Bayesian Divergence-Time Estimation

Tracy Heath

Ecology, Evolution, & Organismal Biology
Iowa State University

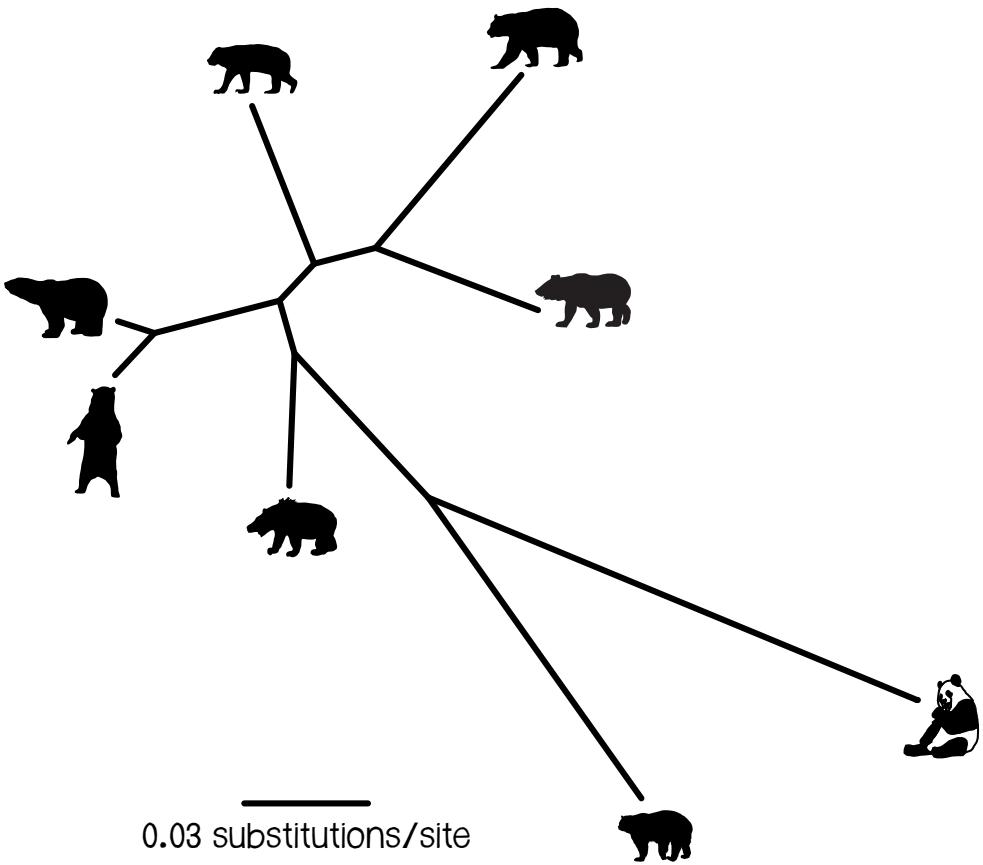
@trayc7

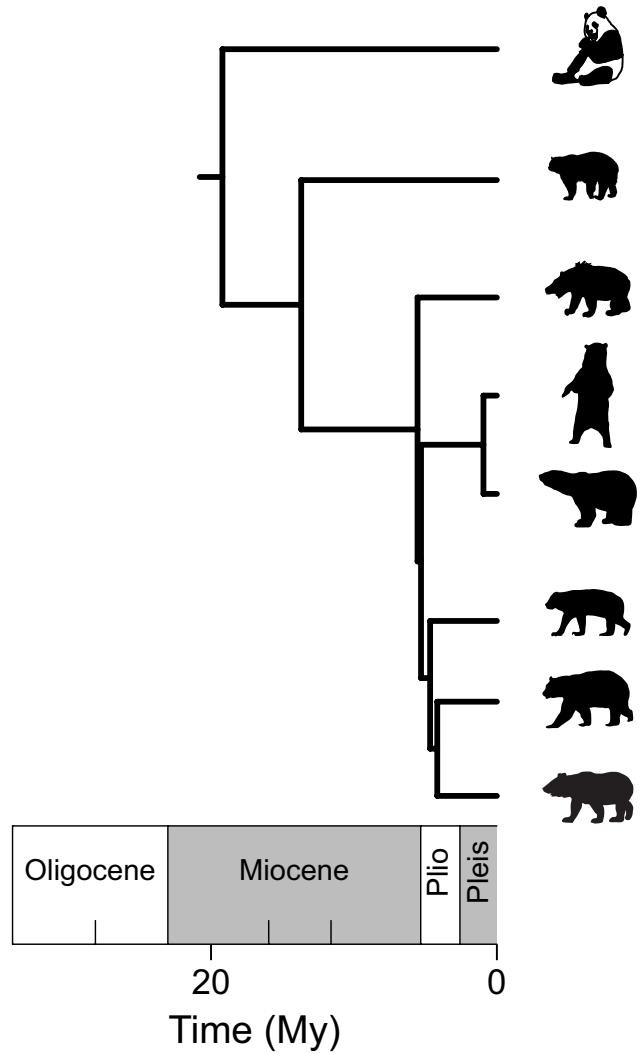
http://phyloworks.org

Workshop on Phylogenomics
Český Krumlov, CZ
January 28, 2019

A Time-Scale for Evolution

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

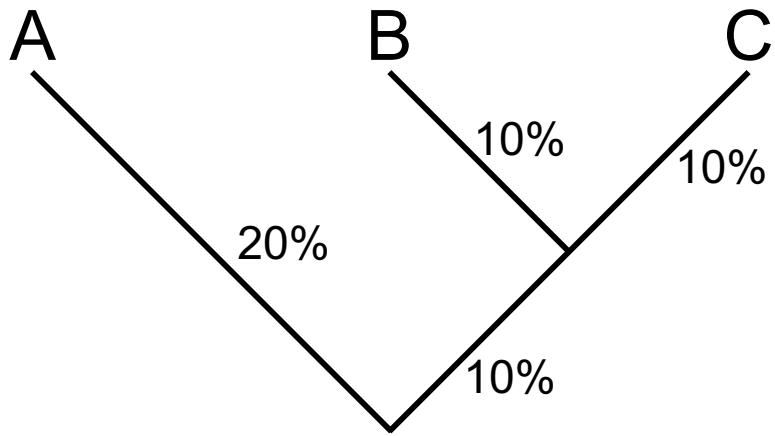




The Global Molecular Clock

Assume that the rate of evolutionary change is constant over time

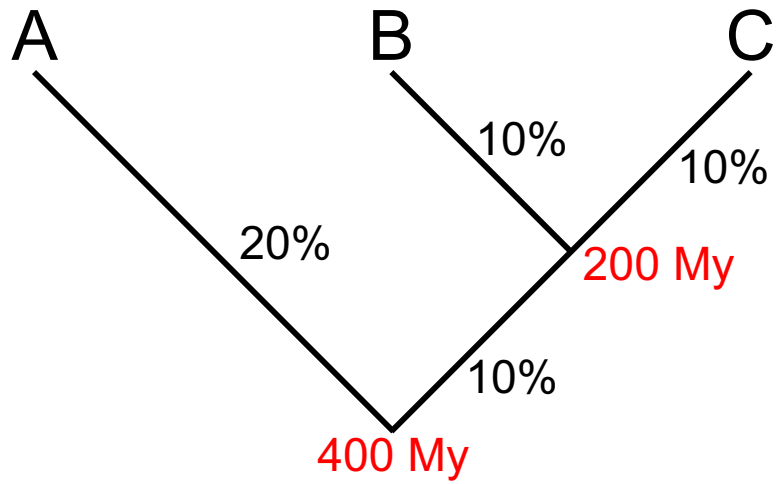
(branch lengths equal percent sequence divergence)



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

The Global Molecular Clock

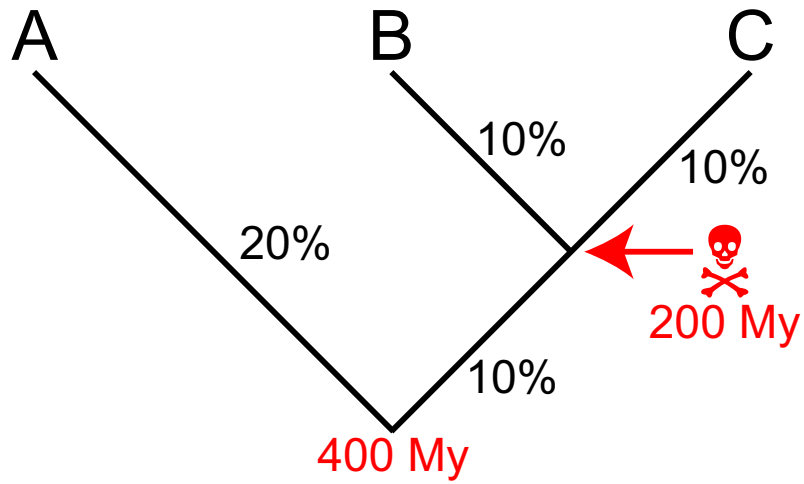
We can date the tree if we know the rate of change is 1% divergence per 10 My



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

The Global Molecular Clock

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



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

Rejecting the Global Molecular Clock

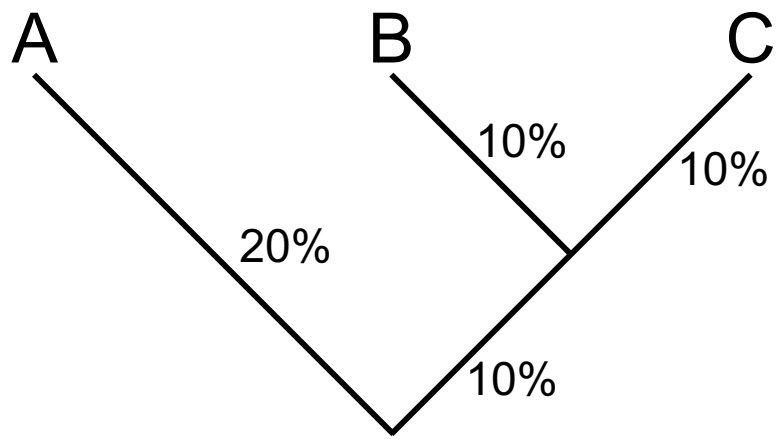
Rates of evolution vary across lineages and over time

Mutation rate:
Variation in

- metabolic rate
- generation time
- DNA repair

Fixation rate:
Variation in

- strength and targets of selection
- population sizes

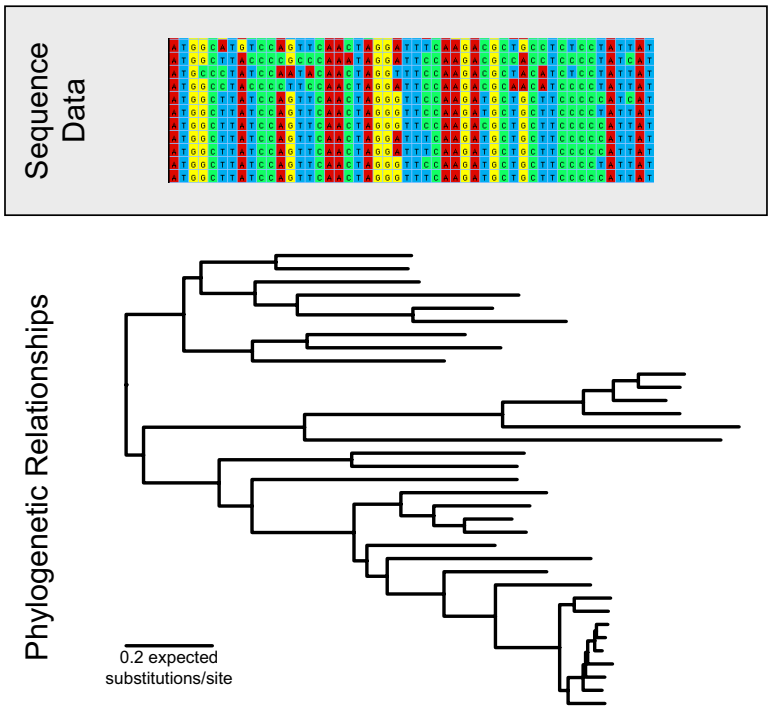


Unconstrained Analysis

Sequence data provide information about **branch lengths**

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

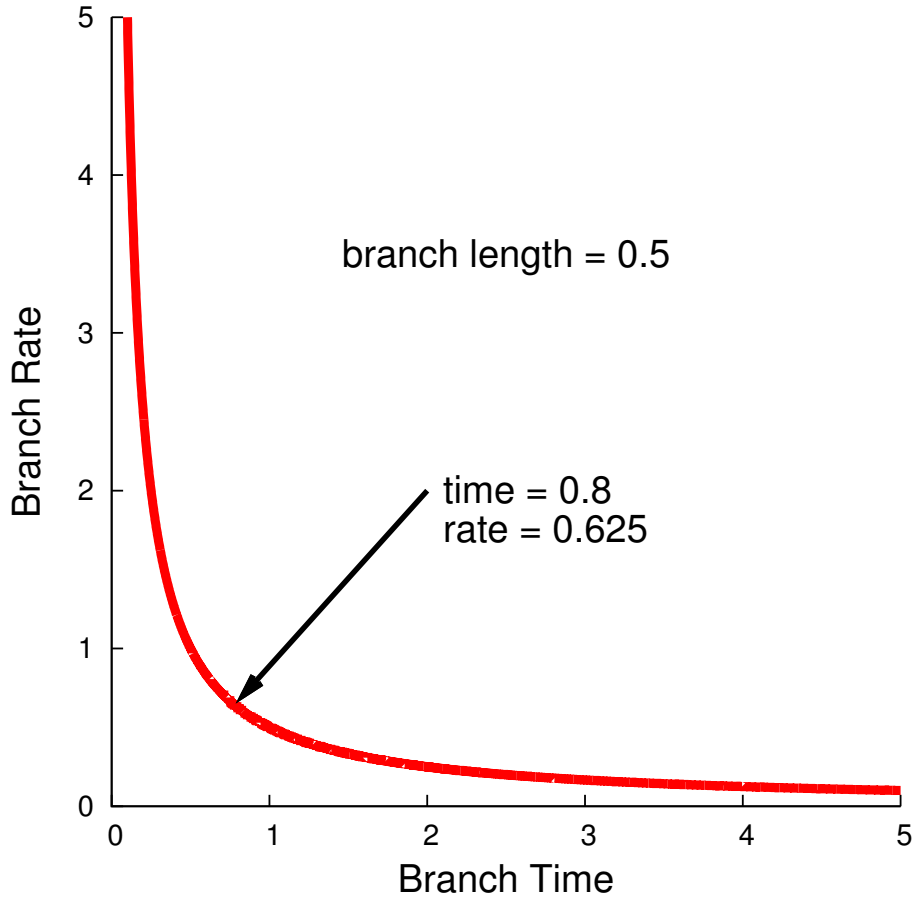
branch length = rate × time



Estimating Rate & Time

The sequence data provide information about branch length

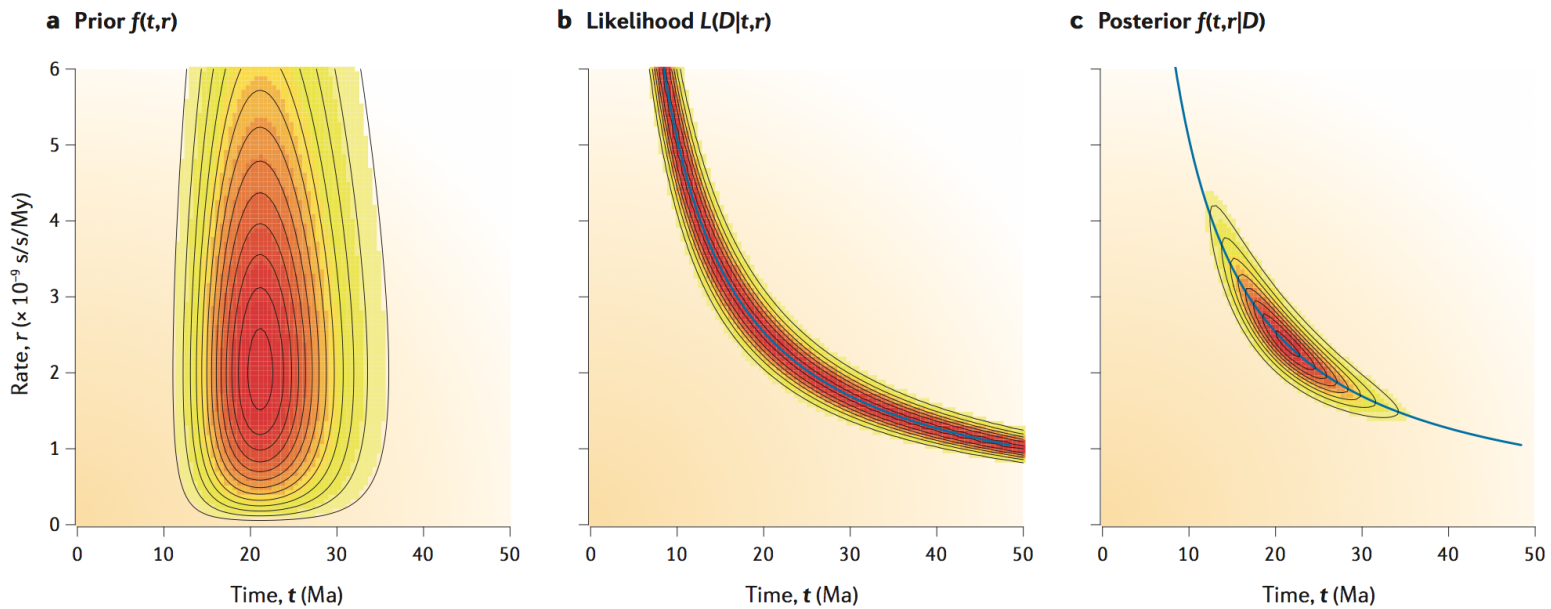
for any possible rate, there's a time that fits the branch length perfectly



(figure based on Thorne & Kishino, 2005)

Estimating Rate & Time

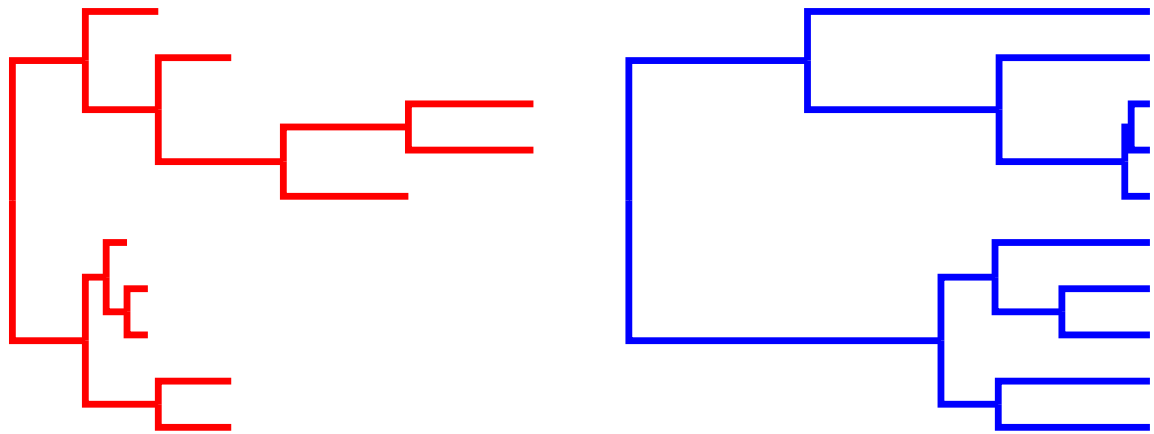
Methods for dating species divergences estimate the substitution rate and time separately



(dos Reis et al. *Nature Reviews Genetics*, 2016)

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

Bayesian Divergence Time Estimation

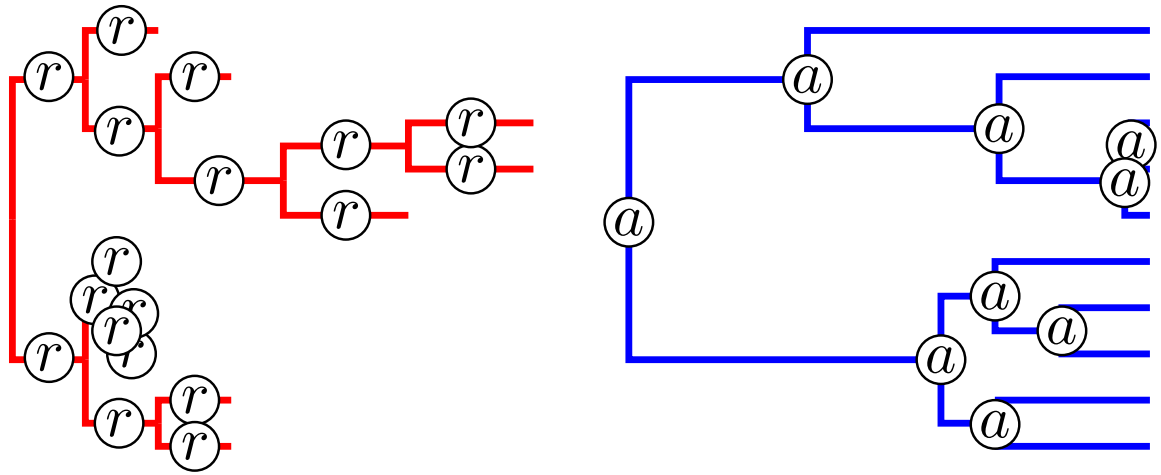


$$\boldsymbol{\mathcal{R}} = (r_1, r_2, r_3, \ldots, r_{2N-2})$$

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

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

Bayesian Divergence Time Estimation



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

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

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

Bayesian Divergence Time Estimation

Posterior probability

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

$\boldsymbol{\mathcal{R}}$	Vector of rates on branches
$\boldsymbol{\mathcal{A}}$	Vector of internal node ages
$\theta_{\mathcal{R}}, \theta_{\mathcal{A}}, \theta_s$	Model parameters
D	Molecular or morphology data
$\mathcal{T}$	Tree topology

Bayesian Divergence Time Estimation

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

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

Modeling Rate Variation

Some models describing lineage-specific substitution rate variation:

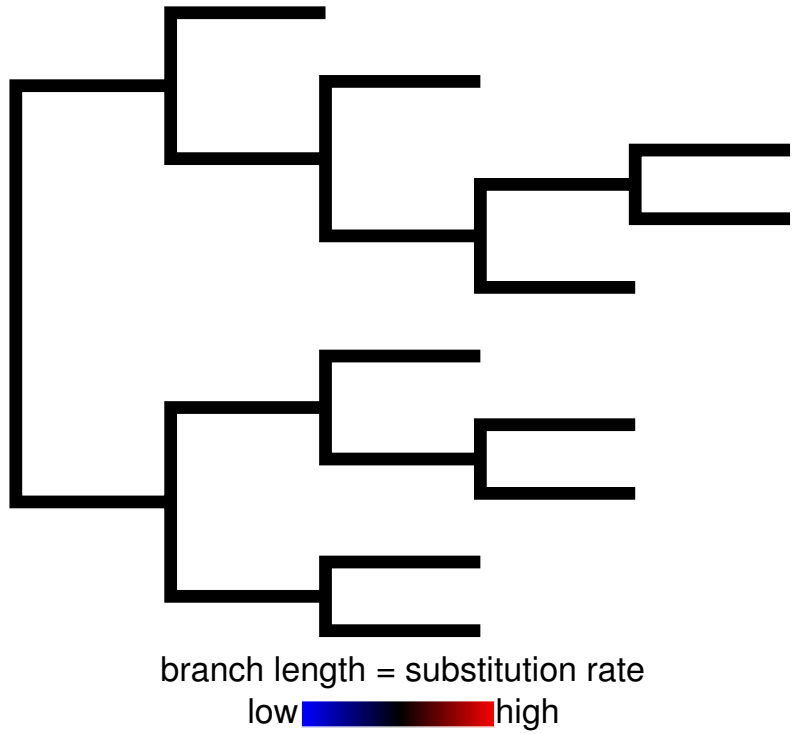
- **Global/strict clock** (Zuckerkandl & Pauling, 1962)
- **Local clocks** (Hasegawa, Kishino & Yano 1989; Kishino & Hasegawa 1990; Yoder & Yang 2000; Yang & Yoder 2003, Drummond and Suchard 2010)
- **Punctuated rate change model** (Huelsenbeck, Larget and Swofford 2000)
- **Autocorrelated rates** (Thorne,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

