

BAYESIAN DIVERGENCE TIME ESTIMATION

Tracy Heath

Department of Integrative Biology, University of California, Berkeley

2013 Applied Phylogenetics Workshop
Bodega Bay, CA USA

OUTLINE

Overview of divergence time estimation

- Relaxed clock models – accounting for variation in substitution rates among lineages
- Tree priors and fossil calibration

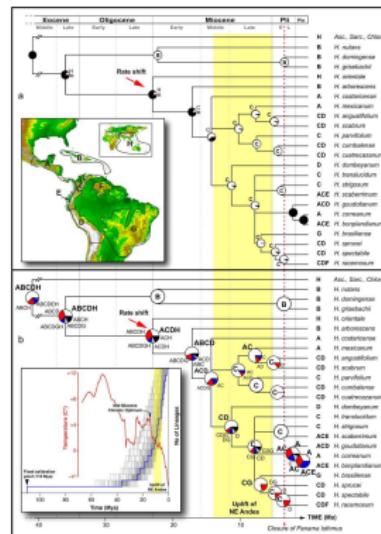
break

BEAST Tutorial:

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

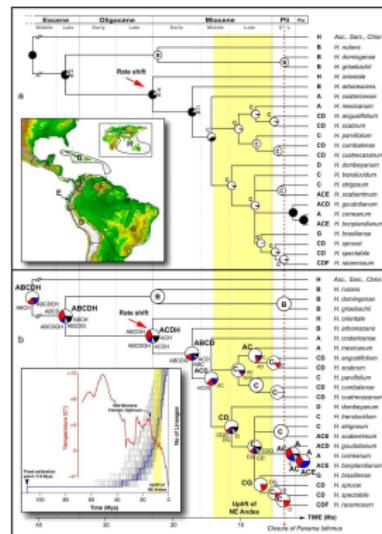
A TIME-SCALE FOR EVOLUTION

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



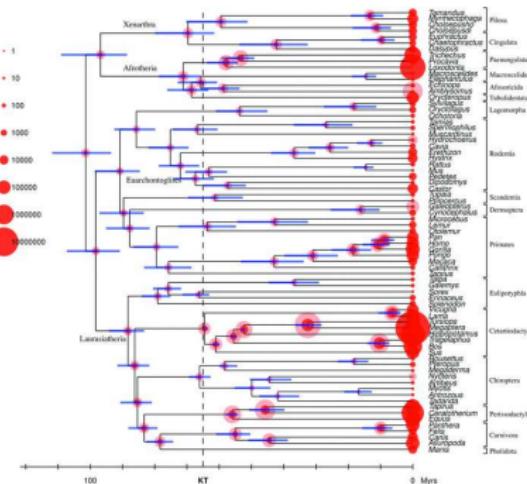
A TIME-SCALE FOR EVOLUTION

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



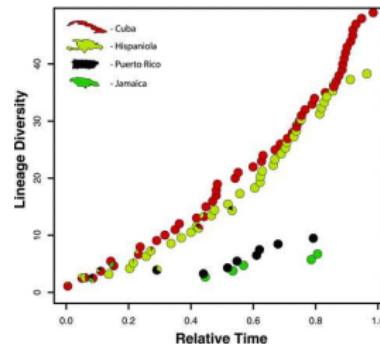
A TIME-SCALE FOR EVOLUTION

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



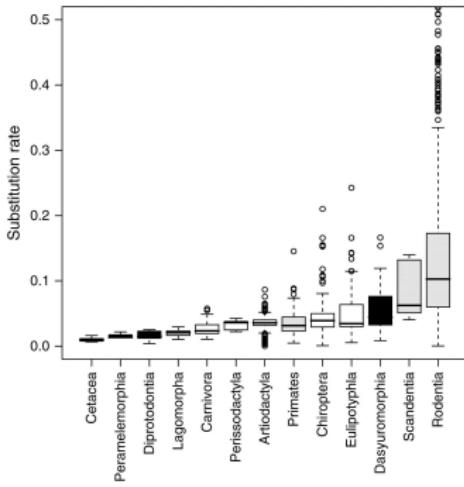
A TIME-SCALE FOR EVOLUTION

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



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



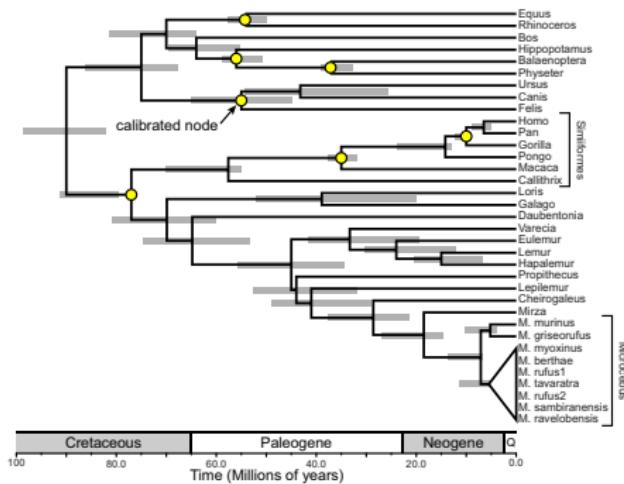
DIVERGENCE TIME ESTIMATION

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

Model how rates are distributed across the tree

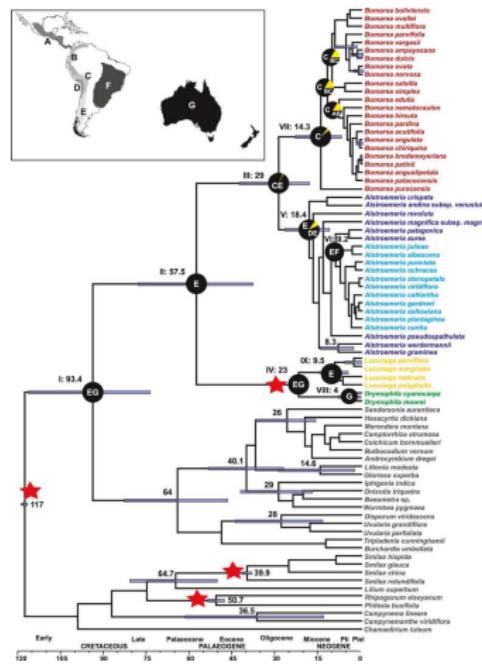
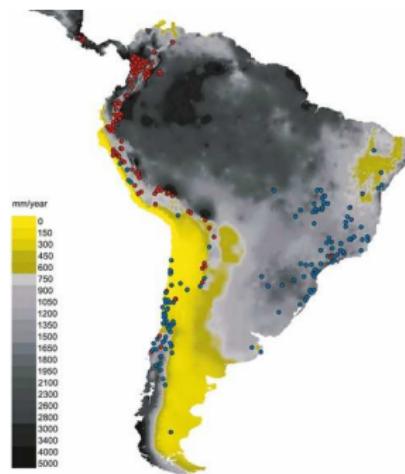
Describe the distribution of speciation events over time

External calibration information for estimates of absolute node times



UNDERSTANDING HISTORICAL BIOGEOGRAPHY

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



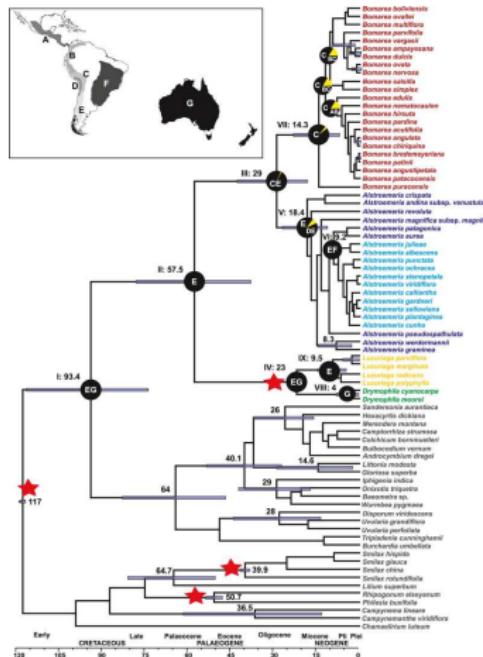
DIVERGENCE TIME ESTIMATION

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



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



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

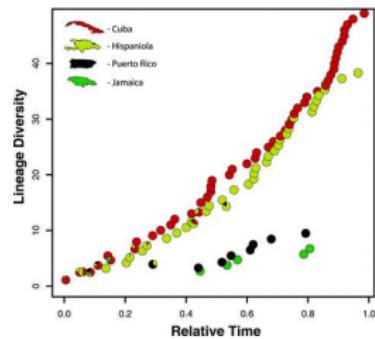
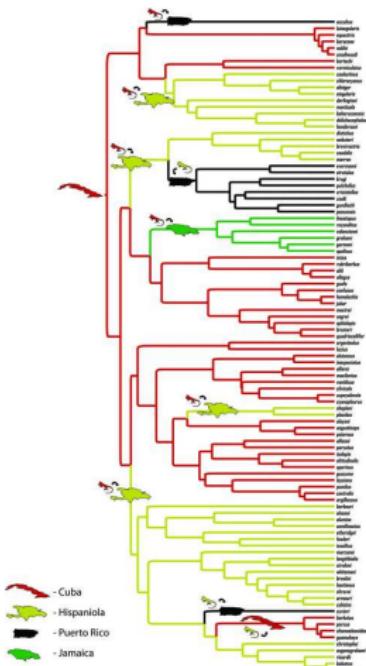
Estimation of relative divergence times



Anolis hendersoni (Image courtesy of L. Mahler)

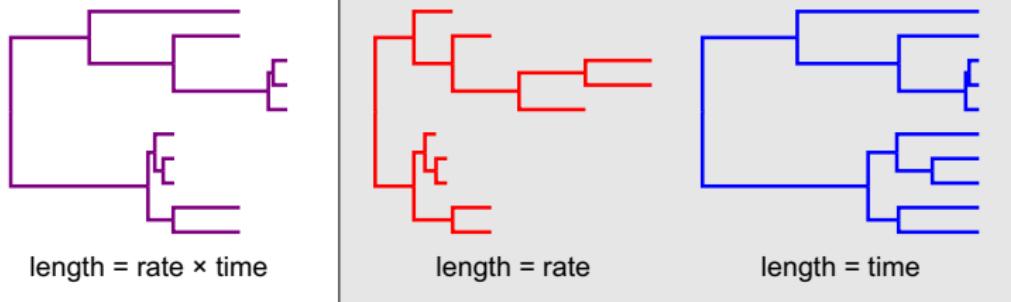
RELATIVE TIMES AND DIVERSIFICATION

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



DIVERGENCE TIME ESTIMATION

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

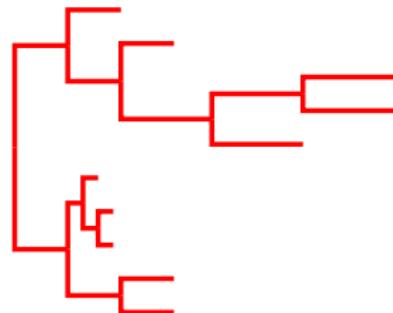


Methods for dating species divergences estimate the **substitution rate** and **time** separately

SUBSTITUTION RATE

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

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

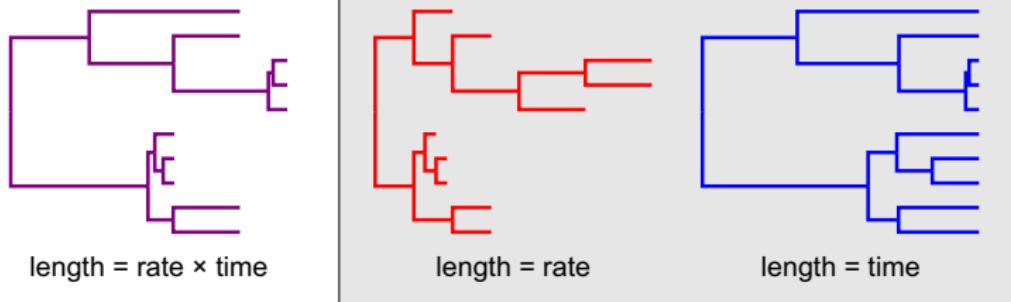


length = subst. rate

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**

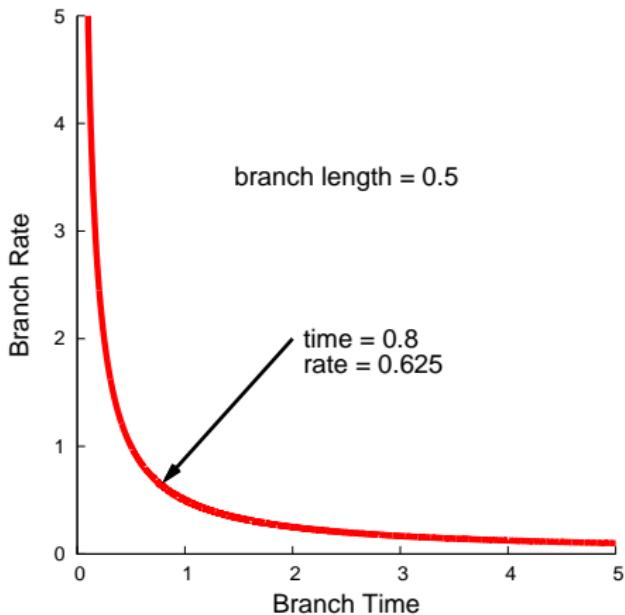


Methods for dating species divergences estimate the **substitution rate** and **time** separately

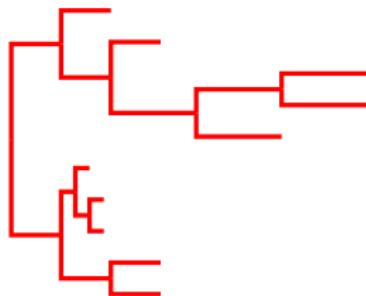
RATES AND TIMES

The sequence data provide information about branch length

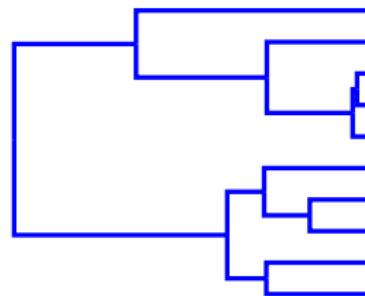
for any possible rate, there's a time that fits the branch length perfectly



BAYESIAN DIVERGENCE TIME ESTIMATION



length = rate



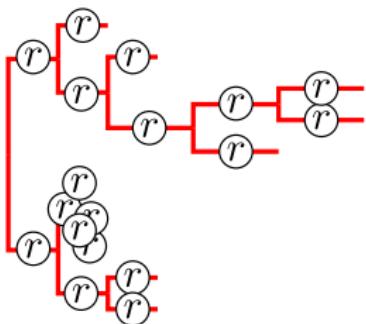
length = time

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

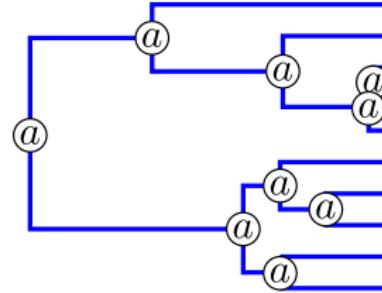
$$\mathcal{A} = (a_1, a_2, a_3, \dots, a_{N-1})$$

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

BAYESIAN DIVERGENCE TIME ESTIMATION



length = rate



length = time

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

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

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

BAYESIAN DIVERGENCE TIME ESTIMATION

Posterior probability

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

\mathcal{R} Vector of rates on branches

\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

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

$f(D | \mathcal{R}, \mathcal{A}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s)$ Likelihood

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

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

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:

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

BAYESIAN DIVERGENCE TIME ESTIMATION

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

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

$$\frac{f(D | \mathcal{R}, \mathcal{A}, \theta_s) f(\mathcal{R} | \theta_{\mathcal{R}}) f(\mathcal{A} | \theta_{\mathcal{A}}) f(\theta_{\mathcal{R}}) f(\theta_{\mathcal{A}}) f(\theta_s)}{f(D)}$$

BAYESIAN DIVERGENCE TIME ESTIMATION

Estimating divergence times relies on 2 main elements:

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

MODELING RATE VARIATION

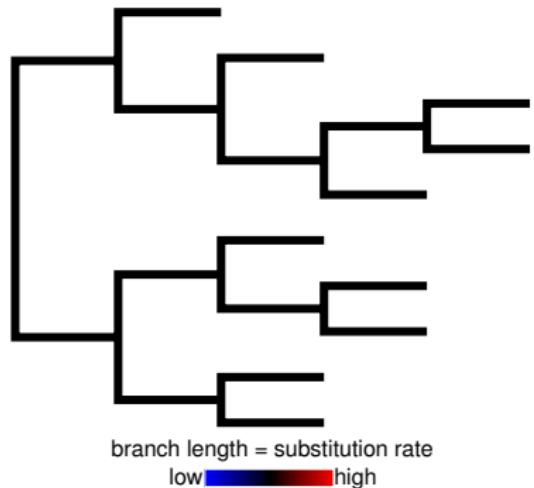
Some models describing lineage-specific substitution rate variation:

- **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)
- **Mixture models on branch rates** (Heath, Holder, Huelsenbeck 2012)

GLOBAL MOLECULAR CLOCK

The substitution rate is constant over time

All lineages share the same rate



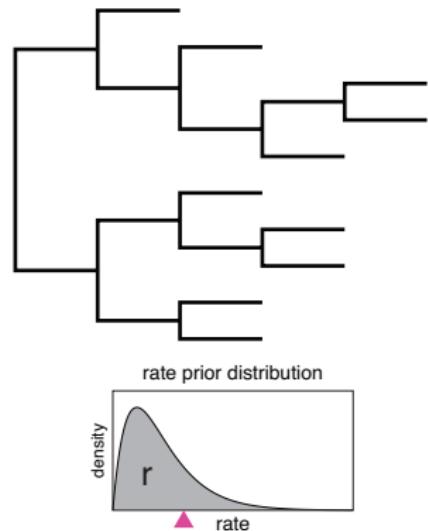
GLOBAL MOLECULAR CLOCK

Assume the clock rate is gamma-distributed

$$\mathcal{R} = (r, r, \dots, r)$$

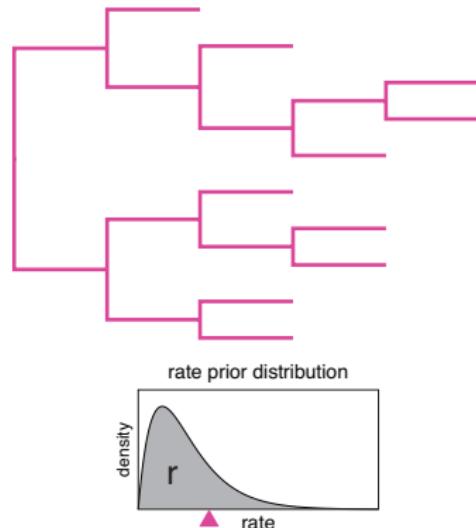
$$r \sim \text{Gamma}(\alpha, \lambda)$$

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



GLOBAL MOLECULAR CLOCK

The sampled rate is applied
to every branch in the tree



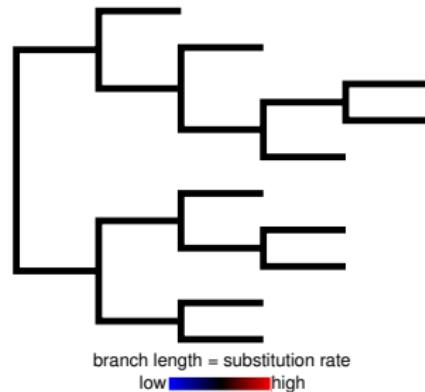
REJECTING THE GLOBAL MOLECULAR CLOCK

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

Mutation rate:

Variation in

- metabolic rate
- generation time
- DNA repair



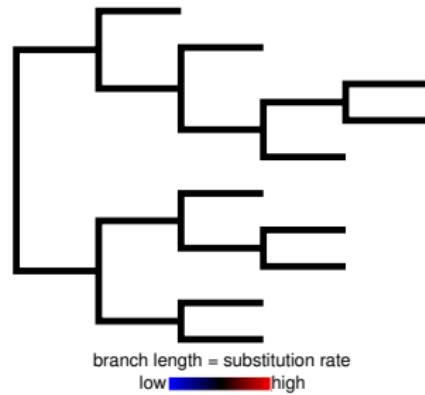
REJECTING THE GLOBAL MOLECULAR CLOCK

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

Fixation rate:

Variability in

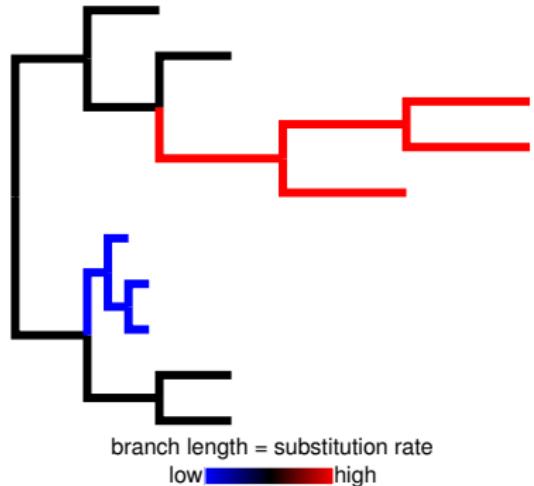
- strength and targets of selection
- population sizes



