Bayesian Divergence-Time Estimation

@trayc7.bsky.social

phylo@iastate.edu

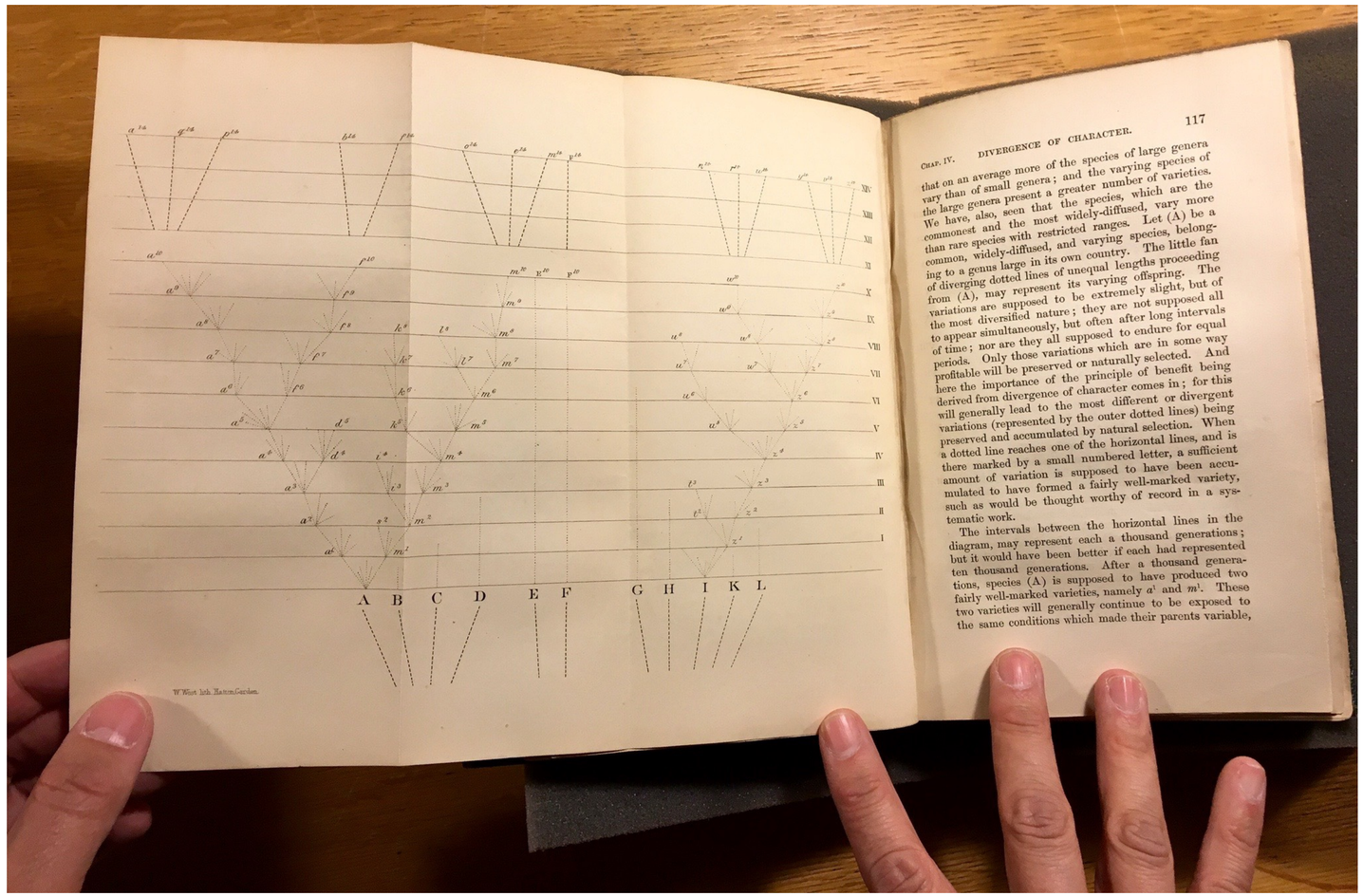
phyloworks.org

Tracy Heath

Ecology, Evolution, & Organismal Biology
Iowa State University

A Time-Scale for Evolution

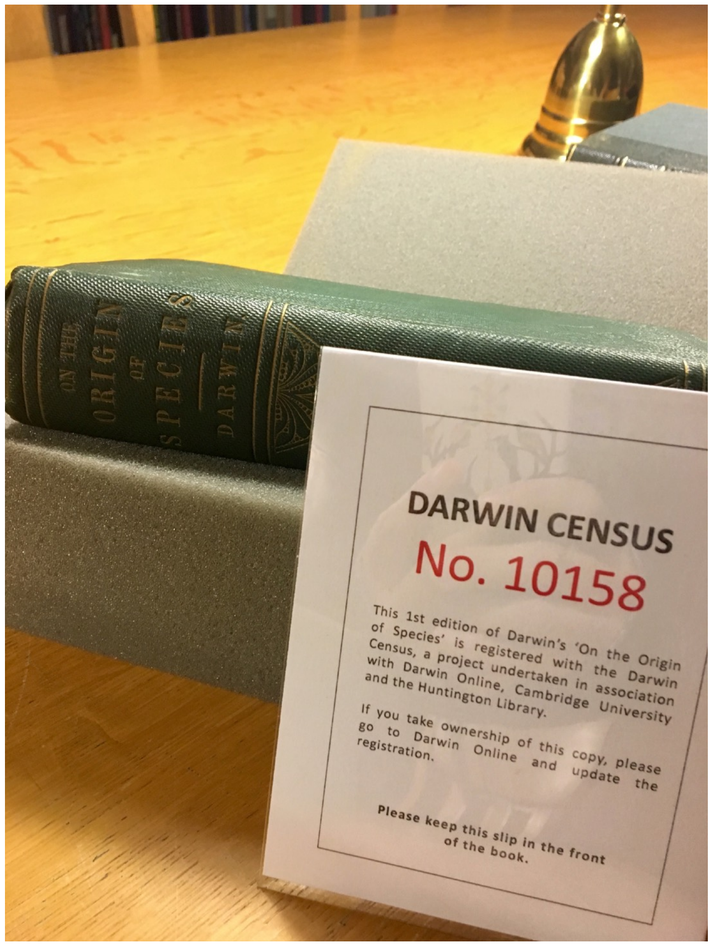
Phylogenies with branch lengths proportional to time provide valuable information about evolutionary history.



CHAP. IV. DIVERGENCE OF CHARACTER. 117

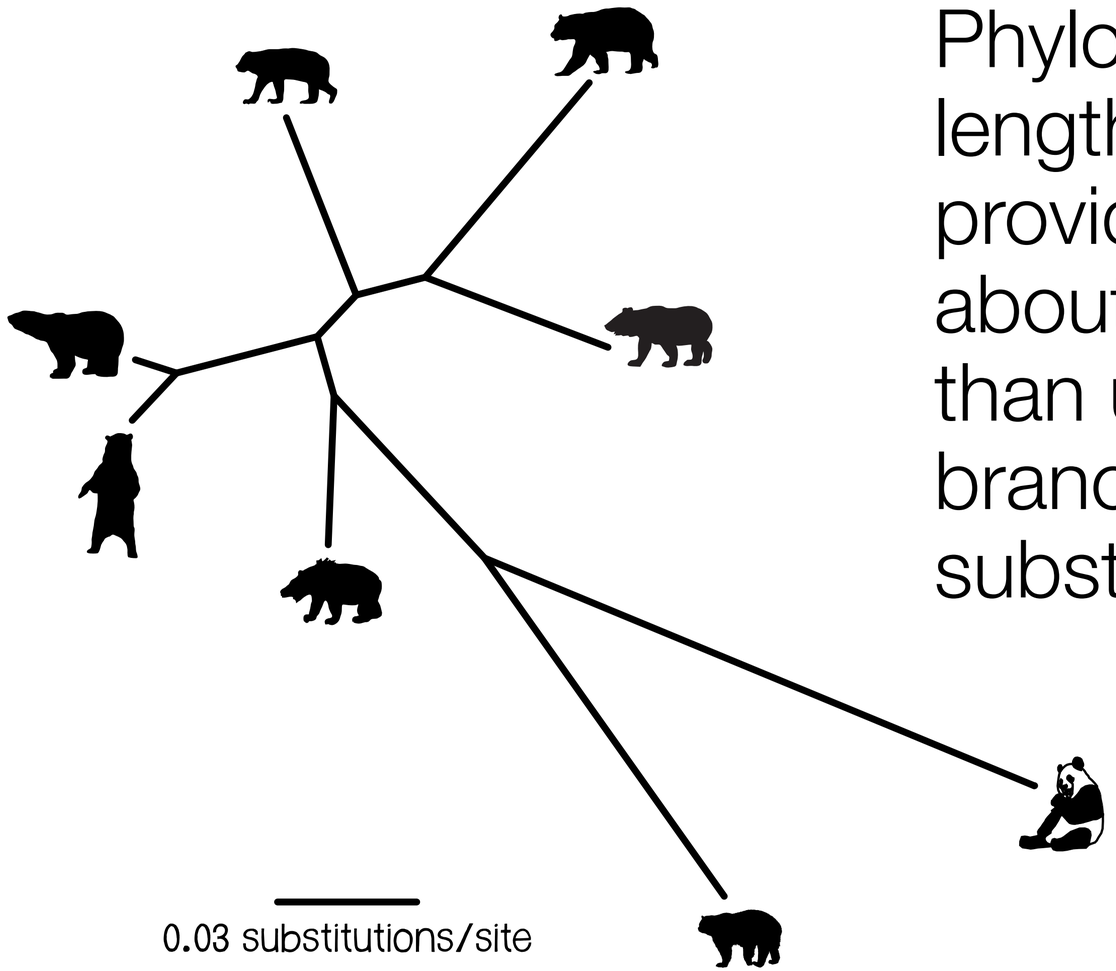
that on an average more of the species of large genera vary than of small genera; and the varying species of the large genera present a greater number of varieties. We have, also, seen that the species, which are the commonest and the most widely-diffused, vary more than rare species with restricted ranges. Let (A) be a common, widely-diffused, and varying species, belonging to a genus large in its own country. The little fan of diverging dotted lines of unequal lengths proceeding from (A), may represent its varying offspring. The variations are supposed to be extremely slight, but of the most diversified nature; they are not supposed all to appear simultaneously, but often after long intervals of time; nor are they all supposed to endure for equal periods. Only those variations which are in some way profitable will be preserved or naturally selected. And here the importance of the principle of benefit being derived from divergence of character comes in; for this will generally lead to the most different or divergent variations (represented by the outer dotted lines) being preserved and accumulated by natural selection. When a dotted line reaches one of the horizontal lines, and is there marked by a small numbered letter, a sufficient amount of variation is supposed to have been accumulated to have formed a fairly well-marked variety, such as would be thought worthy of record in a systematic work.

The intervals between the horizontal lines in the diagram, may represent each a thousand generations; but it would have been better if each had represented ten thousand generations. After a thousand generations, species (A) is supposed to have produced two fairly well-marked varieties, namely a^1 and m^1. These two varieties will generally continue to be exposed to the same conditions which made their parents variable,

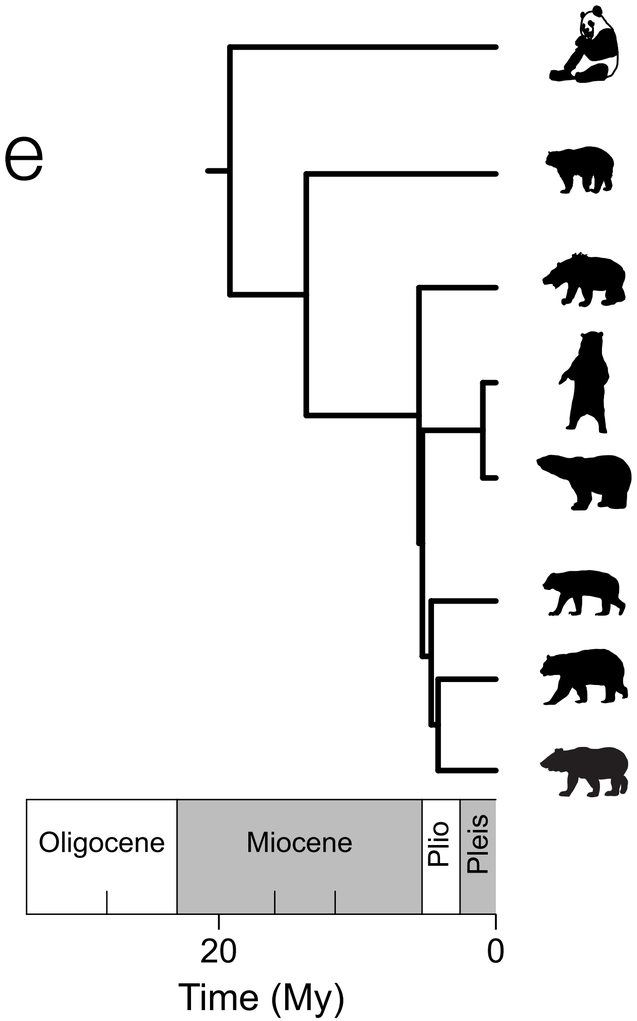


1st edition of Darwin's On the Origin of Species at Grinnell College (Grinnell, Iowa USA)

A Time-Scale for Evolution



Phylogenies with branch lengths proportional to time provide more information about evolutionary history than unrooted trees with branch lengths in units of substitutions/site.

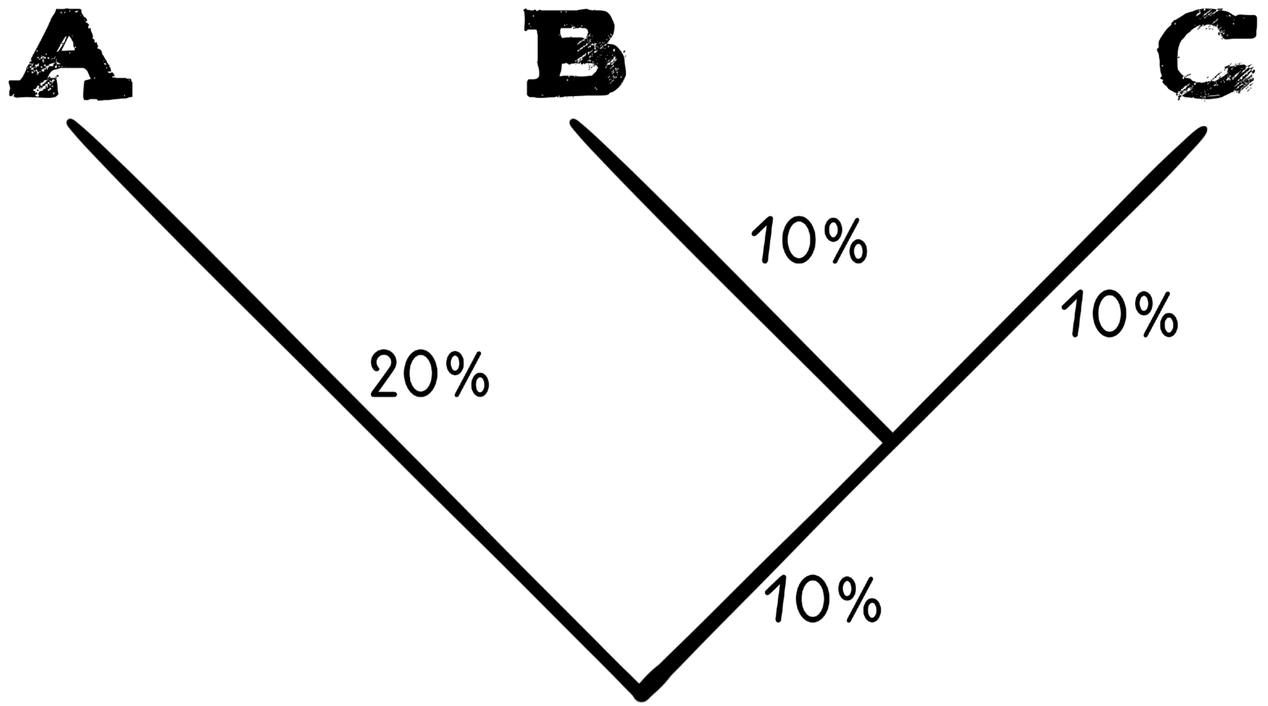


(figure adapted from Heath et al. 2014)

The Global Molecular Clock

Assume that the rate of evolutionary change is constant over time

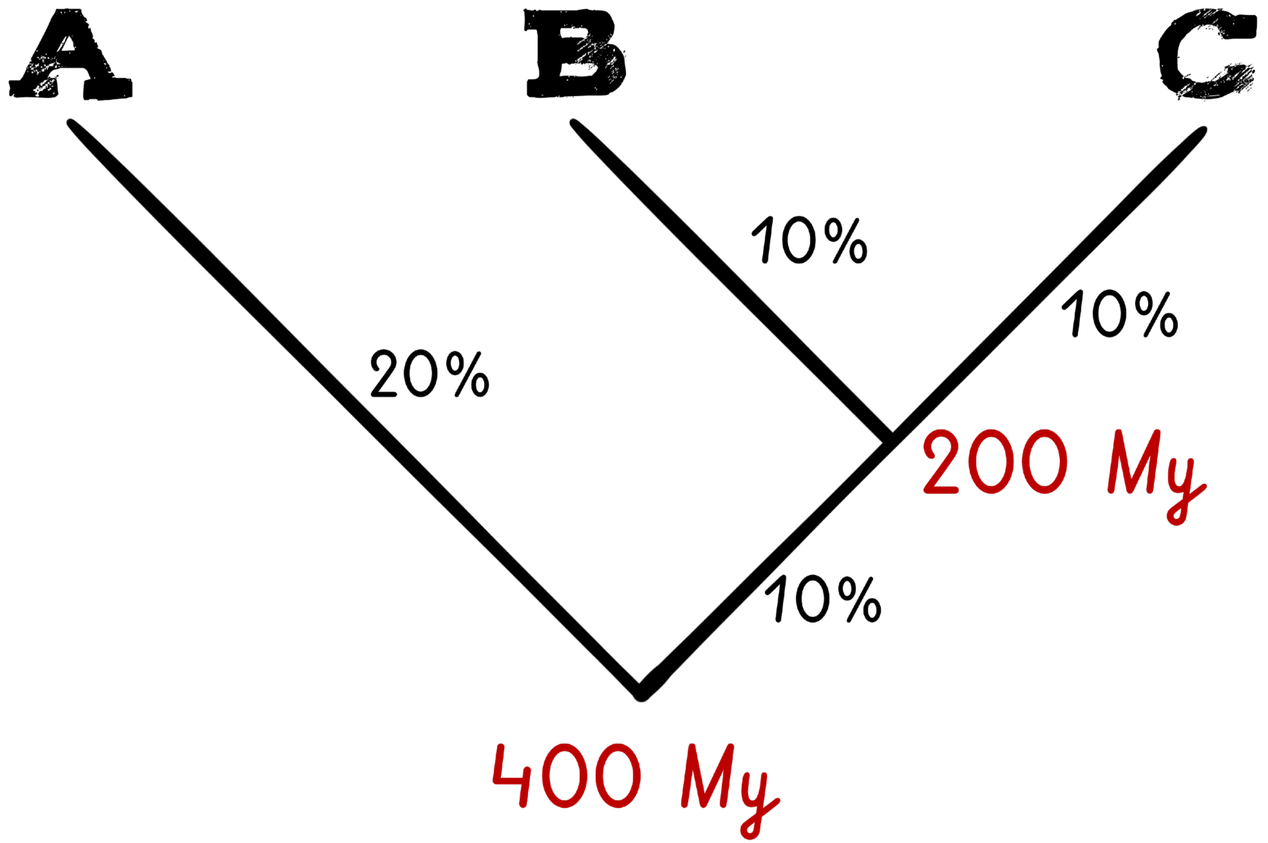
(branch lengths equal percent sequence divergence)



(based on slides by Jeff Thorne http://statgen.ncsu.edu/thorne/compmolevo.html)

The Global Molecular Clock

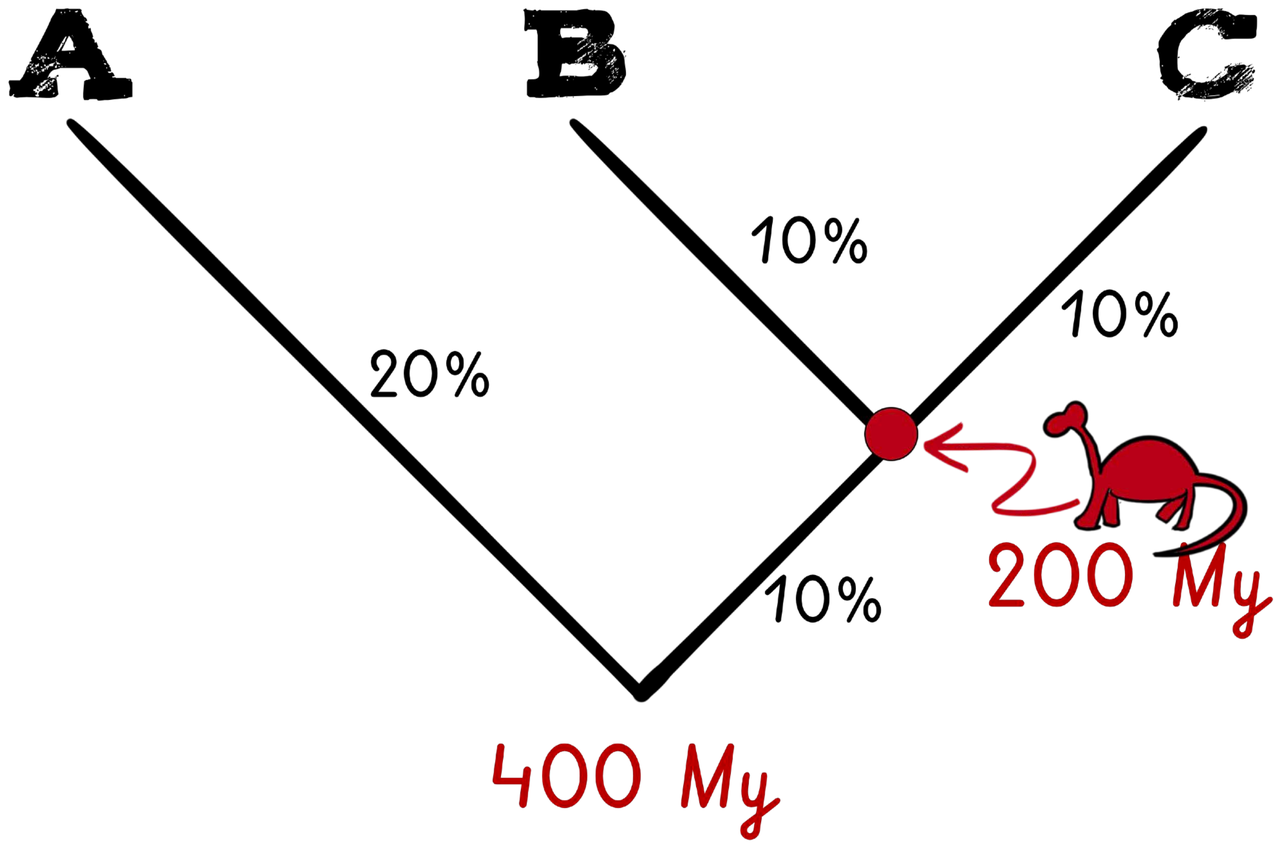
We can date the tree if we know the rate of change is 1% divergence per 10 My



(based on slides by Jeff Thorne http://statgen.ncsu.edu/thorne/compmolevo.html)

The Global Molecular Clock

If we found a fossil of the MRCA of **B** and **C**, we can use it to calculate the rate of change & date the root of the tree



(based on slides by Jeff Thorne http://statgen.ncsu.edu/thorne/compmolevo.html)

