Bayesian Divergence Time Estimation

Tracy Heath

Department of Integrative Biology, University of California, Berkeley

2013 Workshop on Molecular Evolution
Český Krumlov, Czech Republic

Outline

Overview of divergence time estimation

- Relaxed clock models – accounting for variation in substitution rates among lineages
  - Dirichlet process prior for lineage-specific rates

break

- Tree priors and fossil calibration

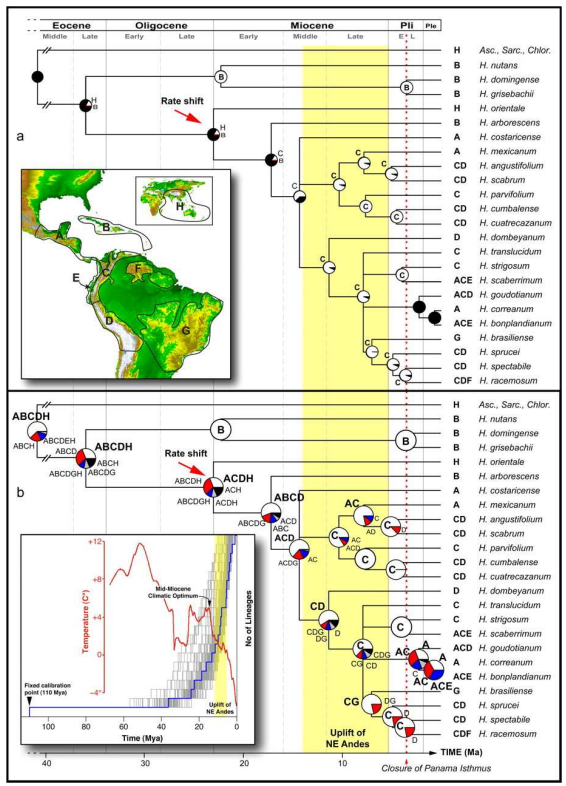
lunch

BEAST Tutorial:

- Walk through: set up BEAST input file in BEAUti and execute BEAST MCMC analysis
- On your own: complete analysis by summarizing output

A Time-Scale for Evolution

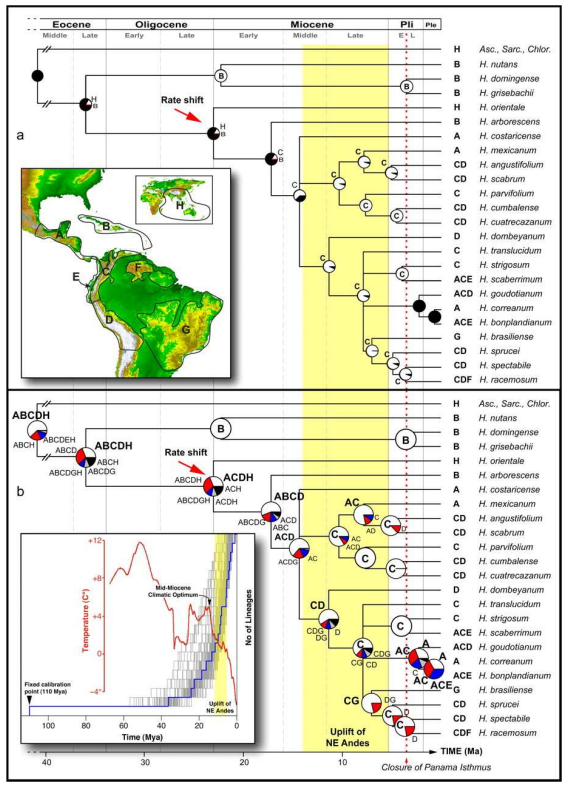
- Reconstruct ancestral ranges
- Environmental or geological correlates to diversification
- Morphological character change over time
- Detect shifts in rates of diversification
- Lineage-specific substitution rate



(Antonelli & Sanmartín. *Syst. Biol.* 2011)

A Time-Scale for Evolution

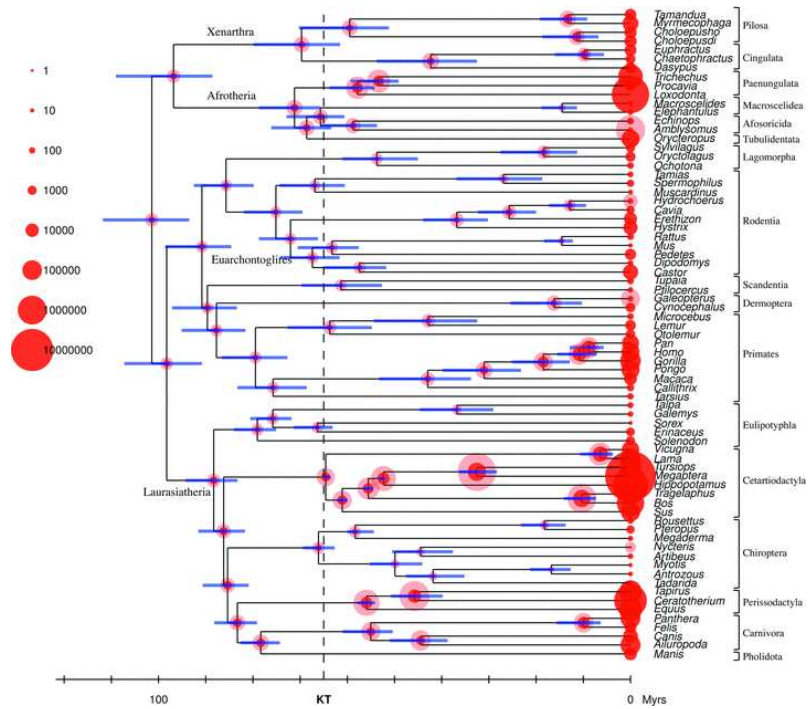
- Reconstruct ancestral ranges
- Environmental or geological correlates to diversification
- Morphological character change over time
- Detect shifts in rates of diversification
- Lineage-specific substitution rate



(Antonelli & Sanmartín. *Syst. Biol.* 2011)

A Time-Scale for Evolution

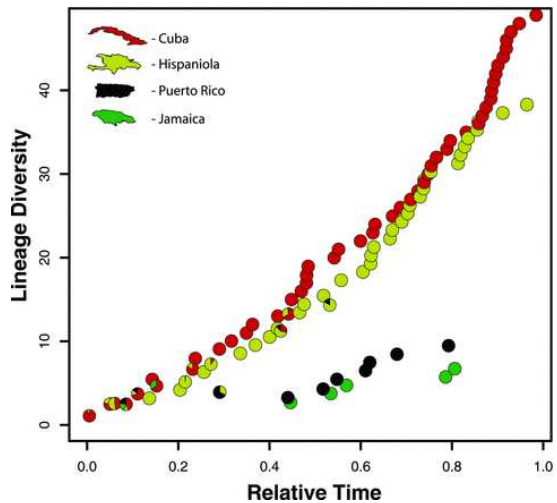
- Reconstruct ancestral ranges
- Environmental or geological correlates to diversification
- Morphological character change over time
- Detect shifts in rates of diversification
- Lineage-specific substitution rate



(Lartillot & Delsuc. *Evolution* 2012)

A Time-Scale for Evolution

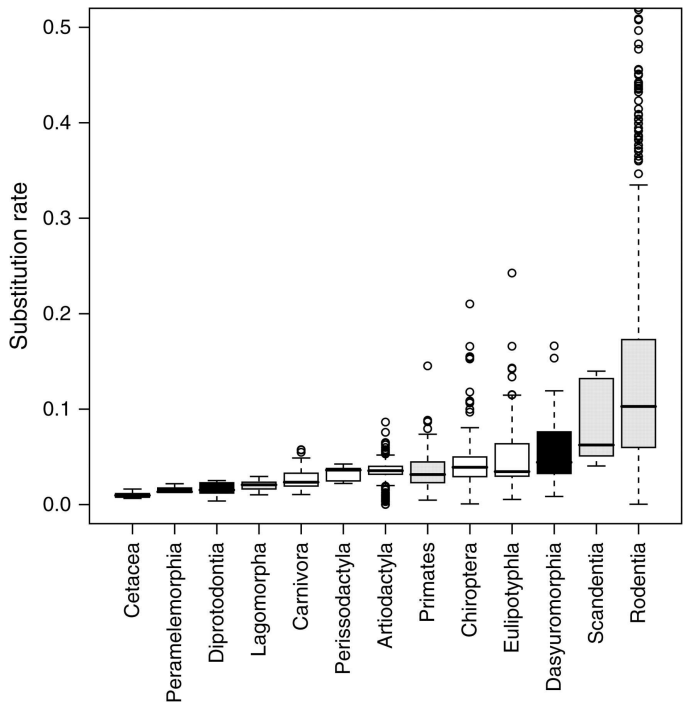
- Reconstruct ancestral ranges
- Environmental or geological correlates to diversification
- Morphological character change over time
- Detect shifts in rates of diversification
- Lineage-specific substitution rate



(Mahler, Revell, Glor, & Losos. *Evolution* 2010)

A Time-Scale for Evolution

- Reconstruct ancestral ranges
- Environmental or geological correlates to diversification
- Morphological character change over time
- Detect shifts in rates of diversification
- Lineage-specific substitution rate



(Nabholz, Glémin, Galtier. *MBE* 2008)

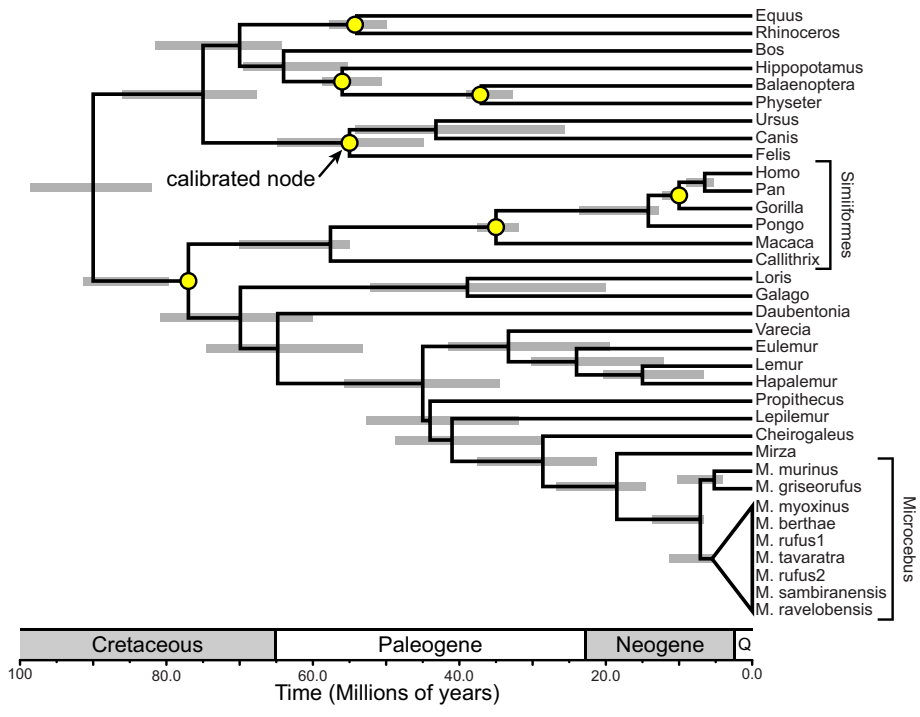
Divergence Time Estimation

Goal: Estimate the ages of interior nodes to understand the timing and rates of evolutionary processes

Model how rates are distributed across the tree

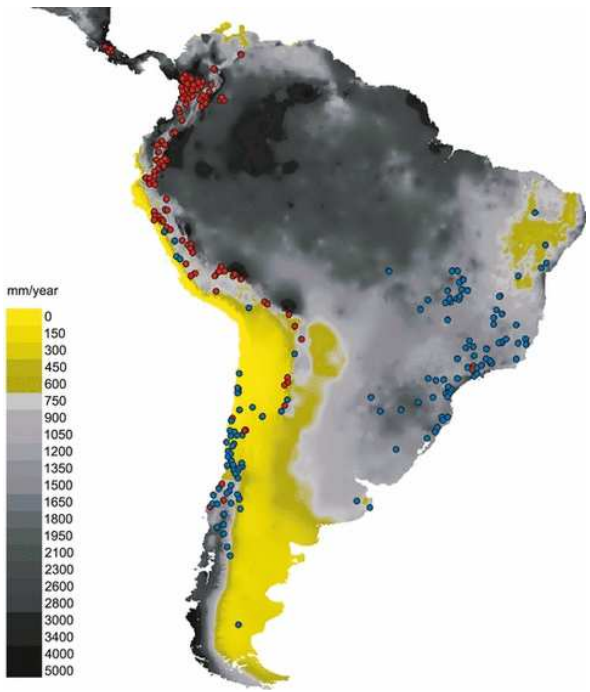
Describe the distribution of speciation events over time

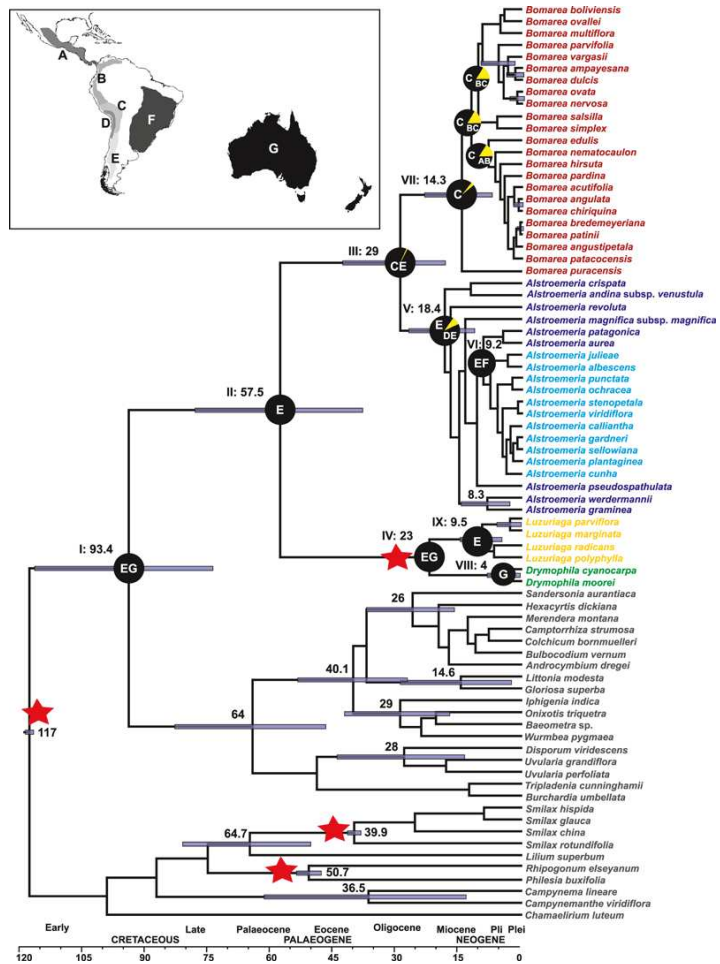
External calibration information for estimates of absolute node times



Understanding Historical Biogeography

"From East Gondwana to Central America: historical biogeography of the Alstroemeriaceae"





(Chacón et al., *J. Biolgeograpy* 2012)

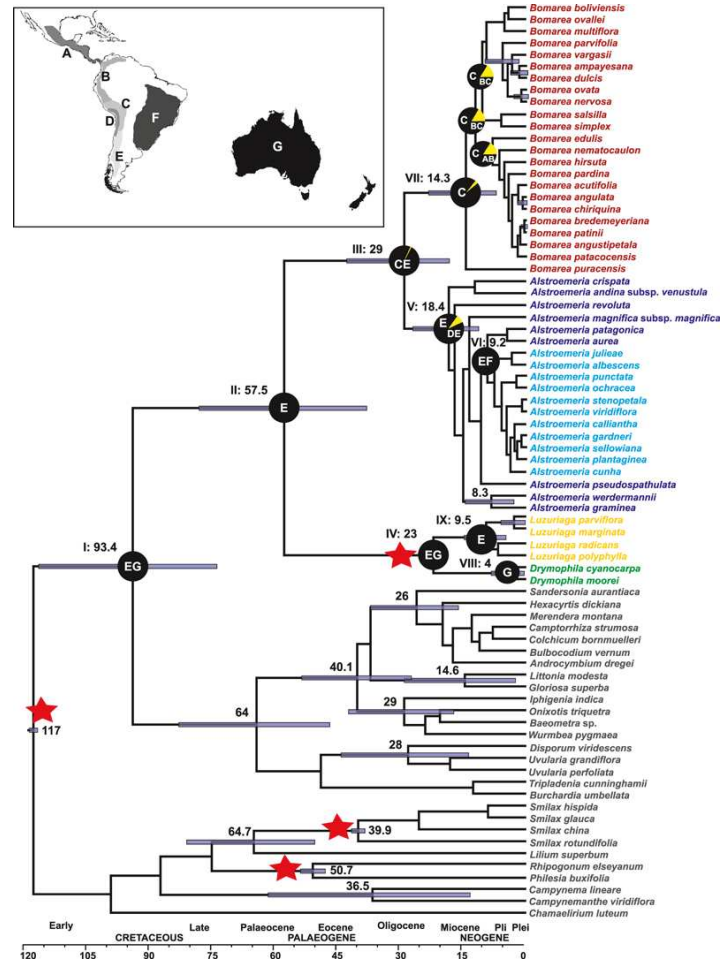
Divergence Time Estimation

Historical biogeography requires external calibration

Model how rates are distributed across the tree

Describe the distribution of speciation events over time

External calibration information for estimates of absolute node times



(Chacón et al., *J. Biolgeograpy* 2012)

Divergence Time Estimation

What about when the fossil record (or other types of calibration information) is poor or absent?

Example: Despite the rich diversity of *Anolis* there are few fossils

There are some amber fossils, but these fossils fall within a narrow time range

Amber *Anolis* fossil (http://www.anoleannals.org/2012/03/06/the-hi-tech-world-of-anole-paleontology/)

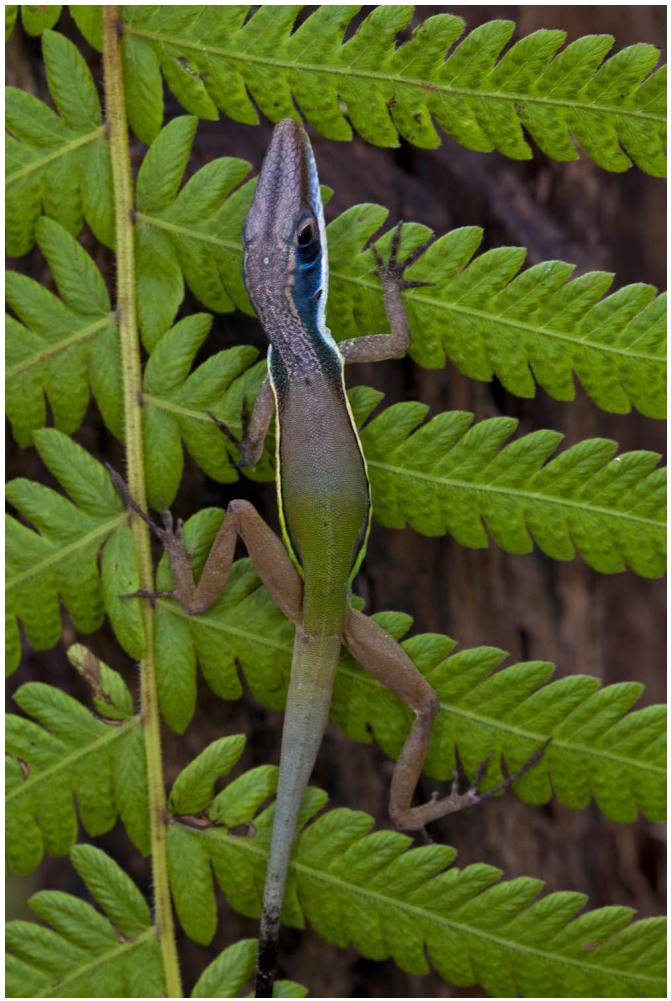
Divergence Time Estimation

What about when the fossil record is poor or absent?

Model how rates are distributed across the tree

Describe the distribution of speciation events over time

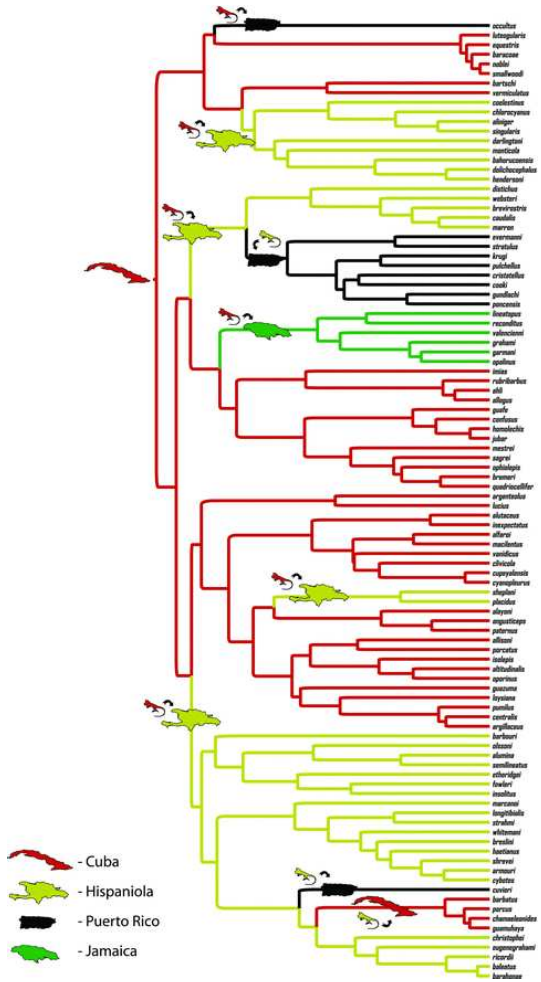
Estimation of relative divergence times

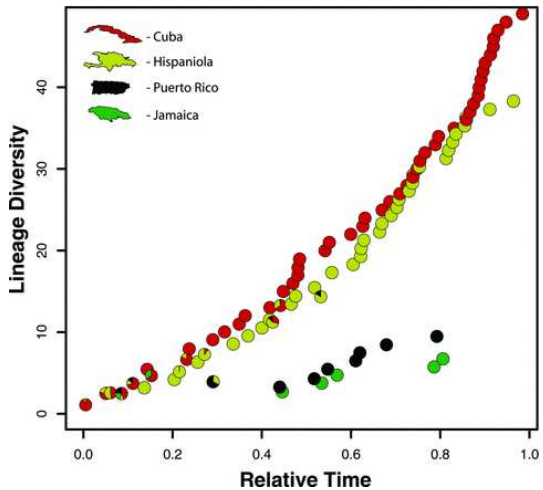


Anolis hendersoni (Image courtesy of L. Mahler)

Relative Times and Diversification

"Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean Anoles"



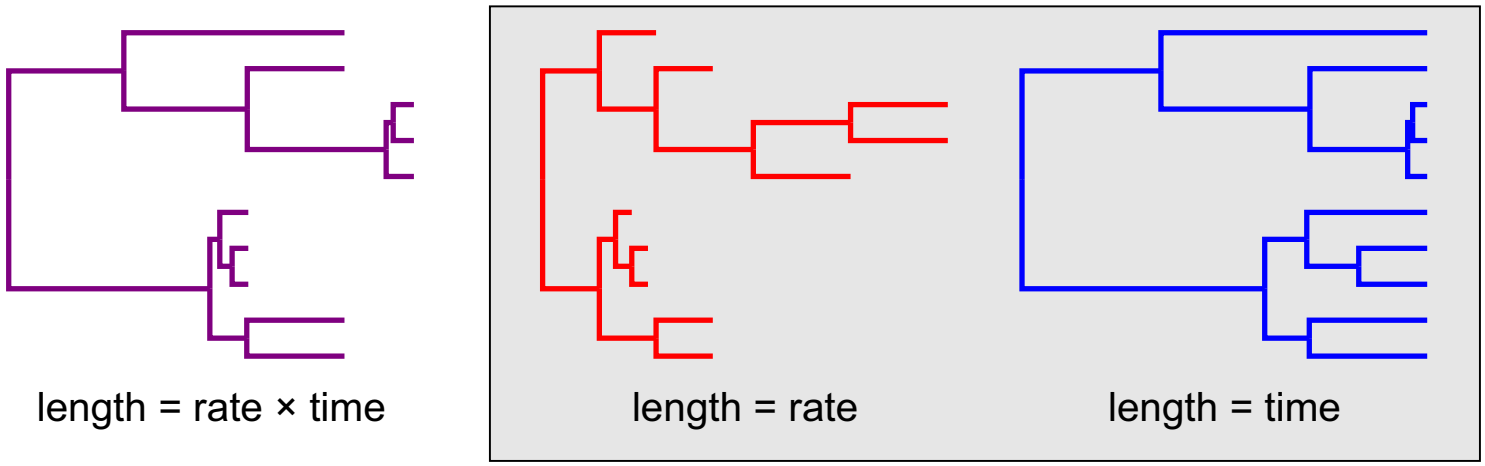



Anolis fowleri (image courtesy of L. Mahler)

(Mahler, Revell, Glor, & Losos. *Evolution* 2010)

Divergence Time Estimation

The expected # of substitutions/site occurring along a branch is the product of the substitution rate and time

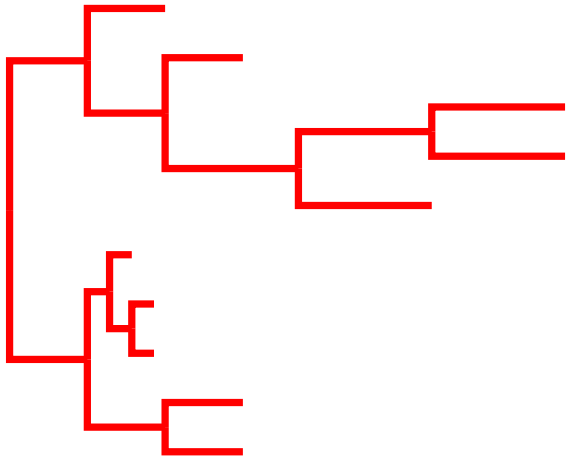


Methods for dating species divergences estimate the substitution rate and time separately

Substitution Rate

Substitution rate: the rate at which mutations are fixed in a population

Depends on: mutation rate, selection, population size, drift

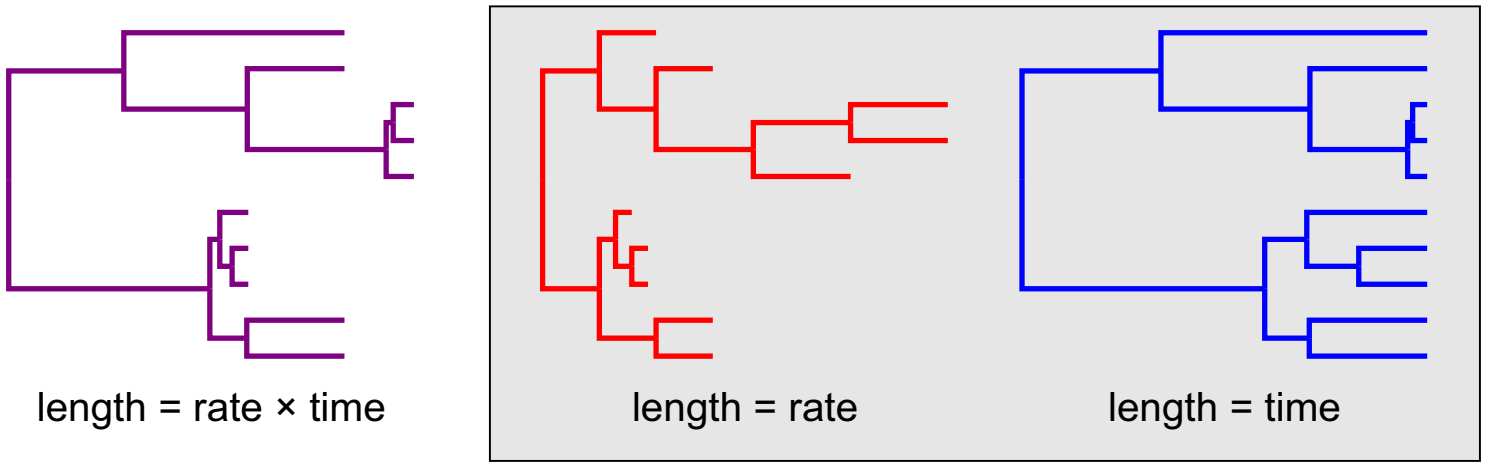


length = subst. rate

Mutation rate measures the rate at which mutations occur over time and is affected by metabolic rate, generation time, DNA repair efficiency

Divergence Time Estimation

The expected # of substitutions/site occurring along a branch is the product of the substitution rate and time

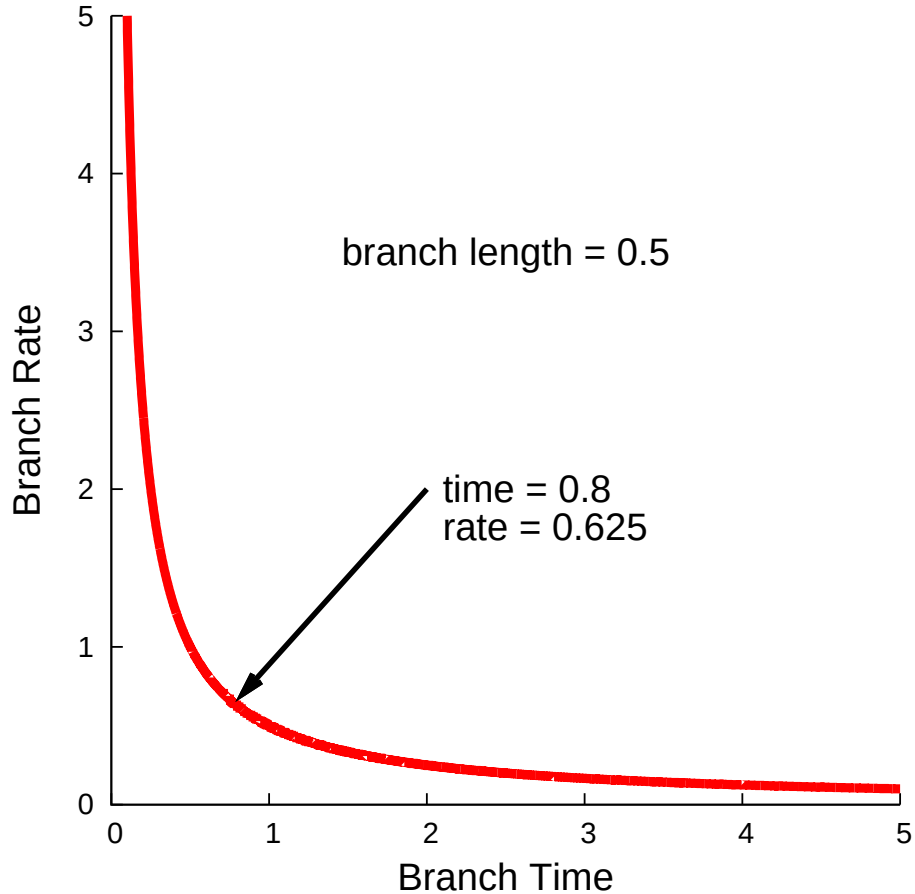


Methods for dating species divergences estimate the substitution rate and time separately

Rates and Times

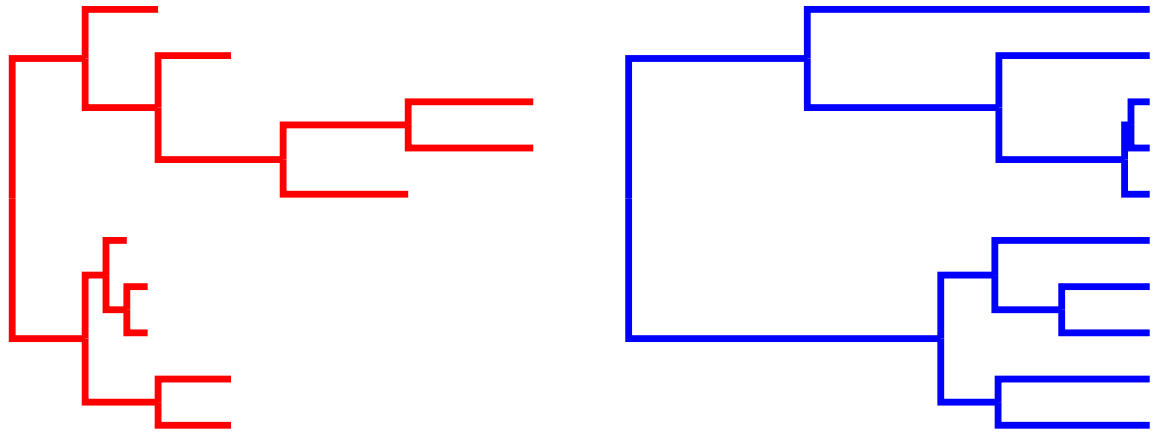
The sequence data provide information about branch length

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



(based on Thorne & Kishino, 2005)

Bayesian Divergence Time Estimation



length = rate

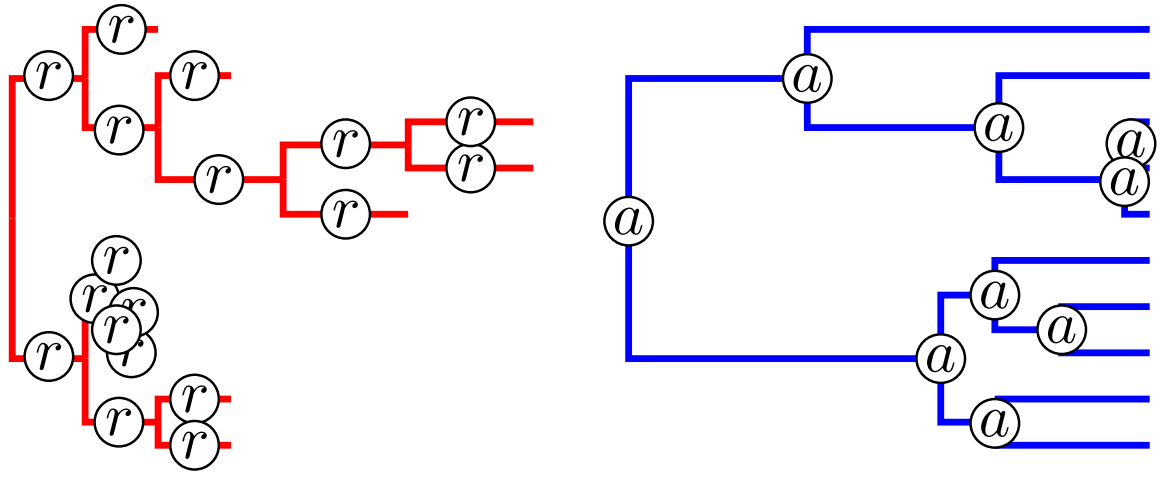
length = time

$$\boldsymbol{\mathcal{R}} = (r_1, r_2, r_3, \ldots, r_{2N-2})$$

$$\boldsymbol{\mathcal{A}} = (a_1, a_2, a_3, \ldots, a_{N-1})$$

$$N = \text{number of tips}$$

Bayesian Divergence Time Estimation



length = rate

length = time

$$\boldsymbol{\mathcal{R}} = (r_1, r_2, r_3, \ldots, r_{2N-2})$$

$$\boldsymbol{\mathcal{A}} = (a_1, a_2, a_3, \ldots, a_{N-1})$$

$$N = \text{number of tips}$$

Bayesian Divergence Time Estimation

Posterior probability

$$f(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s \mid D, \tau)$$

$\boldsymbol{\mathcal{R}}$	Vector of rates on branches
$\boldsymbol{\mathcal{A}}$	Vector of internal node ages
$\theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s$	Model parameters
D	Sequence data
τ	Tree topology (assumed known for the moment)

Bayesian Divergence Time Estimation

$$f(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s \mid D) = \frac{f(D \mid \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s) f(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s)}{f(D)}$$

$f(D \mid \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s)$	Likelihood
$f(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s)$	Joint prior density
$f(D)$	Marginal probability of the data

Bayesian Divergence Time Estimation

The likelihood depends on the node times and the rates of evolution, but not on the processes generating the rates and node times

$$f(D \mid \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s) = f(D \mid \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_s)$$

Bayesian Divergence Time Estimation

Assume that the process governing the ages of nodes operates independently of processes governing mutation, and that the process governing the total rates of substitutions is independent from the mutational parameters that determine relative rates of different substitutions:

$$f(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s) = f(\boldsymbol{\mathcal{R}} \mid \theta_{\mathcal{R}}) f(\boldsymbol{\mathcal{A}} \mid \theta_{\mathcal{A}}) f(\theta_{\mathcal{R}}) f(\theta_{\mathcal{A}}) f(\theta_s)$$

Bayesian Divergence Time Estimation

After enforcing these assumptions, the posterior distribution of the parameters and hyperparameters can be expressed as:

$$f(\boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s \mid D) \;=\; \frac{f\left(D \mid \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{A}}, \theta_s\right) f(\boldsymbol{\mathcal{R}} \mid \theta_{\mathcal{R}}) f(\boldsymbol{\mathcal{A}} \mid \theta_{\mathcal{A}}) f(\theta_{\mathcal{R}}) f(\theta_{\mathcal{A}}) f(\theta_s)}{f(D)}$$

Bayesian Divergence Time Estimation

Estimating divergence times relies on 2 main elements:

- Branch-specific rates: $f(\boldsymbol{\mathcal{R}} \mid \theta_{\mathcal{R}})$
- Node ages: $f(\boldsymbol{\mathcal{A}} \mid \theta_{\mathcal{A}}, \boldsymbol{\mathcal{C}})$

Modeling Rate Variation

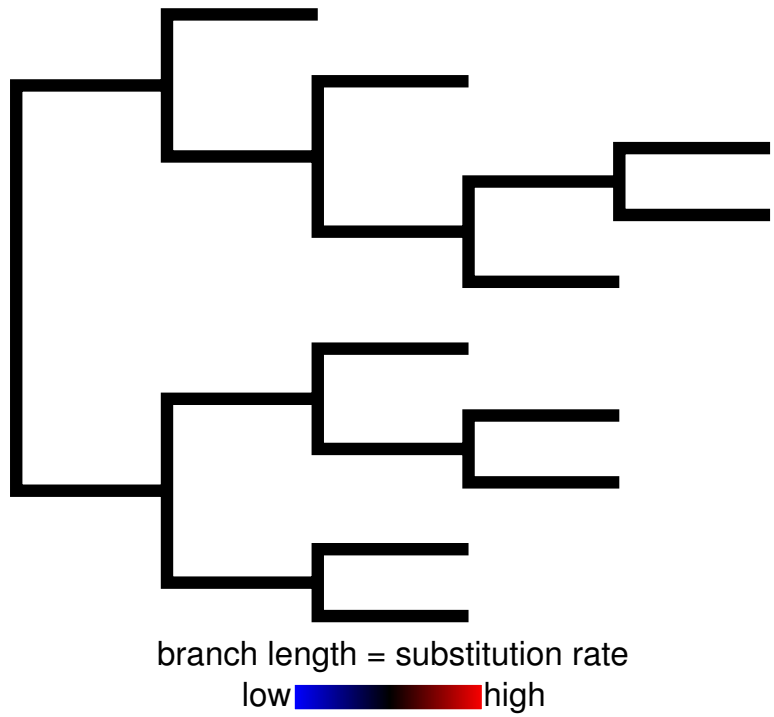
Some models describing lineage-specific substitution rate variation:

- **Global molecular clock** (Zuckerkandl & Pauling, 1962)
- **Local molecular clocks** (Hasegawa, Kishino % Yano 1989; Kishino & Hasegawa 1990; Yoder & Yang 2000; Yang & Yoder 2003, Drummond and Suchard 2010)
- **Compound Poisson process model** (Huelsenbeck, Larget and Swofford 2000)
- **Log-normally distributed autocorrelated rates** (Thorne, Kishino & Painter 1998; Kishino, Thorne & Bruno 2001; Thorne & Kishino 2002)
- **Uncorrelated/independent rates models** (Drummond et al. 2006; Rannala & Yang 2007; Lepage et al. 2007; Heath, Holder, Huelsenbeck 2012)

Global Molecular Clock

The substitution rate is constant over time

All lineages share the same rate



(Zuckerkandl & Pauling, 1962)

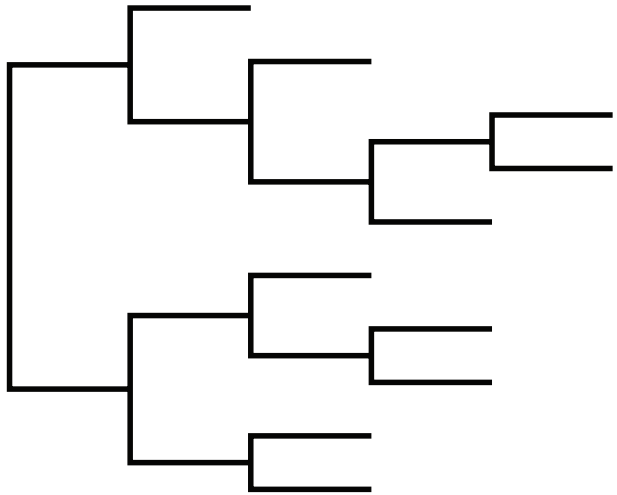
Global Molecular Clock

Assume the clock rate is gamma-distributed

$$\boldsymbol{\mathcal{R}} = (r, r, \ldots, r)$$

$$r \sim \text{Gamma}(\alpha, \lambda)$$

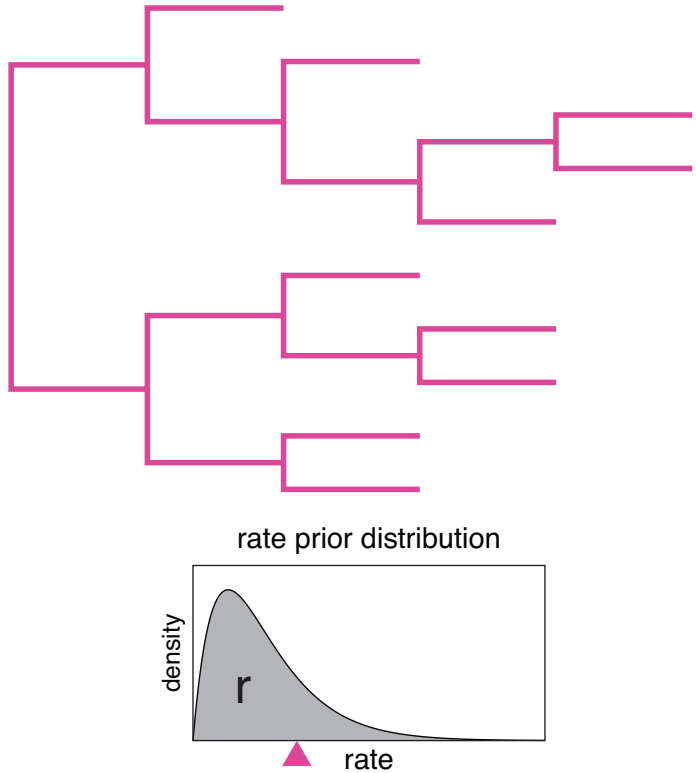
$$f(\boldsymbol{\mathcal{R}} \mid \theta_{\mathcal{R}}) = f(r \mid \alpha, \lambda)$$



(Zuckerkandl & Pauling, 1962)

Global Molecular Clock

The sampled rate is applied to every branch in the tree



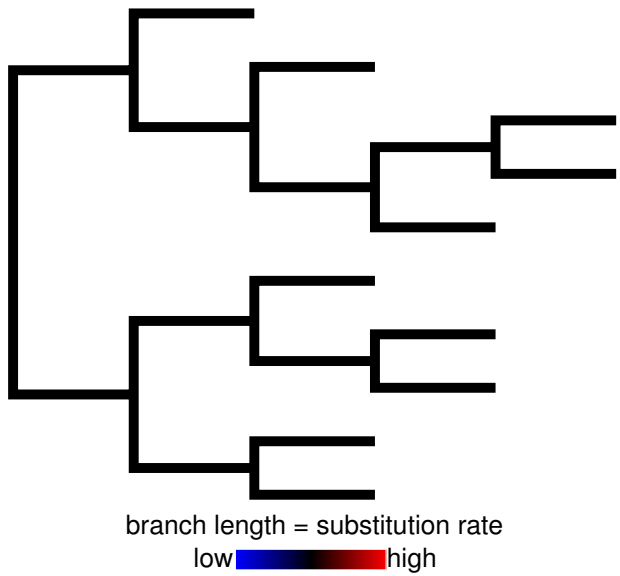
(Zuckerkandl & Pauling, 1962)

Rejecting the Global Molecular Clock

Rates of evolution vary across lineages and over time (and how!)

Mutation rate:
Variation in

- metabolic rate
- generation time
- DNA repair

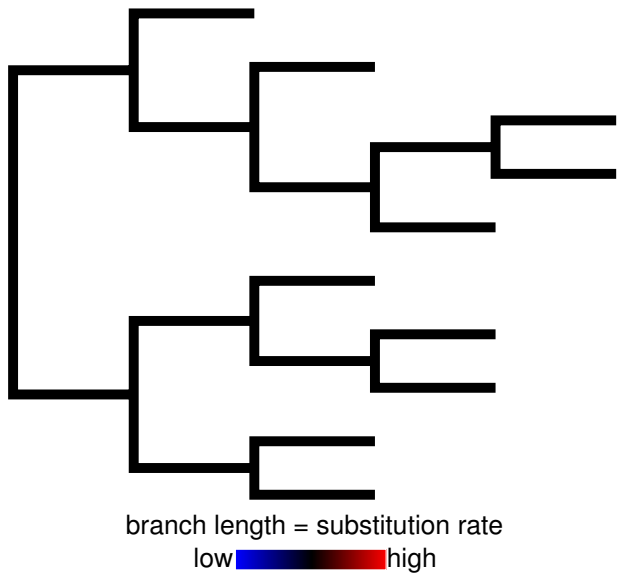


Rejecting the Global Molecular Clock

Rates of evolution vary across lineages and over time (and how!)