LOCAL MOLECULAR CLOCKS

Rate shifts occur infrequently over the tree

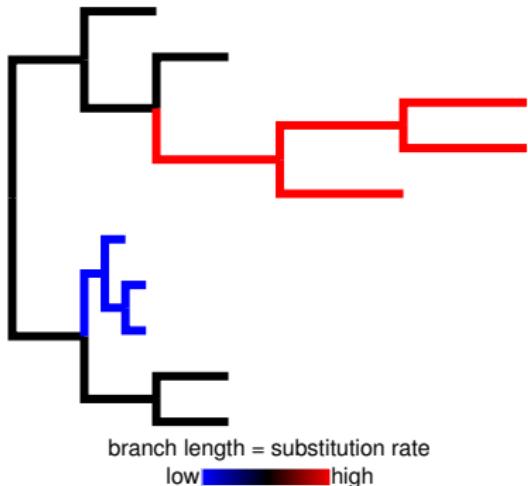
Closely related lineages have equivalent rates (clustered by sub-clades)



LOCAL MOLECULAR CLOCKS

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

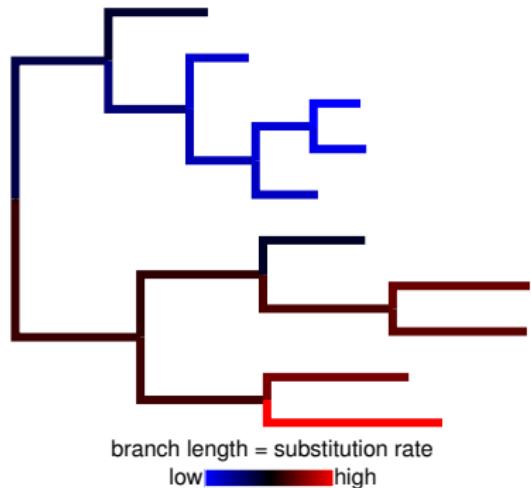
Drummond and Suchard (2010) introduced a Bayesian method that samples over a broad range of possible *random local clocks*



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



AUTOCORRELATED RATES

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

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

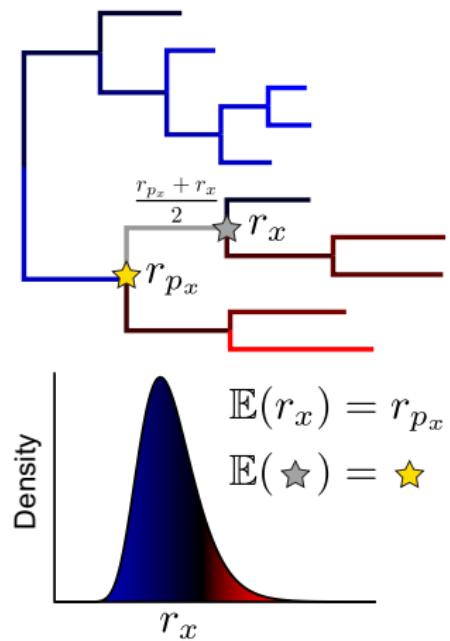
$$\mu = \ln(r_{p_i}) - \frac{\sigma^2}{2}$$

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

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

ψ is the variance parameter

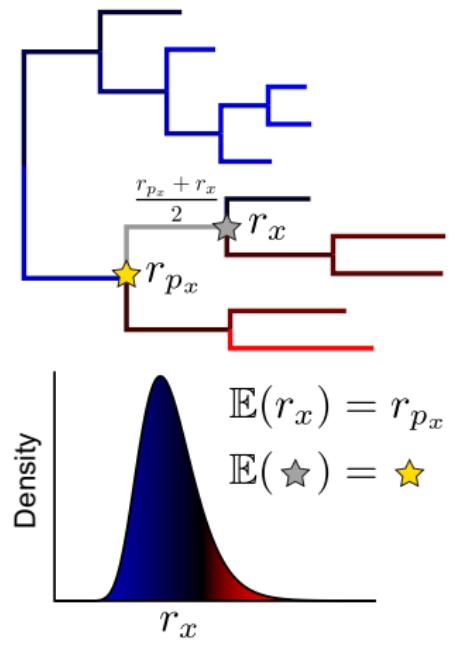
Δt is the difference in time
between the 2 nodes



AUTOCORRELATED RATES

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

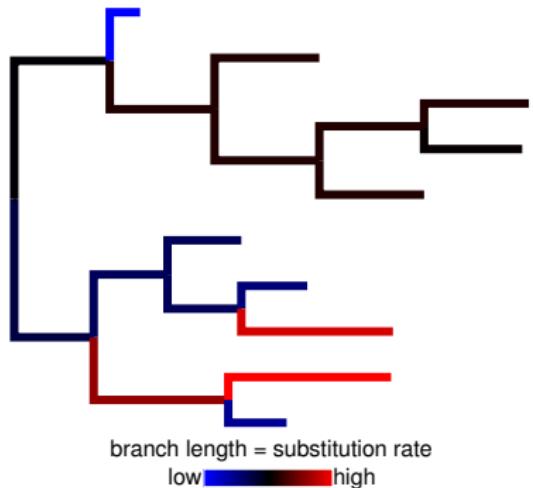
The rate for the branch is equal to the mean of the two subtending nodes



COMPOUND POISSON PROCESS

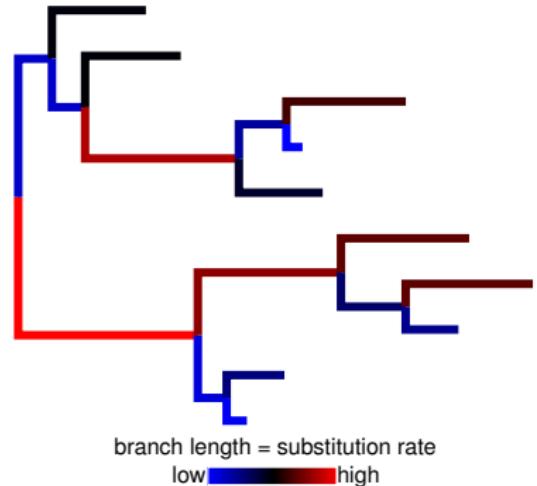
Rate changes occur along lineages according to a point process

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



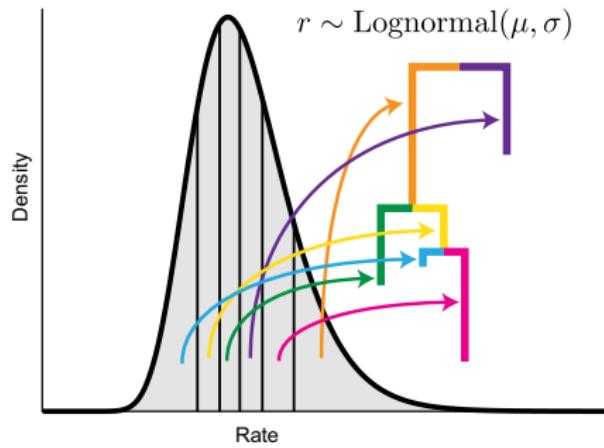
INDEPENDENT/UNCORRELATED RATES

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

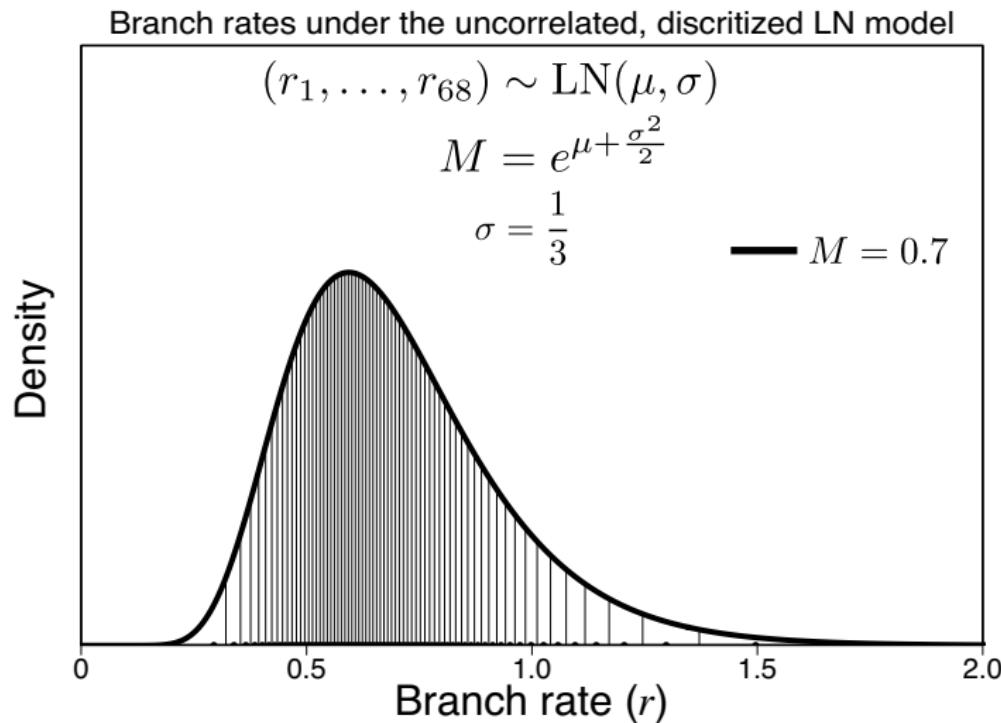


INDEPENDENT/UNCORRELATED RATES

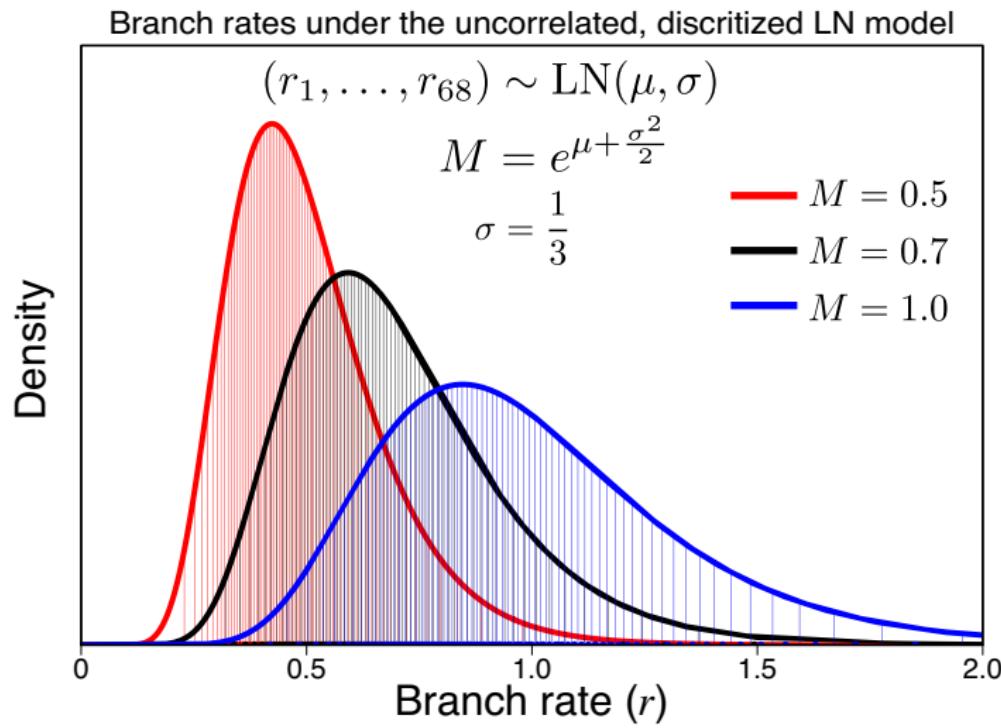
In BEAST, the rates for the branches are drawn from a discretized lognormal distribution



INDEPENDENT/UNCORRELATED RATES



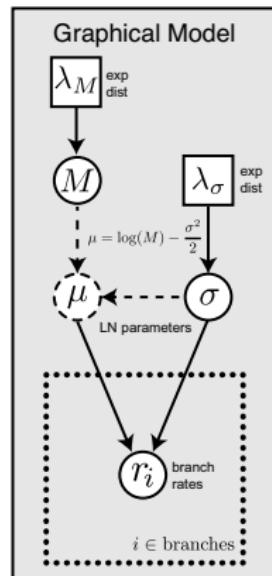
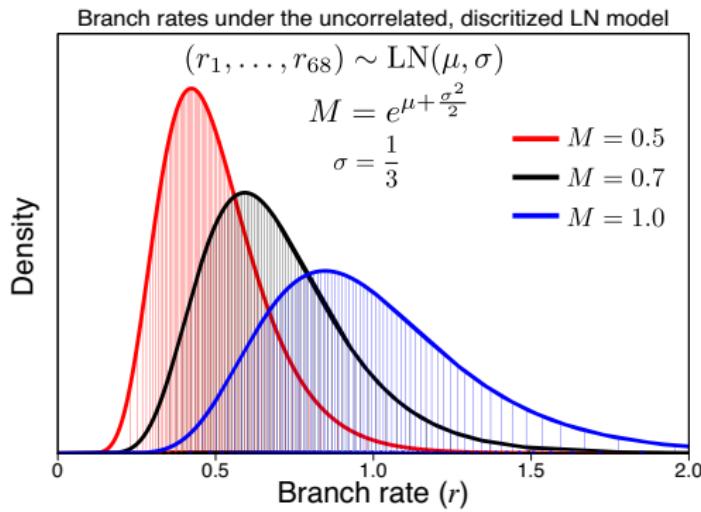
INDEPENDENT/UNCORRELATED RATES



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

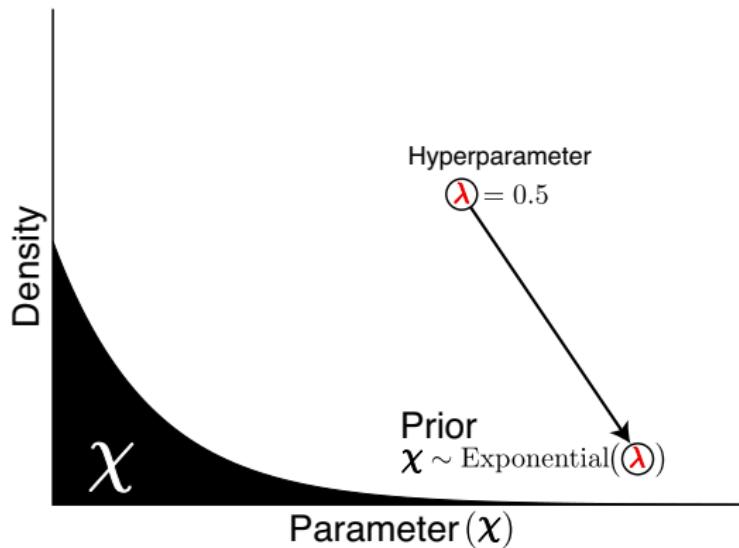


$$\mathbb{E}(M) = \lambda_M^{-1}$$

A HIERARCHICAL BAYESIAN MODEL

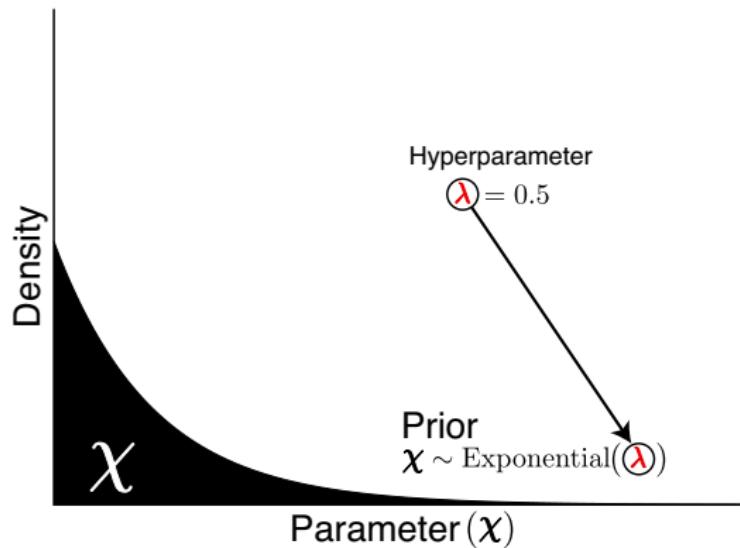
From the bottom up:

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



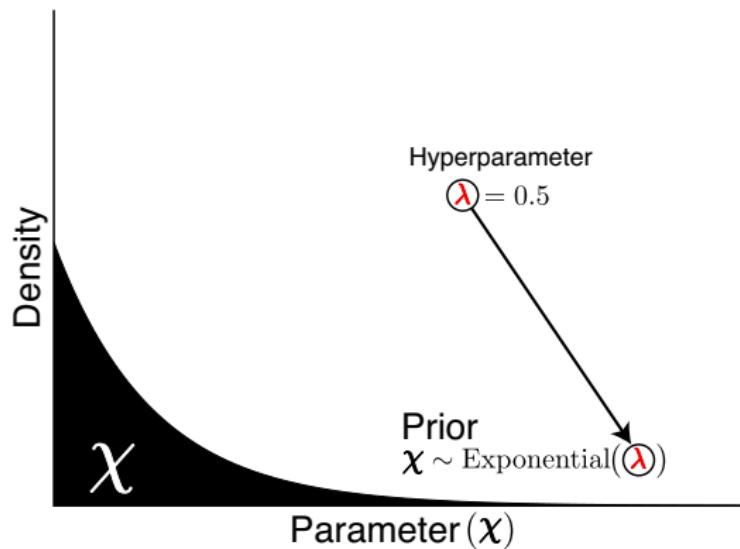
A HIERARCHICAL BAYESIAN MODEL

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



A HIERARCHICAL BAYESIAN MODEL

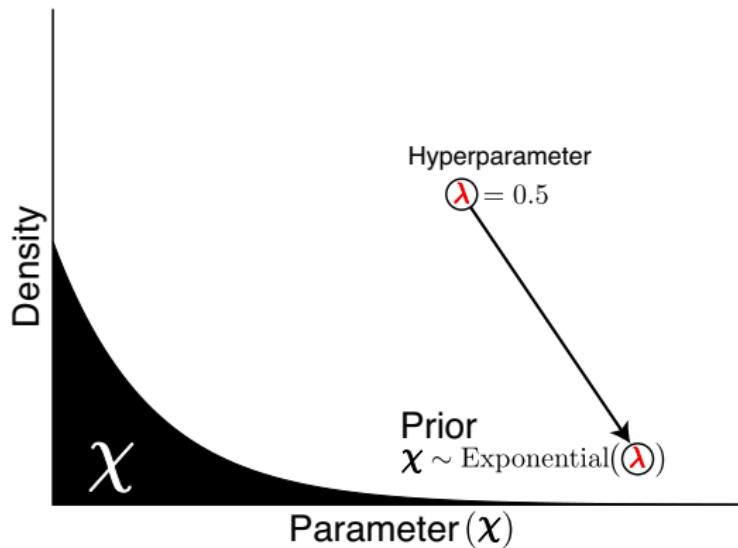
The **exponential** prior on χ has a hyperparameter: λ



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

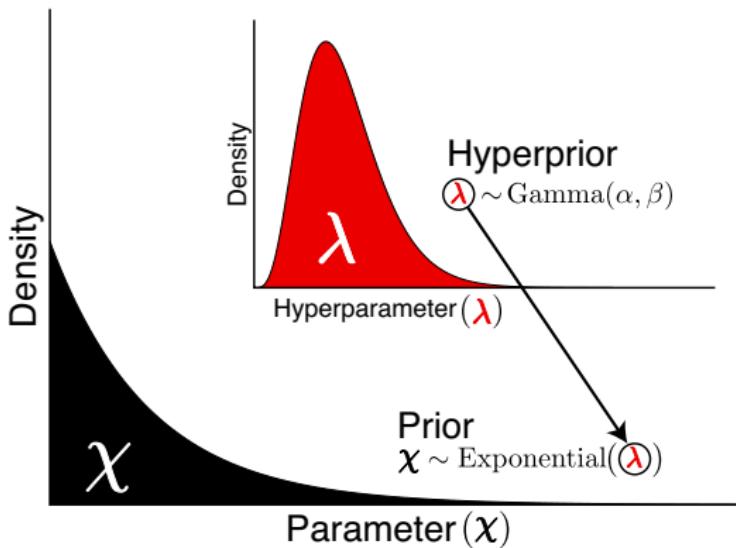


A HIERARCHICAL BAYESIAN MODEL

Hyperprior:

second order prior
placed on a
hyperparameter

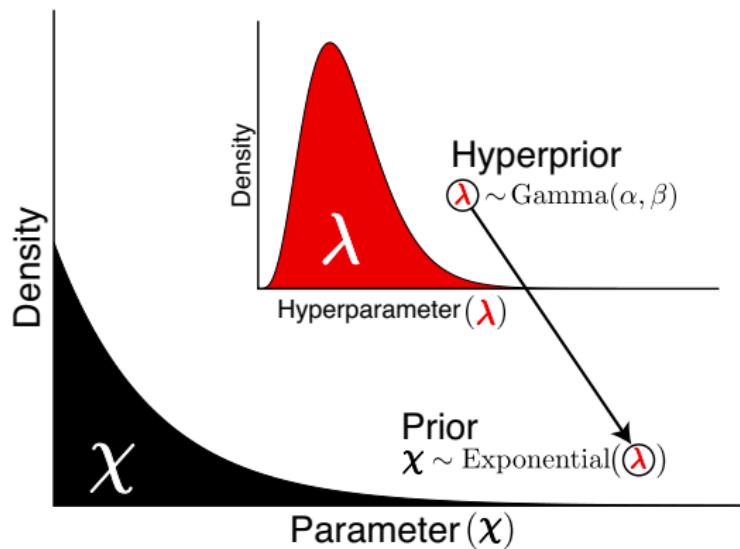
λ becomes a random
variable under the
hierarchical model



