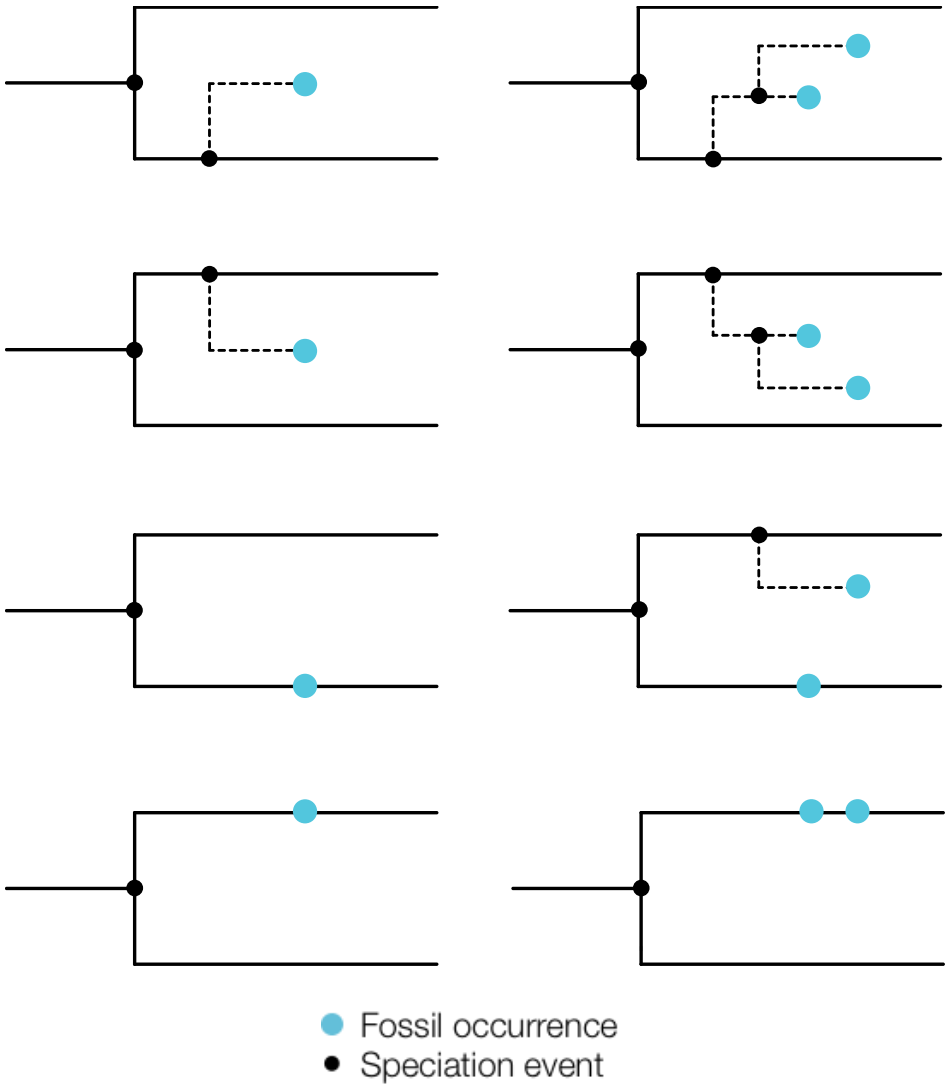
MCMC implementation



A fossil sample can be placed:

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

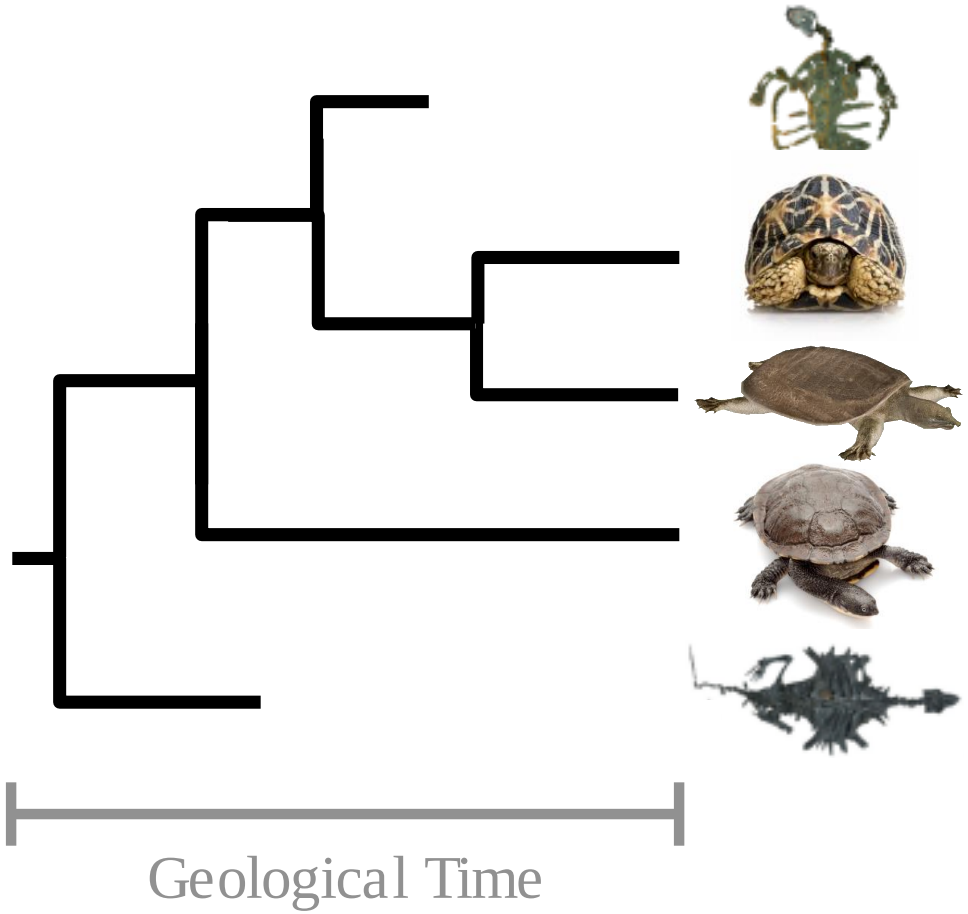
MCMC implementation



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

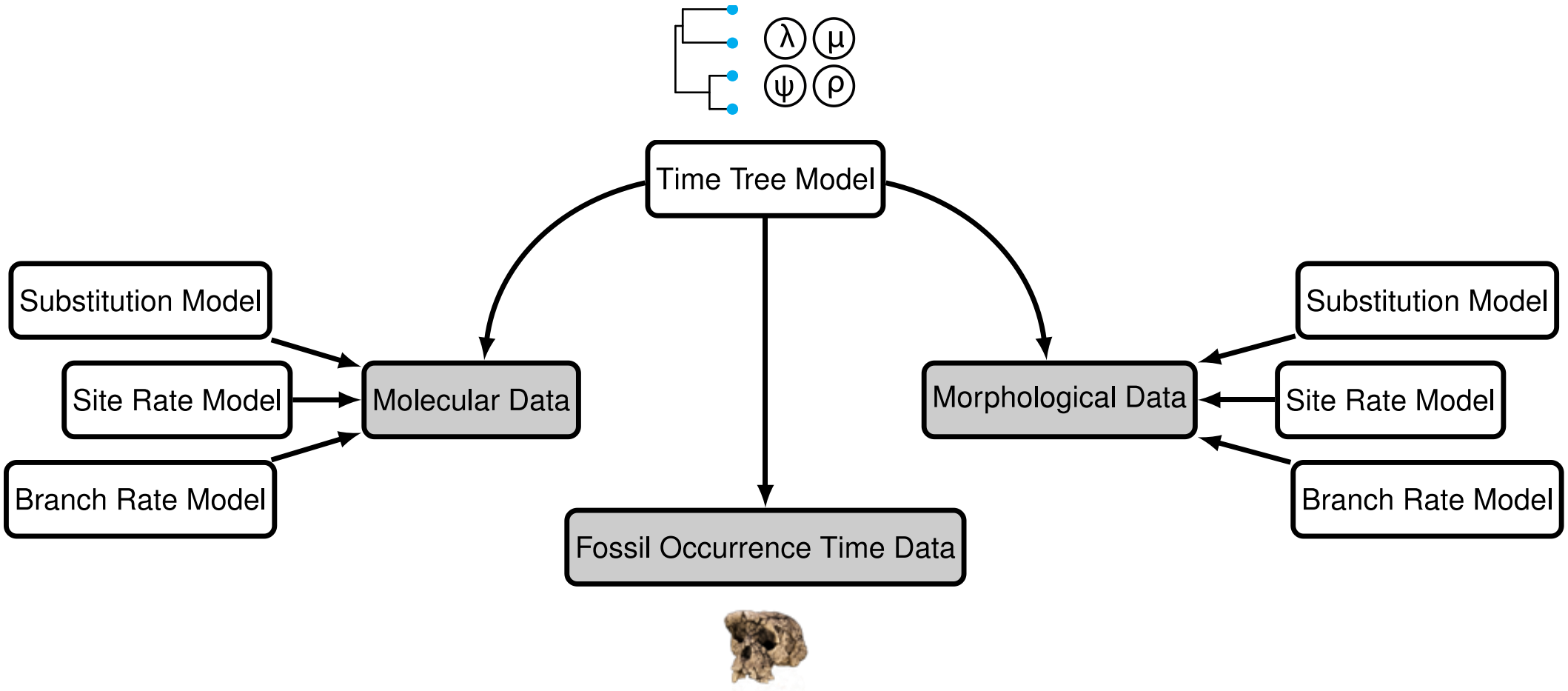
The FBD process gives the probability of each configuration conditioned on the model parameters

