

DATING SPECIES DIVERGENCE TIMES WITH FOSSIL DATA

Tracy A. Heath

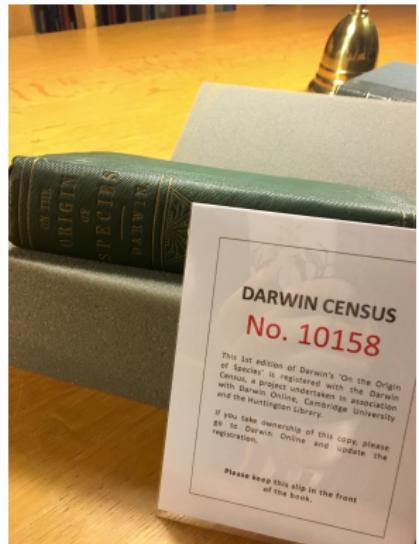
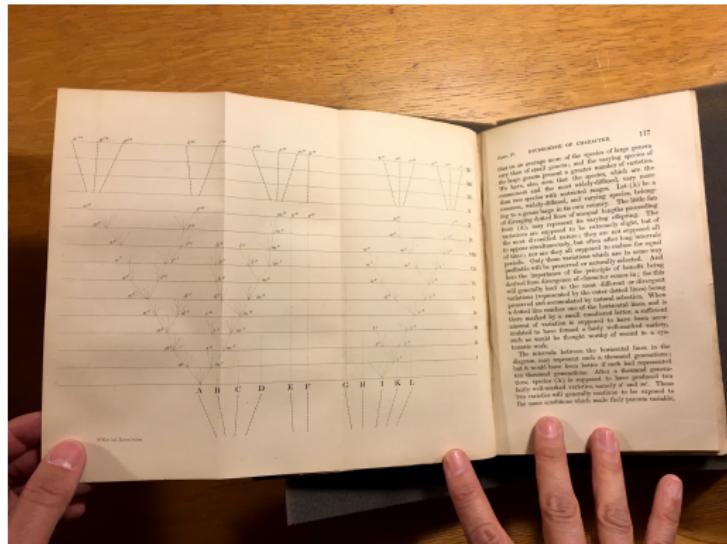
**Ecology, Evolution, & Organismal Biology
Iowa State University**

Computational Molecular Evolution Workshop
Hinxton, Cambridge UK

May 16, 2017

A TIME-SCALE FOR MACROEVOLUTION

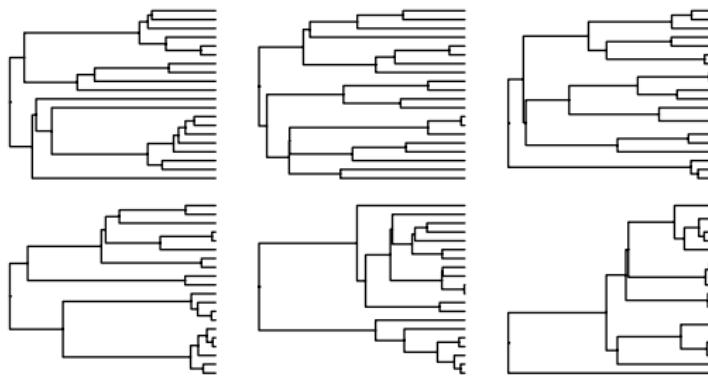
Phylogenies with branch lengths proportional to time provide valuable information about evolutionary history.



1st edition of Darwin's *On the Origin of Species* at Grinnell College (Grinnell, Iowa USA)

PRIORS ON THE TREE AND NODE AGES

Relaxed clock Bayesian analyses require a prior distribution on time trees



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

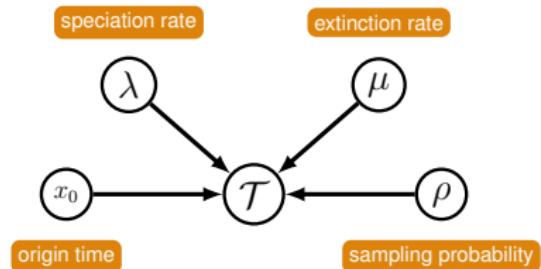
STOCHASTIC BRANCHING PROCESSES

Node-age priors based on stochastic models of lineage diversification

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, x_0

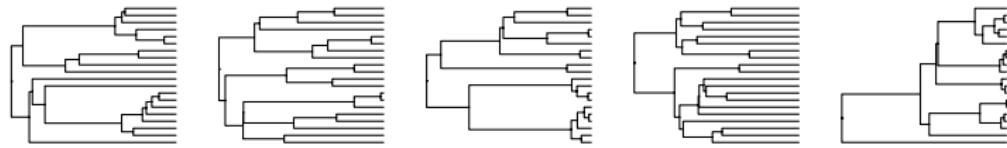


$$f(\mathcal{T} | \lambda, \mu, \rho, x_0)$$

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

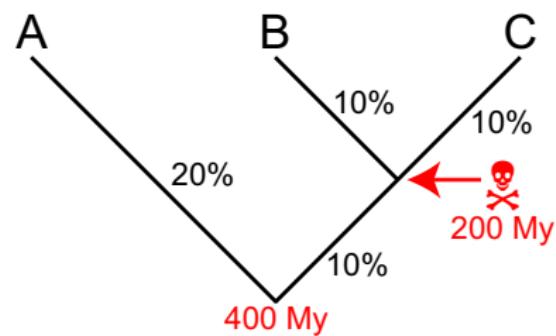


CALIBRATING DIVERGENCE TIMES

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

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

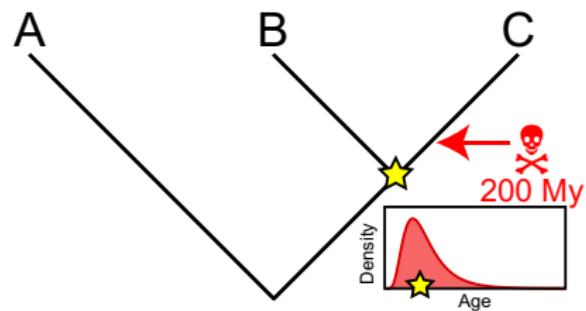
Uncertainty in the placement of fossils



CALIBRATION DENSITIES

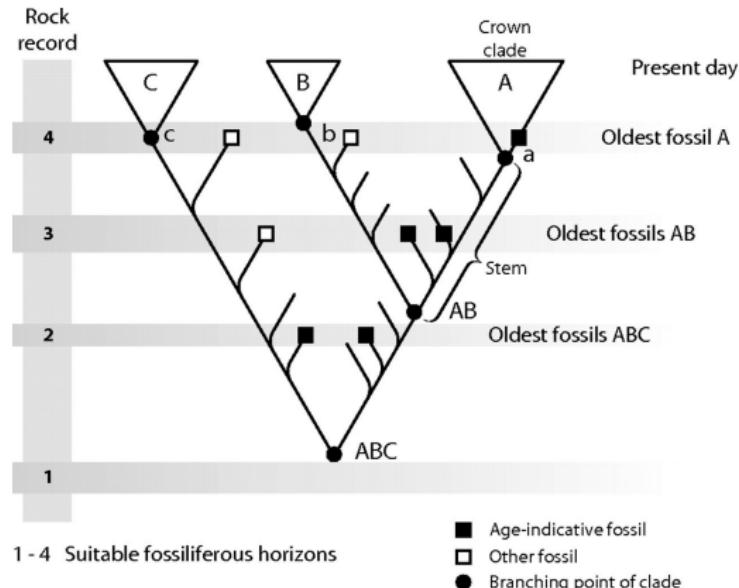
Bayesian inference is well suited to accommodating uncertainty in the age of the calibration node

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

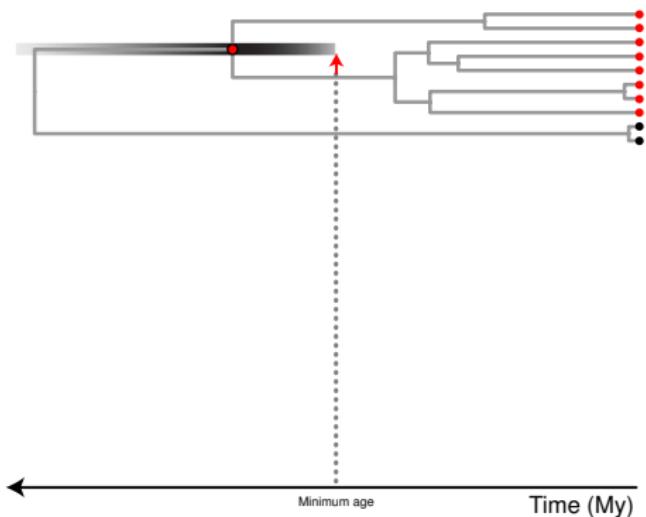


(figure from Benton & Donoghue *Mol. Biol. Evol.* 2007)