Rejecting the Global Molecular Clock

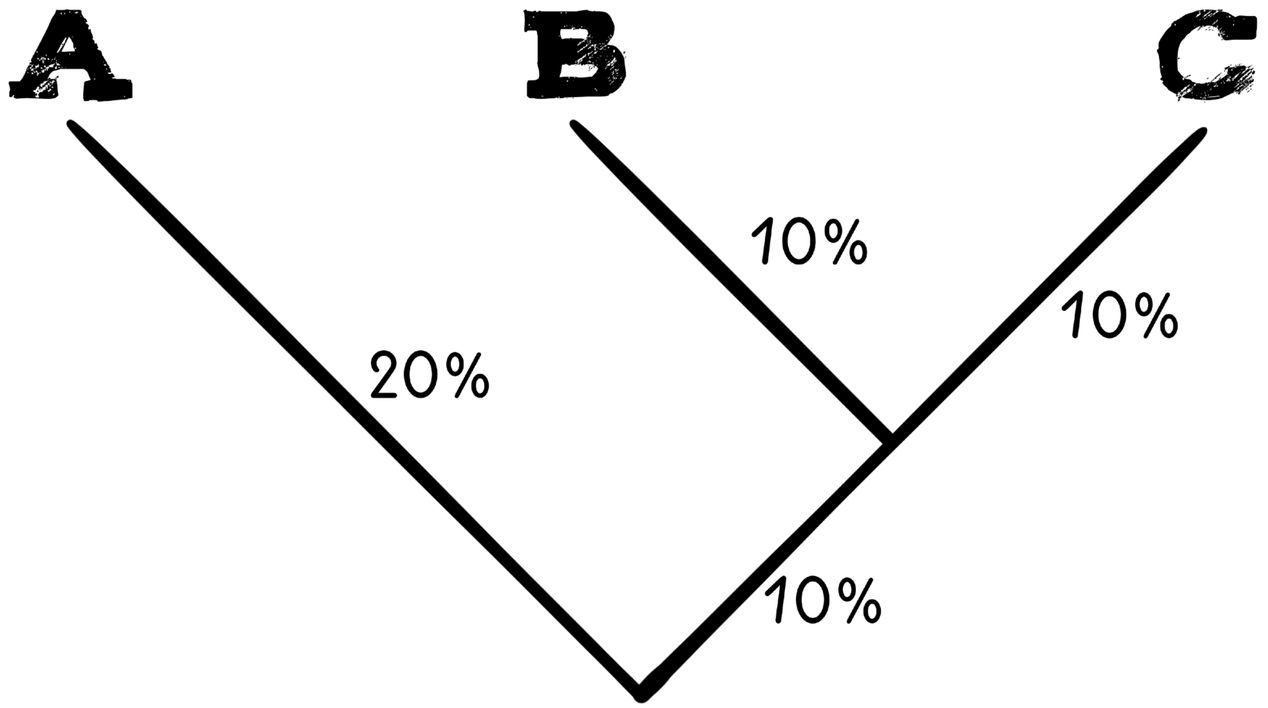
Rates of evolution vary across lineages and over time

Mutation rate

- metabolic rate
- generation time
- DNA repair

Fixation rate

- strengths/targes of selection
- population size

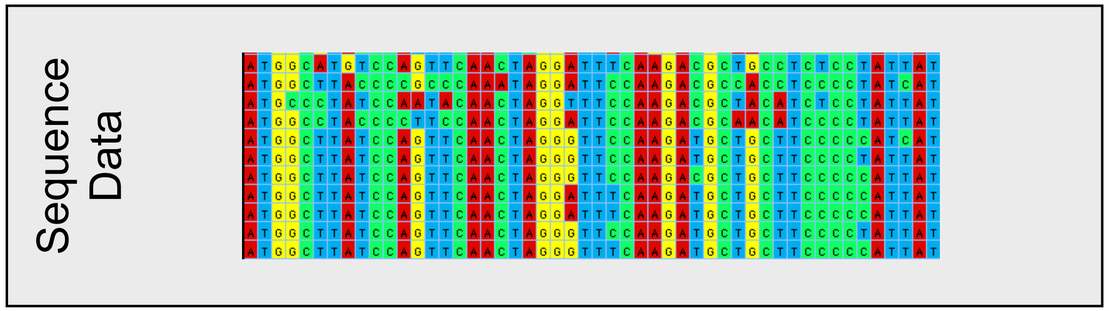


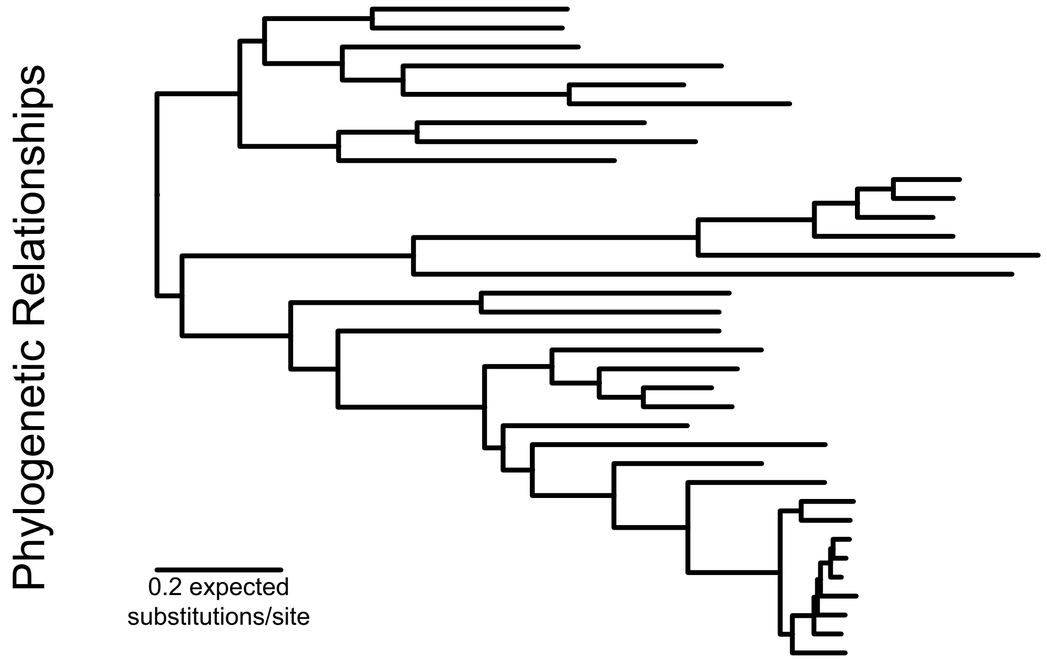
Unconstrained Analysis

Sequence data provide information about **branch lengths**

In units of the **expected # of substitutions per site**

branch length = rate × time

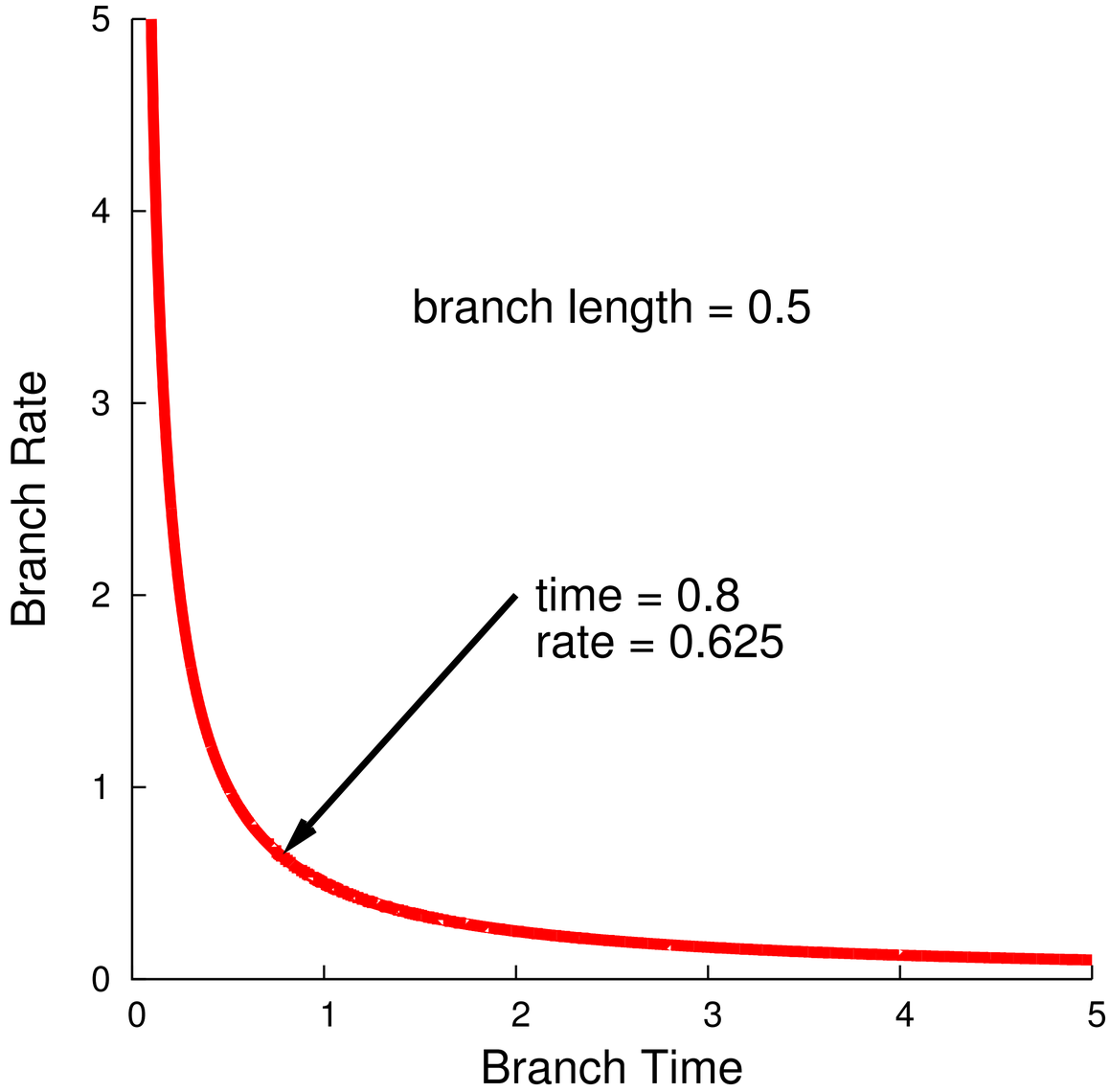




Estimating Rate & Time

The sequence data provide information about branch length

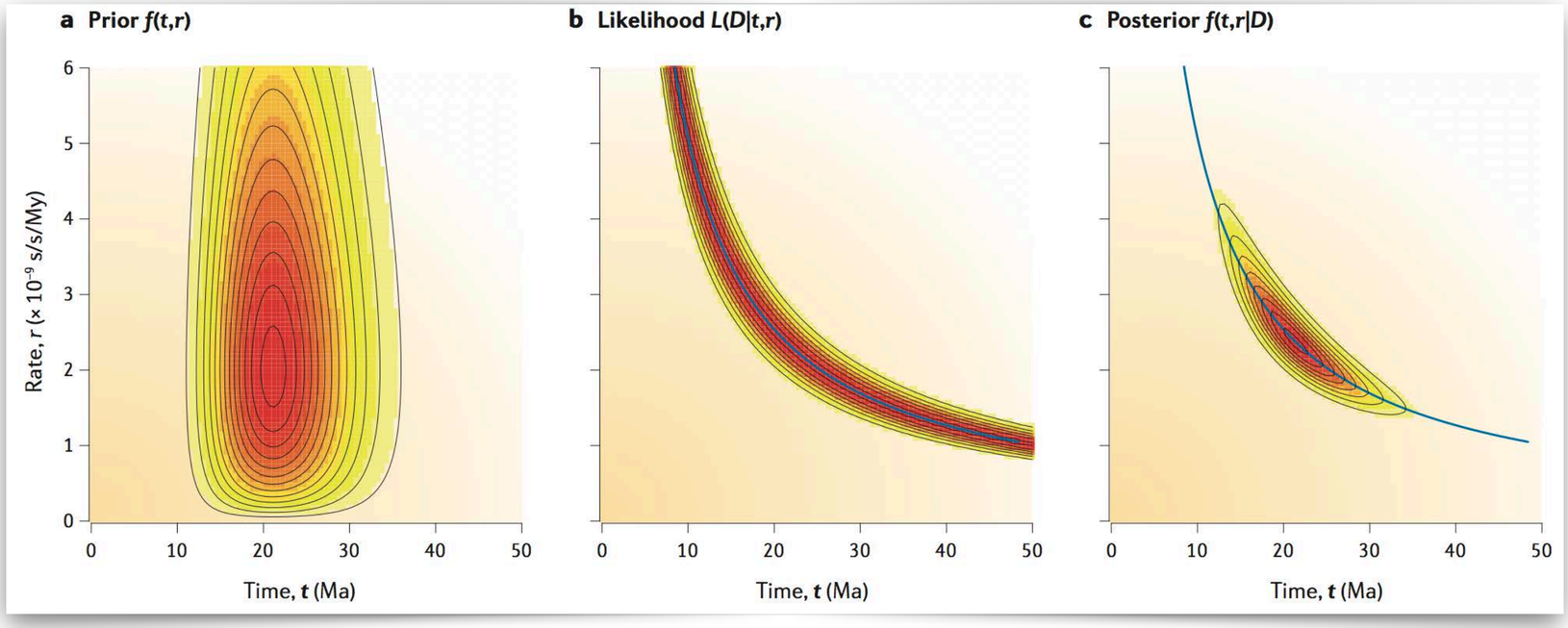
for any possible rate, there's a time that fits the branch length perfectly



(figure based on Thorne & Kishino 2005)

Estimating Rate & Time

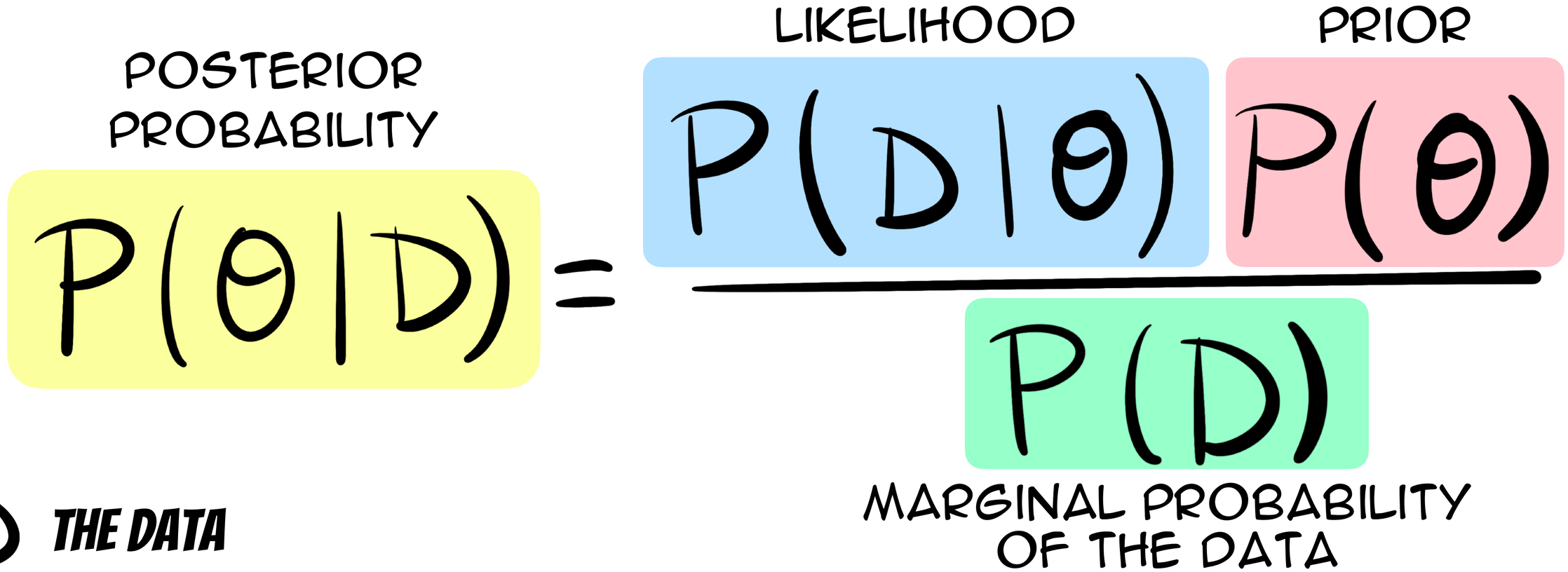
Methods for dating species divergences estimate the **substitution rate** and **time** separately



Tree-time priors for molecular phylogenies are only informative on a **relative** time scale

(figure from dos Reis et al. 2016)

Bayes Theorem

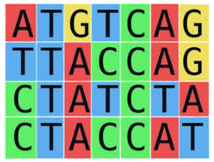


D THE DATA

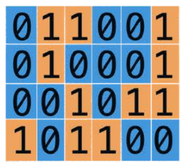
θ THE MODEL

Our Model Components

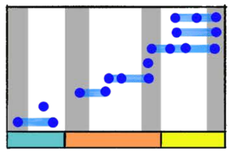
THE DATA



DNA

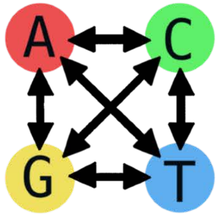


MORPHOLOGY

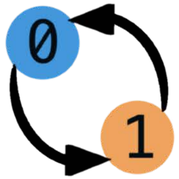


FOSSIL OCCURRENCES

THE MODELS

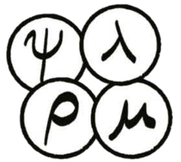


DNA SUBSTITUTION PROCESS

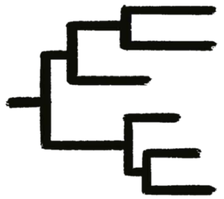


MORPHOLOGICAL CHARACTER CHANGE

BRANCH-SPECIFIC SUBSTITUTION RATES MODEL



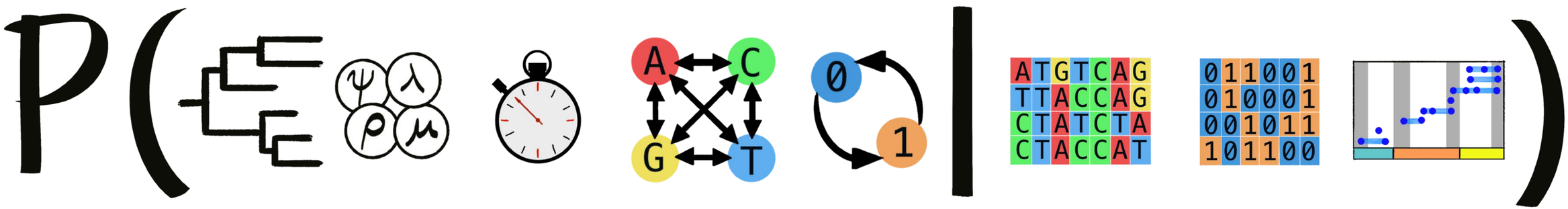
DIVERSIFICATION & SAMPLING PROCESS



TREE TOPOLOGY & BRANCH TIMES

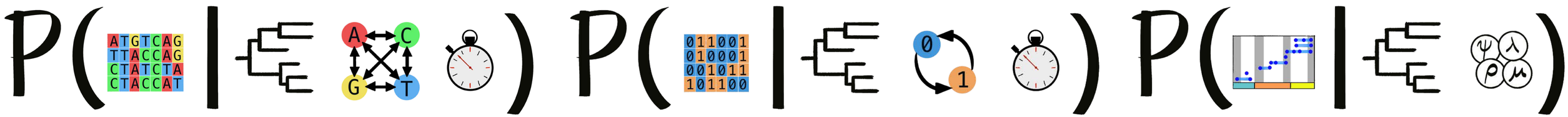
(For an overview, see: Mulvey et al. 2025)

Posterior Probability



(For an overview, see: Mulvey et al. 2025)

Likelihoods



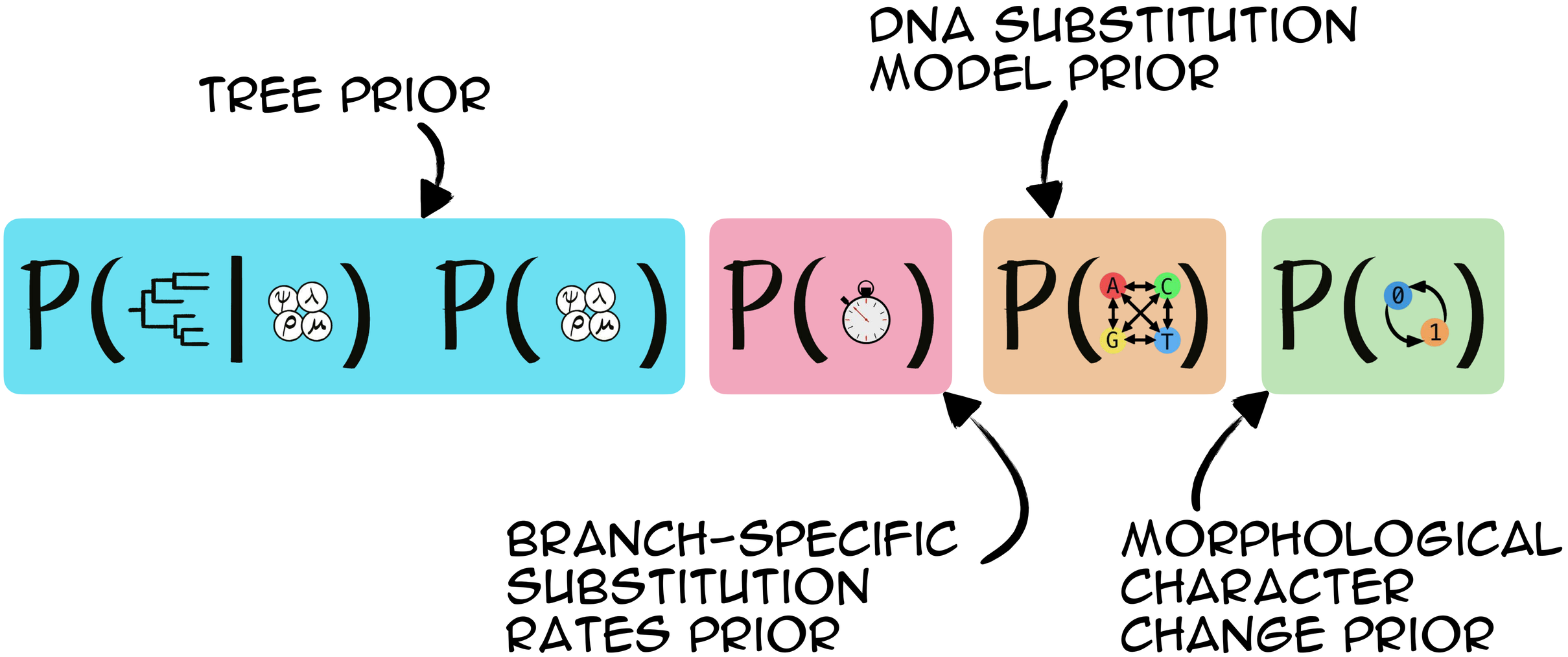
MOLECULAR DATA

MORPHOLOGY

FOSSIL OCCURRENCES

(For an overview, see: Mulvey et al. 2025)

Priors



(For an overview, see: Mulvey et al. 2025)

Bayes Theorem

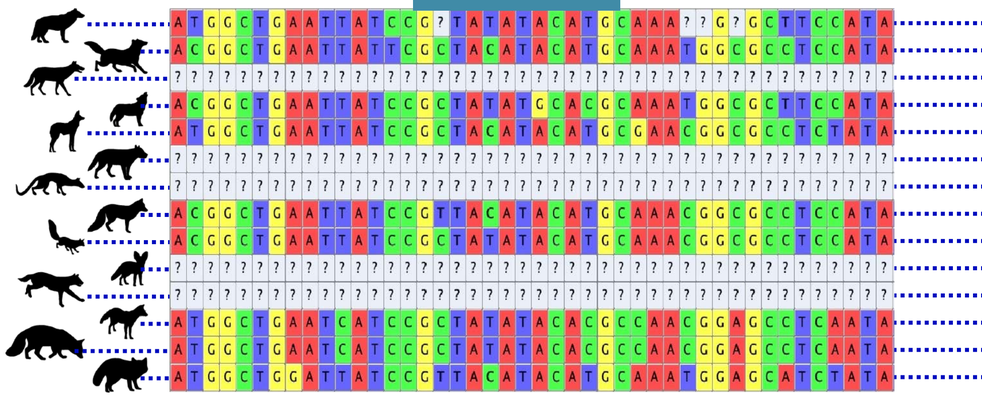
$$P(\text{tree}, \lambda\rho\mu, \text{clock}, \text{Q}, \text{model} \mid \text{data}) =$$

$$\frac{P(\text{data} \mid \text{tree}, \text{Q}, \text{clock})\, P(\text{model} \mid \text{tree}, \text{clock})\, P(\text{clock} \mid \text{tree}, \lambda\rho\mu)\, P(\text{tree} \mid \lambda\rho\mu)\, P(\lambda\rho\mu)\, P(\text{clock})\, P(\text{Q})\, P(\text{rate})}{P(\text{data})}$$