Genomic and/or morphological data

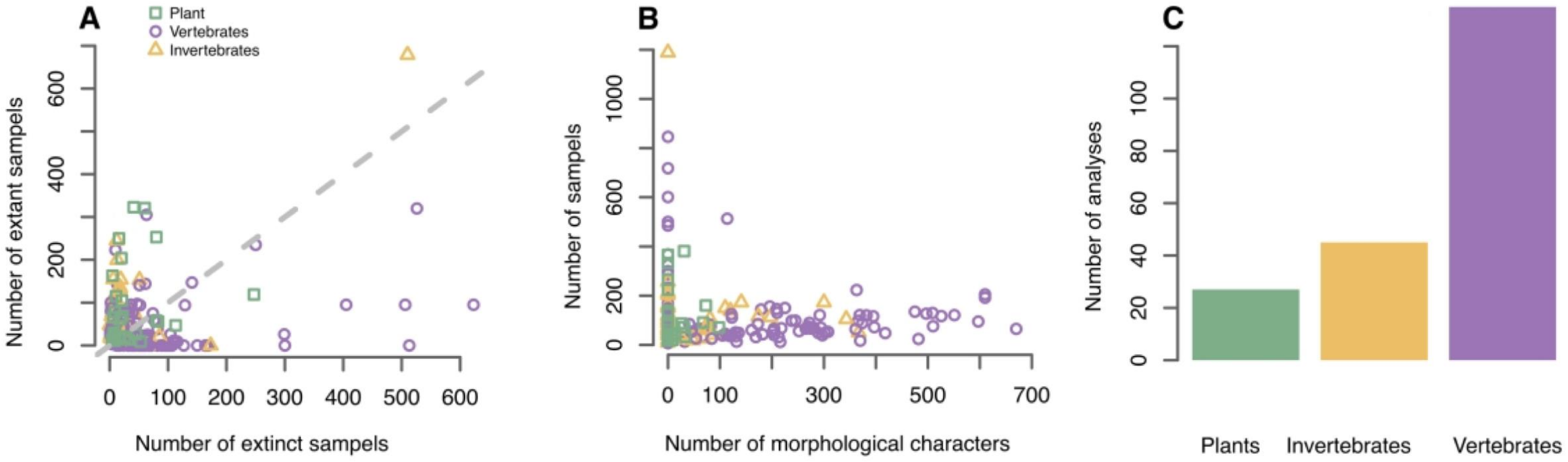


???110211????	????????????????
0001010111001	CATTATTCGCATTA
0101111110114	CATCATTCGAATCA
0011100110005	AATTATCCGAGTAA
1??010000????	????????????????

Total evidence approach



FBD applications

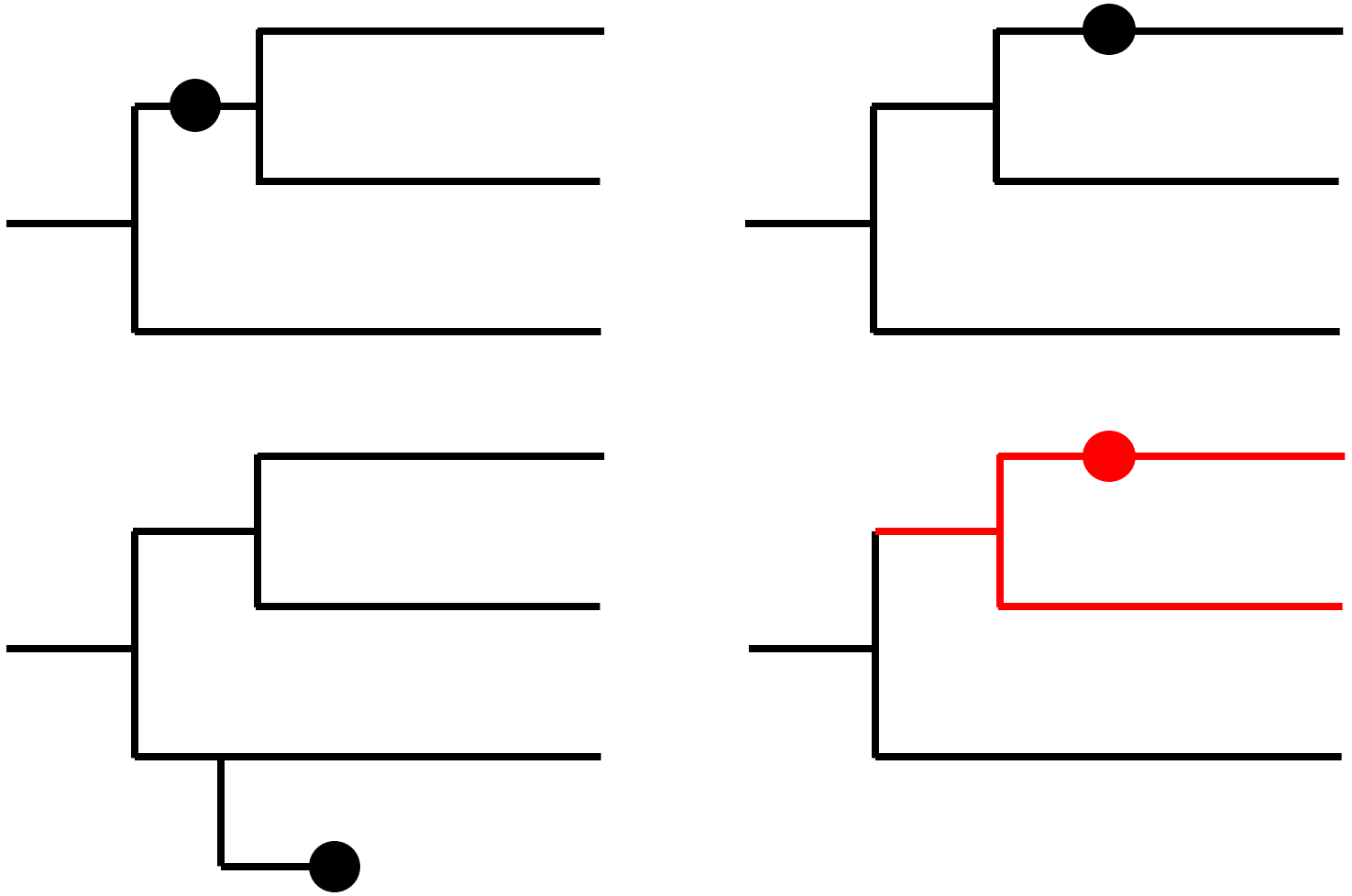


Mulvey et al. (in press)