A HIERARCHICAL BAYESIAN MODEL

Hyperprior:

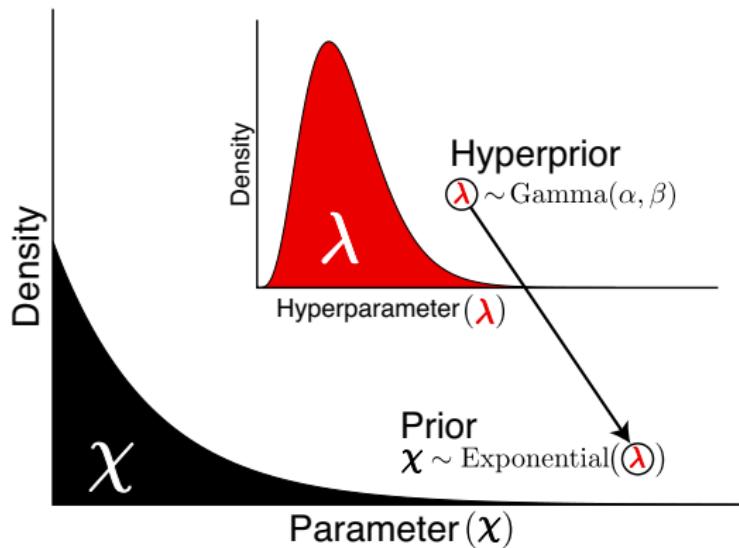
allows for inference
under a richer class of
models



A HIERARCHICAL BAYESIAN MODEL

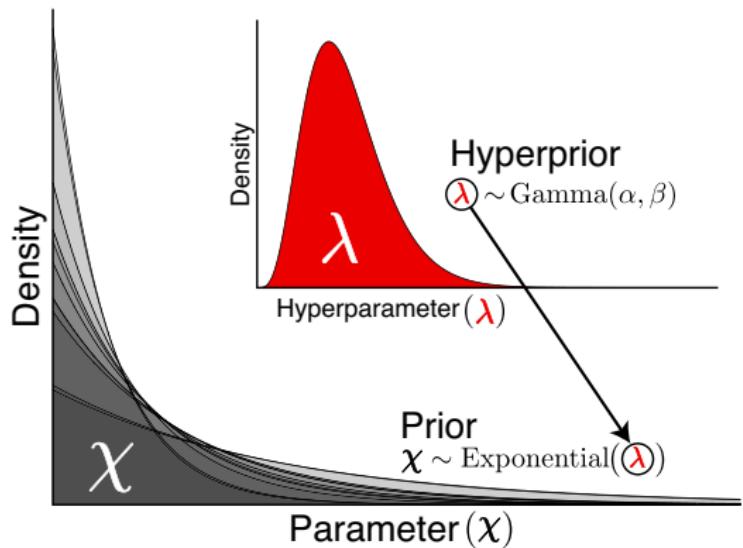
Hyperprior:

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

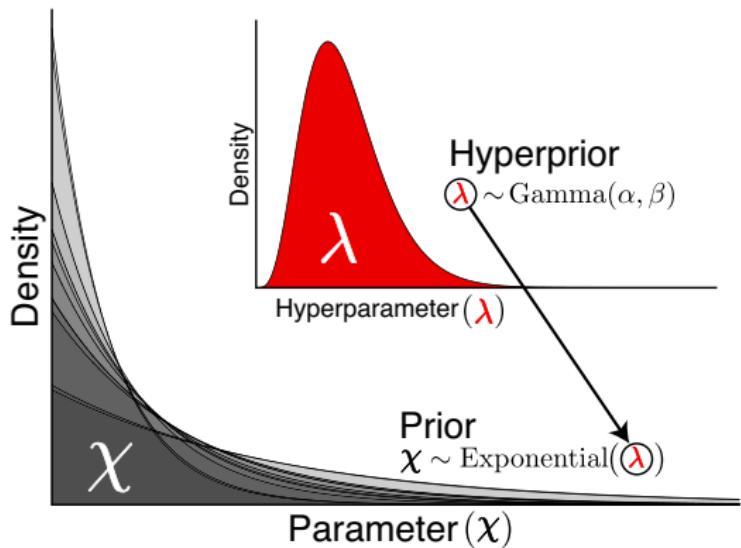


A HIERARCHICAL BAYESIAN MODEL

Hyperprior:

provides estimates of
the hyperparameter

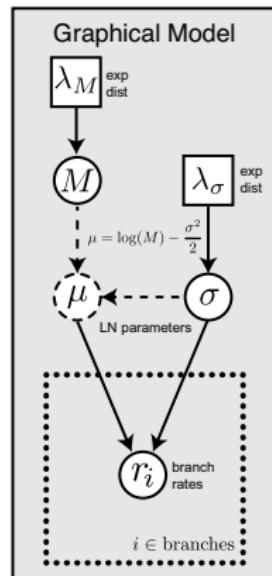
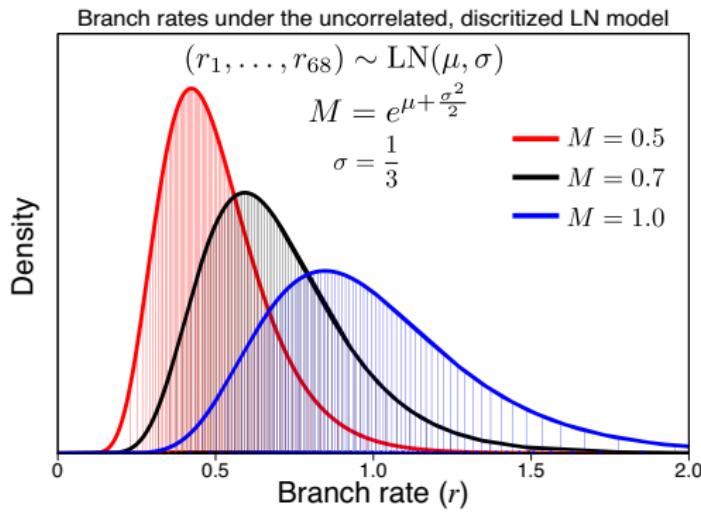
accounts for and
quantifies uncertainty
in the hyperparameter



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



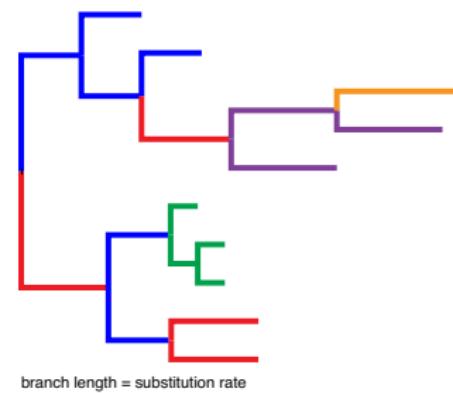
$$\mathbb{E}(M) = \lambda_M^{-1}$$

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

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

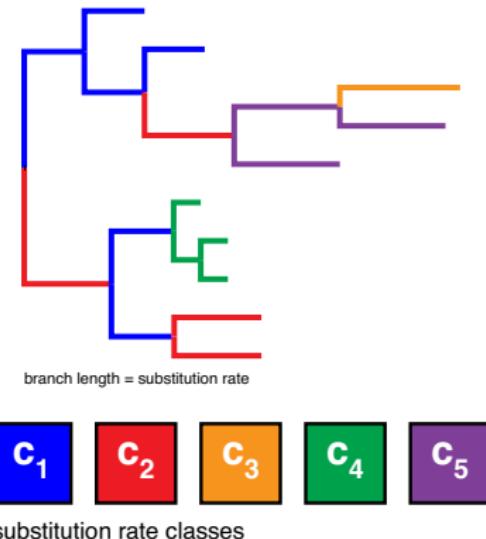


THE DIRICHLET PROCESS PRIOR (DPP)

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

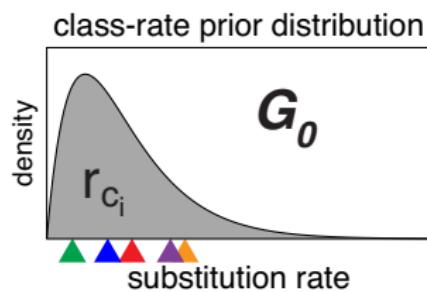
Random variables under the DPP:

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

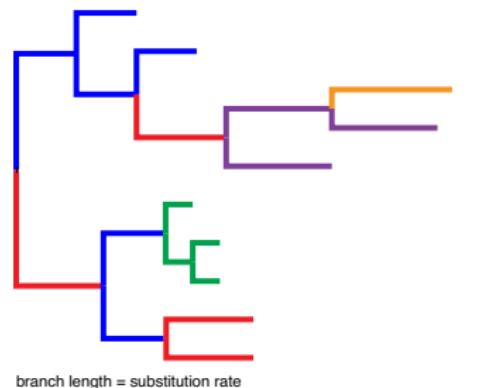


THE DIRICHLET PROCESS PRIOR (DPP)

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

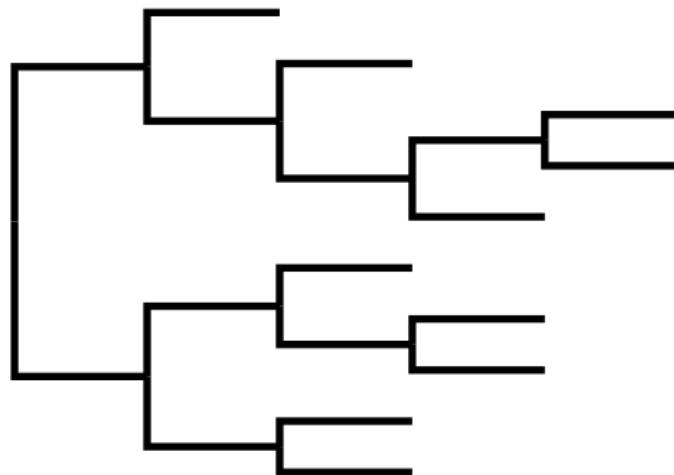


r_{c_i} = the rate value for each class



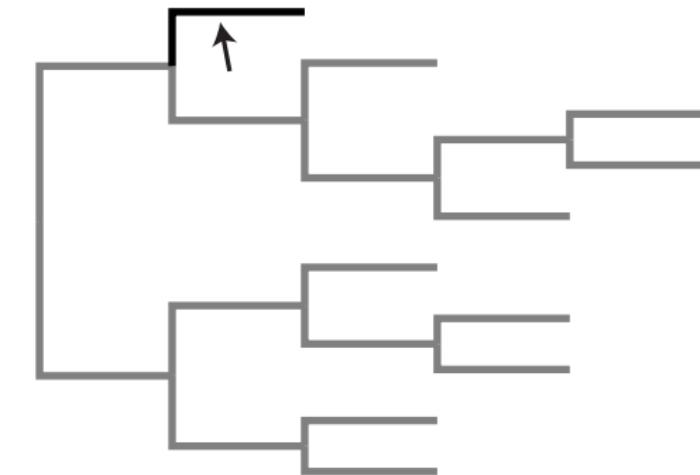
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

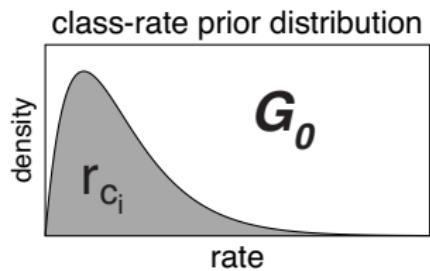


THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

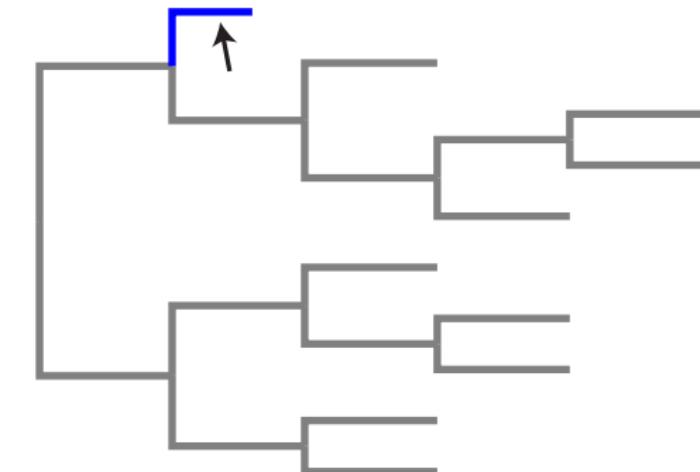


rate classes

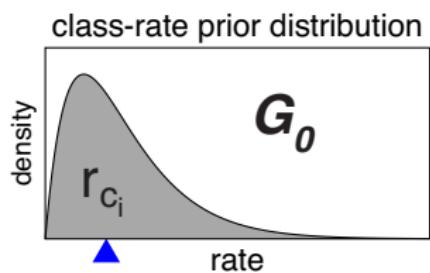


THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

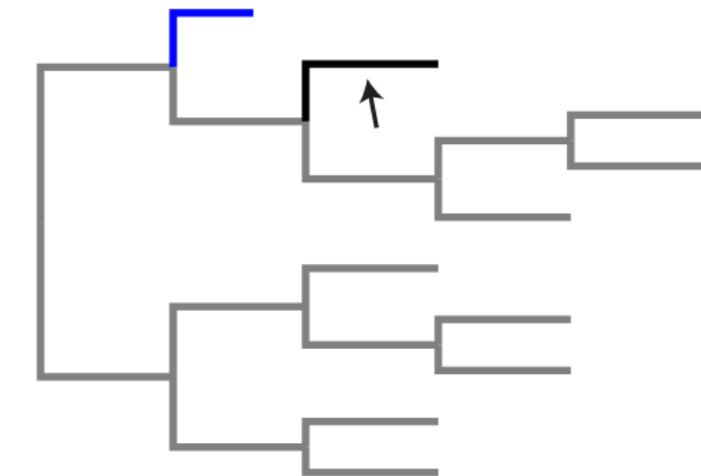


1
c₁
rate classes



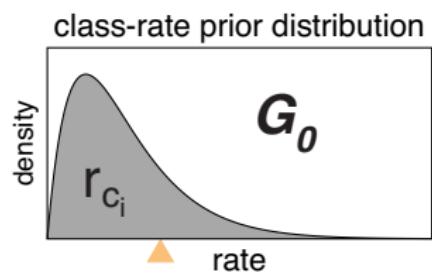
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate



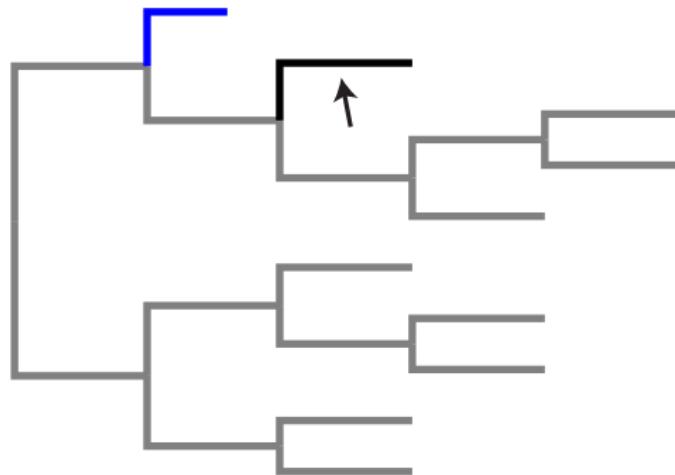
1
c₁
rate classes

α
c_∞
⋮



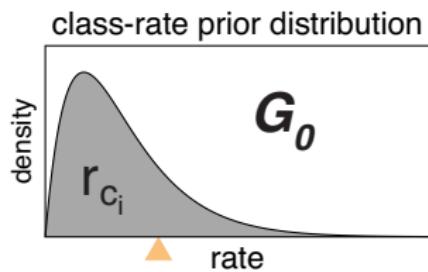
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate



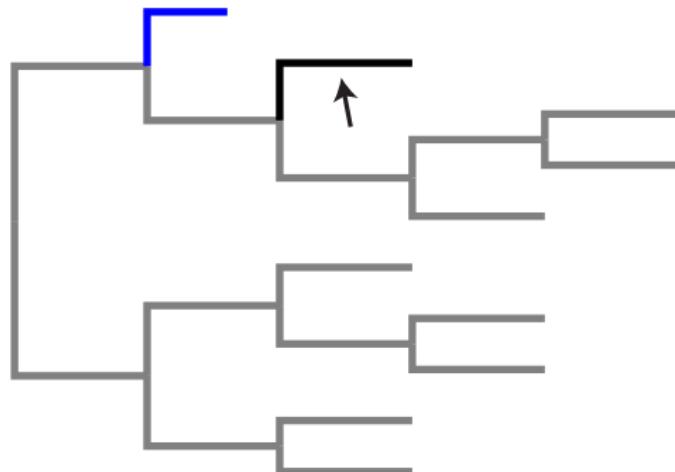
$$c_1 \leftarrow \frac{1}{1+\alpha}$$

rate classes



THE DIRICHLET PROCESS PRIOR (DPP)

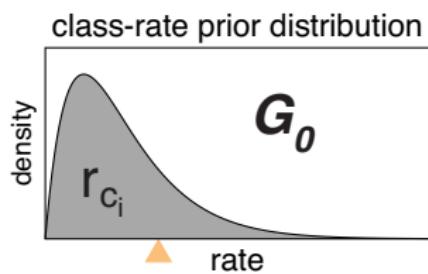
branch length = substitution rate



$$c_1 \leftarrow \frac{1}{1+\alpha}$$

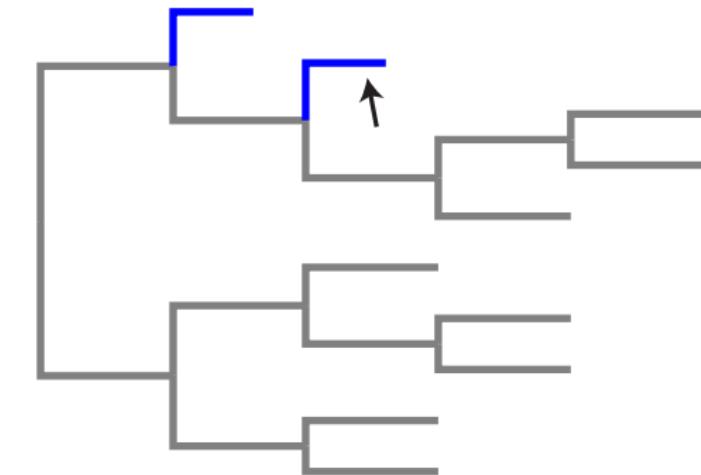
rate classes

$$\frac{\alpha}{1+\alpha} \rightarrow \begin{matrix} c_{\infty} \\ \vdots \end{matrix}$$

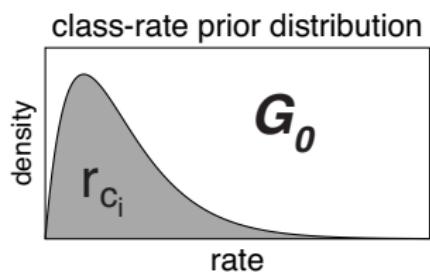


THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

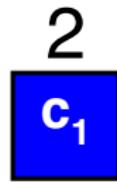
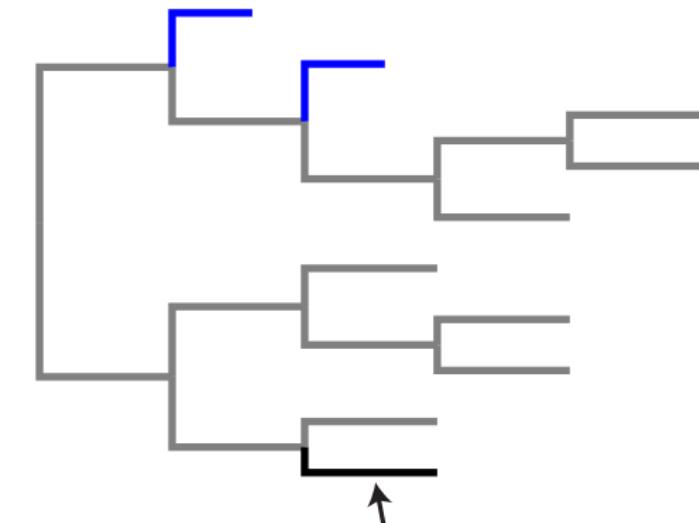


