

Santiago Claramunt

Department of Ecology and Evolutionary Biology



Calibrating molecular clocks using the fossil record and CladeDate



2023 SSB
5TH STANDALONE MEETING
UNAM - MEXICO CITY

Introduce Yourself

- Who are you?
- Experience or interest in divergence time estimation?

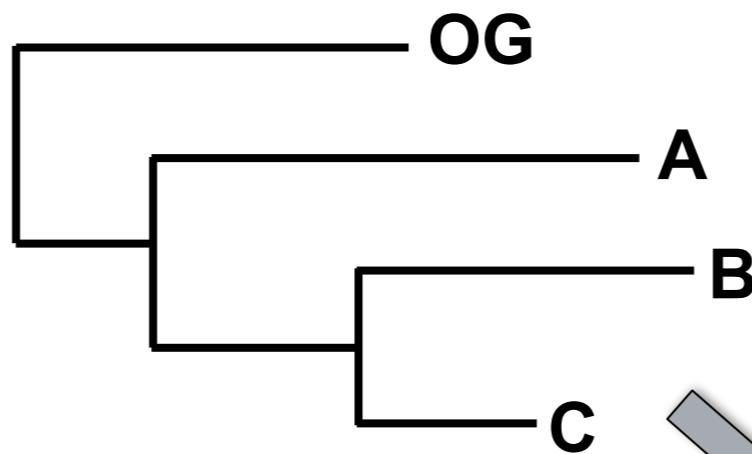
Calibrating molecular clocks using the fossil record and CladeDate

- Time-Tree Estimation Methods
- Clock Calibration Methods
- Choosing Calibration Fossils
- Clade Age Estimation Methods
- The CladeDate Algorithm
- Validation and Comparisons
- Example: a time-tree of modern birds

Basic Sequential Method

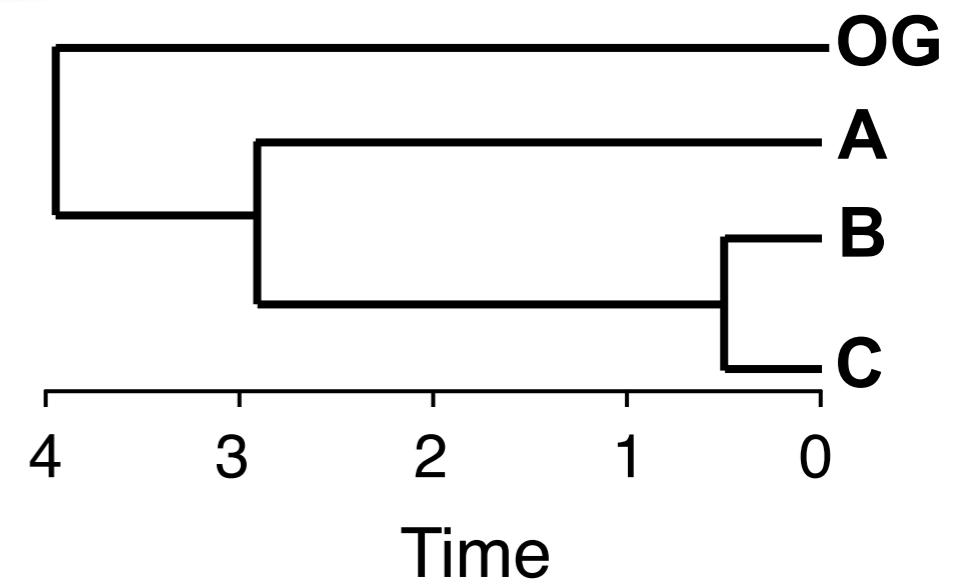
Outgroup **AAGCTTCATA**
Species A **GAGCTTCACA**
Species B **GTGCTTCACG**
Species C **GTGCCTCACG**

Phylogenetic analysis



- Methods:
- averaging
 - likelihood

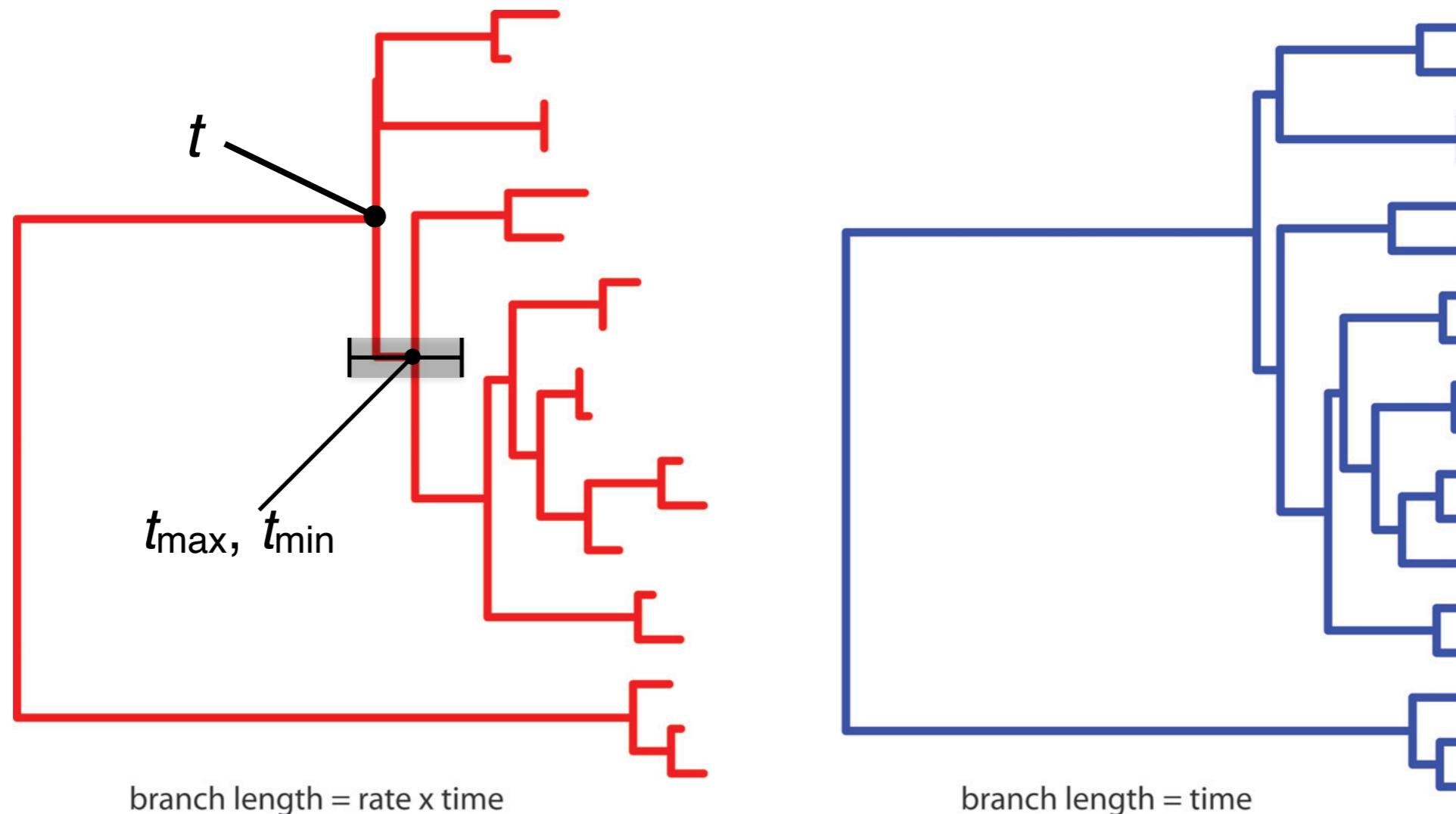
Time-scaling



Basic Sequential Method

Averaging and (penalized) likelihood methods

- Constraining the age of one or multiple nodes provides the calibration information
- Both fixed ages and minimum and maximum bounds are allowed



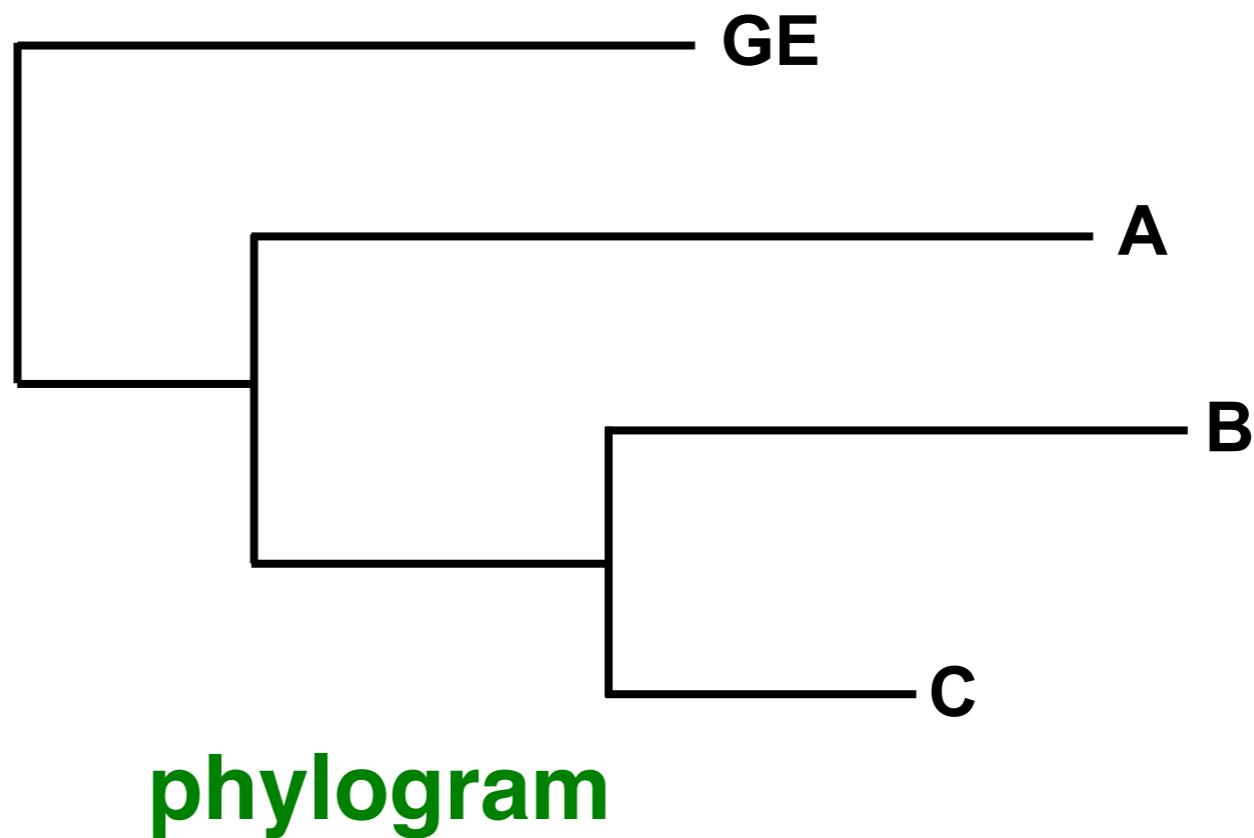
Maximum Likelihood Phylogenetic Inference