Fossil Calibration

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

Reliable **maximum** bounds are typically unavailable

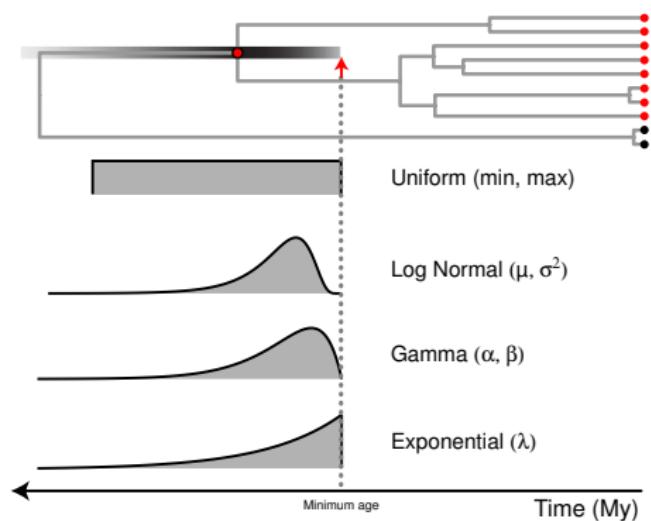


PRIOR DENSITIES ON CALIBRATED NODES

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

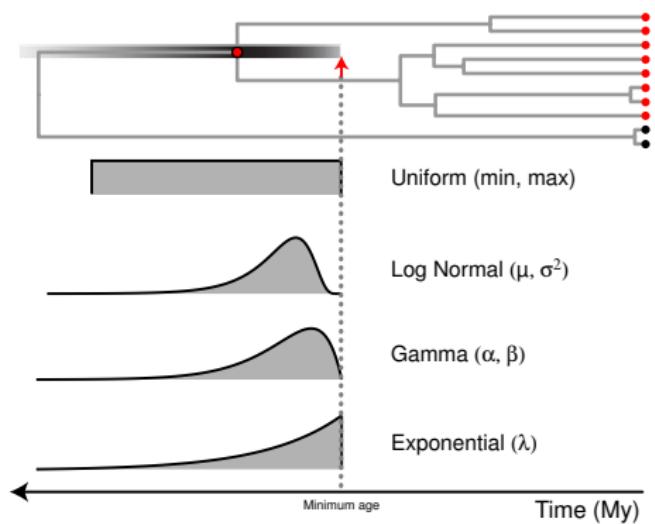


PRIOR DENSITIES ON CALIBRATED NODES

Common practice in Bayesian divergence-time estimation:

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

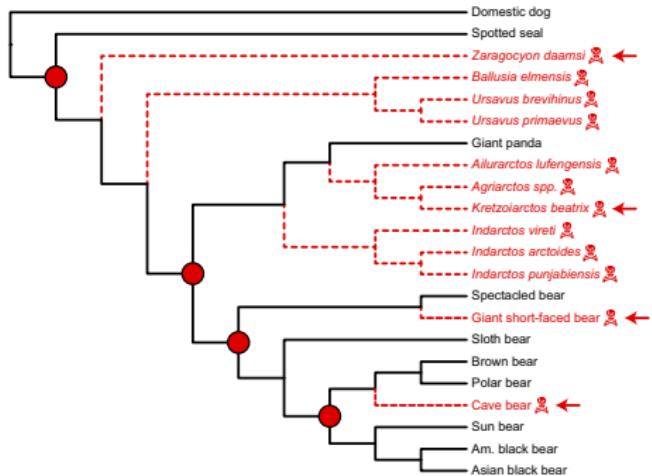
Specifying appropriate densities is a challenge for most molecular biologists



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

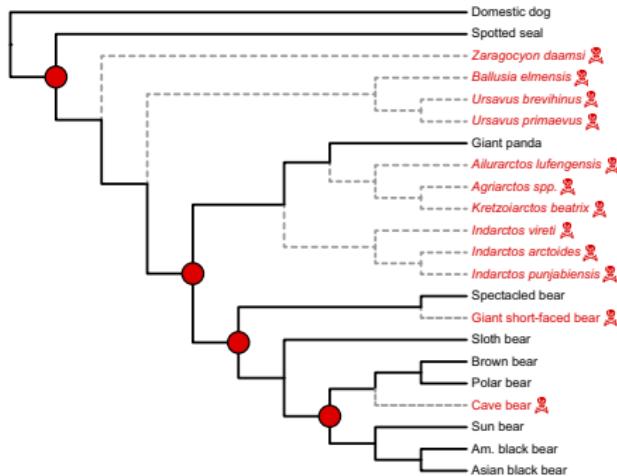
12 fossils are reduced to 4 calibration ages with calibration density methods



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

IMPROVING FOSSIL CALIBRATION

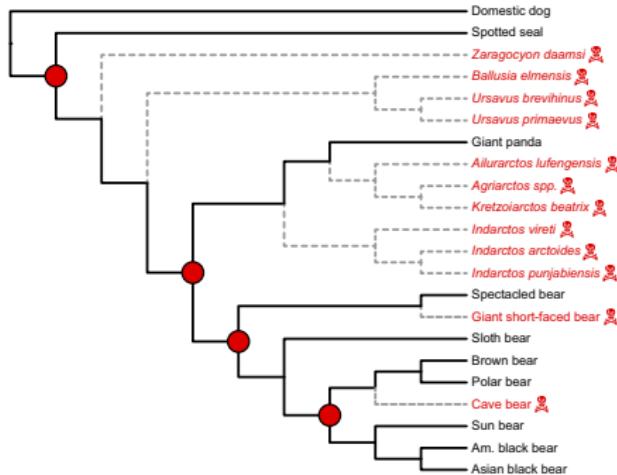
Because fossils are part of the diversification process, we can combine fossil calibration with birth-death models



(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)

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

© The Author(s) 2012. Published by Oxford University Press

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

DOI:10.1093/sysbio/sys058

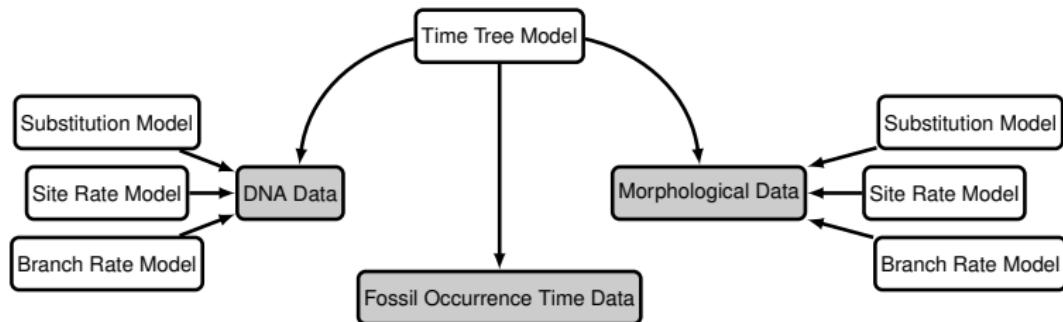
Advance Access publication on June 20, 2012

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

FREDRIK RONQUIST^{1,*}, SERAINA KLOPFSTEIN¹, LARS VILHELMSEN², SUSANNE SCHULMEISTER³, DEBRA L. MURRAY⁴, AND ALEXANDR P. RASNITSYN⁵

