

OUTLINE

Bayesian Fundamentals

- Bayes theorem, priors and posteriors, MCMC
- Example: the birth-death process

Bayesian Inference of Species Divergence Times

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

BAYESIAN INFERENCE

Estimate the **probability** of a hypothesis (model) conditional on observed data.

The probability represents the researcher's **degree of belief**.

Bayes' Theorem specifies the conditional probability of the hypothesis given the data.

BAYES' THEOREM

$$\Pr(\text{Model} \mid \text{Data}) = \frac{\Pr(\text{Data} \mid \text{Model}) \Pr(\text{Model})}{\Pr(\text{Data})}$$

BAYES' THEOREM

posterior probability

$$\Pr(\text{Model} \mid \text{Data}) = \frac{\Pr(\text{Data} \mid \text{Model}) \Pr(\text{Model})}{\Pr(\text{Data})}$$

BAYES' THEOREM

$$\Pr(\text{Model} \mid \text{Data}) = \frac{\Pr(\text{Data} \mid \text{Model}) \Pr(\text{Model})}{\Pr(\text{Data})}$$

↓
likelihood

BAYES' THEOREM

$$\Pr(\text{Model} \mid \text{Data}) = \frac{\Pr(\text{Data} \mid \text{Model}) \Pr(\text{Model})}{\Pr(\text{Data})}$$

↑
prior probability

BAYES' THEOREM

$$\Pr(\text{Model} \mid \text{Data}) = \frac{\Pr(\text{Data} \mid \text{Model}) \Pr(\text{Model})}{\Pr(\text{Data})}$$


↑
marginal probability of the data

BAYES' THEOREM

The posterior probability of a discrete parameter δ conditional on the data D is

$$\Pr(\delta | D) = \frac{\Pr(D | \delta) \Pr(\delta)}{\sum_{\delta} \Pr(D | \delta) \Pr(\delta)}$$

$\sum_{\delta} \Pr(D | \delta) \Pr(\delta)$ is the likelihood **marginalized** over all possible values of δ .

BAYES' THEOREM

The posterior probability **density** a continuous parameter θ conditional on the data D is

$$f(\theta | D) = \frac{f(D | \theta)f(\theta)}{\int_{\theta} f(D | \theta)f(\theta)d\theta}$$

$\int_{\theta} f(D | \theta)f(\theta)d\theta$ is the likelihood **marginalized** over all possible values of θ .

PRIORS

The distribution of θ before any data are collected is the prior

$$f(\theta)$$

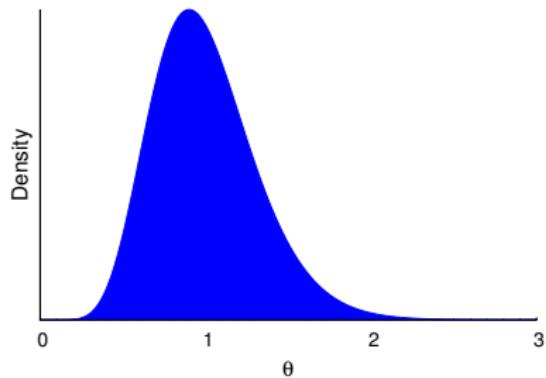
The prior describes your uncertainty in the parameters of your model.

PRIORS

We may assume a gamma-prior distribution on θ with a shape parameter α and a scale parameter β .

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

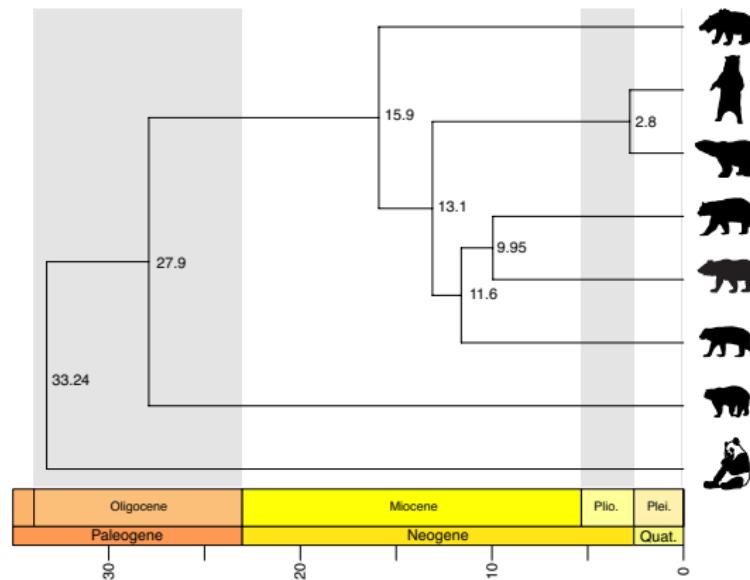
$$f(\theta | \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} \theta^{\alpha-1} e^{-\frac{\theta}{\beta}}$$



This requires us to assign values for α and β based on our **prior belief**, or we can place **hyperpriors** on these parameters if we are uncertain about their values.

EXAMPLE: THE BIRTH-DEATH PROCESS

A time machine allowed us to observe the dates of each speciation event in the history of extant bears

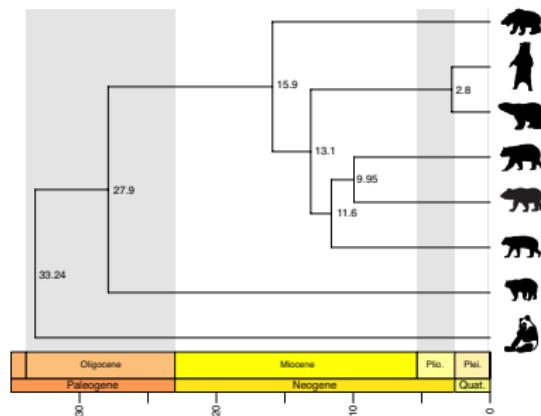


EXAMPLE: THE BIRTH-DEATH PROCESS

We assume that the diversification of bears matches a birth death process with parameters:

λ = speciation rate

μ = extinction rate



EXAMPLE: THE BIRTH-DEATH PROCESS

The birth-death process allows us to compute the probability density of our observed time-tree (Ψ) conditional on any value of speciation (λ) and any value of extinction (μ).

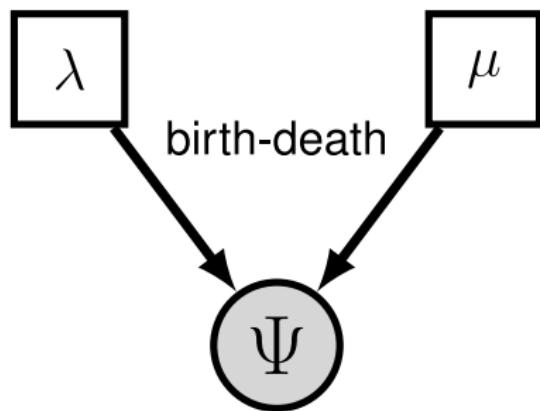
$$f(\Psi | \lambda, \mu)$$

This model states that $\Psi \sim \text{BD}(\lambda, \mu)$

EXAMPLE: THE BIRTH-DEATH PROCESS

Another way of expressing $\Psi \sim BD(\lambda, \mu)$ is with a **probabilistic graphical model**

Our “observed” time-tree is conditioned on some constant value of λ and μ .

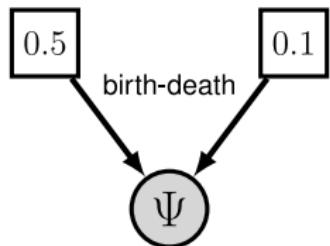


EXAMPLE: THE BIRTH-DEATH PROCESS

If our time machine also allowed us to **know** the rate of speciation and extinction, we can easily calculate the likelihood of our observed tree:

$$f(\Psi | \lambda, \mu) = N! (\lambda - \mu) \lambda^{N-1} \frac{e^{-(\lambda - \mu)x_1}}{\lambda - \mu e^{-(\lambda - \mu)x_1}} \prod_{i=1}^{N-1} \frac{(\lambda - \mu)^2 e^{-(\lambda - \mu)x_i}}{(\lambda - \mu e^{-(\lambda - \mu)x_i})^2}$$

$$f(\Psi | \lambda = 0.5, \mu = 0.1) = 3.49215e-32$$



EXAMPLE: THE BIRTH-DEATH PROCESS

What if we **do not know** λ and μ ?

We can use frequentist or Bayesian methods for estimating their values.

Frequentist methods require us to find the values of λ and μ that **maximize** $f(\Psi | \lambda, \mu)$.

Bayesian methods use **prior distributions** to describe our uncertainty in λ and μ and estimate $f(\lambda, \mu | \Psi)$.

EXAMPLE: THE BIRTH-DEATH PROCESS

We must define prior distributions for λ and μ to estimate the posterior probability density

$$f(\lambda, \mu | \psi, y, z) = \frac{f(\psi | \lambda, \mu)f(\lambda | y)f(\mu | z)}{\int_{\lambda} \int_{\mu} f(\psi | \lambda, \mu)f(\lambda | y)f(\mu | z)d\lambda d\mu}$$

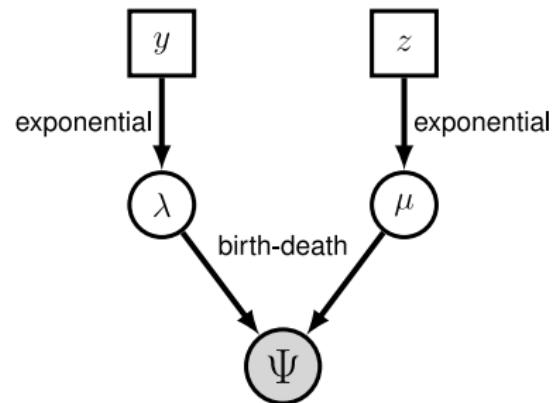
Now y and z are the parameters of the prior distributions on λ and μ .

EXAMPLE: THE BIRTH-DEATH PROCESS

We can choose **exponential** prior distributions for λ and μ .
Now y and z represent the rate parameters of the exponential priors.

$$\begin{aligned}\lambda &\sim \text{Exponential}(y) \\ \mu &\sim \text{Exponential}(z)\end{aligned}$$

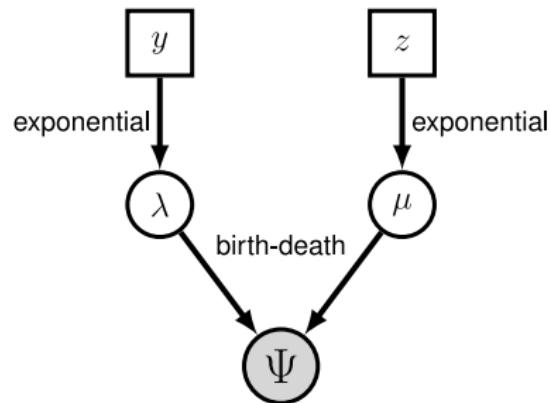
$$\begin{aligned}f(\lambda | y) &= ye^{-y\lambda} \\ f(\mu | z) &= ze^{-z\mu}\end{aligned}$$



EXAMPLE: THE BIRTH-DEATH PROCESS

Now that we have a defined model, how do we estimate the posterior probability density?