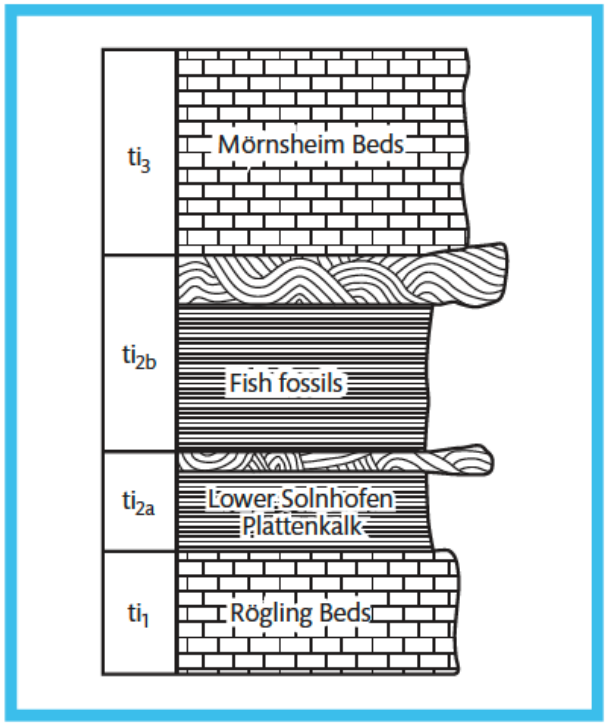
Fossils can inform age without characters

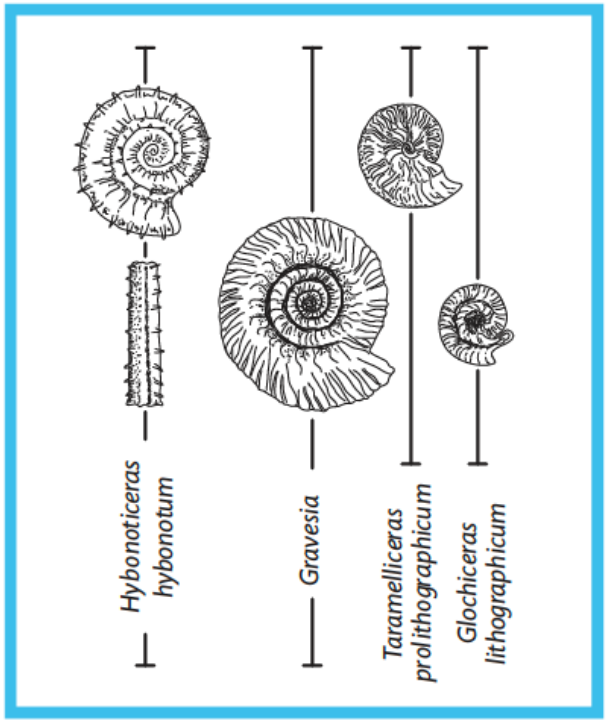
It is possible to insert fossil tips with no sequence data into the phylogeny

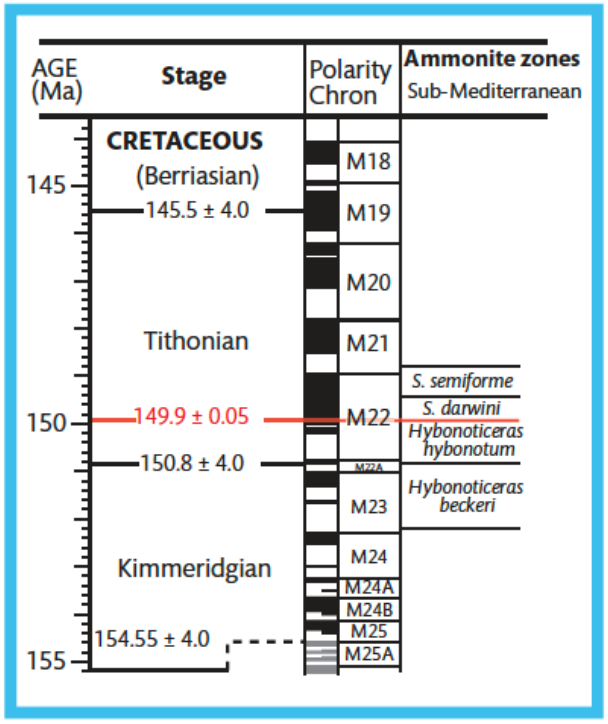
This calibrates the tree but incorporates uncertainty in fossil placement and age



