

Age estimation under Fossilized Birth Death

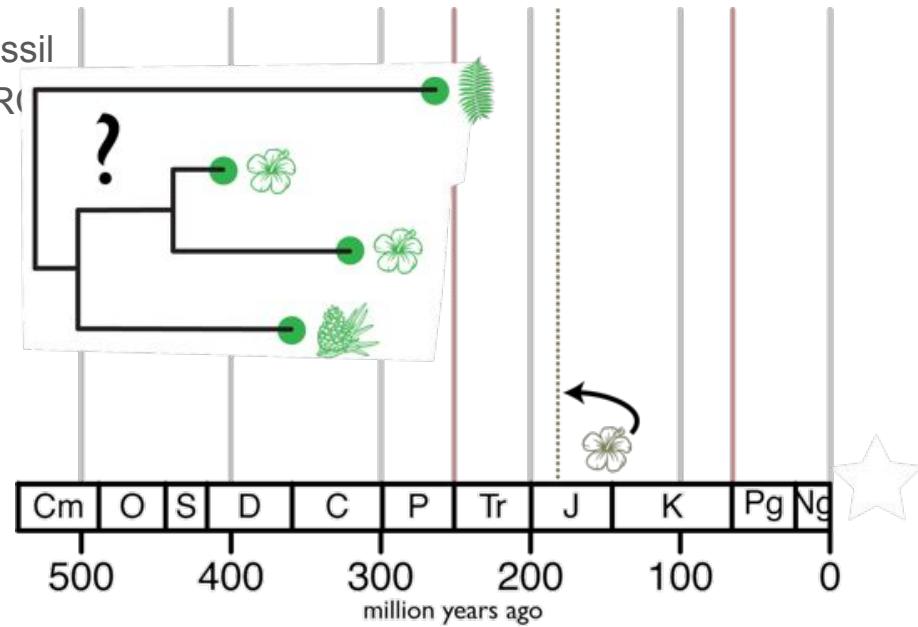
Ben Muddiman
Jenn Wagner

IB 290 - *Phylogenetics Seminar - Fall 2019*

Node-based Dating

Overview

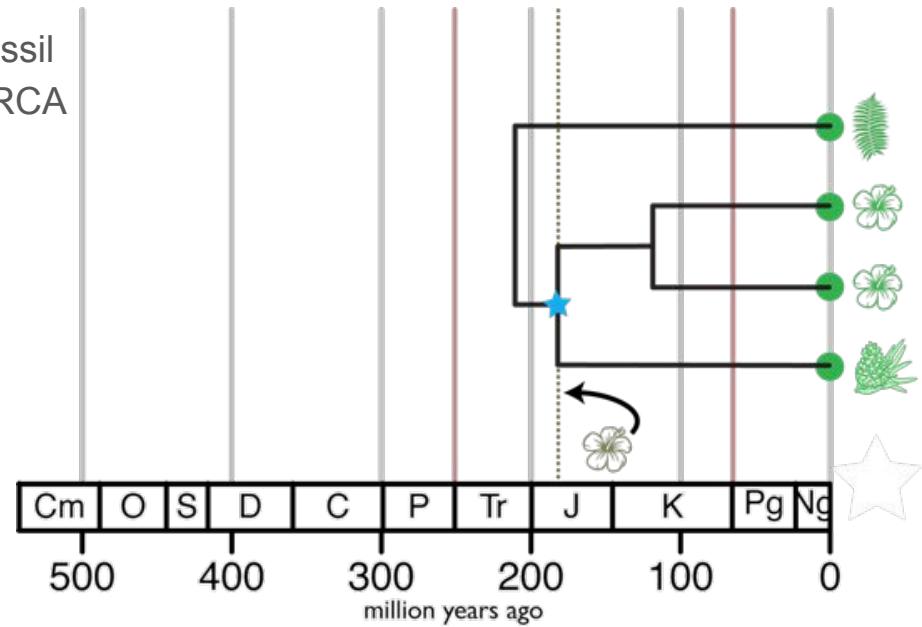
- **Ad hoc fossil placement:** An expert assigns a fossil to a specific node — this node is the proposed MRC b/w that fossil and extant taxa of the clade.



Node-based Dating

Overview

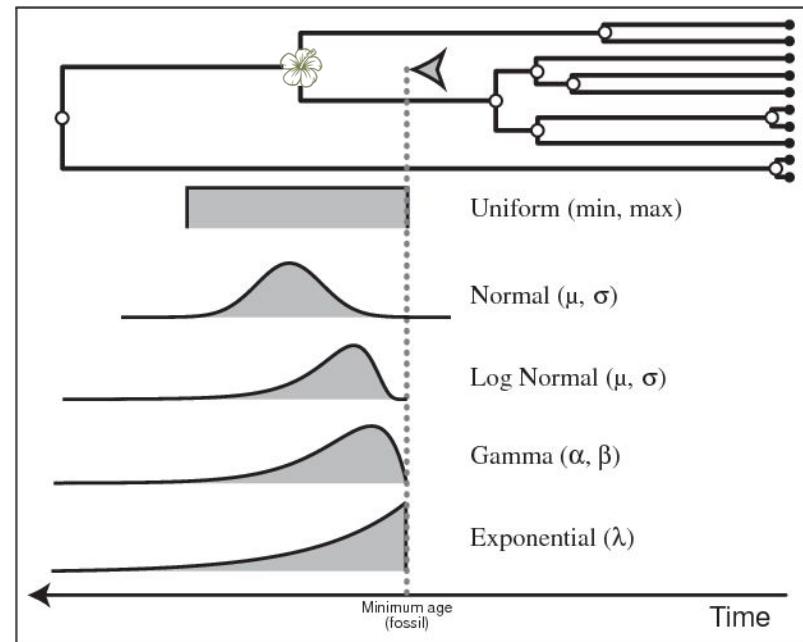
- **Ad hoc fossil placement:** An expert assigns a fossil to a specific node — this node is the proposed MRCA b/w that fossil and extant taxa of the clade.



Node-based Dating

Overview

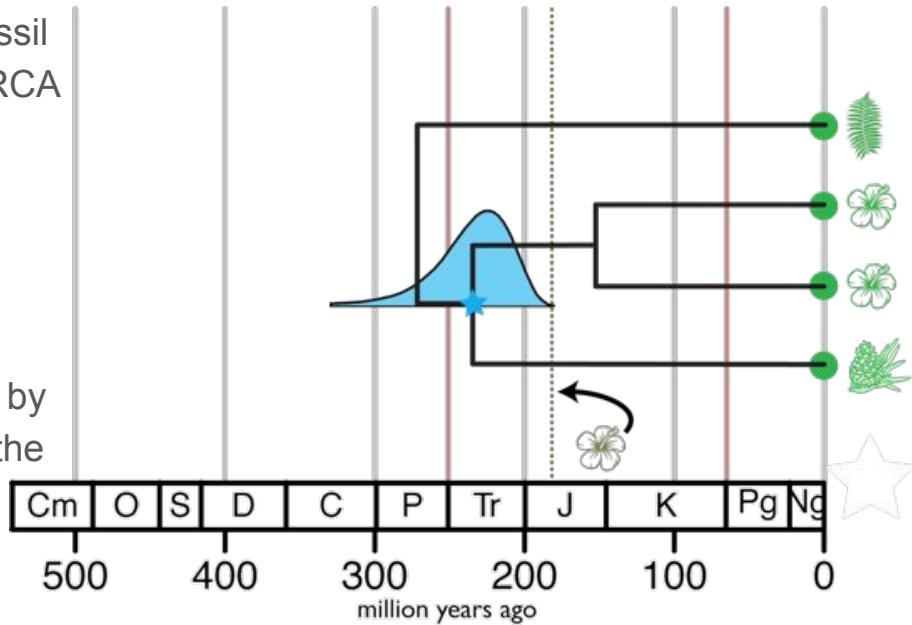
- **Ad hoc fossil placement:** An expert assigns a fossil to a specific node — this node is the proposed MRCA b/w that fossil and extant taxa of the clade.
- **Age assignment:** The age of the fossil is used to assign an age to the node. A probability density distribution of ages is assigned to the node. There are a variety of shapes that distribution can take (exponential, log-normal, uniform), ideally justified by some knowledge of the most likely correct age of the fossil/clade.