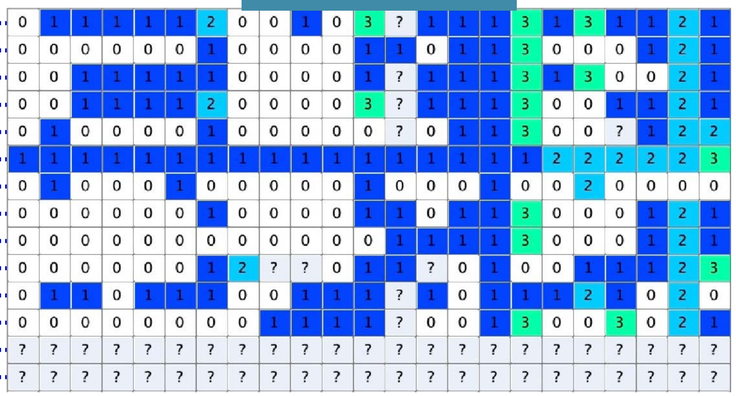
(For an overview, see: Mulvey et al. 2025)

Data Sources Informing Phylogenetic Models

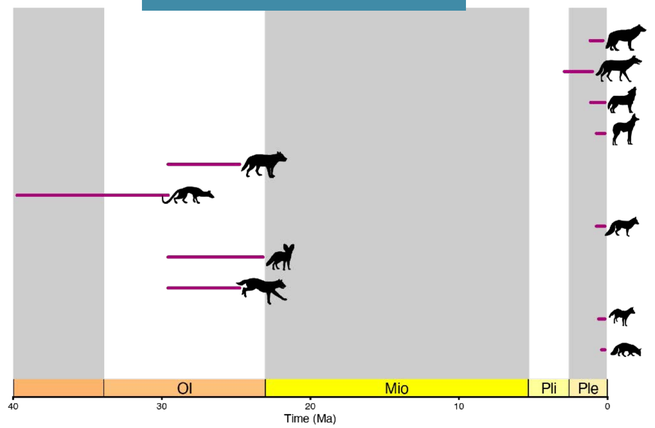
MOLECULAR



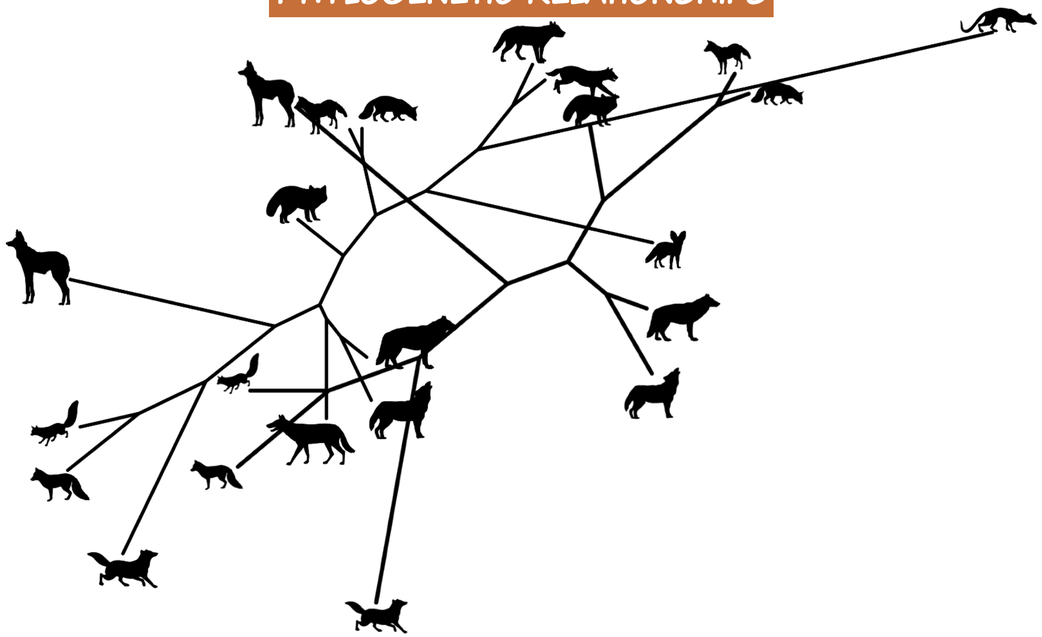
MORPHOLOGICAL



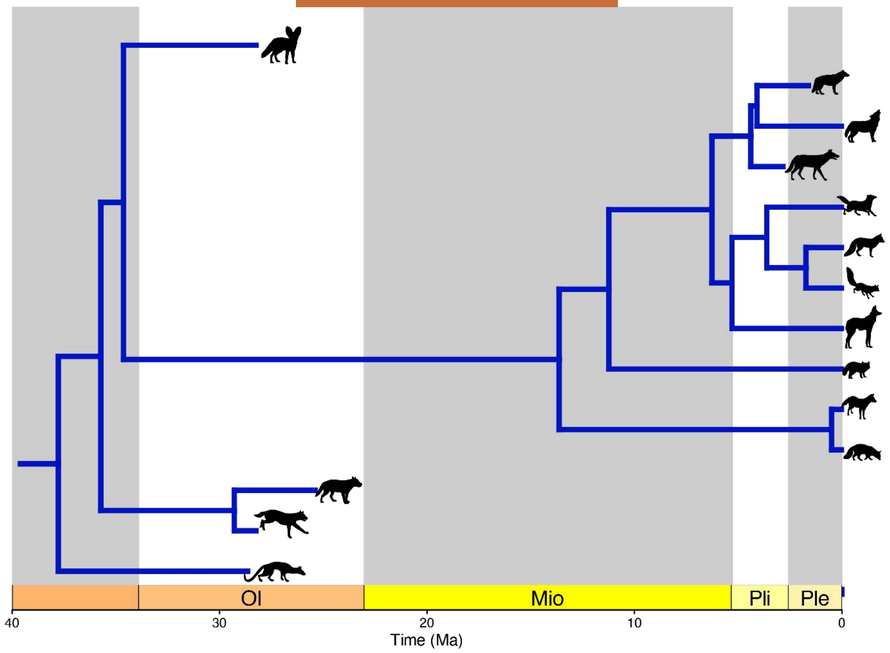
OCCURRENCE TIMES



PHYLOGENETIC RELATIONSHIPS



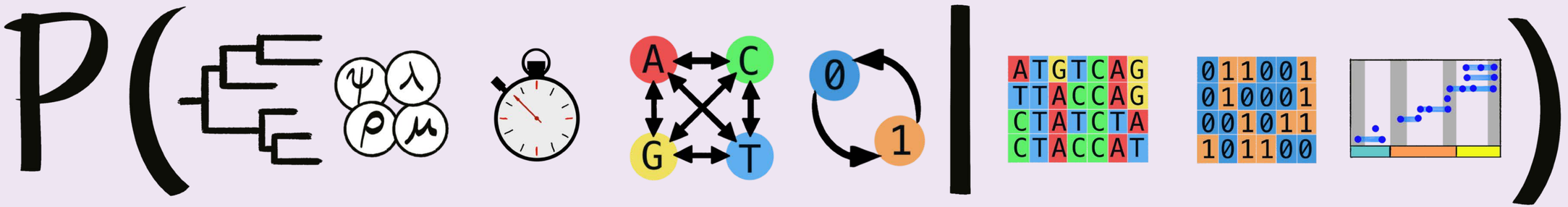
DATED PHYLOGENY



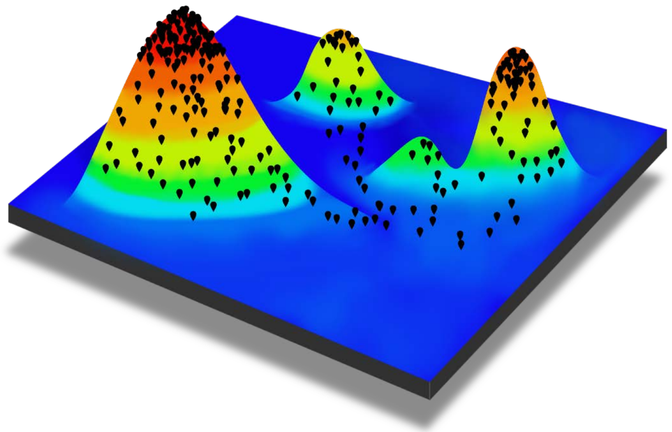
Statistical Inference under Phylogenetic Models

Bayesian inference methods are useful for estimating phylogenetic parameters and testing hypotheses

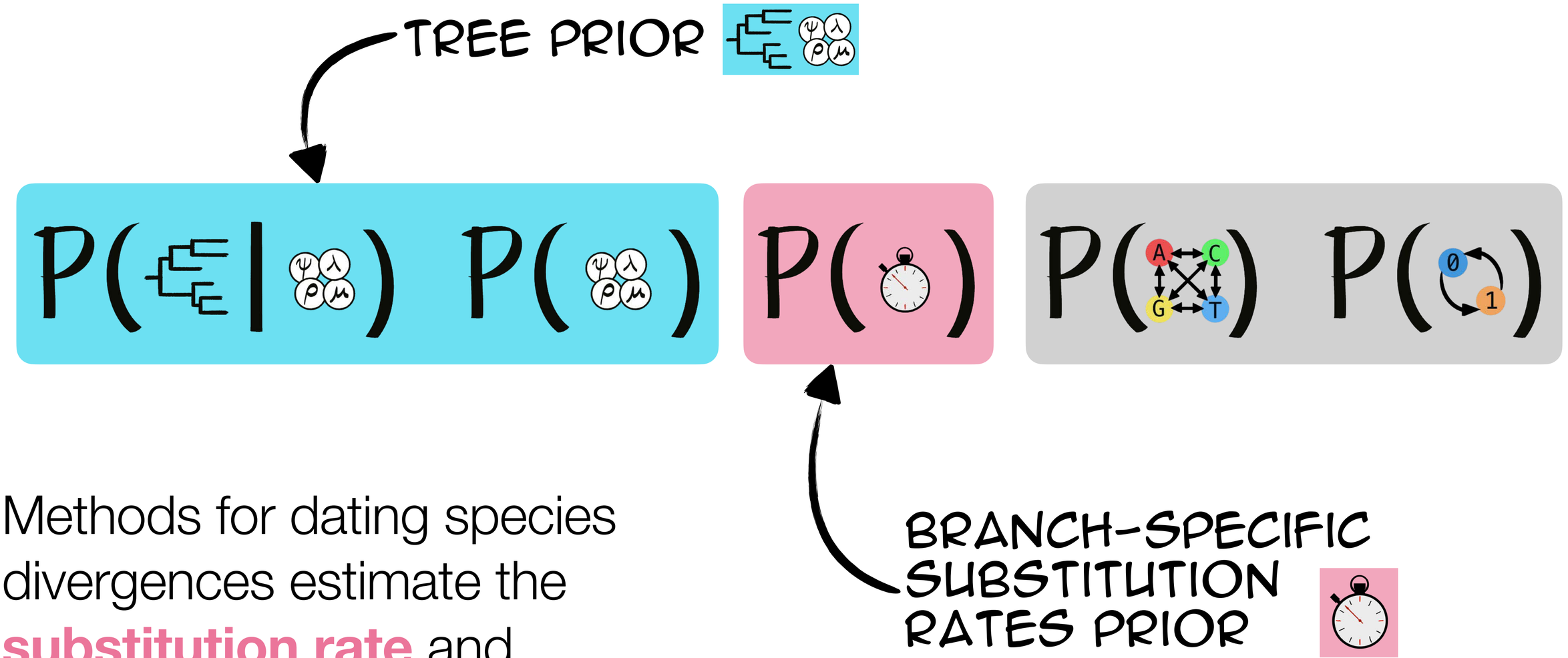
Posterior Probability



We use Markov chain Monte Carlo (MCMC) and other approaches to sample the joint posterior probability density of the model parameters conditional on the data



Essential Priors for Estimating Divergence Times



Methods for dating species divergences estimate the **substitution rate** and **time** separately

(For an overview, see: Mulvey et al. 2025)

Modeling Branch Substitution Rates

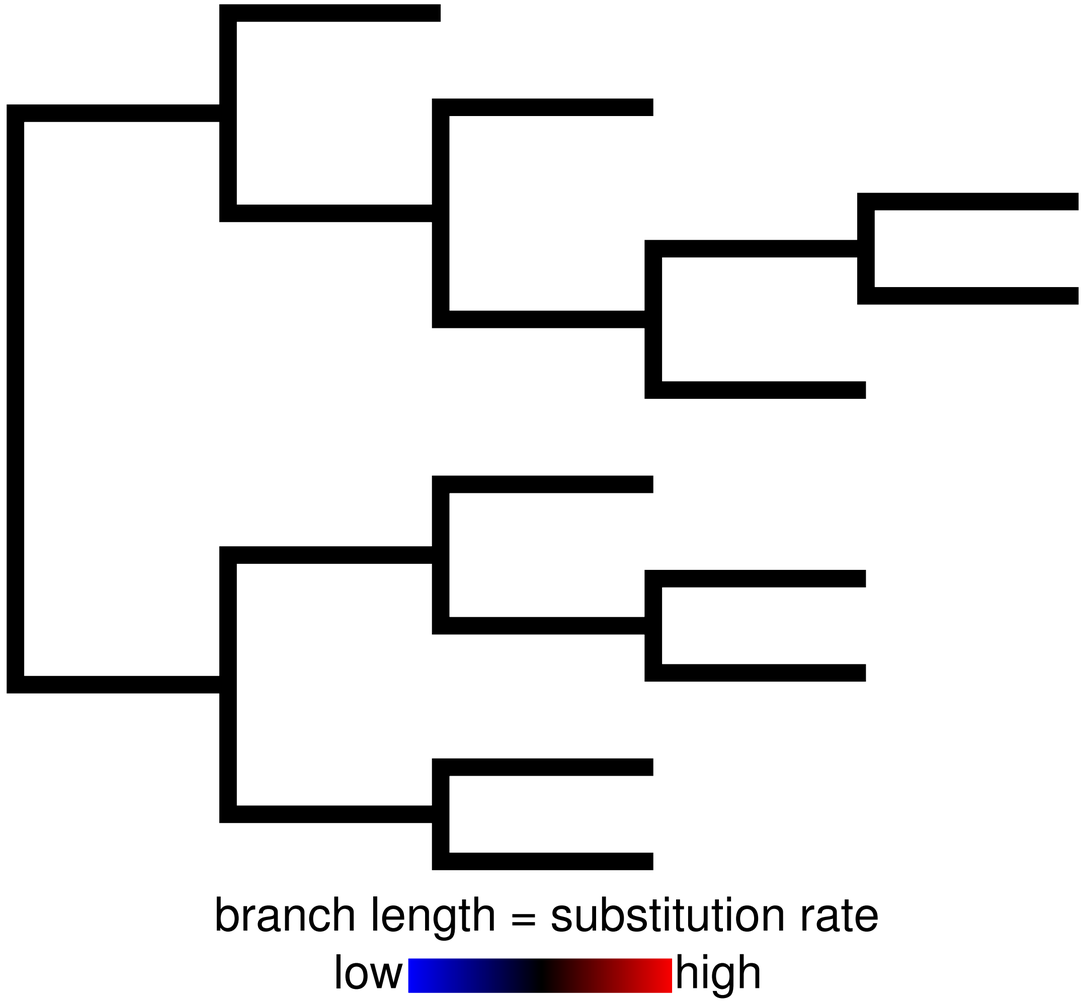
Models describing lineage-specific substitution rate variation:

- Global/strict clock (Zuckerkandl & Pauling, 1962, 1965)
- Local molecular clocks (Hasegawa, Kishino & Yano 1989; Kishino & Hasegawa 1990; Yoder & Yang 2000; Yang & Yoder 2003, Drummond and Suchard 2010)
- Punctuated rate change model (Huelsenbeck, Larget and Swofford 2000)
- Autocorrelated rates (Thorne, Kishino & Painter 1998; Kishino, Thorne & Bruno 2001; Thorne & Kishino 2002; Lepage et al. 2007)
- Time-dependent substitution rates (Membrebe et al. 2019)
- Mixture models on branch rates (Heath, Holder, Huelsenbeck 2012)
- Uncorrelated/independent rates models (Drummond et al. 2006; Rannala & Yang 2007; Lepage et al. 2007)

Global Molecular Clock

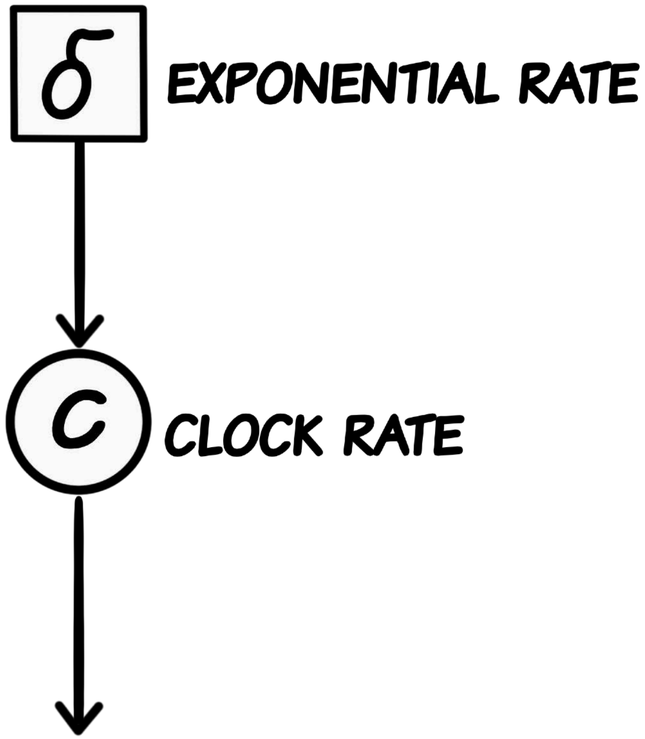
The substitution rate is constant over time

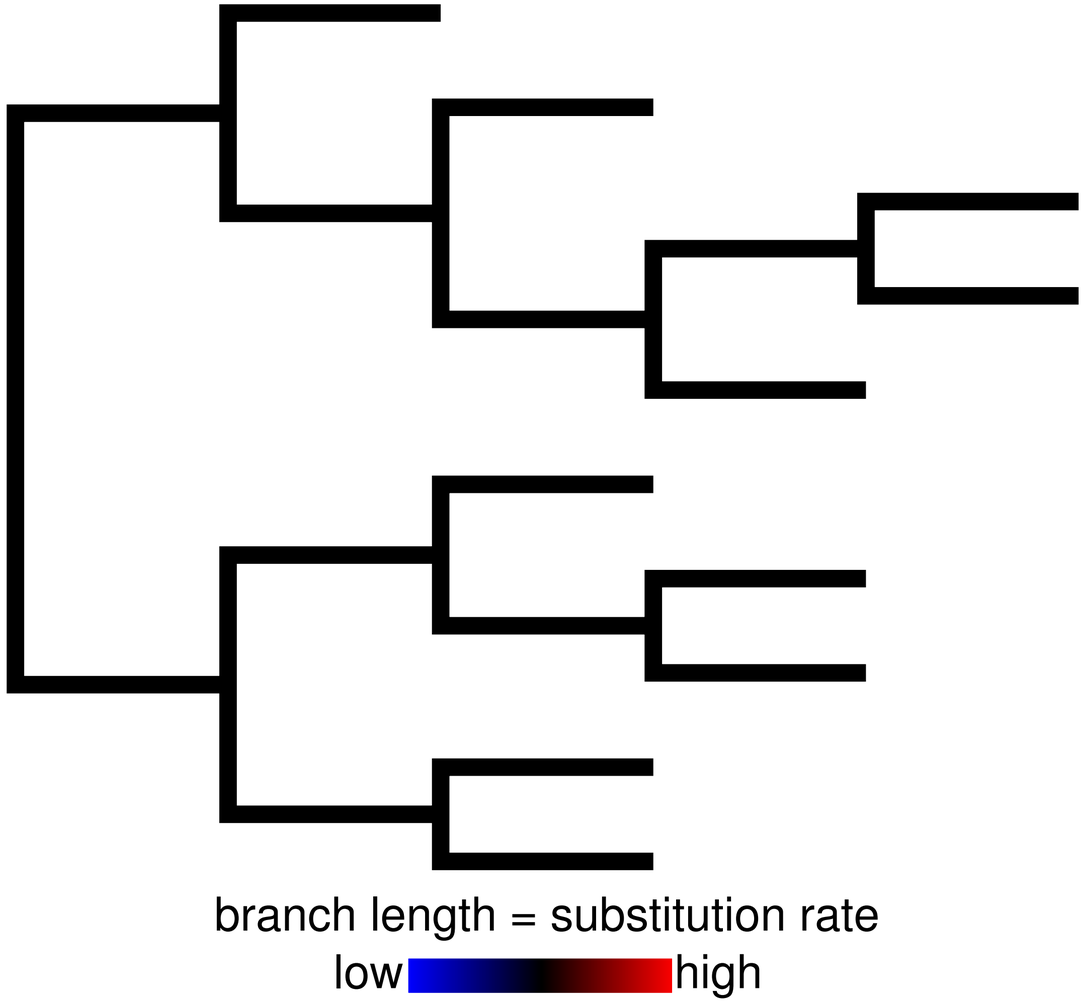
All lineages share the same rate



Global Molecular Clock

$$c \sim \text{Exponential}(\delta)$$

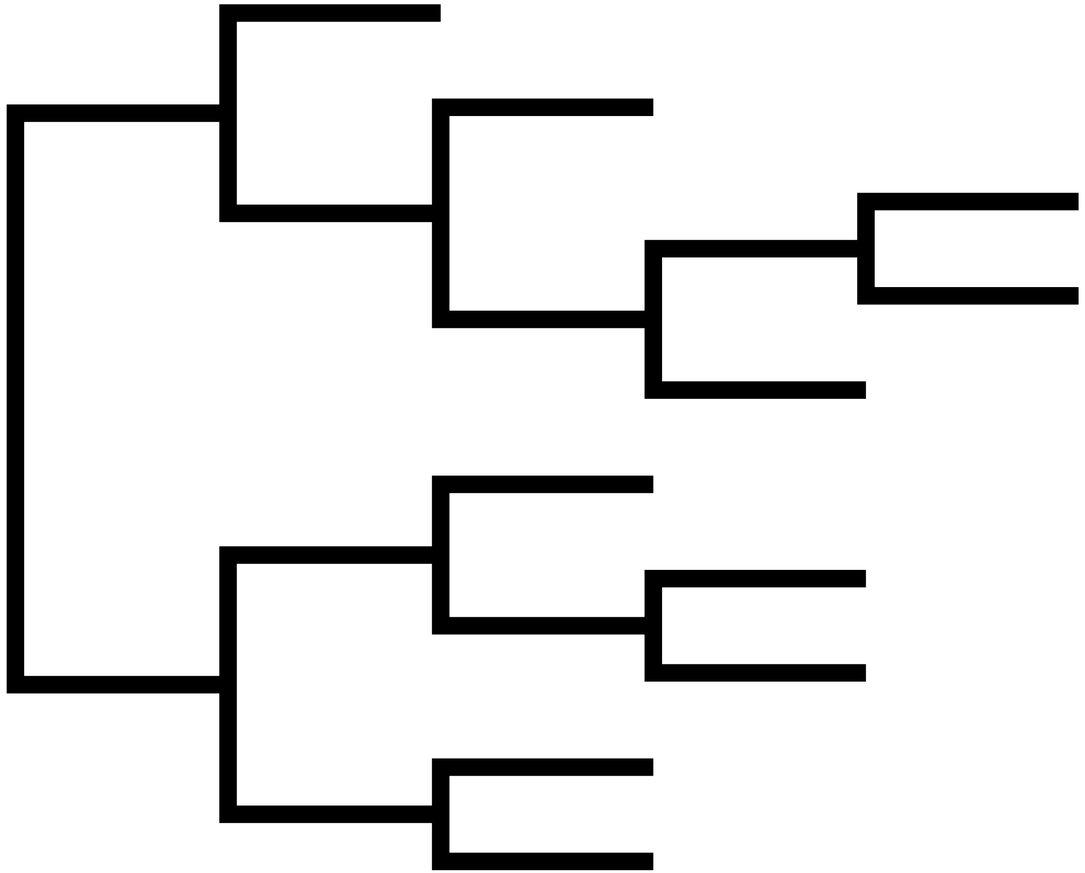




Global Molecular Clock

$$c \sim \text{Gamma}(\alpha, \beta)$$

DRAW THE GRAPHICAL MODEL!