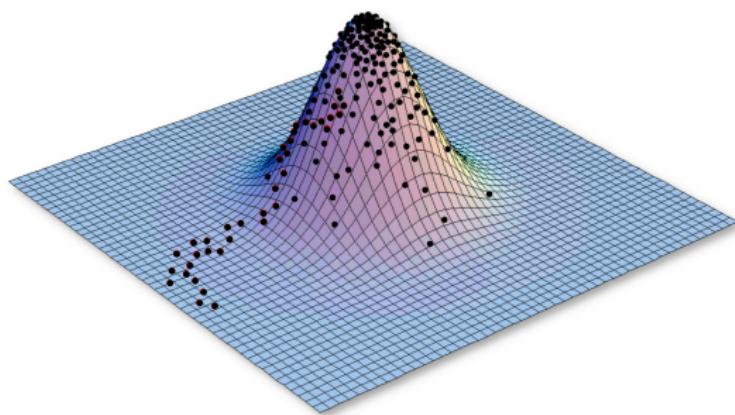
$$\begin{aligned}\lambda &\sim \text{Exponential}(y) \\ \mu &\sim \text{Exponential}(z)\end{aligned}$$



$$f(\lambda, \mu | \Psi, y, z) = \frac{f(\Psi | \lambda, \mu) f(\lambda | y) f(\mu | z)}{\int_{\lambda} \int_{\mu} f(\Psi | \lambda, \mu) f(\lambda | y) f(\mu | z) d\lambda d\mu}$$

MARKOV CHAIN MONTE CARLO (MCMC)

An algorithm for approximating the posterior distribution



Metropolis, Rosenbluth, Rosenbluth, Teller, Teller. 1953. Equations of state calculations by fast computing machines. *J. Chem. Phys.*

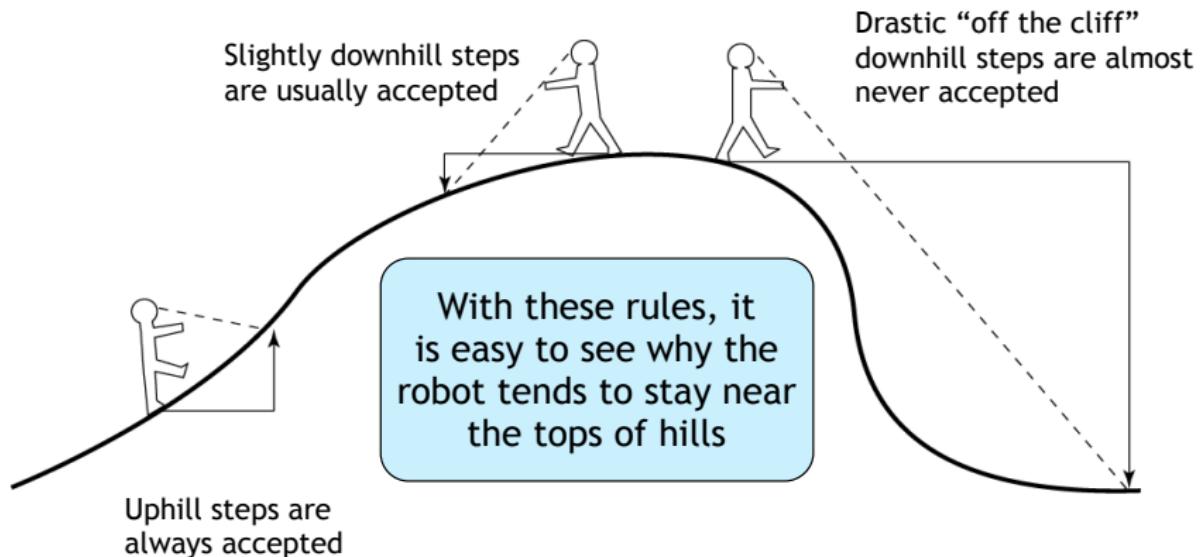
Hastings. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*.

MARKOV CHAIN MONTE CARLO (MCMC)

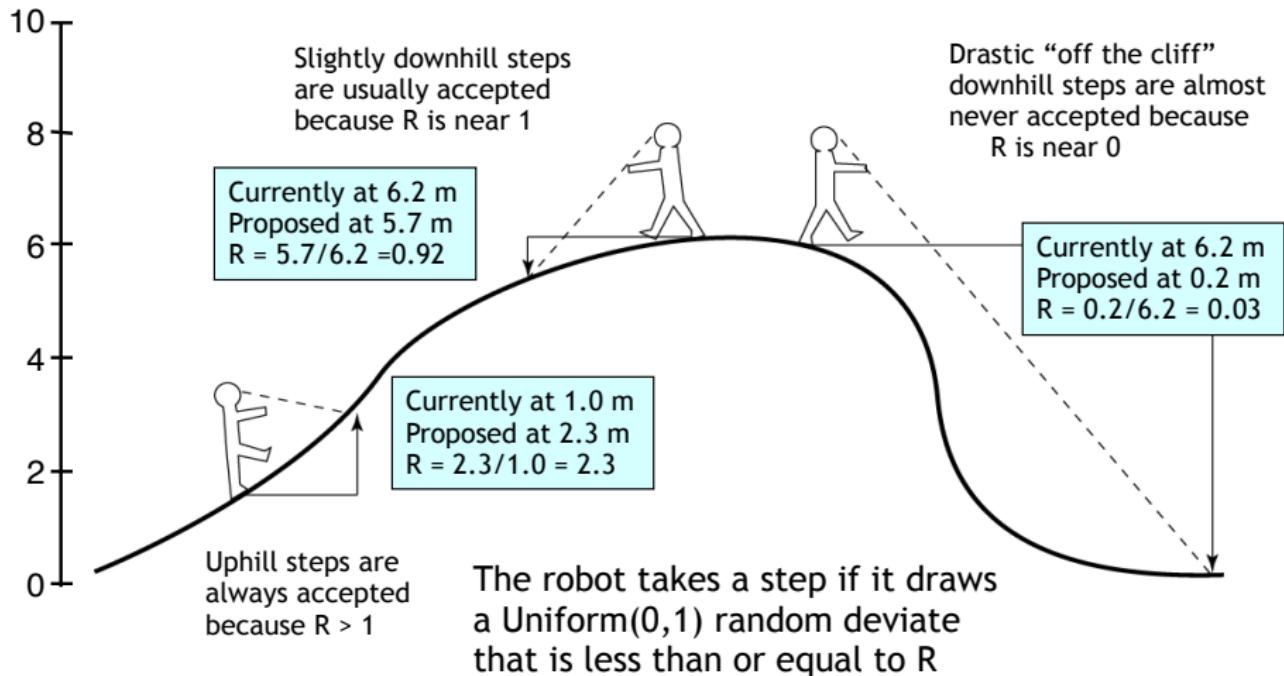
More on MCMC from Paul Lewis and his lecture on Bayesian phylogenetics

Slides source: https://molevol.mbl.edu/index.php/Paul_Lewis

MCMC robot's rules



(Actual) MCMC robot rules



Cancellation of marginal likelihood

When calculating the ratio R of posterior densities, the marginal probability of the data cancels.

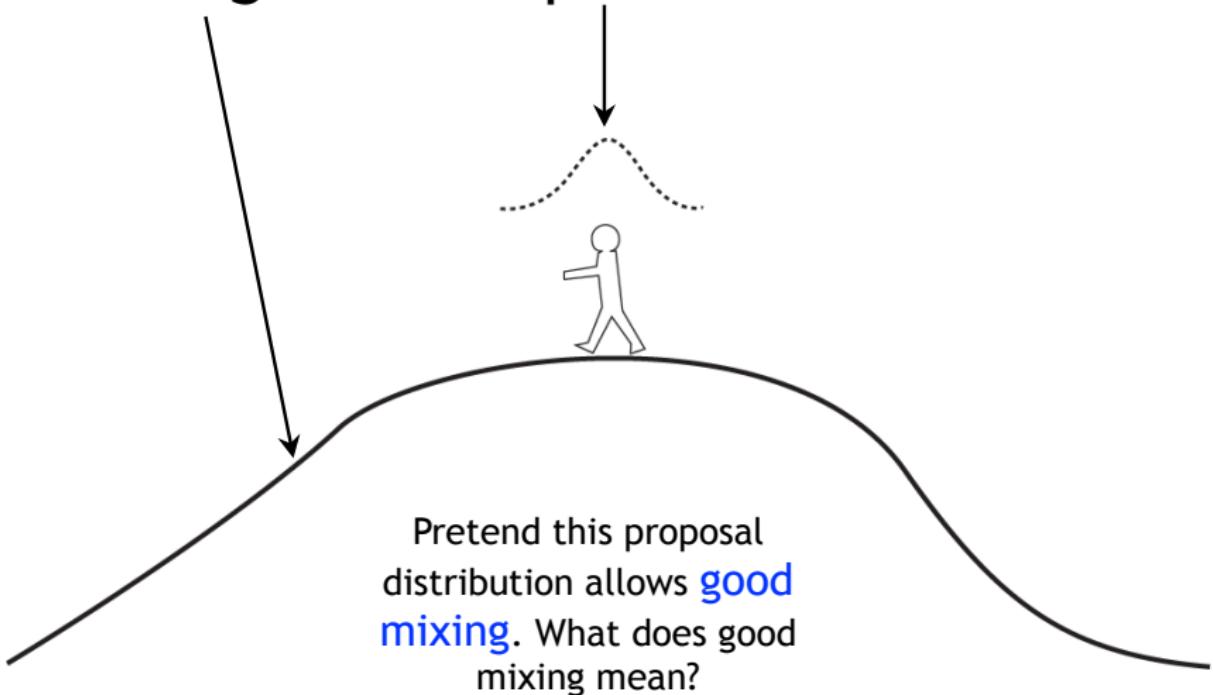
$$\frac{f(\theta^*|D)}{f(\theta|D)} = \frac{\frac{f(D|\theta^*)f(\theta^*)}{\cancel{f(D)}}}{\frac{f(D|\theta)f(\theta)}{\cancel{f(D)}}} = \frac{f(D|\theta^*)f(\theta^*)}{f(D|\theta)f(\theta)}$$

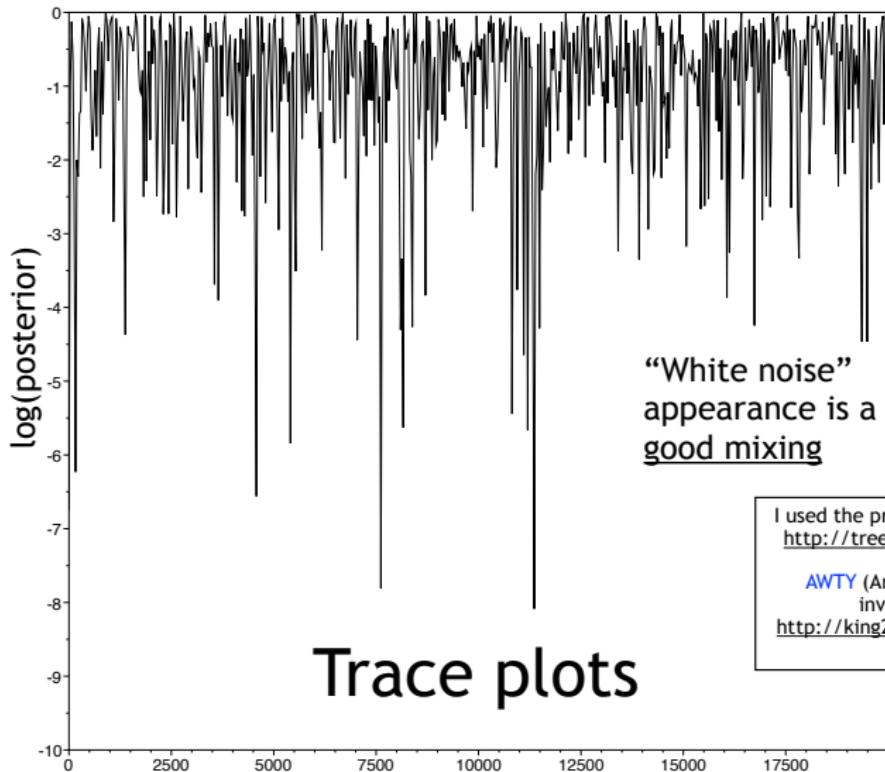
Posterior
odds

Likelihood
ratio

Prior odds

Target vs. Proposal Distributions





I used the program **Tracer** to create this plot:
<http://tree.bio.ed.ac.uk/software/tracer/>

AWTY (Are We There Yet?) is useful for
investigating convergence:
http://king2.scs.fsu.edu/CEBProjects/awty/_awty_start.php

