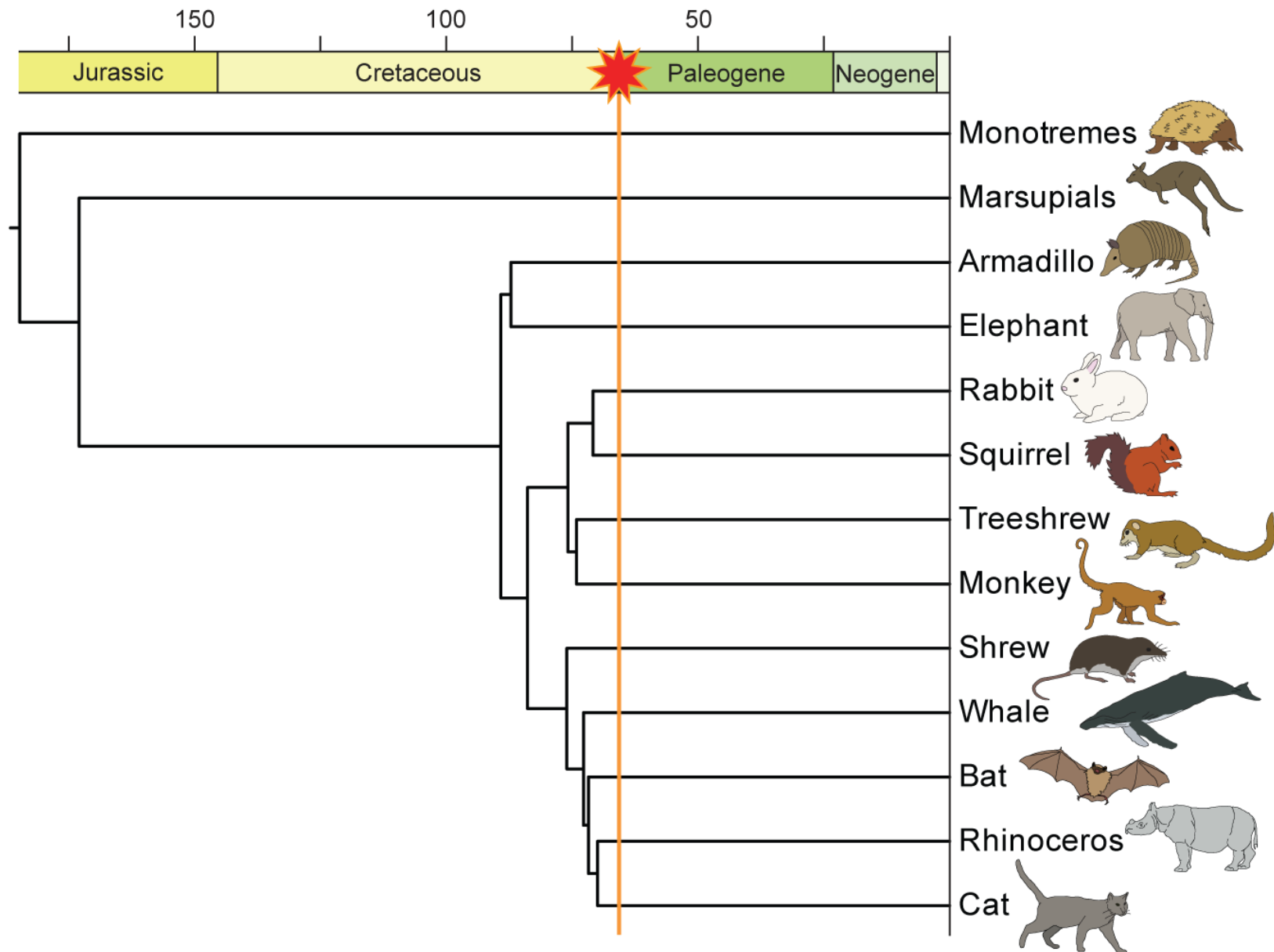

Lecture 2.3

Molecular Dating

Evolutionary timescales

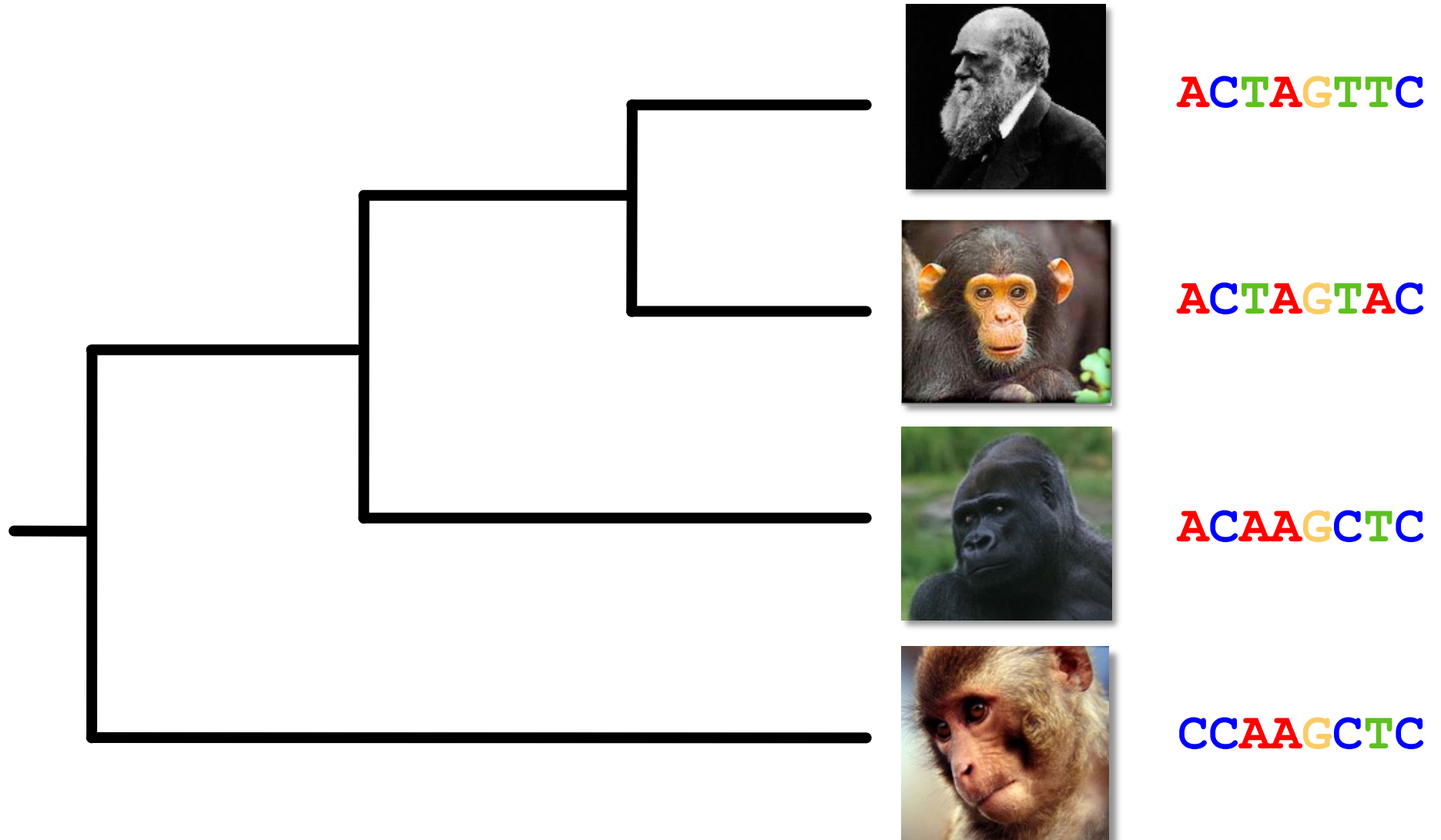


Dated phylogenetic trees

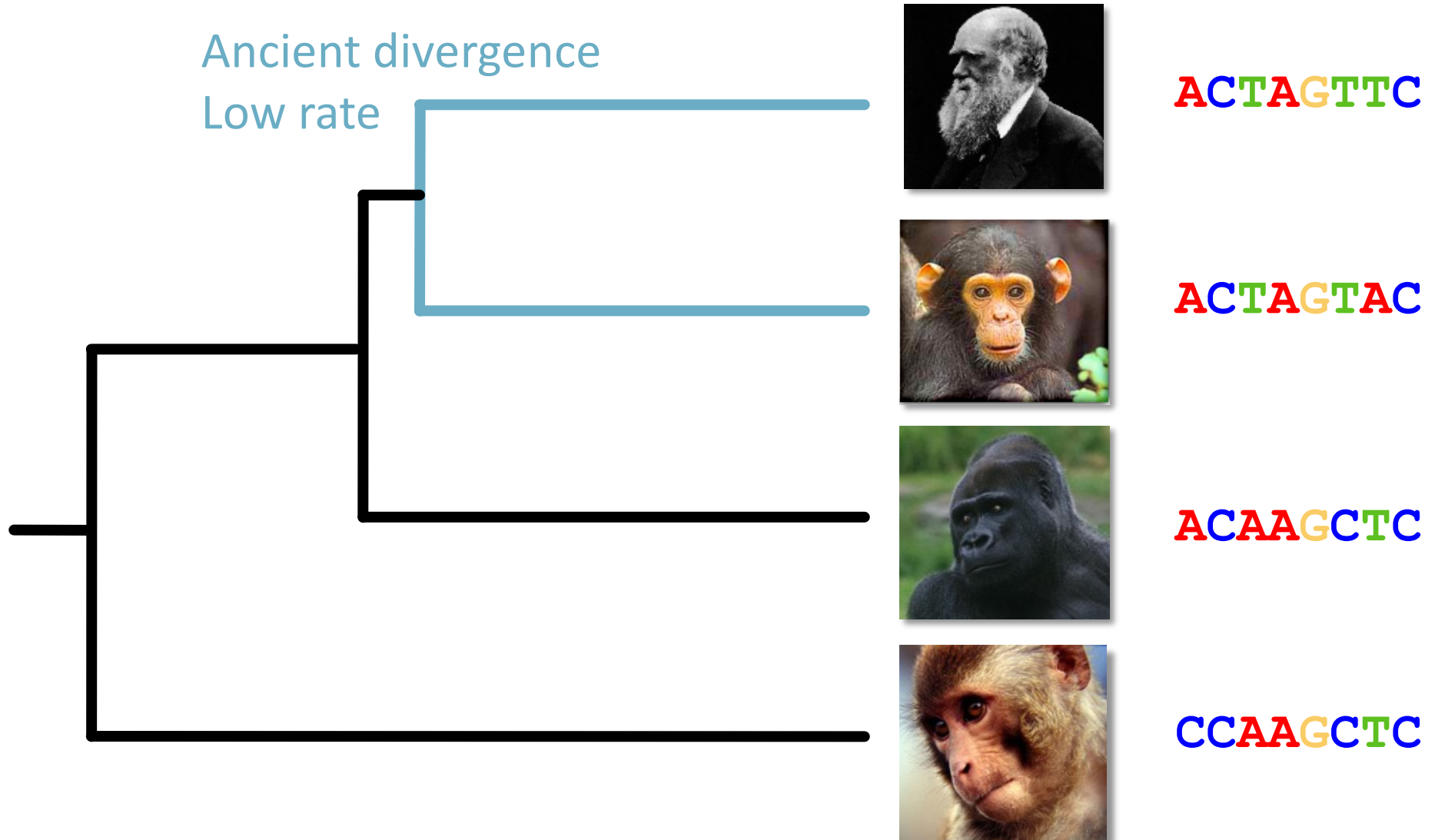
- Species delimitation
- Diversification rates
- Trait evolution
- Biogeography
- Epidemiology
- Demographic history