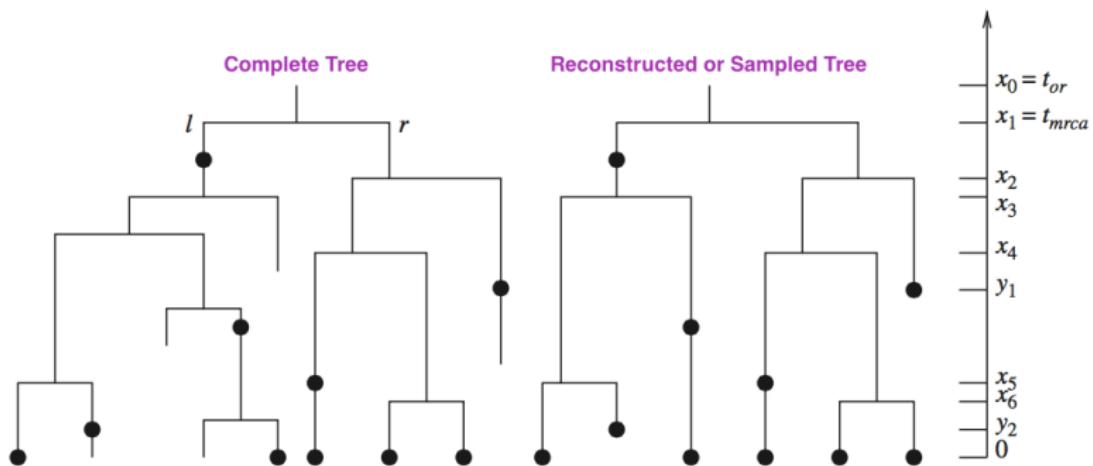
COMBINING FOSSIL & EXTANT DATA

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



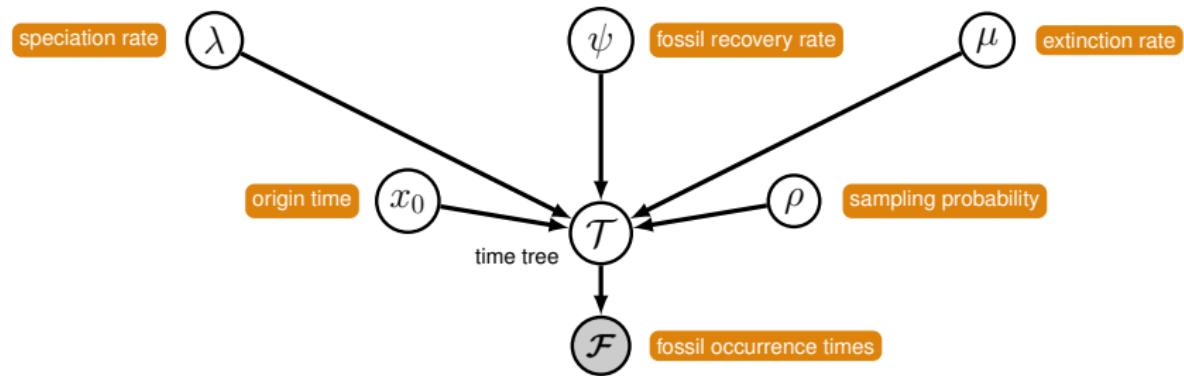
MODELING THE TREE & OCCURRENCE TIMES

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



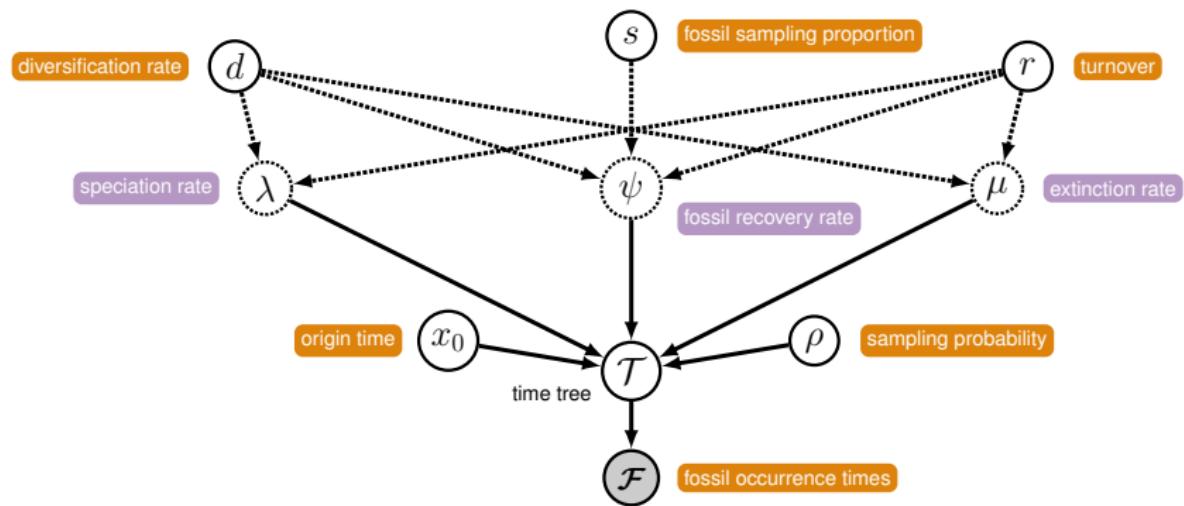
PARAMETERS OF THE FBD

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



PARAMETERS OF THE FBD

We re-parameterize the model so that we are directly estimating the diversification rate, turnover and fossil sampling proportion



$$\lambda = \frac{d}{1-r} \quad \mu = \frac{rd}{1-r} \quad \psi = \frac{s}{1-s} \frac{rd}{1-r}$$

THE FOSSILIZED BIRTH-DEATH PROCESS (FBD)

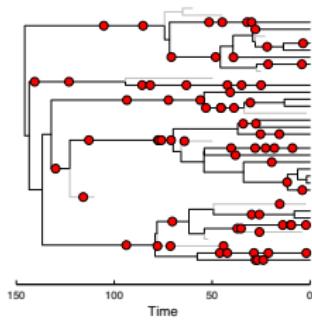
Improving statistical inference of absolute node ages

Eliminates the need to specify arbitrary calibration densities

Useful for 'total-evidence' analyses

Better capture our statistical uncertainty in species divergence dates

All reliable fossils associated with a clade are used

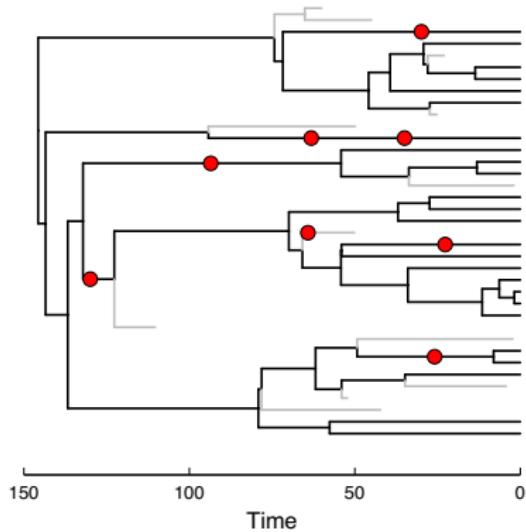


The fossilized birth–death process for coherent calibration of divergence-time estimates

Tracy A. Heath^{a,b}, John P. Huelsenbeck^{a,c}, and Tanja Stadler^{d,e,1}

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)

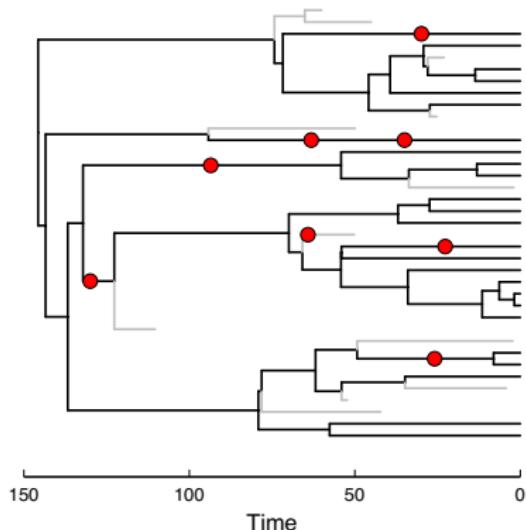
THE FOSSILIZED BIRTH-DEATH PROCESS (FBD)

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)

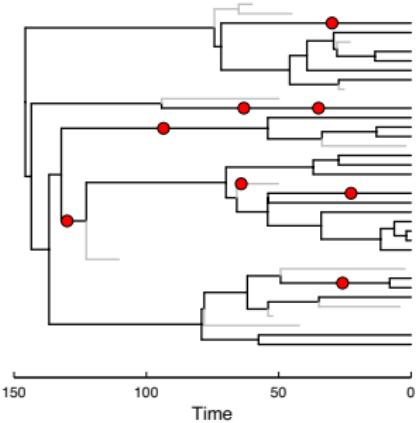
SAMPLED ANCESTORS

Sampled lineages with sampled descendants

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

On the probability of ancestors in the fossil record

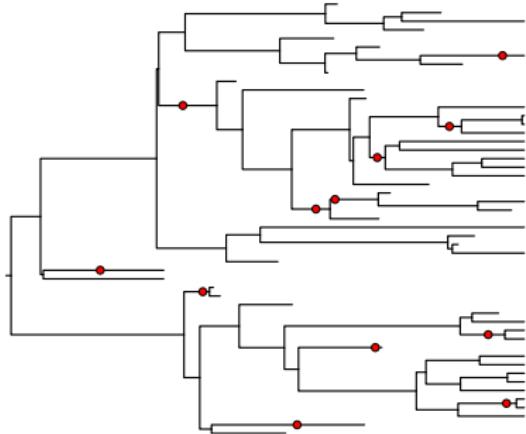
Mike Foote



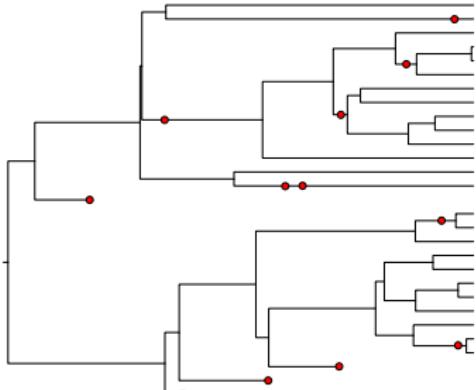
There is a non-zero probability of sampling ancestor-descendant relationships from the fossil record

