

# BAYESIAN DIVERGENCE TIME ESTIMATION

Tracy Heath

Ecology, Evolution, & Organismal Biology  
Iowa State University



@tracyc7

<http://phyloworks.org>

SSB Workshop at Evolution 2015  
Guarujá, Brazil

# TAKE-HOME MESSAGES

What I hope to emphasize here:

- Bayes' theorem is a beautiful thing
- The substitution rate & time are confounded parameters
- To estimate branch *time* we need separate models for the rate along the branch & the time duration of the branch
- Sequence data *alone* are not informative for absolute time (in years)
- To infer absolute times, additional data (e.g., fossils or biogeography) are needed
- It's very important to have a good understanding of *all* data (including fossils) used for divergence-time estimation

# 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  $\delta$  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  $\delta$ .

## BAYES' THEOREM

The posterior probability **density** a continuous parameter  $\theta$  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  $\theta$ .

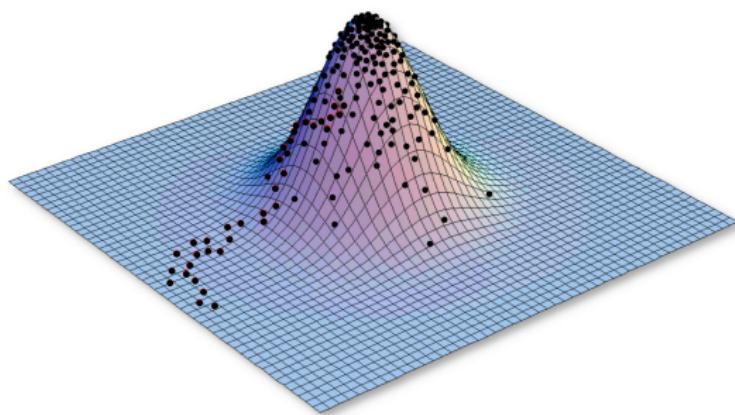
## ESTIMATING THE POSTERIOR PROBABILITY

Once we have a model defined that represents  $f(\theta | D)$ , how do we compute the posterior probability?

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

# 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.*

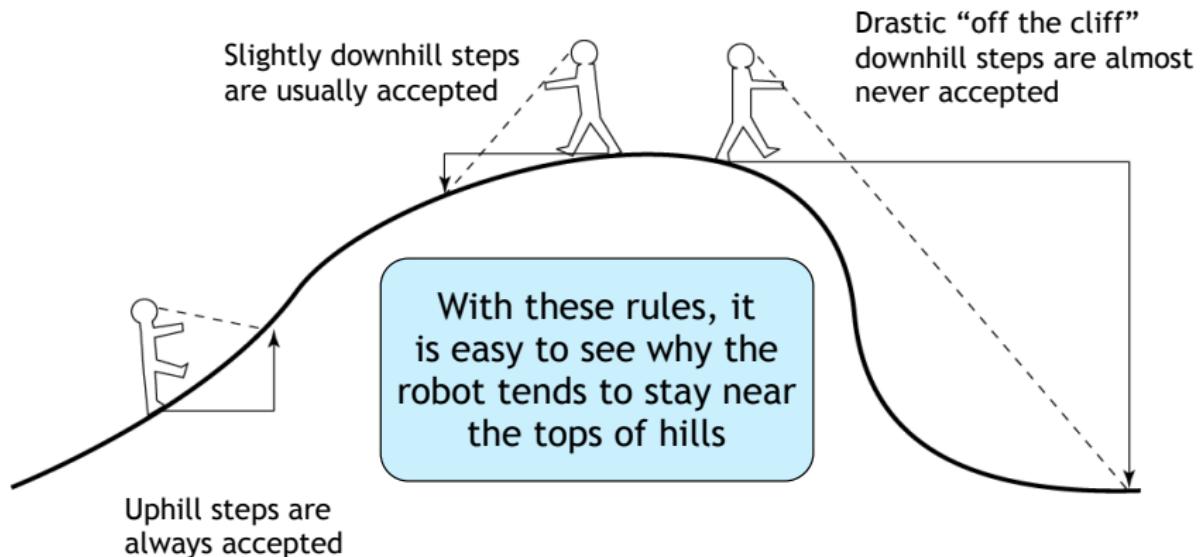
Hastings. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*.

# MARKOV CHAIN MONTE CARLO (MCMC)

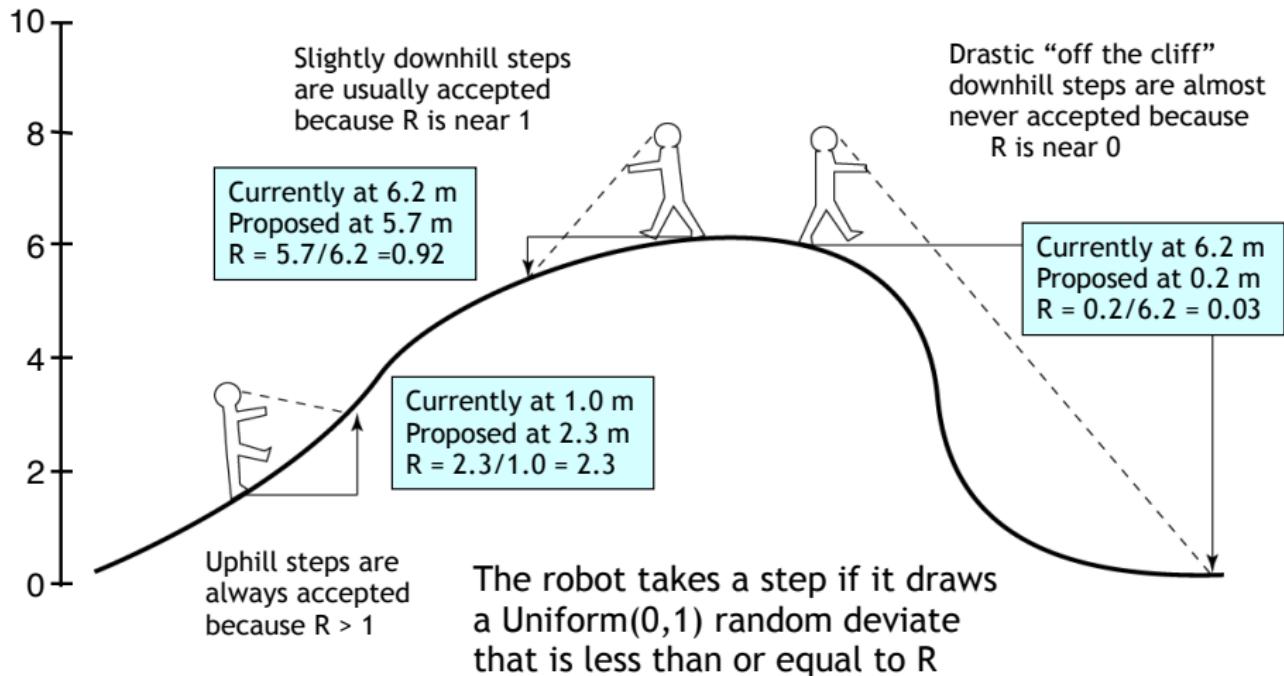
More on MCMC from Paul Lewis—our esteemed SSB President—and his lecture on Bayesian phylogenetics

Slides source: [https://molevol.mbl.edu/index.php/Paul\\_Lewis](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.

