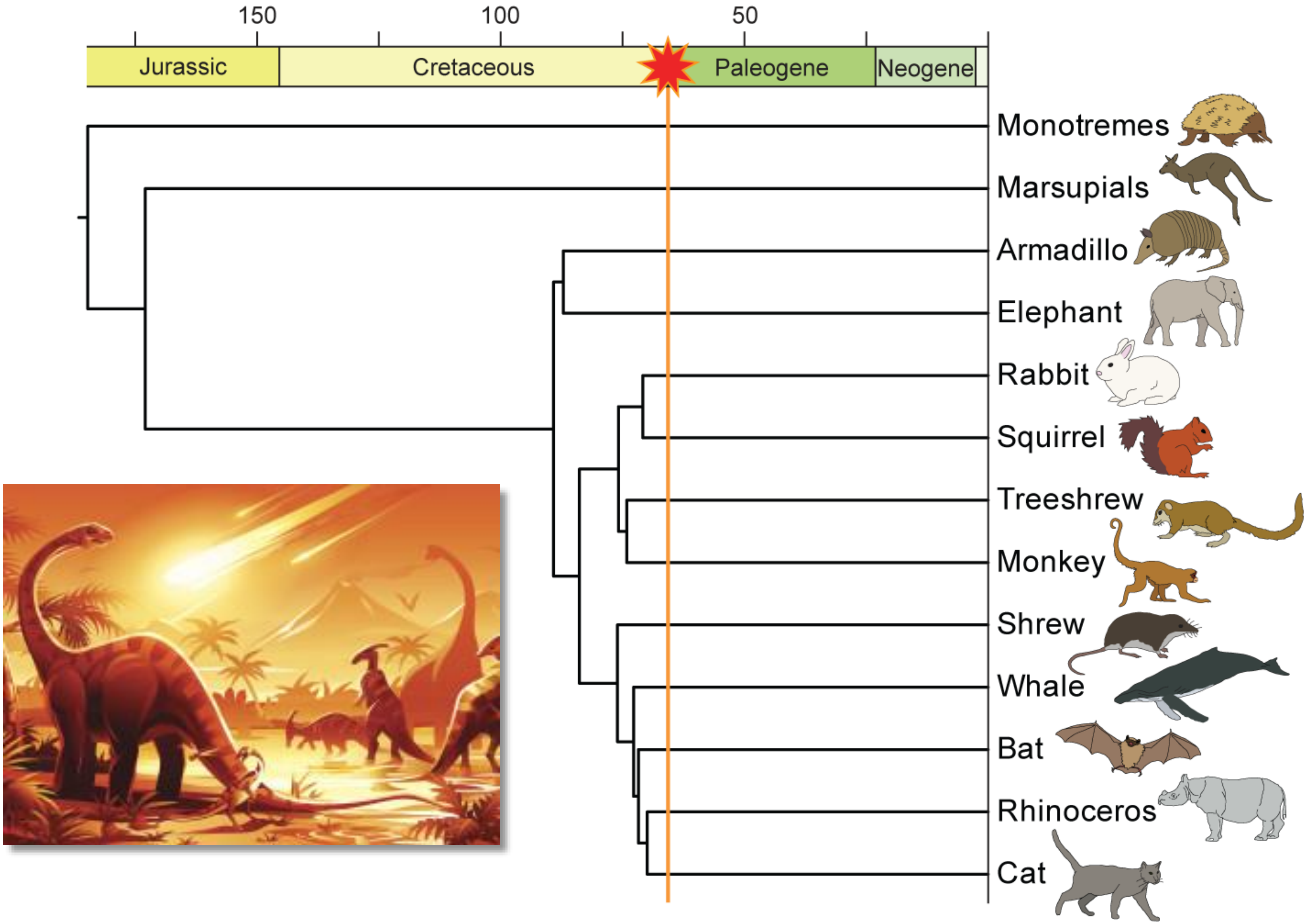
Molecular dating

Relaxed Phylogenetics and Dating with Confidence

Alexei J Drummond, Simon Y. W Ho, Matthew J Phillips, Andrew Rambaut

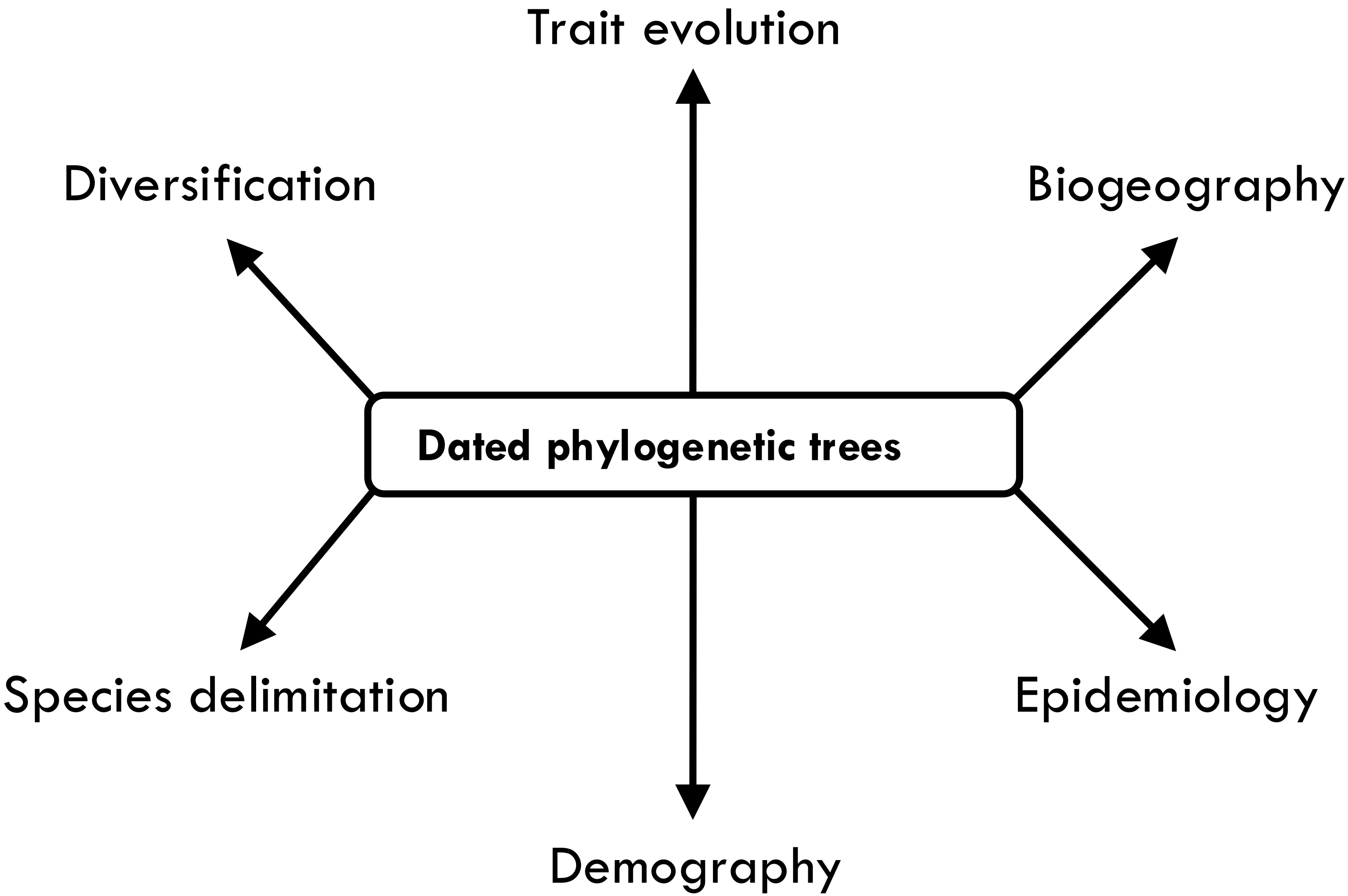
Published: March 14, 2006 • https://doi.org/10.1371/journal.pbio.0040088

