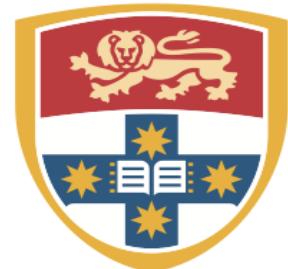
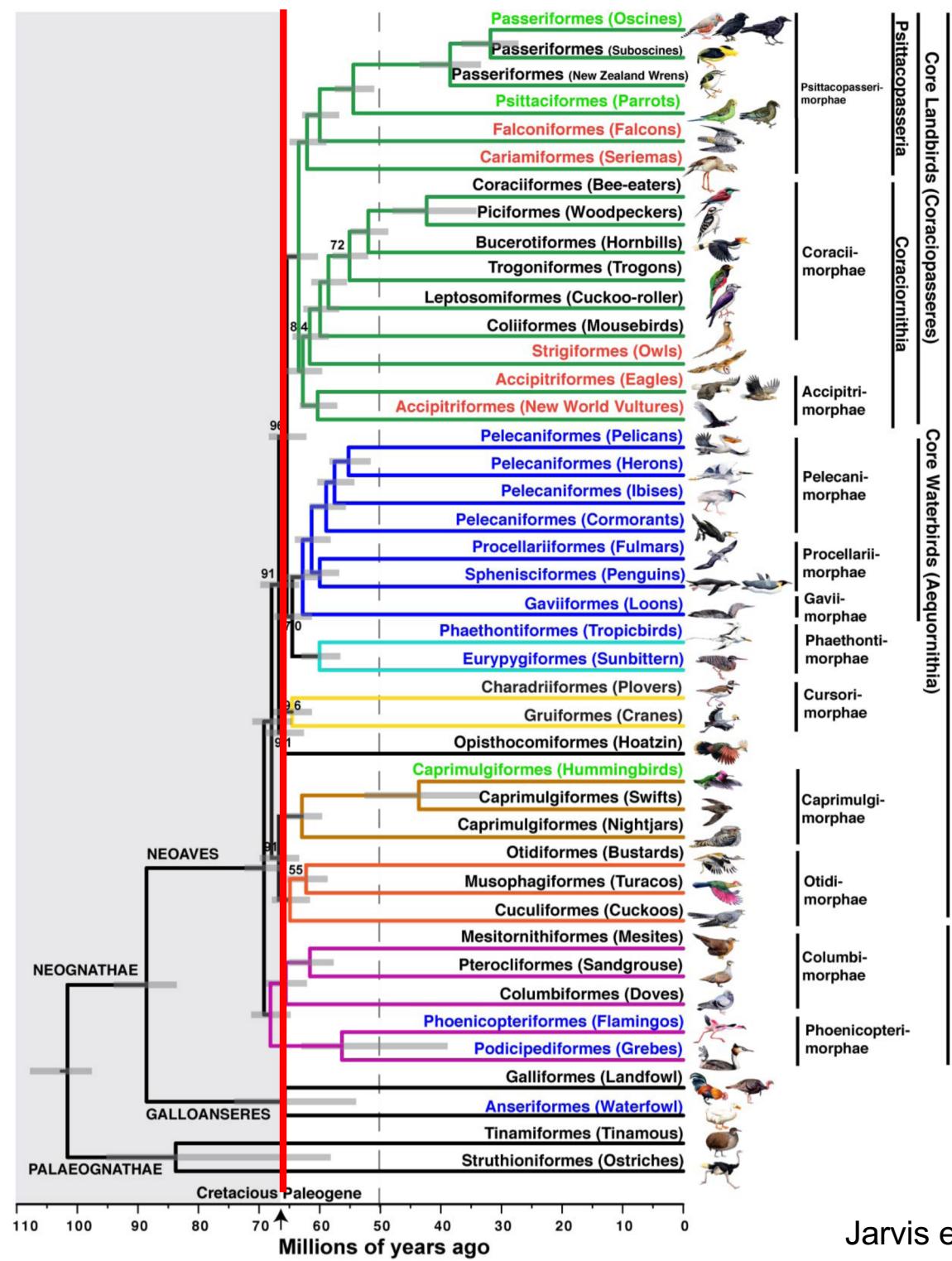


# Molecular Dating

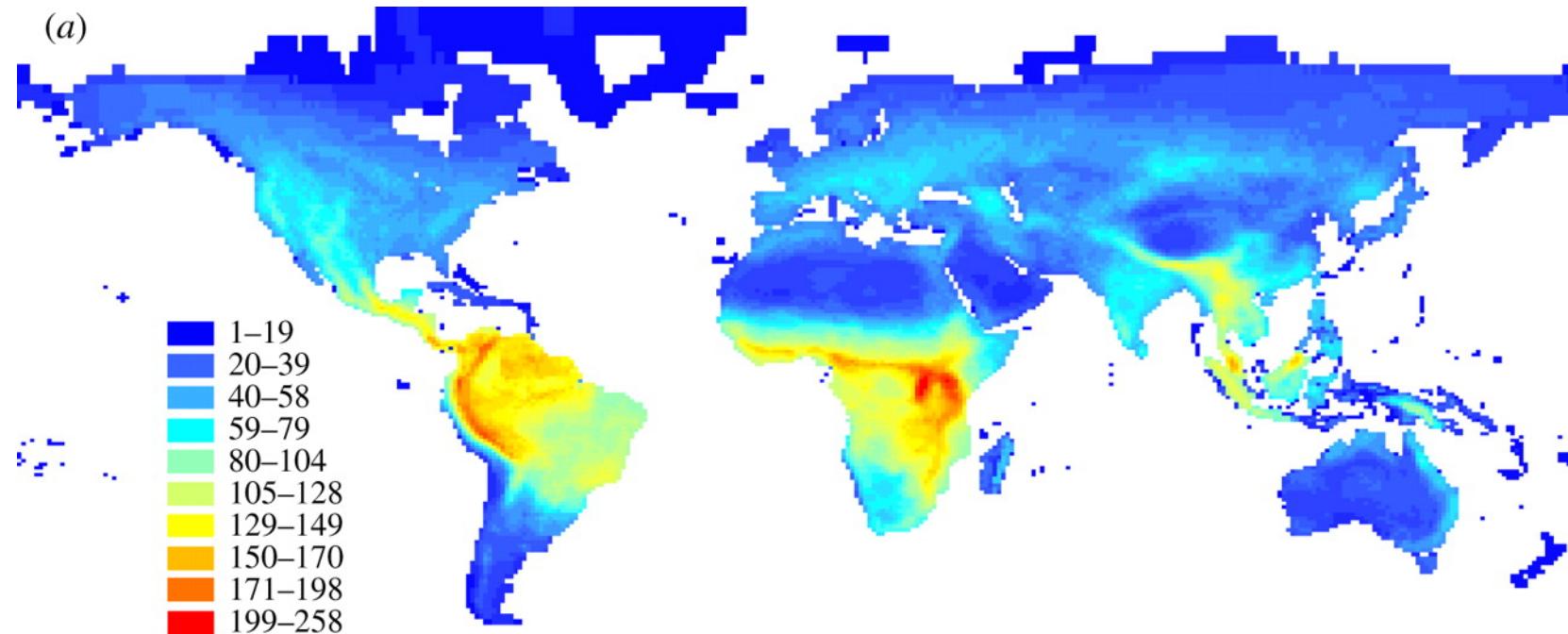
Simon Ho



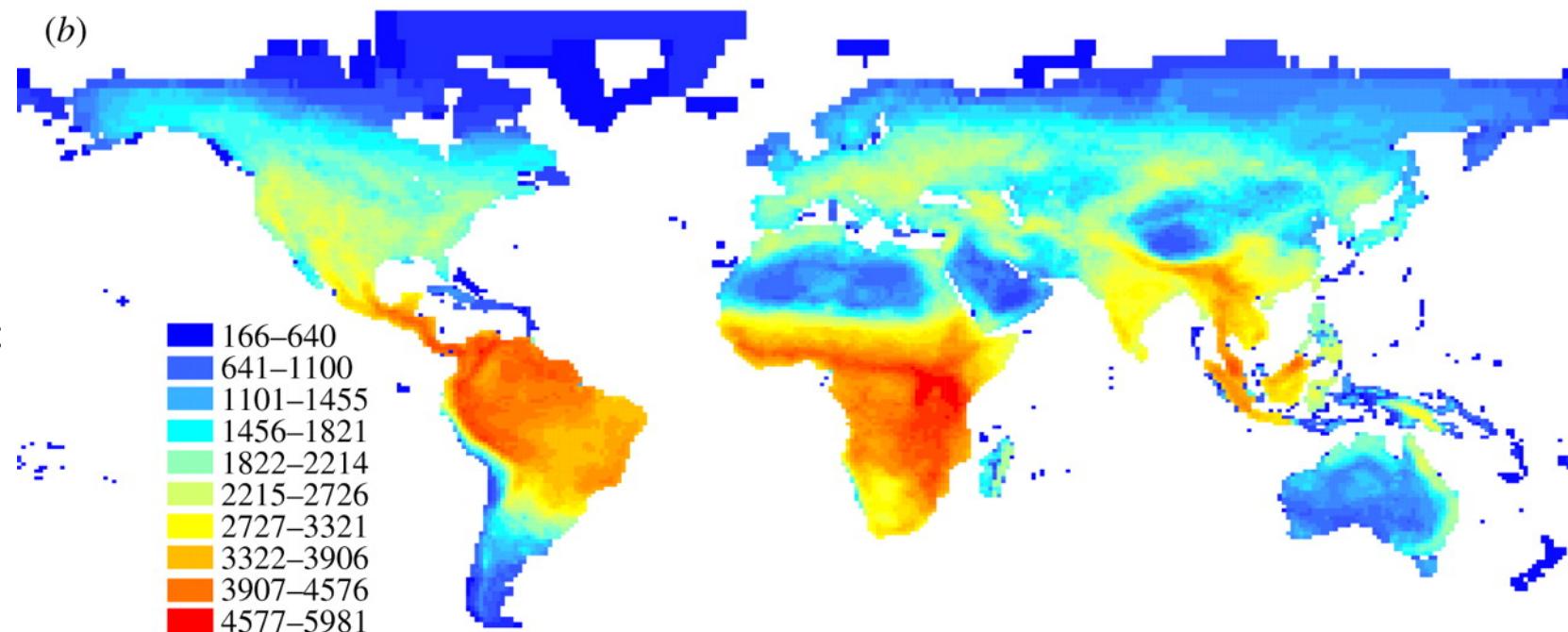
THE UNIVERSITY OF  
**SYDNEY**

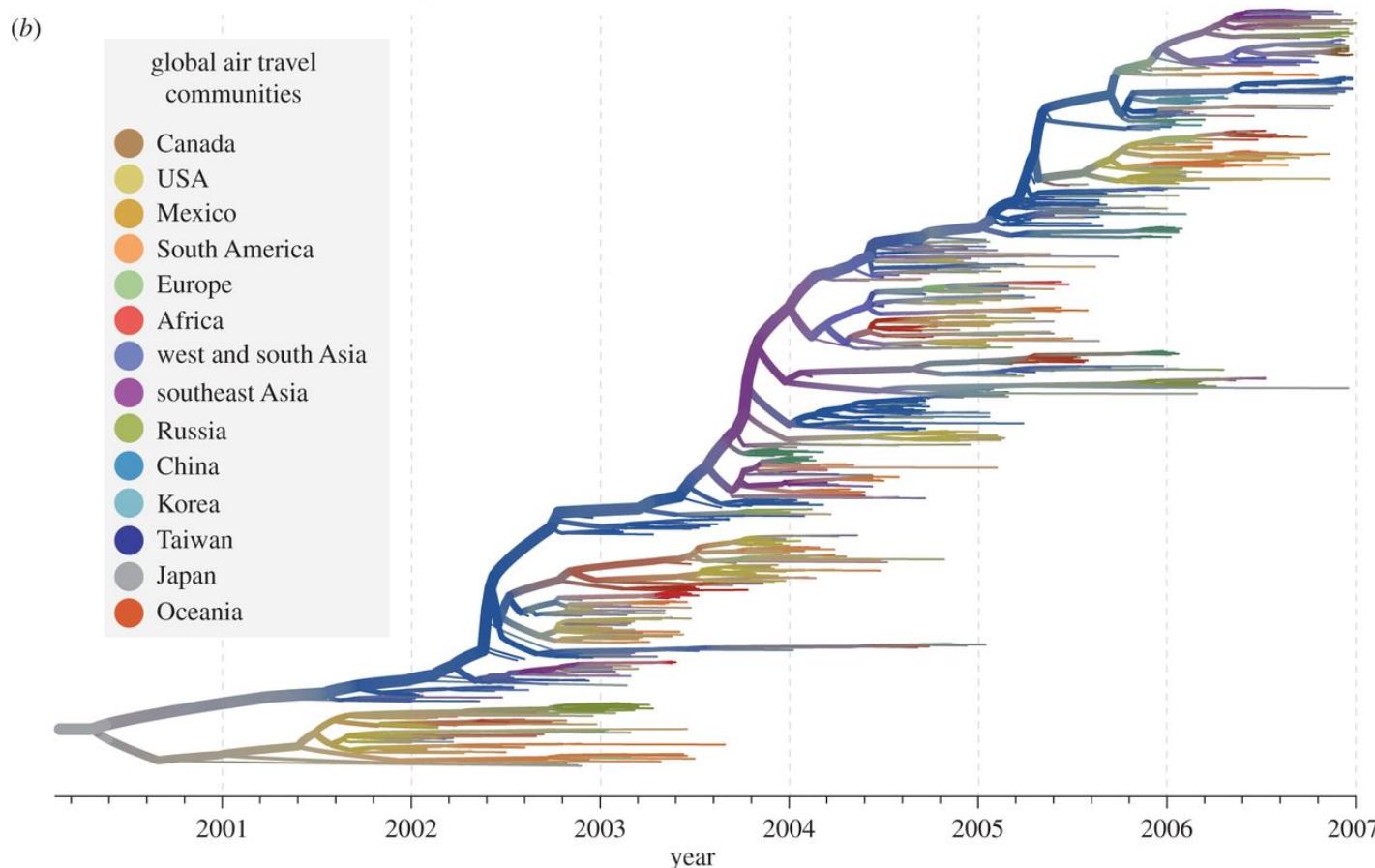
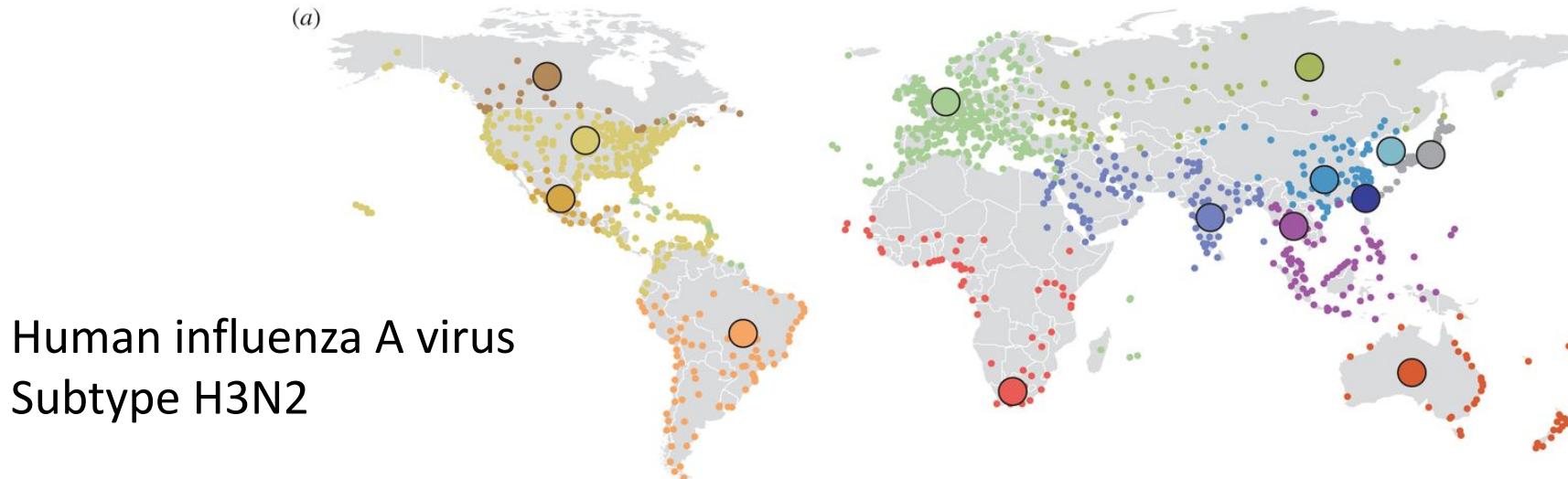