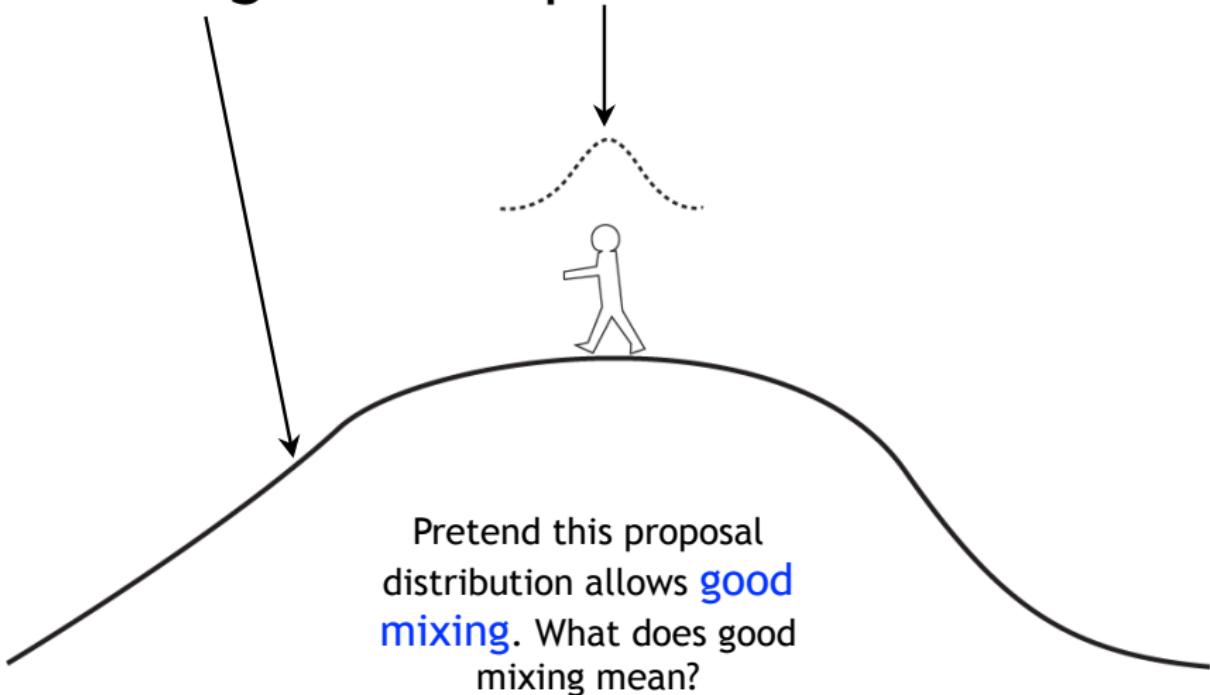
$$\frac{f(\theta^*|D)}{f(\theta|D)} = \frac{\frac{f(D|\theta^*)f(\theta^*)}{\cancel{f(D)}}}{\frac{f(D|\theta)f(\theta)}{\cancel{f(D)}}} = \frac{f(D|\theta^*)f(\theta^*)}{f(D|\theta)f(\theta)}$$

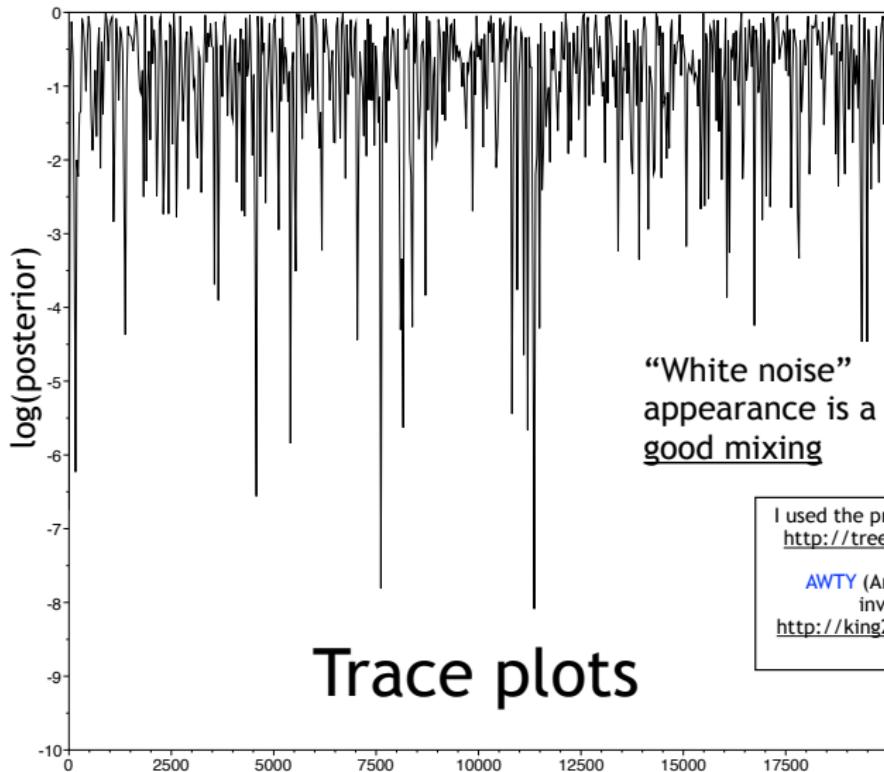
Posterior  
odds

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](http://king2.scs.fsu.edu/CEBProjects/awty/_awty_start.php)