Node-based Dating

Overview

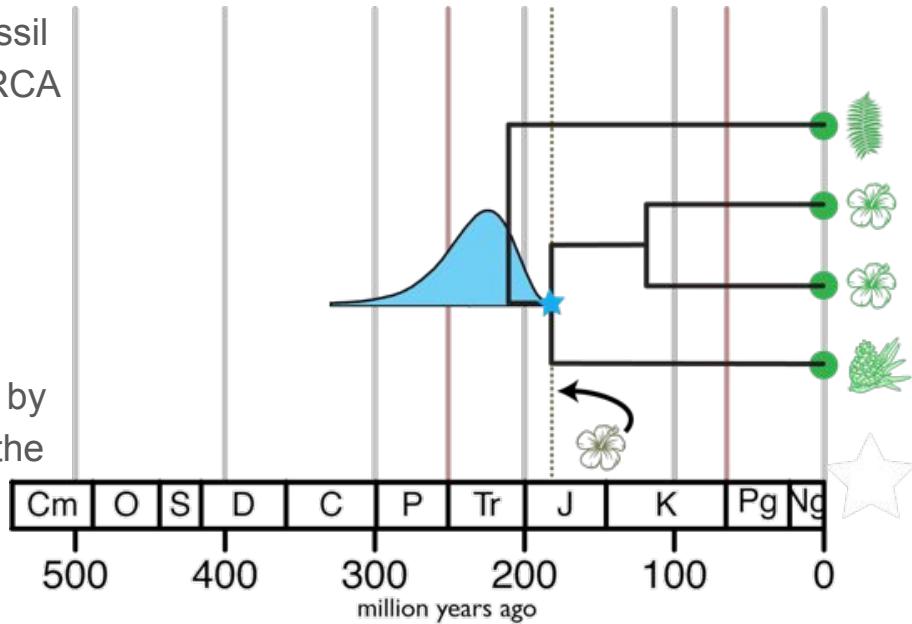
- **Ad hoc fossil placement:** An expert assigns a fossil to a specific node — this node is the proposed MRCA b/w that fossil and extant taxa of the clade.
- **Age assignment:** The age of the fossil is used to assign an age to the node. A probability density distribution of ages is assigned to the node. There are a variety of shapes that distribution can take (exponential, log-normal, uniform), ideally justified by some knowledge of the most likely correct age of the fossil/clade.
 - Where the fossil falls in that distribution will have a large effect on age reconstruction.



Node-based Dating

Overview

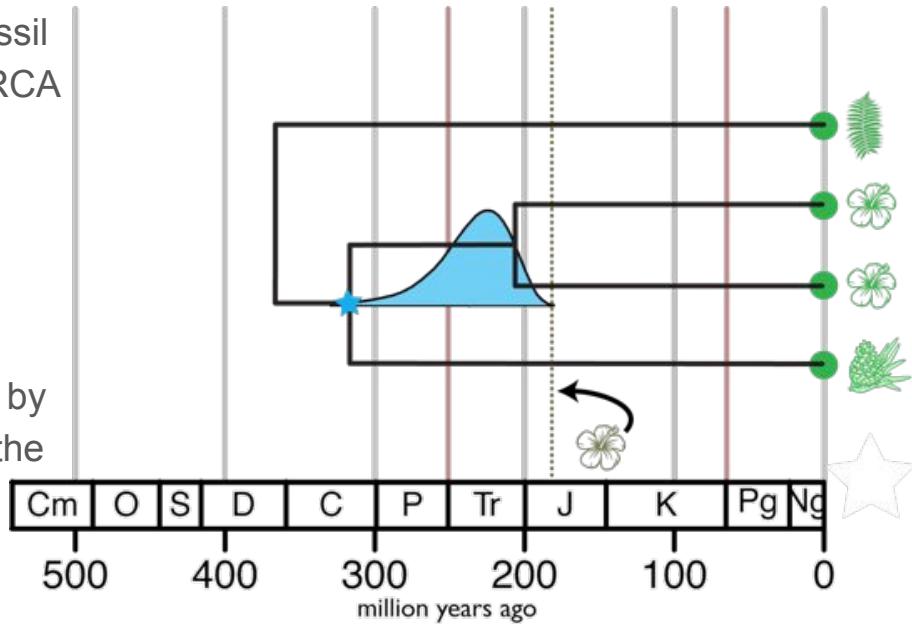
- **Ad hoc fossil placement:** An expert assigns a fossil to a specific node — this node is the proposed MRCA b/w that fossil and extant taxa of the clade.
- **Age assignment:** The age of the fossil is used to assign an age to the node. A probability density distribution of ages is assigned to the node. There are a variety of shapes that distribution can take (exponential, log-normal, uniform), ideally justified by some knowledge of the most likely correct age of the fossil/clade.
 - Where the fossil falls in that distribution will have a large effect on age reconstruction.



Node-based Dating

Overview

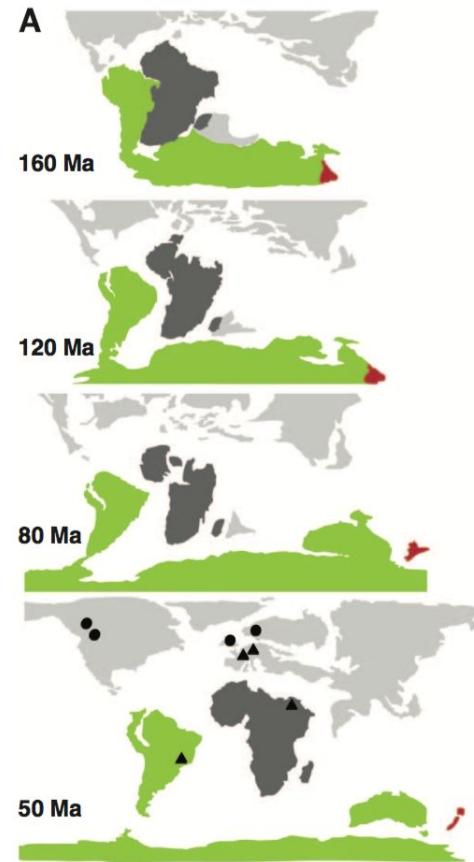
- **Ad hoc fossil placement:** An expert assigns a fossil to a specific node — this node is the proposed MRCA b/w that fossil and extant taxa of the clade.
- **Age assignment:** The age of the fossil is used to assign an age to the node. A probability density distribution of ages is assigned to the node. There are a variety of shapes that distribution can take (exponential, log-normal, uniform), ideally justified by some knowledge of the most likely correct age of the fossil/clade.
 - Where the fossil falls in that distribution will have a large effect on age reconstruction.



Node-based Dating

Overview

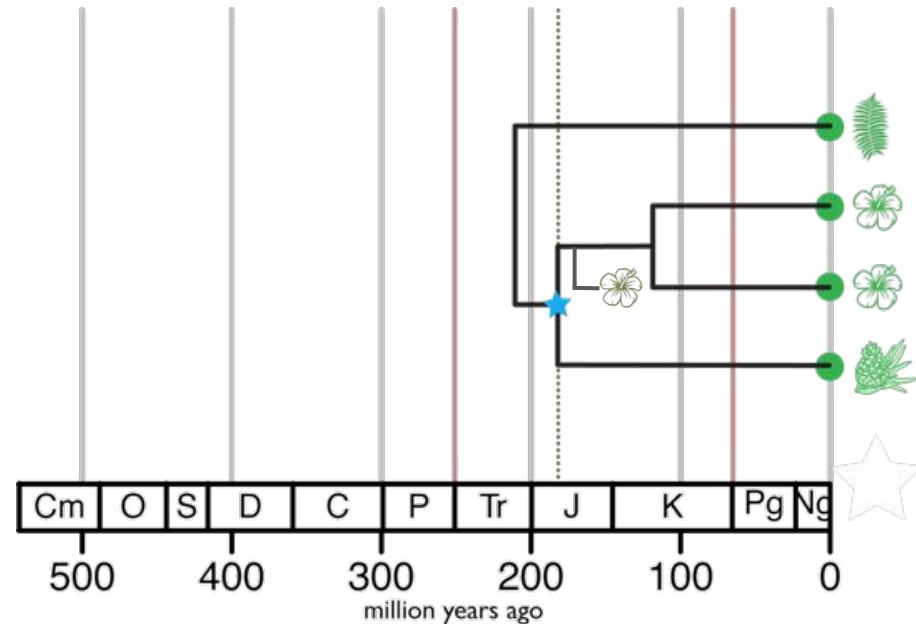
- **Ad hoc fossil placement:** An expert assigns a fossil to a specific node — this node is the proposed MRCA b/w that fossil and extant taxa of the clade.
- **Age assignment:** The age of the fossil is used to assign an age to the node. A probability distribution of ages is assigned to the node. The shape of this distribution can take on a variety of shapes (exponential, log-normal, uniform), ideally justified by some knowledge of the most likely correct age of the fossil.
- Note: Fossils are not the only things that can inform node age estimations. The timings of biogeographic events (e.g. continents splitting up) are an example of another potential age-calibration data source.