SAMPLED ANCESTORS

Complete FBD Tree



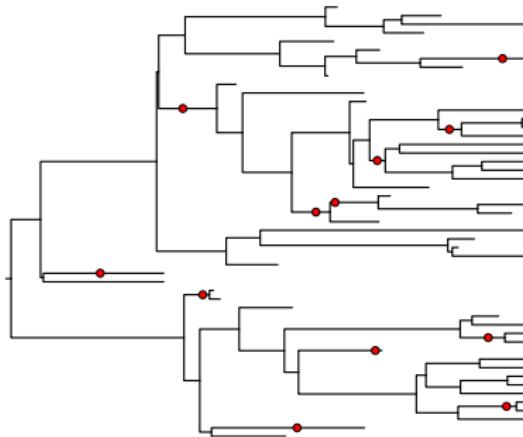
Reconstructed FBD Tree



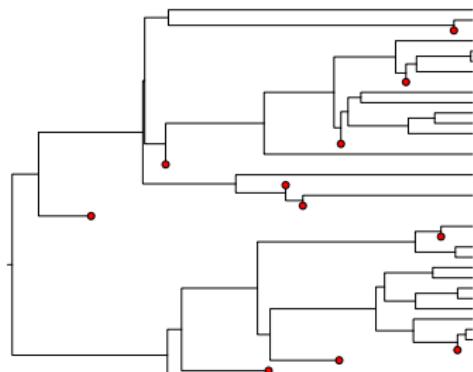
Because fossils & living taxa are assumed to come from a single diversification process, there is a non-zero probability of sampled ancestors

SAMPLED ANCESTORS

Complete FBD Tree

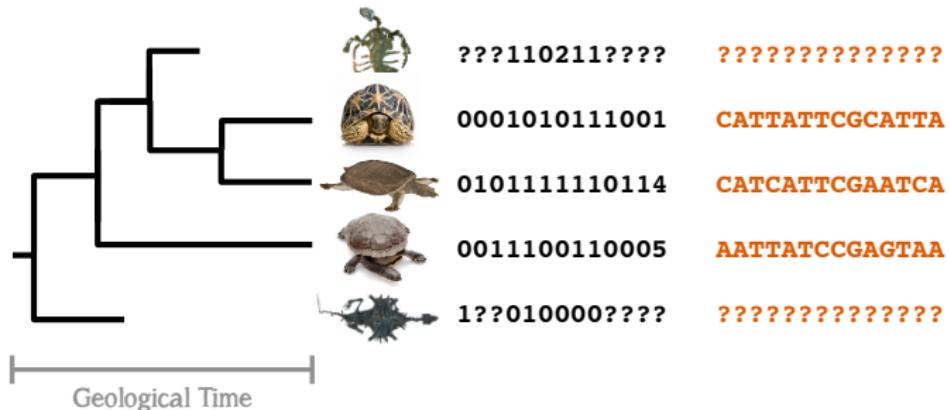
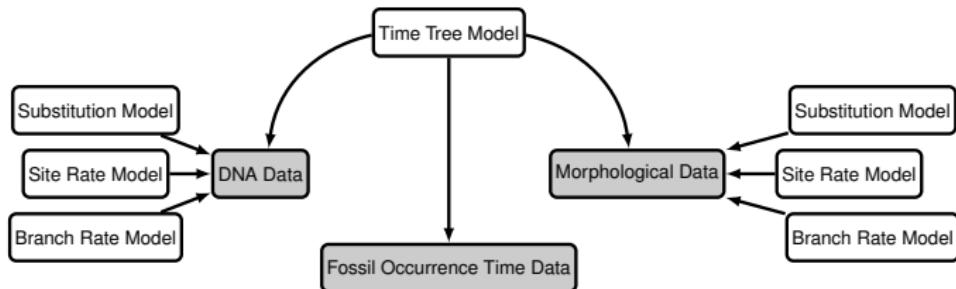


No Sampled Ancestor Tree

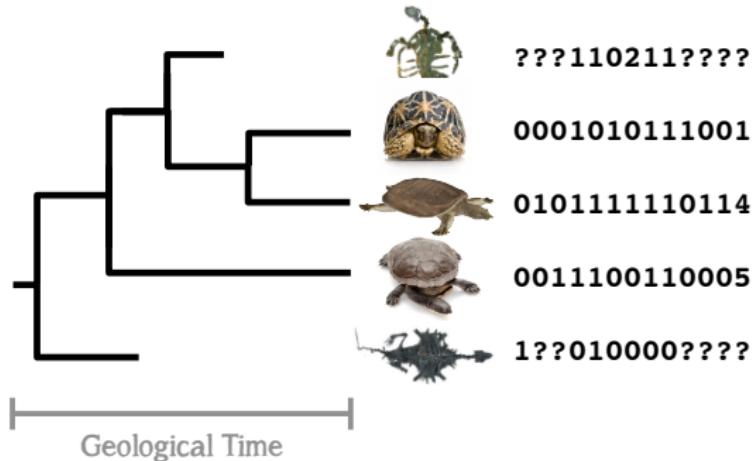


If all fossils are forced to be on separate lineages, this induces additional speciation events and will, in turn, influence rate & node-age estimates.

COMBINING FOSSIL & EXTANT DATA



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

MODELING MORPHOLOGICAL CHARACTER CHANGE

The Lewis Mk model

Assumes a character can take
 k states

T1 0

T2 0

T3 1

T4 2

T5 2

T6 1

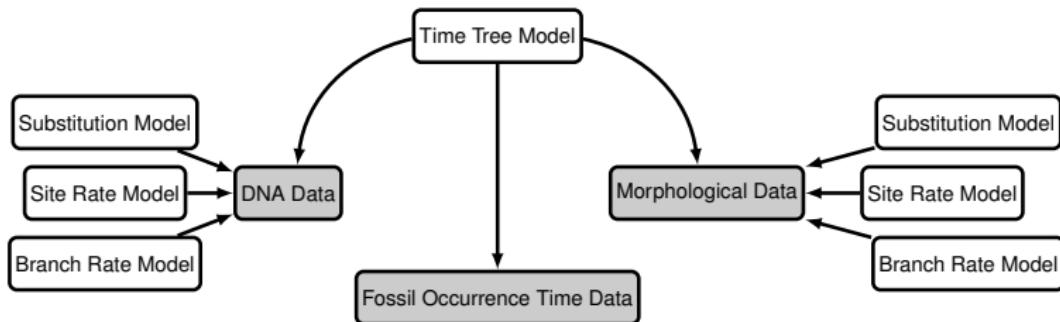
T7 1

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

(Lewis. *Systematic Biology* 2001)

