Phylogenetics

Bayesian timetrees: clock models and node dating

Rachel Warnock, Laura Mulvey
rachel.warnock@fau.de, laura.l.mulvey@fau.de

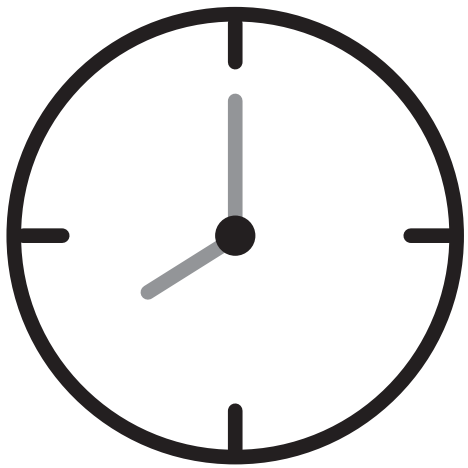
September 5, 2022

Analytical Paleobiology Workshop, Erlangen 2022

Part 5 objectives

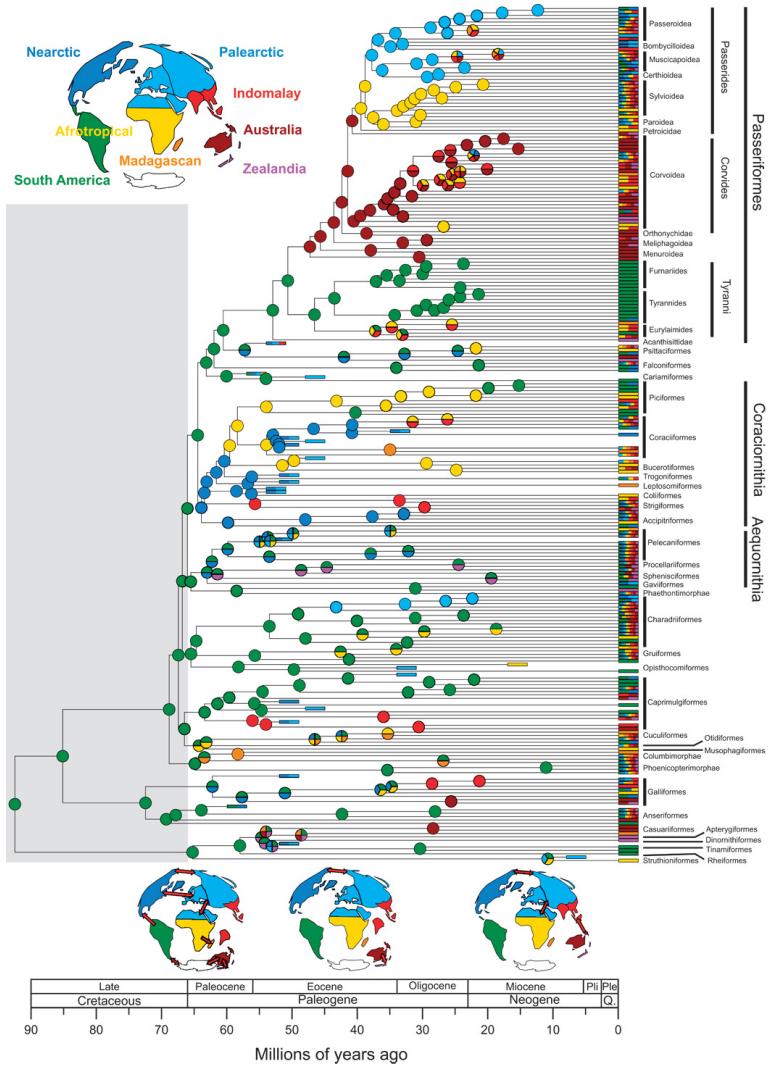
Unbelievably brief introduction to:

- the molecular clock hypothesis
- a framework for Bayesian molecular dating
- strict and relaxed clock models
- node dating



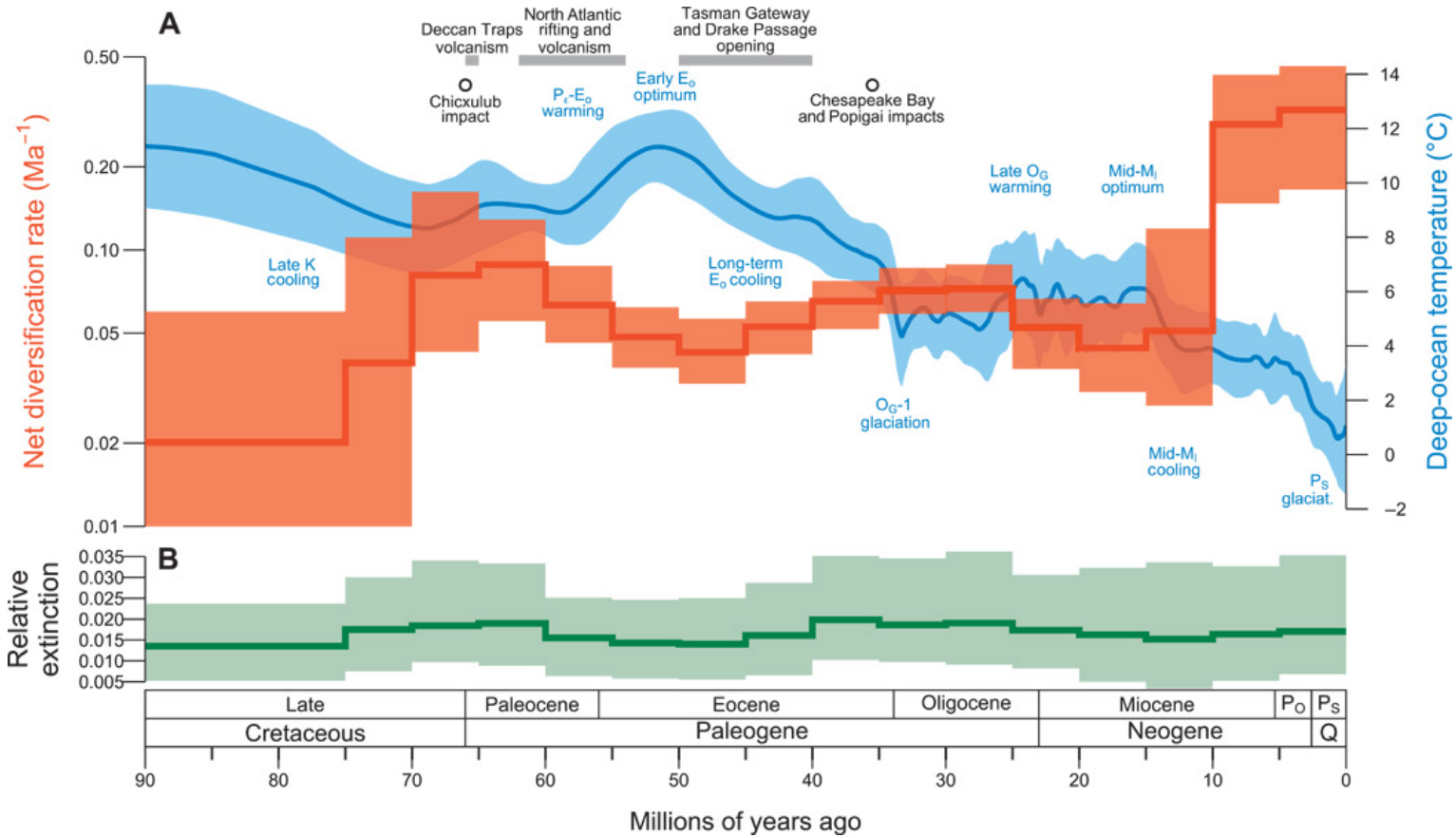
The molecular clock

Telling evolutionary time: motivation



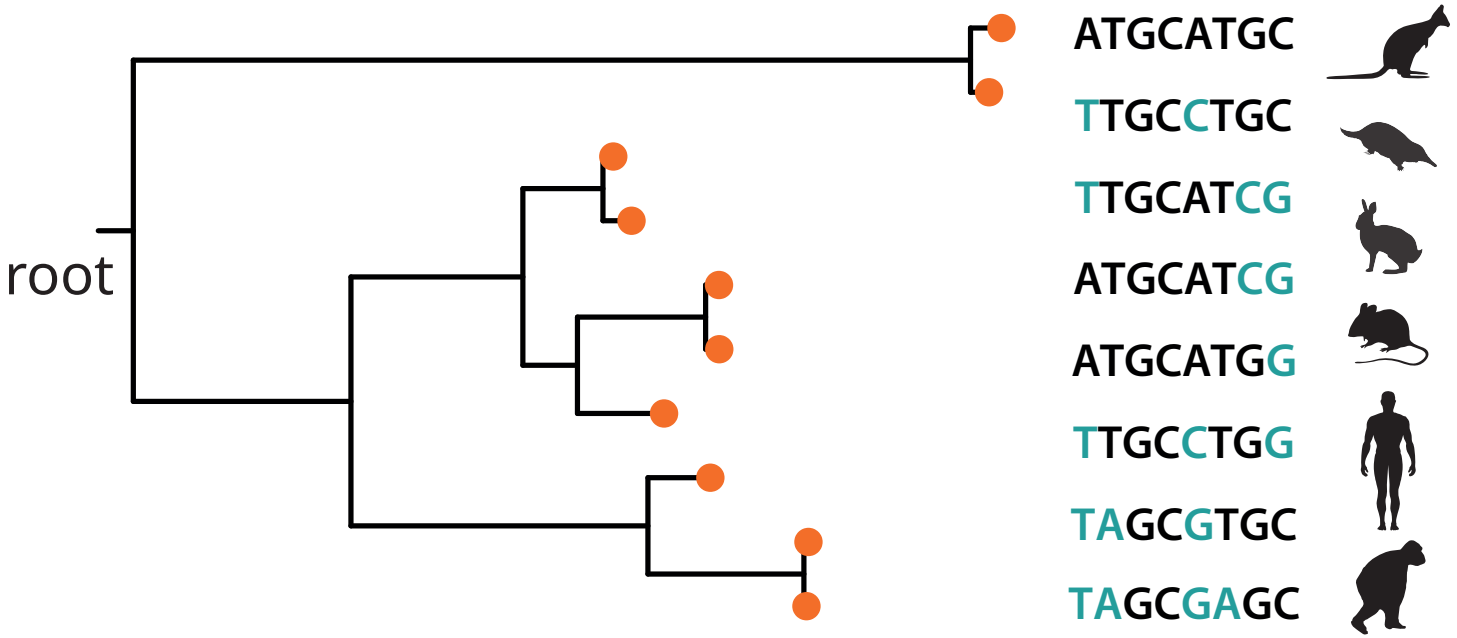
Claramunt et al. 2015 *Science Advances* — A new time tree reveals Earth history's imprint on the evolution of modern birds