The Molecular Clock

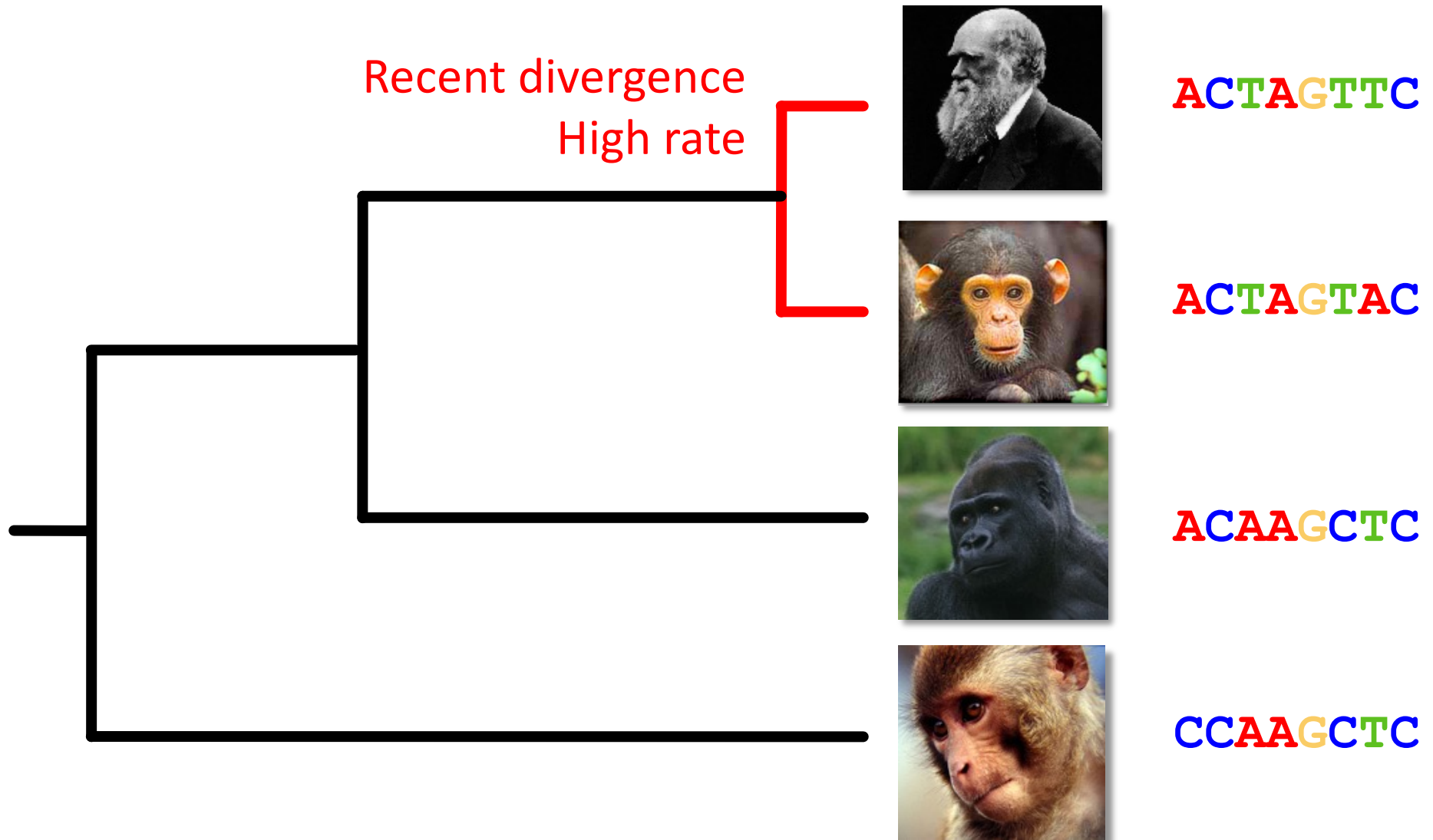
Rates and times



Rates and times



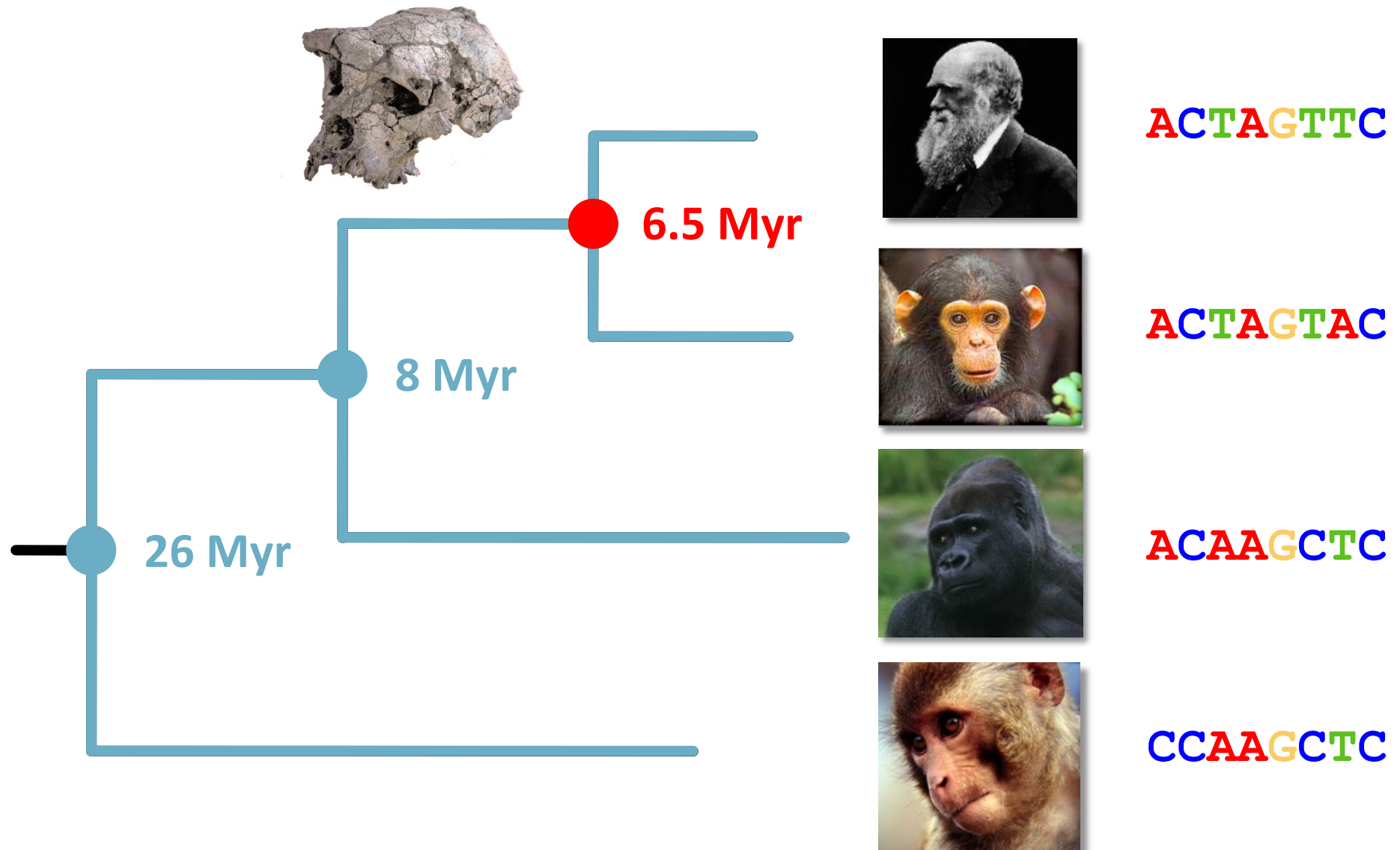
Rates and times



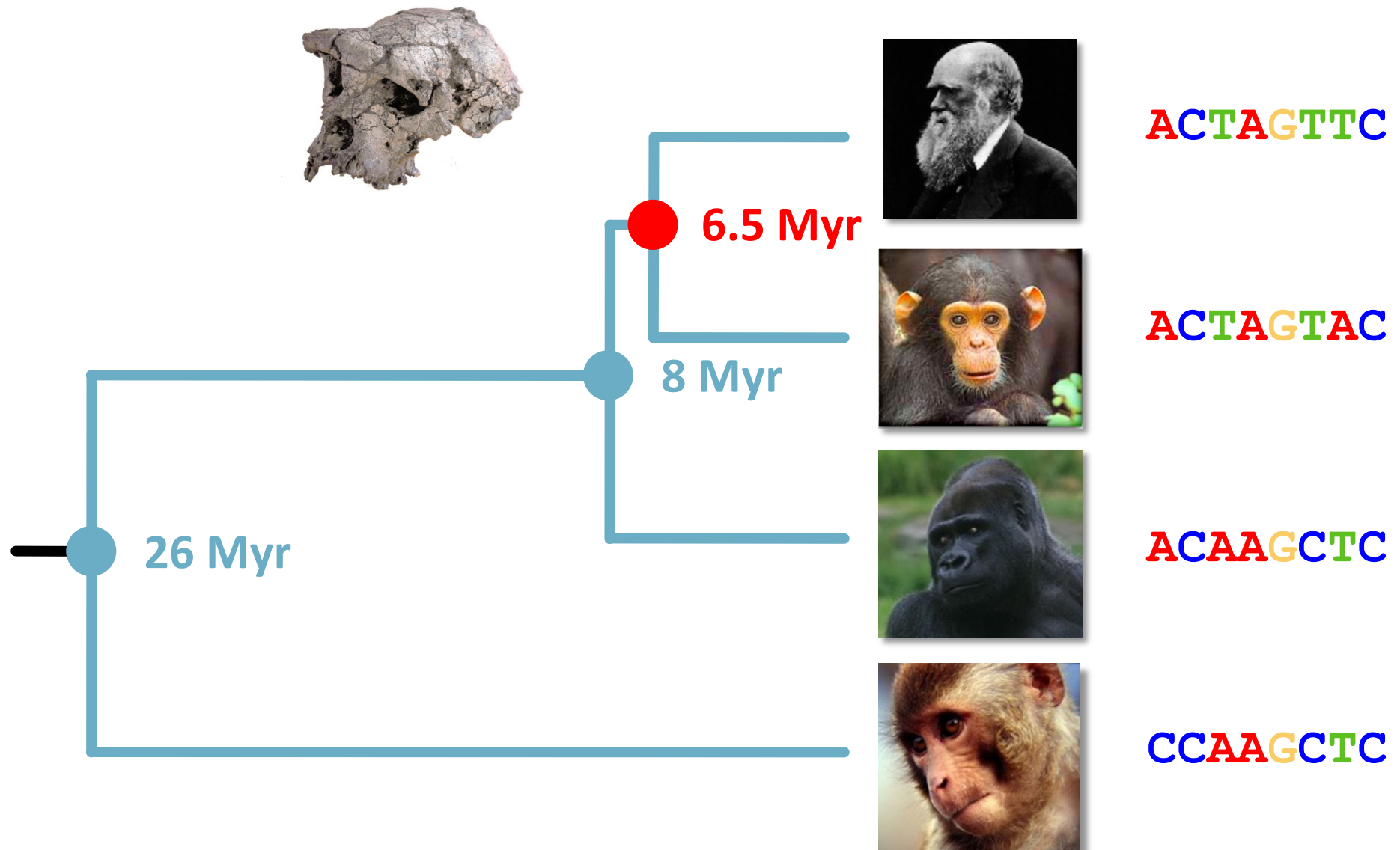
Rates and times

- Rates and times are **non-identifiable**
- Likelihood only depends on their product
 - Branch lengths in substitutions per site
- To separate rate and time, we need (prior) information about one or the other