branch length = substitution rate

low high

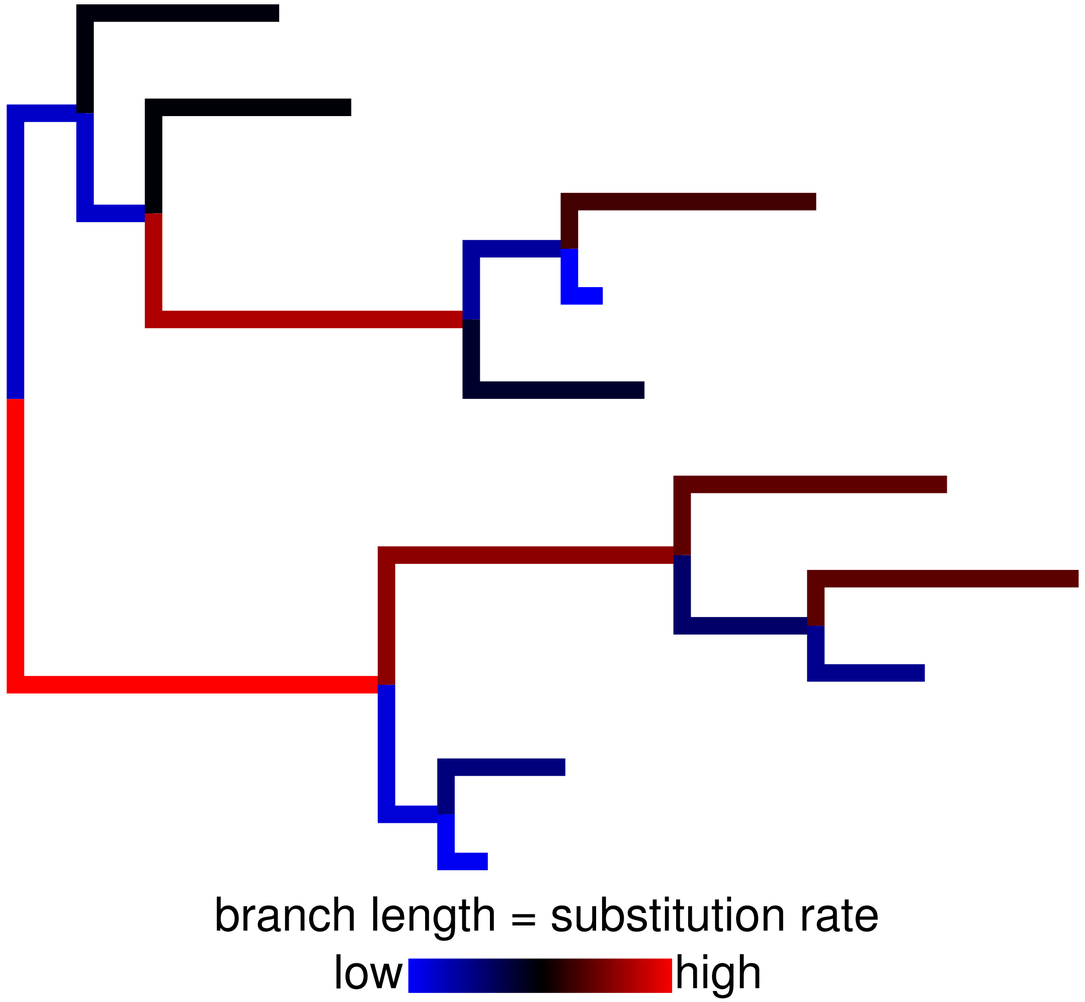
Modeling Rate Variation

To accommodate variation in substitution rates 'relaxed-clock' models estimate lineage-specific substitution rates

- Global/strict clock
- **Local molecular clocks**
- **Punctuated rate change**
- **Autocorrelated rates**
- **Time-dependent substitution rates**
- **Mixture models on branch rates**
- **Uncorrelated/independent rates**

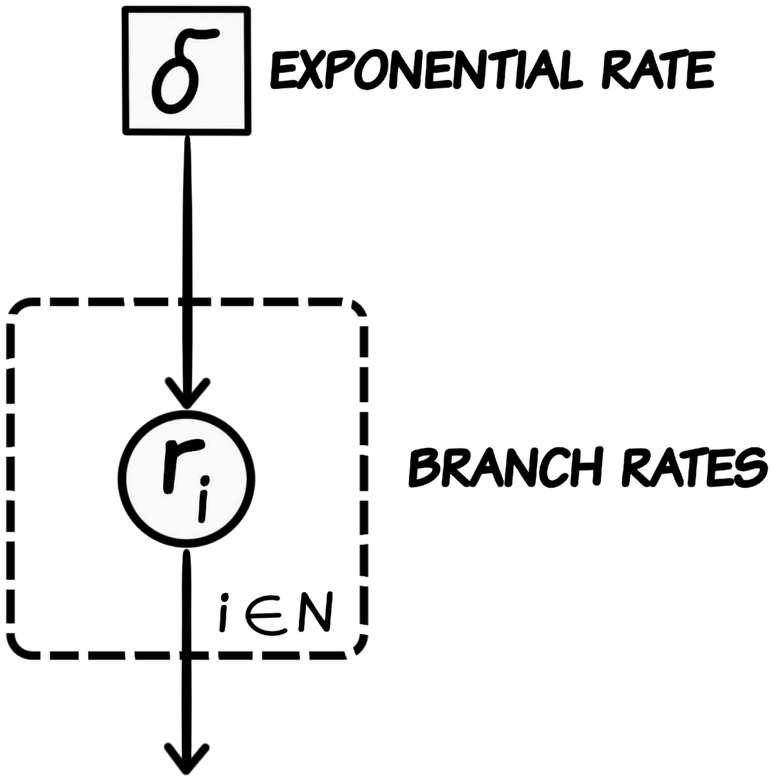
Independent/Uncorrelated Rates

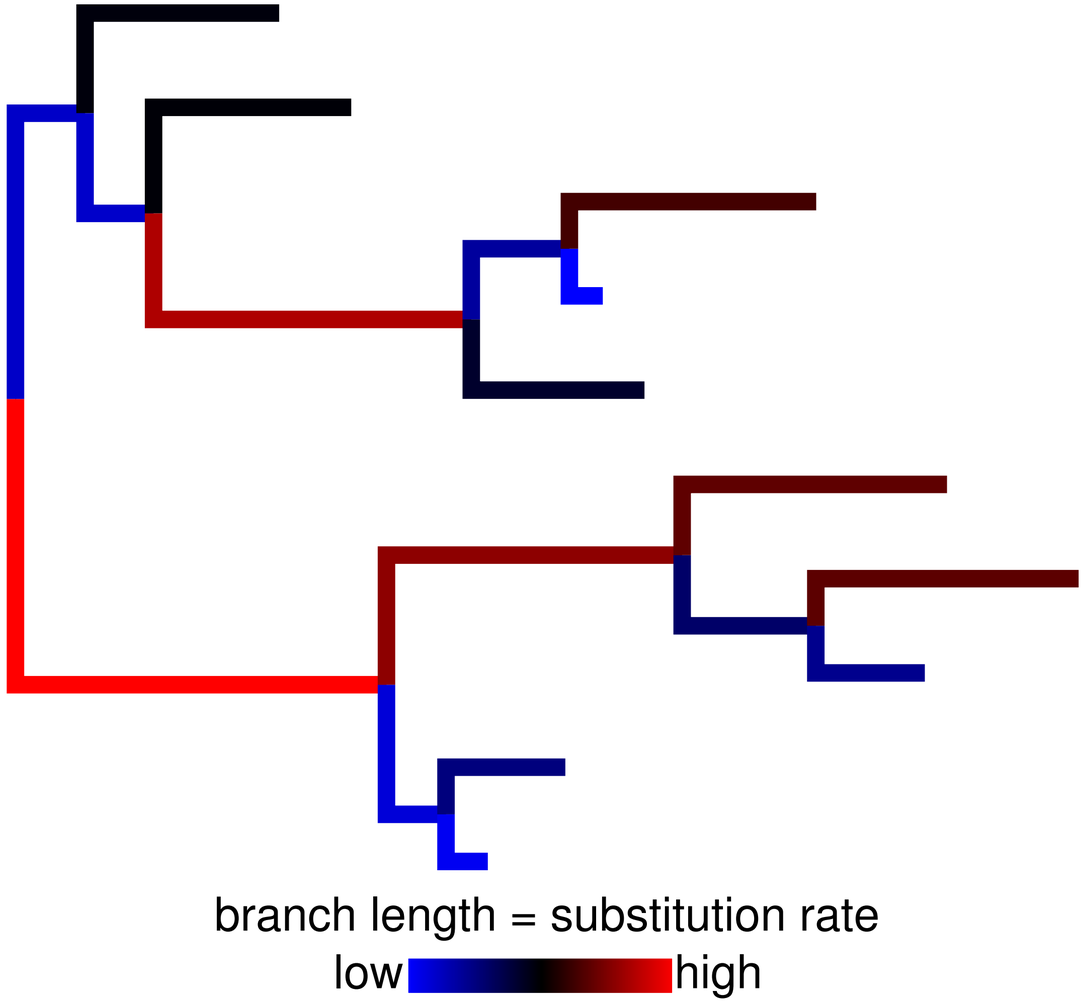
Lineage-specific rates are uncorrelated when the rate assigned to each branch is independently drawn from an underlying distribution



Independent/Uncorrelated Rates

$r_i \sim \text{Exponential}(\delta)$

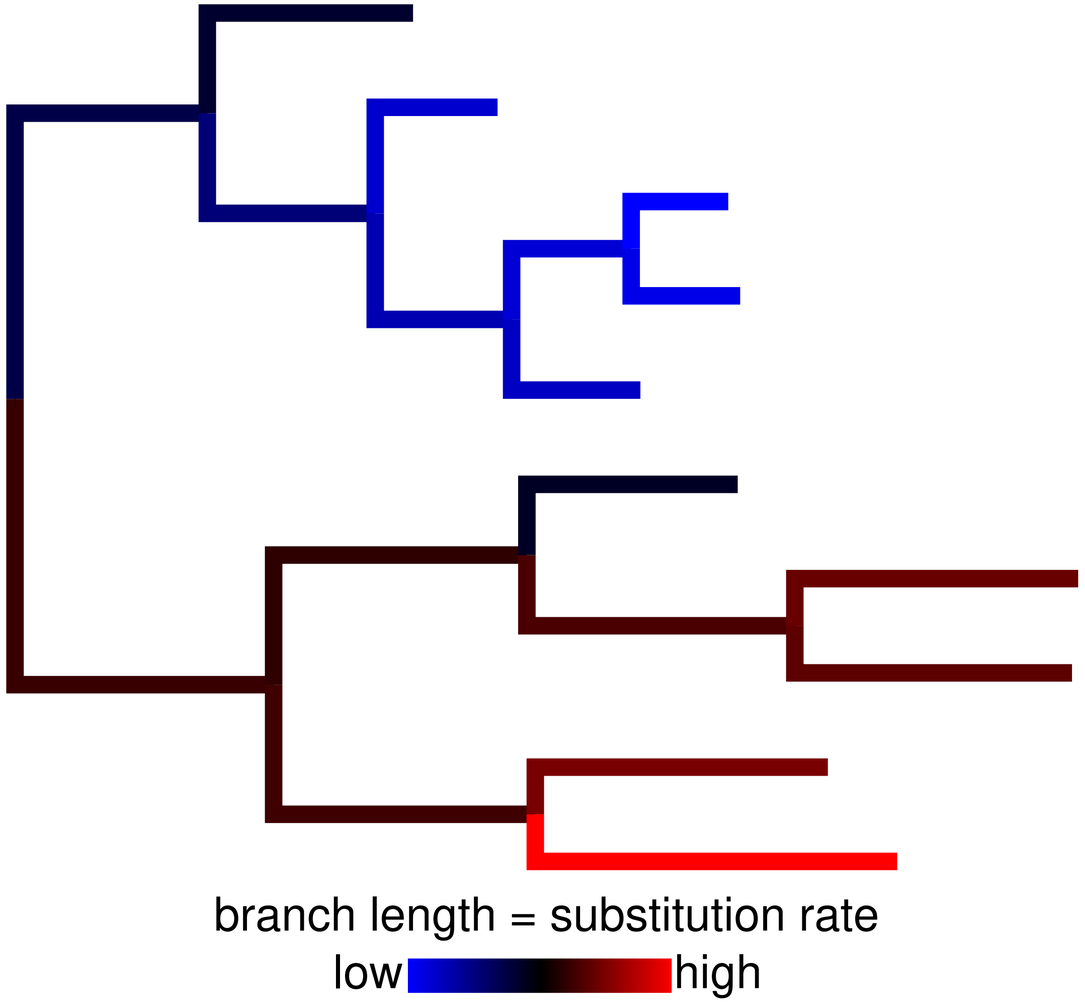




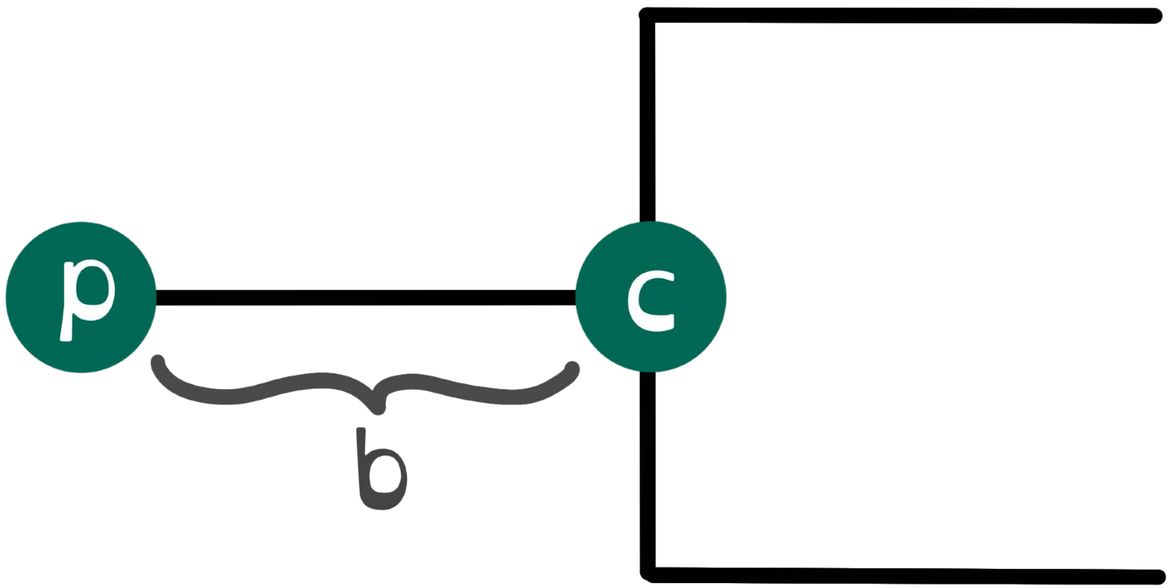
Autocorrelated Rates

Substitution rates evolve gradually over time – closely related lineages have similar rates

The rate at a node is drawn from a distribution with a mean equal to the parent rate



Autocorrelated Rates



$$r_c \sim \text{Lognormal}(\mu_c, \sigma_c)$$

$$\sigma_c := \nu t_b$$

$$\mu_c := \ln(r_p) - \frac{\sigma_c^2}{2}$$

$$r_b := \frac{r_p + r_c}{2}$$

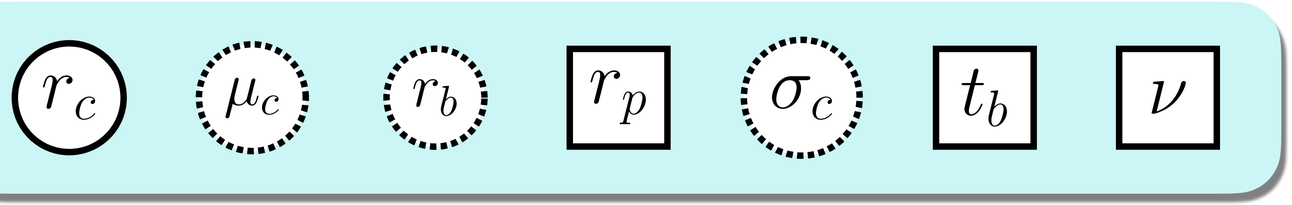
$\nu =$ variance parameter

$t_b =$ time duration of branch

p = parent node

c = child node

b = branch



Autocorrelated Rates

DRAW THE GRAPHICAL MODEL!

$$r_c \sim \text{Lognormal}(\mu_c, \sigma_c)$$

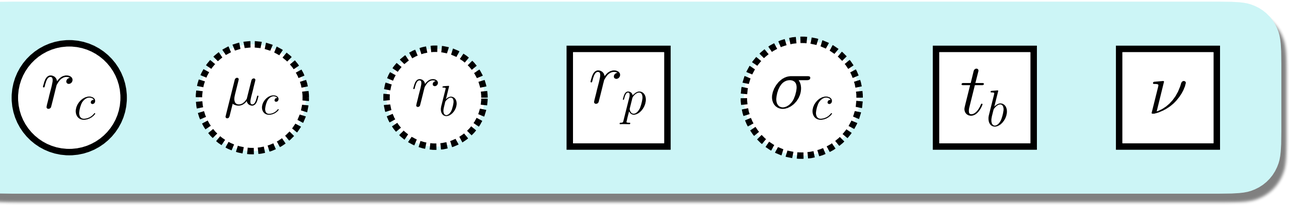
$$\sigma_c := \nu t_b$$

$$\mu_c := \ln(r_p) - \frac{\sigma_c^2}{2}$$

$$r_b := \frac{r_p + r_c}{2}$$

ν = variance parameter

t_b = time duration of branch



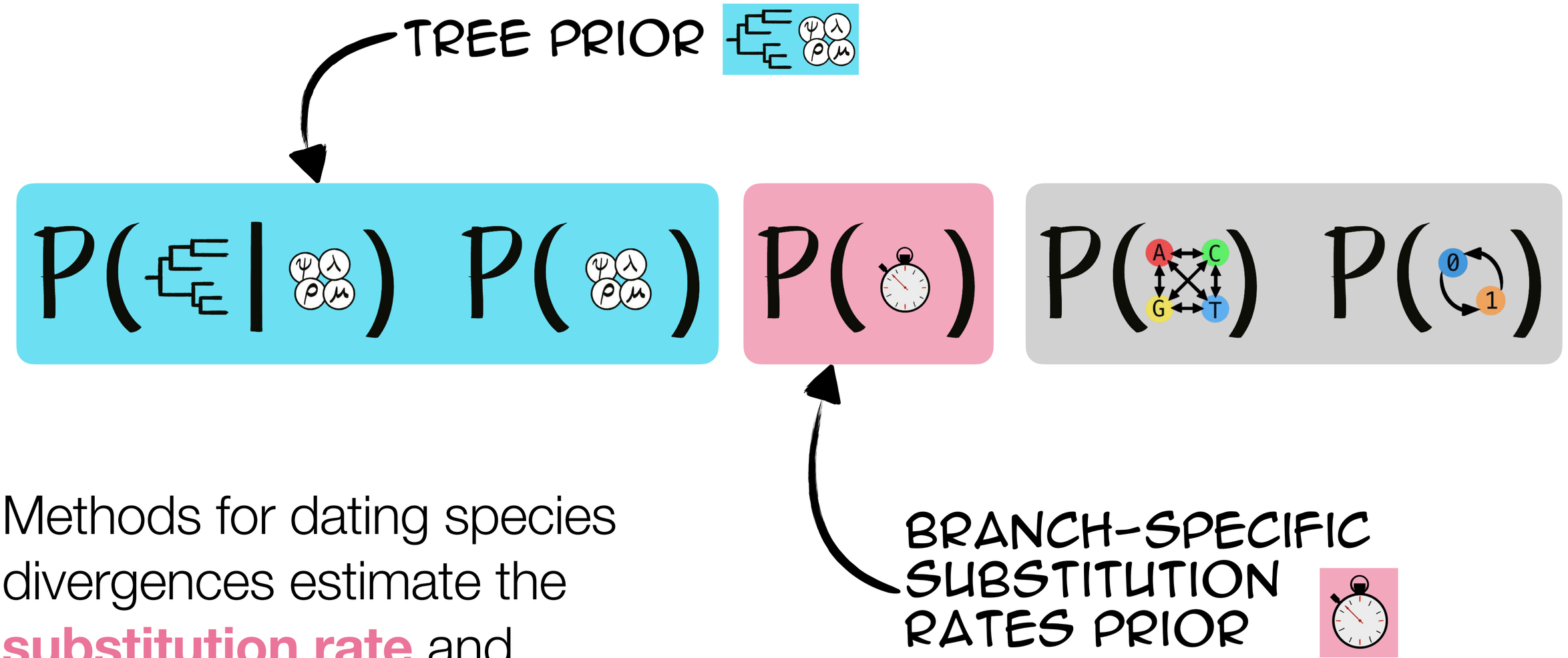
Modeling Rate Variation

These are only a subset of the available models for branch-rate variation

- **Global/strict clock**
- **Local molecular clocks**
- **Punctuated rate change**
- **Autocorrelated rates**
- **Time-dependent substitution rates**
- **Mixture models on branch rates**
- **Uncorrelated/independent rates**

Considering model selection, uncertainty, & plausibility is very important for Bayesian divergence time analysis

Essential Priors for Estimating Divergence Times

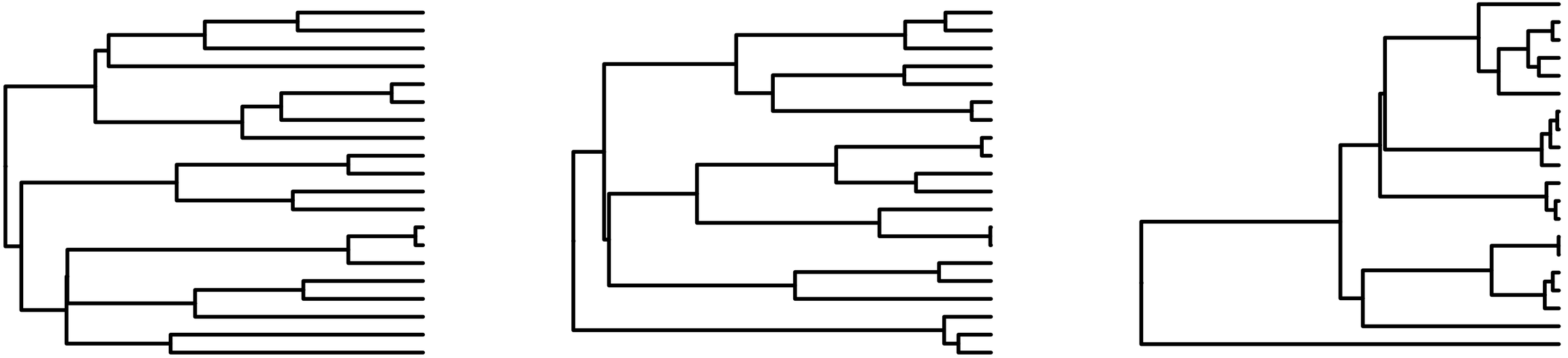


Methods for dating species divergences estimate the **substitution rate** and **time** separately

(For an overview, see: Mulvey et al. 2025)