2
c₁
rate classes



THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

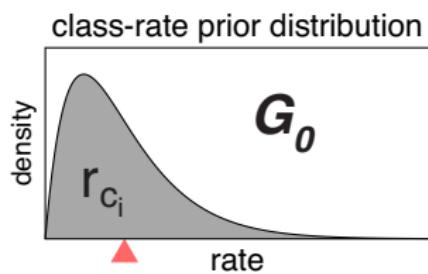


rate classes

α

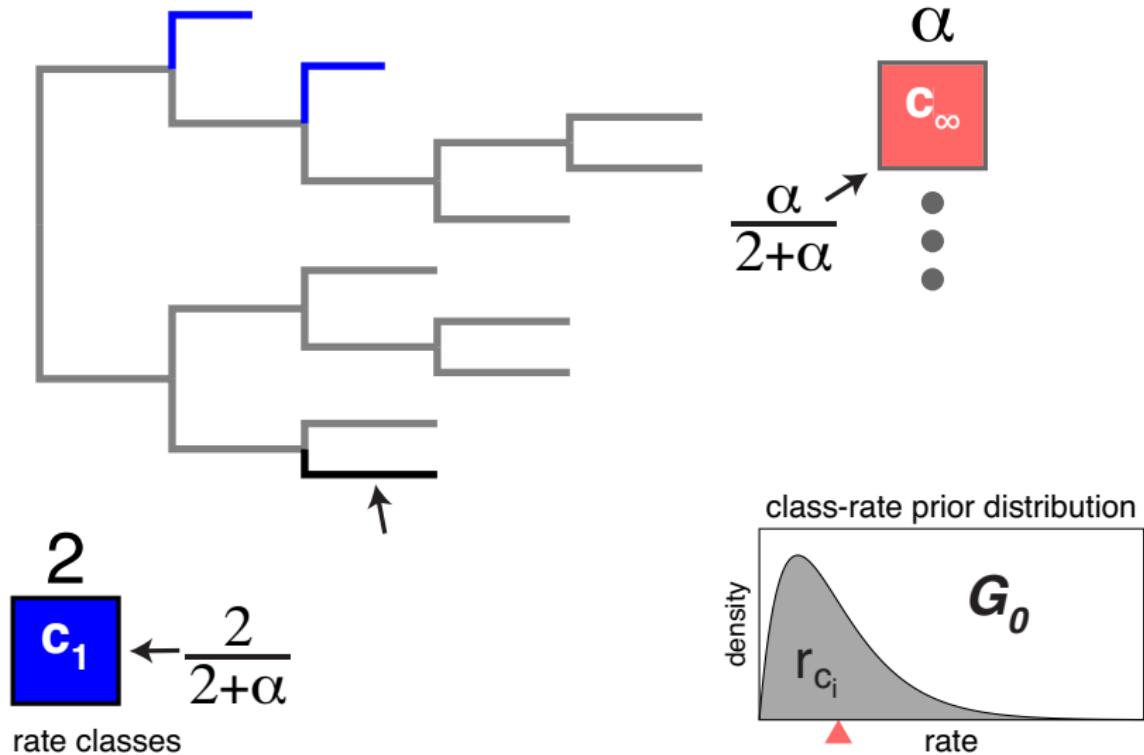
c_{∞}

⋮



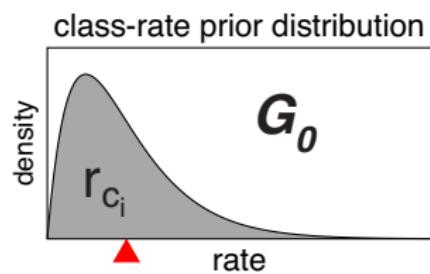
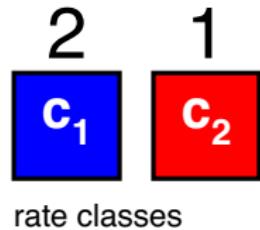
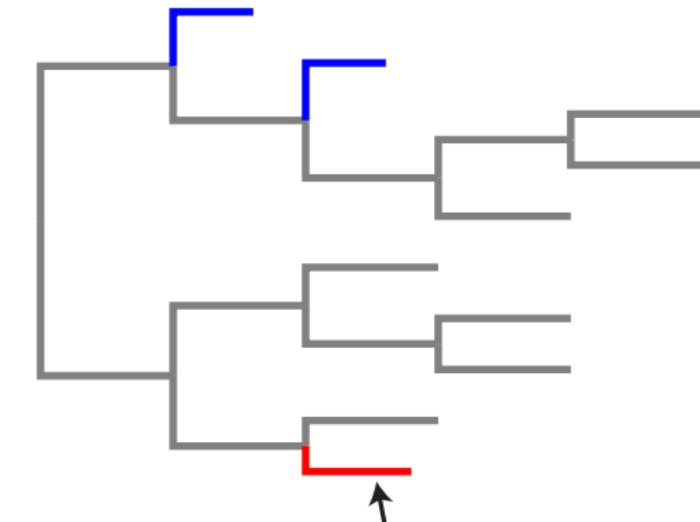
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate



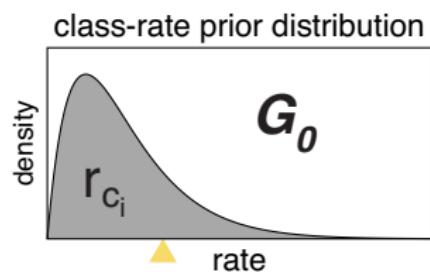
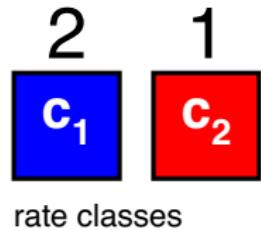
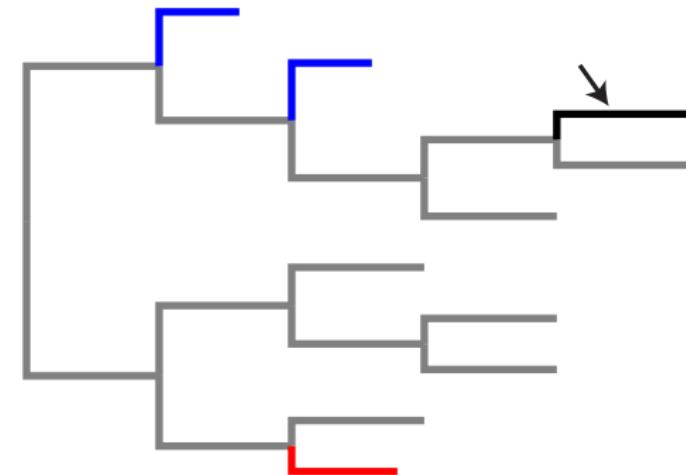
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate



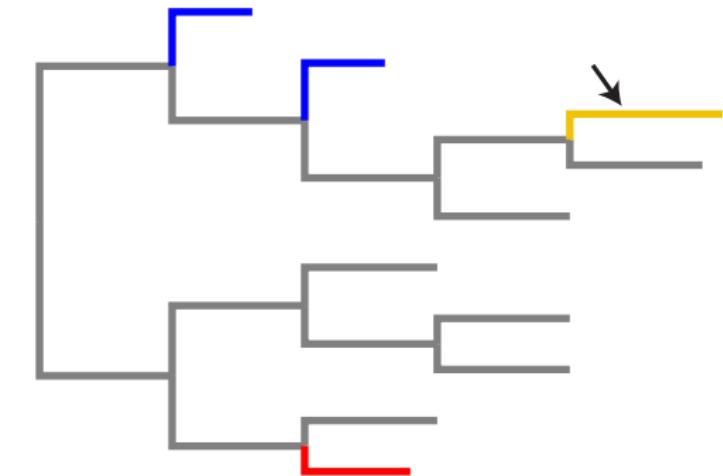
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

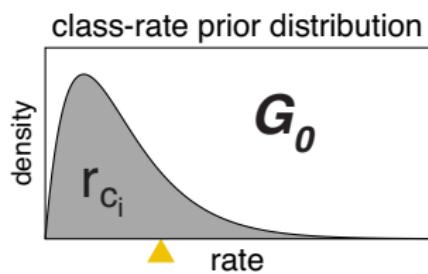


THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

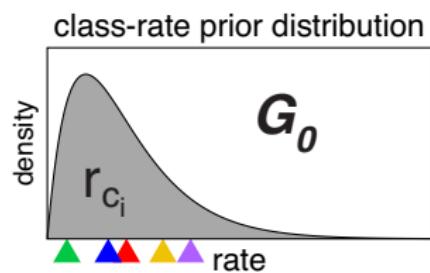
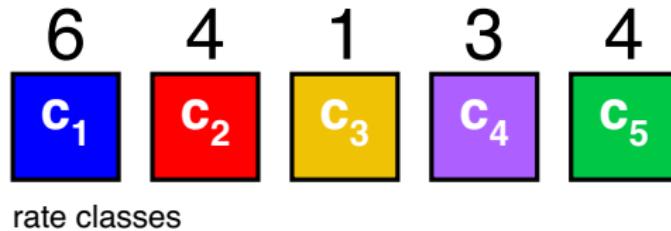
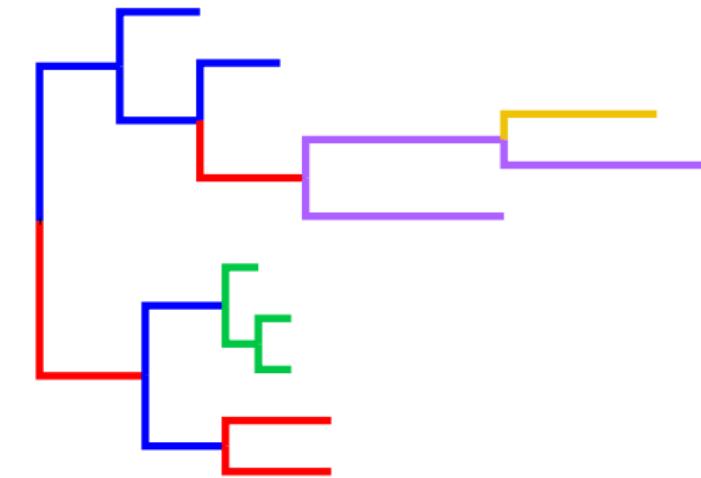


2 1 1
c₁ c₂ c₃
rate classes



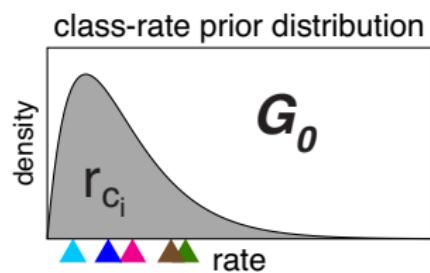
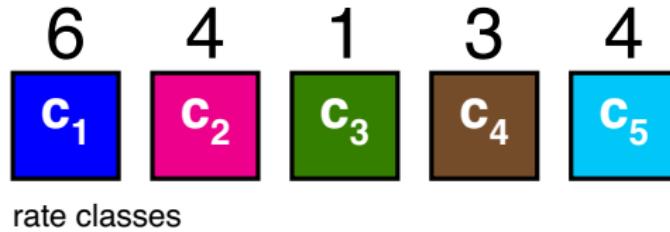
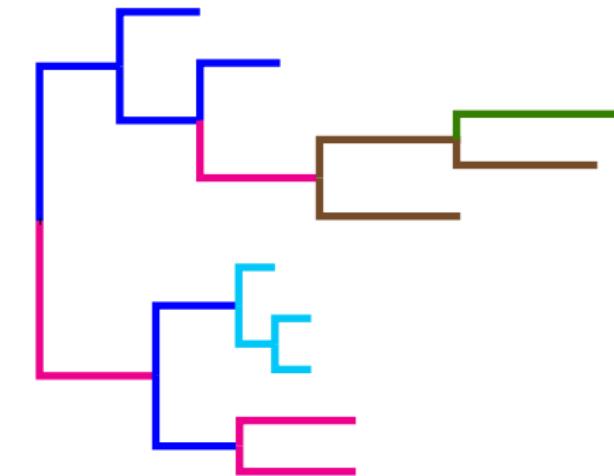
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate



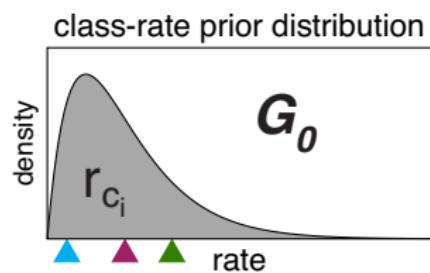
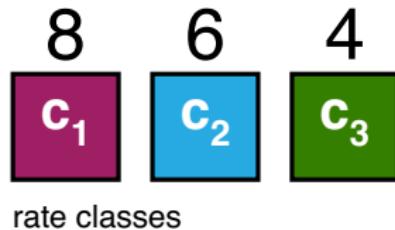
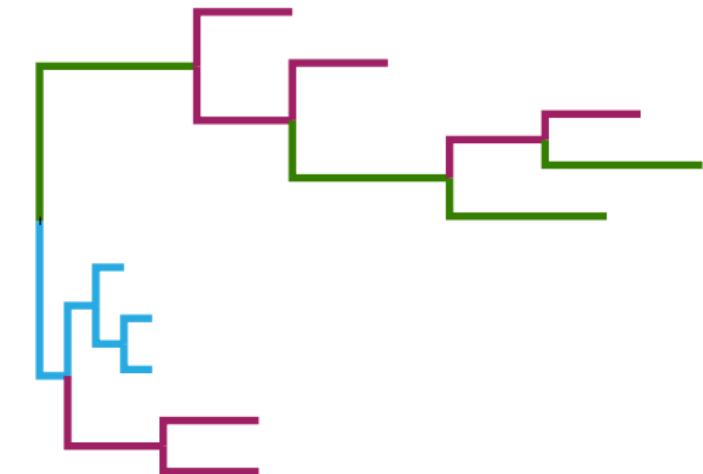
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate



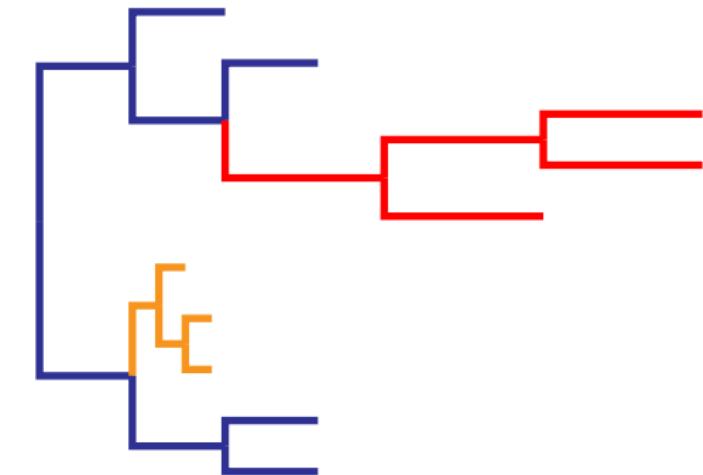
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

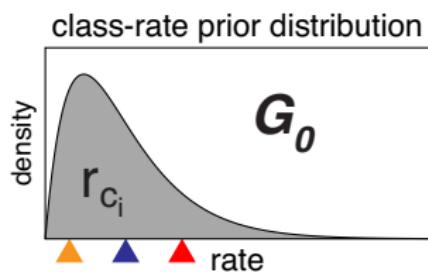
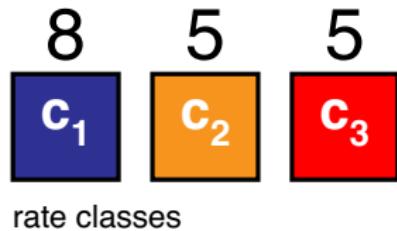


THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

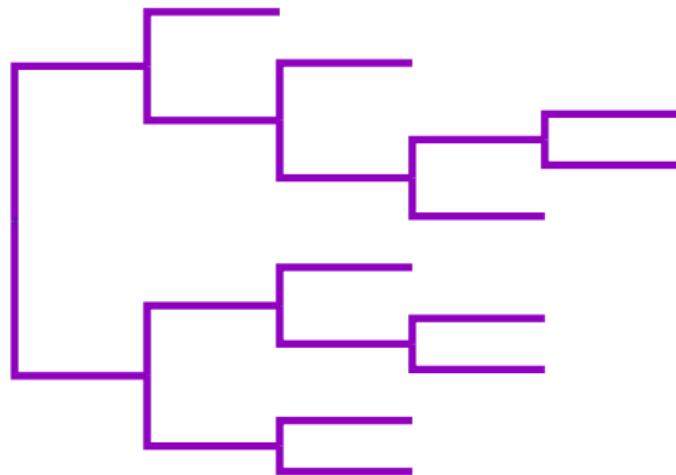


Local molecular
clock



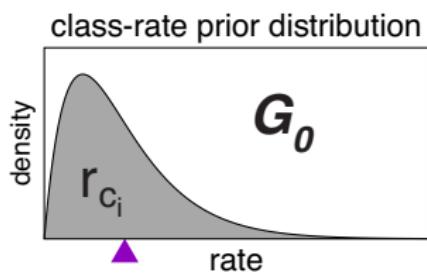
THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate



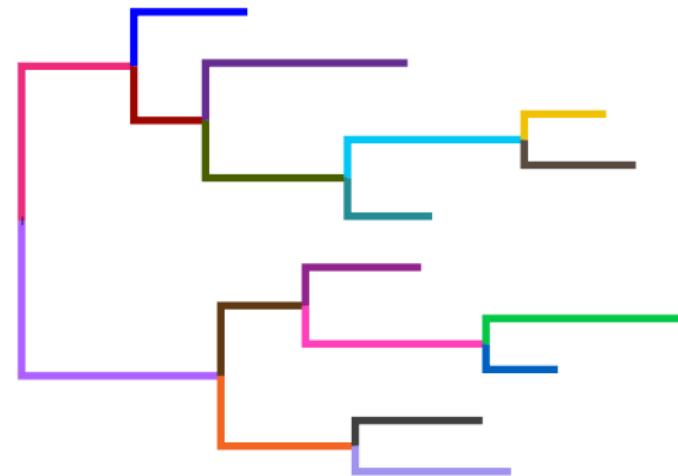
Global molecular clock

18
c₁
rate classes

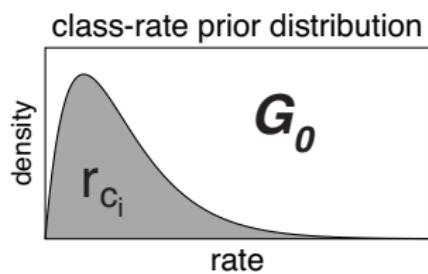
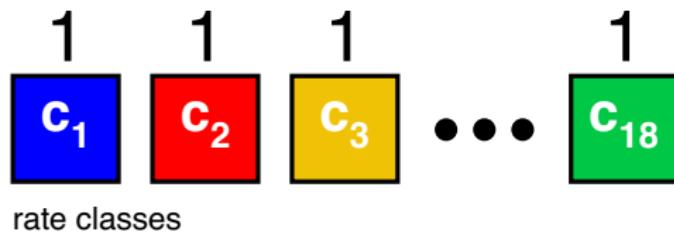


THE DIRICHLET PROCESS PRIOR (DPP)

branch length = substitution rate

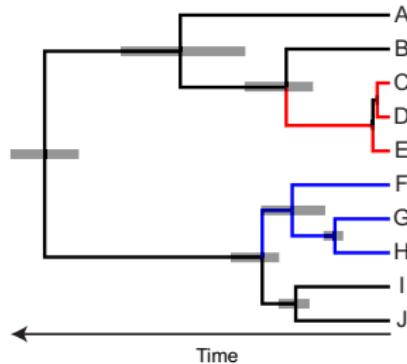


Independent
rates



BAYESIAN INFERENCE UNDER THE DPP

Current implementation: DPPDiv



Availability:

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

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

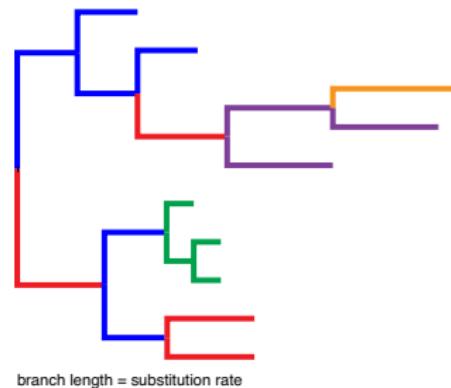
LINEAGE-SPECIFIC SUBSTITUTION RATES

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

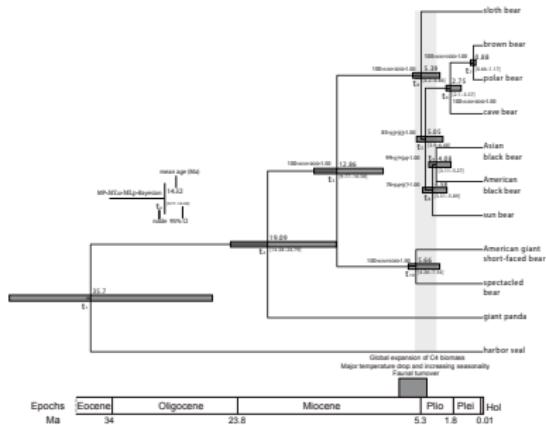
Efficient MCMC implementations



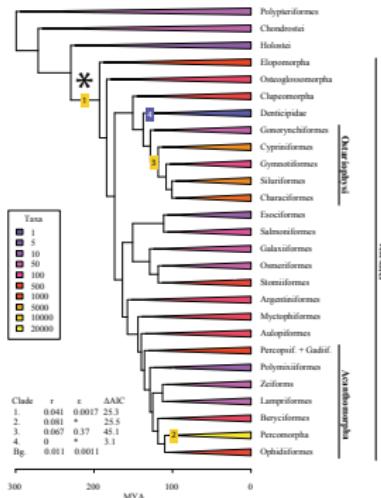
substitution rate classes

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.