Fossil age uncertainty

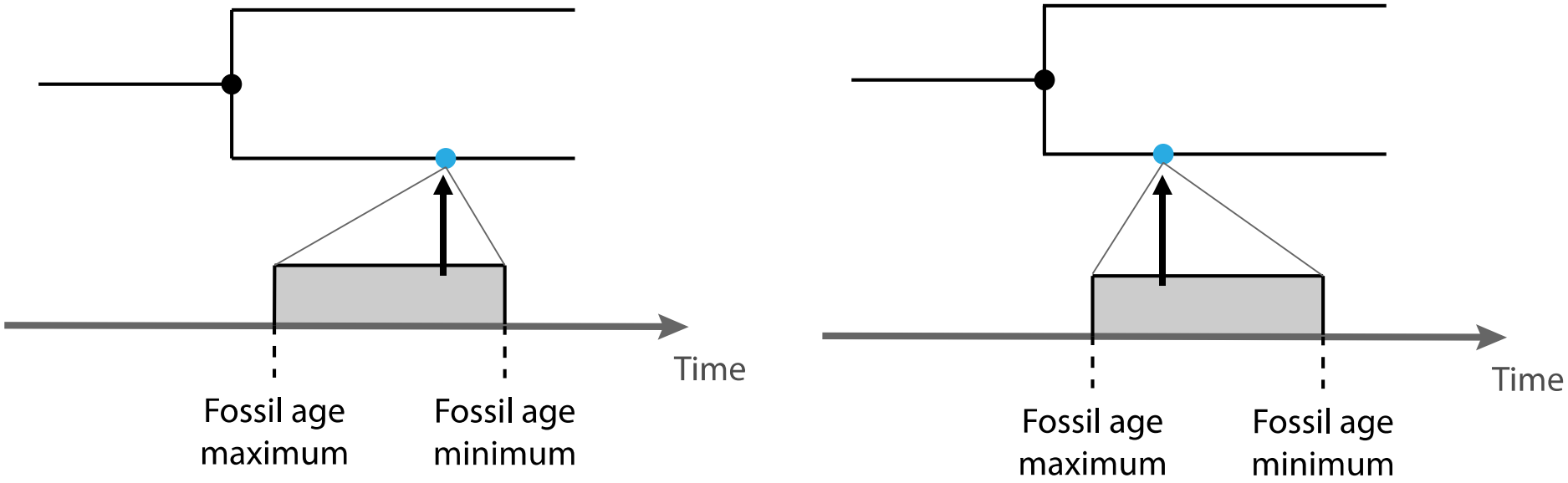






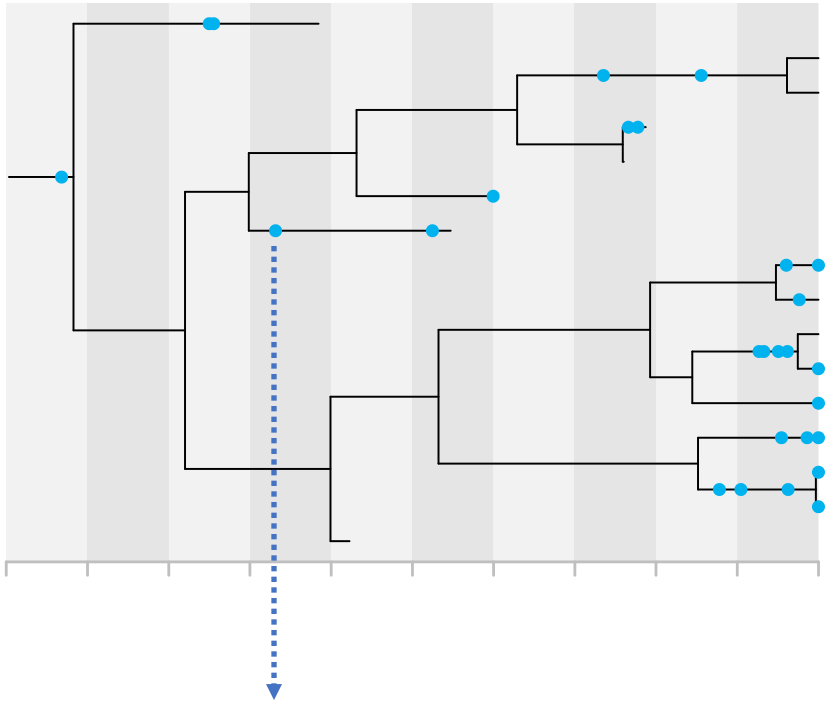
Benton et al. (2009)

Integrating the uncertainty

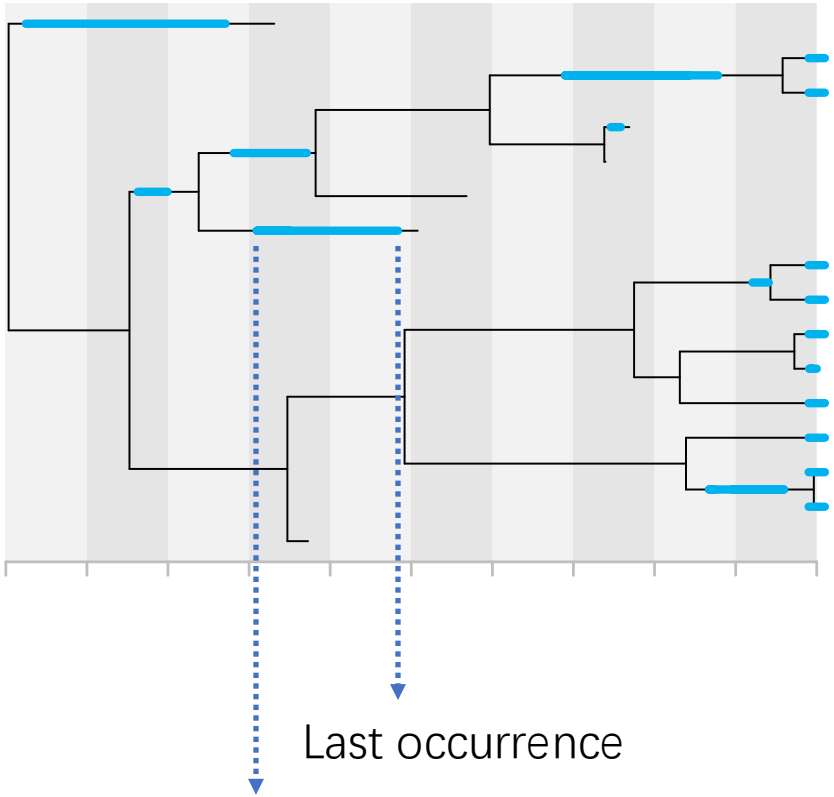


Fossil age uncertainty can (and should!)
be sampled as part of the MCMC

Specimen-level data vs. ranges



Fossil specimen
(one occurrence)