Trace plots

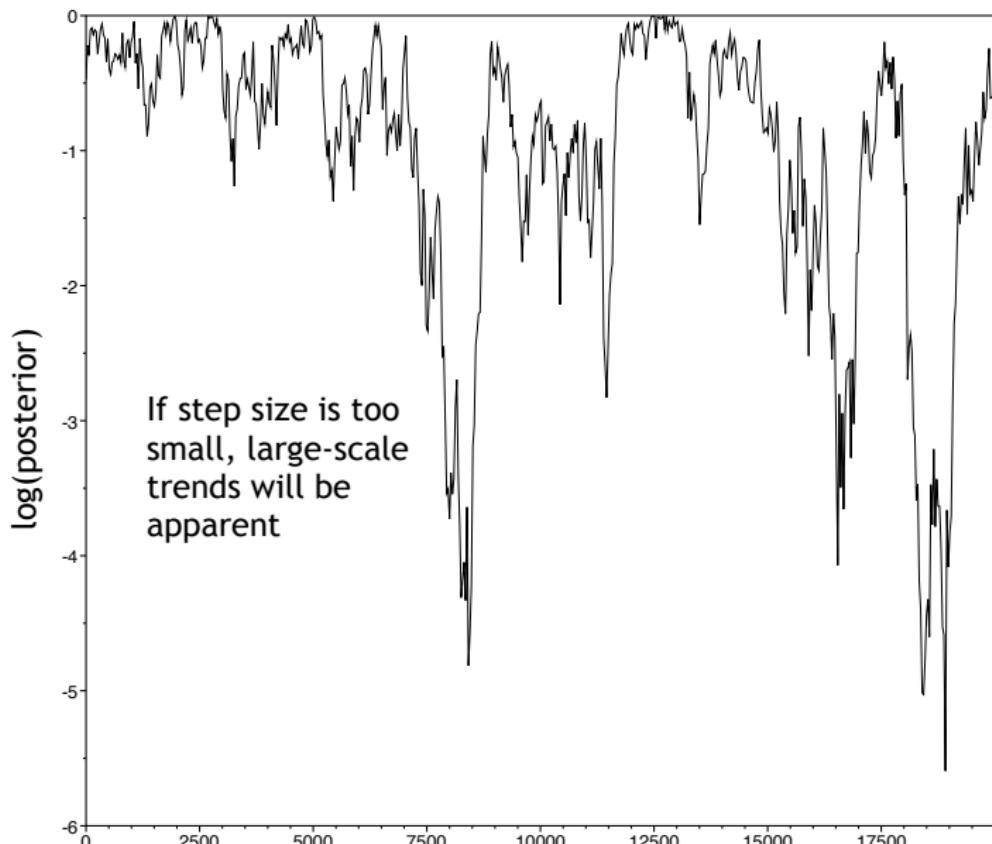
Target vs. Proposal Distributions

Proposal distributions
with **smaller variance**...



Disadvantage: robot takes
smaller steps, more time
required to explore the
same area

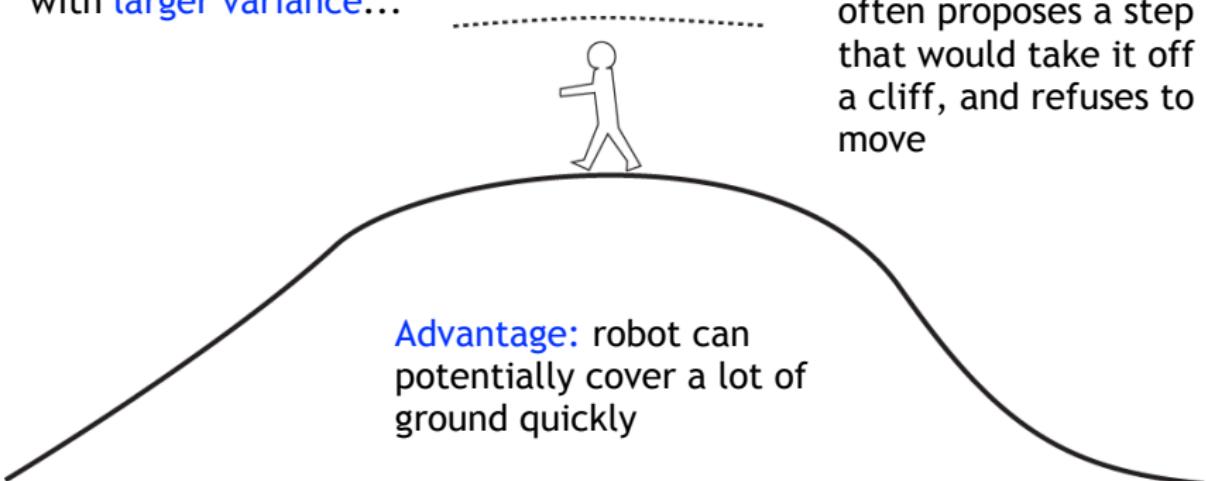
Advantage: robot seldom
refuses to take proposed
steps

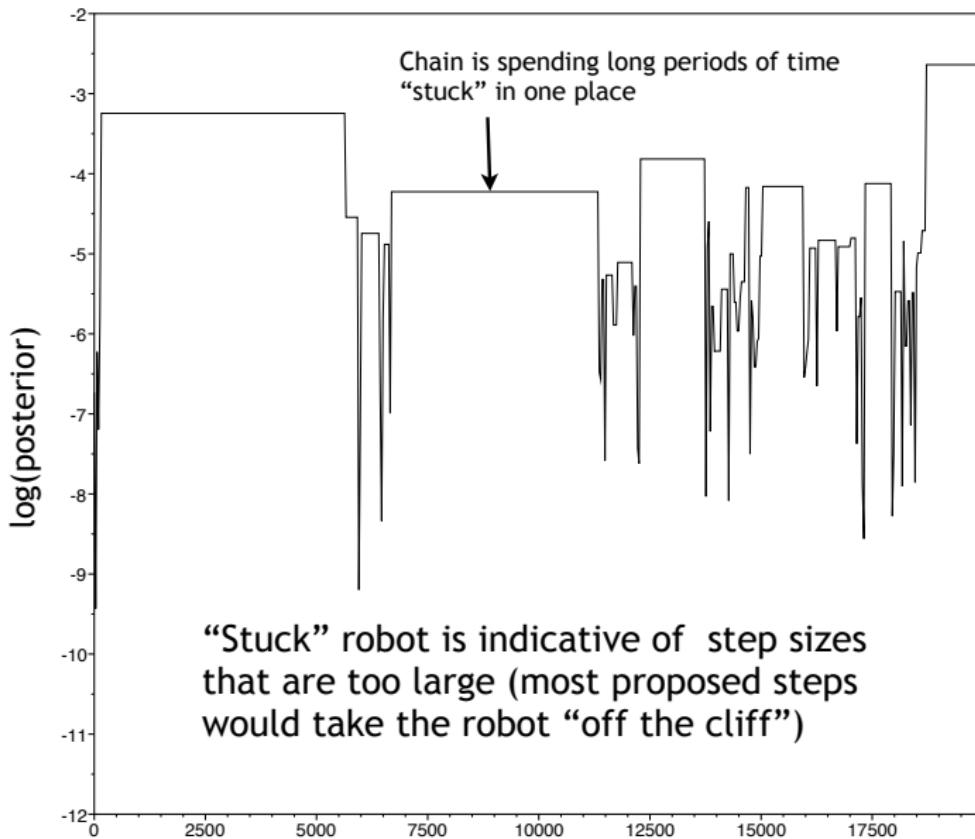


Target vs. Proposal Distributions

Proposal distributions
with **larger variance**...

Disadvantage: robot
often proposes a step
that would take it off
a cliff, and refuses to
move





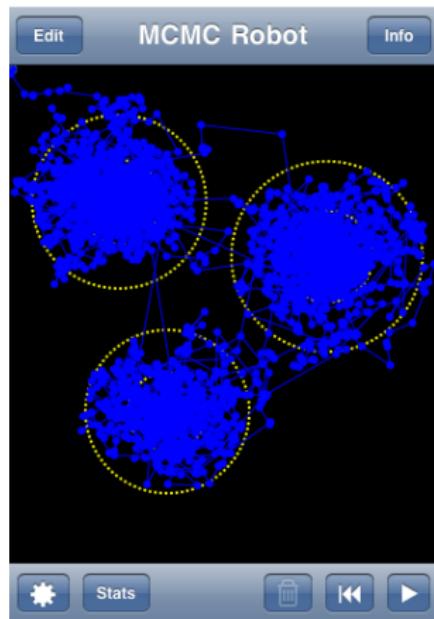
MARKOV CHAIN MONTE CARLO (MCMC)

Thanks, Paul!

Slides source: https://molevol.mbl.edu/index.php/Paul_Lewis

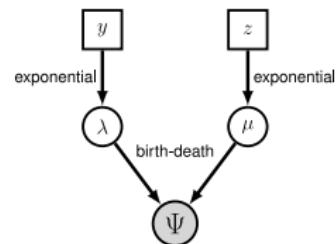
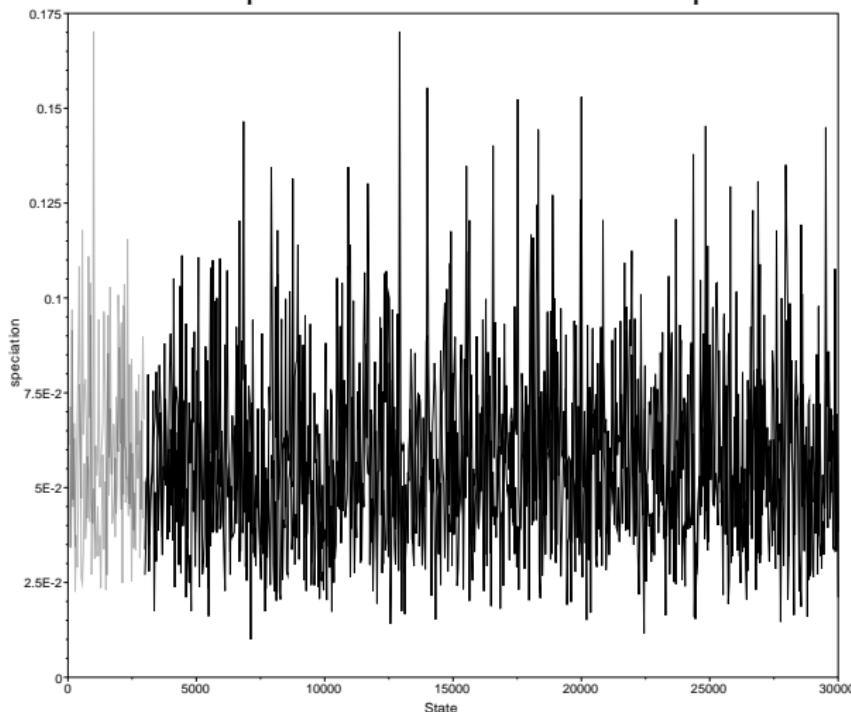
See MCMCRobot, a helpful software program for learning MCMC by Paul Lewis