## Trace plots

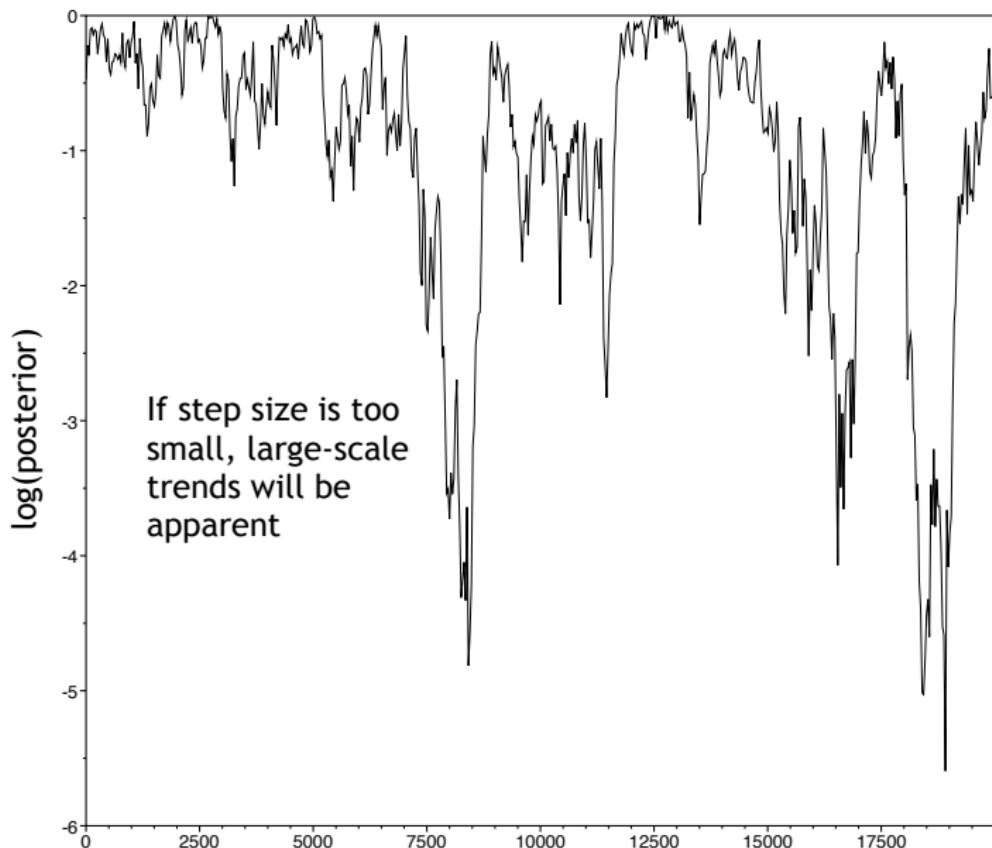
# Target vs. Proposal Distributions

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



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

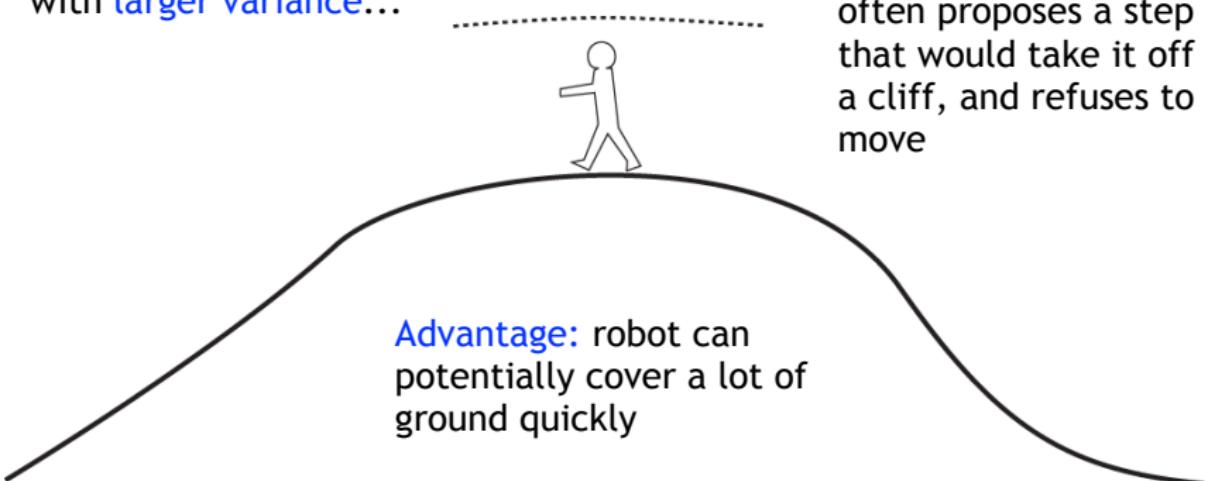
**Advantage:** robot seldom  
refuses to take proposed  
steps

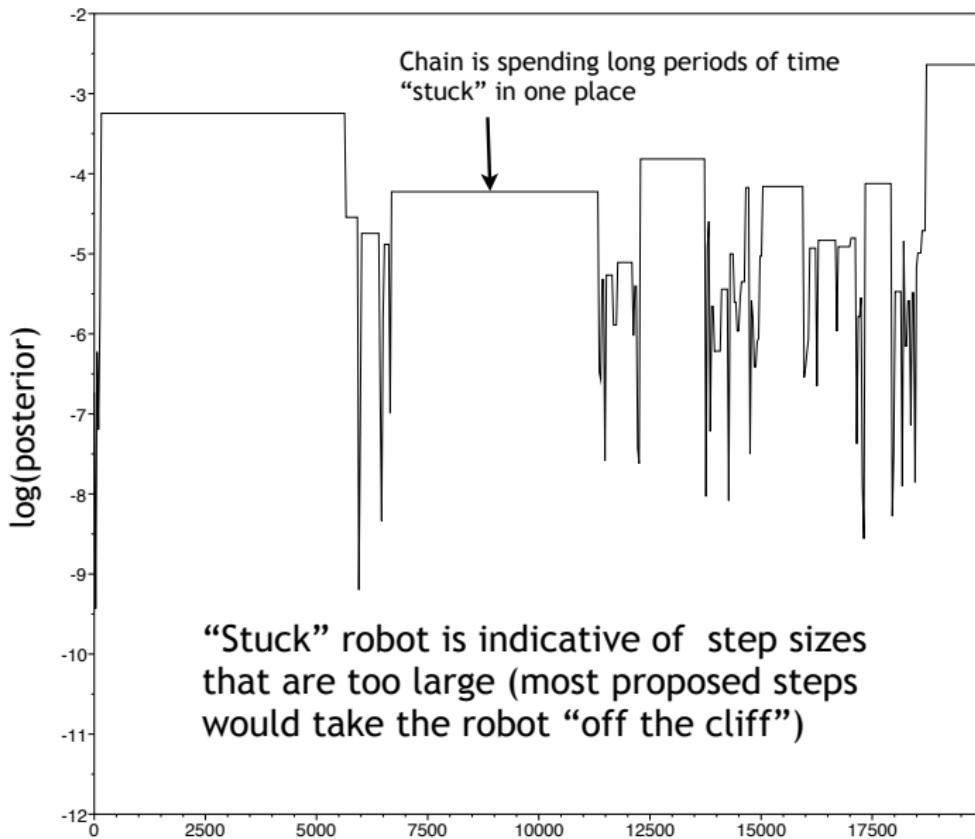


# Target vs. Proposal Distributions

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

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

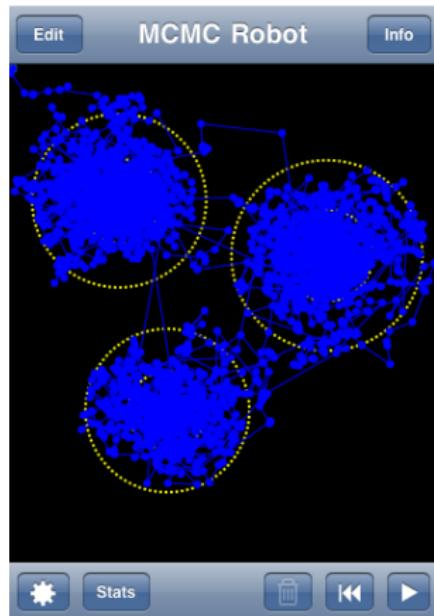




# MARKOV CHAIN MONTE CARLO (MCMC)

Thanks, Paul!

Slides source: [https://molevol.mbl.edu/index.php/Paul\\_Lewis](https://molevol.mbl.edu/index.php/Paul_Lewis)



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

<http://www.mcmcrobot.org>



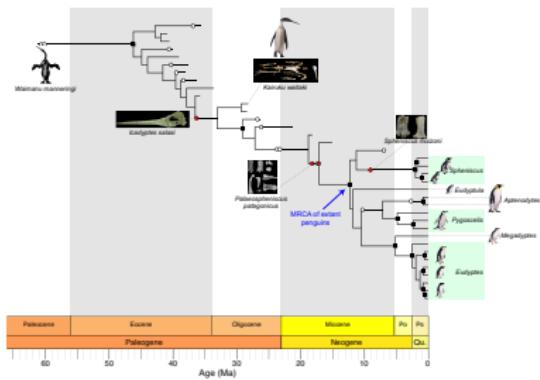
# DIVERGENCE TIME ESTIMATION

**Goal:** Estimate the branch lengths in units proportional to time 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