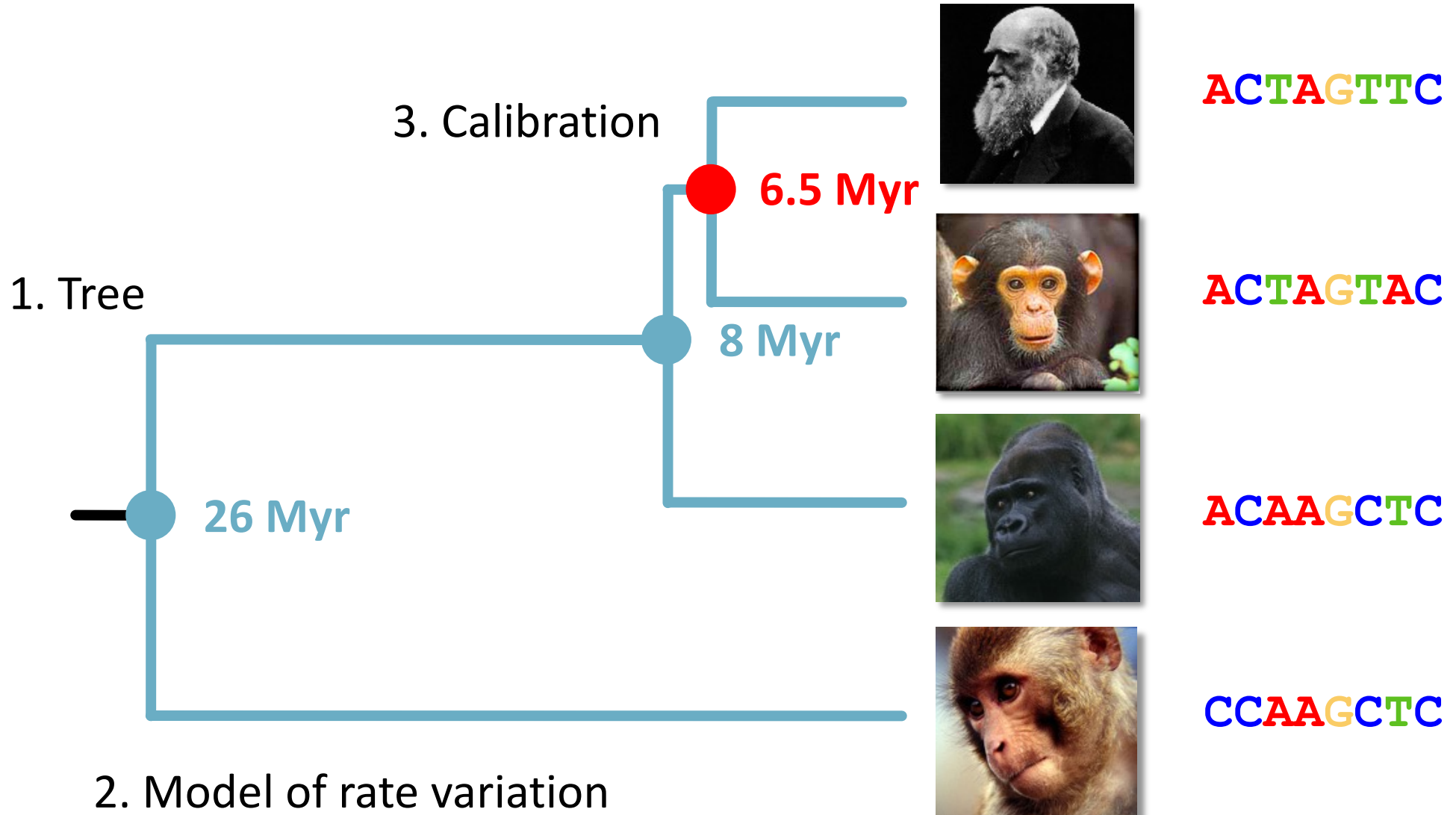
The molecular clock



The molecular clock

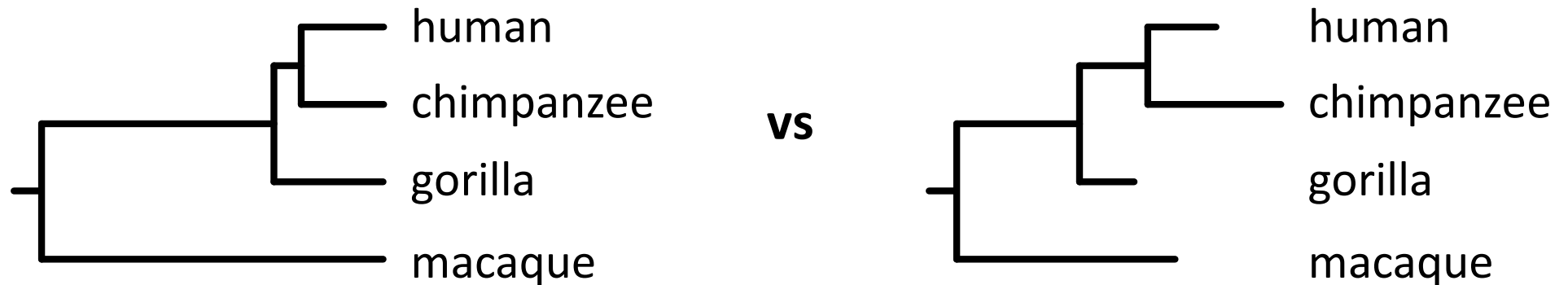


Sources of error



Testing for clocklike evolution

- Likelihood-ratio test

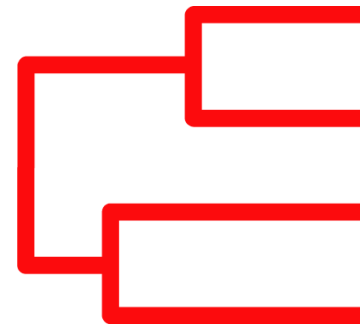


The assumption of rate constancy can be relaxed

Molecular-Clock Models

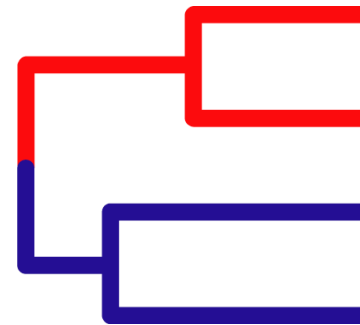
Molecular-clock models

Strict or 'global' molecular clock



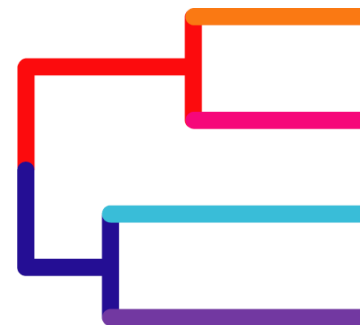
$$k = 1$$

Multi-rate clocks



$$1 < k < n$$

Relaxed clocks



$$k = n$$

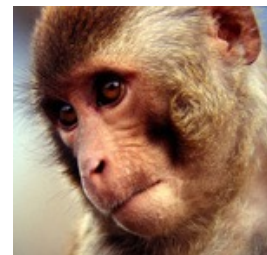
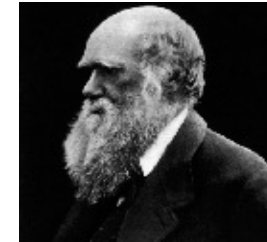
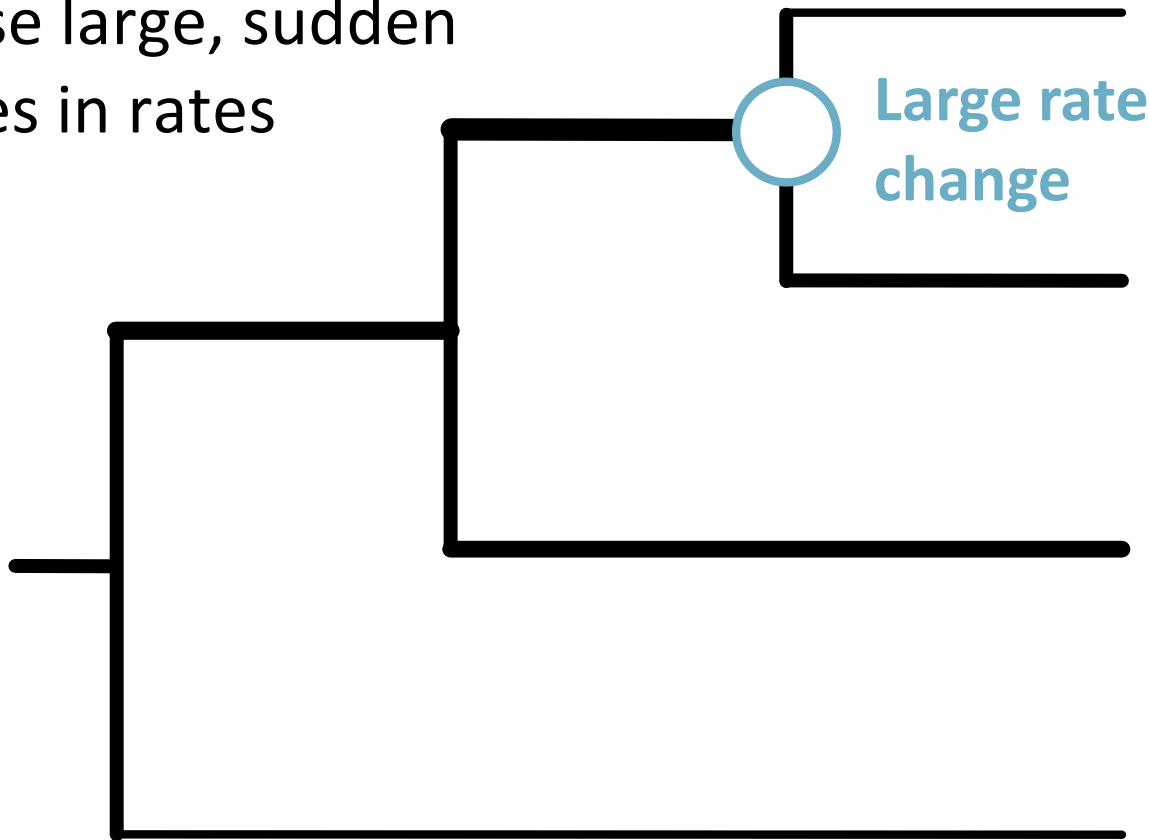
Relaxed clocks

- We know that life-history characteristics:
 - Have effects on rates of molecular evolution