<http://www.mcmcrobot.org>



EXAMPLE: THE BIRTH-DEATH PROCESS

The trace-plot of the MCMC samples for speciation rate

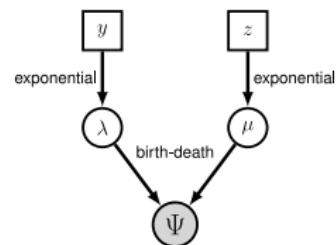
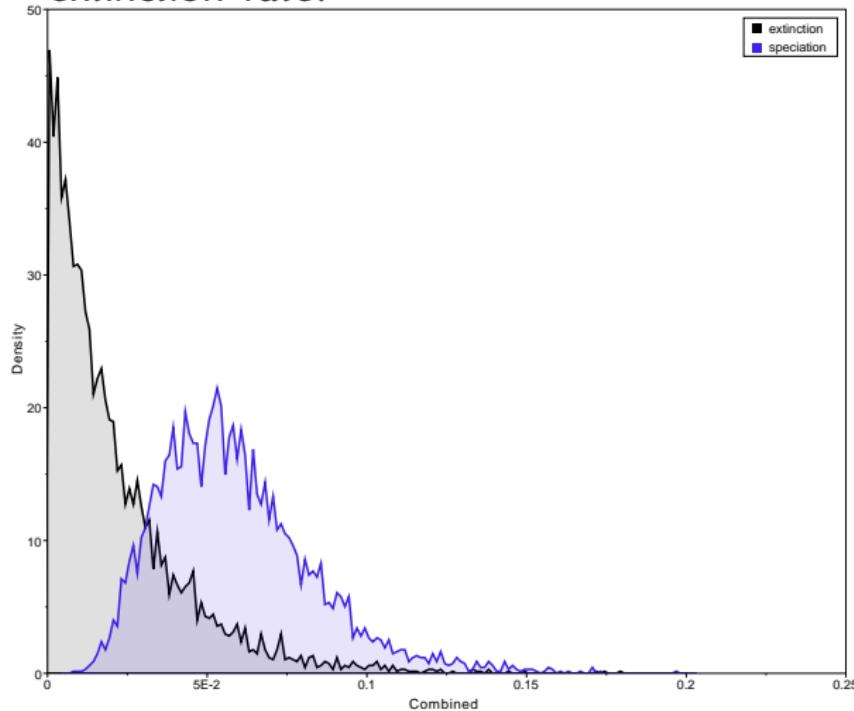


$$y = 1.0$$

$$z = 10.0$$

EXAMPLE: THE BIRTH-DEATH PROCESS

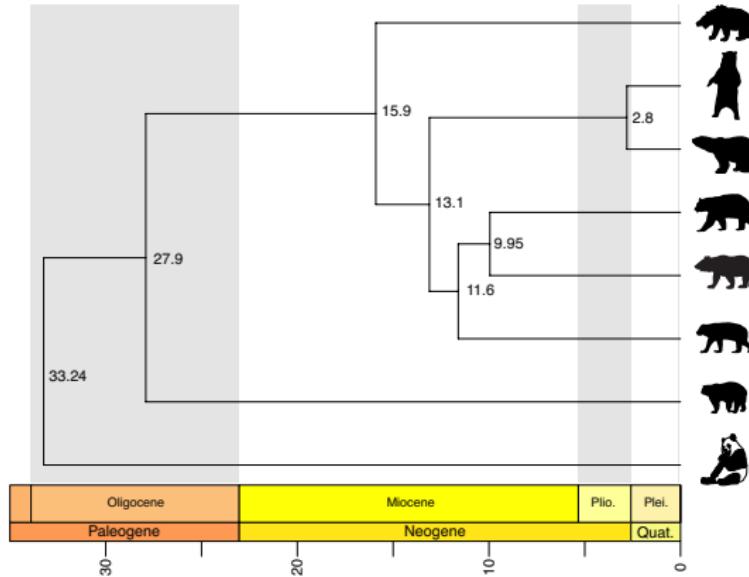
Marginal posterior densities of the speciation rate and extinction rate.



$$y = 1.0$$
$$z = 10.0$$

EXAMPLE: THE BIRTH-DEATH PROCESS

Alas, we do not have a time machine and we **do not know** lineage divergence times without error.



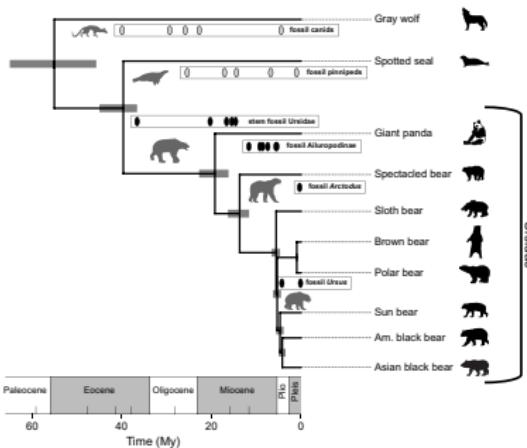
DIVERGENCE TIME ESTIMATION

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

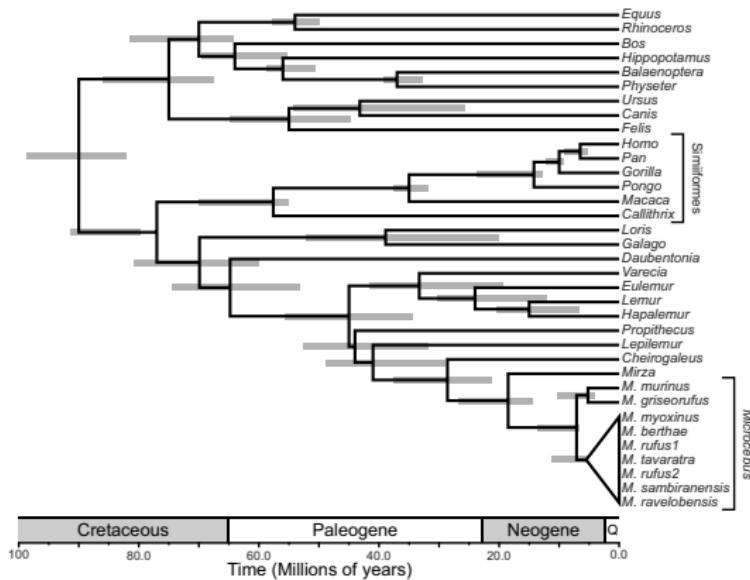
External calibration information for estimates of absolute node times



(Figure from Heath et al., PNAS 2014)

A TIME-SCALE FOR EVOLUTION

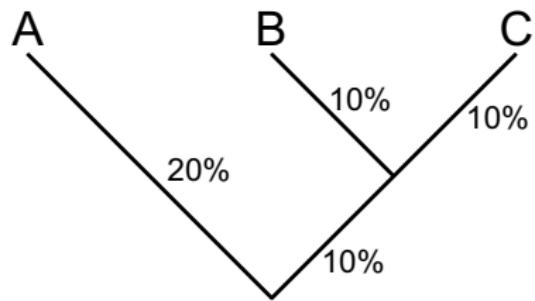
Phylogenetic trees can provide both topological information and temporal information



THE GLOBAL MOLECULAR CLOCK

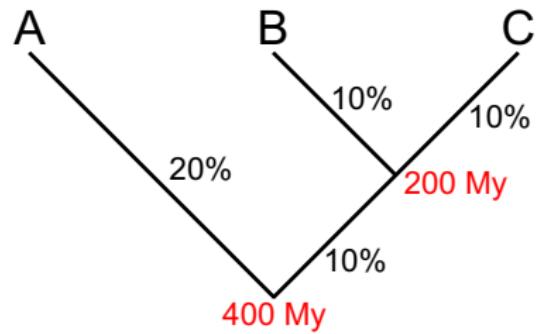
Assume that the rate of evolutionary change is constant over time

(branch lengths equal percent sequence divergence)



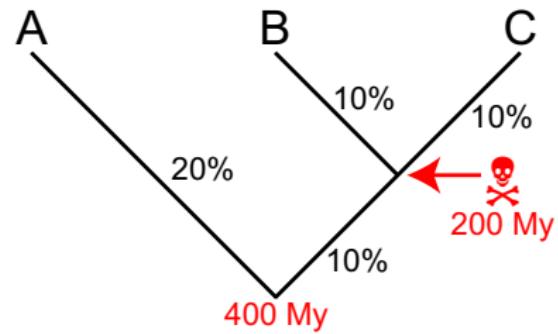
THE GLOBAL MOLECULAR CLOCK

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



THE GLOBAL MOLECULAR CLOCK

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



REJECTING THE GLOBAL MOLECULAR CLOCK

Rates of evolution vary across lineages and over time

Mutation rate:

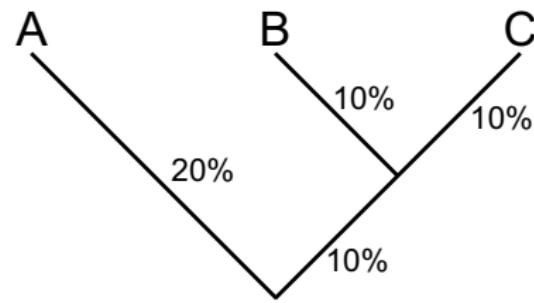
Variation in

- metabolic rate
- generation time
- DNA repair

Fixation rate:

Variation in

- strength and targets of selection
- population sizes

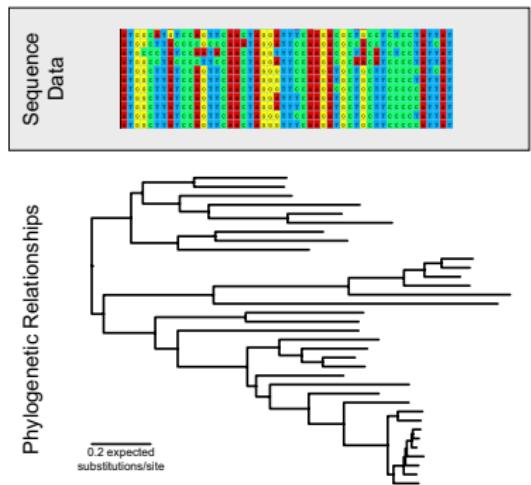


UNCONSTRAINED ANALYSIS

Sequence data provide information about **branch lengths**

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

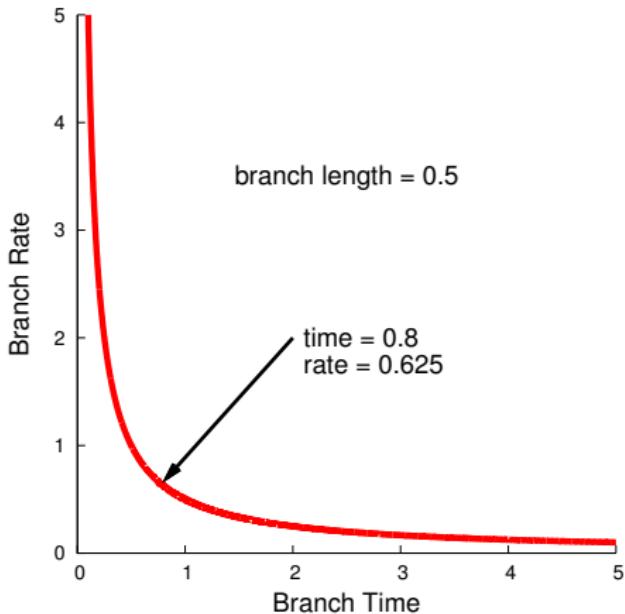
$$\text{branch length} = \text{rate} \times \text{time}$$



RATE AND TIME

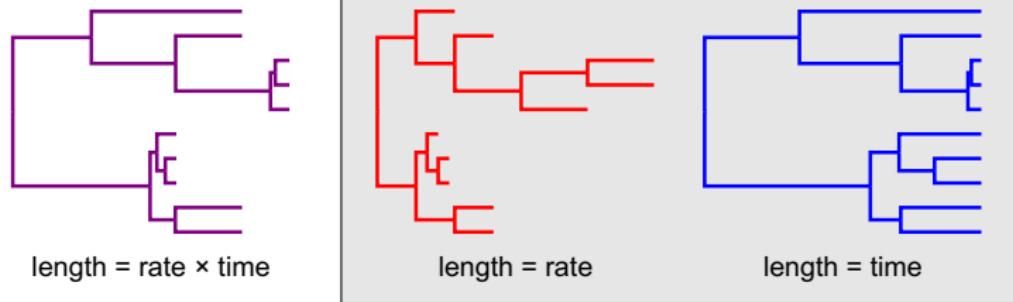
The sequence data provide information about branch length

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



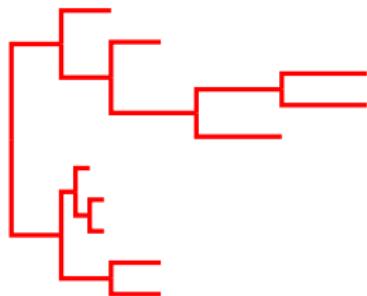
RATE AND TIME

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

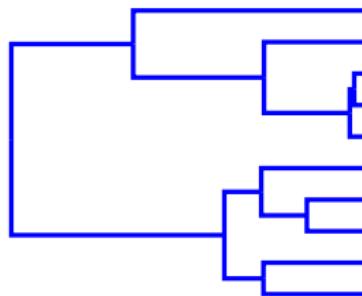


Methods for dating species divergences estimate the **substitution rate** and **time** separately