Last occurrence

First occurrence

Varying rates through time
(skyline; Stadler 2011)

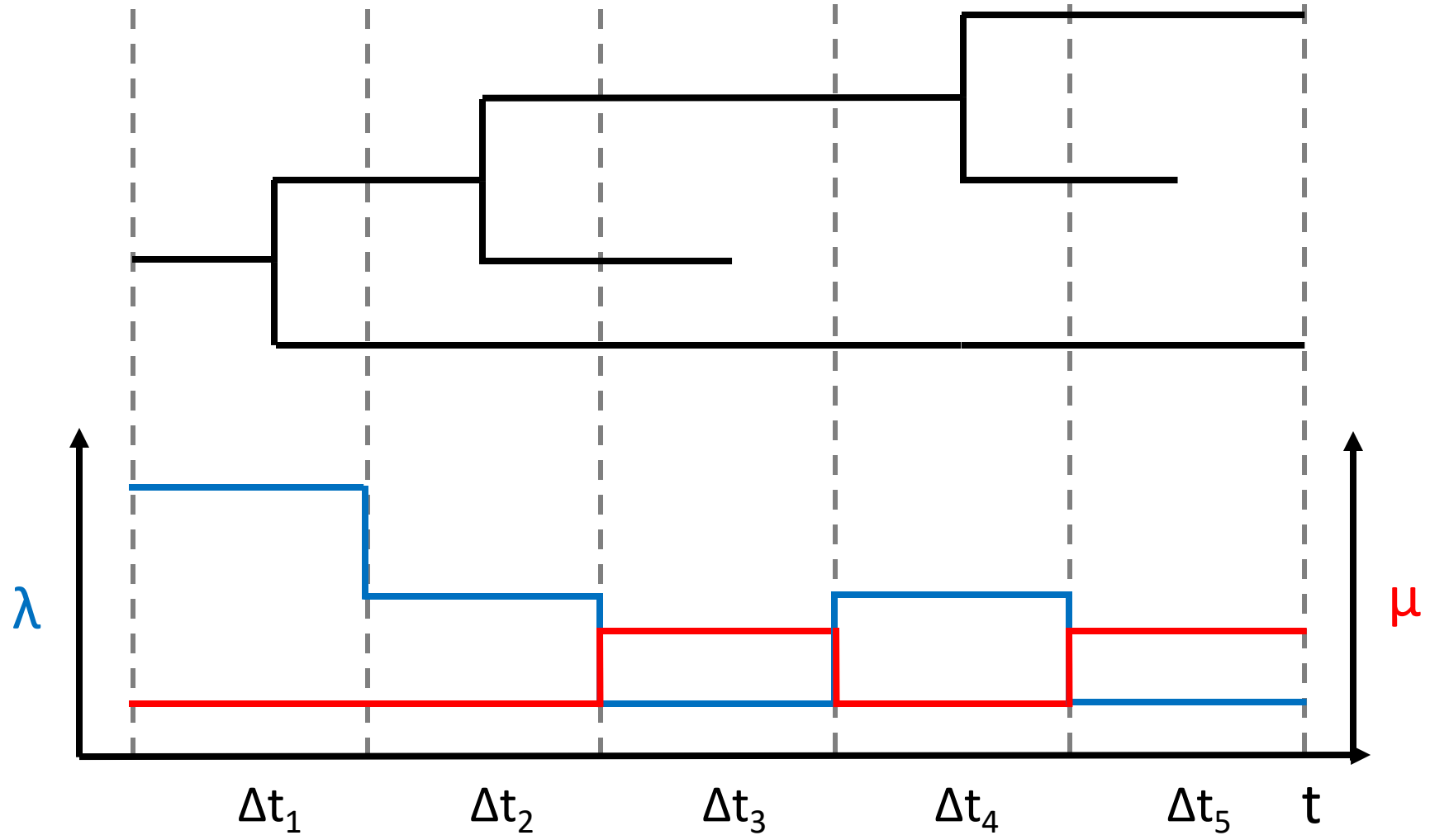
Designating or inferring “types”
(multi-type; Kühnert et al. 2016;
Barido-Sottani et al. 2020)

Incorporating GLMs on
parameters
(GLM; Valenzuela Agüí et al. in
prep)