Telling evolutionary time: motivation



Claramunt et al. 2015 *Science Advances* — A new time tree reveals Earth history's imprint on the evolution of modern birds

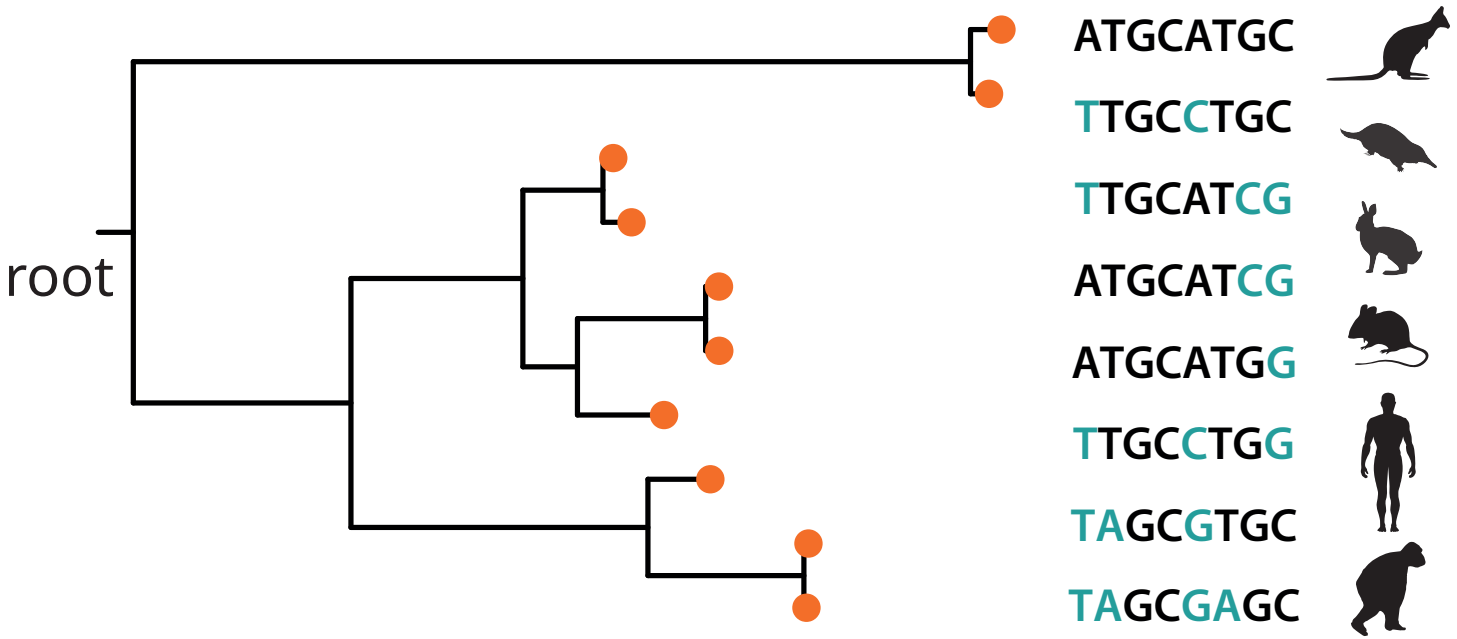
Molecular (or morphological) characters are not independently informative about time



branch lengths = genetic distance

$v = rt$

Molecular (or morphological) characters are not independently informative about time

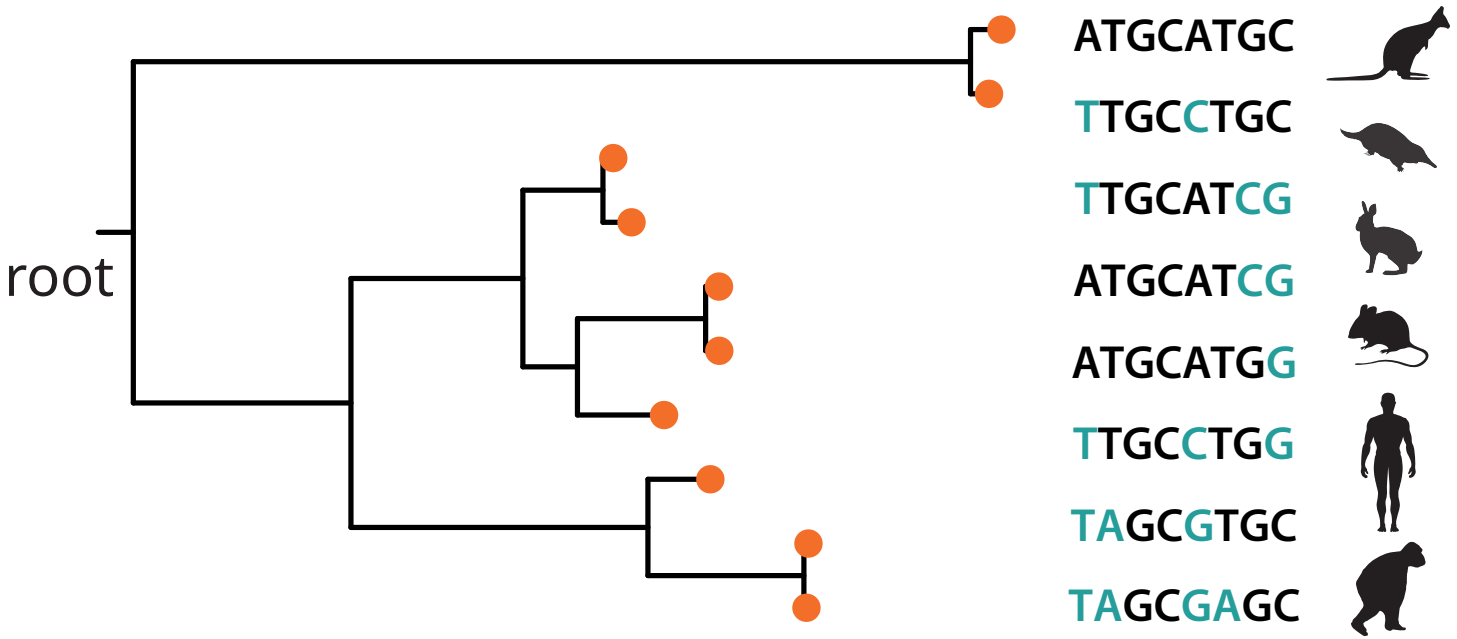


branch lengths = genetic distance

$v = rt$

Slow rate, long interval OR fast rate, short interval?

Molecular (or morphological) characters are not independently informative about time

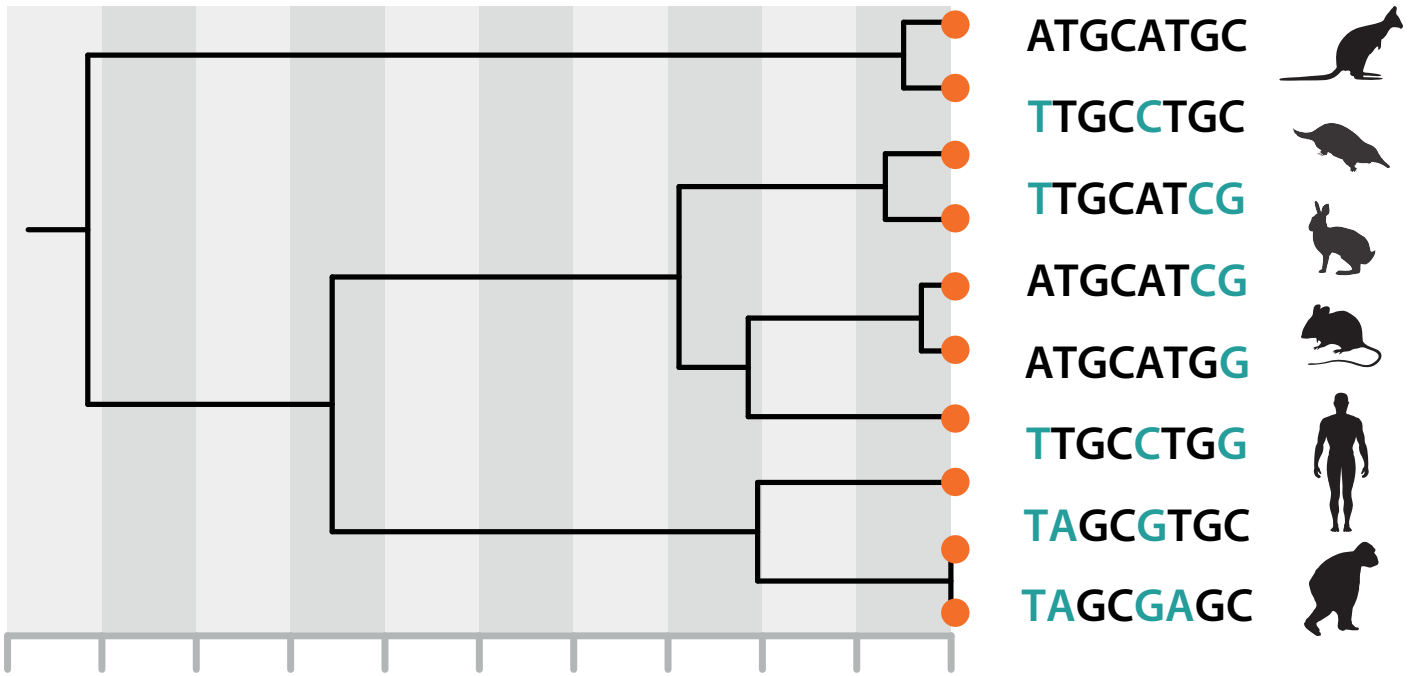


branch lengths = genetic distance

$v = rt$

Goal: to disentangle evolutionary rate and time.