Node-based Dating

Known issues

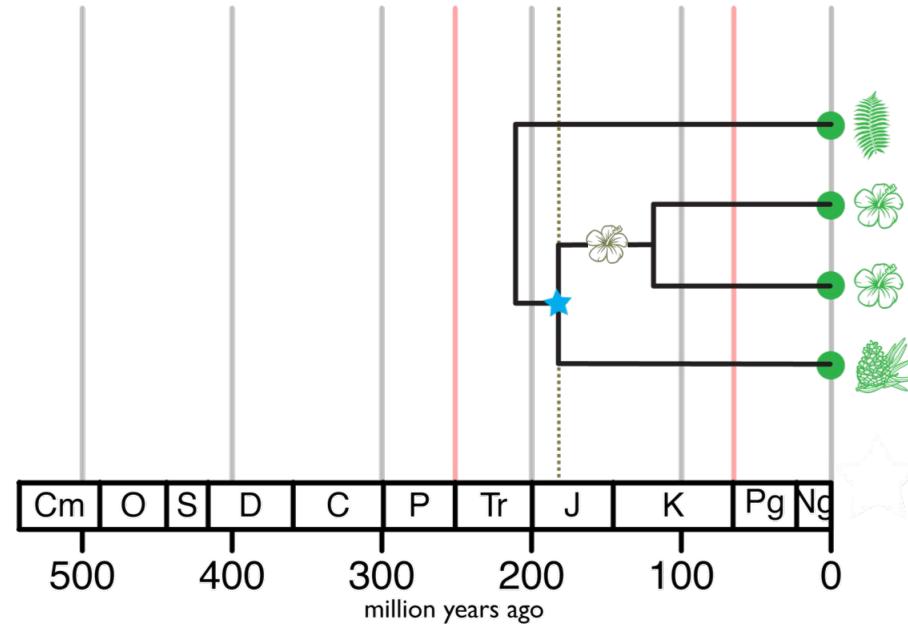
- *Ad hoc* fossil placement requires prior phylogenetic information
- The distribution used for age estimation at a node has major effects on results; and justification of a chosen distribution is challenging
- Your cladogenesis model (e.g. birth-death) and node age calibration density model both affect age reconstructions, and they can interact in confounding ways
- Fossils are placed at specific nodes, even though they likely sit on branches of unknown lengths
- Establishing age maxima of clades typically requires negative evidence



Node-based Dating

Known issues

- *Ad hoc* fossil placement requires prior phylogenetic information
- The distribution used for age estimation at a node has major effects on results; and justification of a chosen distribution is challenging
- Your cladogenesis model (e.g. birth-death) and node age calibration density model both affect age reconstructions, and they can interact in confounding ways
- Fossils are placed at specific nodes, even though they likely sit on branches of unknown lengths
- Establishing age maxima of clades typically requires negative evidence



Node-based Dating

Known issues

- *Ad hoc* fossil placement requires prior phylogenetic information
- The distribution used for age estimation at a node has major effects on results; and justification of a chosen distribution is challenging
- Your cladogenesis model (e.g. birth-death) and node age calibration density model both affect age reconstructions, and they can interact in confounding ways
- Fossils are placed at specific nodes, even though they likely sit on branches of unknown lengths
- Establishing age maxima of clades typically requires negative evidence

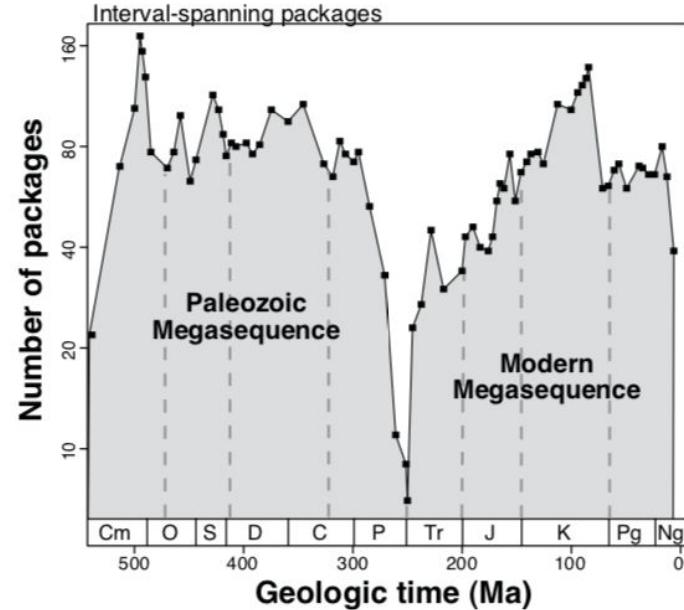
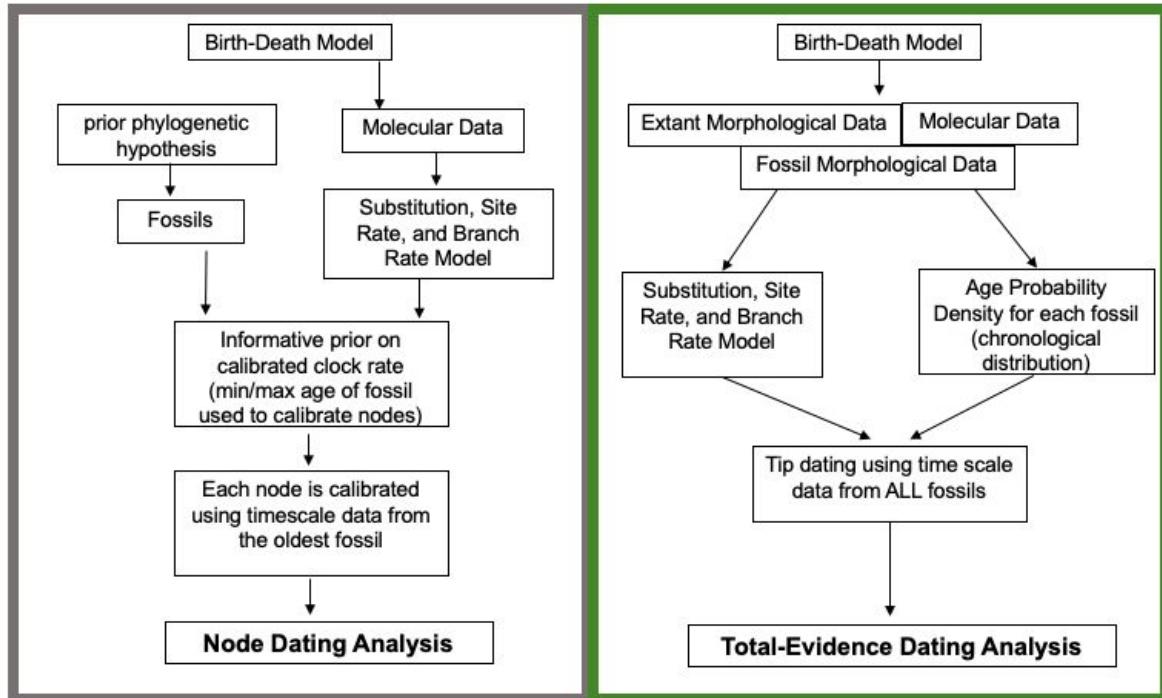


Figure 10—Total number of through-ranging marine sedimentary packages plotted on a logarithmic ordinate. Sequence boundaries from Figure 9 are shown by vertical dashed lines. There are two major Phanerozoic depositional cycles, or “megasequences,” in North America. Note the volatile Paleozoic plateau and the post-Paleozoic increase on a logarithmic scale.

Peters 2008

Total Evidence Dating (TED)

- No longer *ad hoc* placement of fossils
- **Tip-dating** using fossils to inform the phylogenetic position
- Doesn't require a prior phylogenetic hypothesis
- Uses all fossils, not just the oldest
- Simultaneously estimates fossil placement, topology, and evolutionary timescale using ALL available data



Total Evidence Dating (TED)

O'Reilly et al. 2015

- Incorporating stratigraphic information by correlating sections
- If time is known above or below the strata, it can be used to place a distribution on the fossils in question

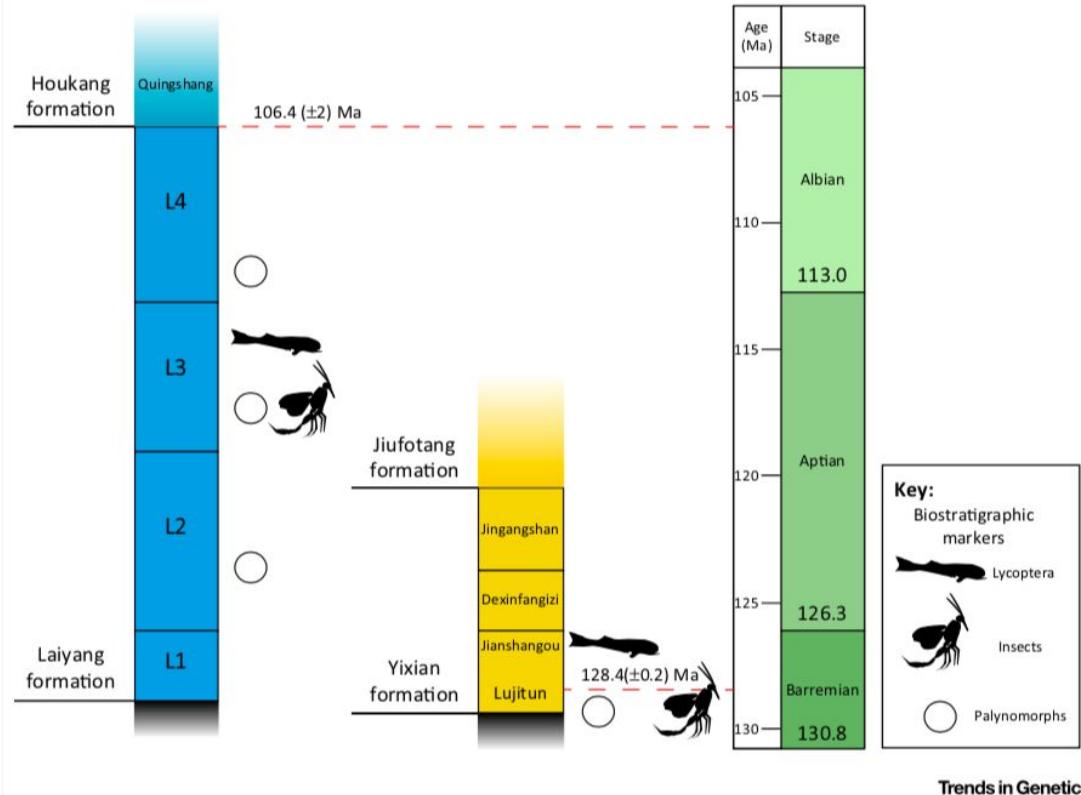


Figure I. Construction of a Tip-Calibration for *P. laiangensis* Based on Stratigraphic Correlation Between the Unit of Provenance, The Laiyang Formation, and the Yixian Formation of China.