MODELING RATE VARIATION

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

- Global molecular clock
- Local molecular clocks
- Compound Poisson process model
- Log-normally distributed autocorrelated rates
- Uncorrelated/independent rates models
- Dirichlet process prior

Model selection and model uncertainty are critically important for Bayesian divergence time analysis

BAYESIAN DIVERGENCE TIME ESTIMATION

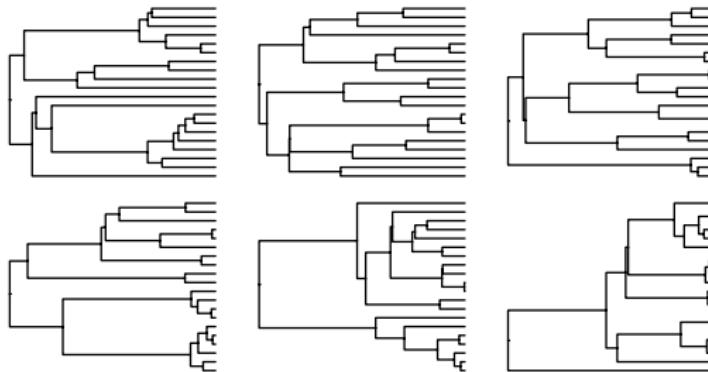
Estimating divergence times relies on 2 main elements:

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

PRIORS ON NODE TIMES

Relaxed clock Bayesian analyses require a prior distribution on node times

$$f(\mathcal{A} | \theta_{\mathcal{A}})$$



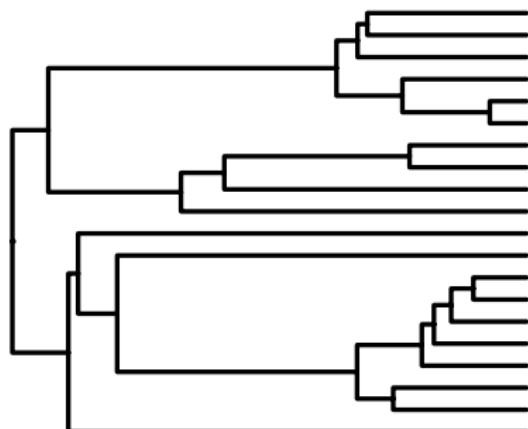
Different node-age priors make different assumptions about the timing of divergence events

GENERIC NODE TIME PRIORS

Assumed to be vague or uninformative by not making assumptions about biological processes

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

(conditioned on root age)

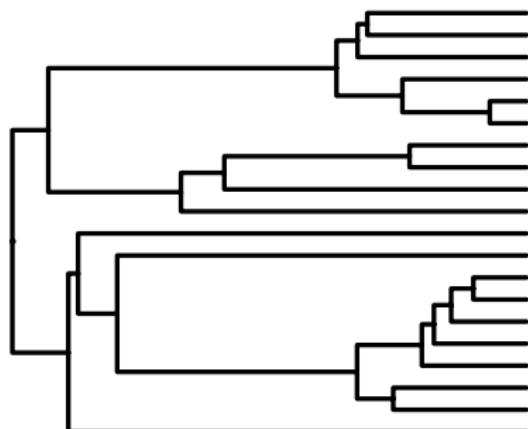


GENERIC NODE TIME PRIORS

Assumed to be vague or uninformative by not making assumptions about biological processes

Dirichlet prior: ages of the interior nodes on a single path spanning the age of the root node to one of the tip nodes are sampled from a flat Dirichlet distribution

(conditioned on root age)



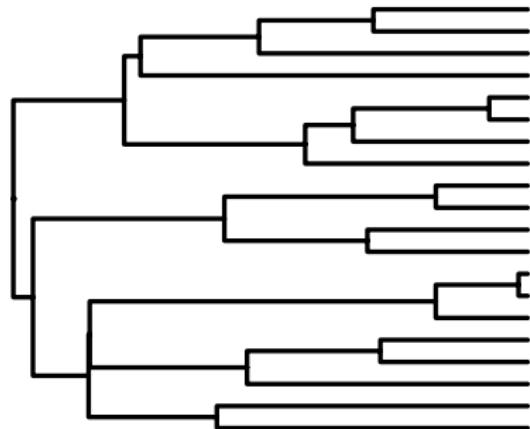
STOCHASTIC BRANCHING PROCESSES

Node-age priors based on stochastic models of lineage diversification

Yule process: assumes a constant rate of speciation, \mathcal{S} , across lineages

A pure birth process—every node leaves extant descendants (no extinction)

Leads to an exponential waiting-time between speciation events



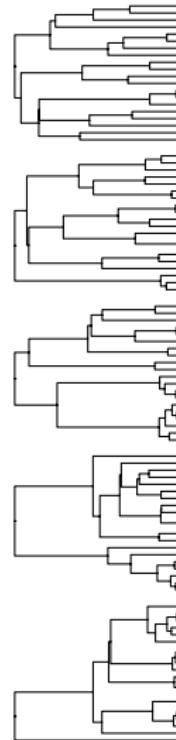
$$f(\mathcal{A} | \mathcal{S})$$

STOCHASTIC BRANCHING PROCESSES

Different values of \mathcal{S} and \mathcal{E} lead to different trees

Bayesian inference under these models can be very sensitive to the values of these parameters

Using hyperpriors on \mathcal{S} and \mathcal{E} accounts for uncertainty in these hyperparameters



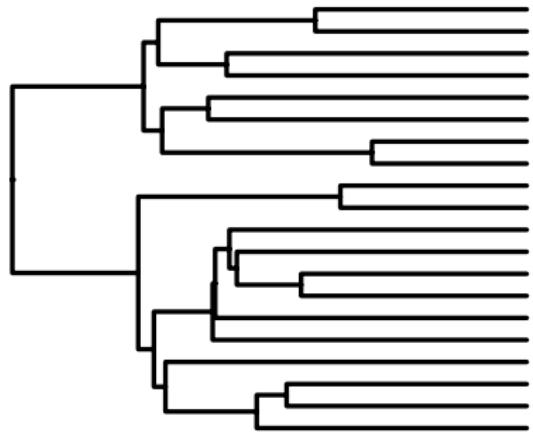
STOCHASTIC BRANCHING PROCESSES

Node-age priors based on stochastic models of lineage diversification

Birth-death-sampling

process: an extension of the constant-rate birth-death model that accounts for random sampling of tips

Conditions on a probability of sampling a tip, ρ

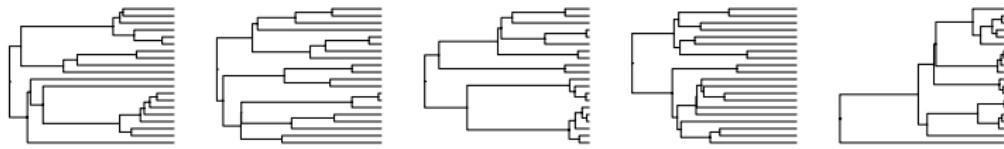


$$f(\mathcal{A} | \mathcal{S}, \mathcal{E}, \rho, N)$$

PRIORS ON NODE TIMES

Sequence data are only informative on *relative*

Node-time priors cannot lead to precise estimates of
absolute node ages

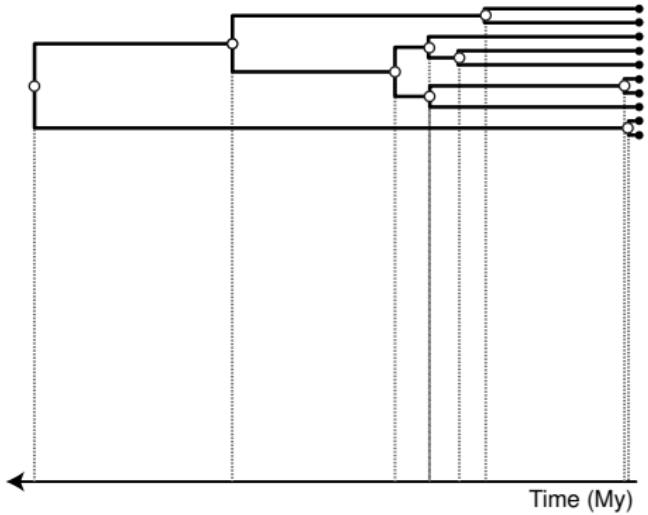


We need external information (like fossils) to *calibrate* or scale the tree to absolute time

$$f(\mathcal{A} | \theta_{\mathcal{A}}, \mathcal{C})$$

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

Time (My)

Fossil Calibration

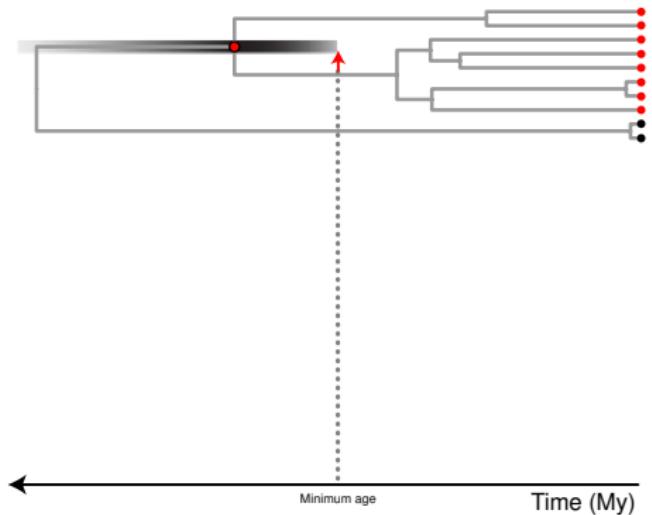


Fossil taxa are assigned to monophyletic clades



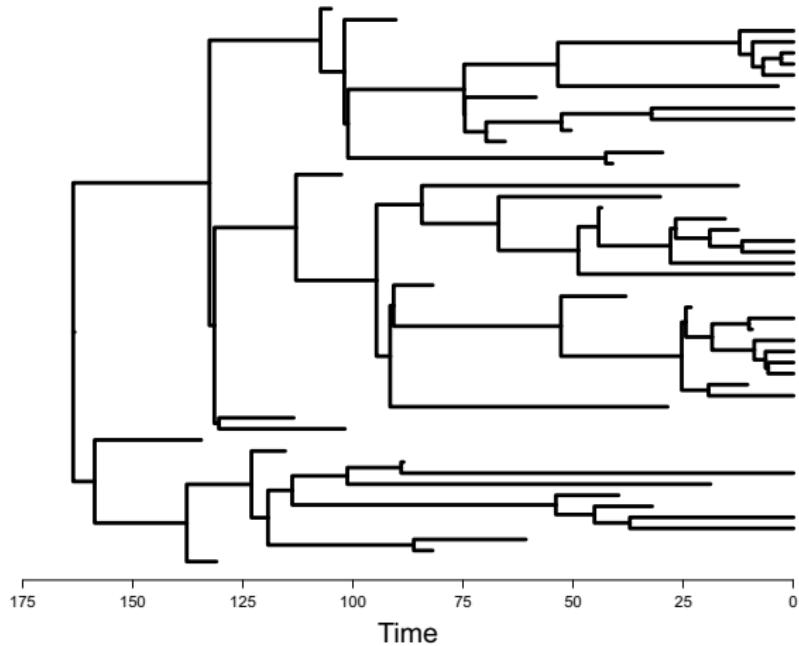
Fossil Calibration

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



MODELING BRANCHING PROCESSES

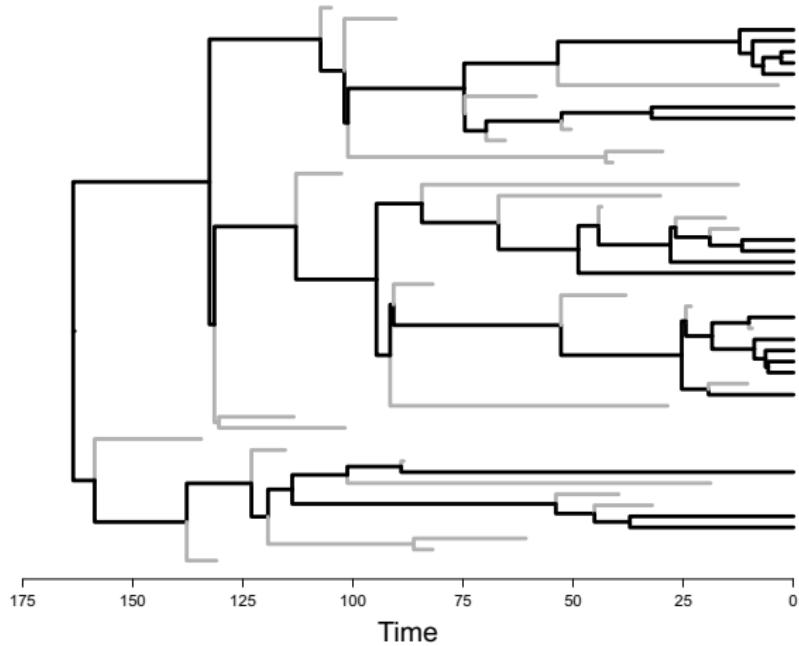
Assume constant
rates of
speciation (S)
and extinction
(E)
(20 extant taxa)



MODELING BRANCHING PROCESSES

Assume constant
rates of
speciation (S)
and extinction
(E)

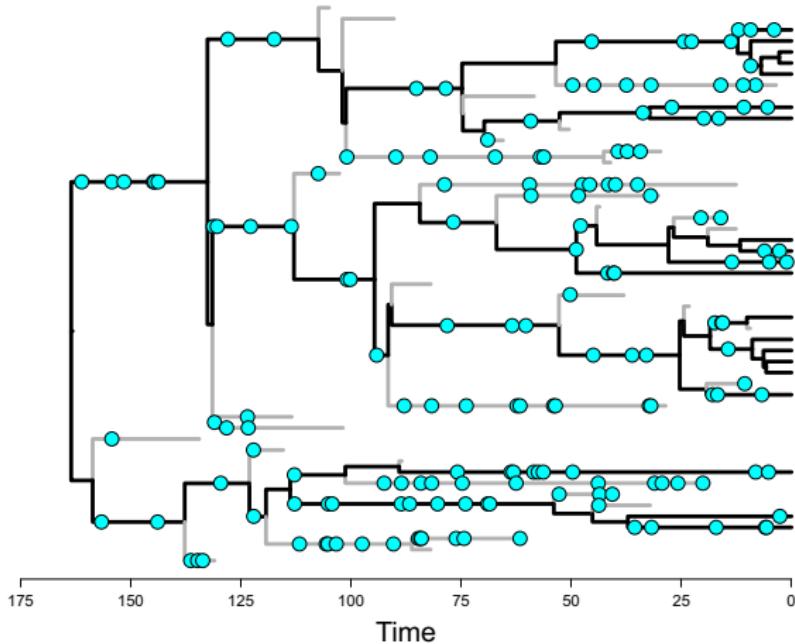
(20 extant taxa)



MODELING TAPHONOMIC PROCESSES

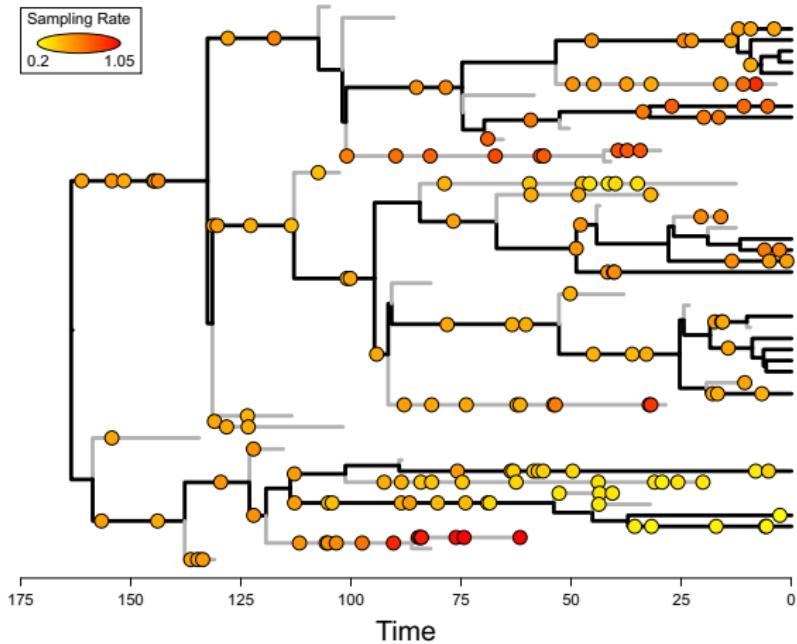
Fossilization events were generated according to a Poisson process