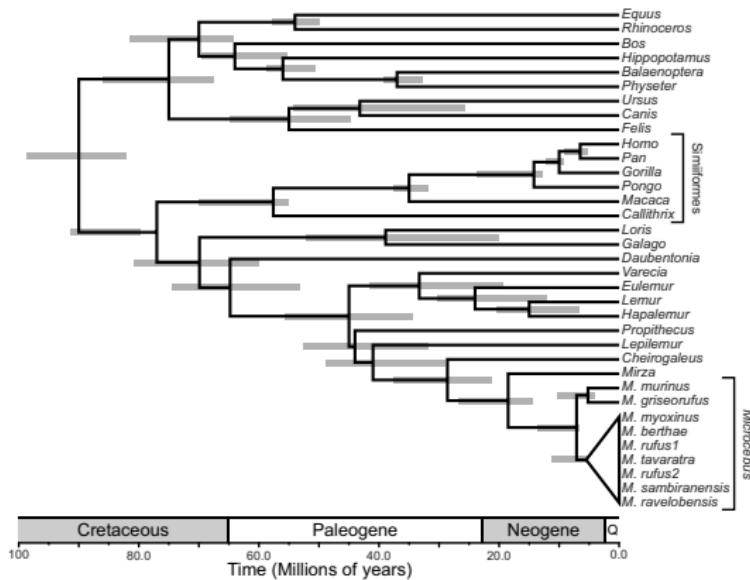
External calibration information for estimates of absolute node times



(Figure adapted from Gavryushkina et al., arXiv:1506.04797)

# A TIME-SCALE FOR EVOLUTION

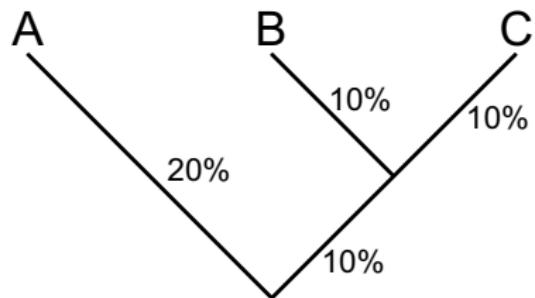
Phylogenetic trees can provide both topological information and temporal information



# THE GLOBAL MOLECULAR CLOCK

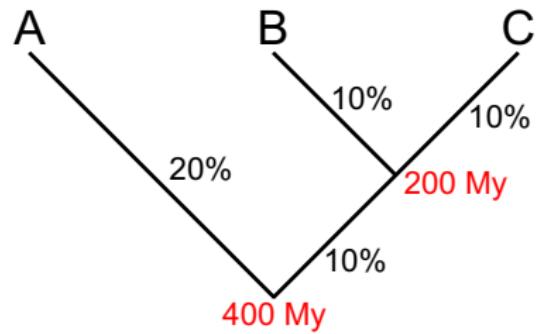
Assume that the rate of evolutionary change is constant over time

(branch lengths equal percent sequence divergence)



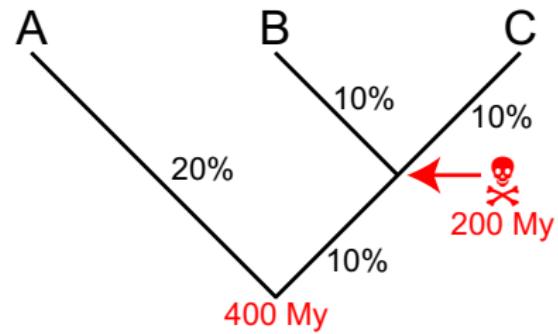
# THE GLOBAL MOLECULAR CLOCK

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:

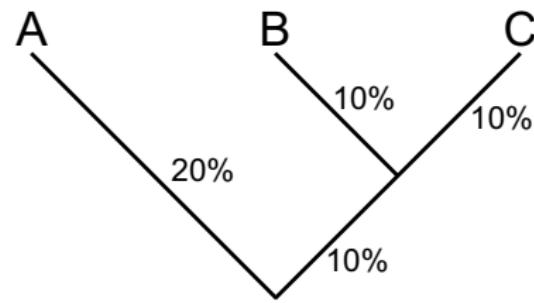
Variation in

- metabolic rate
- generation time
- DNA repair

## Fixation rate:

Variation in

- strength and targets of selection
- population sizes

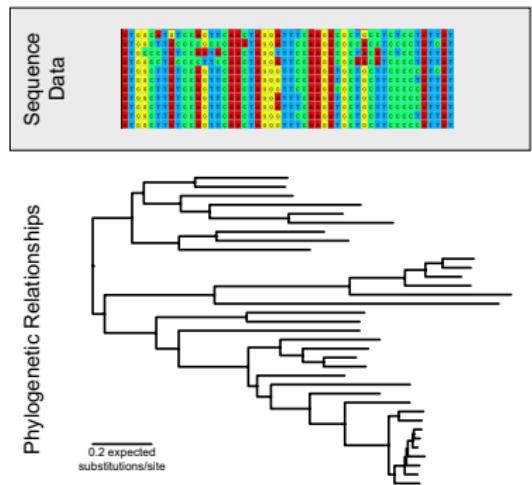


# UNCONSTRAINED ANALYSIS

Sequence data provide information about **branch lengths**

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

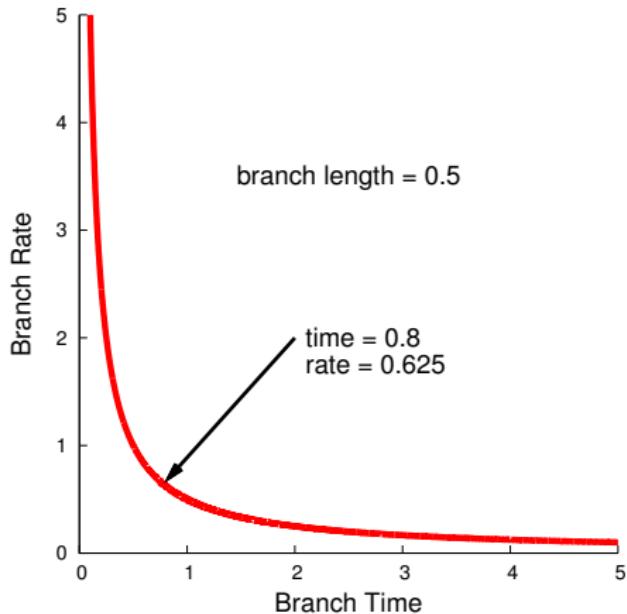
$$\text{branch length} = \text{rate} \times \text{time}$$



# RATE AND TIME

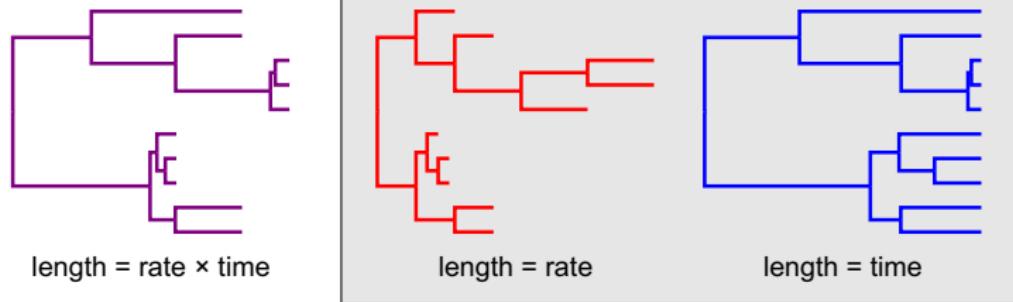
The sequence data provide information about branch length