Evolutionary timescales



dos Reis *et al.* (2012) *Proc Roy Soc B*

Evolutionary timescales



Trait evolution

Diversification

Biogeography

Dated phylogenetic trees

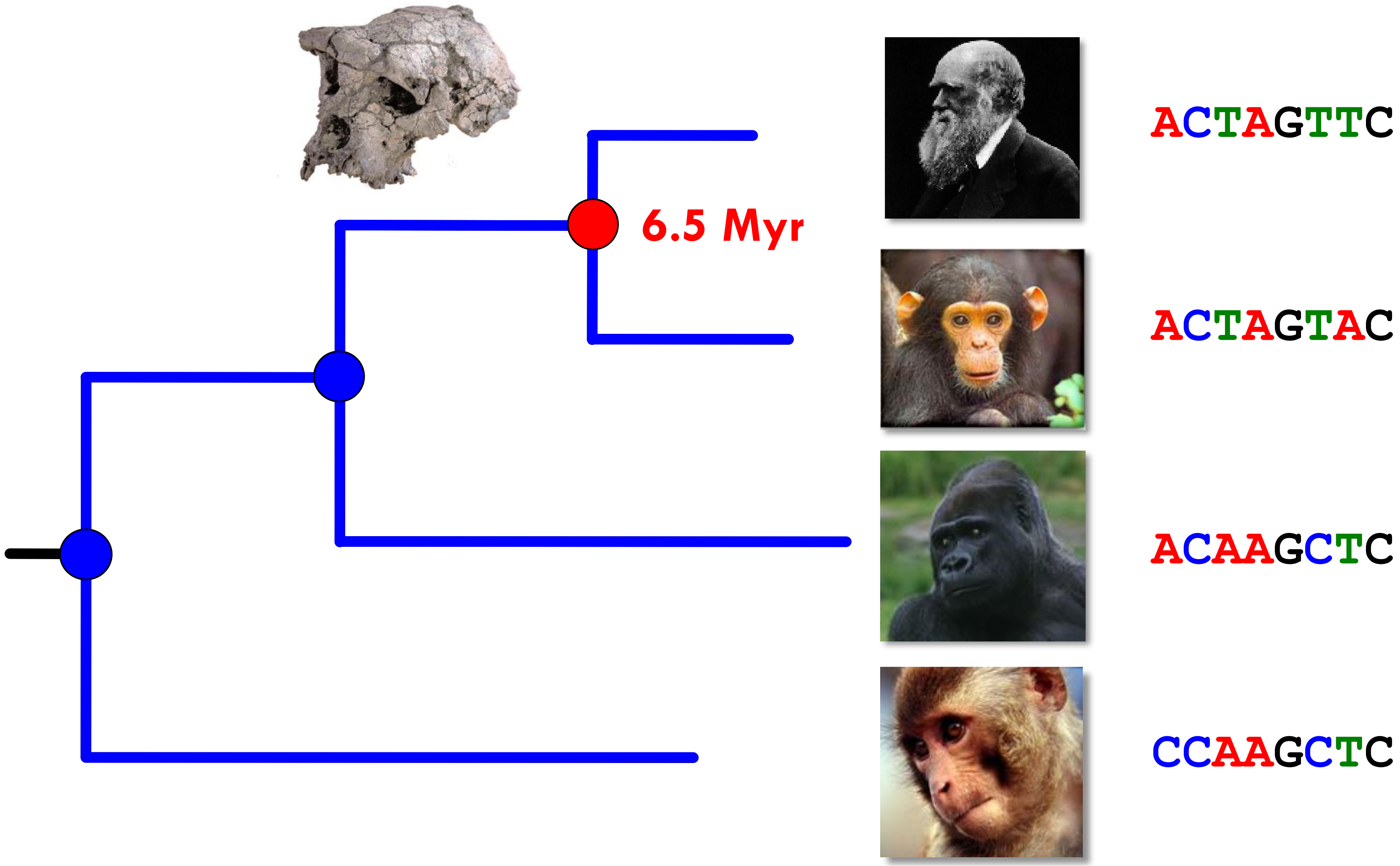
Species delimitation

Epidemiology

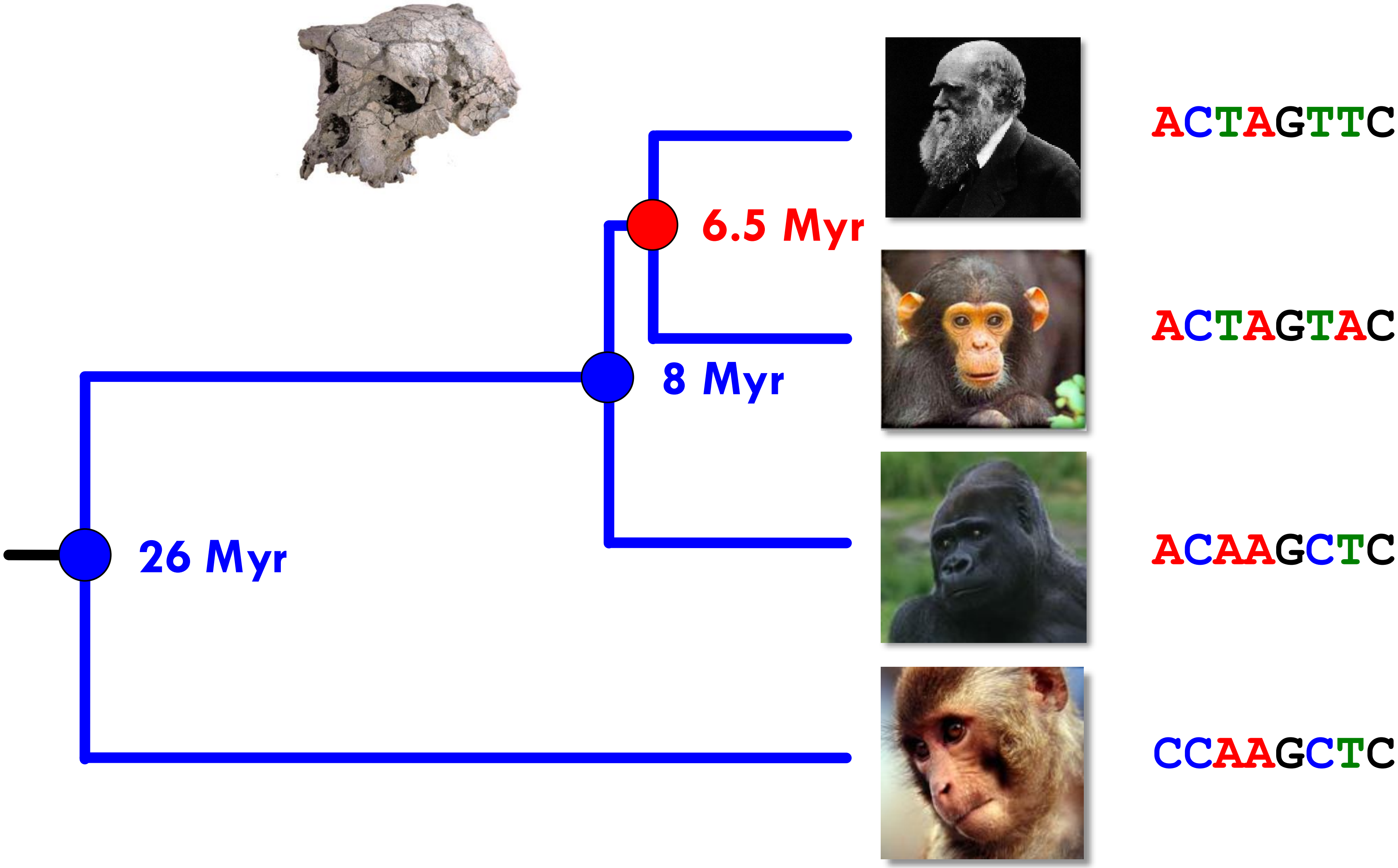
Demography

The Molecular Clock

The molecular clock



The molecular clock



6.5 Myr

8 Myr

26 Myr

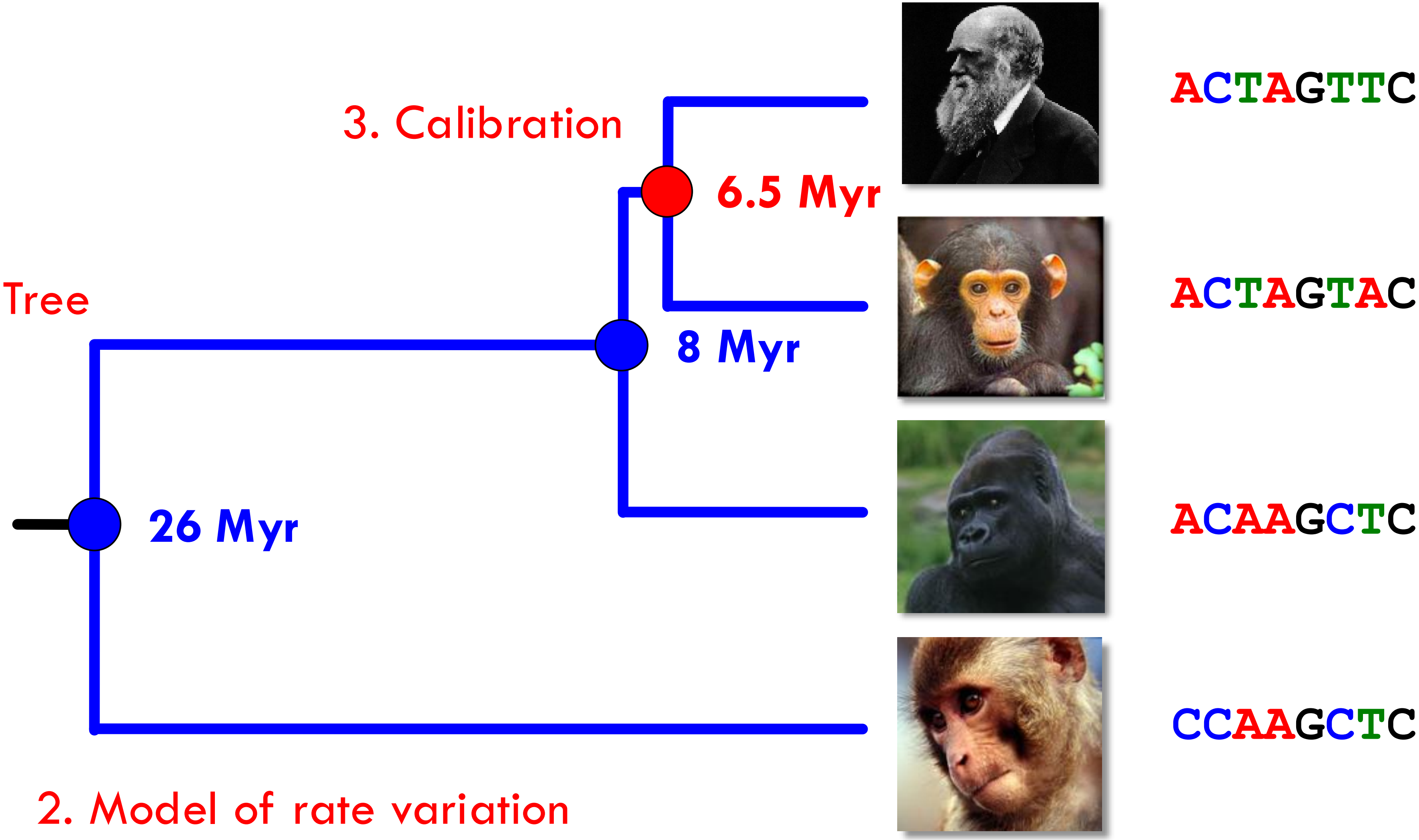
ACTAGTTC

ACTAGTAC

ACAAGCTC

CCAAGCTC

Sources of error



1. Tree

2. Model of rate variation

3. Calibration

ACTAGTTC
ACTAGTAC
ACAAGCTC
CCAAGCTC

A brief history

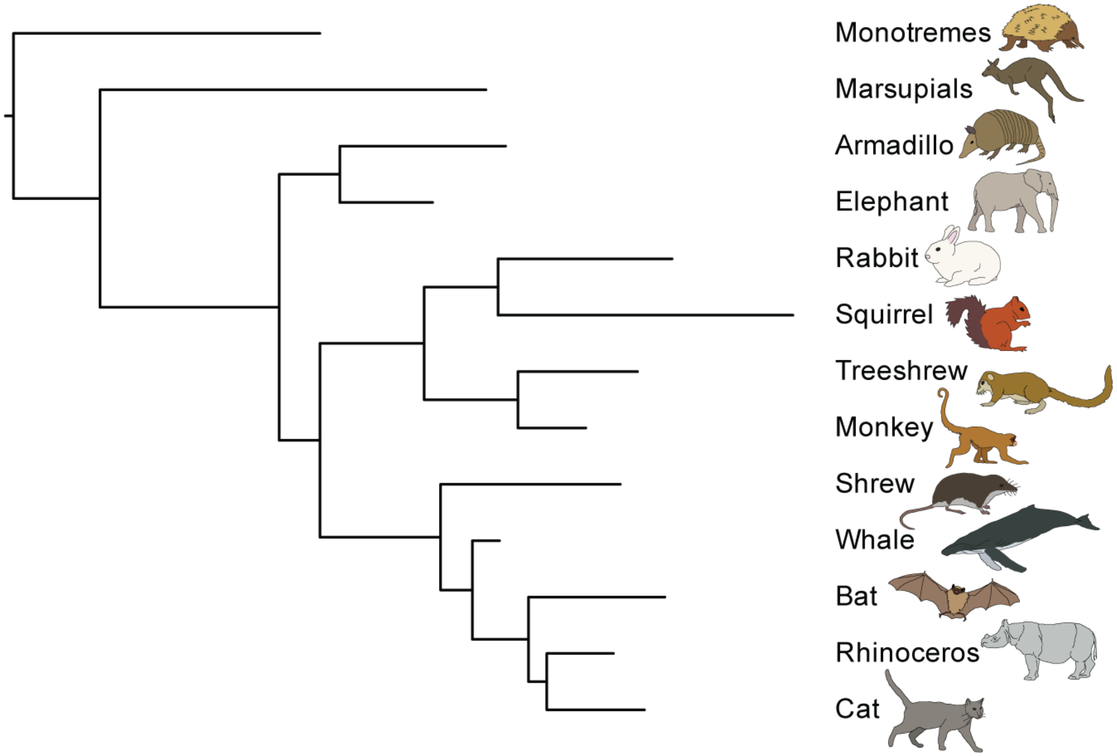
- Emile Zuckerkandl and Linus Pauling

1962 Assumed constant rate among species to estimate timing of globin gene duplications

1965 Introduced the term ‘molecular evolutionary clock’

Departures from the clock

- Rates vary among lineages because of differences in:
  - Exposure to mutagens
  - Metabolic rate
  - Generation time