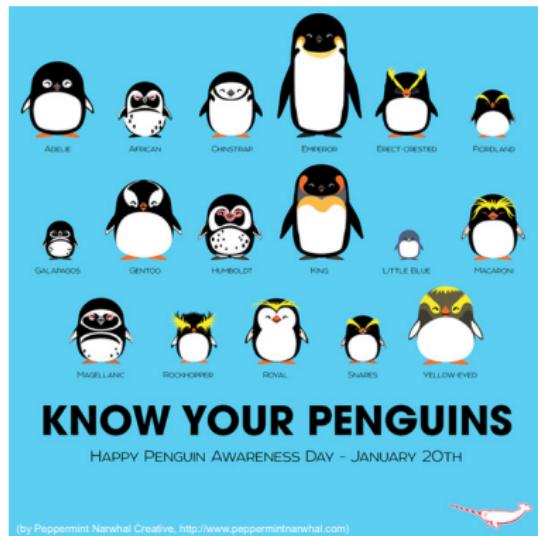
'TOTAL-EVIDENCE' ANALYSIS

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



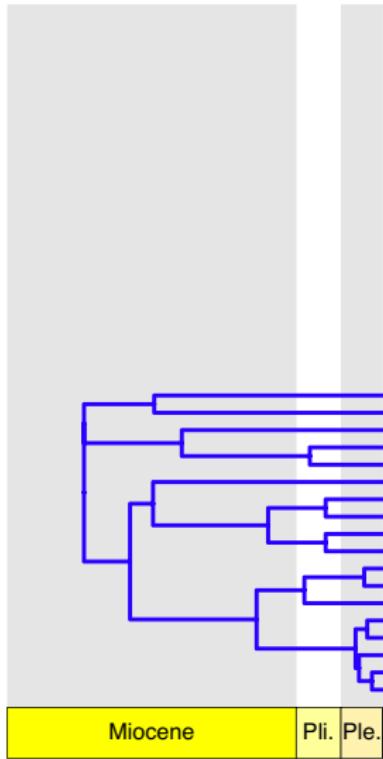
PENGUIN DIVERSITY IN DEEP TIME

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

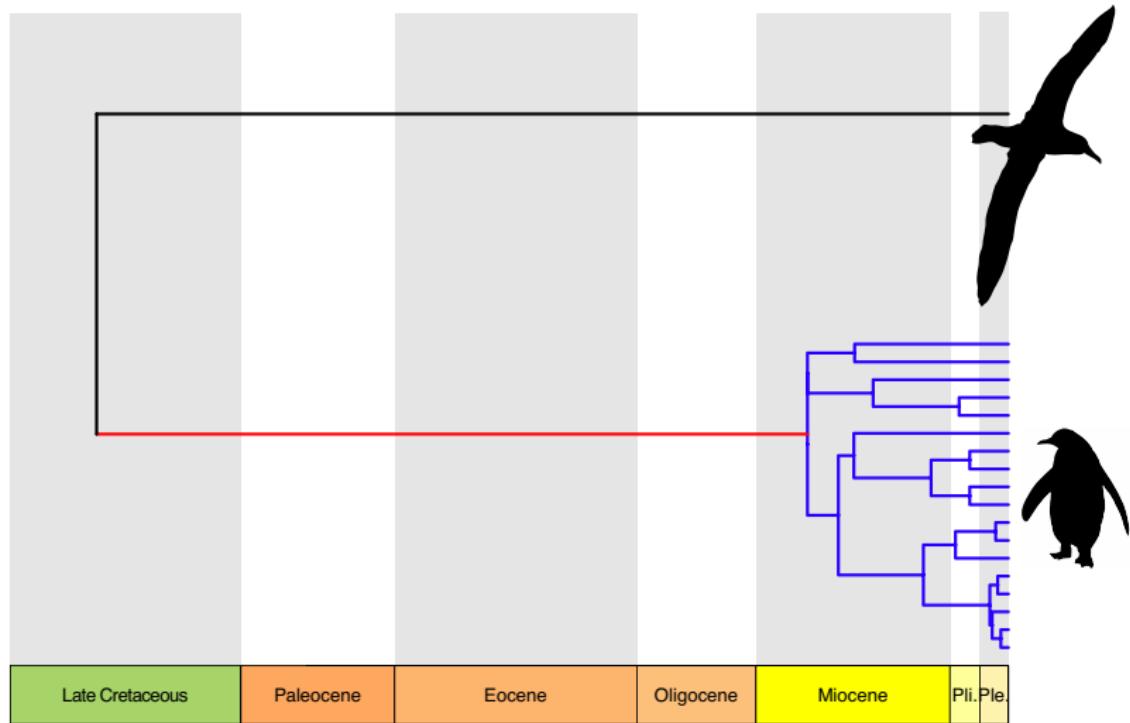


Artistic reconstructions by: Stephanie Abramowicz for Scientific American
Fonyea, R.E. and D.T. Ksepka. The Strangest Bird. *Scientific American* 307, 56 – 61 (2012)

PENGUIN DIVERSITY

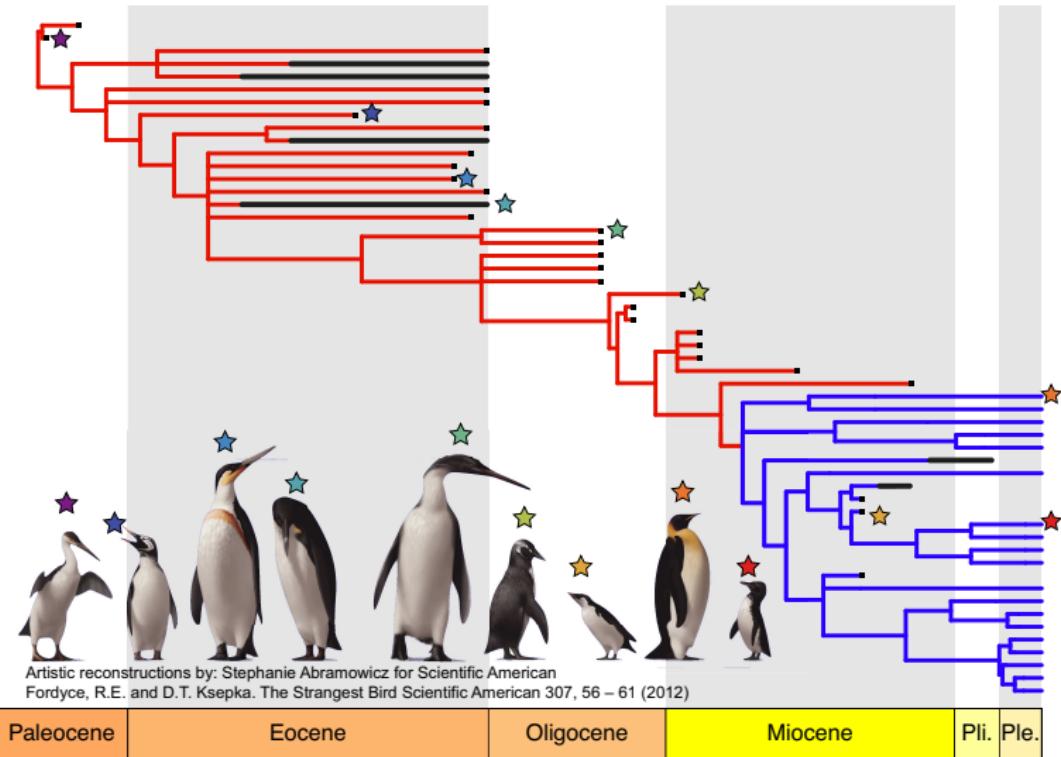


PENGUIN DIVERSITY

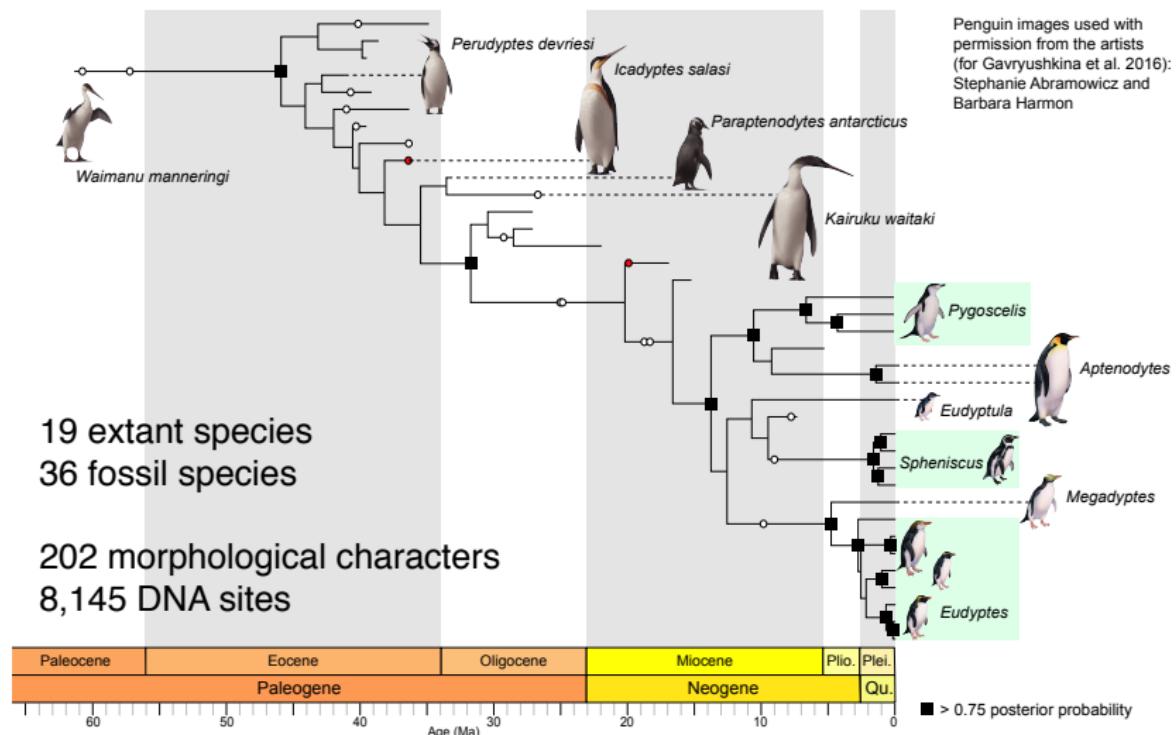


(silhouette images from <http://phylopic.org>)