 - Are usually heritable to some degree
- Treat molecular rate as a heritable trait
- Relaxed clocks generally assume that closely related species share similar rates



Likelihood-based relaxed clocks

- Allow a different rate in each branch
- Penalise large, sudden changes in rates

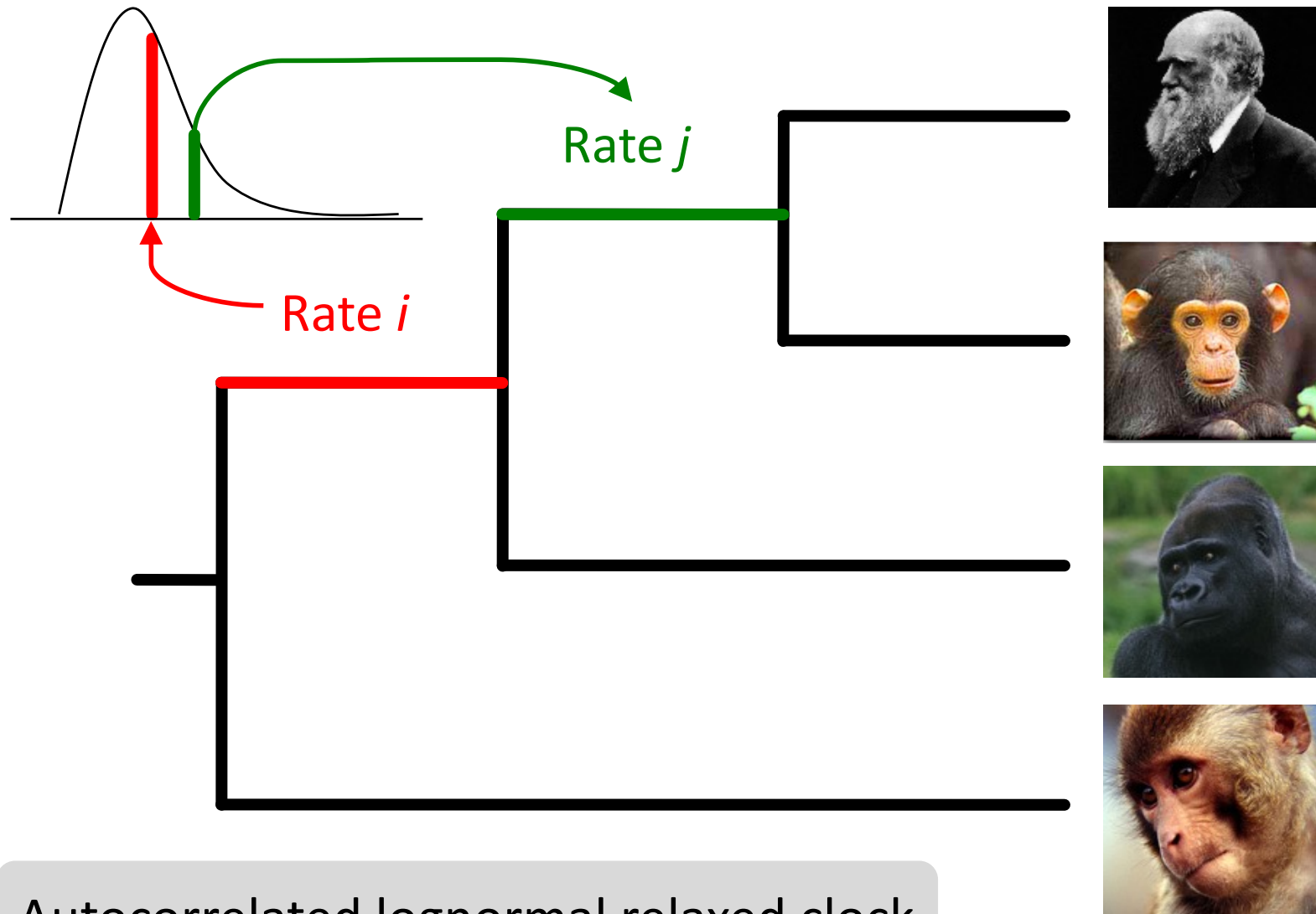


$$\text{Penalised likelihood} = \ln L - \lambda \Phi(r)$$

Bayesian relaxed clocks

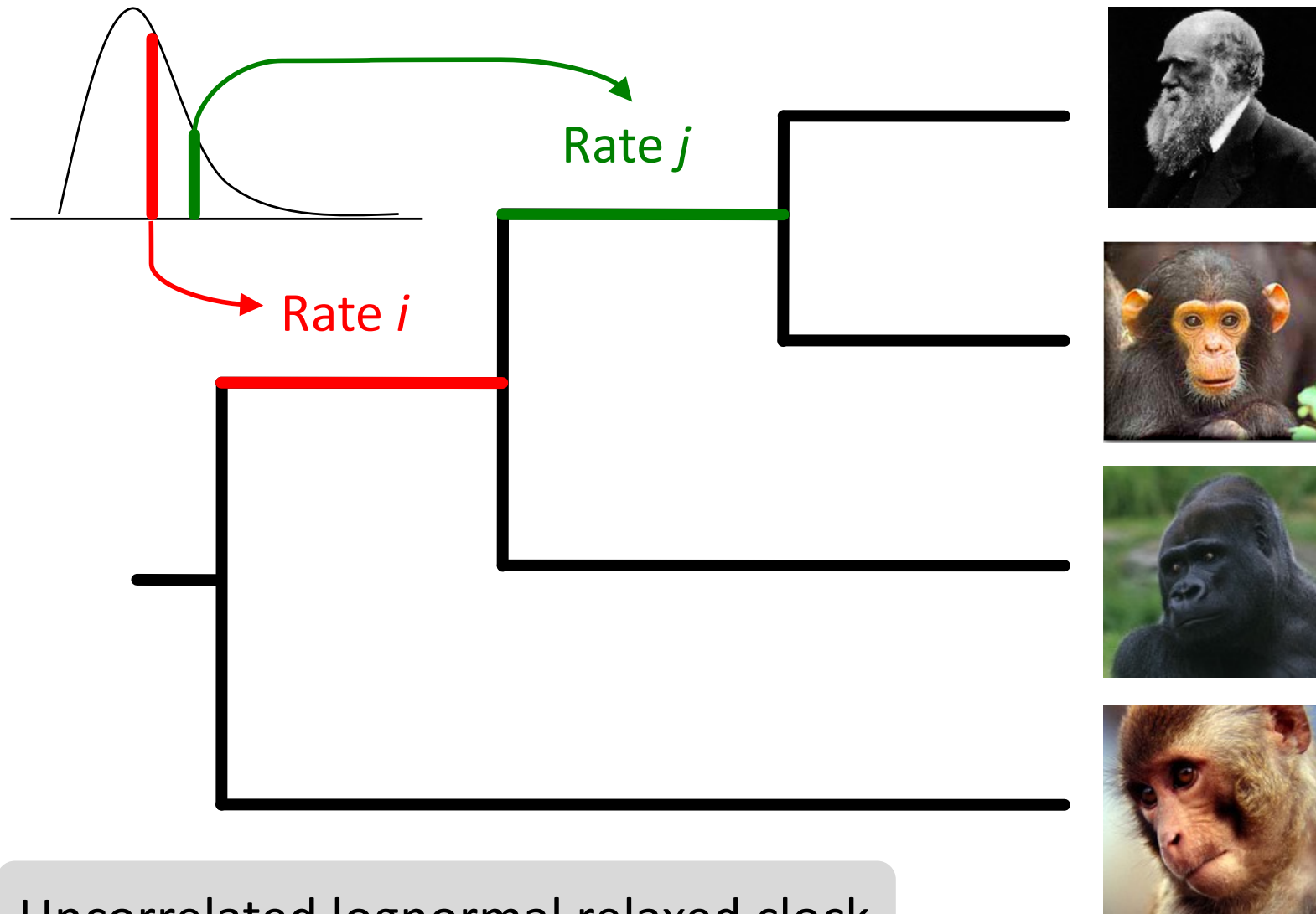
- Allow a different rate in each branch
- Statistical models of rates among branches
- Rates can be autocorrelated or uncorrelated
 - **Autocorrelated**
rates in neighbouring branches are related
 - **Uncorrelated**
rates identically and independently distributed among branches