  - Population size
  - Strength and direction of selection

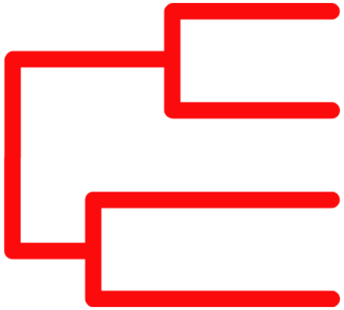


The assumption of rate constancy can be relaxed

Molecular-Clock Models

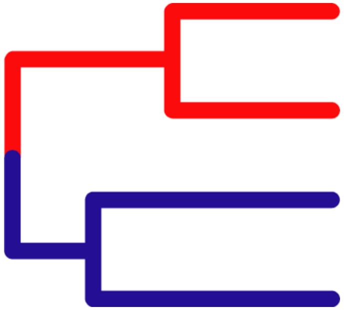
Molecular-clock models

Strict or ‘global’ molecular clock



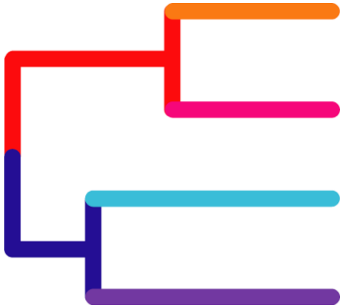
$k = 1$

Multi-rate clocks



$1 < k < n$

Relaxed clocks



$k = n$

Multi-rate clocks

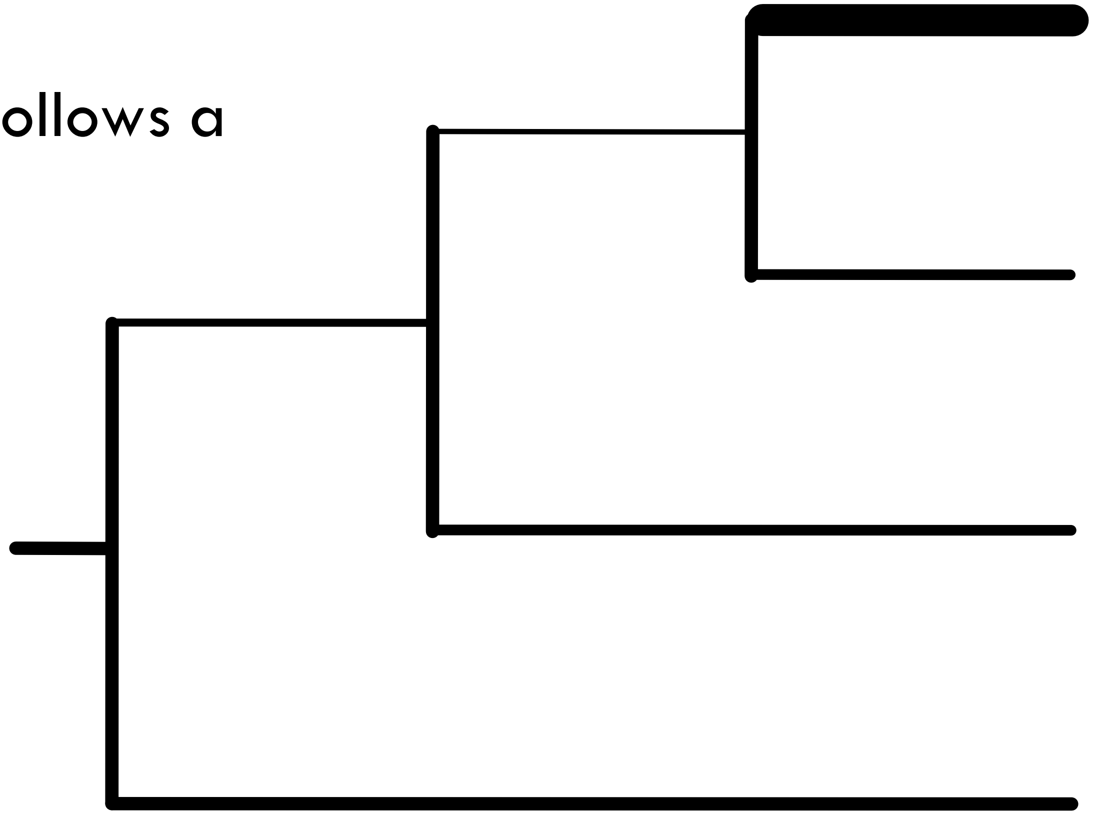
- Small number of rates
  - More than 1 rate (*i.e.*, not a strict clock)
  - Fewer than number of branches (*i.e.*, not a relaxed clock)
- **Local clock**
  - Same rate shared by neighbouring branches
- **Discrete clock**
  - Small number of branch rates, distributed across tree

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

LSD dating

- Variation in rates follows a gaussian model



$$b_i = \omega\left(t_i - t_{a(i)}\right) + \epsilon_i$$

$$\epsilon_i = N\left(0, \sigma_i^2\right)$$

To et al. (2015, *Syst. Biol.*)