BAYESIAN DIVERGENCE TIME ESTIMATION



length = rate



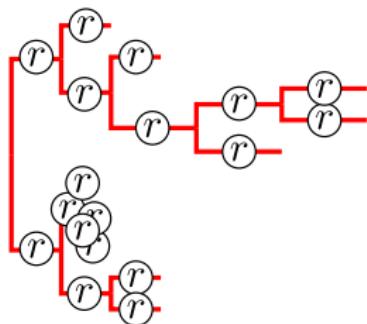
length = time

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

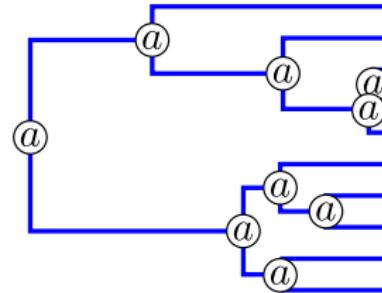
$$\mathcal{A} = (a_1, a_2, a_3, \dots, a_{N-1})$$

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

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

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

\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

Ψ Tree topology

BAYESIAN DIVERGENCE TIME ESTIMATION

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

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

$f(D \mid \mathcal{R}, \mathcal{A}, \theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s)$	Likelihood
$f(\mathcal{R} \mid \theta_{\mathcal{R}})$	Prior on rates
$f(\mathcal{A} \mid \theta_{\mathcal{A}})$	Prior on node ages
$f(\theta_s)$	Prior on substitution parameters
$f(D)$	Marginal probability of the data

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

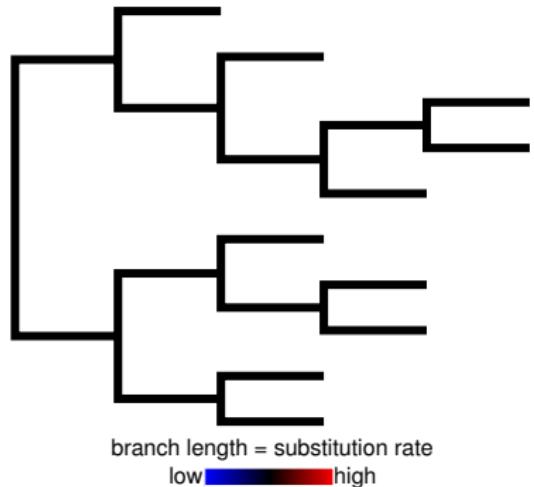
Some models describing lineage-specific substitution rate variation:

- **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)
- **Punctuated rate change 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



RELAXED-CLOCK MODELS

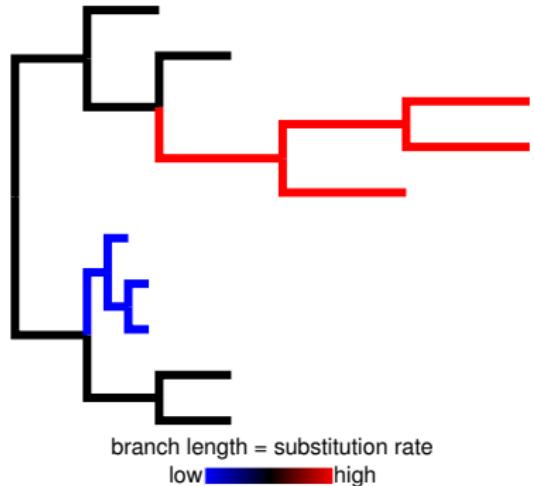
To accommodate variation in substitution rates
'relaxed-clock' models estimate lineage-specific substitution rates

- Local molecular clocks
- Punctuated rate change model
- Log-normally distributed autocorrelated rates
- Uncorrelated/independent rates models
- Mixture models on branch rates

LOCAL MOLECULAR CLOCKS

Rate shifts occur infrequently over the tree

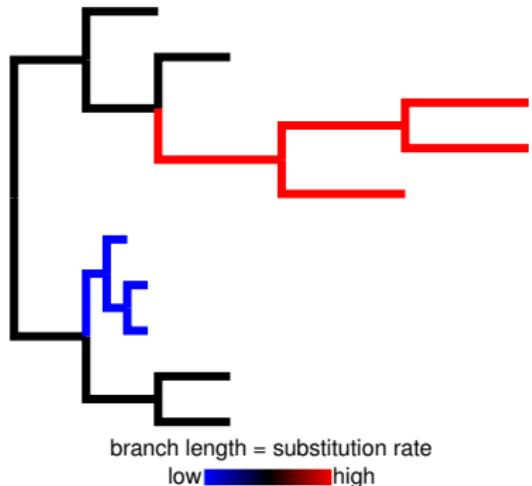
Closely related lineages have equivalent rates (clustered by sub-clades)



LOCAL MOLECULAR CLOCKS

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

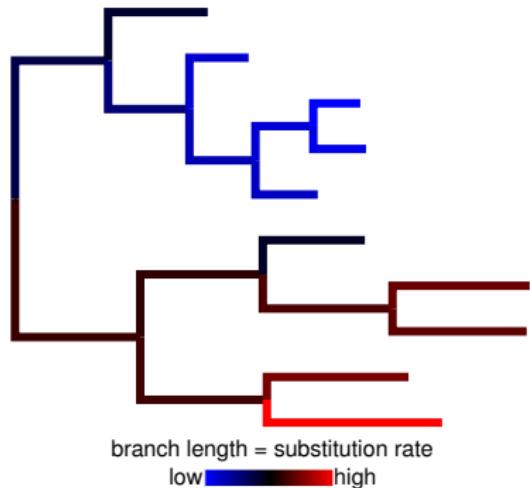
Drummond and Suchard (2010) introduced a Bayesian method that samples over a broad range of possible *random local clocks*



AUTOCORRELATED RATES

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

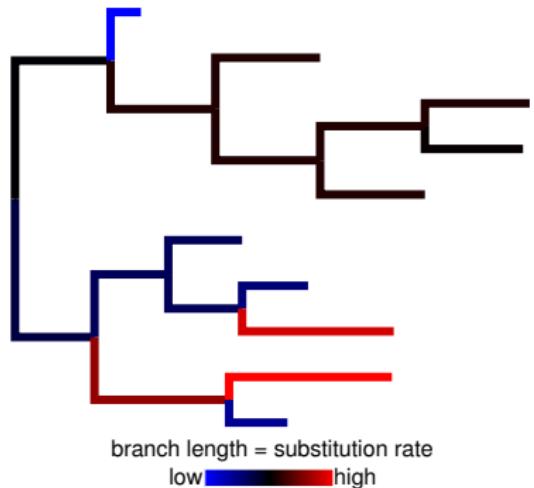
The rate at a node is drawn from a lognormal distribution with a mean equal to the parent rate



PUNCTUATED RATE CHANGE

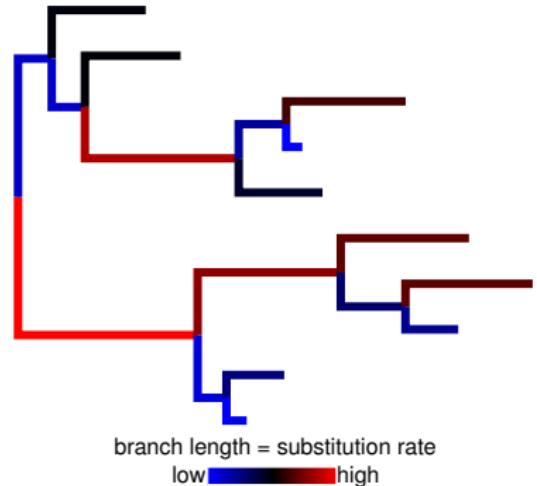
Rate changes occur along lineages according to a point process

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



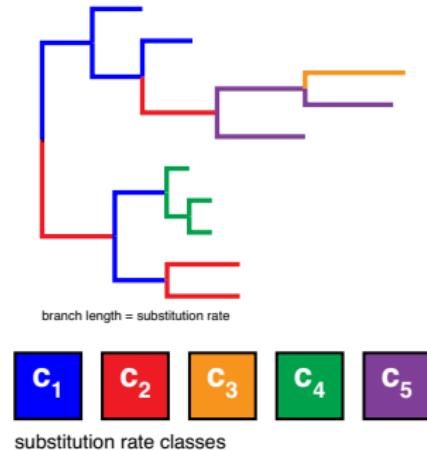
INFINITE MIXTURE MODEL

Dirichlet process prior:

Branches are partitioned
into distinct rate categories

Random variables under the
DPP informed by the **data**:

- the number of rate classes
- the assignment of branches to classes
- the rate value for each class



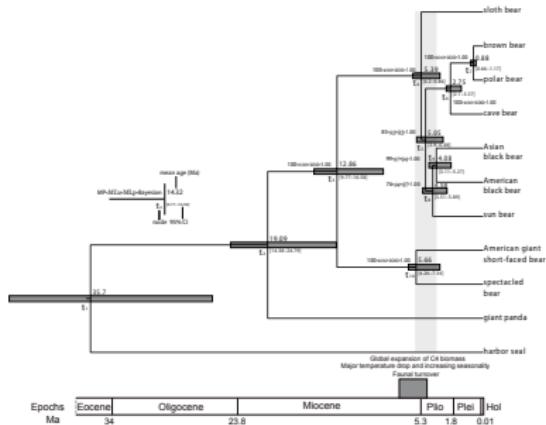
MODELING RATE VARIATION

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

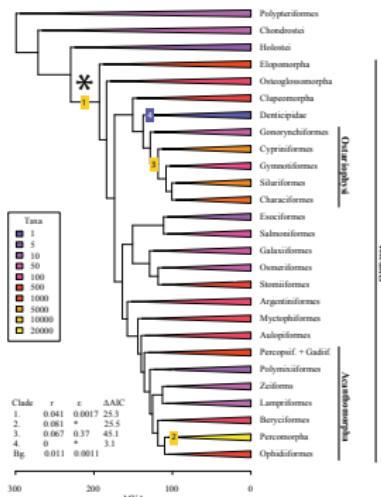
- Global molecular clock
- Local molecular clocks
- Punctuated rate change model
- Log-normally distributed autocorrelated rates
- Uncorrelated/independent rates models
- Dirichlet process prior

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
- Punctuated rate change model
- Log-normally distributed autocorrelated rates
- Uncorrelated/independent rates models
- Dirichlet process prior

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



BAYESIAN DIVERGENCE TIME ESTIMATION

Estimating divergence times relies on 2 main elements:

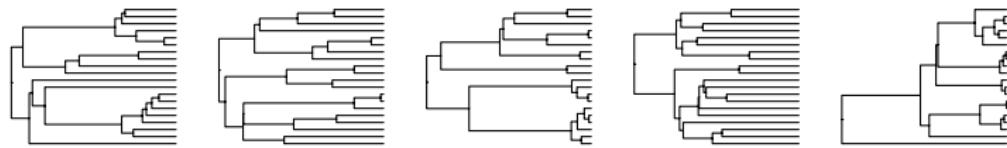
- Branch-specific rates: $f(\mathcal{R} | \theta_{\mathcal{R}})$
- Node ages: $f(\mathcal{A} | \theta_{\mathcal{A}}, \mathcal{C})$

<http://bayesiancook.blogspot.com/2013/12/two-sides-of-same-coin.html>

PRIORS ON NODE TIMES

Relaxed clock Bayesian analyses require a prior distribution on node times

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



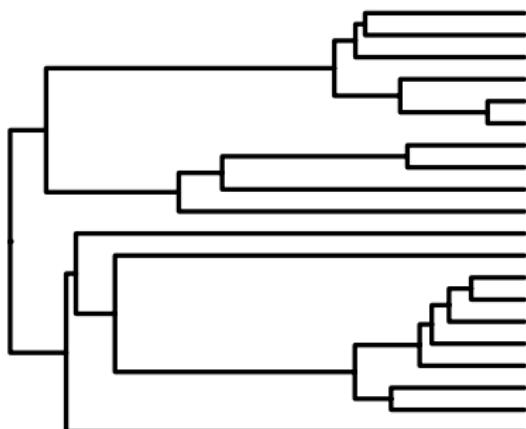
Different node-age priors make different assumptions about the timing of divergence events

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)