Global Clock

The substitution rate is constant over time

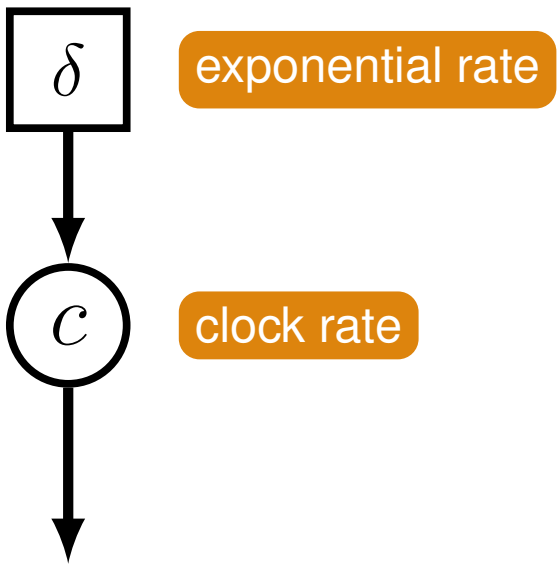
All lineages share the same rate

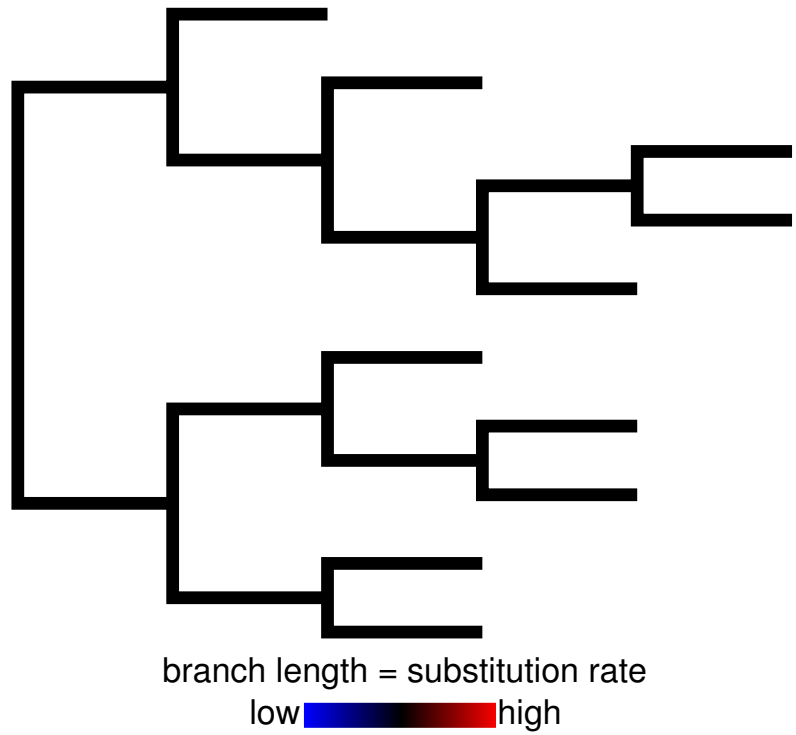
(Zuckerkandl & Pauling, 1962)



Global Clock

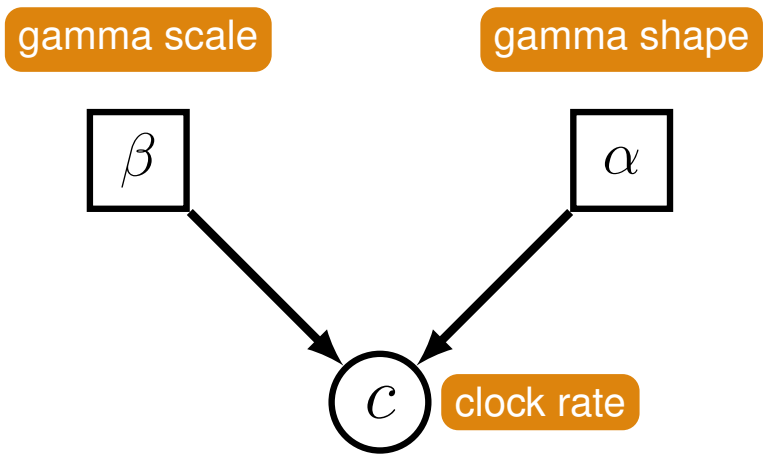
$c \sim \text{Exponential}(\delta)$

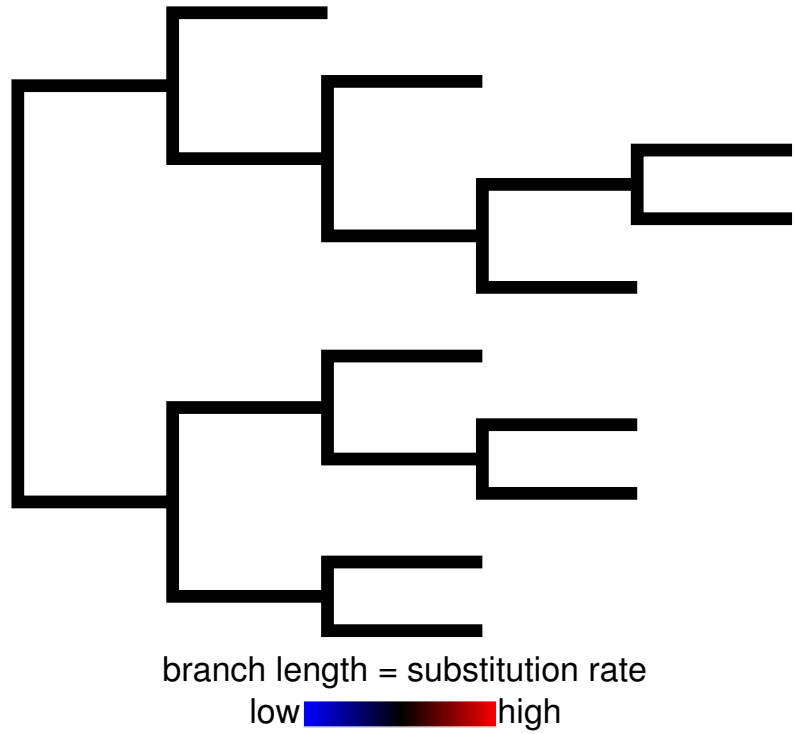




Global Clock

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





Relaxed-Clock Models

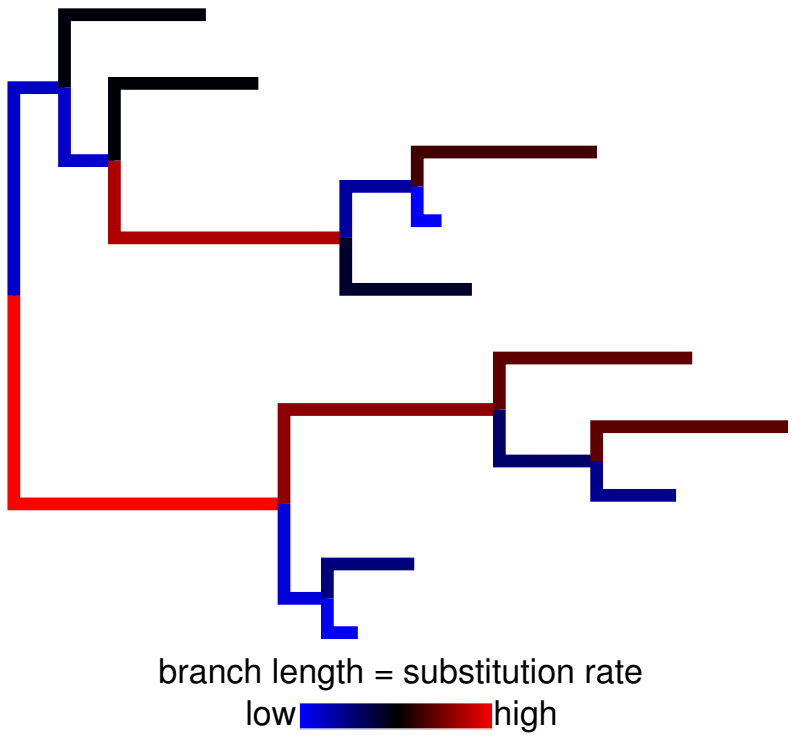
To accommodate variation in substitution rates 'relaxed-clock' models estimate lineage-specific substitution rates

- **Local clocks**
- **Punctuated rate change model**
- **Autocorrelated rates**
- **Mixture models on branch rates**
- **Uncorrelated/independent rates models**

Independent/Uncorrelated Rates

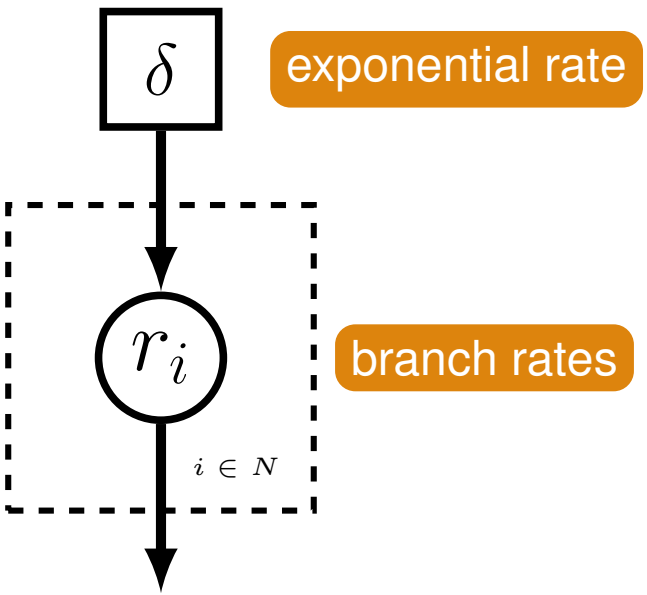
Lineage-specific rates are uncorrelated when the rate assigned to each branch is independently drawn from an underlying distribution

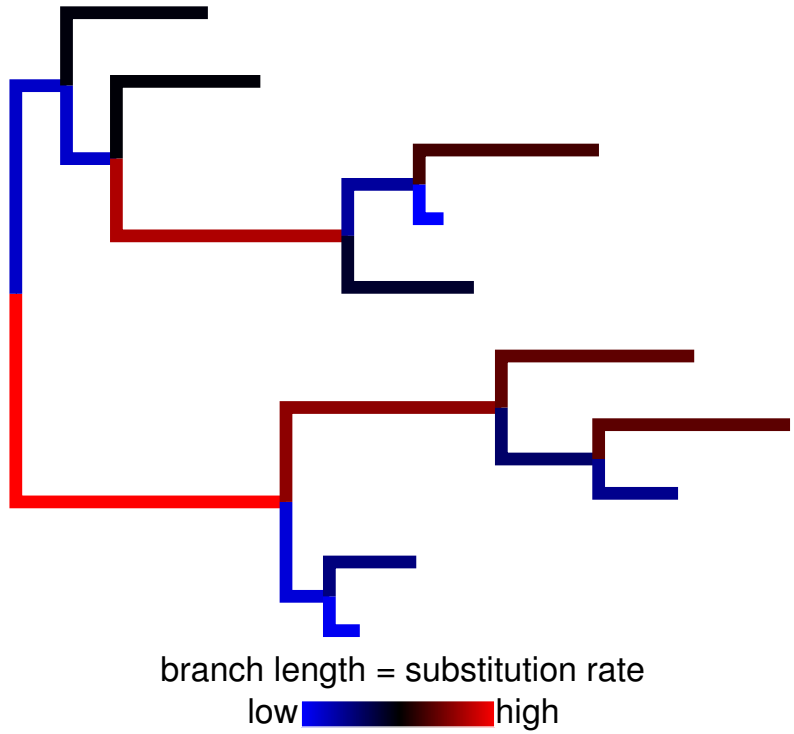
(Drummond et al. 2006; Rannala & Yang 2007; Lepage et al. 2007)



Independent/Uncorrelated Rates

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



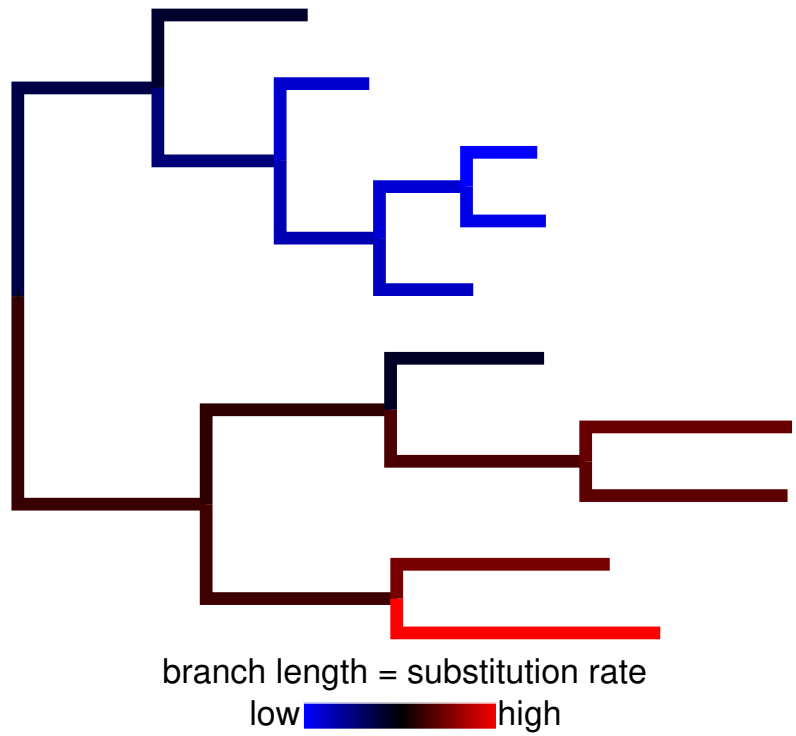


Autocorrelated Rates

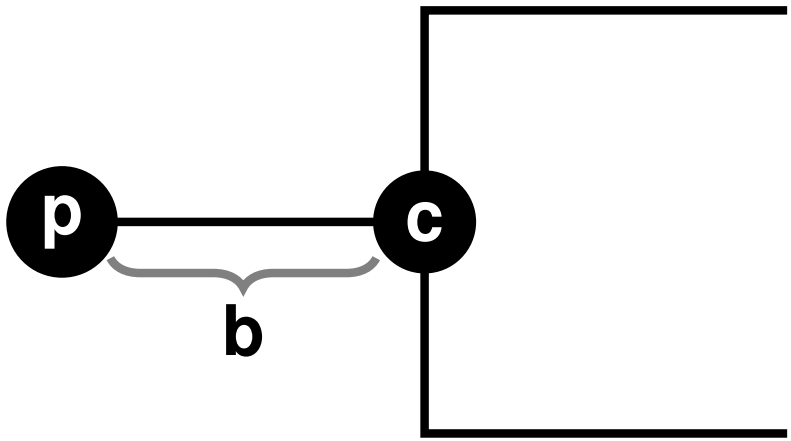
Substitution rates evolve gradually over time — closely related lineages have similar rates

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

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



Autocorrelated Rates

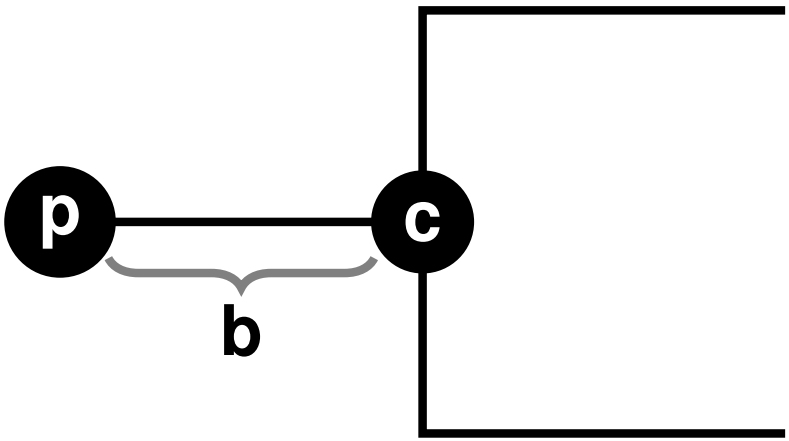


p = parent node

c = child node

b = branch

Autocorrelated Rates



p = parent node

c = child node

b = branch

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

$$\sigma_c := \nu t_b$$

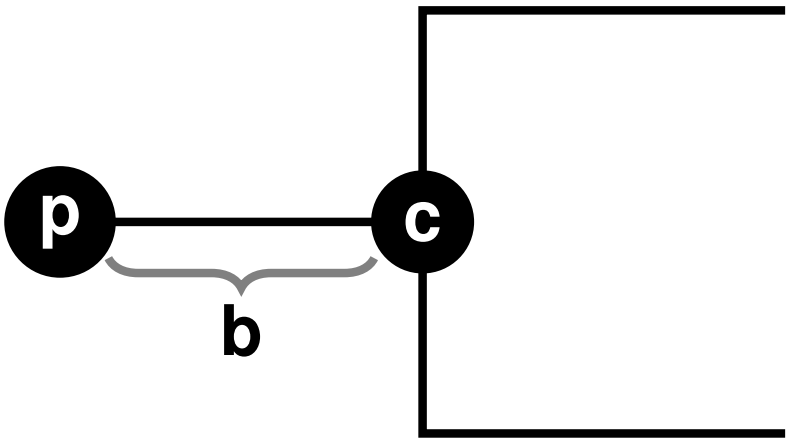
$$\mu_c := \ln(r_p) - \frac{\sigma_c^2}{2}$$

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

ν = variance parameter

t_b = time duration of branch

Autocorrelated Rates



p = parent node

c = child node

b = branch

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

$$\sigma_c := \nu t_b$$

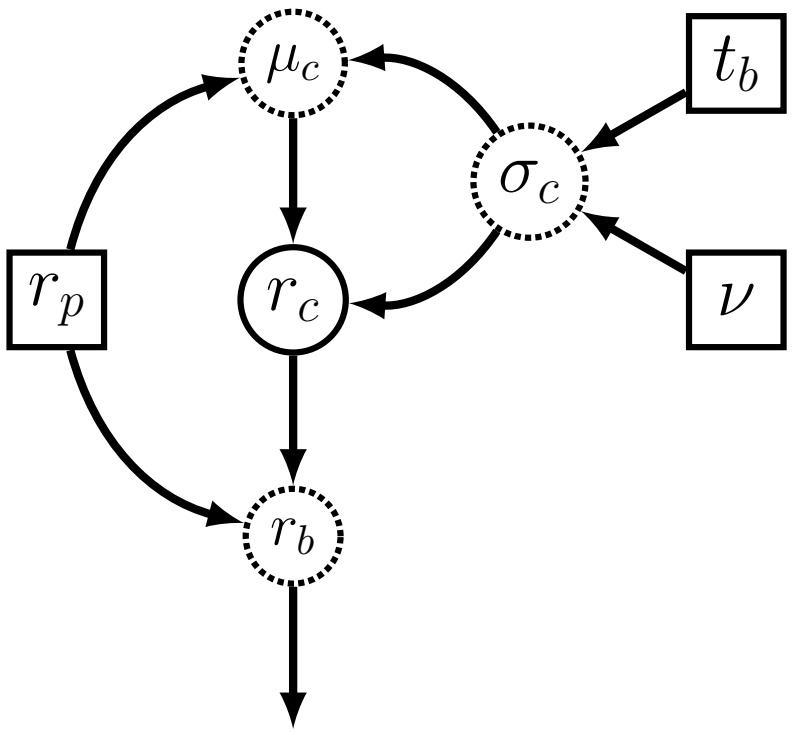
$$\mu_c := \ln(r_p) - \frac{\sigma_c^2}{2}$$

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

ν = variance parameter

t_b = time duration of branch

Autocorrelated Rates



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

$$\sigma_c := \nu t_b$$

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

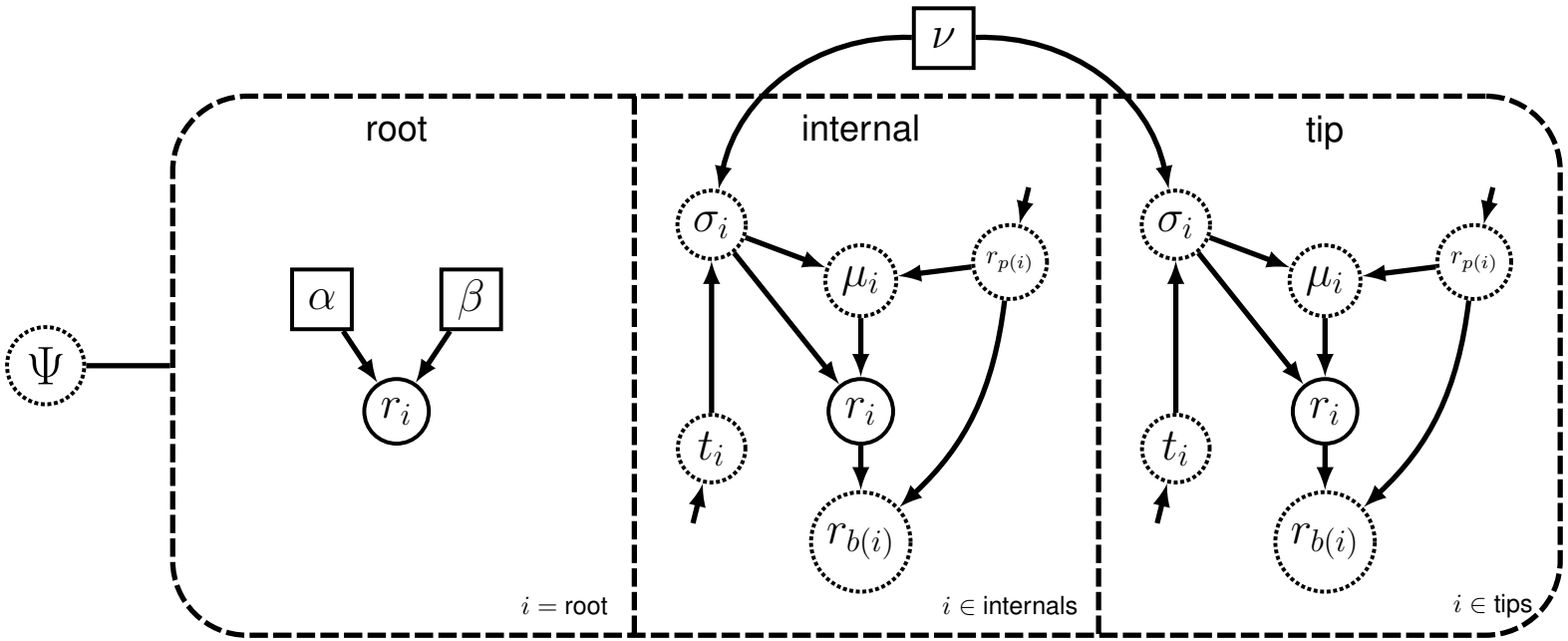
$$r_b := \frac{r_p + r_c}{2}$$

$\nu =$ variance parameter

$t_b =$ time duration of branch

Autocorrelated Rates

The rate associated with each node is a stochastic node, drawn from a distribution centered on its parent node



There is a gamma prior distribution on the rate at the root node

Modeling Rate Variation

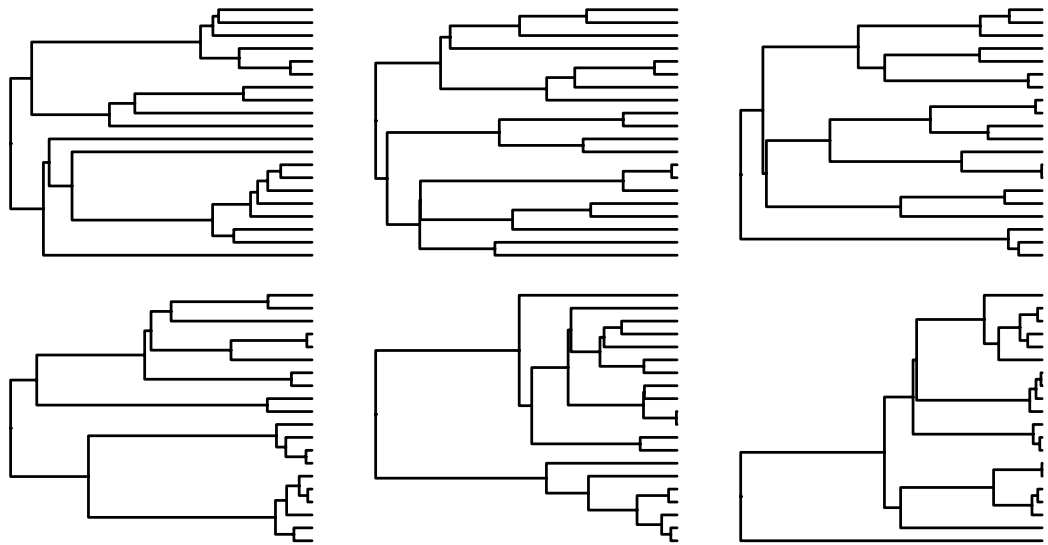
These are only a subset of the available models for branch-rate variation