Bayesian relaxed clocks



Autocorrelated lognormal relaxed clock

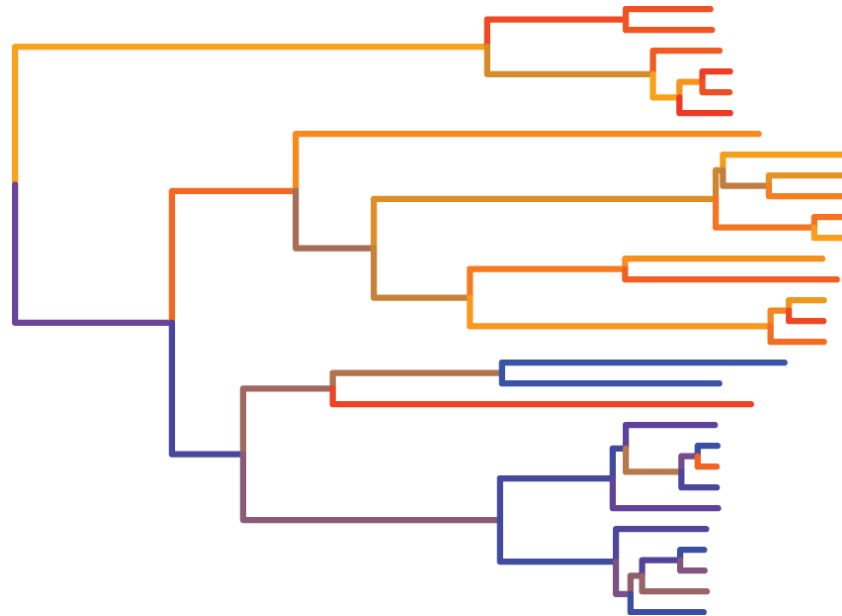
Bayesian relaxed clocks



Uncorrelated lognormal relaxed clock

Rate autocorrelation

- Little evidence of rate autocorrelation in real data
- Uncorrelated relaxed clock probably appropriate in most cases
- Compare estimates from auto- and uncorrelated relaxed clocks



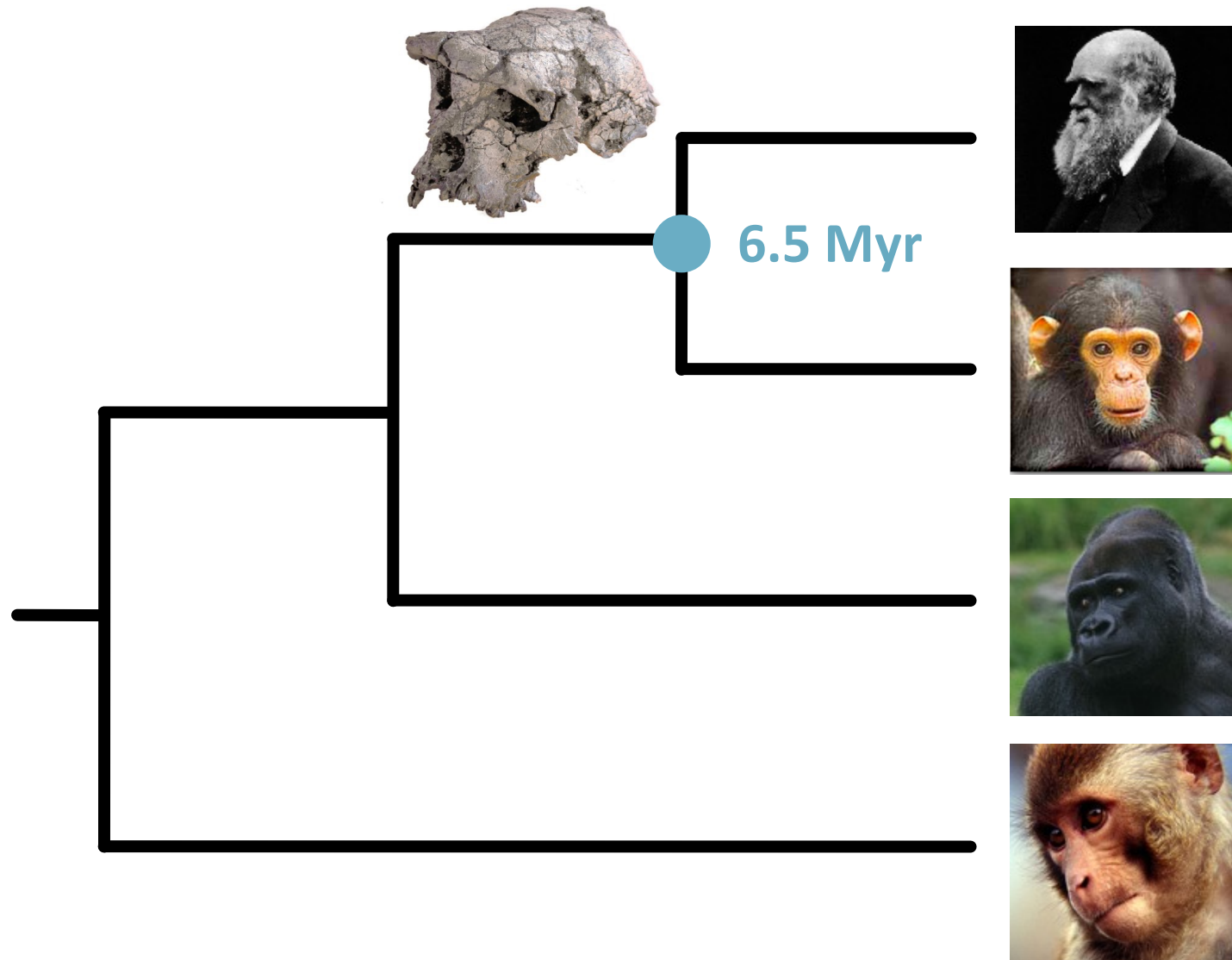
Calibrating the molecular clock

- Rates and times are **non-identifiable**
- Likelihood only depends on their product
 - Branch lengths in substitutions per site
- To separate rate and time, we need (prior) information about one or the other

Calibrating the molecular clock

- Information about **substitution rate**
 - Use to fix rate or to specific prior distribution of rate
- Information about **node times**
 - Fossil record
 - Biogeography
 - Sampling times
 - Documented pedigree

Calibration: Fossil record



Calibration: Fossil record

1. Use fossil data to inform priors on node times

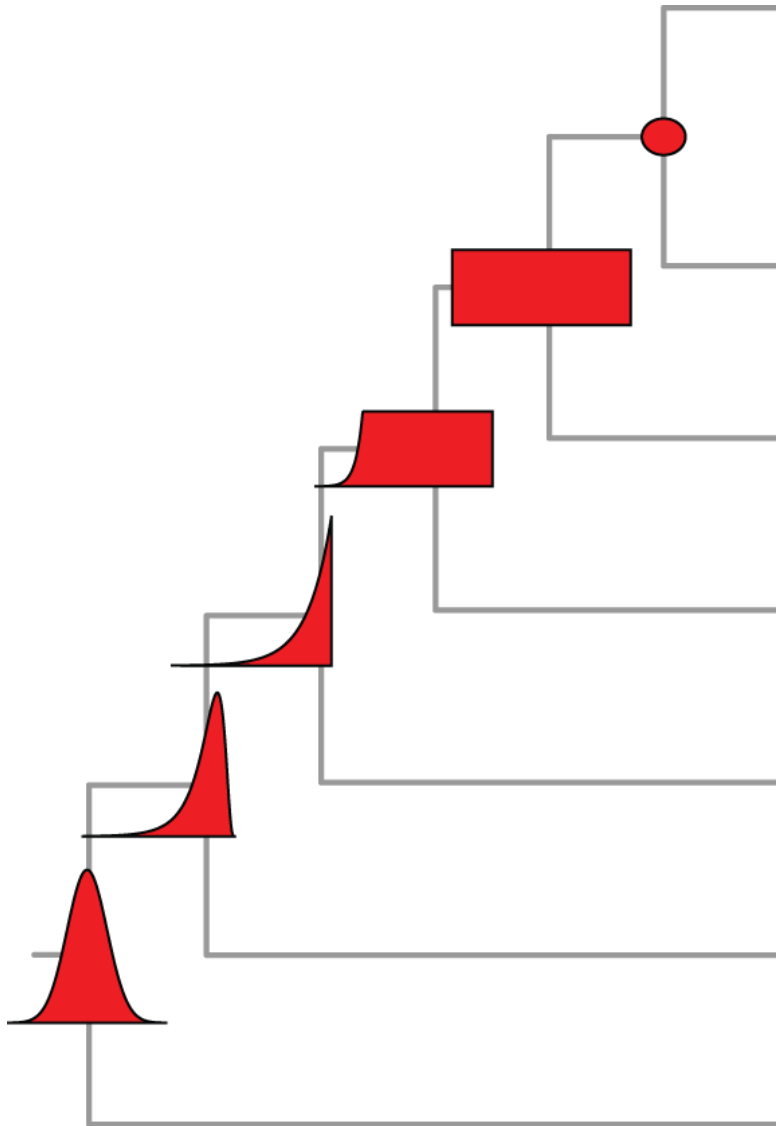
- Minimum age of a node based on oldest fossil assignable to any of its descendent lineages
- Prior distribution of node age specified by user

2. Use fossil directly in the analysis

- Model diversification process use fossil occurrence data
- Include fossil taxa in the data matrix (total-evidence dating)

Calibration Priors

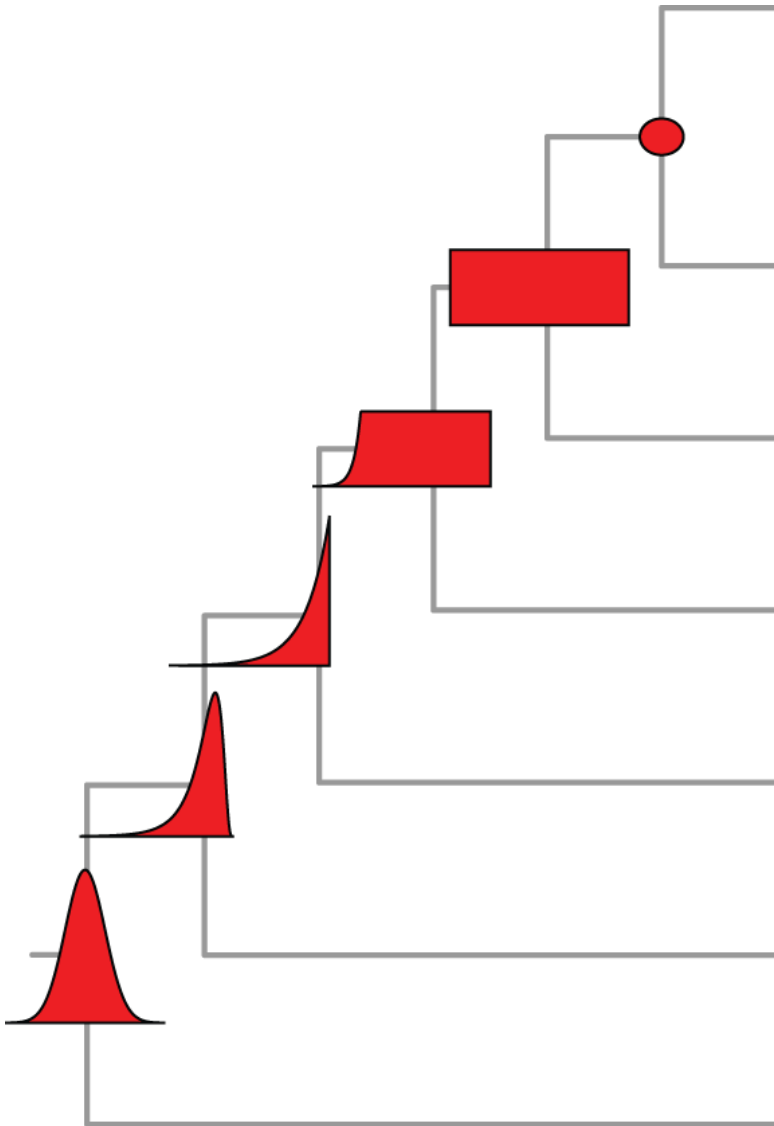
Calibrations



Point calibration

- Ignores uncertainty due to preservational biases, isotopic dating errors, *etc.*