Total Evidence Dating (TED)

- Placing an age probability density on the fossil
- Probability density functions
 - Depending on investigator's knowledge, the prior used for the chronological distribution of each fossil tip can vary

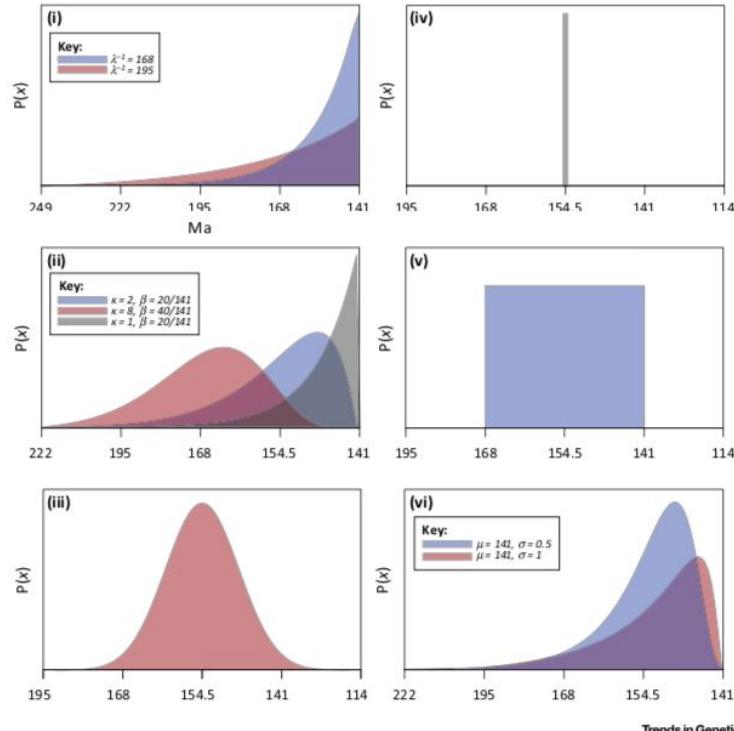


Figure I. Six Alternative Probability Density Functions Commonly Used to Encapsulate Prior Knowledge of the Chronological Distribution of a Fossil Tip. Here the calibration of the fossil taxon *Eoxyela* is used to demonstrate the characteristics of the different distributions.

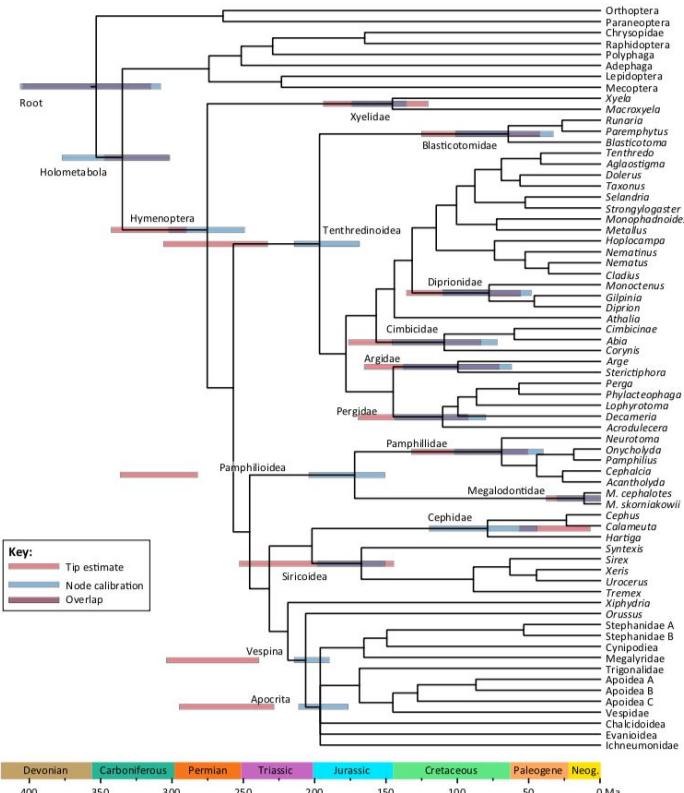
Total Evidence Dating (TED)

Benefits

- Uses all fossils, not just the oldest fossil
- Enables you to use fossils with less diagnostic features and/or poor preservation
- Constructing all parts of the tree at once decreases risk of bias



Total Evidence Dating (TED)



Draw-backs

- Doesn't account for evolutionary relationships that include ancestor-descendent pairs
 - Poorly preserved fossils could have belonged to separate lineage or could be an ancestor
- Stem-ward slippage
- Consistently yields older age reconstructions (systematic bias issue)

Fossilized Birth Death (FBD)

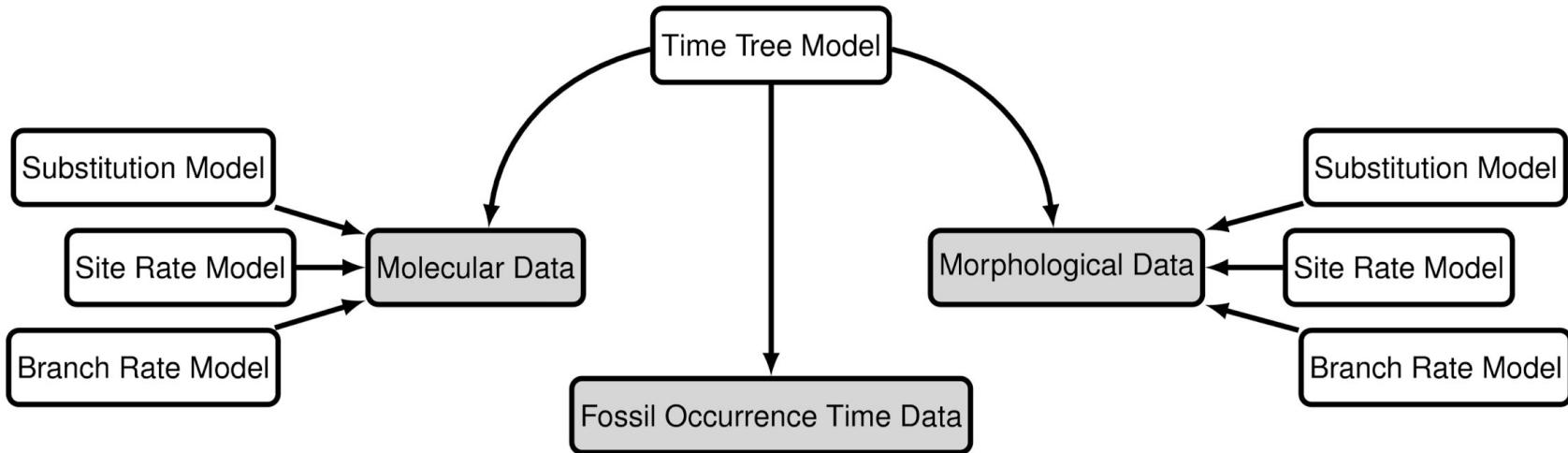


Figure 1. Modular components of the graphical model used in the “combined-evidence” analysis described in this tutorial.

Fossilized Birth Death (FBD)