Mammal  
species  
richness



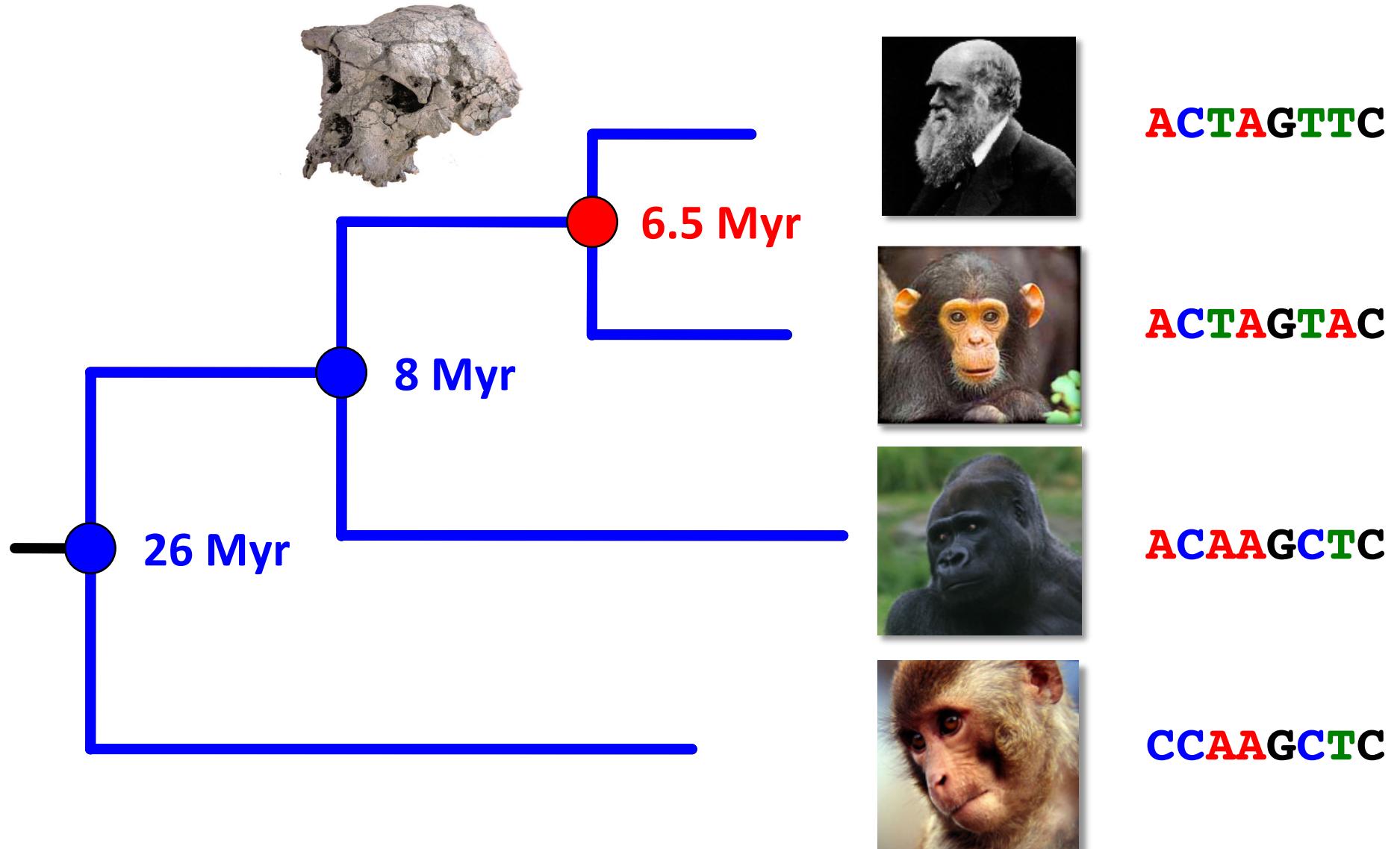
Mammal  
phylogenetic  
diversity



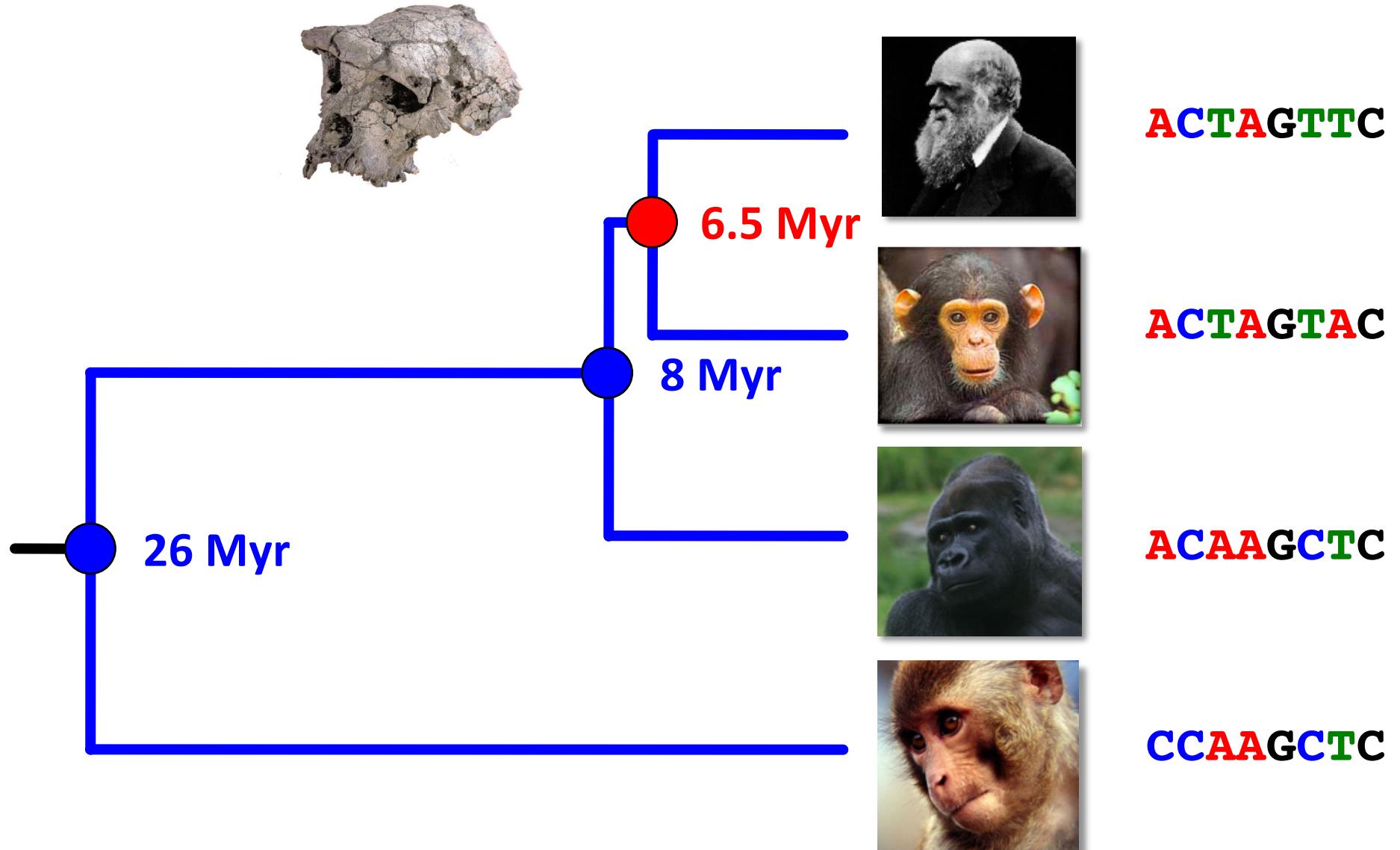


# The Molecular Clock

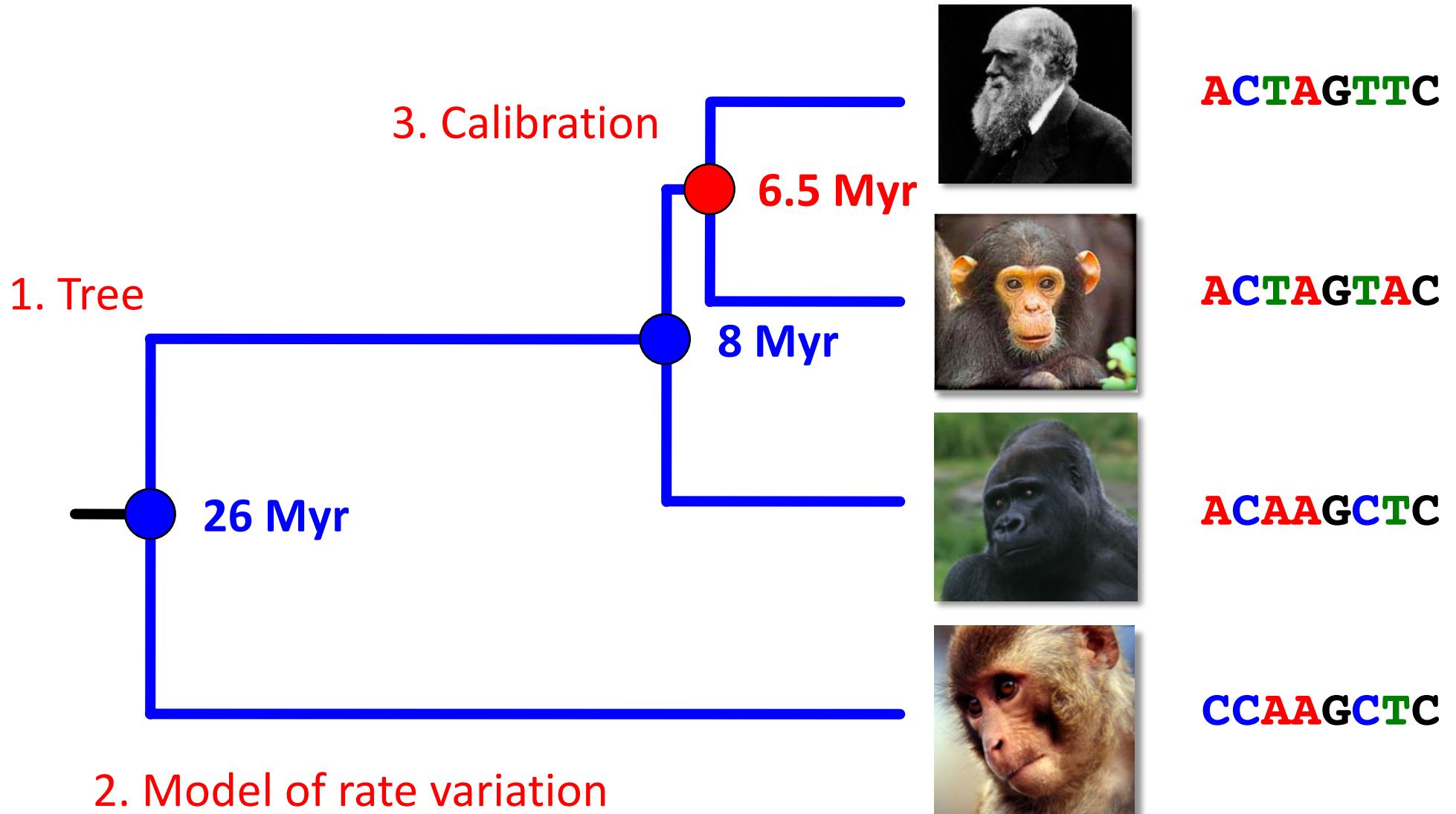
# The molecular clock



# The molecular clock



# Sources of error



# A brief history

---

- Emile Zuckerkandl and Linus Pauling

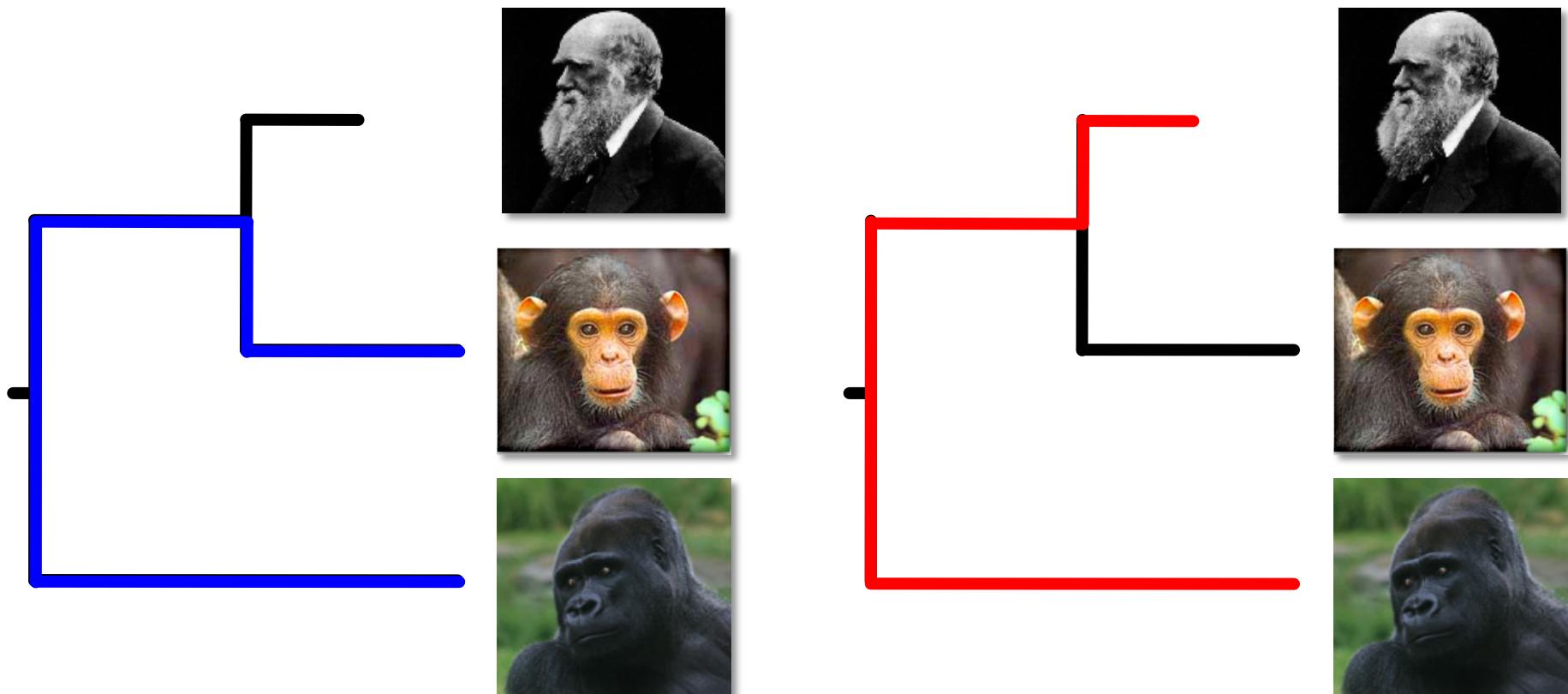
**1962**      Assumed constant rate among species to estimate timing of globin gene duplications

**1965**      Introduced the term ‘molecular evolutionary clock’



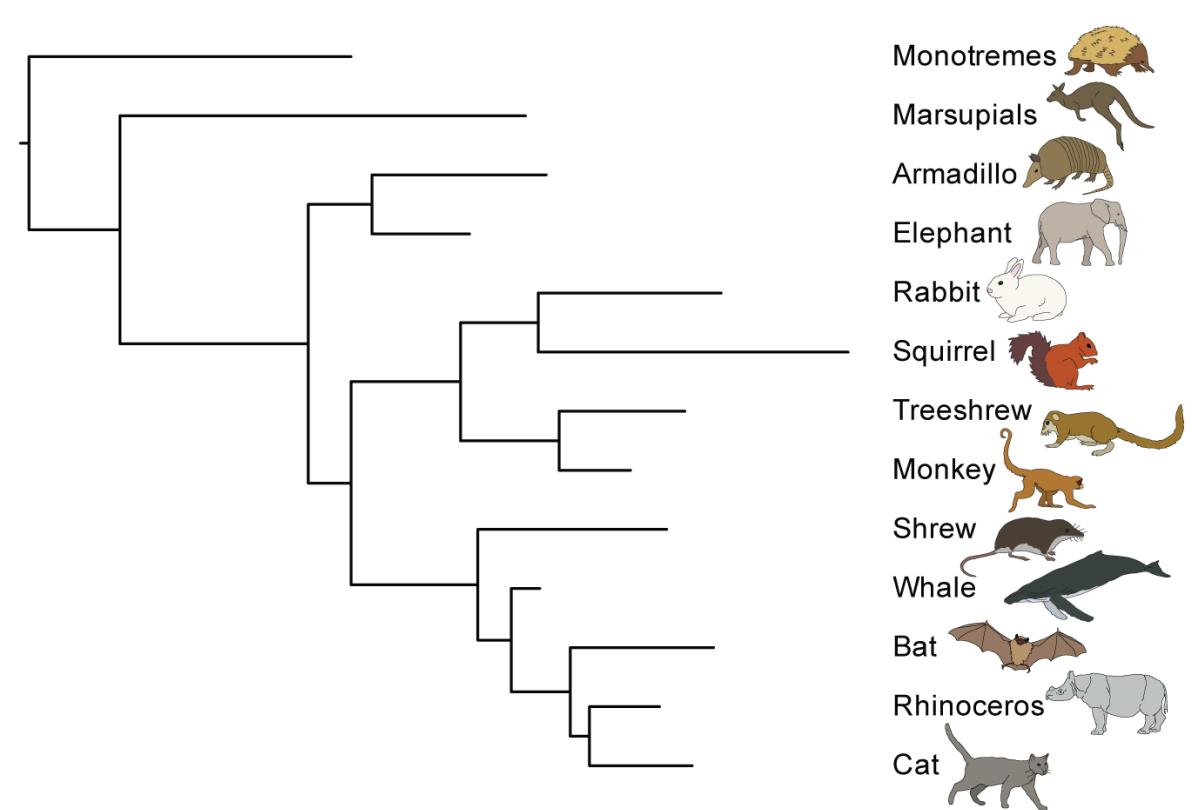
# Testing for clocklike evolution

- Relative-rates test (Fitch, 1976)