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

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

Priors on the Tree and Node Ages

Relaxed clock Bayesian analyses require a prior distribution on time trees



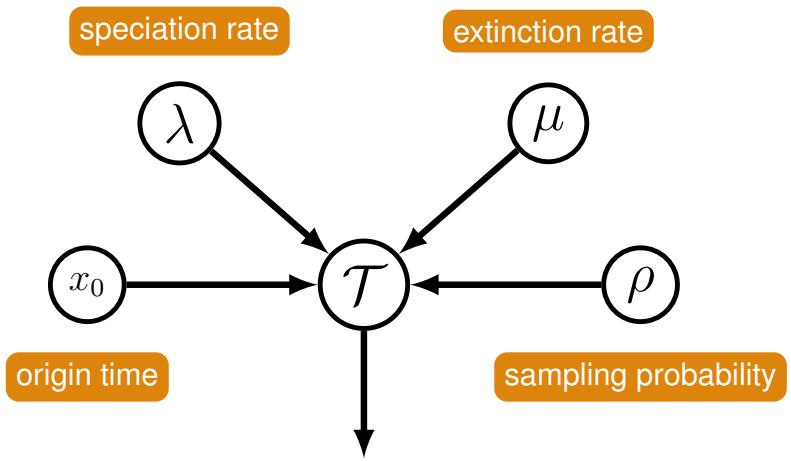
Different tree priors make different assumptions about the timing of divergence events and shape of the tree topology

Stochastic Branching Processes

Tree priors based on stochastic models of lineage diversification

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

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

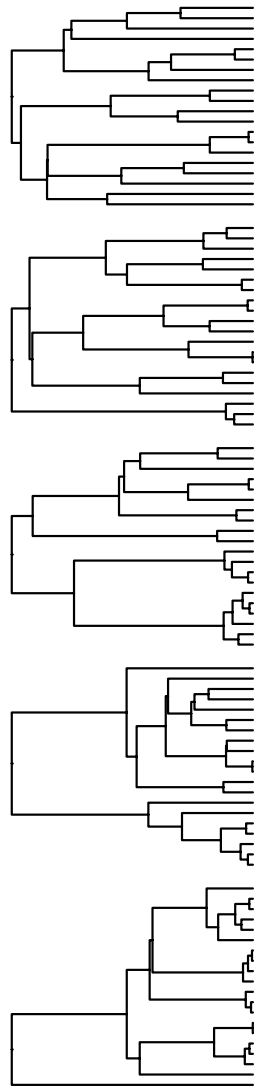


Stochastic Branching Processes

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

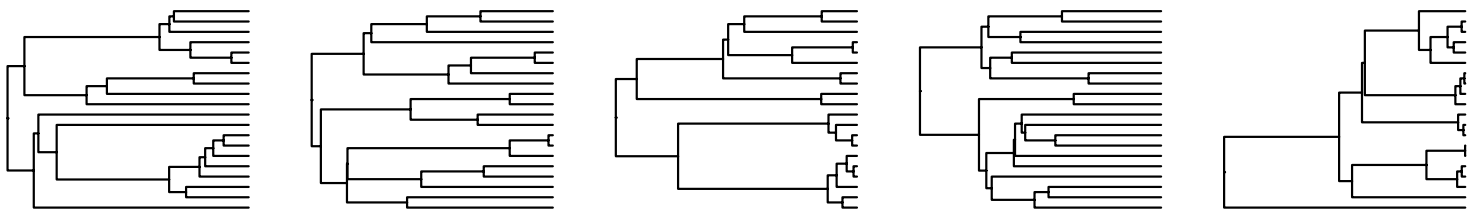
Using hyperpriors on λ and μ (or d and r) accounts for uncertainty in these hyperparameters



Priors on the Tree and Node Ages

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

Most tree priors cannot give precise estimates of *absolute* node ages



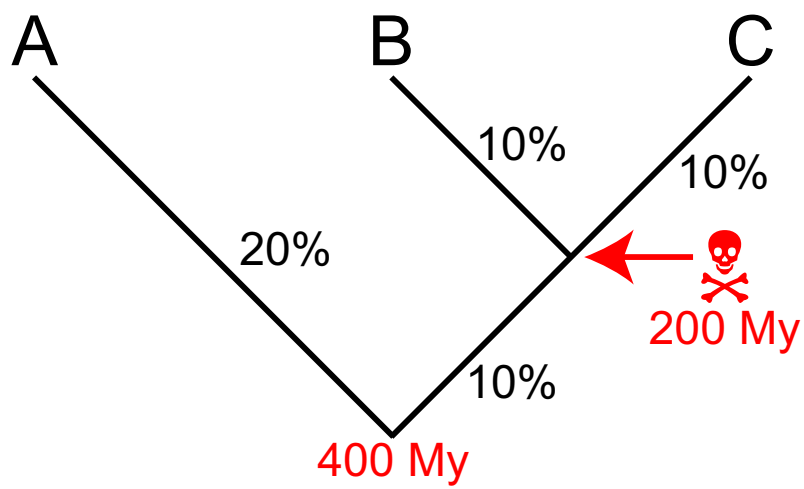
We need additional data (like fossils) to provide absolute time scale

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

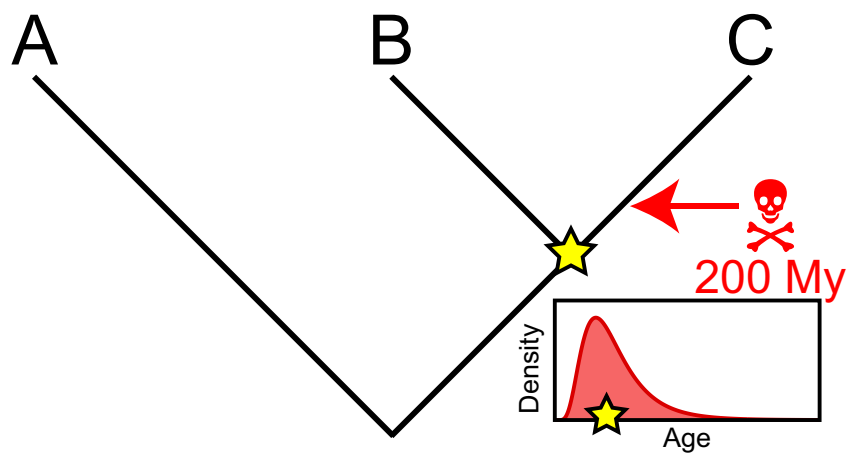
Uncertainty in the placement of fossils



Calibration Densities

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

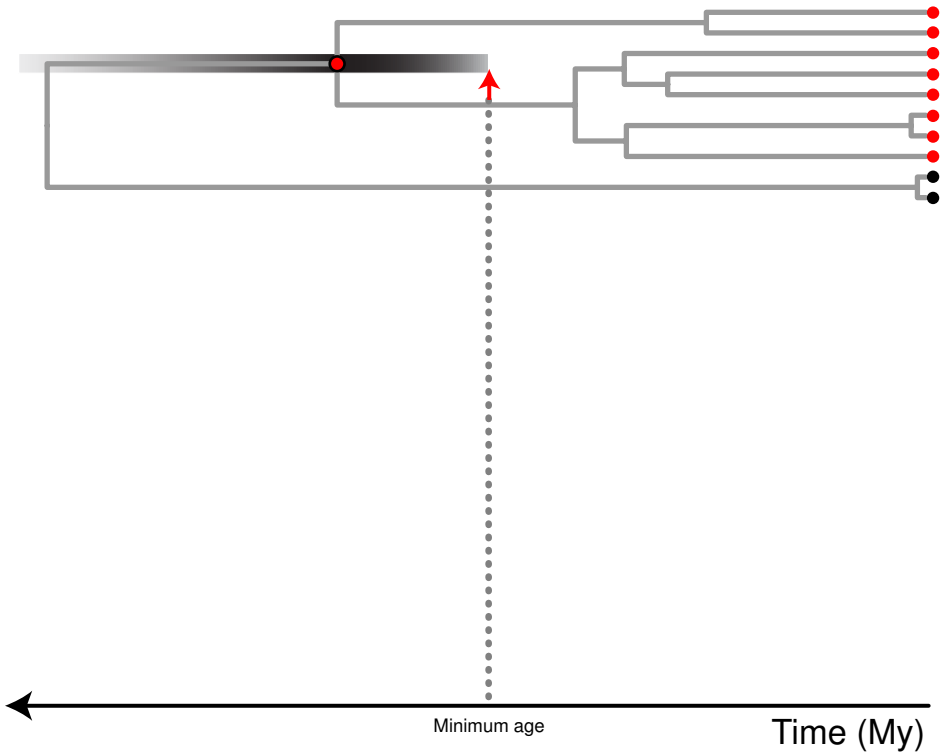
Divergence times are calibrated by placing parametric densities on internal nodes offset by age estimates from the fossil record



Fossil Calibration

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

Reliable **maximum** bounds are typically unavailable

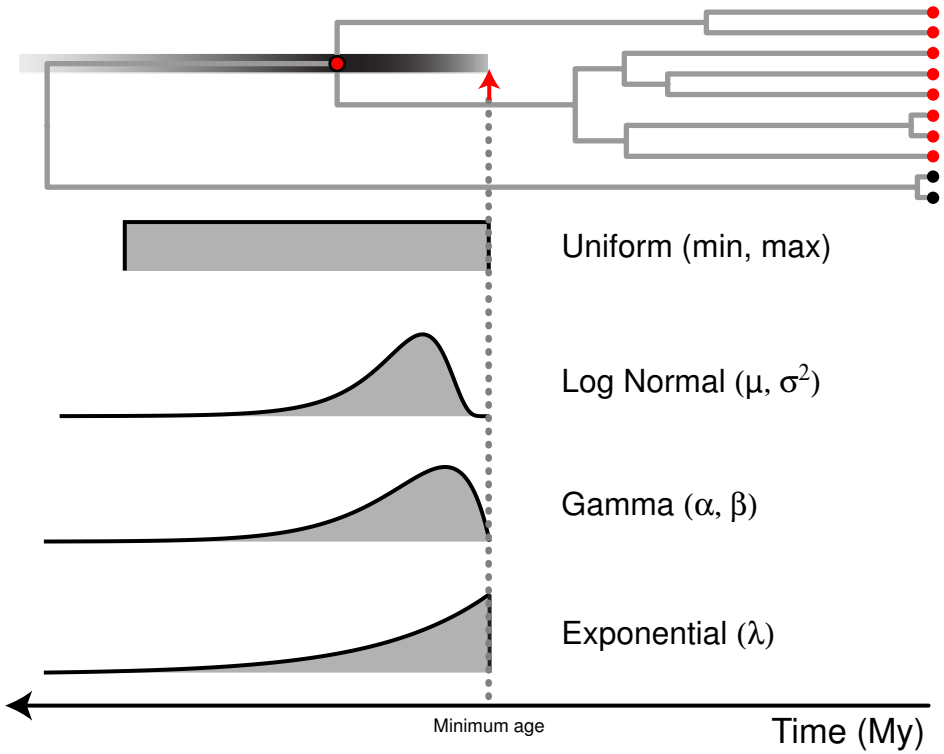


Densities on Calibrated Nodes

Common practice in Bayesian divergence-time estimation:

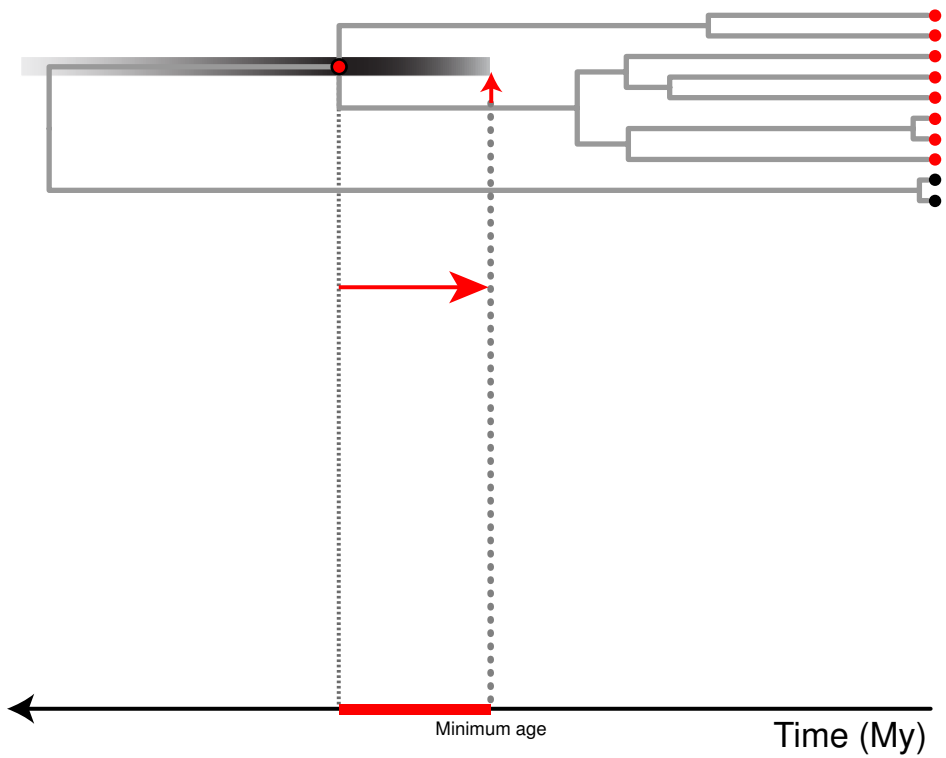
Parametric distributions are typically off-set by the age of the oldest fossil assigned to a clade

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



Densities on Calibrated Nodes

Typically interpreted as a "prior" on the waiting time between the divergence event and the age of the oldest fossil

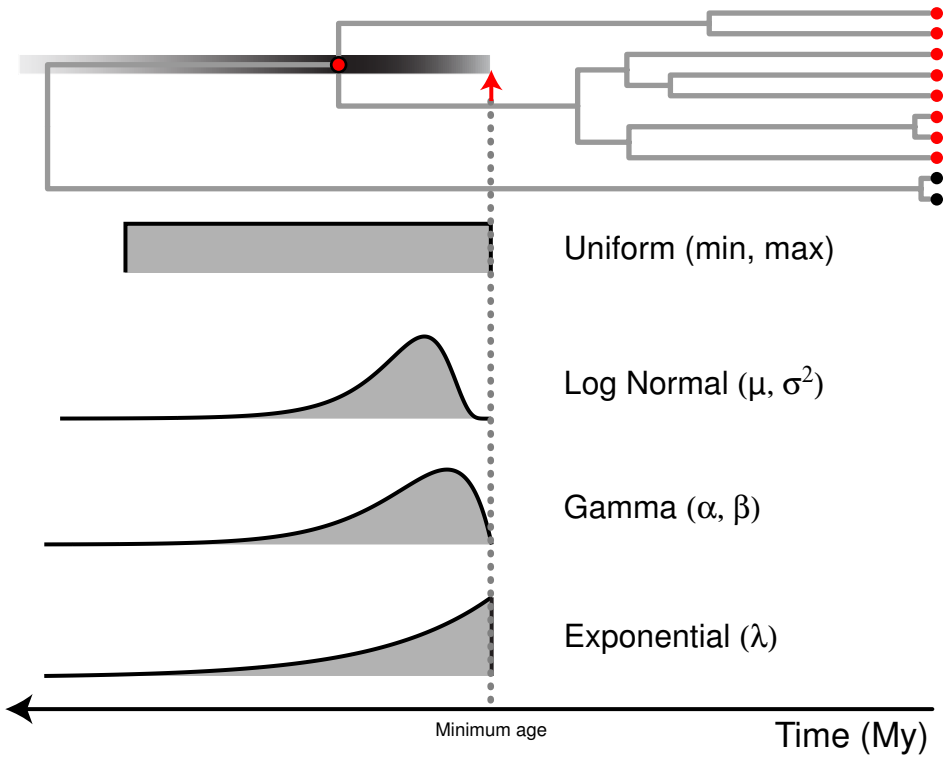


Densities on Calibrated Nodes

Common practice in Bayesian divergence-time estimation:

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

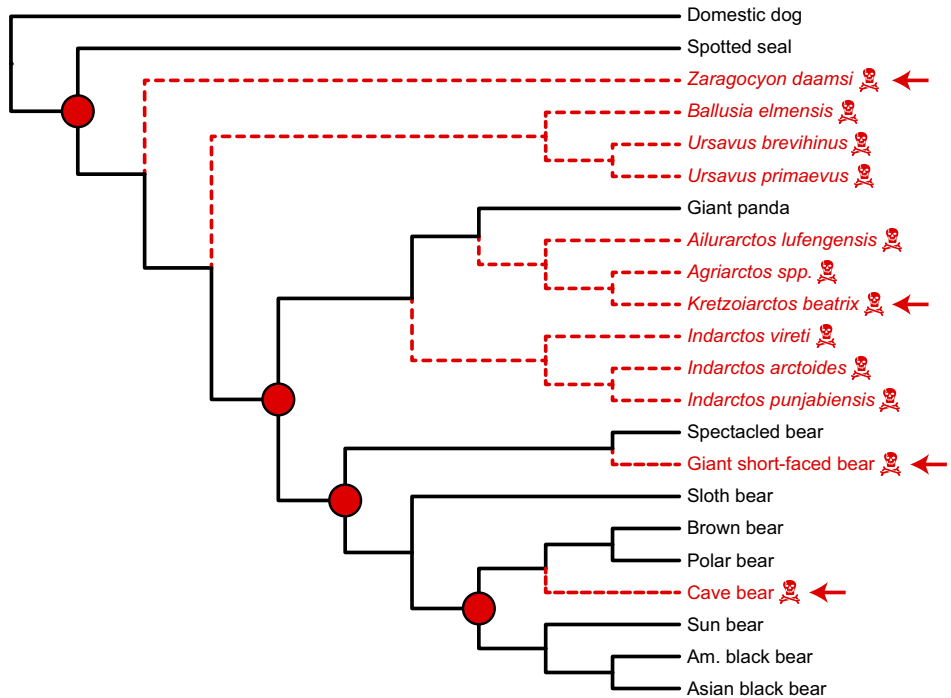
Specifying appropriate densities is a challenge for most molecular biologists



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



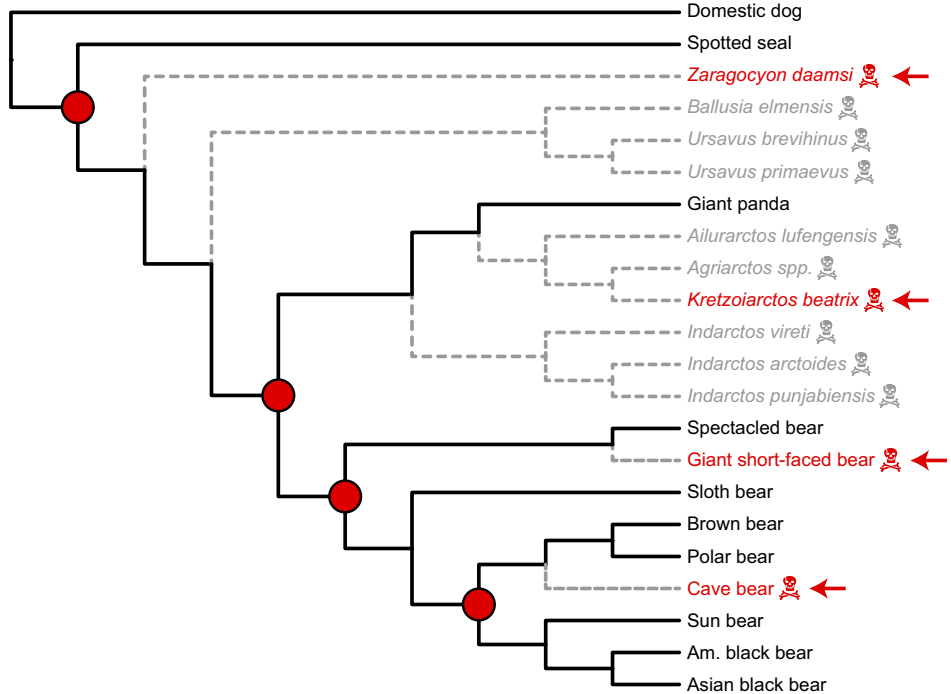
(Krause et al. *BMC Evol. Biol.* 2008; Abella et al. *PLoS ONE* 2012)

Improving Fossil Calibration

We want to use all of the available fossils

Example: Bears

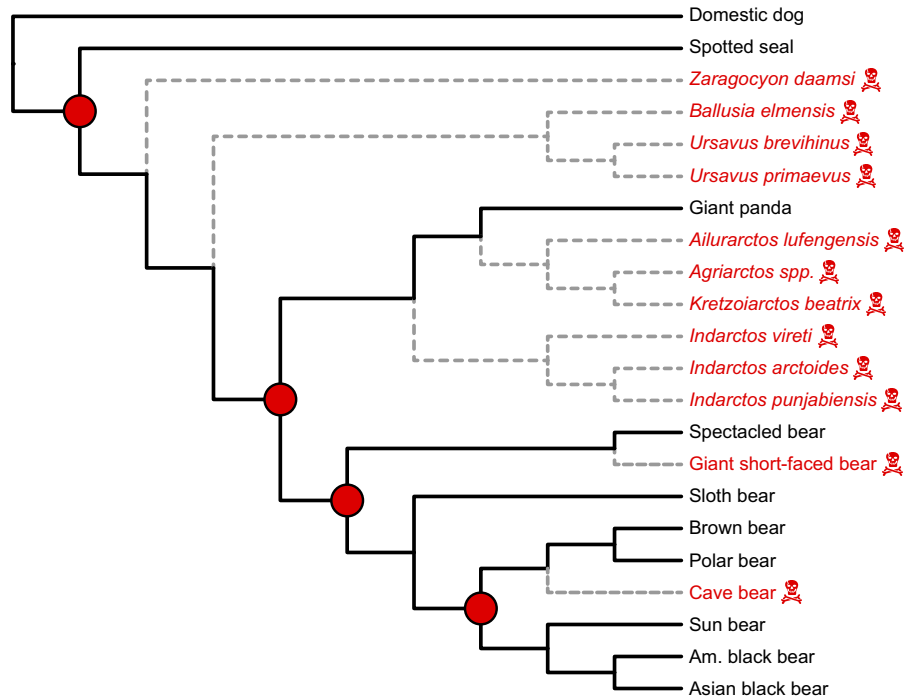
12 fossils are reduced to 4 calibration ages with calibration density methods



(Krause et al. *BMC Evol. Biol.* 2008; Abella et al. *PLoS ONE* 2012)

Improving Fossil Calibration

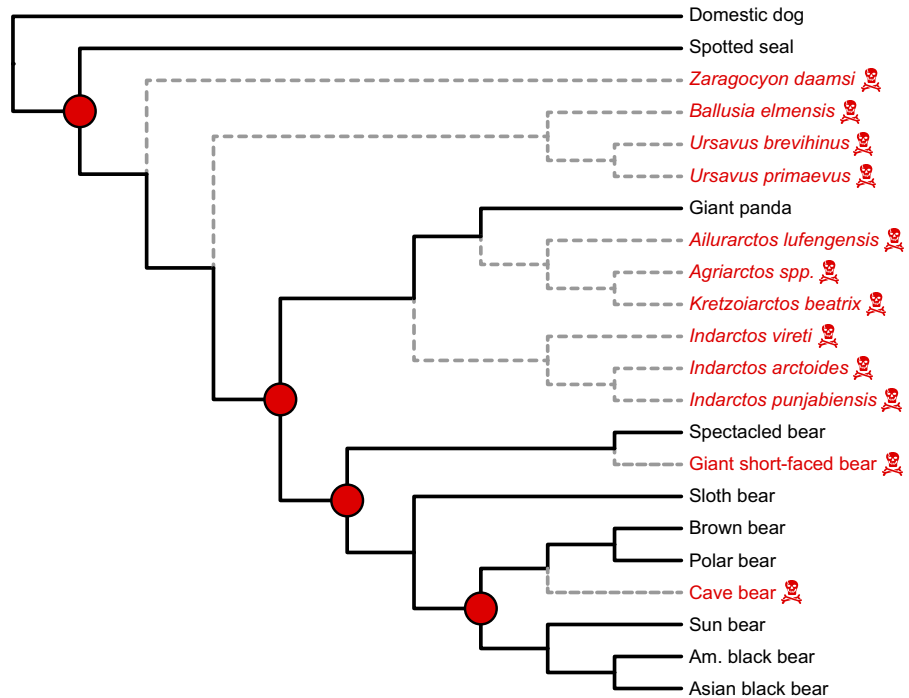
Because fossils are part of the diversification process, we can combine fossil calibration with birth-death models



(Krause et al. *BMC Evol. Biol.* 2008; Abella et al. *PLoS ONE* 2012)

Improving Fossil Calibration

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



(Krause et al. *BMC Evol. Biol.* 2008; Abella et al. *PLoS ONE* 2012)

Paleontology & Neontology

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

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

Paleontology & Neontology

Biology and Philosophy **19:** 687–720, 2004.

The role of fossils in phylogeny reconstruction: Why is it so difficult to integrate paleobiological and neontological evolutionary biology?