Outgroup	AAGCTTCATA
Species A	GAGCTTCACA
Species B	GTGCTTCACG
Species C	GTGCCTCACG

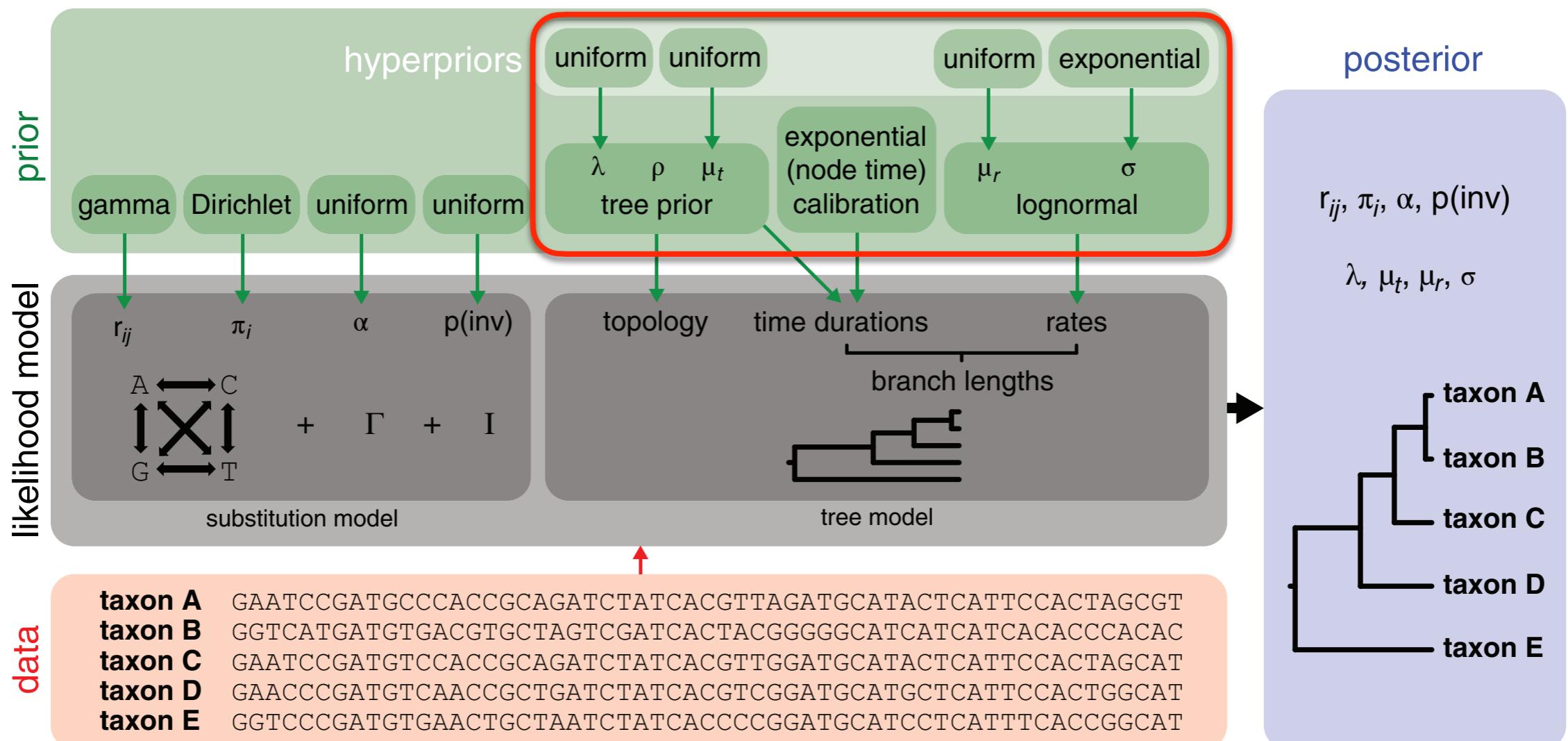


Probabilistic model of
molecular evolution

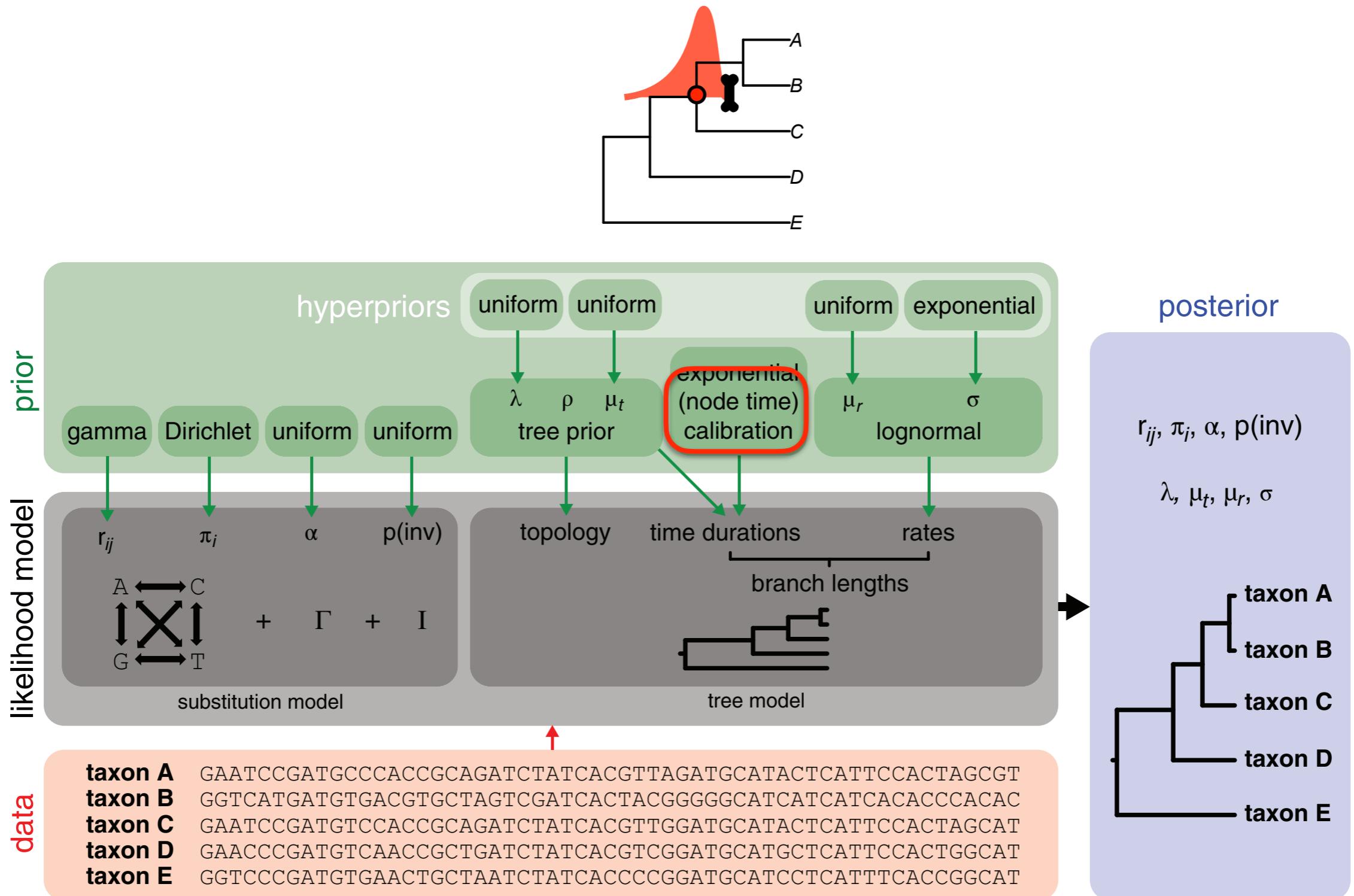
Maximum Likelihood
optimization
 $P(d \mid H_1)$