# 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

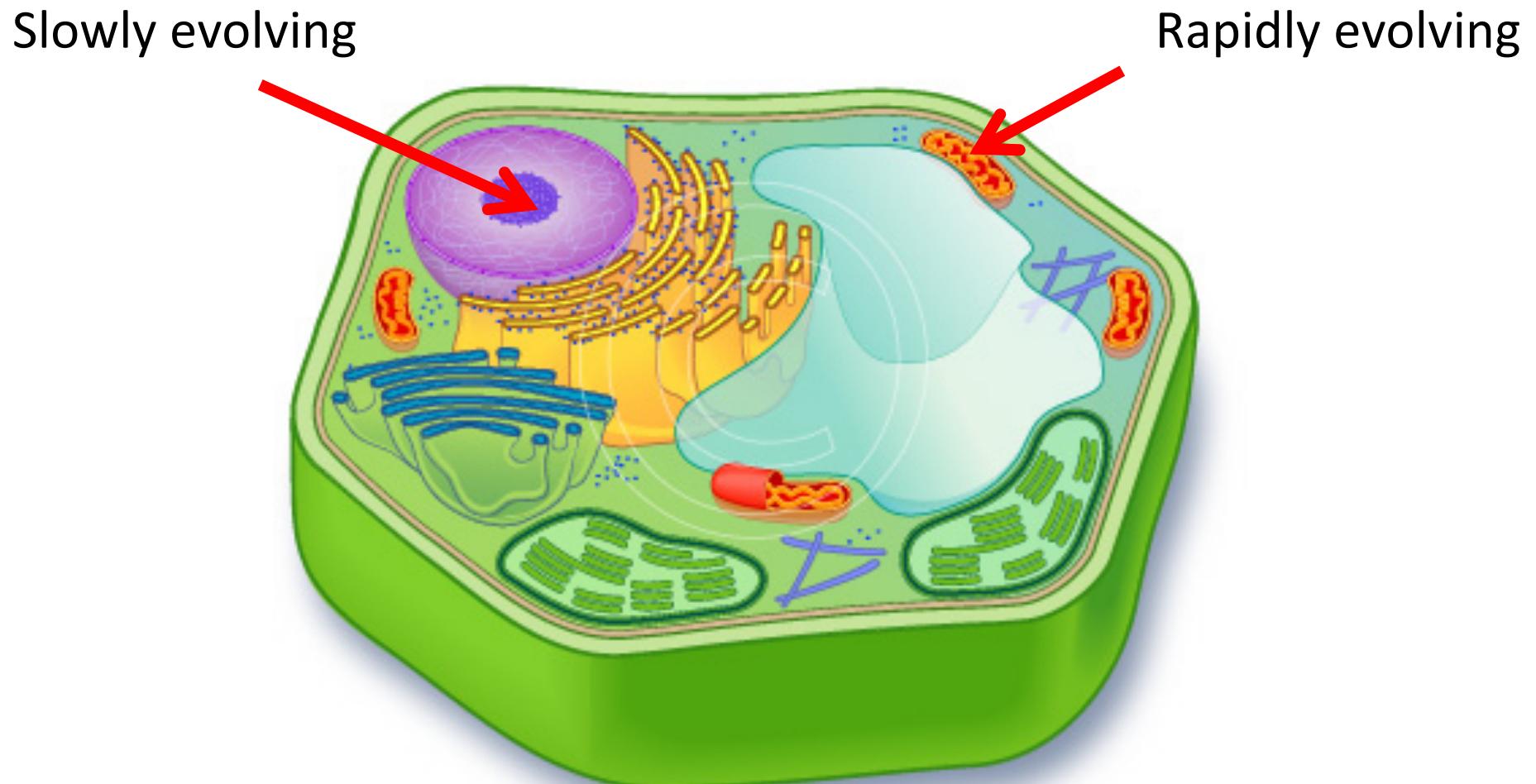


The assumption of rate constancy can be relaxed

# Molecular-Clock Models

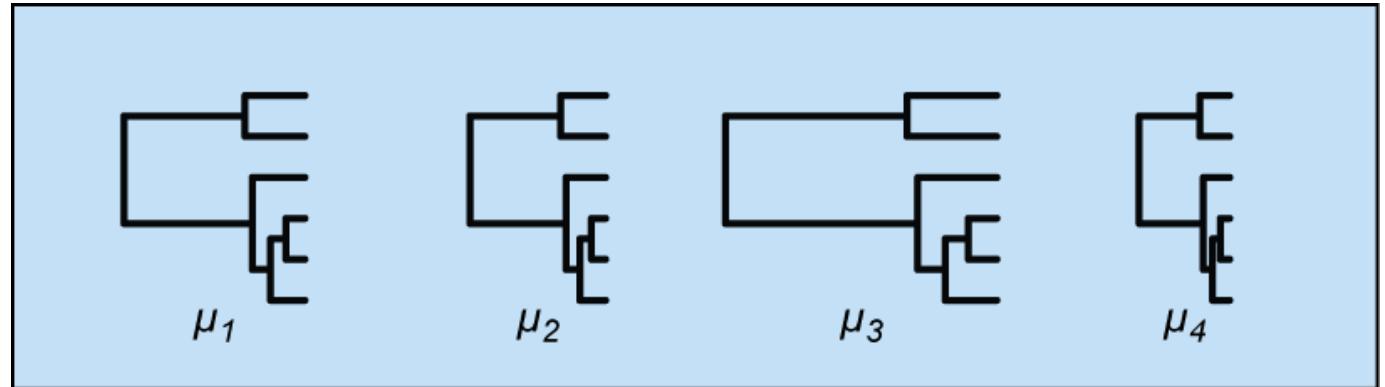
# Rate variation across loci

---

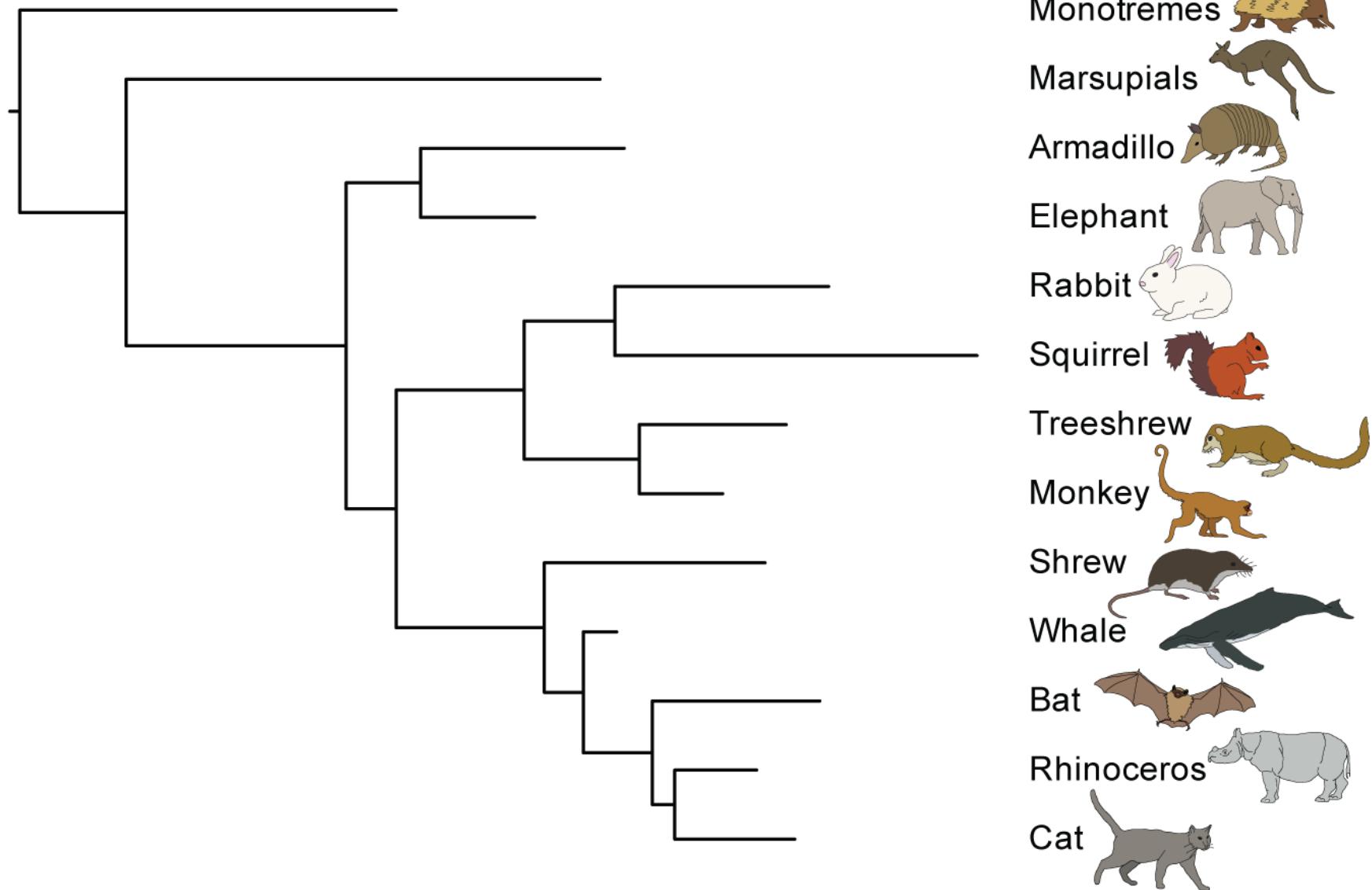


# Rate variation across loci

- Across loci  
**(gene effects)**

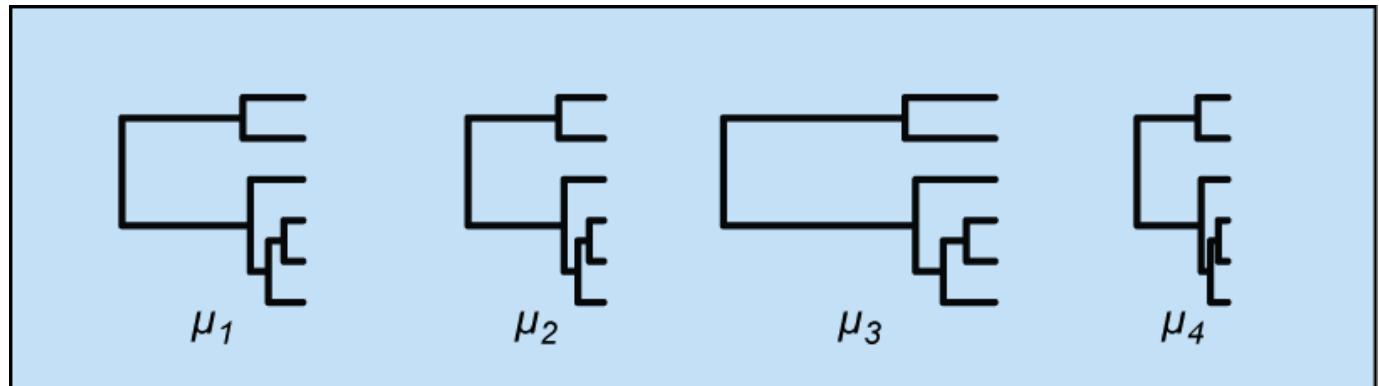


# Rate variation across lineages

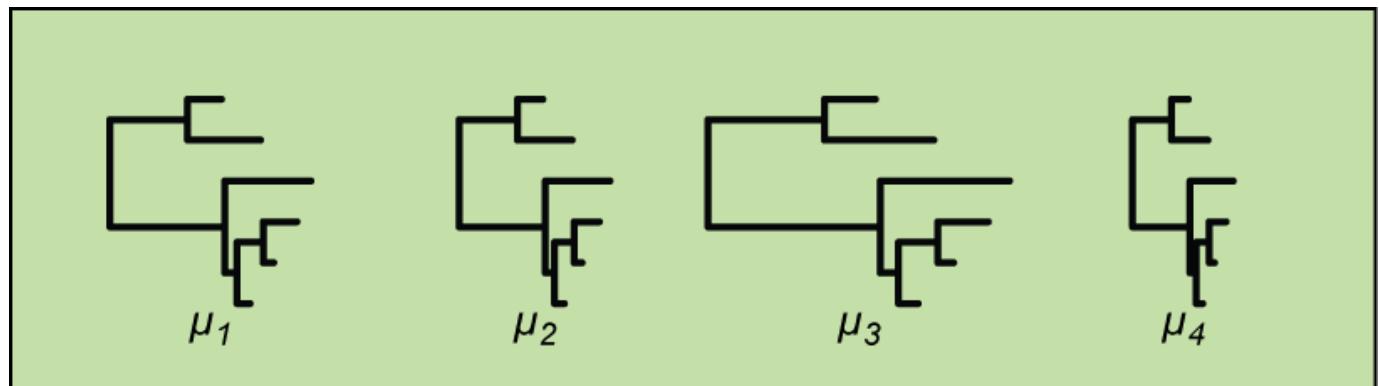


# Rate variation across lineages

- Across loci  
**(gene effects)**

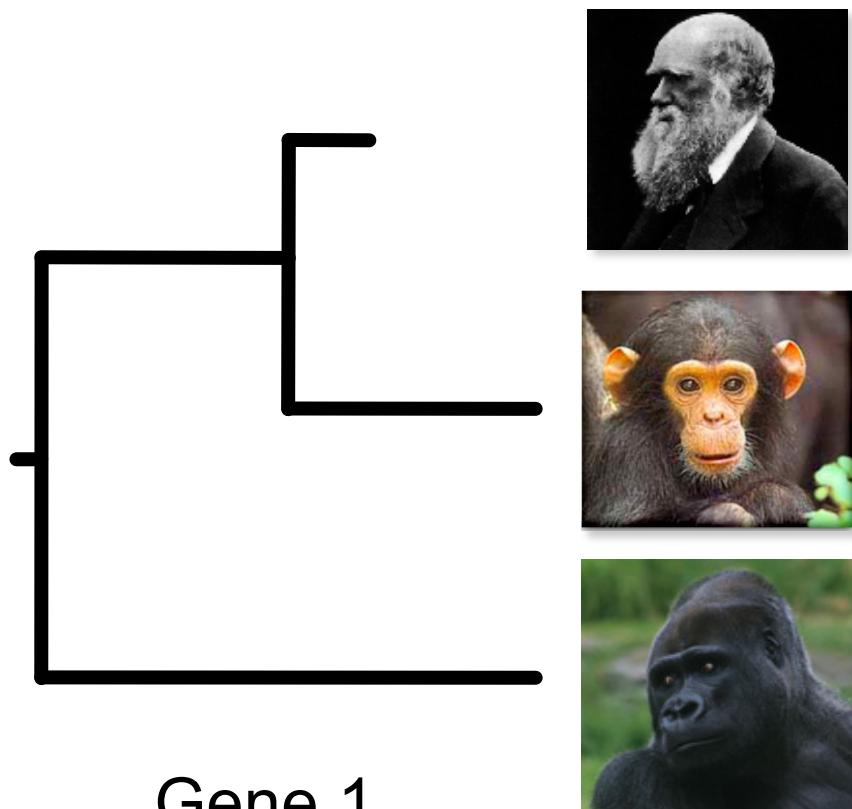


- Across lineages  
**(lineage effects)**

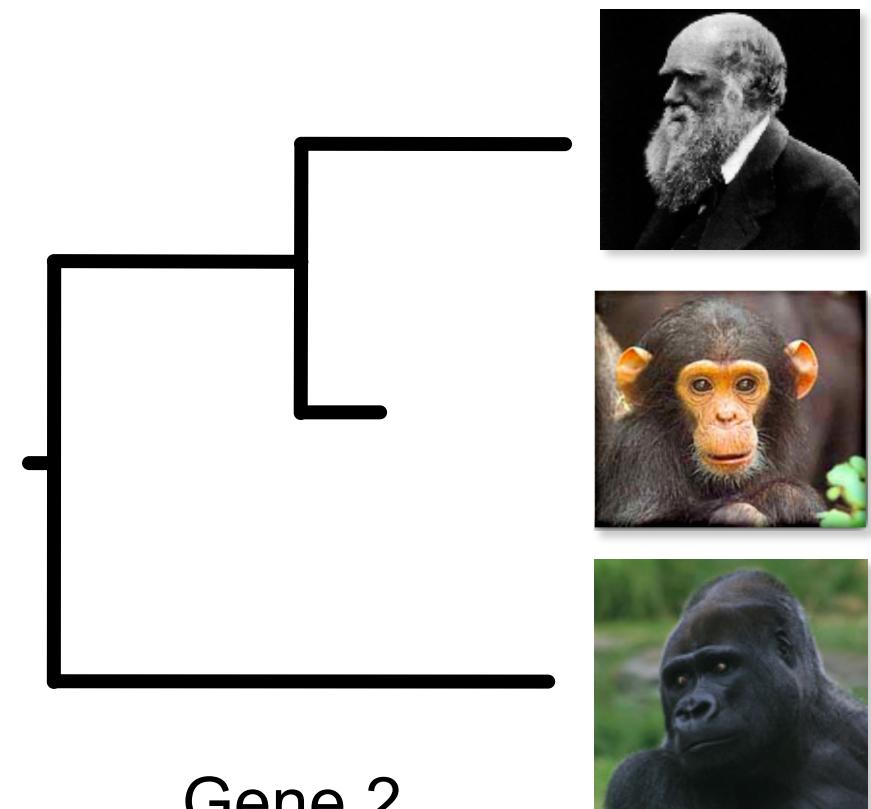


# Gene-by-lineage interactions

---



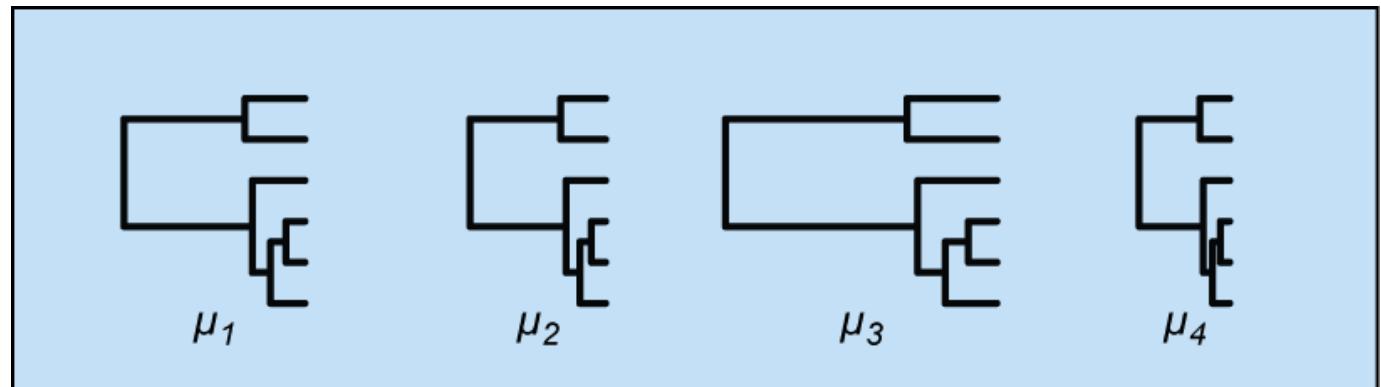
Gene 1



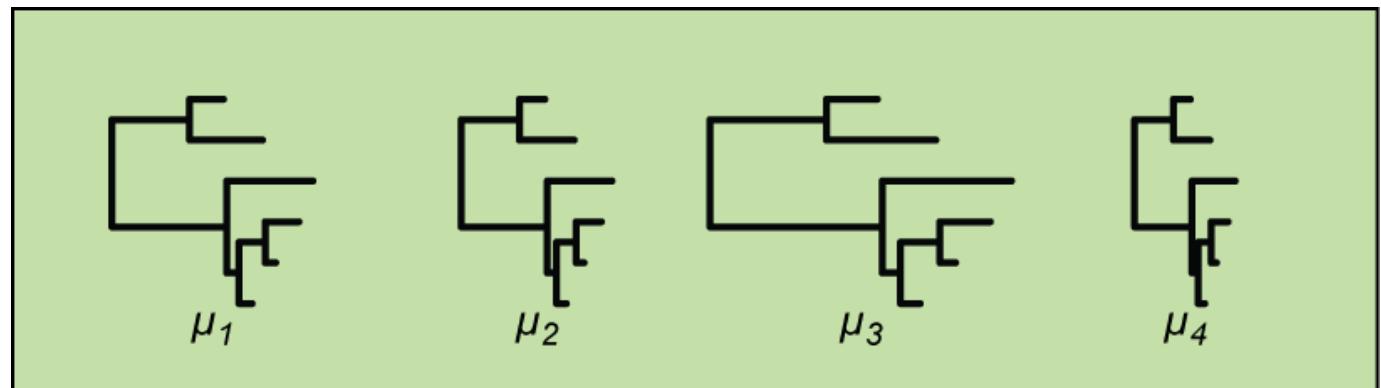
Gene 2

# Evolutionary rate variation

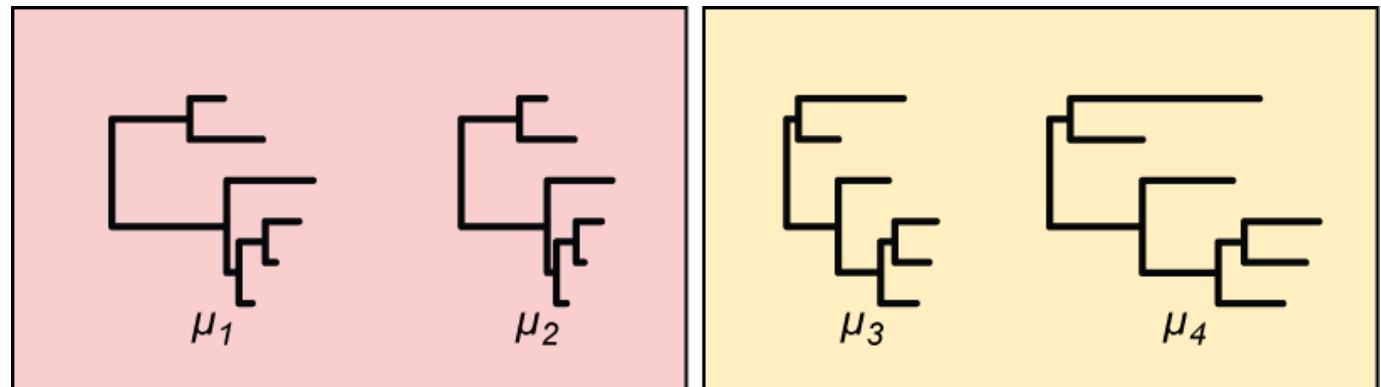
- Across loci  
**(gene effects)**



- Across lineages  
**(lineage effects)**



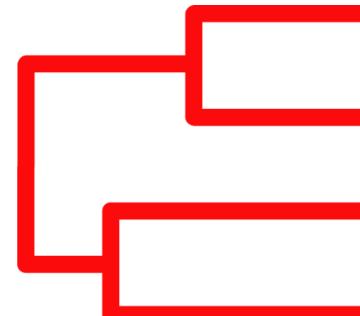
- Gene-by-lineage interaction  
**(residual effects)**



# Molecular-clock models

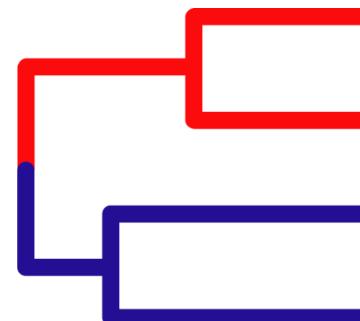
---

Strict or ‘global’ molecular clock



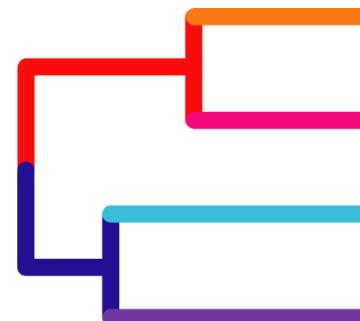
$$k = 1$$

Multi-rate clocks



$$1 < k < n$$

Relaxed clocks



$$k = n$$

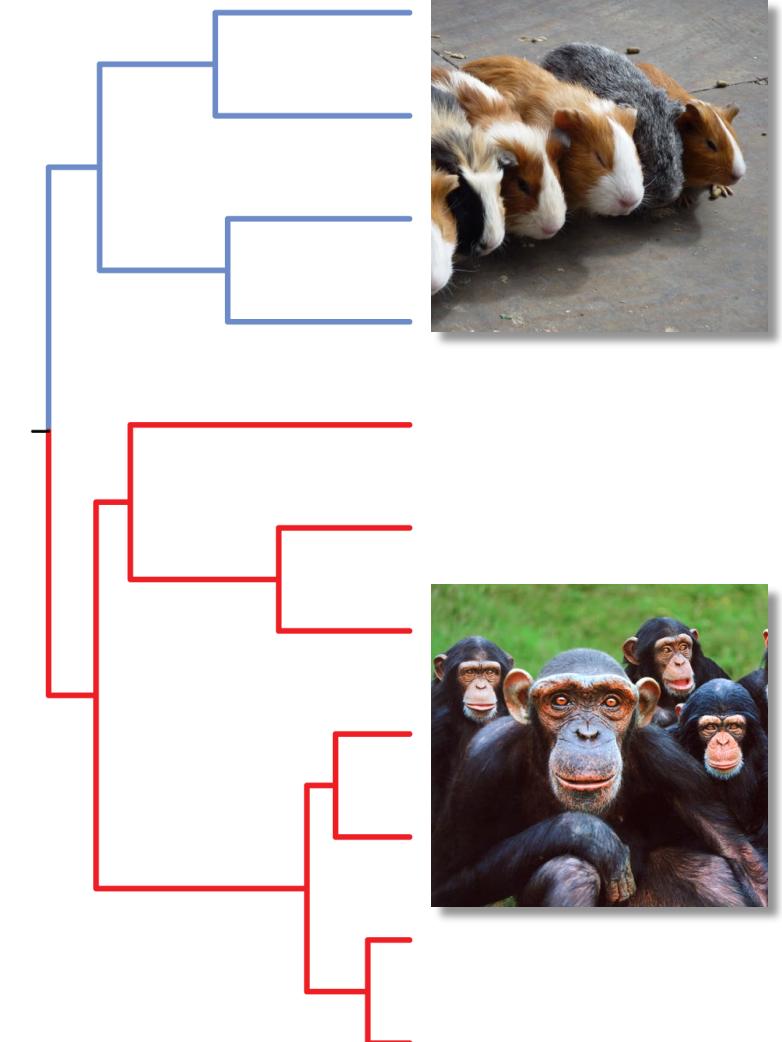
# Multi-rate clocks

---

- 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

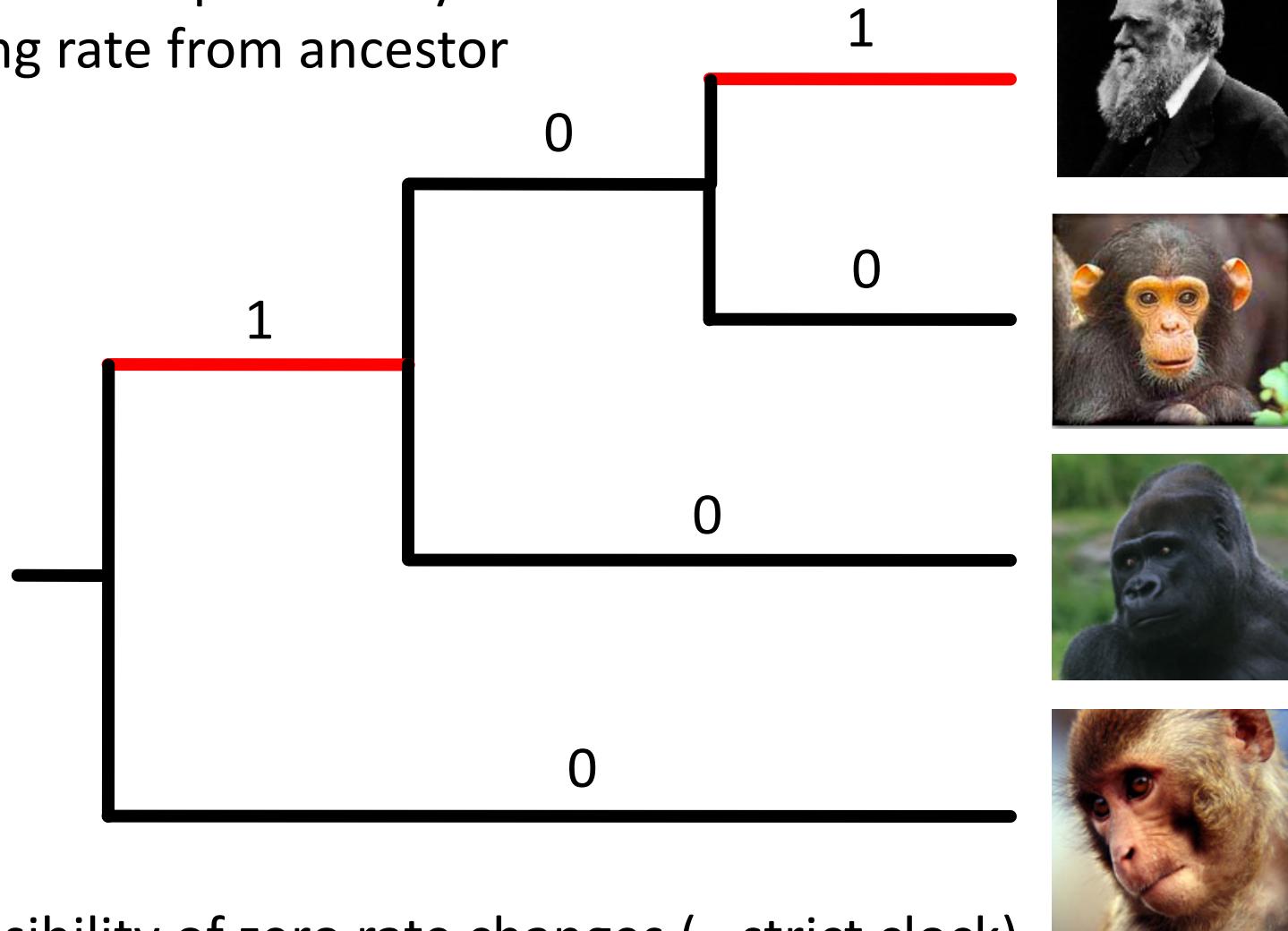
# Local clocks

- **User-defined local clock**
  - Fixed tree topology
- **Random local clock**
  - Each branch has a probability of inheriting rate from ancestor
  - Tree estimated