Fixation rate:
Variability in

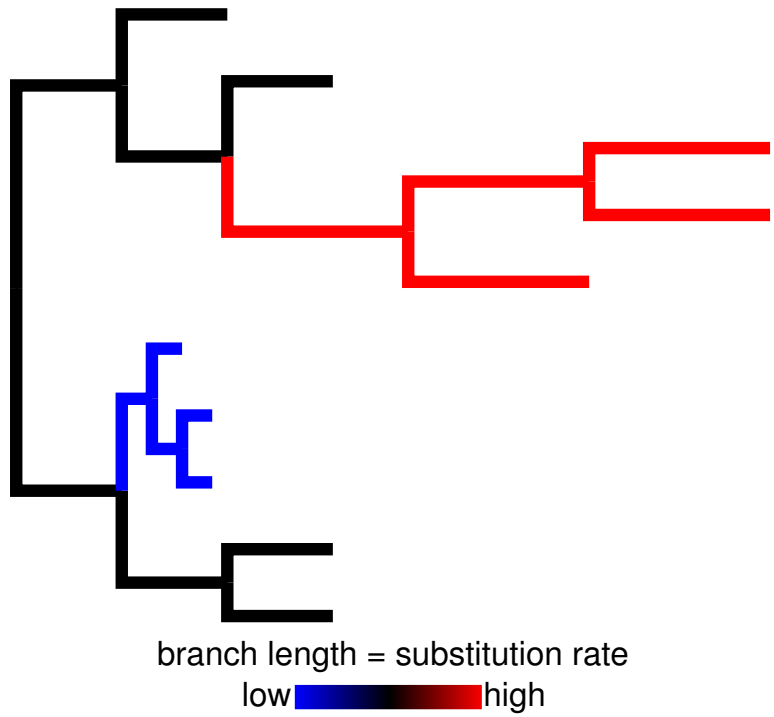
- strength and targets of selection
- population sizes



Local Molecular Clocks

Rate shifts occur infrequently over the tree

Closely related lineages have equivalent rates (clustered by sub-clades)

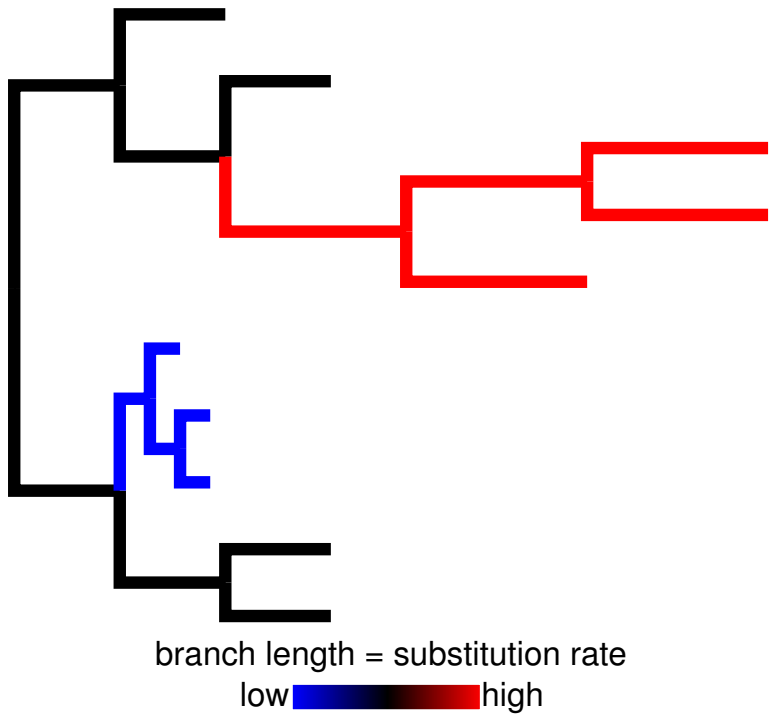


(Yang & Yoder 2003, Drummond and Suchard 2010)

Local Molecular Clocks

Most methods for estimating local clocks required specifying the number and locations of rate changes *a priori*

Drummond and Suchard (2010) introduced a Bayesian method that samples over a broad range of possible *random local clocks*

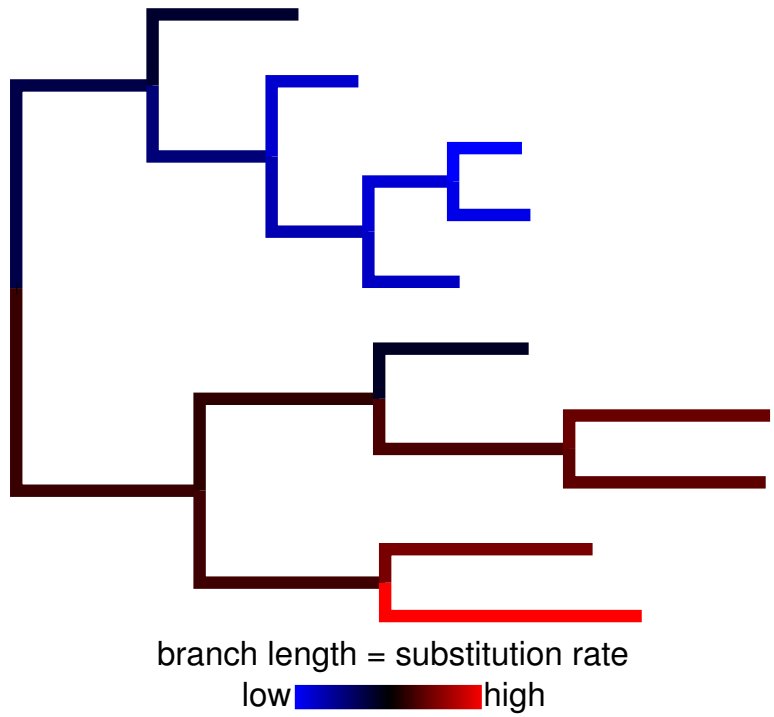


(Yang & Yoder 2003, Drummond and Suchard 2010)

Autocorrelated Rates

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

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



(Thorne, Kishino & Painter 1998; Kishino, Thorne & Bruno 2001)

Autocorrelated Rates

$$\boldsymbol{\mathcal{R}} = (r_1, r_2, \ldots, r_{2N-1})$$

$$\sigma^2 = \psi * \Delta t$$

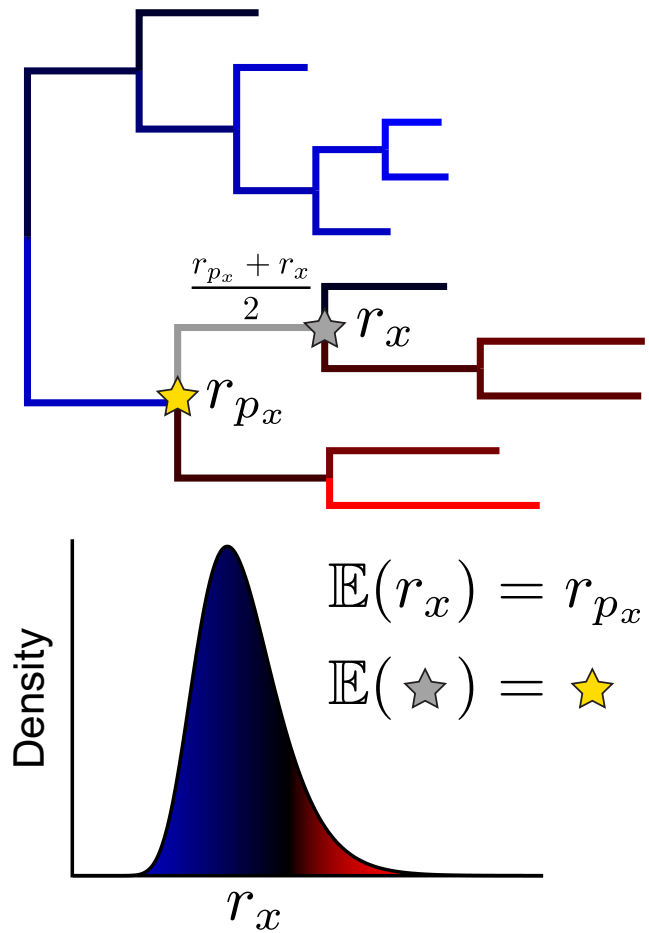
$$\mu = ln(r_{p_i}) - \frac{\sigma^2}{2}$$

$$r_i \sim \text{Lognormal}(\mu, \sigma^2)$$

$$f(\boldsymbol{\mathcal{R}} \mid \theta_{\mathcal{R}}) = f(\boldsymbol{\mathcal{R}} \mid \psi, \boldsymbol{\mathcal{A}}, r_{root})$$

ψ is the variance parameter

Δt is the difference in time between the 2 nodes

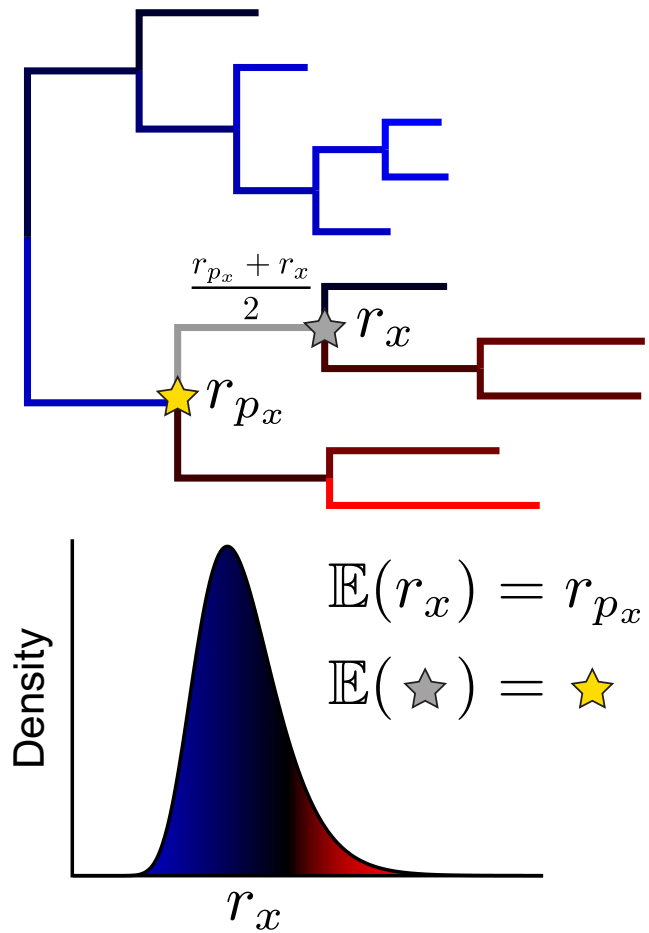


(Thorne, Kishino & Painter 1998; Kishino, Thorne & Bruno 2001)

Autocorrelated Rates

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

The rate for the branch is equal to the mean of the two subtending nodes

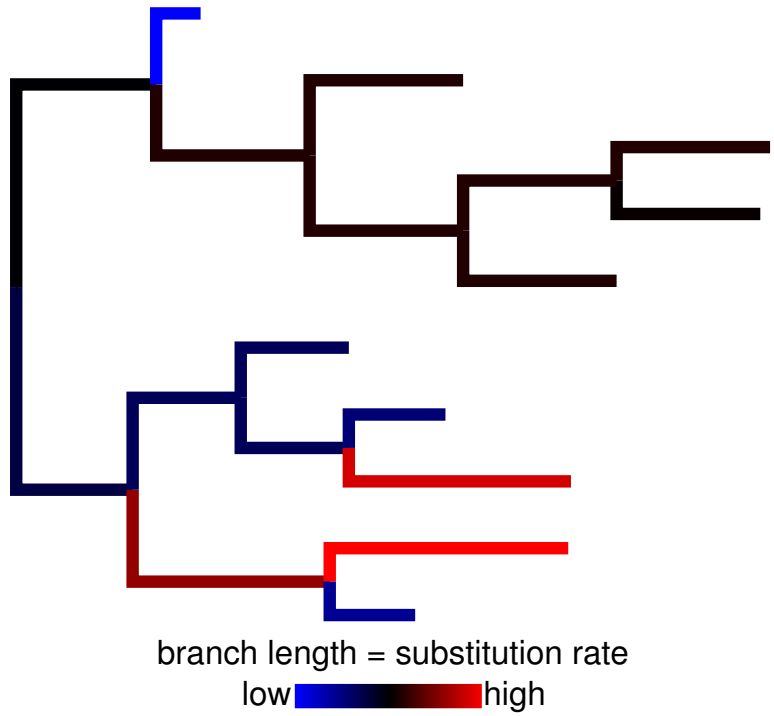


(Thorne, Kishino & Painter 1998; Kishino, Thorne & Bruno 2001)

Compound Poisson Process

Rate changes occur along lineages according to a point process

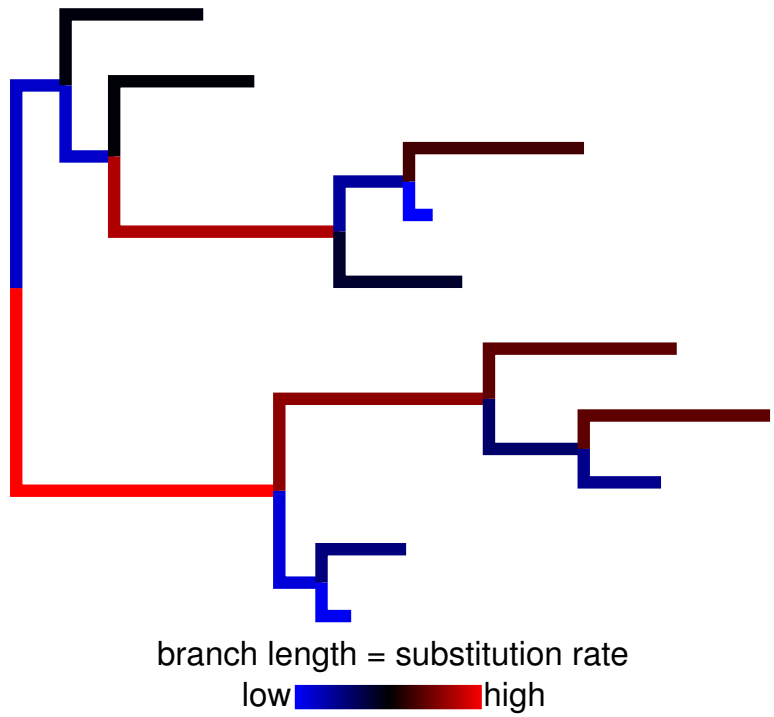
At rate-change events, the new rate is a product of the parent's rate and a Γ-distributed multiplier



(Huelsenbeck, Larget and Swofford 2000)

Independent/Uncorrelated Rates

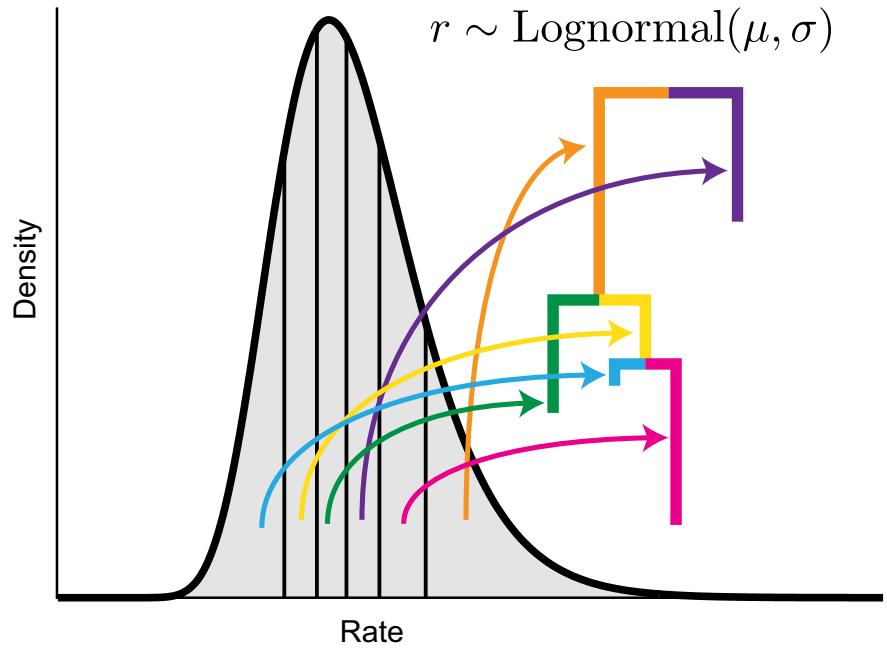
Lineage-specific rates are uncorrelated when the rate assigned to each branch is independently drawn from an underlying distribution



(Drummond et al. 2006)

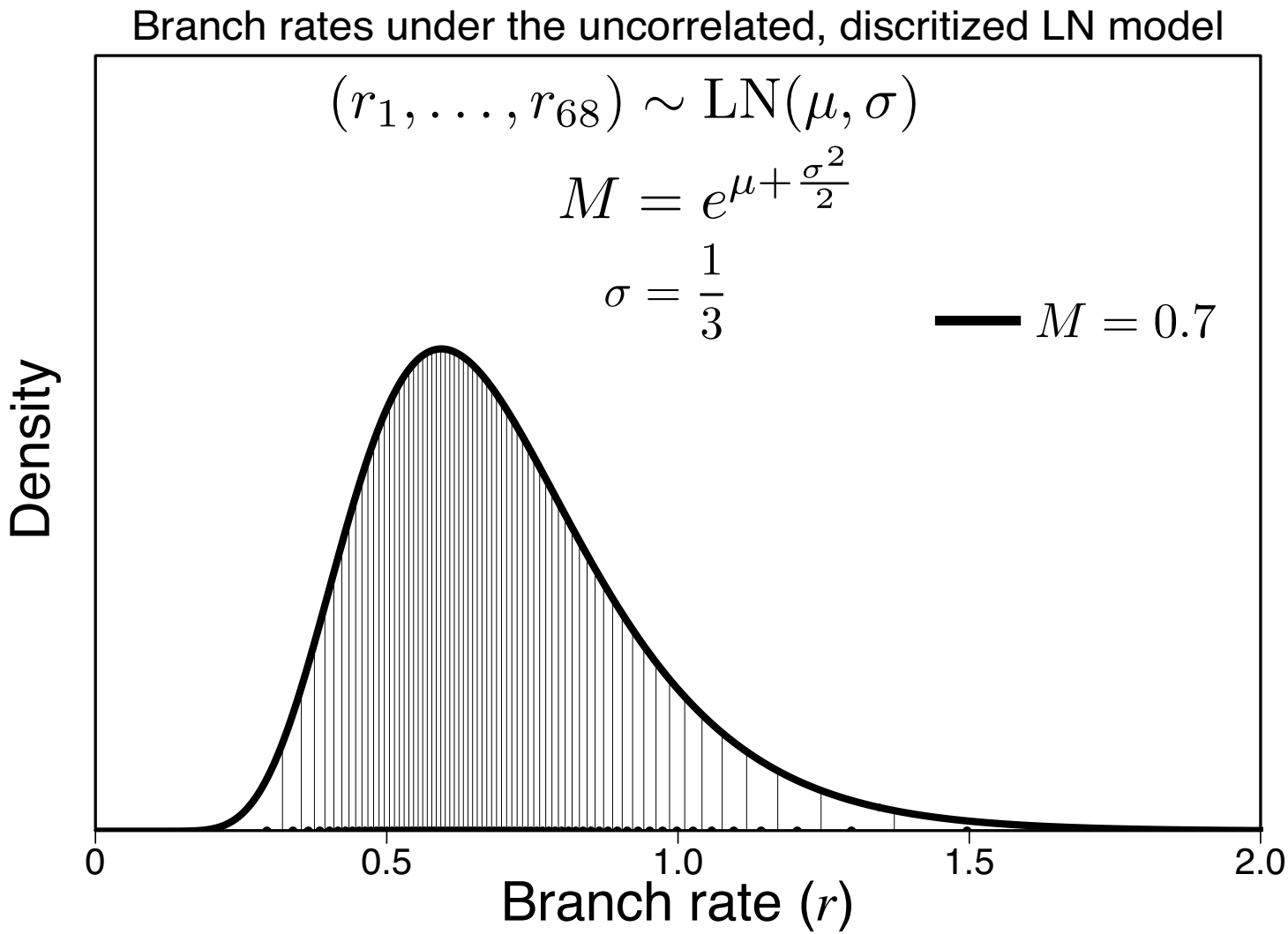
Independent/Uncorrelated Rates

In BEAST, the rates for the branches are drawn from a discretized lognormal distribution



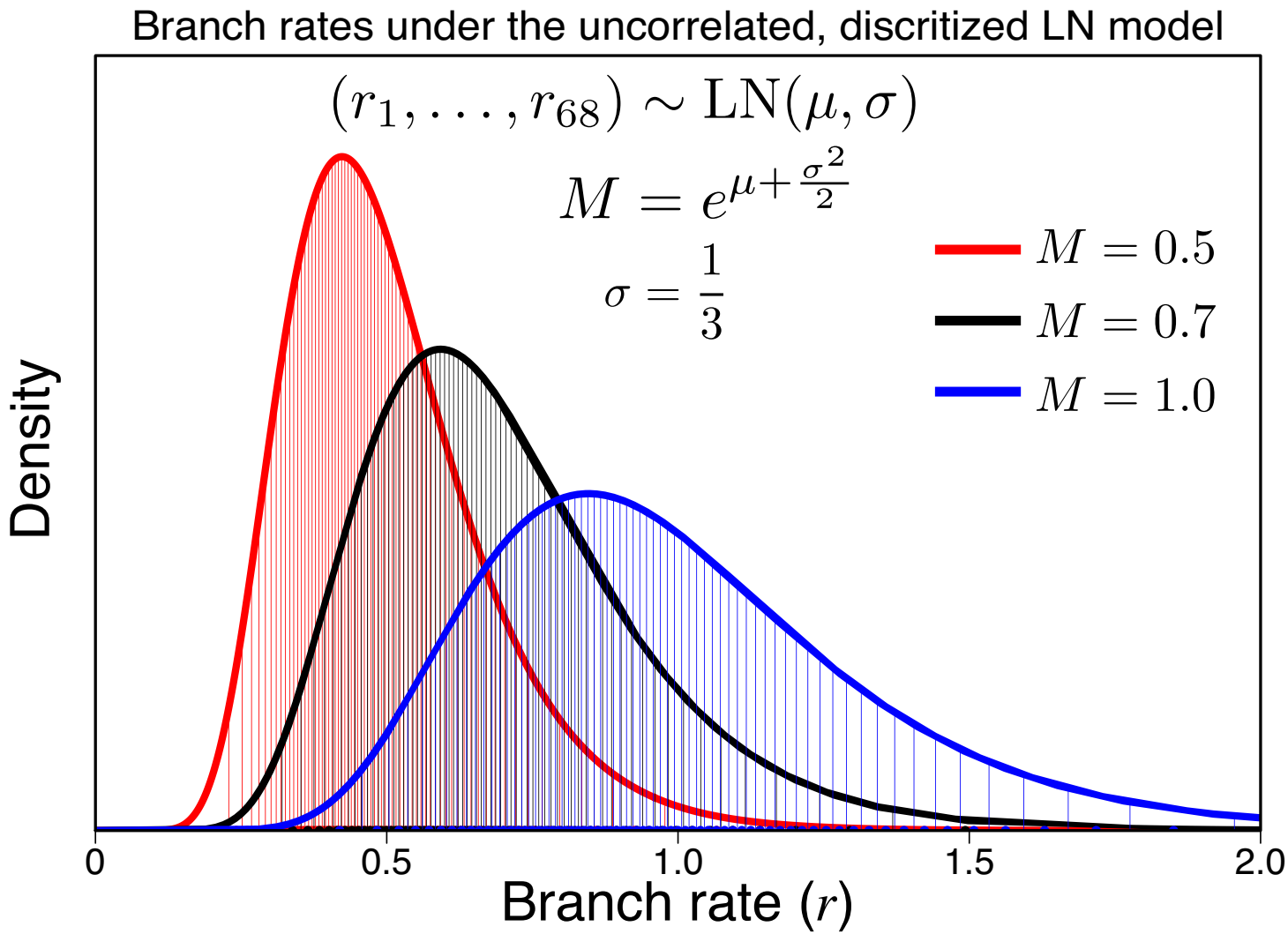
(Drummond et al. 2006)

Independent/Uncorrelated Rates



(Drummond et al. 2006)

Independent/Uncorrelated Rates

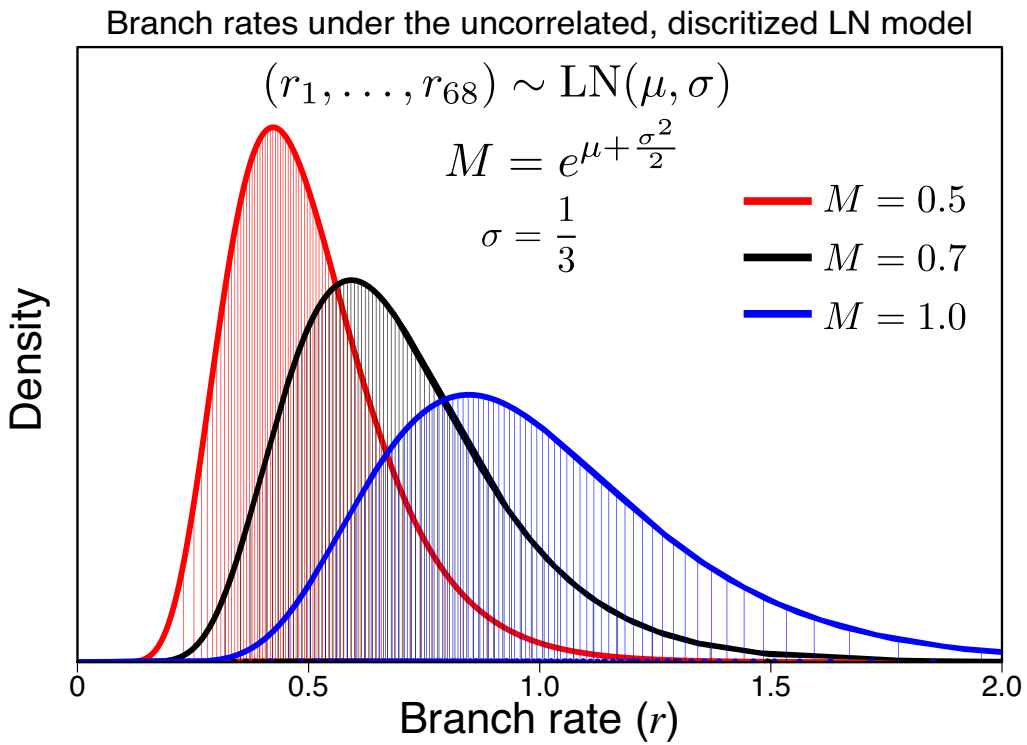


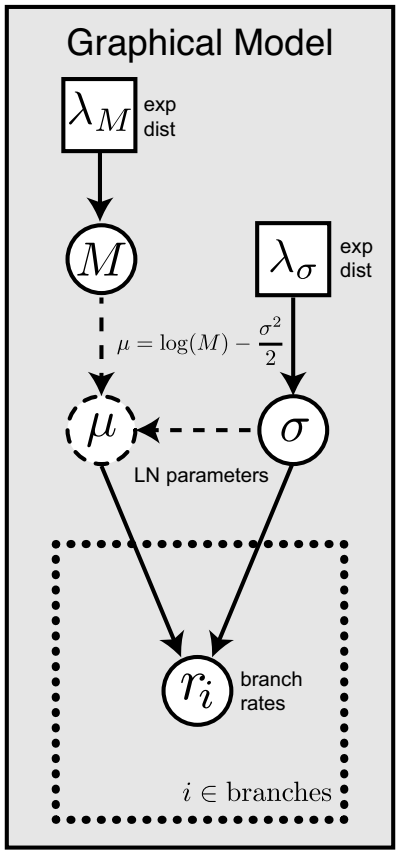
(Drummond et al. 2006)

Independent/Uncorrelated Rates

It is necessary to sample the parameters of the base distribution when assuming a discretized model

We can do this using a hierarchical model



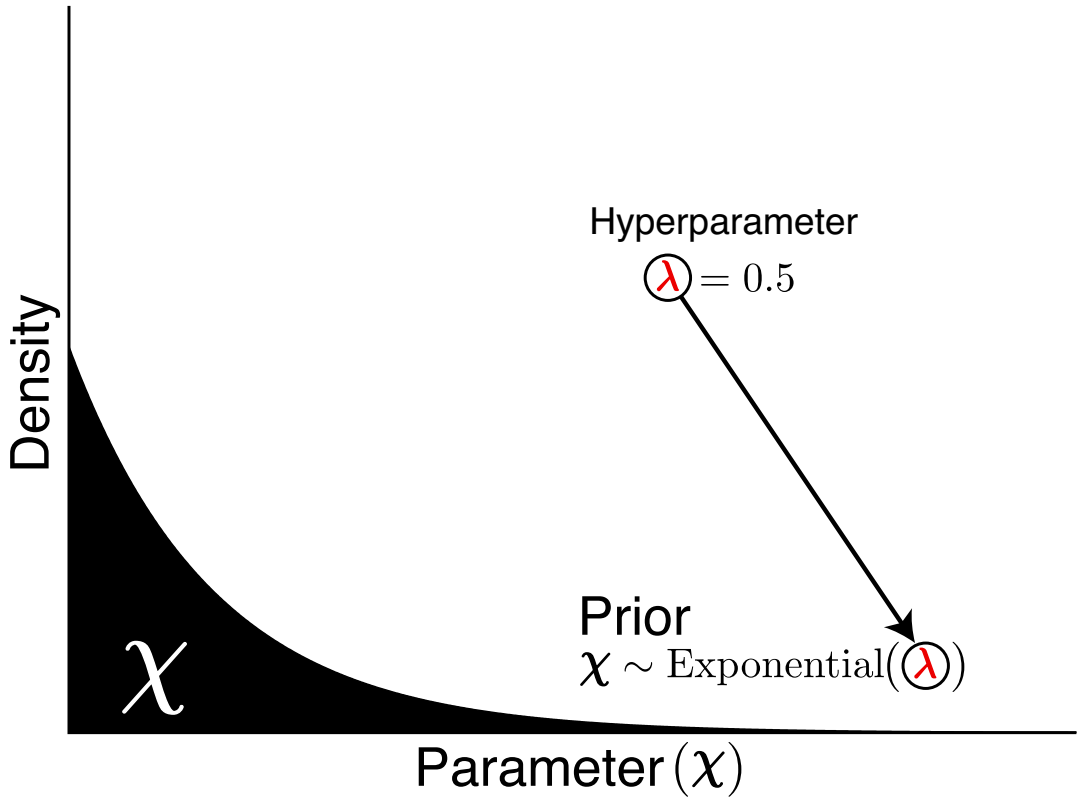


$$\mathbb{E}(M) = \lambda_M^{-1}$$

A Hierarchical Bayesian Model

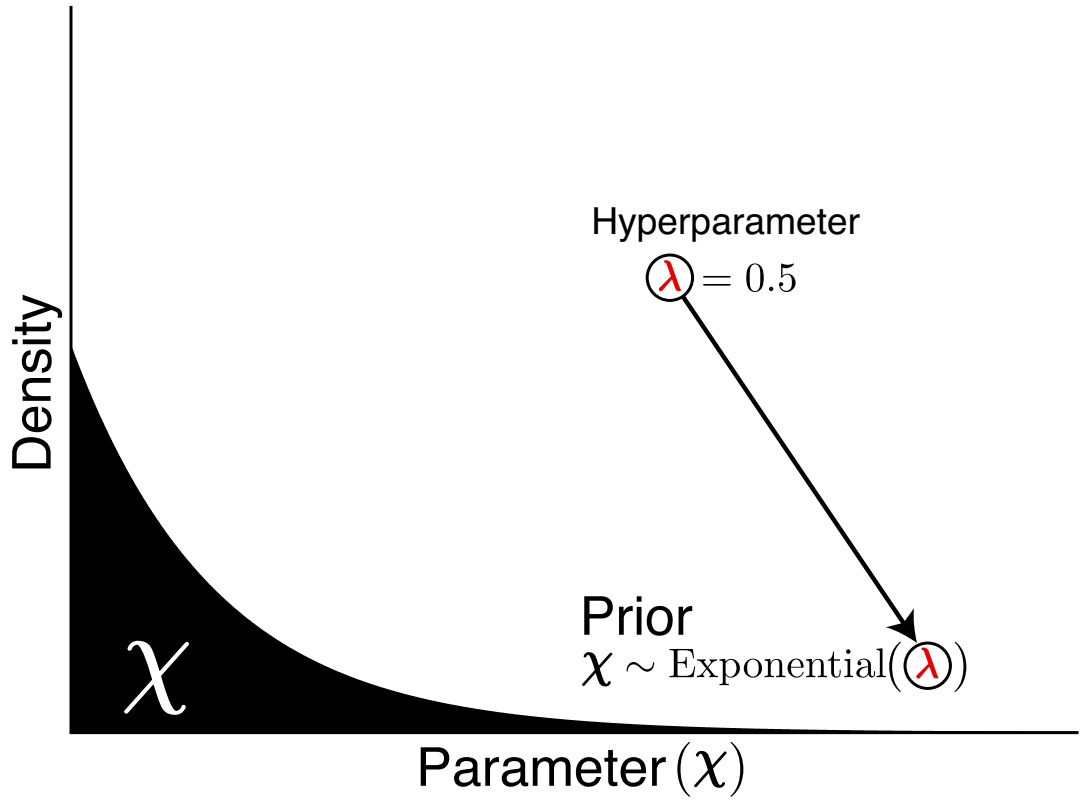
From the bottom up:

The parameter χ is assumed to be drawn from an exponential distribution



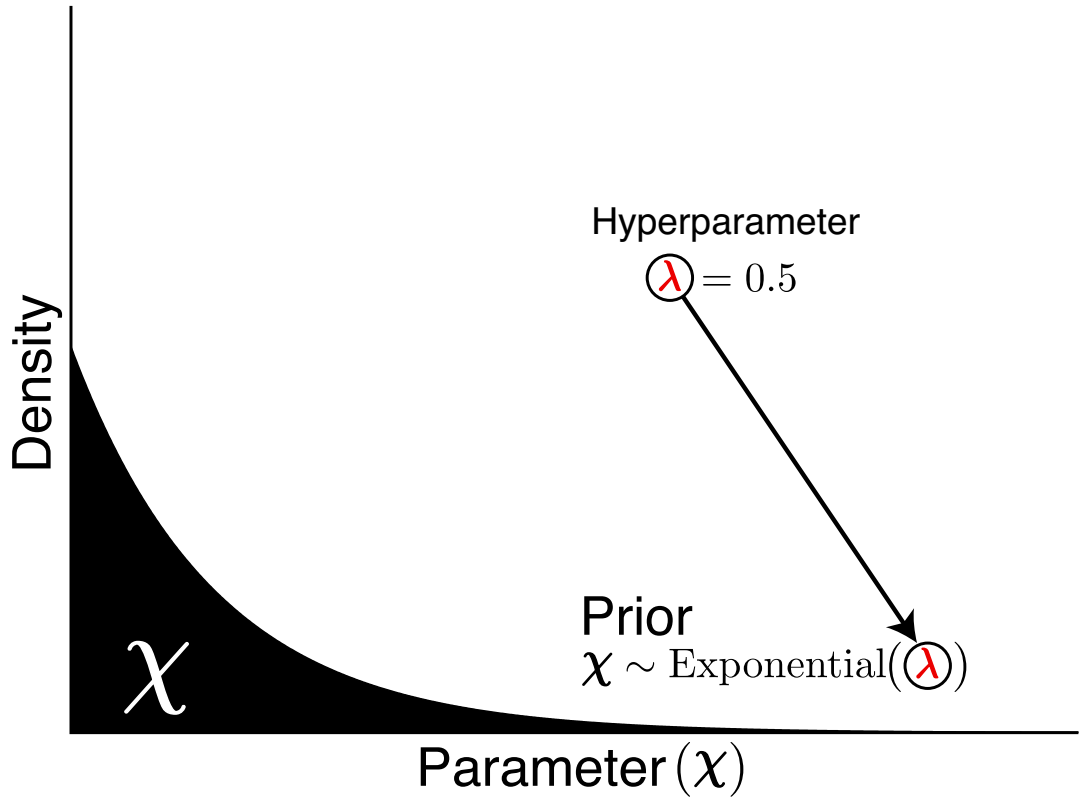
A Hierarchical Bayesian Model

In Bayesian inference, a parameter describing a prior distribution is called a **hyperparameter**



A Hierarchical Bayesian Model

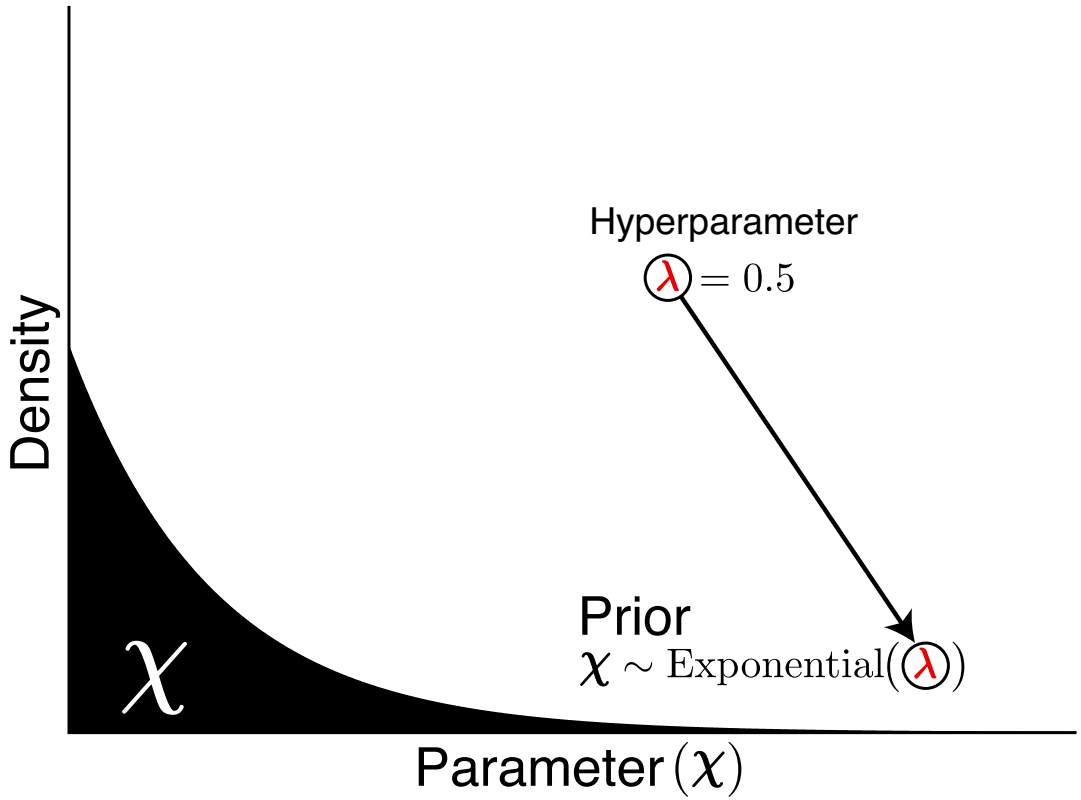
The **exponential** prior on χ has a hyperparameter: λ



A Hierarchical Bayesian Model

λ represents the **rate** of the exponential distribution

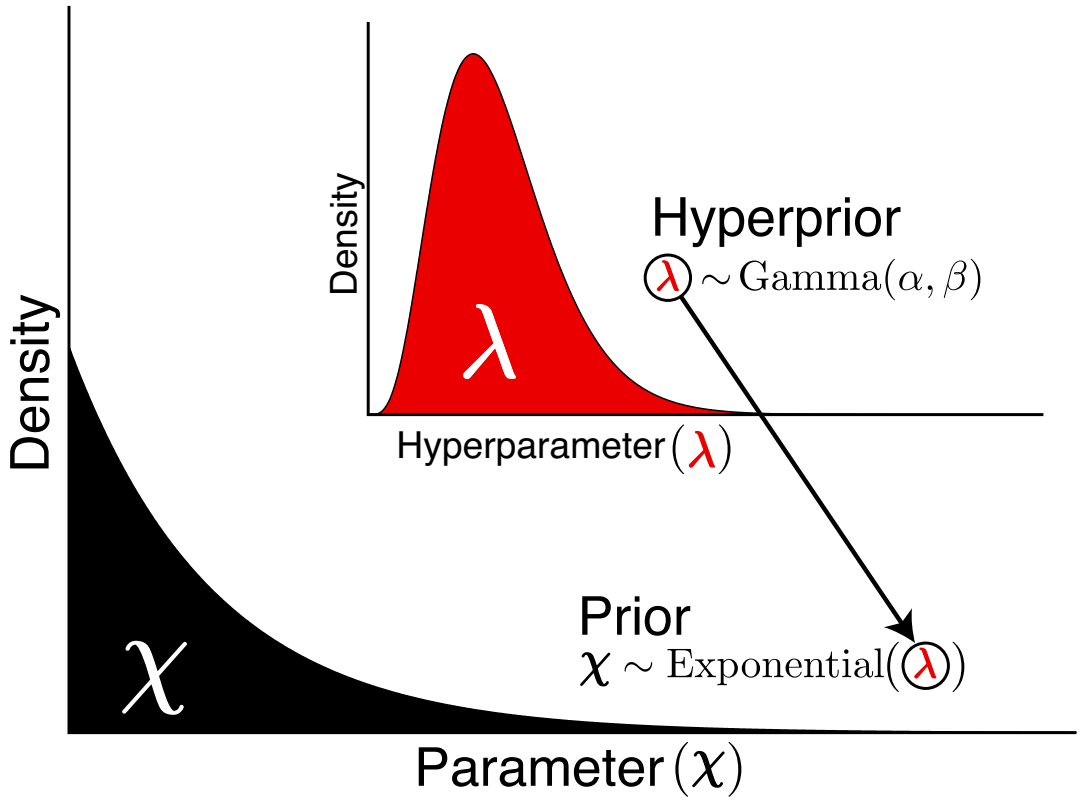
In a non-hierarchical model, the user is required to specify the value of λ



A Hierarchical Bayesian Model

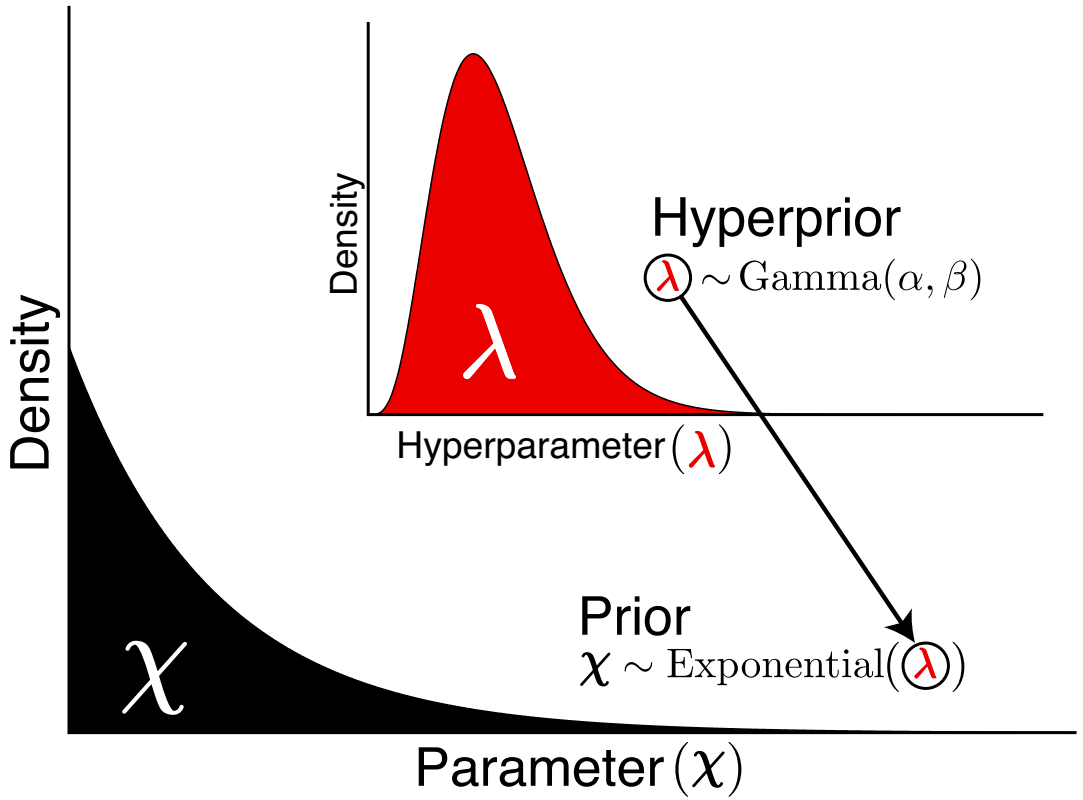
Hyperprior:
second order prior placed on a hyperparameter

λ becomes a random variable under the hierarchical model



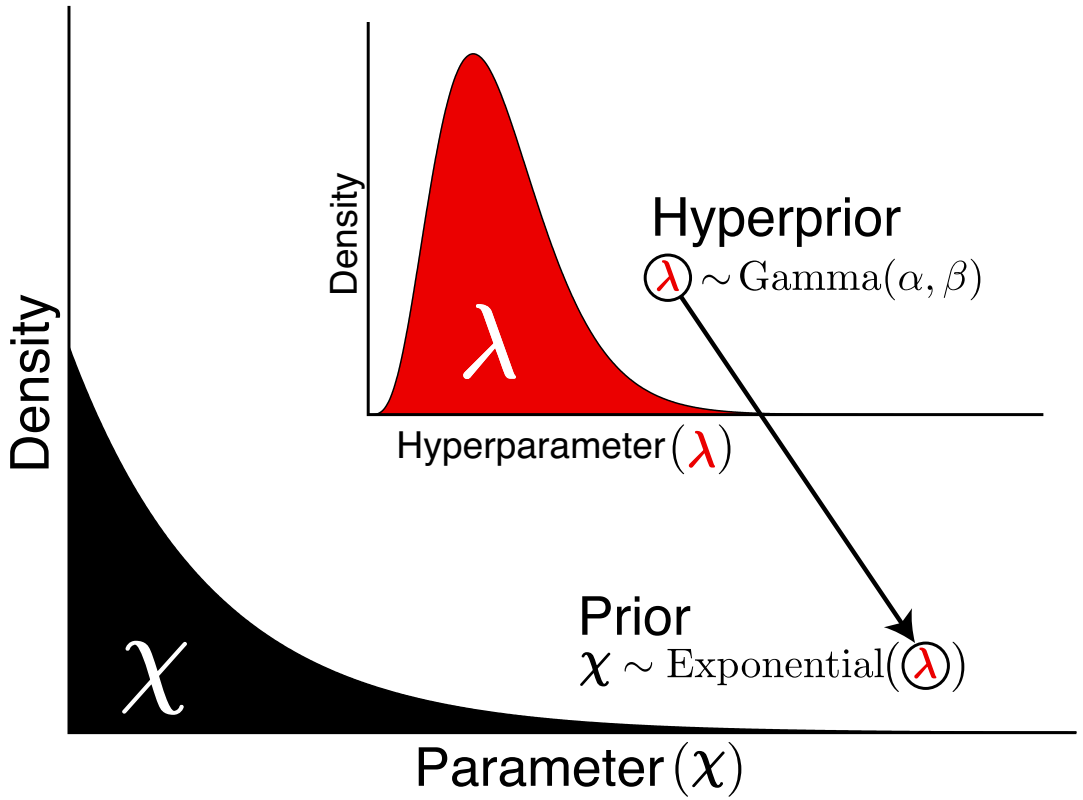
A Hierarchical Bayesian Model

Hyperprior:
allows for inference under a richer class of models



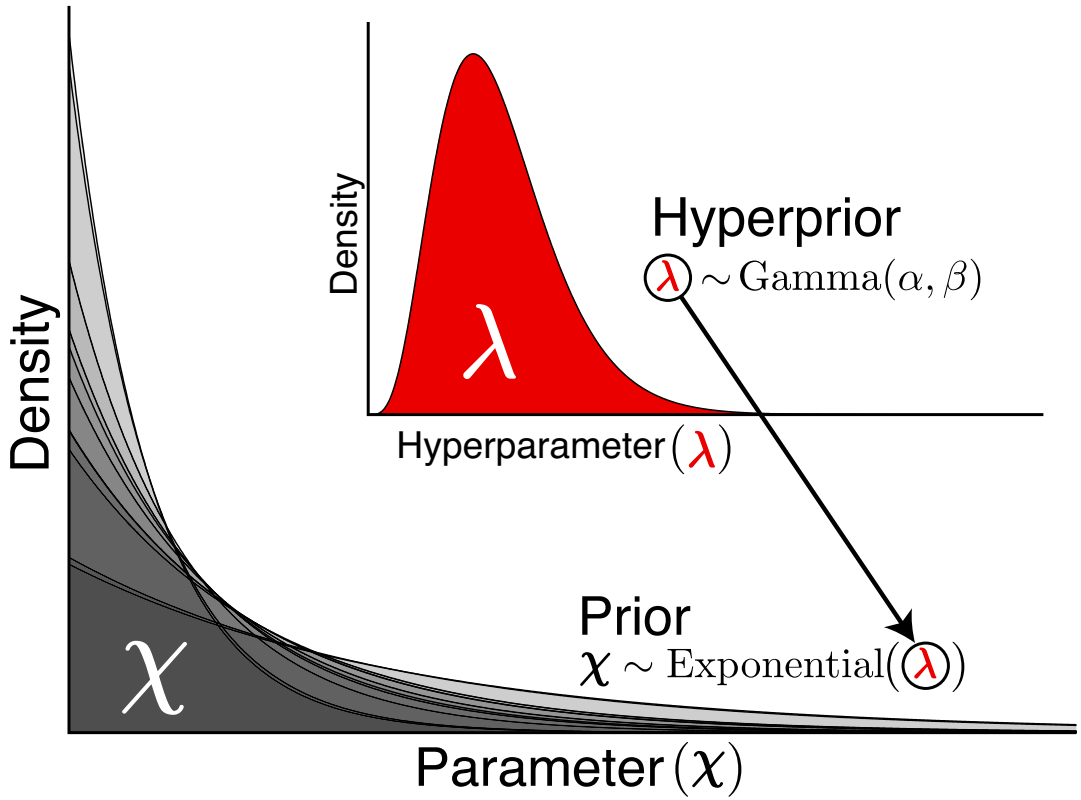
A Hierarchical Bayesian Model

Hyperprior:
frees the user from the difficulty of specifying the value of λ



A Hierarchical Bayesian Model

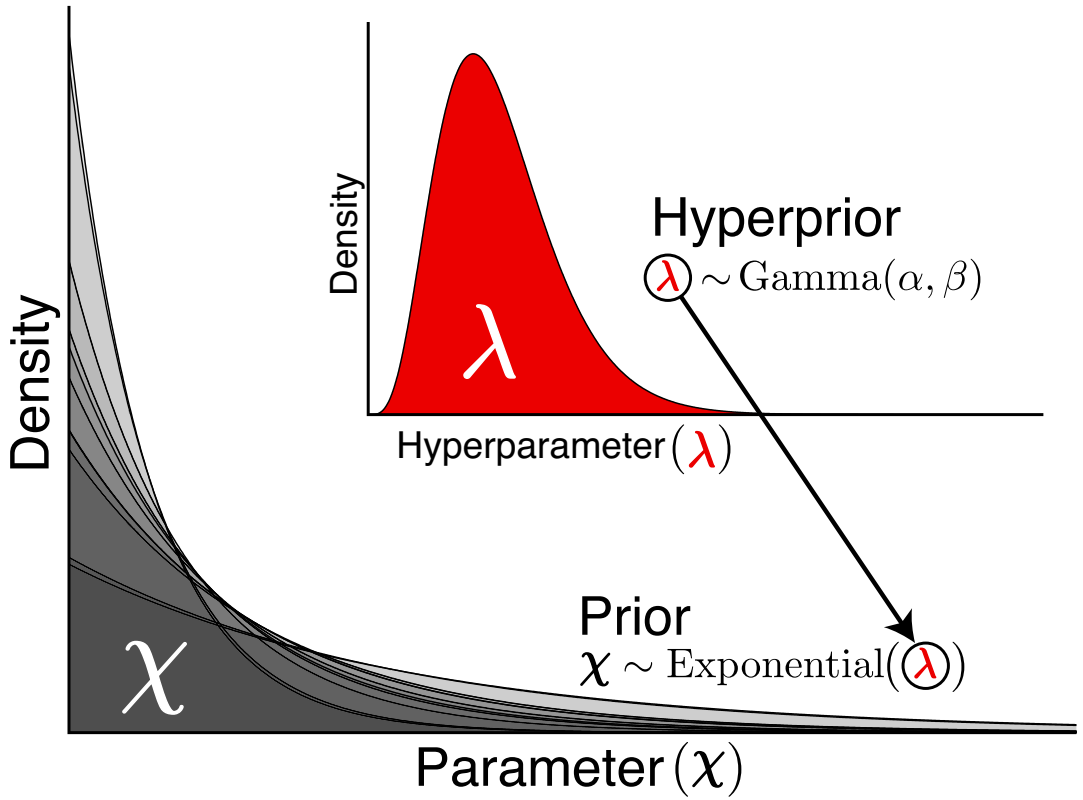
Hyperprior:
values of χ are sampled by MCMC from a mixture of exponential distributions