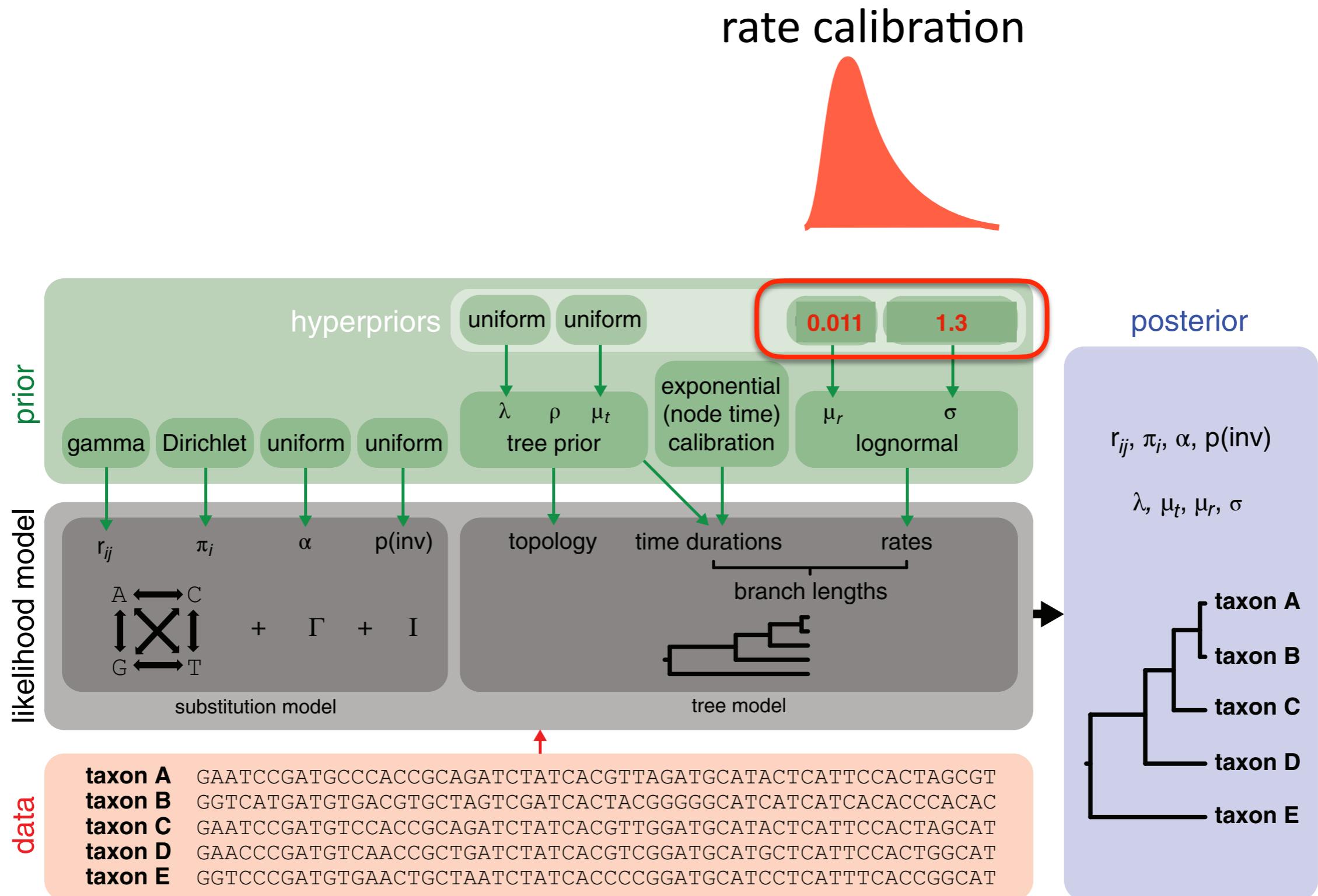
Bayesian Time Tree Estimation



node calibration

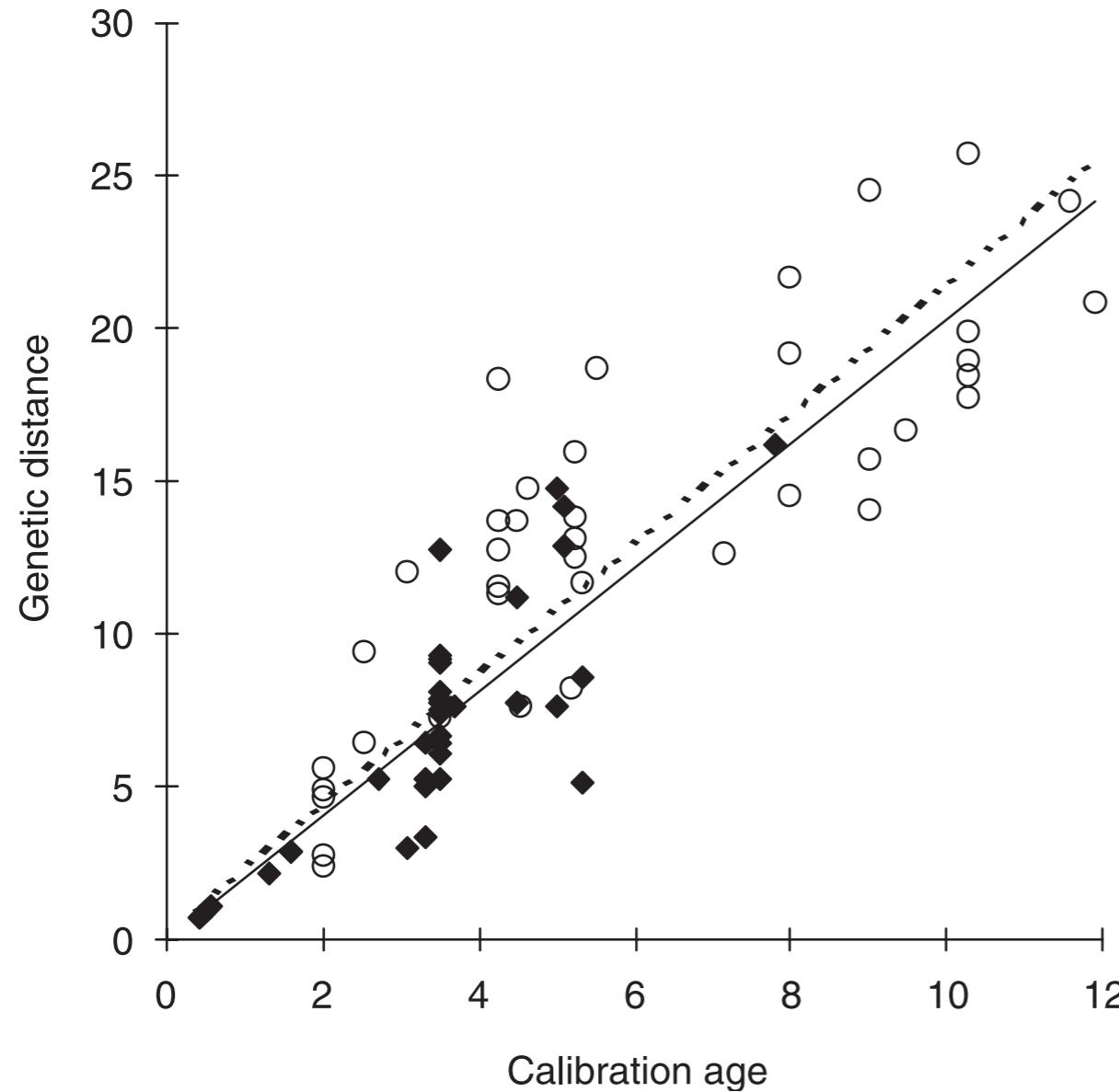


Molecular-clock Rate Priors

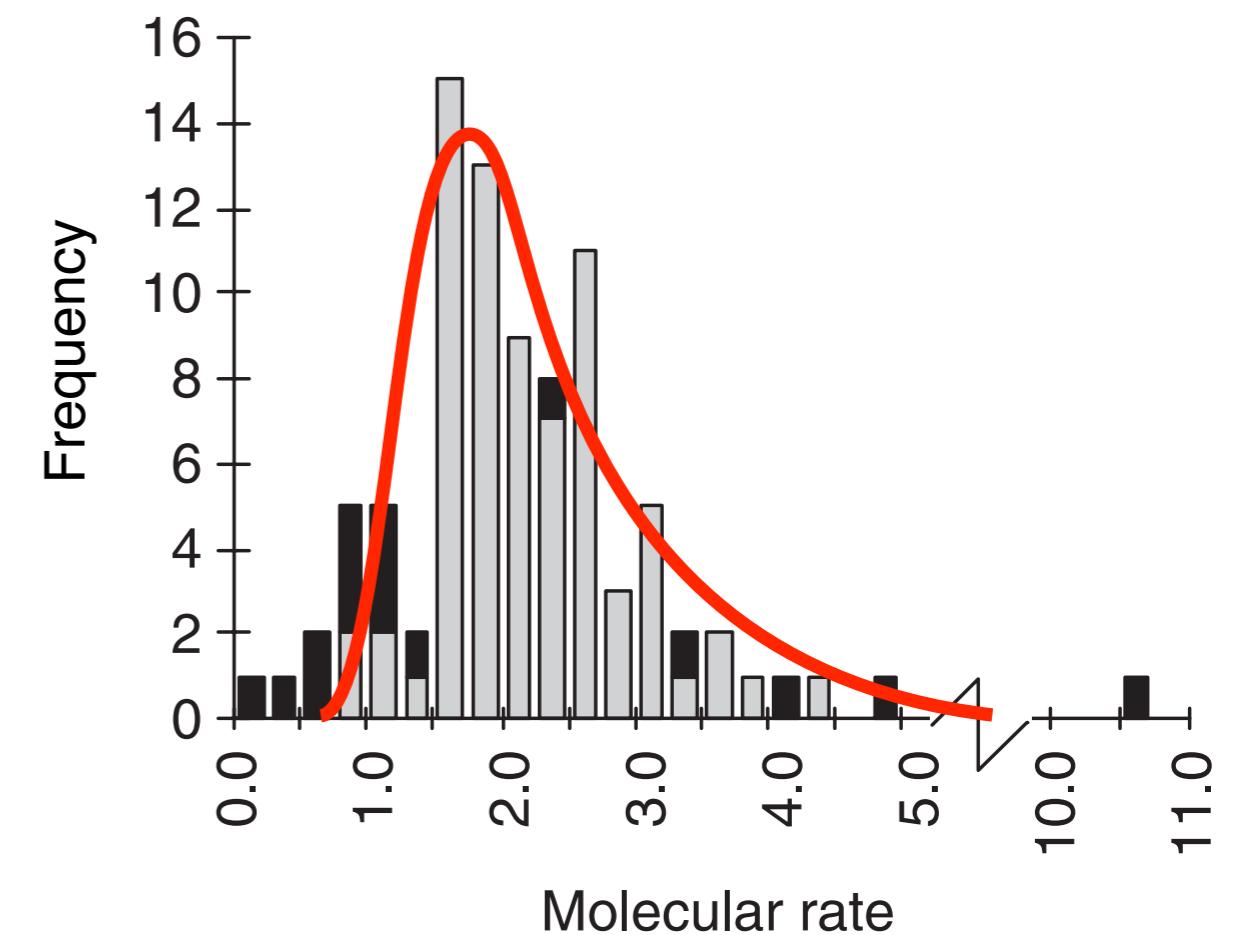


Molecular-clock Rate Priors

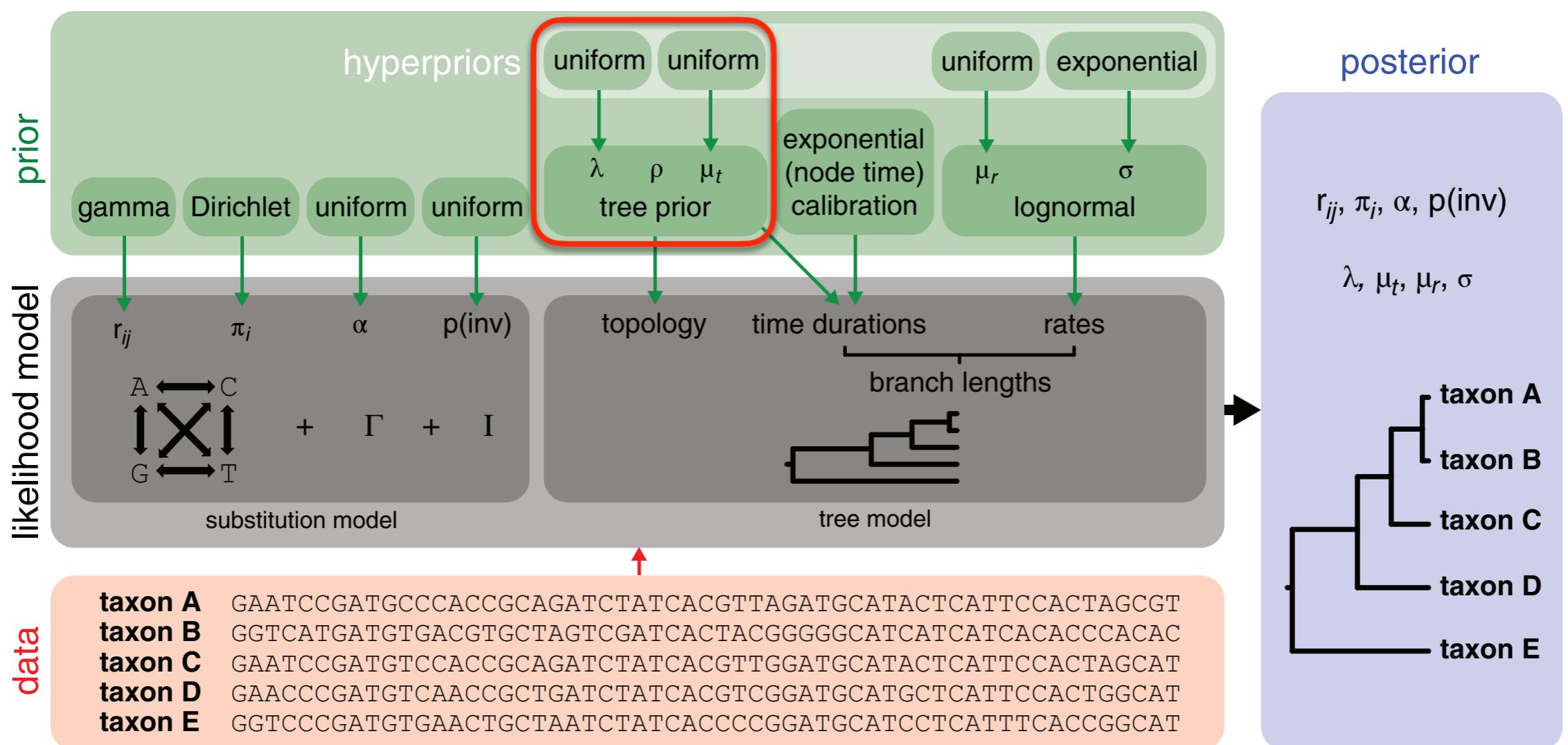