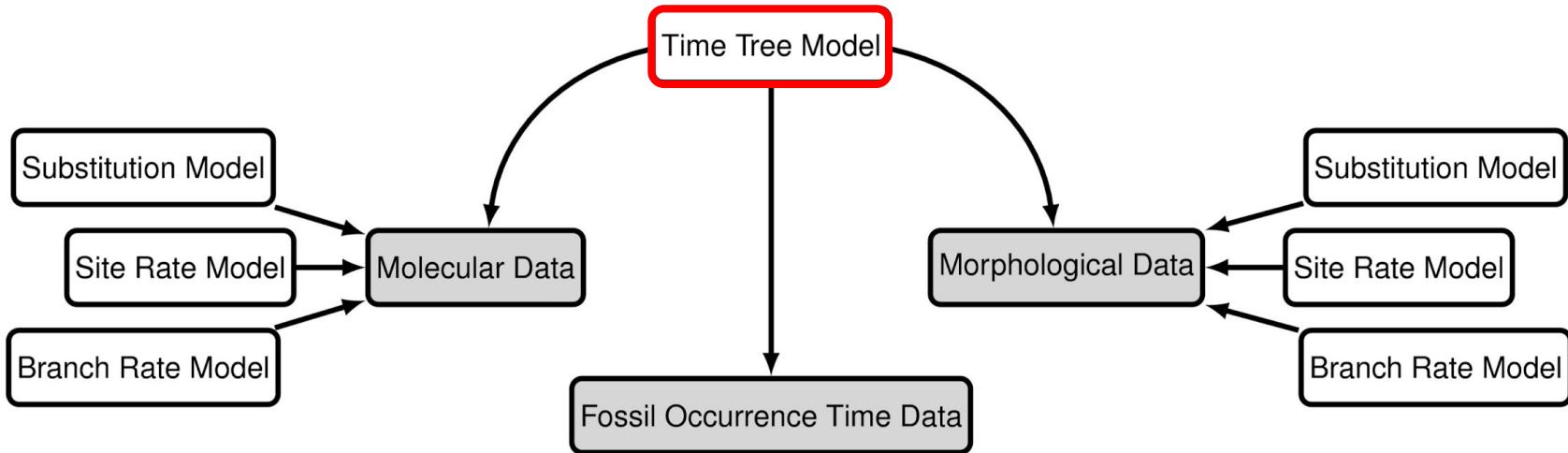
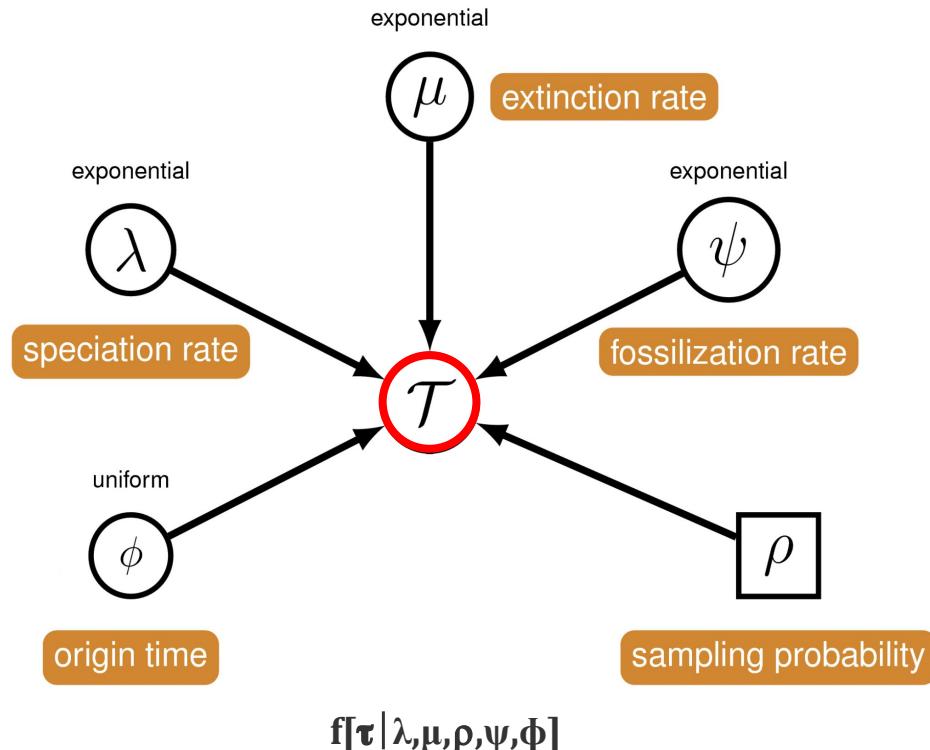


Figure 1. Modular components of the graphical model used in the “combined-evidence” analysis described in this tutorial.

Fossilized Birth Death (FBD)

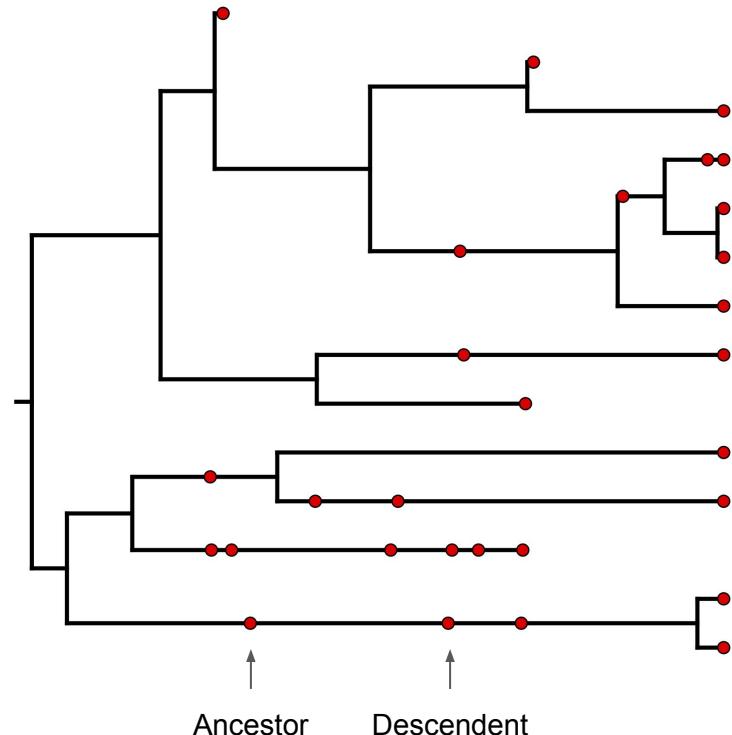
- The FBD models has five core parameters: speciation rate (λ), extinction rate (μ), fossilization rate (fossil discovery rate; ψ), sampling probability (extant taxa; ρ), and origin time (stem; ϕ) → these are used to model τ
 - τ = tree topology, divergence times, fossil occurrence times, and the times at which the fossils attach to the tree
- Each parameter has its own prior; however, you do not have separate priors for individual fossils or nodes. Rather, ages are estimated as a part of the overall model.  This represents a significant departure from traditional Node-based or TED methods.
- Fossils can either be placed *a priori* or their phylogenetic placement can be estimated using TED (preferred).



Fossilized Birth Death (FBD)

Benefits of FBD

- Fossil and extant taxa are modeled using the same underlying processes
- FBD does not require *ad hoc* calibration priors, as the age estimation is included in the overall model
- No unintended interaction b/w age priors and cladogenesis model
- Fossils can appear in an Ancestor-Descendent (A-D) fashion (on a lineage) or as a Speciation Event (branching)



Fossilized Birth Death (FBD)

Benefits of FBD (cont.)

- FBD can be combined with TED
 - Note: when run w/o total evidence, FBD marginalizes over all possible fossil placements, meaning some prior placement of fossils is required
- The addition of more fossils to the tree improves date estimates even if the exact placement of those fossils is not known
- Allows for the implication of a variety of MCMC methods to improve modeling
 - E.g. add-branch moves, delete-branch moves, a node-slider mechanism, and a subtree-pruning-and-regraft (SPR) moves mechanism

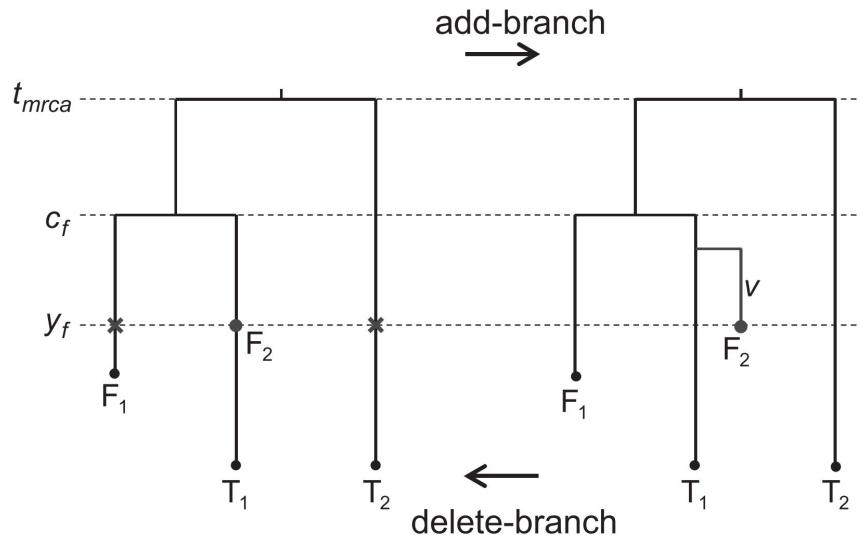
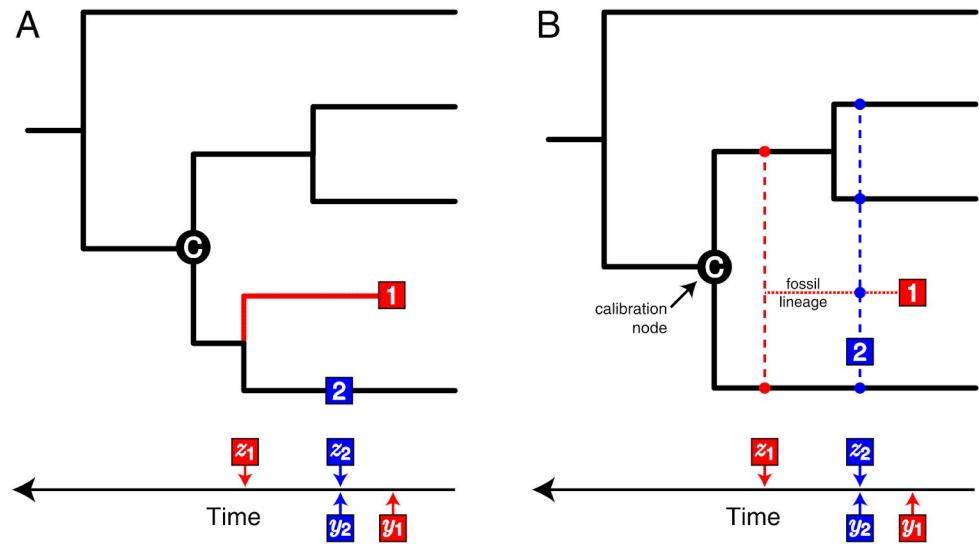


FIGURE 2. F_1 and F_2 are two fossils, T_1 and T_2 are extant taxa. F_2 is either split from the lineage of T_1 using the add-branch move (left to right) or merged to the lineage of T_1 using the delete-branch move (right to left). The two crosses on lineages of F_1 and T_2 at time y_f represent two alternative positions to which we could move fossil F_2 using a SPR move.