A Hierarchical Bayesian Model

Hyperprior:
provides estimates of the hyperparameter

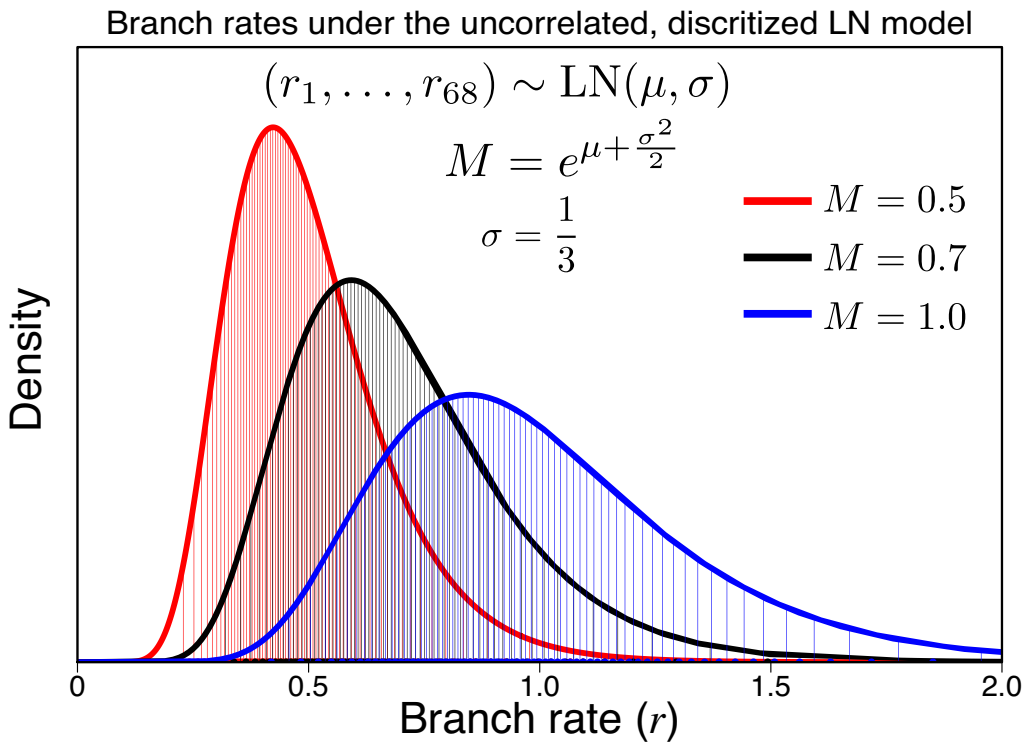
accounts for and quantifies uncertainty in the hyperparameter

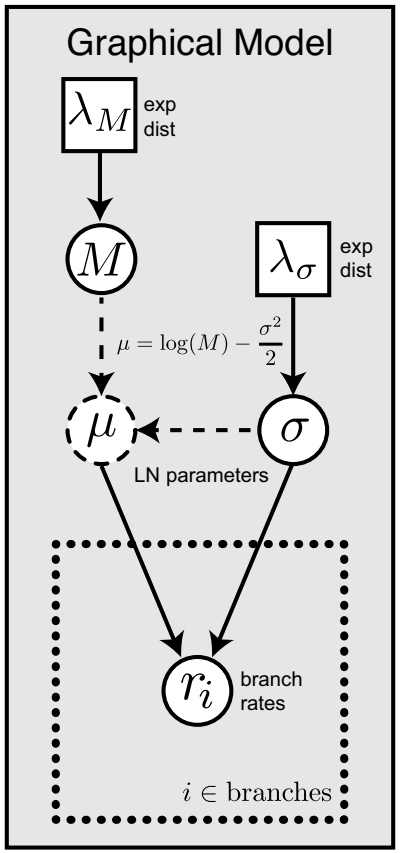


Independent/Uncorrelated Rates

It is necessary to sample the parameters of the base distribution when assuming a discretized model

We can do this using a hierarchical model

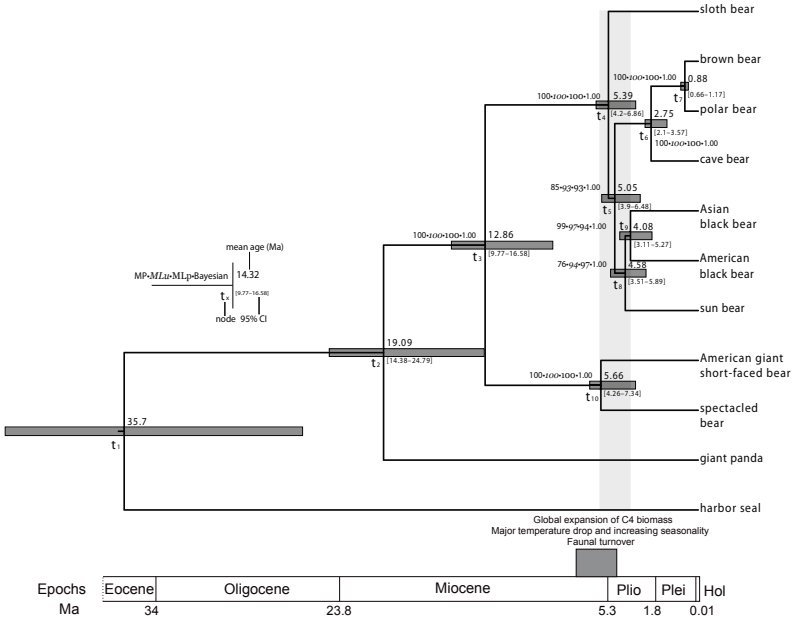




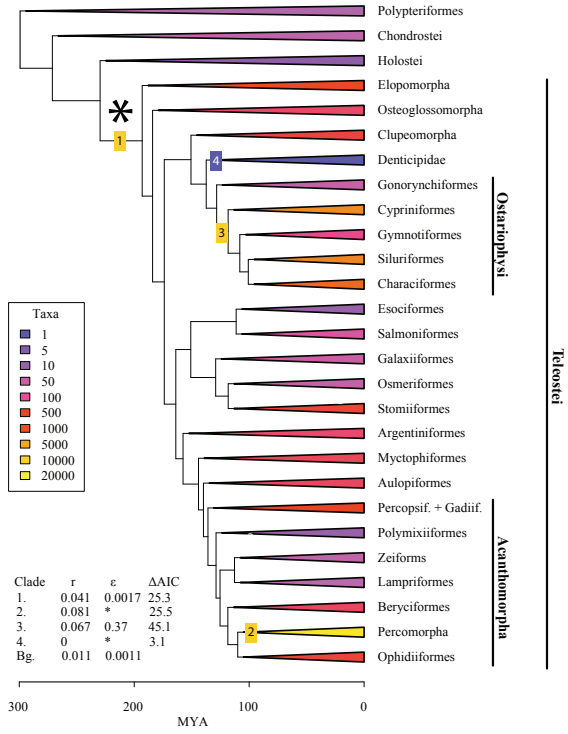
$$\mathbb{E}(M) = \lambda_M^{-1}$$

Modeling Rate Variation

Are our models appropriate across all data sets?



Krause et al., 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. BMC Evol. Biol. 8.



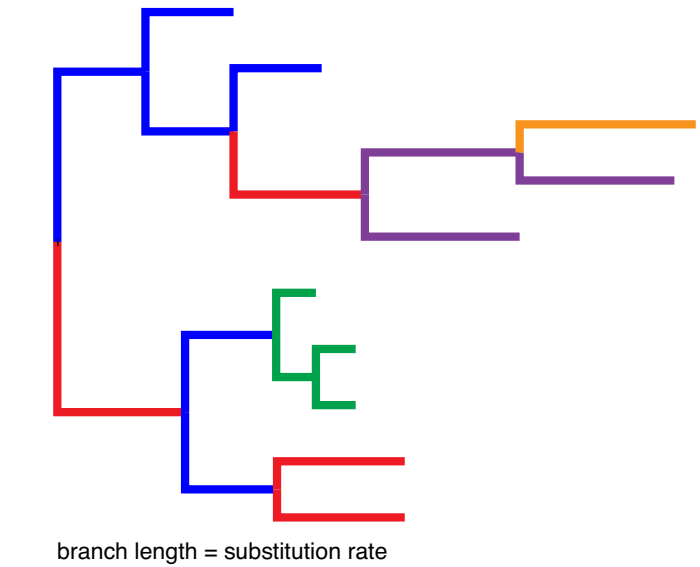
Santini et al., 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. BMC Evol. Biol. 9.

The Dirichlet Process Prior (DPP)

A stochastic process that models data as a mixture of distributions and can identify latent classes present in the data

Branches are assumed to be clustered into distinct substitution rate classes

$(r_1, \ldots, r_{2N-2}) \sim \text{DPP}(\alpha, G_0)$



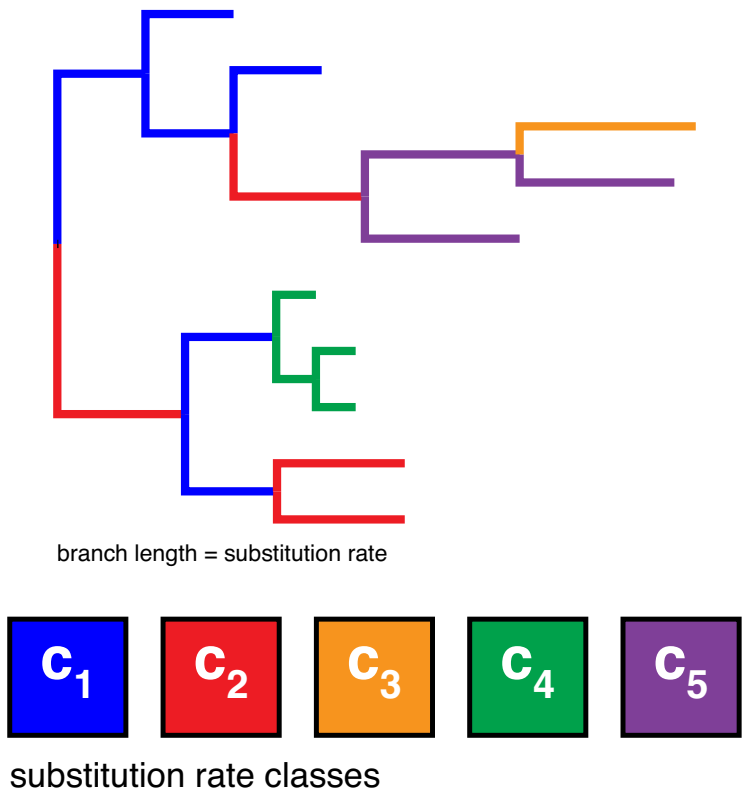
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

The concentration parameter: α
controls partitioning of branches into specific rate categories

Random variables under the DPP:

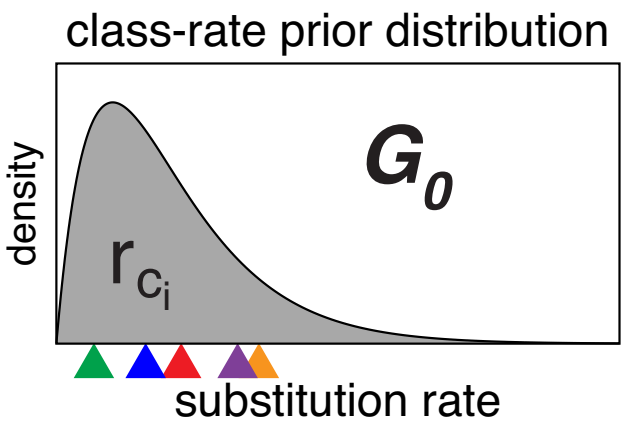
- k = the number of rate classes
- the assignment of branches to classes



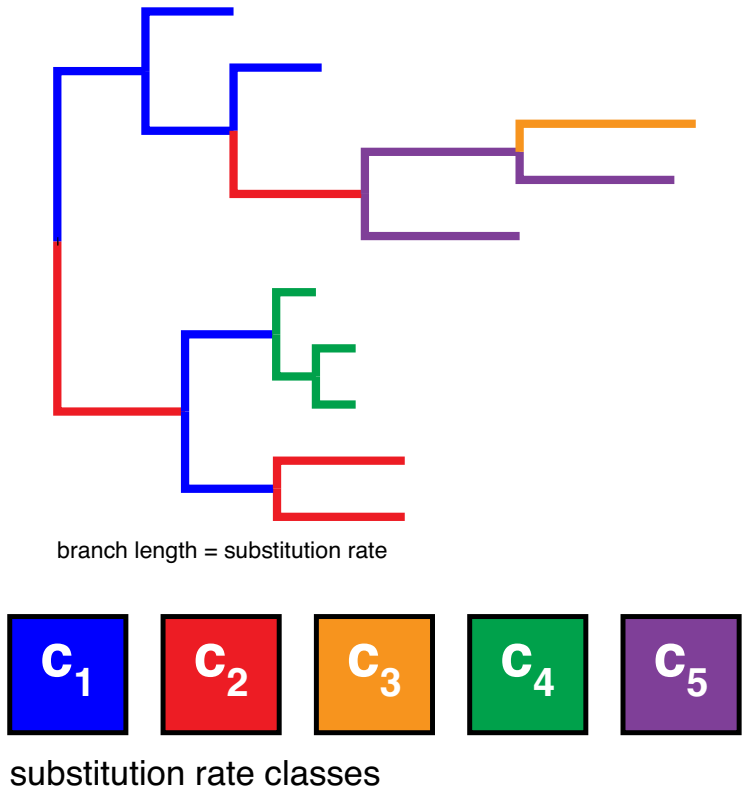
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

G_0 represents the parametric distribution from which substitution rates are drawn for each category



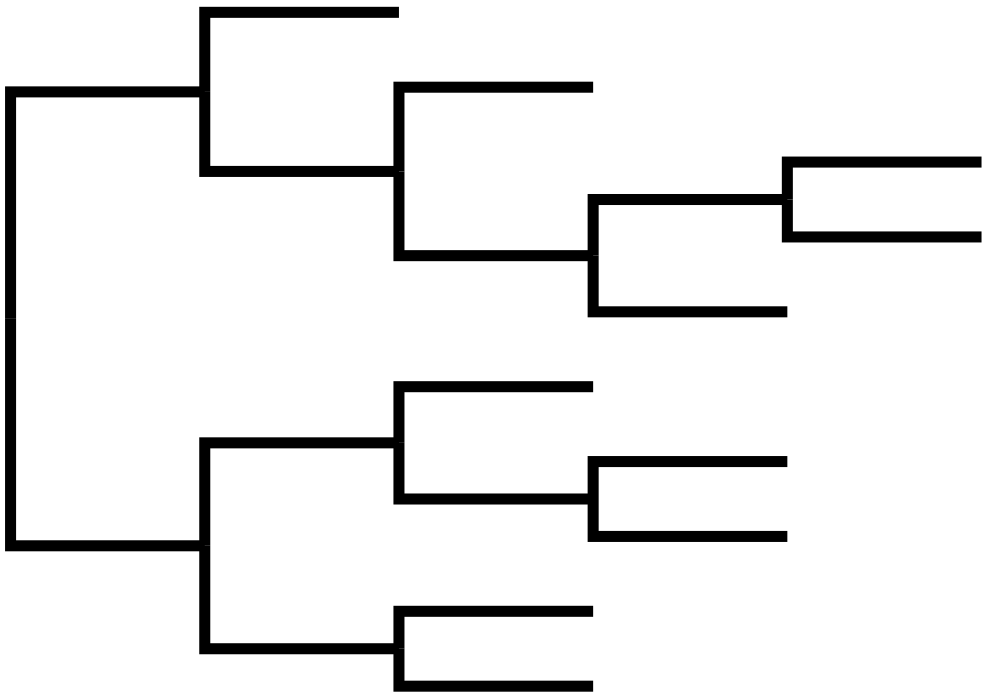
r_{c_i} = the rate value for each class



Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

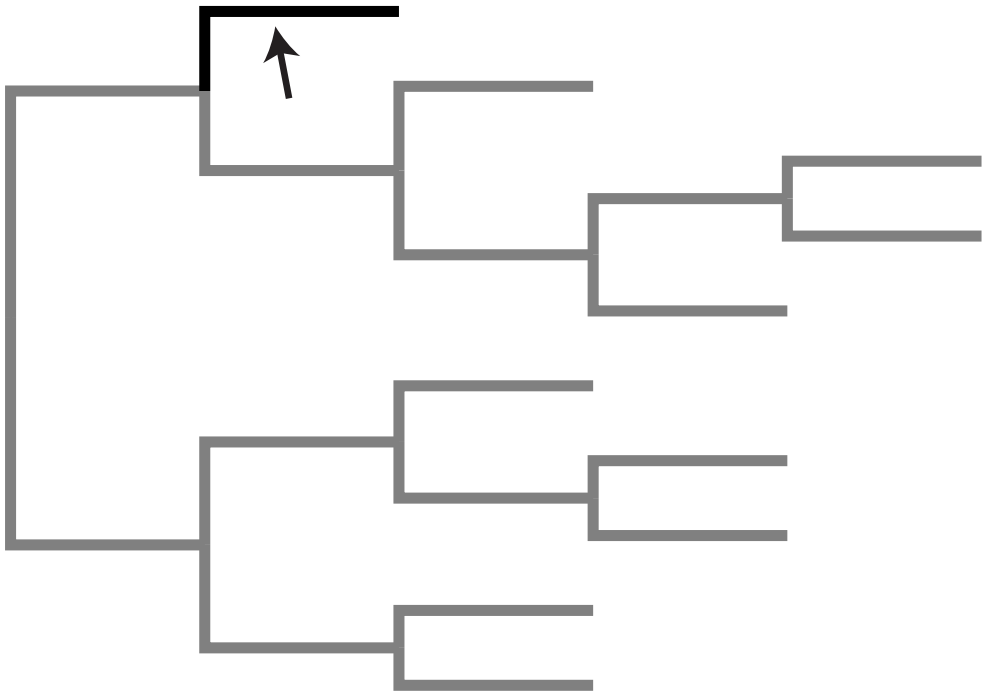
branch length = substitution rate



Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

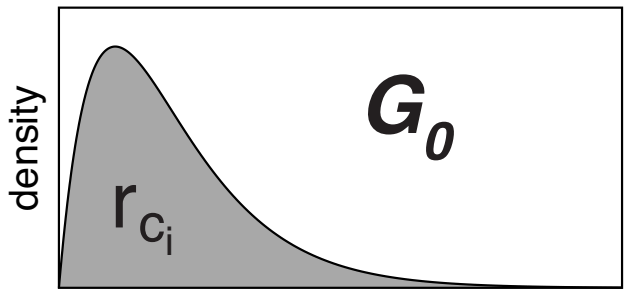
The Dirichlet Process Prior (DPP)

branch length = substitution rate



rate classes

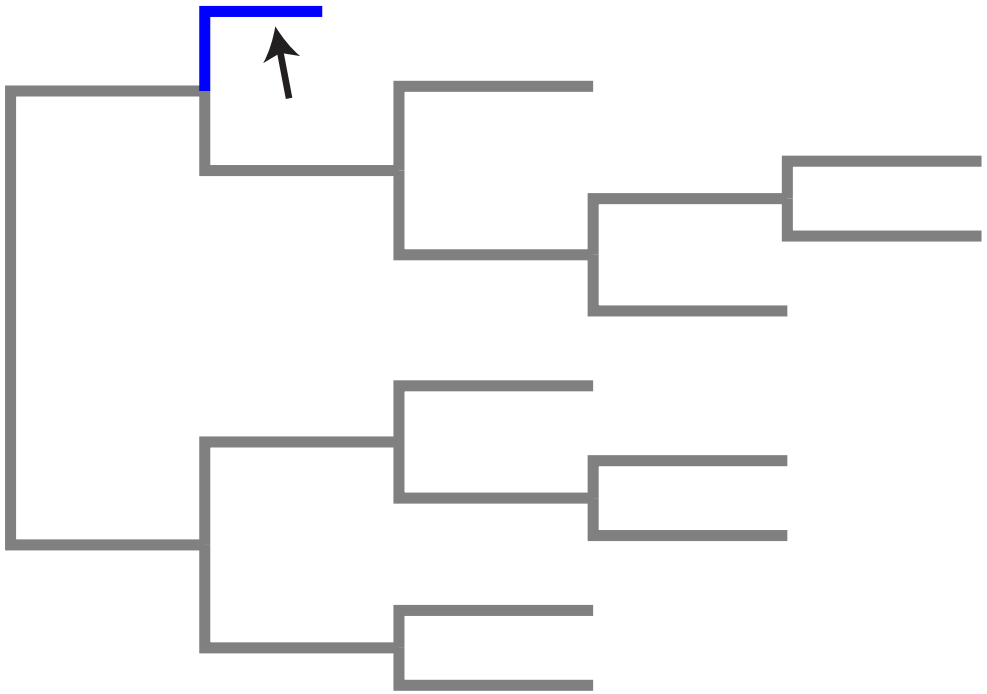
class-rate prior distribution

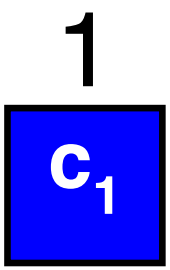


Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

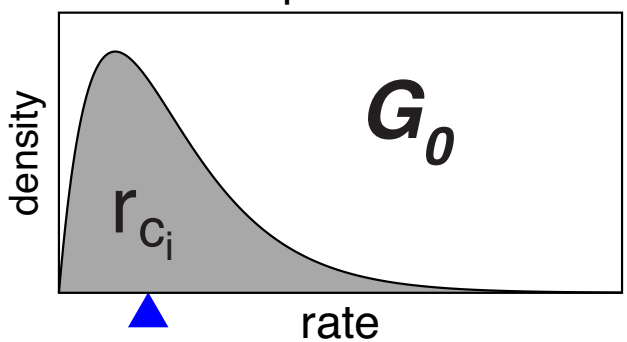
branch length = substitution rate





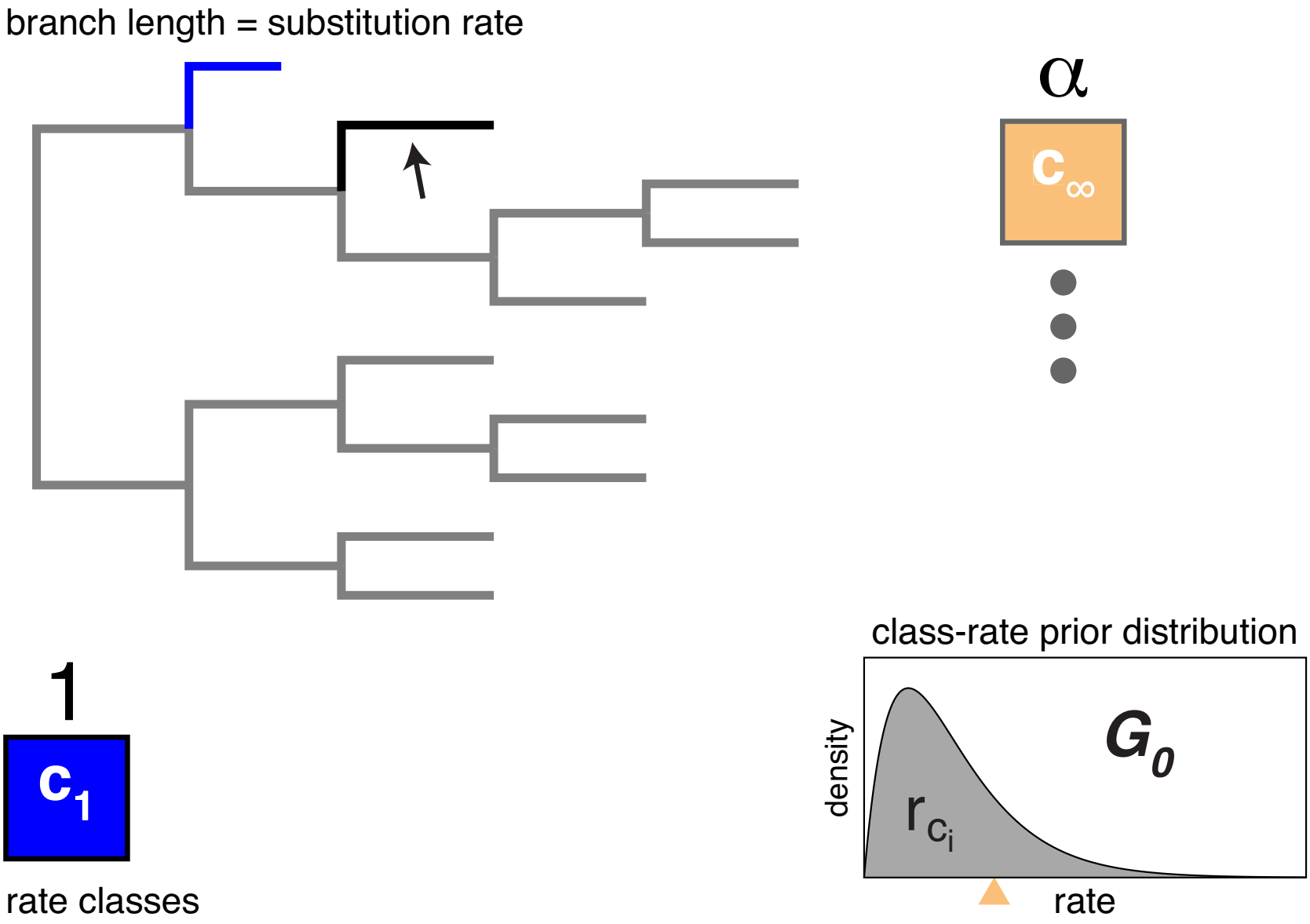
rate classes

class-rate prior distribution



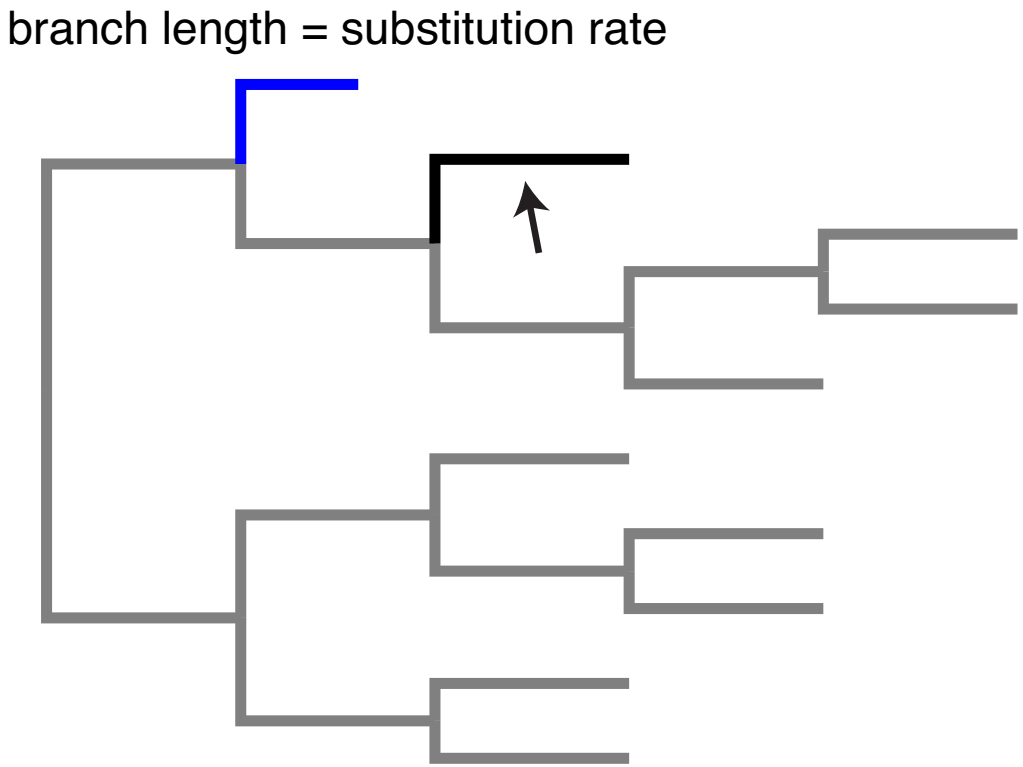
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

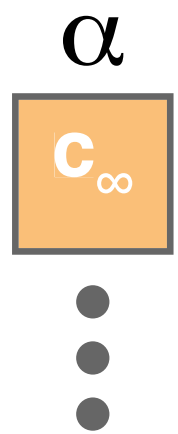
The Dirichlet Process Prior (DPP)

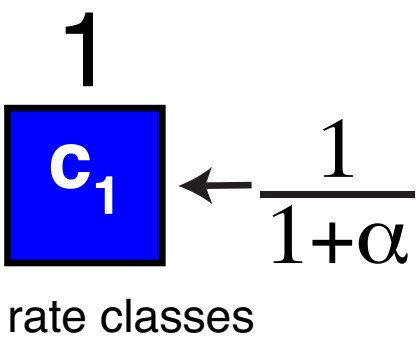


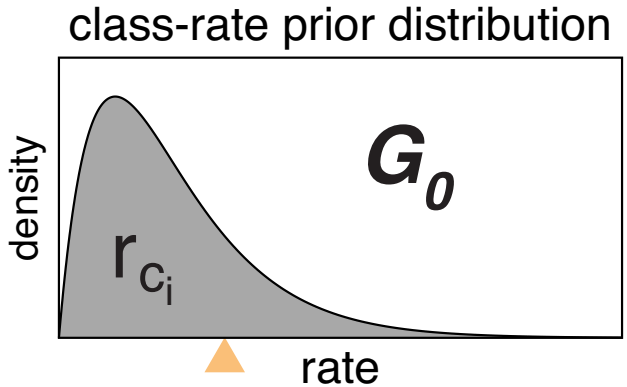
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)



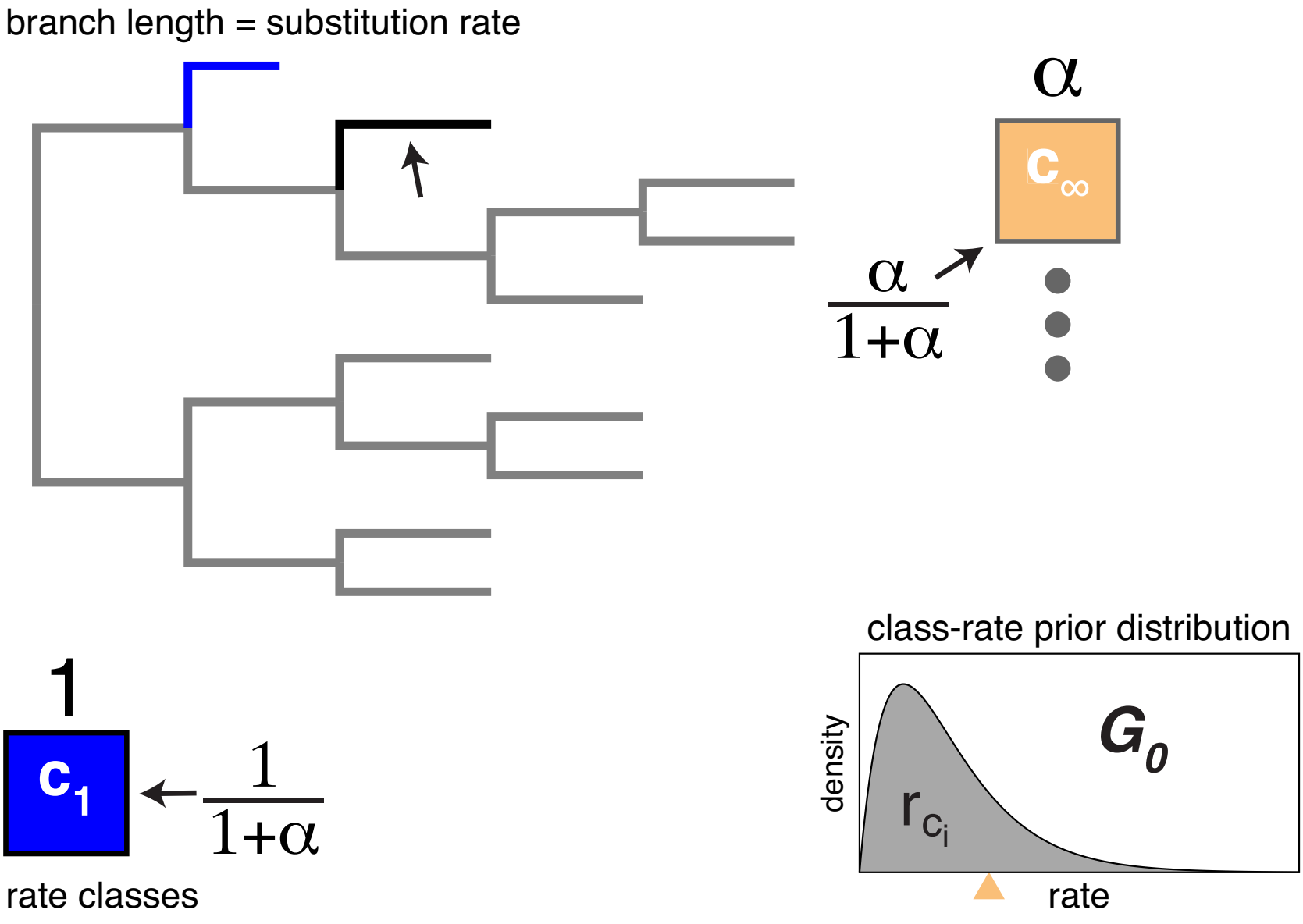






Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

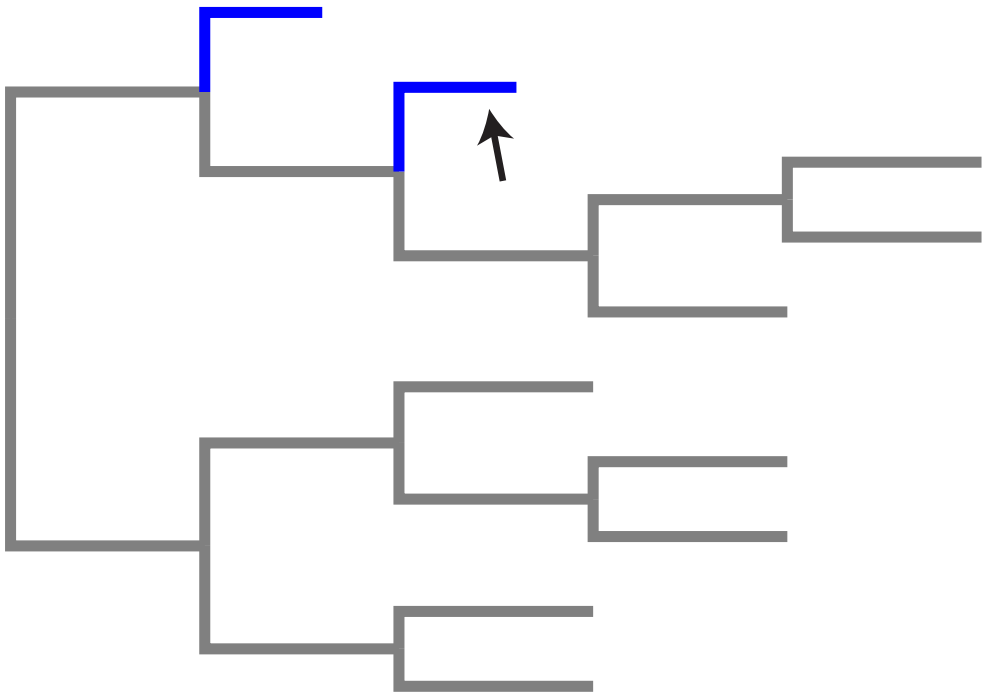
The Dirichlet Process Prior (DPP)

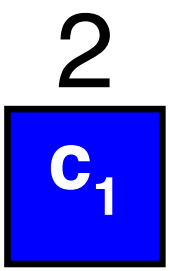


Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

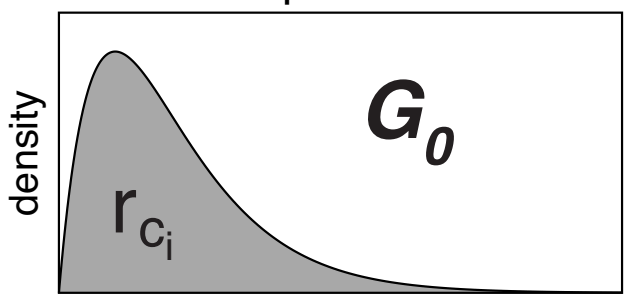
branch length = substitution rate





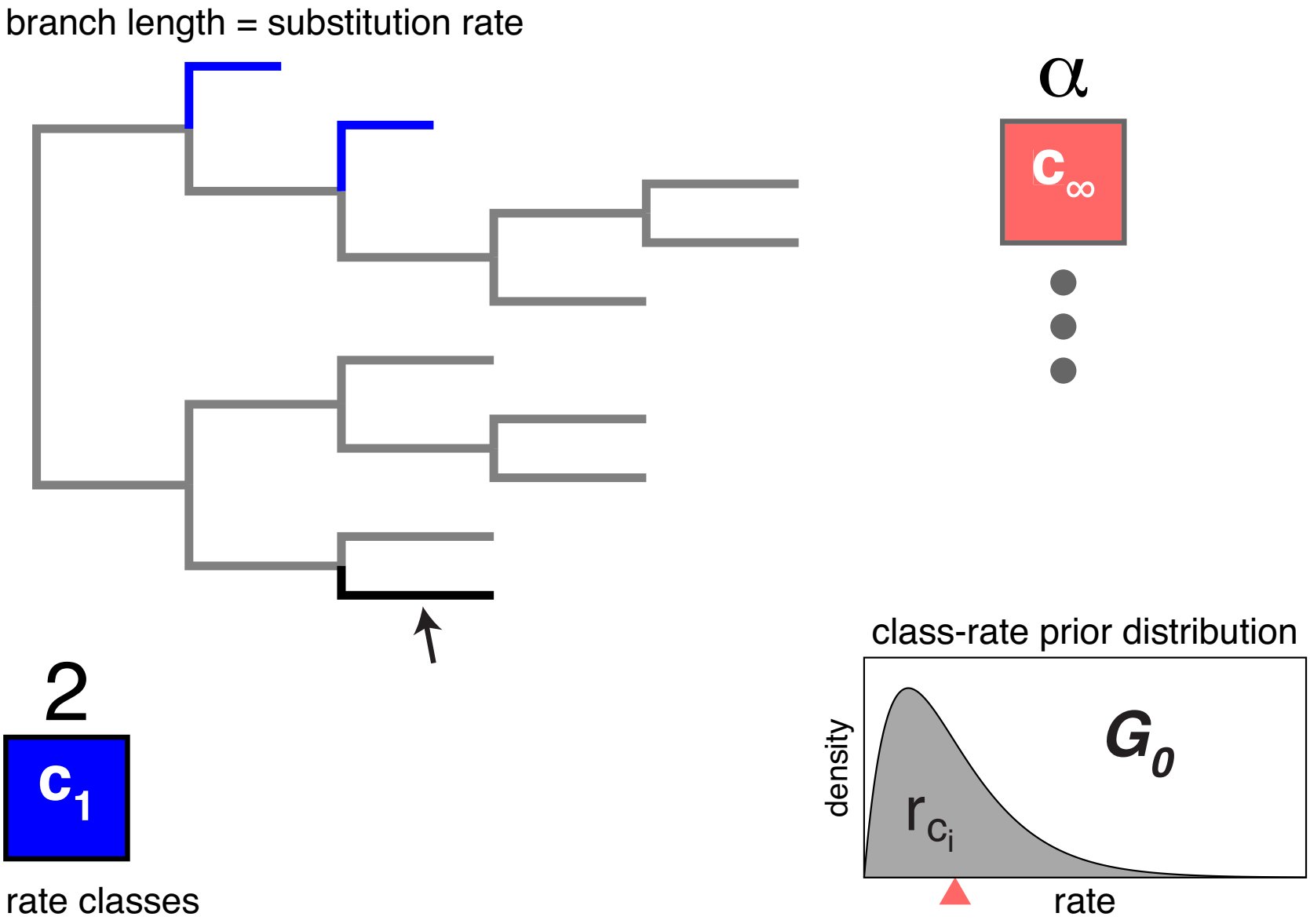
rate classes

class-rate prior distribution



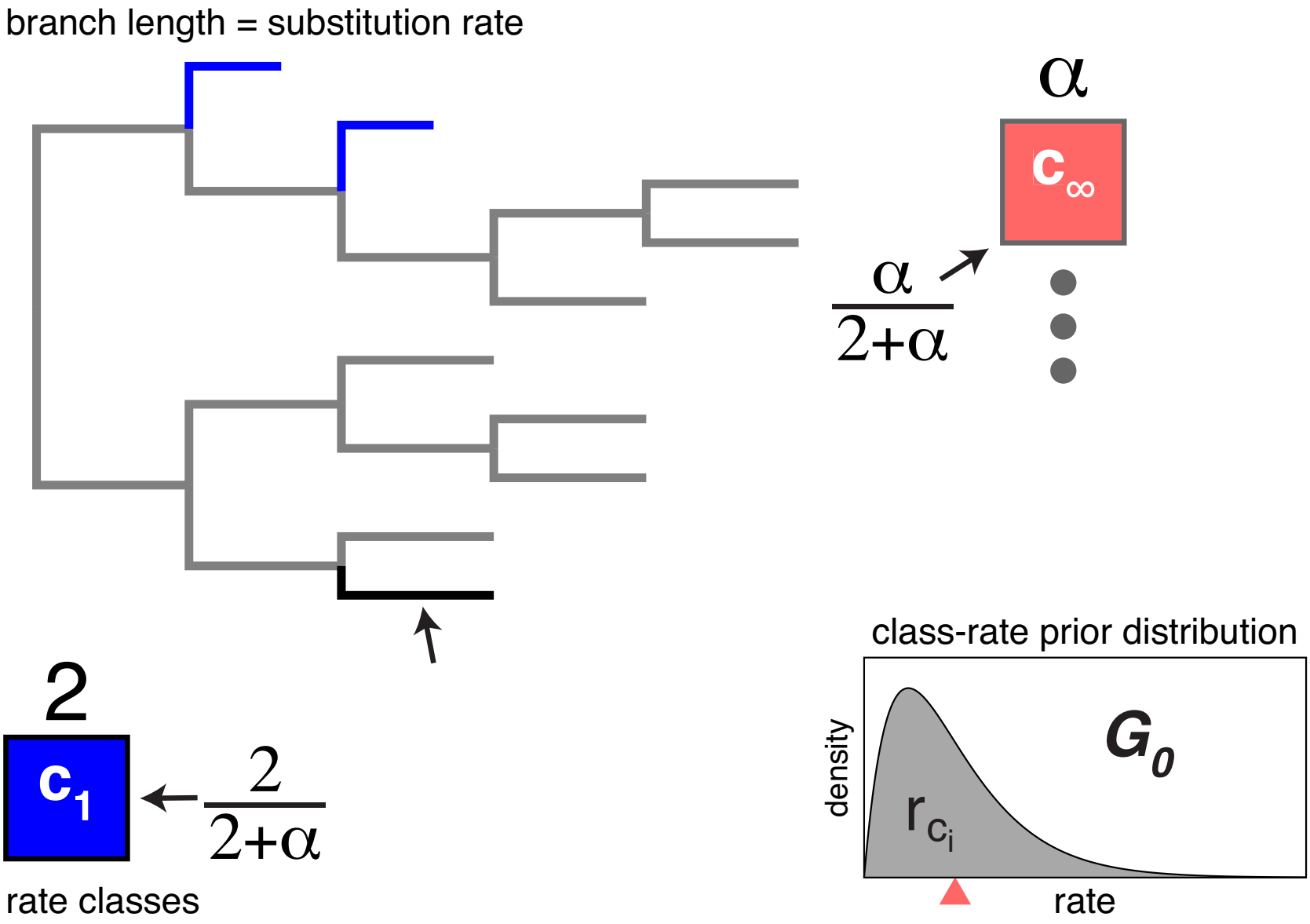
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)



Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

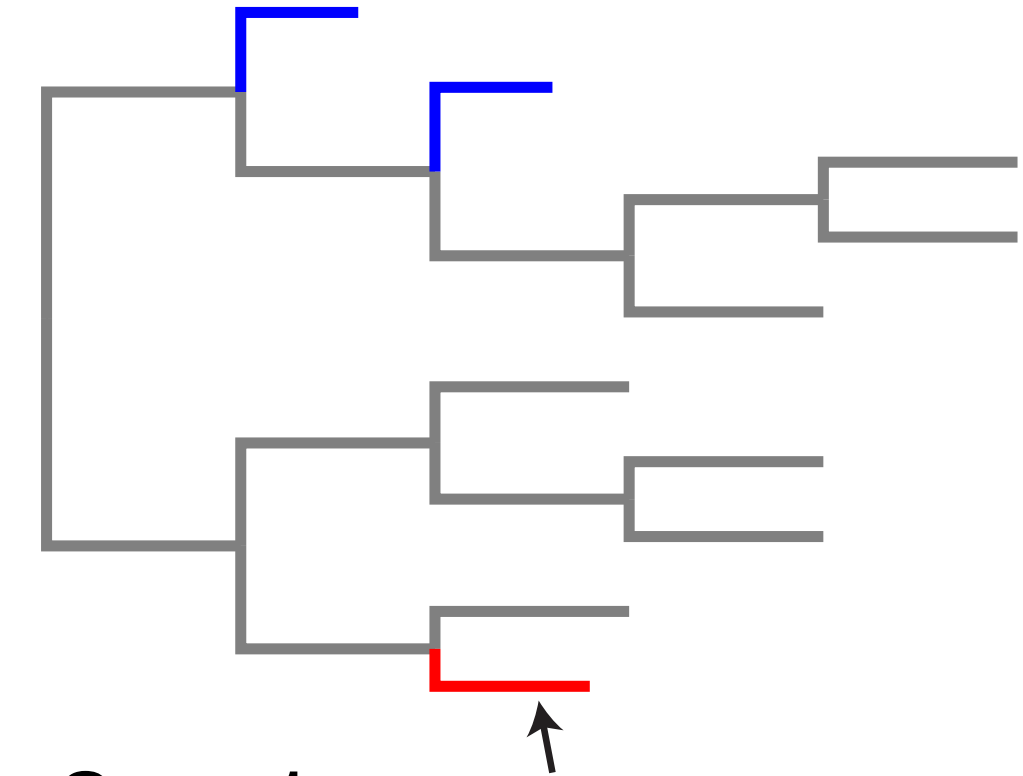
The Dirichlet Process Prior (DPP)

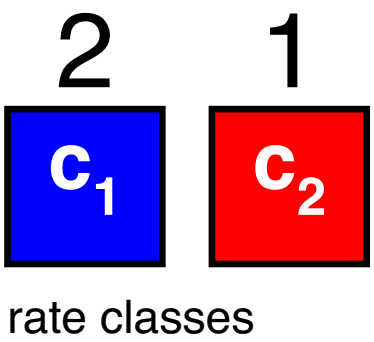


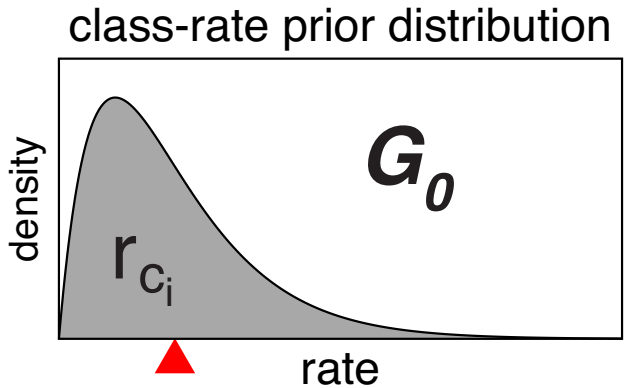
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

branch length = substitution rate

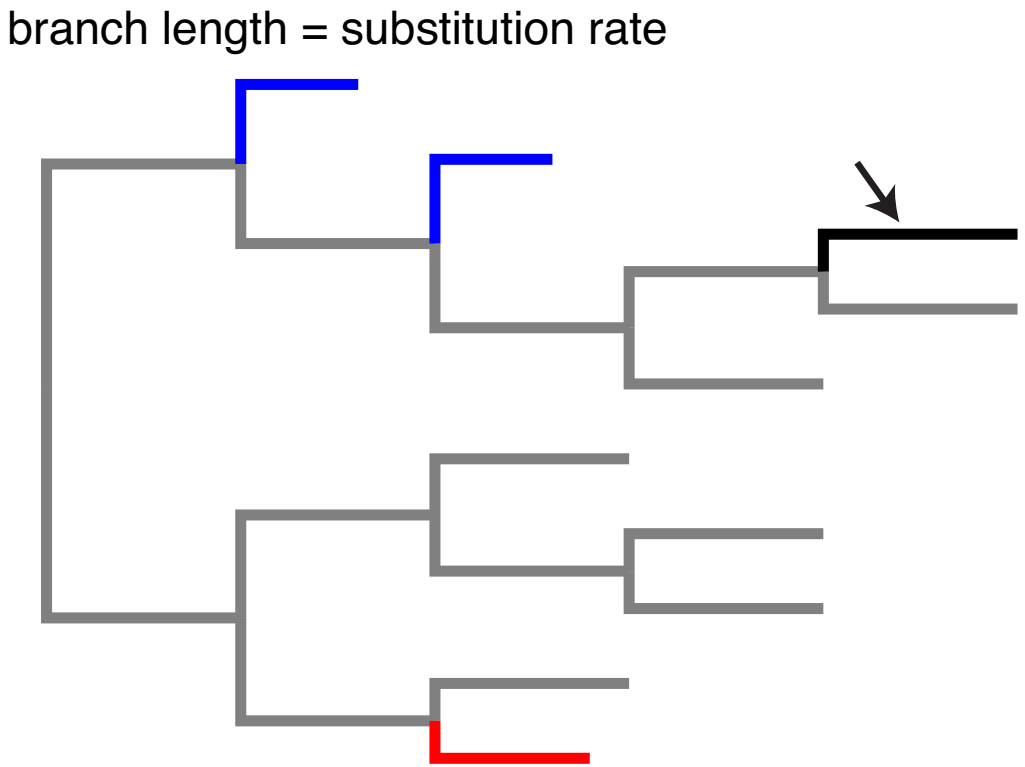


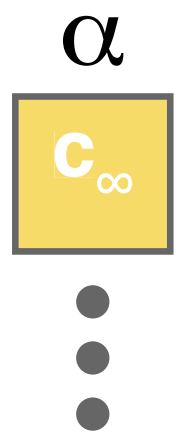


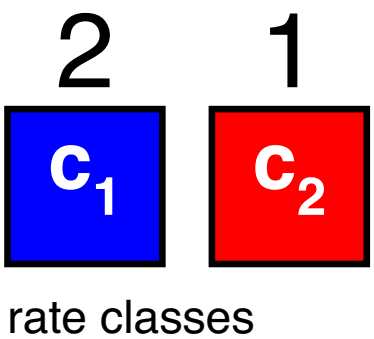


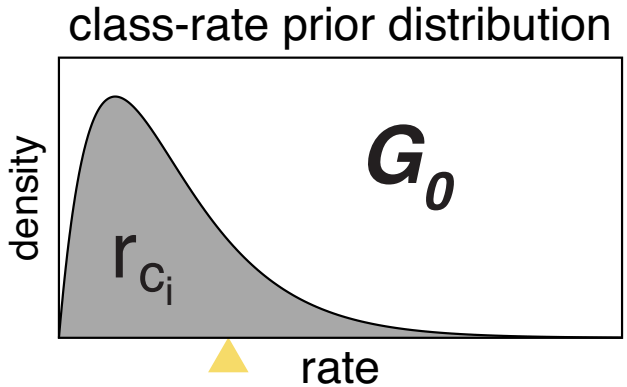
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)





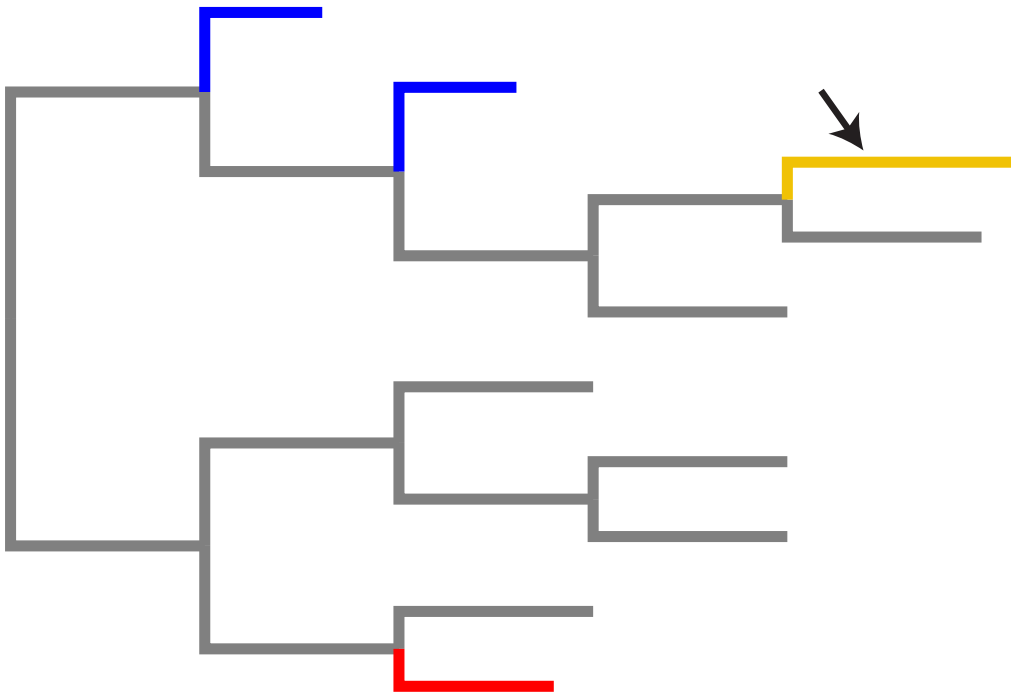


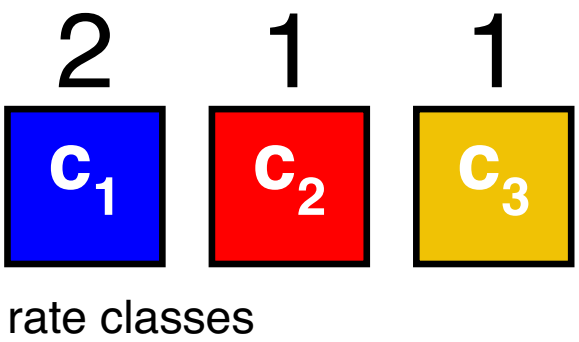


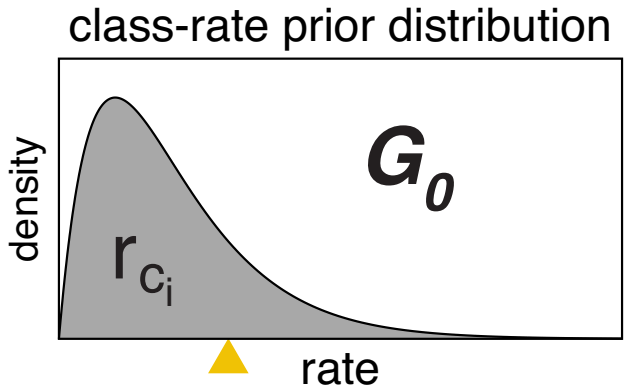
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

branch length = substitution rate



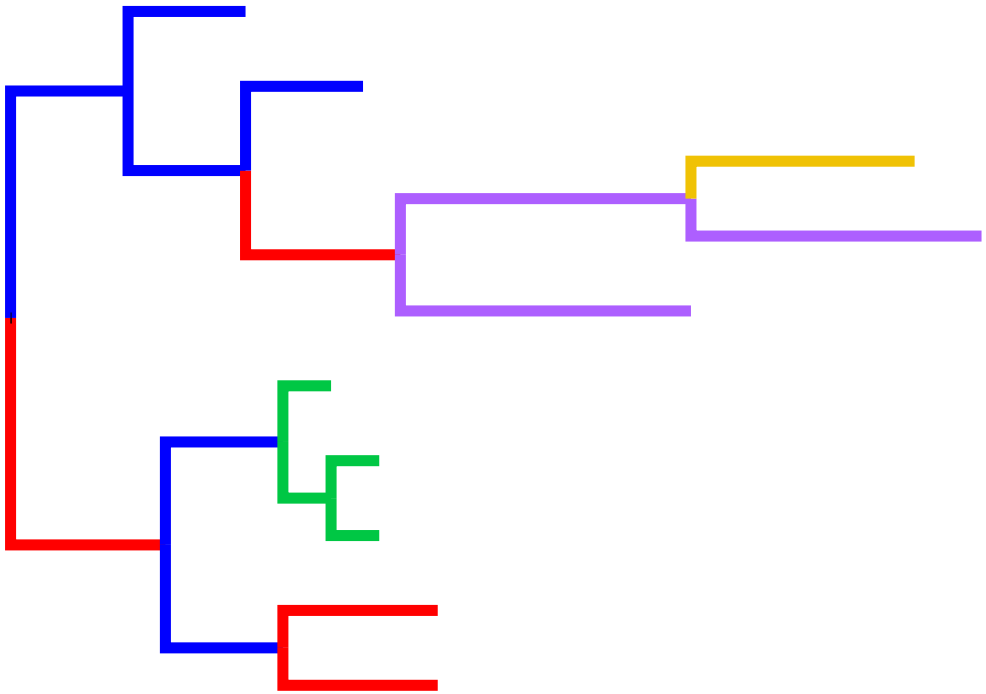


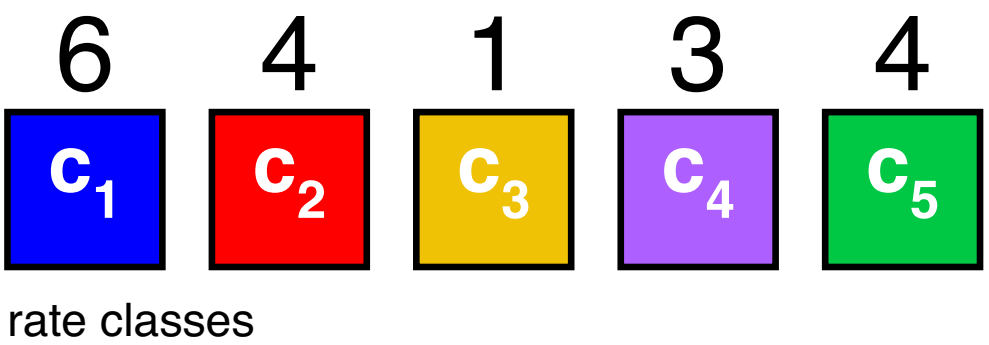


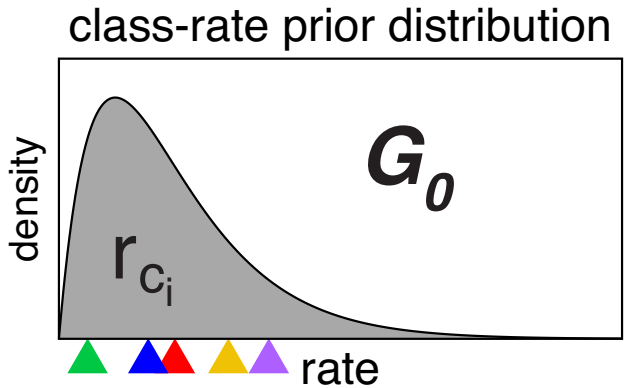
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

branch length = substitution rate



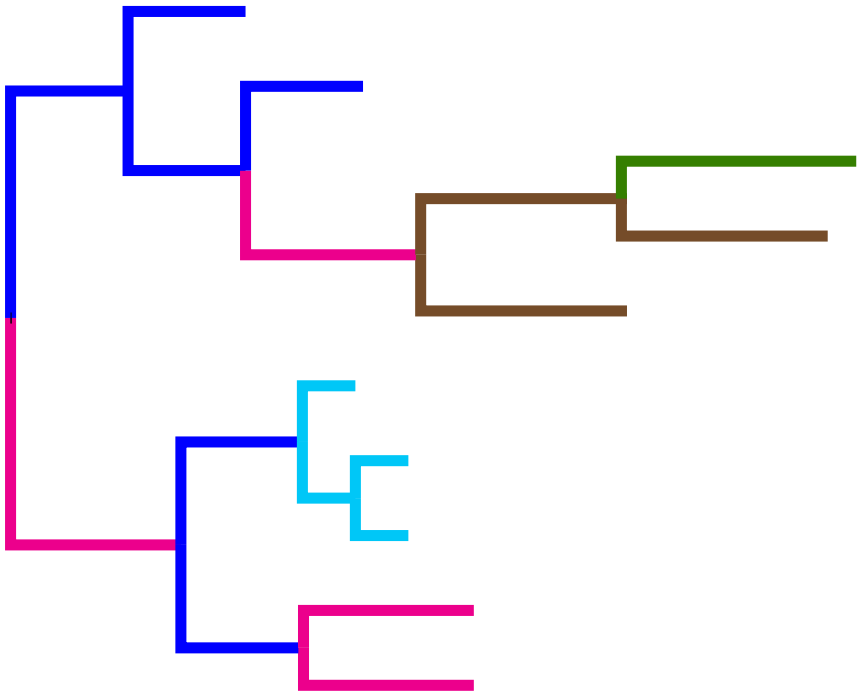


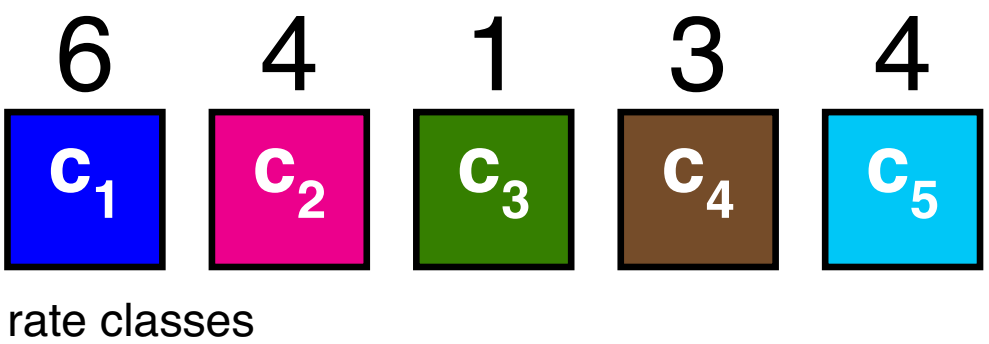


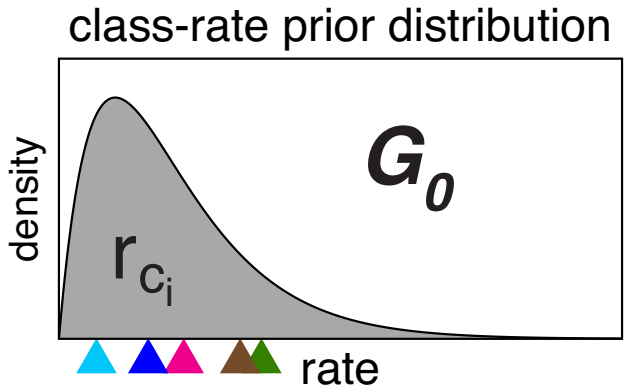
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

branch length = substitution rate



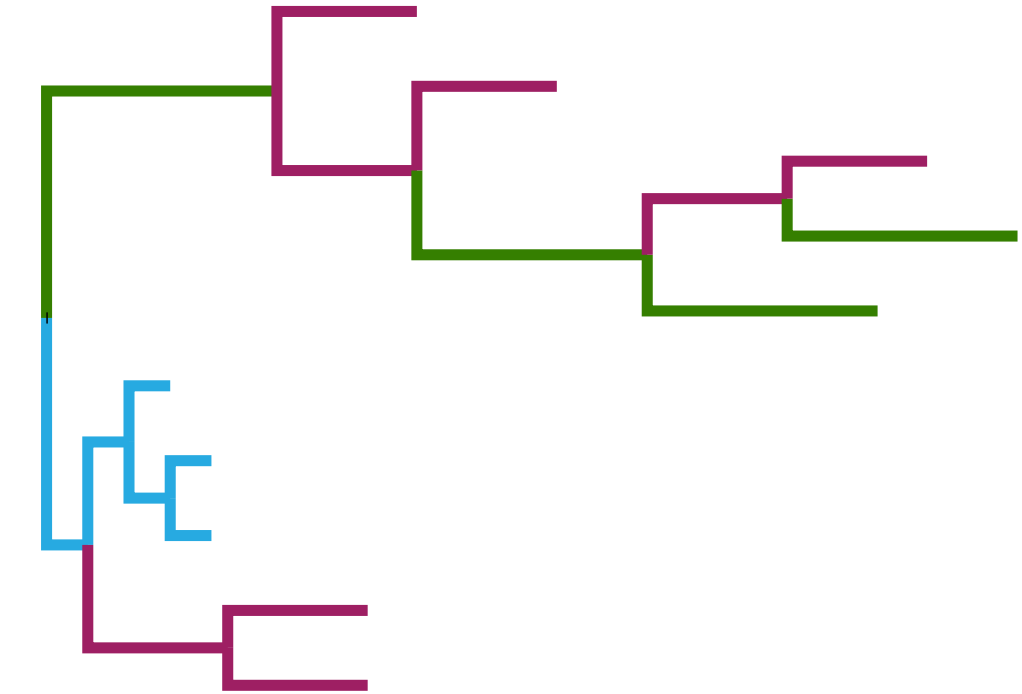


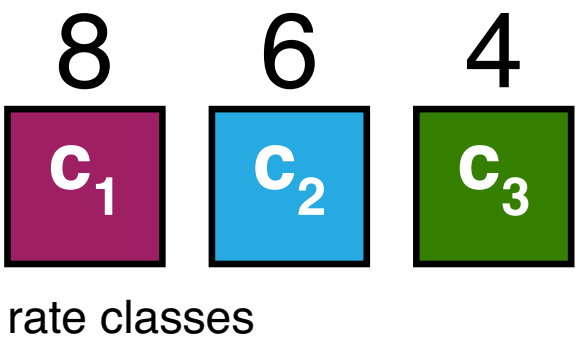


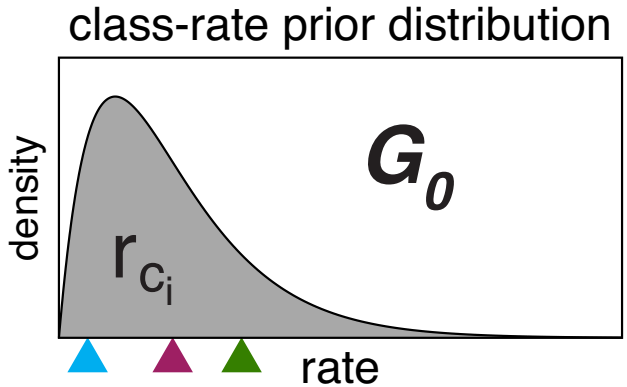
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

branch length = substitution rate



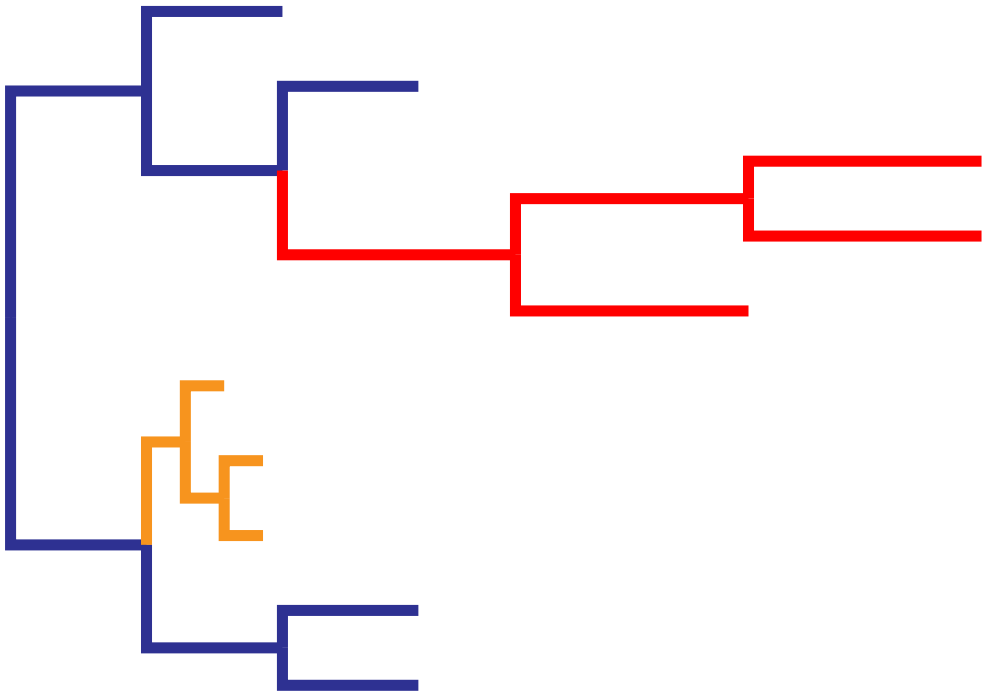




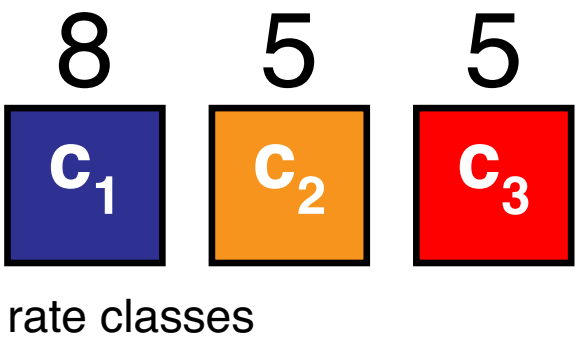
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

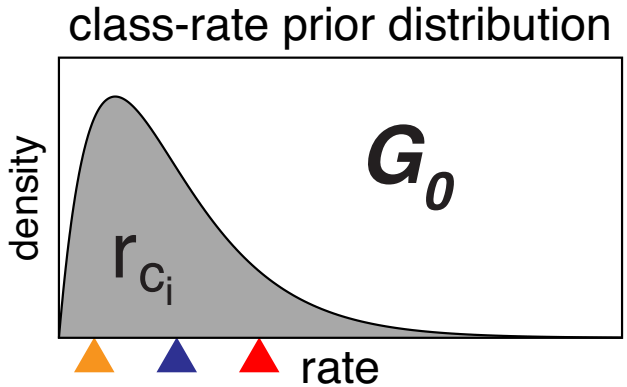
The Dirichlet Process Prior (DPP)

branch length = substitution rate



Local molecular clock

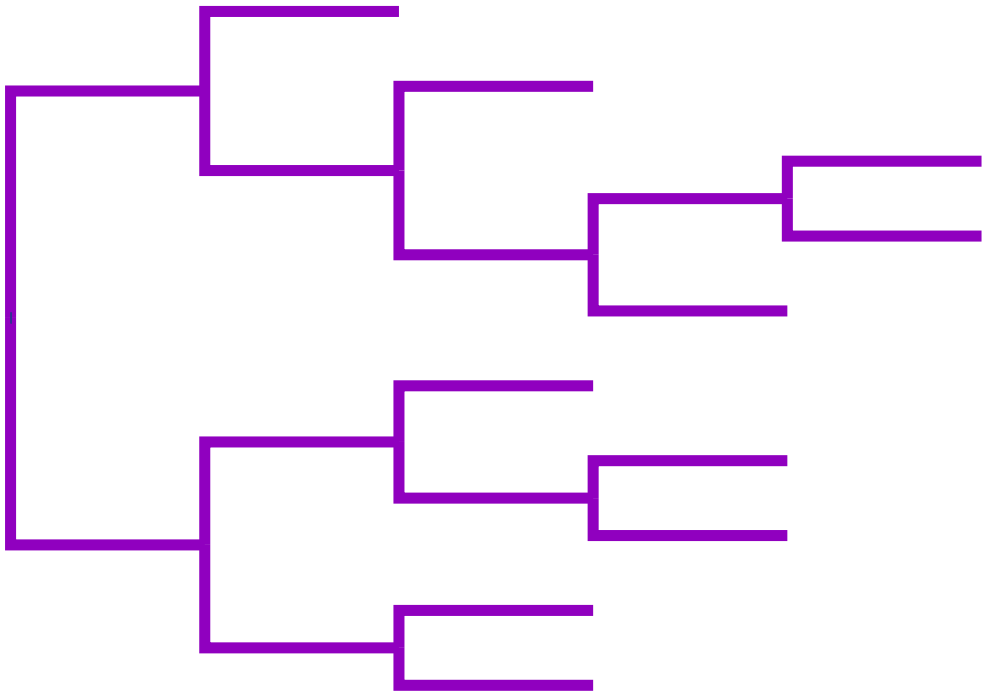




Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

The Dirichlet Process Prior (DPP)

branch length = substitution rate

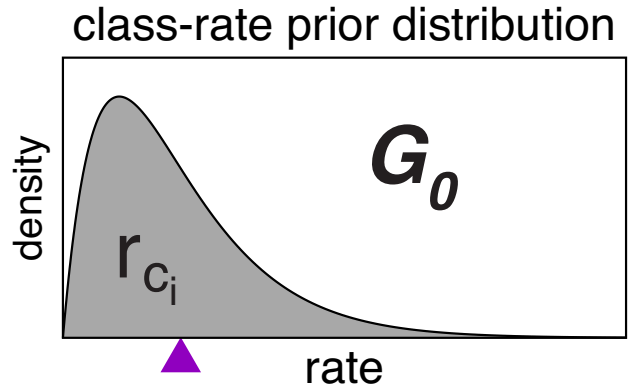


Global molecular clock

18

c_1

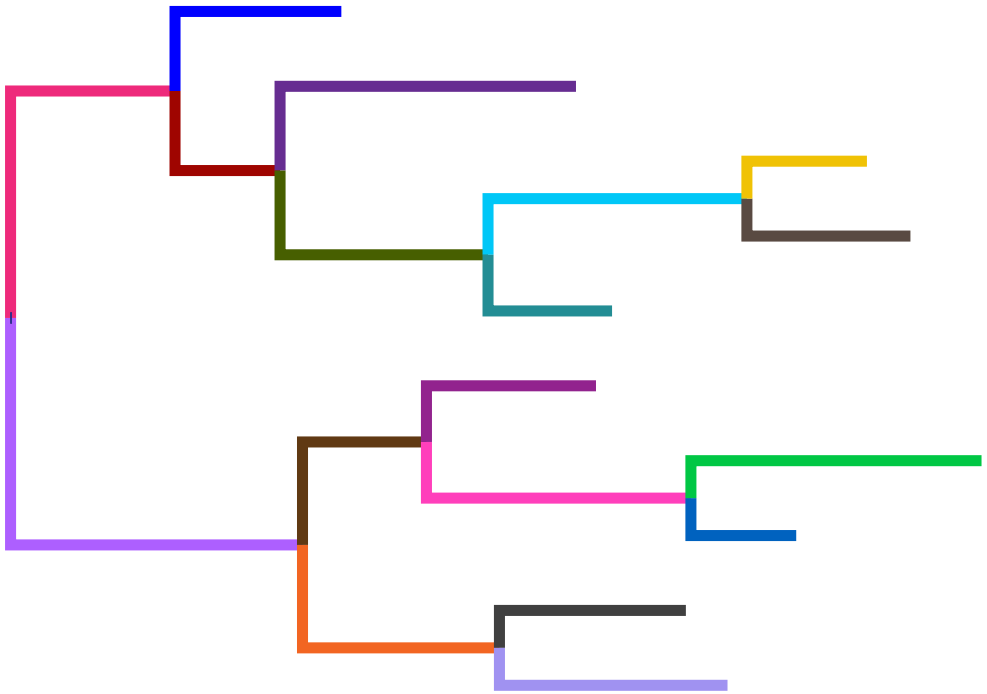
rate classes



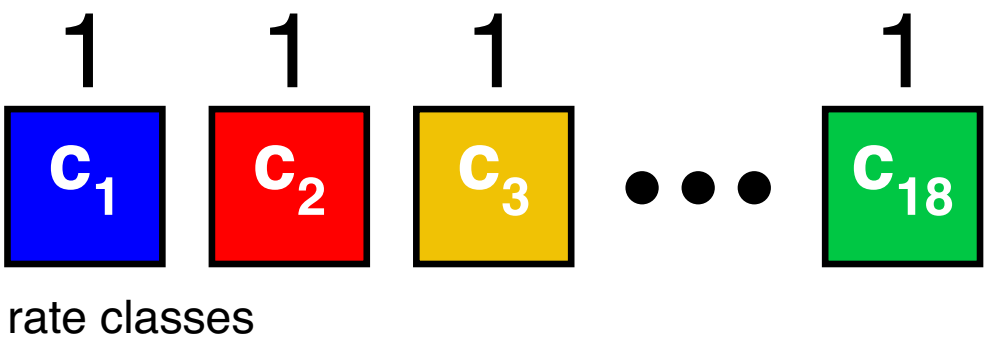
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

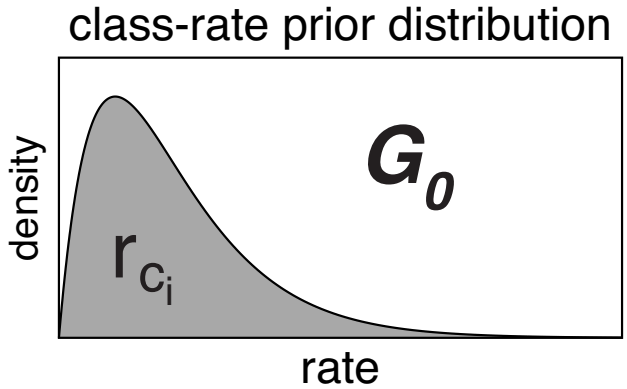
The Dirichlet Process Prior (DPP)

branch length = substitution rate



Independent rates

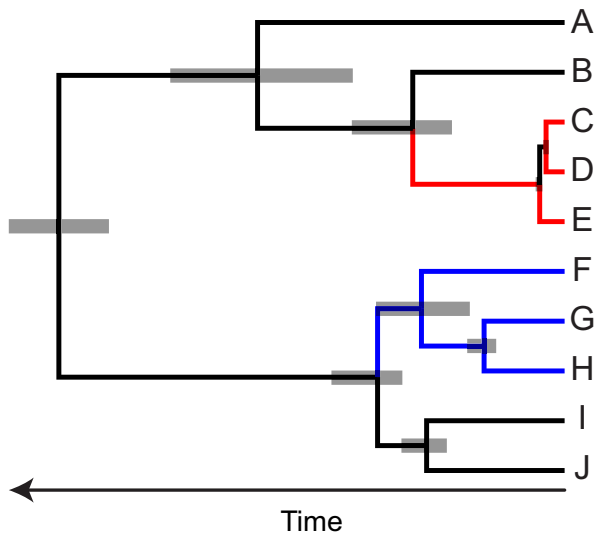




Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

Bayesian Inference Under the DPP

Current implementation: `DPPDiv`



Availability:

`http://phylo.bio.ku.edu/content/tracy-heath-dppdiv`

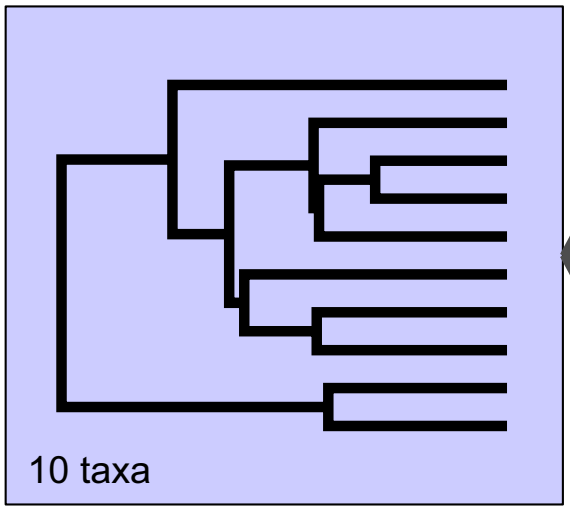
*with optimized and paralleized versions by Diego Darriba, Tomáš Flouri, & Alexis Stamatakis

Simulations: Data Generation

Simulate substitution rate variation under 6 different models

Simulate data GTR + Γ (2000 bases)

Simulate tree topologies and branching times under a birth-death process



Global Molecular Clock (GMC)

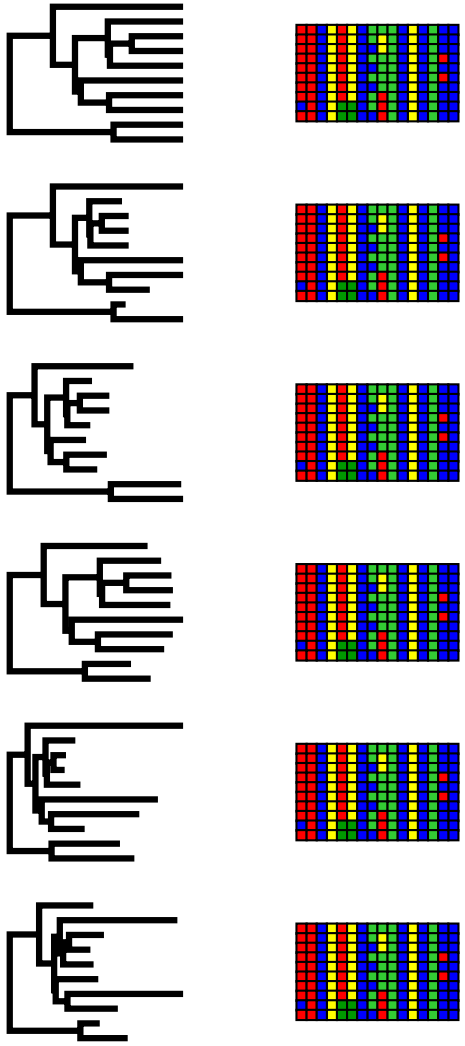
Local Molecular Clocks (LMC)

Compound Poisson process (CPP)

Log-Norm dist Autocorrelated Rates (AR-LN)

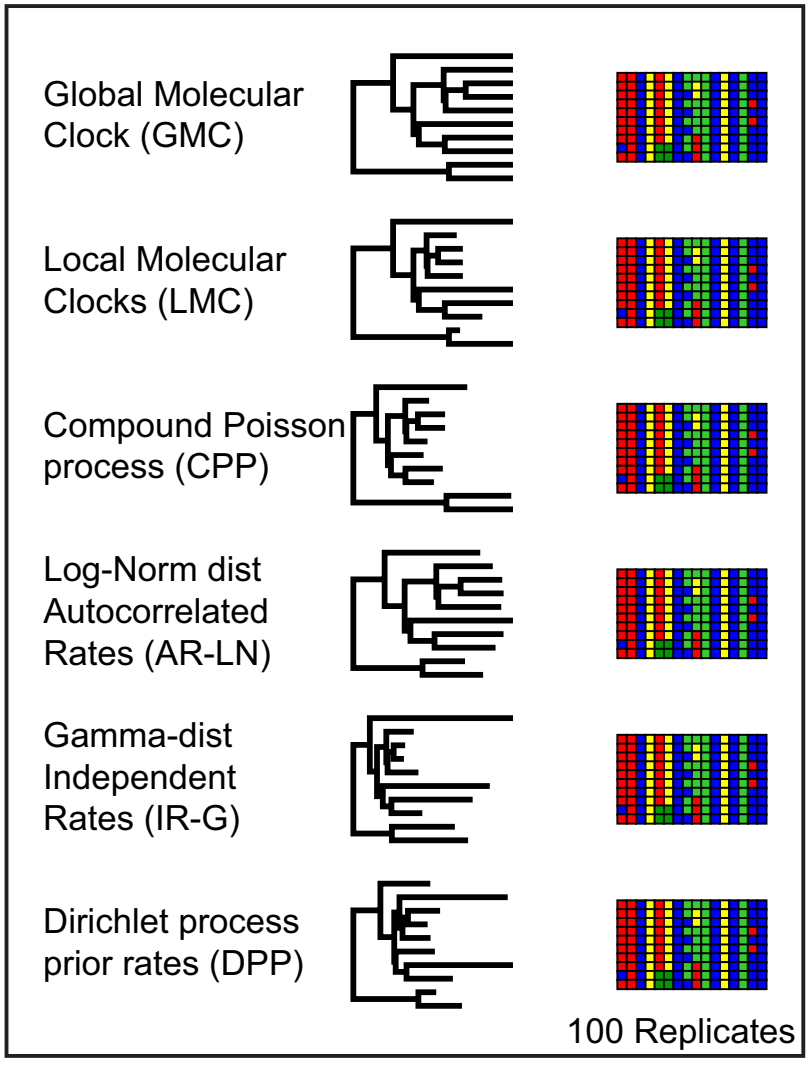
Gamma-dist Independent Rates (IR-G)

Dirichlet process prior rates (DPP)



100 Replicates

Simulations: Analysis

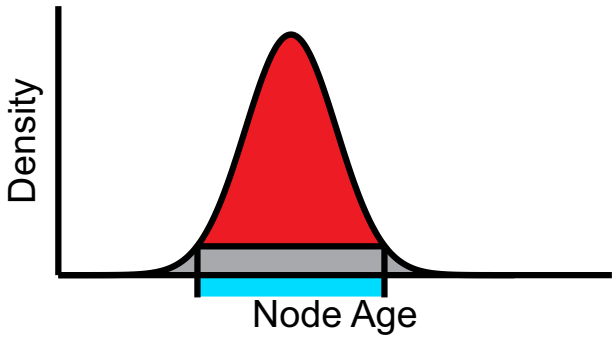


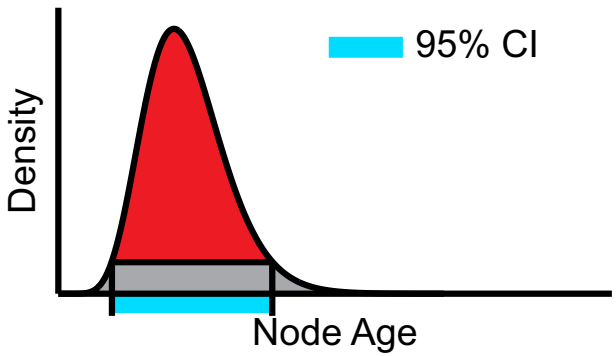
Models of rate variation:

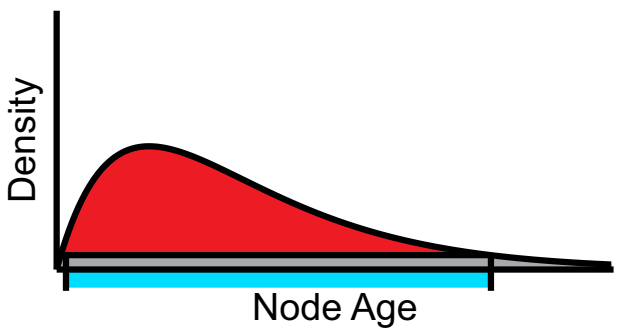
- Dirichlet process prior
  - Gamma-dist hyperprior on α, expected value: $\mathrm{E}[\alpha] = 1.93$
- Global molecular clock
- Independent rates (Gamma-distributed)

Relative node ages

95% Credible Interval (CI)



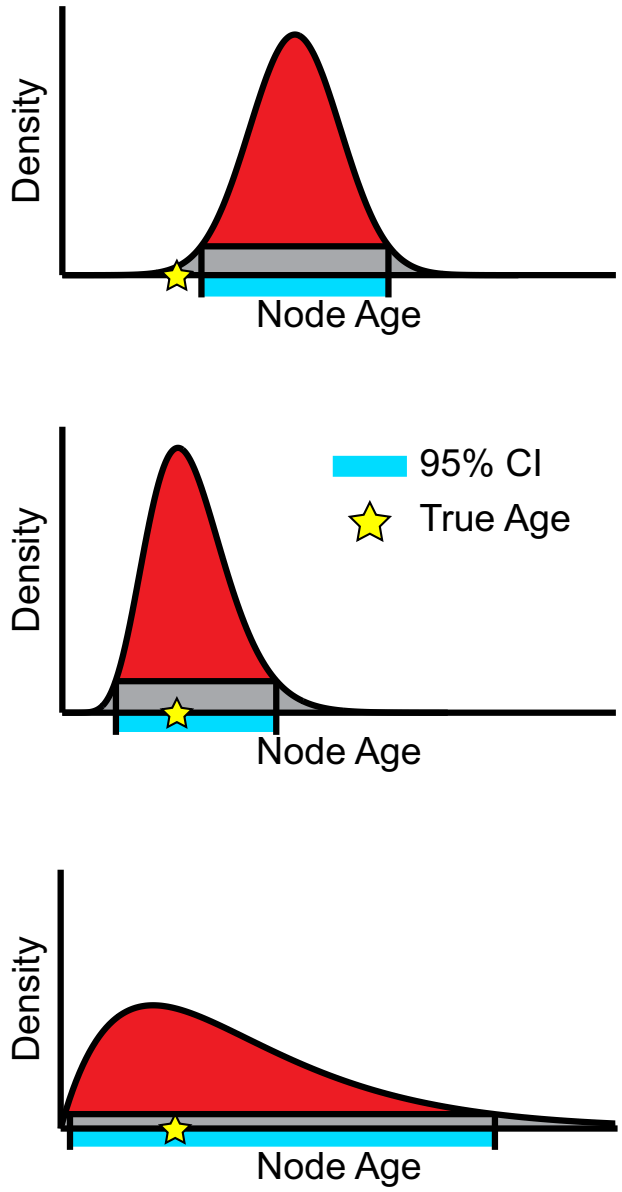




A measure of uncertainty

Approximation of the interval containing 95% of the highest posterior density (HPD)

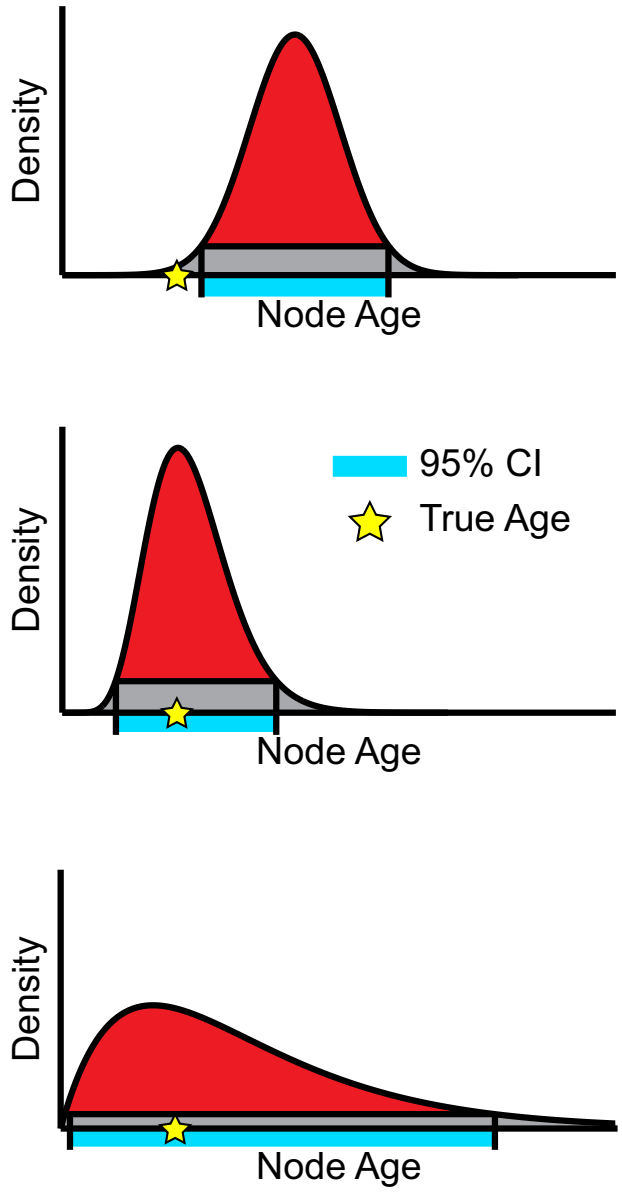
Bayesian Analysis of Simulated Data



Coverage Probability:

The proportion of the time the 95% credible interval (CI) contains the true value is a measure of accuracy

Bayesian Analysis of Simulated Data



Power:

An estimator can have high coverage probability, but reduced power when 95% CIs are very large

Branch Rate: Accuracy

The DPP and Independent Rates models had higher coverage for estimates of branch rates, depending on the simulation model