# Random local clock

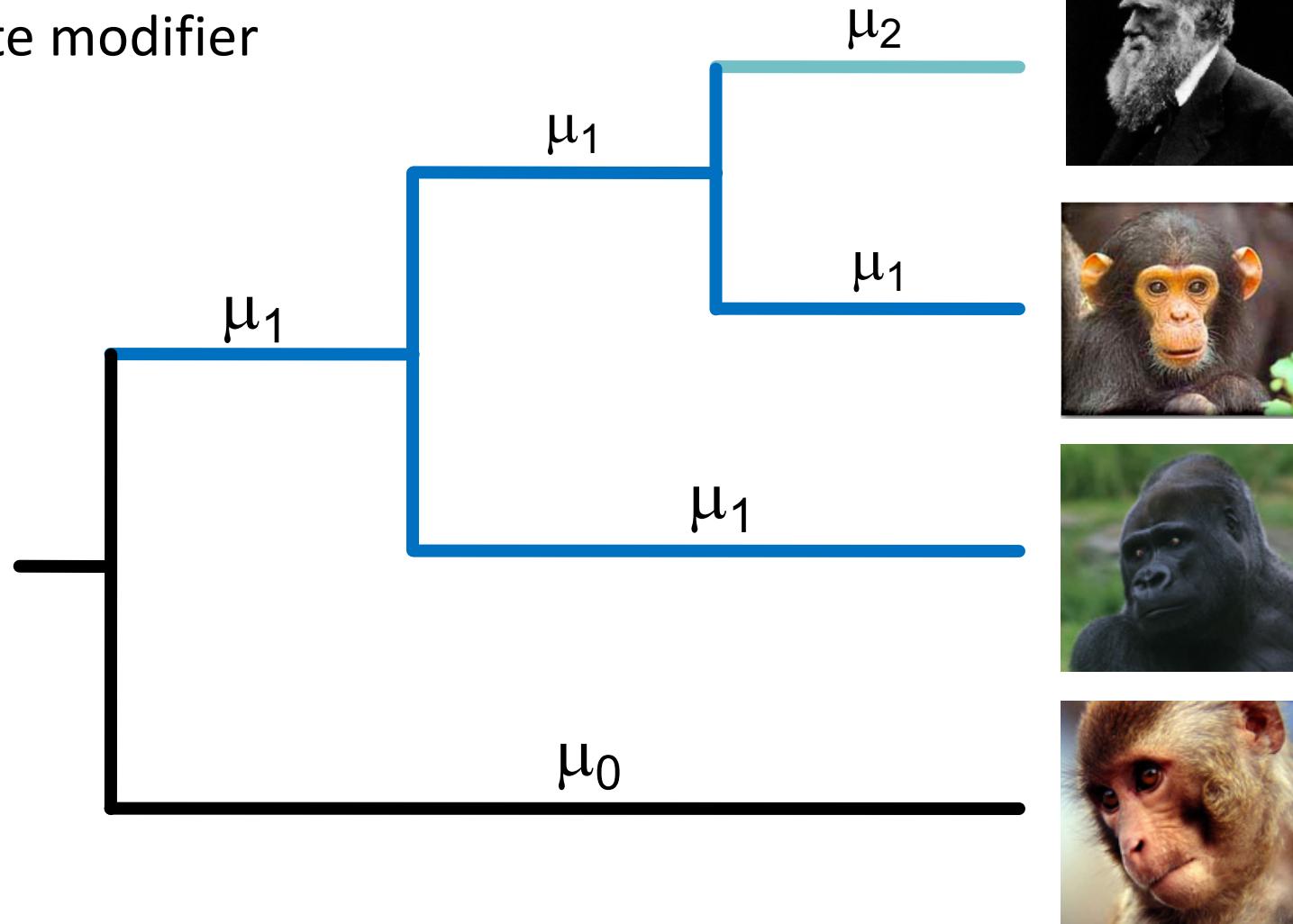
Each branch has a probability  
of inheriting rate from ancestor



Includes possibility of zero rate changes (= strict clock)

# Random local clock

Otherwise multiplied by a relative rate modifier



# Relaxed clocks

---

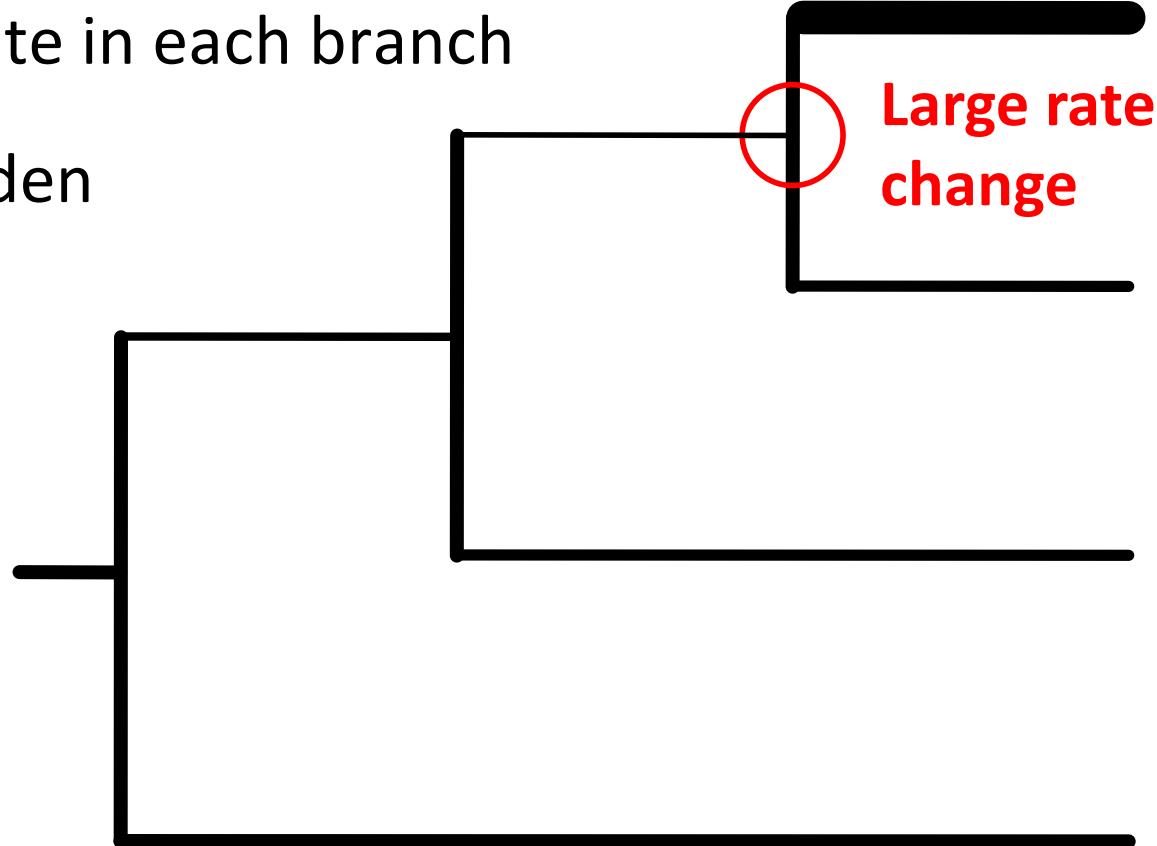
- We know that life-history characteristics:
  - Have effects on rates of molecular evolution
  - Are usually heritable to some degree
- Treat molecular rate as a heritable trait
- Early implementations of relaxed clocks assumed that closely related species share similar rates



The Persistence of Memory, ©Salvador Dalí

# Likelihood-based relaxed clocks

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



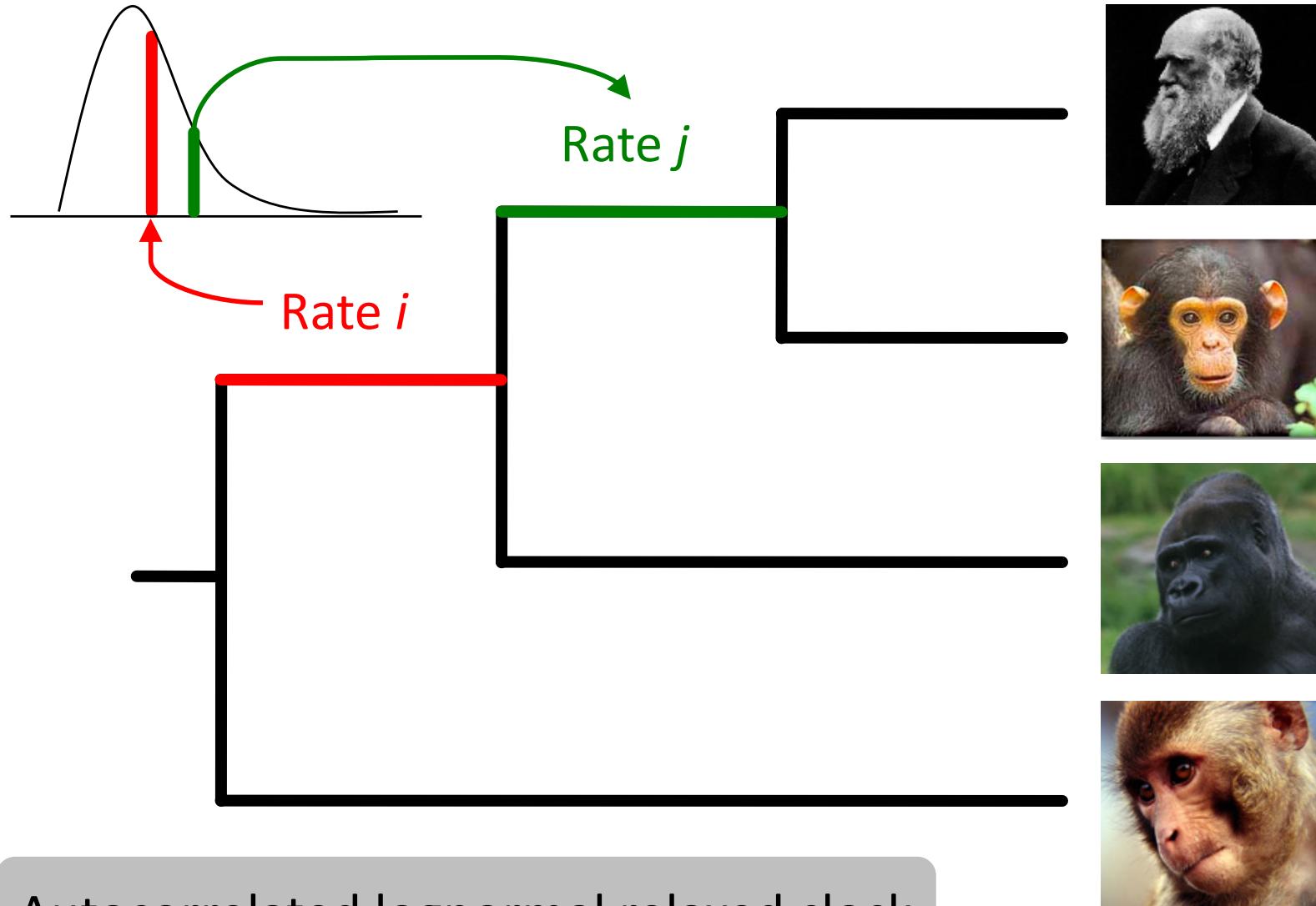
$$\text{Penalised likelihood} = \ln L - \lambda \Phi(r)$$

# Bayesian relaxed clocks

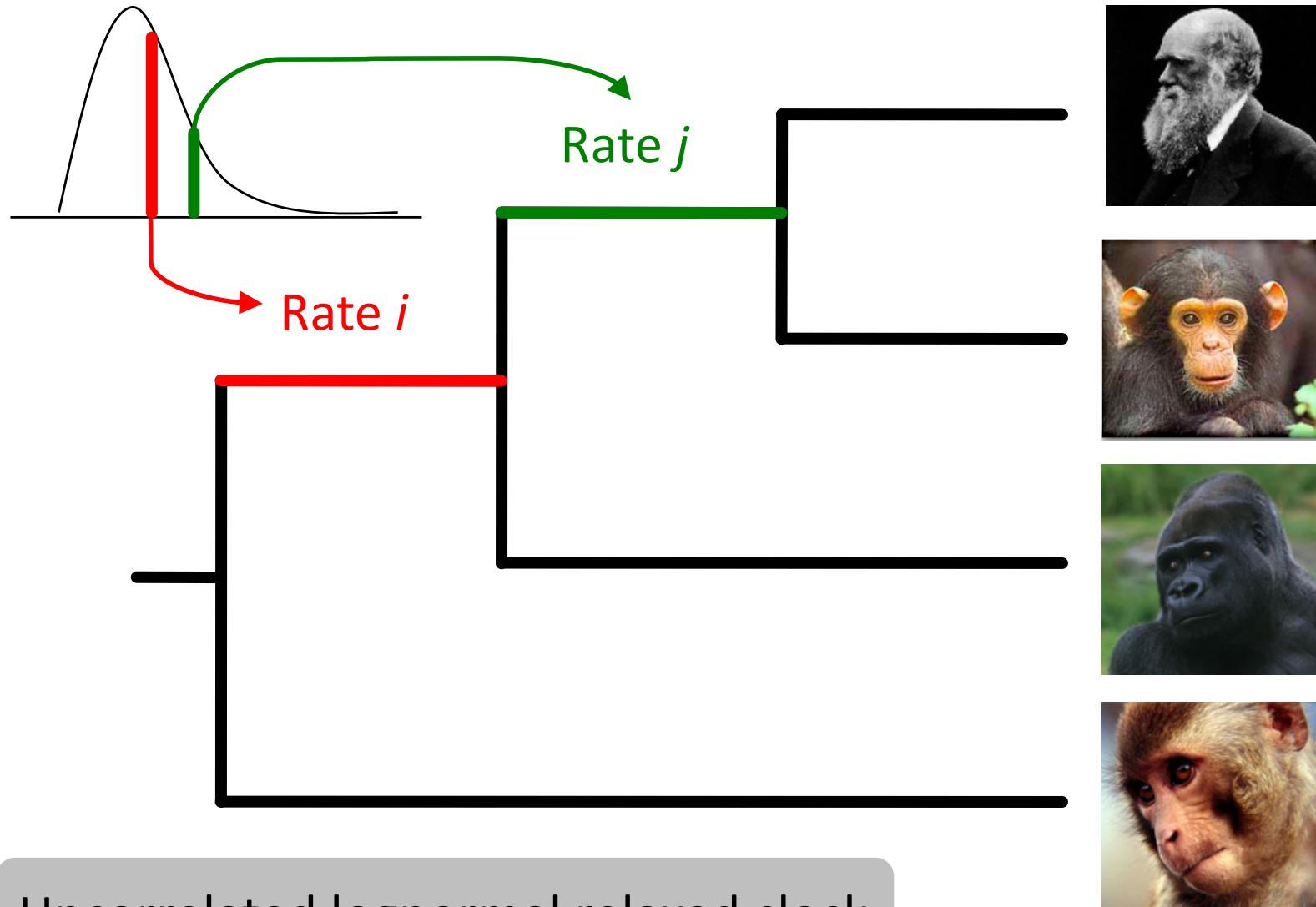
---

- 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 or ‘memory-less’**  
rates identically and independently distributed among branches

# Bayesian relaxed clocks



# Bayesian relaxed clocks



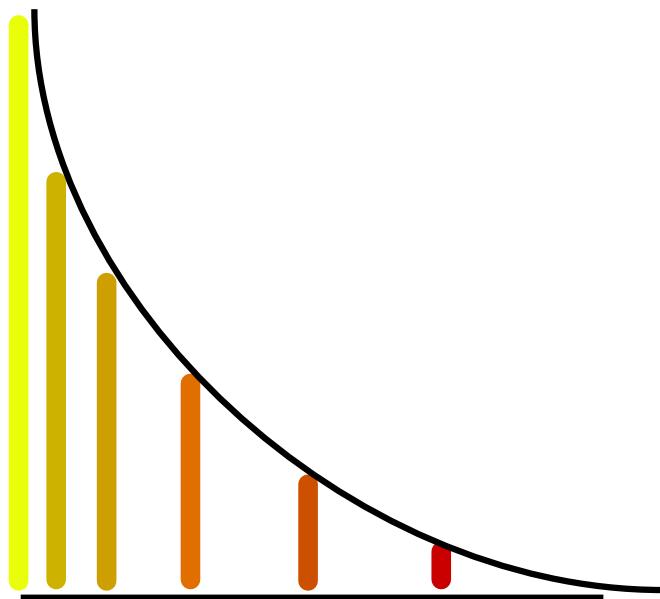
Uncorrelated lognormal relaxed clock

# Uncorrelated relaxed clock

---

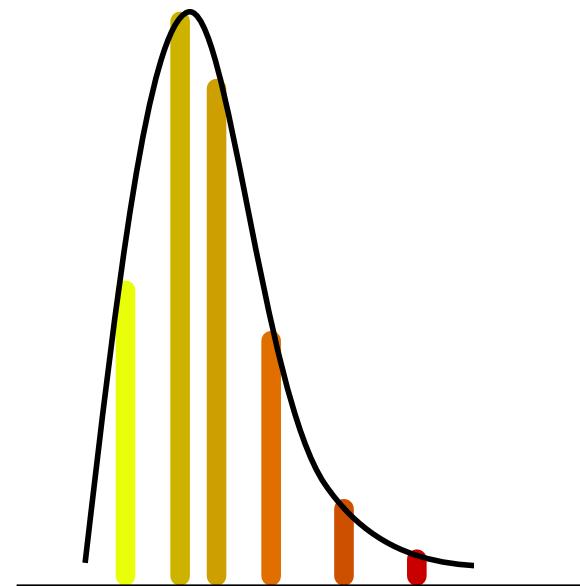
## Exponential distribution

Most rates are quite low



## Lognormal distribution

Most rates cluster around mean



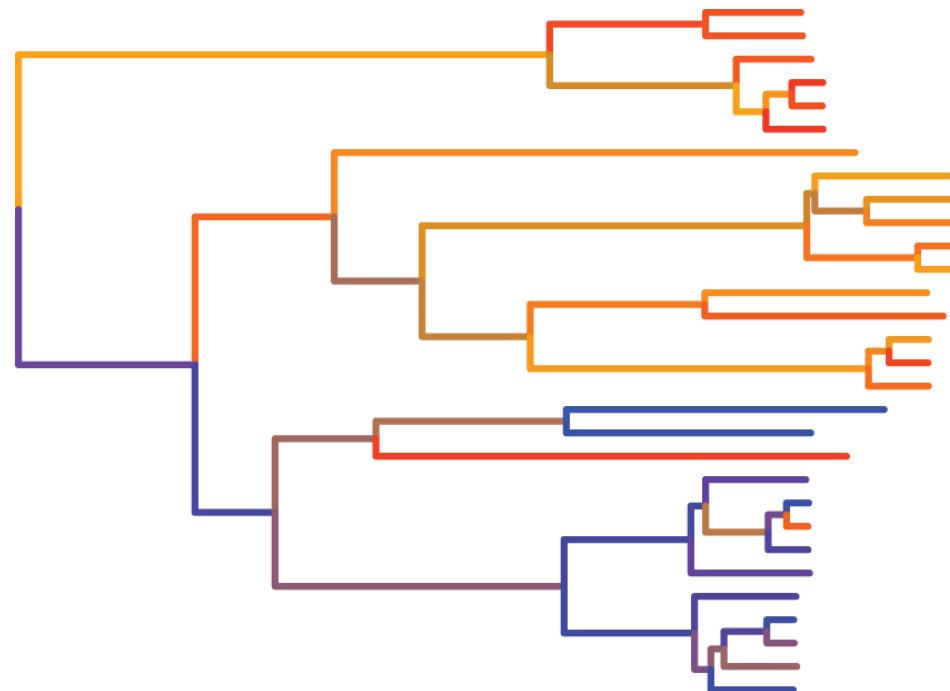
# Bayesian relaxed clocks

---

- 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

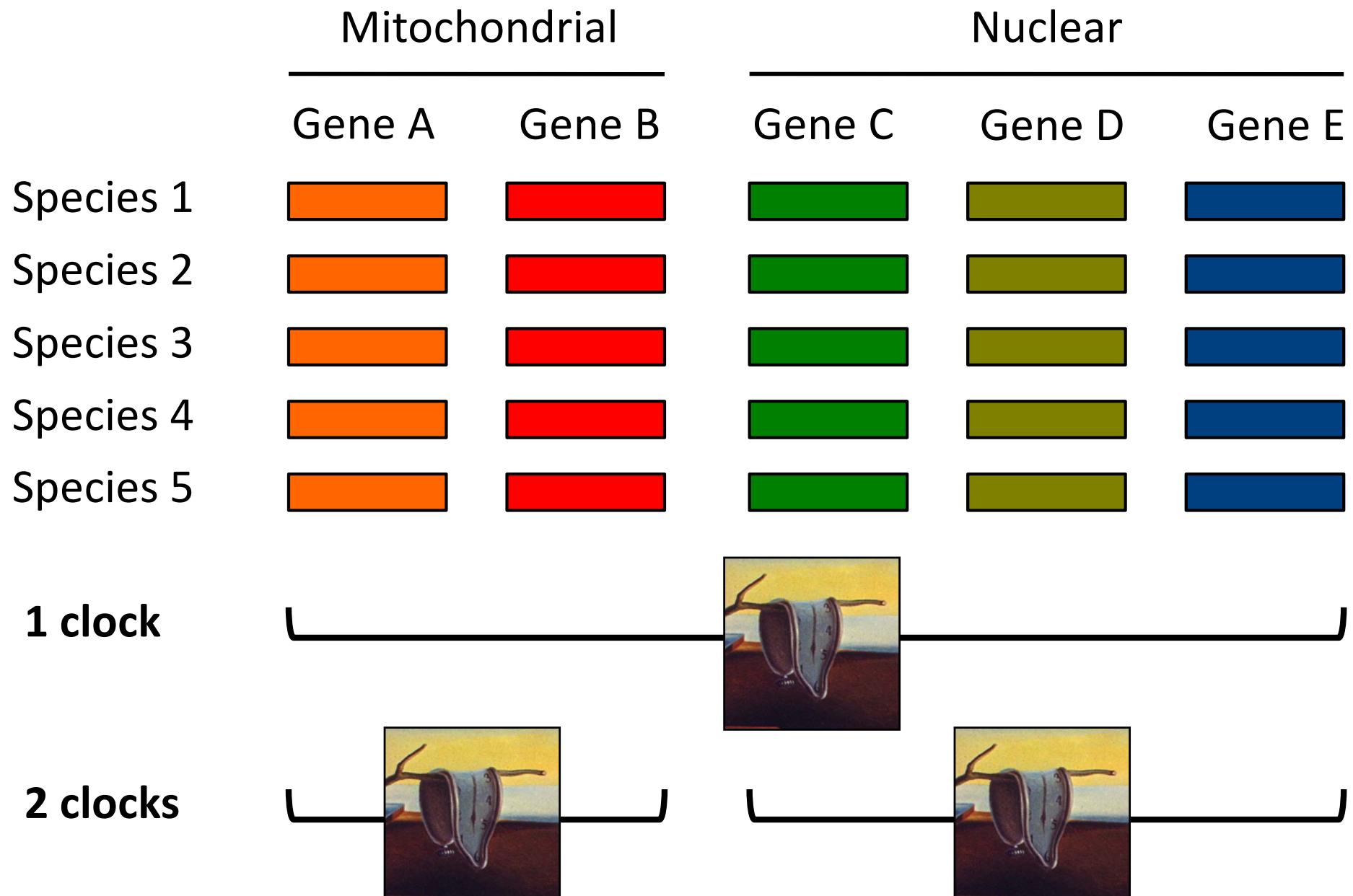
# Rate autocorrelation

- Little evidence of rate autocorrelation in real data
- Uncorrelated relaxed clock probably appropriate in most cases