Fossilized Birth Death (FBD)

Benefits of FBD (cont.)

- No unintended interaction b/w age priors and cladogenesis model
- FBD can be combined with TED
 - Note: when run w/o total evidence, FBD marginalizes over all possible fossil placements, meaning some prior placement of fossils is required
- Allows for the implication of a variety of MCMC methods to improve modeling
 - E.g. add-branch moves, delete-branch moves, a node-slider mechanism, and a subtree-pruning-and-regraft (SPR) moves mechanism



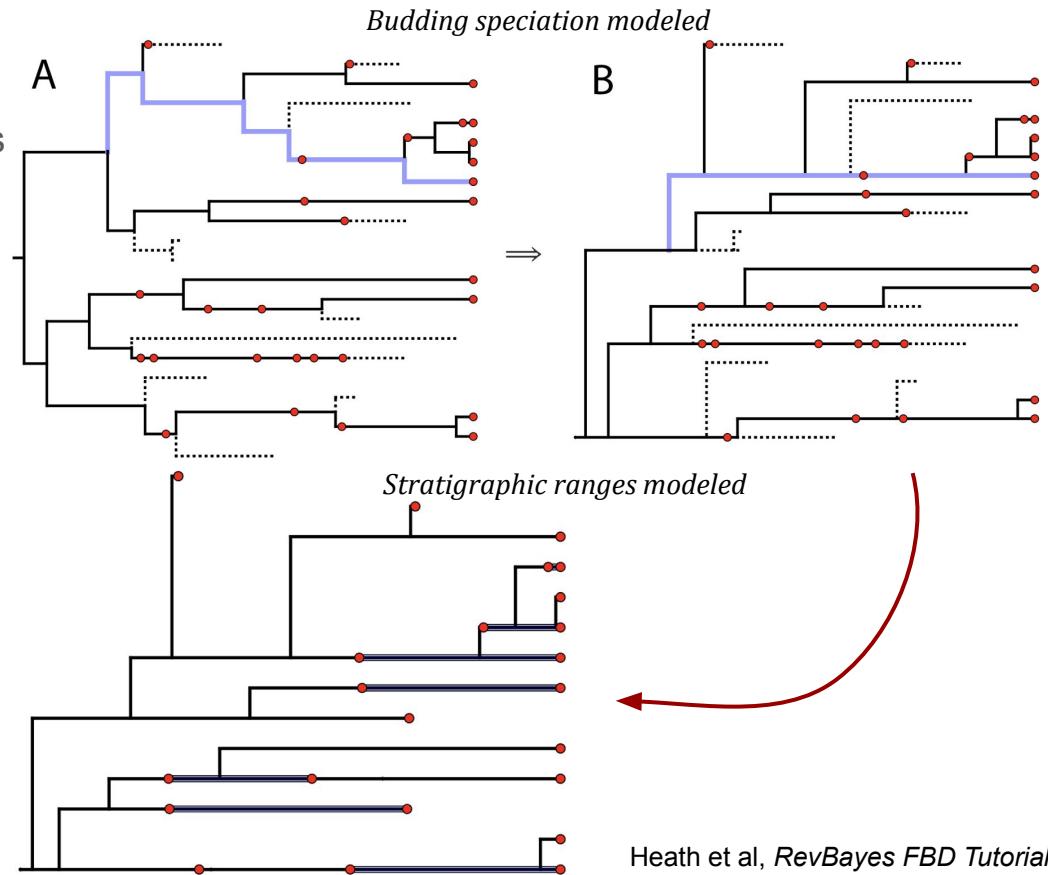
Fossilized Birth Death (FBD)

Recent developments

- Assigning stratigraphic ranges to species using a “budding” speciation model (Fossilized Birth Death Range Process)
- Modeling diversified vs random sampling

Known issues

- Model support and comparison in a Bayesian framework are a challenging
 - Bayes Factors may be problematic, but there are options like Posterior/Prior comparison and Predictive Posterior/Posterior comparison



Fossilized Birth Death (FBD)

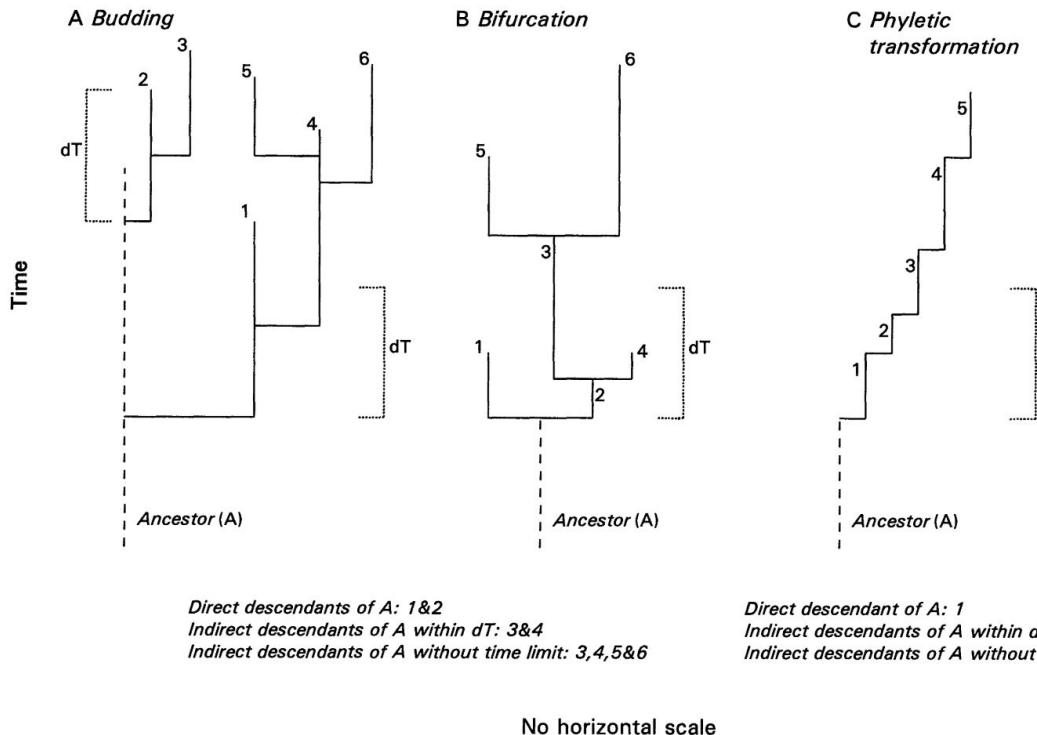


FIGURE 1. Birth-death models. A, Budding cladogenesis. A species may persist after it leaves a descendant. B, Bifurcation. A species is considered to terminate and leave two descendants at the point of branching. C, Phyletic transformation. A species either changes into a new species (pseudoextinction; transitions to species 1–5) or the lineage truly terminates (extinction; end of species 5). Horizontal lines are not meant to imply sudden morphological transitions; models in A and C should not be construed as punctuated equilibrium and punctuated anagenesis.

Random vs Diversified Sampling (Zhang et al. 2016)

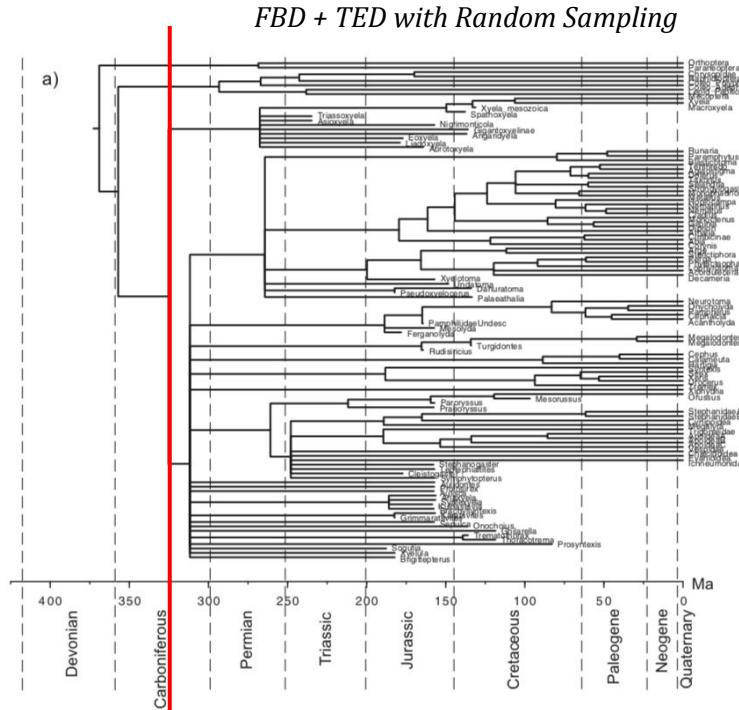


FIGURE 6. Majority-rule consensus tree, a) including all fossils and b) including only extant taxa, from total-evidence dating analysis under the piecewise-constant FBD prior with random sampling and under the IGR relaxed-clock model. Node bars indicate HPD intervals of estimated divergence times (cf. Table 5, pcFBD_Rnd).