this example has 162 fossilization events



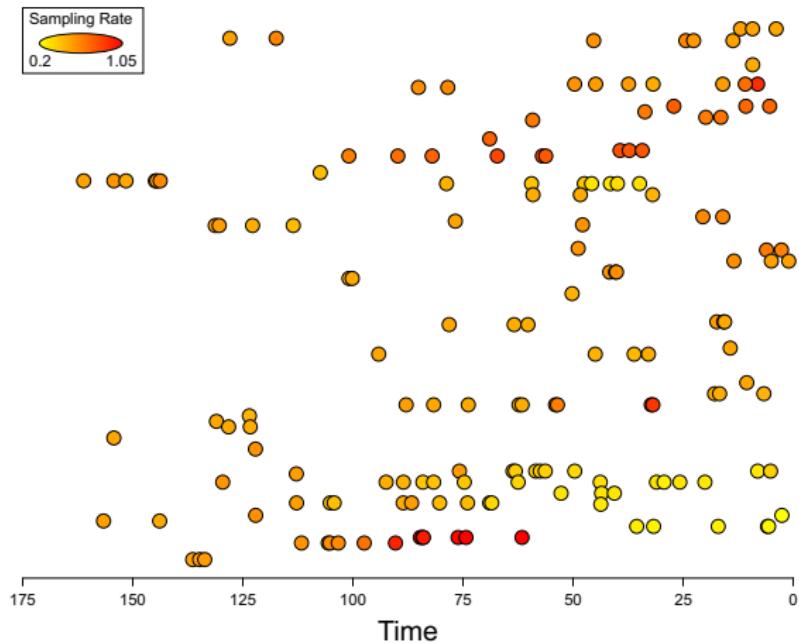
MODELING TAPHONOMIC PROCESSES

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



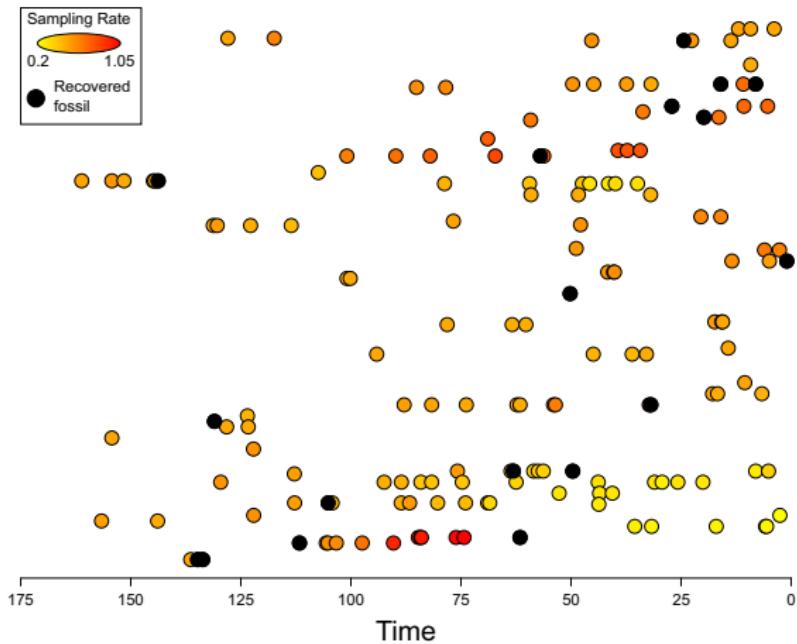
MODELING TAPHONOMIC PROCESSES

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



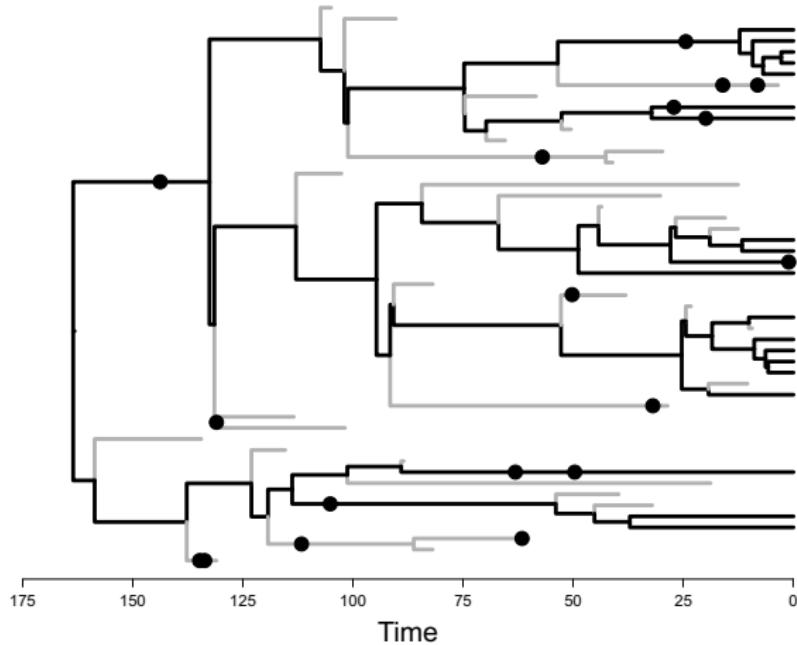
MODELING TAPHONOMIC PROCESSES

18 fossils were
“recovered” in
proportion to
their sampling
rates



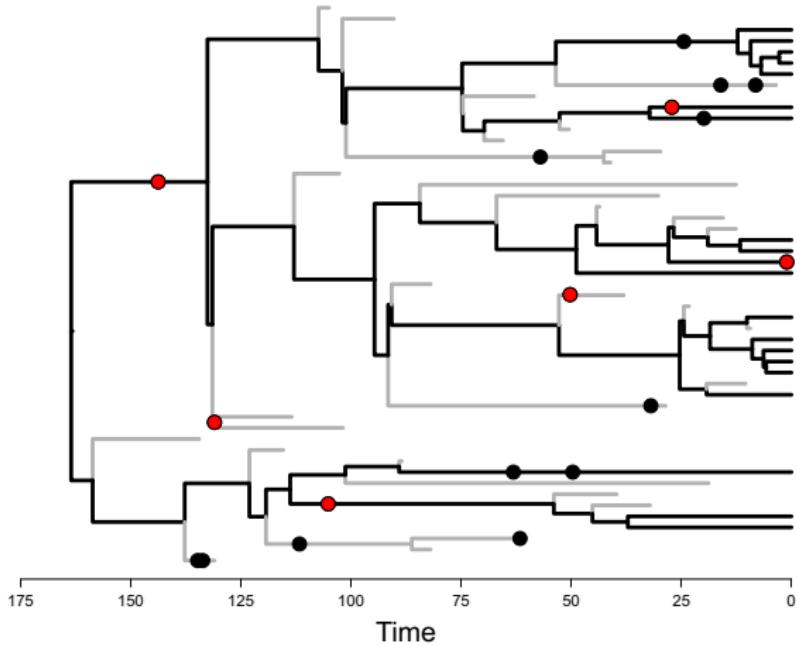
RECOVERED FOSSILS

Assume we
know the true
phylogenetic
placement of the
recovered fossils



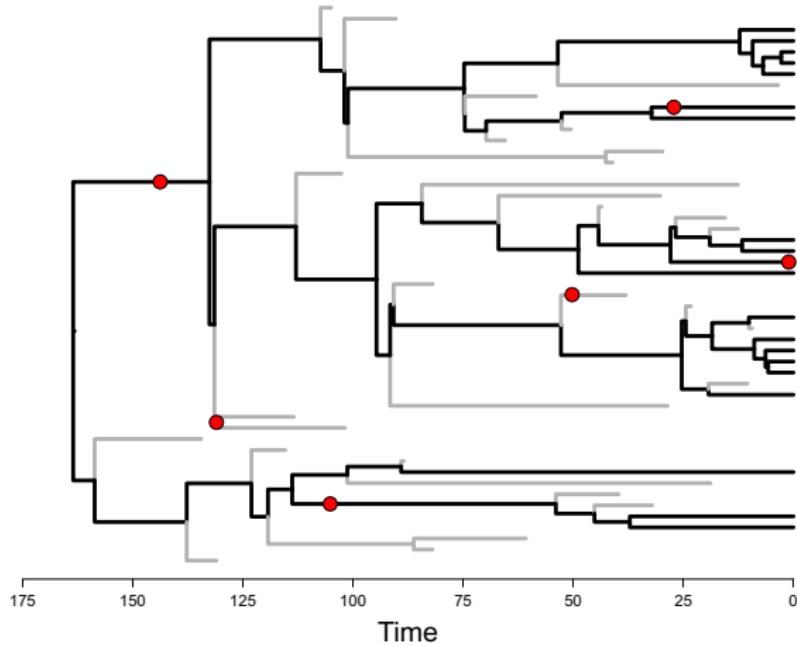
CALIBRATION FOSSILS

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



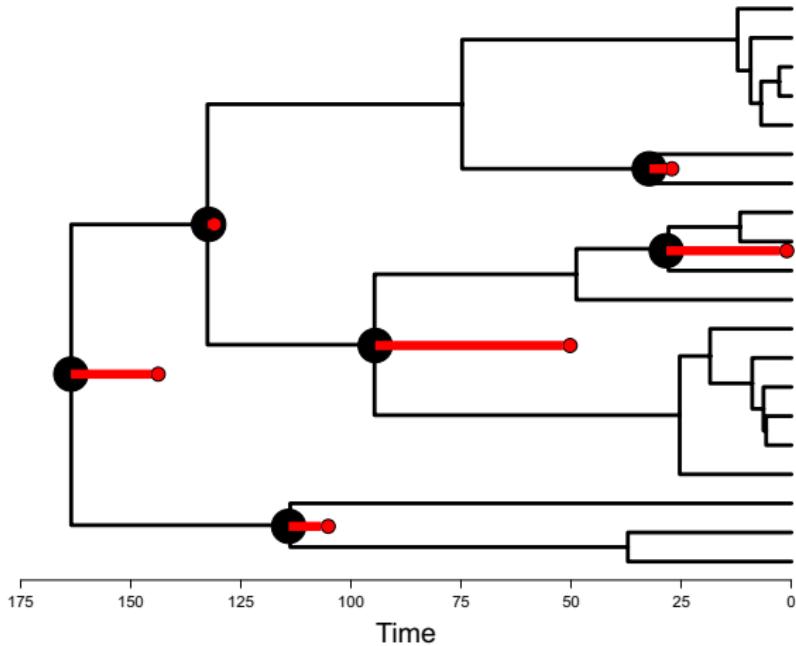
CALIBRATION FOSSILS

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



CALIBRATION FOSSILS

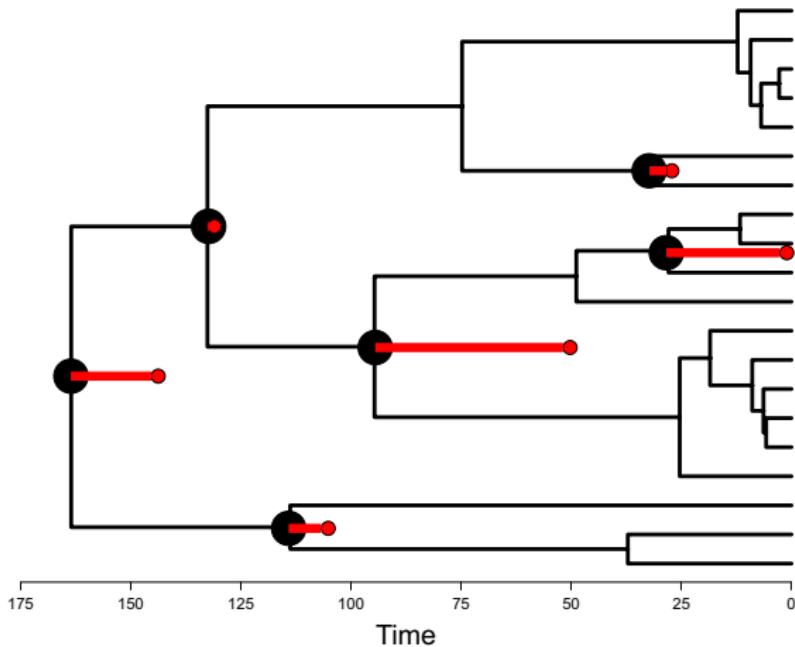
Only the oldest fossil assigned to a given node can be used for calibration



CALIBRATION FOSSILS

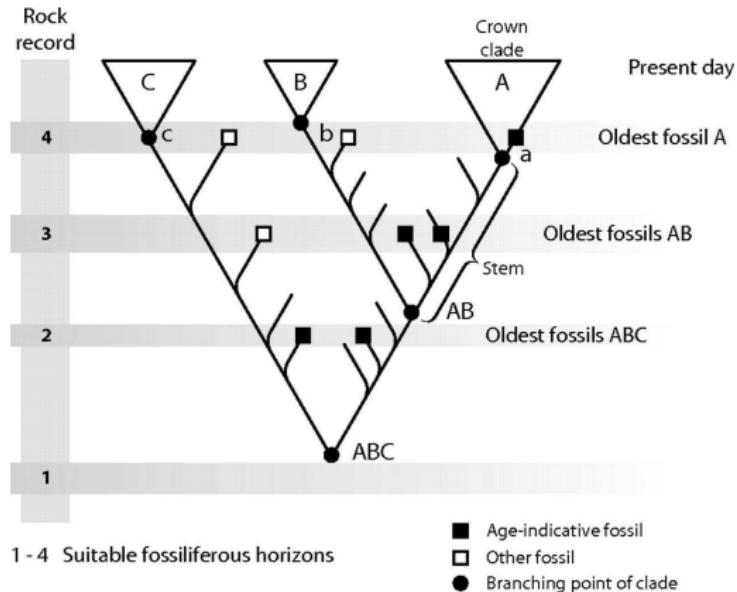
Taphonomic bias

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



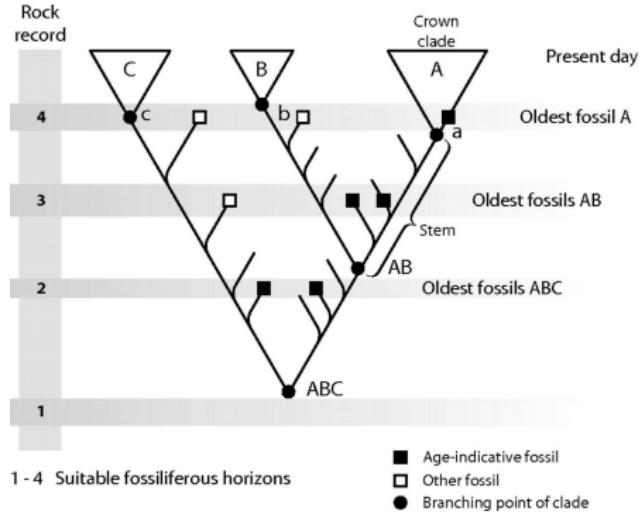
ASSIGNING FOSSILS TO CLADES

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



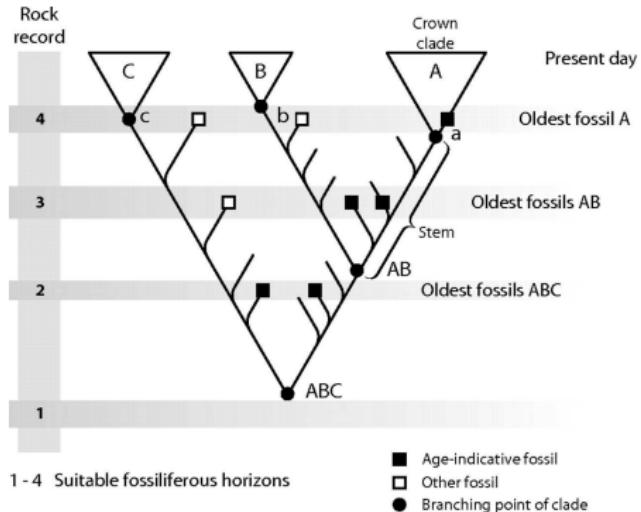
ASSIGNING FOSSILS TO CLADES

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



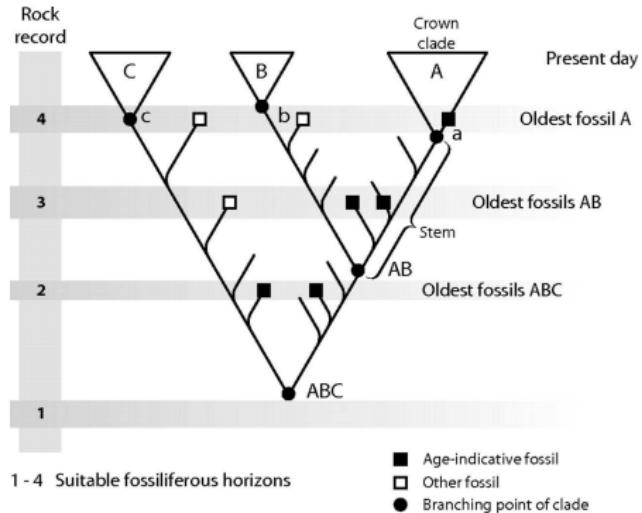
ASSIGNING FOSSILS TO CLADES

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



ASSIGNING FOSSILS TO CLADES

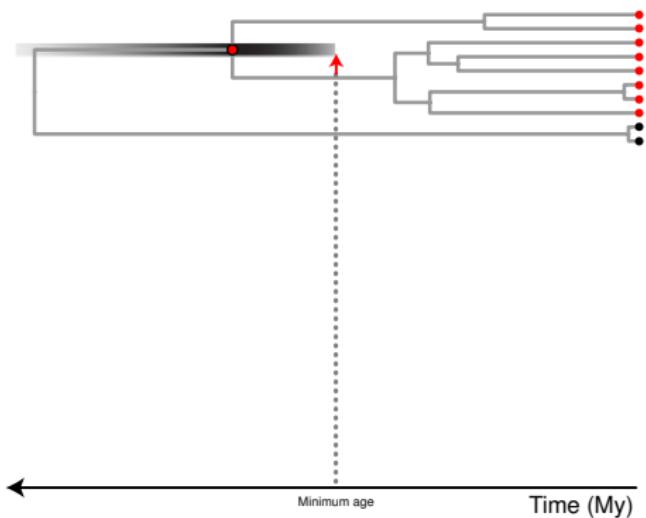
Fossiliferous horizons: the sources in the rock record for relevant fossils



Fossil Calibration

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

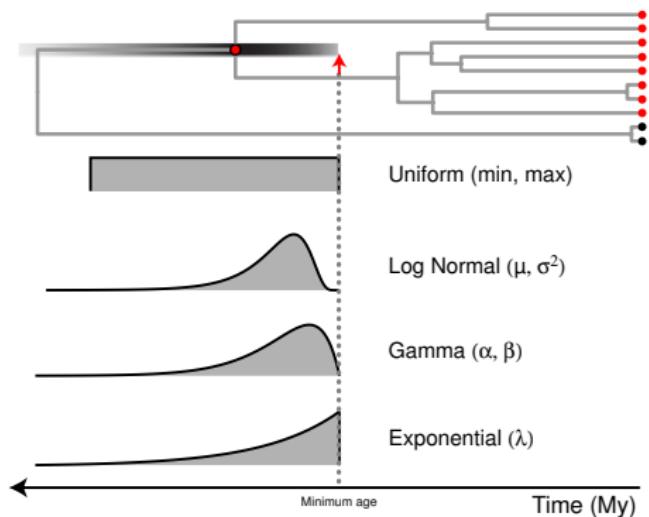
Reliable **maximum** bounds are typically unavailable



PRIOR DENSITIES ON CALIBRATED NODES

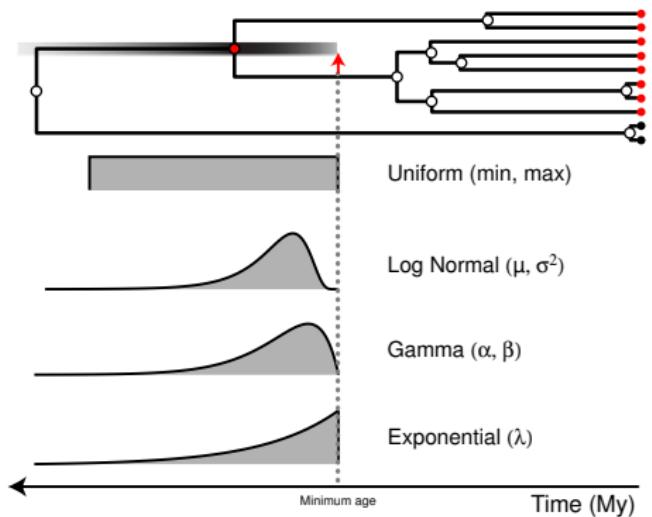
Parametric distributions are typically off-set by the age of the oldest fossil assigned to a clade