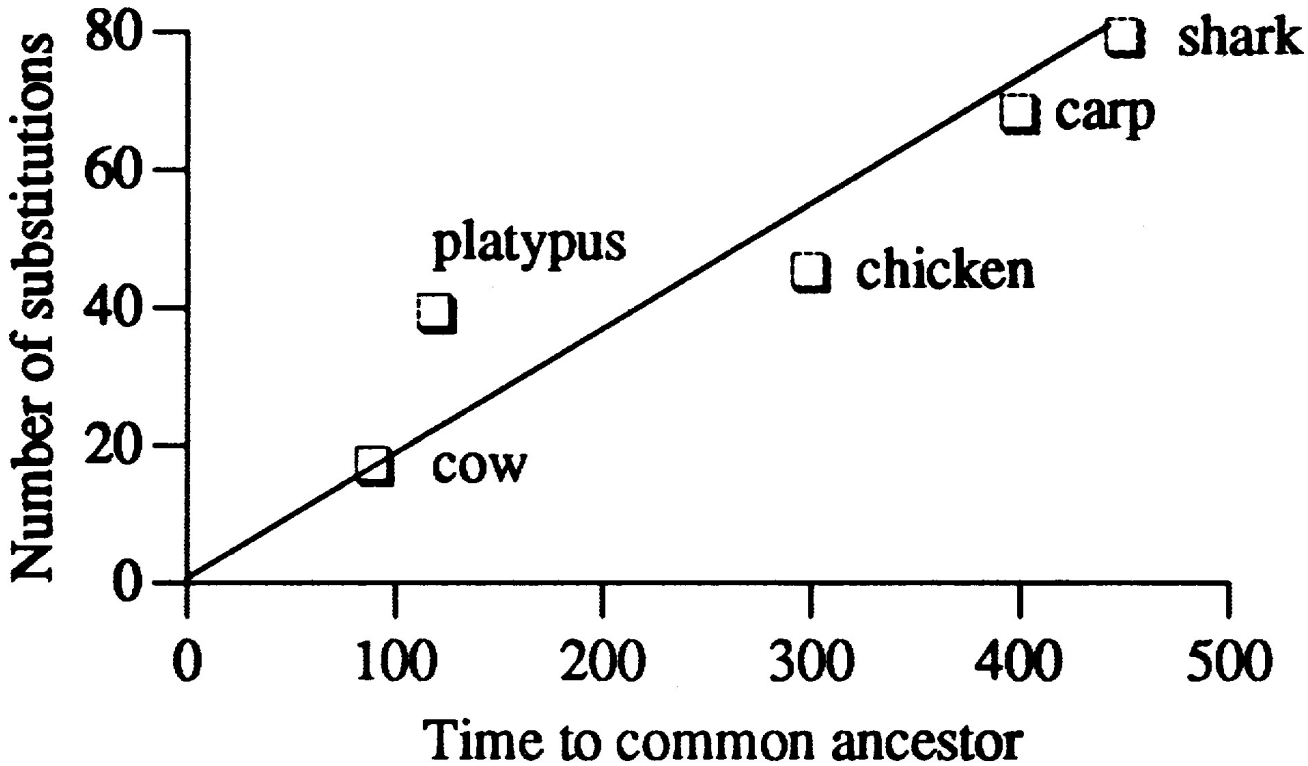
Molecular (or morphological) characters are not independently informative about time



branch lengths = time

Goal: to disentangle evolutionary rate and time.

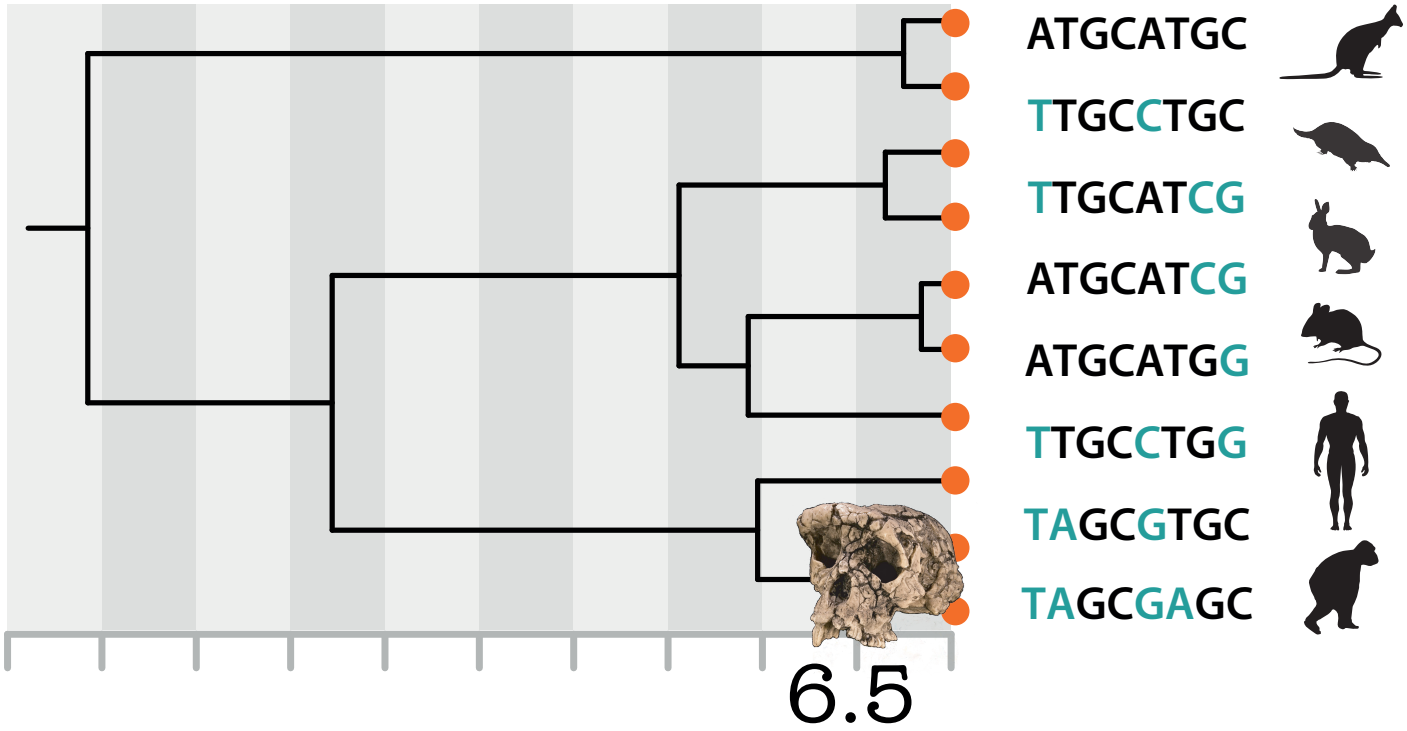
The molecular clock hypothesis



Zuckerkandl & Pauling (1965) — Molecules as documents of evolutionary history.

Morgan (1998) — A history of the molecular clock.

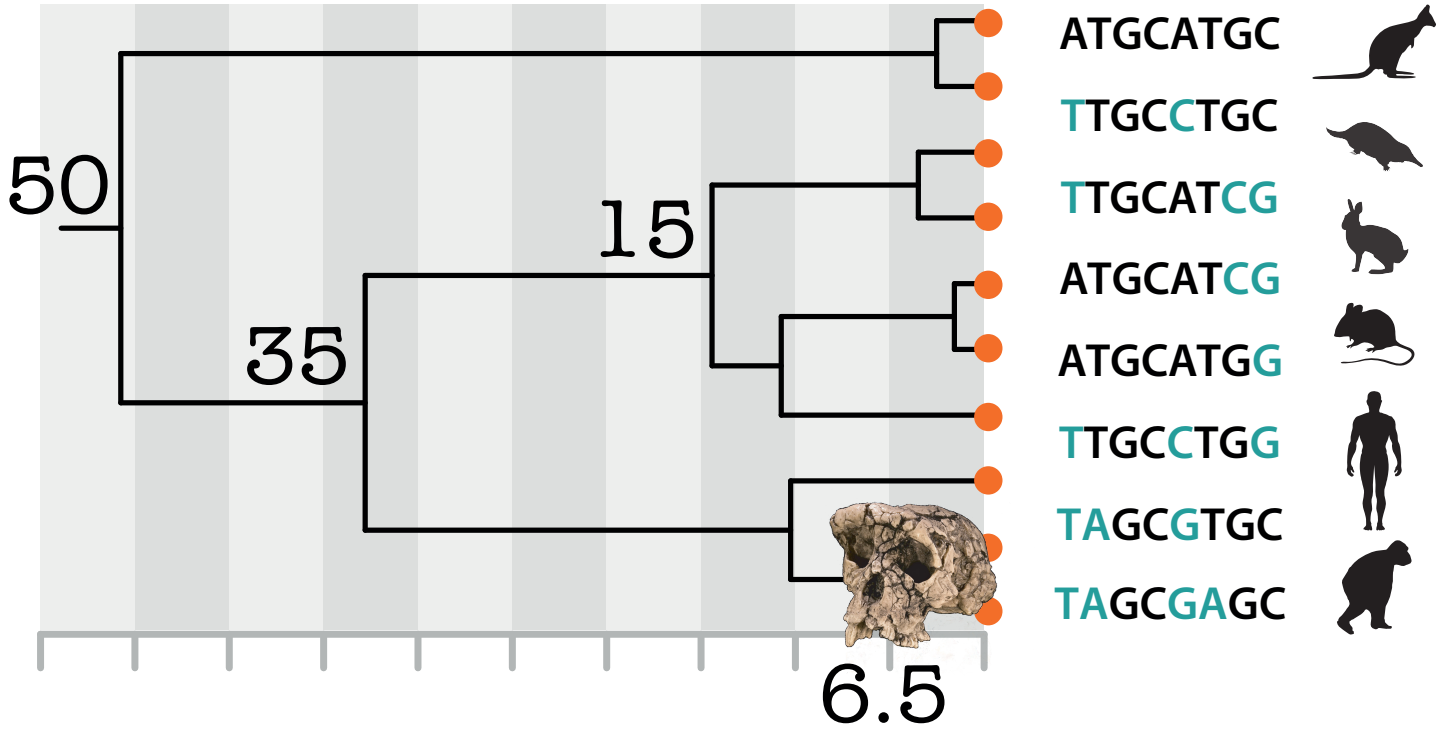
If we have independent evidence of time, we can calibrate the substitution rate



branch lengths = time

Temporal evidence of divergence for one species pair let's us calibrate the average rate of molecular evolution...