TODD GRANTHAM
Department of Philosophy, College of Charleston, Charleston, SC 29424, USA
(e-mail: granthamt@cofc.edu)

Combining Fossil & Extant Data

Statistical methods provide a way to integrate paleontological & neontological data

Syst. Biol. 50(6):913–925, 2001

A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data

Paul O. Lewis

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

Syst. Biol. 61(6):973–999, 2012

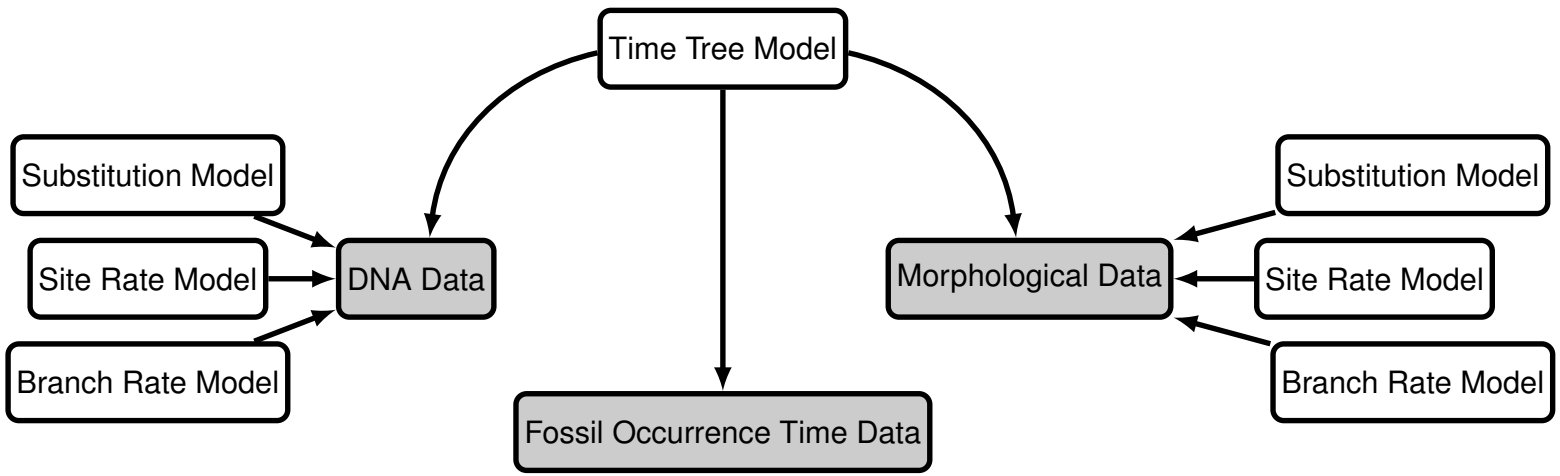
DOI:10.1093/sysbio/sys058
Advance Access publication on June 20, 2012

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

Fredrik Ronquist[1,*], Seraina Klopfstein[1], Lars Vilhelmsen[2], Susanne Schulmeister[3], Debra L. Murray[4], and Alexandr P. Rasnitsyn[5]

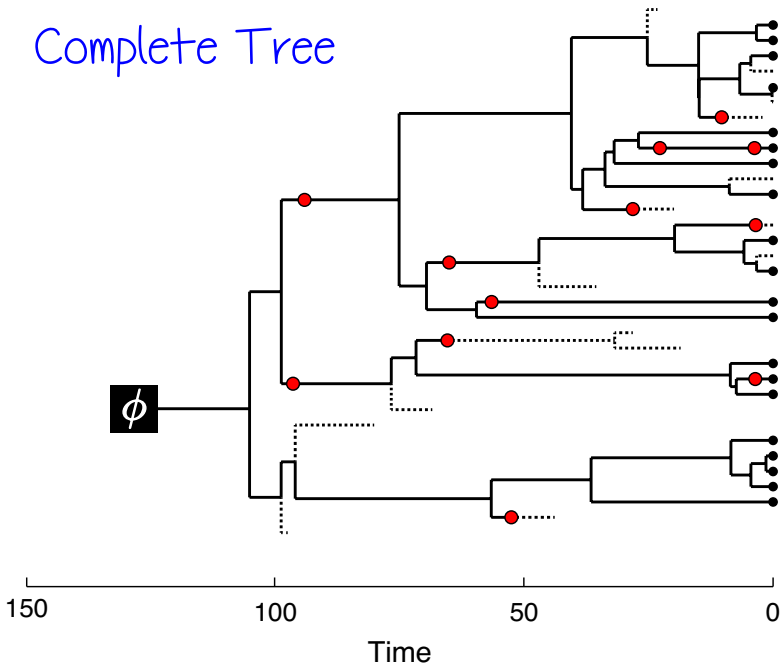
Combining Fossil & Extant Data

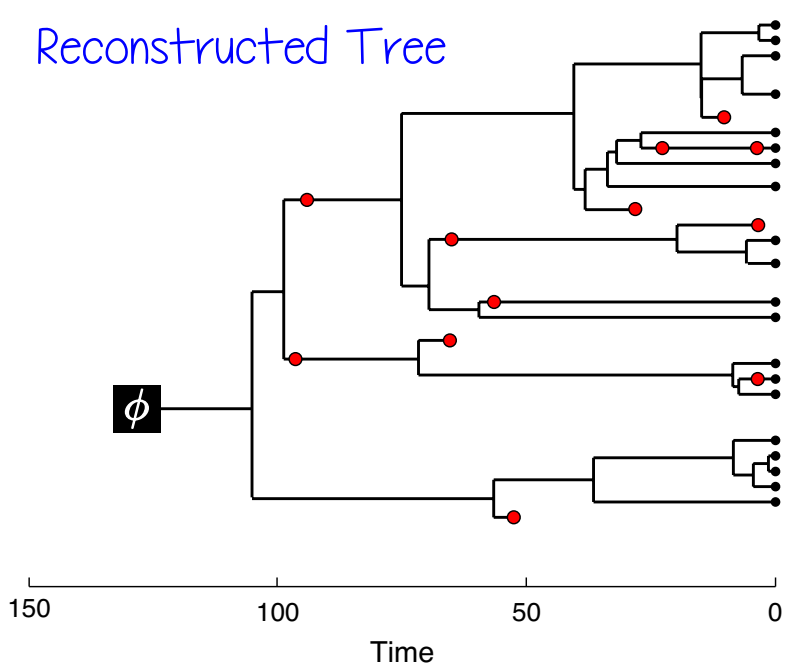
Combine models for sequence evolution, morphological change, & fossil recovery to jointly estimate the tree topology, divergence times, & lineage diversification rates



Modeling the Tree & Occurrence Times

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

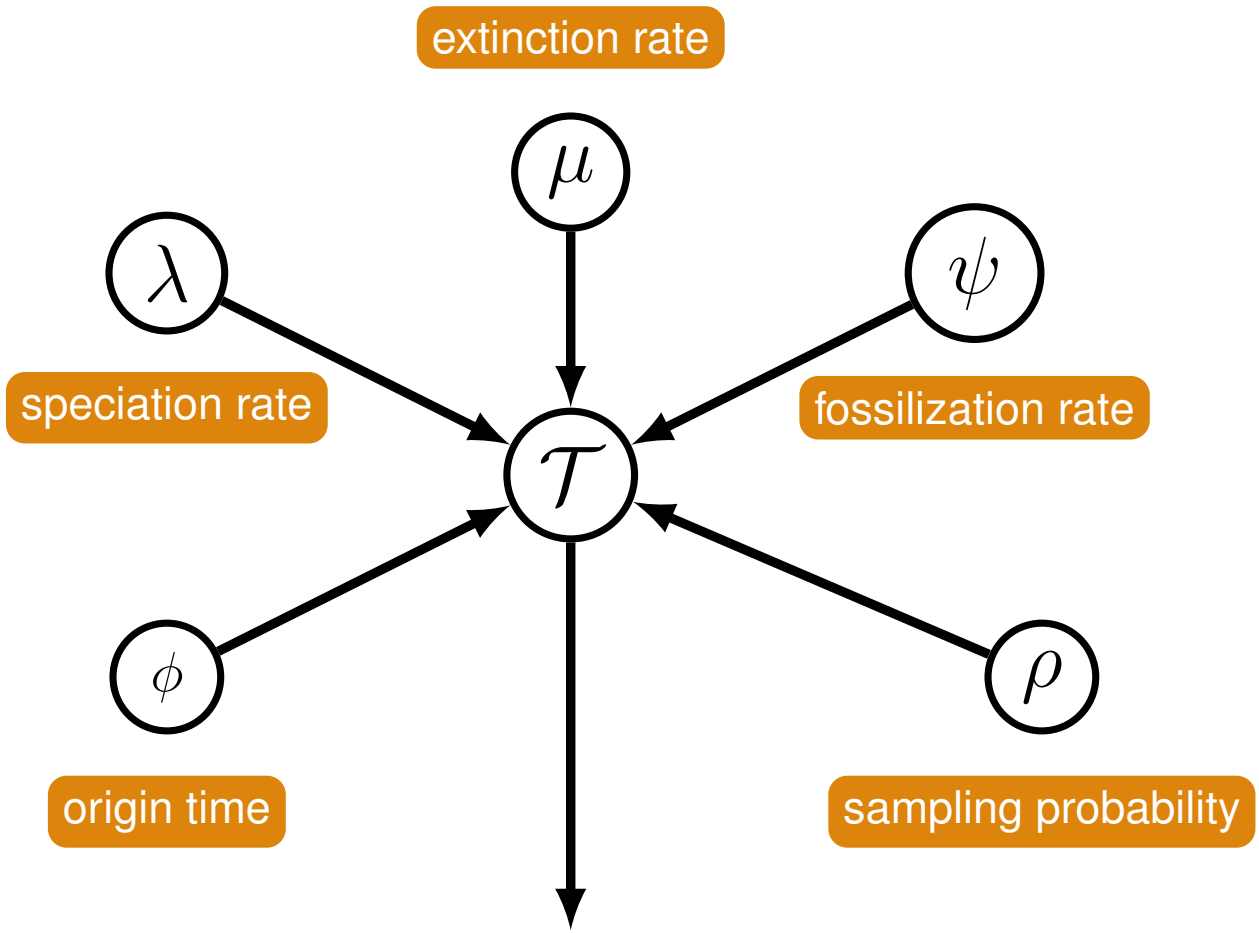




(Stadler. *Journal of Theoretical Biology* 2010)

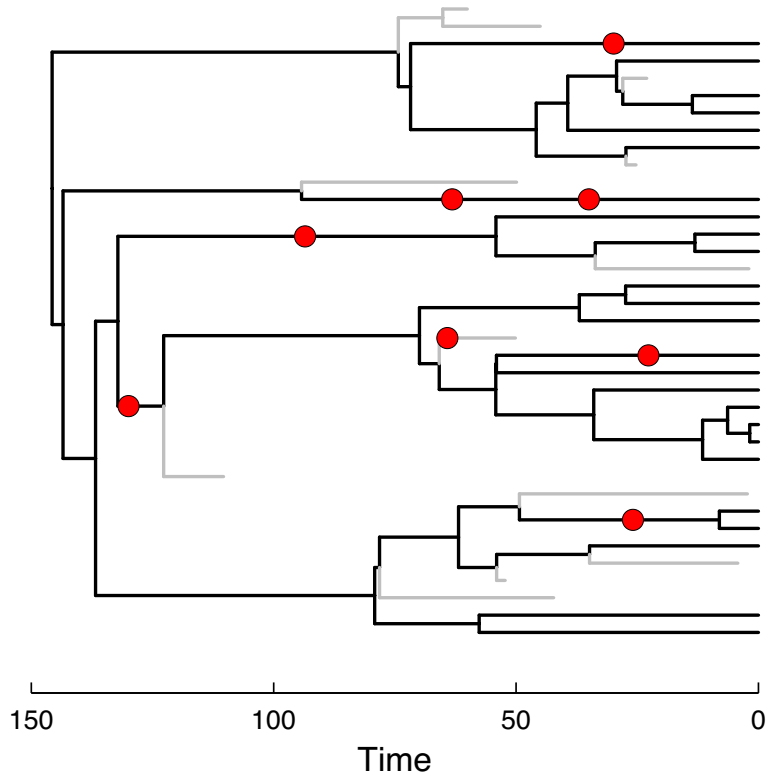
Parameters of the FBD

This graph shows the conditional dependence structure of the FBD model, which is a generating process for a sampled, dated time tree and fossil occurrences



The Fossilized Birth-Death Process (FBD)

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



(Heath, Huelsenbeck, Stadler. *PNAS* 2014)

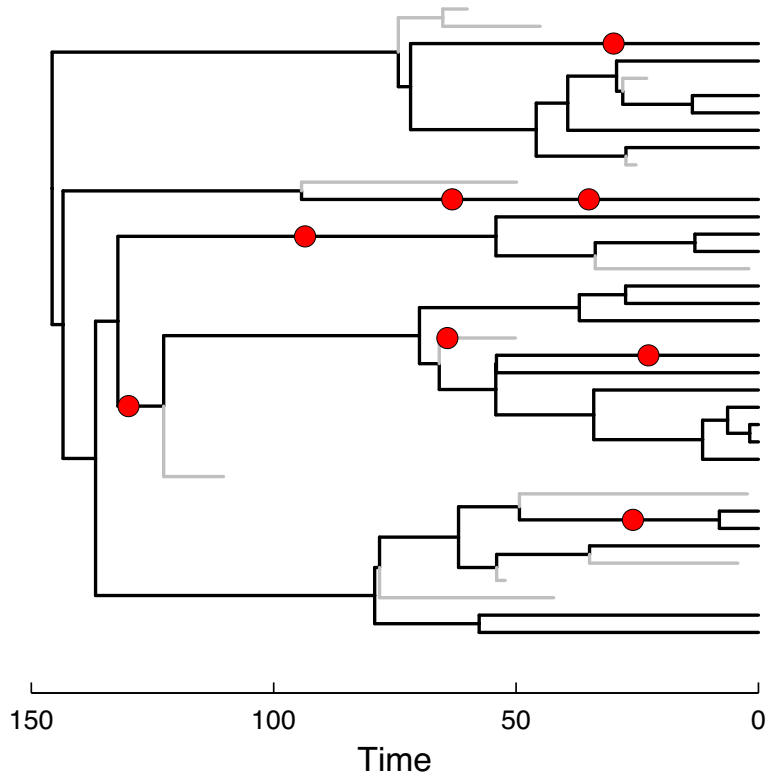
The Fossilized Birth-Death Process (FBD)

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

λ = speciation

μ = extinction

ψ = fossilization/recovery



(Heath, Huelsenbeck, Stadler. *PNAS* 2014)

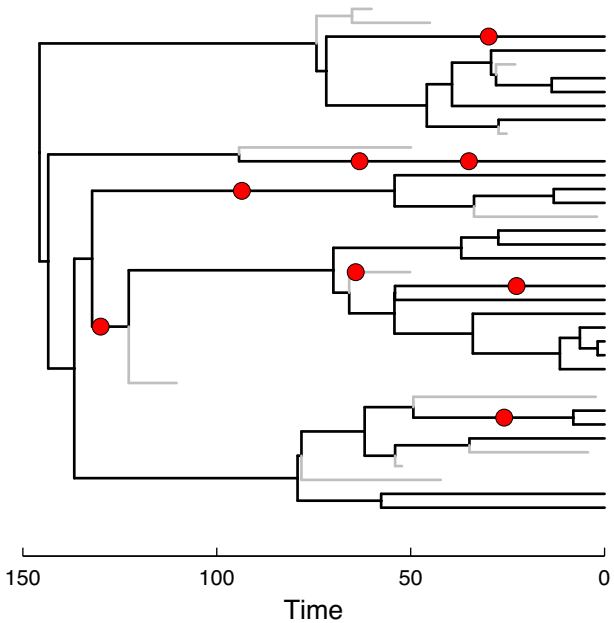
Sampled Ancestors

Sampled lineages with sampled descendants

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

On the probability of ancestors in the fossil record

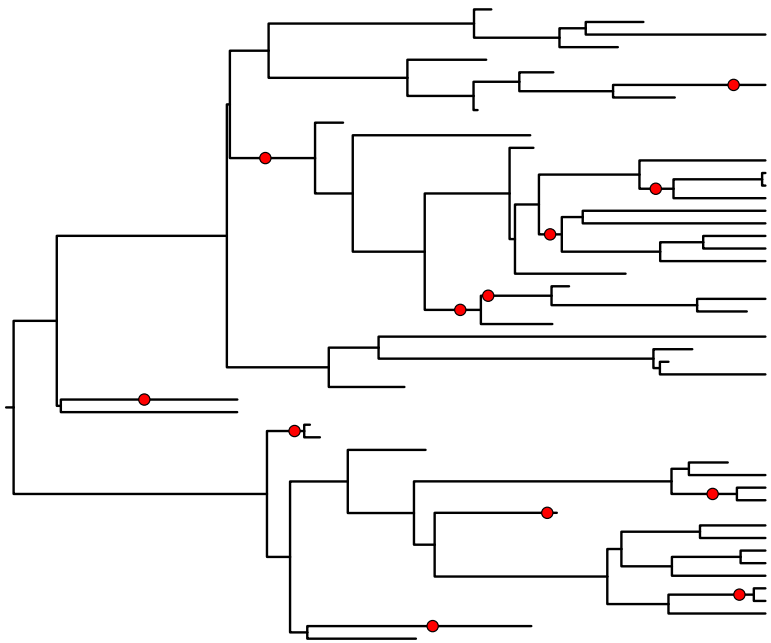
Mike Foote



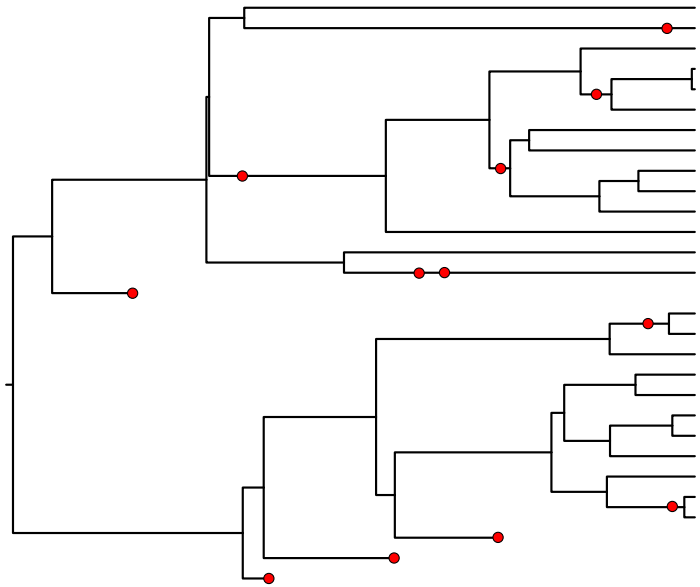
There is a non-zero probability of sampling ancestor-descendant relationships from the fossil record

Sampled Ancestors

Complete FBD Tree



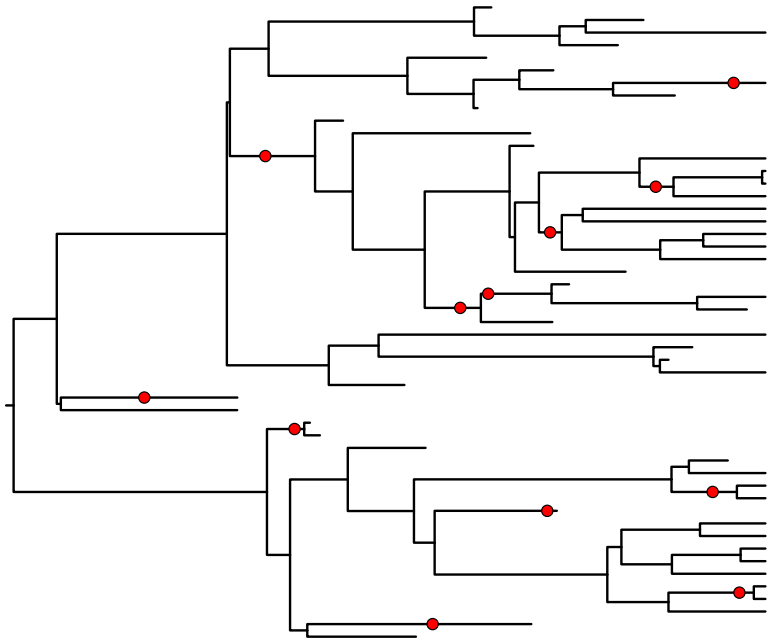
Reconstructed FBD Tree



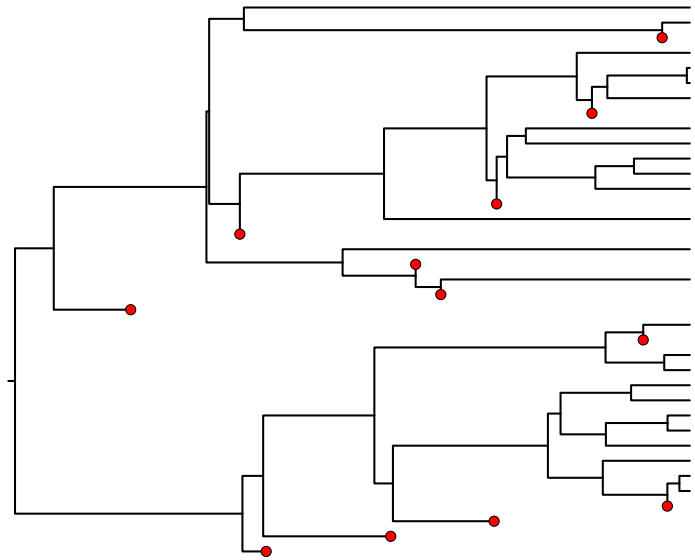
Because fossils & living taxa are assumed to come from a single diversification process, there is a non-zero probability of sampled ancestors