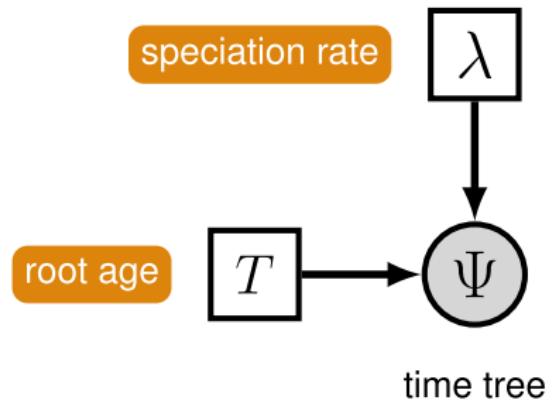


STOCHASTIC BRANCHING PROCESSES

Node-age priors based on stochastic models of lineage diversification

Yule process: assumes a constant rate of speciation, across lineages

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

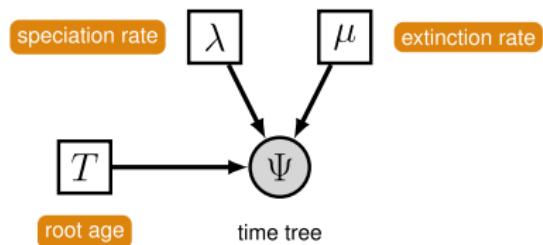


STOCHASTIC BRANCHING PROCESSES

Node-age priors based on stochastic models of lineage diversification

Constant-rate birth-death

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

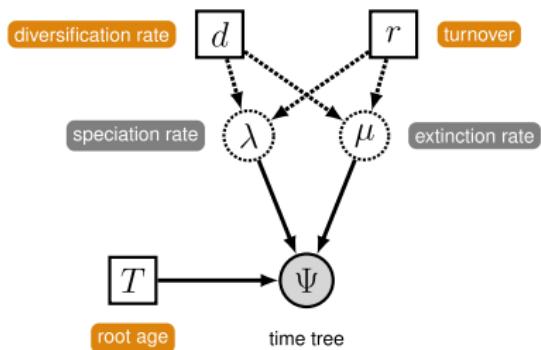


STOCHASTIC BRANCHING PROCESSES

Node-age priors based on stochastic models of lineage diversification

Constant-rate birth-death

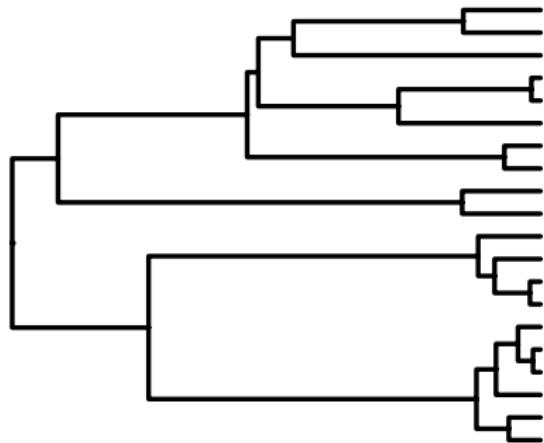
process: at any point in time a lineage can speciate at rate λ or go extinct with a rate of μ



STOCHASTIC BRANCHING PROCESSES

Node-age priors based on stochastic models of lineage diversification

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

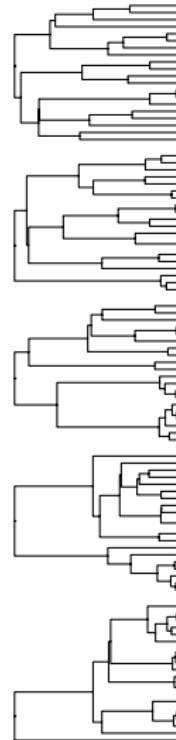


STOCHASTIC BRANCHING PROCESSES

Different values of λ and μ lead to different trees

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

Using hyperpriors on λ and μ accounts for uncertainty in these hyperparameters



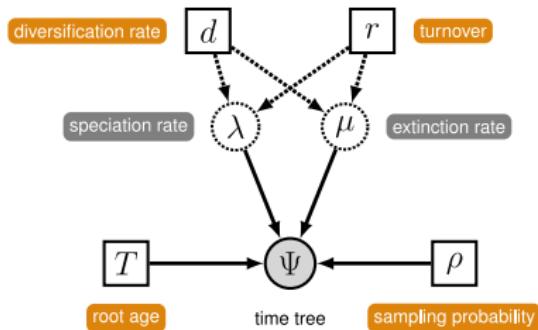
STOCHASTIC BRANCHING PROCESSES

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

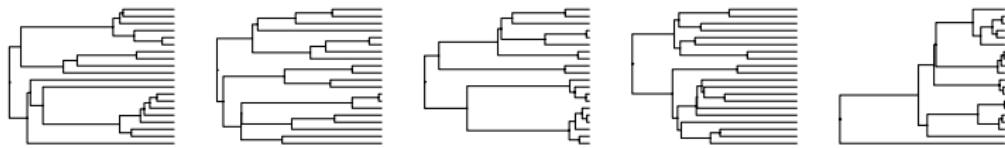
Conditions on a probability of sampling a tip, ρ



PRIORS ON NODE TIMES

Sequence data are only informative on *relative* rates & times

Node-time priors cannot give precise estimates of *absolute* node ages



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

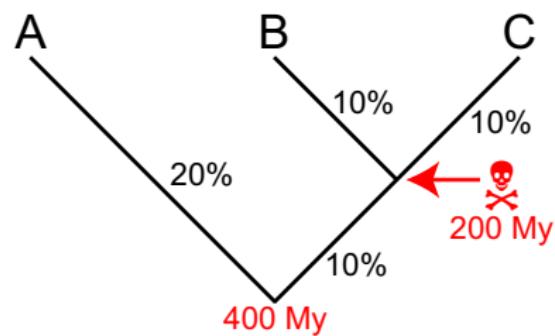


CALIBRATING DIVERGENCE TIMES

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

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

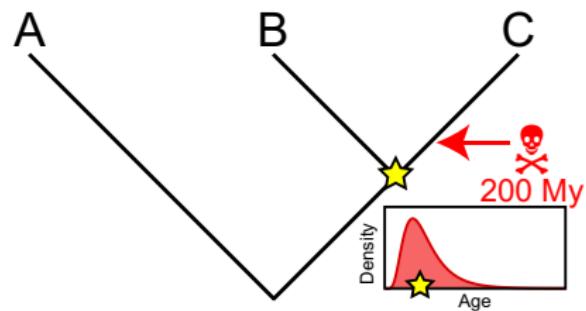
Uncertainty in the placement of fossils



CALIBRATION DENSITIES

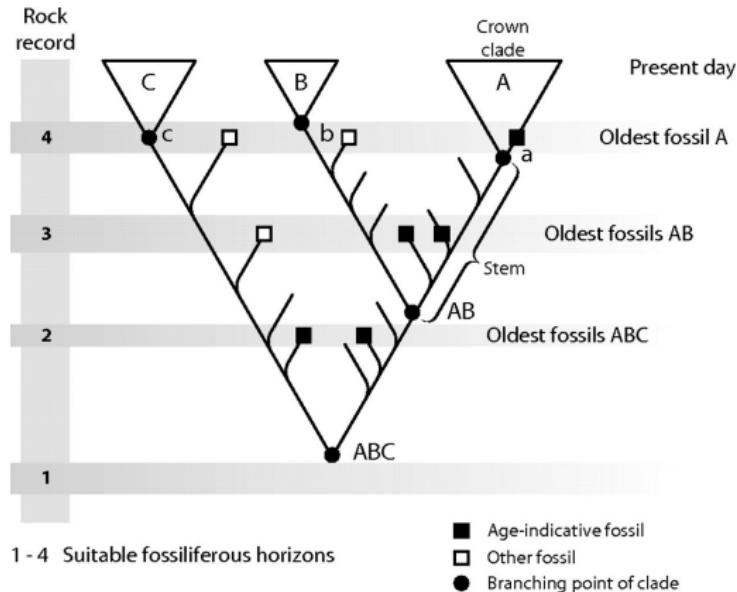
Bayesian inference is well suited to accommodating uncertainty in the age of the calibration node

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



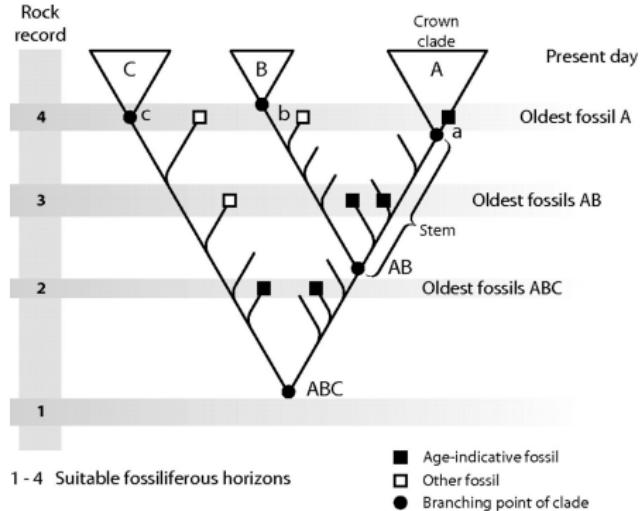
ASSIGNING FOSSILS TO CLADES

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



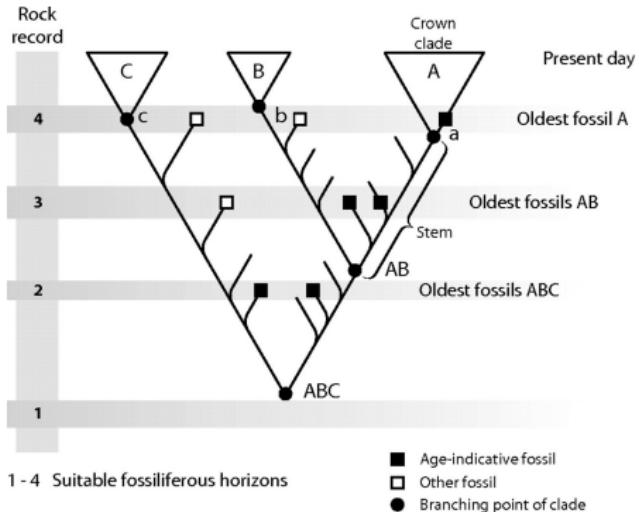
ASSIGNING FOSSILS TO CLADES

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



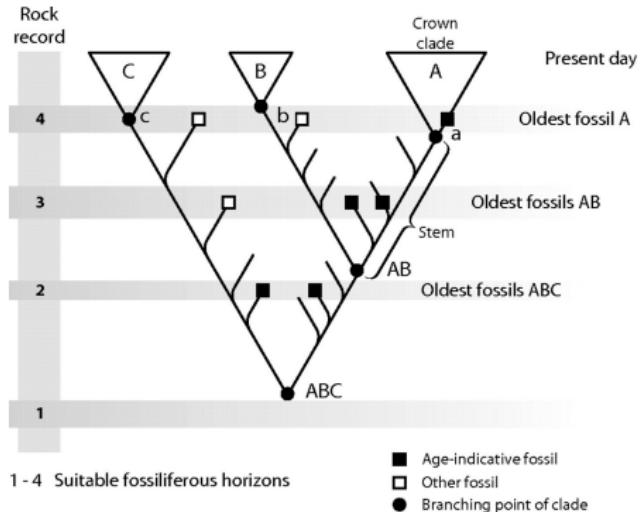
ASSIGNING FOSSILS TO CLADES

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



ASSIGNING FOSSILS TO CLADES

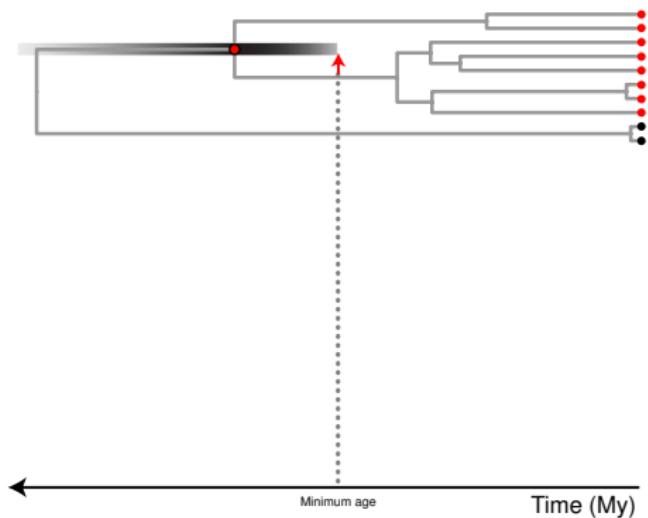
Fossiliferous horizons: the sources in the rock record for relevant fossils