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



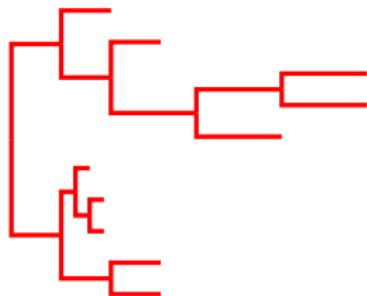
# RATE AND TIME

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

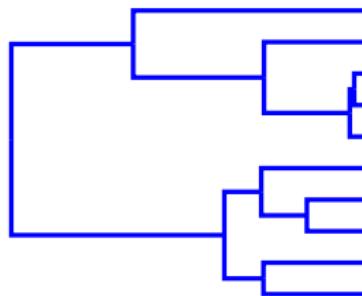


Methods for dating species divergences estimate the **substitution rate** and **time** separately

# BAYESIAN DIVERGENCE TIME ESTIMATION



length = rate



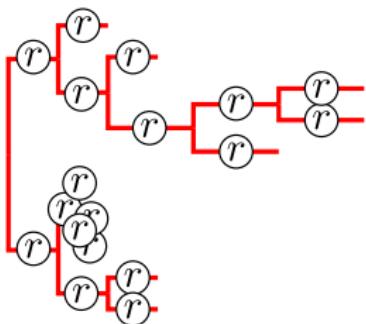
length = time

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

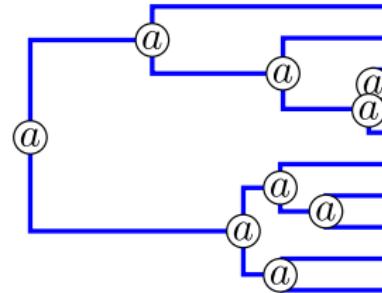
$$\mathcal{A} = (a_1, a_2, a_3, \dots, a_{N-1})$$

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

# 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

$\Psi$  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)

## RELAXED-CLOCK MODELS

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

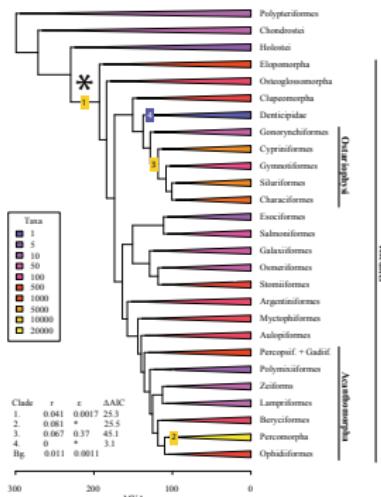
- Local molecular clocks
- Punctuated rate change model
- Log-normally distributed autocorrelated rates
- Uncorrelated/independent rates models
- Mixture models on branch rates

# 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

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

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:

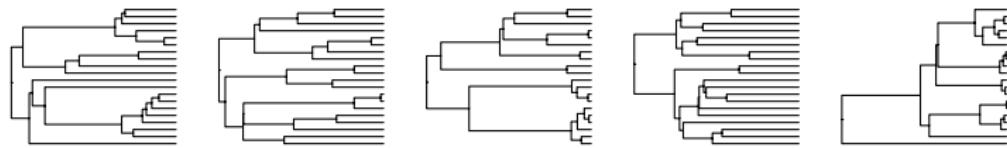
- Branch-specific rates:  $f(\mathcal{R} | \theta_{\mathcal{R}})$
- Node ages:  $f(\mathcal{A} | \theta_{\mathcal{A}}, \mathcal{C})$

<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

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



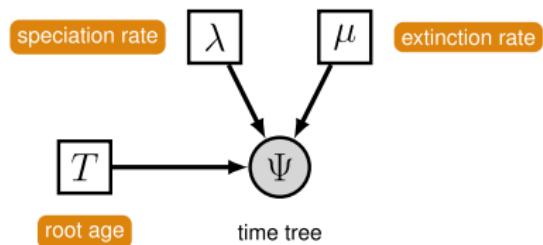
Different node-age priors make different assumptions about the timing of divergence events

# STOCHASTIC BRANCHING PROCESSES

Node-age priors based on stochastic models of lineage diversification

## Constant-rate birth-death

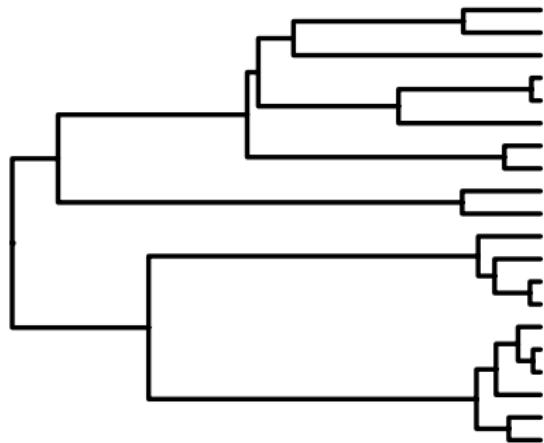
**process:** at any point in time a lineage can speciate at rate  $\lambda$  or go extinct with a rate of  $\mu$



# 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  $\lambda$  or go extinct with a rate of  $\mu$

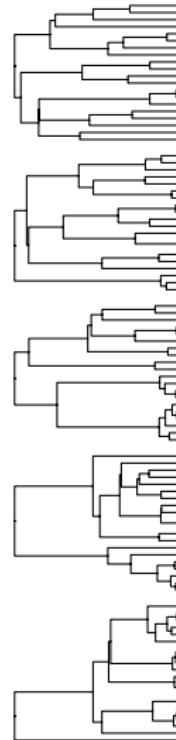


# STOCHASTIC BRANCHING PROCESSES

Different values of  $\lambda$  and  $\mu$  lead to different trees

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

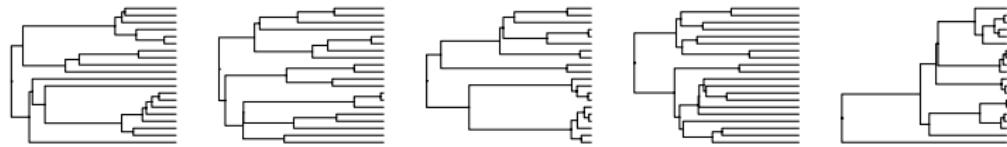
Using hyperpriors on  $\lambda$  and  $\mu$  accounts for uncertainty in these hyperparameters