Sampled Ancestors

Complete FBD Tree

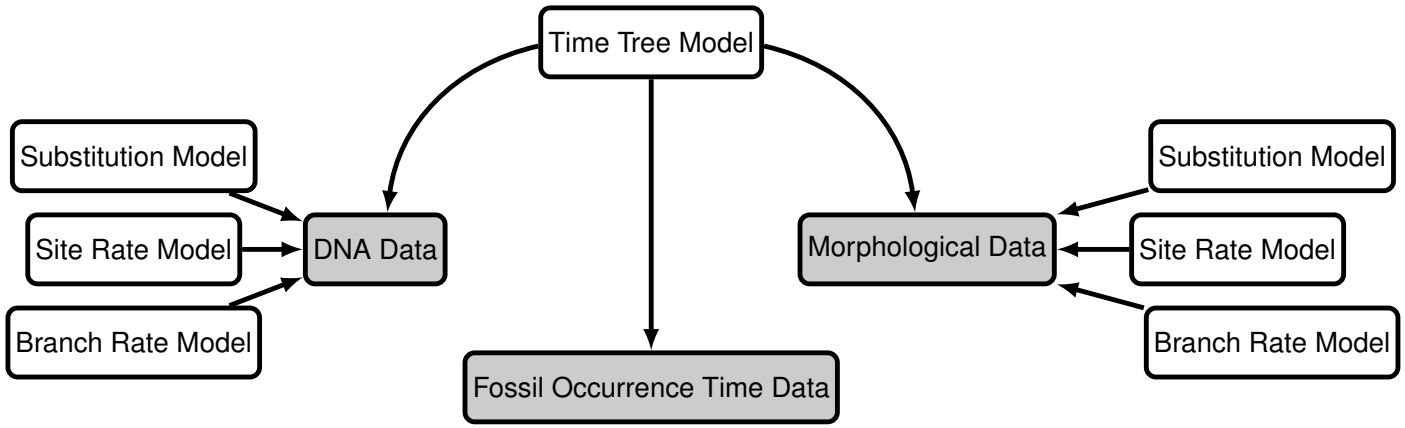


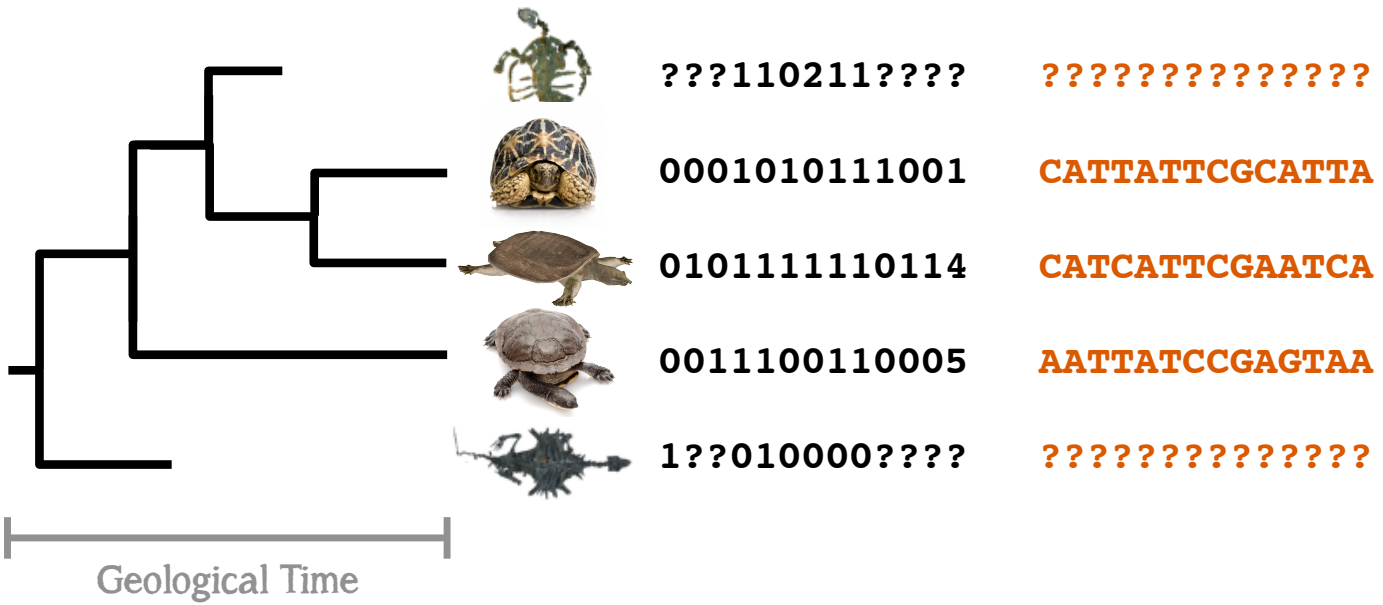
No Sampled Ancestor Tree



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

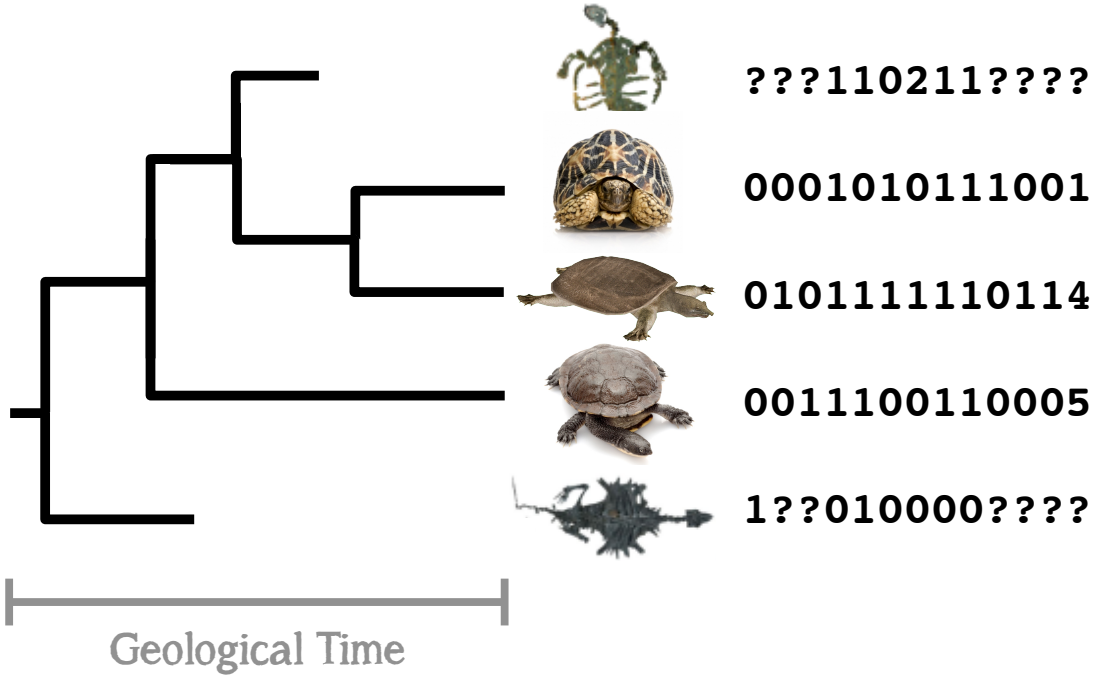
Combining Fossil & Extant Data





(turtle tree image by M. Landis)

Modeling Morphological Character Change



Syst. Biol. 50(6):913–925, 2001

A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data

Paul O. Lewis

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

(turtle tree image by M. Landis)

Modeling Morphological Character Change

The Lewis Mk model

Assumes a character can take k states

Transition rates between states are equal (symmetric)

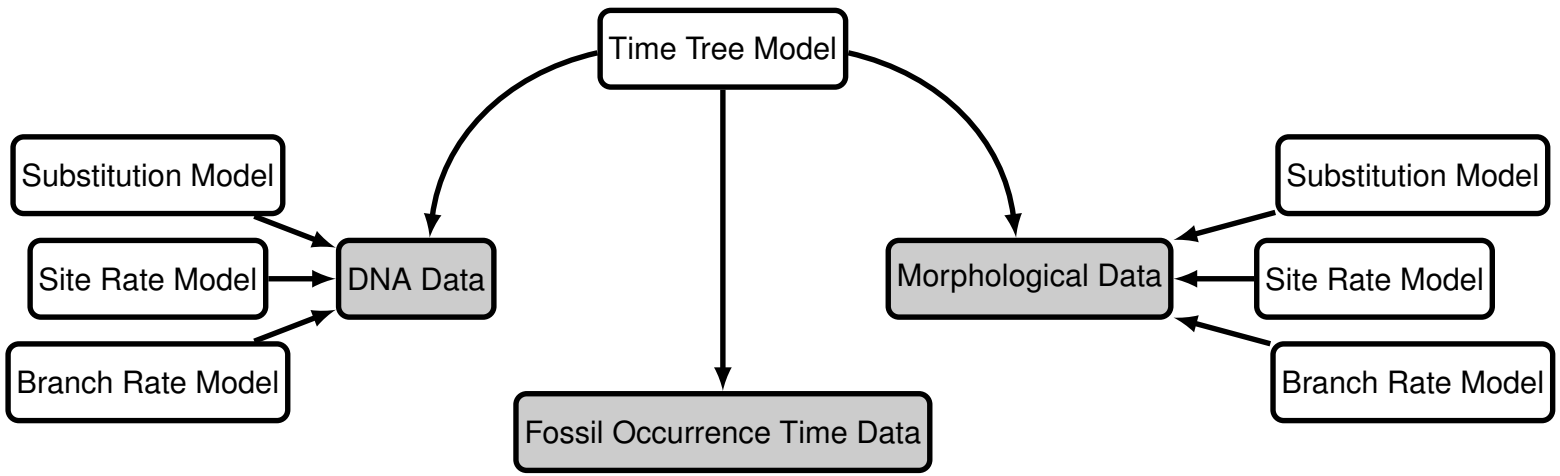
$$Q = \alpha \begin{bmatrix} 1-k & 1 & \dots & 1 \\ \vdots & 1-k & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1-k \end{bmatrix}$$

T1	0
T2	0
T3	1
T4	2
T5	2
T6	1
T7	1

(Lewis. *Systematic Biology* 2001)

"Total-Evidence" Analysis

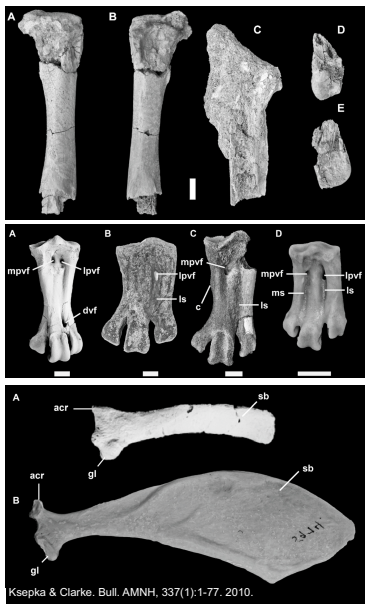
Integrating models of molecular and morphological evolution with improved tree priors enables joint inference of the tree topology (extant & extinct) and divergence times

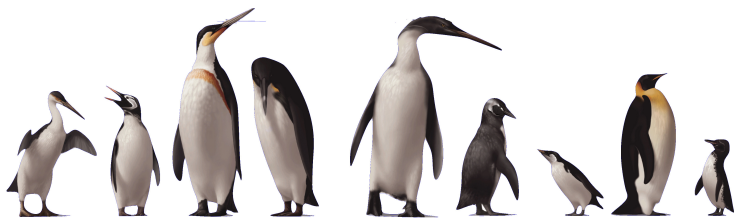


Penguin Diversity in Deep Time

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

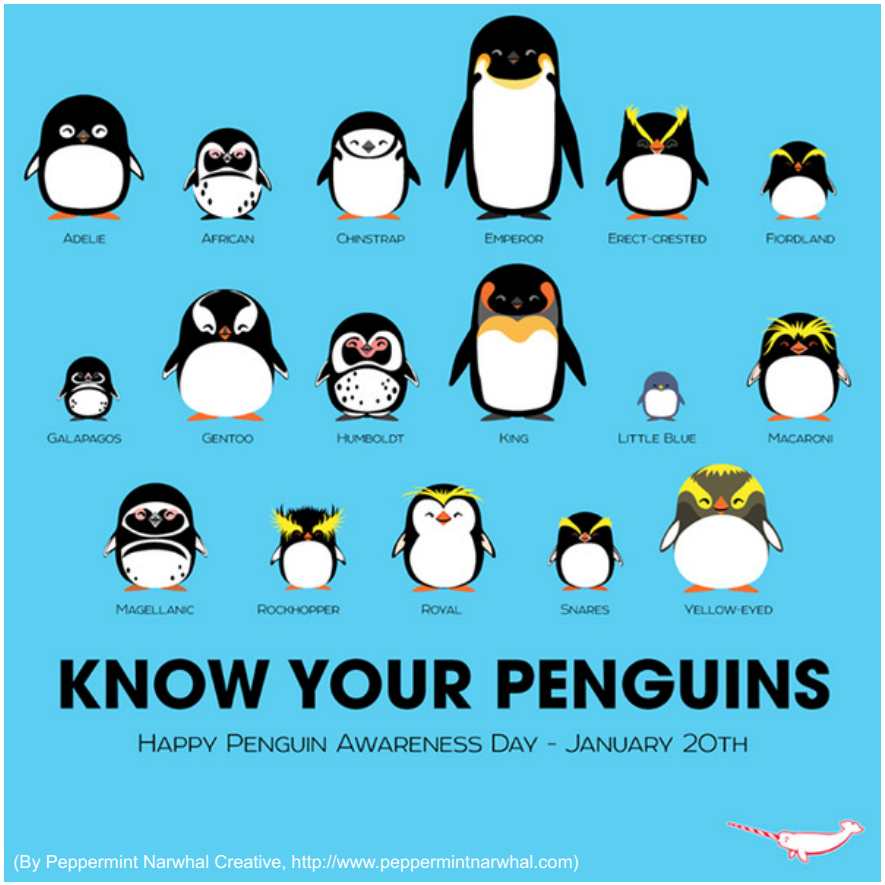
"Penguin Party" by Kate Dzikiewicz

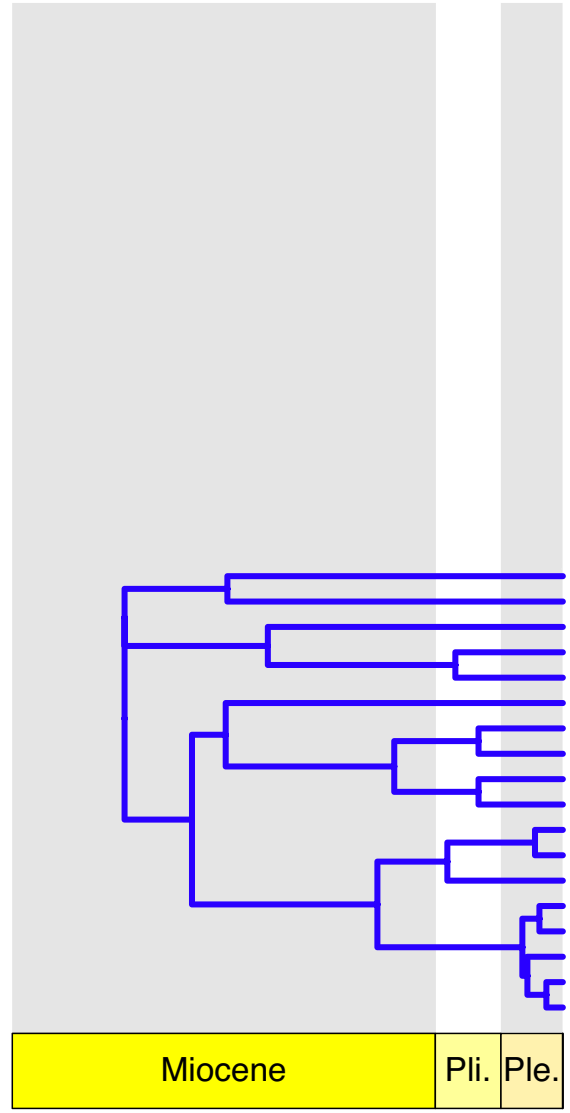




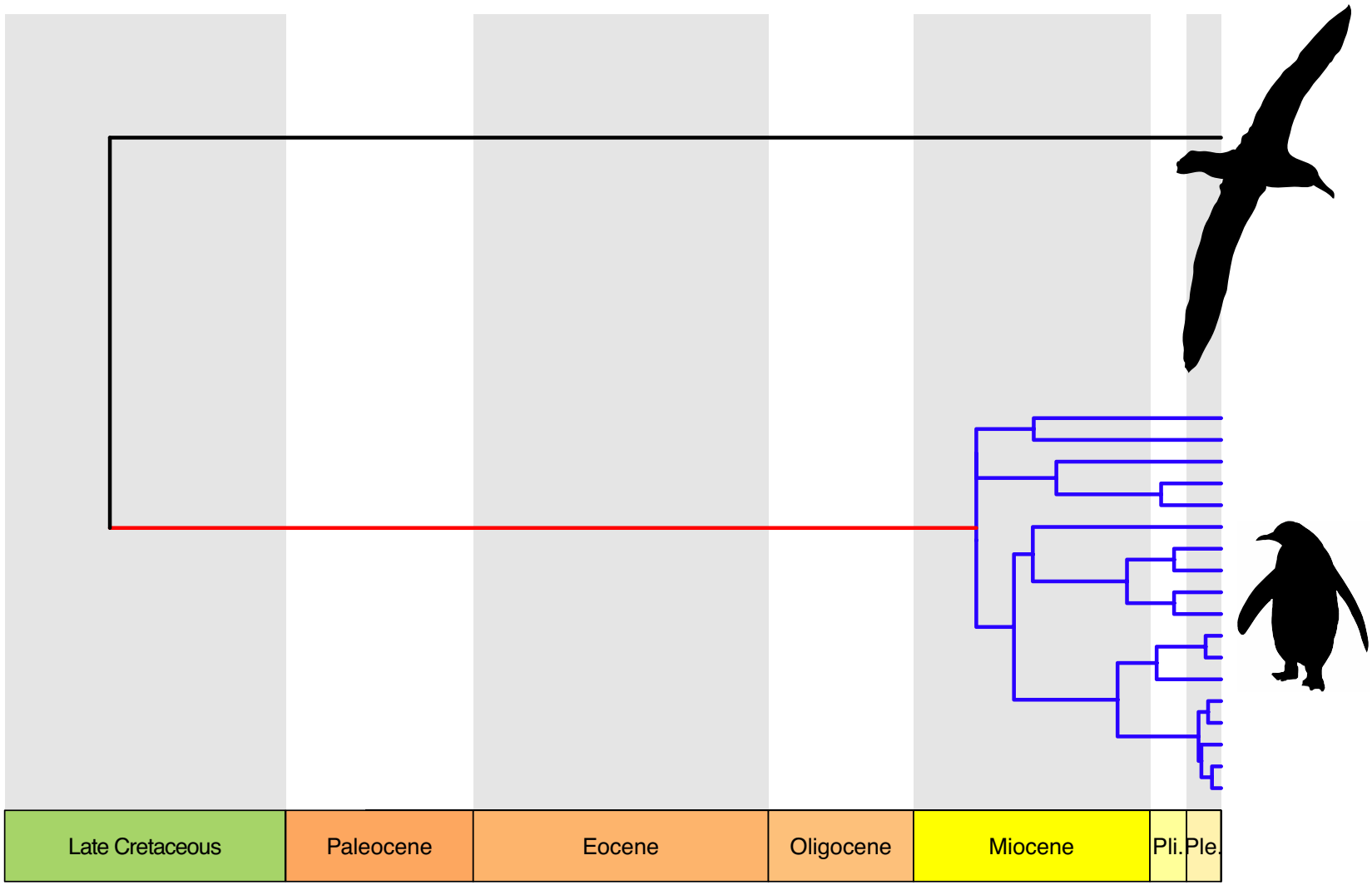
Artistic reconstructions by: Stephanie Abramowicz for Scientific American
Fordyce, R.E. and D.T. Ksepka. The Strangest Bird Scientific American 307, 56 – 61 (2012)

Penguin Diversity





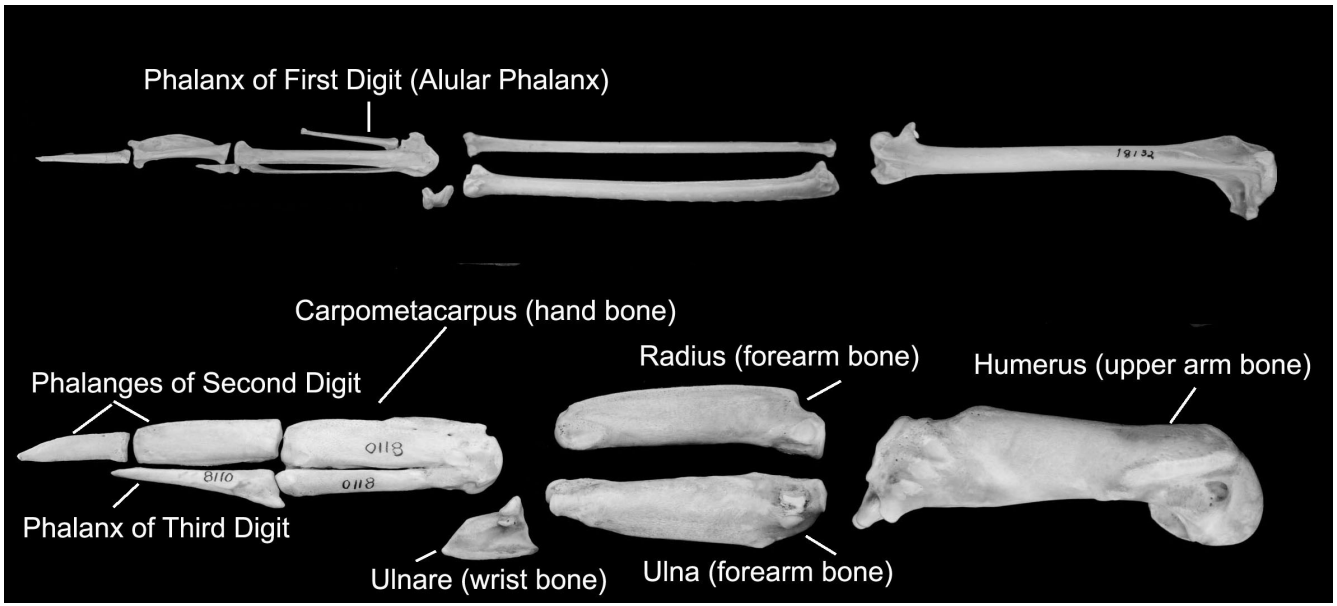
Penguin Diversity



(silhouette images from `http://phylopic.org`)

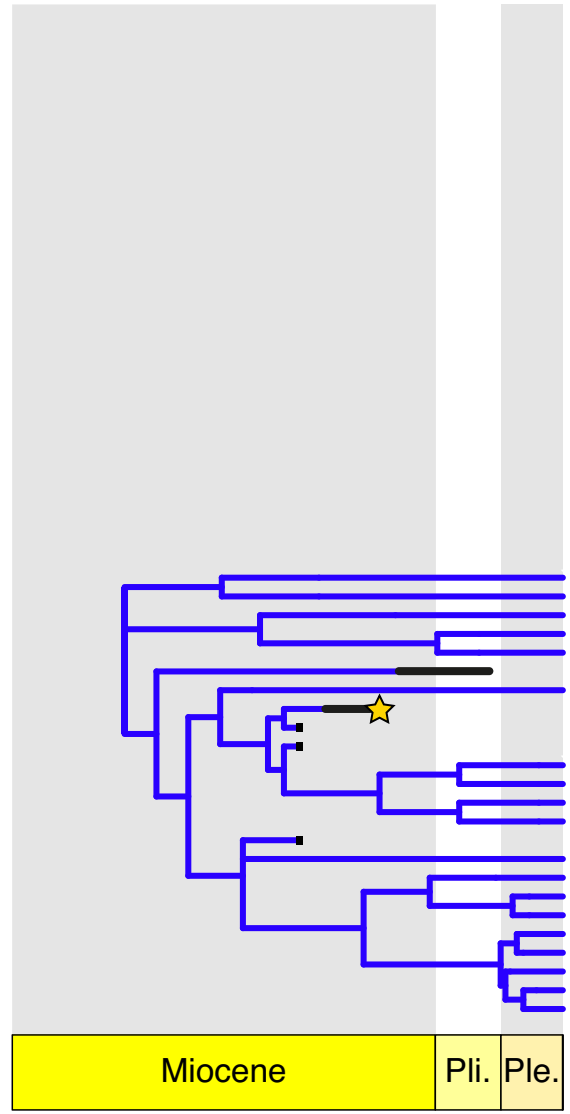
What Makes a Penguin a Penguin?