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



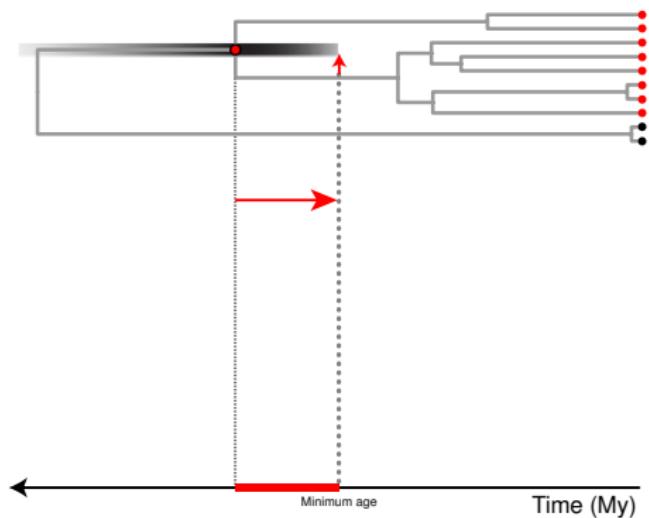
PRIOR DENSITIES ON CALIBRATED NODES

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



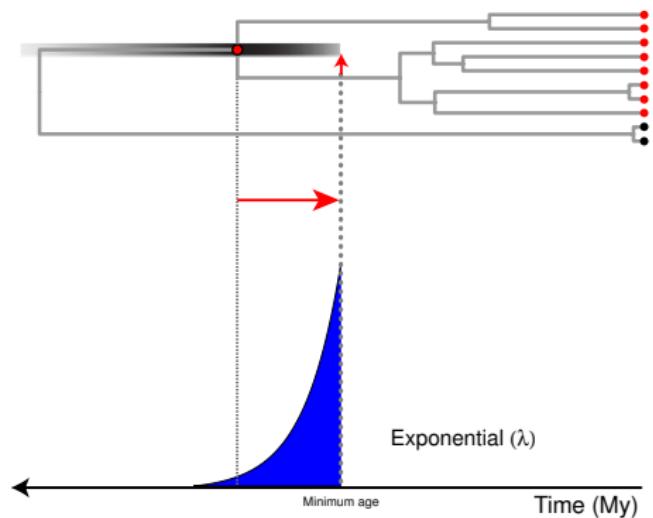
PRIOR DENSITIES ON CALIBRATED NODES

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



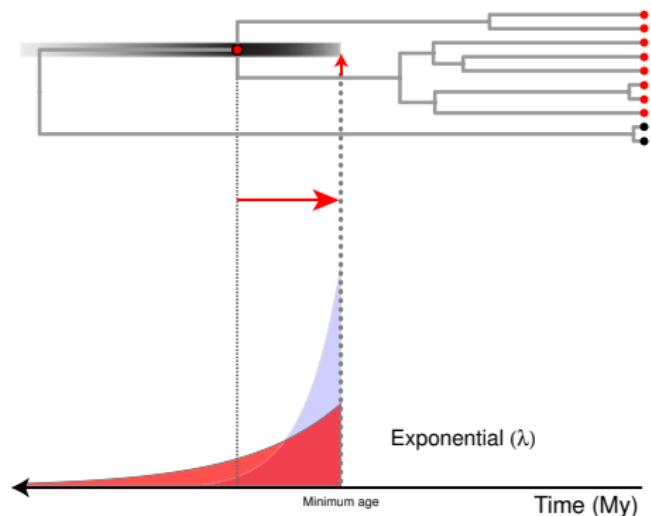
PRIOR DENSITIES ON CALIBRATED NODES

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



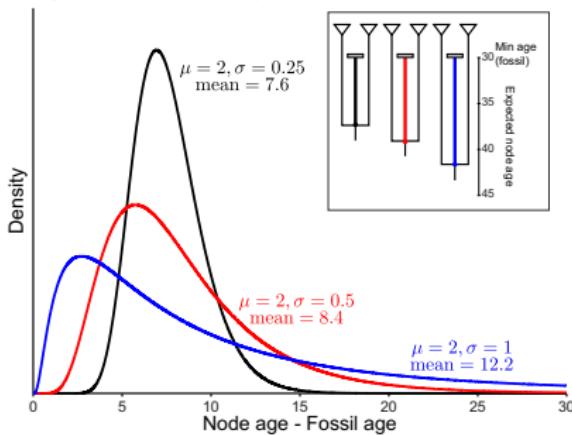
PRIOR DENSITIES ON CALIBRATED NODES

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

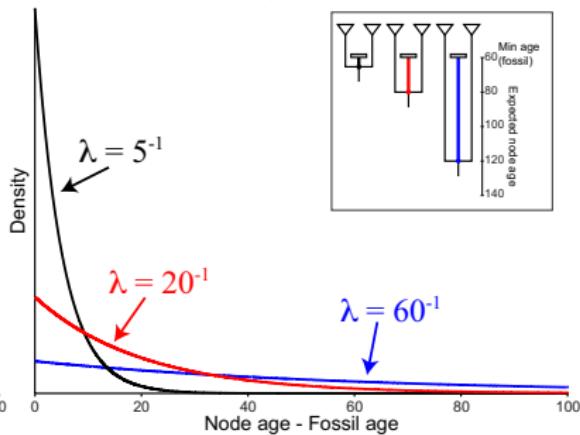


PRIOR DENSITIES ON CALIBRATED NODES

Lognormal prior density

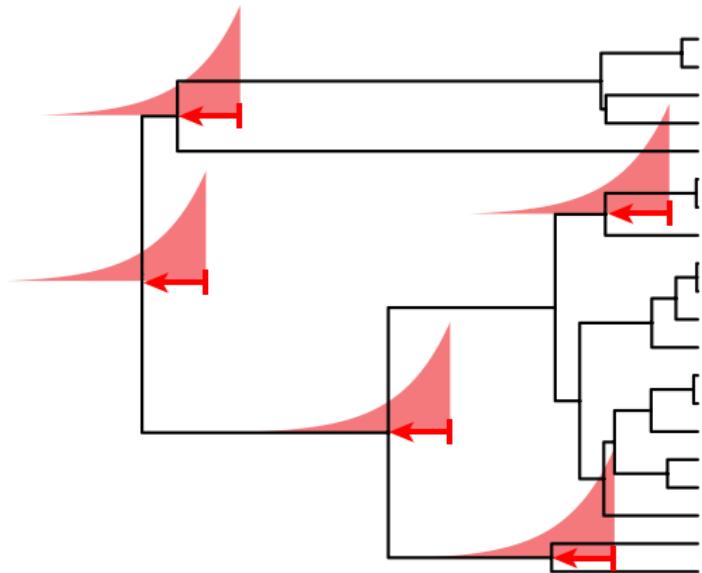


Exponential prior density



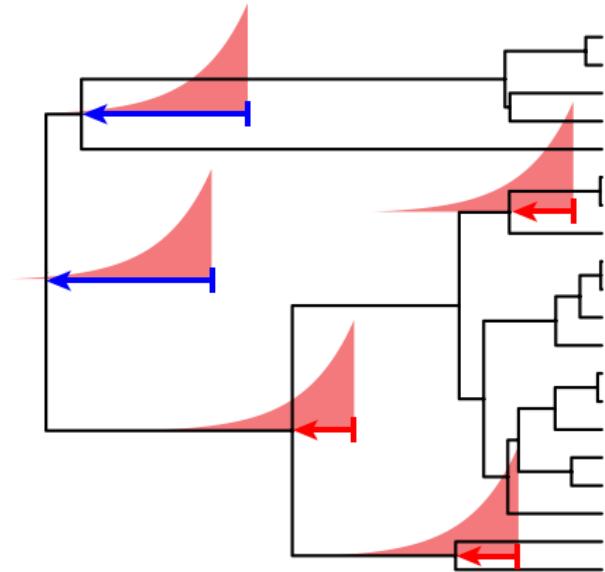
PRIORS ON MULTIPLE CALIBRATIONS

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



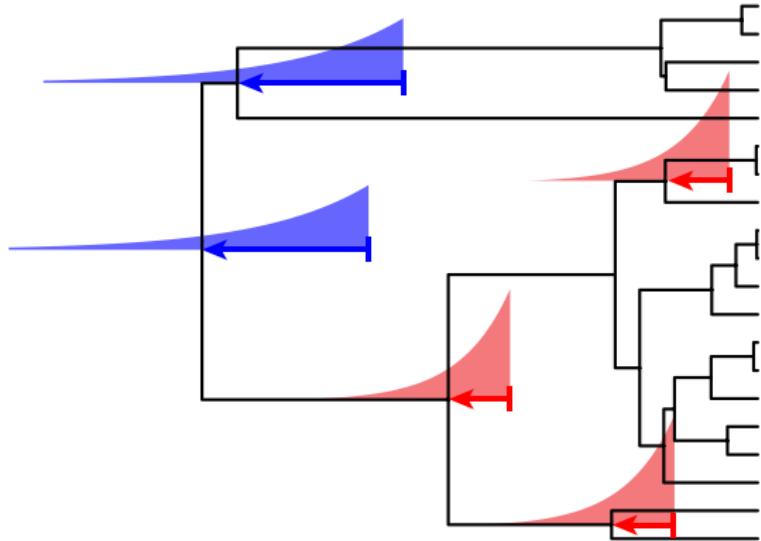
PRIORS ON MULTIPLE CALIBRATIONS

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



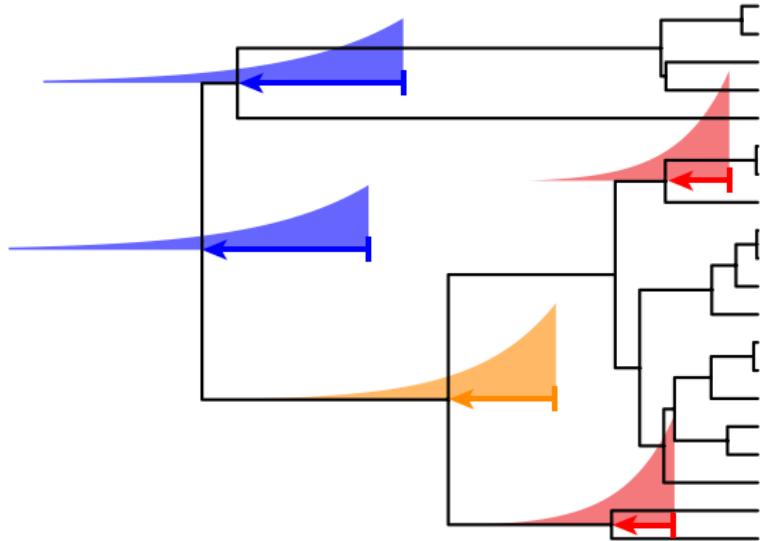
PRIORS ON MULTIPLE CALIBRATIONS

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



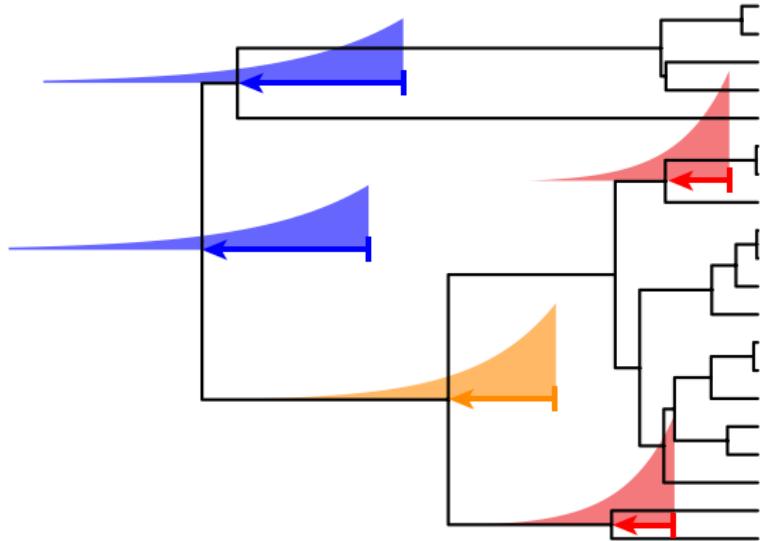
PRIORS ON MULTIPLE CALIBRATIONS

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

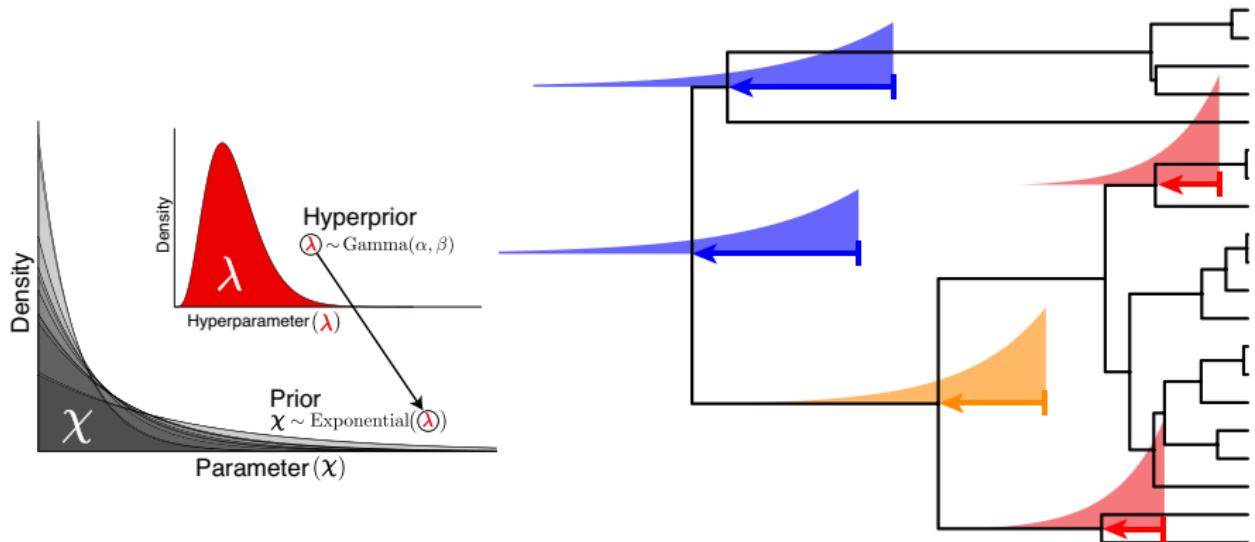


PRIORS ON MULTIPLE CALIBRATIONS

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



PRIORS ON MULTIPLE CALIBRATIONS

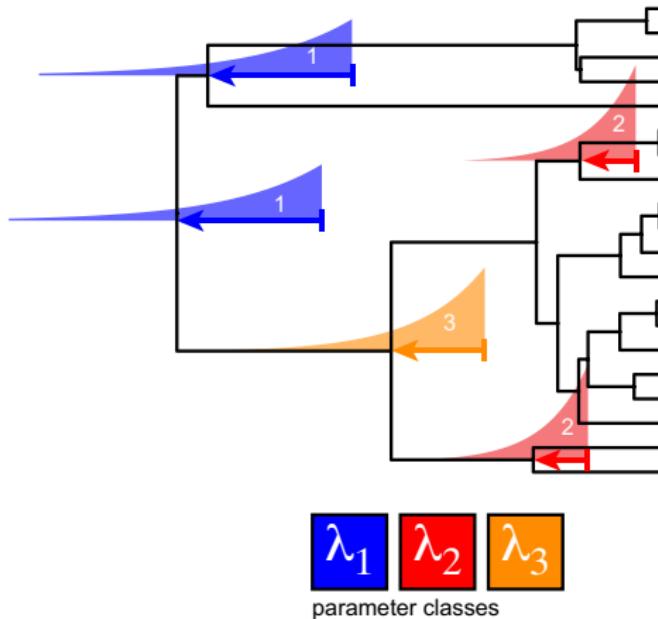


HYPERTPRIOR ON CALIBRATED NODES

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

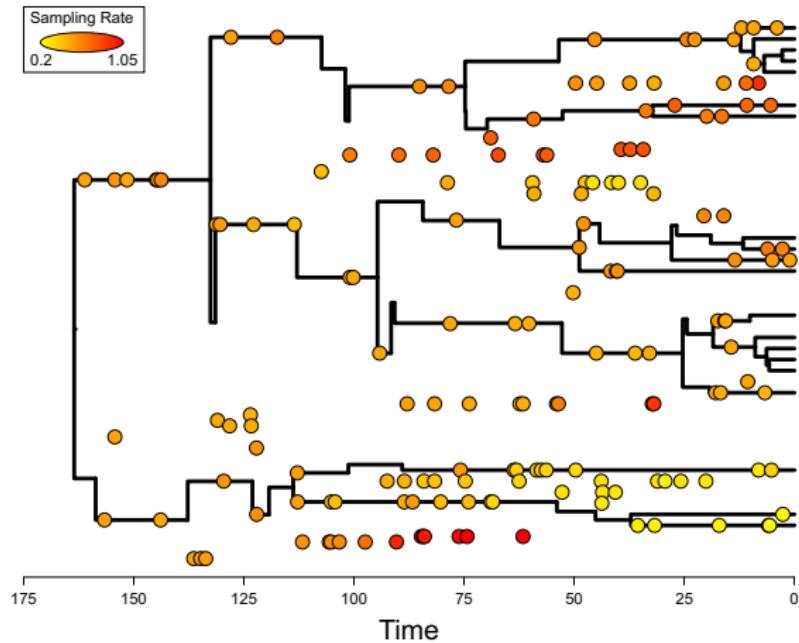
Sample the time from the MRCA to the fossil from a mixture of different exponential distributions

Account for uncertainty in values of λ



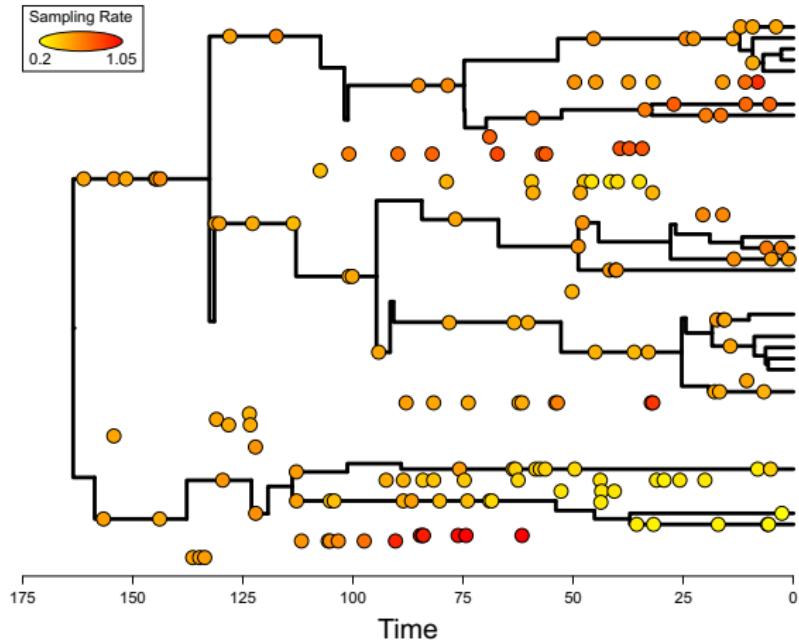
COMPLEX MODELS OF MACROEVOLUTION

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



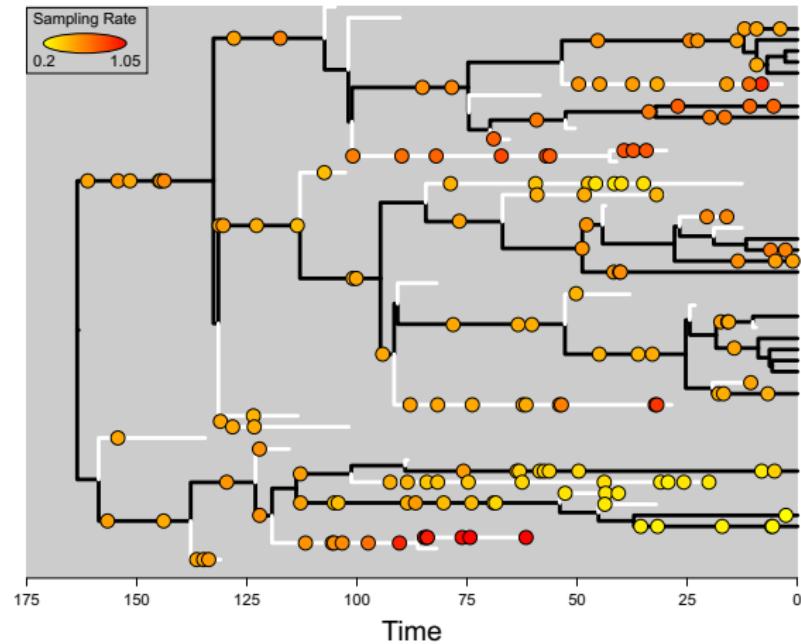
COMPLEX MODELS OF MACROEVOLUTION

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



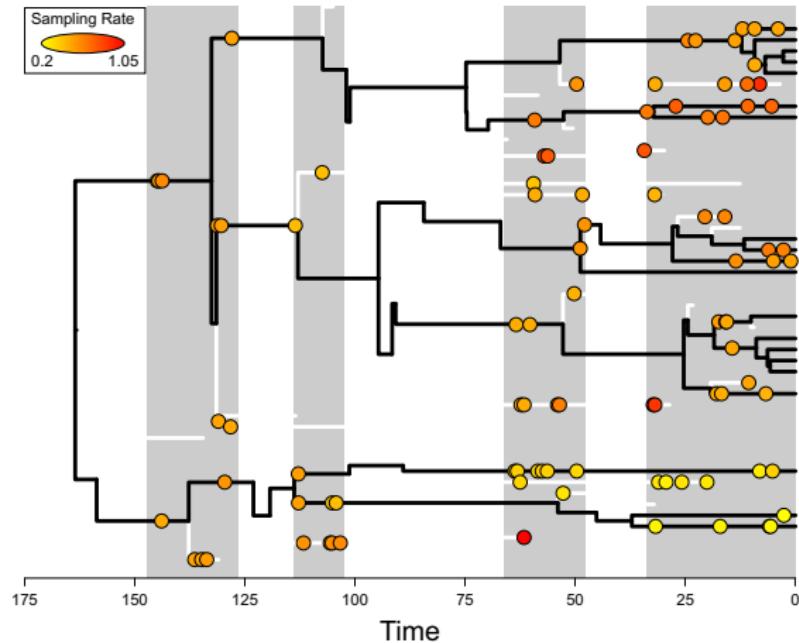
CALIBRATING THE DIVERSIFICATION MODEL

Fossils provide observations of the complex diversification process back in time



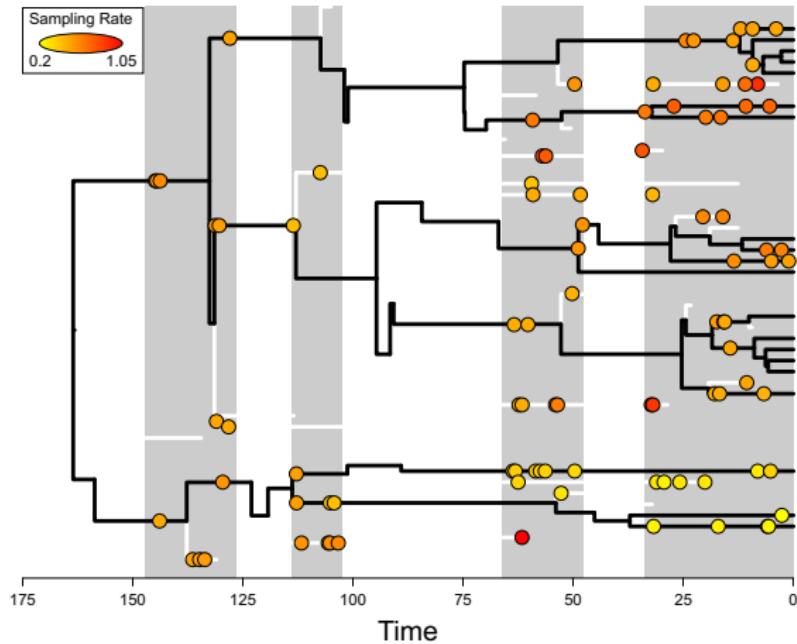
CALIBRATING THE DIVERSIFICATION MODEL

The rock record leads to sporadic observations of lineages through time



CALIBRATING THE DIVERSIFICATION MODEL

We can incorporate aspects of taphonomy into the tree model to calibrate the tree without calibration densities