Priors on the Tree & Times

Relaxed clock Bayesian analyses require a prior distribution on time trees



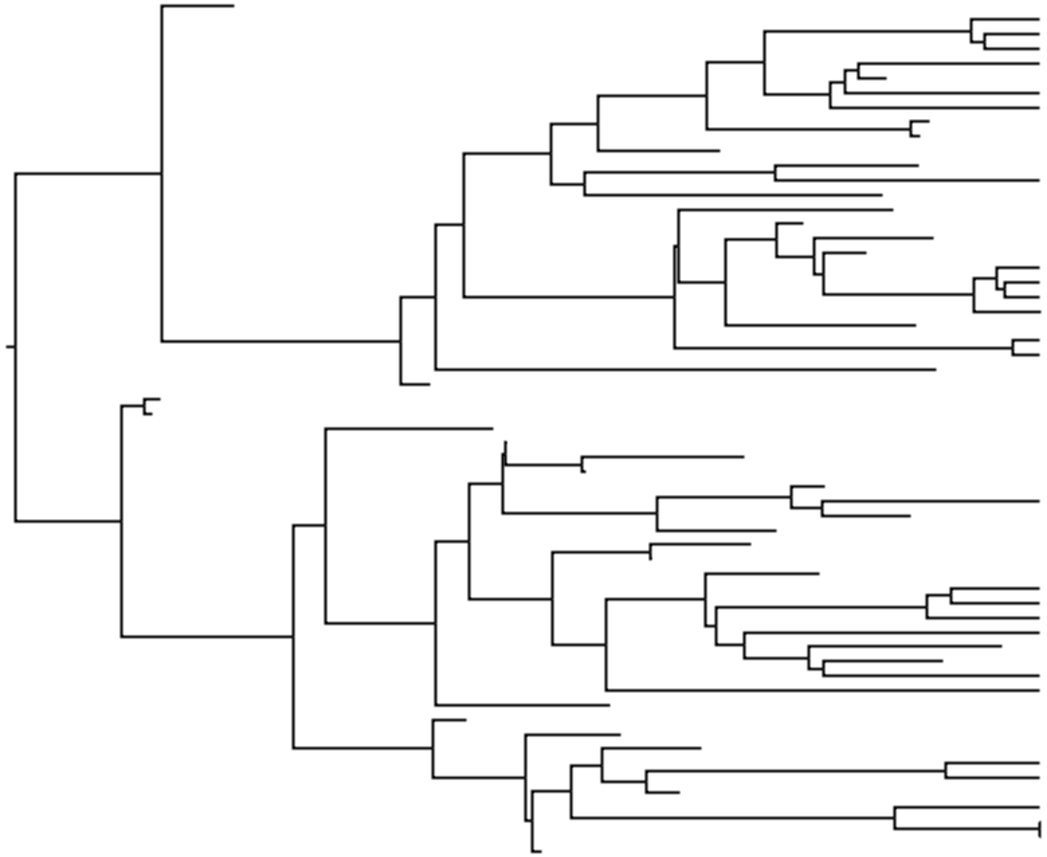
Different tree priors make different assumptions about the timing of divergence events and shape of the tree topology

The Birth-Death Process

A stochastic branching process where new lineages arise at rate λ = births and go extinct at rate μ = deaths (Kendall 1948)

λ = speciation

μ = extinction



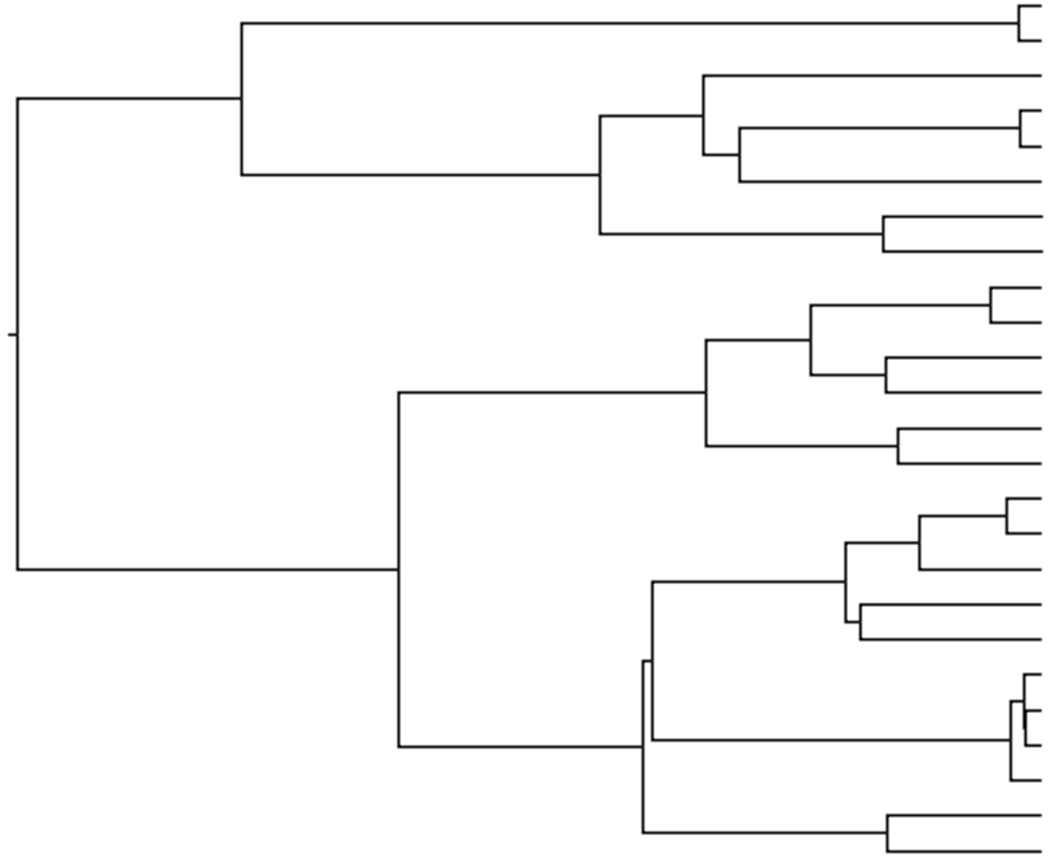
Most commonly used tree priors are based on the birth-death process

The Yule Process

A stochastic branching process where new lineages arise at rate λ = births and go extinct at rate μ = deaths (Kendall 1948)

λ = speciation

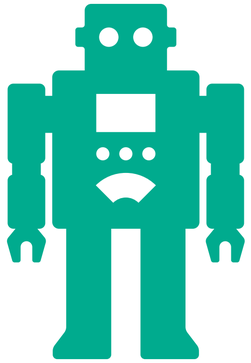
μ = extinction **= 0**



Most commonly used tree priors are based on the birth-death process

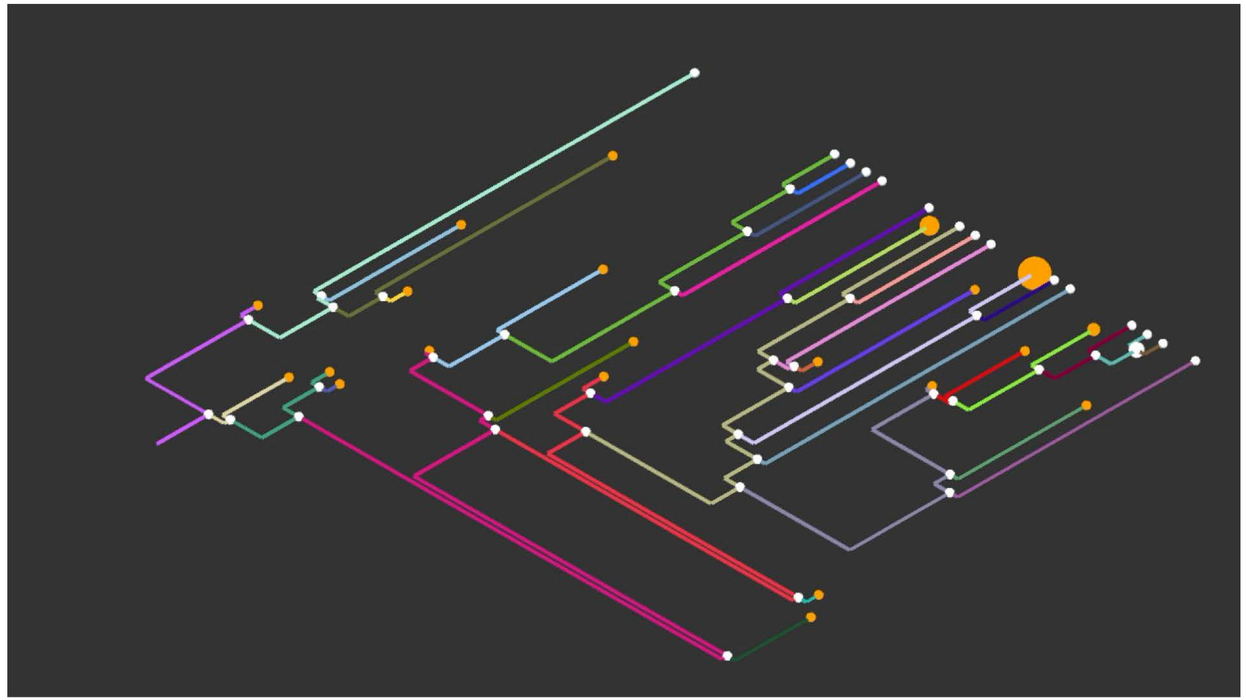
The Birth-Death Process

A stochastic branching process where new lineages arise at rate λ = births and go extinct at rate μ = deaths (Kendall 1948)



λ = speciation

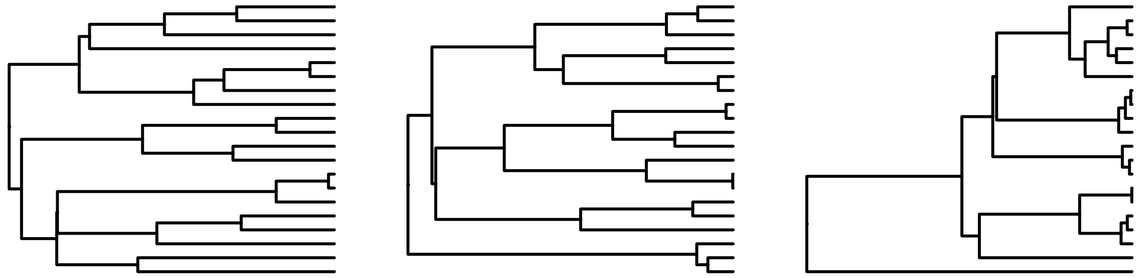
μ = extinction



Most commonly used tree priors are based on the birth-death process

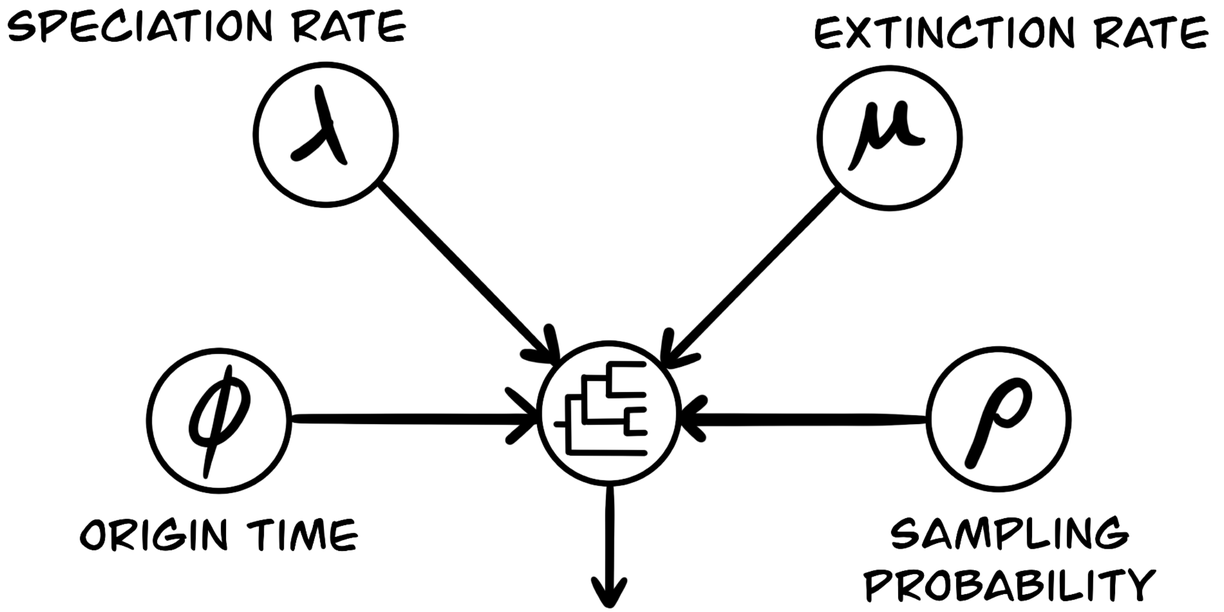
The Birth-Death Process

Tree priors based on stochastic models of lineage diversification



Birth-death-sampling process:
at any point in time a lineage can speciate at rate λ or go extinct with a rate of μ

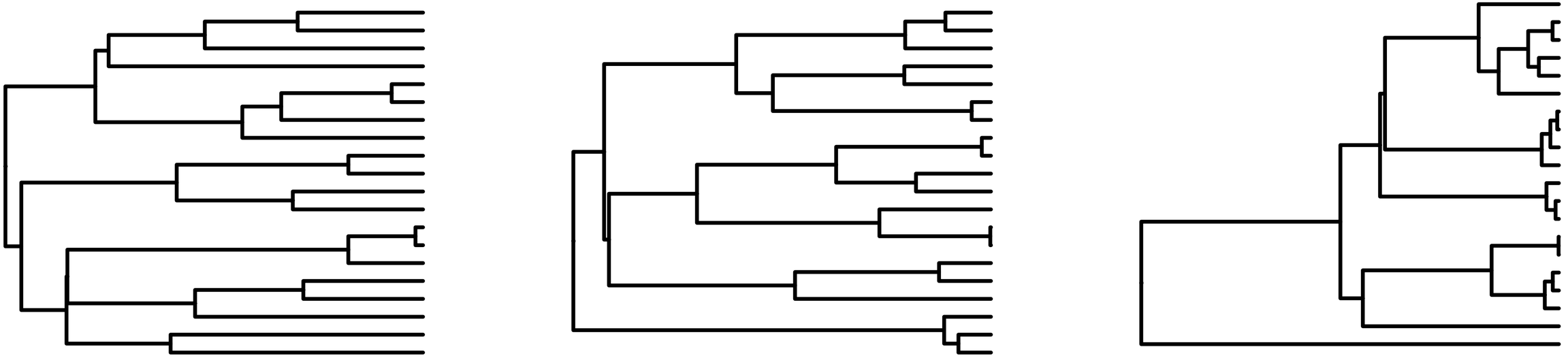
Conditions on a probability of sampling a tip ρ, and the origin time of the process ϕ



Priors on the Tree & Times

Sequence data are only informative on relative rates & times

Most tree priors cannot give precise estimates of absolute node ages



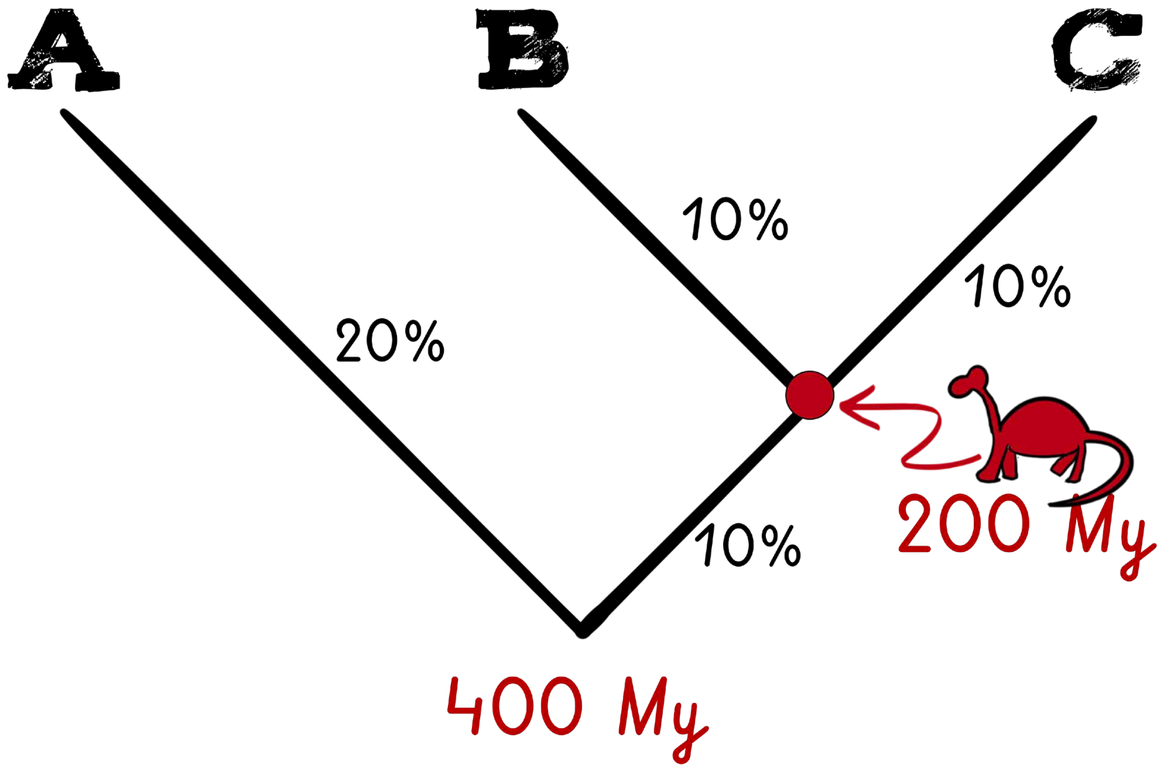
We need additional data (like fossils) to provide an absolute time scale

Dating Nodes

Fossils (or other data) are necessary to estimate ***absolute*** node ages

There is **no information** in the sequence data for absolute time

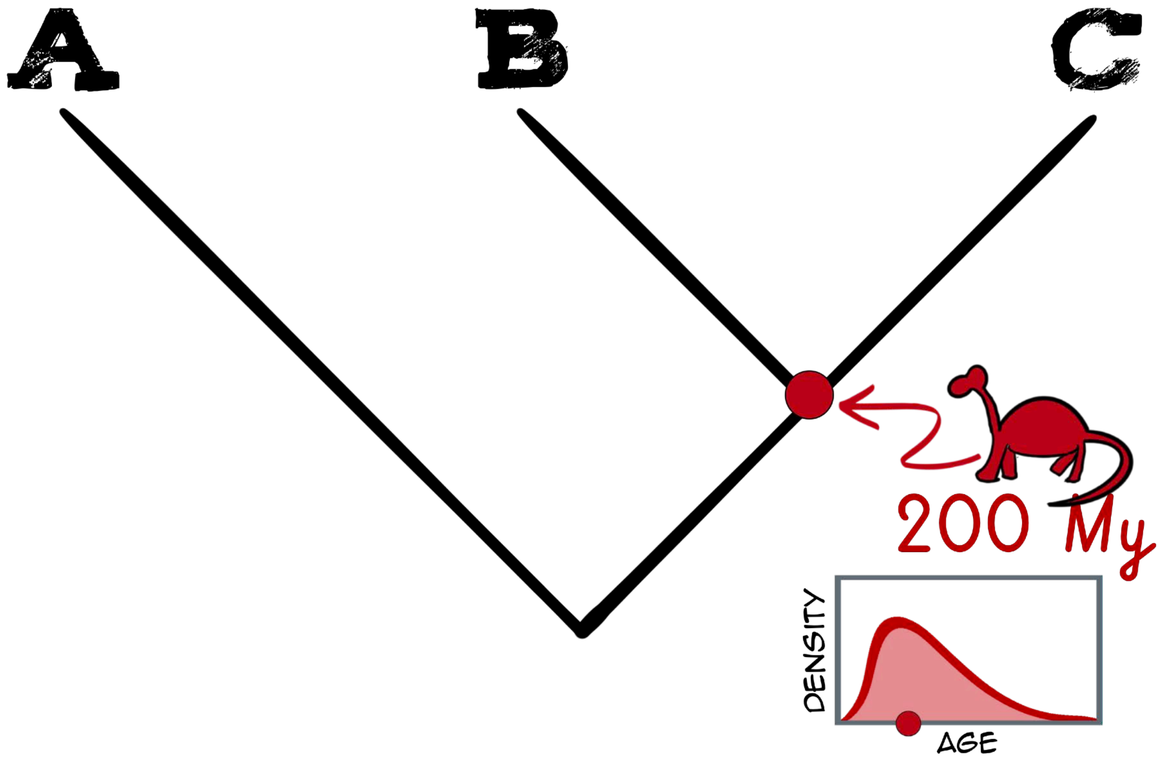
The placements and ages of fossil taxa are always uncertain



Dating Nodes

Bayesian inference is well suited to accommodating uncertainty in the age of the calibration node

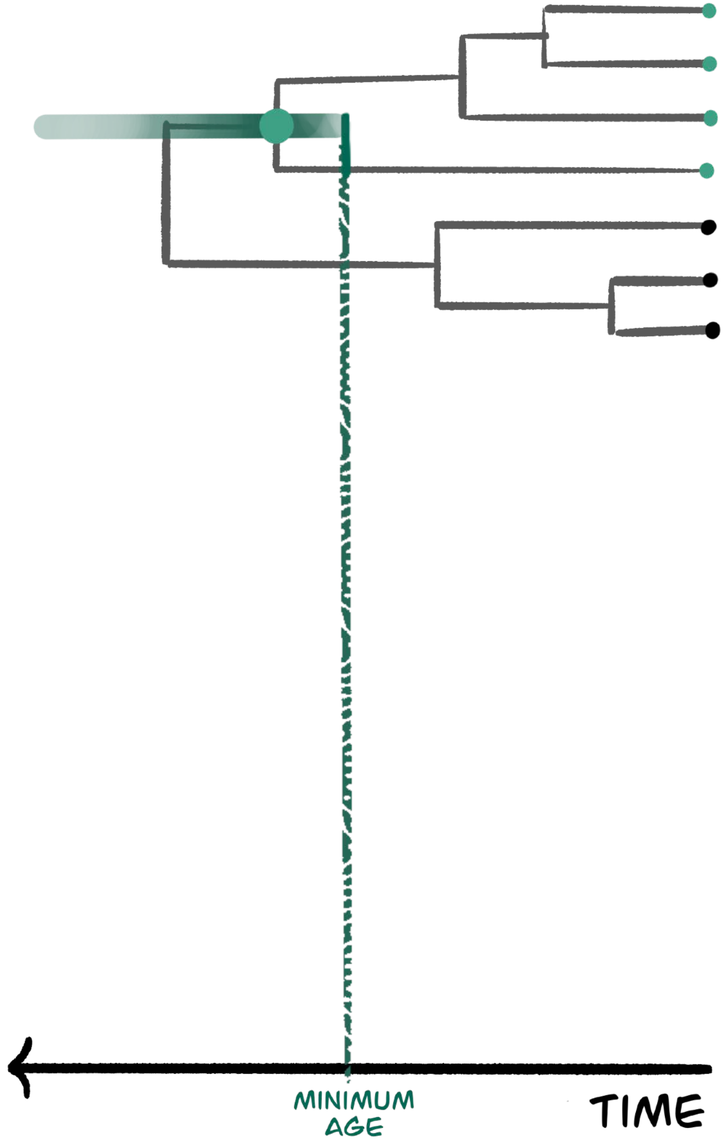
Divergence times are calibrated by placing parametric densities on internal nodes offset by age estimates from the fossil record



Dating Nodes with Fossils

Age estimates from fossils can provide **minimum** time constraints for internal nodes