# 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

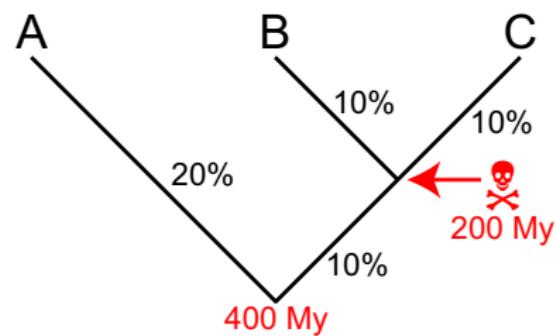


# CALIBRATING DIVERGENCE TIMES

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

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

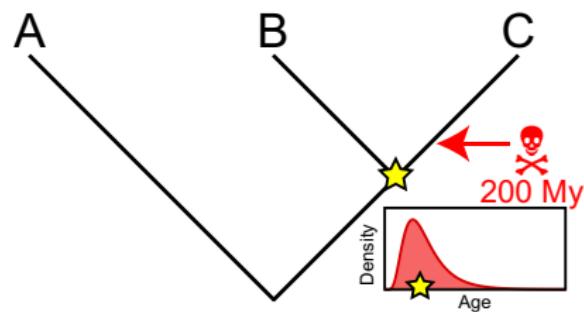
Uncertainty in the placement of fossils



# CALIBRATION DENSITIES

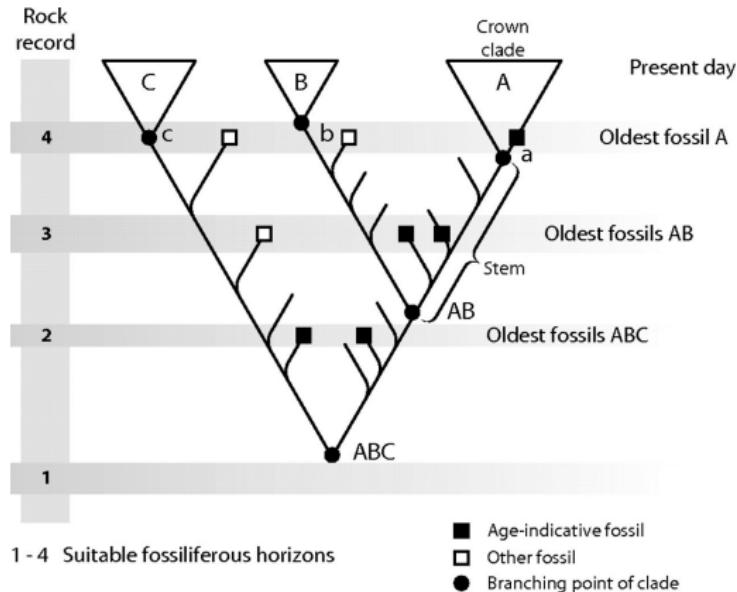
Bayesian inference is well suited to accommodating uncertainty in the age of the calibration node

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



# ASSIGNING FOSSILS TO CLADES

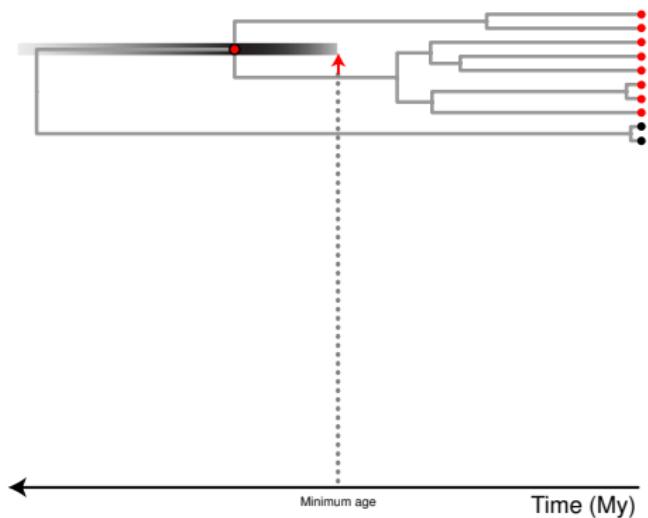
Misplaced fossils can affect node age estimates throughout the tree – if the fossil is older than its presumed MRCA



# Fossil Calibration

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

Reliable **maximum** bounds are typically unavailable

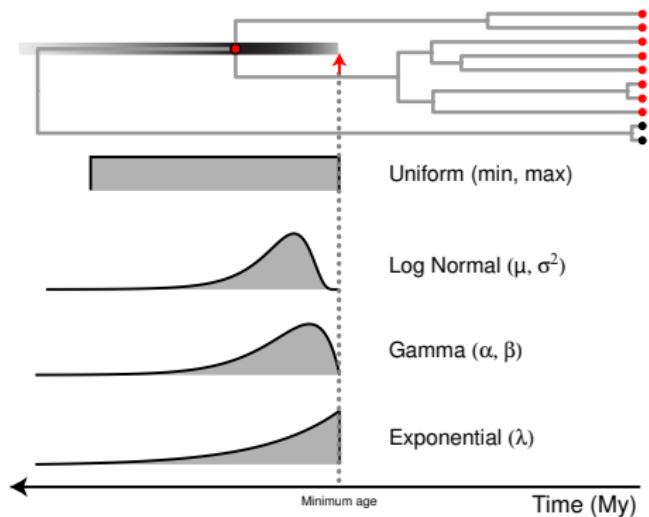


# PRIOR DENSITIES ON CALIBRATED NODES

Common practice in Bayesian divergence-time estimation:

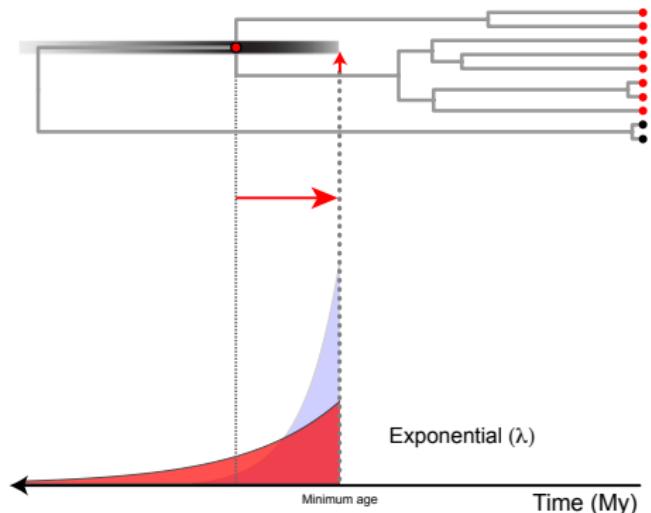
Parametric distributions are typically off-set by the age of the oldest fossil assigned to a clade

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

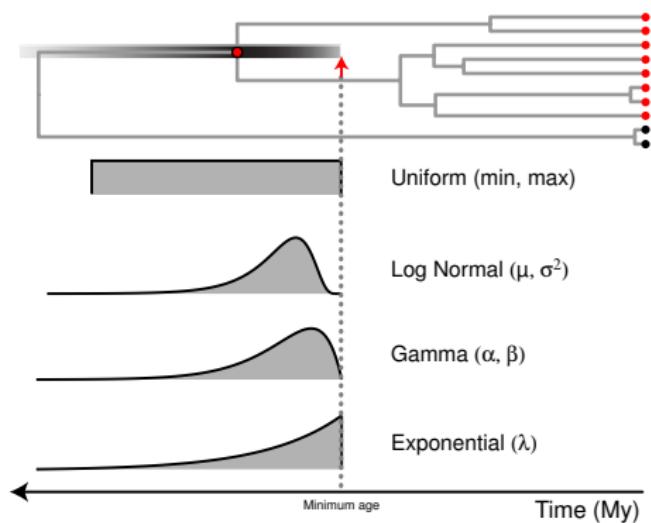


# PRIOR DENSITIES ON CALIBRATED NODES

Common practice in Bayesian divergence-time estimation:

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

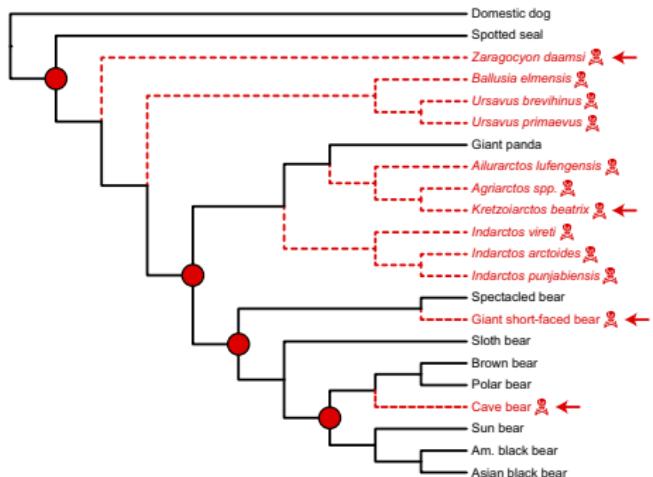
Specifying appropriate densities is a challenge for most molecular biologists