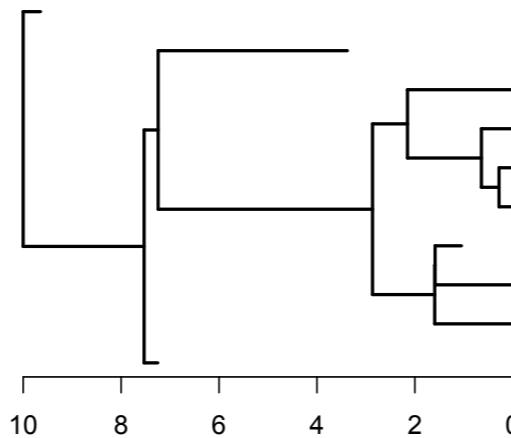
- Empirical data on mitochondrial divergence in birds



Lognormal(mean=0.011, s.d.=1.3)



Branching-process Priors

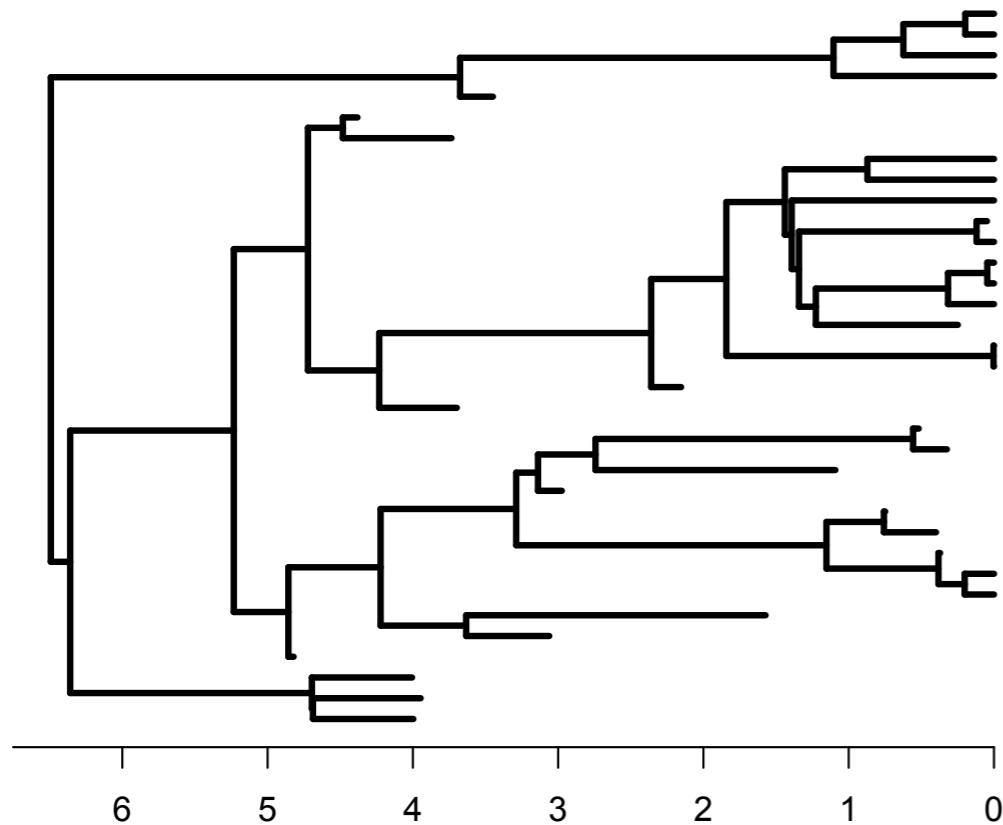


Branching-process Priors

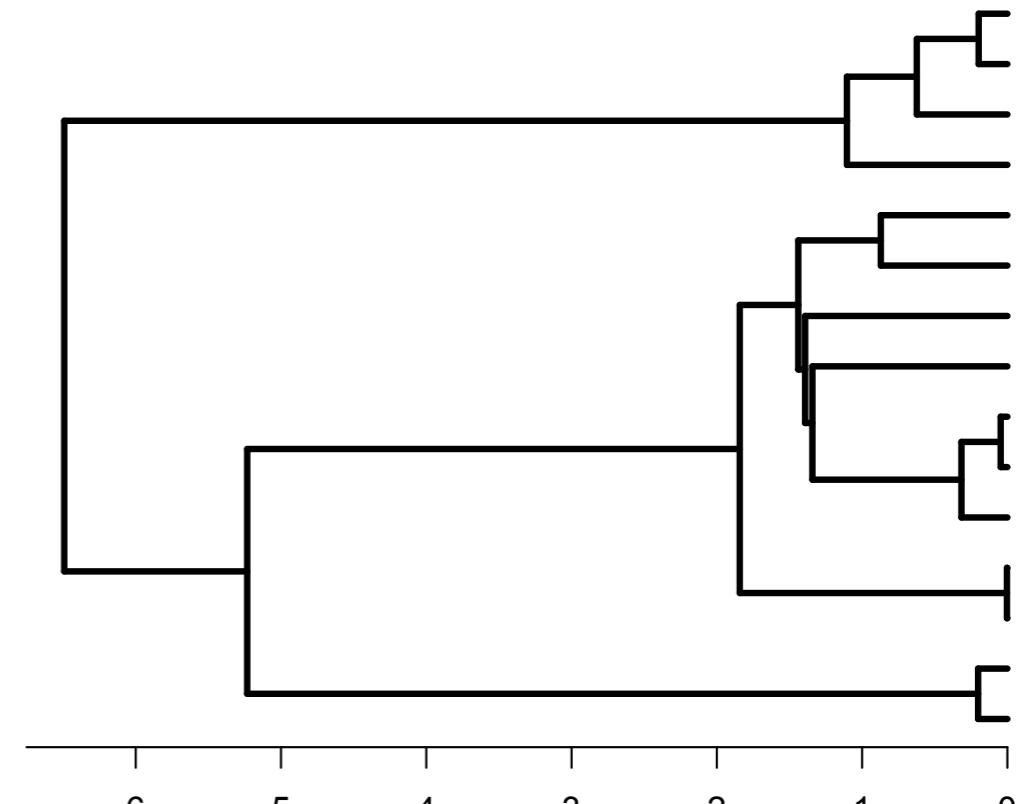
Birth-death Models (contemporaneous terminals)