If we have independent evidence of time, we can calibrate the substitution rate

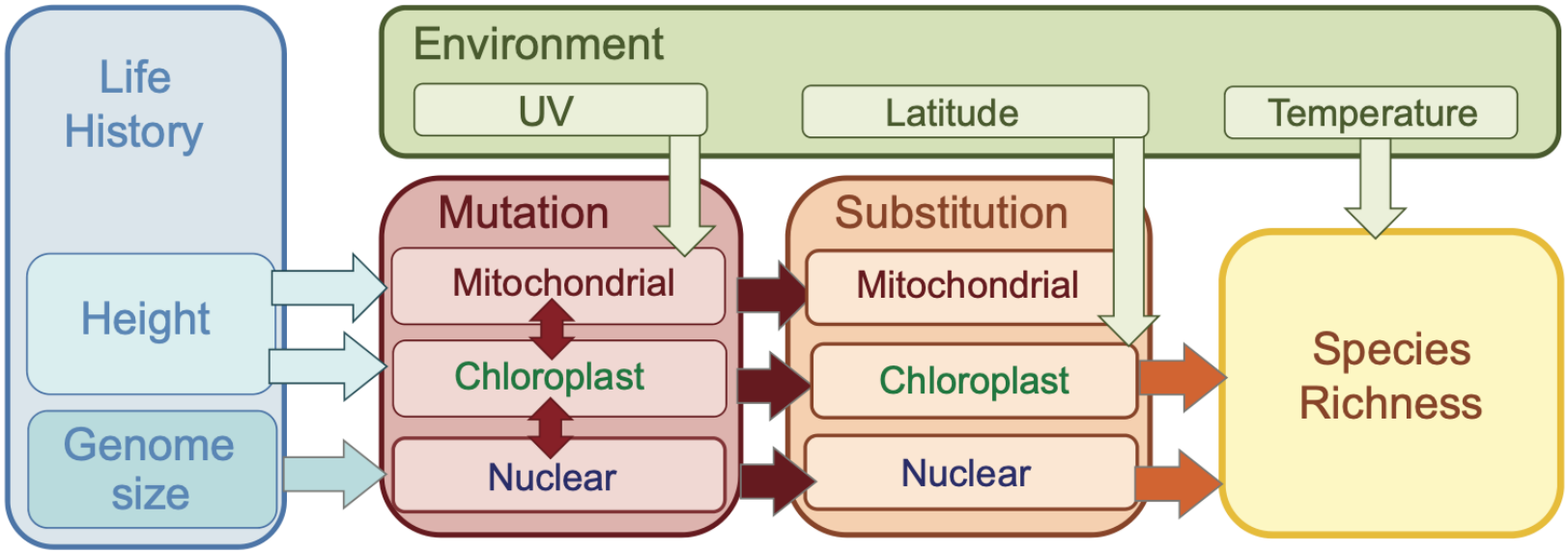


branch lengths = time

...and use this to extrapolate the divergence times for other species pairs.

The molecular clock: challenges

Many variables contribute to variation in the substitution rate.



Bromham et al. (2015).

The molecular clock: challenges

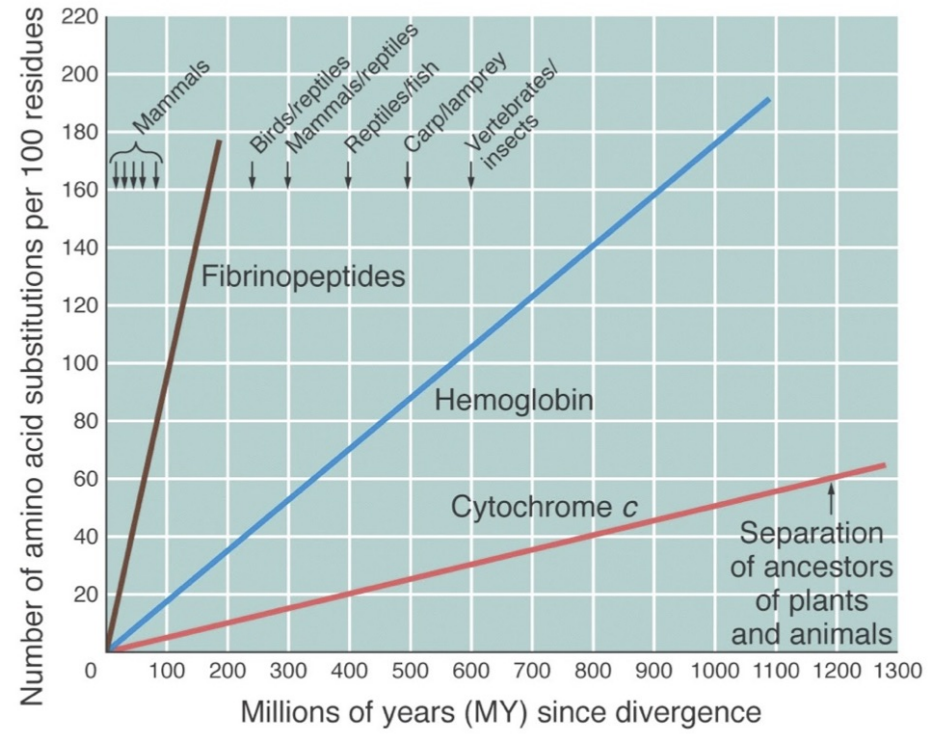
The molecular clock is not constant over time.

- Rates vary across taxa / time / genes / sites within the same gene

The molecular clock: challenges

The molecular clock is not constant over time.

- Rates vary across taxa / time / genes / sites within the same gene



Variation in rate makes different genes useful for different timescales.

The molecular clock: challenges

The molecular clock is not constant over time.

- Rates vary across taxa / time / genes / sites within the same gene

Calibrations are rarely known precisely.

The molecular clock: challenges

The molecular clock is not constant over time.

- Rates vary across taxa / time / genes / sites within the same gene

Calibrations are rarely known precisely.

$\rightarrow$ we need a flexible statistical framework that deals well with uncertainty.

Bayesian inference

Bayes' theorem

$$P(\text{parameters} \mid \text{data}, \text{model}) = \frac{P(\text{data} \mid \text{parameters}, \text{model})\, P(\text{parameters} \mid \text{model})}{P(\text{data} \mid \text{model})}$$

posterior

likelihood

priors

marginal probability of the data

Bayes' theorem

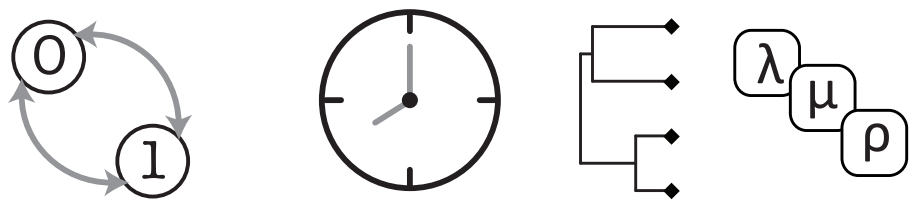
P (data | parameters, model) ← the model used to calculate the **likelihood**.

P (parameters | model) ← this represents our **prior knowledge** of the model parameters.

P (parameters | data, model) ← the **posterior** reflects our combined knowledge based on the likelihood and the priors.

Bayesian phylogenetic dating requires three model components

- The **substitution model** ← describes how sites evolve over time.
- The **clock model** ← describes how evolutionary rates vary across the tree.
- The **tree model** ← describes how trees grow over time. Temporal evidence is included here.



Bayesian phylogenetic dating

The data

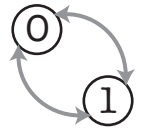
AND/OR

0101... ATTG...
1101... TTGC...
0100... ATTC...