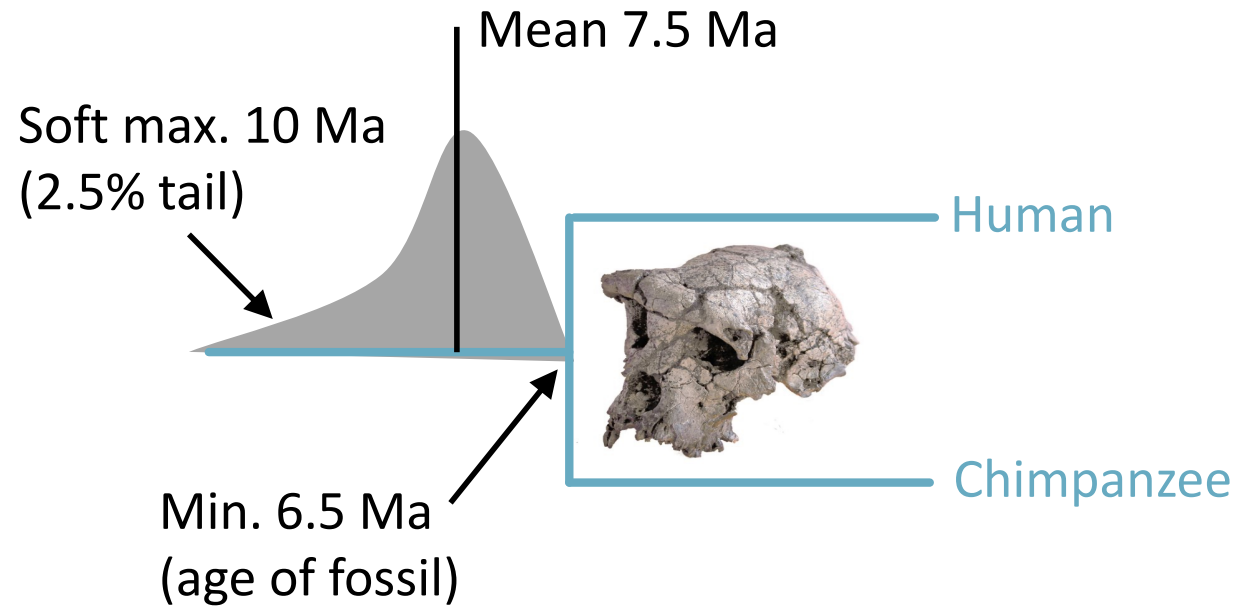
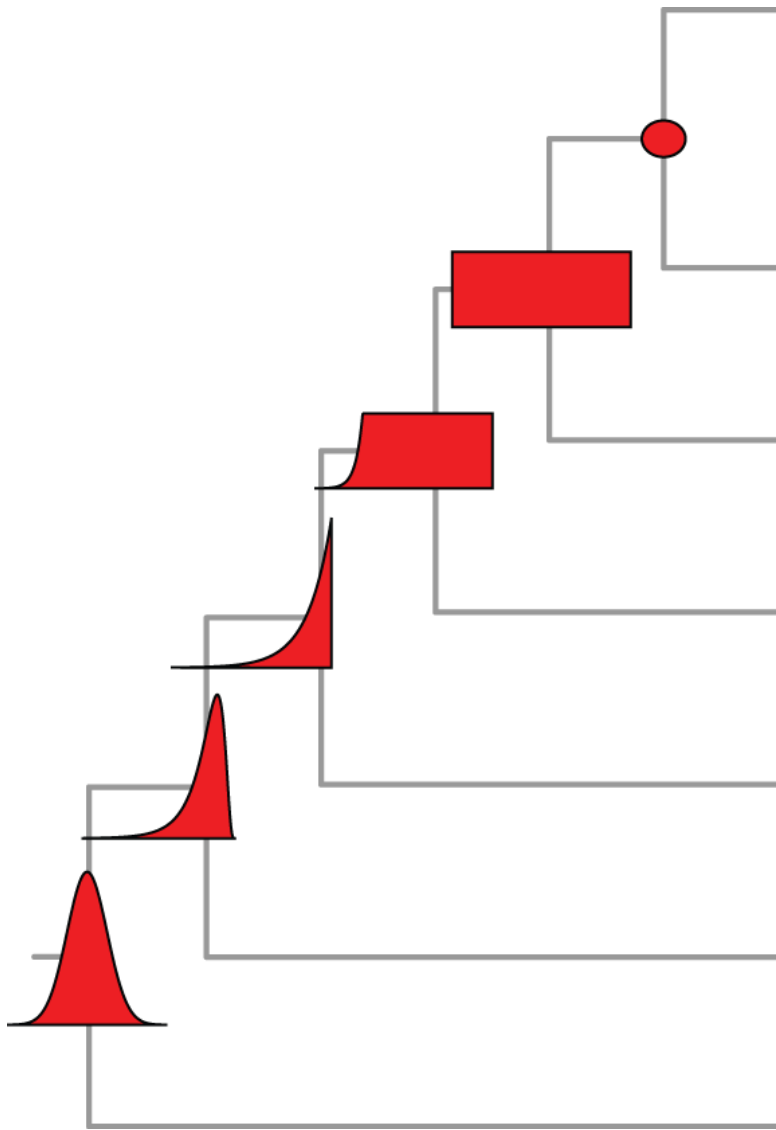
Calibrations



Uniform prior

- Combination of hard minimum and maximum bounds
- Does not effectively use information at hand
- Difficult to choose useful maximum bounds

Calibrations

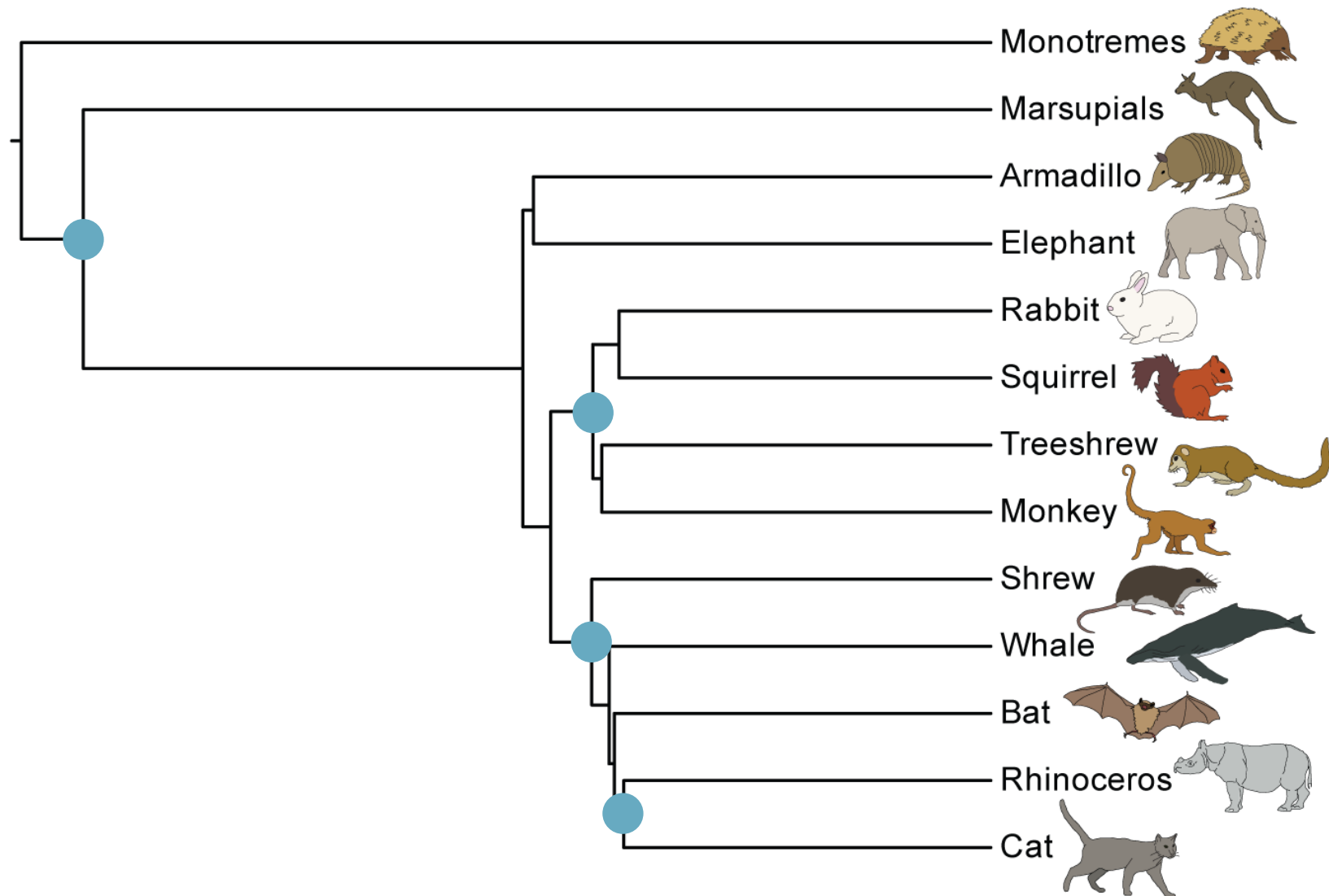


Lognormal prior

- Need 3 values: minimum, mean, and stdev
- Perhaps the most appropriate for fossils