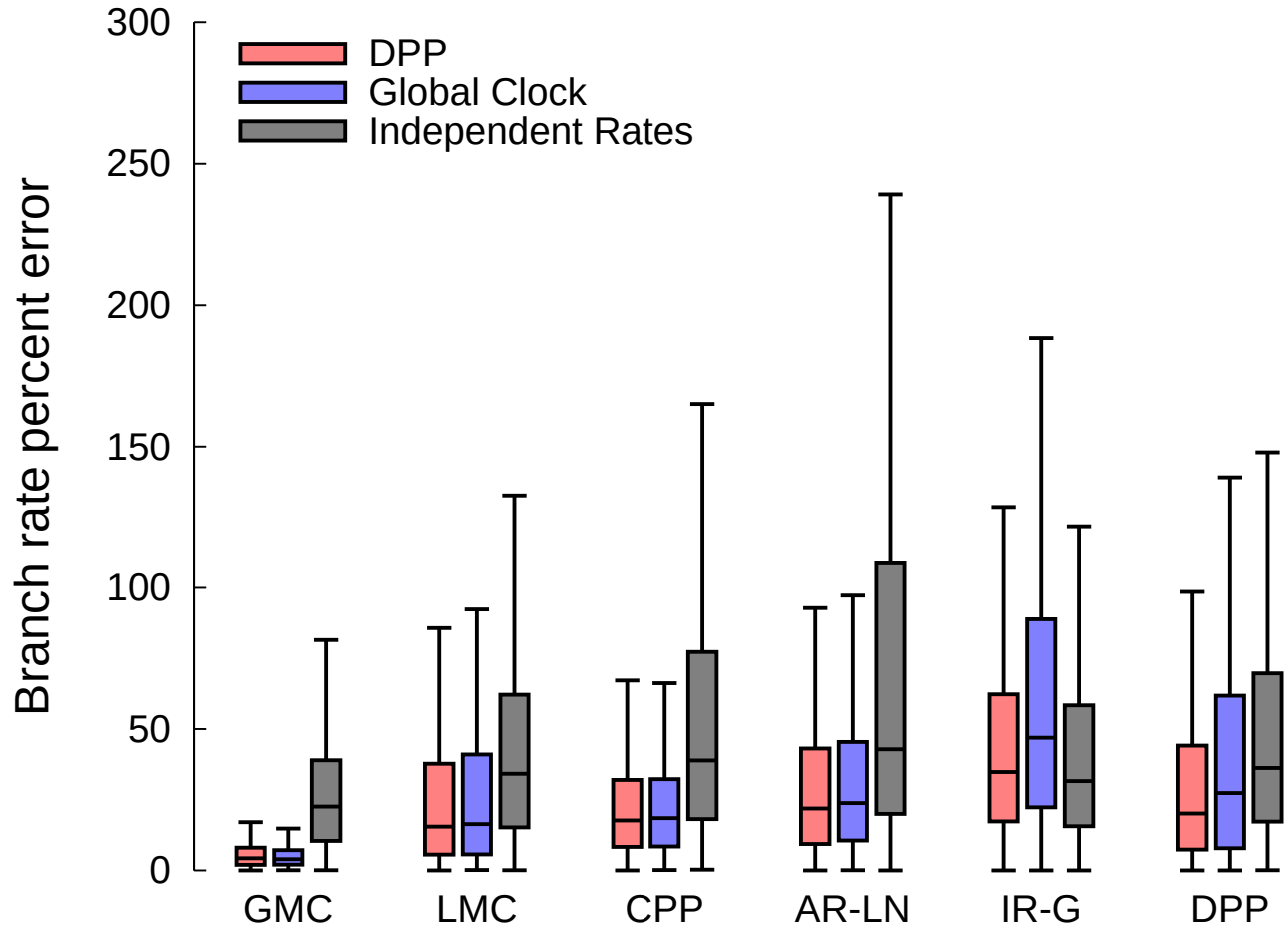
Rate Simulation	Coverage probability* DPP	Independent Rates	Global Clock
GMC – global molecular clock	**0.988**	0.963	0.920
LMC – local molecular clocks	**0.908**	**0.908**	0.398
CPP – compound Poisson	0.807	**0.861**	0.318
AR-LN – autocorrelated rates	0.801	**0.844**	0.257
IR-G – independent rates	0.874	**0.939**	0.126
DPP – Dirichlet process	**0.912**	0.908	0.292

*Accuracy: proportion of time the 95% credible interval covers the true branch rate

Branch Rate: Percent Error

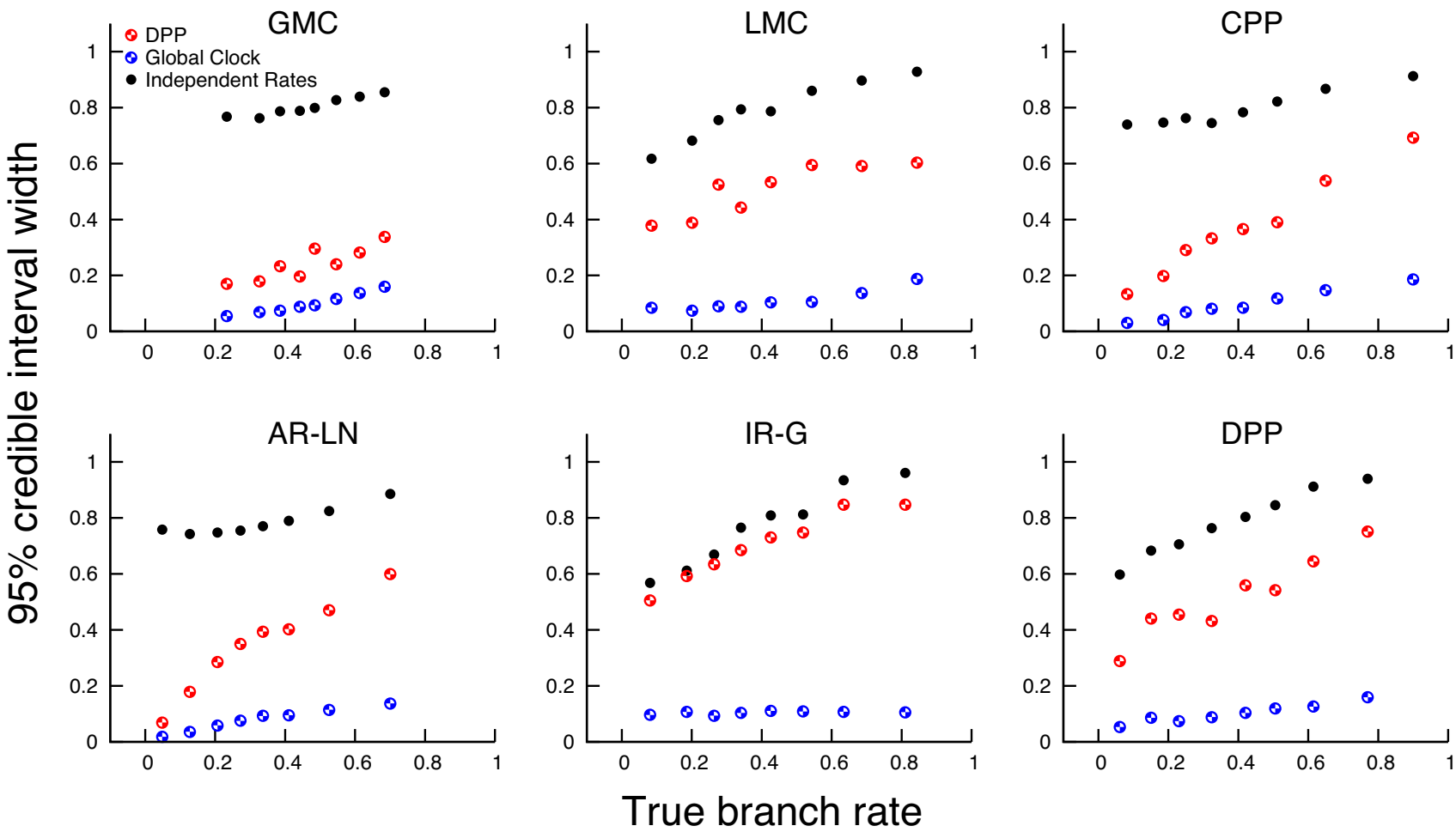
The percent error in mean branch rate estimates

$$\% \text{ Error} = \frac{|\widehat{r_i} - r_i|}{r_i} \times 100\%$$



Branch Rate: Power

95% CI size compared to TRUE branch rate



Analyses: **DPP**, **Global clock**, and **Independent rates**

Node Age: Accuracy

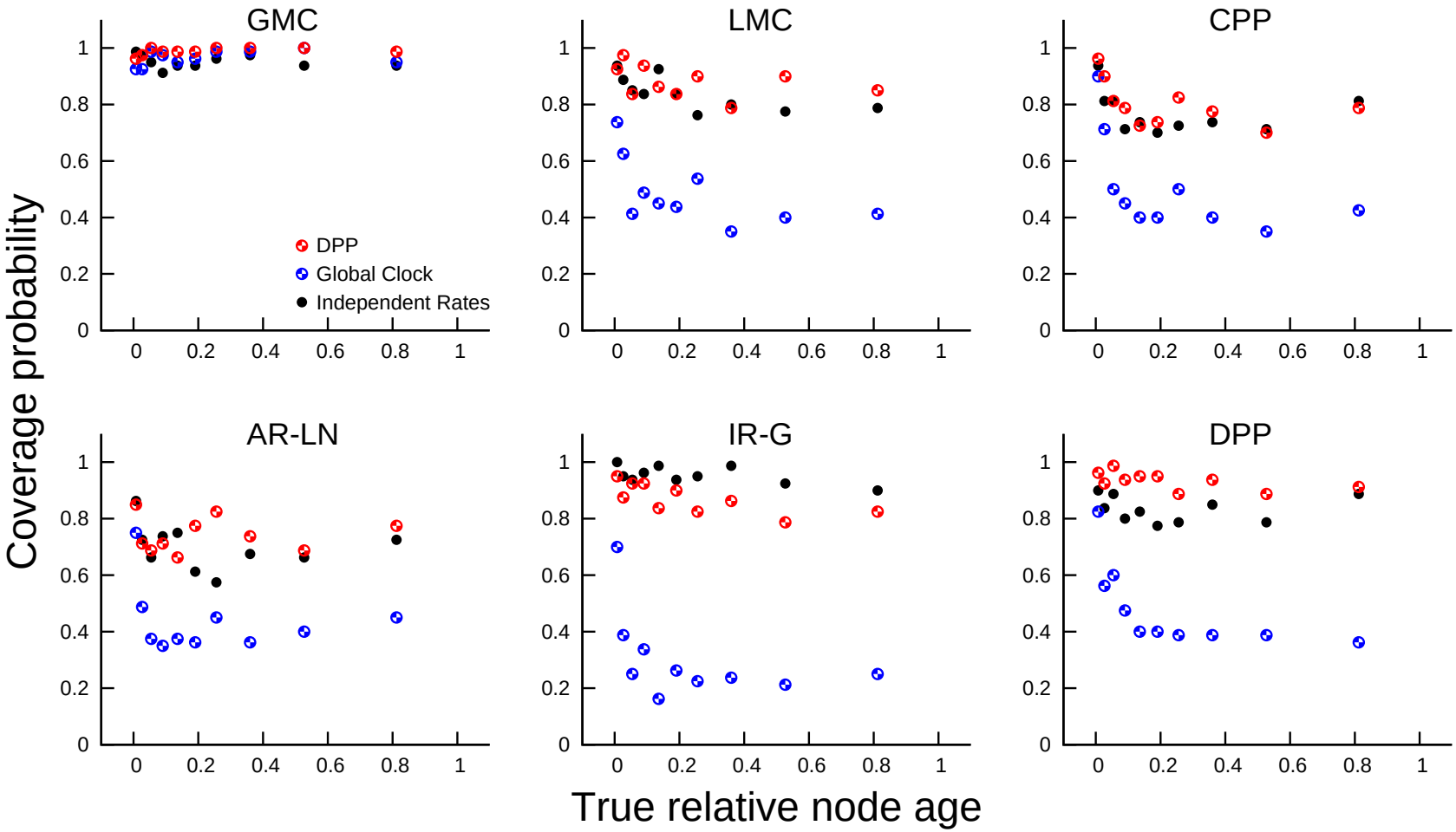
Node age estimates under DPP are more accurate compared to an independent rate model and the global molecular clock

Rate Simulation	Coverage probability* DPP	Independent Rates	Global Clock
GMC – global molecular clock	**0.989**	0.951	0.965
LMC – local molecular clocks	**0.881**	0.840	0.485
CPP – compound Poisson	**0.801**	0.770	0.504
AR-LN – autocorrelated rates	**0.743**	0.699	0.436
IR-G – independent rates	0.871	**0.954**	0.303
DPP – Dirichlet process	**0.934**	0.834	0.479

*Accuracy: proportion of time the 95% credible interval covers the true node age

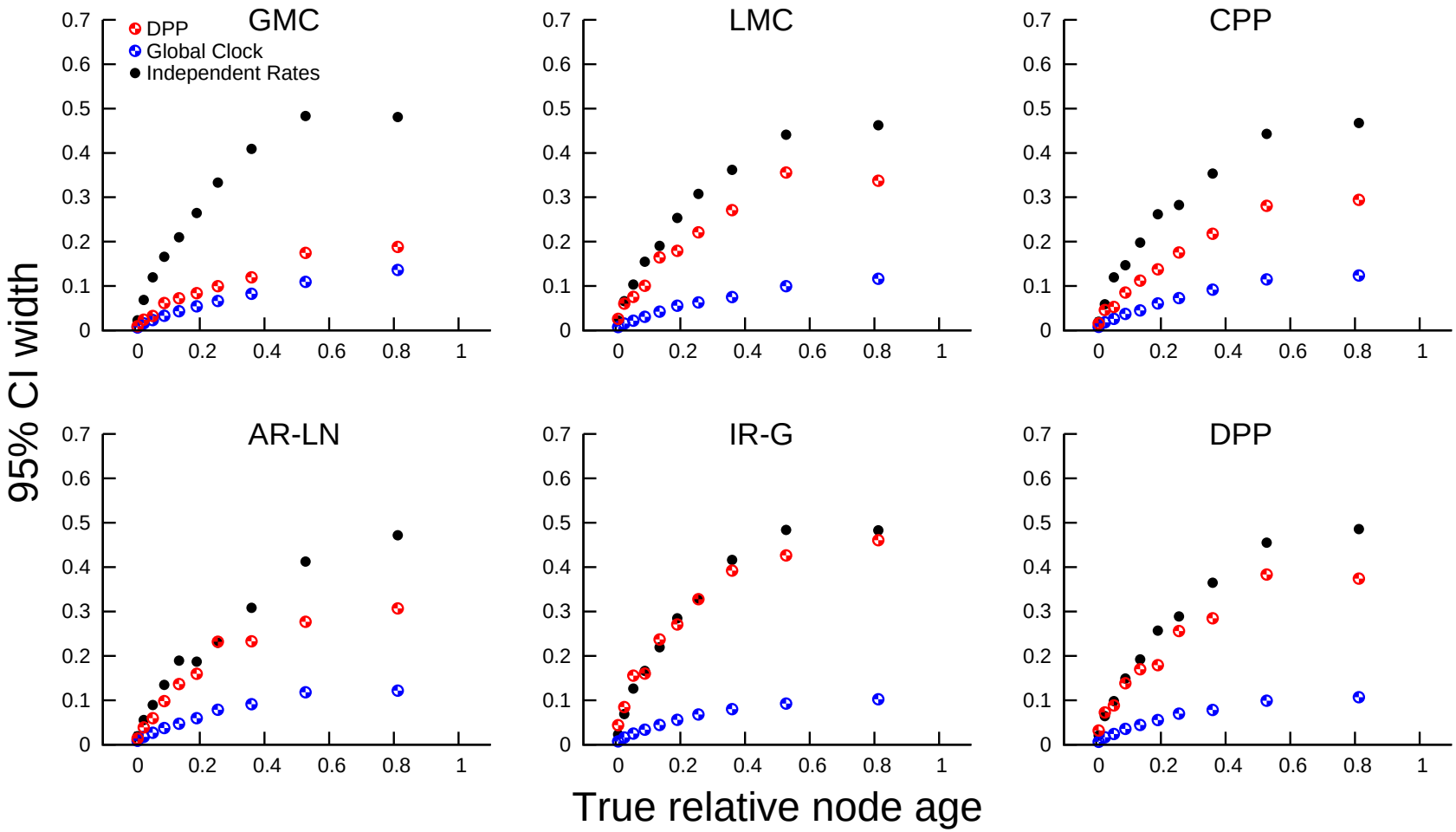
Heath, Holder, Huelsenbeck. 2012 *MBE* 29:939-955.

Node Age: Coverage Probability



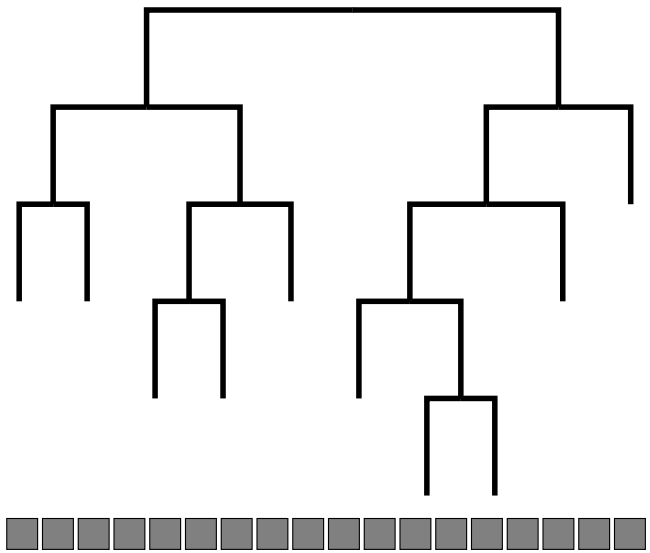
Analyses: **DPP**, **Global clock**, and **Independent rates**

Node Age: Power



Analyses: **DPP**, **Global clock**, and **Independent rates**

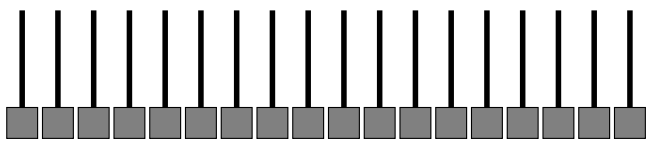
Summarizing MCMC under the DPP



MEAN PARTITION:

Identified from MCMC samples under the DPP

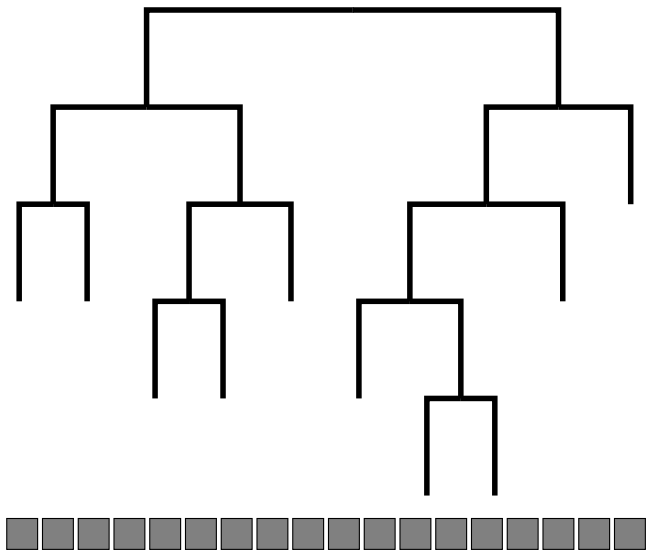
Summarizing MCMC under the DPP



MEAN PARTITION:

Identified from MCMC samples under the DPP

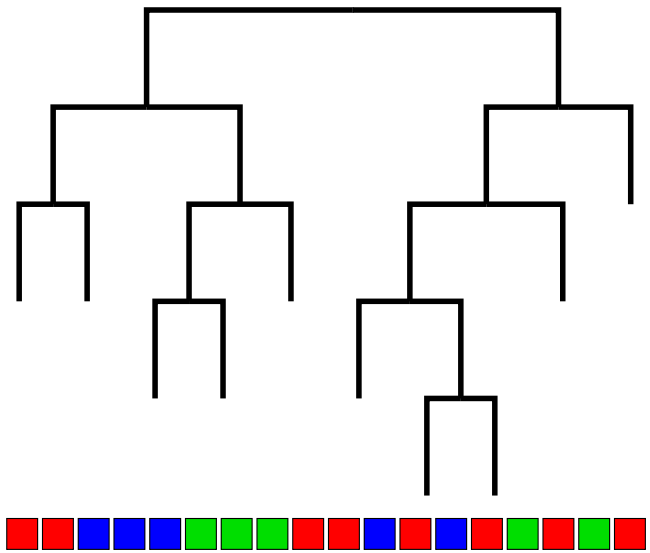
Summarizing MCMC under the DPP



MEAN PARTITION:

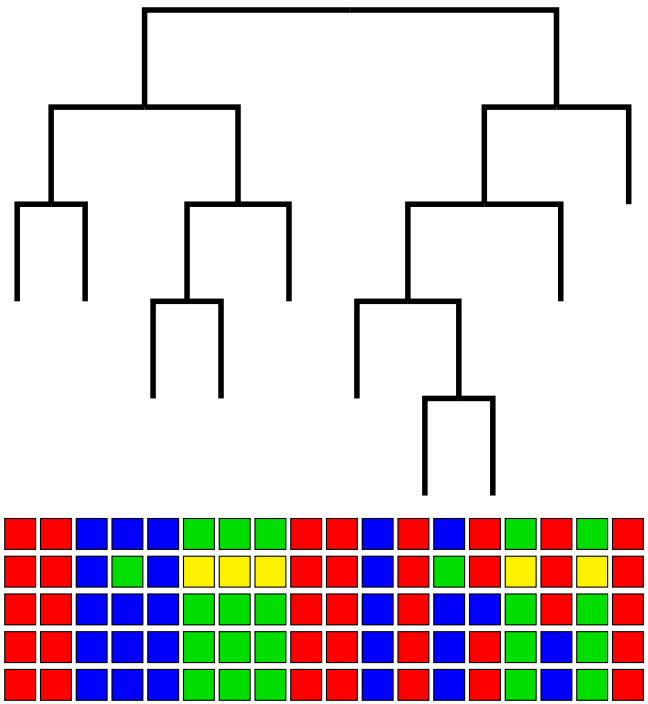
Identified from MCMC samples under the DPP

Summarizing MCMC under the DPP



MCMC samples different branch-partition assignments

Summarizing MCMC under the DPP



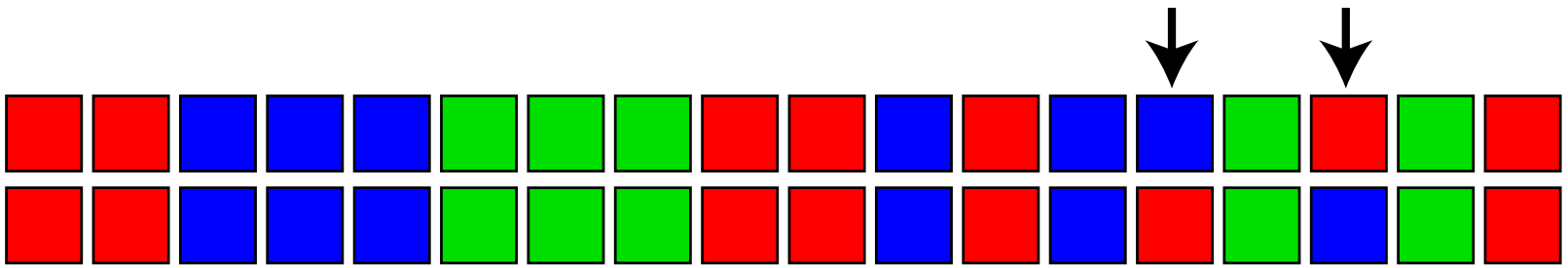
MEAN PARTITION:

Identified from MCMC samples of different branch-partition assignments under the DPP

Summarizing MCMC under the DPP

PARTITION DISTANCE:

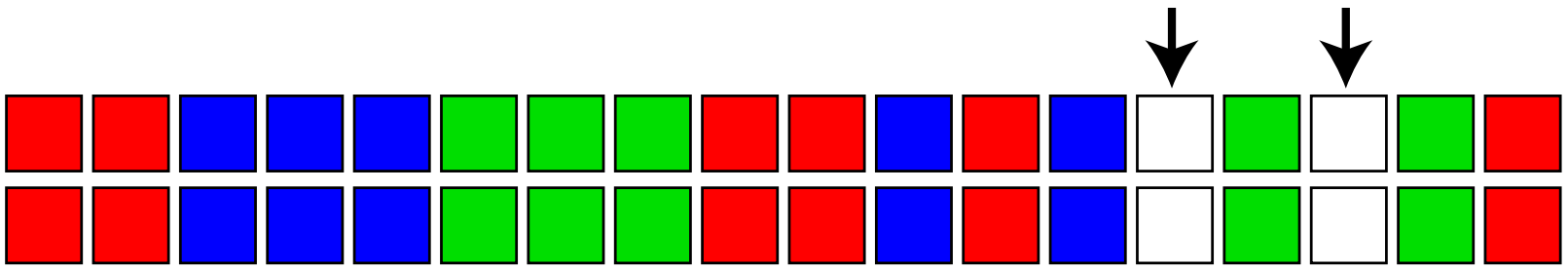
The minimum number of elements that must be removed to make 2 identical partitions



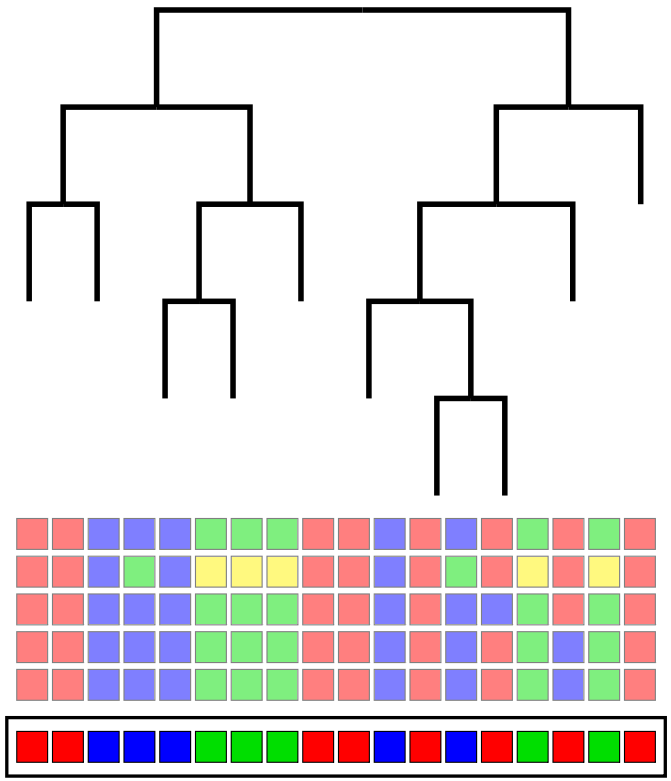
Summarizing MCMC under the DPP

PARTITION DISTANCE:

The minimum number of elements that must be removed to make 2 identical partitions



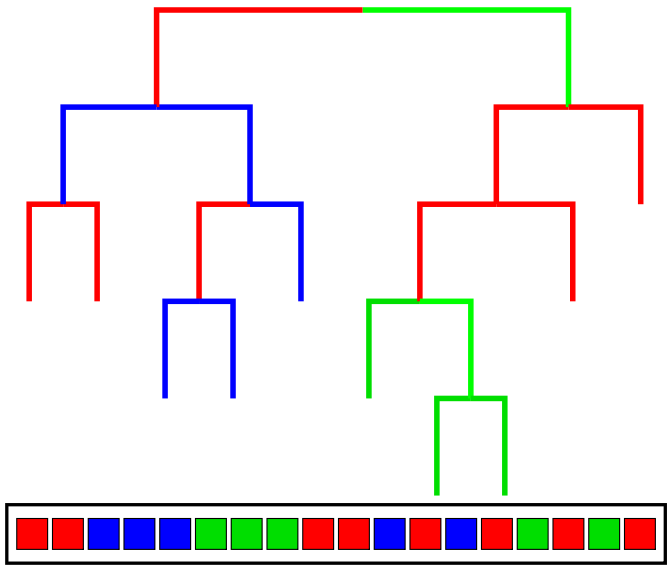
Summarizing MCMC under the DPP



MEAN PARTITION:

The set of branch-partition assignments that minimizes the sum of squared distances to all of the partition sets sampled by MCMC

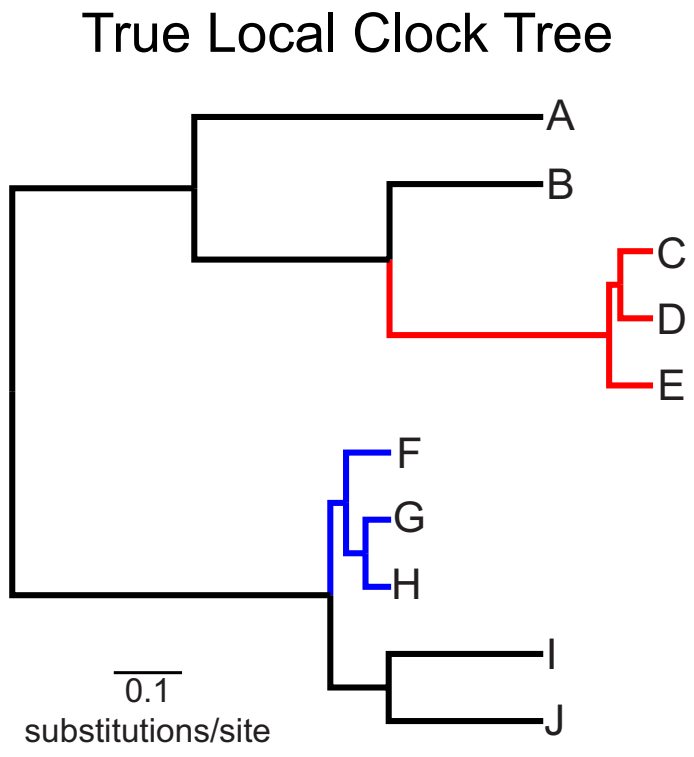
Summarizing MCMC under the DPP



MEAN PARTITION:

Useful for identifying lineages that may share similar properties

Summarizing MCMC under the DPP

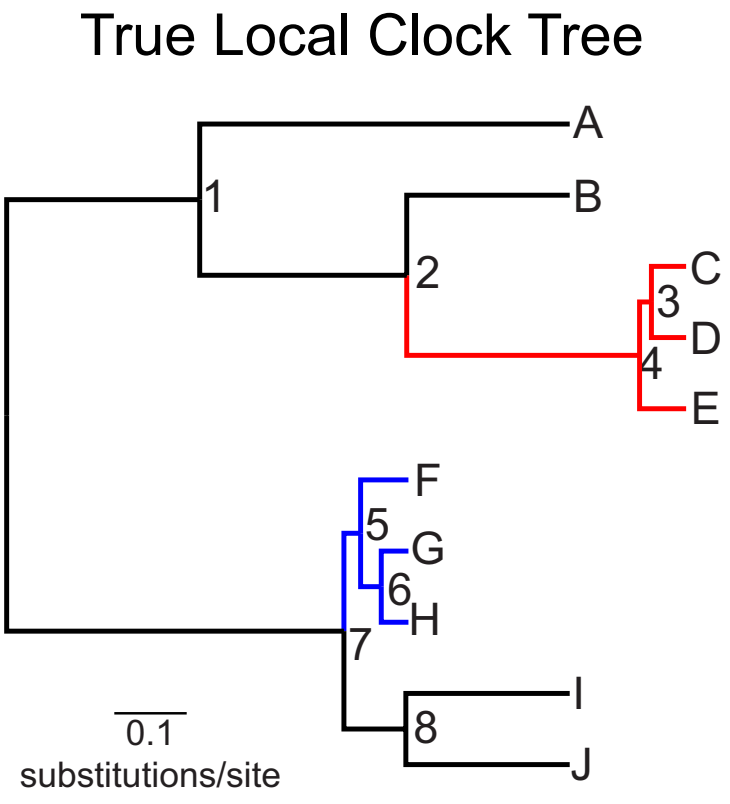


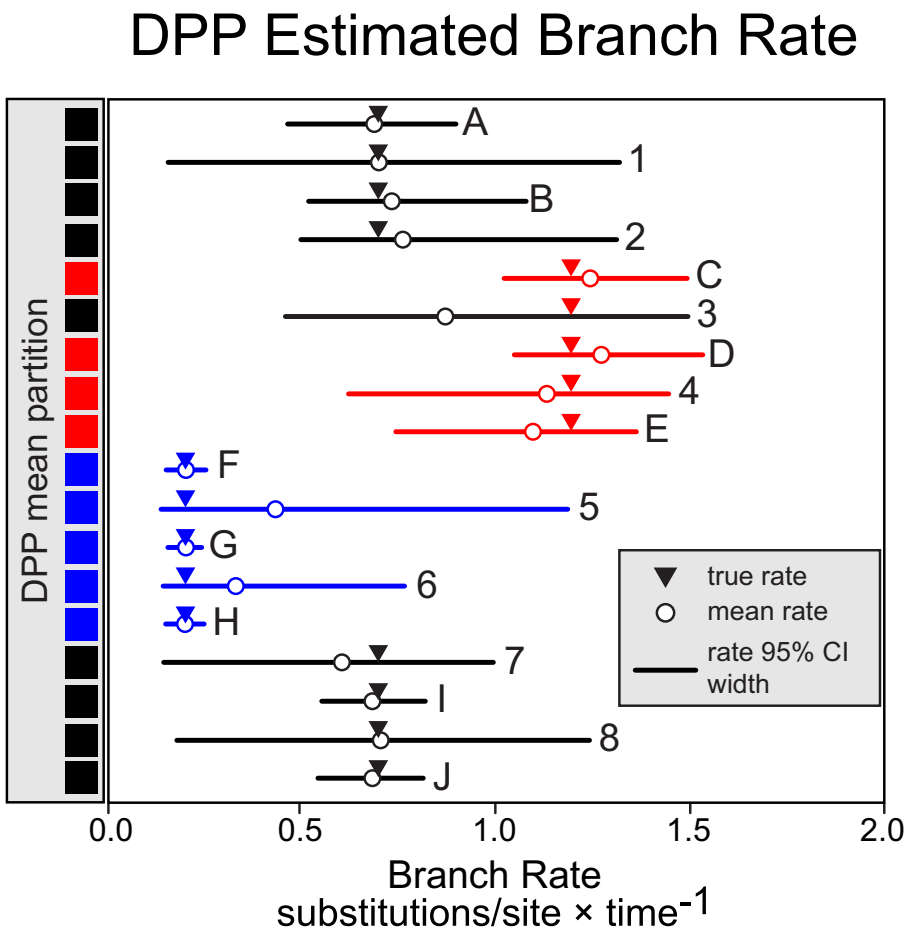
Branch lengths generated under a local molecular clock (LMC)

3 rate categories (substitutions/site*time^{-1}):

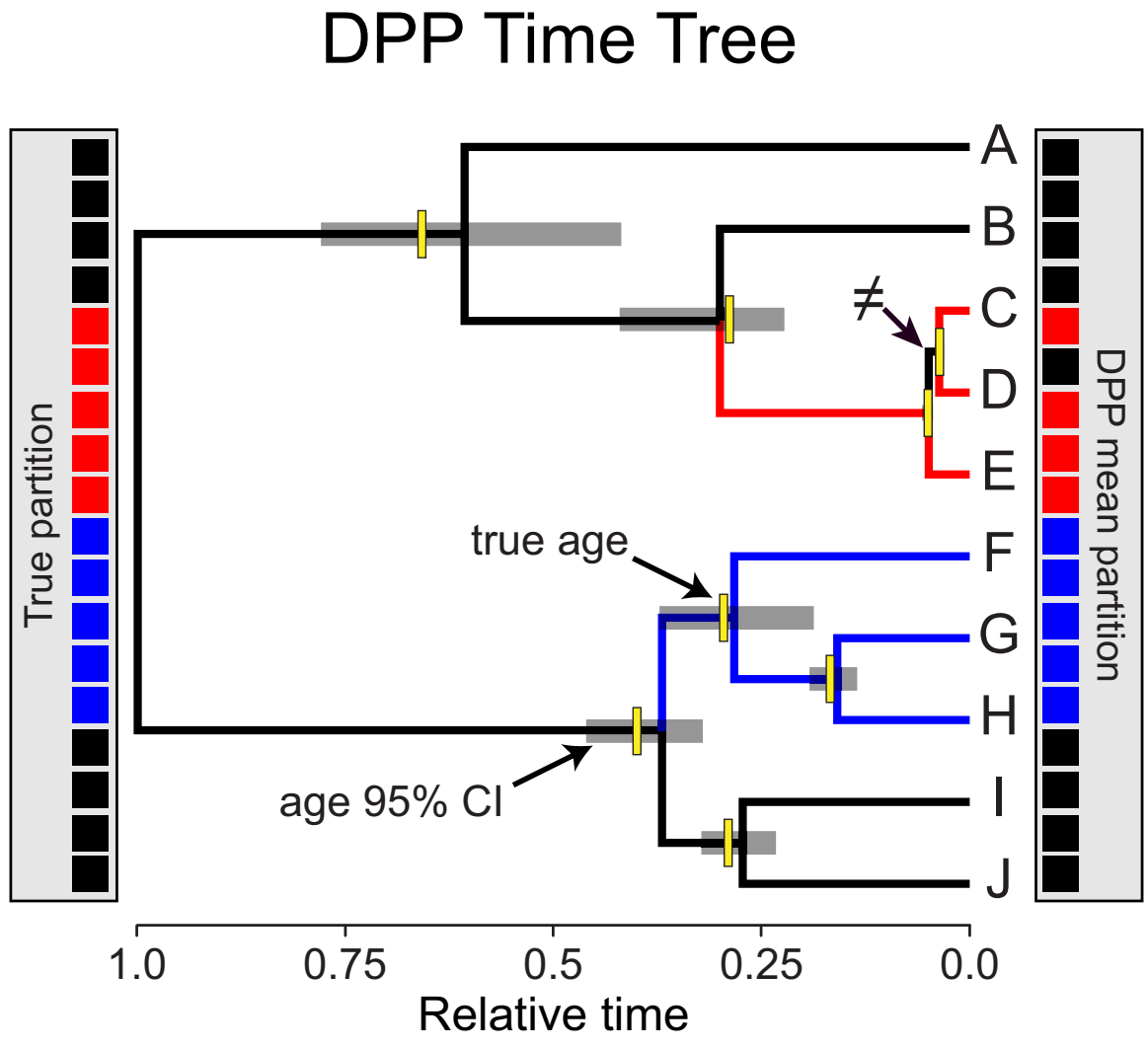
- **0.2**
- **0.7**
- **1.2**

Summarizing MCMC under the DPP





Summarizing MCMC under the DPP

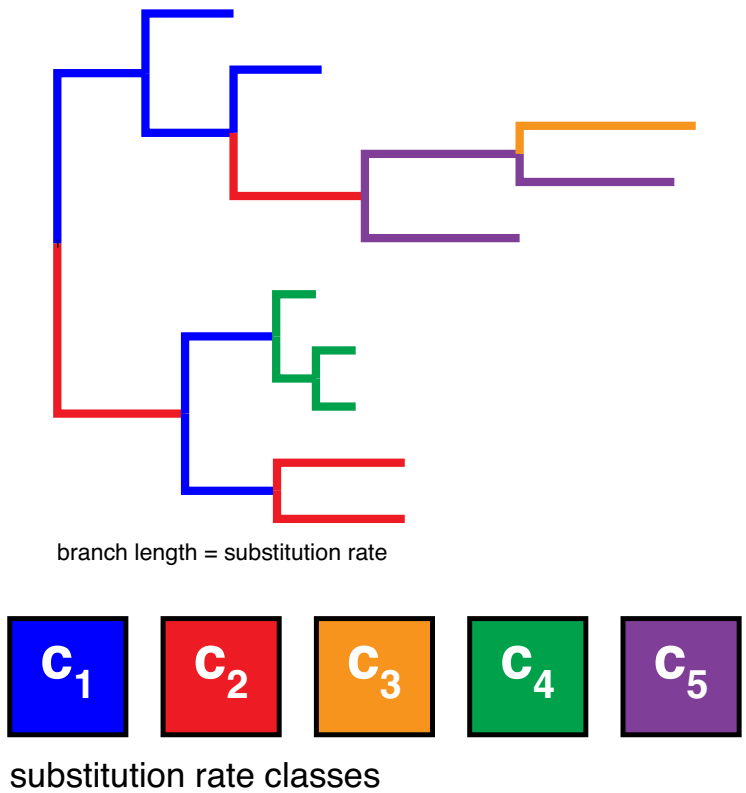


Lineage-Specific Substitution Rates

DPP provides robust estimates of branch-rate and node-age without significant loss in power

The flexibility of the DPP allows it to encompass different branch-wise models of substitution rate variation

Including cases in which distant branches have equivalent (or nearly equivalent) rates



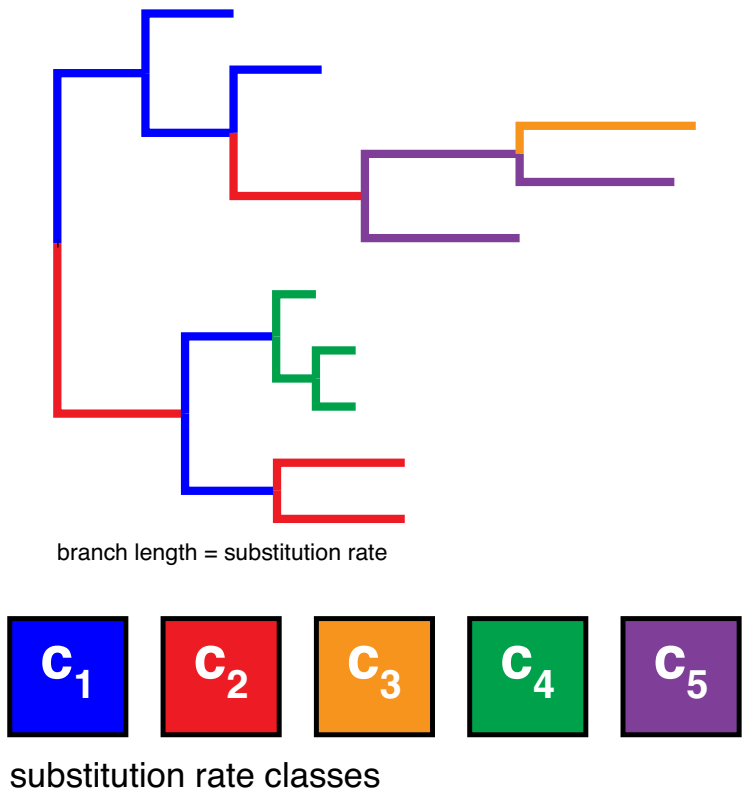
Availability: `DPPDiv @ http://phylo.bio.ku.edu/content/tracy-heath-dppdiv`

Lineage-Specific Substitution Rates

DPP provides robust estimates of branch-rate and node-age without significant loss in power

The mean branch partition found under the DPP allows for the identification of latent classes

Efficient MCMC implementations



Availability: `DPPDiv @ http://phylo.bio.ku.edu/content/tracy-heath-dppdiv`

let's take a break...