Bayesian relaxed clocks

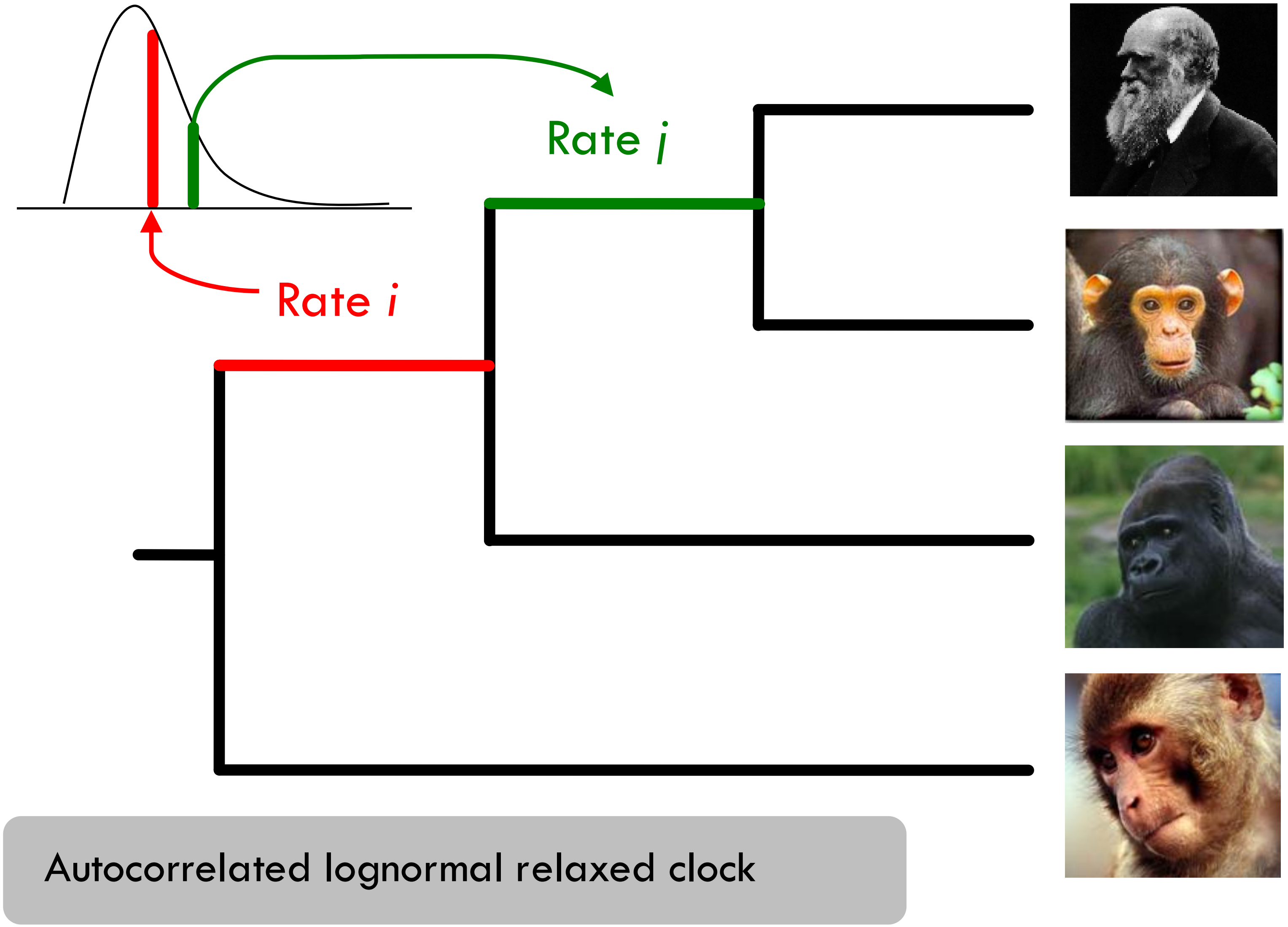
- Allow a different rate in each branch
- Statistical models of rates among branches
- Rates can be autocorrelated or uncorrelated
  - **Autocorrelated**
    rates in neighbouring branches are related
  - **Uncorrelated**
    rates identically and independently distributed among branches

Relaxed Phylogenetics and Dating with Confidence

Alexei J Drummond, Simon Y. W Ho, Matthew J Phillips, Andrew Rambaut

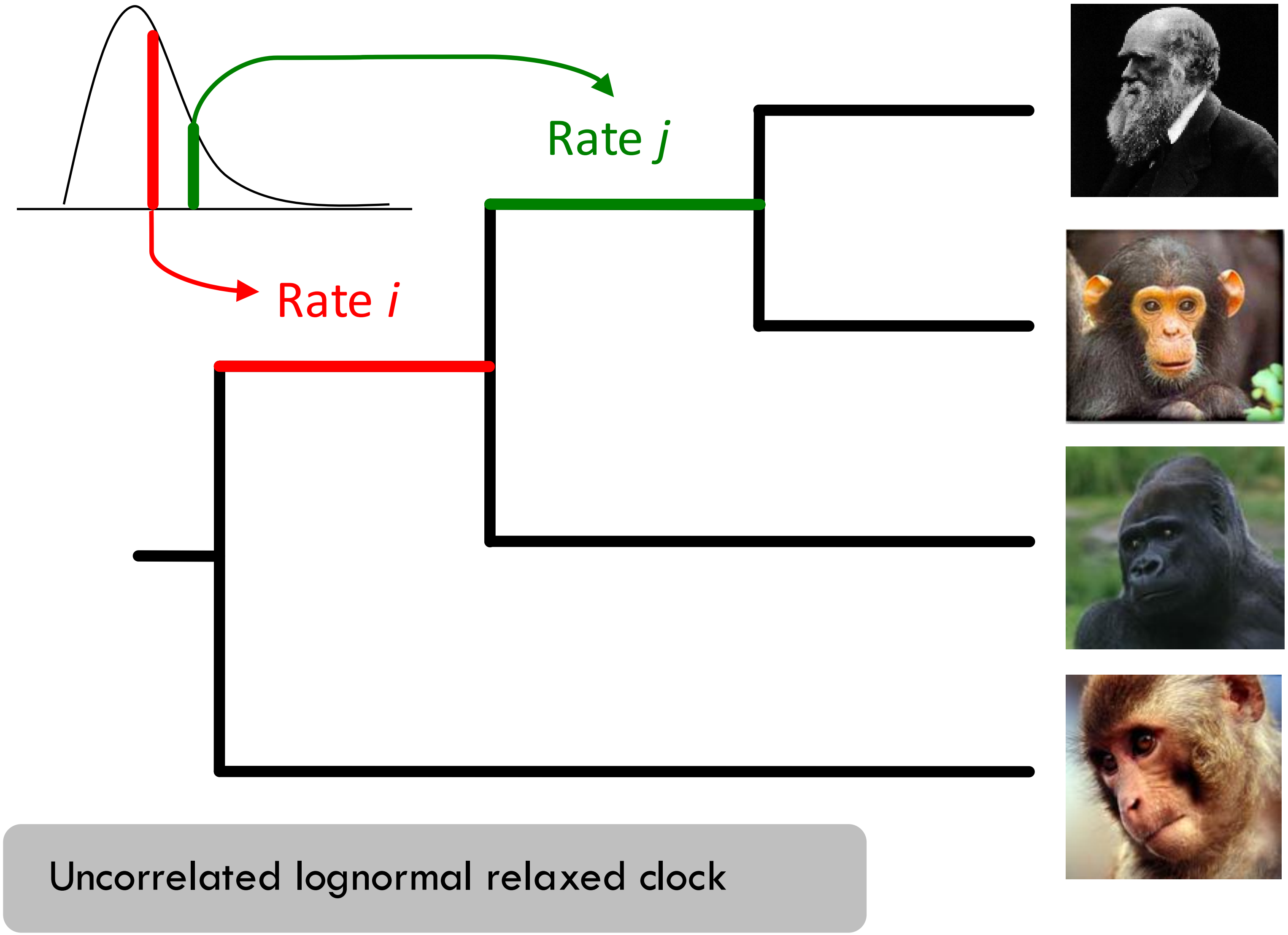
Published: March 14, 2006 • https://doi.org/10.1371/journal.pbio.0040088

Bayesian relaxed clocks



Autocorrelated lognormal relaxed clock

Bayesian relaxed clocks



Uncorrelated lognormal relaxed clock

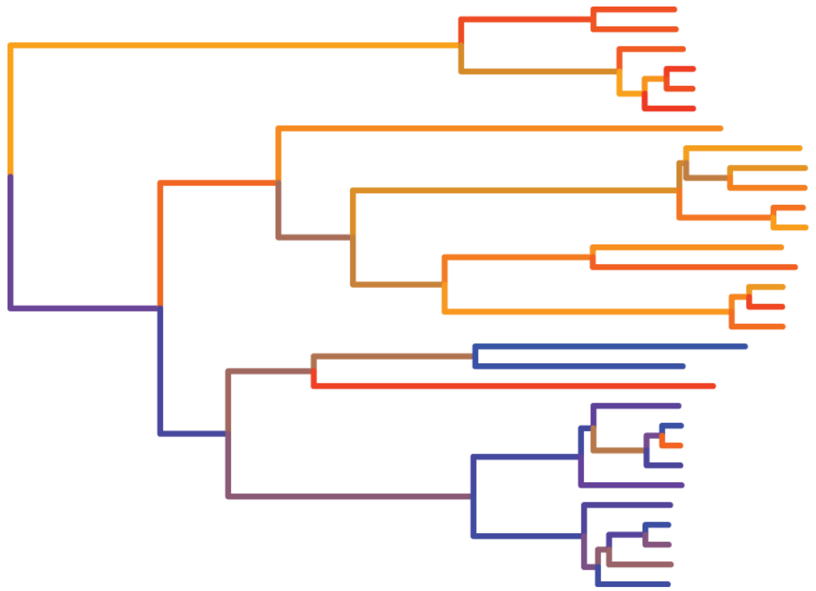
Bayesian relaxed clocks

- In the uncorrelated lognormal relaxed clock,
  two statistics can be obtained:

  1. **Coefficient of variation of rates**
     Measures the rate variation among branches
     A value of 0 indicates clocklike evolution
  2. **Covariance of rates**
     Measures autocorrelation of rates between adjacent branches