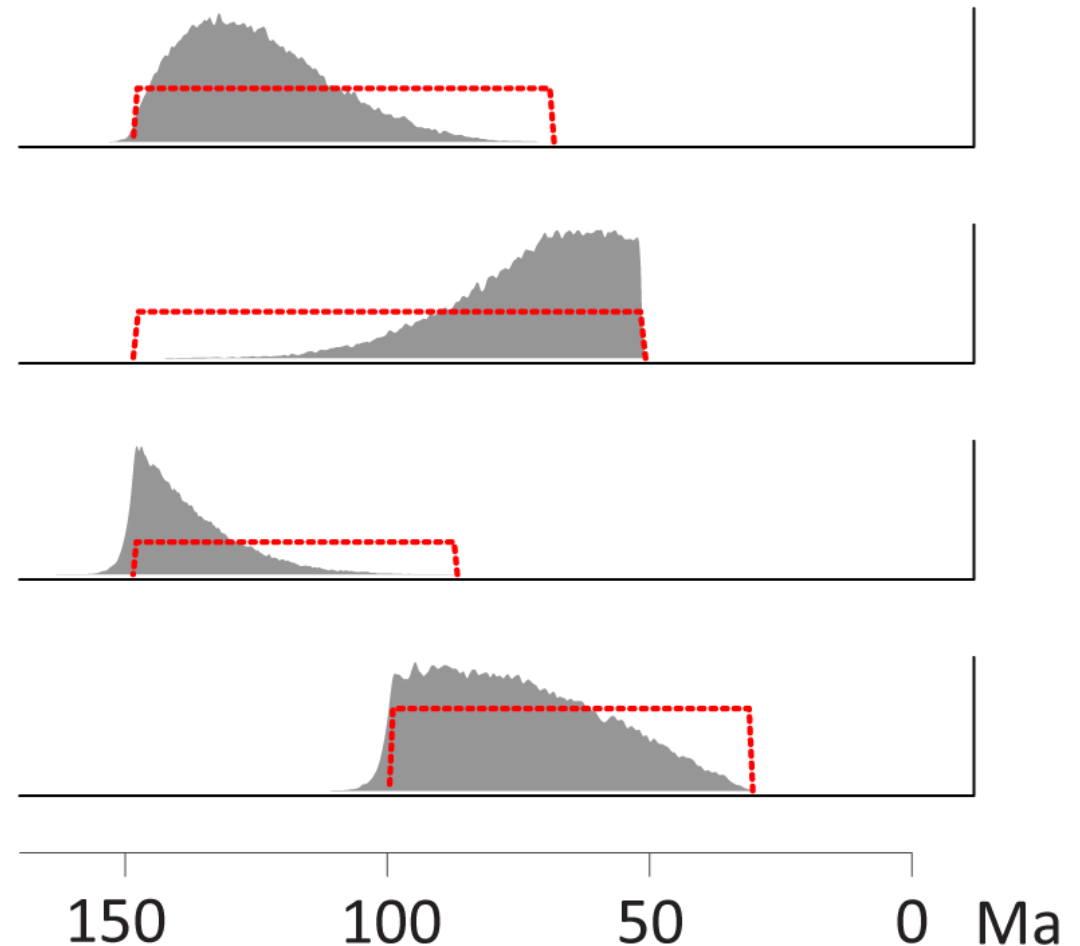
Multiple calibrations

- Use multiple calibrations if possible

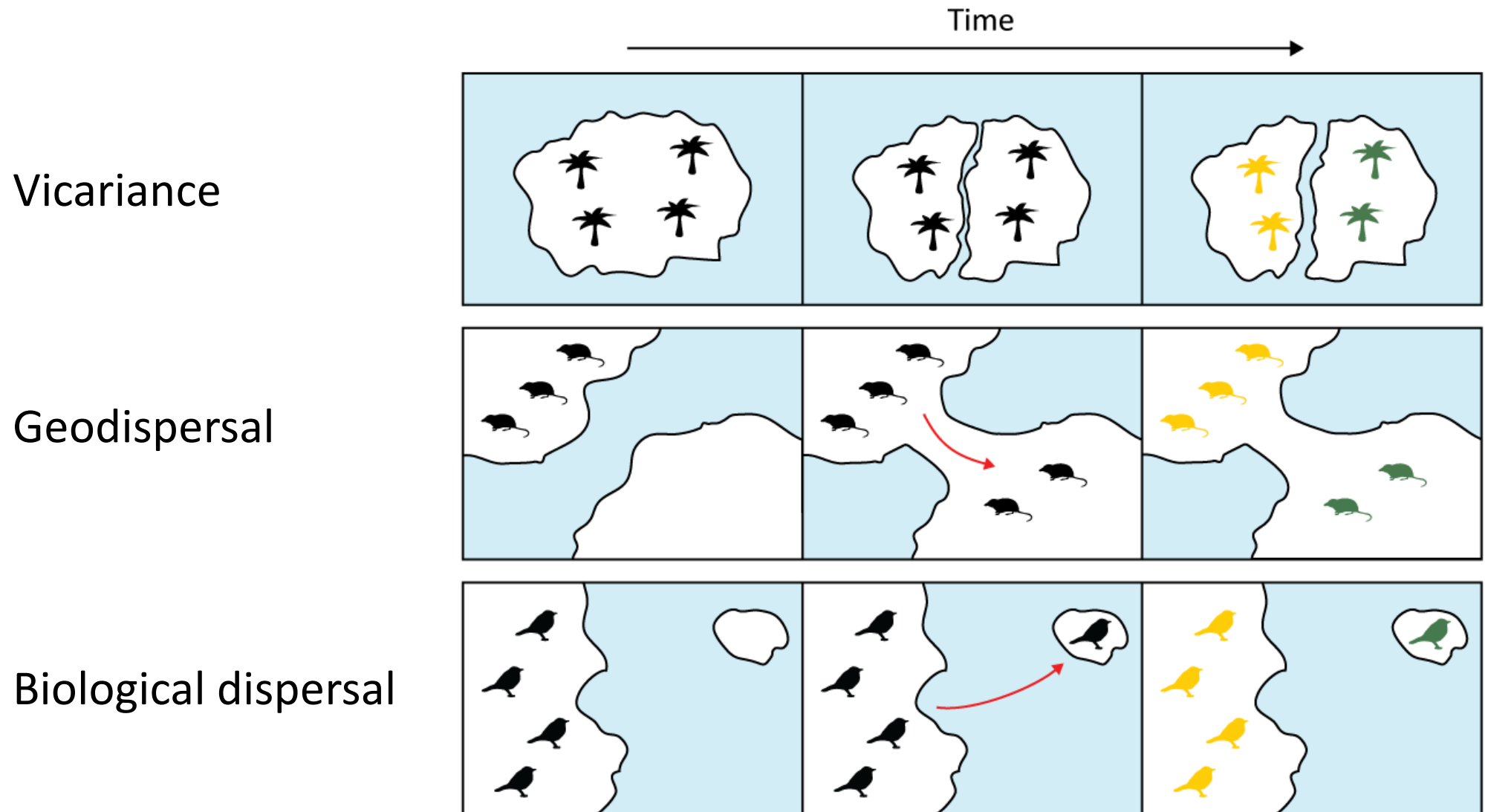


Multiple calibrations

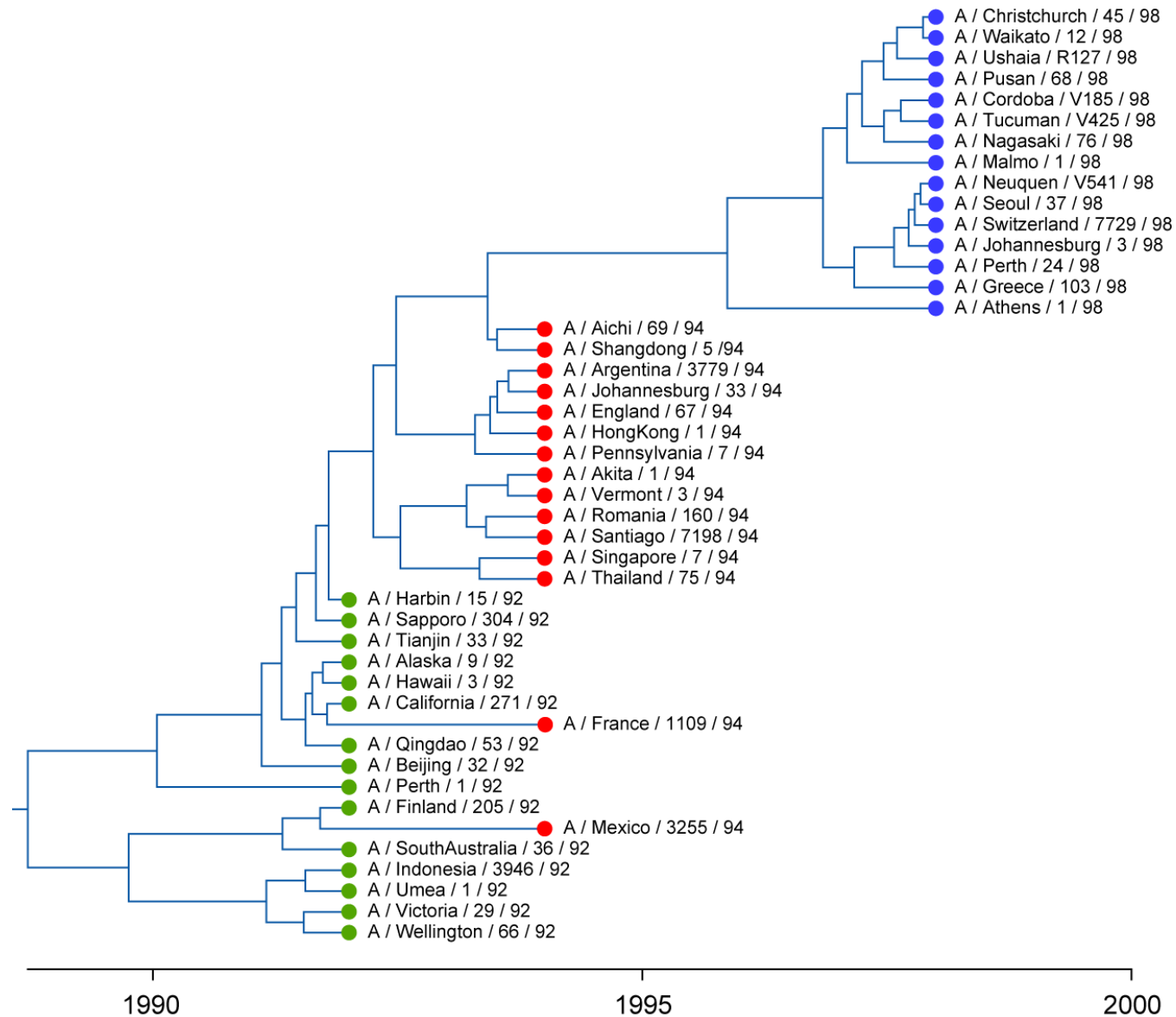
- Priors on node ages are the joint product of the tree prior and the user-specified calibration priors
- These priors can interact
- Marginal priors can differ from user-specified priors