# Partitioned Molecular Clocks

# Data partitioning



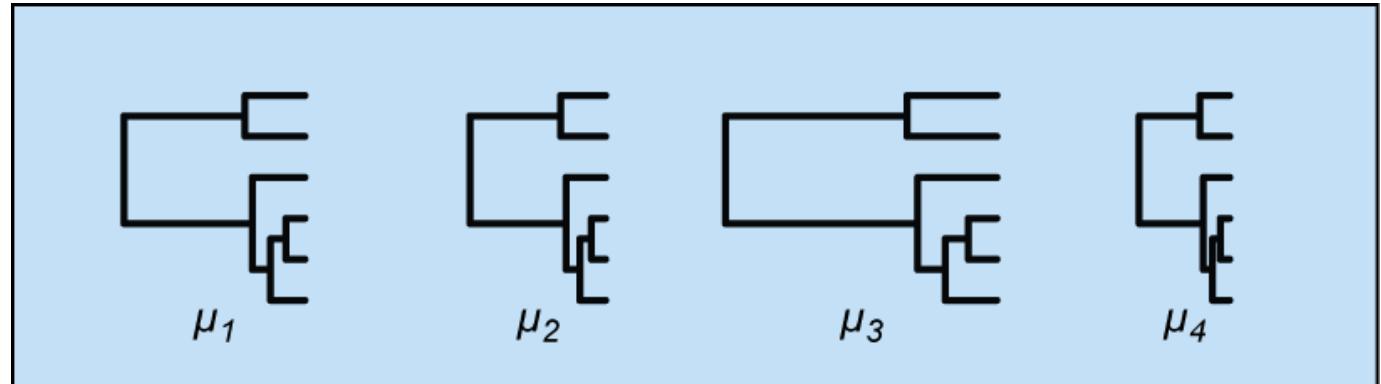
# Multiple clock models

---

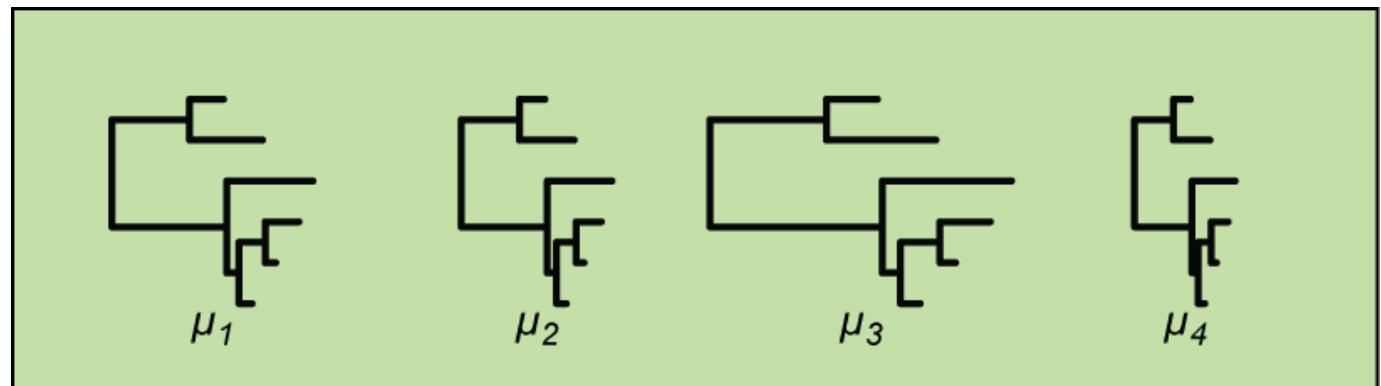
- Simple approach
  - Single relaxed-clock model
  - Scaling factor for each gene
- Multiple relaxed-clock models
  - Separate relaxed-clock model for each data subset
  - Potentially many parameters

# Multiple clock models

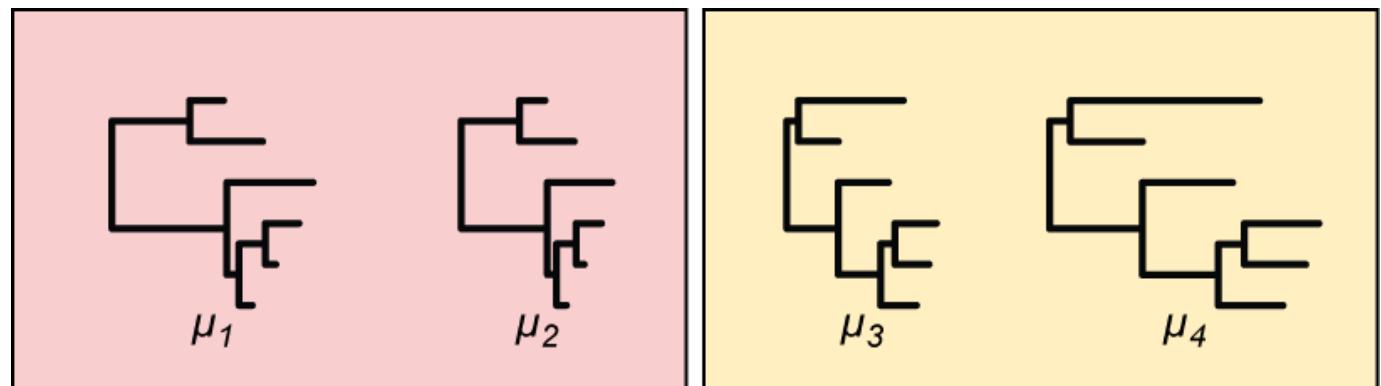
- Across loci  
**(gene effects)**

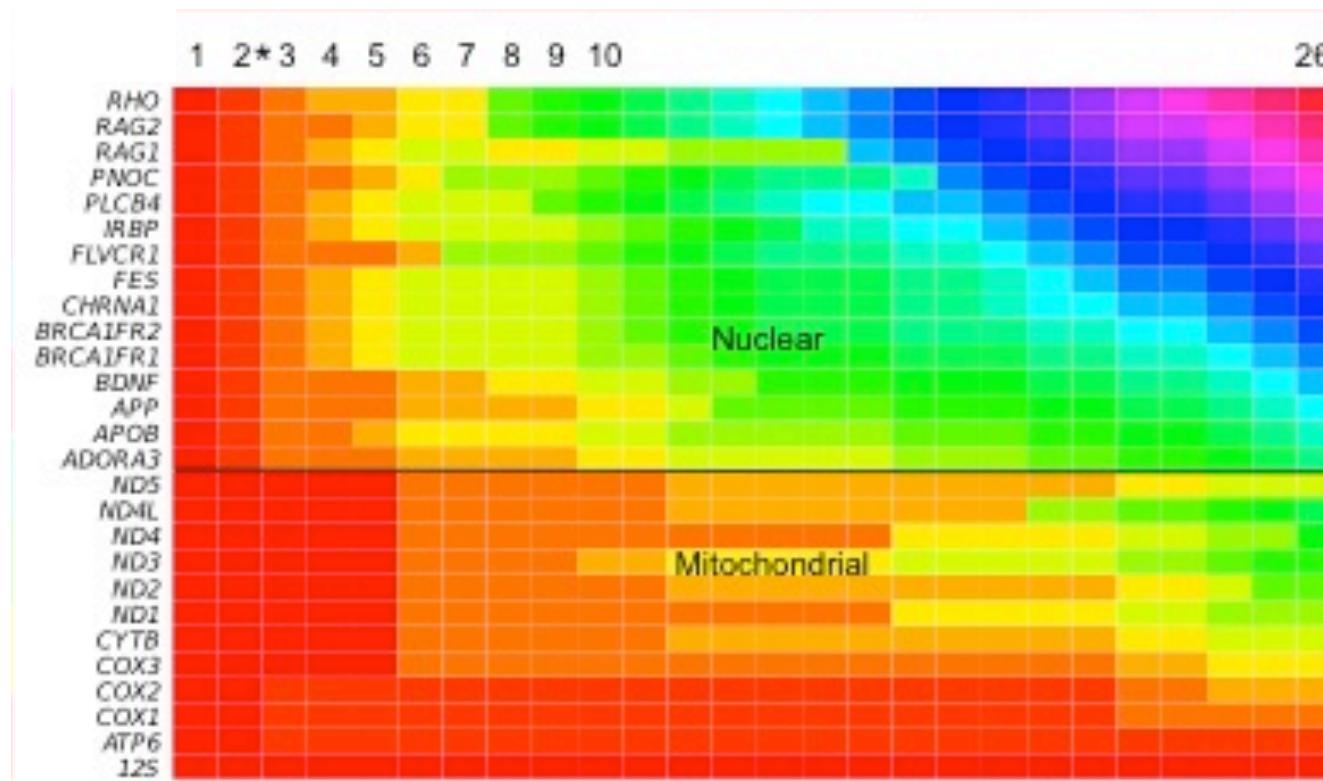


- Across lineages  
**(lineage effects)**

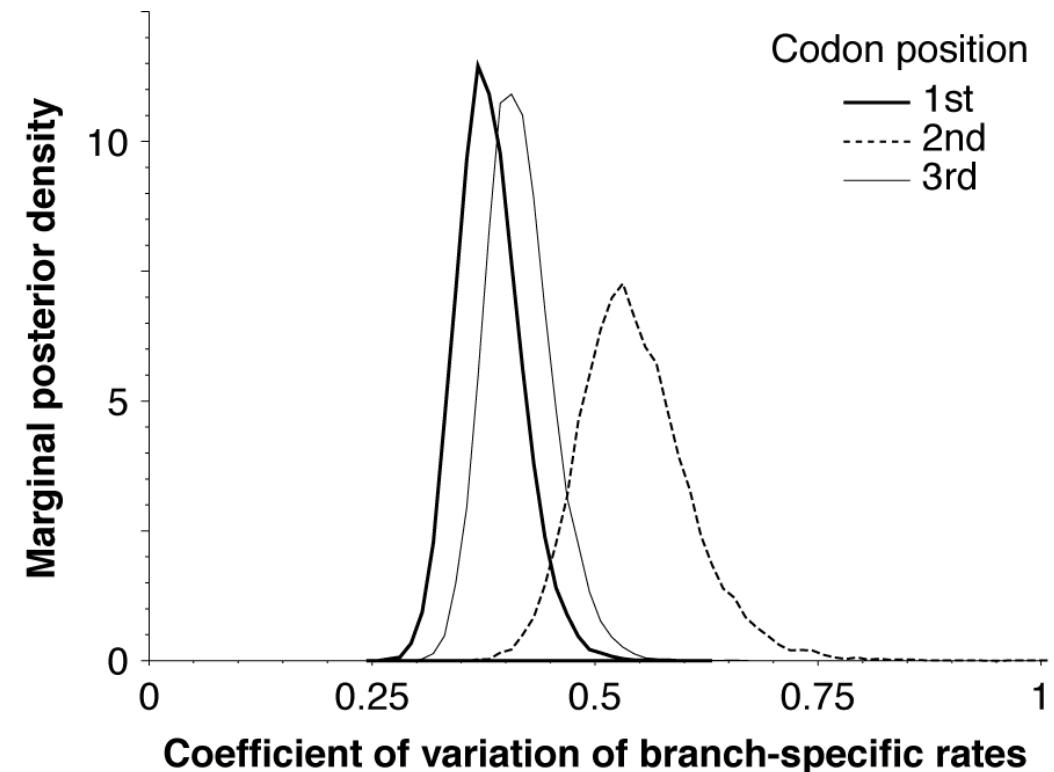
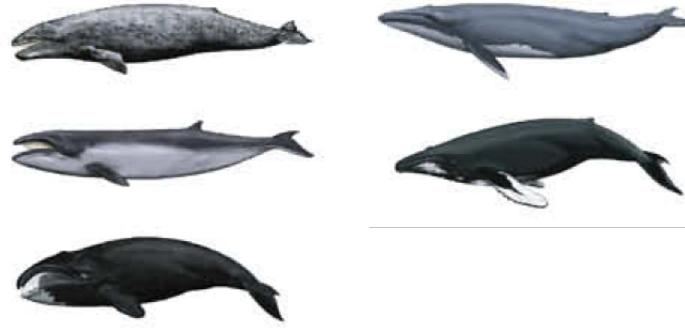
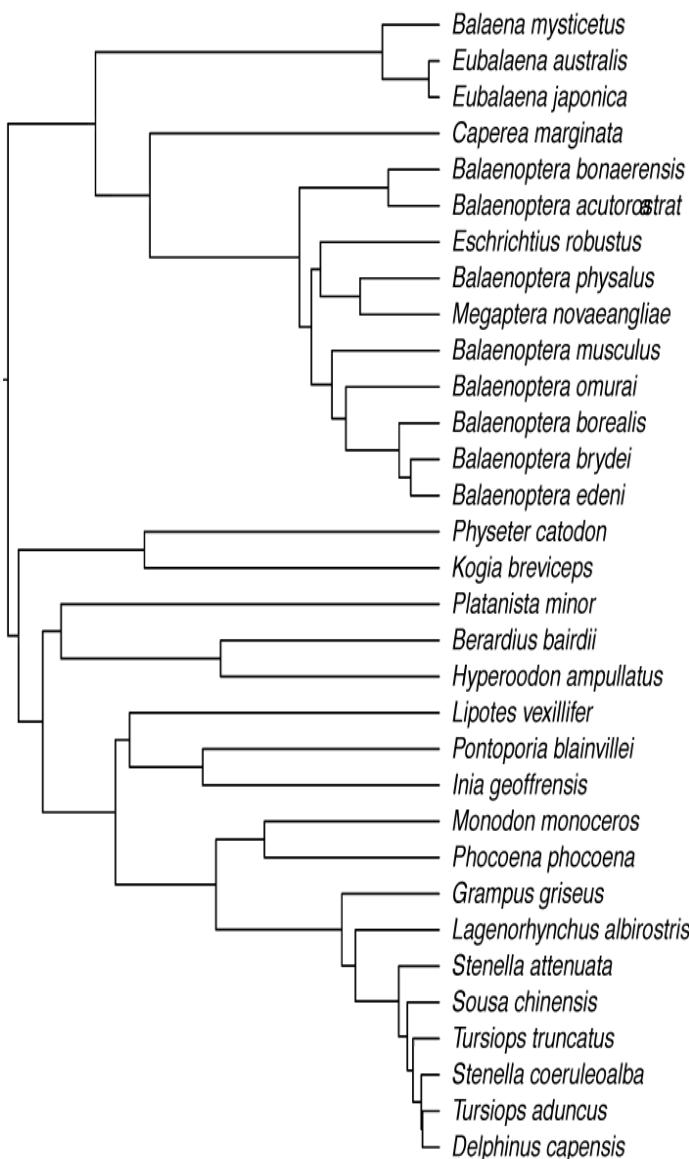


- Gene-by-lineage interaction  
**(residual effects)**



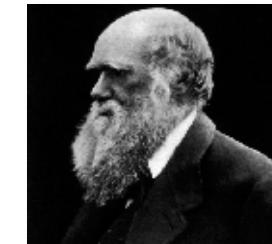
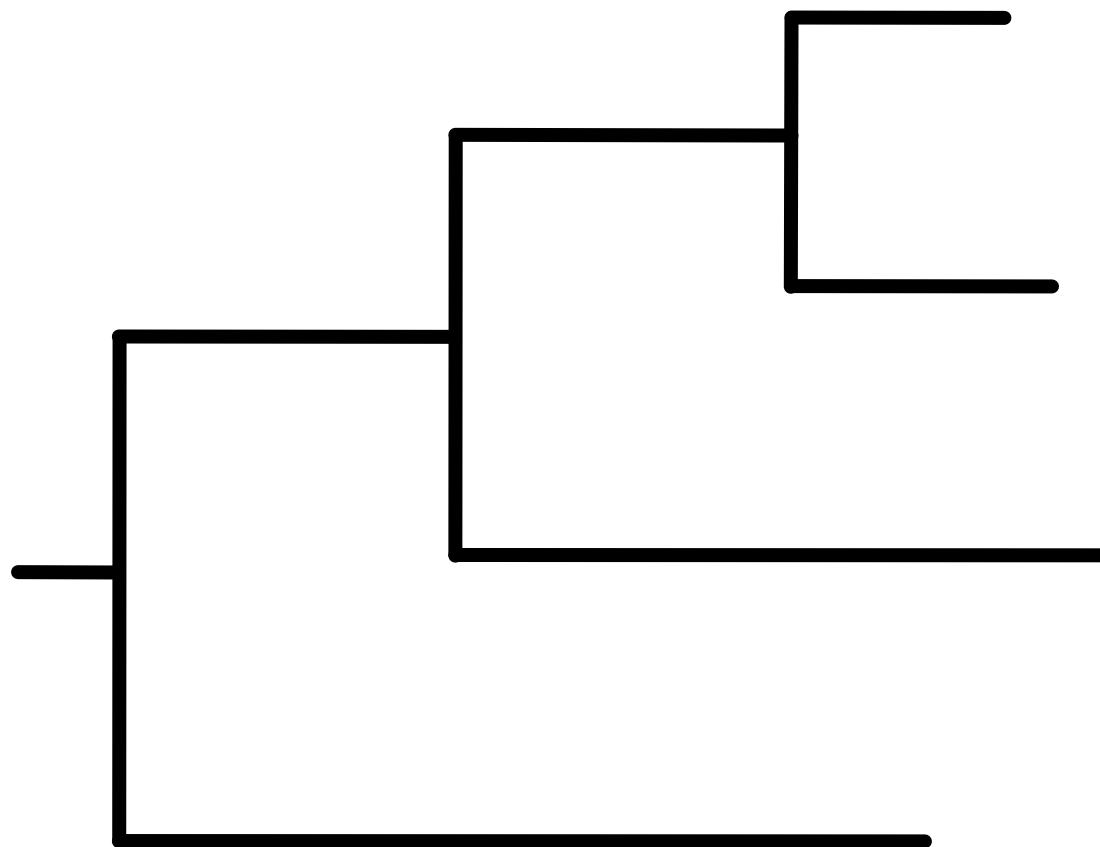