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



Ho, Duchêne, & Duchêne (2015) *Mol Ecol Resour*

Calibrating the Molecular Clock

Calibrating the molecular clock

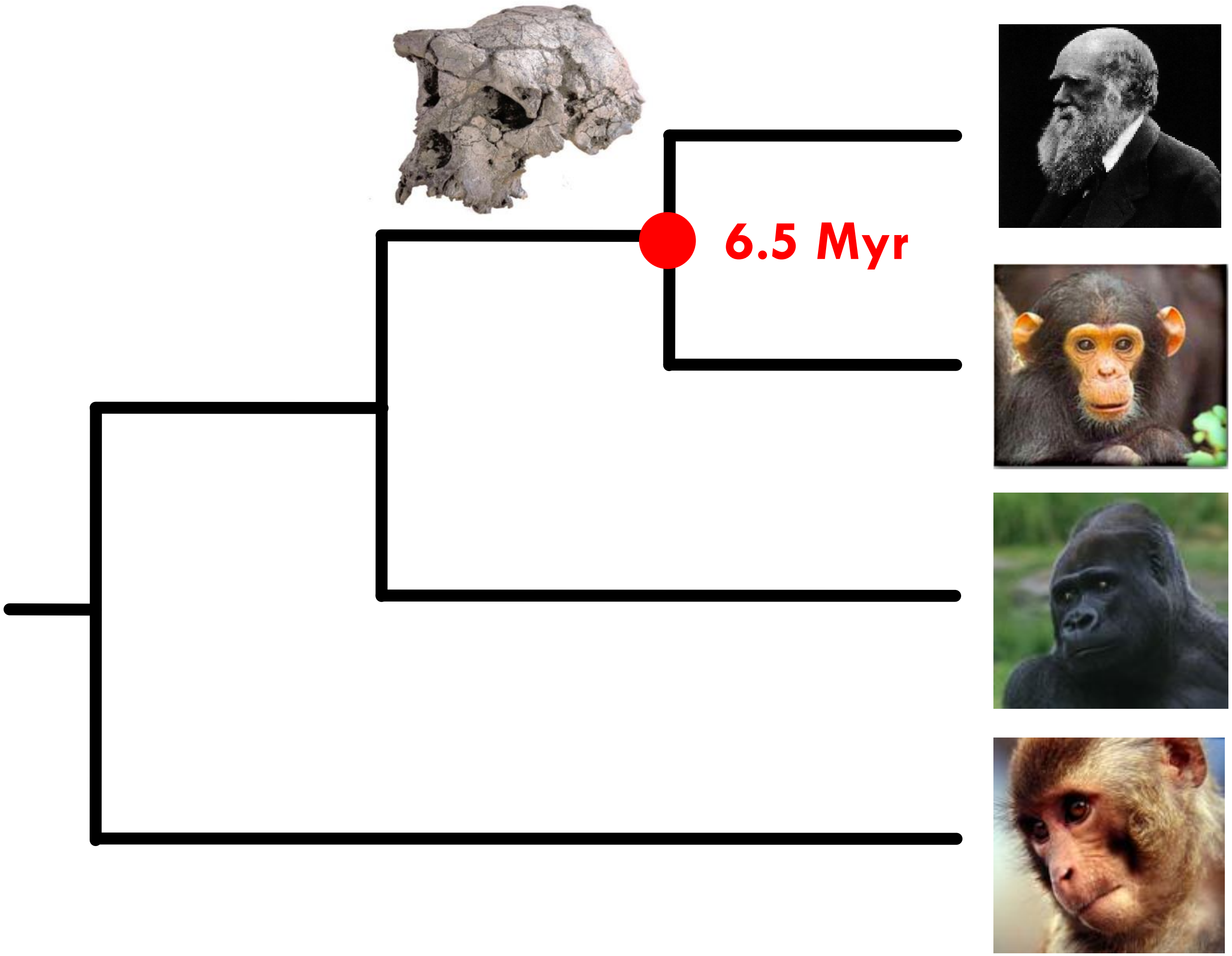
- Likelihood in dating depends on the product of rates and time
  - Branch lengths in substitutions per site
- But alone, rates and times are **non-identifiable**
- To separate rate and time, we need (prior) information about one or the other

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



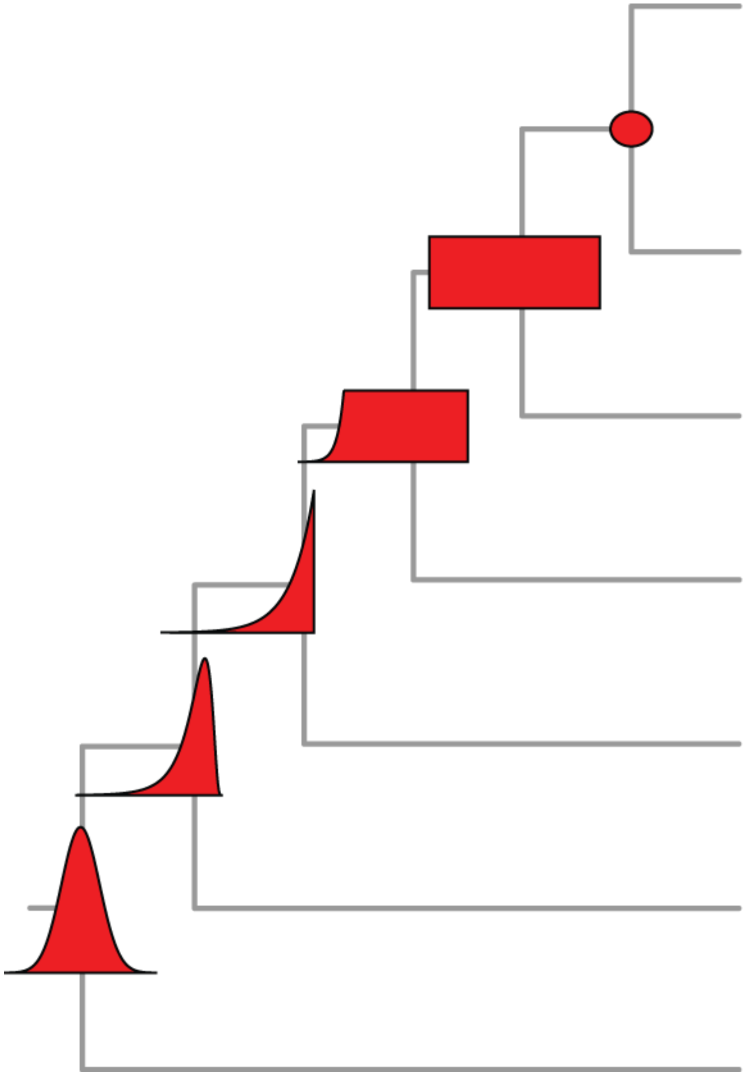
Choosing fossil calibrations

1. Museum numbers of specimen that demonstrate all the relevant characters and provenance data
2. Apomorphy-based diagnosis or phylogenetic analysis of the specimen
3. Explicit statements on the reconciliation of morphological and molecular data sets
4. Locality and stratigraphic level from which the calibrating fossil was collected
5. Reference to a published radioisotopic age and/or numeric timescale and details of numeric age selection

Parham *et al*. (2012) *Syst Biol*

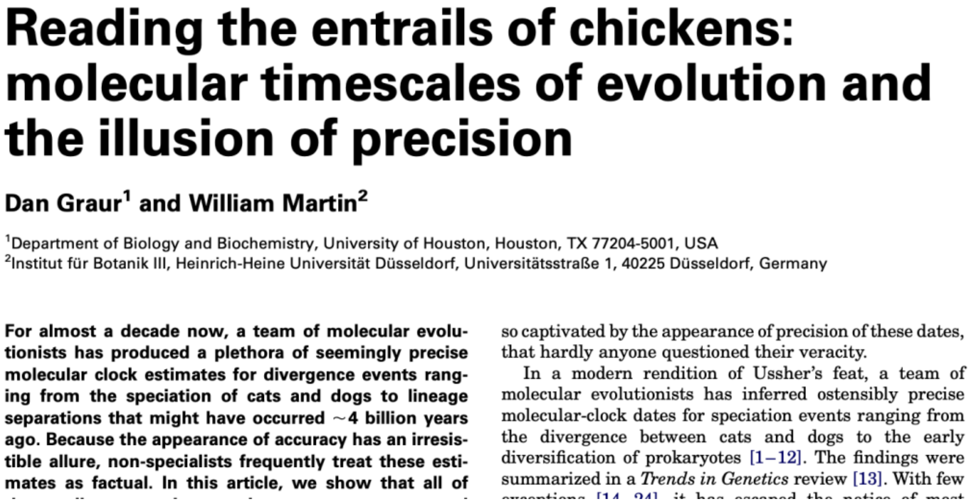
Calibration Priors

Calibrations



Point calibration

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



Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision

Dan Graur[1] and William Martin[2]