Characters

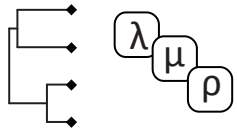
Fossil ages

The model



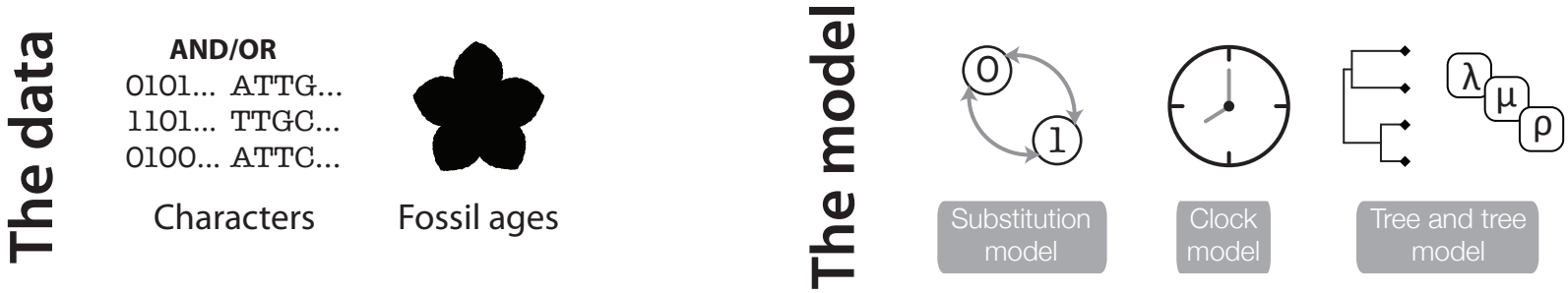
Substitution model

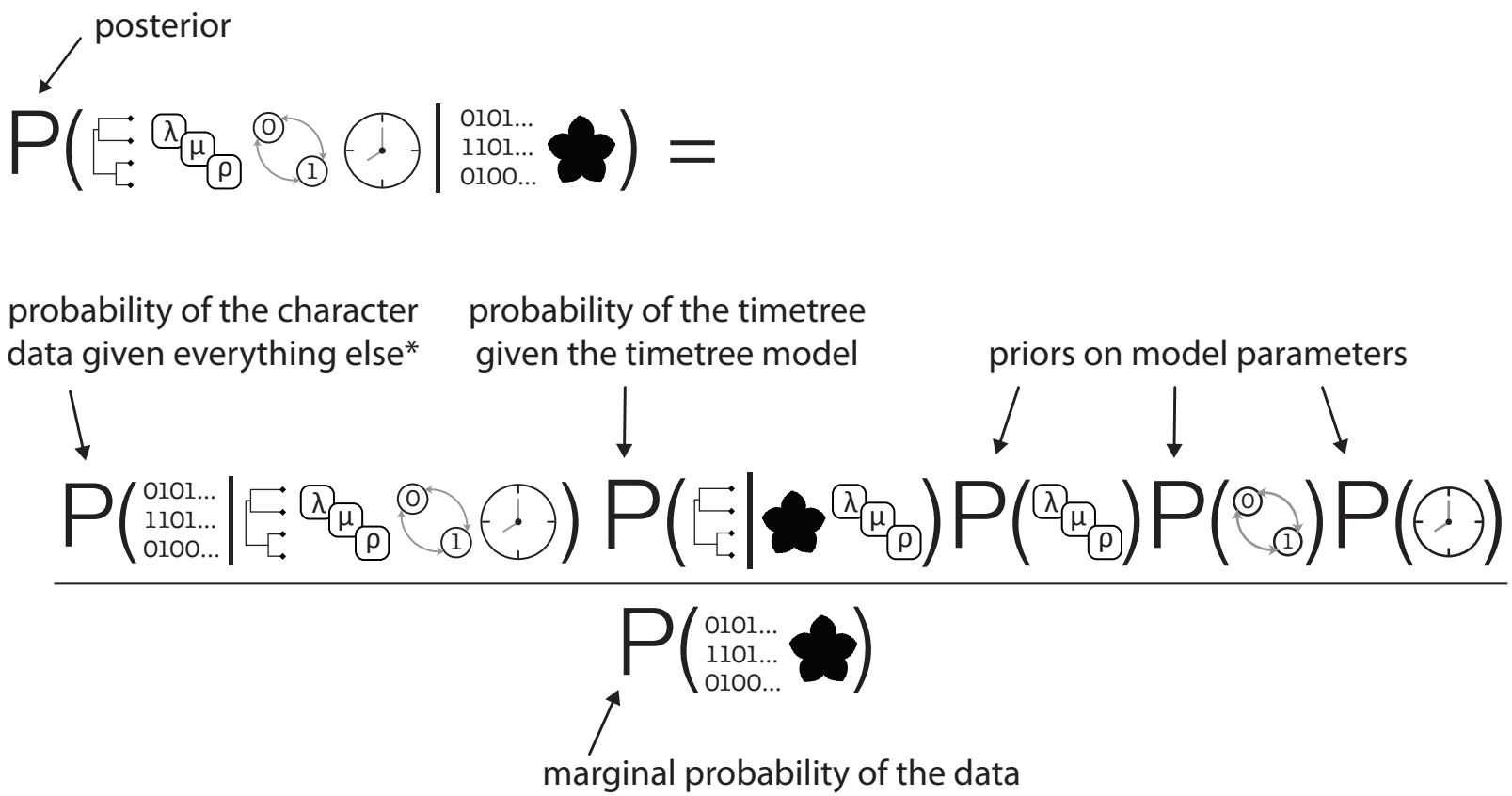
Clock model



Tree and tree model

Bayesian phylogenetic dating

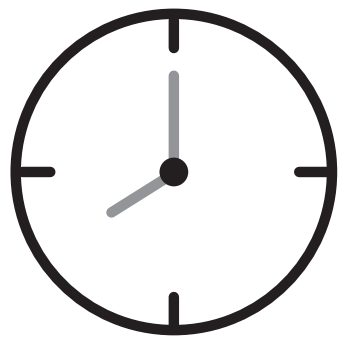




*the timetree, the parameters and the tripartite model

Molecular clock models

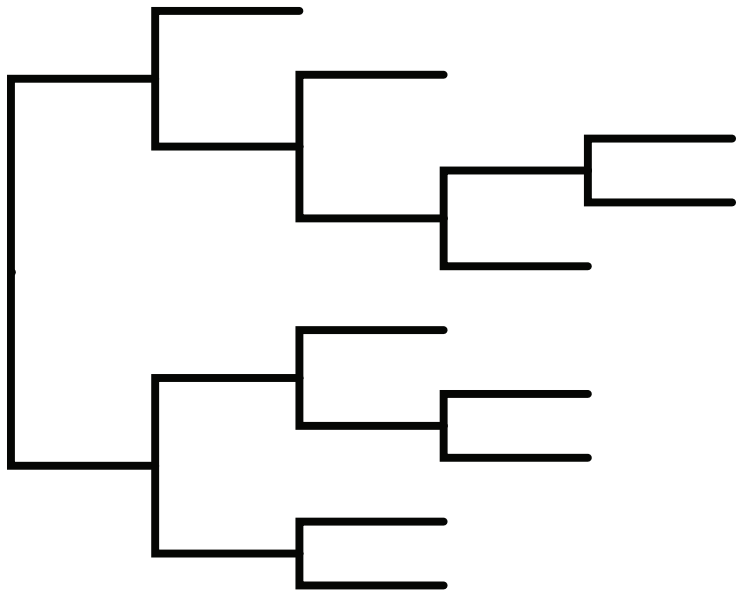
The **clock model** ← describes how evolutionary rates vary across the tree.



The strict / constant molecular clock model

Assumptions:

- The substitution rate is constant over time.
- All lineages share the same rate.



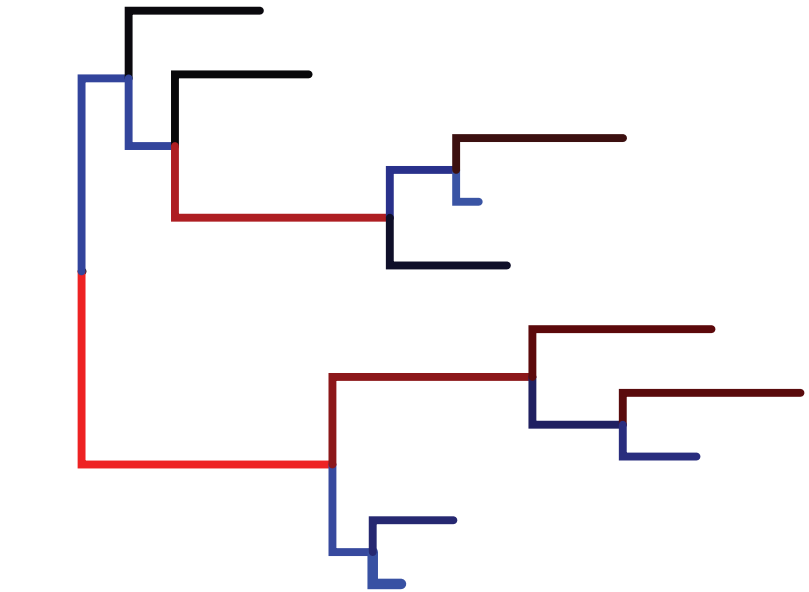
branch length = substitution rate

low high

Relaxed clock models

Assumptions:

- Lineage-specific rates are independent (i.e., uncorrelated).
- The rate assigned to each branch is drawn independently from the underlying distribution.



branch length = substitution rate

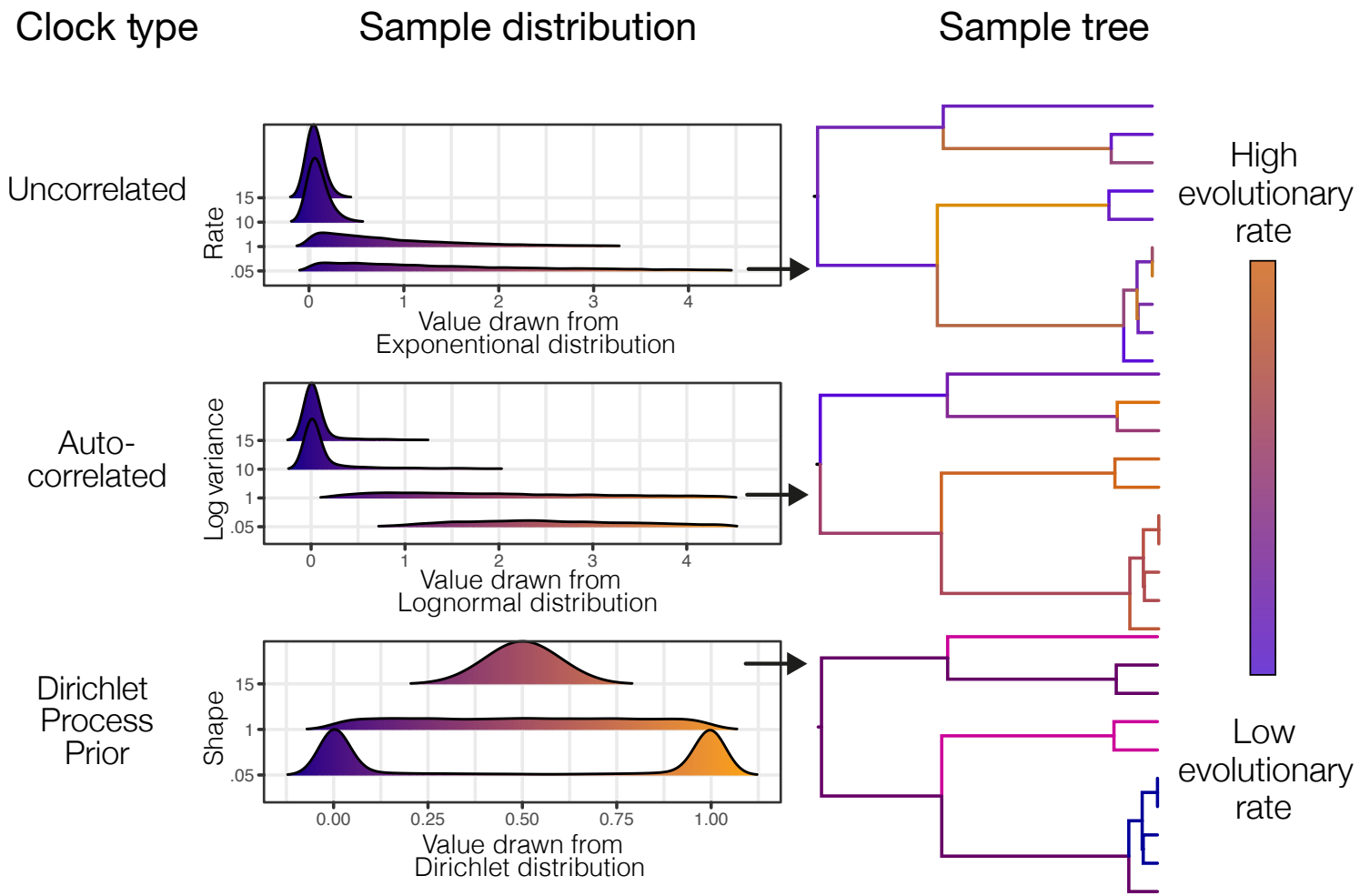
low high

Many different clock models

- Strict clock
- Uncorrelated clock (= the favourite)
- Autocorrelated clock
- Local clocks
- Mixture models

See Warnock & Wright (2020) for an overview.