Fossil Calibration

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

Reliable **maximum** bounds are typically unavailable

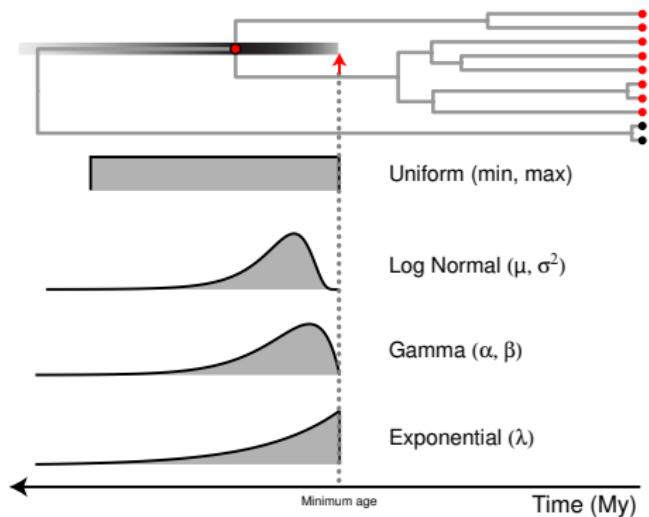


PRIOR DENSITIES ON CALIBRATED NODES

Common practice in Bayesian divergence-time estimation:

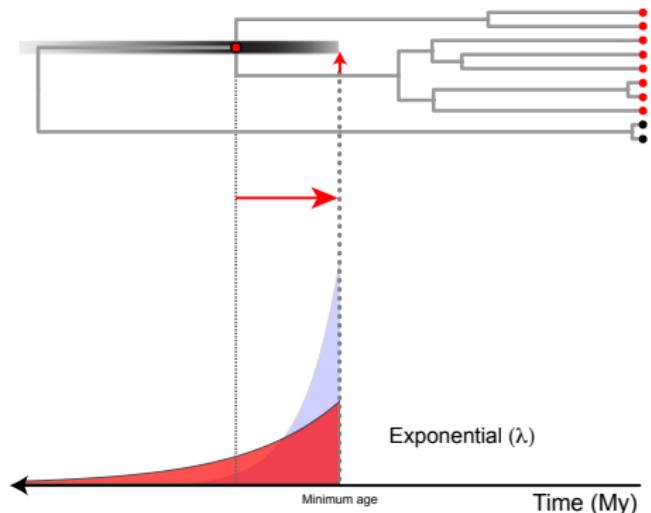
Parametric distributions are typically off-set by the age of the oldest fossil assigned to a clade

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



PRIOR DENSITIES ON CALIBRATED NODES

Calibration densities describe the waiting time between the divergence event and the age of the oldest fossil

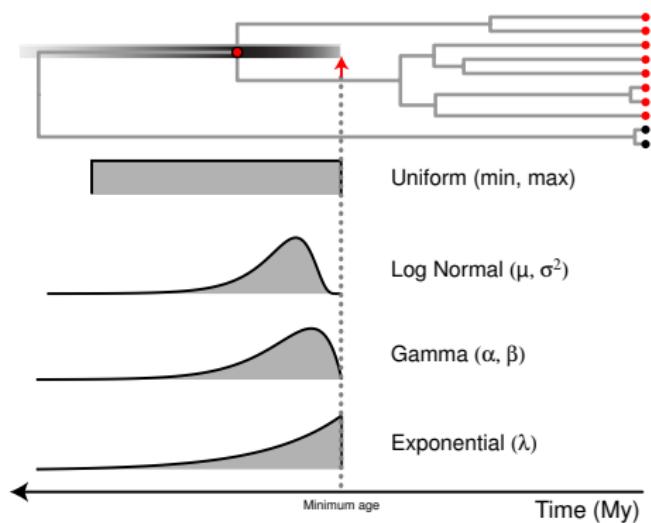


PRIOR DENSITIES ON CALIBRATED NODES

Common practice in Bayesian divergence-time estimation:

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

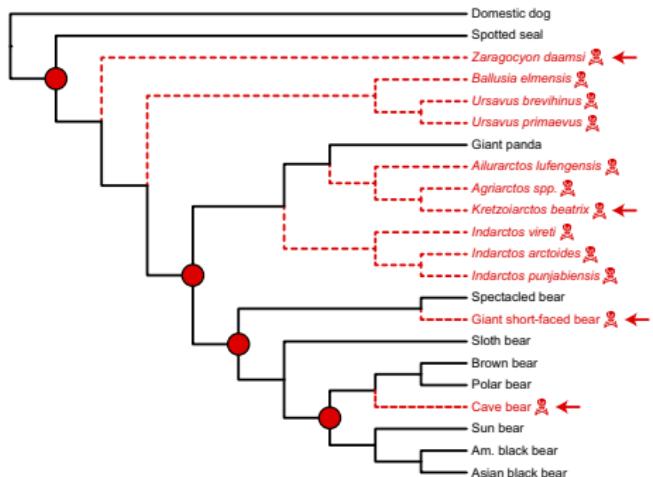
Specifying appropriate densities is a challenge for most molecular biologists



IMPROVING FOSSIL CALIBRATION

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

Calibration densities do not account for diversification of fossils

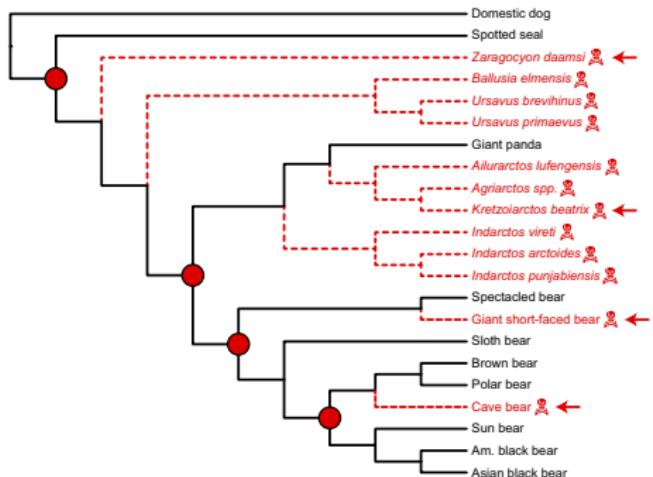


IMPROVING FOSSIL CALIBRATION

We want to use all of the available fossils

Example: Bears

12 fossils are reduced to 4 calibration ages with calibration density methods

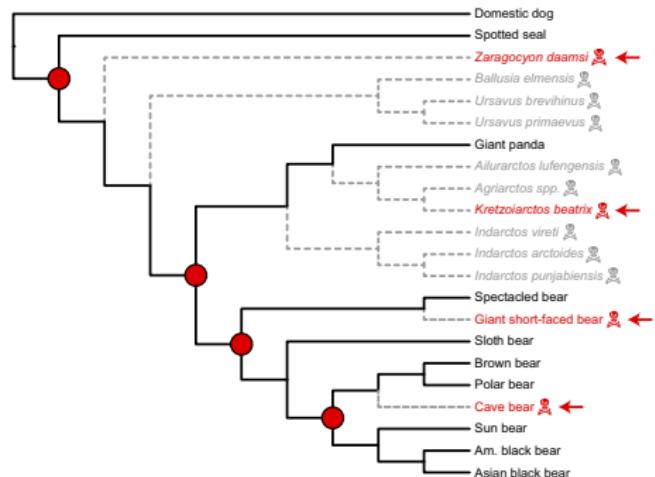


IMPROVING FOSSIL CALIBRATION

We want to use all of the available fossils

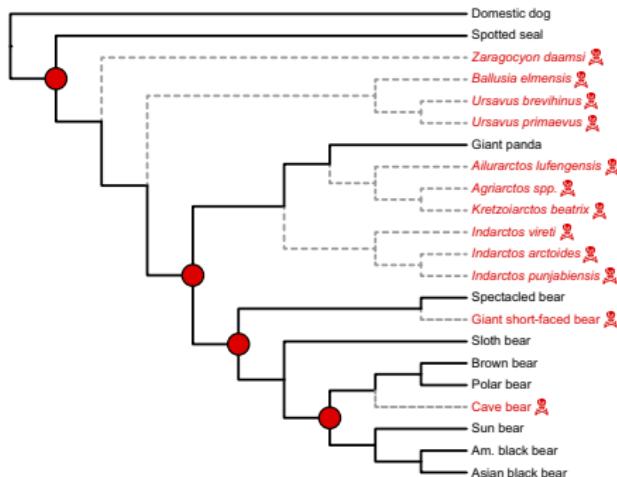
Example: Bears

12 fossils are reduced to 4 calibration ages with calibration density methods



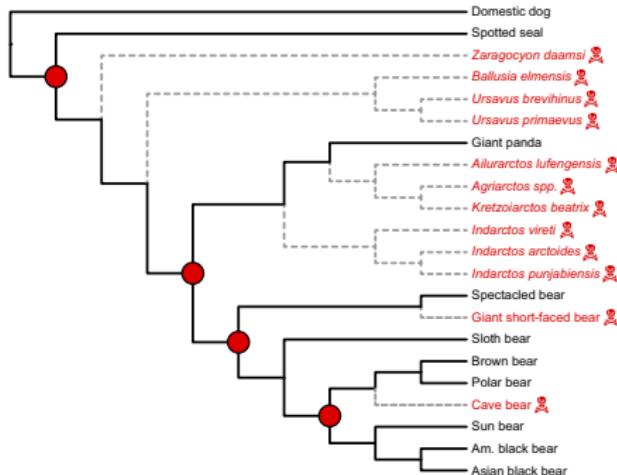
IMPROVING FOSSIL CALIBRATION

Because fossils are part of the diversification process, we can combine fossil calibration with birth-death models



IMPROVING FOSSIL CALIBRATION

This relies on a branching model that accounts for **speciation, extinction, and rates of fossilization, preservation, and recovery**



THE FOSSILIZED BIRTH-DEATH PROCESS (FBD)

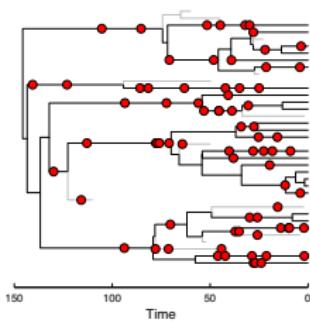
Improving statistical inference of absolute node ages

Eliminates the need to specify arbitrary calibration densities

Better capture our statistical uncertainty in species divergence dates

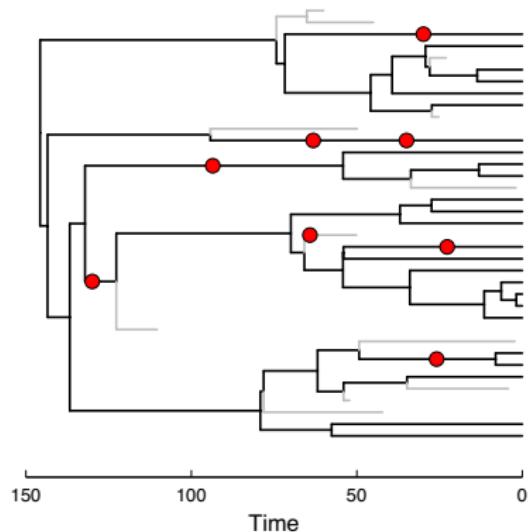
All reliable fossils associated with a clade are used

Useful for calibration or 'total-evidence' dating



THE FOSSILIZED BIRTH-DEATH PROCESS (FBD)