PENGUIN DIVERSITY IN DEEP TIME

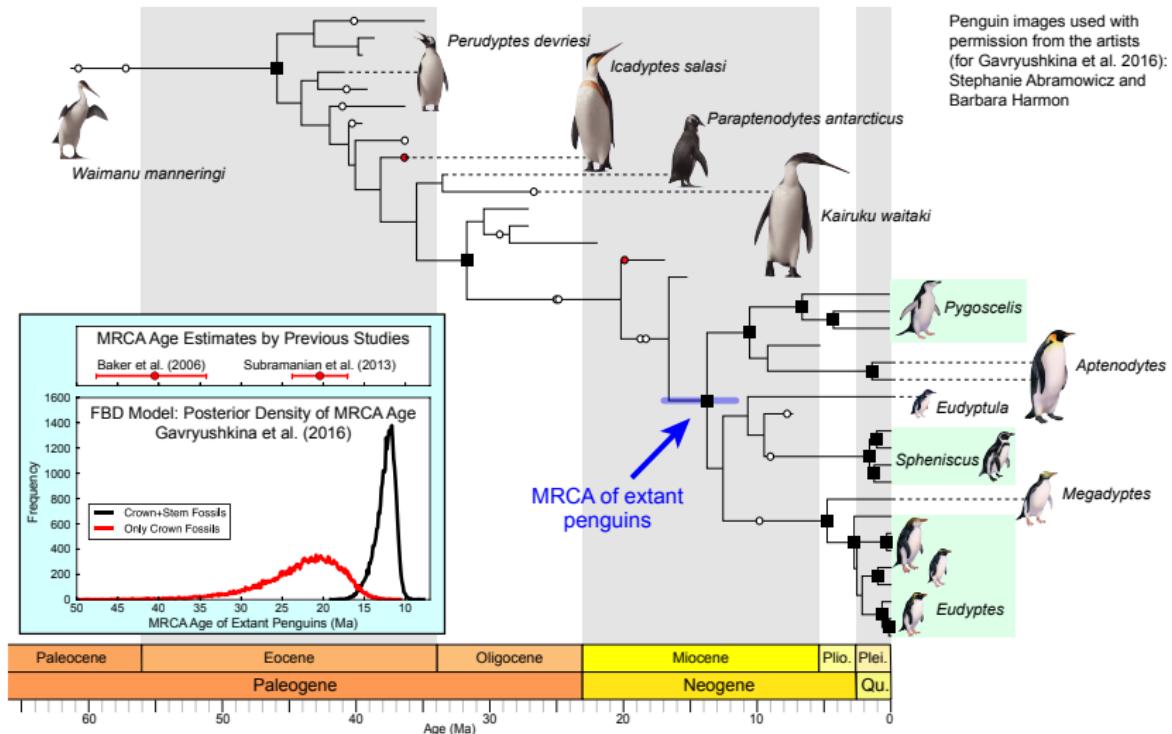


PENGUIN DIVERSITY IN DEEP TIME



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

PENGUIN DIVERSITY IN DEEP TIME



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



Taming the BEAST

Tutorial: Dating Species Divergences with the Fossilized Birth-Death Process

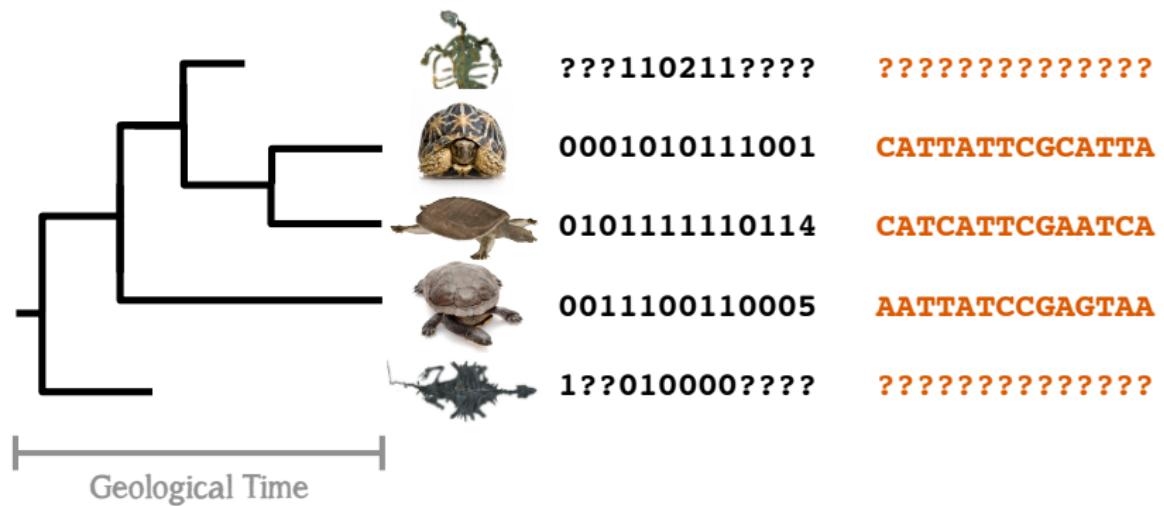
Download the tutorial files from:

<https://taming-the-beast.github.io/tutorials/FBD-tutorial>

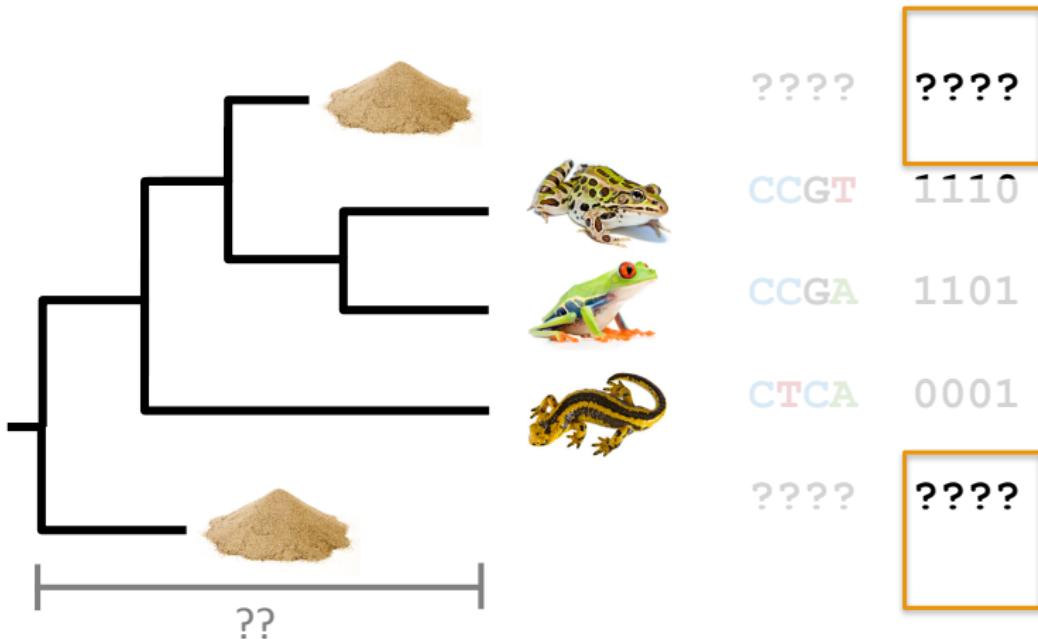
Additional BEAST 2 tutorials are available from the
Taming the BEAST website:

<https://taming-the-beast.github.io>

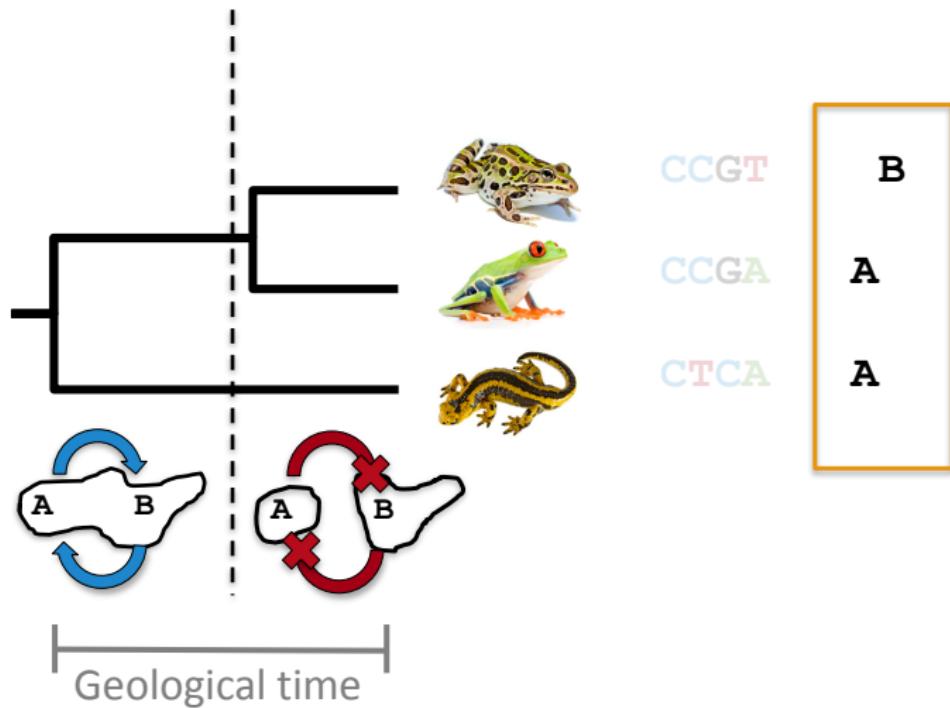
MOLECULES + MORPHOLOGY + FOSSILS



...but I study amphibians...



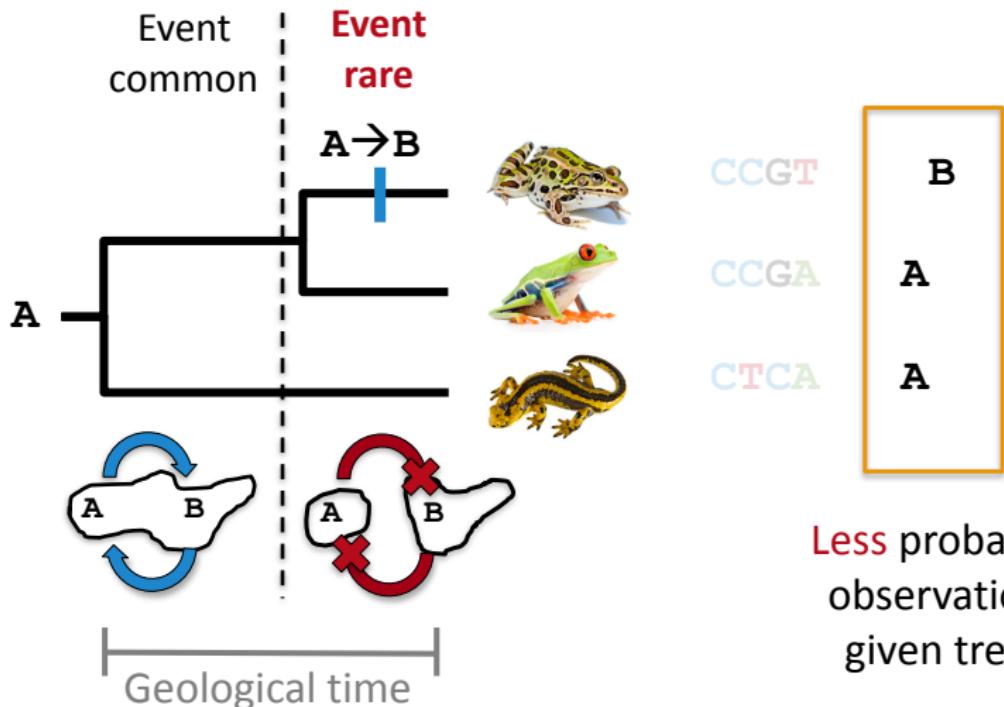
Molecules + biogeography + paleogeography



+ Paleogeography Landis, 2016

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

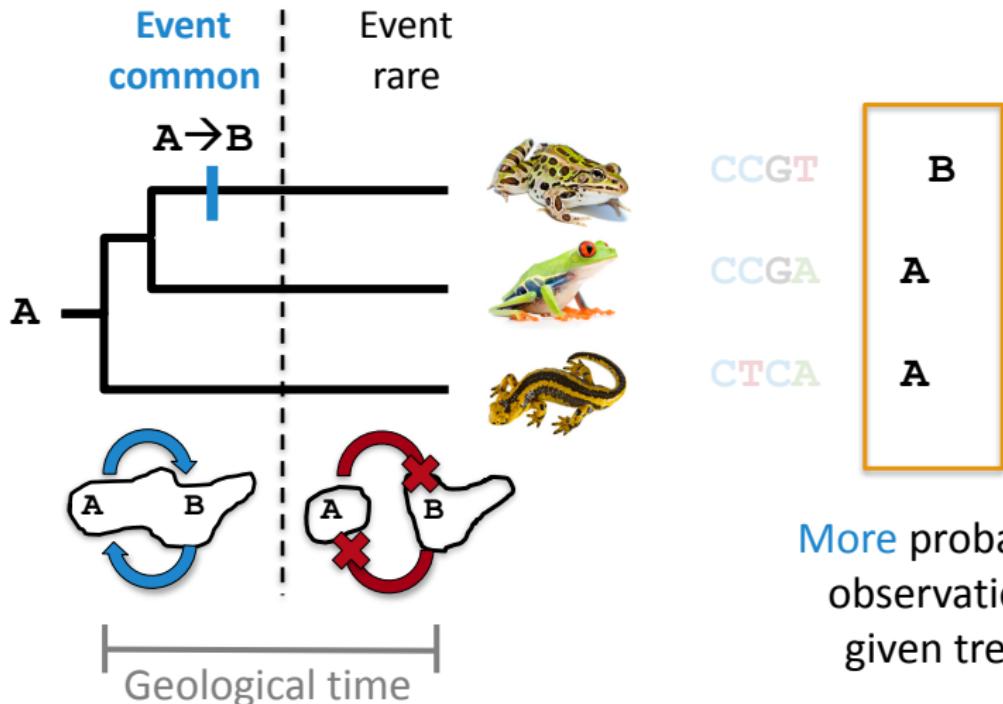
Events should occur *before* areas split



+ Paleogeography Landis, 2016

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

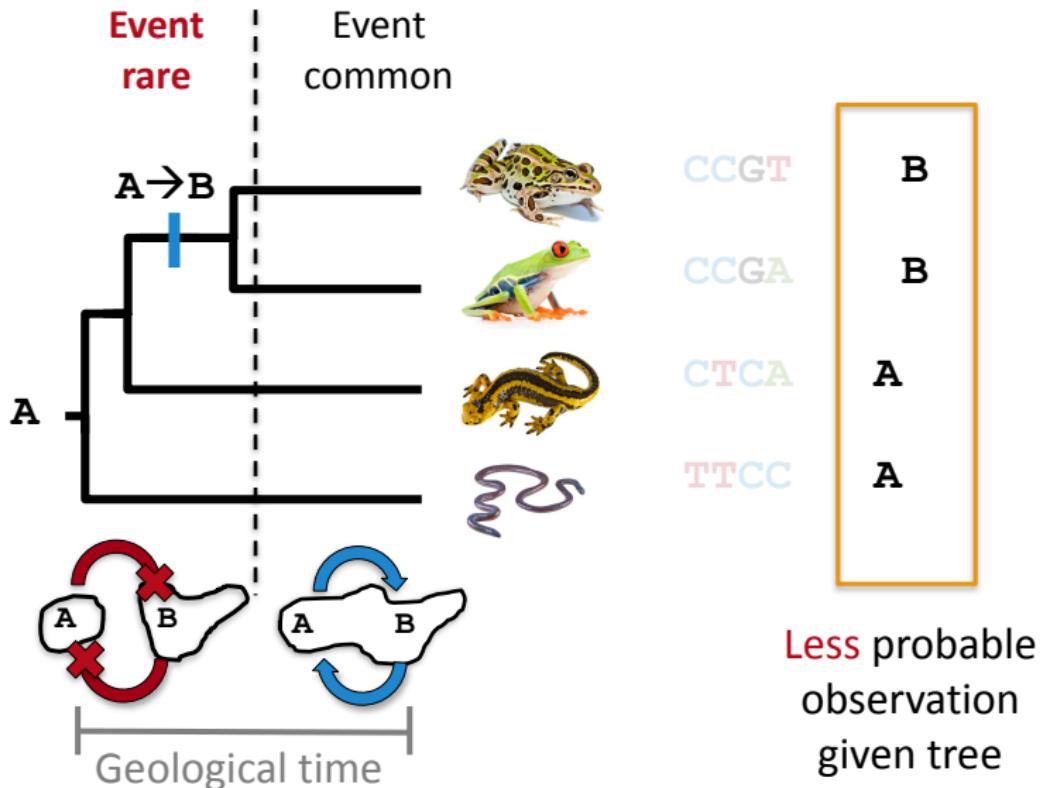
Events should occur *before areas split*