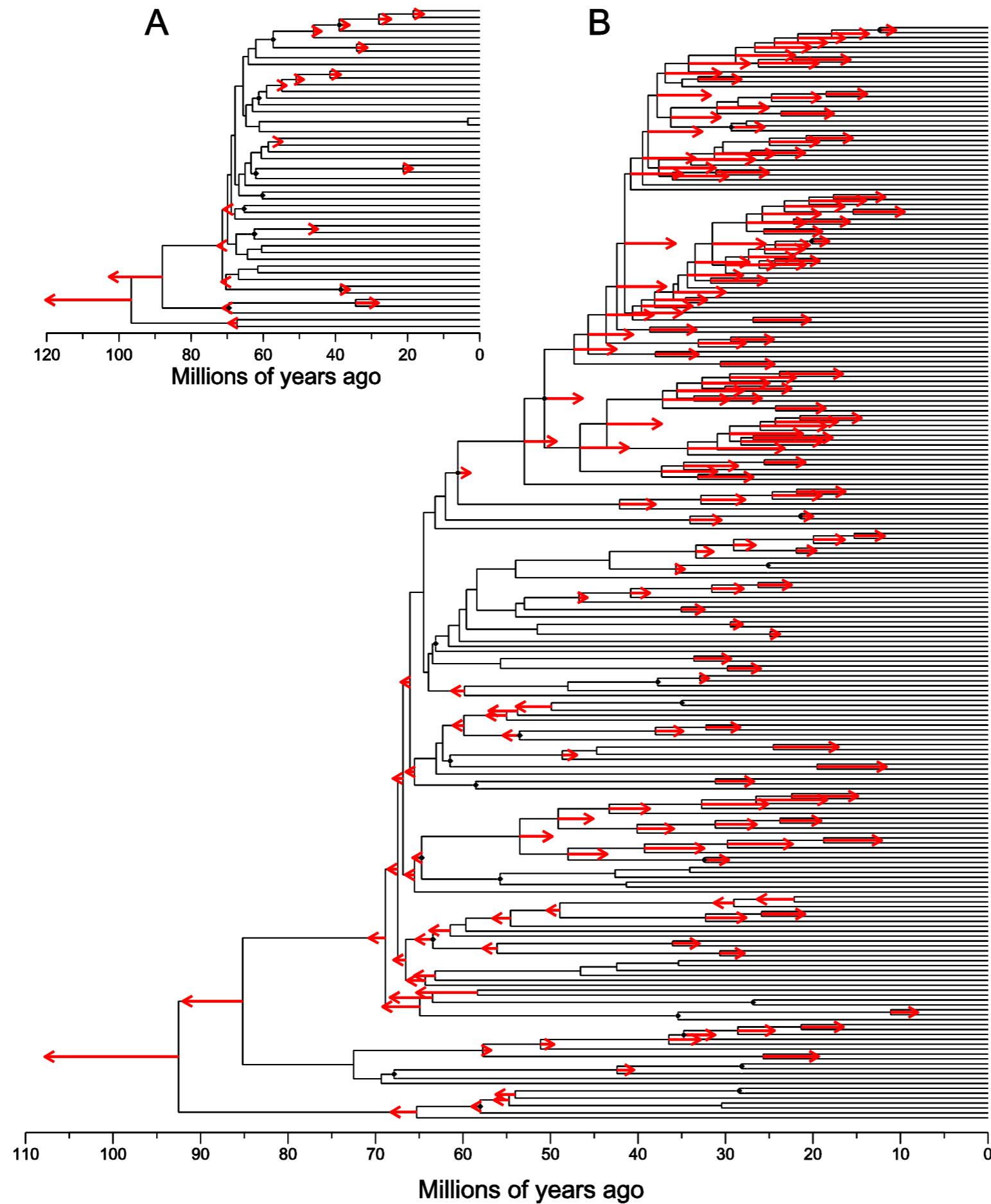
- Constant, per-lineage, rate of speciation and extinction
- More short branches expected towards the tip of the tree.
- Priors on speciation and extinction rates

With extinct lineages



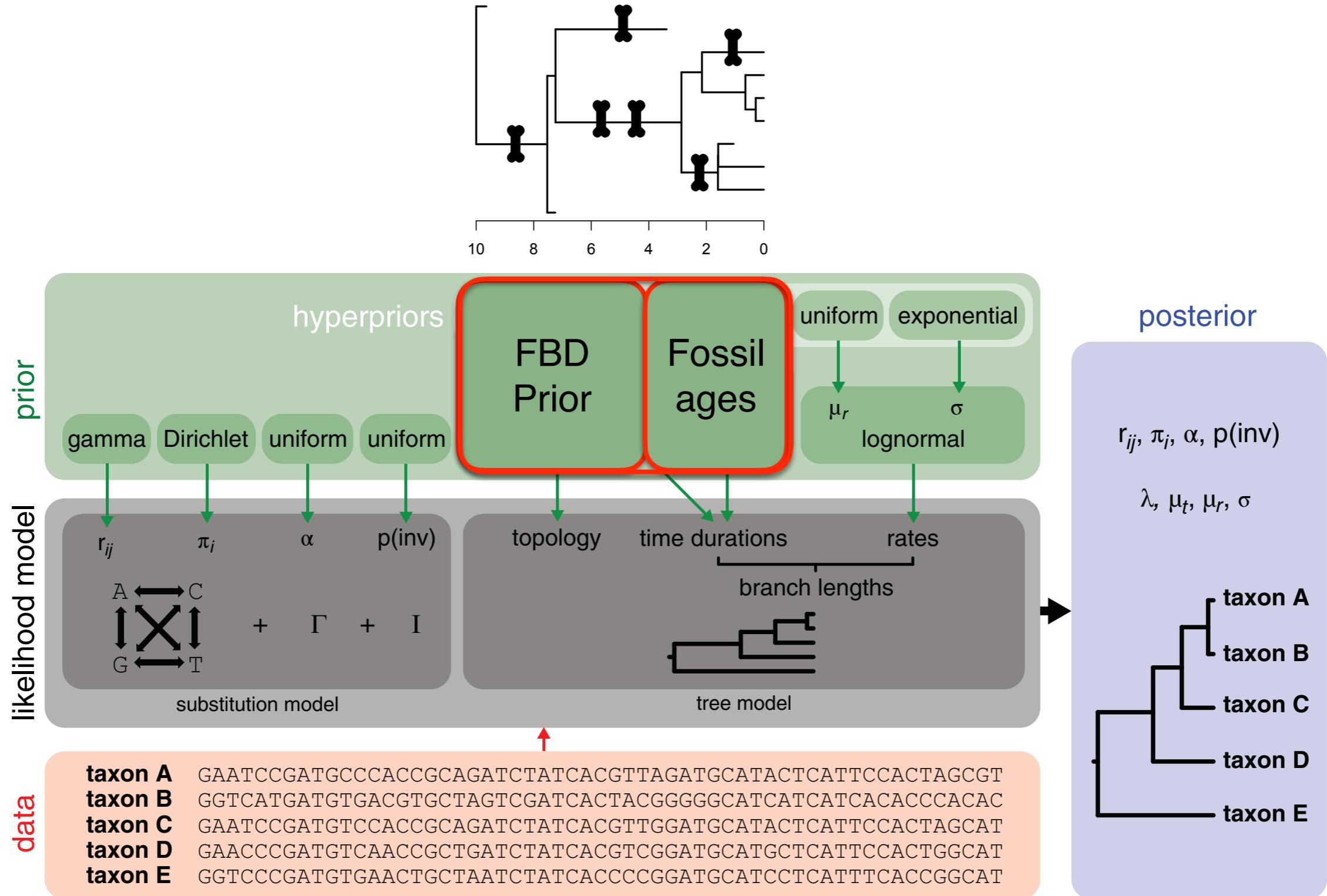
Without extinct lineages





Fossilized Birth-death Process

- Integrates fossils into the diversification process prior by modelling fossil sampling
- Uses all fossils instead of just the oldest per clade