# Multiple clock models



# Calibrating the Molecular Clock

# The molecular clock



**ACTAGTTC**



**ACTAGTAC**



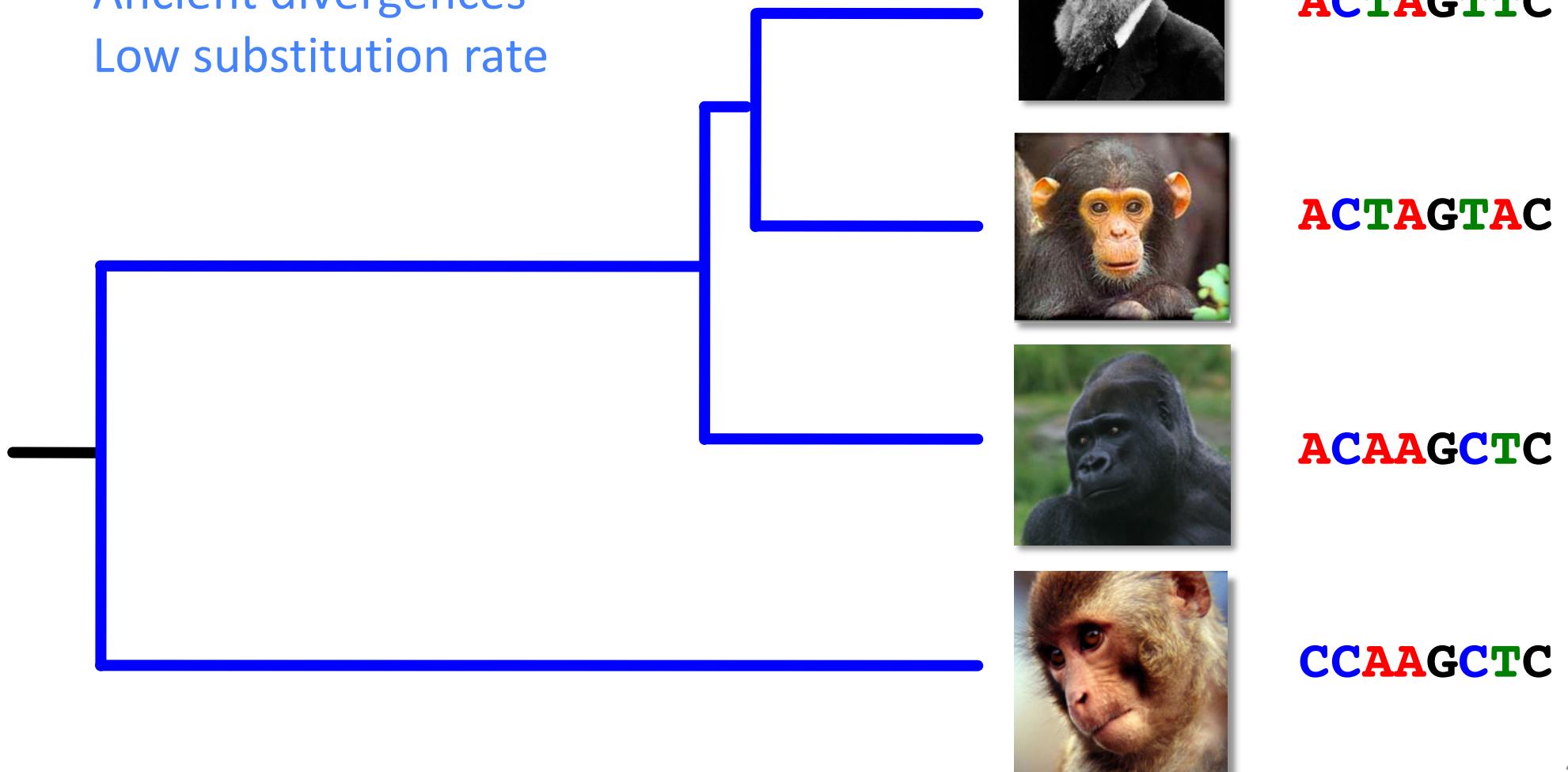
**ACAAGCTC**



**CCAAGCTC**

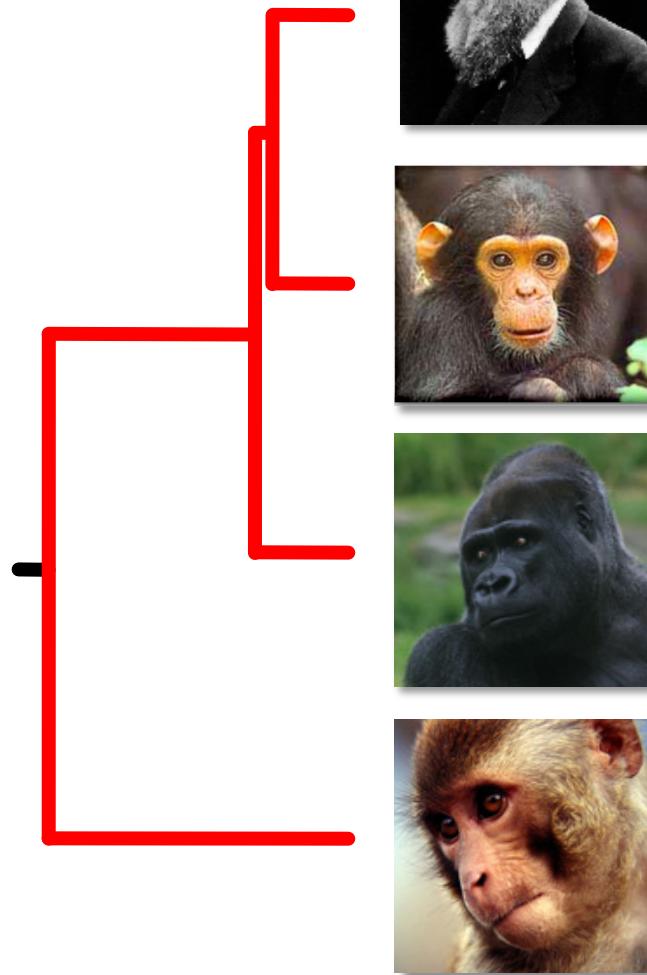
# The molecular clock

Ancient divergences  
Low substitution rate



# The molecular clock

Recent divergences  
High substitution rate



**ACTAGTTC**

**ACTAGTAC**

**ACAAGCTC**

**CCAAGCTC**

# Calibrating the molecular clock

---

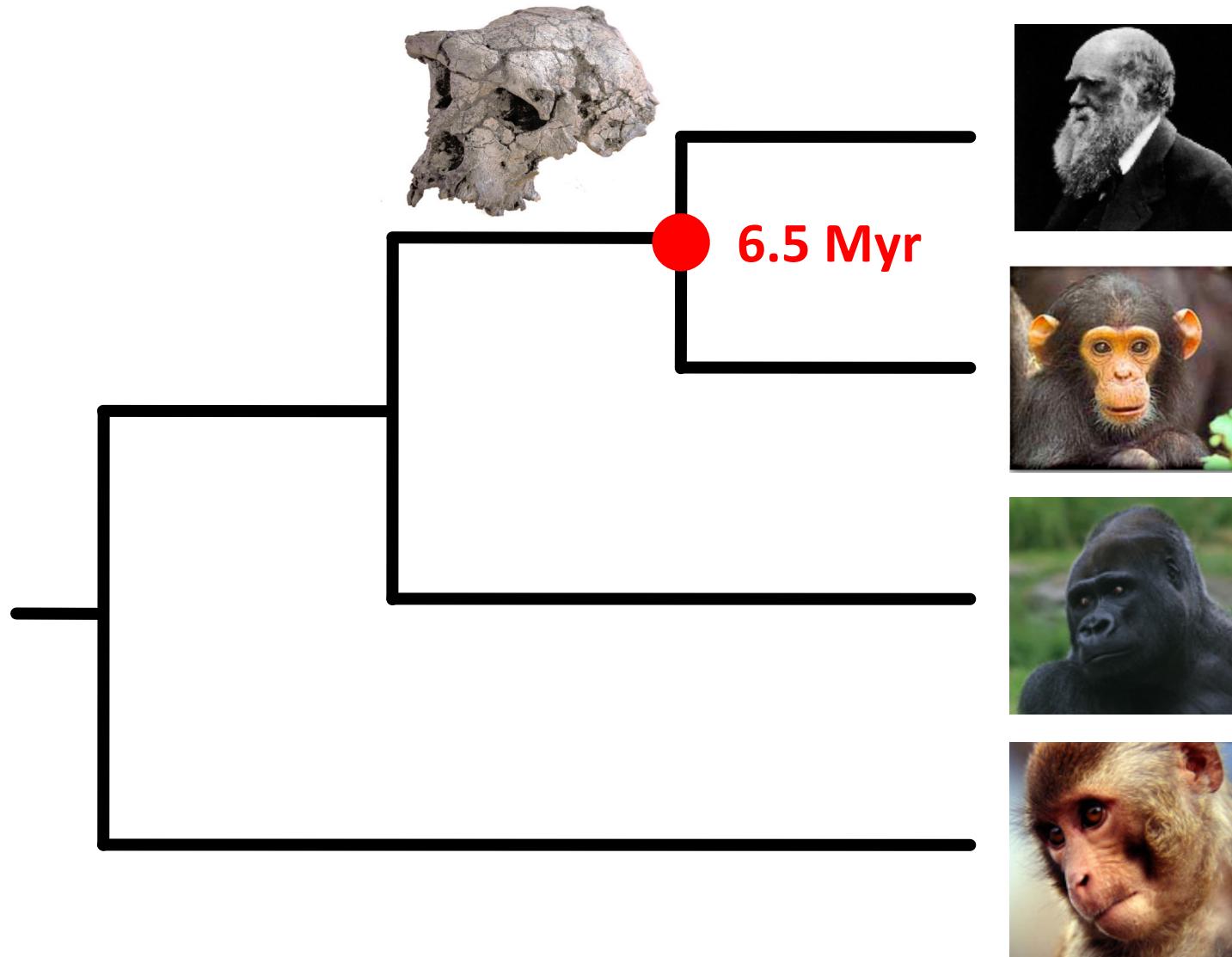
- Rates and times are **non-identifiable**
- Likelihood only depends on their product
  - Branch lengths in substitutions per site
- To separate rate and time, we need (prior) information about one or the other

# Calibrating the molecular clock

---

- Information about **substitution rate**
  - Use to specify prior density of rate
- Information about **node times**
  - Fossil record
  - Biogeography
  - Sampling times

# 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 density of node age specified by user

## 2. Use fossil directly in the analysis

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

# Choosing fossil calibrations

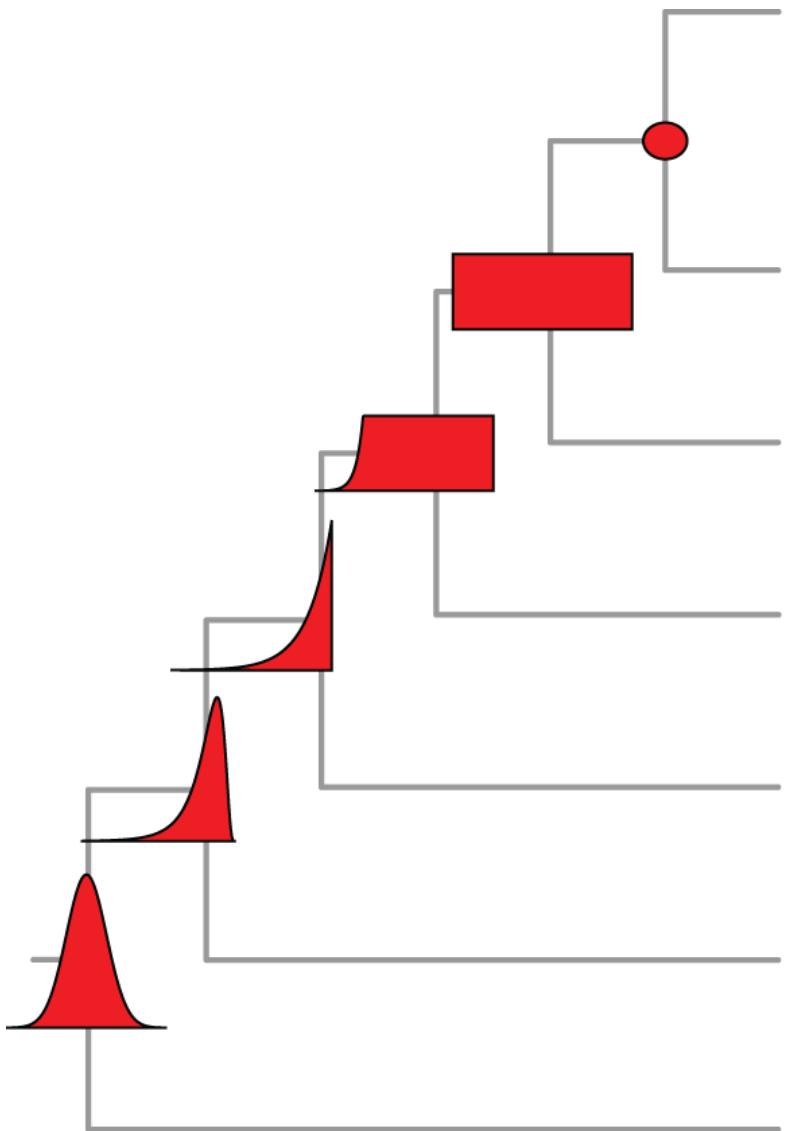
---

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

# Calibration Priors

# Calibrations

---

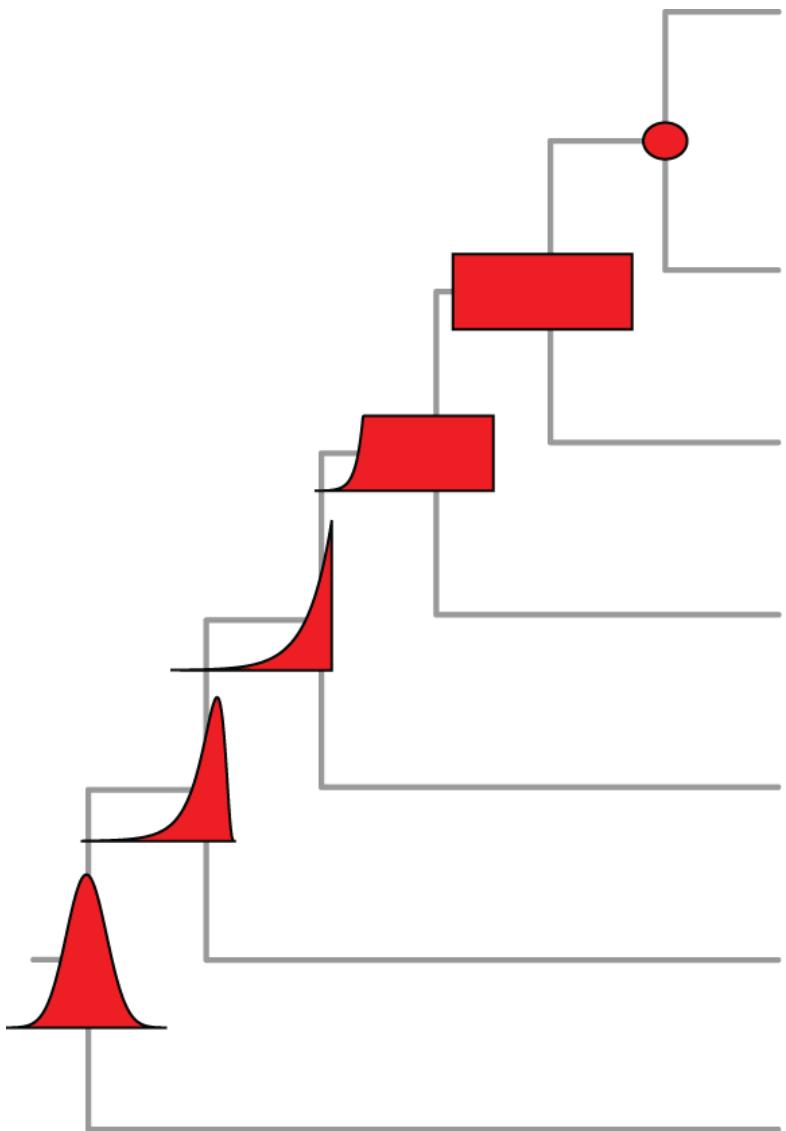


## Point calibration

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

# Calibrations

---

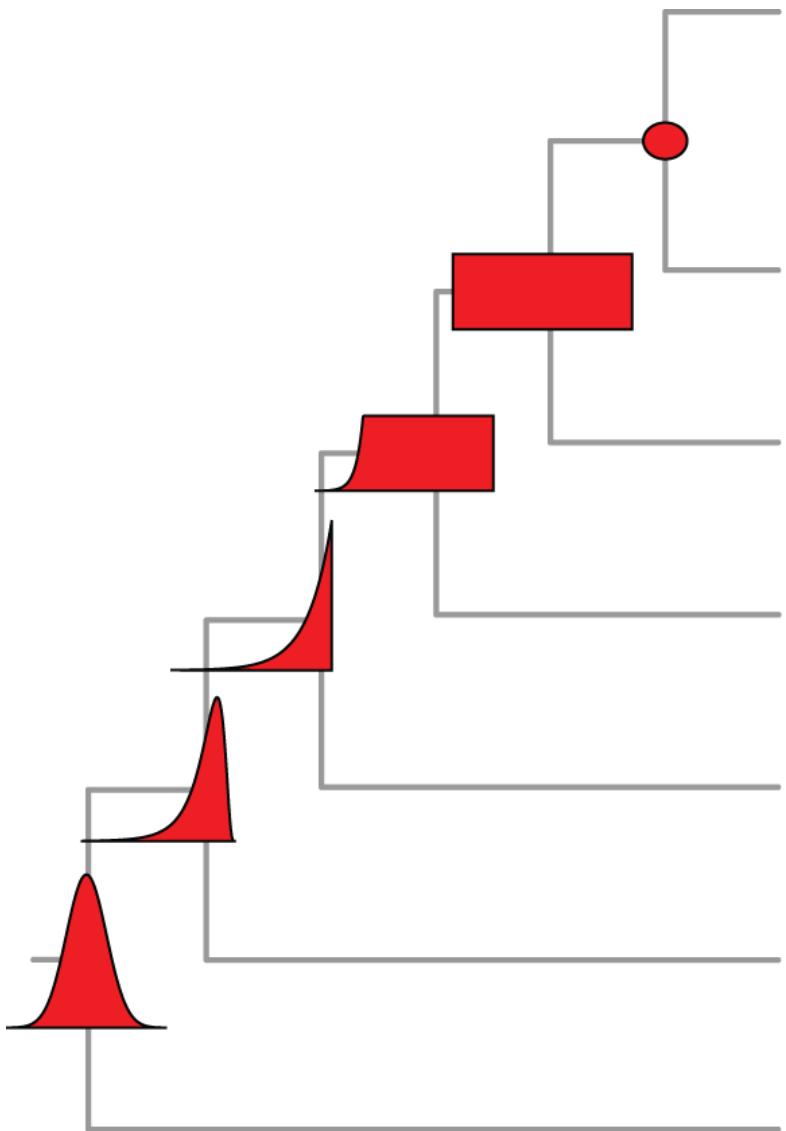


## Uniform prior

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

# Calibrations

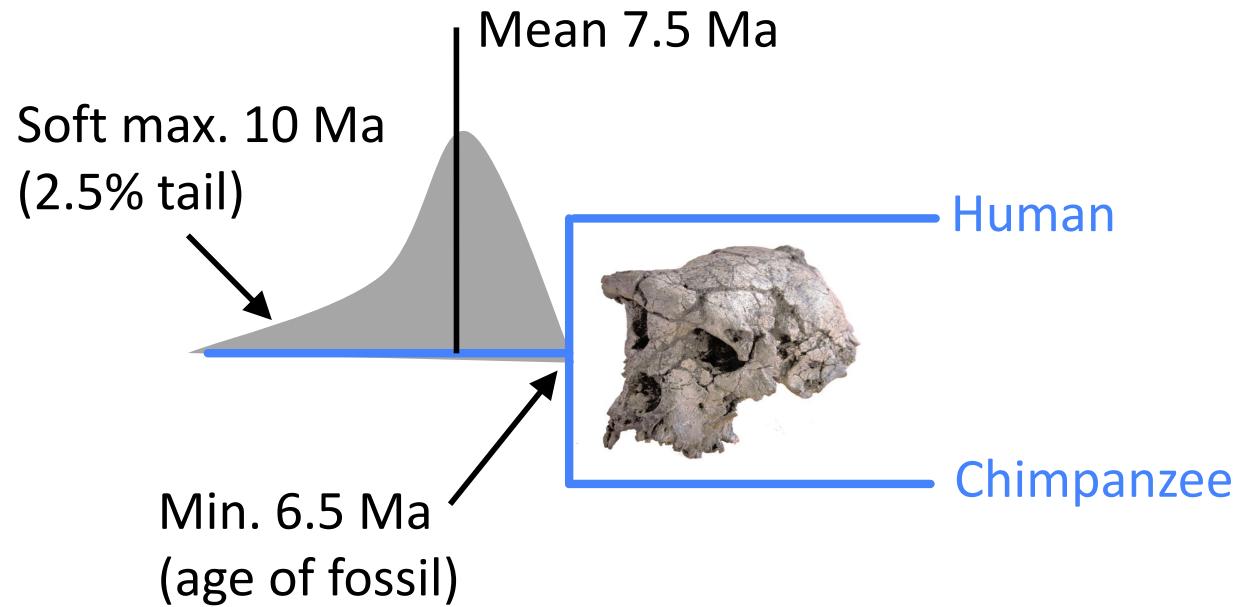
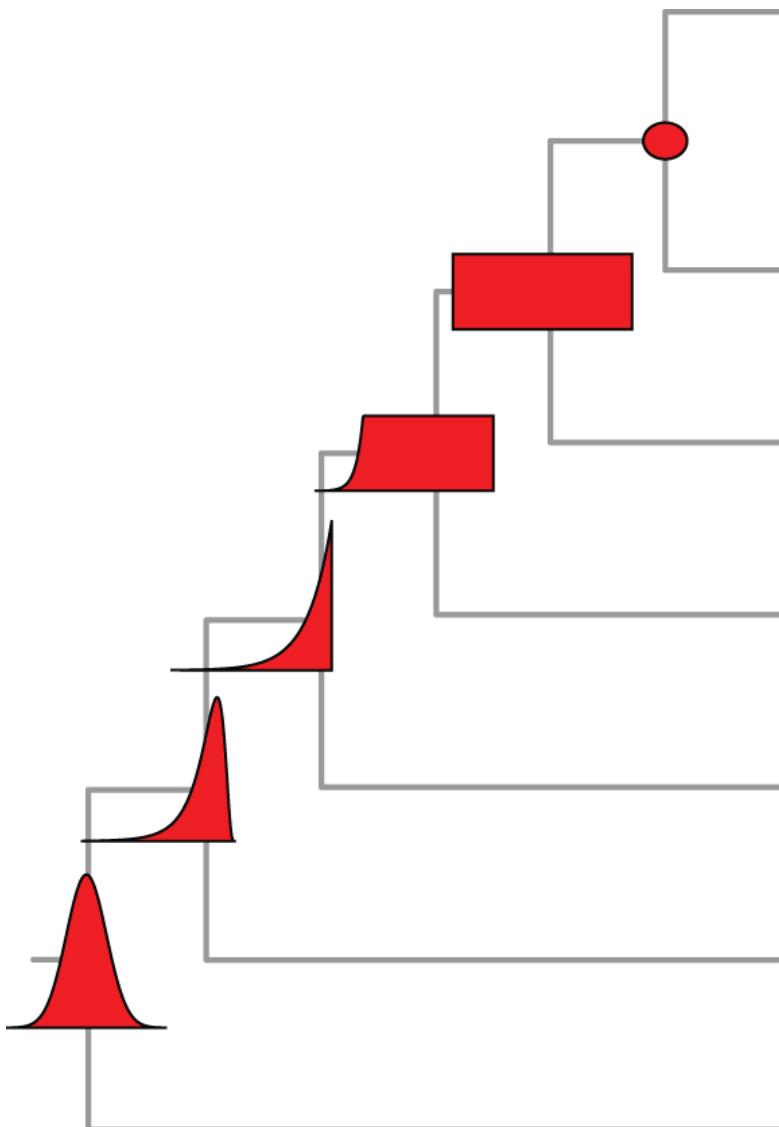
---