Reliable **maximum** bounds are typically unavailable

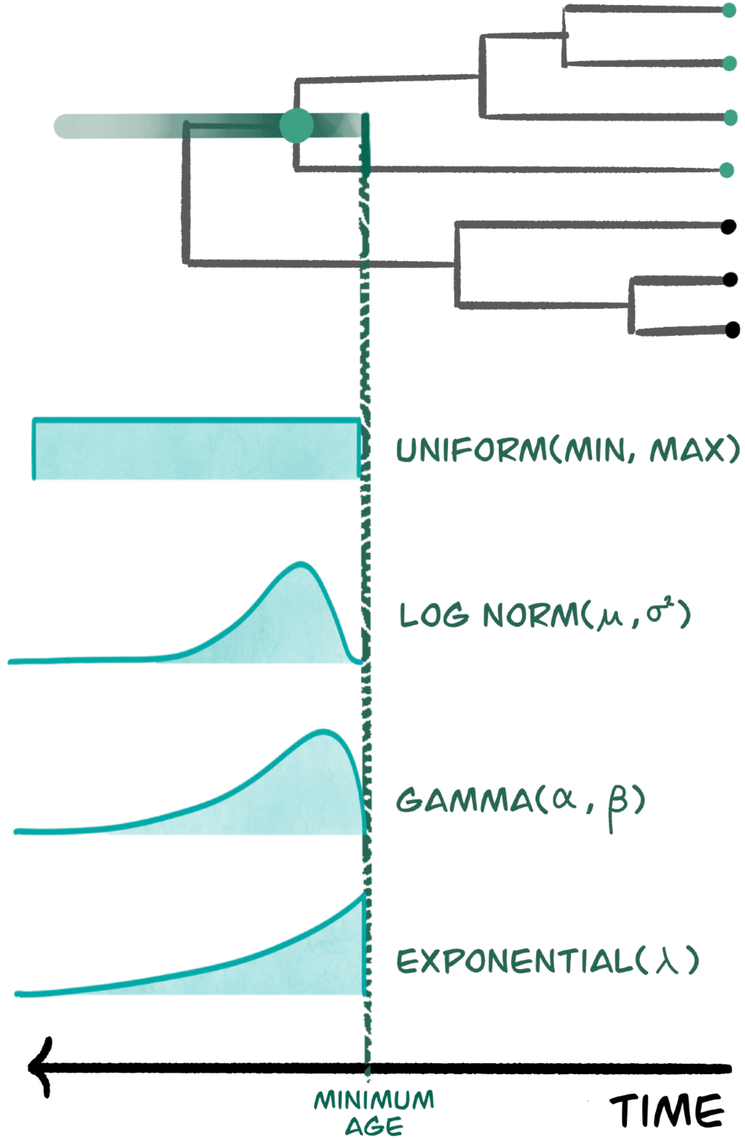


Node Calibrations

Common practice in Bayesian divergence-time estimation

Parametric distributions are typically off-set by the age of the oldest fossil assigned to a clade

These calibration densities do not (necessarily) require specification of maximum bounds

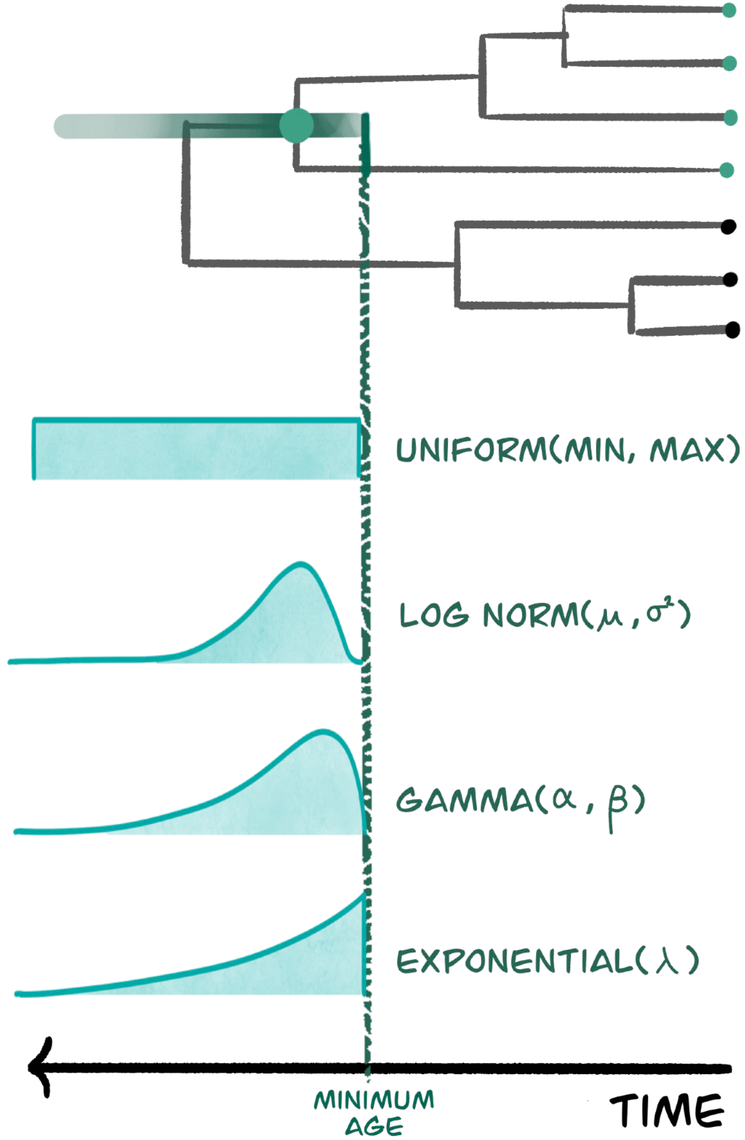


Node Calibrations

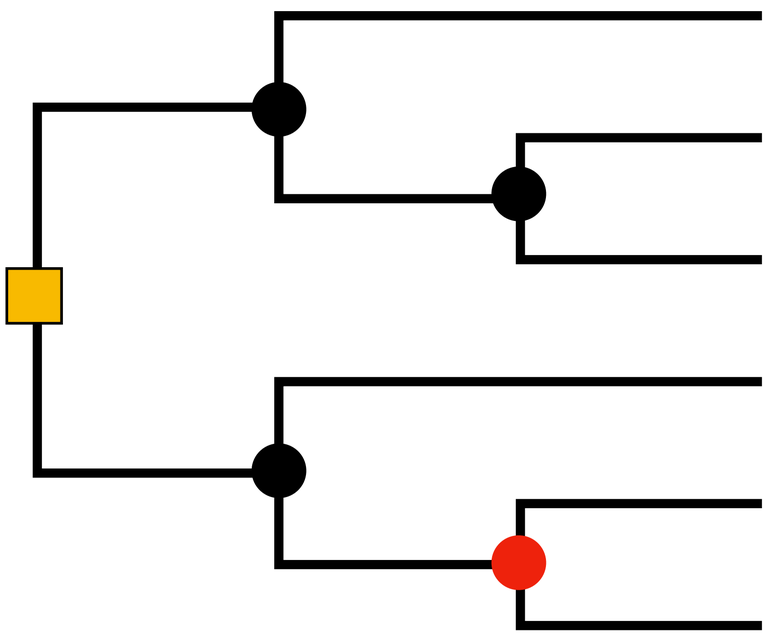
Estimates of absolute node ages are driven primarily by the calibration density

Specifying appropriate densities is a challenge for most molecular biologists

Interpreting the specification and impact of calibration densities is not straightforward



Node Calibrations as Priors



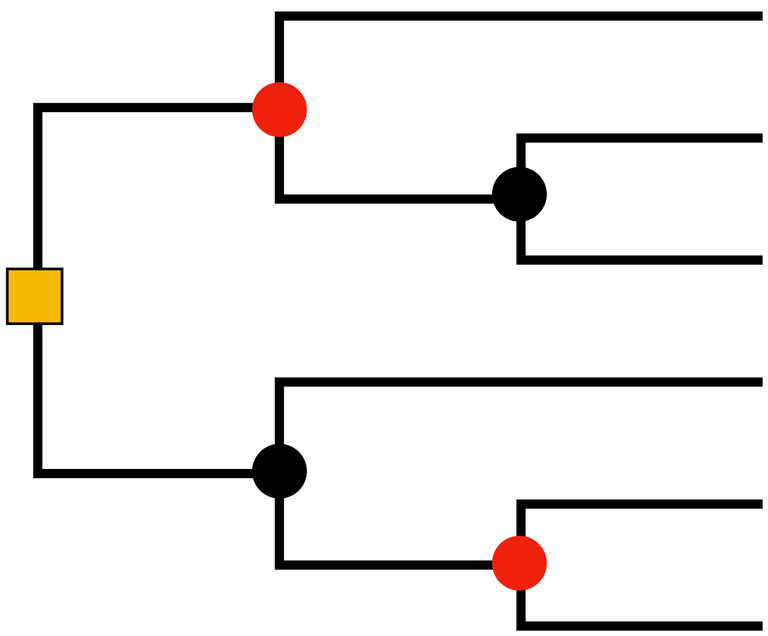
If node ages are considered priors, then one must condition the uncalibrated nodes on the ages of the calibrated nodes

The tree probability is constructed to simultaneously account for uncertainty in the ages of both calibrated and uncalibrated nodes

$$f(\text{tree}) = f(\bullet\ \bullet\ \bullet\ \text{tree} \mid \bullet, \lambda, \mu, \rho) f(\bullet \mid \theta_c)$$

(Rannala & Yang 2005; Heled & Drummond 2012)

Node Calibrations as Priors



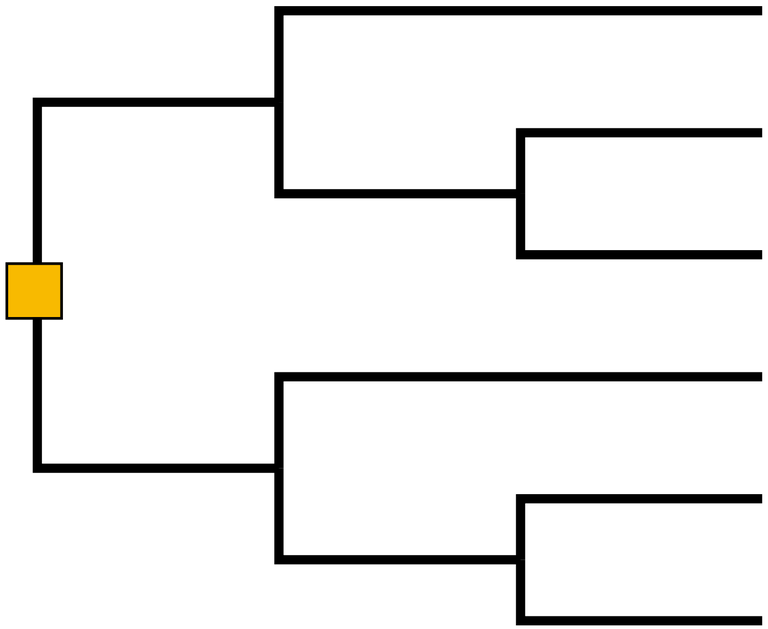
For multiple calibrations:

Rannala (2016) showed that the conditional prior leads to counterintuitive topologically inconsistent realized priors

dos Reis (2016) demonstrated that this prior is computationally intractable

$$f(\tau) = f(\tau \mid \lambda, \mu, \rho) f(\bullet \mid \theta_{c1})(\bullet \mid \theta_{c2})$$

Node Calibrations as Likelihoods



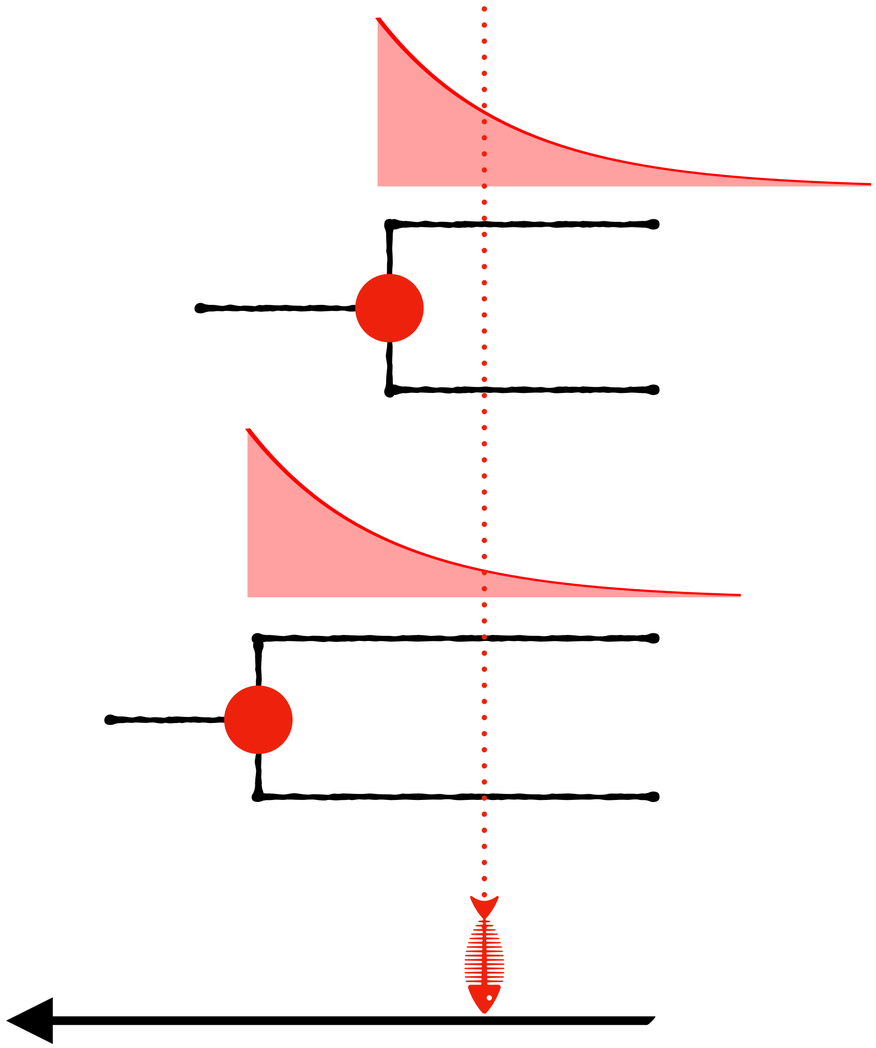
It is conceptually simpler to interpret the node calibrations as the likelihood of the fossil data

This works easily for multiple calibrations

$$f(\text{tree} \mid \lambda, \mu, \rho, \text{fossil}) \propto f(\text{fossil} \mid \text{tree}, \theta_c) f(\text{tree} \mid \lambda, \mu, \rho)$$

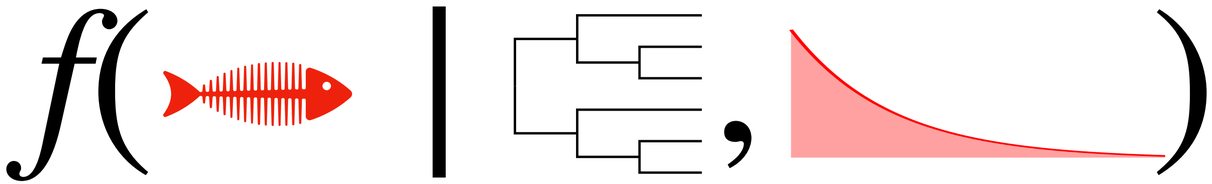
(Rannala 2016; Pett & Heath 2020)

Node Calibrations as Likelihoods



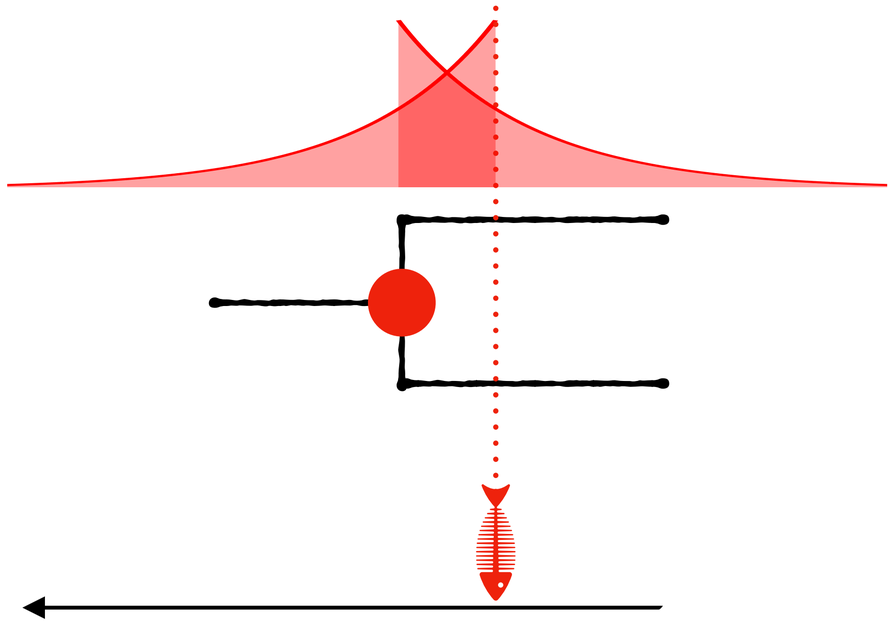
It is conceptually simpler to interpret the node calibrations as the likelihood of the fossil data

This works easily for multiple calibrations



(Pett & Heath 2020)

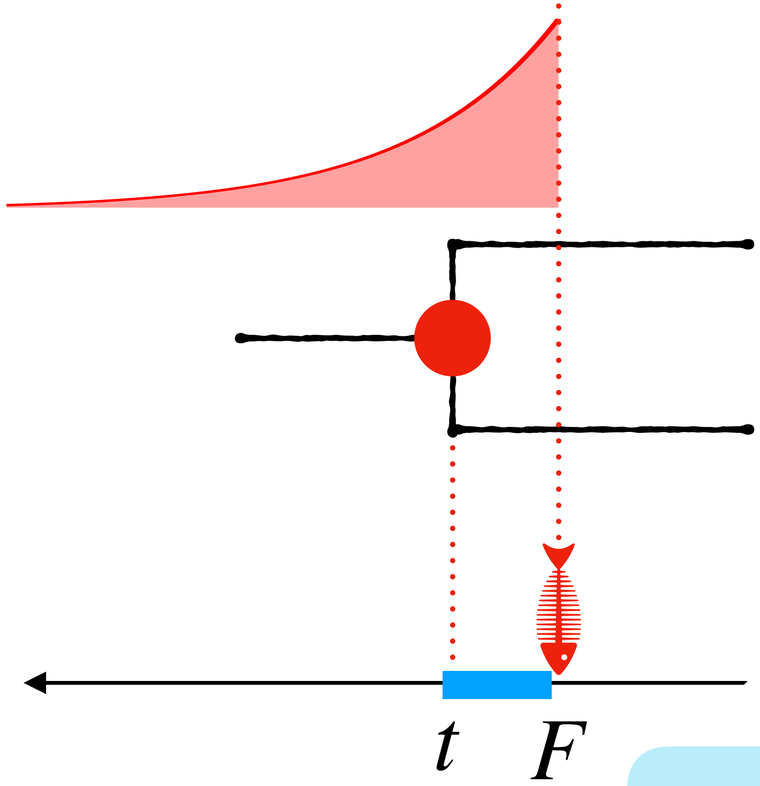
Node Calibrations as Likelihoods

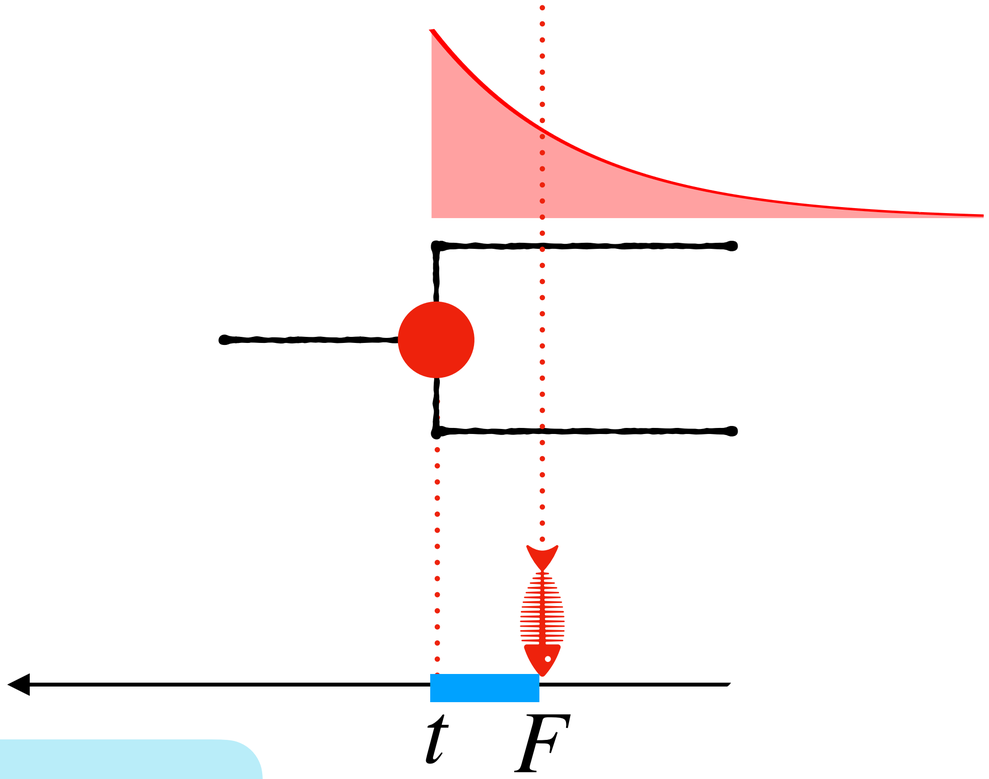


It is common practice to treat node calibrations as likelihoods, but this is typically misinterpreted as a prior

(Pett & Heath 2020)

Misinterpretation of Node Calibrations





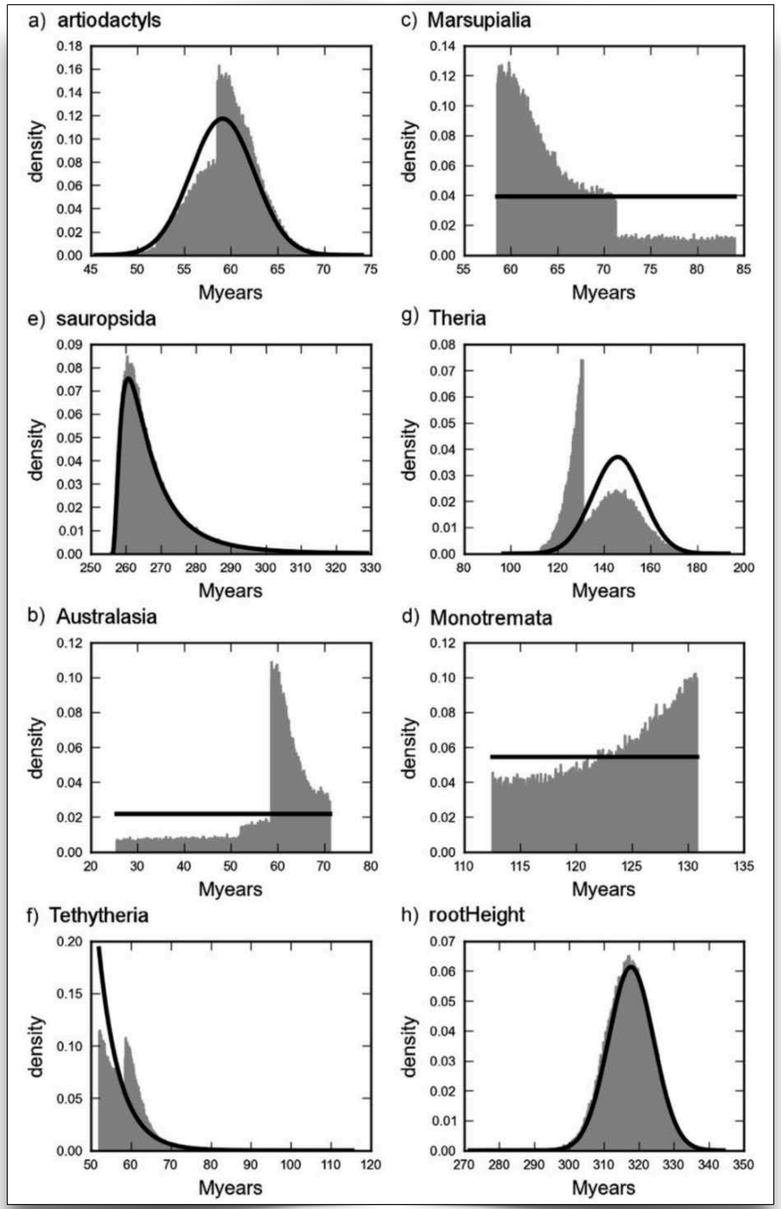
THE MATH STILL WORKS

$$f(t \mid F, \lambda) = \lambda e^{-\lambda(t-F)}$$

$$f(F \mid t, \lambda) = \lambda e^{-\lambda(t-F)}$$

Node Calibrations as "Priors"

Despite this being a conceptual issue, it leads to ***misinterpreting*** the marginal posterior as an "effective" or "implicit" prior