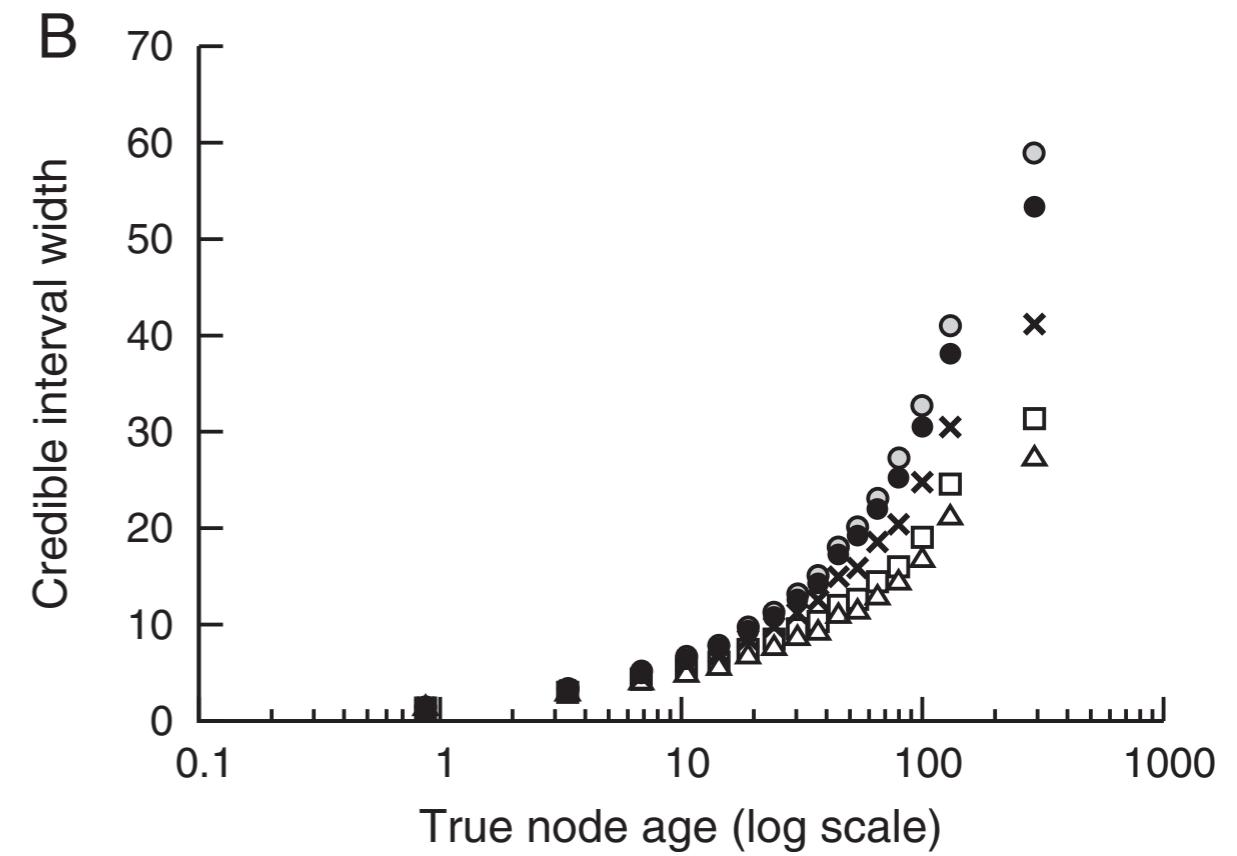
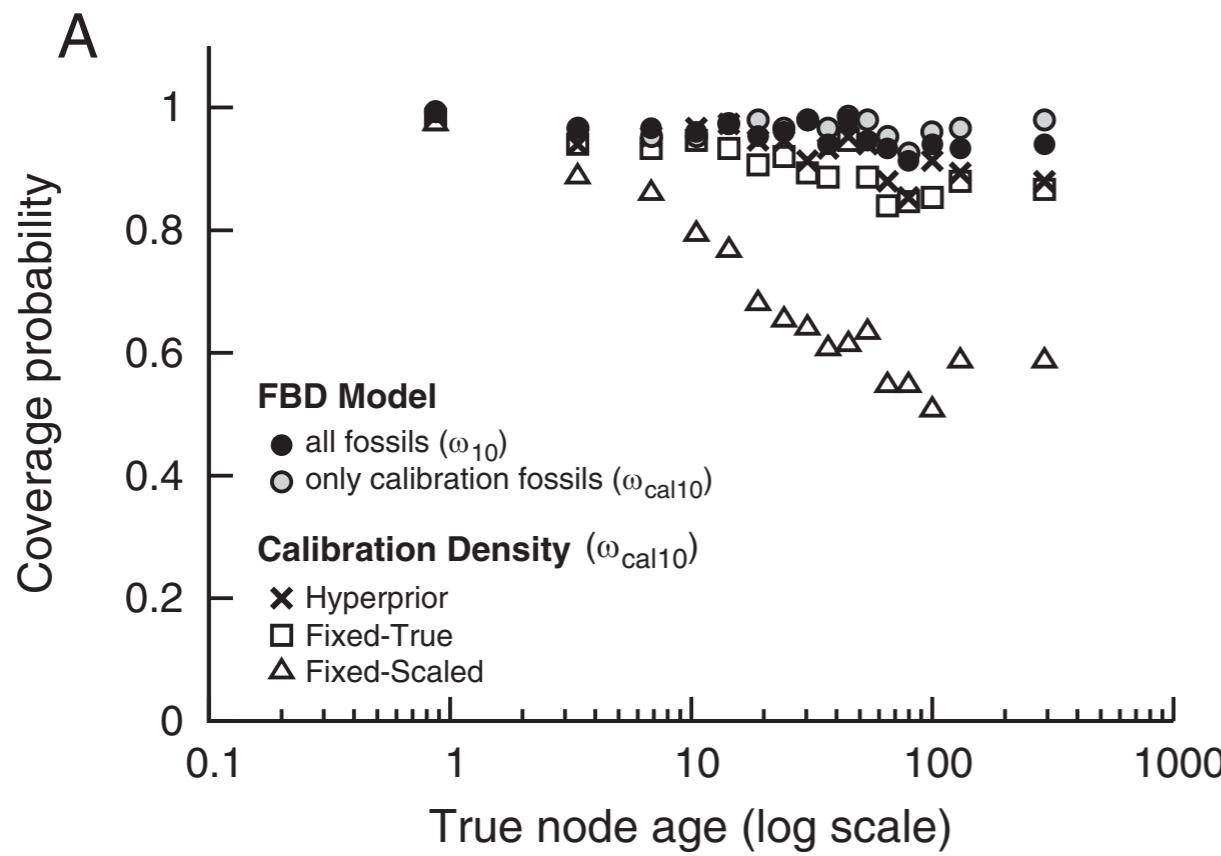


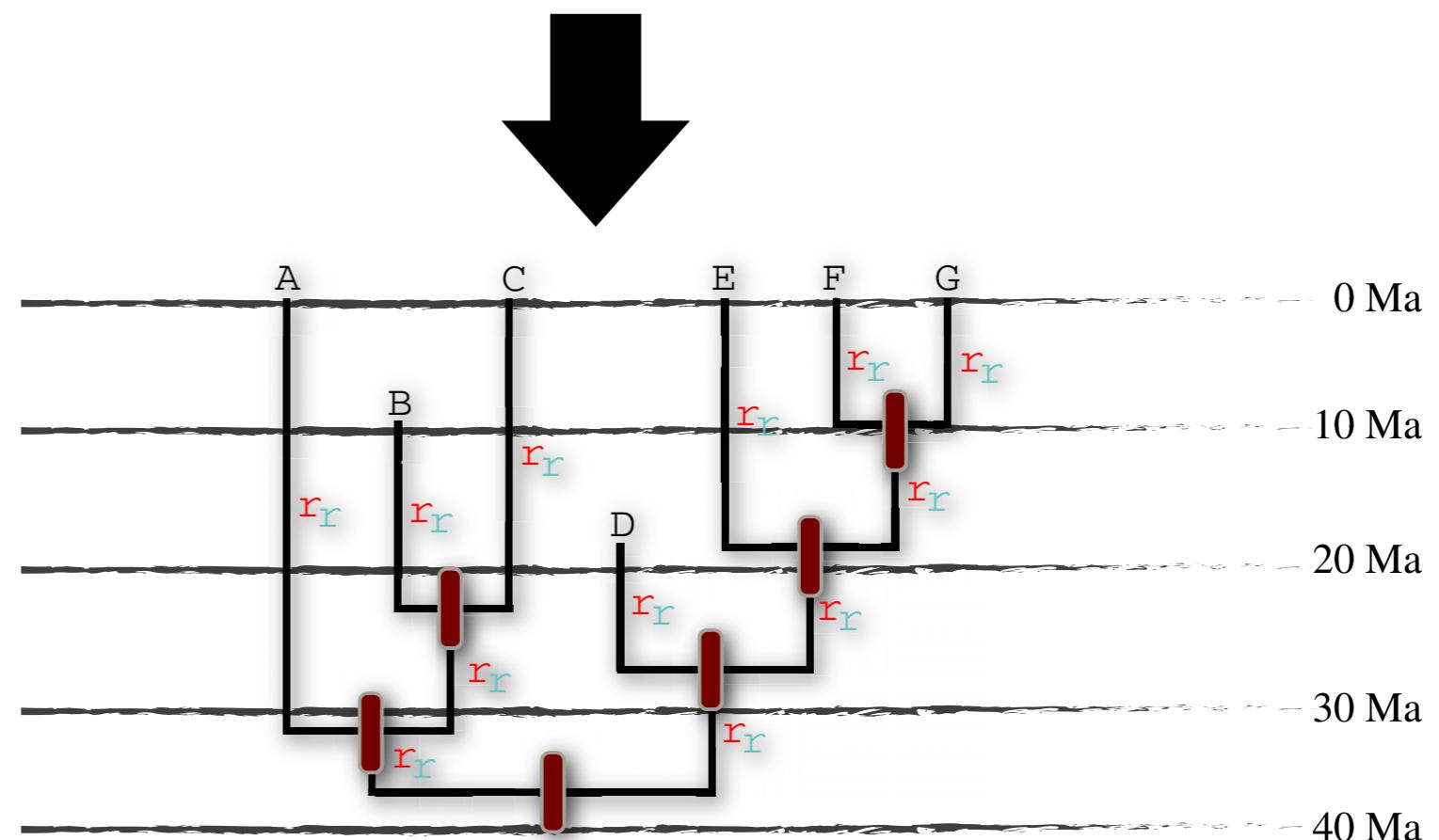
Fig. 2. The results for 100 replicate trees simulated under the FBD model with $\mu/\lambda = 0.5$, $\lambda - \mu = 0.01$. Node age estimates are summarized for analyses under the FBD model by using all available fossils sampled randomly (10%; ω_{10}) from the total number of simulated fossils (●). These results are compared with divergence time estimates on the set of calibration fossils (ω_{cal10}) under the FBD model (○), with a hyperprior on calibration density parameters (×), with a fixed calibration density where the expected value is equal to the true node age (□), and for a fixed calibration density scaled based on the age of the fossil (△). Both the coverage probability and precision (95% CI width) are shown as a function of the true node age (log scale), where the nodes were binned so that each bin contained 150 nodes and the statistics were computed within each bin. (A) The coverage probability is the proportion of nodes in which the true value falls within the 95% CI. (B) The average size of the 95% CIs for each bin was computed to evaluate precision.

Total Evidence Tip Dating

- Dataset combines molecular data for extant species and morphological data for both extant and fossils species
- Molecular and morphological models are used to estimate branch lengths.