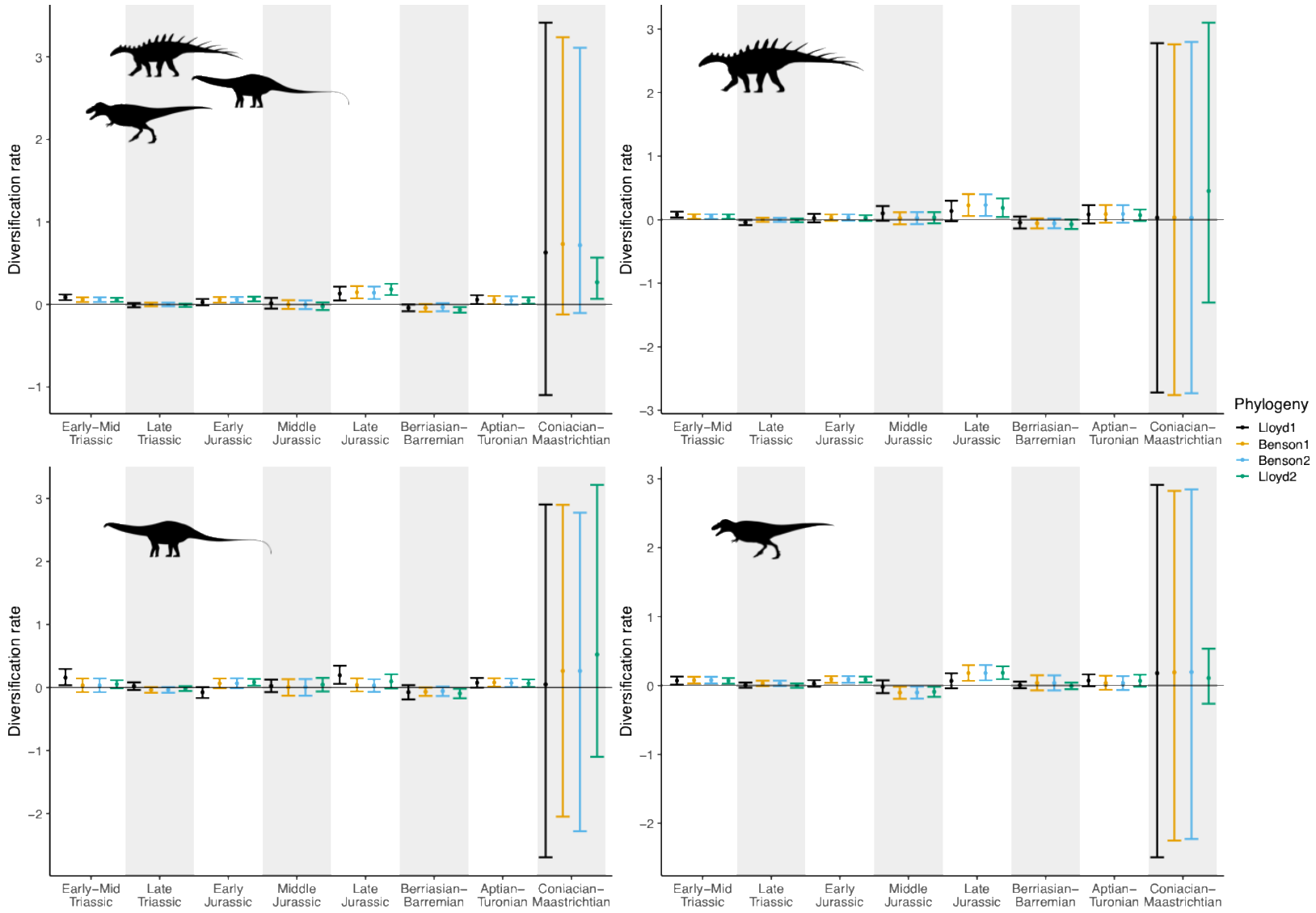
Adding occurrences to
population model
(occurrence-BD; Andréoletti et al.
2022)

Modelling age information as
ranges
(sRanges; Stadler et al. 2018)