Bayesian Divergence Time Estimation

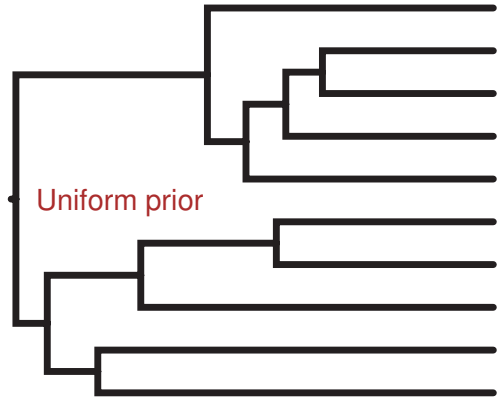
Estimating divergence times relies on 2 main elements:

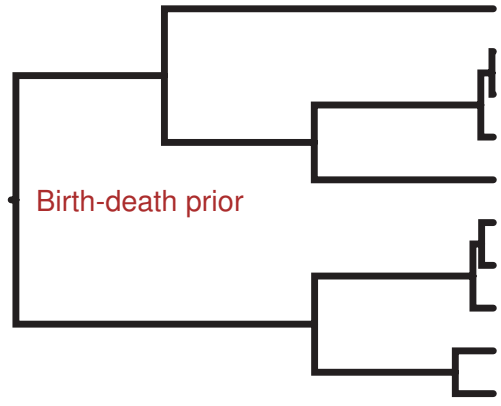
- Branch-specific rates: $f(\boldsymbol{\mathcal{R}} \mid \theta_{\mathcal{R}})$
- Node ages: $f(\boldsymbol{\mathcal{A}} \mid \theta_{\mathcal{A}}, \boldsymbol{\mathcal{C}})$

Priors on Node Times

Relaxed clock Bayesian analyses require a prior distribution on node times

Uniform prior: the time at a given node has equal probability across the interval between the time of the parent node and the time of the oldest daughter node



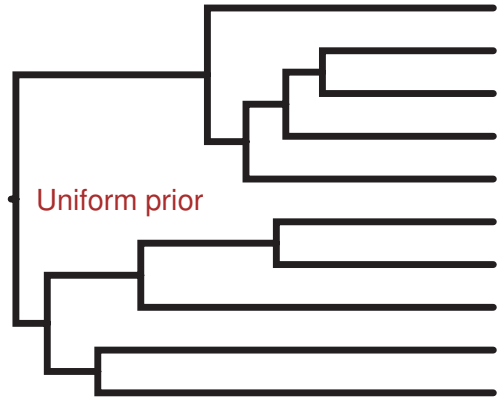


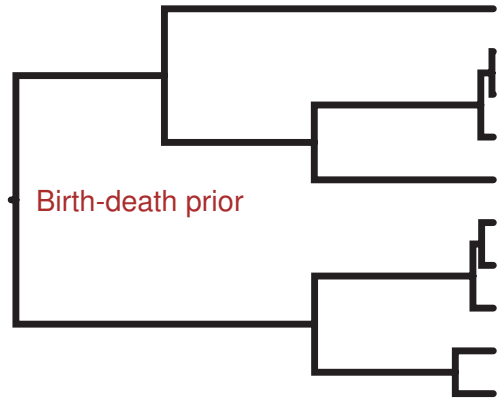
Priors on Node Times

Relaxed clock Bayesian analyses require a prior distribution on node times

Birth-death prior: node times are sampled from a stochastic process with parameters for speciation, $\mathcal{S}$, and extinction, $\mathcal{E}$, (and in some cases taxon sampling)

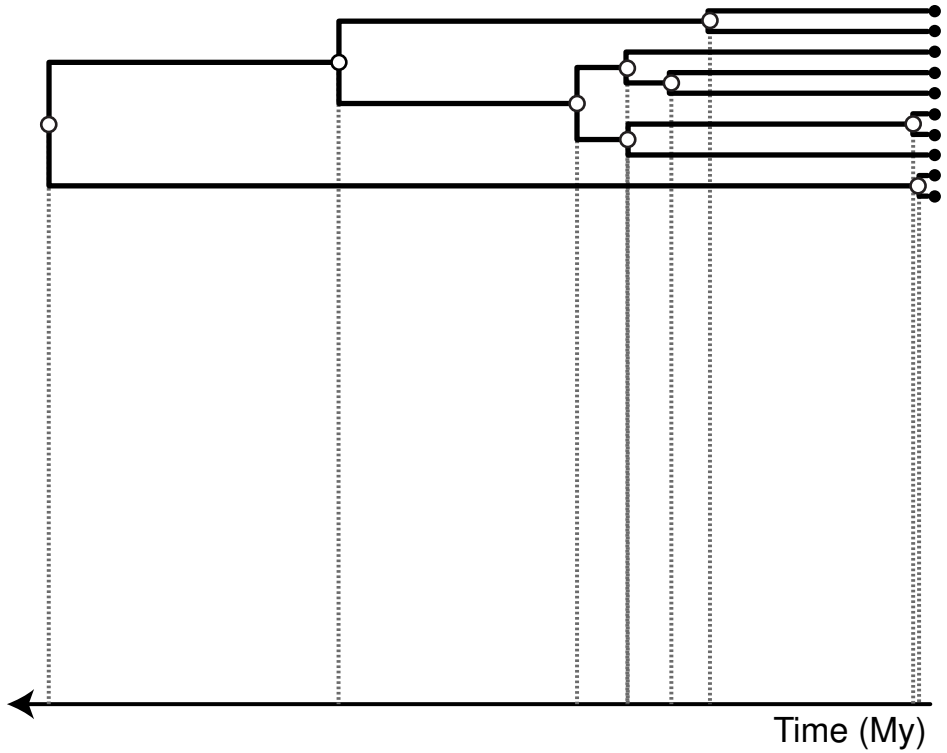
$$f(\mathcal{A} \mid \mathcal{S}, \mathcal{E})$$



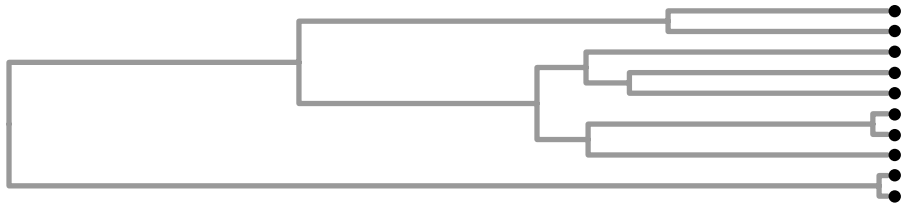


Fossil Calibration

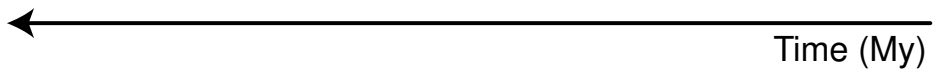
Fossil and geological data can be used to estimate the absolute ages of ancient divergences



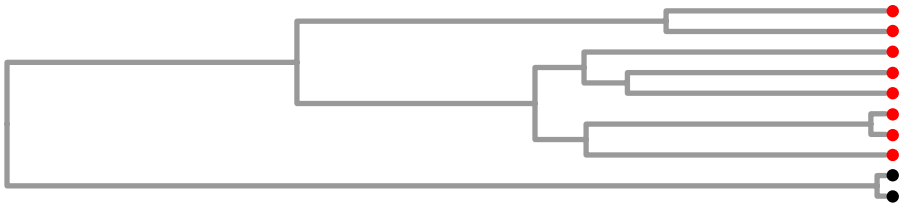
Fossil Calibration



The ages of extant taxa are known



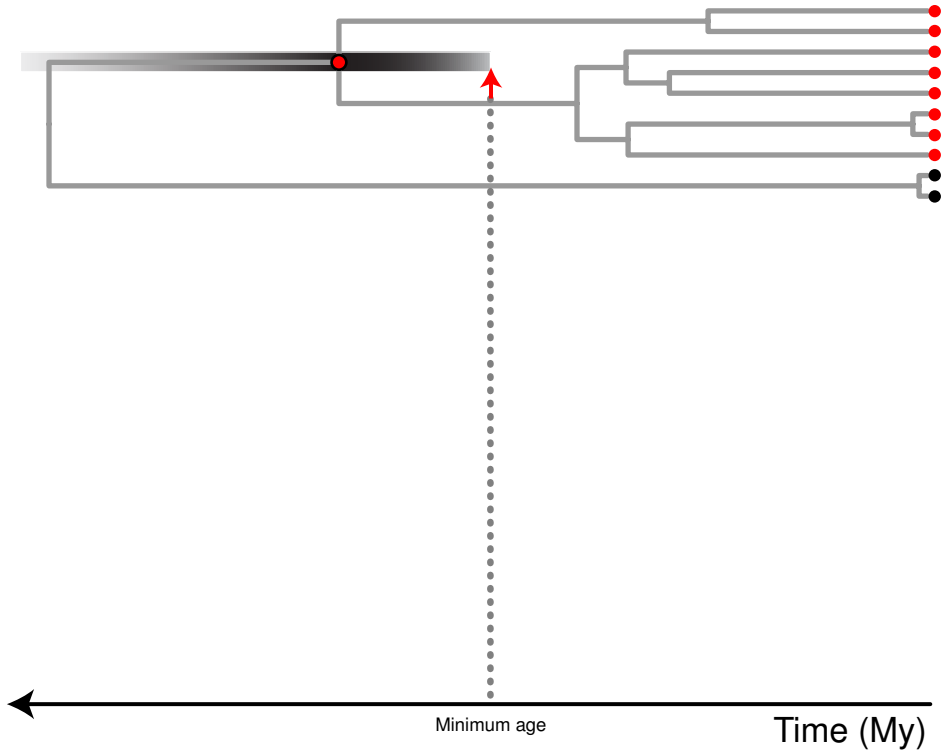
Fossil Calibration



Fossil taxa are assigned to monophyletic clades

Fossil Calibration

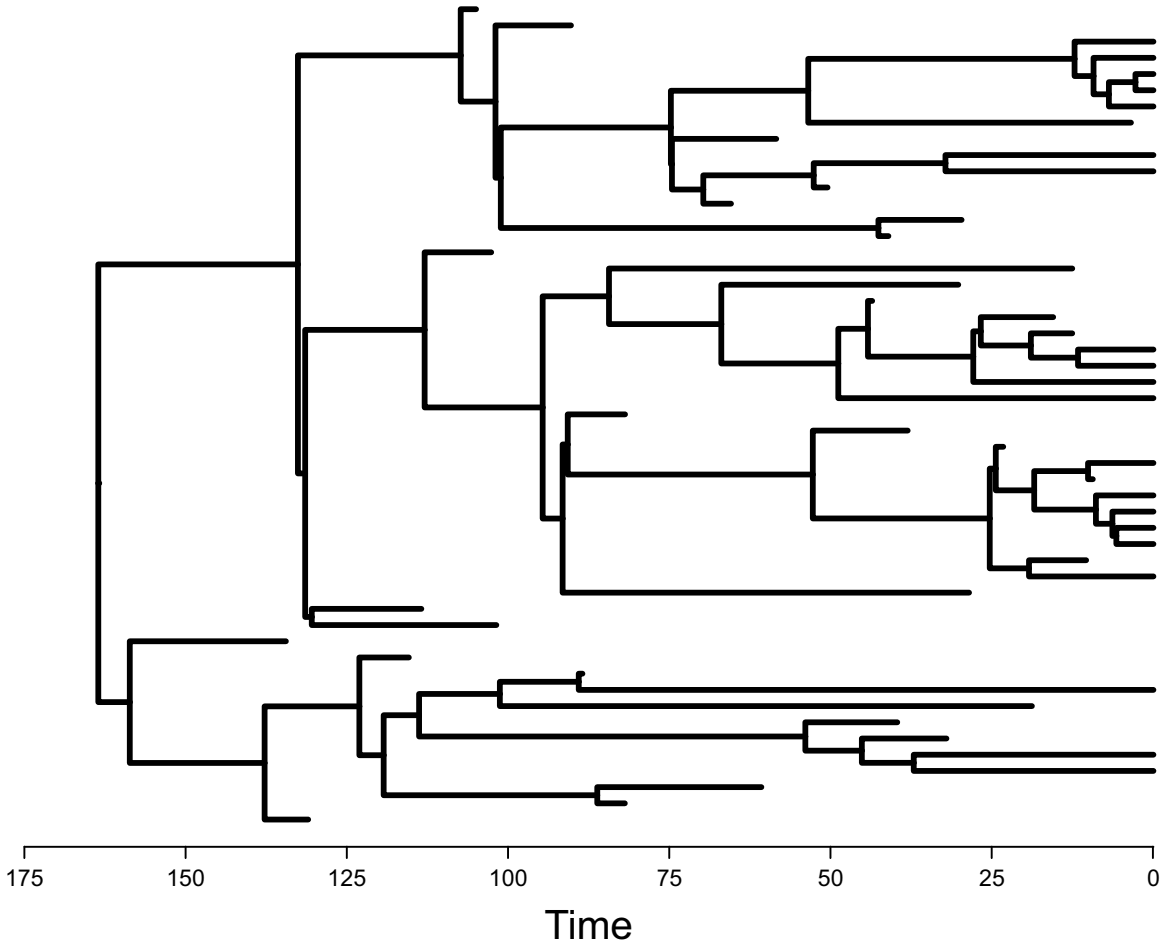
Fossil taxa are assigned to monophyletic clades and constrain the age of the MRCA



Modeling Branching Processes

Assume constant rates of speciation ($\mathcal{S}$) and extinction ($\mathcal{E}$)

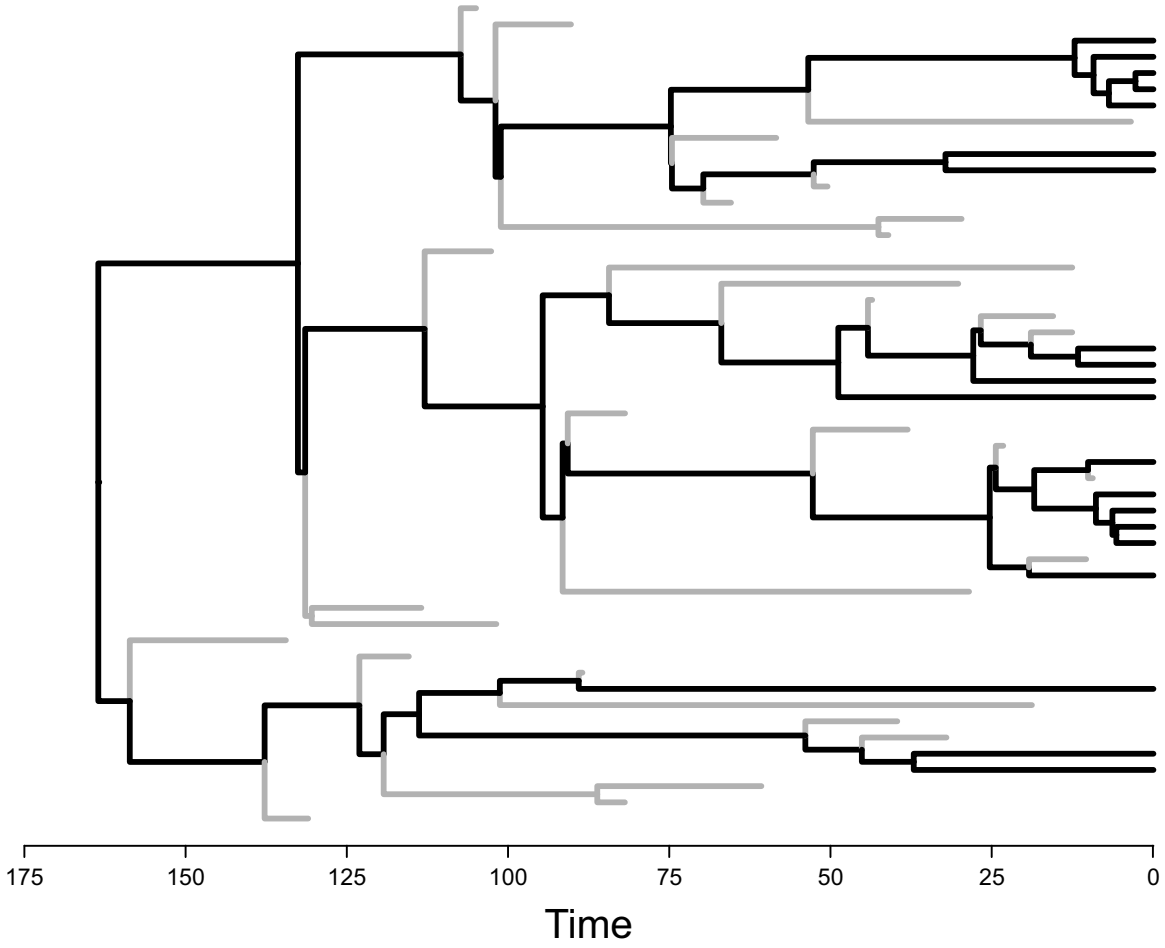
(20 extant taxa)



Modeling Branching Processes

Assume constant rates of speciation ($\mathcal{S}$) and extinction ($\mathcal{E}$)

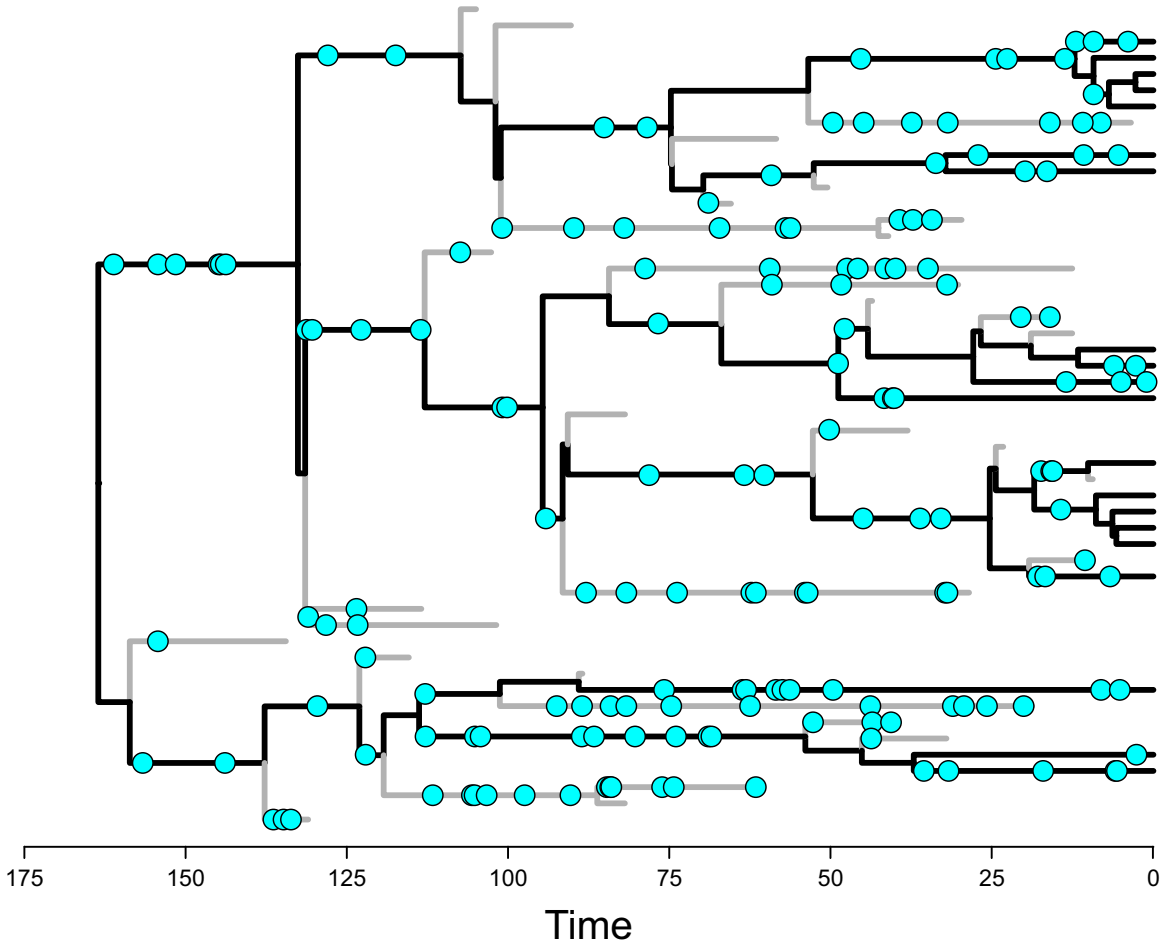
(20 extant taxa)



Modeling Taphonomic Processes

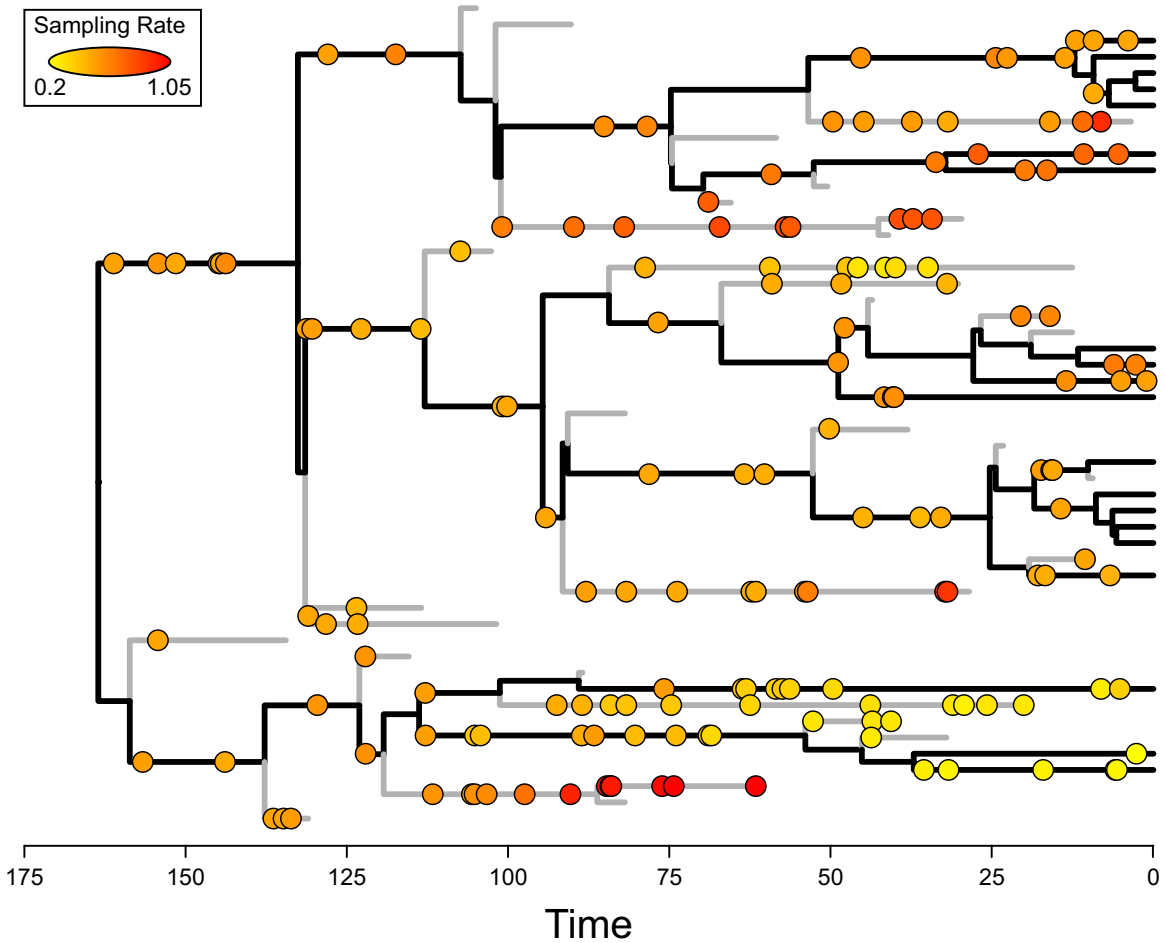
Fossilization events were generated according to a Poisson process

this example has 162 fossilization events



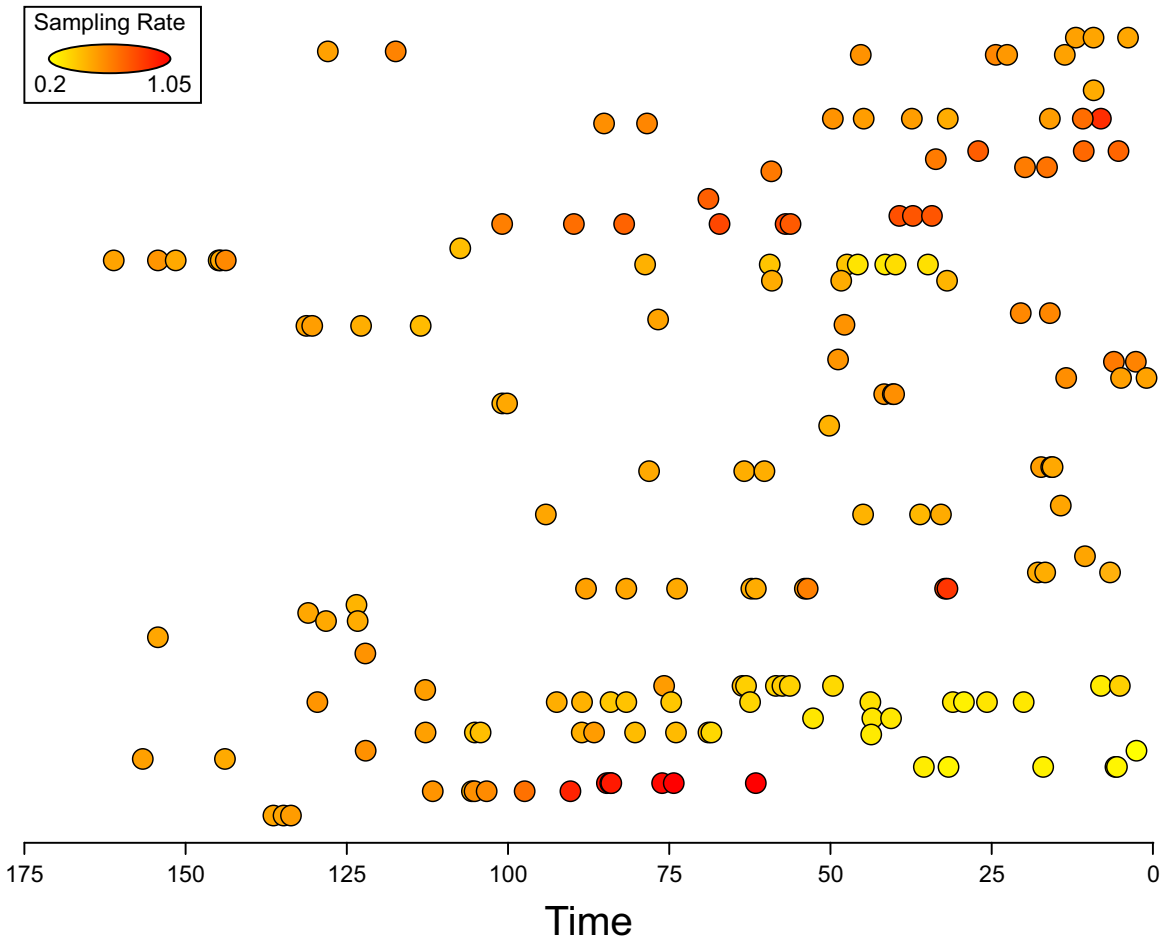
Modeling Taphonomic Processes

The fossil sampling rate was evolved under an autocorrelated Brownian motion model



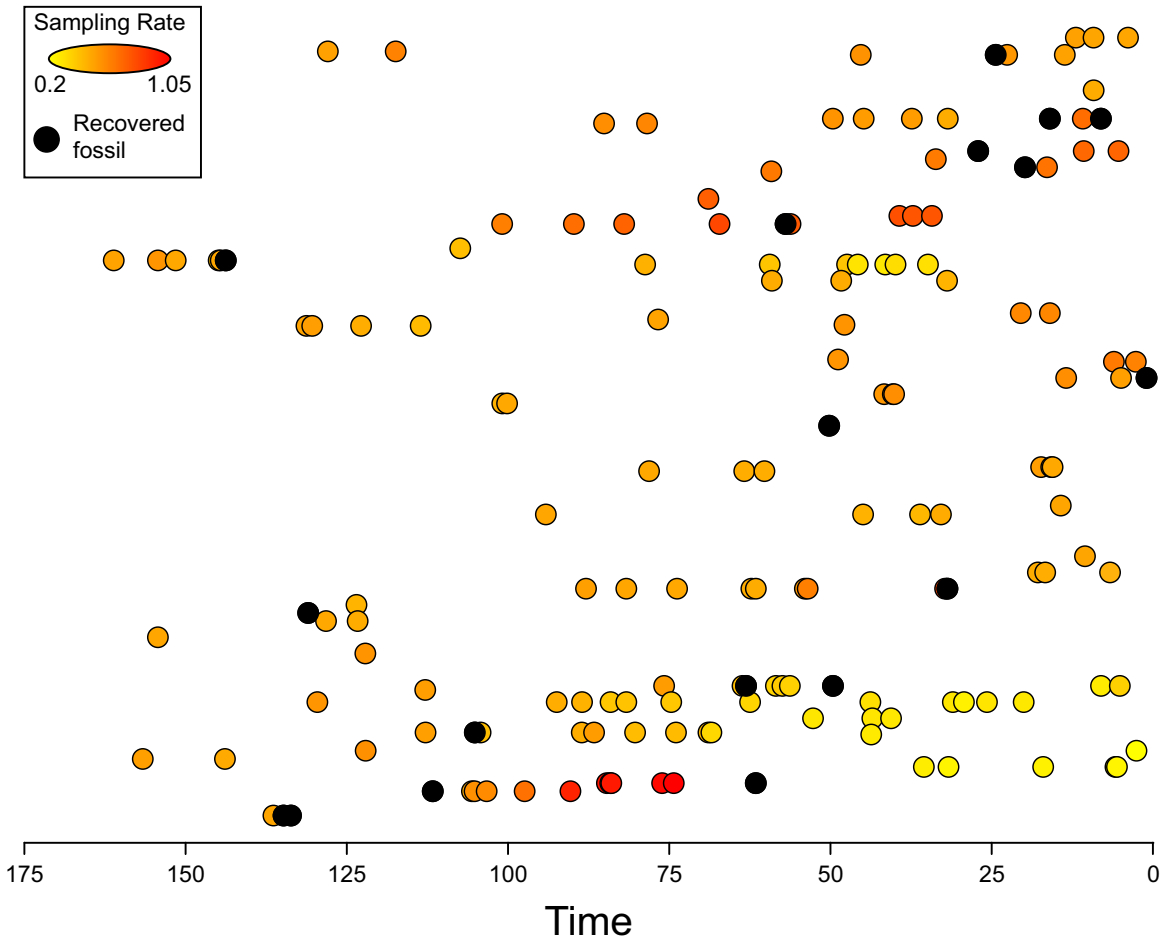
Modeling Taphonomic Processes

The fossil sampling rate was evolved under an autocorrelated Brownian motion model



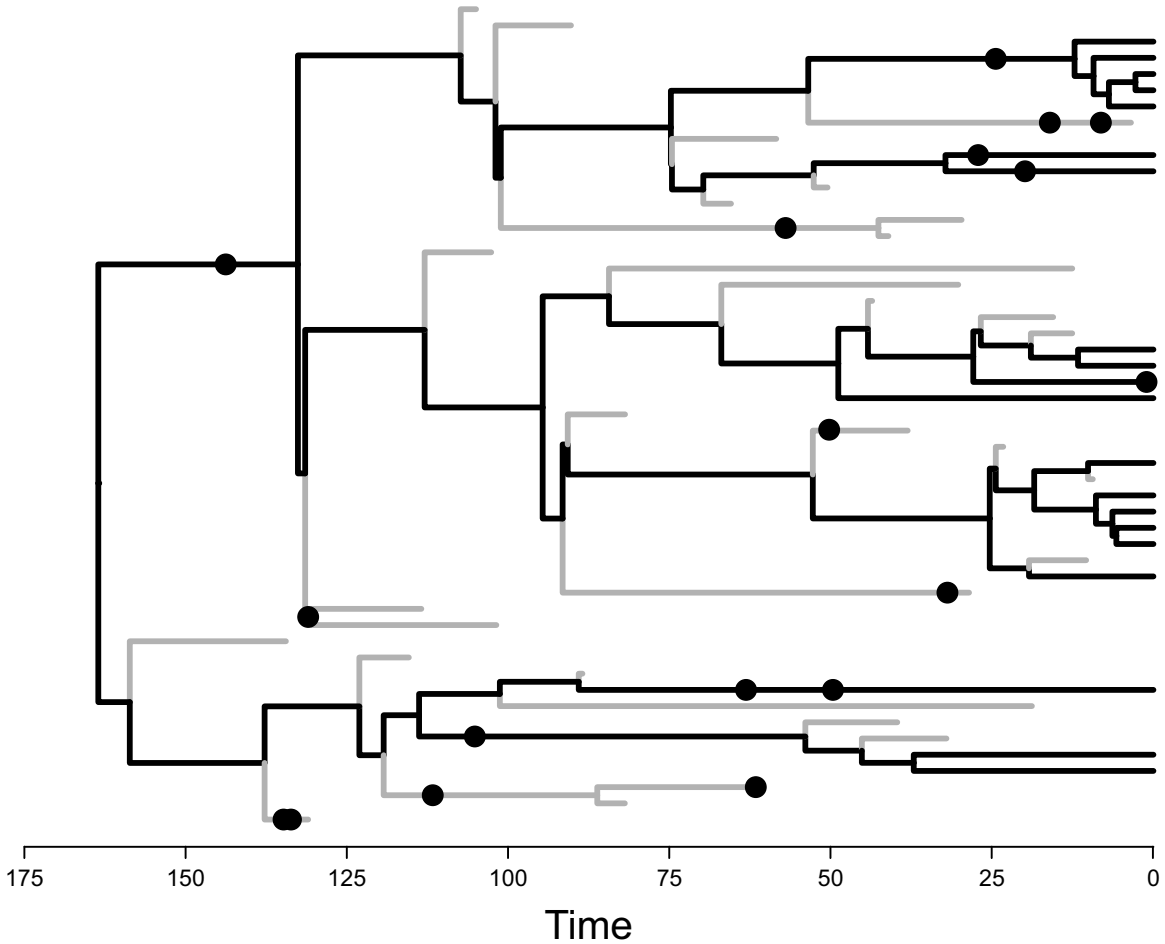
Modeling Taphonomic Processes

18 fossils were "recovered" in proportion to their sampling rates



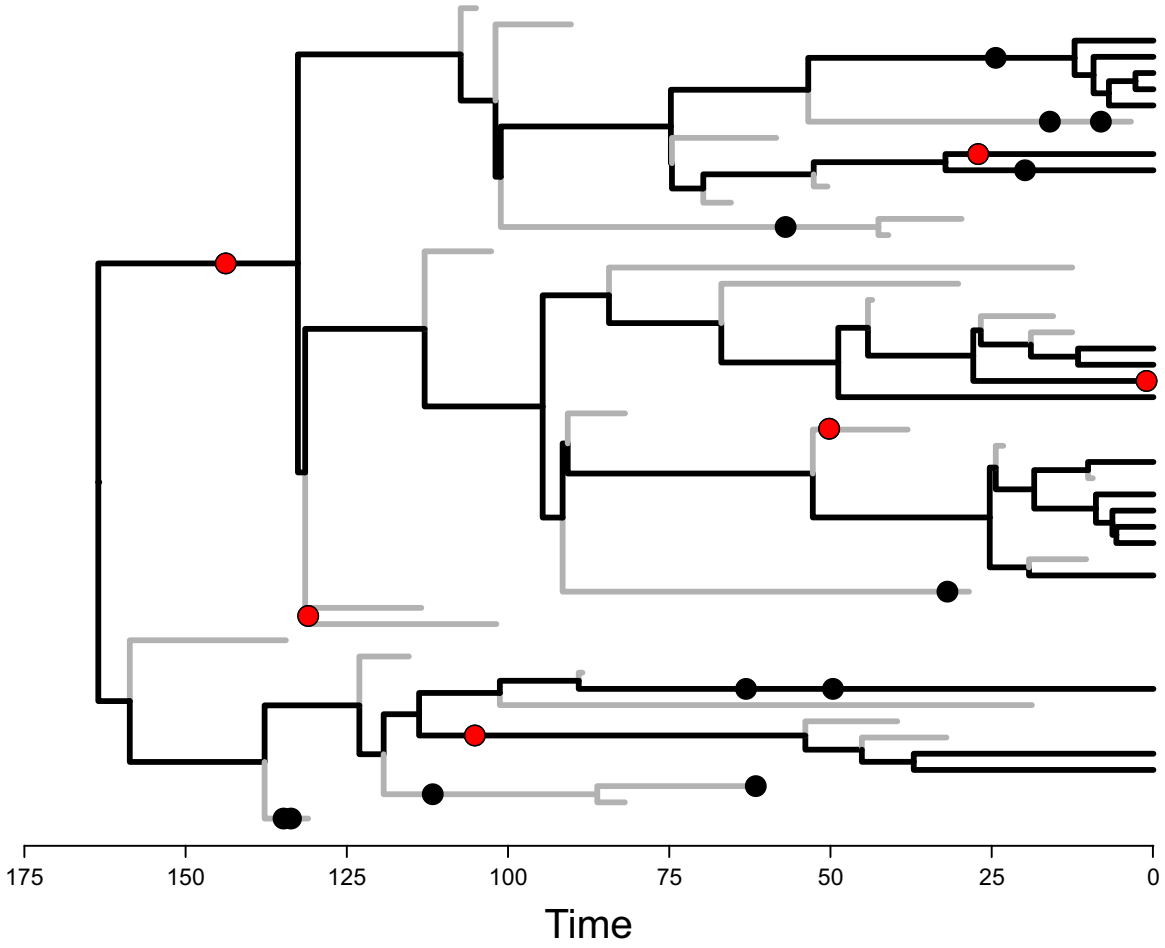
Recovered Fossils

Assume we know the true phylogenetic placement of the recovered fossils



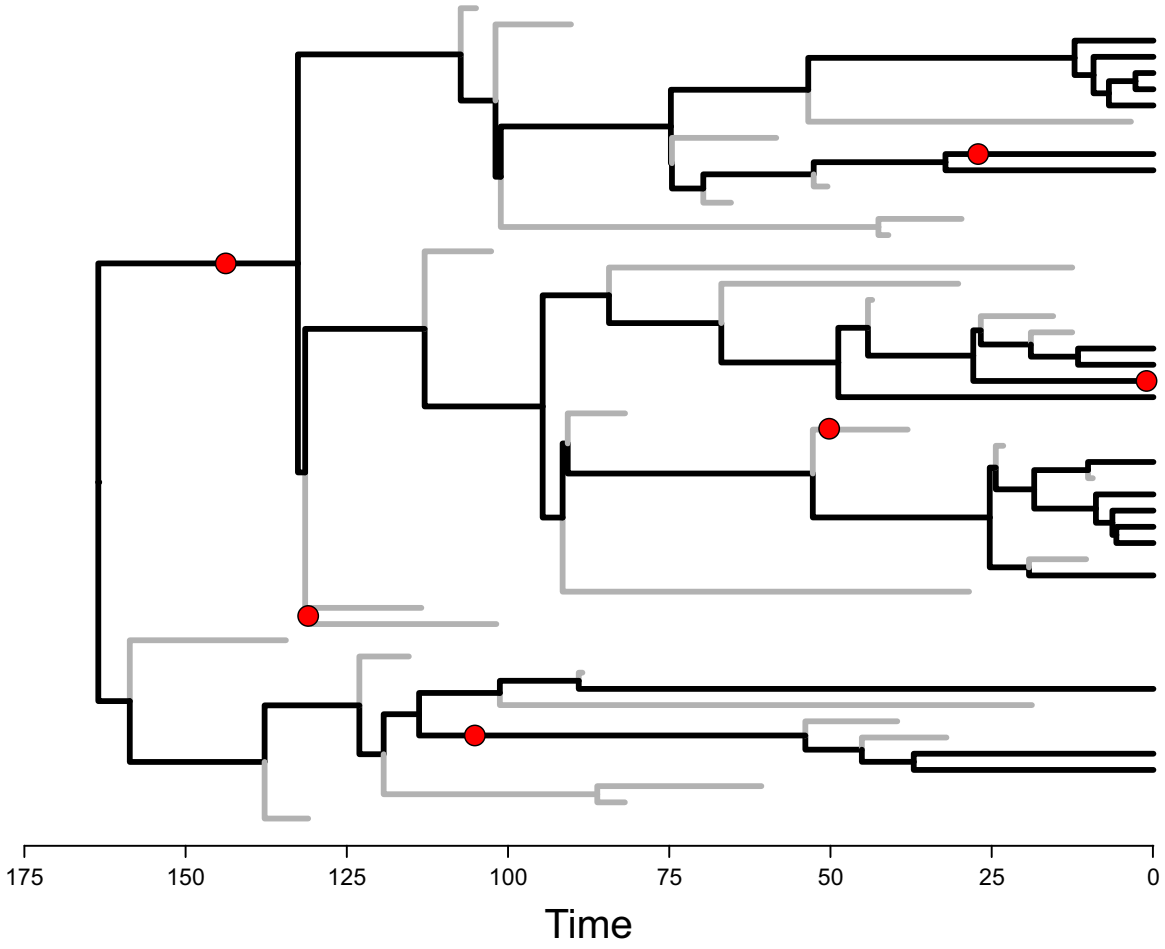
Calibration Fossils

Only the oldest fossil assigned to a given node can be used for calibration



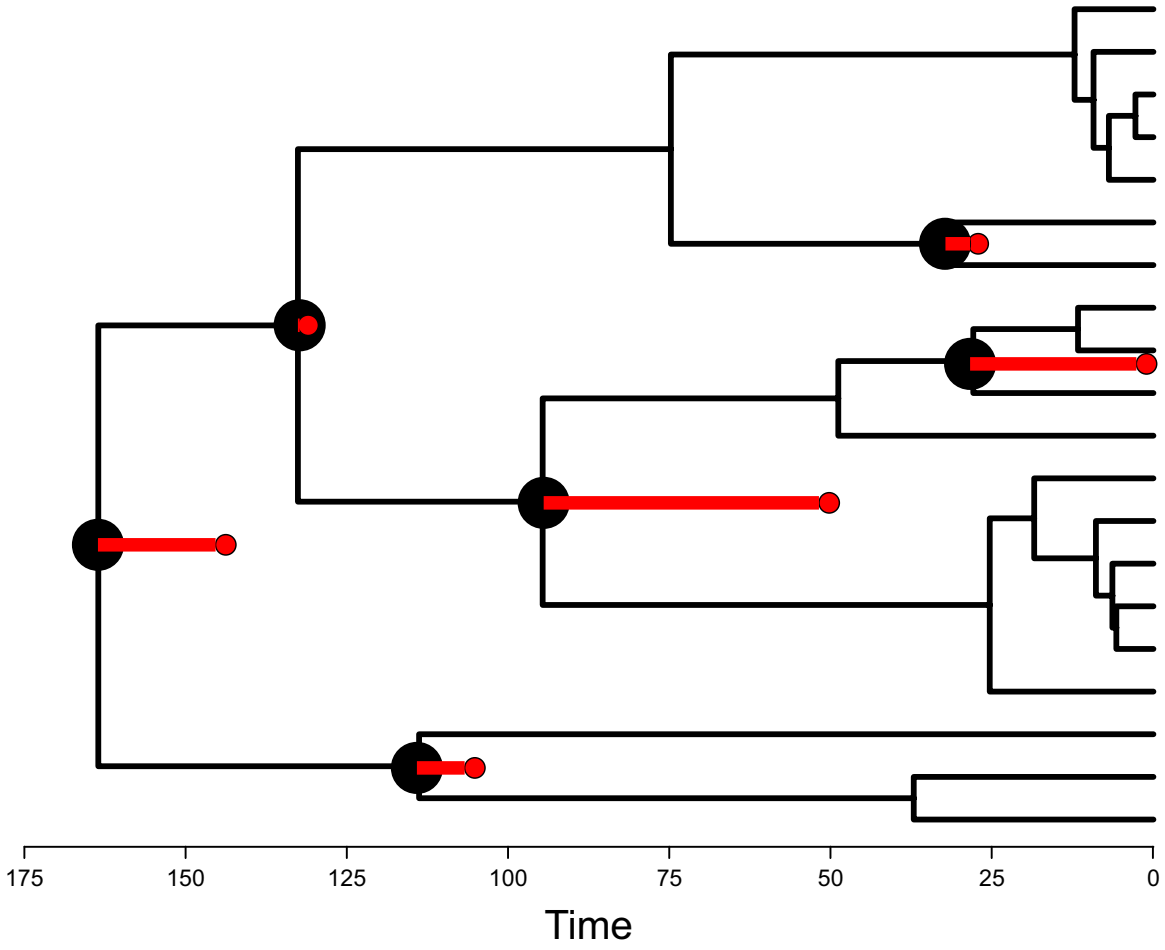
Calibration Fossils

Only the oldest fossil assigned to a given node can be used for calibration



Calibration Fossils

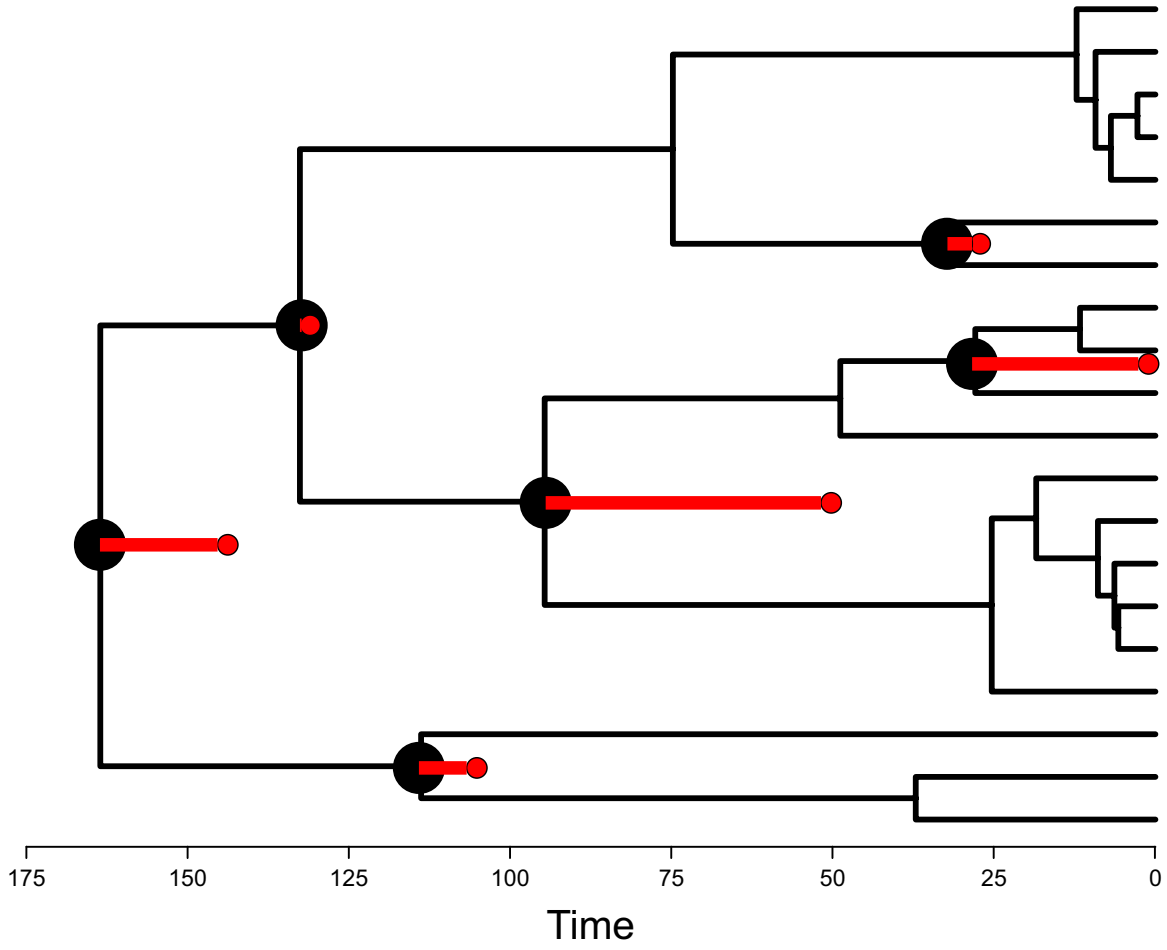
Only the oldest fossil assigned to a given node can be used for calibration



Calibration Fossils

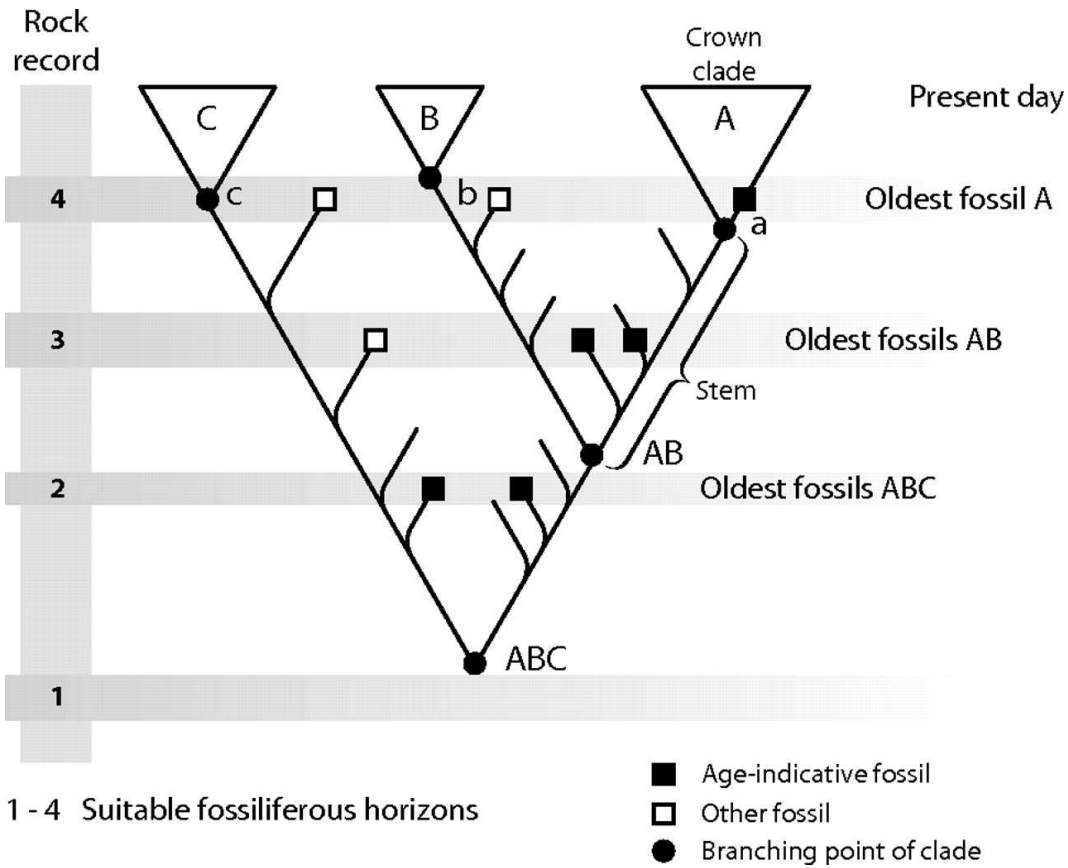
Taphonomic bias

- disparity in fossilization and preservation
- geographical distribution
- recovery bias
- identification



Assigning Fossils to Clades

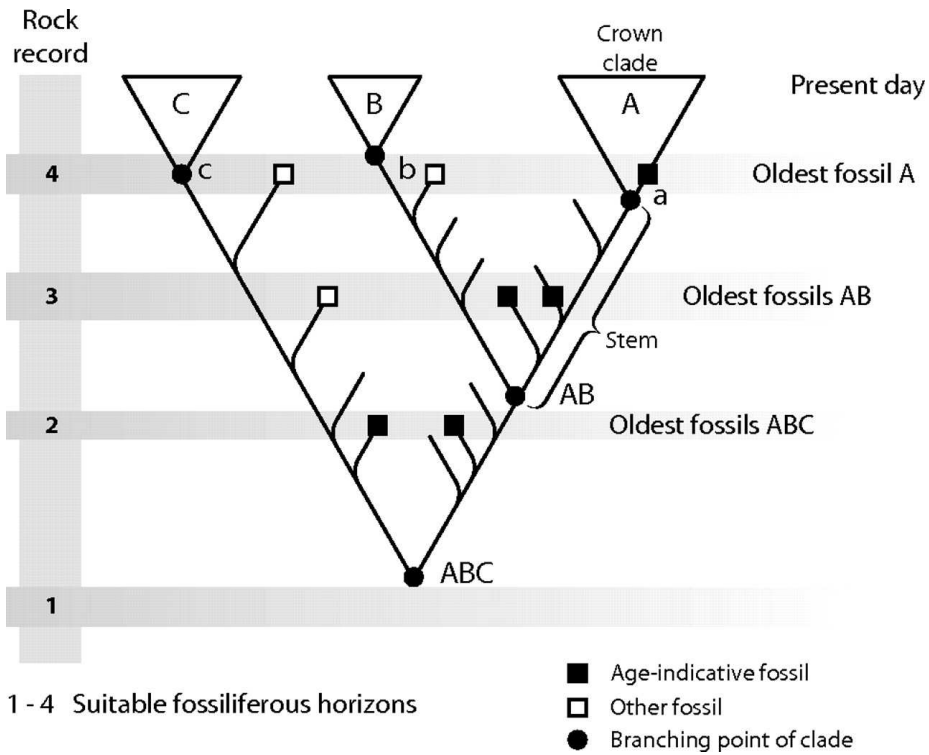
Misplaced fossils can affect node age estimates throughout the tree — if the fossil is older than its presumed MRCA