(a)

species A	CTTAGATGCTACCAA--AATA	1110-00110
species B	?????????????????????????	1110-00110
species C	CTTAGATGCTACCAAA-TAC-	1321-00200
species D	?????????????????????????	1301-00001
species E	CTTAGATGCTTCCAA--ACAT	1320-10200
species F	ATTAGATGCTACCAATATAAG	1333010210
species G	ATTAGATGCTACCAA--ATAT	1332000100

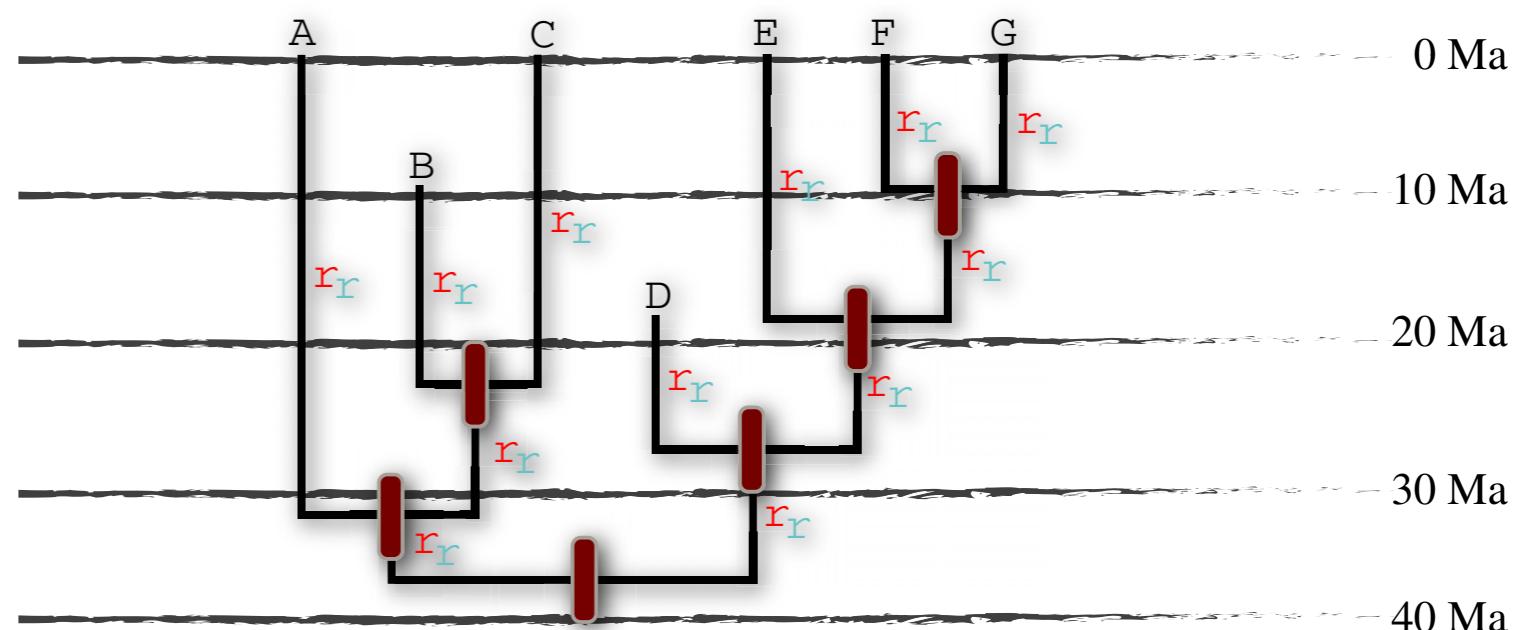


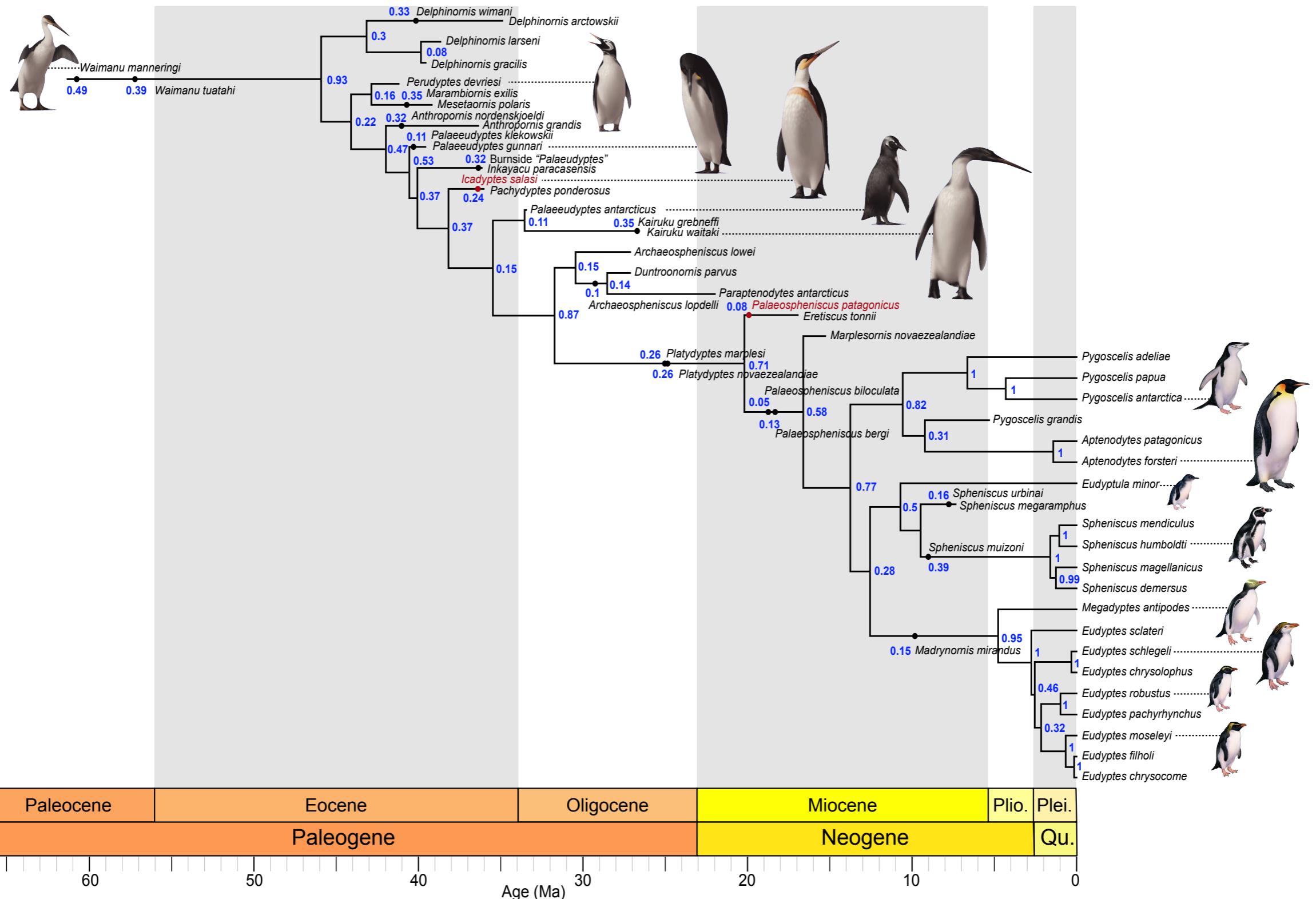
Fossilized Birth-death Total Evidence Dating

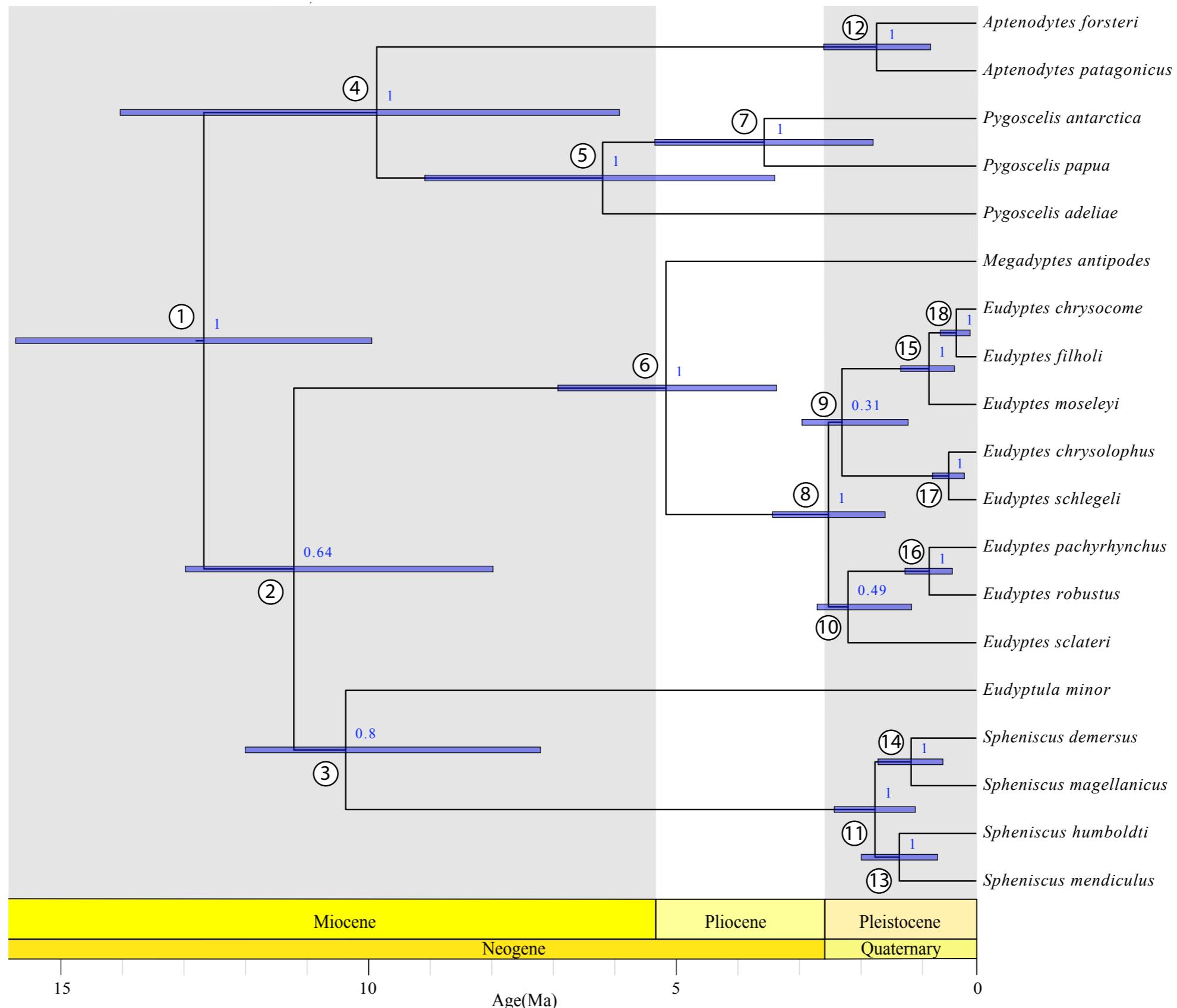
- Integrates the fossilized birth-death process with total-evidence tip dating
- Fossils are terminals in the analysis and also part of the diversification-fossilization-recovery process