## Exponential prior

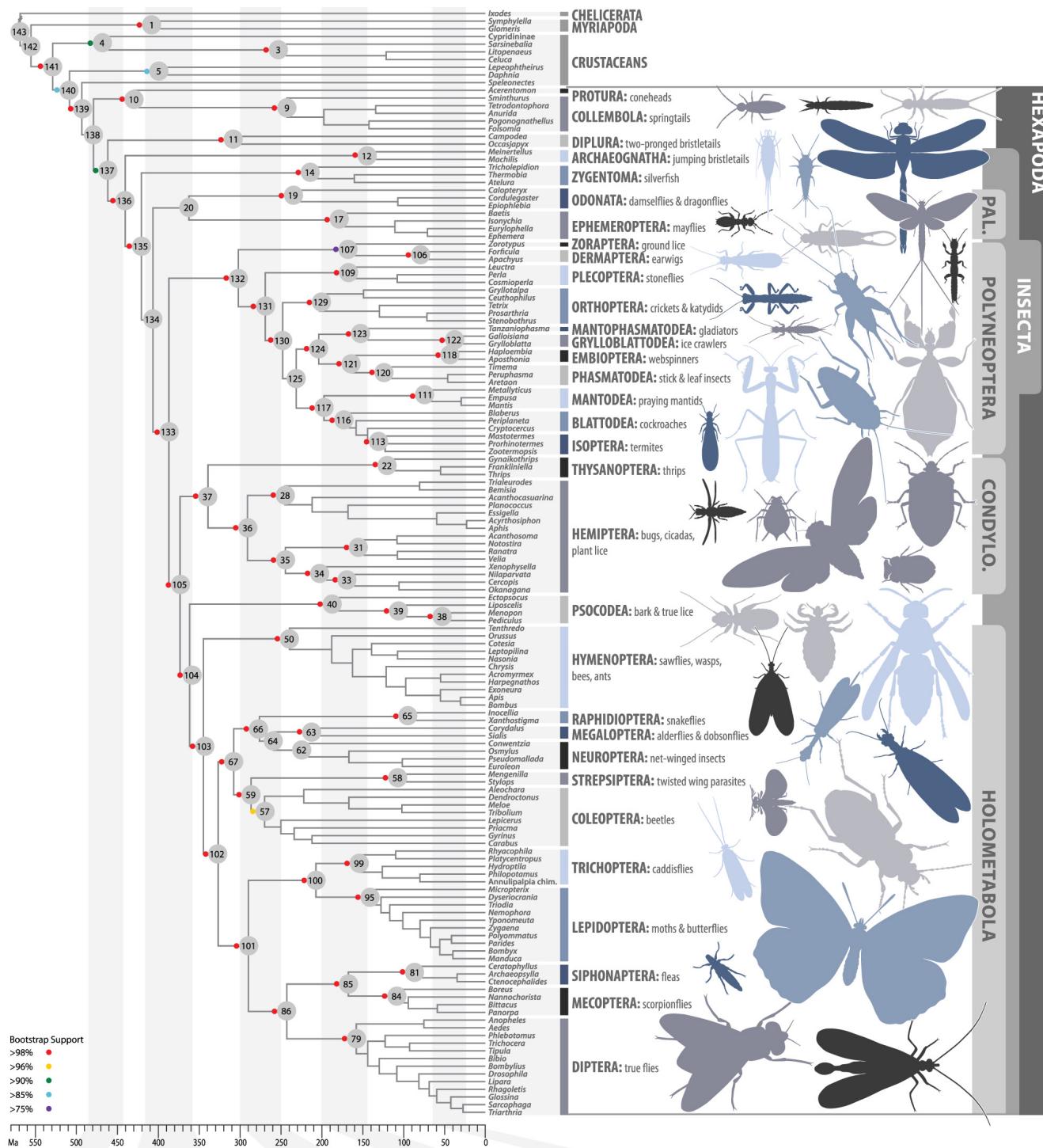
- Need 2 values: minimum and mean
- Strong assumption about relationship of fossil taxon to internal node

# Calibrations



## Lognormal prior

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



- Misof et al. (2014)
- Lognormal priors for ages of 20 nodes
- Arbitrary values:  
Mean = 2  
St. dev. = 0.5

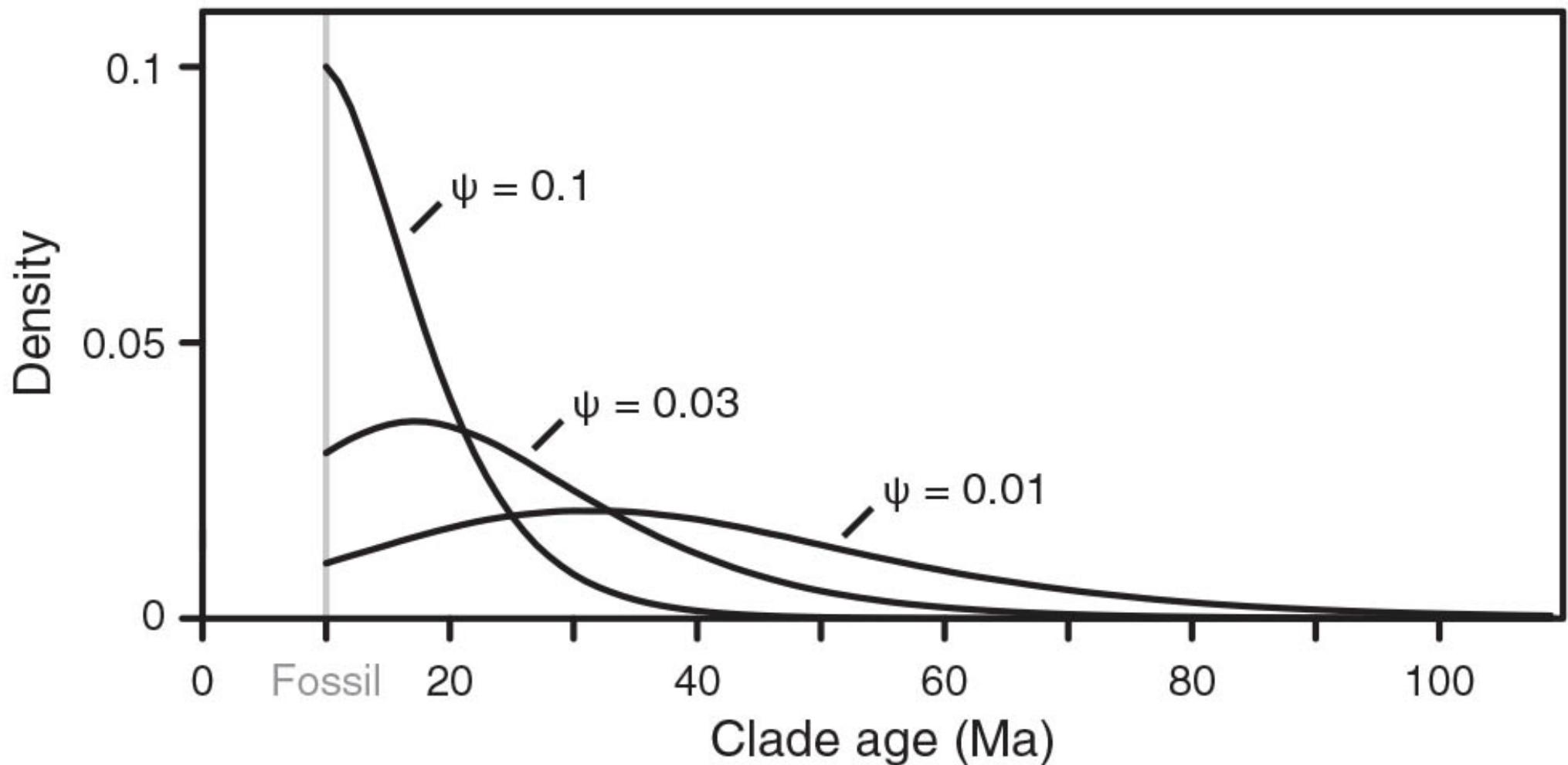
# CladeAge add-on package

---

- Calculates calibration densities based on:
  - First occurrence age
  - Net diversification rate  
(speciation rate – extinction rate)
  - Turnover rate  
(extinction rate / speciation rate)
  - Fossil sampling rate
  - Optional: Sampling gap