# 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

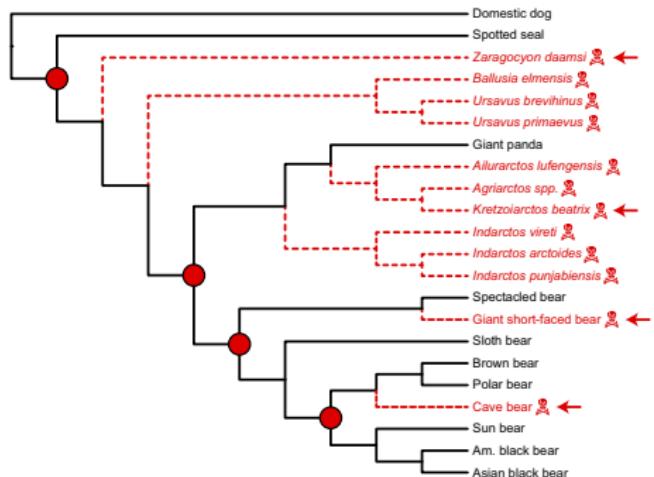


# IMPROVING FOSSIL CALIBRATION

We want to use all of the available fossils

## Example: Bears

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

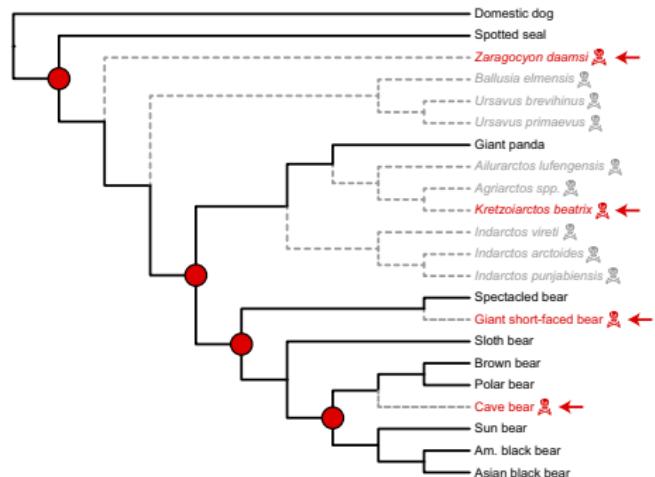


# IMPROVING FOSSIL CALIBRATION

We want to use all of the available fossils

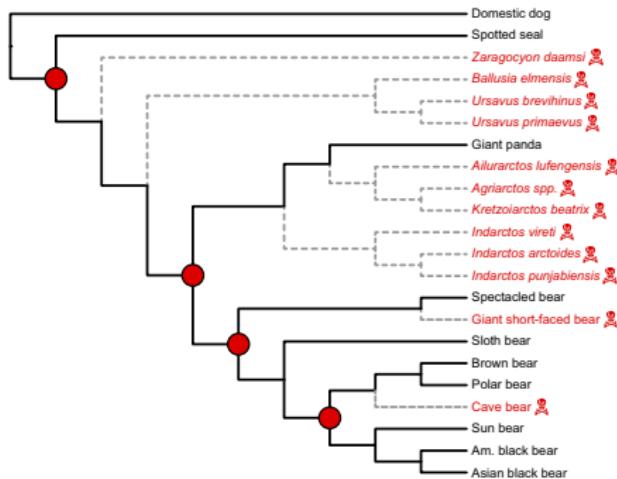
## Example: Bears

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



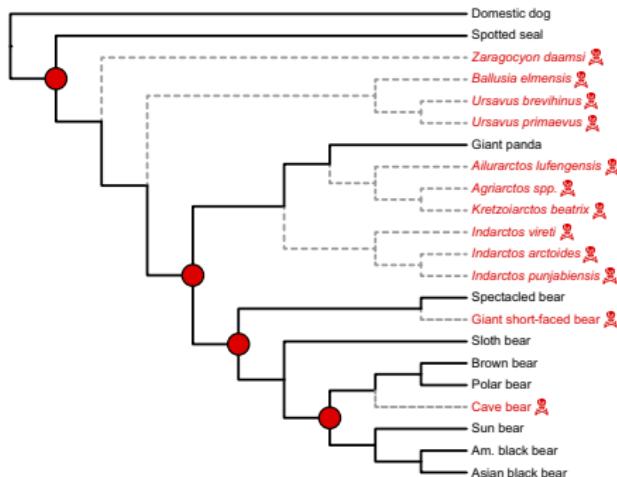
# IMPROVING FOSSIL CALIBRATION

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)

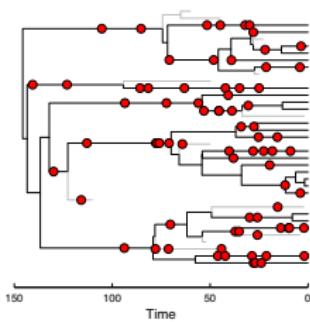
Improving statistical inference of absolute node ages

Eliminates the need to specify arbitrary calibration densities

Better capture our statistical uncertainty in species divergence dates

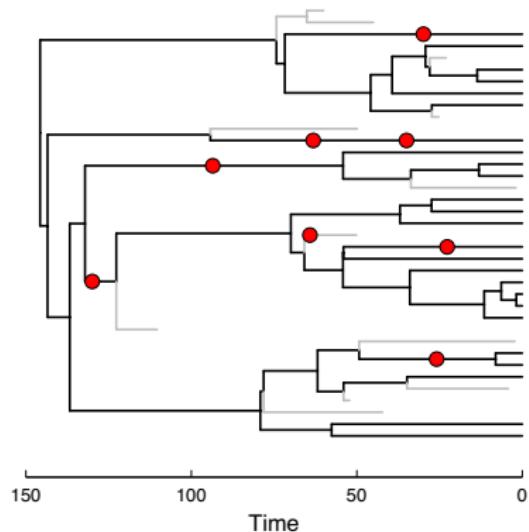
All reliable fossils associated with a clade are used

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



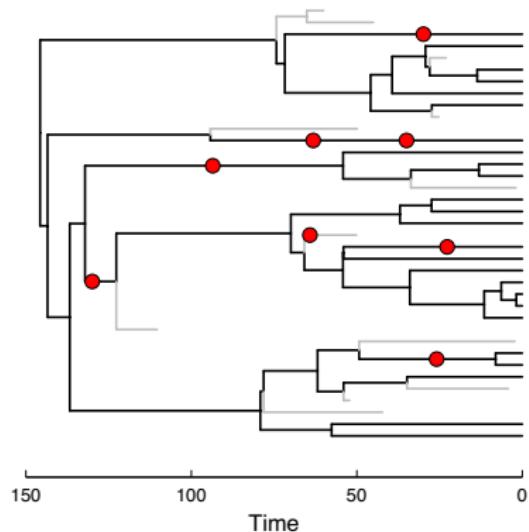
# THE FOSSILIZED BIRTH-DEATH PROCESS (FBD)