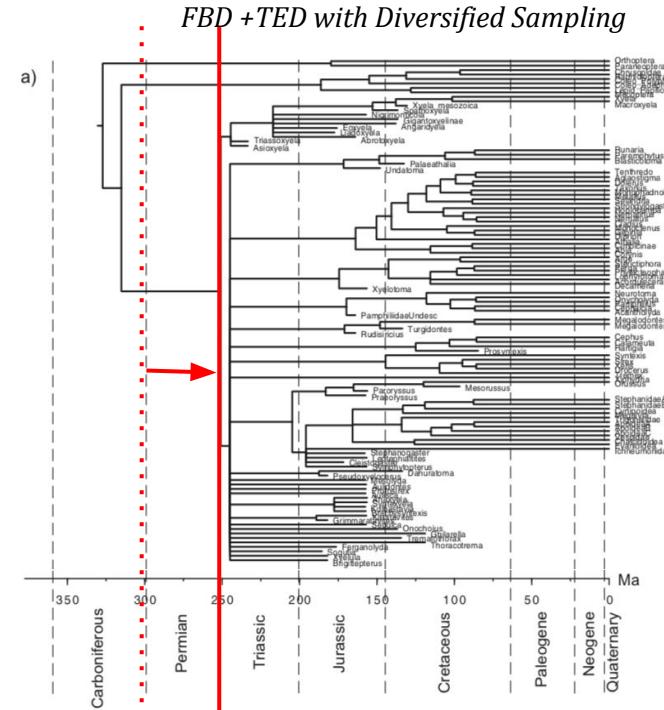


FIGURE 7. Majority-rule consensus tree, a) including all fossils and b) including only extant taxa, from total-evidence dating analysis under the piecewise-constant FBD prior with diversified sampling and under the IGR relaxed-clock model. Node bars indicate HPD intervals of estimated divergence times (cf. Table 5, pcFBD_Div).

- Most sampling for phylogenetic analysis is likely diversified rather than random
- Modeling diversified instead of random sampling yields to overall younger age estimates; this may resolve some of the “rocks vs clocks” discrepancy

Random vs Diversified Sampling (Zhang et al. 2016)

Ronquist et al 2012

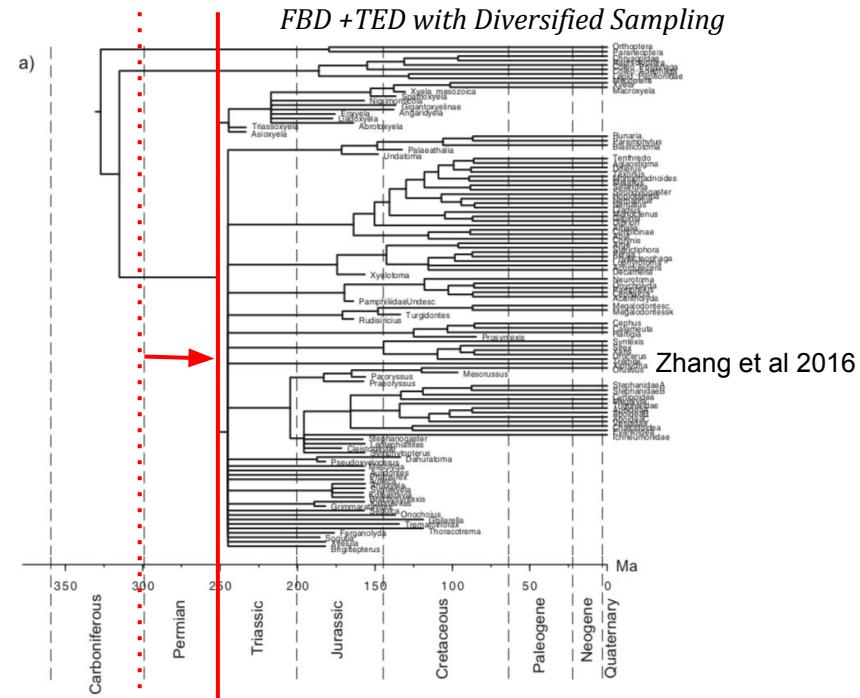
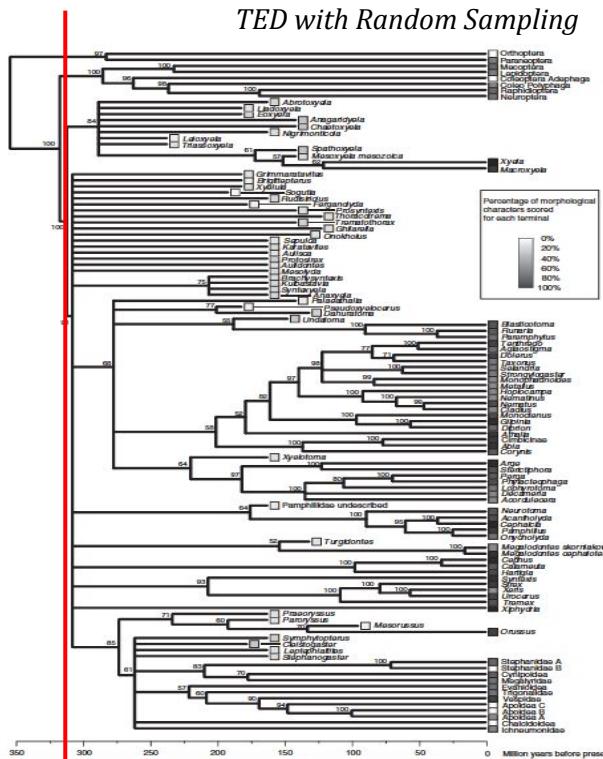


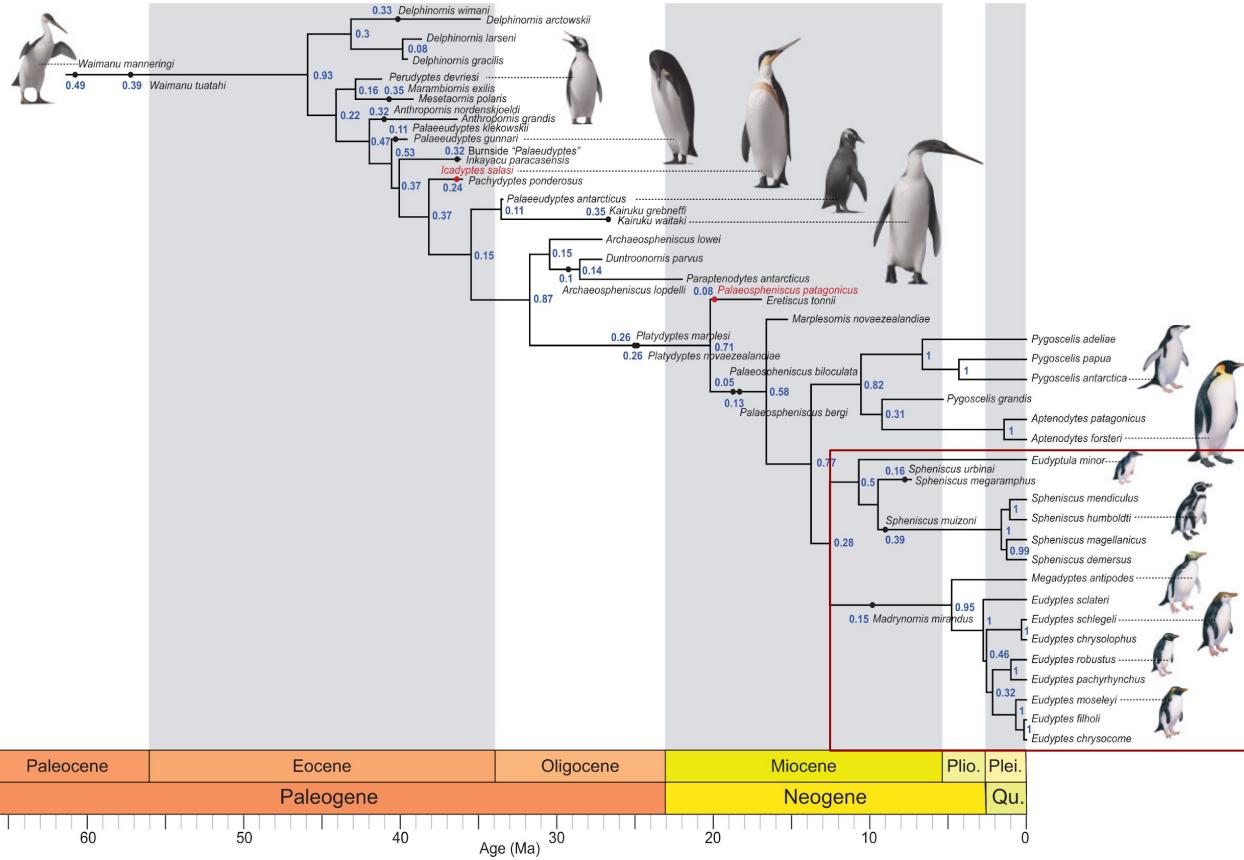
FIGURE 7. Majority-rule consensus tree, a) including all fossils and b) including only extant taxa, from total-evidence dating analysis under the piecewise-constant FBD prior with diversified sampling and under the IGR relaxed-clock model. Node bars indicate HPD intervals of estimated divergence times (cf. Table 5, pcFBD_Div).