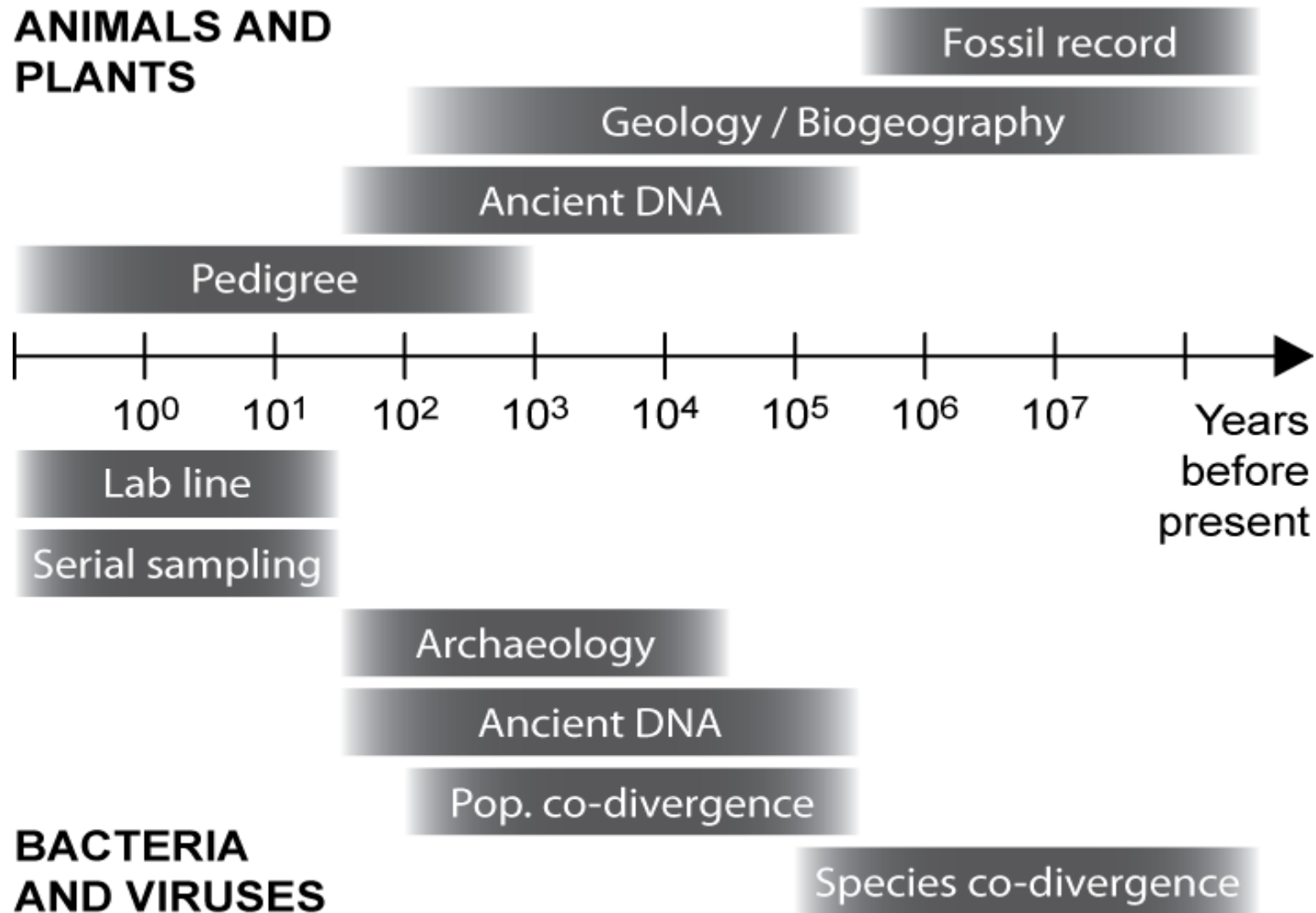
Calibration: Biogeography



Calibration: Sampling times



Calibrations



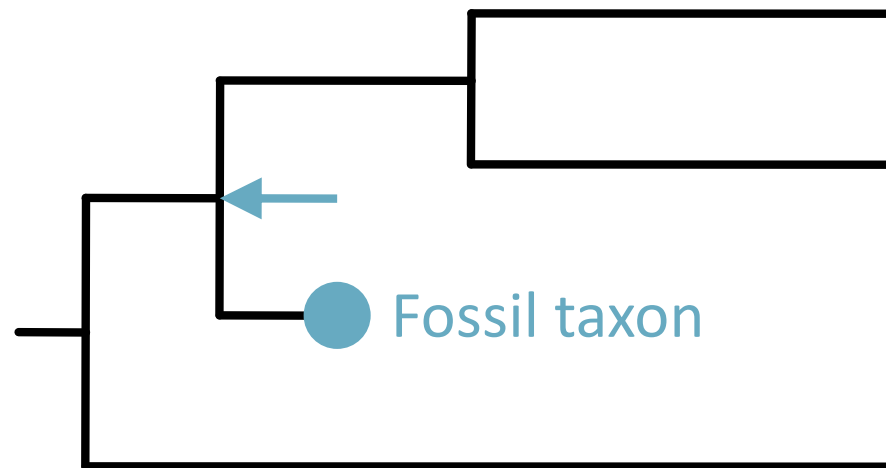
Total-Evidence Dating

Total-evidence dating

- Combined morphological and molecular data set
- Fossil taxa included
 - Phylogenetic placement estimated using morphological data
 - Age acts as a calibration by constraining the age of parent node
- Birth-death tree prior

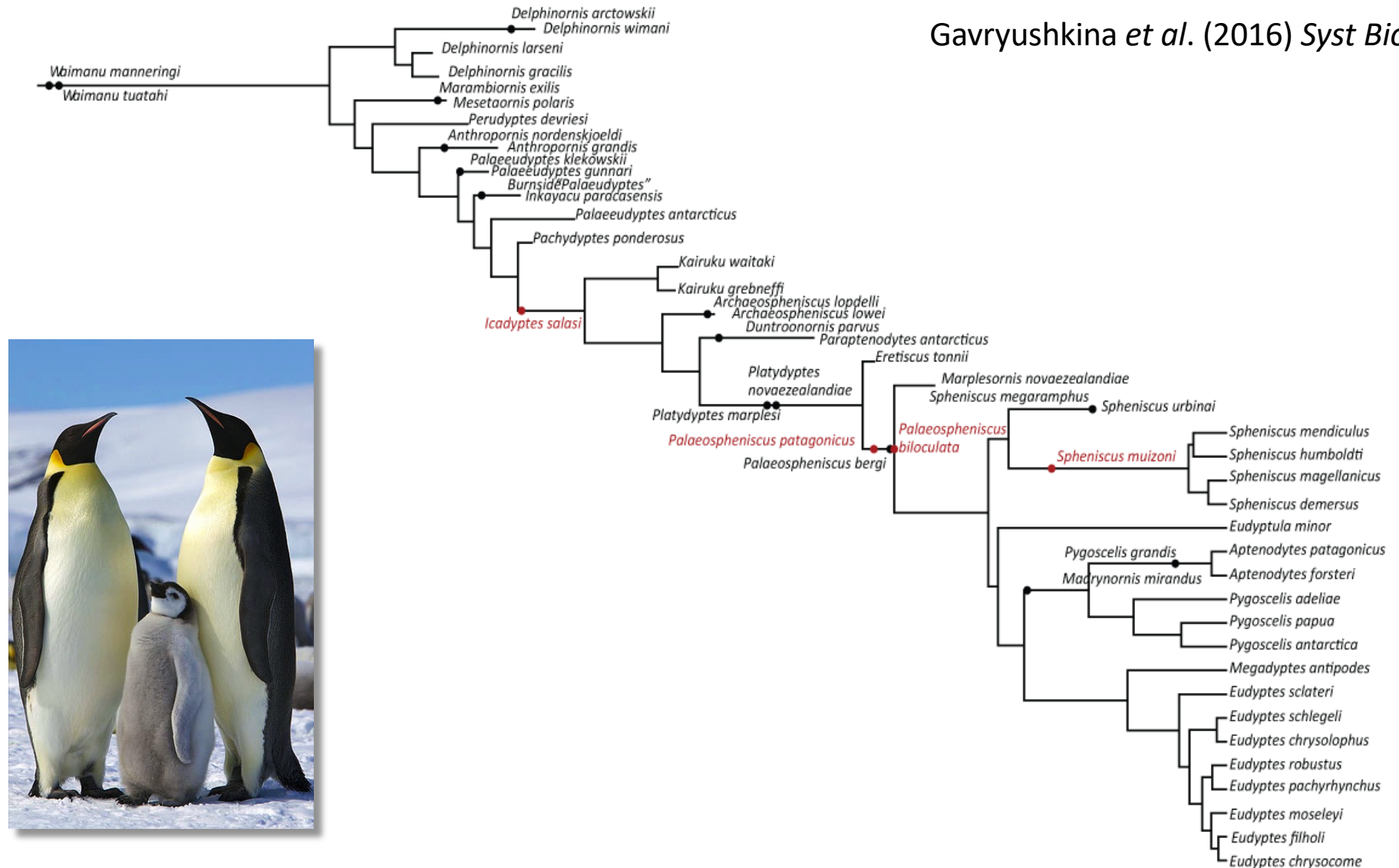
Total-evidence dating

- Avoids the need to construct priors for ages of internal nodes
- Can include fossil taxa with uncertain phylogenetic position
- Can provide sufficient calibration without the need to include maximum age constraints



Total-evidence dating

Gavryushkina *et al.* (2016) *Syst Biol*



Useful references

- **A practical guide to molecular dating**
Sauquet (2013) *Comptes Rendus Palevol*, 12: 355–367.
- **Bayesian molecular clock dating of species divergences in the genomics era**
dos Reis, Donoghue, & Yang (2016) *Nature Reviews Genetics*, 17: 71–80.
- **Molecular-clock methods for estimating evolutionary rates and timescales**
Ho & Duchêne (2014) *Molecular Ecology*, 23: 5947–5965.