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

$\lambda$  = speciation

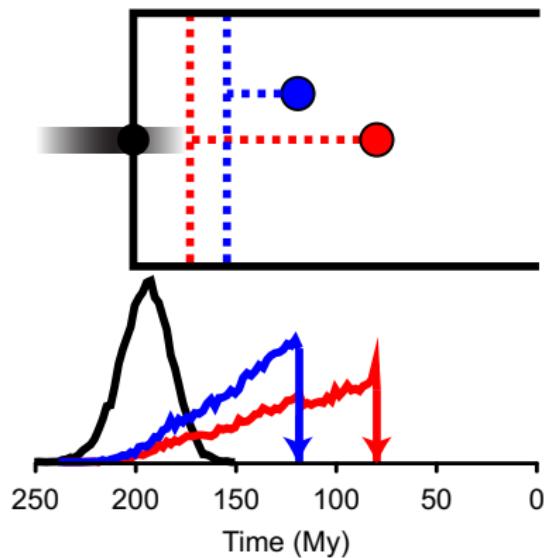
$\mu$  = extinction

$\psi$  = fossilization/recovery



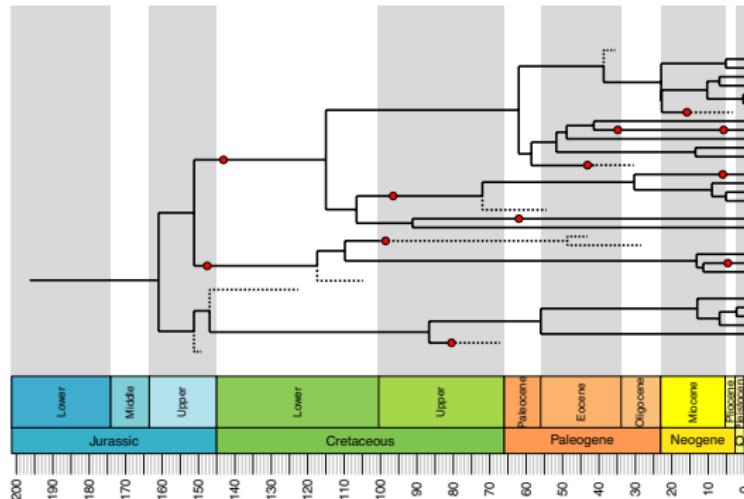
# THE FOSSILIZED BIRTH-DEATH PROCESS (FBD)

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.

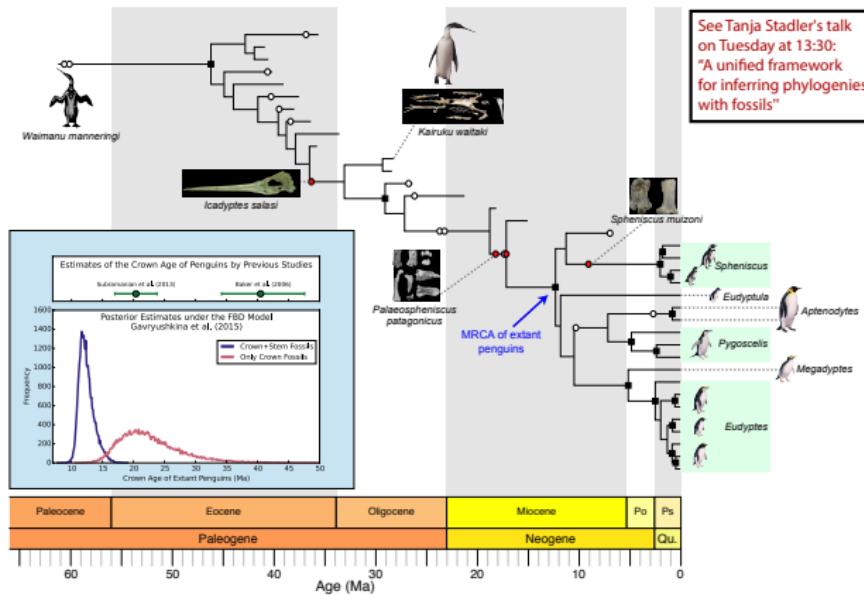


With character data for both fossil & extant species, we account for uncertainty in fossil placement

# 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

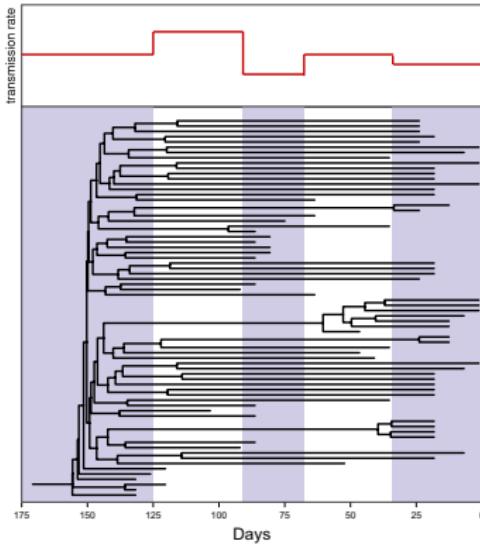


(Figure adapted from Gavryushkina et al., arXiv:1506.04797)

# 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)**

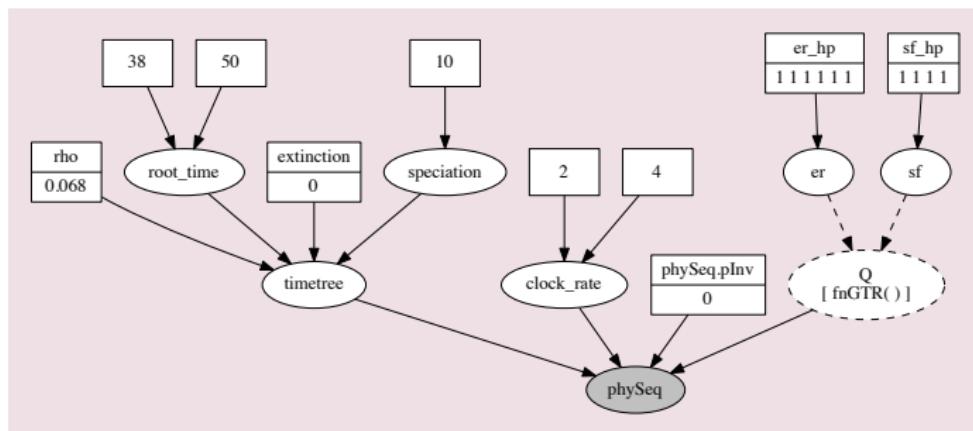
Tanja Stadler<sup>a,1,2</sup>, Denise Kühnert<sup>b,c,1</sup>, Sebastian Bonhoeffer<sup>a</sup>, and Alexei J. Drummond<sup>b,c</sup>

# RevBayes Demo: A SIMPLE ANALYSIS

**RevBayes** – Fully integrative Bayesian inference of phylogenetic parameters using probabilistic graphical models and an interpreted language

<http://RevBayes.com>

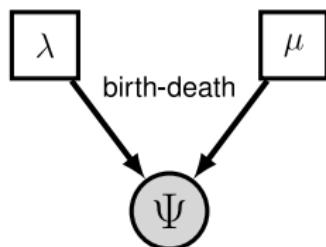
Graphical model: Strict clock, pure birth process, GTR



# GRAPHICAL MODELS IN RevBAYES

Graphical models provide tools for visually & computationally representing complex, parameter-rich probabilistic models

We can depict the conditional dependence structure of various parameters and other random variables



Höhna, Heath, Boussau, Landis, Ronquist, Huelsenbeck. 2014.  
**Probabilistic Graphical Model Representation in Phylogenetics.**  
*Systematic Biology.* (doi: 10.1093/sysbio/syu039)