Flattened, solid wing-bones



(image courtesy of D. Ksepka https://fossilpenguins.wordpress.com)

Fossil Penguin Diversity



(*S. urbinai* holotype fossil, 5-7 MYA, image by Martin Chávez)

Penguins in the Oligocene

Kairuku

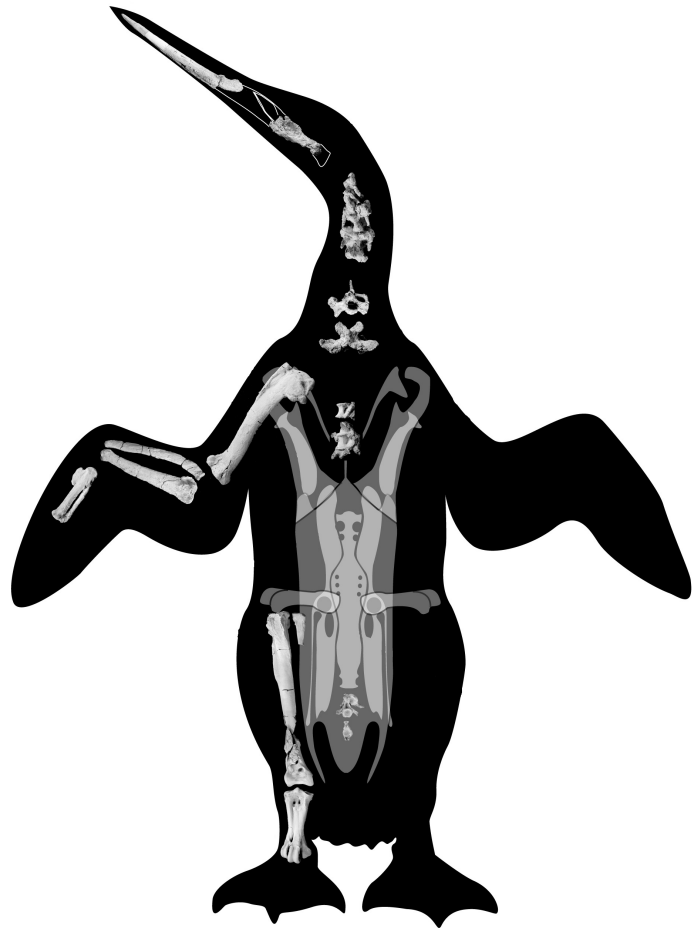
- ~1.5 m tall
- slender, with narrow bill
- scapula & pygostyle are more similar to non-penguins
- ~27 Mya

(Ksepka, Fordyce, Ando, & Jones, *J. Vert. Paleo.* 2012)

Penguins in the Paleocene

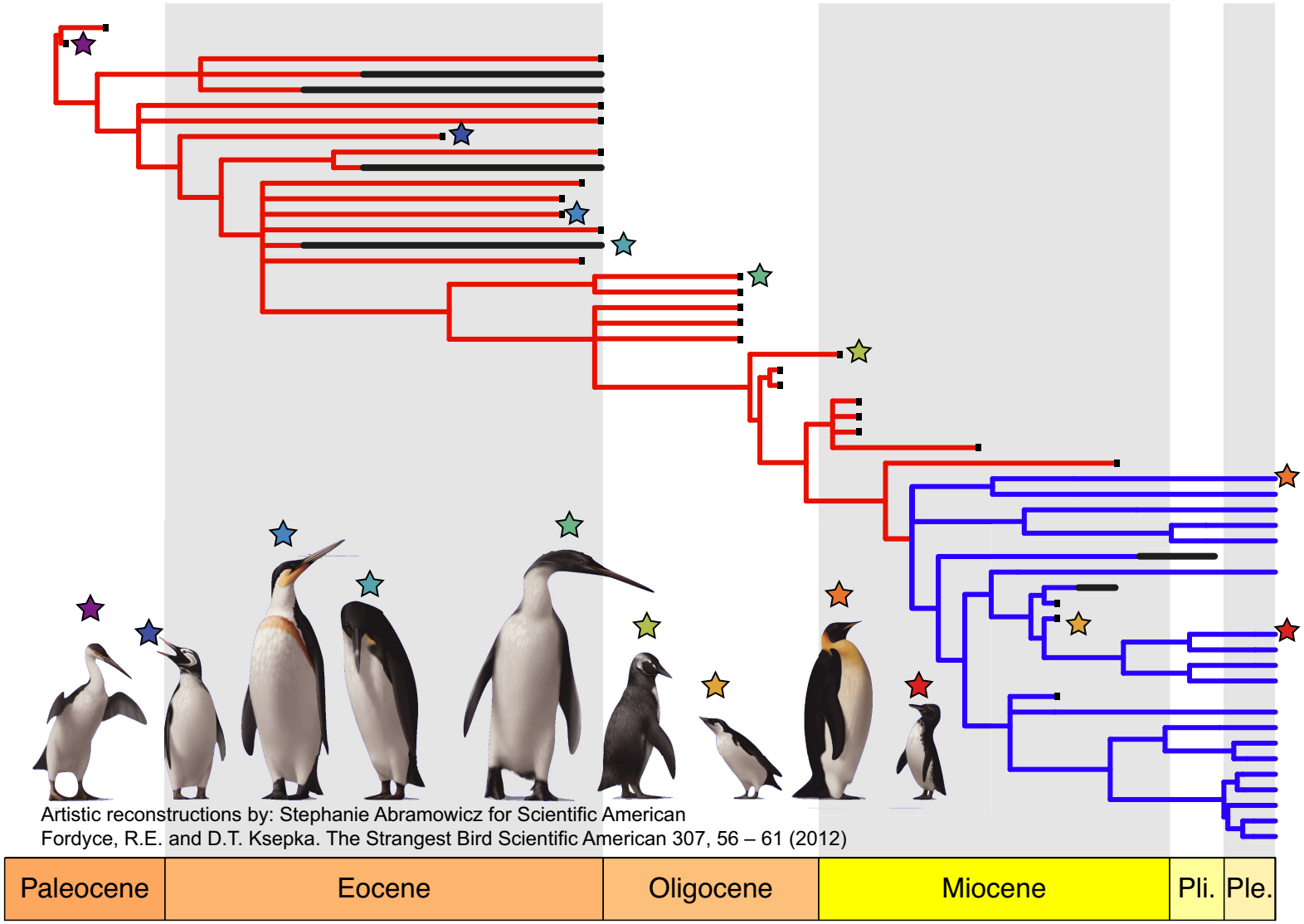
Waimanu

- oldest known penguin species
- intermediate wing morphology
- ~58–61.6 Mya



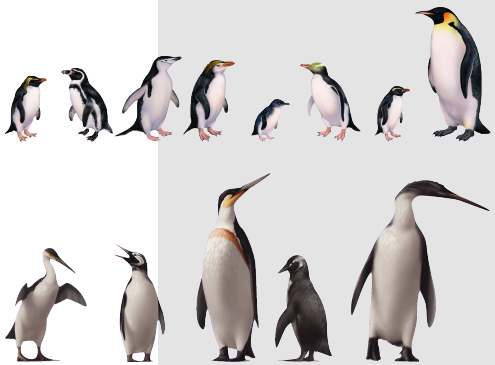
(Slack et al., *Mol. Biol. Evol.* 2006)

Penguin Diversity in Deep Time



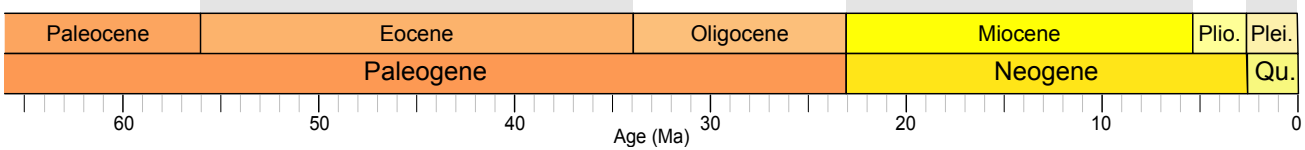
Artistic reconstructions by: Stephanie Abramowicz for Scientific American
Fordyce, R.E. and D.T. Ksepka. The Strangest Bird Scientific American 307, 56 – 61 (2012)

Penguin Diversity in Deep Time



19 extant species
36 fossil species

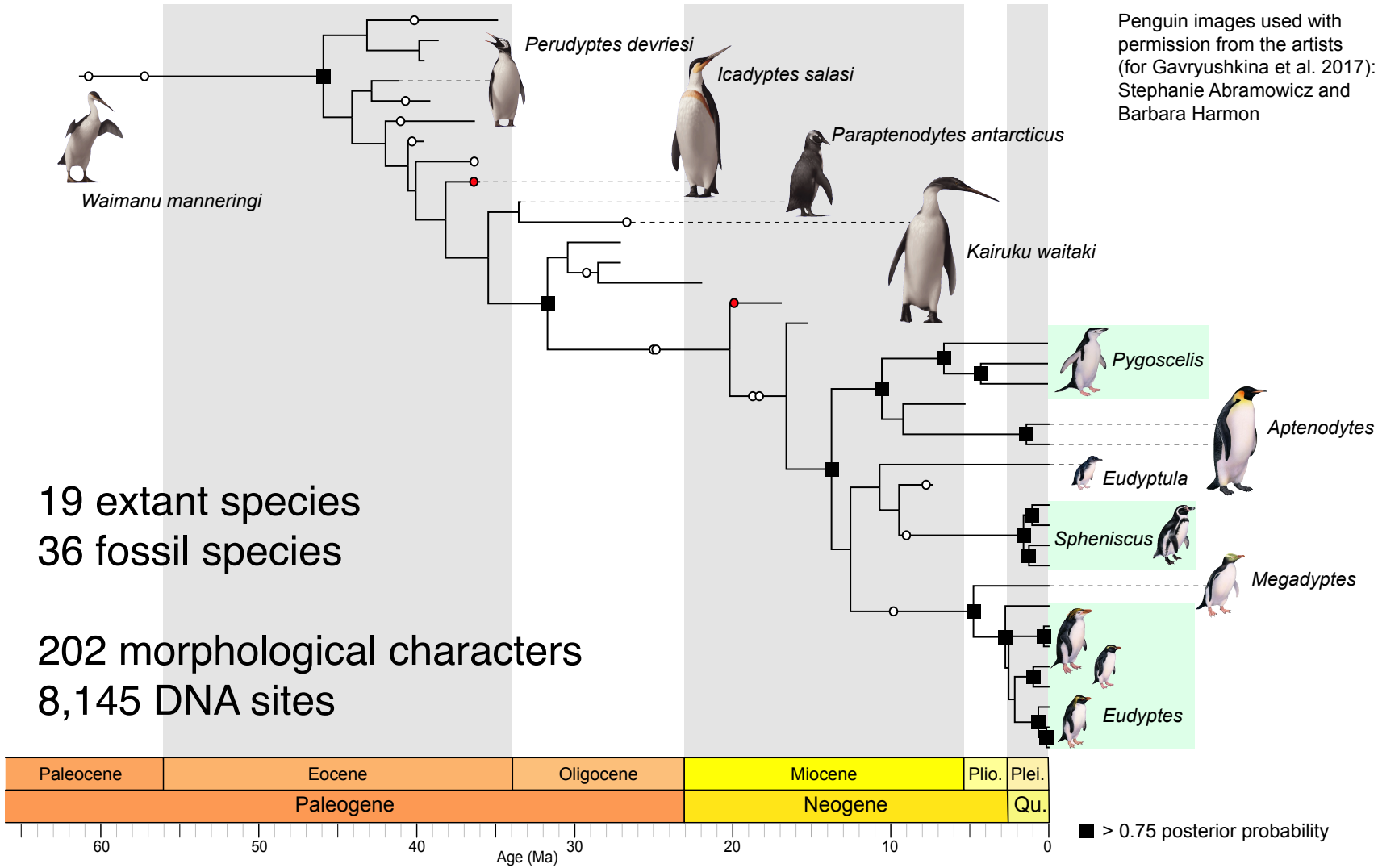
202 morphological characters
8,145 DNA sites



Penguin images used with permission from the artists (for Gavryushkina et al. 2017): Stephanie Abramowicz and Barbara Harmon

(Gavryushkina, Heath, Ksepka, Welch, Stadler, Drummond. 2017. *Syst. Biol.*)

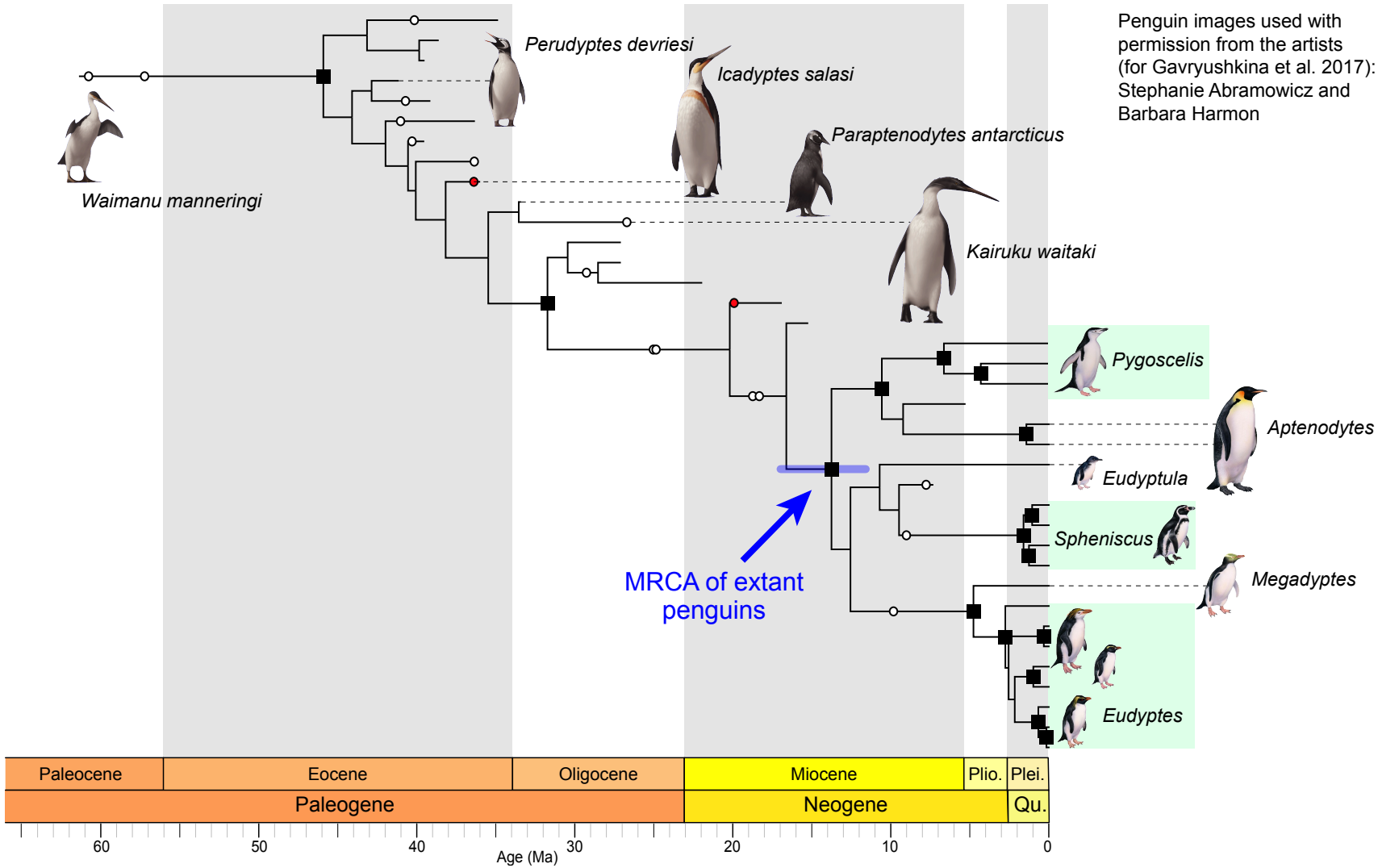
Penguin Diversity in Deep Time



Penguin images used with permission from the artists (for Gavryushkina et al. 2017): Stephanie Abramowicz and Barbara Harmon

(Gavryushkina, Heath, Ksepka, Welch, Stadler, Drummond. 2017. *Syst. Biol.*)

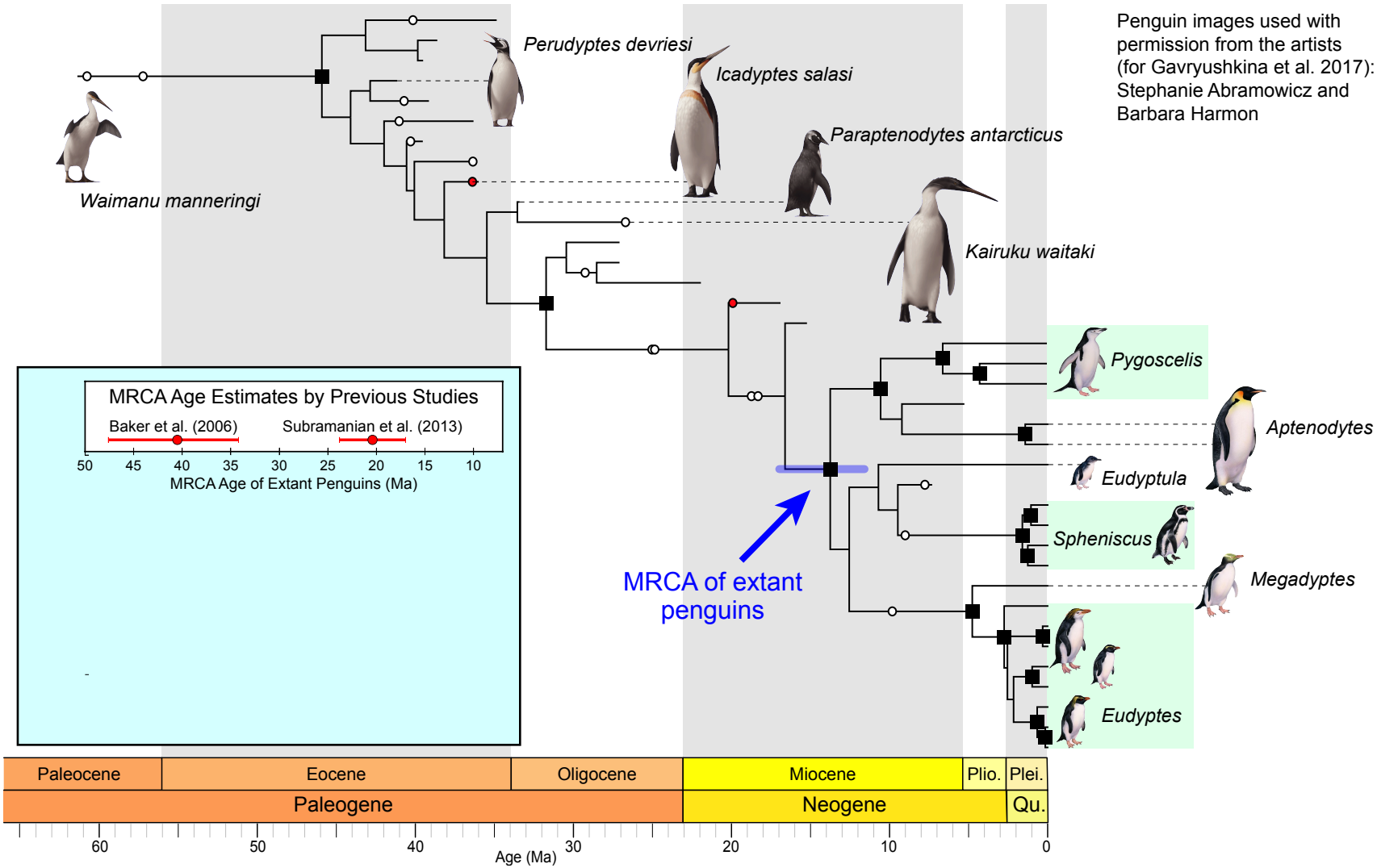
Penguin Diversity in Deep Time



Penguin images used with permission from the artists (for Gavryushkina et al. 2017): Stephanie Abramowicz and Barbara Harmon

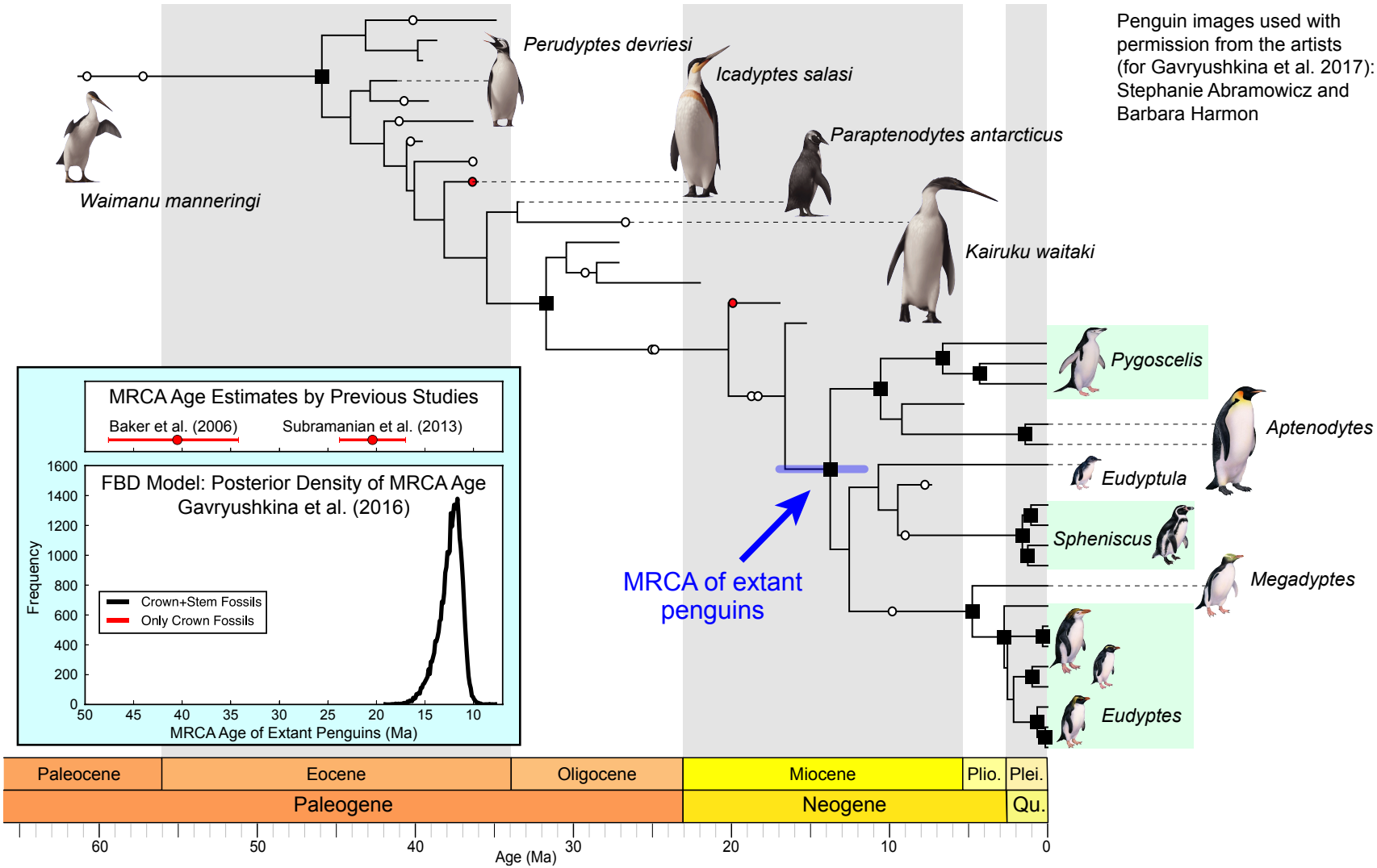
(Gavryushkina, Heath, Ksepka, Welch, Stadler, Drummond. 2017. *Syst. Biol.*)

Penguin Diversity in Deep Time



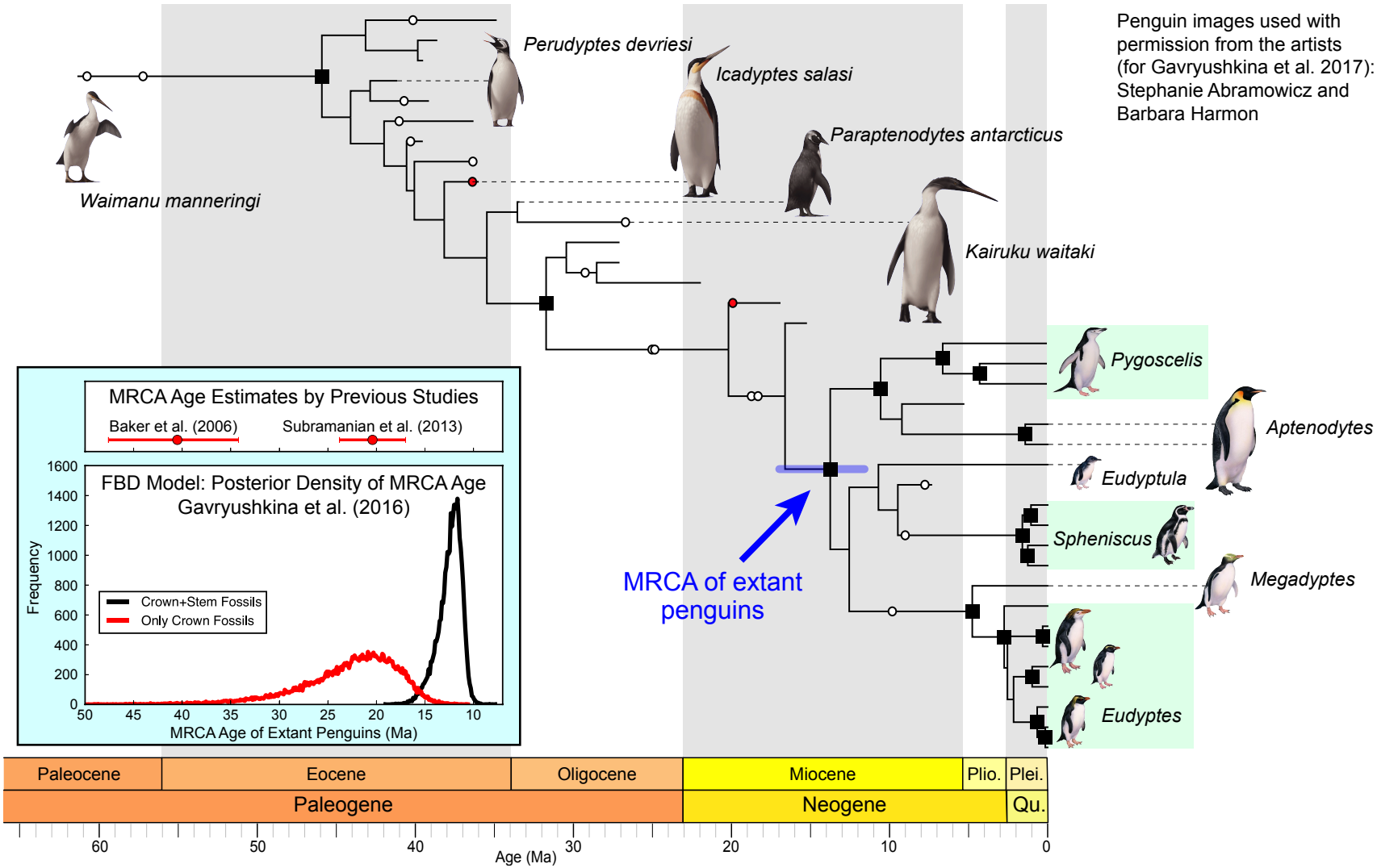
(Gavryushkina, Heath, Ksepka, Welch, Stadler, Drummond. 2017. *Syst. Biol.*)

Penguin Diversity in Deep Time



(Gavryushkina, Heath, Ksepka, Welch, Stadler, Drummond. 2017. *Syst. Biol.*)