- The younger ages given by FBD + TED with diversified sampling may also point to a partial resolution of the overall issue of TED analyses yielding older dates
- This can be seen when comparing the Ronquist et al and Zhang et al *Hymenoptera* phylogenies

Bayesian Total Evidence Dating Reveals the Recent Crown Radiation in Penguins

Gavryushkina et al. 2017

FBD estimated ~12.7 Ma



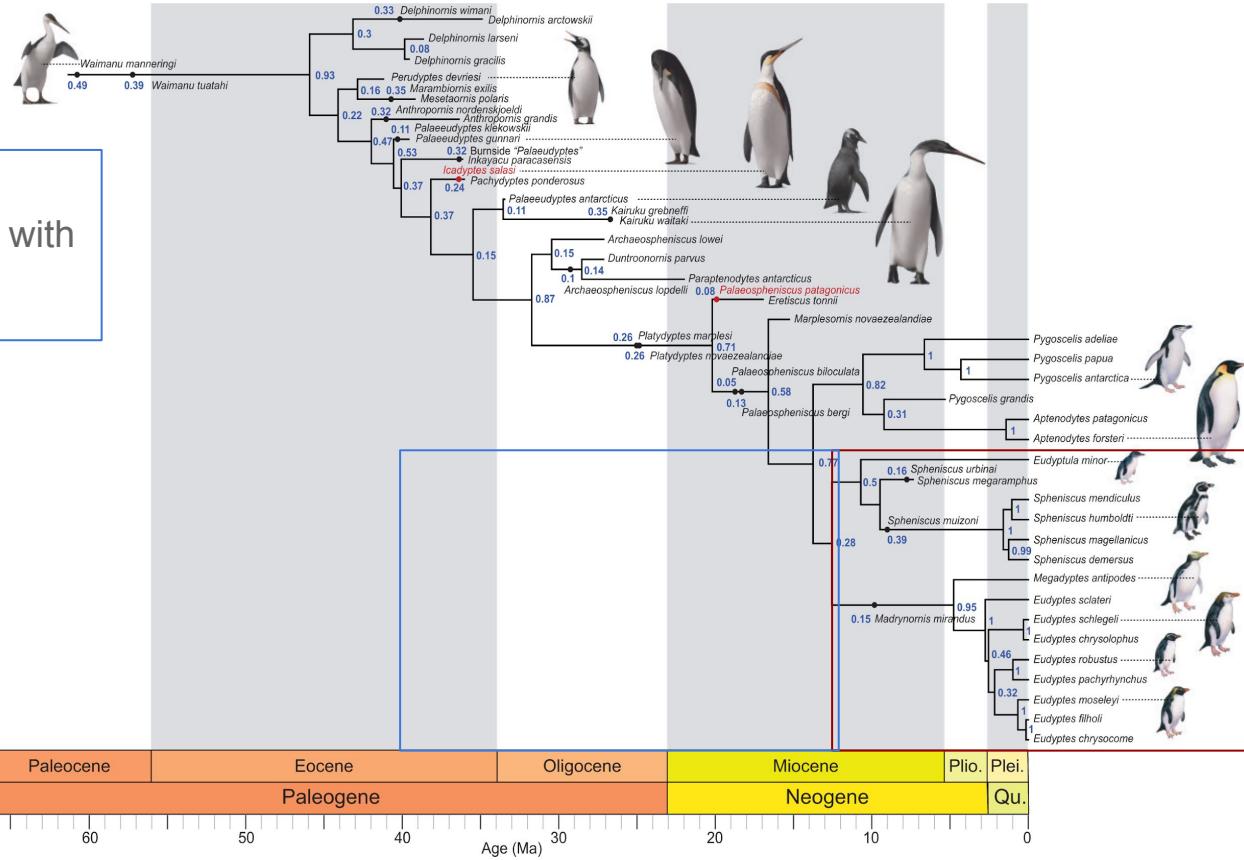
Bayesian Total Evidence Dating Reveals the Recent Crown Radiation in Penguins

Gavryushkina et al. 2017

FBD estimated ~12.7 Ma

Previous Results

- **~40.5 Ma** (Baker et al. 2006)
 - penalized likelihood approach with secondary node calibrations



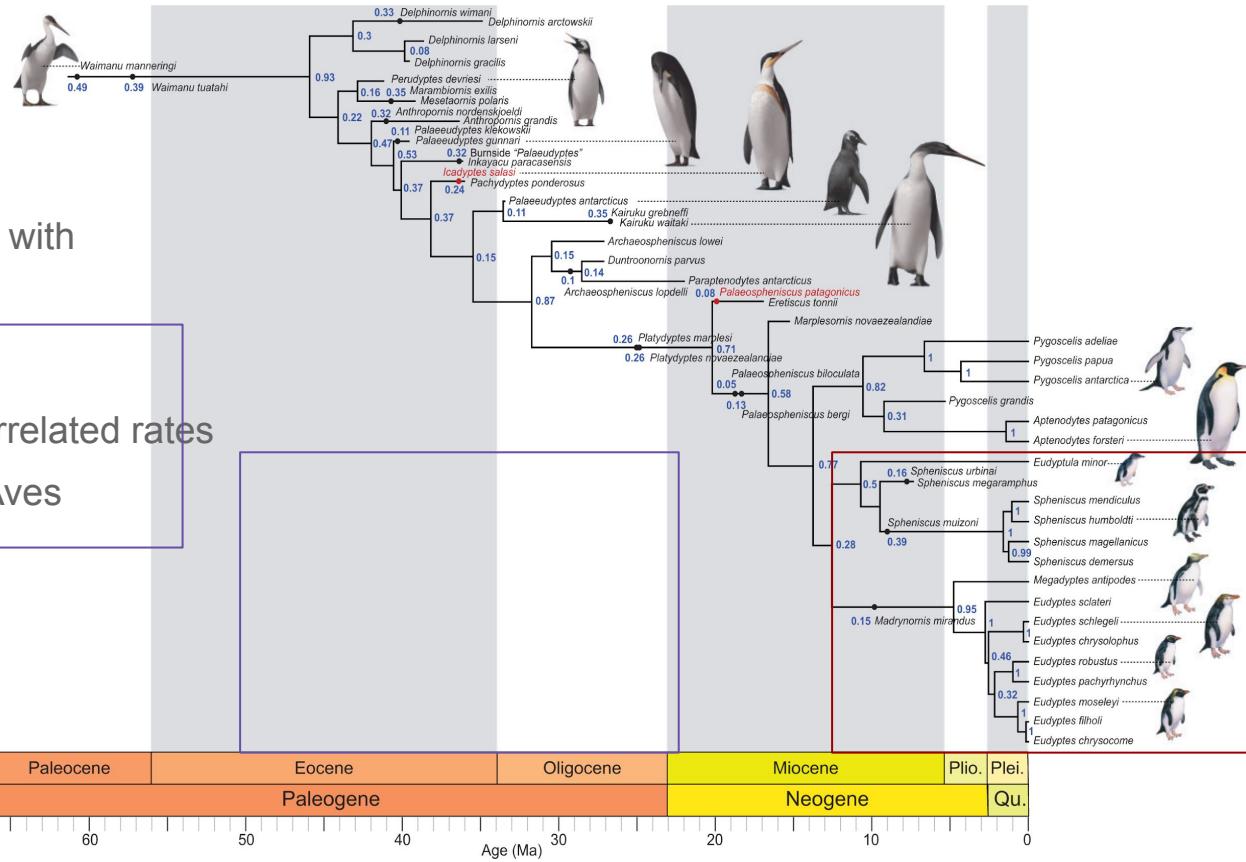
Bayesian Total Evidence Dating Reveals the Recent Crown Radiation in Penguins

Gavryushkina et al. 2017

FBD estimated ~12.7 Ma

Previous Results

- ~40.5 Ma (Baker et al. 2006)
 - penalized likelihood approach with secondary node calibrations
- ~50 Ma. (Brown et al. 2008)
 - Bayesian approach with uncorrelated rates & 20 calibrations throughout Aves



Bayesian Total Evidence Dating Reveals the Recent Crown Radiation in Penguins

Gavryushkina et al. 2017

FBD estimated ~12.7 Ma

Previous Results

- ~40.5 Ma (Baker et al. 2006)
 - penalized likelihood approach with secondary node calibrations
 - ~50 Ma. (Brown et al. 2008)
 - Bayesian approach with uncorrelated rates & 20 calibrations throughout Aves
 - ~20 Ma (Subramanian et al. 2013)
 - Bayesian analysis with node calibration densities using 4 fossil penguin taxa