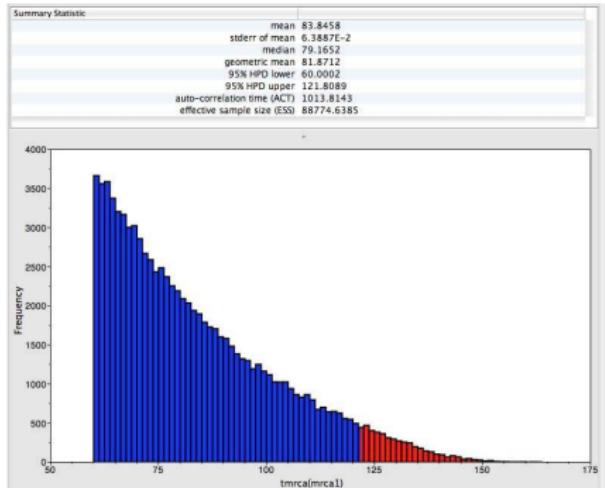
CRITICISM OF RELAXED CLOCK METHODS

- 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

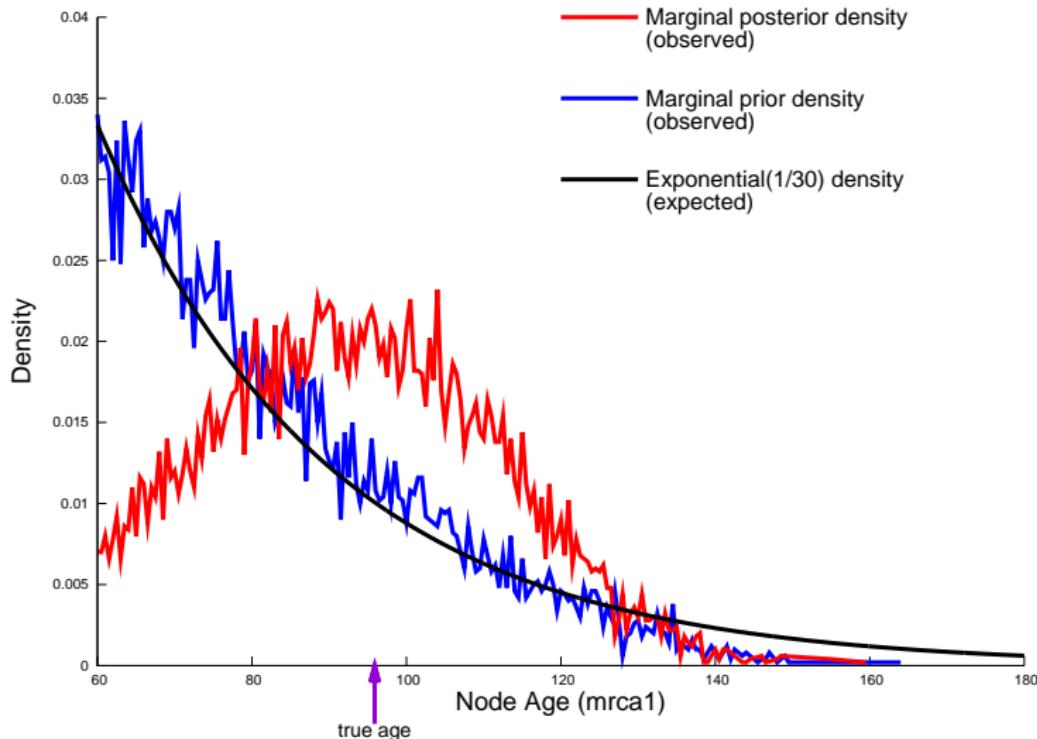
MCMC UNDER THE PRIOR

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

Allows you to assess your prior specification and examine prior sensitivity

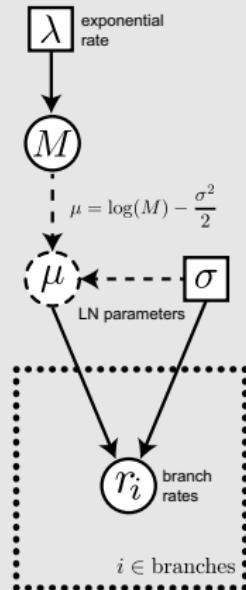


MCMC UNDER THE PRIOR

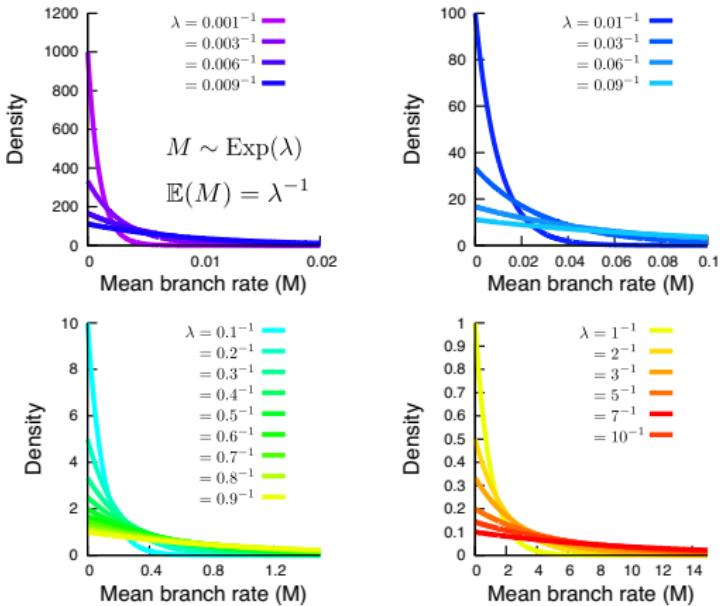


SENSITIVITY TO THE PRIOR

a) Graphical Model

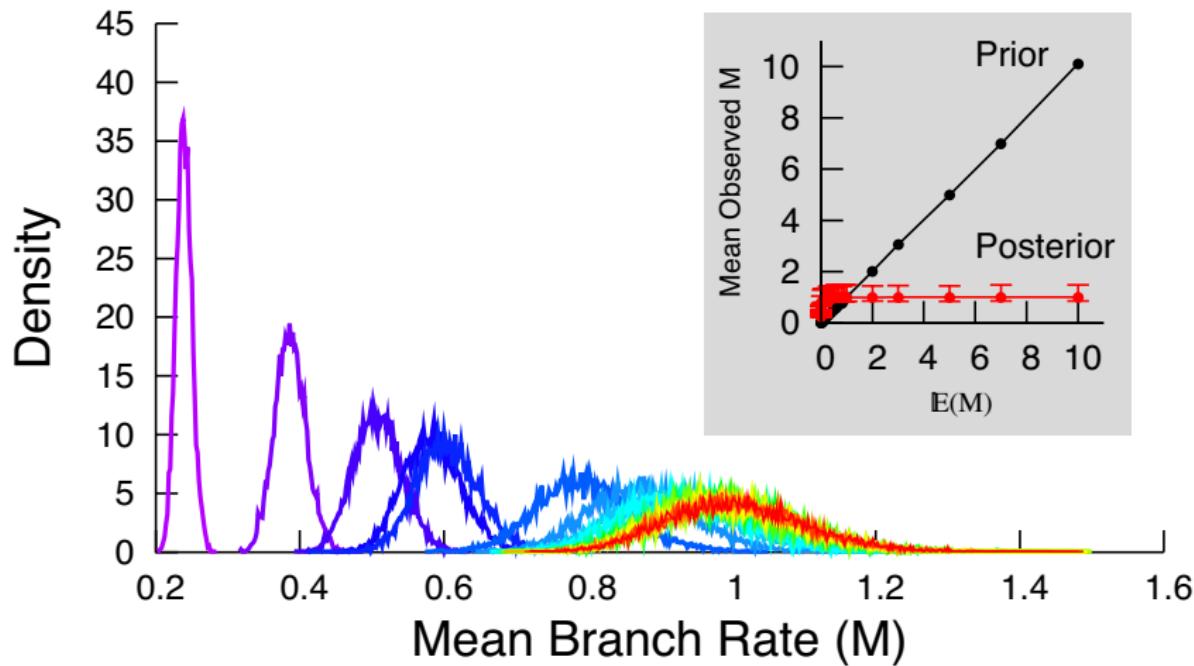


b) Exponential hyperprior densities on expected rate (M)

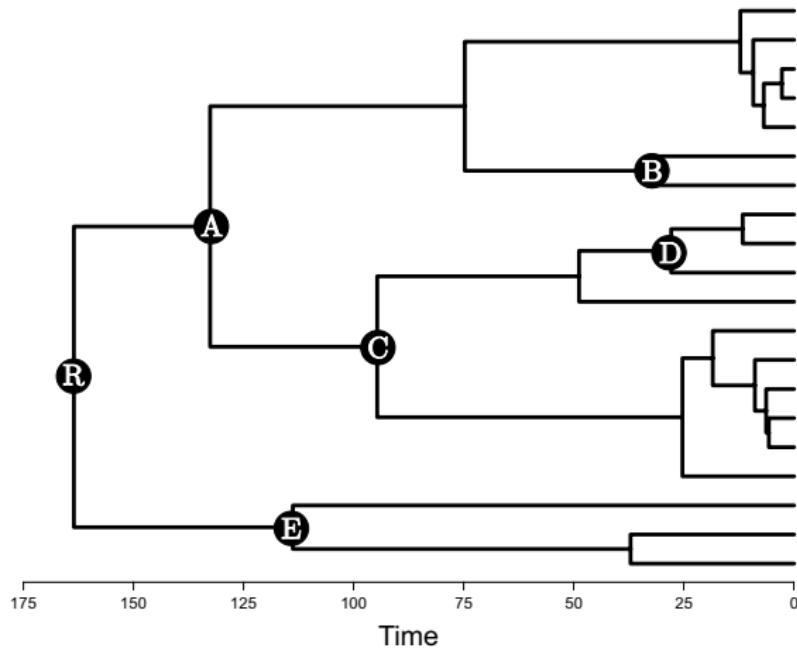


SENSITIVITY TO THE PRIOR

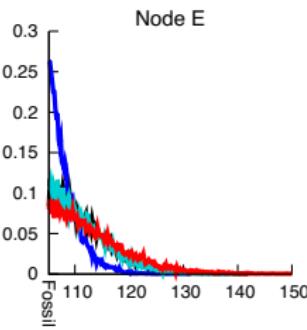
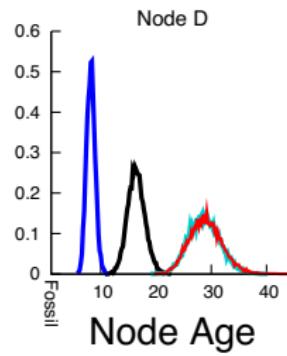
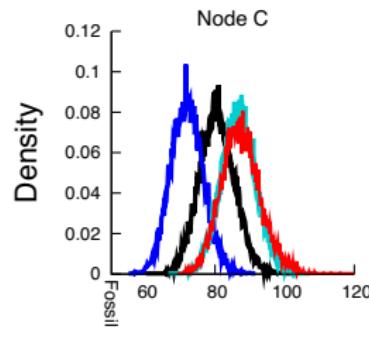
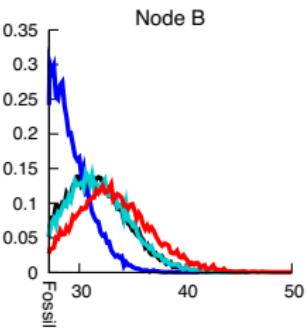
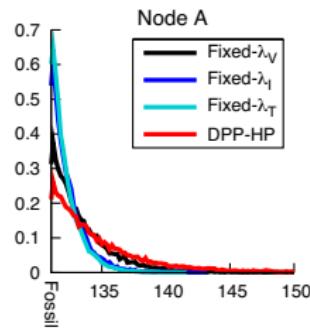
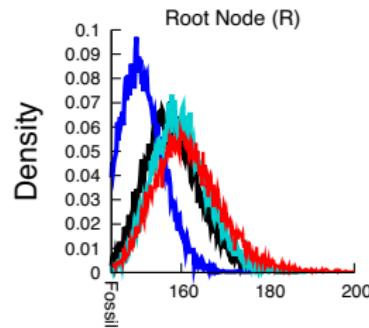
Marginal posterior densities of mean branch rate



CALIBRATED NODE AGE ESTIMATES



SENSITIVITY TO THE CALIBRATION PRIOR



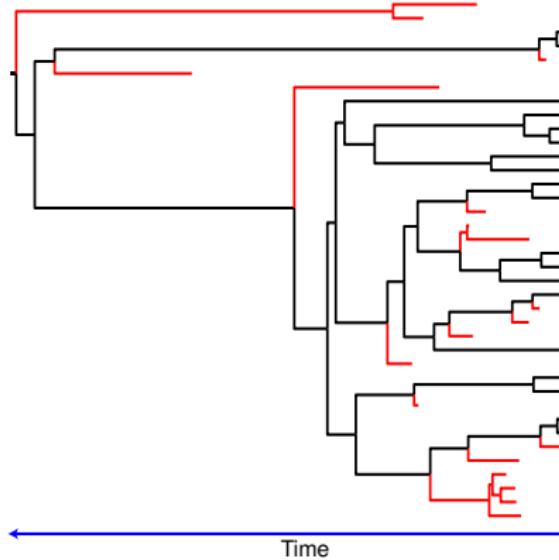
Node Age

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

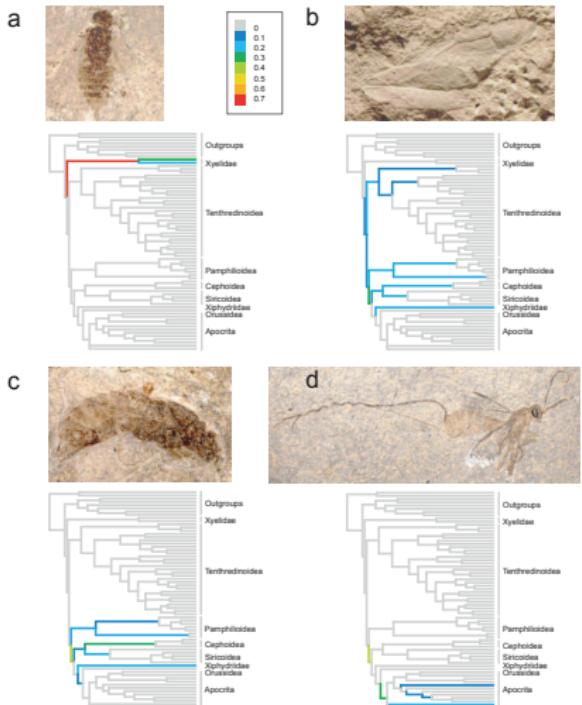


FOSSIL TIP DATING

Fredrik Ronquist and his colleagues implemented tip dating in MrBayes

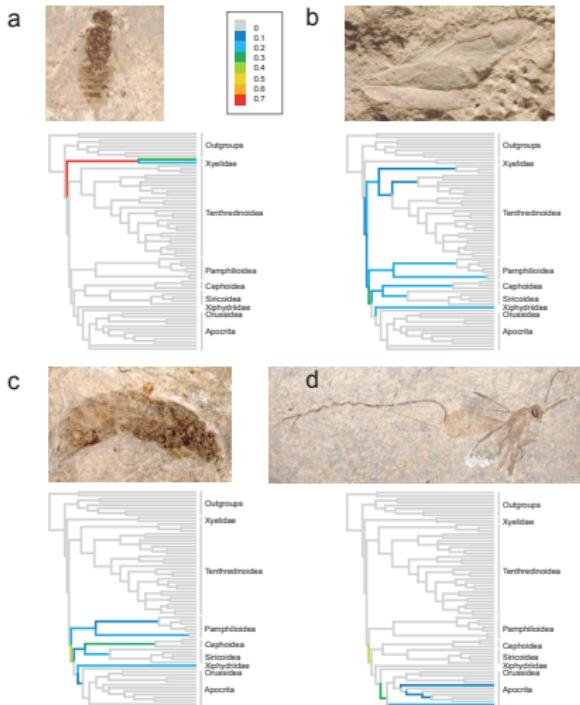
Early radiation of Hymenoptera

- 66 extant taxa
- 45 fossil taxa
- 7 genes, $\sim 5\text{kB}$ (extant taxa only)
- 343 morphological characters (12% complete for fossils)



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

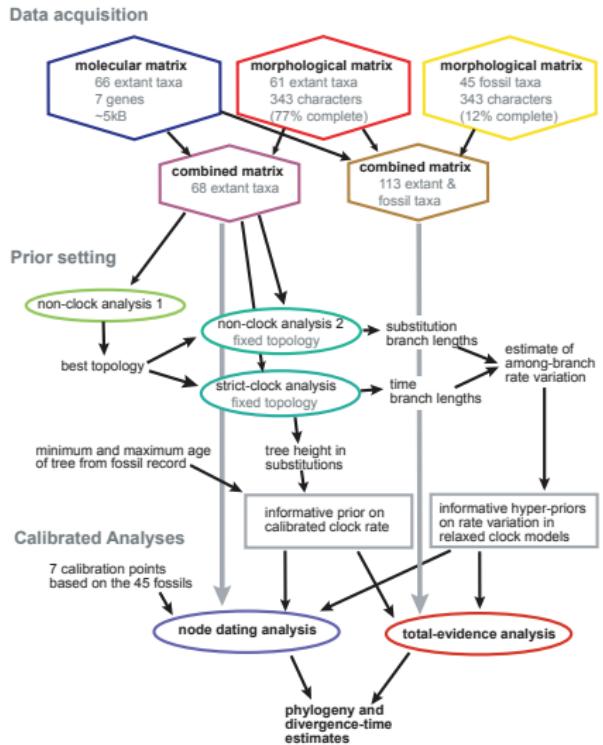


Fossil Tip Dating

Thorough analysis is necessary for this kind of dataset

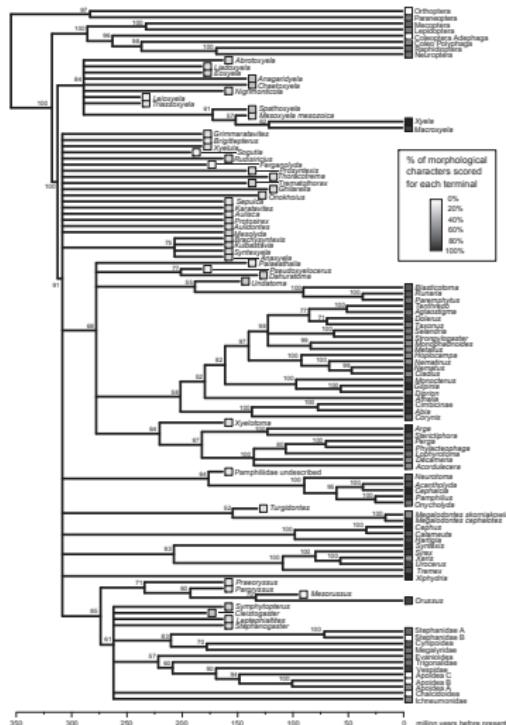
Ronquist et al. used Bayes factors to choose a relaxed clock model (this is rarely done, but really important)

Compared node dating and tip dating



Fossil Tip Dating

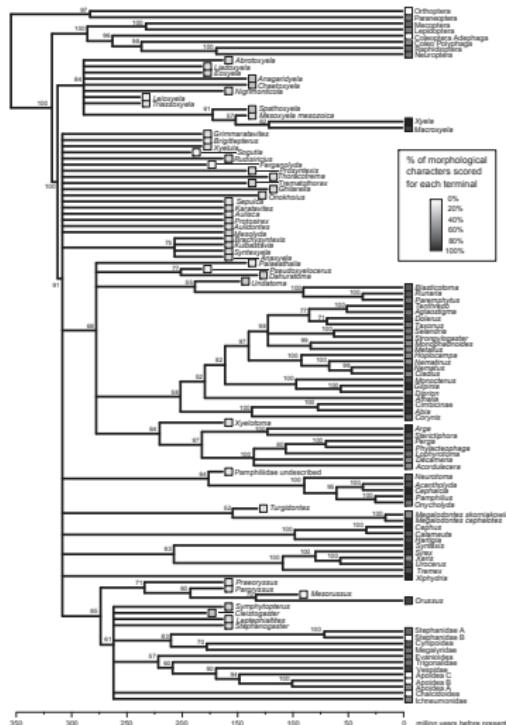
- Resulted in a fairly unresolved phylogeny, but fossils significantly contribute to estimates of node ages
 - Posteriors on node times are less sensitive to priors compared with node dating
 - Higher precision for divergence time estimates



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

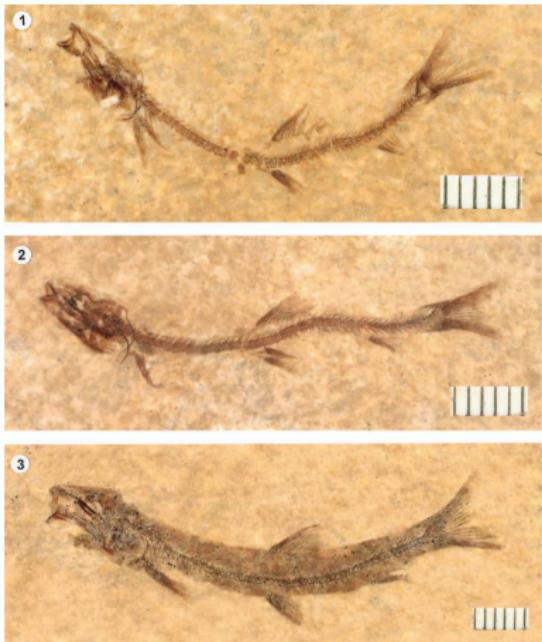


Fossil Tip Dating

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

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