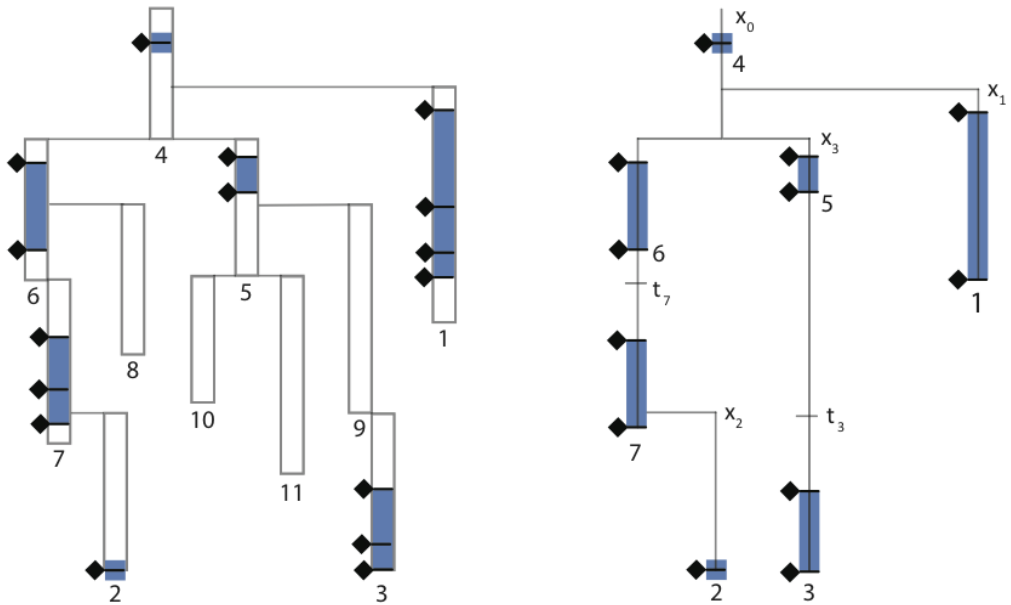
Penguin Diversity in Deep Time



(Gavryushkina, Heath, Ksepka, Welch, Stadler, Drummond. 2017. *Syst. Biol.*)

Fossil Data & Phylogenies

Through collaboration with paleontologists, we are building models to account for the structure of the fossil record and the nature of paleontological data



The fossilized birth-death model for the analysis of stratigraphic range data under different speciation modes

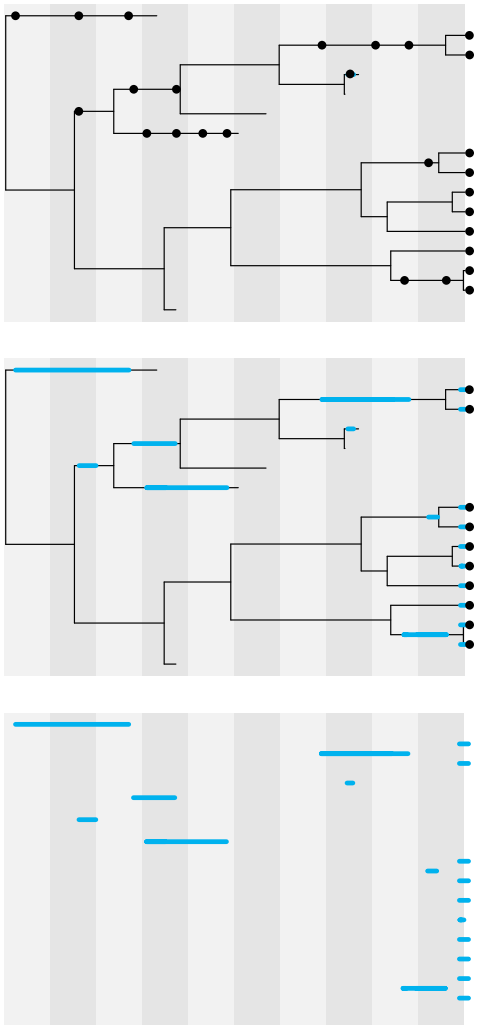
Tanja Stadler[a,b,*], Alexandra Gavryushkina[a,b], Rachel C.M. Warnock[a,b], Alexei J. Drummond[c], Tracy A. Heath[d]

Journal of Theoretical Biology 447:41-55 (2018)

Fossil Data & Phylogenies

The FBD model can accommodate different kinds of paleontological data

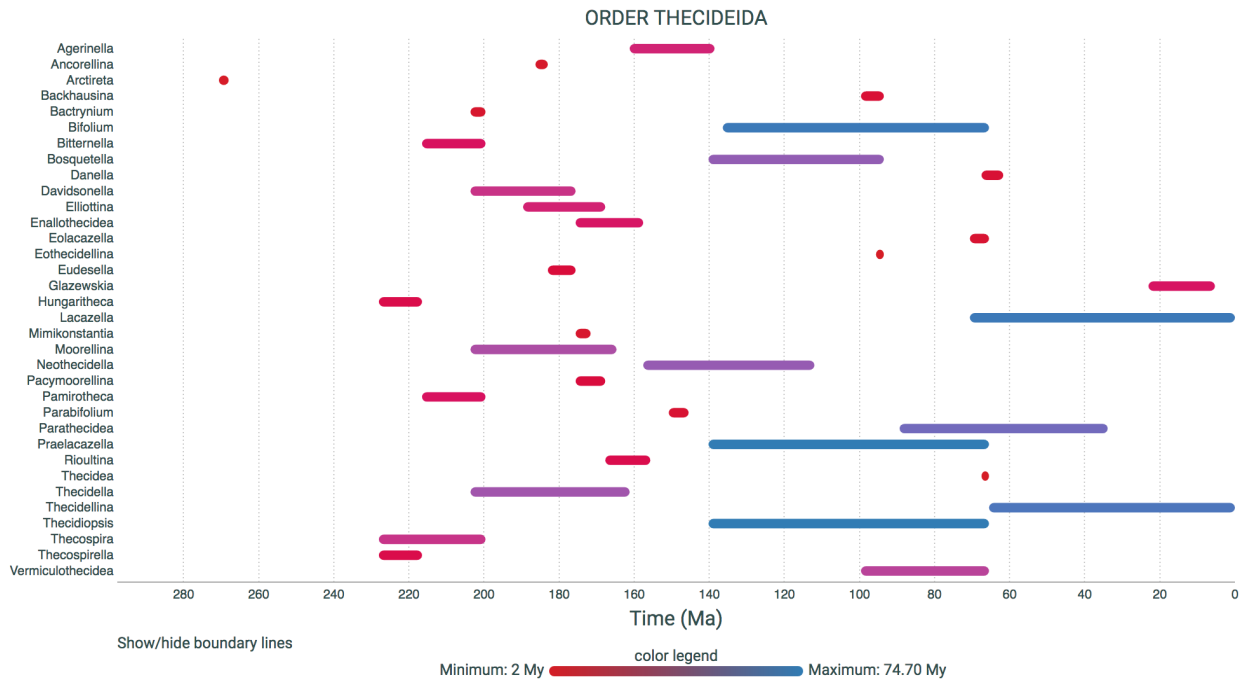
- specimen-level sampling
- when the fossil data are only coded for first and last occurrences (stratigraphic ranges)
- when only stratigraphic range data are available



(figure courtesy of R. Warnock)

FBD for Stratigraphic Range Data

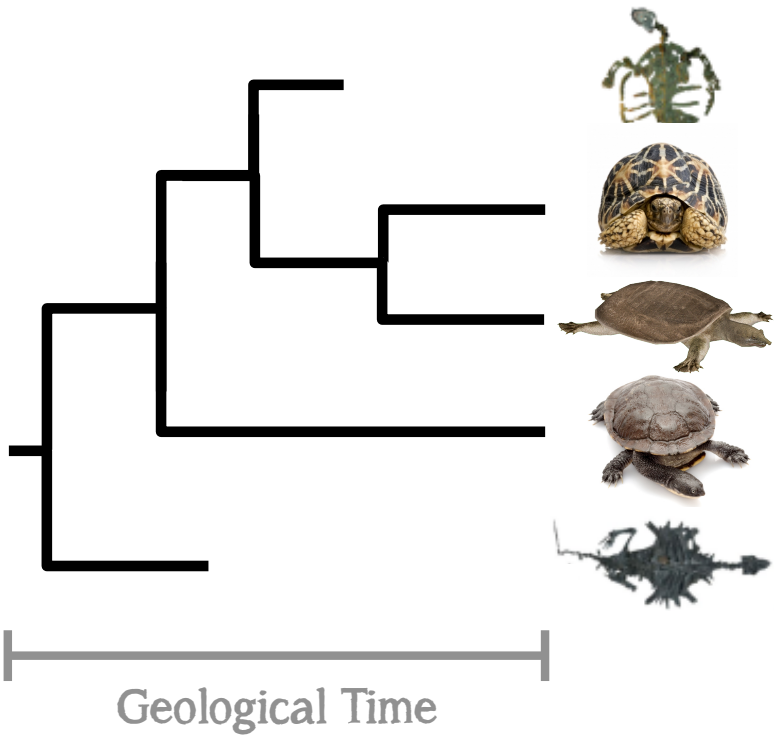
Estimate rates of speciation, extinction, & fossil recovery when no phylogenetic data are available



(figure from `https://gerardofurtado.com/sr/sr.html`, using data from Sepkoski 2002)

The FBD model explicitly accounts for incomplete species sampling, as well as uncertainty in speciation and extinction times and the phylogeny

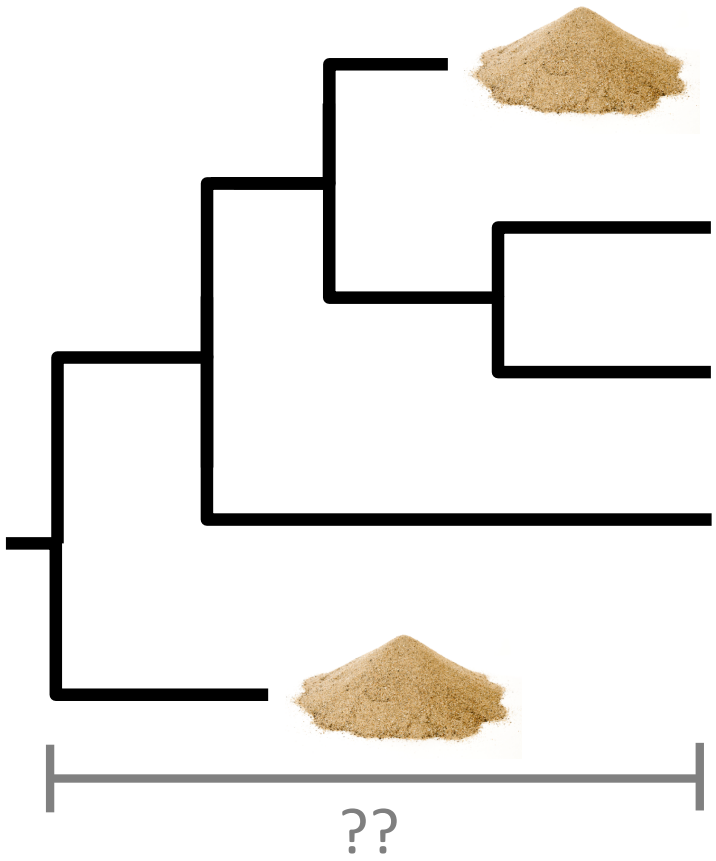
Molecules + Morphology + Fossils



???110211????	????????????????
0001010111001	CATTATTCGCATTA
0101111110114	CATCATTCGAATCA
0011100110005	AATTATCCGAGTAA
1??010000????	????????????????

(based on slides by M. Landis)

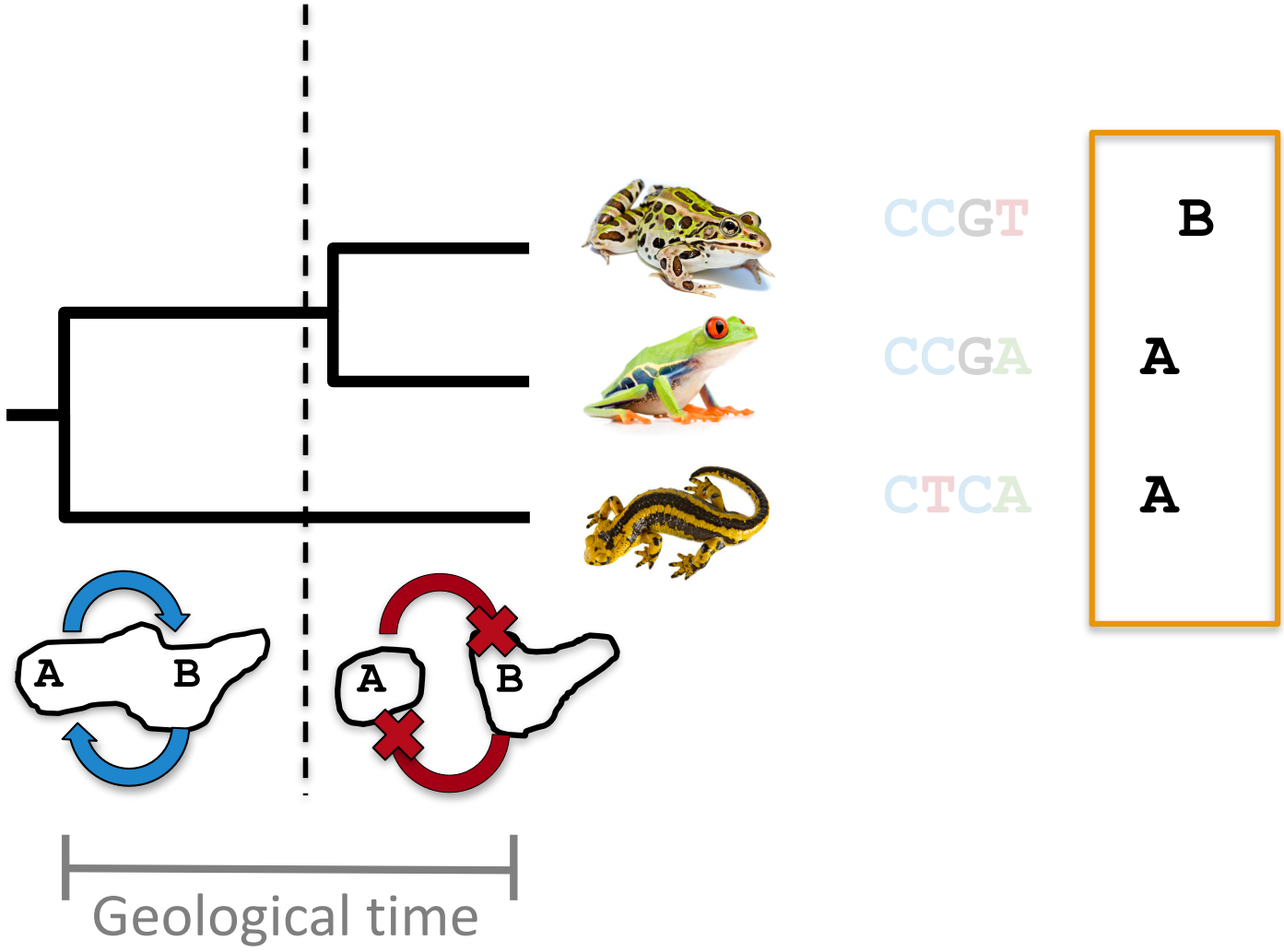
...but I study amphibians...



????	????
CCGT	1110
CCGA	1101
CTCA	0001
????	????

(slides courtesy of M. Landis; http://bit.ly/2alHqB4)

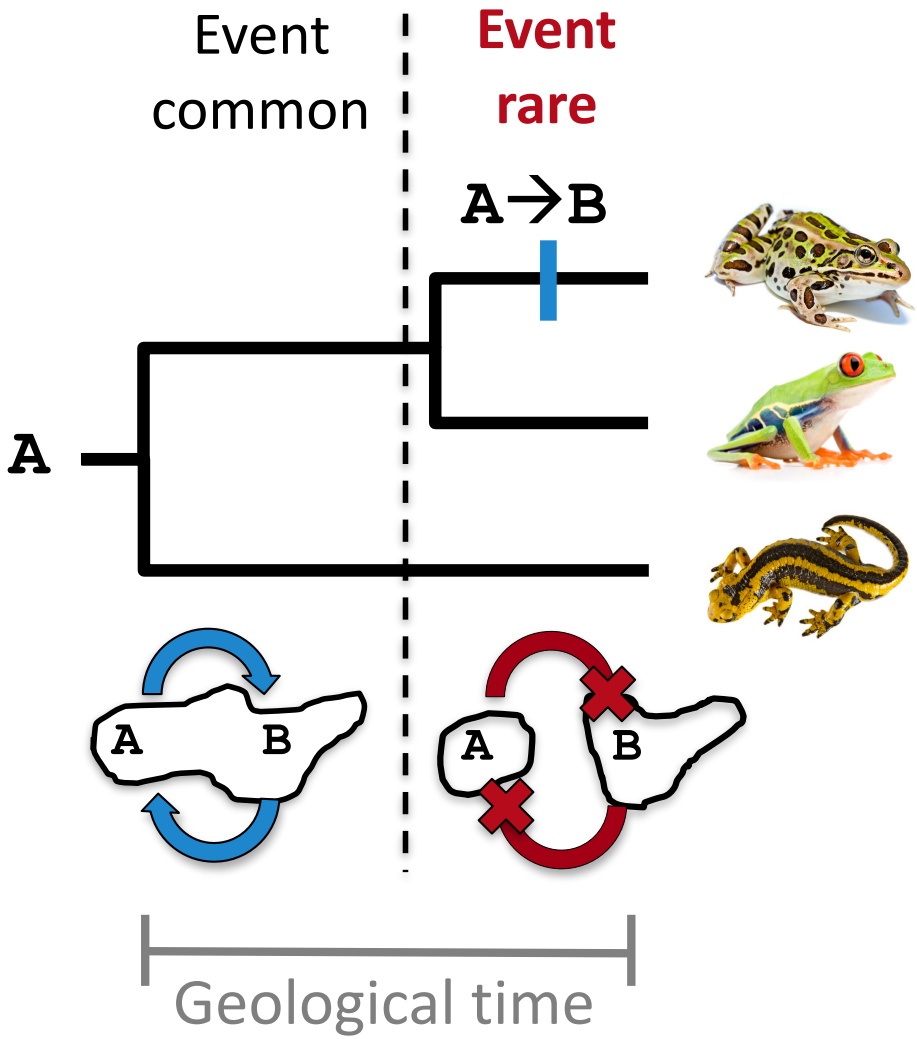
Molecules + biogeography + paleogeography



+ Paleogeography Landis, 2016

(slides courtesy of M. Landis; http://bit.ly/2alHqB4)

Events should occur *before areas split*



CCGT B

CCGA A

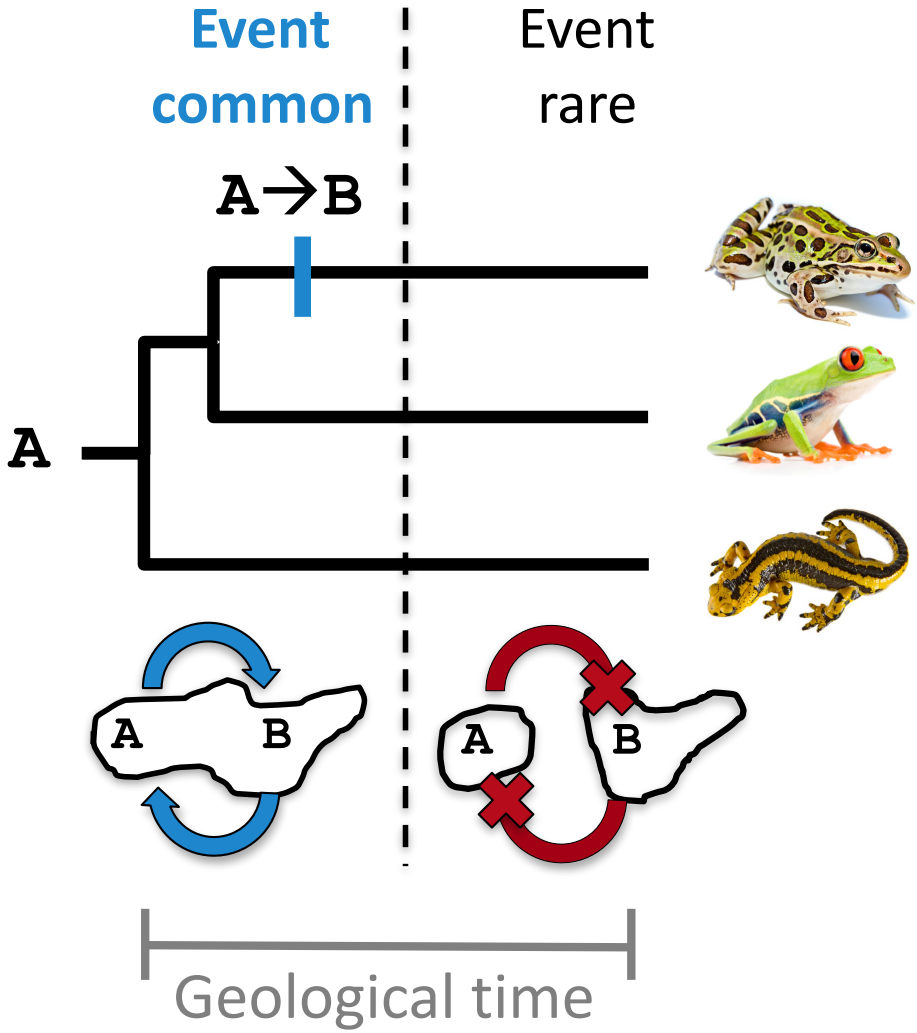
CTCA A

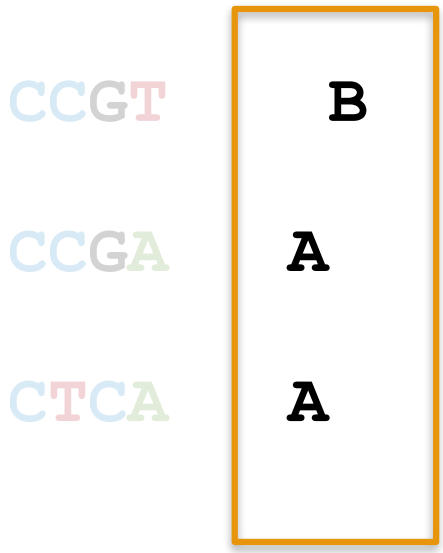
Less probable observation given tree

+ Paleogeography Landis, 2016

(slides courtesy of M. Landis; http://bit.ly/2alHqB4)

Events should occur *before areas split*



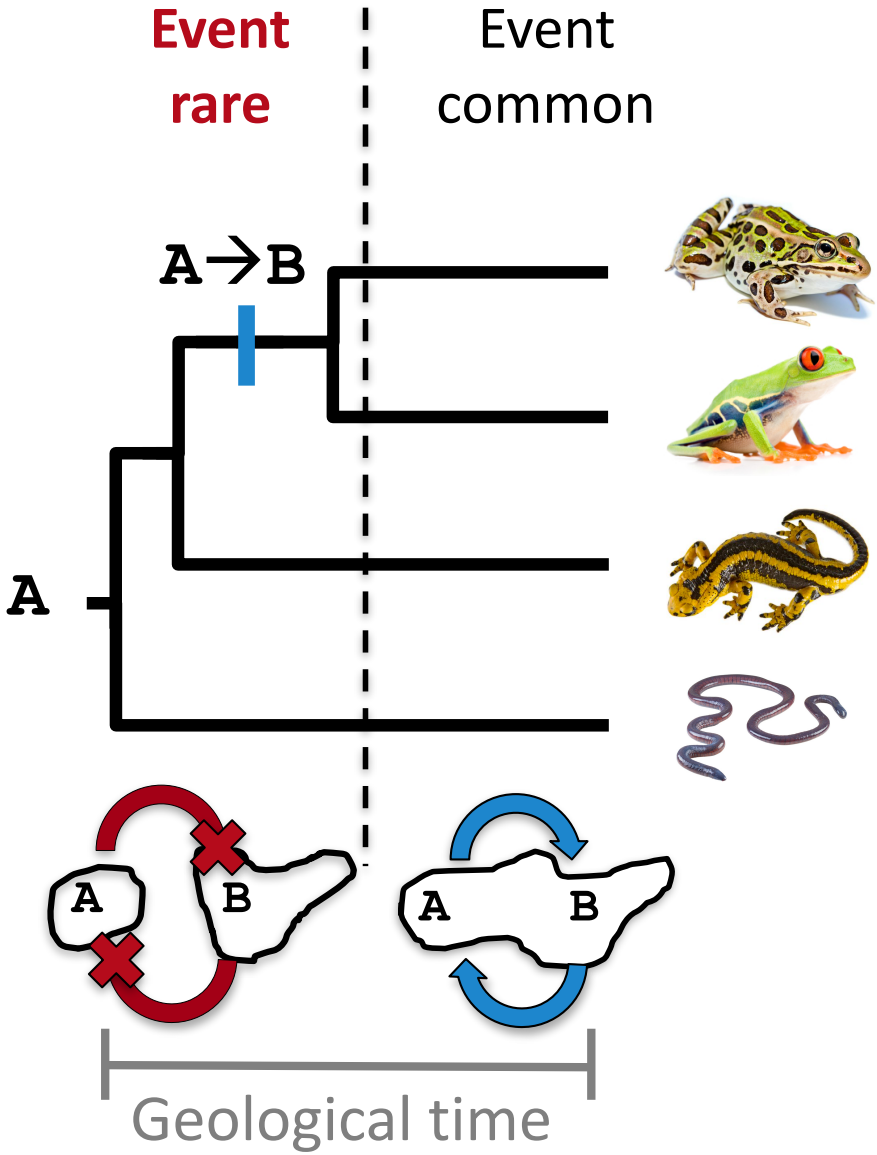


More probable observation given tree

+ Paleogeography Landis, 2016

(slides courtesy of M. Landis; http://bit.ly/2alHqB4)