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



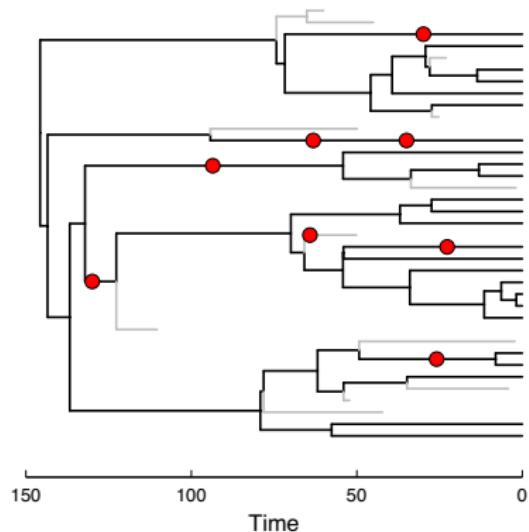
THE FOSSILIZED BIRTH-DEATH PROCESS (FBD)

The probability of the tree and fossil observations under a birth-death model with rate parameters:

λ = speciation

μ = extinction

ψ = fossilization/recovery



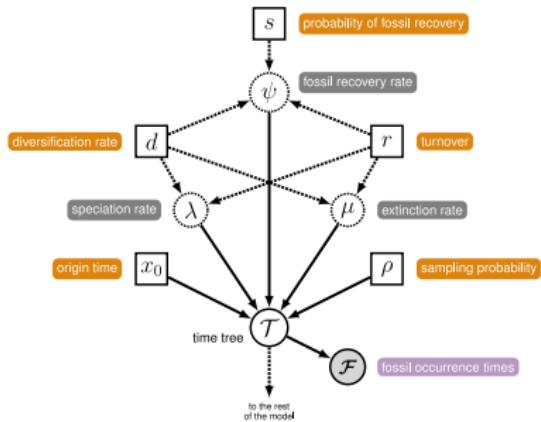
THE FOSSILIZED BIRTH-DEATH PROCESS (FBD)

The probability of the tree and fossil observations under a birth-death model with rate parameters:

λ = speciation

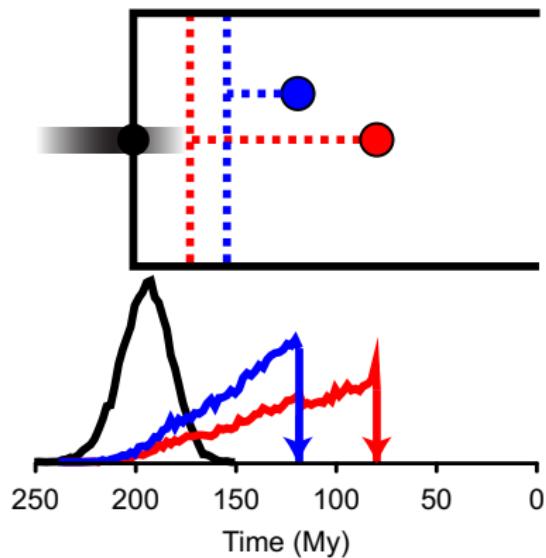
μ = extinction

ψ = fossilization/recovery



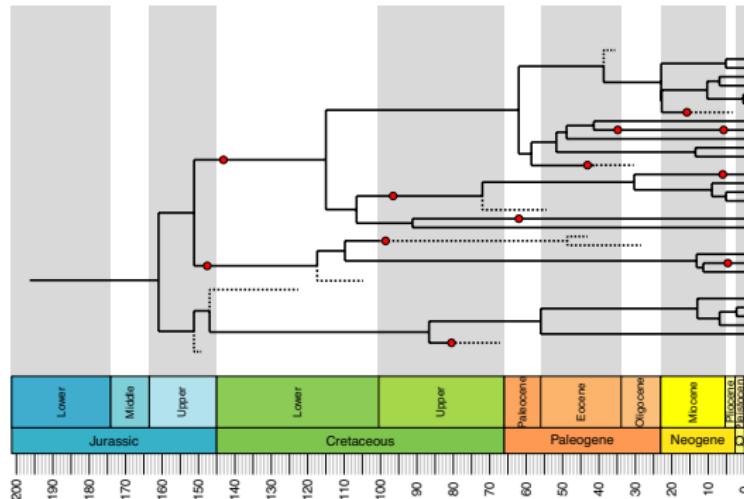
THE FOSSILIZED BIRTH-DEATH PROCESS (FBD)

We use MCMC to sample realizations of the diversification process, integrating over the topology—including placement of the fossils—and speciation times



INFERRING FBD TREES

Extensions of the fossilized birth-death process accommodate variation in fossil sampling, non-random species sampling, & shifts in diversification rates.

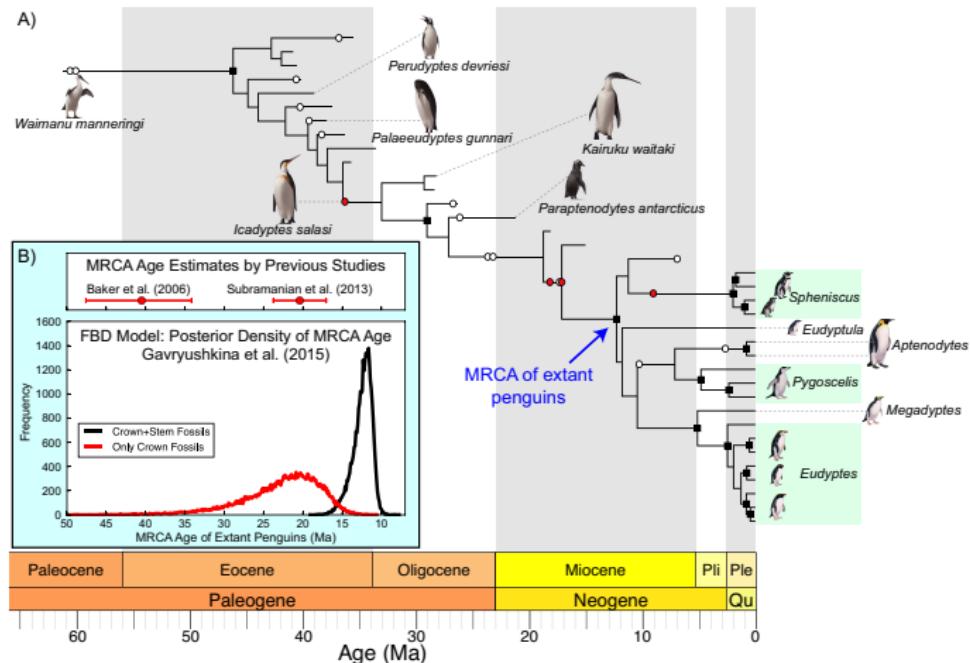


With character data for both fossil & extant species, we account for uncertainty in fossil placement

DATING THE CROWN-PENGUIN RADIATION

Analysis of morphology + DNA for fossil & extant taxa

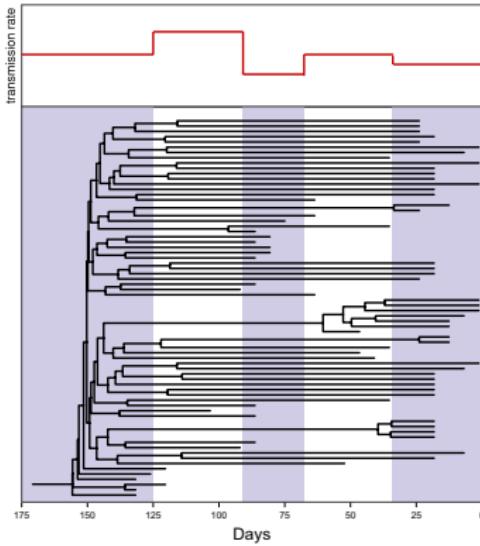
Earlier age for crown MRCA is more consistent with the fossil record



SKYLINE BIRTH-DEATH PROCESS

A piecewise shifting model where parameters change over time

Used to estimate epidemiological parameters of an outbreak



Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV)

Tanja Stadler^{a,1,2}, Denise Kühnert^{b,c,1}, Sebastian Bonhoeffer^a, and Alexei J. Drummond^{b,c}

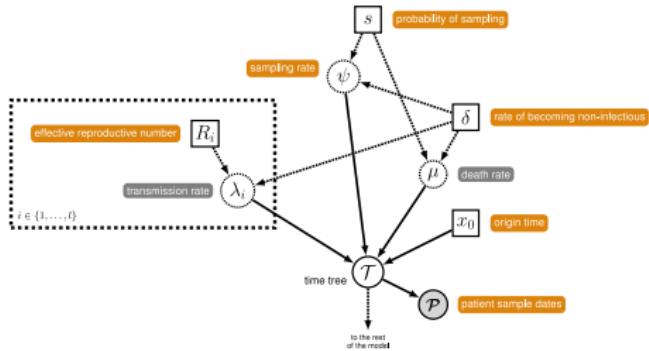
SKYLINE BIRTH-DEATH PROCESS

l is the number of parameter intervals

R_i is the effective reproductive number for interval $i \in l$

δ is the rate of becoming non-infectious

s is the probability of sampling an individual after becoming non-infectious



$$R_i = \frac{\lambda_i}{\mu + \psi}, \quad \delta = \mu + \psi, \quad s = \frac{\psi}{\mu + \psi}$$

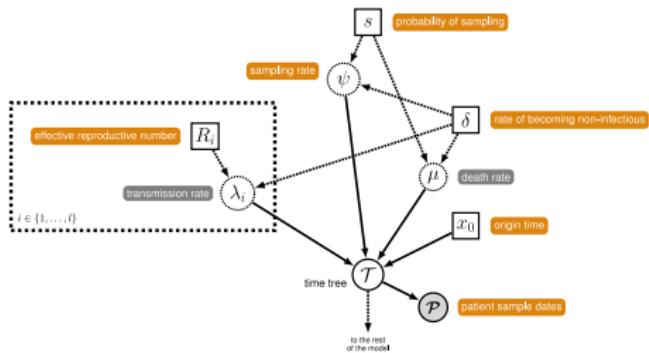
SKYLINE BIRTH-DEATH PROCESS

l is the number of parameter intervals

λ_i is the transmission rate for interval $i \in l$

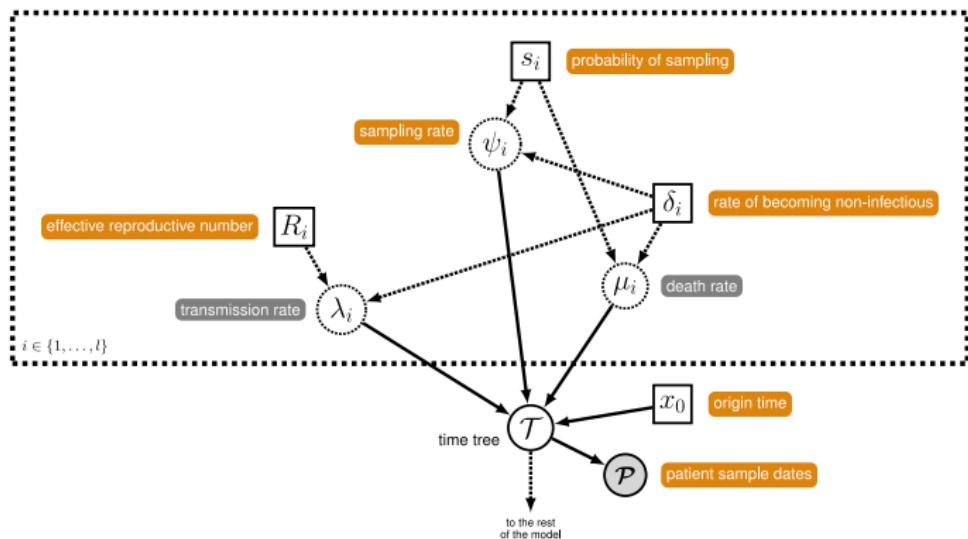
μ is the viral lineage death rate

ψ is the rate each individual is sampled



$$\lambda_i = R_i \delta, \quad \mu = \delta - s\delta, \quad \psi = s\delta$$

SKYLINE BIRTH-DEATH PROCESS



SKYLINE BIRTH-DEATH PROCESS

A decline in R over the history of HIV-1 in the UK is consistent with the introduction of effective drug therapies

After 1998 R decreased below 1, indicating a declining epidemic