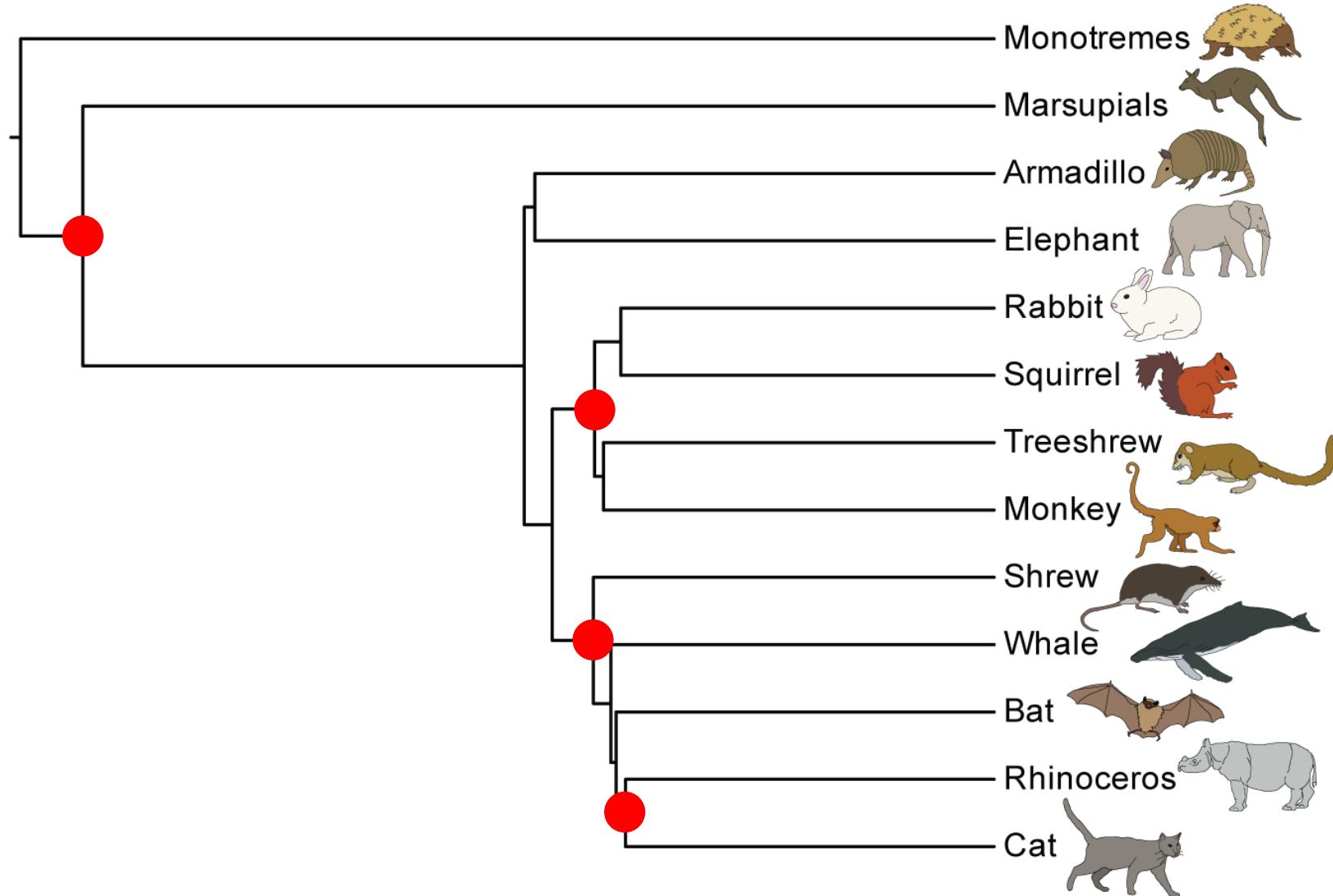


# CladeAge add-on package



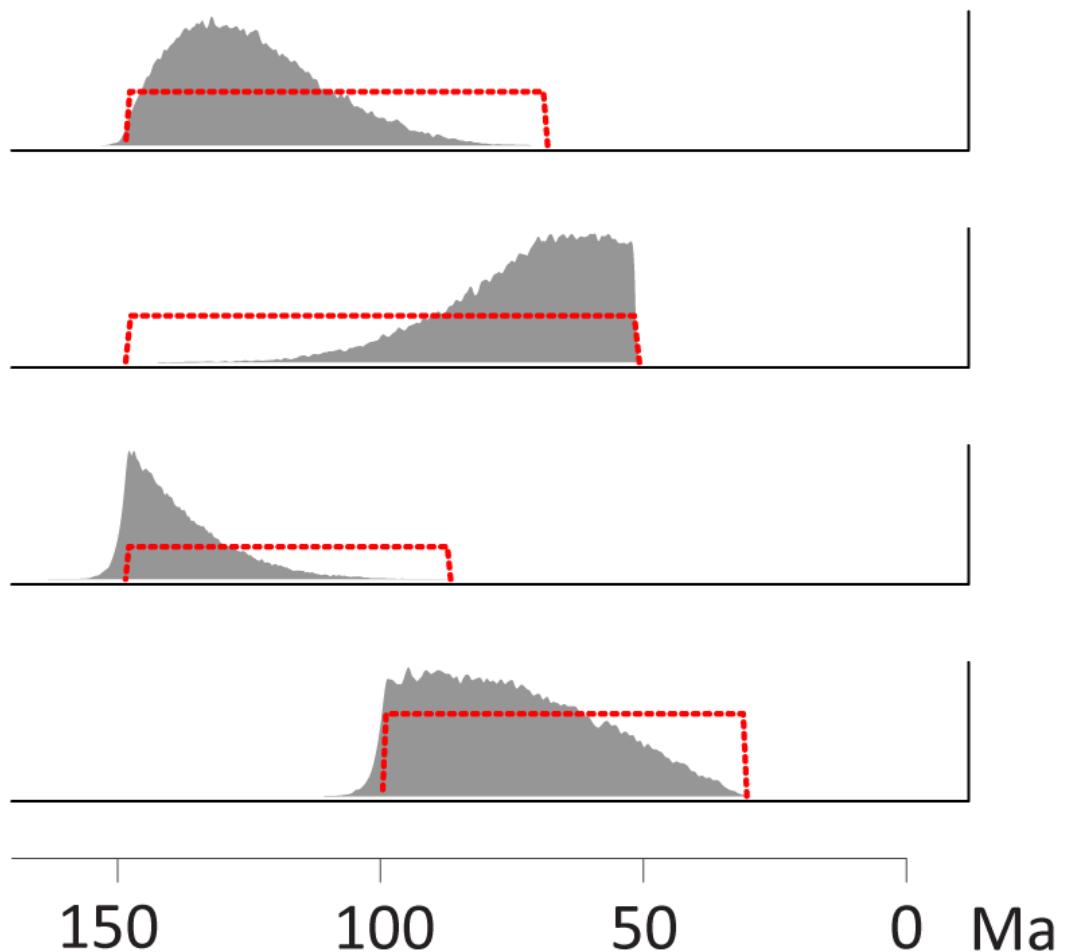
# Multiple calibrations

- Use multiple calibrations if possible



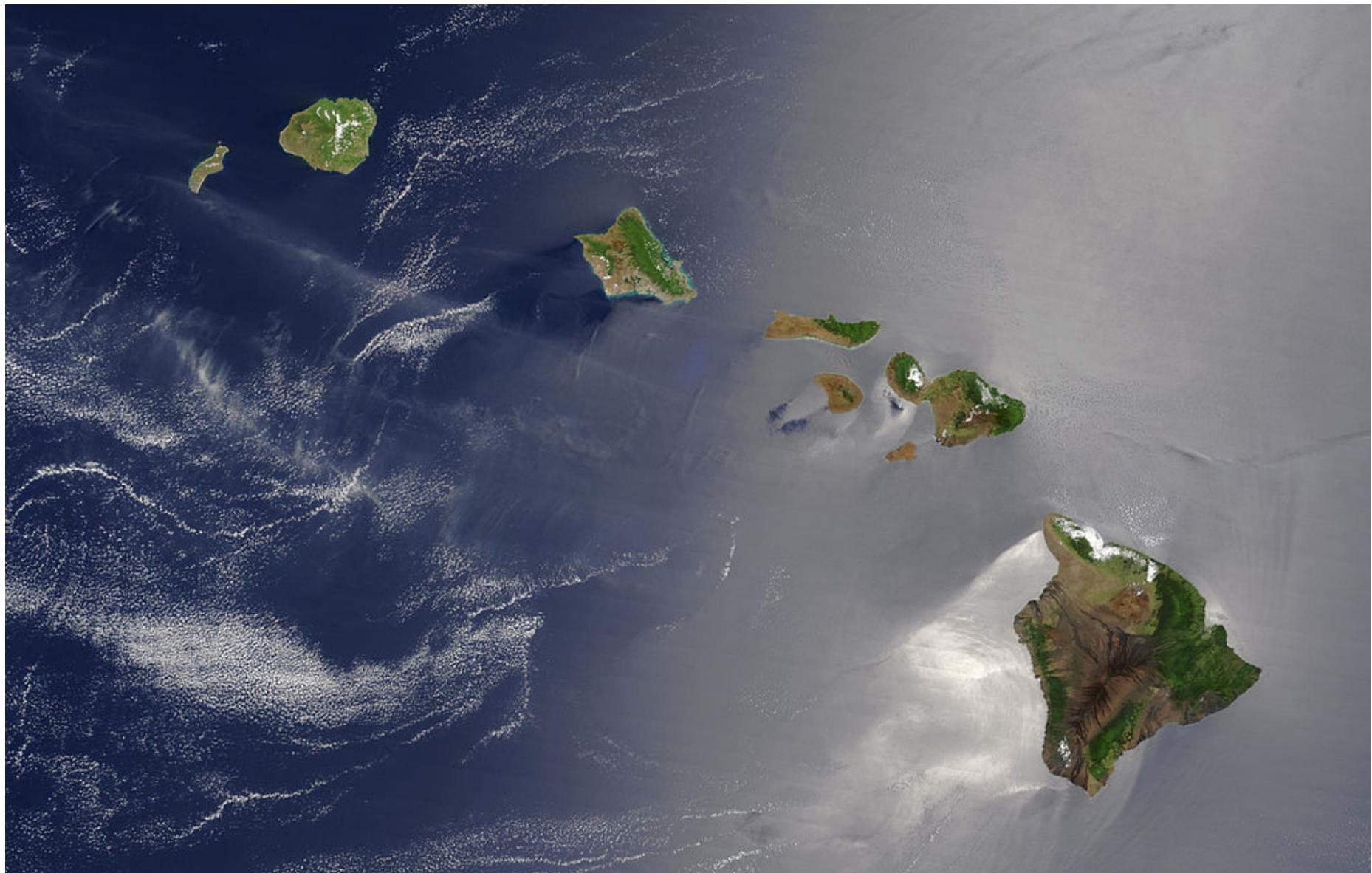
# Multiple calibrations

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



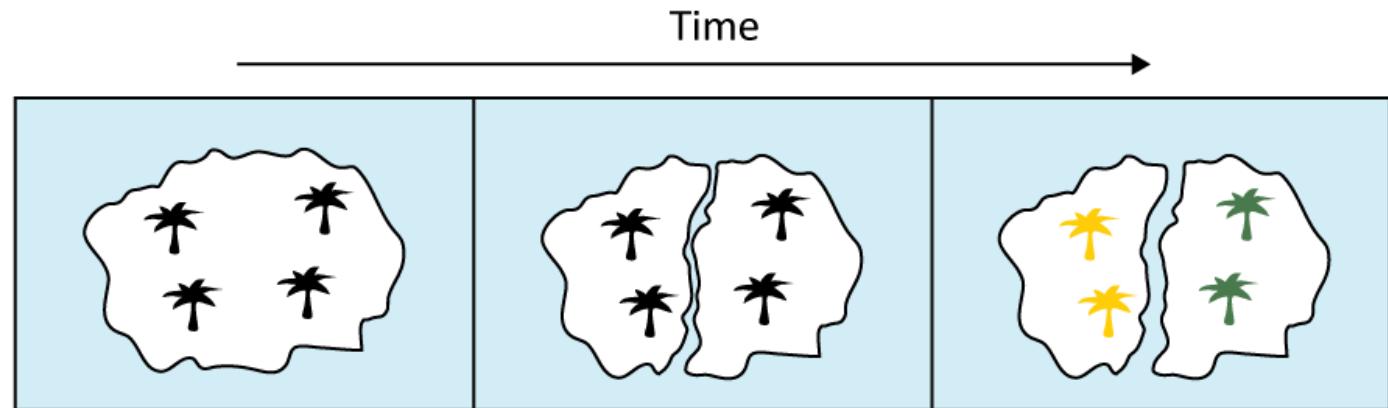
# Calibration: Biogeography

---

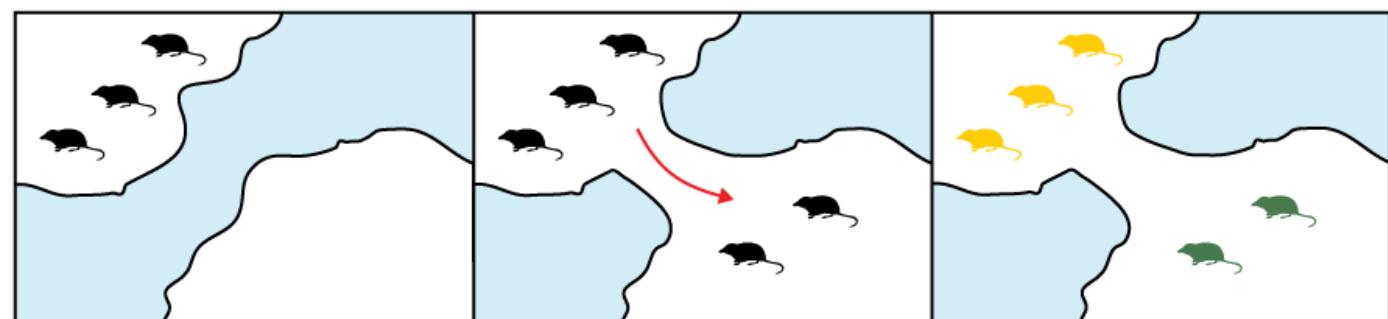


# Calibration: Biogeography

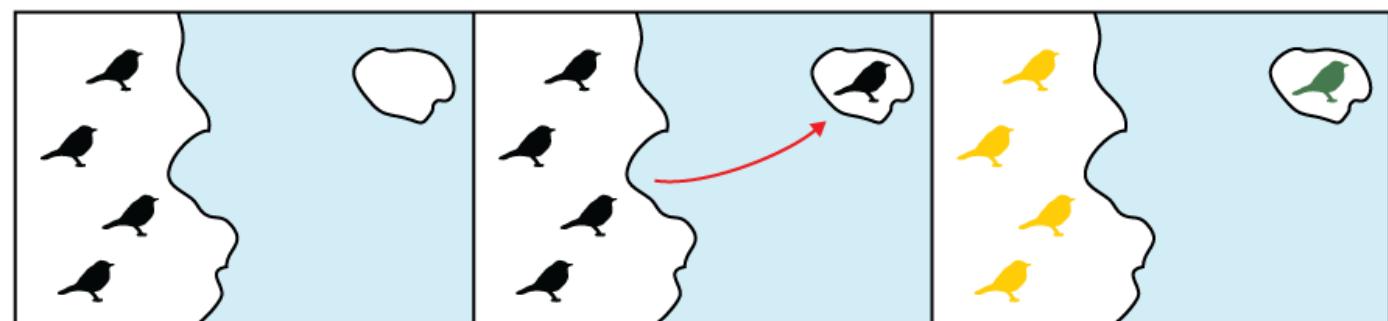
Vicariance



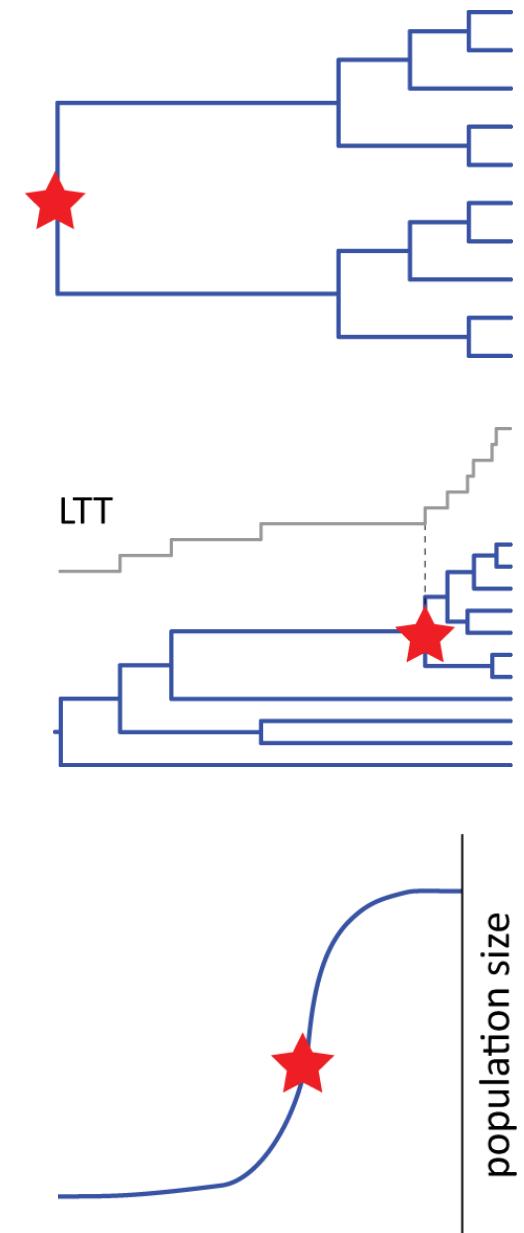
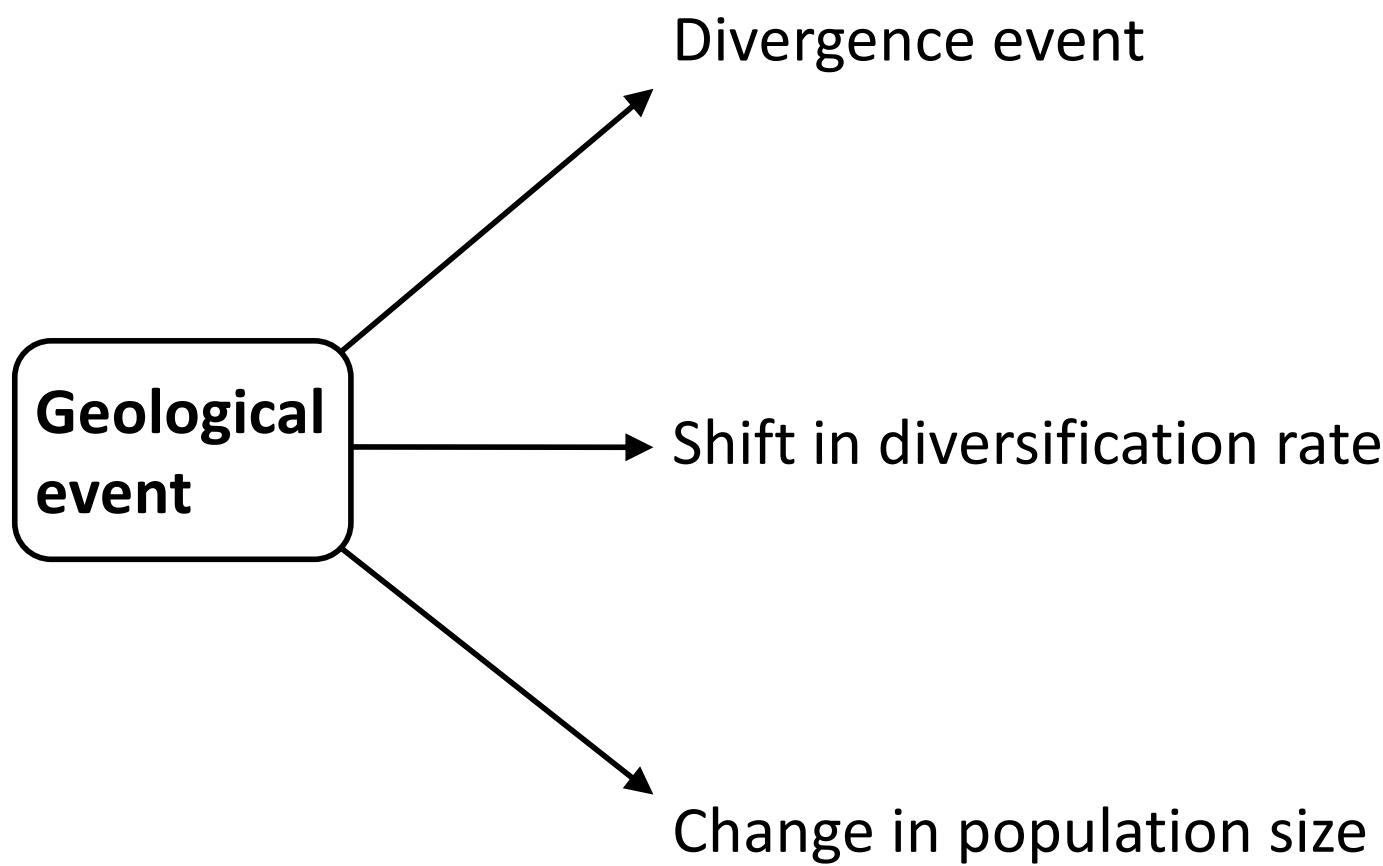
Geodispersal



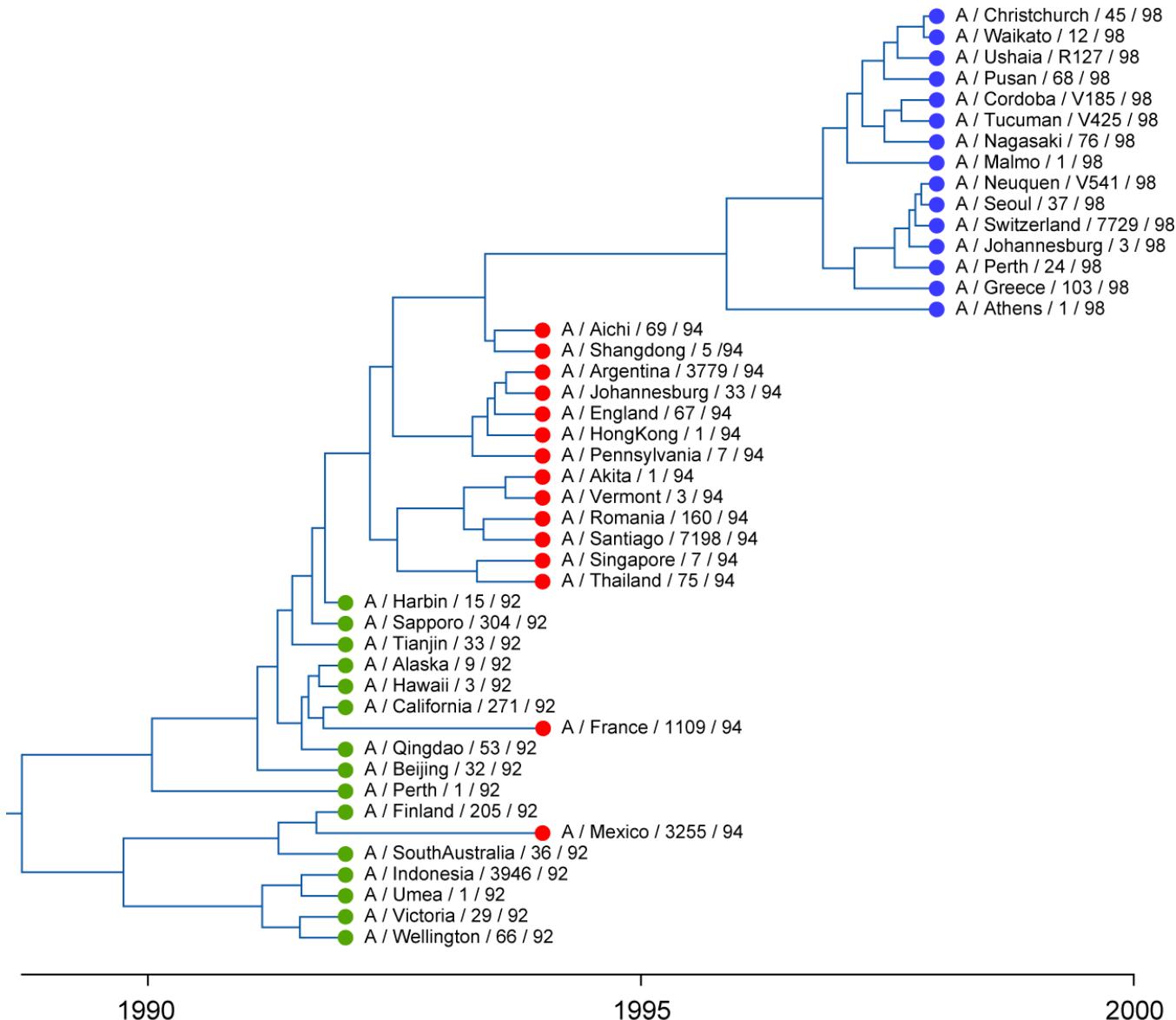
Biological dispersal



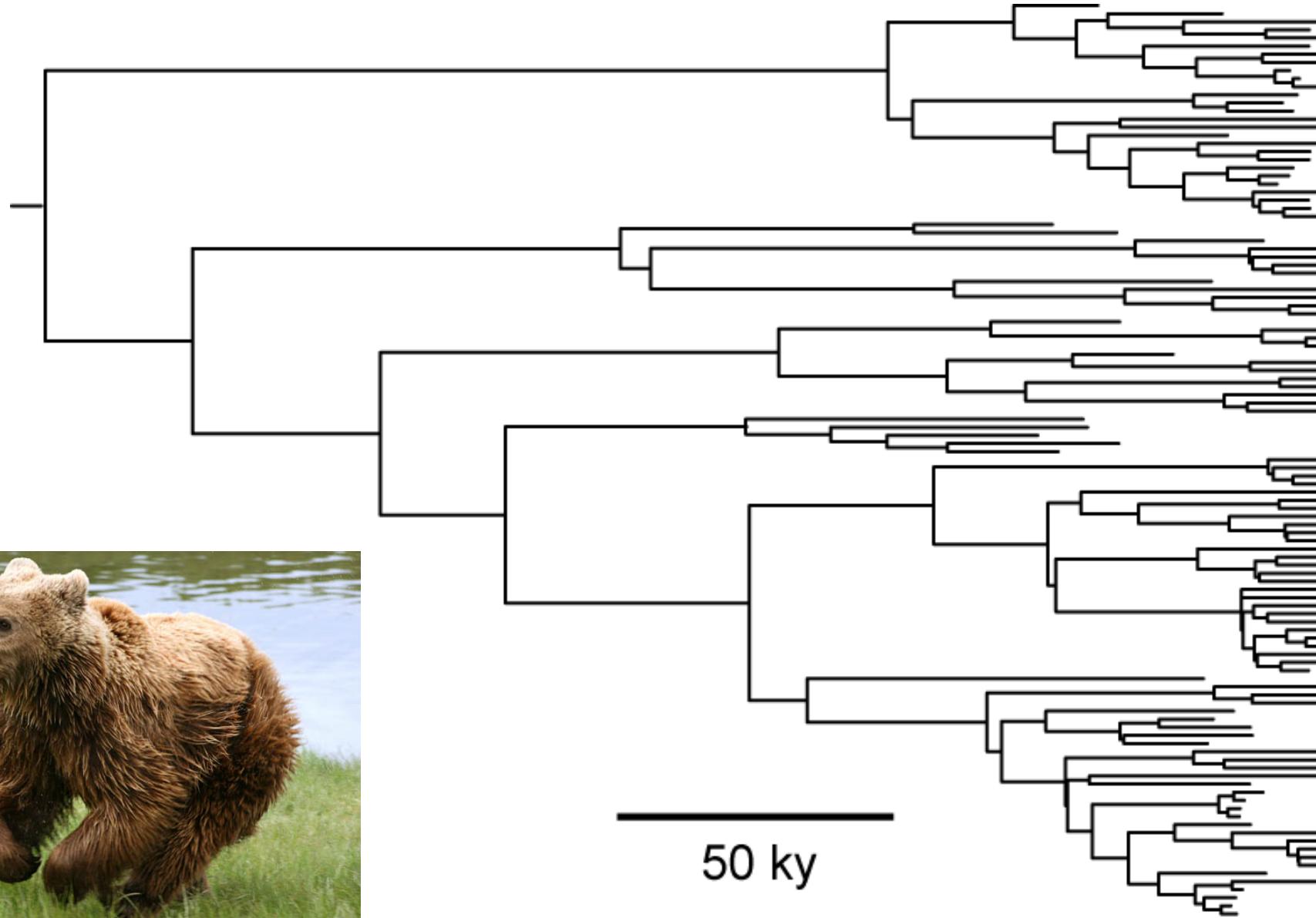
# Calibration: Biogeography



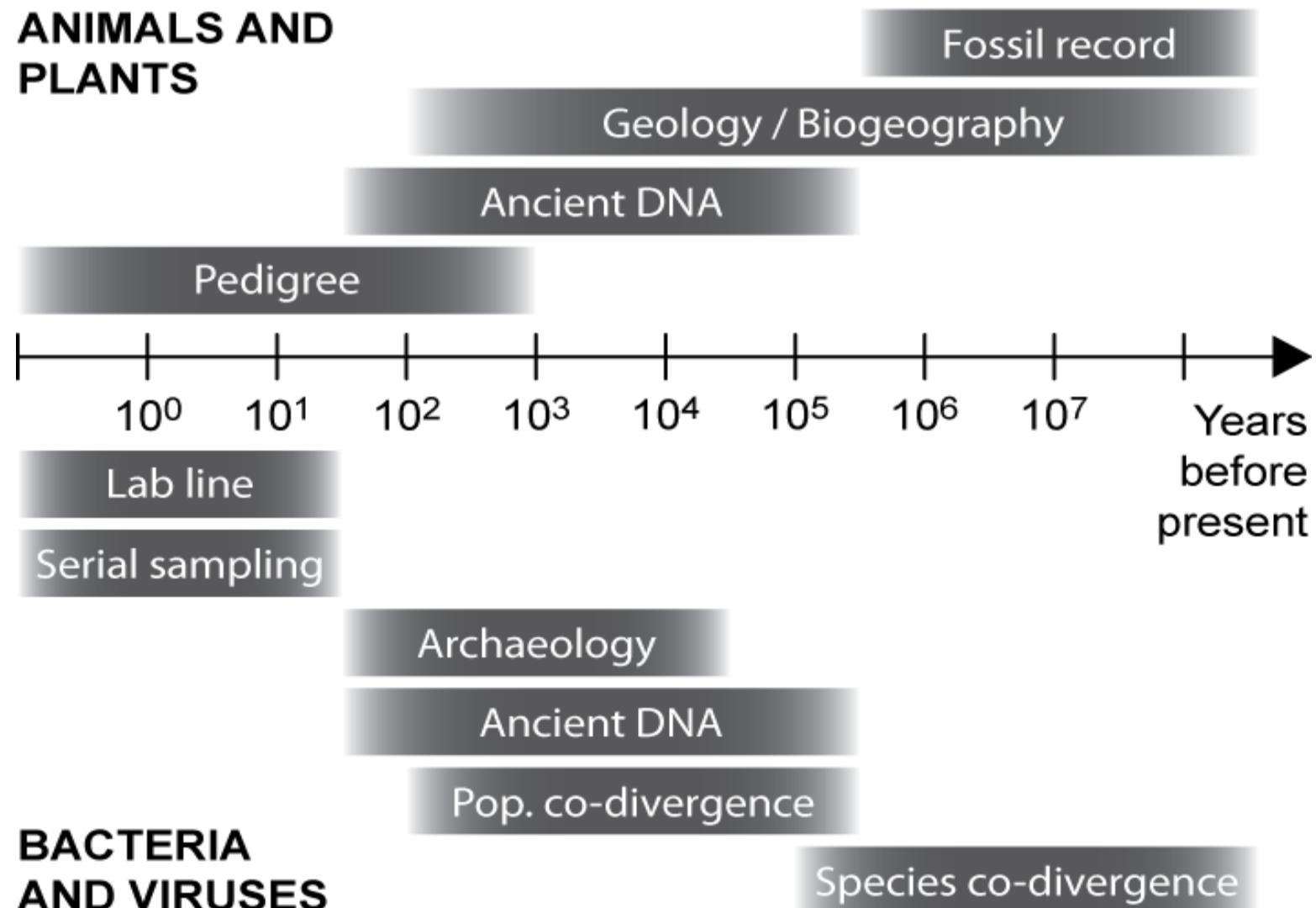
# Calibration: Sampling times



# 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

---

- **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.
- **Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors**  
Warnock *et al.* (2014) *Proceedings of the Royal Society B*, 282: 20141013.
- **Accounting for uncertainty in phylogenetic estimation of evolutionary 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.