Many different clock models

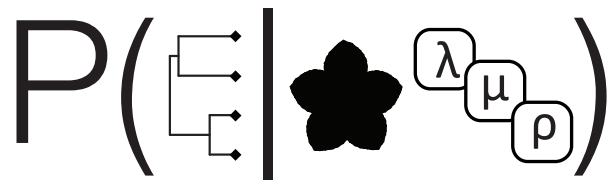


Warnock & Wright (2020)

Tree models and node dating

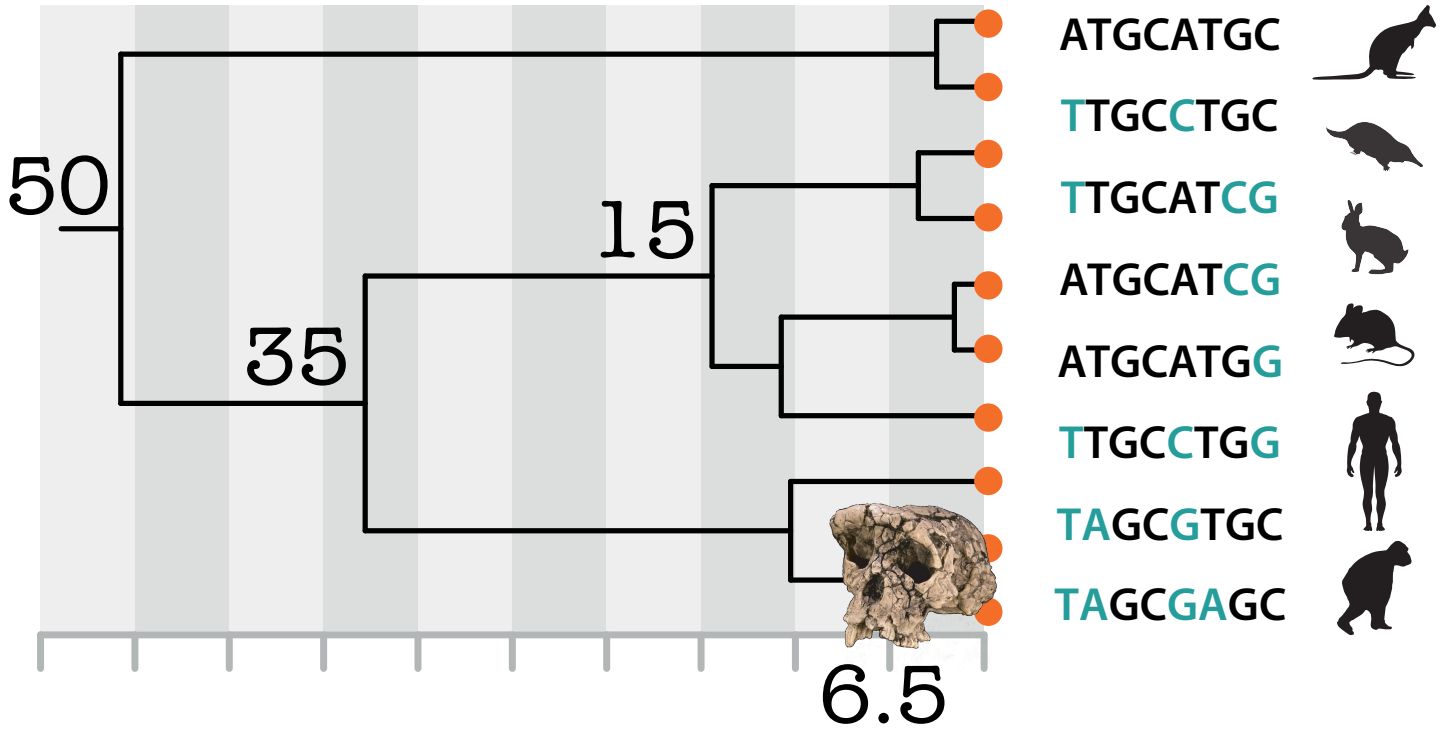
The tree model

- Describes the dynamics (speciation / transmission / replication) of the tree generating process over time.
- Gives rise to the **tree prior**:



- How likely is the phylogeny given the tree model?
- Calibration information is either combined with or incorporated into the tree model.

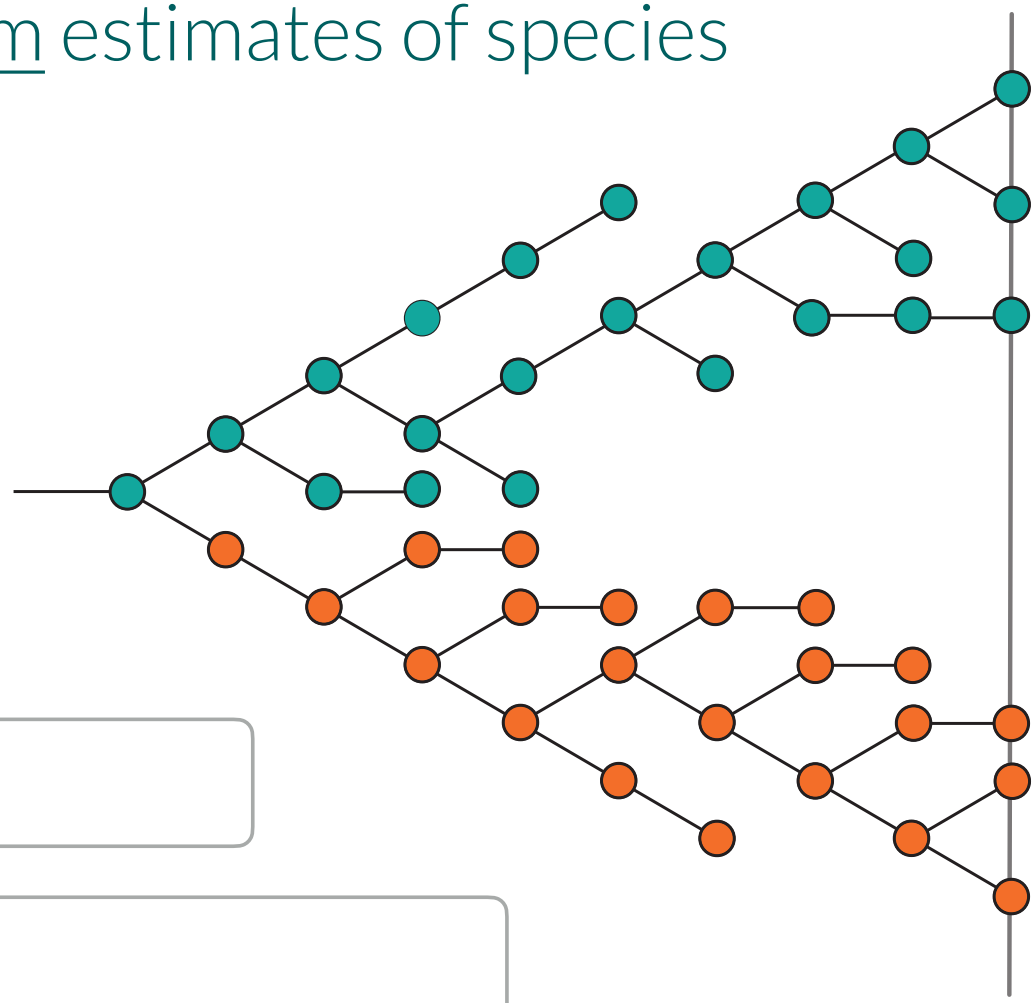
If we have independent evidence of time, we can calibrate the substitution rate



branch lengths = time

What evidence are we really able to recover from the fossil record?

Fossils provide minimum estimates of species divergence time



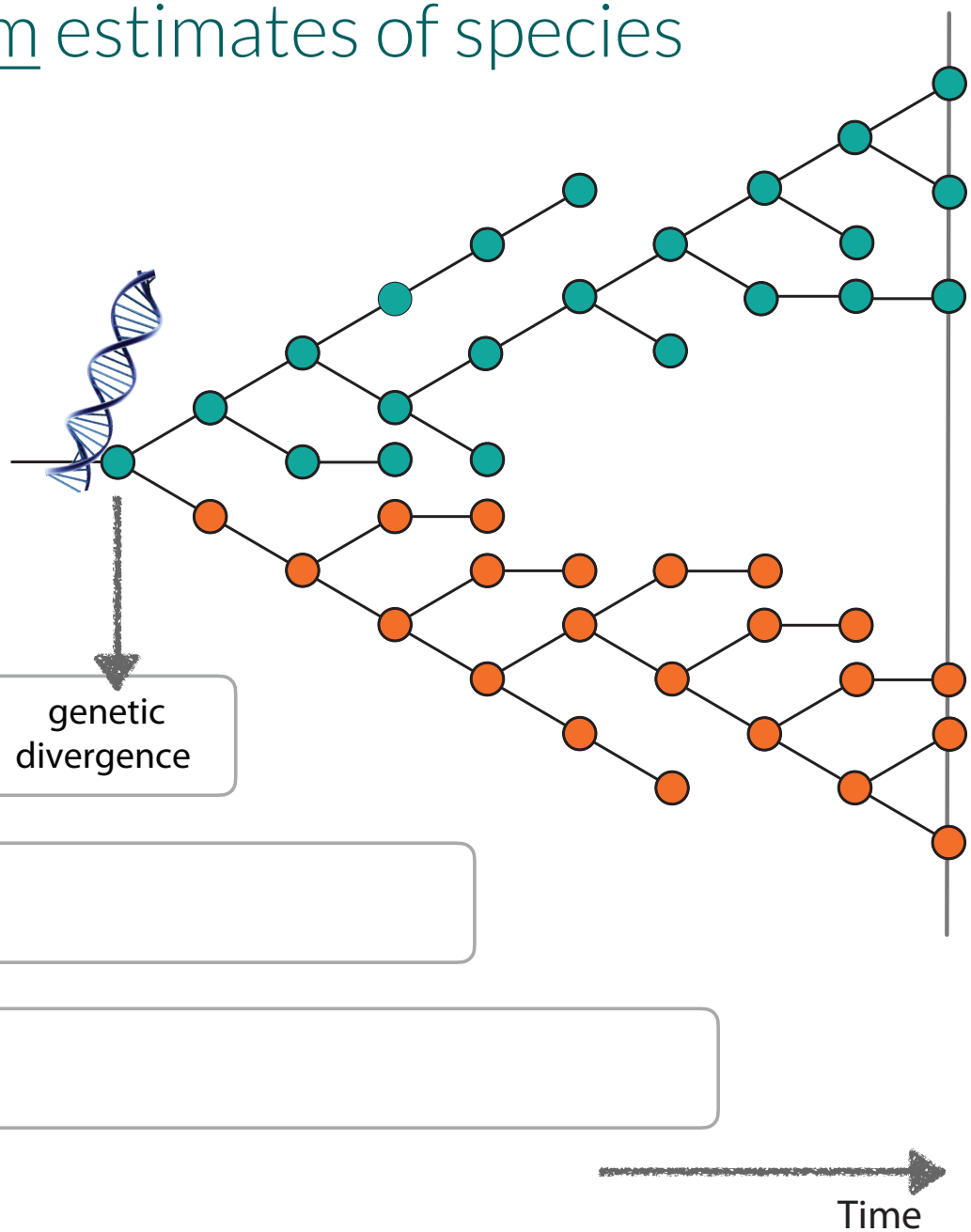
Molecular evolution:

Morphological evolution:

Fossil preservation:

Time

Fossils provide minimum estimates of species divergence time

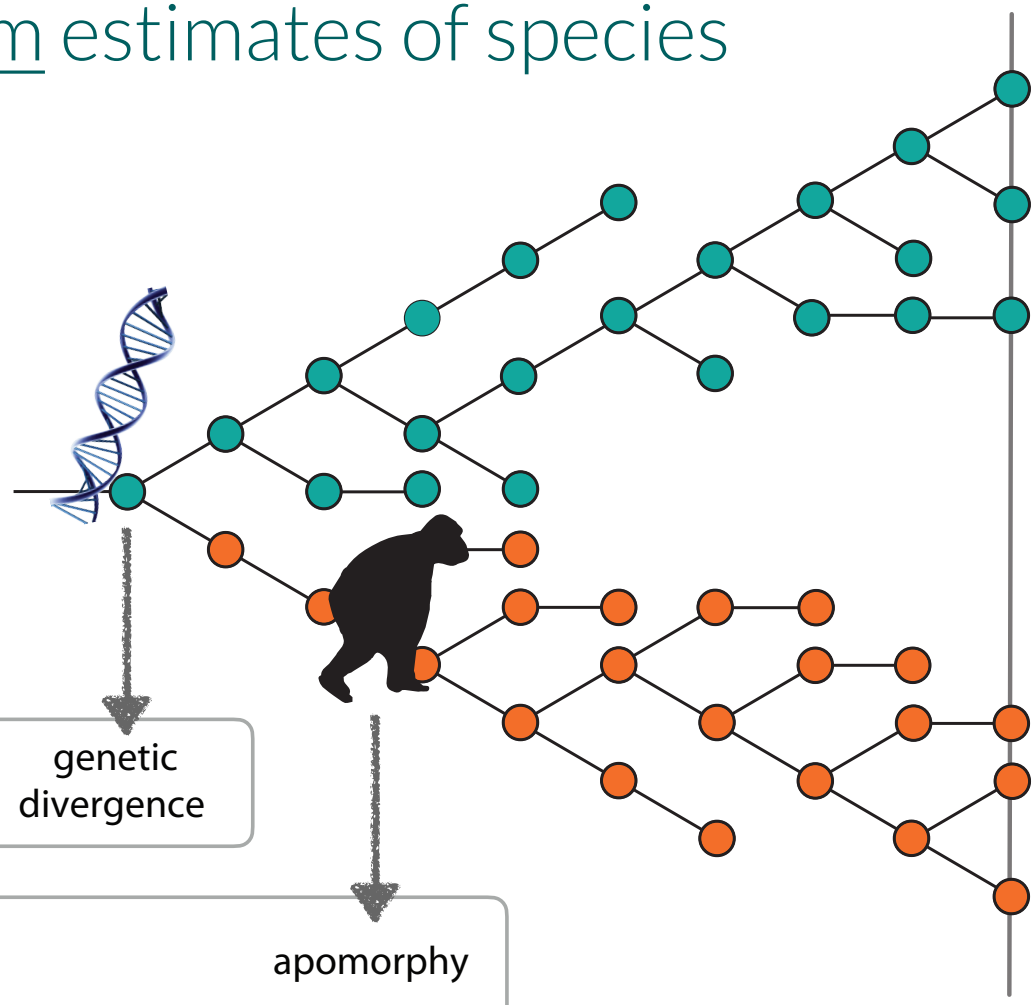


Molecular evolution:

Morphological evolution:

Fossil preservation:

Fossils provide minimum estimates of species divergence time



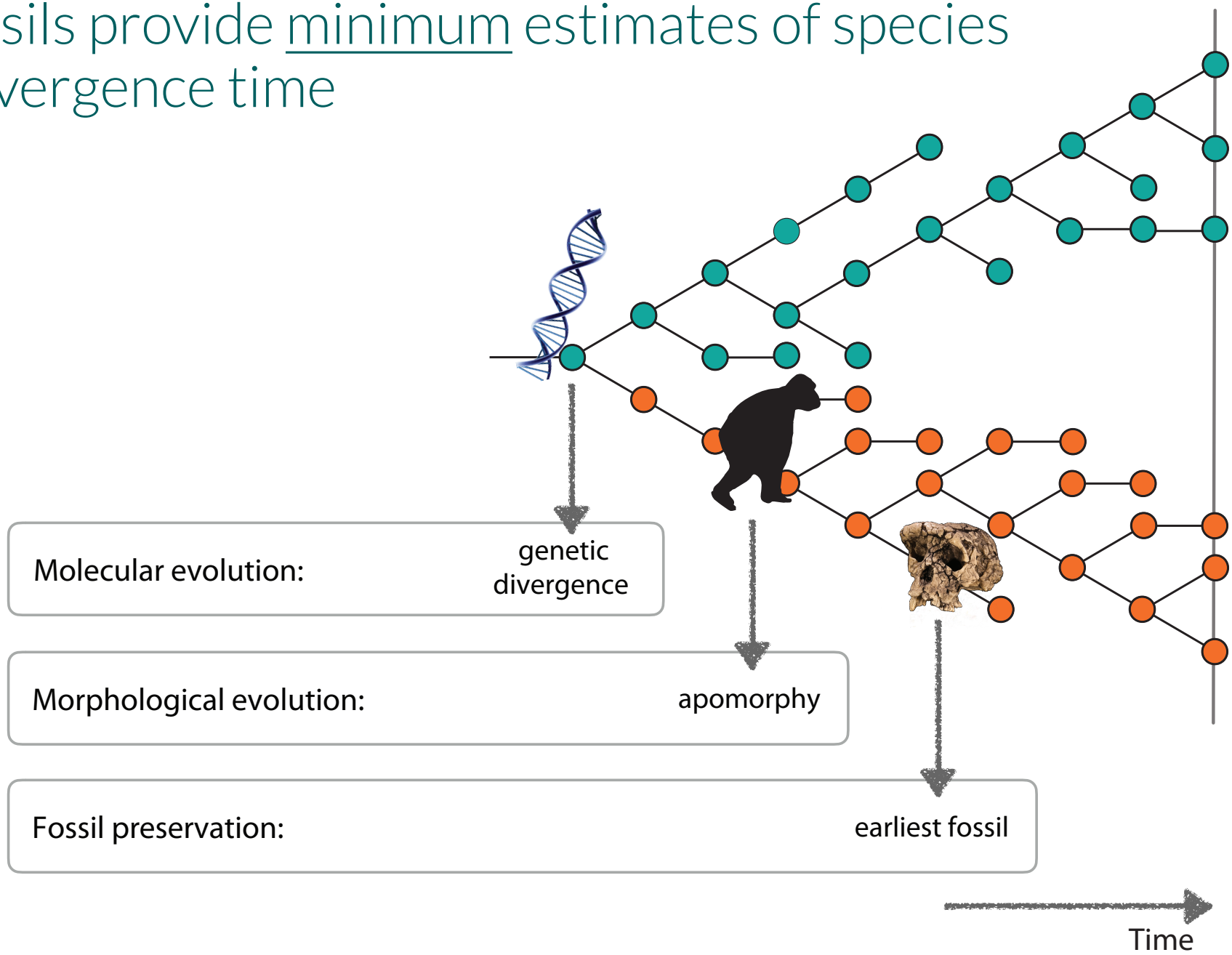
Molecular evolution: genetic divergence

Morphological evolution: apomorphy

Fossil preservation:

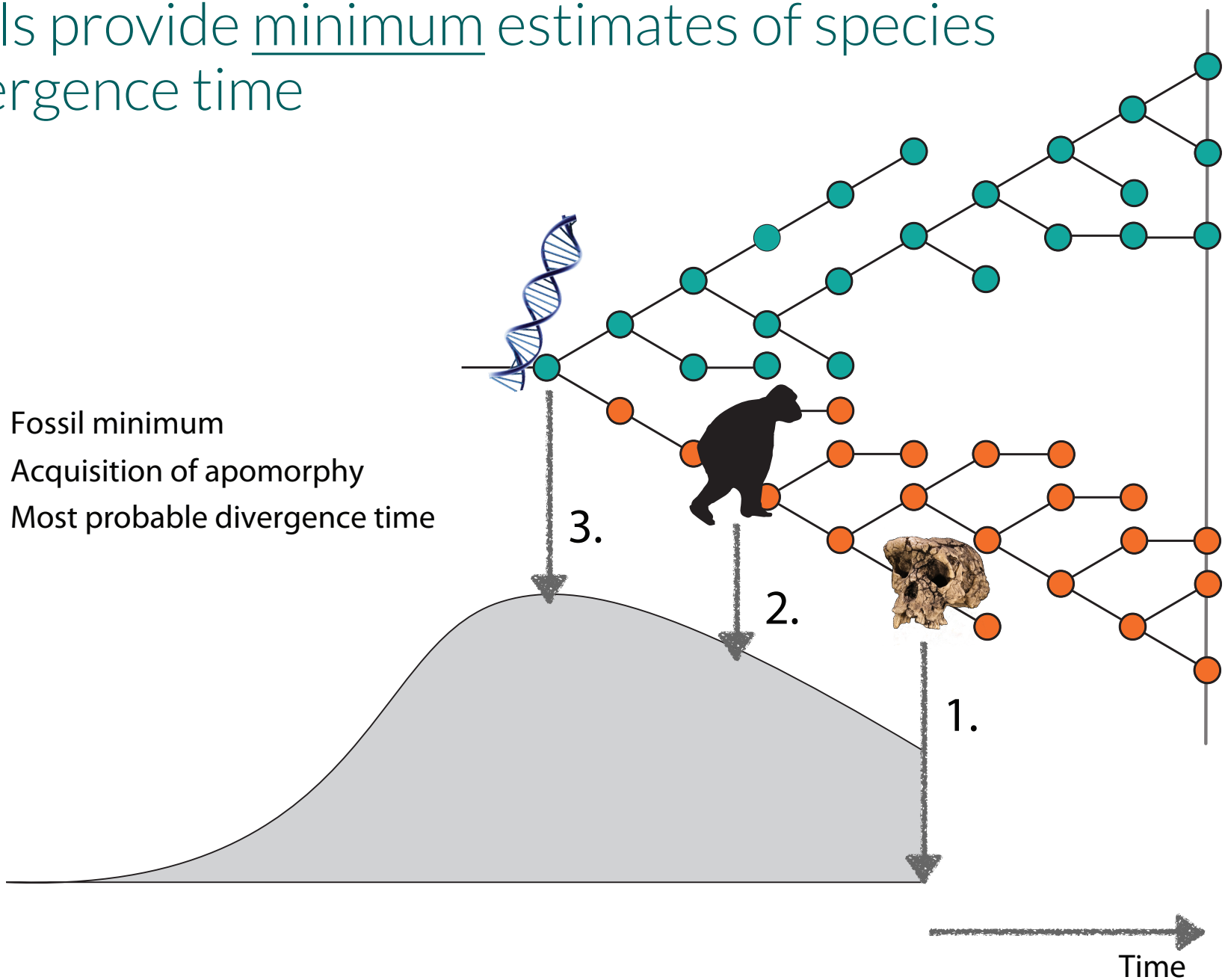
Time

Fossils provide minimum estimates of species divergence time

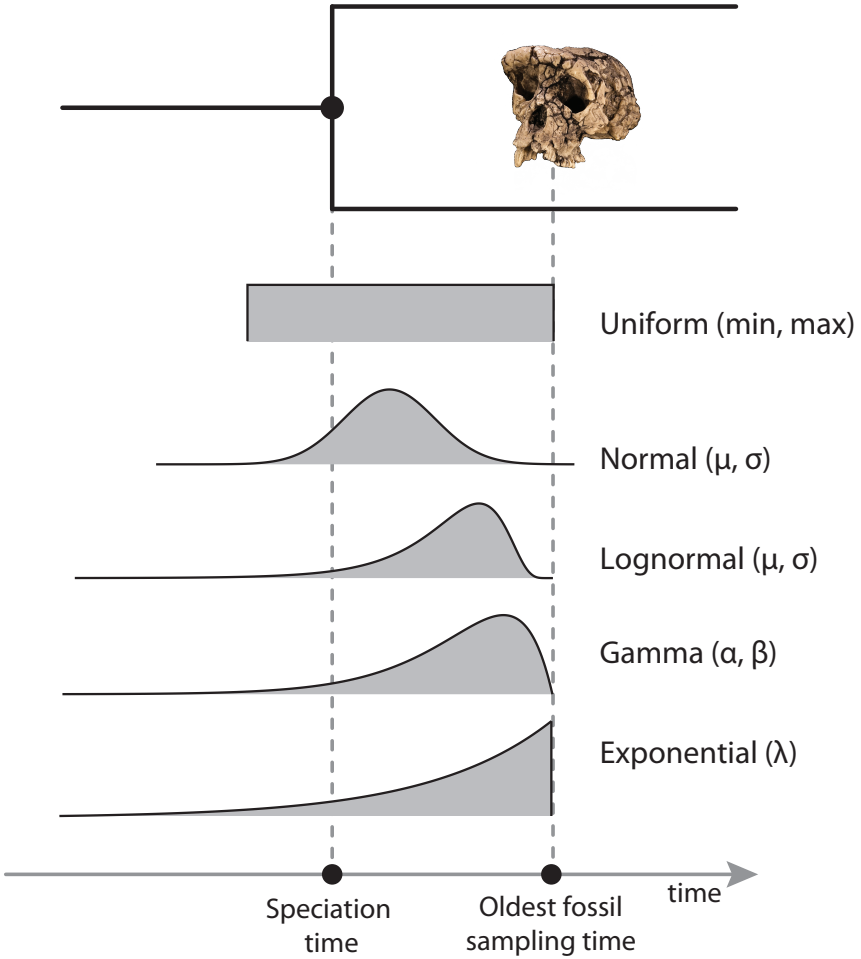


Fossils provide minimum estimates of species divergence time

1. Fossil minimum
2. Acquisition of apomorphy
3. Most probable divergence time



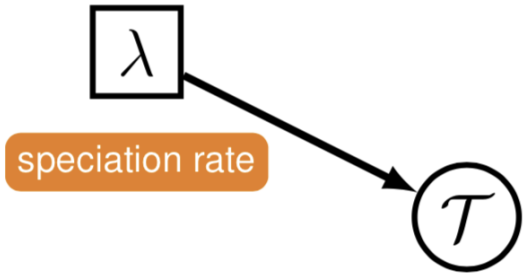
Node dating



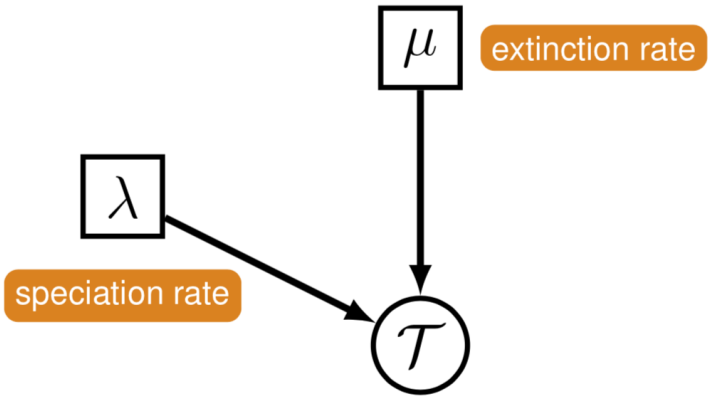
- We used a birth-death model to describe the tree generating process, given we only observe extant species.
- Then we separately apply a calibration density to constrain internal node ages.

Image adapted from Heath (2012) *Systematic Biology*

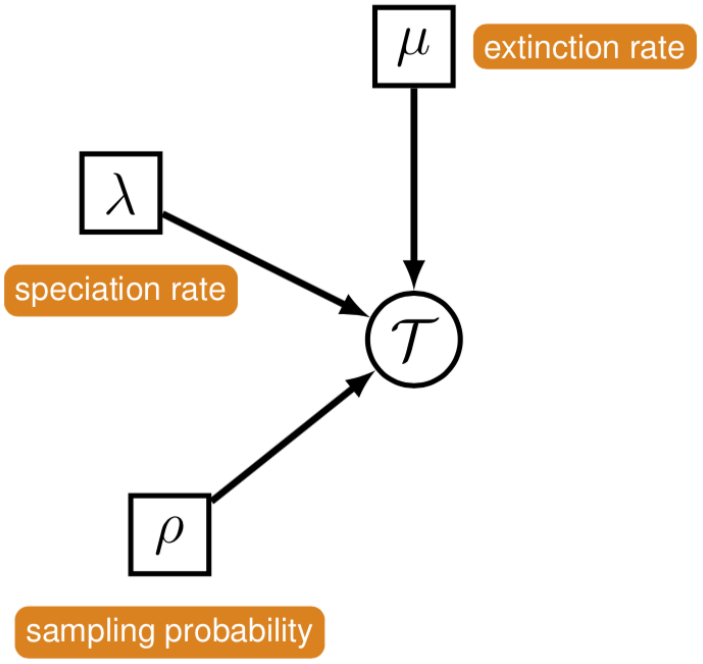
The tree prior (for the non-fossil calibrated nodes)



Pure birth process
i.e. no extinction



Birth-death process



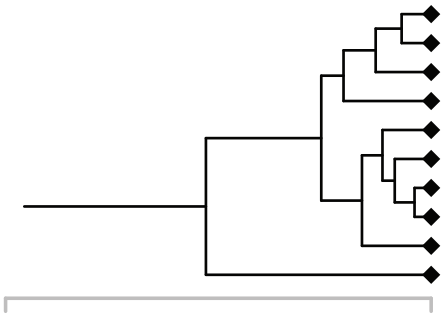
Birth-death sampling process

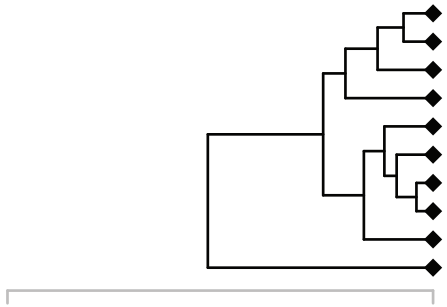
Complete versus reconstructed tree

The complete outcome of the diversification and sampling processes

The reconstructed tree

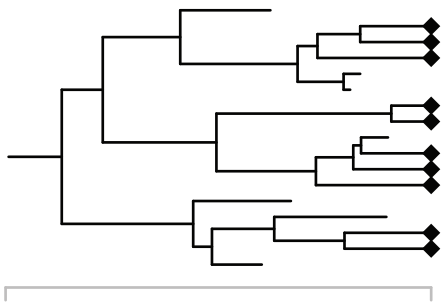
Model parameters

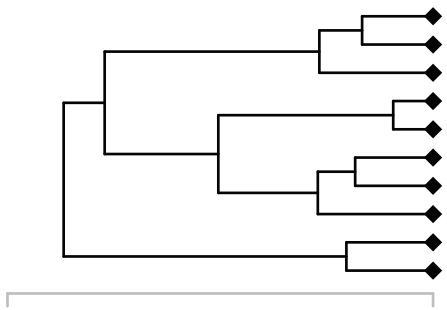




speciation (λ) = 0.1

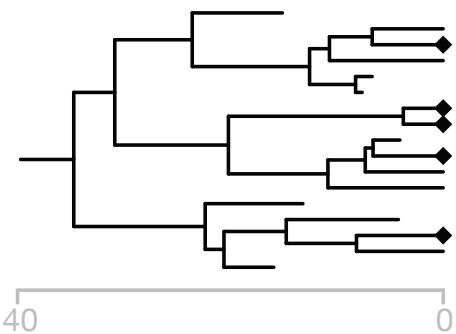
Pure birth process

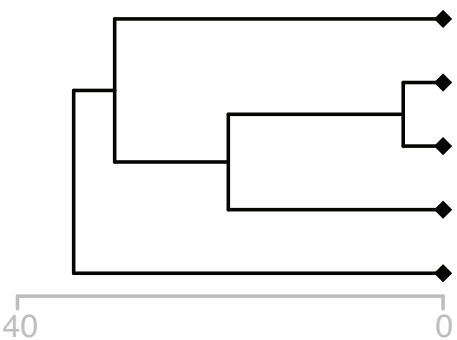




speciation (λ) = 0.1
extinction (μ) = 0.05

Birth-death process





speciation (λ) = 0.1
extinction (μ) = 0.05
extant sampling (ρ) = 0.6

Birth-death sampling process

Take homes

Sequences do not directly contain information about time. To date phylogenetic trees we need to separate rate and time.

The molecular clock hypotheses allows us calibrate the substitution rate and to date phylogenetic trees.

Bayesian inference is a flexible statistical framework that allows us to integrate prior knowledge with models that describe evolutionary processes.

Suggested reading

Understanding the tripartite approach to Bayesian divergence time estimation — Warnock, Wright (2020)

This goal of this review paper is to provide an introduction to the substitution, clock and tree models.

Break → Exercise