(figure from Benton & Donoghue *Mol. Biol. Evol.* 2007)

Assigning Fossils to Clades

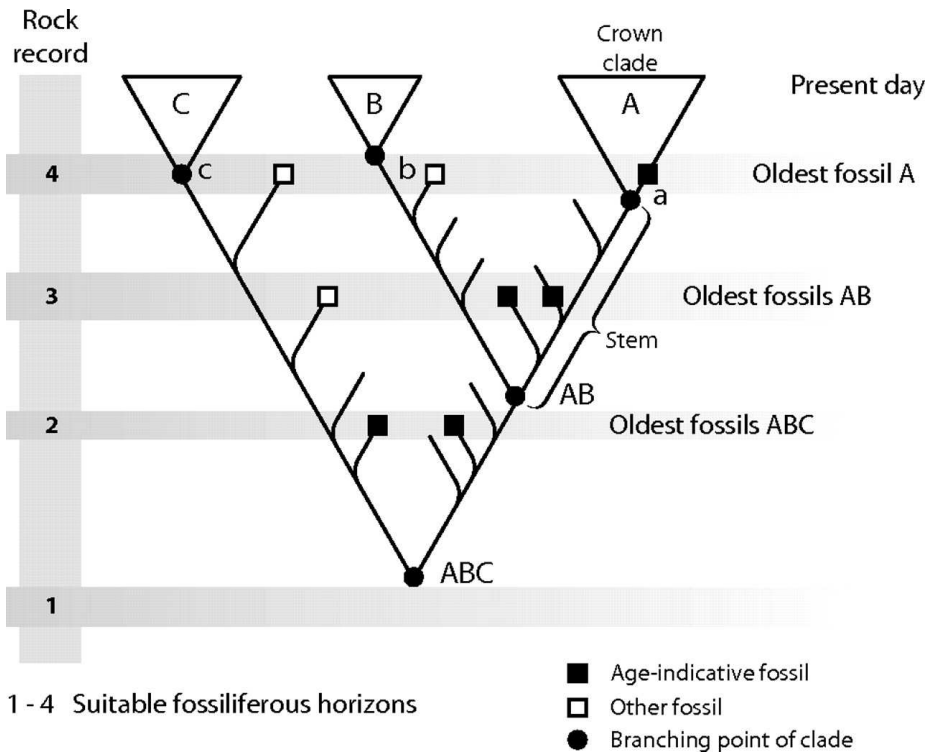
Crown clade: all living species and their most-recent common ancestor (MRCA)



(figure from Benton & Donoghue *Mol. Biol. Evol.* 2007)

Assigning Fossils to Clades

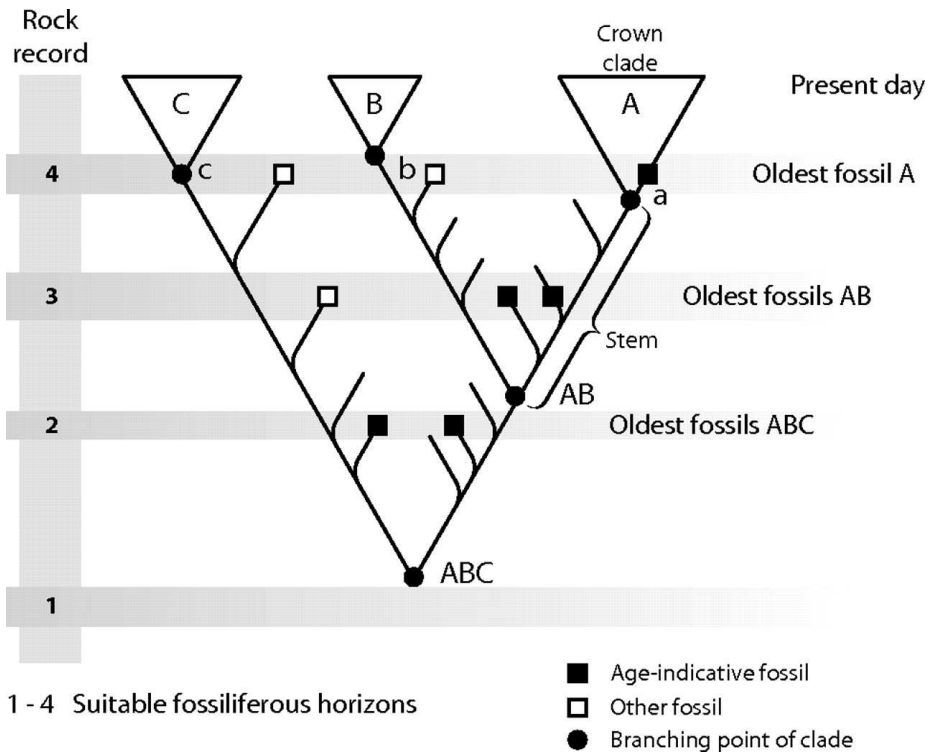
Stem lineages: purely fossil forms that are closer to their descendant crown clade than any other crown clade



(figure from Benton & Donoghue *Mol. Biol. Evol.* 2007)

Assigning Fossils to Clades

Fossiliferous horizons: the sources in the rock record for relevant fossils

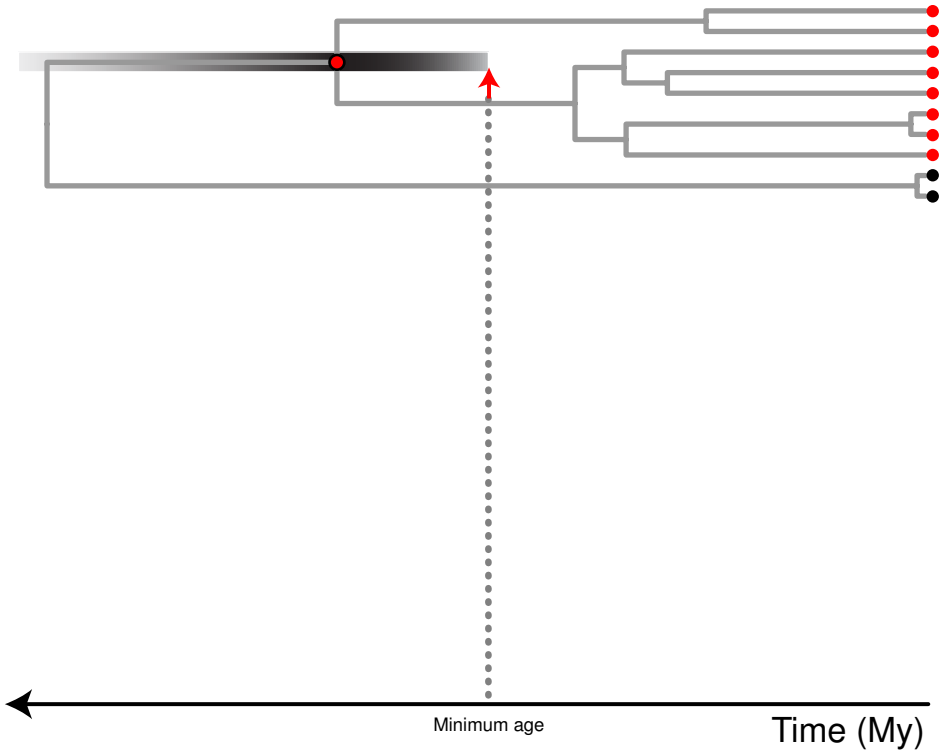


(figure from Benton & Donoghue *Mol. Biol. Evol.* 2007)

Fossil Calibration

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

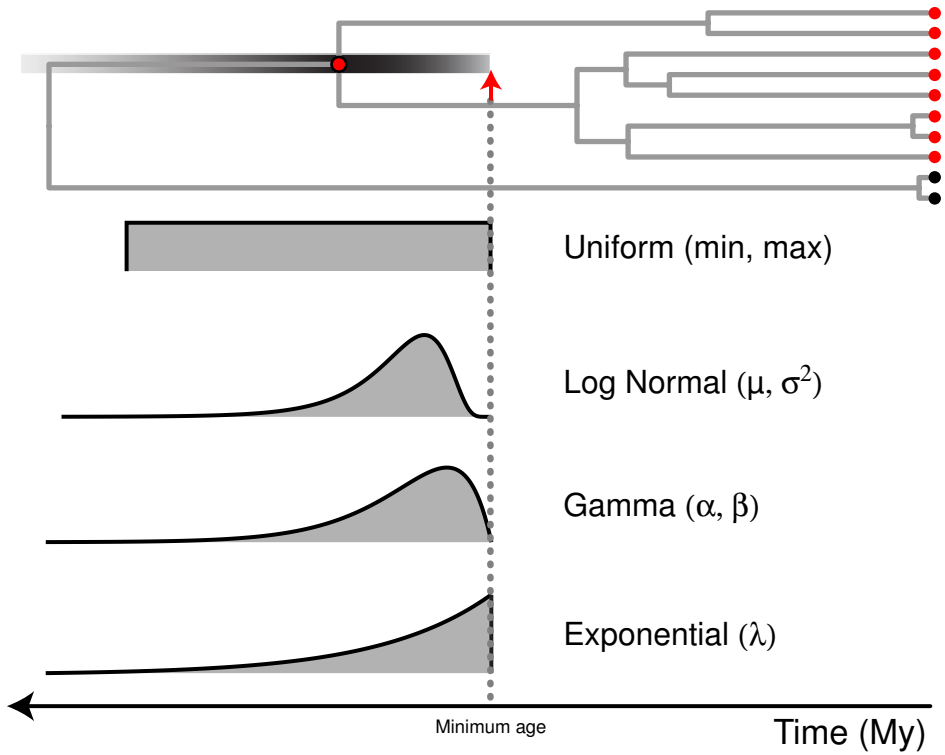
Reliable **maximum** bounds are typically unavailable



Prior Densities on Calibrated Nodes

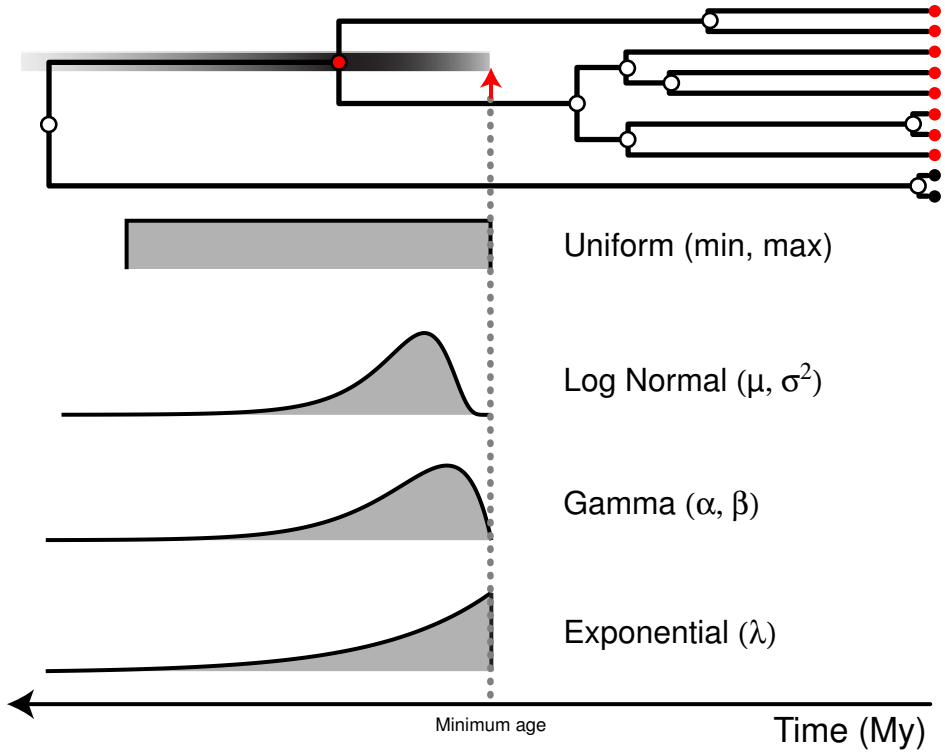
Parametric distributions are typically off-set by the age of the oldest fossil assigned to a clade

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



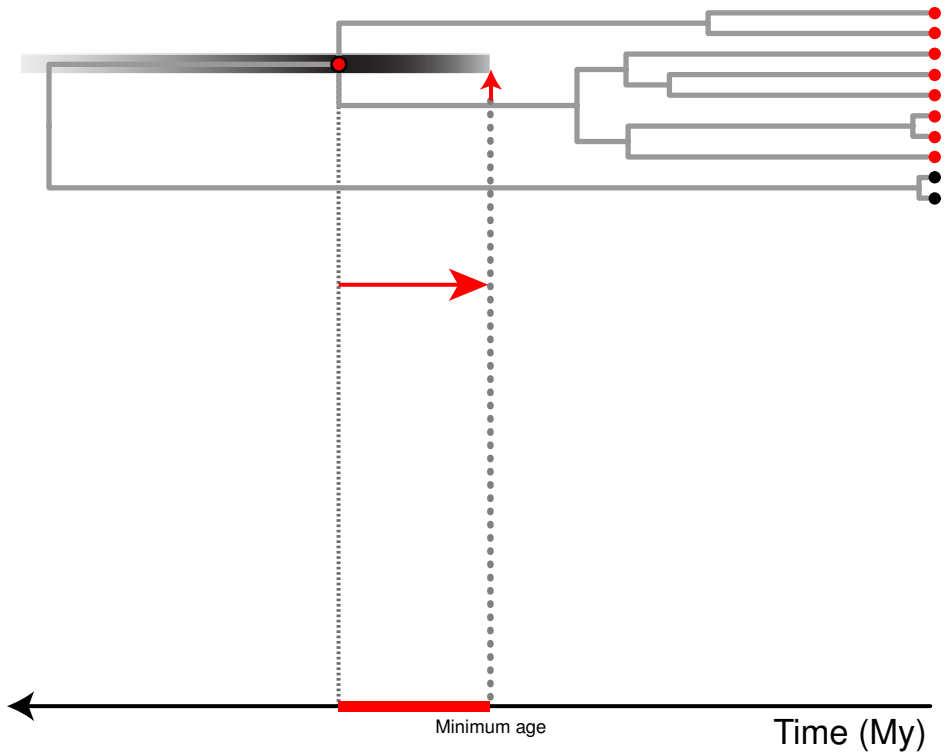
Prior Densities on Calibrated Nodes

Describe the waiting time between the divergence event and the age of the oldest fossil



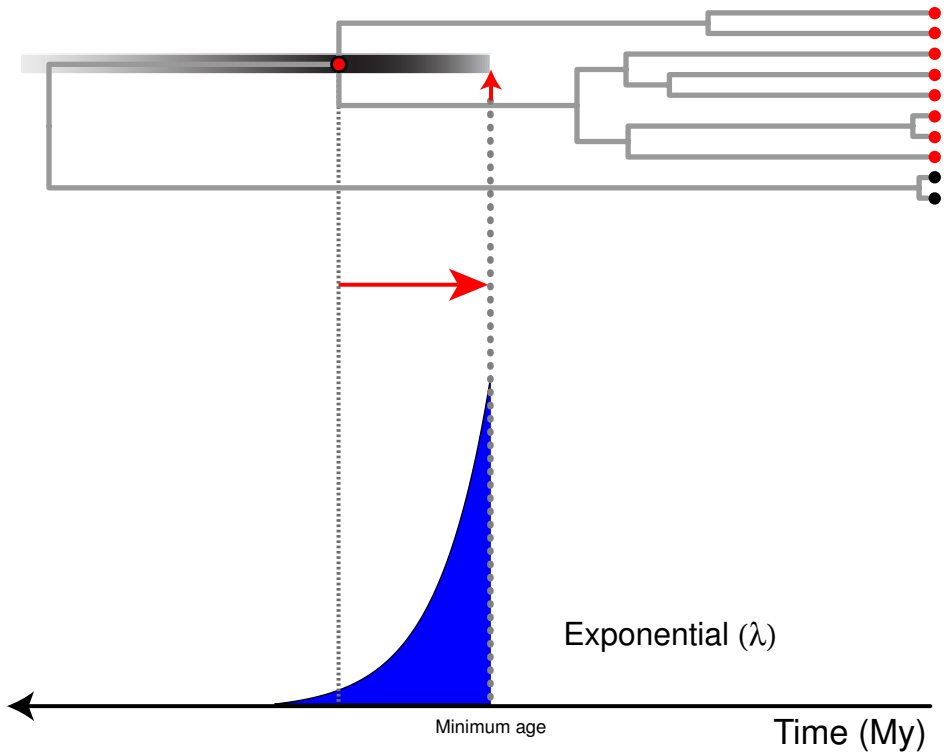
Prior Densities on Calibrated Nodes

Describe the waiting time between the divergence event and the age of the oldest fossil



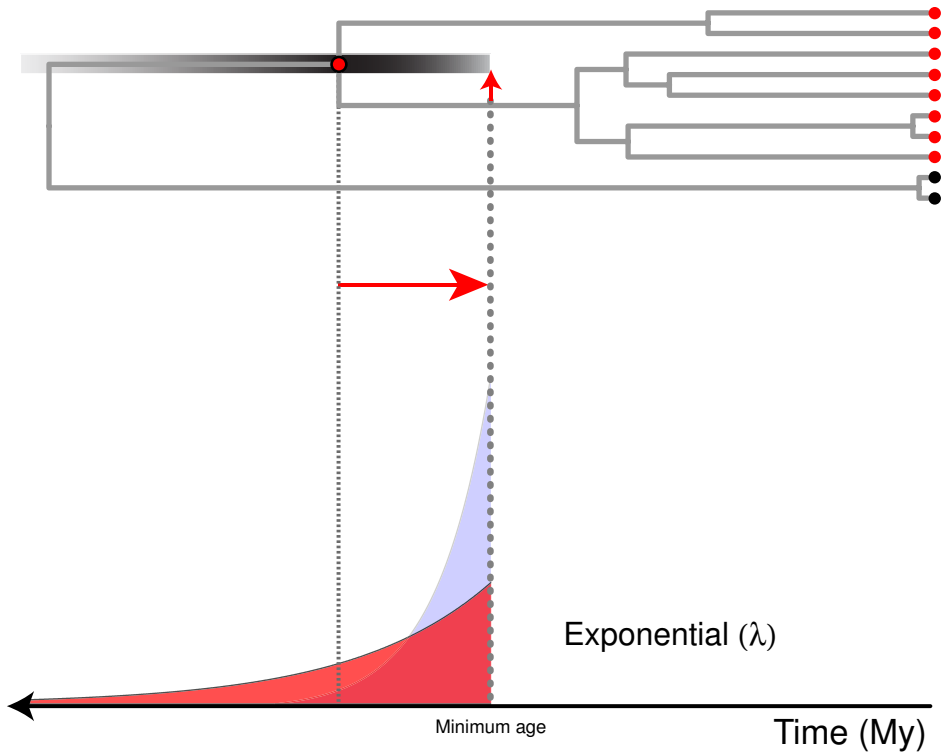
Prior Densities on Calibrated Nodes

Overly **informative** priors can bias node age estimates to be too young

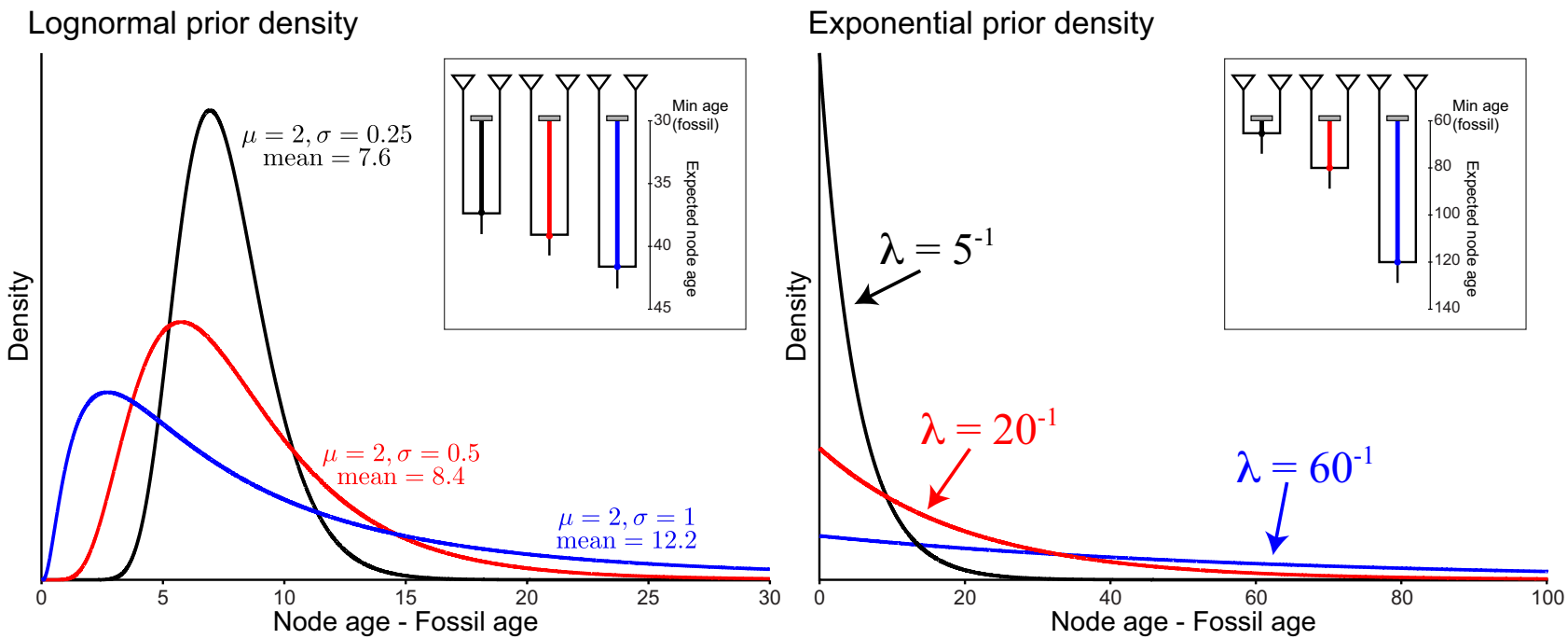


Prior Densities on Calibrated Nodes

Uncertainty in the age of the MRCA of the clade relative to the age of the fossil may be better captured by **vague** prior densities

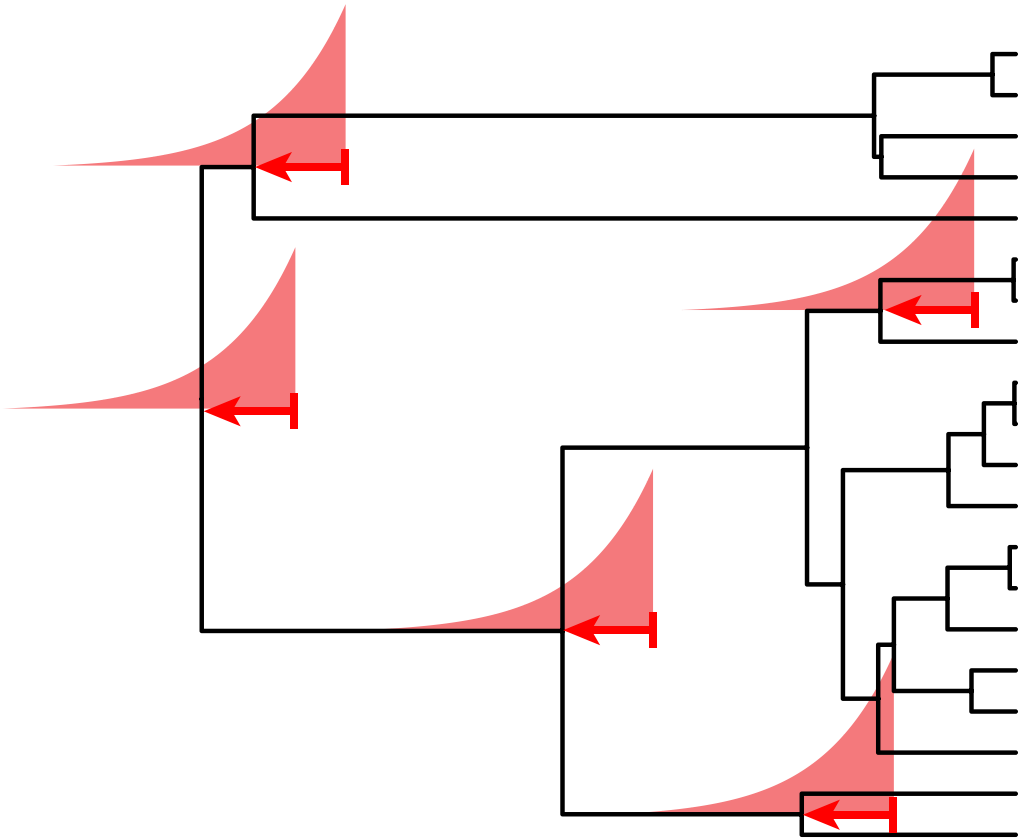


Prior Densities on Calibrated Nodes



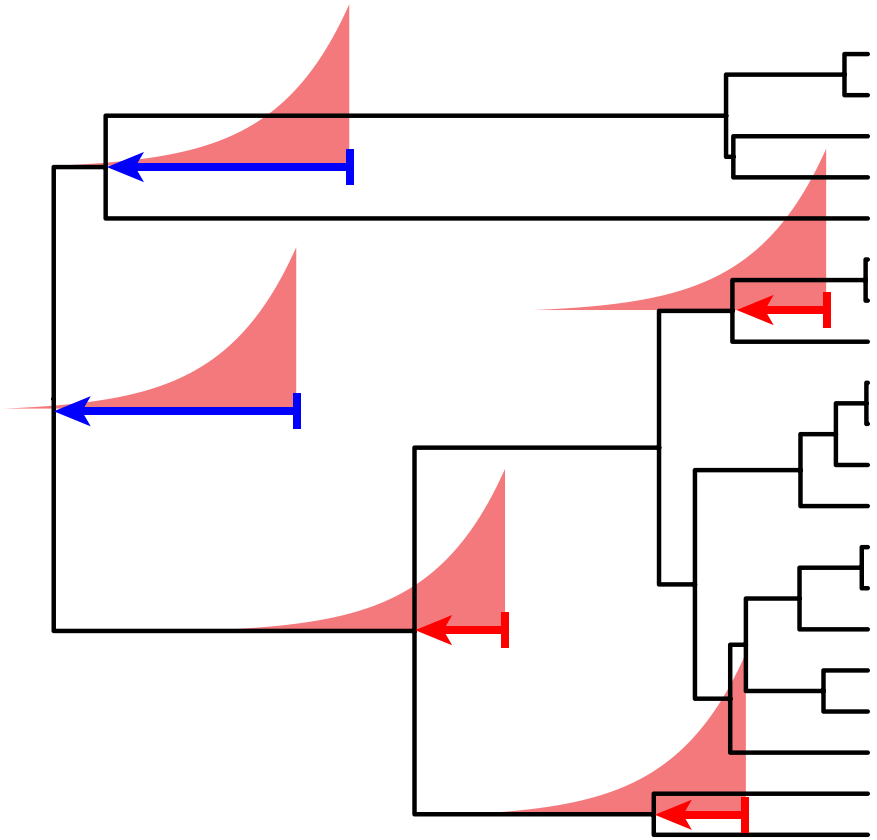
Priors on Multiple Calibrations

It is unlikely that multiple fossil calibrations can be characterized by a single prior density



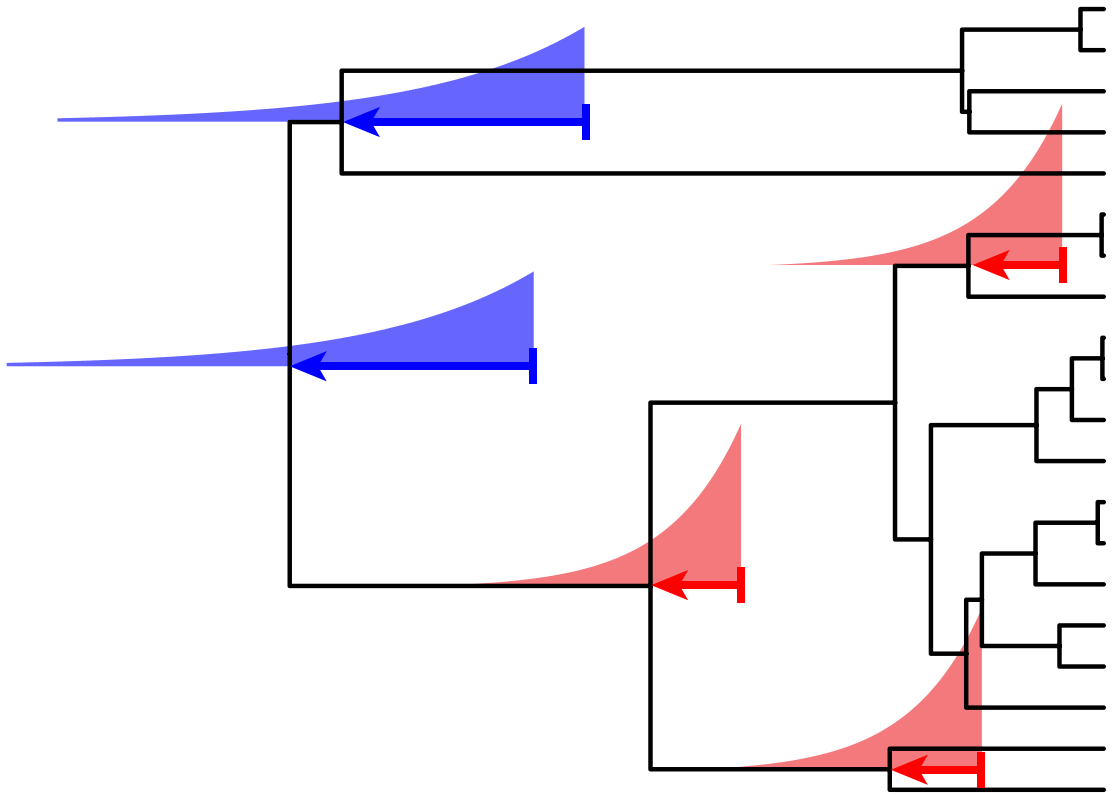
Priors on Multiple Calibrations

An appropriate prior for some nodes can also be an overly **informative** prior for other nodes



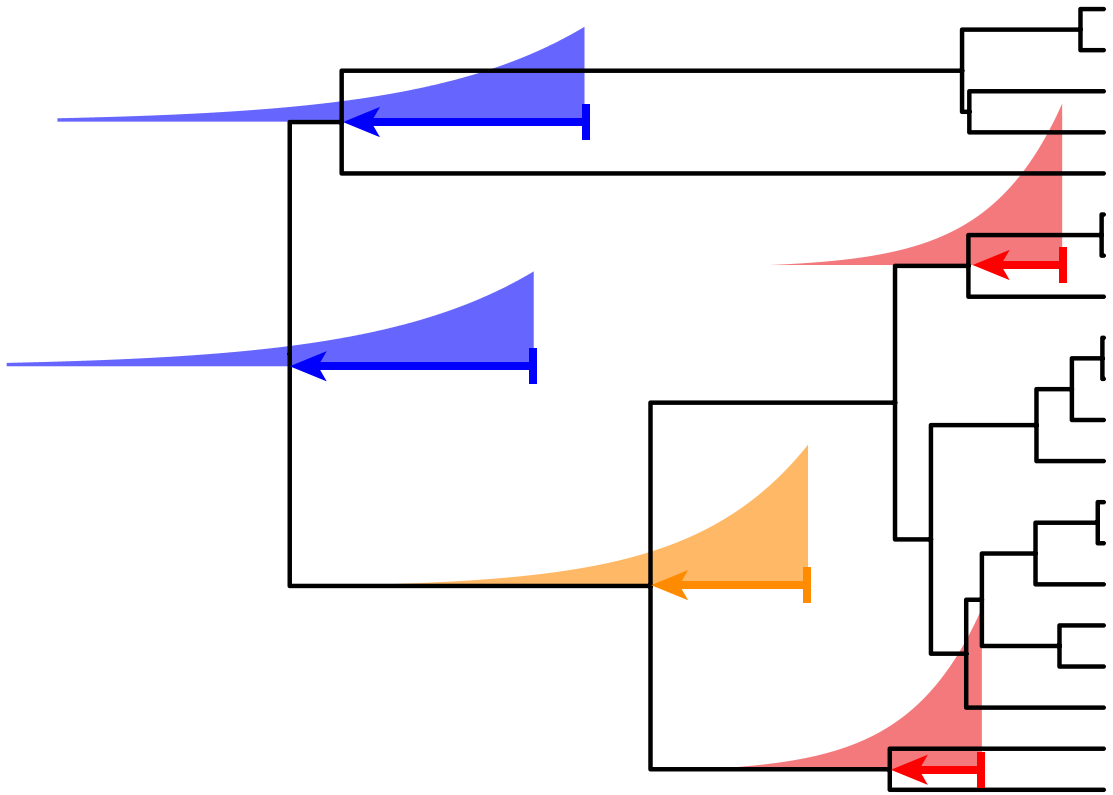
Priors on Multiple Calibrations

Our knowledge of the fossil and rock records indicate that there is variation in the precision of geological data as minimum age constraints



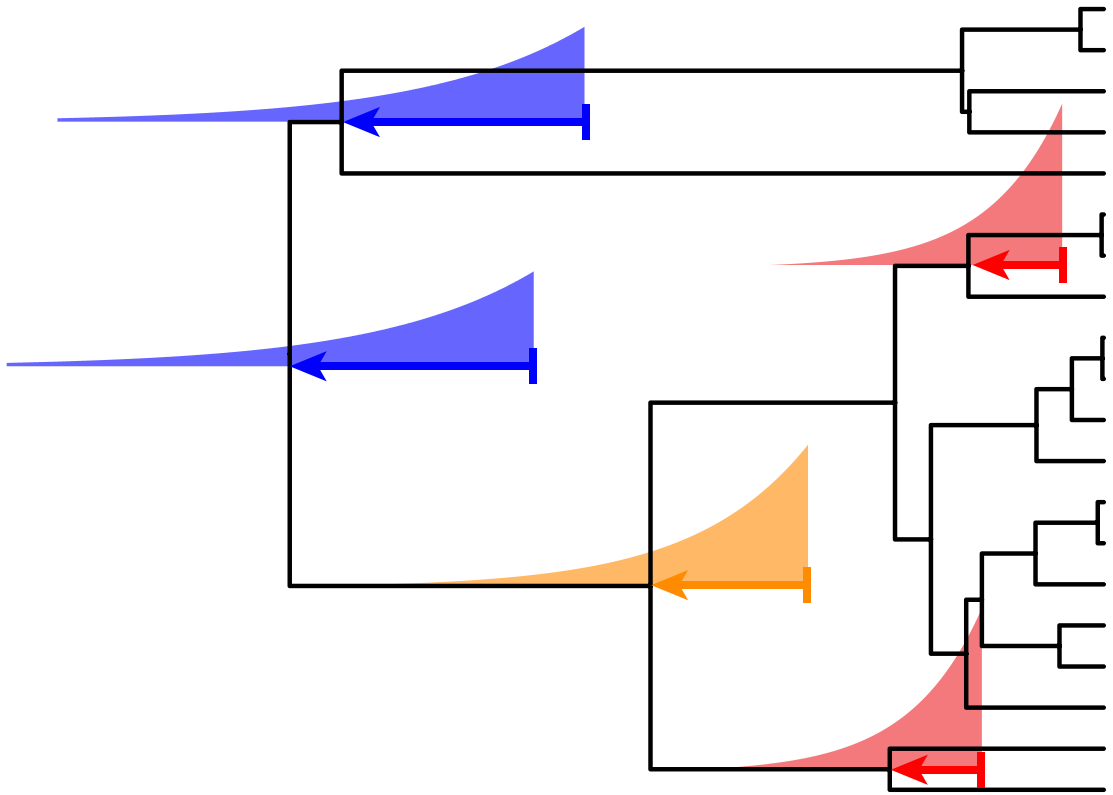
Priors on Multiple Calibrations

Uncertainty in the time difference can be better captured by **vague** prior densities

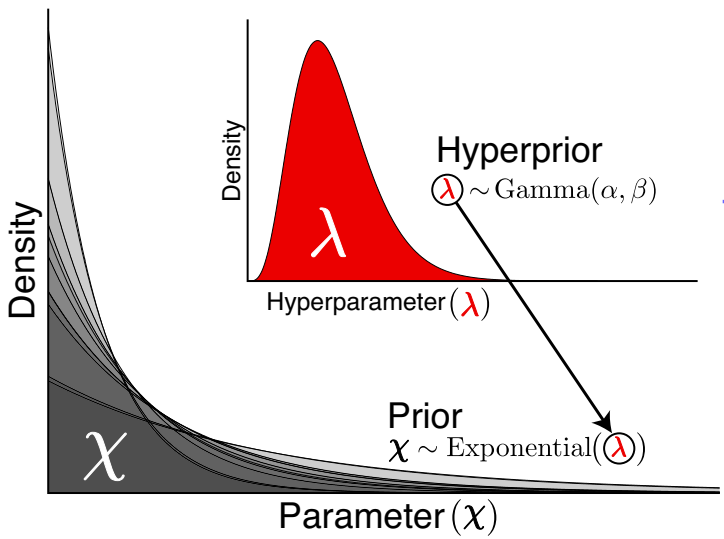


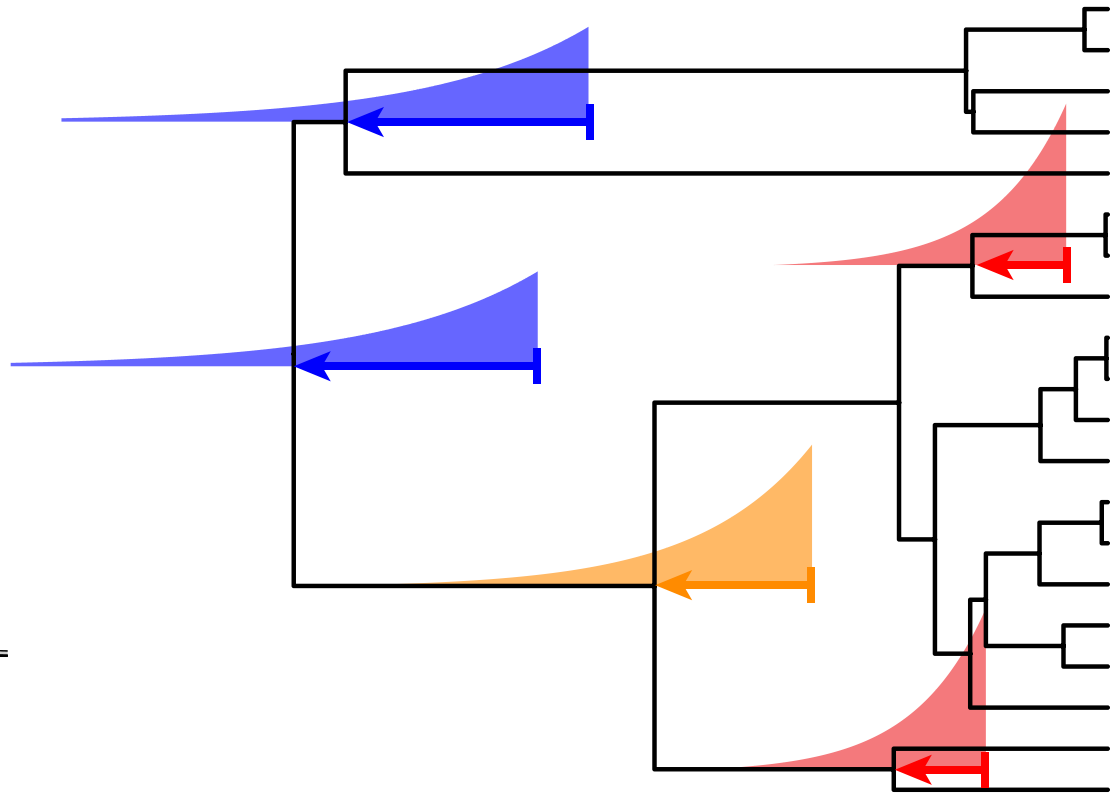
Priors on Multiple Calibrations

Specifying appropriate prior densities for a range of minimum age constraints is a challenge for most molecular biologists



Priors on Multiple Calibrations



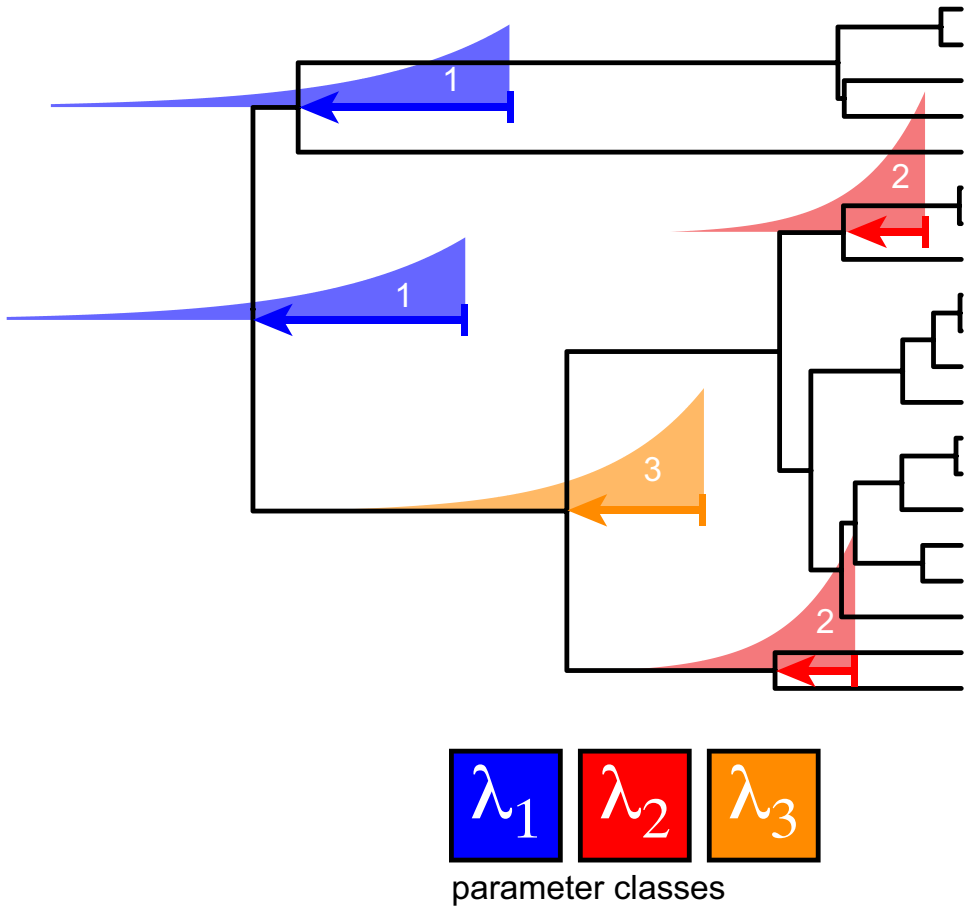


Hyperprior on Calibrated Nodes

Dirichlet process prior on rate-parameters of exponential prior densities on multiple calibrated nodes

Sample the time from the MRCA to the fossil from a mixture of different exponential distributions

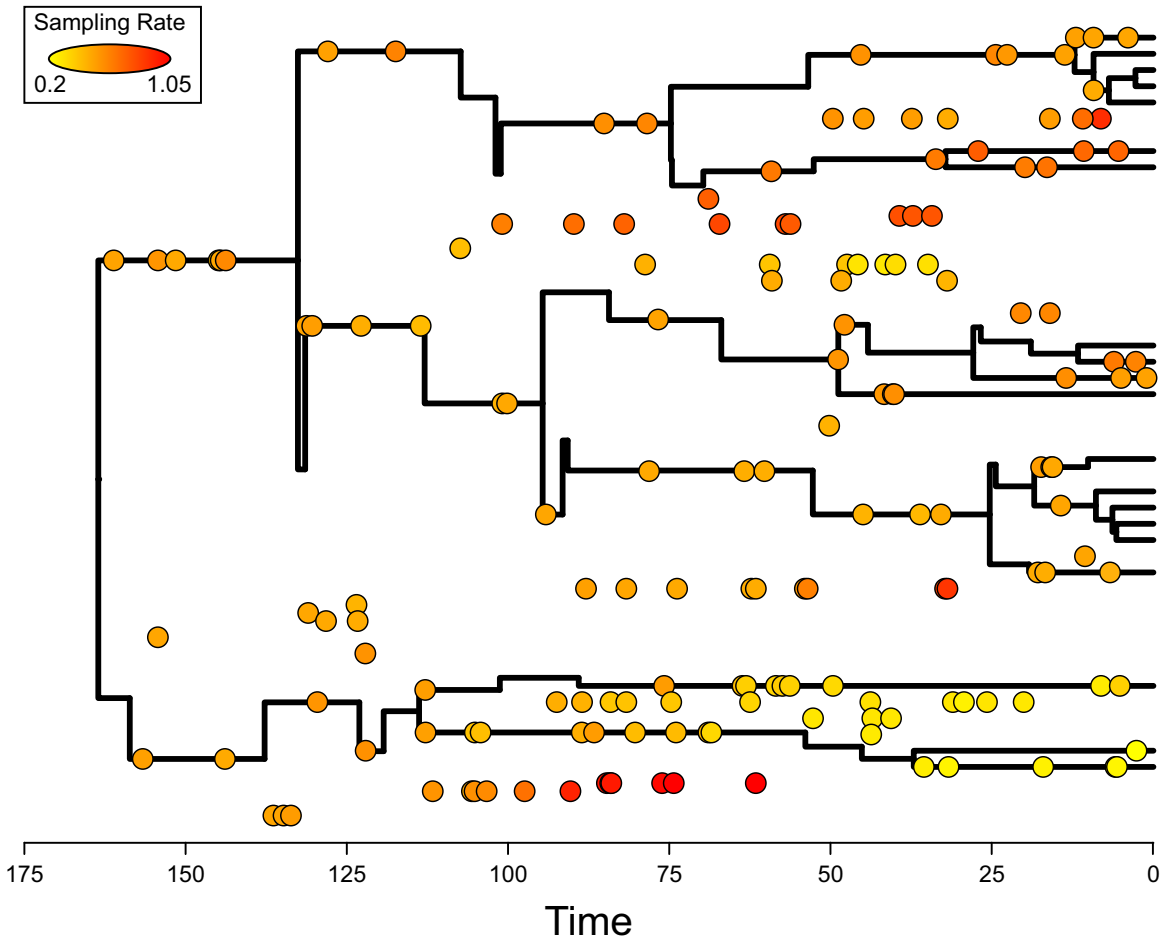
Account for uncertainty in values of λ



(Heath, *Syst. Biol.* 2012)