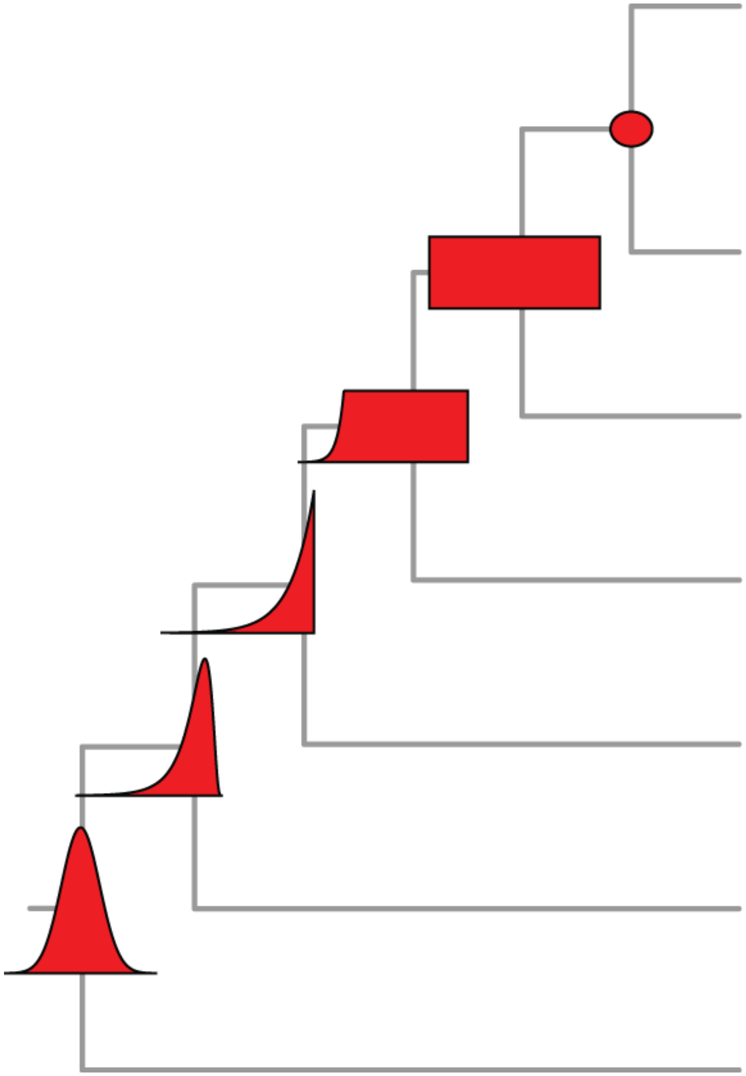
[1]Department of Biology and Biochemistry, University of Houston, Houston, TX 77204-5001, USA
[2]Institut für Botanik III, Heinrich-Heine Universität Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

For almost a decade now, a team of molecular evolutionists has produced a plethora of seemingly precise molecular clock estimates for divergence events ranging from the speciation of cats and dogs to lineage separations that might have occurred ~4 billion years ago. Because the appearance of accuracy has an irresistible allure, non-specialists frequently treat these estimates as factual. In this article, we show that all of

so captivated by the appearance of precision of these dates, that hardly anyone questioned their veracity.

In a modern rendition of Ussher's feat, a team of molecular evolutionists has inferred ostensibly precise molecular-clock dates for speciation events ranging from the divergence between cats and dogs to the early diversification of prokaryotes [1–12]. The findings were summarized in a *Trends in Genetics* review [13]. With few

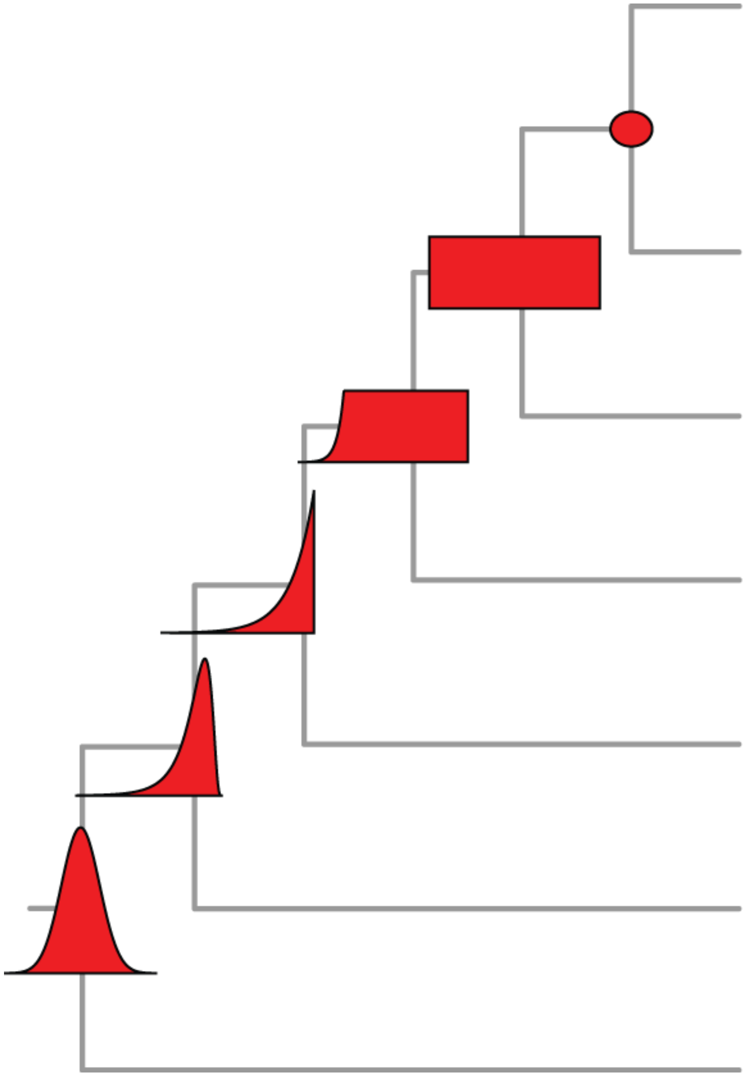
Calibrations



Uniform prior

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

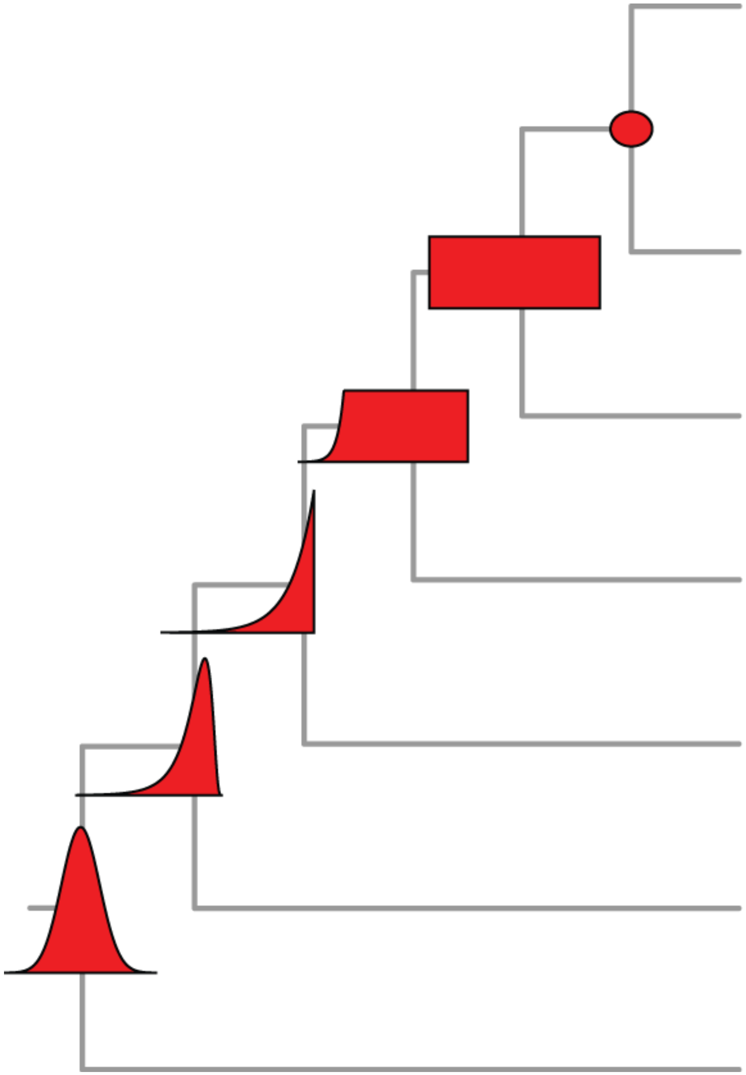
Calibrations

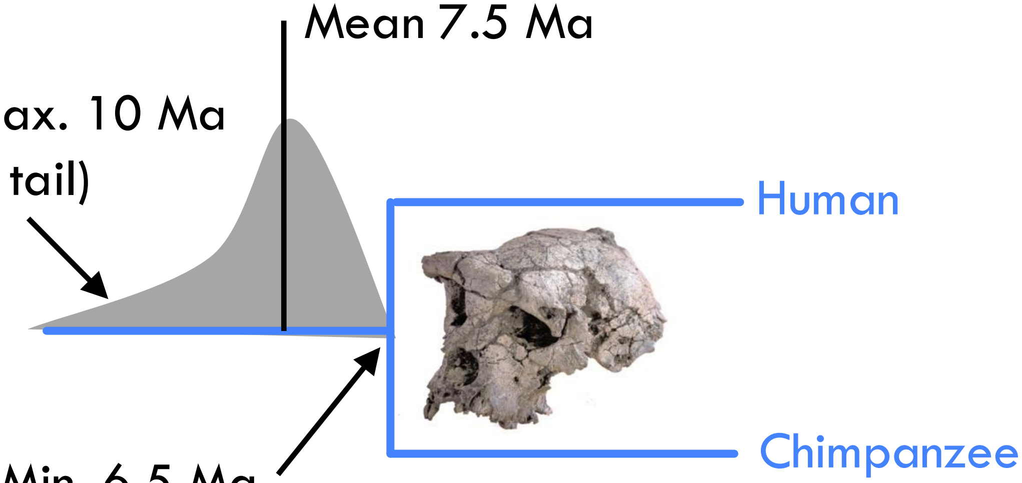


Exponential prior

- Need 2 values: minimum and mean
- Strong assumption about relationship of fossil taxon to internal node

Calibrations





Mean 7.5 Ma

Soft max. 10 Ma
(2.5% tail)