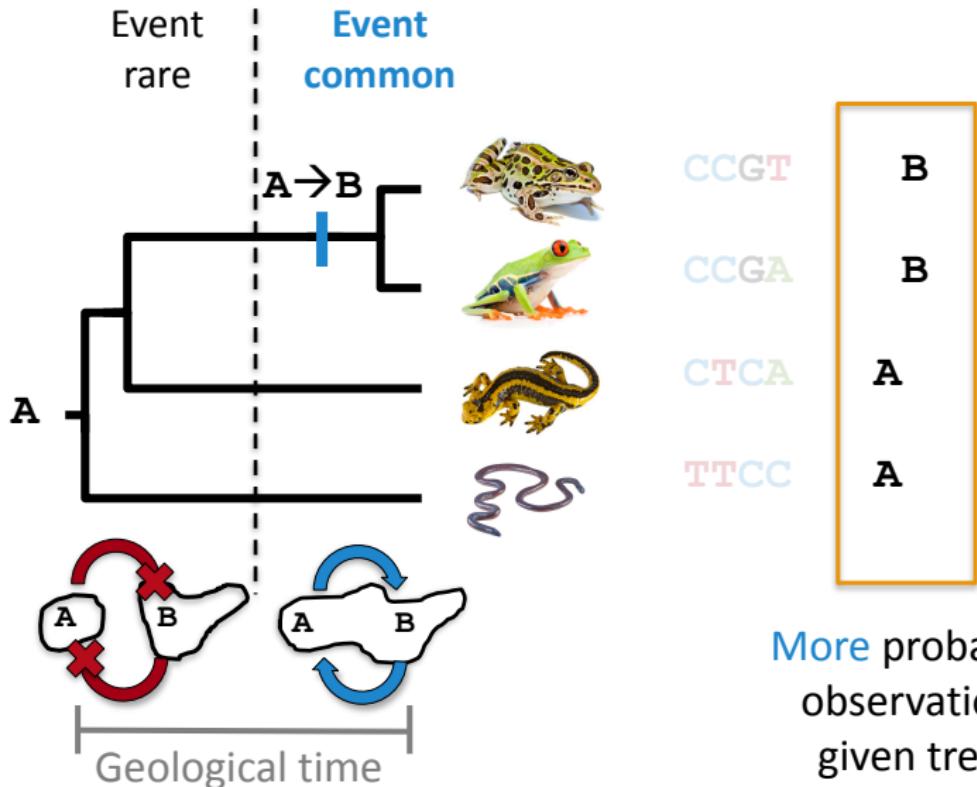
+ Paleogeography Landis, 2016

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

Events should occur *after* areas merge



Events should occur *after* areas merge



BIOGEOGRAPHIC DATING



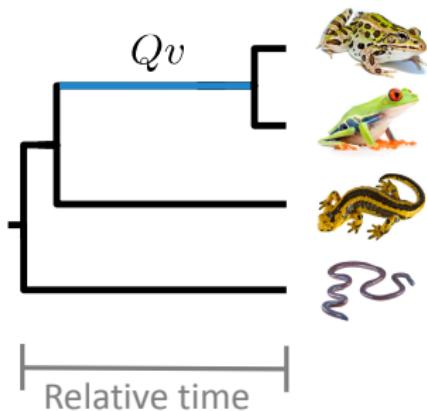
BIOGEOGRAPHIC DATING



BIOGEOGRAPHIC DATING

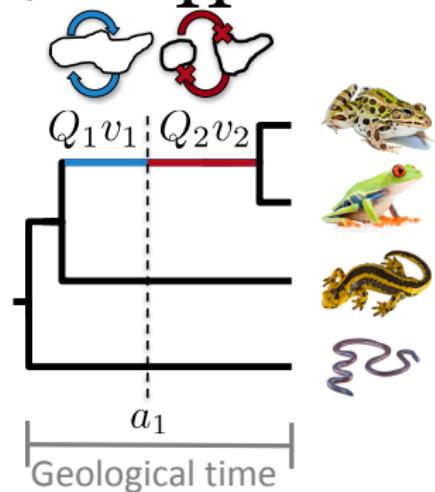
Constant Model

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



Epoch Model

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

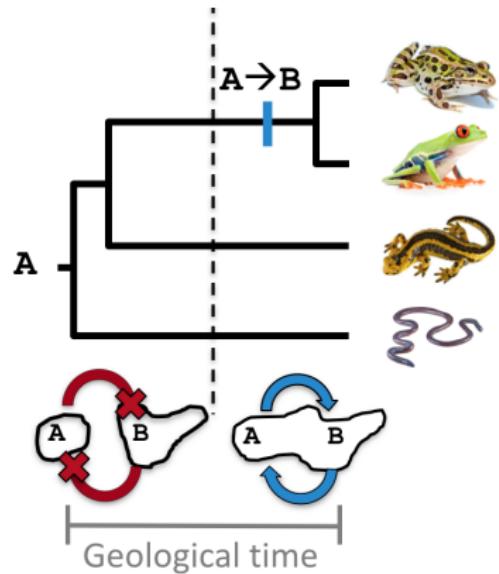


Epoch model Ree et al., 2005
 Bielejec et al., 2014

BIOGEOGRAPHIC DATING

Fossil-free calibration

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



Landis. In Press. "Biogeographic Dating of Speciation Times Using Paleogeographically Informed Processes". *Systematic Biology*, doi: 10.1093/sysbio/syw040.

BIOGEOGRAPHIC DATING

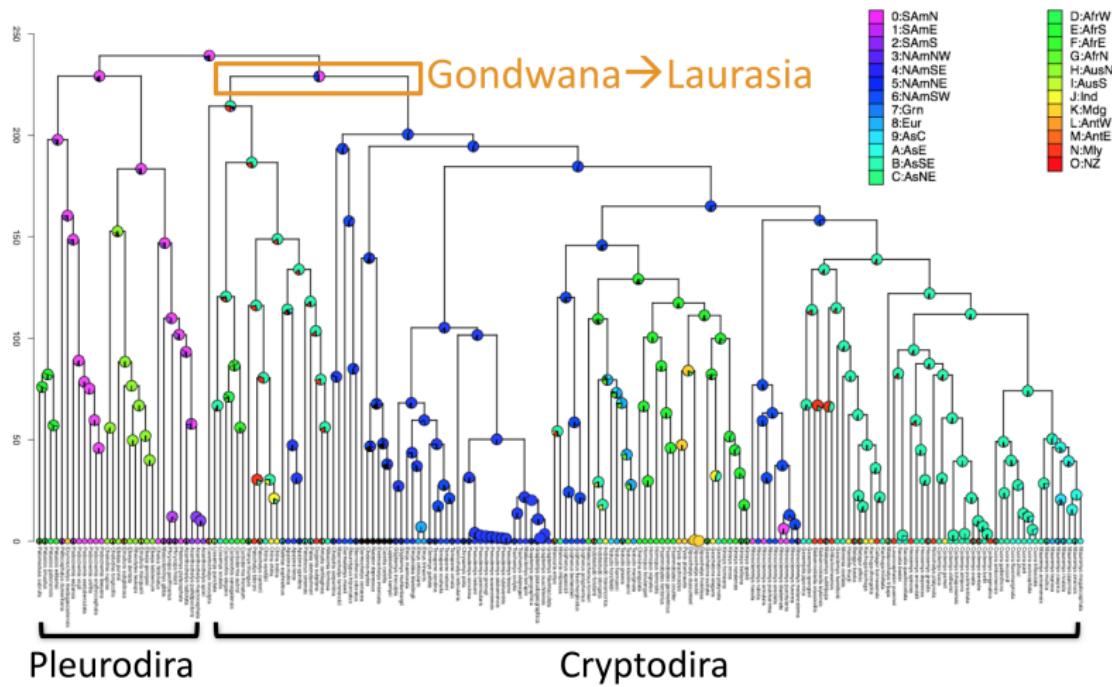
25 areas, 26 time-slices, 540–0Ma

strong=share land, **weak**=nearby land, **none**=all pairs

Connectivity model constructed using literature review and
GPlates (<http://www.gplates.org/>)

DATING + ANCESTRAL AREA RECONSTRUCTION

Ancestral area estimates (+G)



(image by M. Landis)