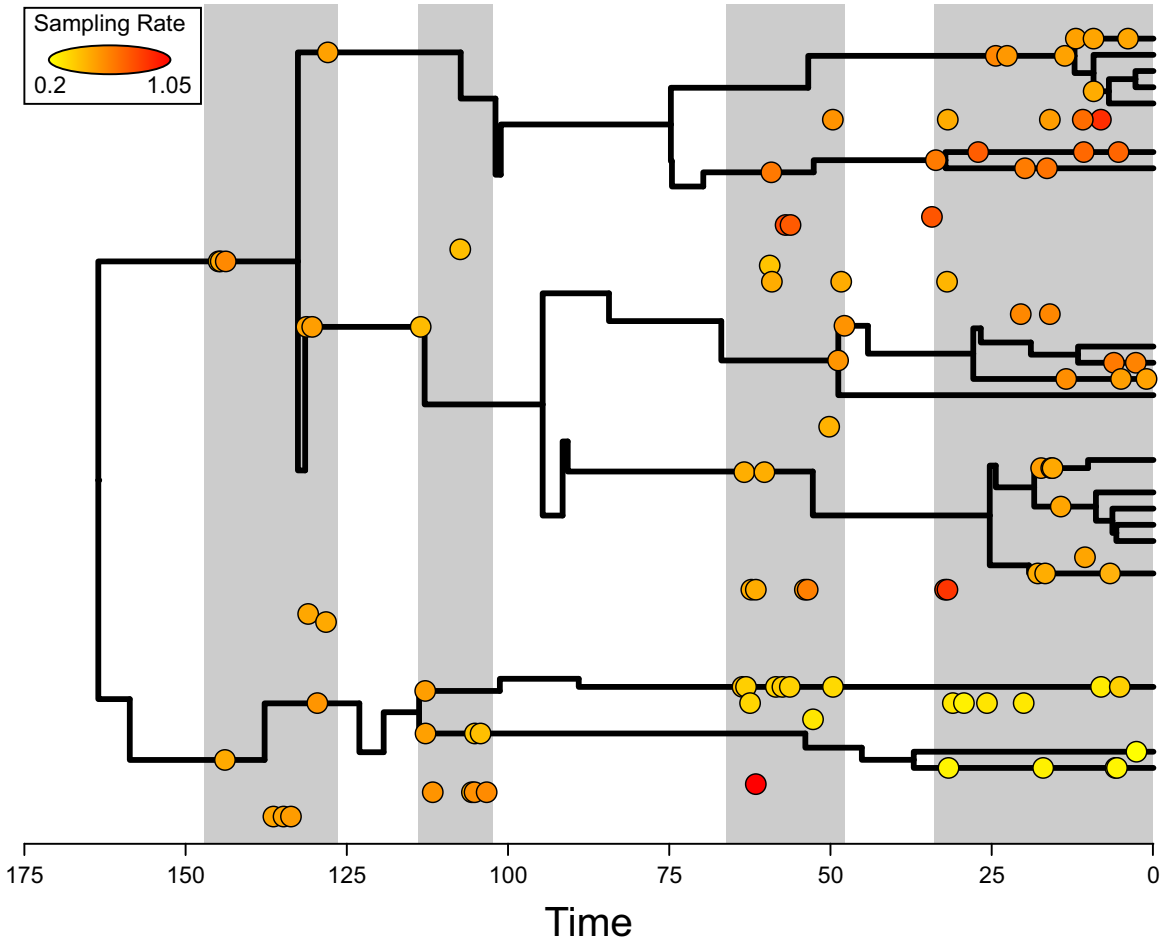
Complex Models of Macroevolution

Modeling branching patterns AND fossilization, preservation, and recovery for use as priors for divergence time estimation



Complex Models of Macroevolution

Incorporate more information from the fossil and rock records and construct better and more realistic tree priors



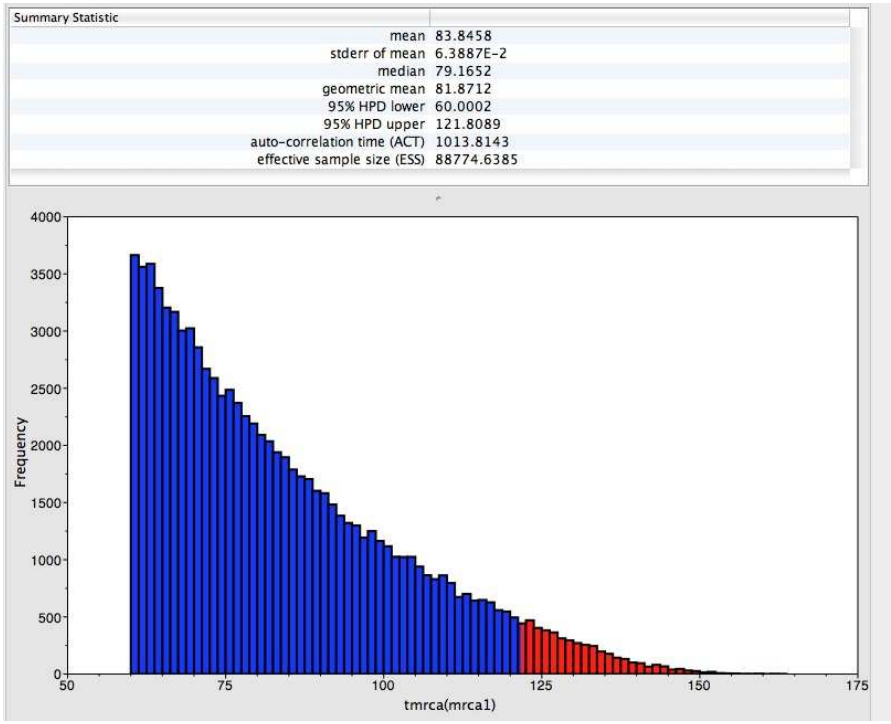
Criticism of Relaxed Clock Methods

- Dependent on and sensitive to fossil calibrations – fossil age estimates and node assignment are not without error
- Models are not biologically realistic
- Different methods/models can produce very different estimates of the same divergence times
- Priors are too informative
- Studies comparing methods have produced conflicting and unclear results

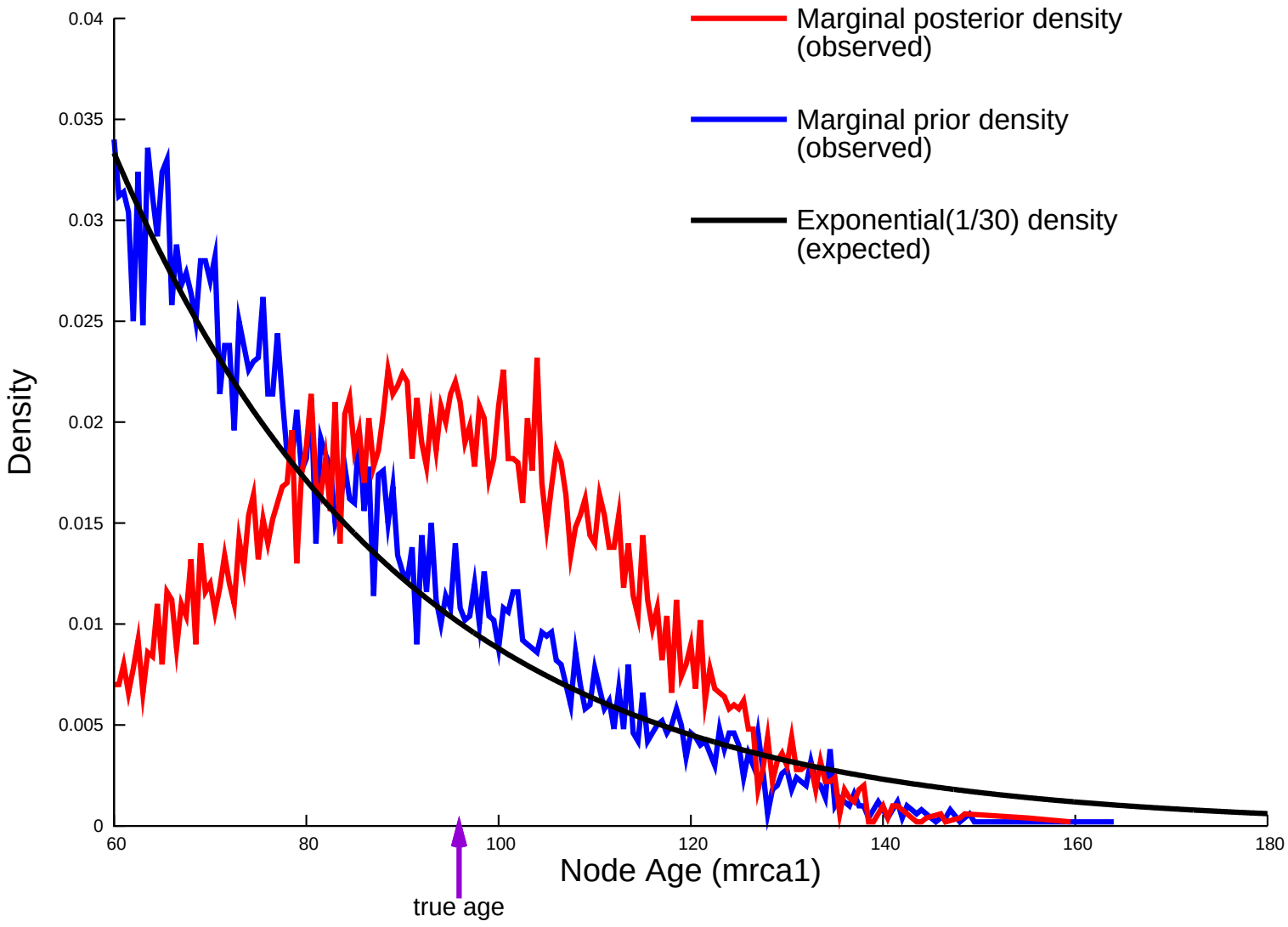
MCMC Under the Prior

It is critical for any Bayesian analysis to sample under the prior

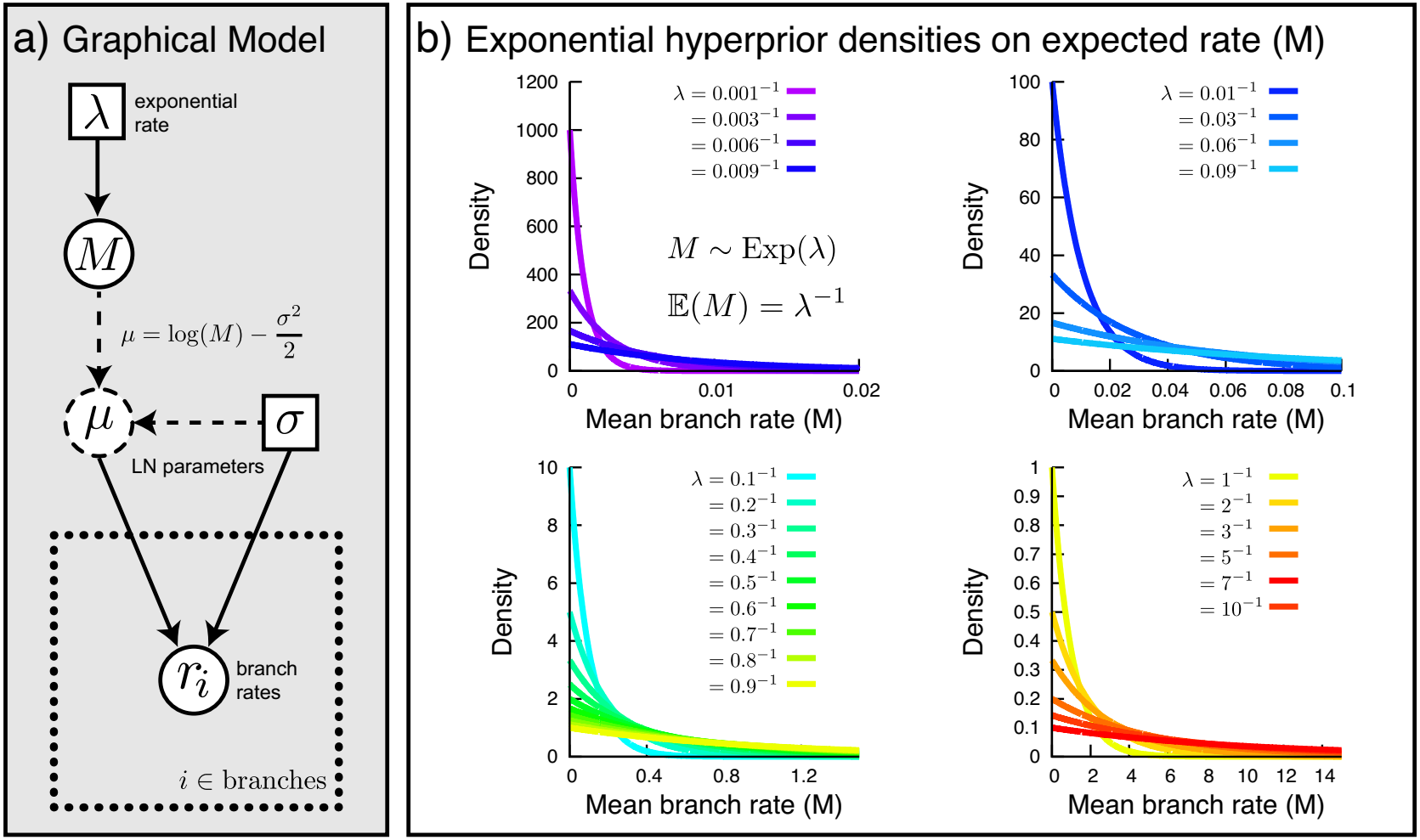
Allows you to assess your prior specification and examine prior sensitivity



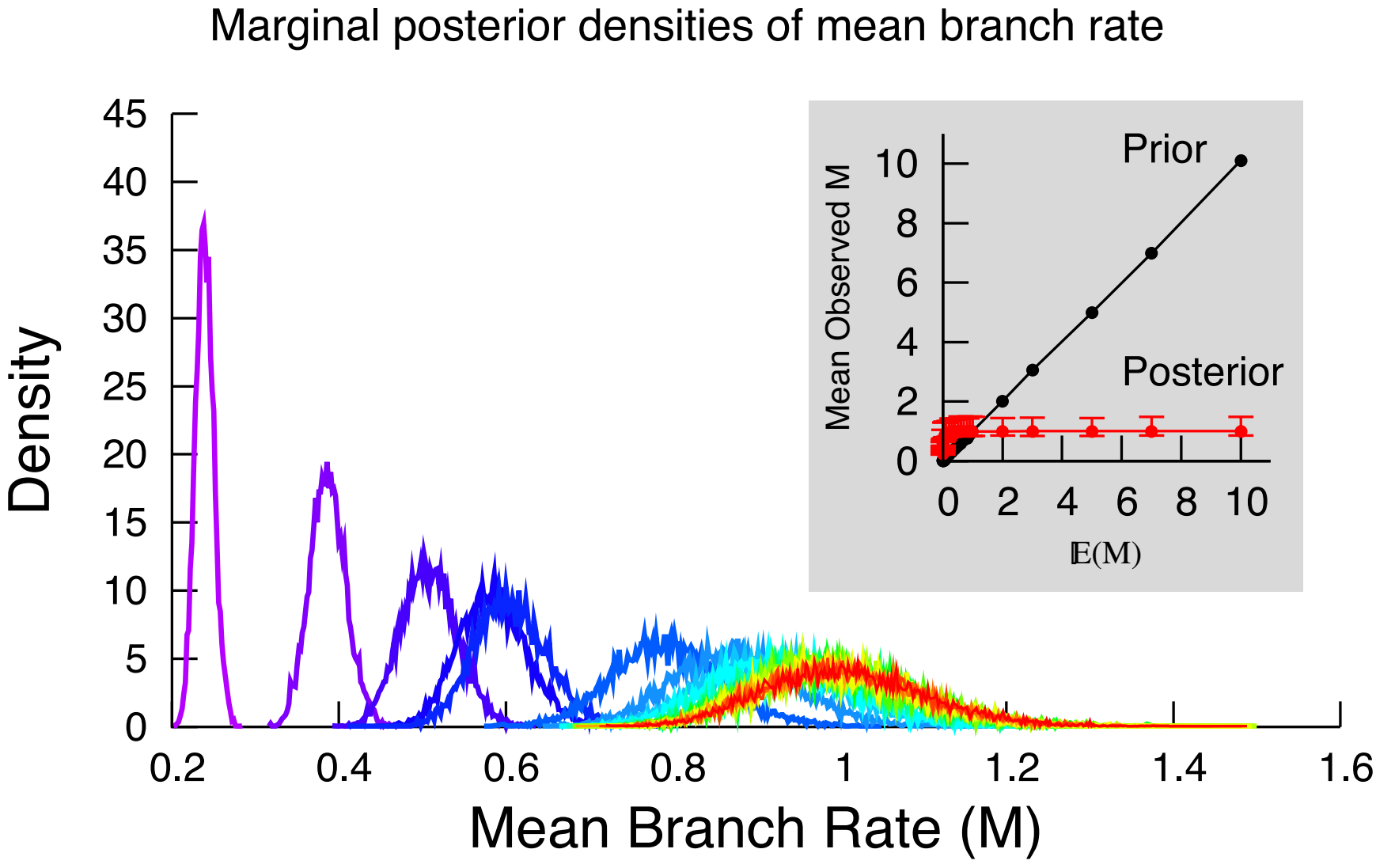
MCMC Under the Prior



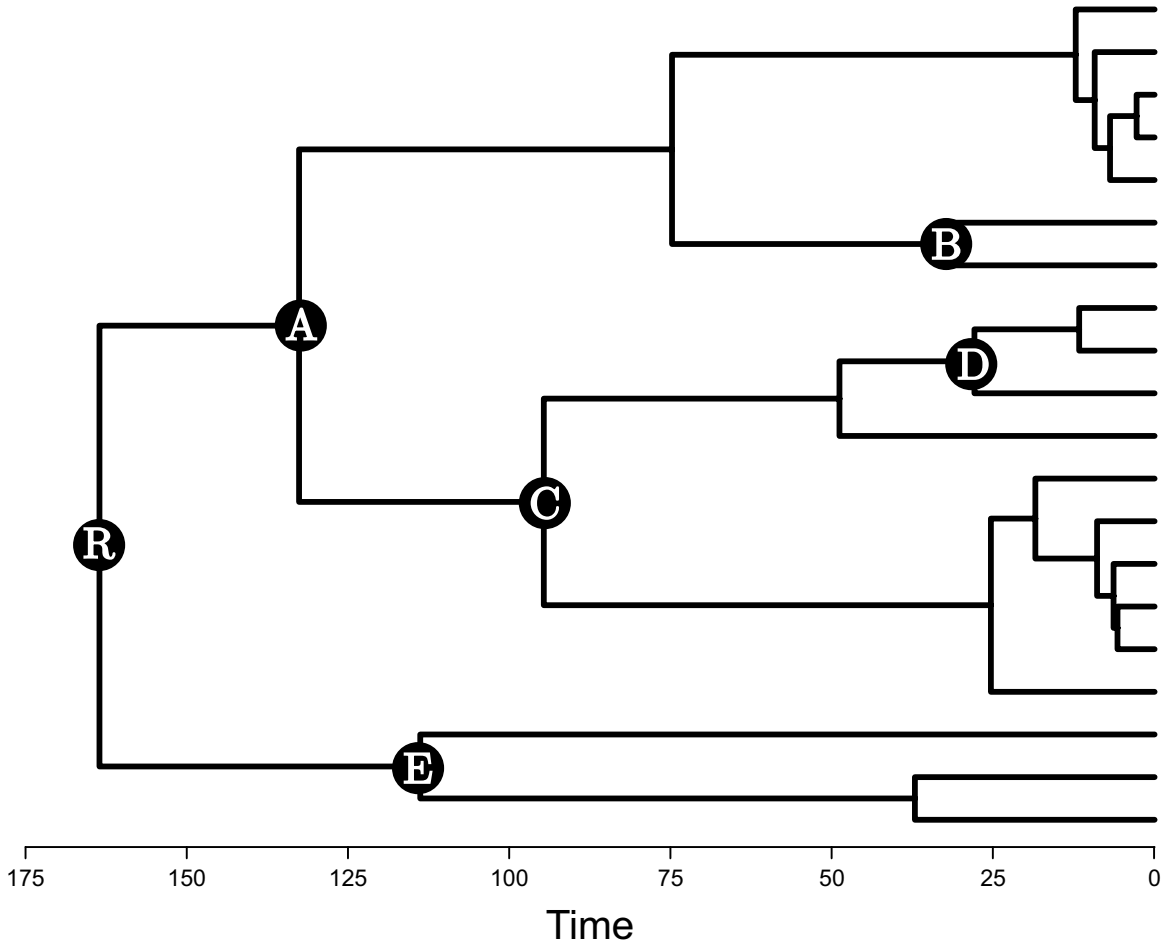
Sensitivity to the Prior



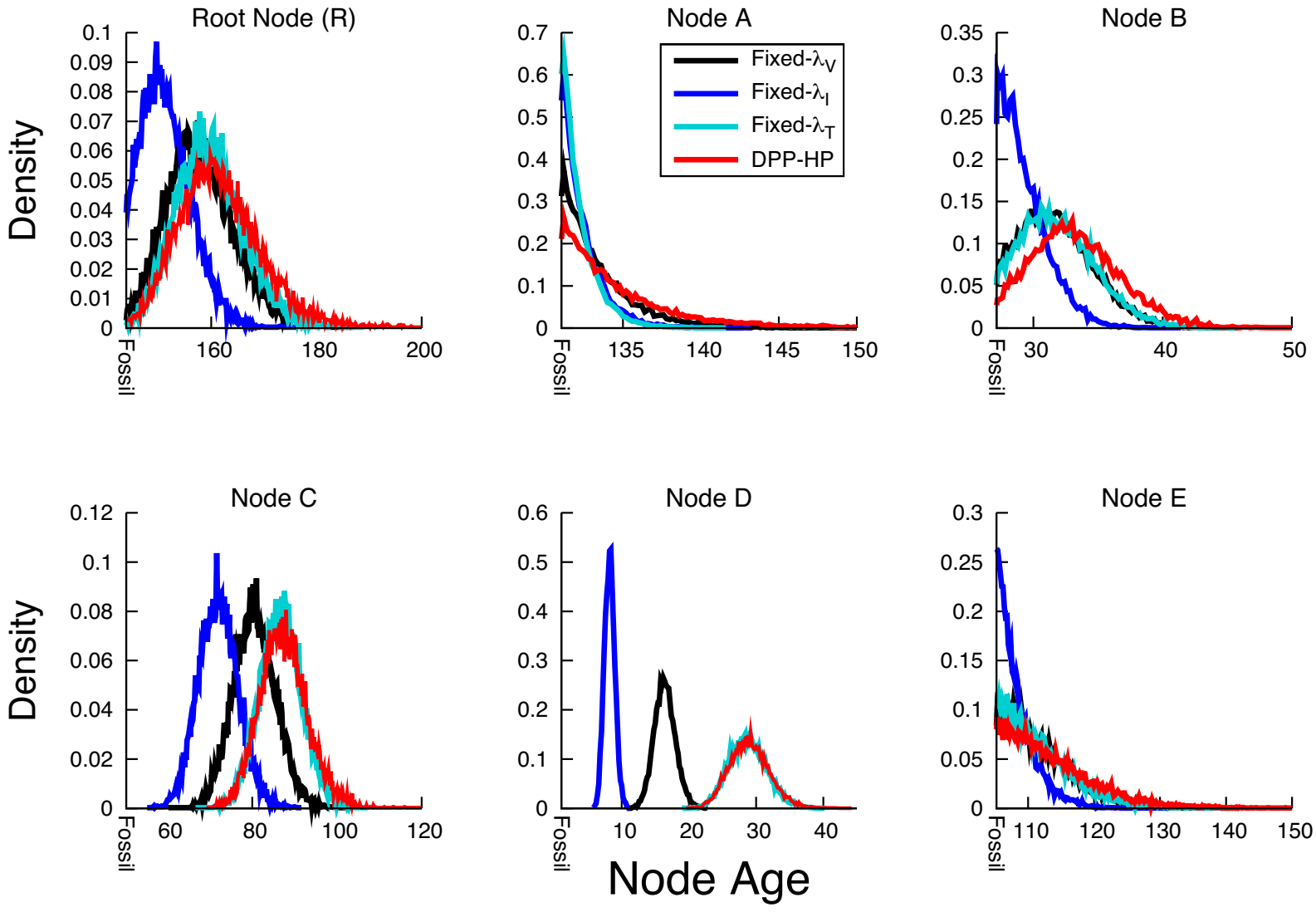
Sensitivity to the Prior



Calibrated Node Age Estimates



Sensitivity to the Calibration Prior

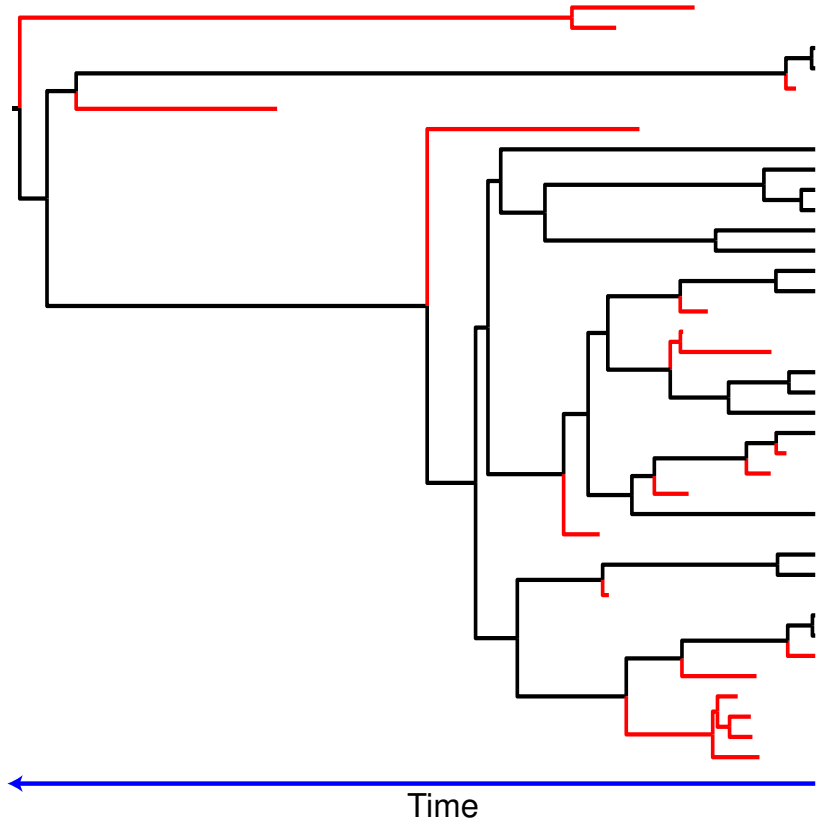


Fossil Tip Dating

Ideally, we would like to include all of the available data

Account for uncertainty in the placement of fossil lineages

Keep all fossil data, not just the oldest descendant for a given node

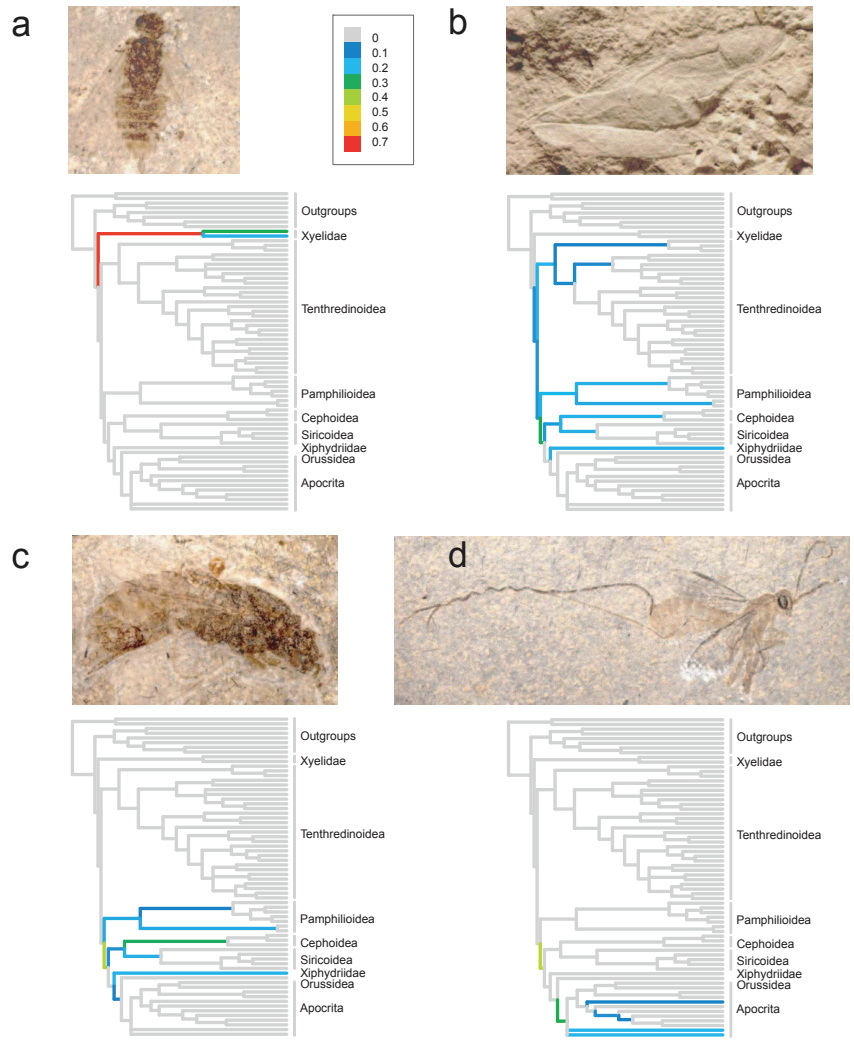


Fossil Tip Dating

Fredrik Ronquist and his colleagues implemented tip dating in MrBayes

Early radiation of Hymenoptera

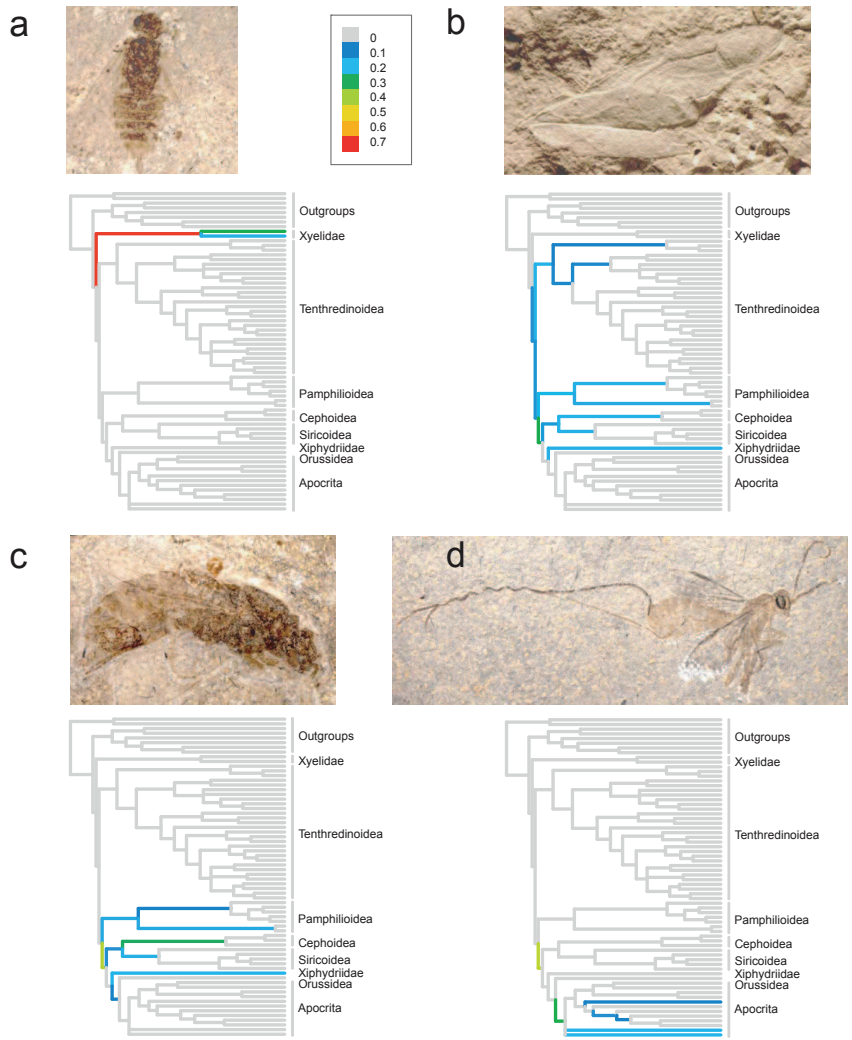
- 66 extant taxa
- 45 fossil taxa
- 7 genes, ~ 5kB (extant taxa only)
- 343 morphological characters (12% complete for fossils)



(Ronquist, Klopfstein, et al. *Syst. Biol.* 2012. doi: 10.1093/sysbio/sys058)

Fossil Tip Dating

- Hymenoptera fossils are mostly poorly-preserved impression fossils, difficult to place phylogenetically
- With node dating, their set of 45 fossils are reduced to 9 calibration points
- They developed a, presumably, vague uniform prior on node times



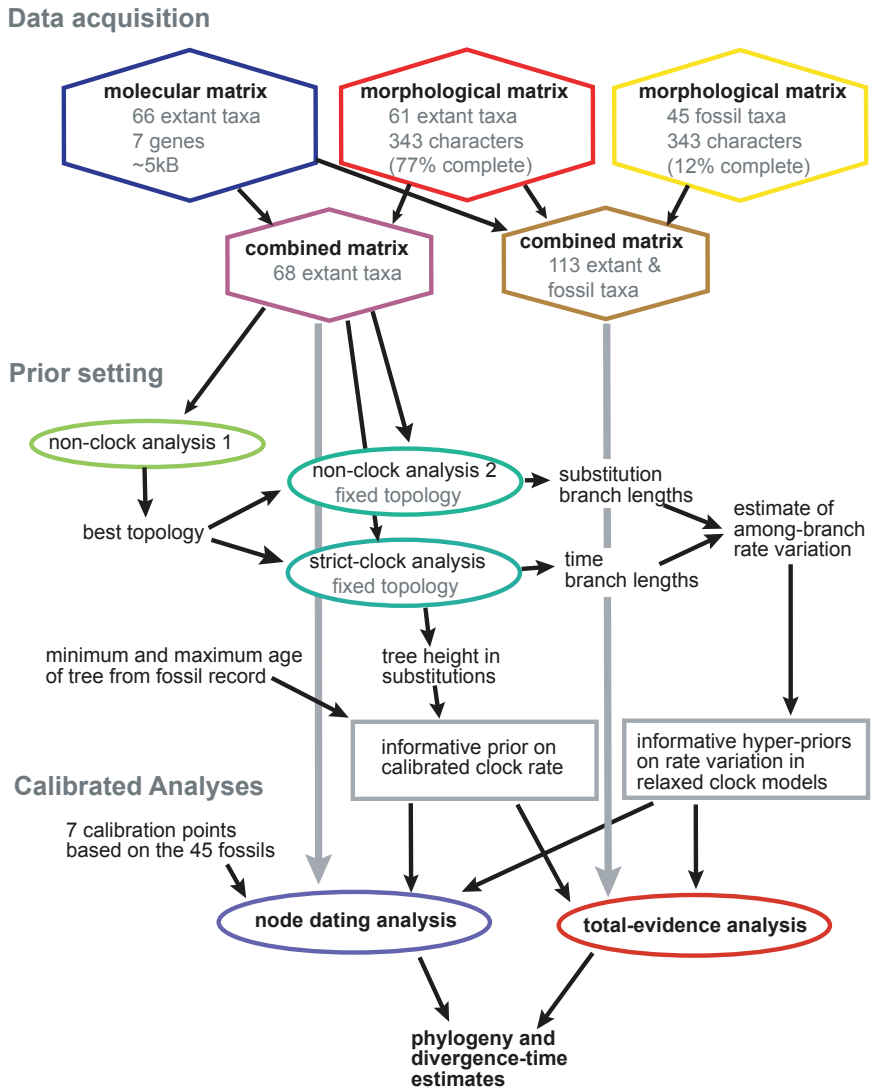
(Ronquist, Klopfstein, et al. *Syst. Biol.* 2012. doi: 10.1093/sysbio/sys058)

Fossil Tip Dating

Thorough analysis is necessary for this kind of dataset

Ronquist et al. used Bayes factors to choose a relaxed clock model (this is rarely done, but really important)

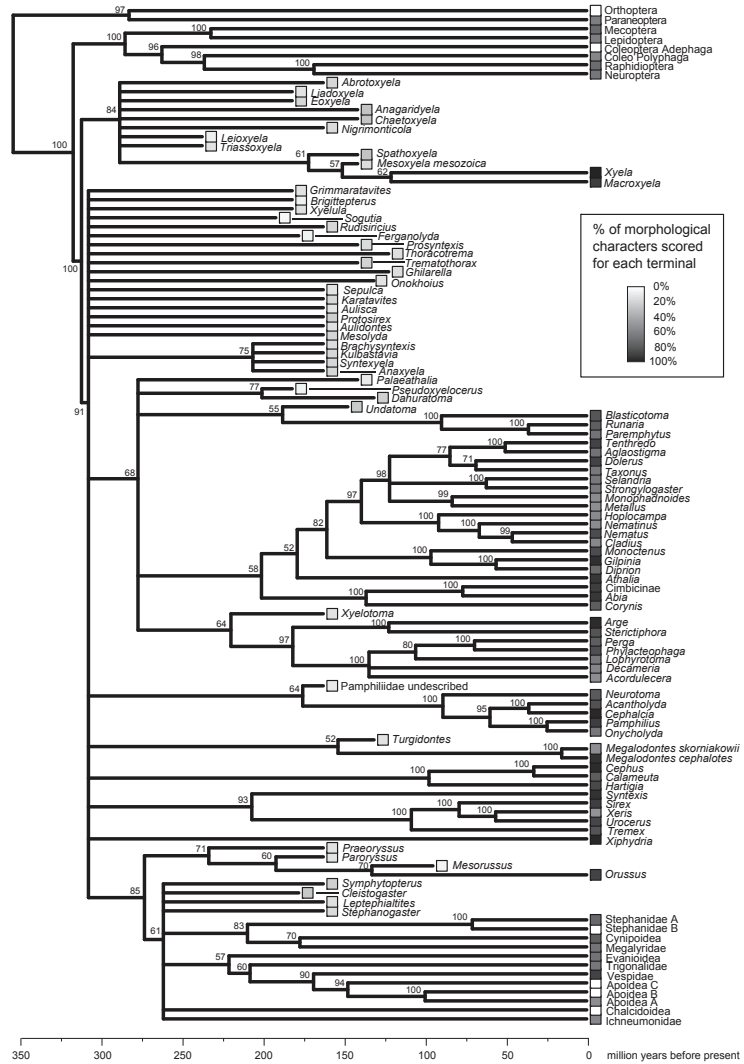
Compared node dating and tip dating



(Ronquist, Klopfstein, et al. *Syst. Biol.* 2012. doi: 10.1093/sysbio/sys058)

Fossil Tip Dating

- Resulted in a fairly unresolved phylogeny, but fossils significantly contribute to estimates of node ages
- Posteriors on node times are less sensitive to priors compared with node dating
- Higher precision for divergence time estimates

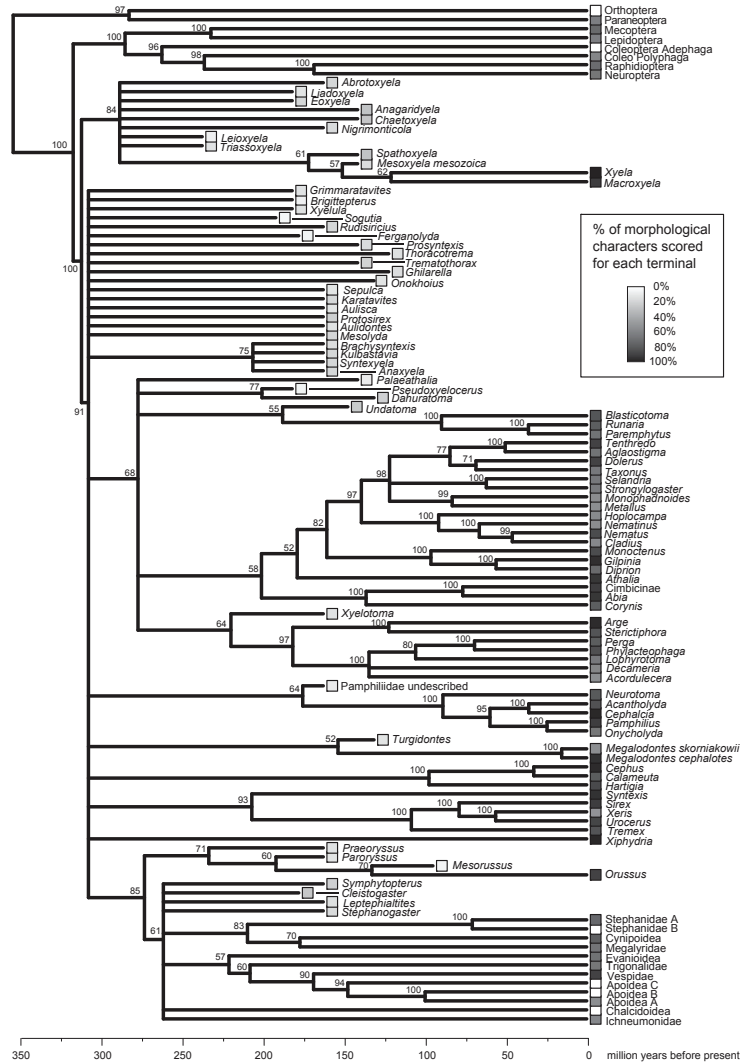


(Ronquist, Klopfstein, et al. *Syst. Biol.* 2012. doi: 10.1093/sysbio/sys058)

Fossil Tip Dating

The Hymenoptera crown group dates back to the Carboniferous, approximately 309 Ma (95% interval: 291–347 Ma)

And diversified into major extant lineages much earlier than previously thought, well before the Triassic



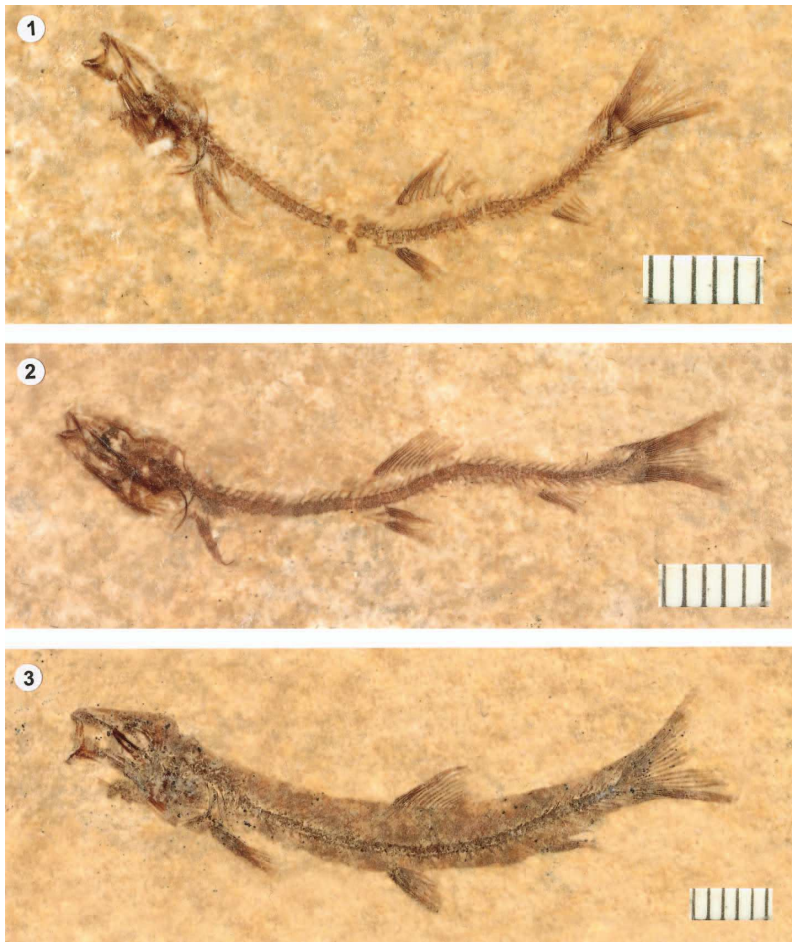
(Ronquist, Klopfstein, et al. *Syst. Biol.* 2012. doi: 10.1093/sysbio/sys058)

Fossil Tip Dating

In groups with rich fossil records, tip dating is an ideal approach

Allows for dating trees with more of the available fossils

Investigate questions (i.e. historical biogeography, character evolution) with extinct lineages



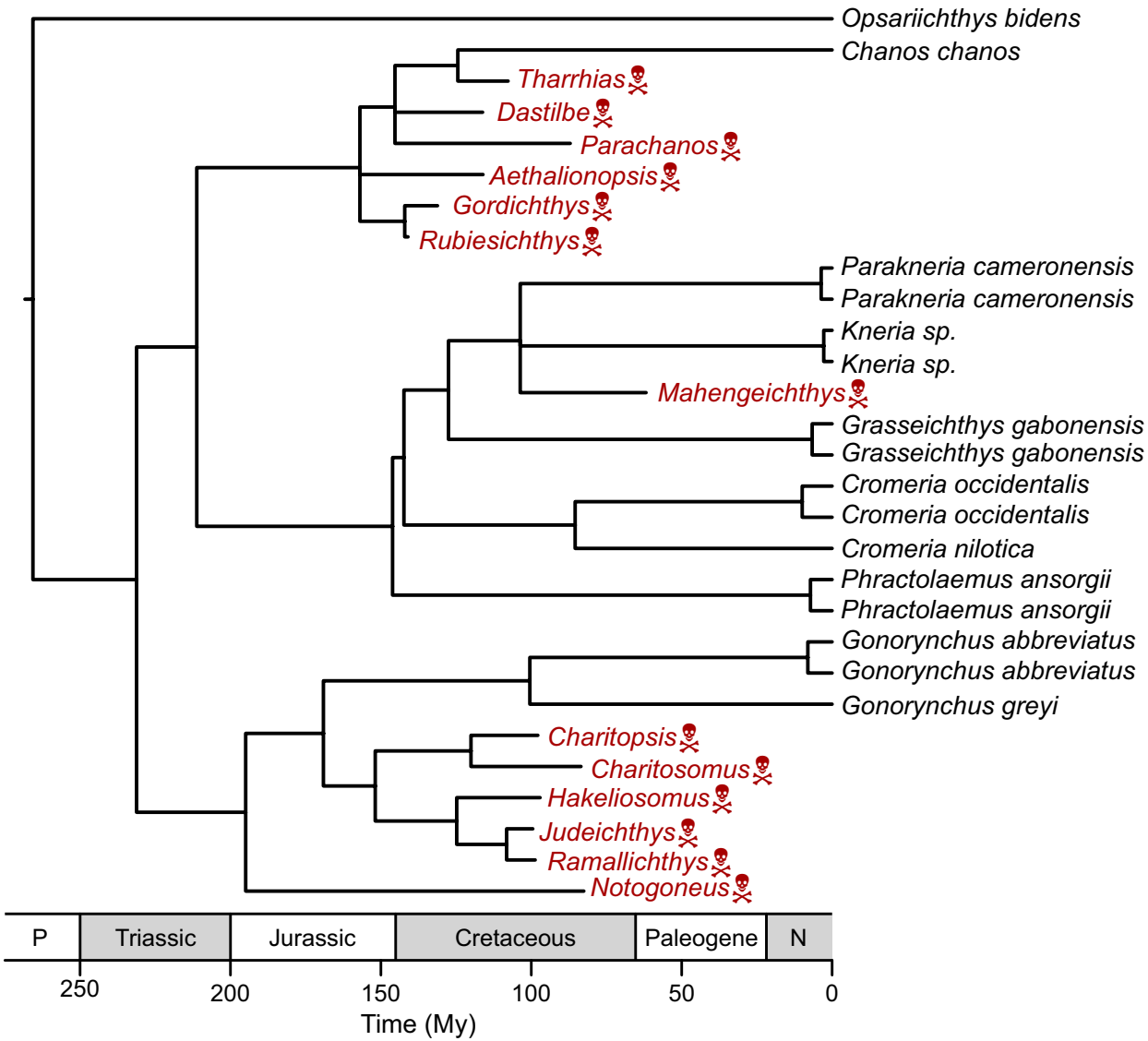
Notogoneus osculus – early growth series

illustrating the ontogeny of the scale covering

(Grande & Grande *J. Paleont.* 2008)

Fossil Tip Dating

Gonorynchiformes



Preliminary work with: Matt Davis (Field Museum) and Charles Linkem (KU/UW)

Fossil Tip Dating

Fossil tip-dating methods are available in MrBayes and BEAST, though our understanding of how well these methods work is still incomplete