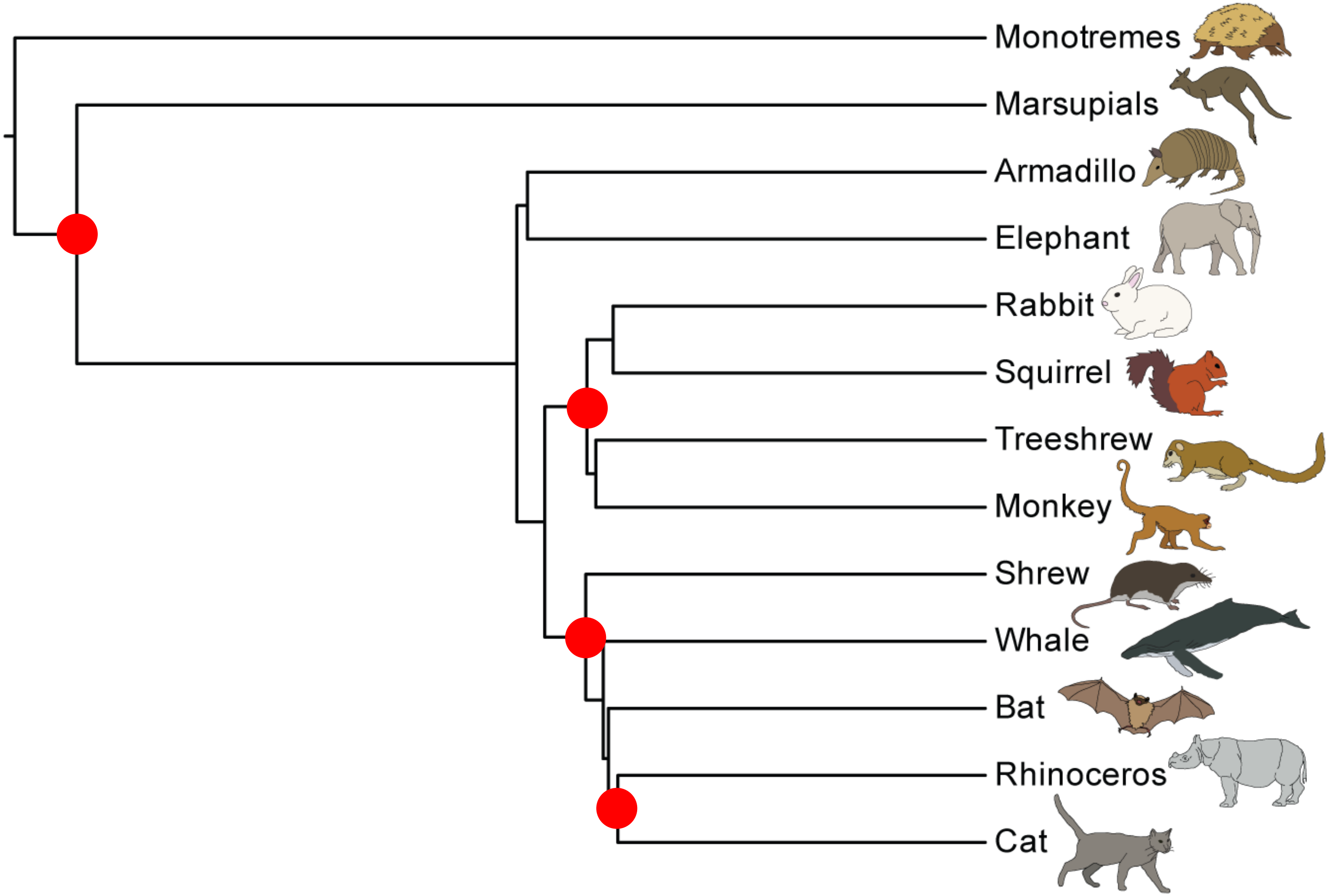
Min. 6.5 Ma
(age of fossil)