Events should occur *after areas merge*



CCGT B

CCGA B

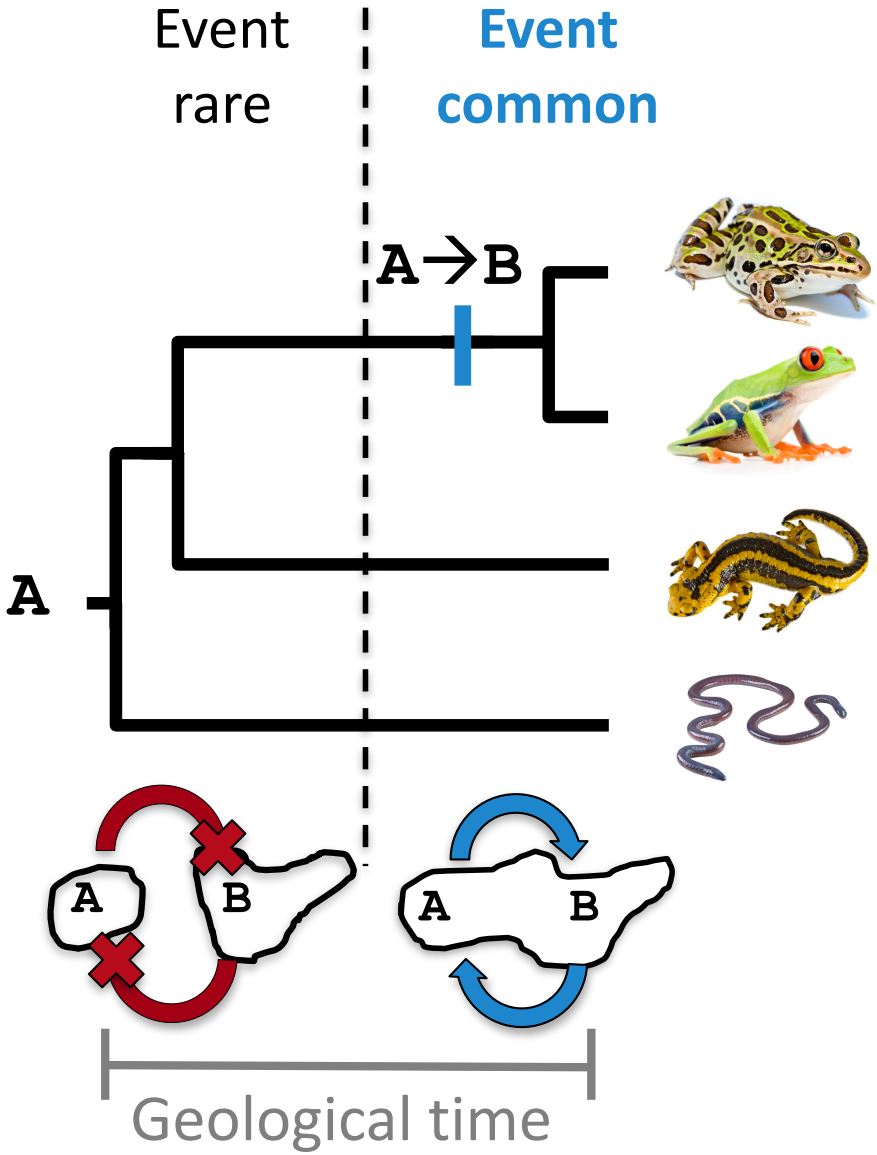
CTCA A

TTCC A

Less probable observation given tree

(slides courtesy of M. Landis; http://bit.ly/2alHqB4)

Events should occur *after areas merge*



CCGT	B
CCGA	B
CTCA	A
TTCC	A

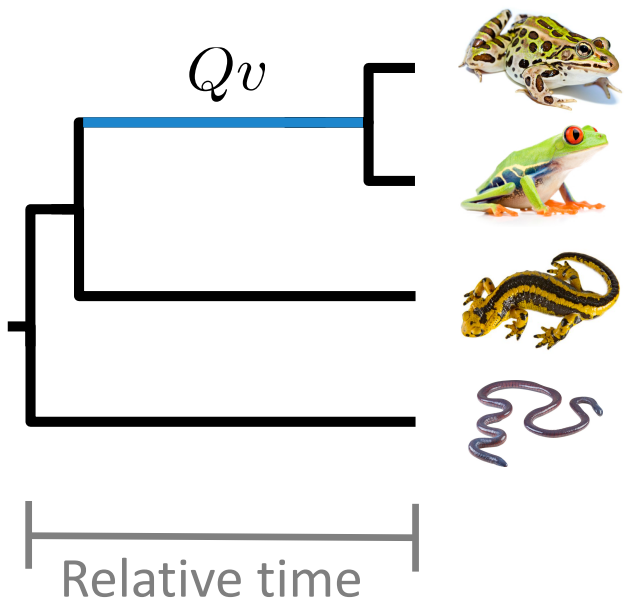
More probable observation given tree

(slides courtesy of M. Landis; http://bit.ly/2alHqB4)

Biogeographic Dating

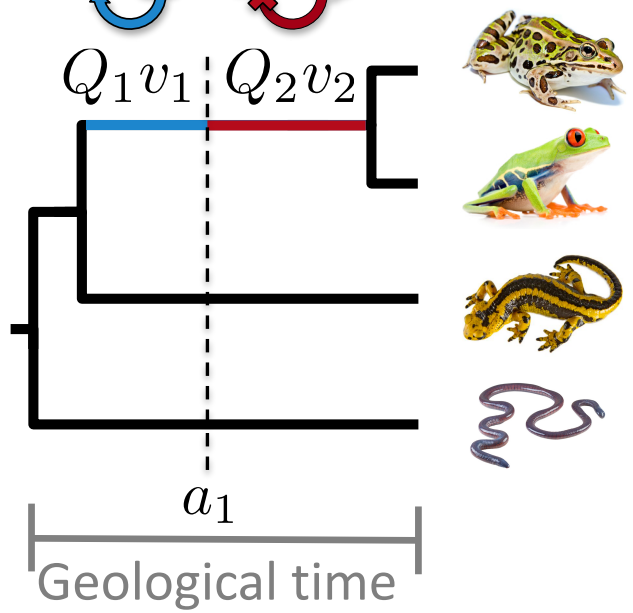
Constant Model

$$\mathbf{P}_{ij}(v) = \exp\{Qv\}$$



Epoch Model

$$\mathbf{P}_{ij}(\mathbf{v}) = \prod \exp\{Q_k v_k\}$$



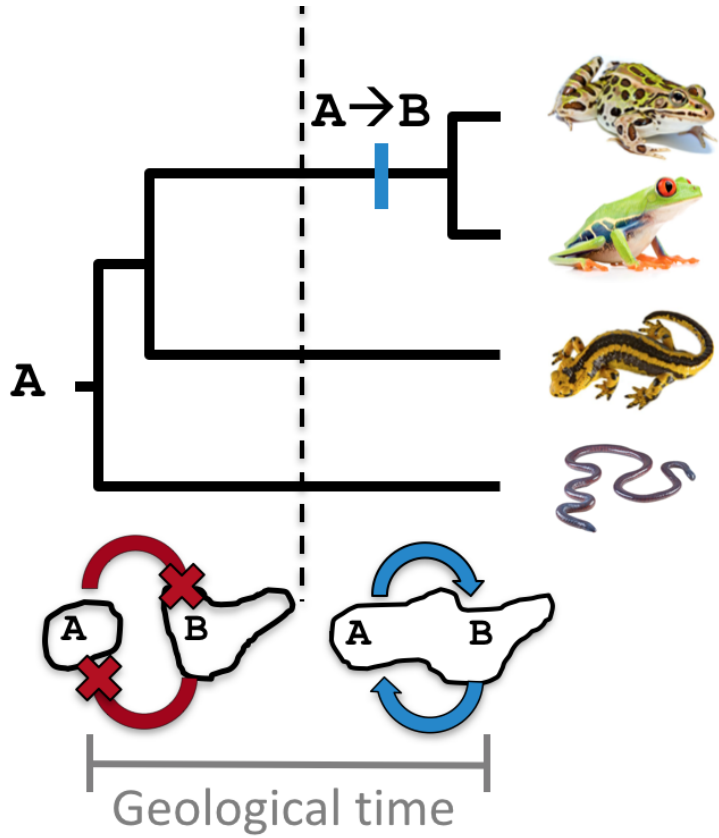
Epoch model Ree et al., 2005
Bielejec et al., 2014

(image by M. Landis)

Biogeographic Dating

Fossil-free calibration

- data: molecular sequences
- data: biogeographic ranges
- empirical paleogeographic model that alters the rates of biogeographic change over time

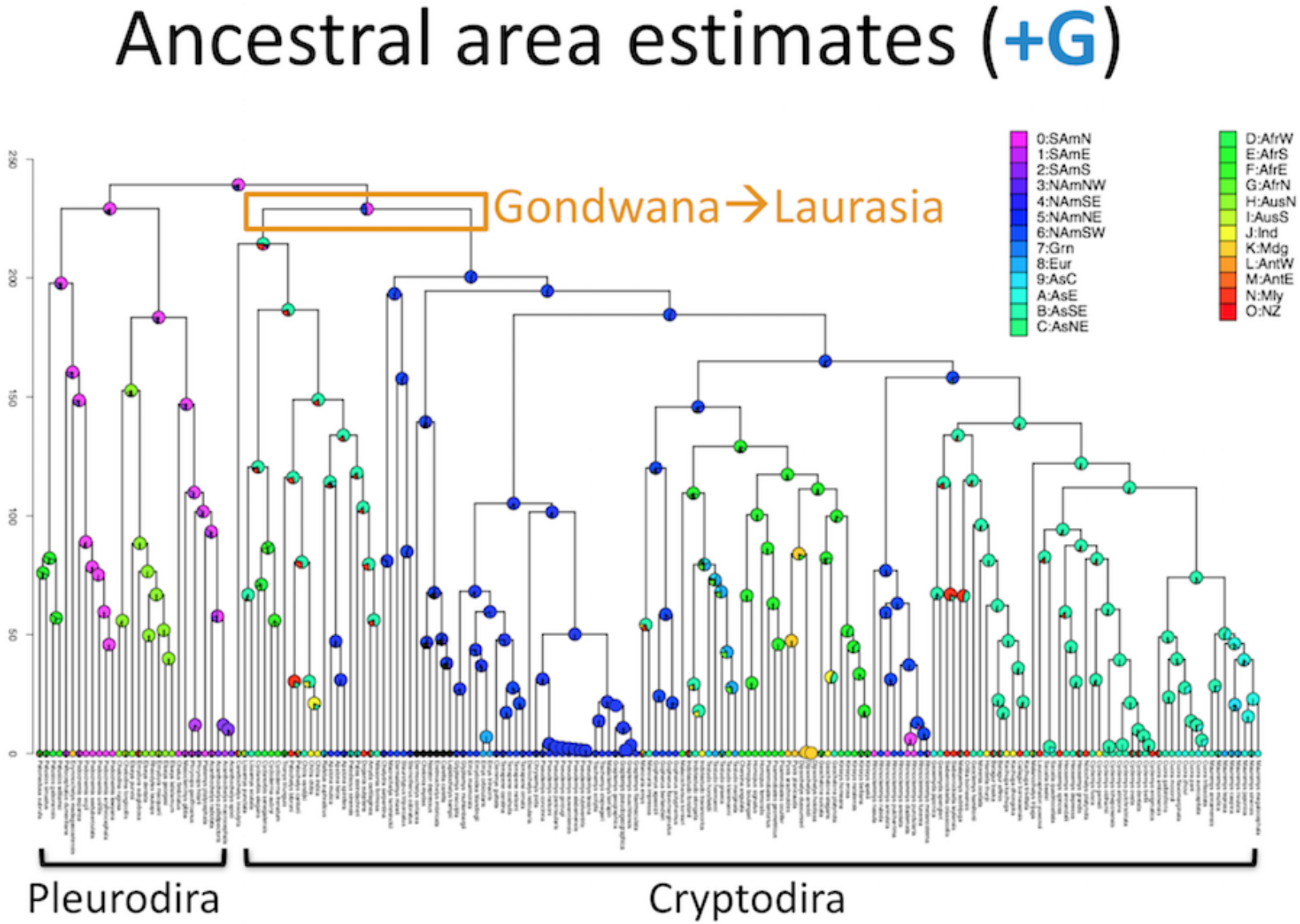


Landis. 2017. "Biogeographic Dating of Speciation Times Using Paleogeographically Informed Processes". *Systematic Biology*

doi: 10.1093/sysbio/syw040.

(image by M. Landis)

Dating + Ancestral Area Reconstruction



(image by M. Landis)

Bayesian Analysis under Complex Models

The models I have described so far are primarily available in two software packages

RevBayes

revbayes.com

flexible model construction using graphical models and an interpreted language for modeling, simulation, and Bayesian inference in evolutionary biology, particularly phylogenetics

tutorials:
revbayes.com/tutorials

BEAST 2

beast2.org

a modular program for using Bayesian inference to infer rooted trees for addressing questions in population genetics, phylogenetics, epidemiology, and macroevolution

tutorials:
taming-the-beast.org

Large Datasets & Bayesian Inference

Alignments with thousands of loci and/or thousands of taxa are not practical with fully Bayesian methods, particularly under complex, hierarchical models.

Large datasets lead to long mixing times because the number of parameters requires many proposals in order to effectively sample the joint posterior density

What to do?

- Simplify the problem (in ways that do not also sacrifice accuracy)

Pecos

Approximate Computation of the Likelihood

Computing the full likelihood of an alignment on a tree is computationally intensive

Bayesian inference under complex models from large datasets requires the likelihood to be calculated millions of times

The practical solution is to compute the likelihood faster, which is possible if we can get a reasonable approximation

Approximate Likelihood Calculation on a Phylogeny for Bayesian Estimation of Divergence Times FREE

Mario dos Reis, Ziheng Yang ✉

Molecular Biology and Evolution, Volume 28, Issue 7, 1 July 2011, Pages 2161–2172,

Approximate Computation of the Likelihood

A 2-step analysis:

1. branch lengths are first estimated on the unrooted tree along with elements needed for approximation method
2. estimate divergence times using MCMC, with inputs from 1 for calculating the approximate likelihood for each M-H move

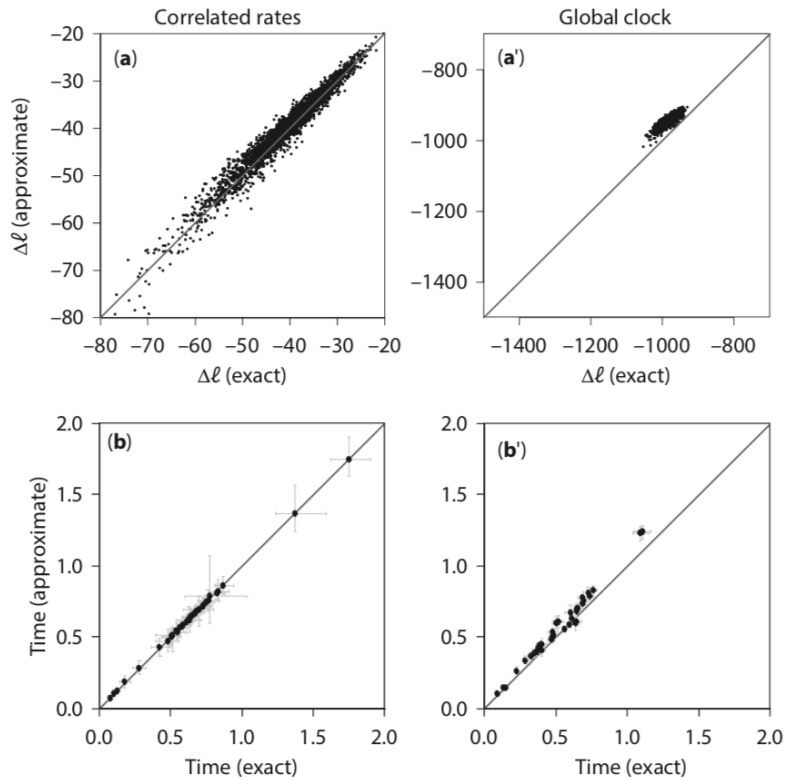


Fig. 10.14 (**a** and **a'**) The log likelihood (or the difference from the highest log likelihood at the MLEs) and (**b** and **b'**) posterior estimates (means and 95% CIs) of divergence times calculated using the exact and approximate likelihood methods under the relaxed clock and strict clock models. The mitochondrial data for 36 mammalian species are analysed. Redrawn from the plots for the arcsine transform in Figures 3 and 5 of dos Reis and Yang (2011).

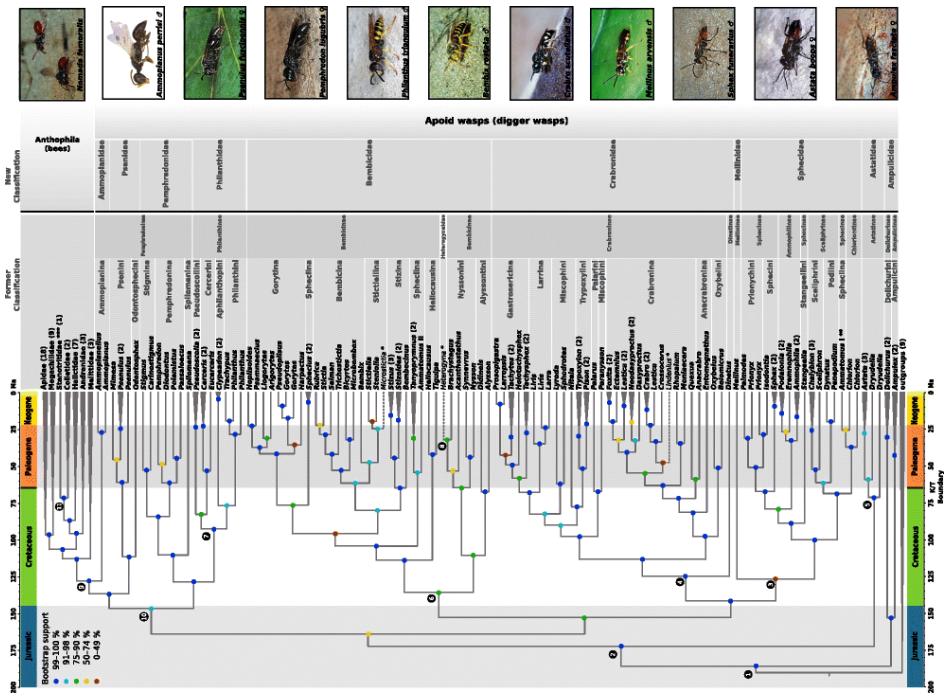
(figure from Yang. 2014. *Molecular Evolution: A Statistical Approach*)

Bayesian Dating with Big(ger) Data

Using approximate likelihood computation in MCMCTree

Apoid Wasps & Bees

- 174 taxa
- 284,607 bp (195 single-copy protein-coding genes)
- 4 fossil calibrations



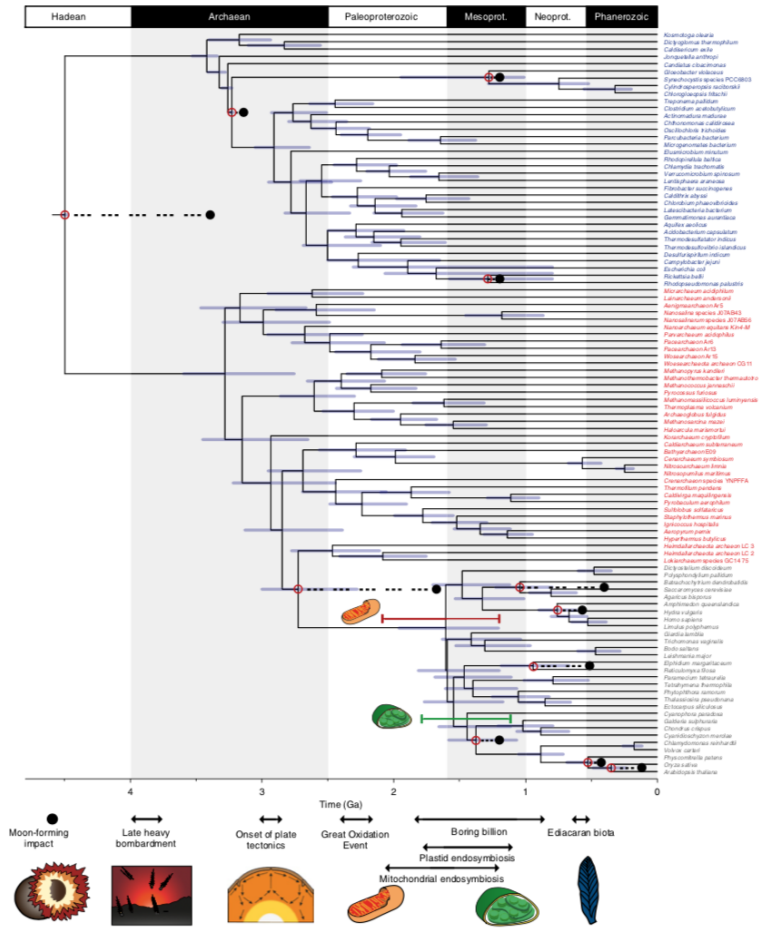
Phylogenomic analysis of Apoidea sheds new light on the sister group of bees (Sann et al. 2018. *BMC Evol. Biol.*)

Bayesian Dating with Big(ger) Data

Using approximate likelihood computation in MCMCTree

Life

- 102 taxa
- 14,745 bp (29 universally distributed, protein-coding genes)
- 11 fossil calibrations



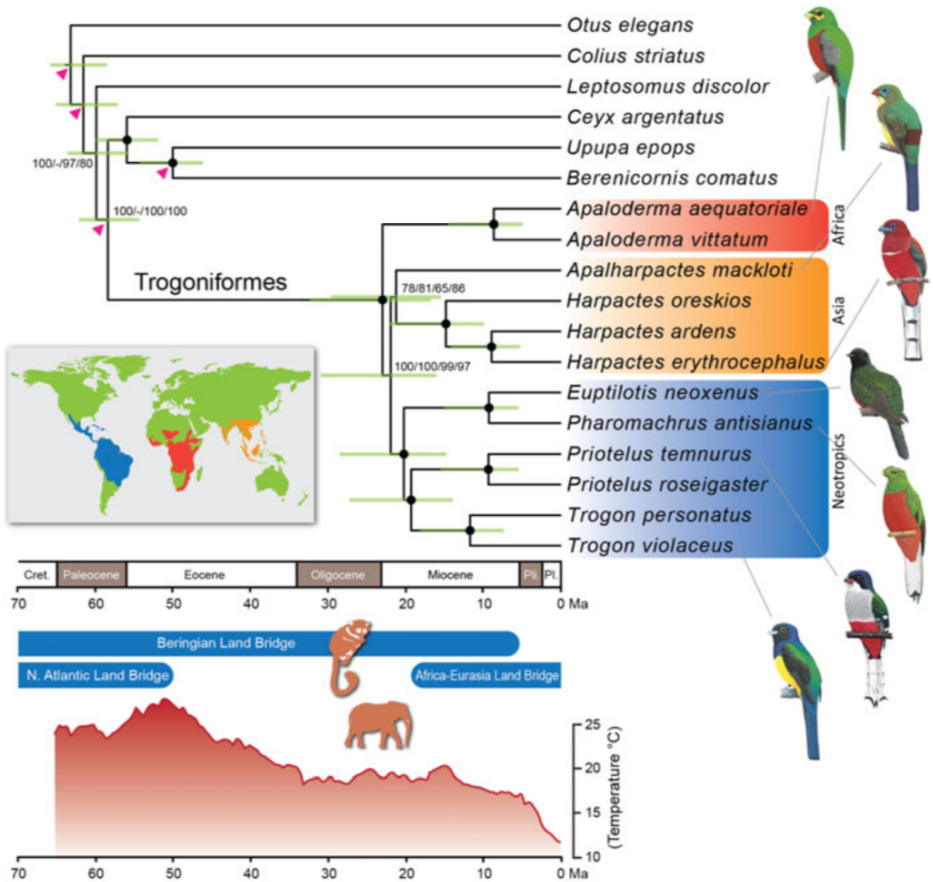
Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin (Betts et al. 2018. *NE&E*)

Bayesian Dating with Big(ger) Data

Using approximate likelihood computation in MCMCTree

Trogons

- 12 ingroup + 6 outgroup taxa
- 2.9 Million bp (3581 UCE loci)
- 4 fossil calibrations



Rapid Laurasian diversification of a pantropical bird family during the Oligocene-Miocene transition (Oliveros et al. in press. *Ibis*)