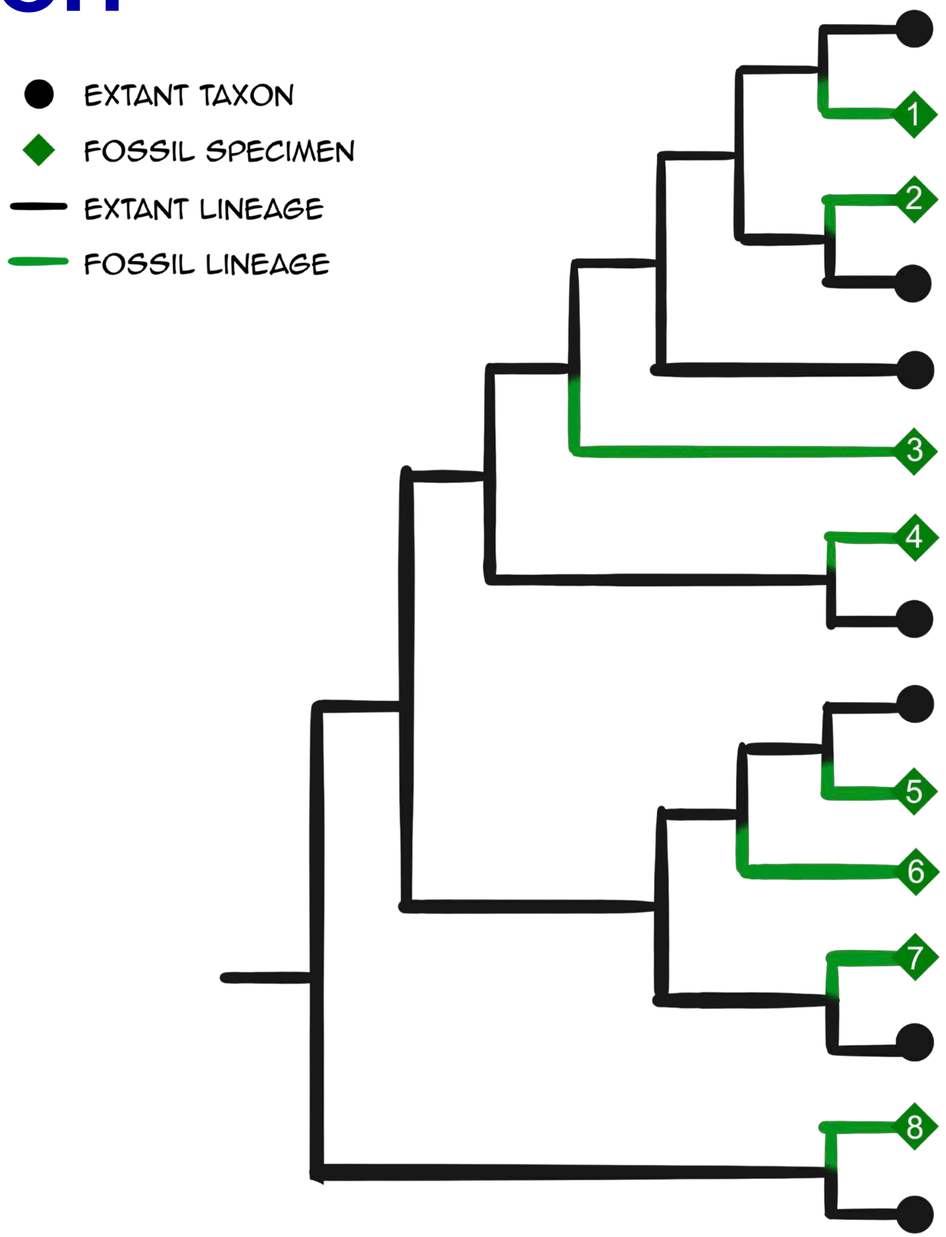


(figure from Heled & Drummond 2012)

Improving Fossil Calibration

We would prefer to eliminate the need for ad hoc calibration densities

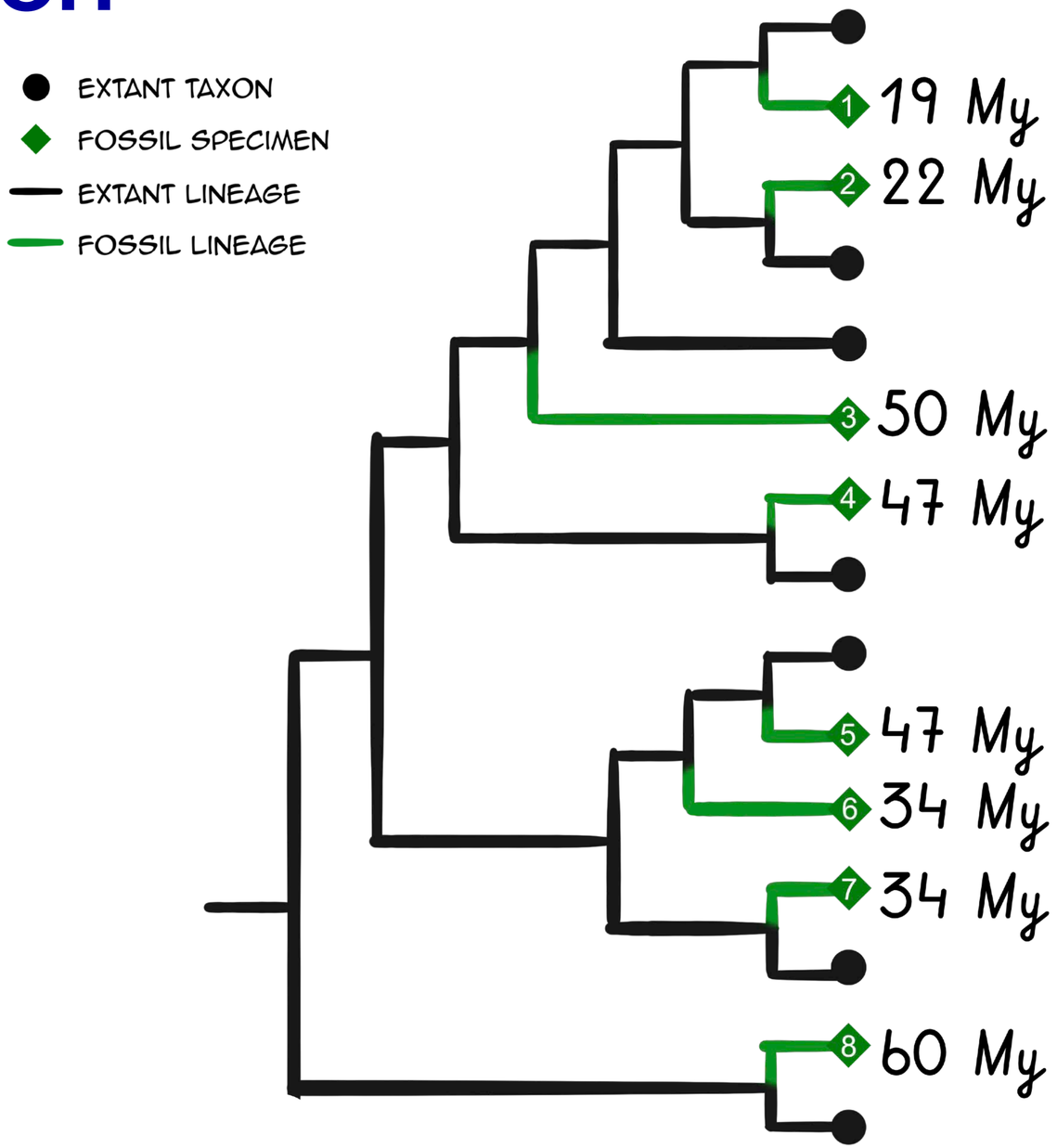
Calibration densities do not account for diversification of fossils



Improving Fossil Calibration

Example:
8 dated fossils ranging from 60 Million years to 19 Million years

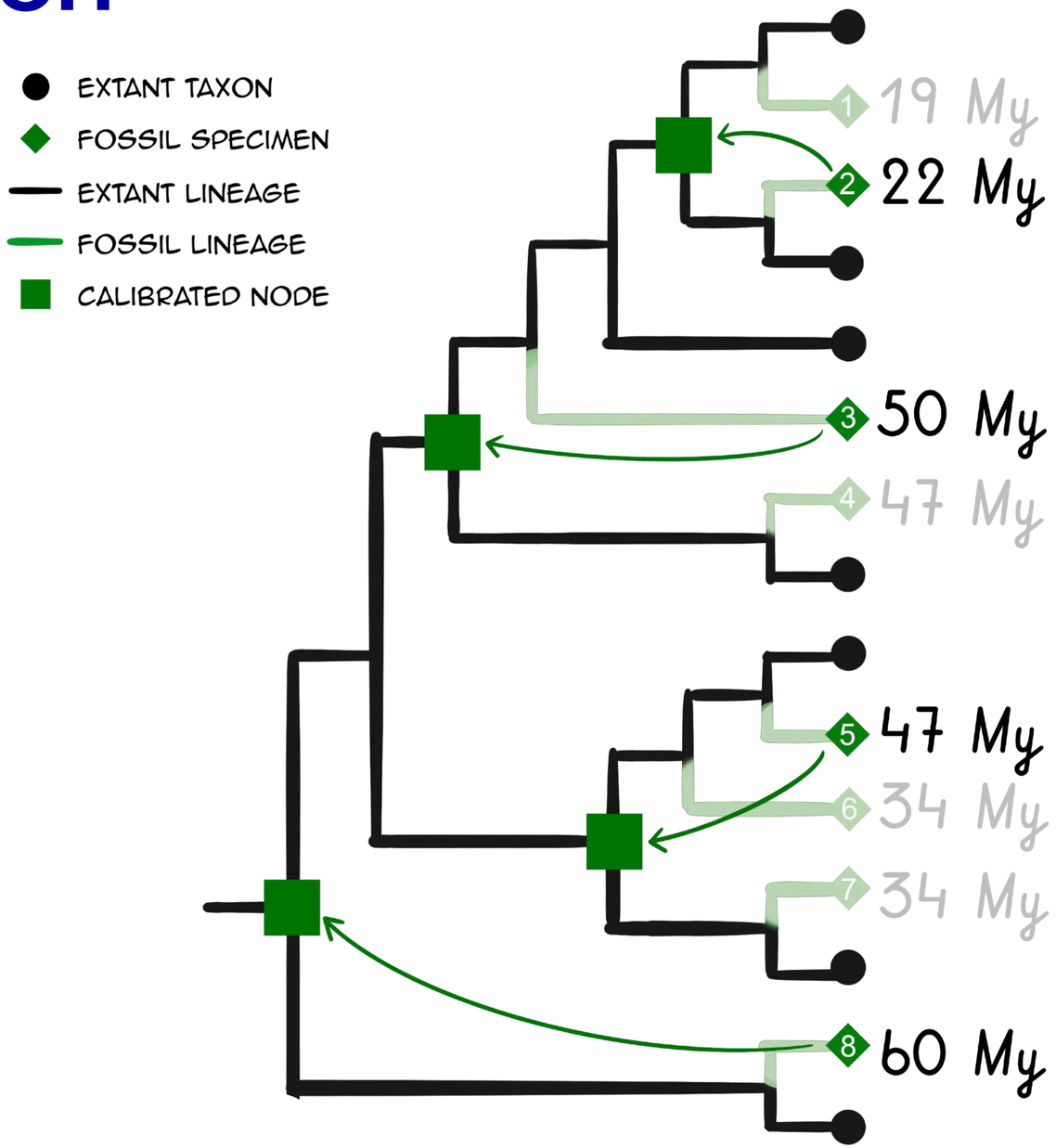
Only 4 nodes can be constrained using calibration densities



Improving Fossil Calibration

Example:
8 dated fossils ranging from 60 to 19 Million years

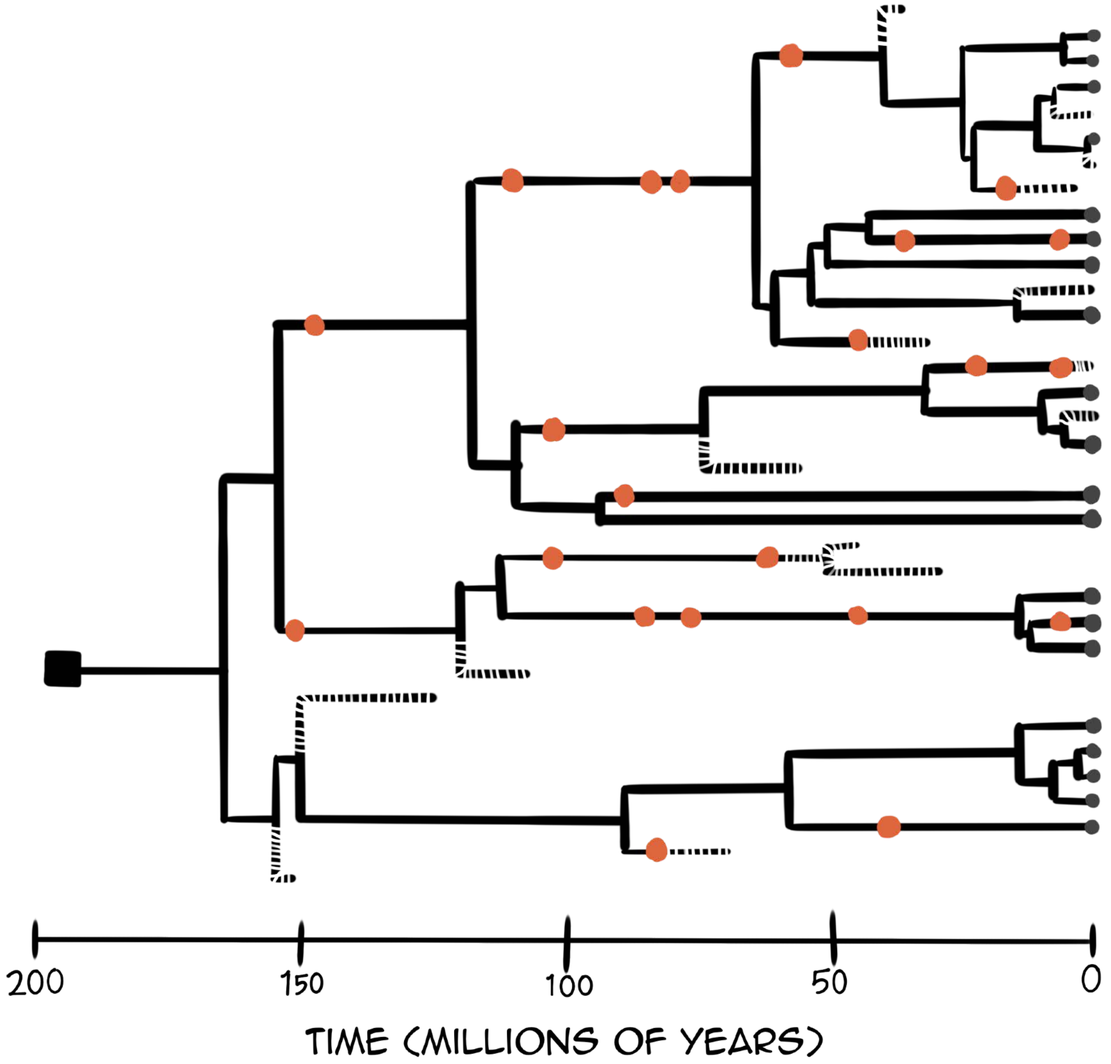
Only 4 nodes can be constrained using calibration densities



Improving Fossil Calibration

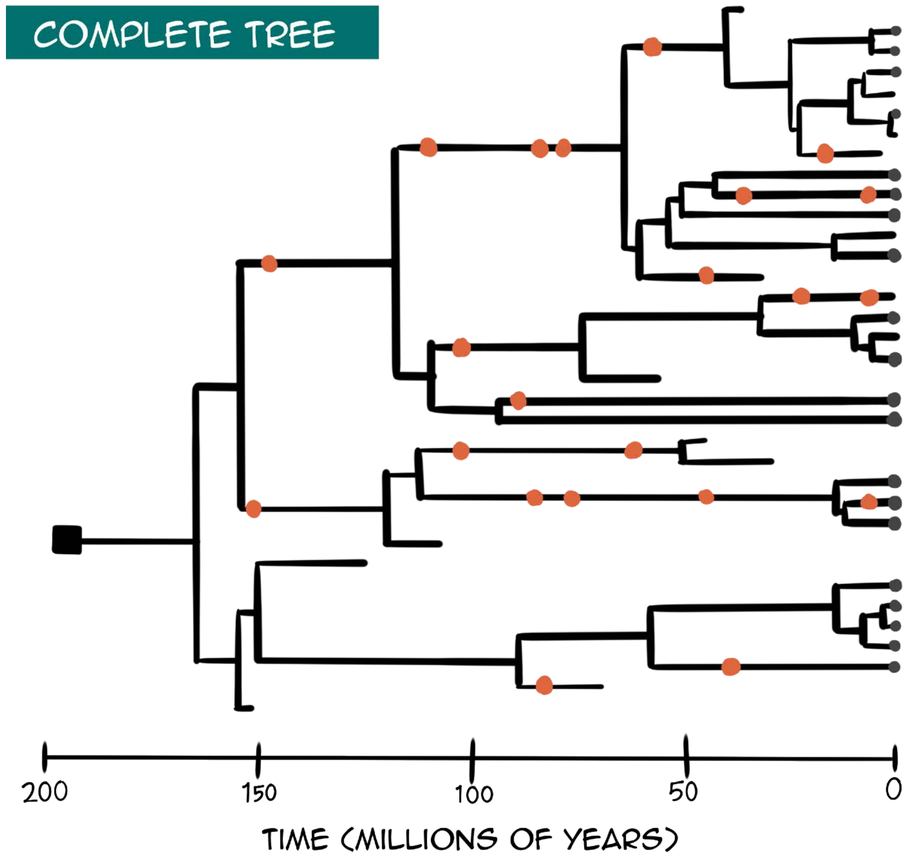
We want to use all of the available fossils

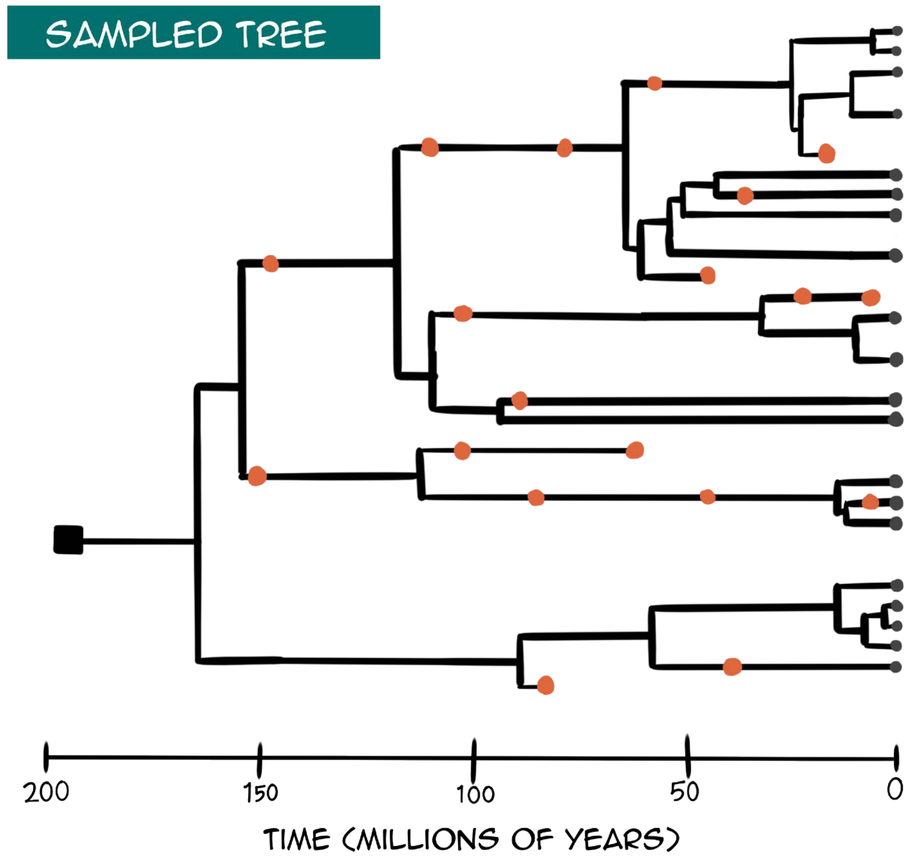
Because fossils are part of the same diversification process that gave rise to extant taxa, we can include them in the birth-death model



Modeling the Tree & Occurrence Times

Stadler (2010) introduced a generating model for a serially sampled time tree – this is the *fossilized birth-death process*

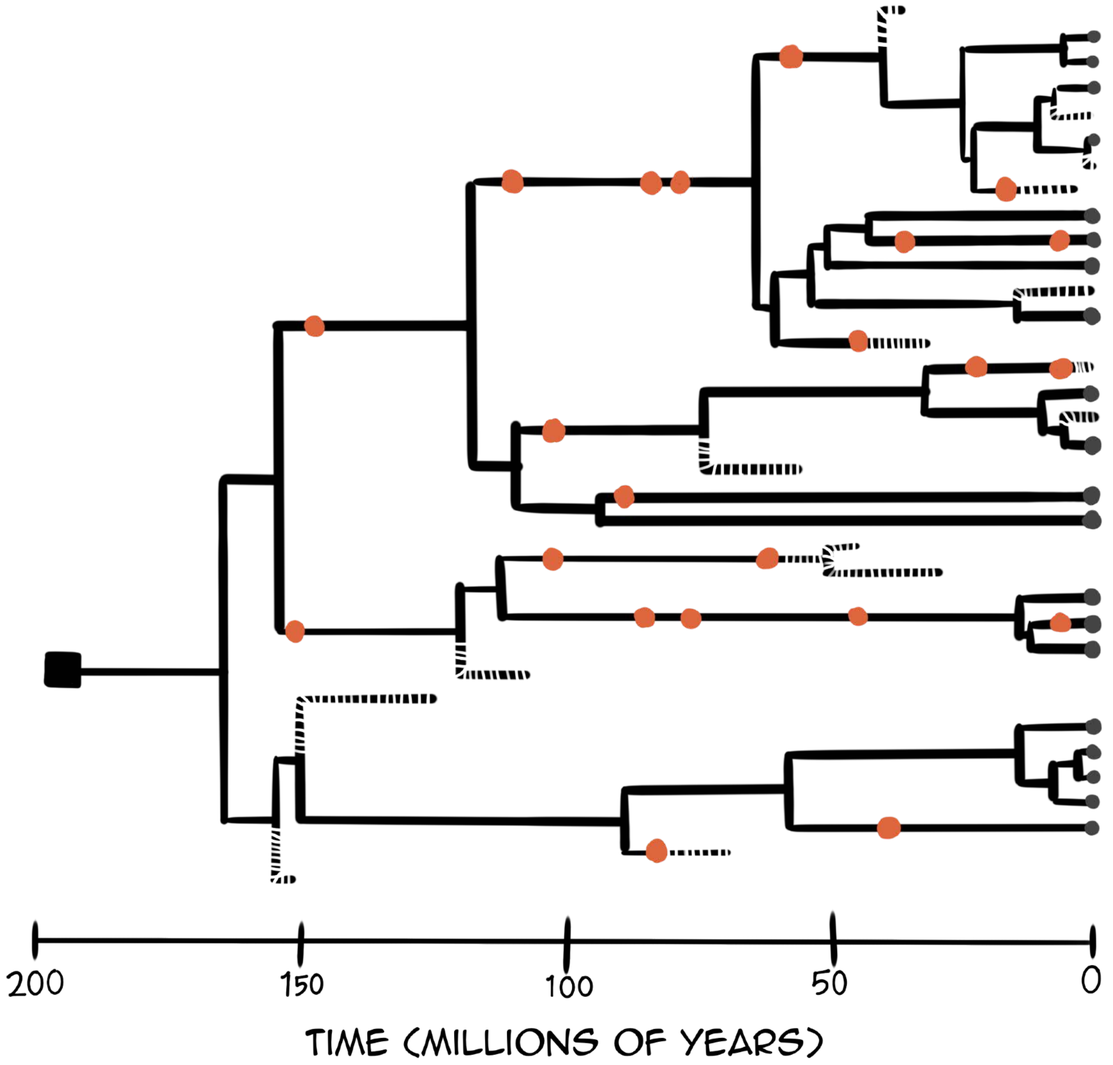




(figure based on Stadler 2010)