Lognormal prior

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

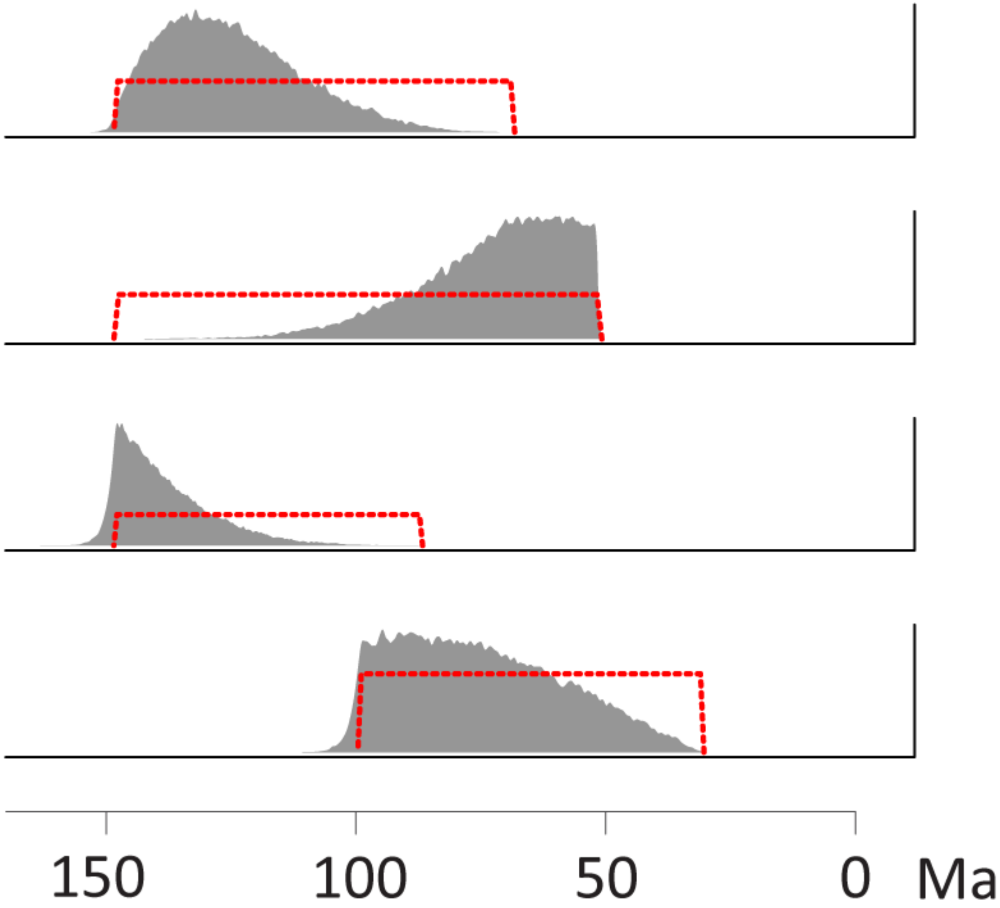
Multiple calibrations

- Use multiple calibrations if possible



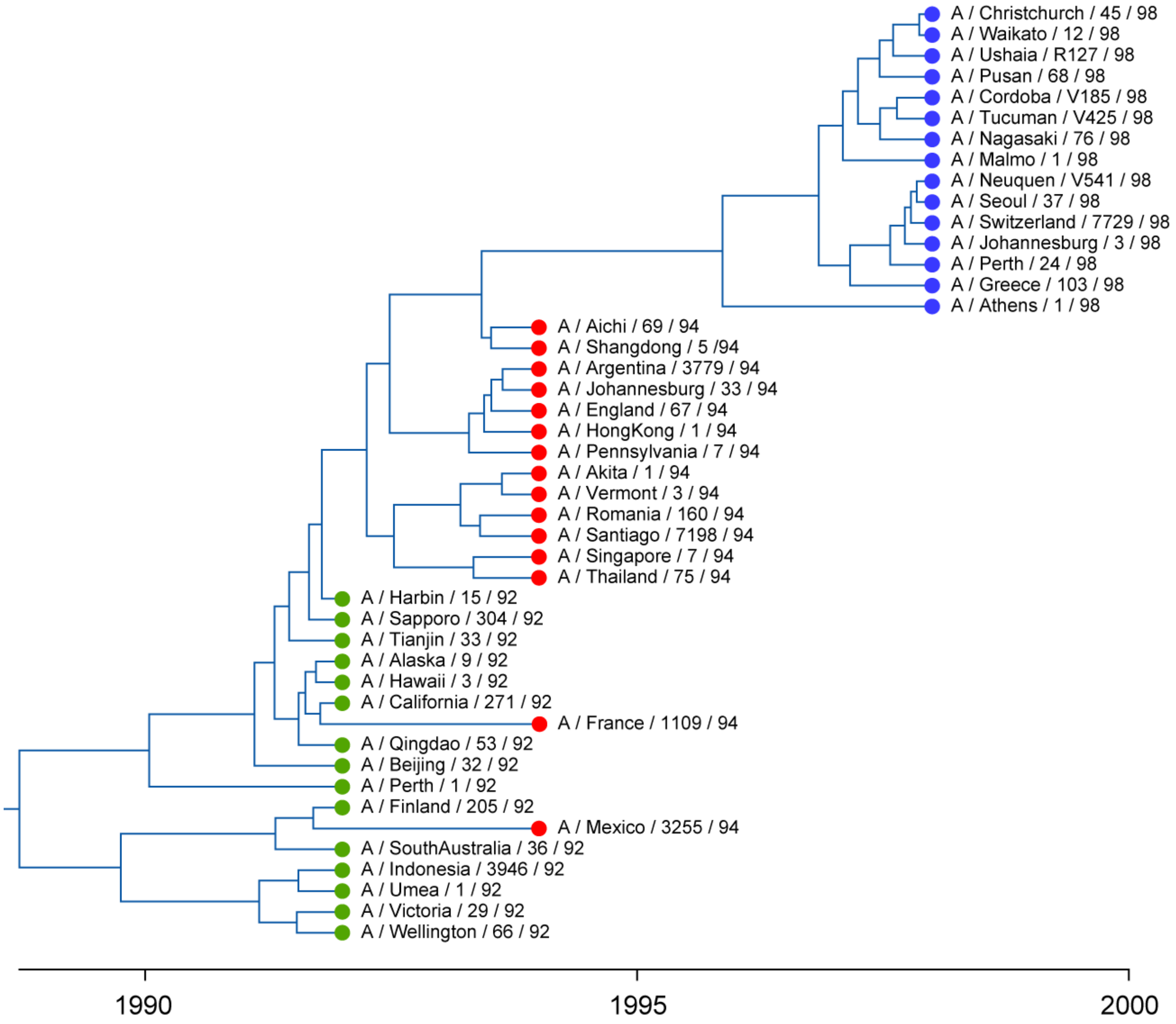
Multiple calibrations

- 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

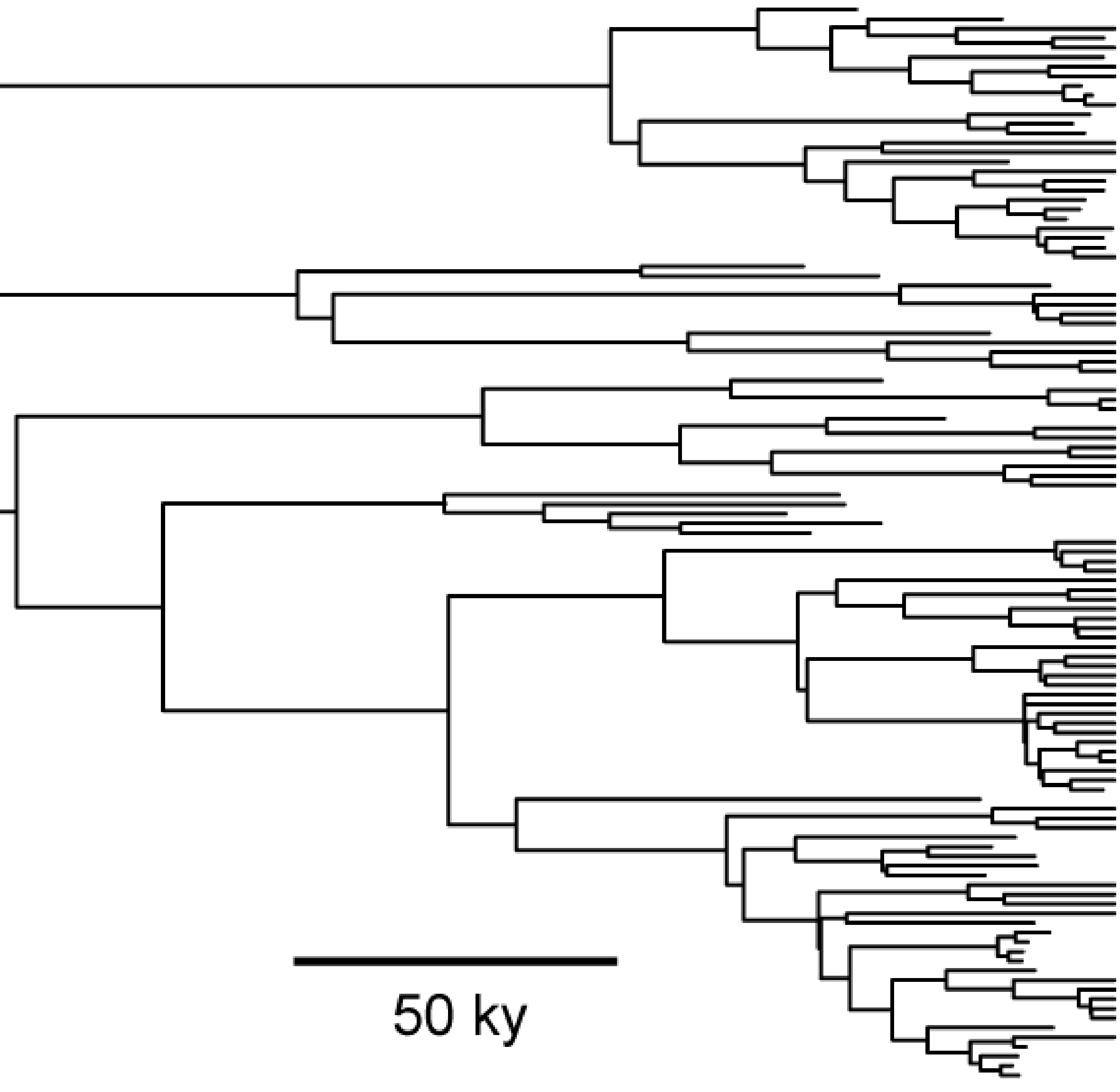


Warnock *et al.* (2014) *Proc Roy Soc B*

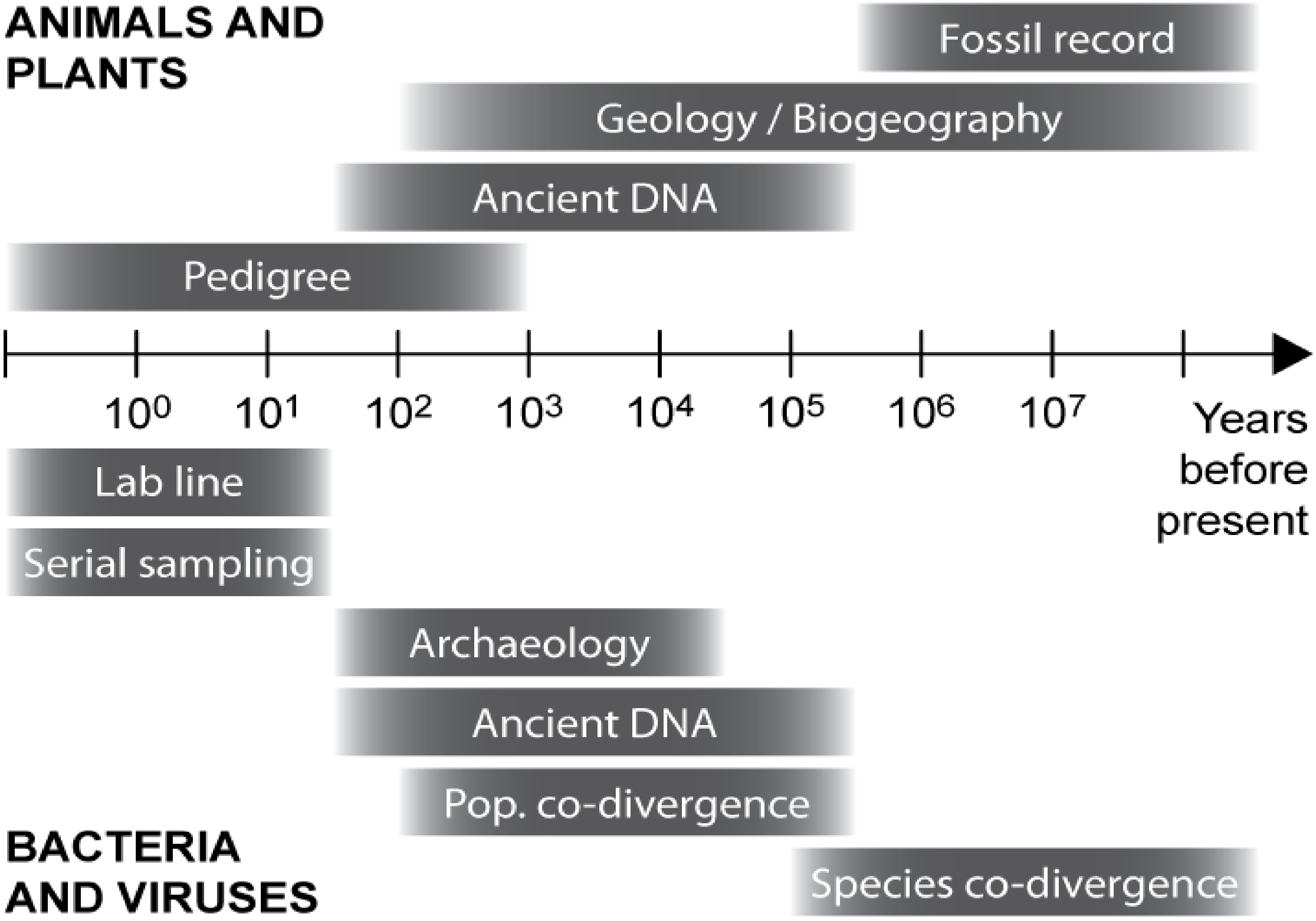
Calibration: Sampling times



Calibration: Sampling times



Calibrations



Choosing calibrations

- Use multiple calibrations if possible
- The age estimates for poorly supported clades should be interpreted carefully
- Careful selection of clock models can improve the estimates

Useful references

- **Calibration uncertainty in molecular dating analyses: there is no substitution for the prior evaluation of time priors**
  Warnock *et al*. (2014) *Proceedings of the Royal Society B*, 282: 20141013.
- **Time-dependent rates of molecular evolution**
  Ho *et al*. (2011) *Molecular Ecology*, 20: 3087–3101.
- **Accounting for uncertainty in phylogenetic estimation of evolutonary divergence times**
  Ho & Phillips (2009) *Systematic Biology*, 58: 367–380.
- **Best practices for justifying fossil calibrations**
  Parham *et al*. (2012) *Systematic Biology*, 61: 346–359.
- **Biogeographic calibrations for the molecular clock**
  Ho *et al*. (2015) *Biology Letters*, 11: 20150194.

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.