Skyline inference

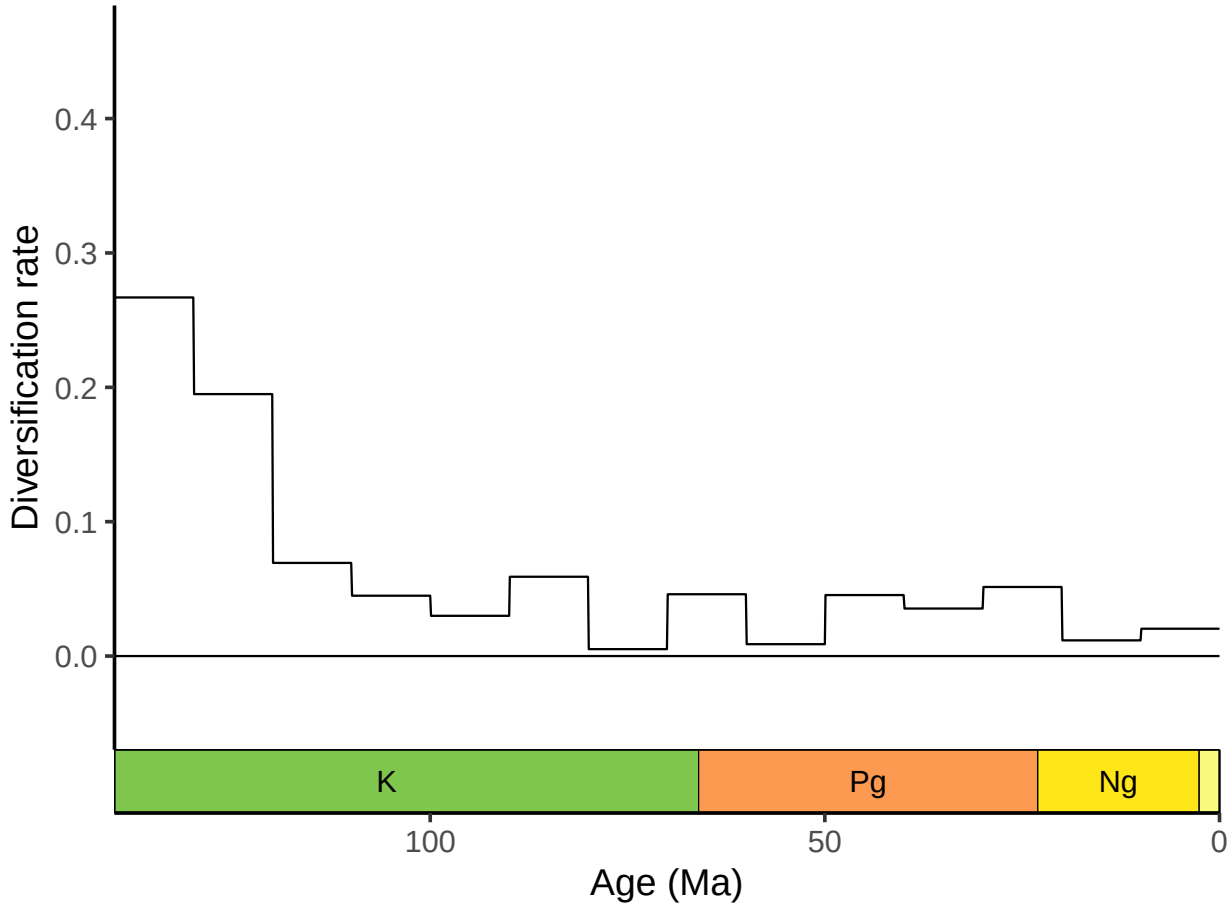


Dinosaurs before the K-Pg

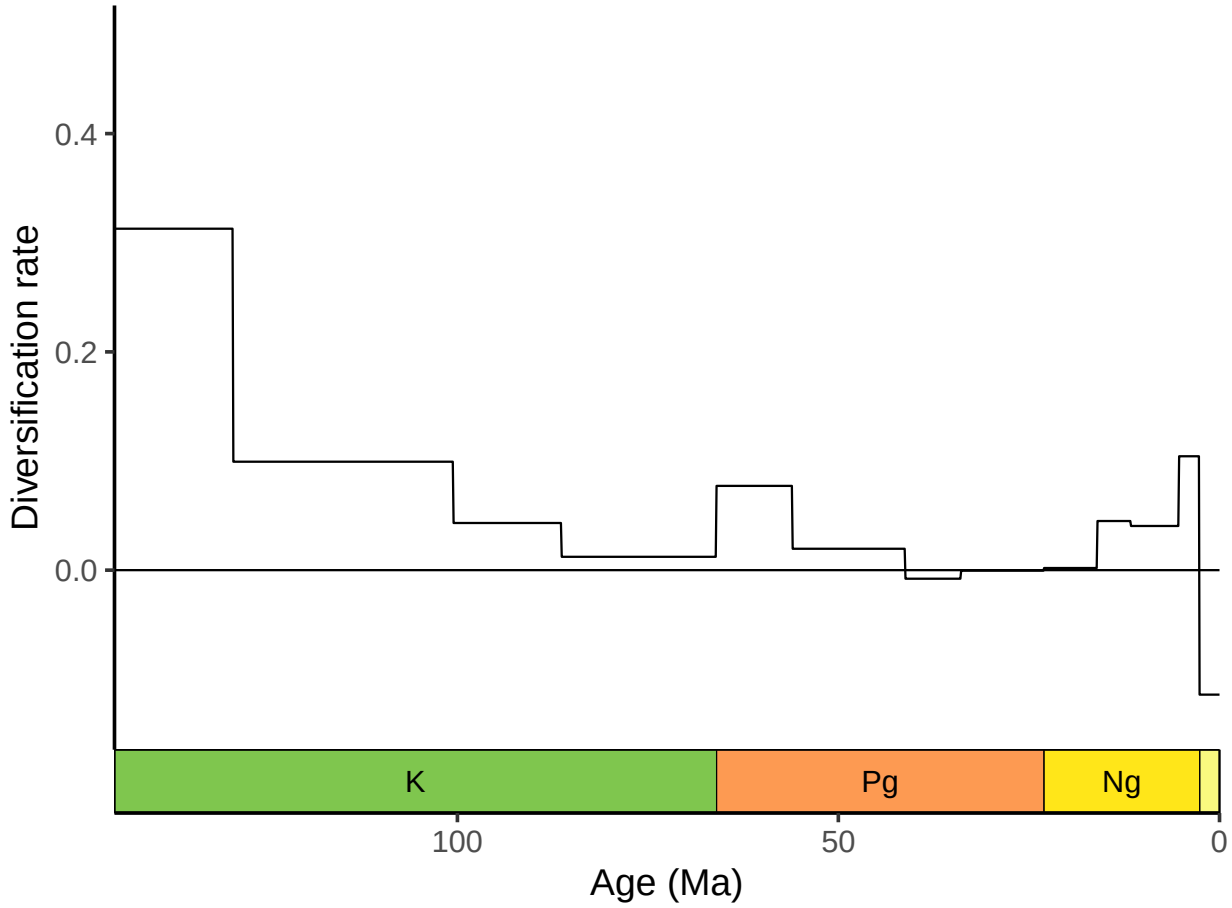


Allen et al. (2024)

Angiosperms in the Sino-Himalayas



Equal time bins



Geological time bins

Allen et al. (2024)

Final exercise