The Fossilized Birth-Death Process (FBD)

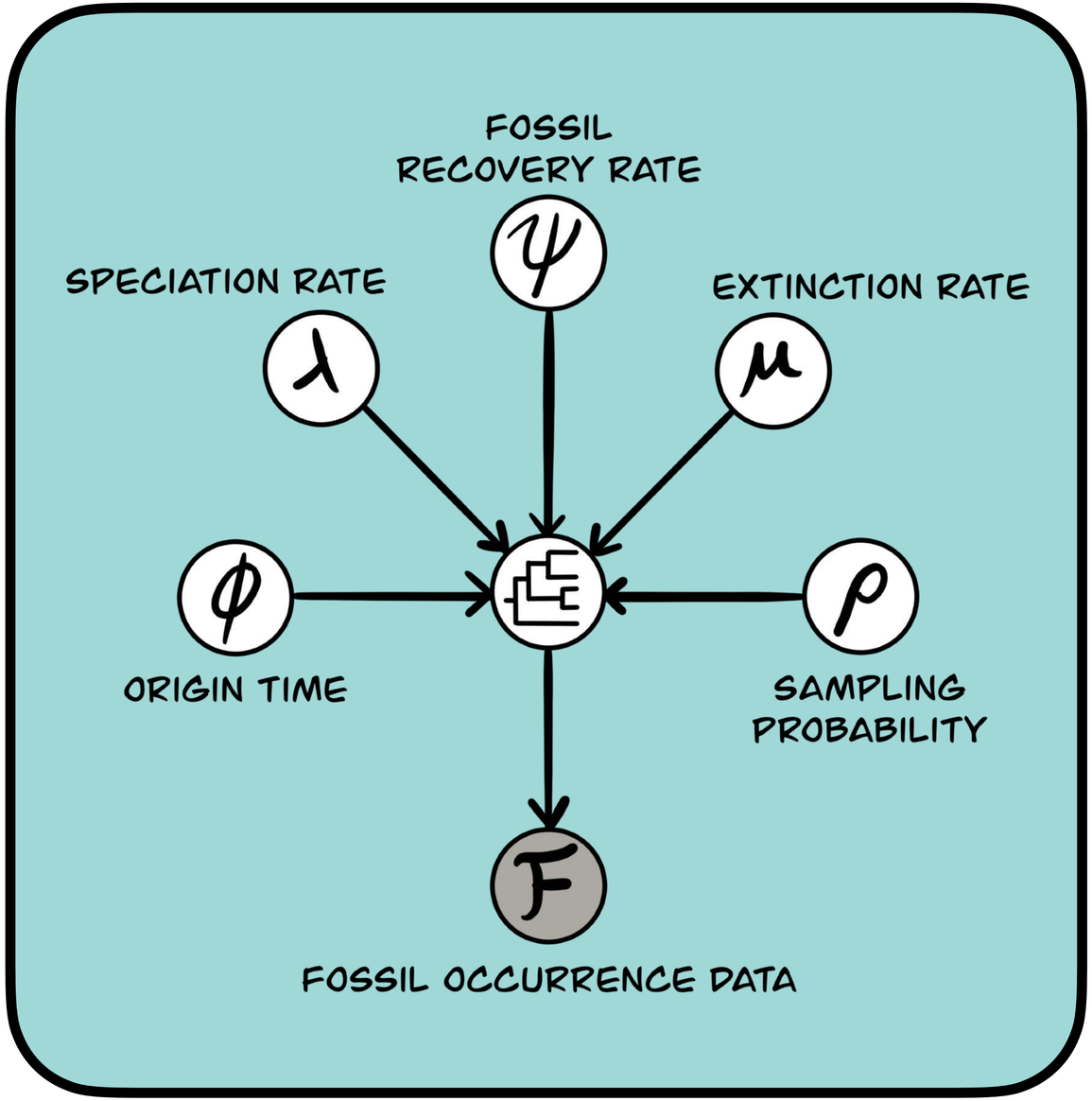
Recovered fossil specimens provide historical observations of the diversification process that generated the tree of extant species



The Fossilized Birth-Death Process (FBD)

The probability of the tree topology, divergence times, and fossil occurrences is conditioned on

- the origin time
- speciation rate
- extinction rate
- fossil recovery rate
- probability of sampling an extant species



Sampled Ancestors

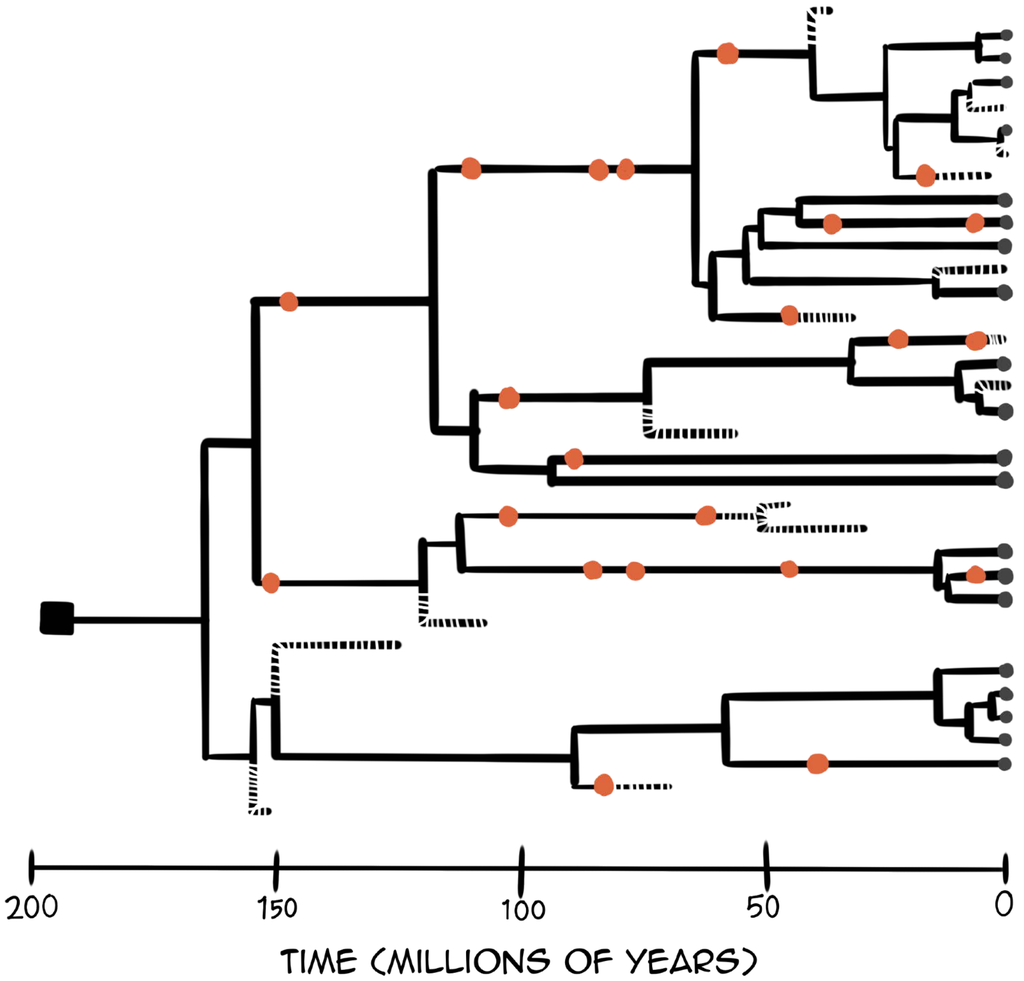
Sampled lineages with sampled decedents

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

On the probability of ancestors in the fossil record

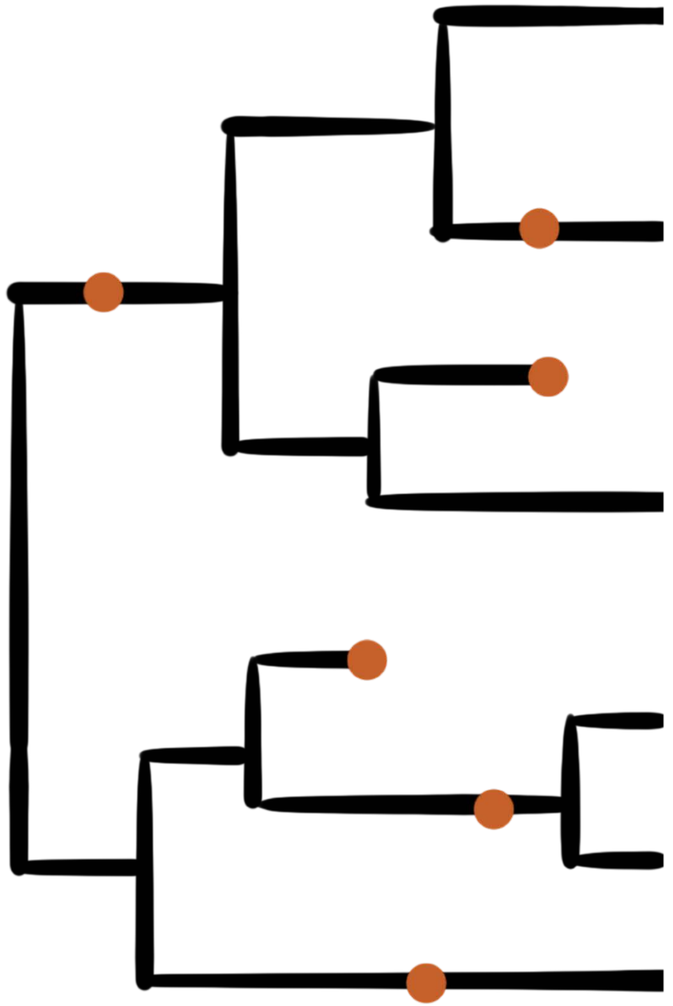
Mike Foote

There is a non-zero probability of sampling ancestor-defendant relationships from the fossil record



Sampled Ancestors

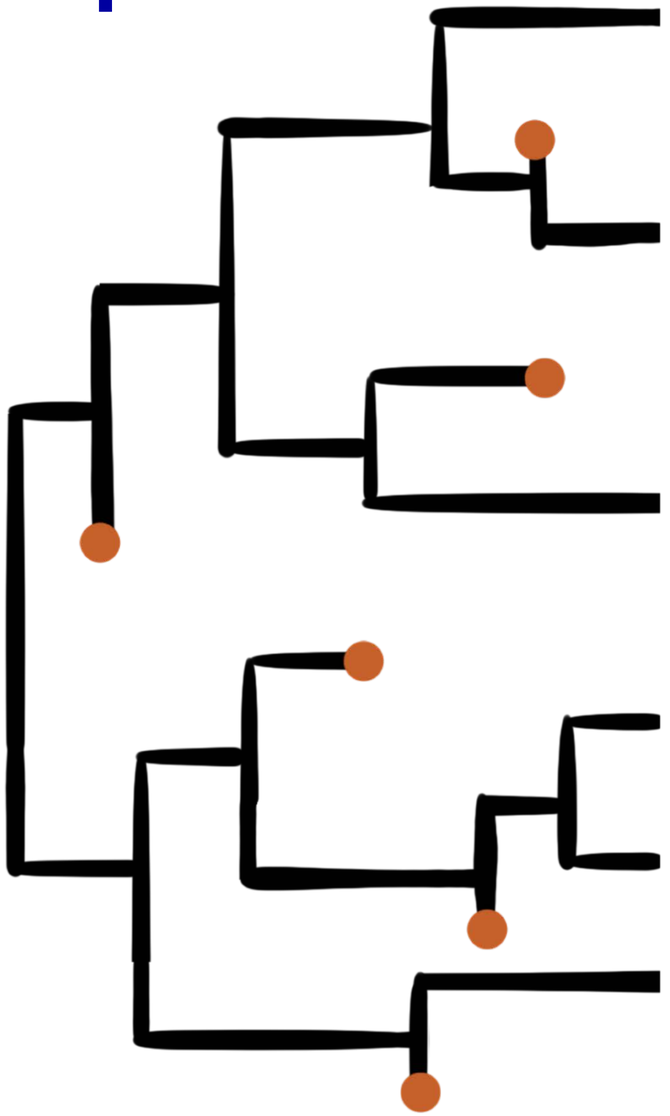
Because fossil & living taxa are assumed to come from a single diversification process, there is a non-zero probability of sampled ancestors



(Pett & Heath 2020)

~~Sampled Ancestors~~ Fossils as "Tips"

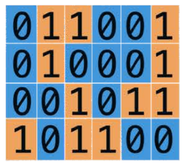
If all fossils, including sampled ancestors, are forced to be on separate lineages, this induces additional speciation events and will, in turn, influence rate & node-age estimates



(Pett & Heath 2020)

Combining Fossil & Extant Data

Statistical methods provide a way to integrate paleontological and neontological data

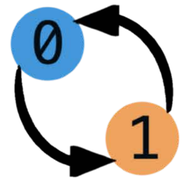


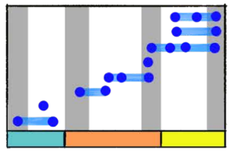
Syst. Biol. 50(6):913–925, 2001

A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data

PAUL O. LEWIS

Department of Ecology and Evolutionary Biology, The University of Connecticut, Storrs, Connecticut 06269-3043 , USA;
E-mail: paul.lewis@uconn.edu



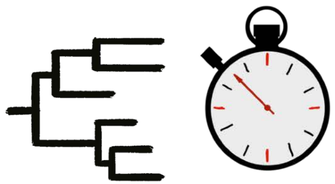


Syst. Biol. 61(6):973–999, 2012

DOI:10.1093/sysbio/sys058
Advance Access publication on June 20, 2012

A Total-Evidence Approach to Dating with Fossils, Applied to the Early Radiation of the Hymenoptera

FREDRIK RONQUIST[1,*], SERAINA KLOPFSTEIN[1], LARS VILHELMSEN[2], SUSANNE SCHULMEISTER[3], DEBRA L. MURRAY[4], AND ALEXANDR P. RASNITSYN[5]



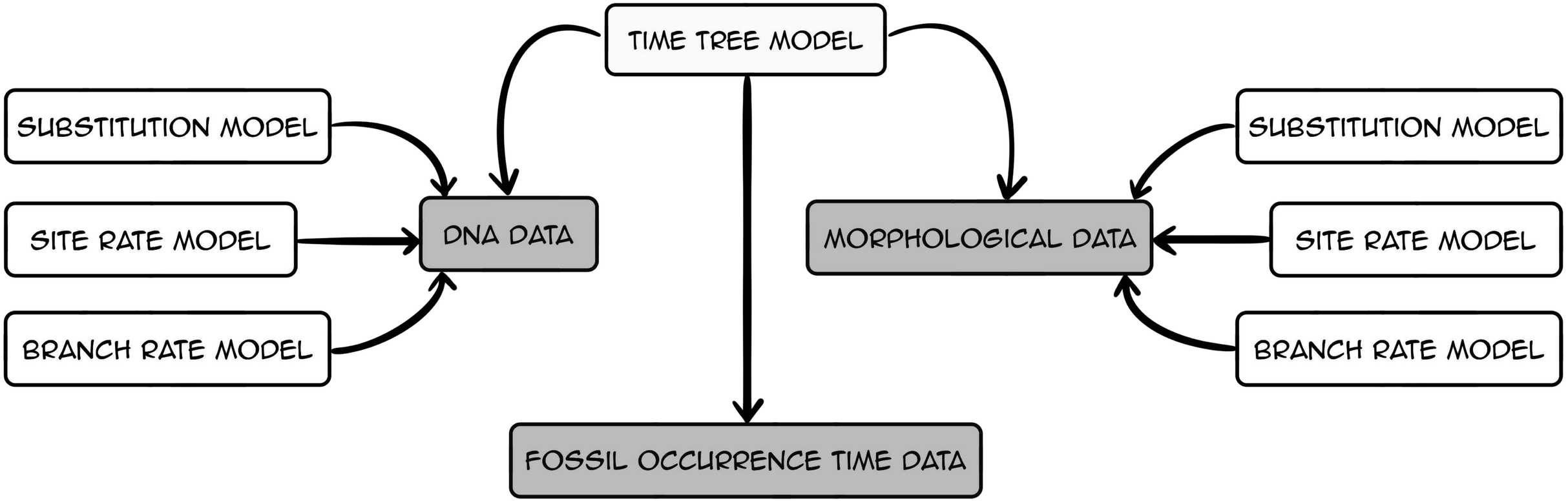
Paleontology & Neontology

"Except during the interlude of the [Modern] Synthesis, there has been limited communication historically among the disciplines of evolutionary biology, particularly between students of evolutionary history (paleontologists and systematists) and those of molecular, population, and organismal biology. **There has been increasing realization that barriers between these subfields must be overcome if a complete theory of evolution and systematics is to be forged.**"

Reaka-Kudla, M.L. & Colwell, R.: in E.C. Dudley (ed.), *The Unity of Evolutionary Biology: Proceedings of the Fourth International Congress of Systematic & Evolutionary Biology,* Discorides Press, Portland, OR, p. 16.

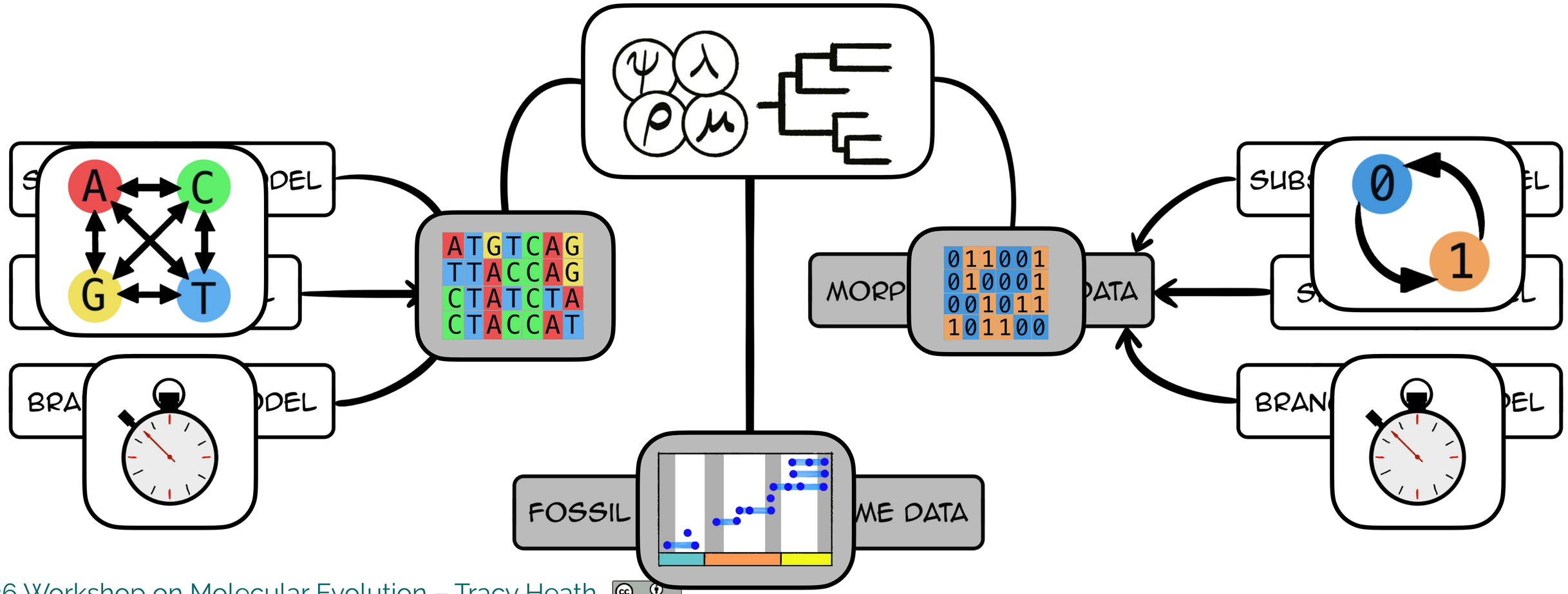
Combining Fossil & Extant Data

Combine models for sequence evolution, morphological change, & fossil recovery to jointly estimate the tree topology, divergence times, & lineage diversification rates



Combining Fossil & Extant Data

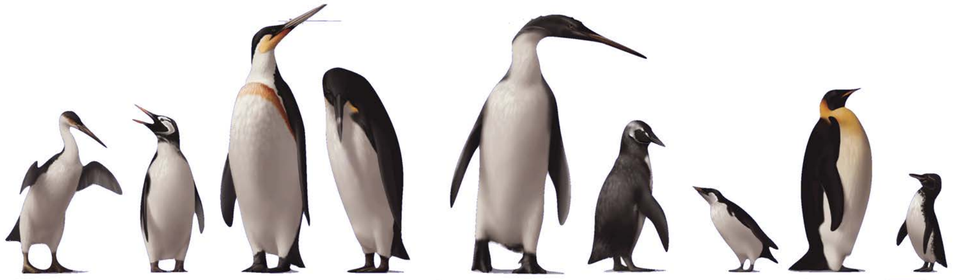
Combine models for sequence evolution, morphological change, & fossil recovery to jointly estimate the tree topology, divergence times, & lineage diversification rates



Penguin Diversity in Deep Time

Does our understanding of penguin evolution improve when we consider both extant and fossil taxa?

"Penguin Party" by Kate Dzikiewicz



Artistic reconstructions by: Stephanie Abramowicz for Scientific American
Fordyce, R.E. and D.T. Ksepka. The Strangest Bird Scientific American 307, 56 – 61 (2012)

Penguin Evolution

New penguin fossils shed light on the evolution of smaller body size in modern penguins

Illustrations by Simone Giovanardi

NEWLY DESCRIBED GIANT PENGUIN WEIGHED APPROXIMATELY 160 KG

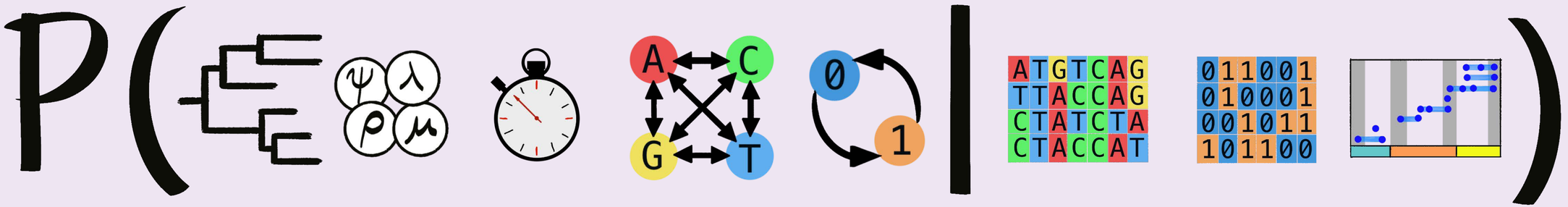
EMPEROR PENGUIN

(Thomas et al. *Proc. Roy Soc*. B 2020; Cole et al, *Nature Comm*. 2022; Ksepka et al. *J. Paleontology* 2023)

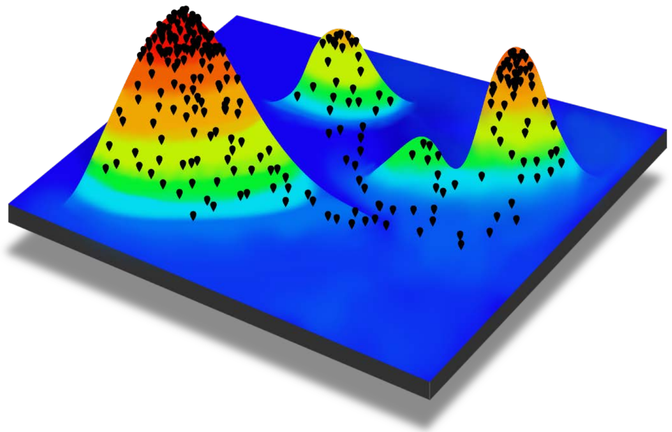
Statistical Inference under Phylogenetic Models

Bayesian inference methods are useful for estimating phylogenetic parameters and testing hypotheses

Posterior Probability



We use Markov chain Monte Carlo (MCMC) and other approaches to sample the joint posterior probability density of the model parameters conditional on the data



Tutorial

https://revbayes.github.io/tutorials/fbd_simple