(a)

species A	CTTAGATGCTACCAA--AATA	1110-00110
species B	???????????????????????	1110-00110
species C	CTTAGATGCTACCAAA-TAC-	1321-00200
species D	???????????????????????	1301-00001
species E	CTTAGATGCTTCCAA--ACAT	1320-10200
species F	ATTAGATGCTACCAATATAAG	1333010210
species G	ATTAGATGCTACCAA--ATAT	1332000100







But Node-Based Calibrations Still Critical

Can be used in simpler sequential methods for large-scale genomic datasets

- Fast results
- Too many species
- Genomic dataset

Fewer assumptions about processes governing fossils and lineages

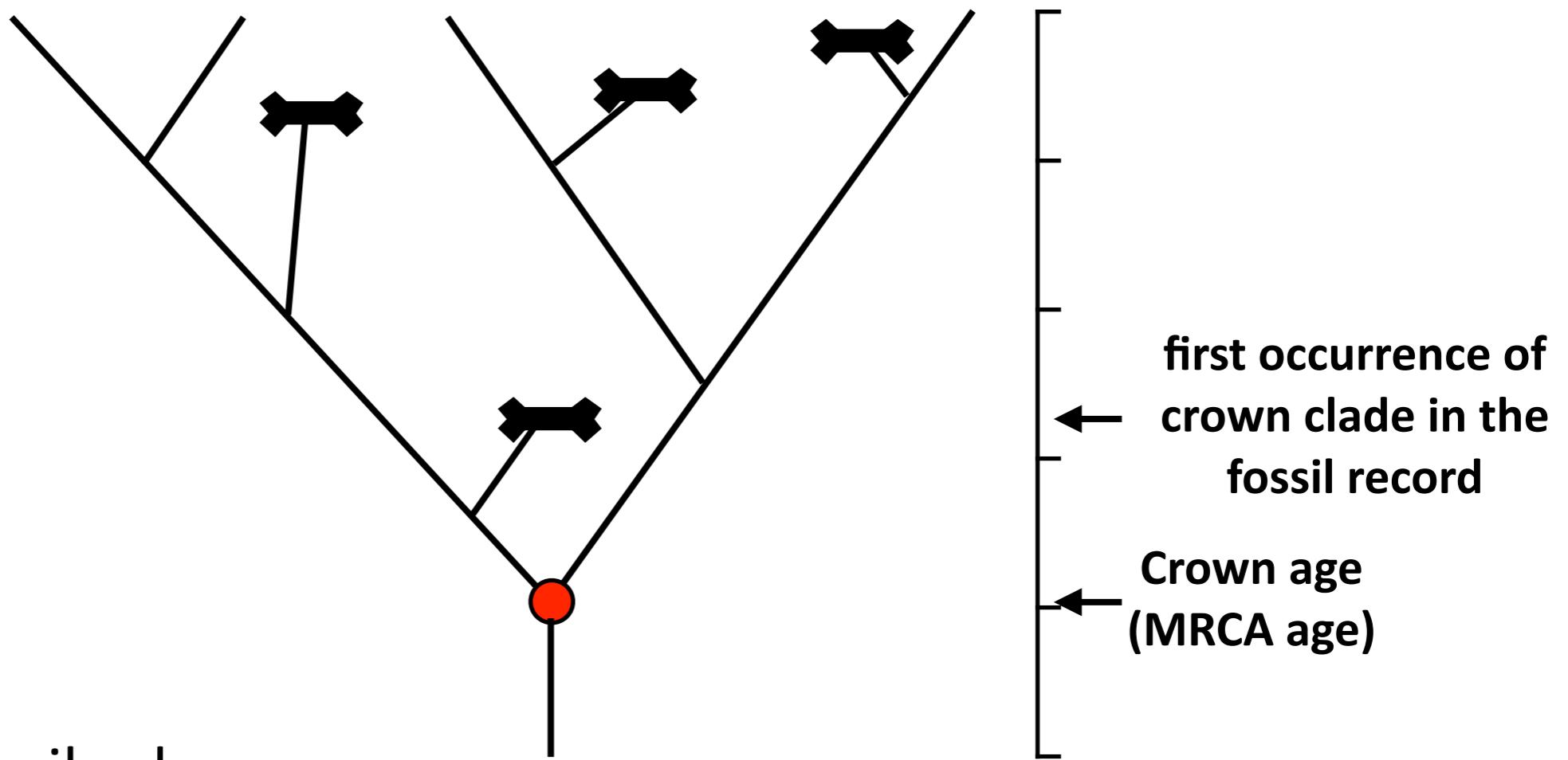
- Diversification rates
- Evolutionary rates
- Fossils sampling

Total evidence tip-dating using fossilized birth-death models still needs a prior for the age of the entire tree

- An informative tree age prior may improve the estimation

Choosing Calibration Fossils

- Formally described
- Composed of multiple bones
- Affinities established by phylogenetic analysis
- Low age uncertainty



Conflictus antarcticus

65.7 Ma

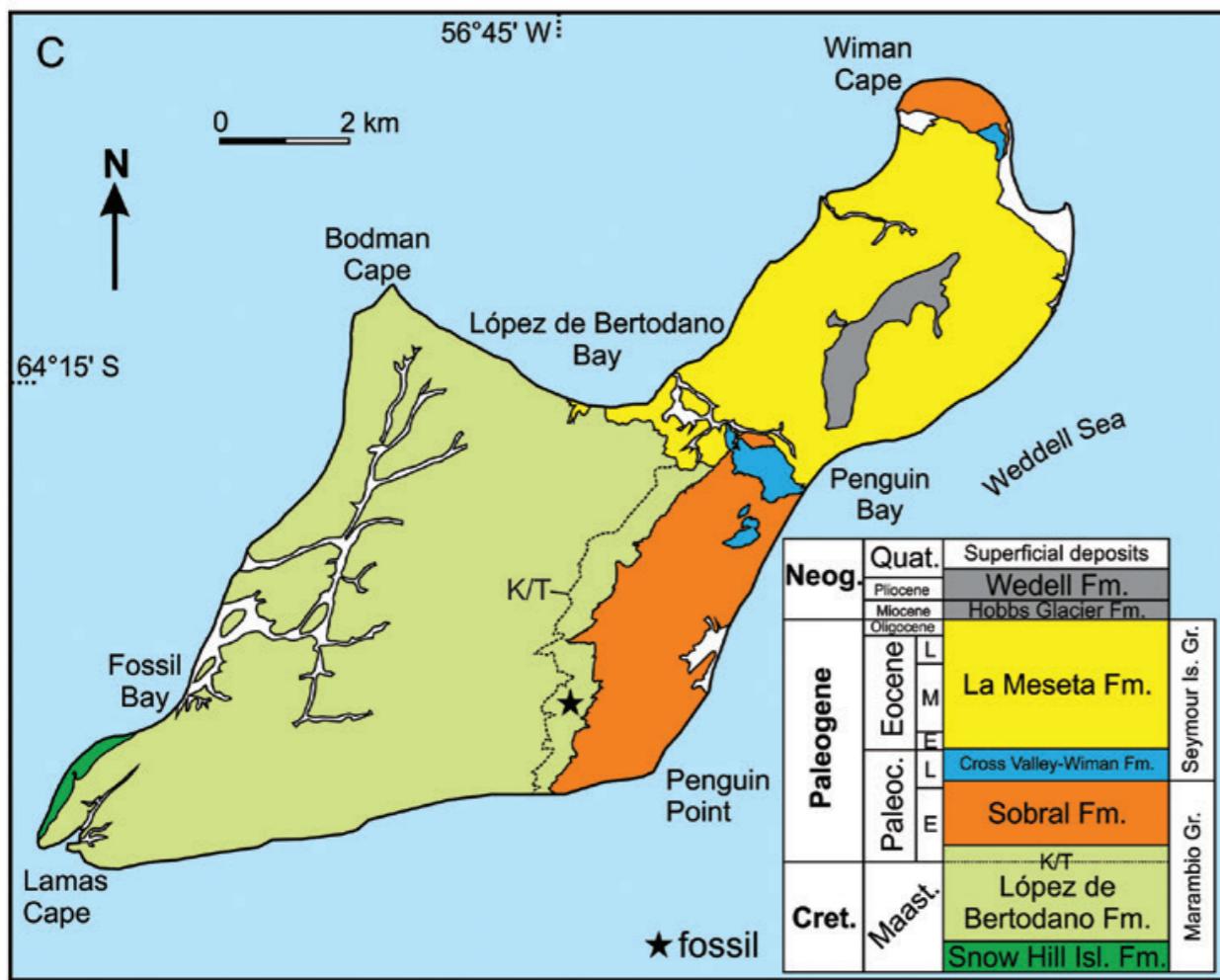
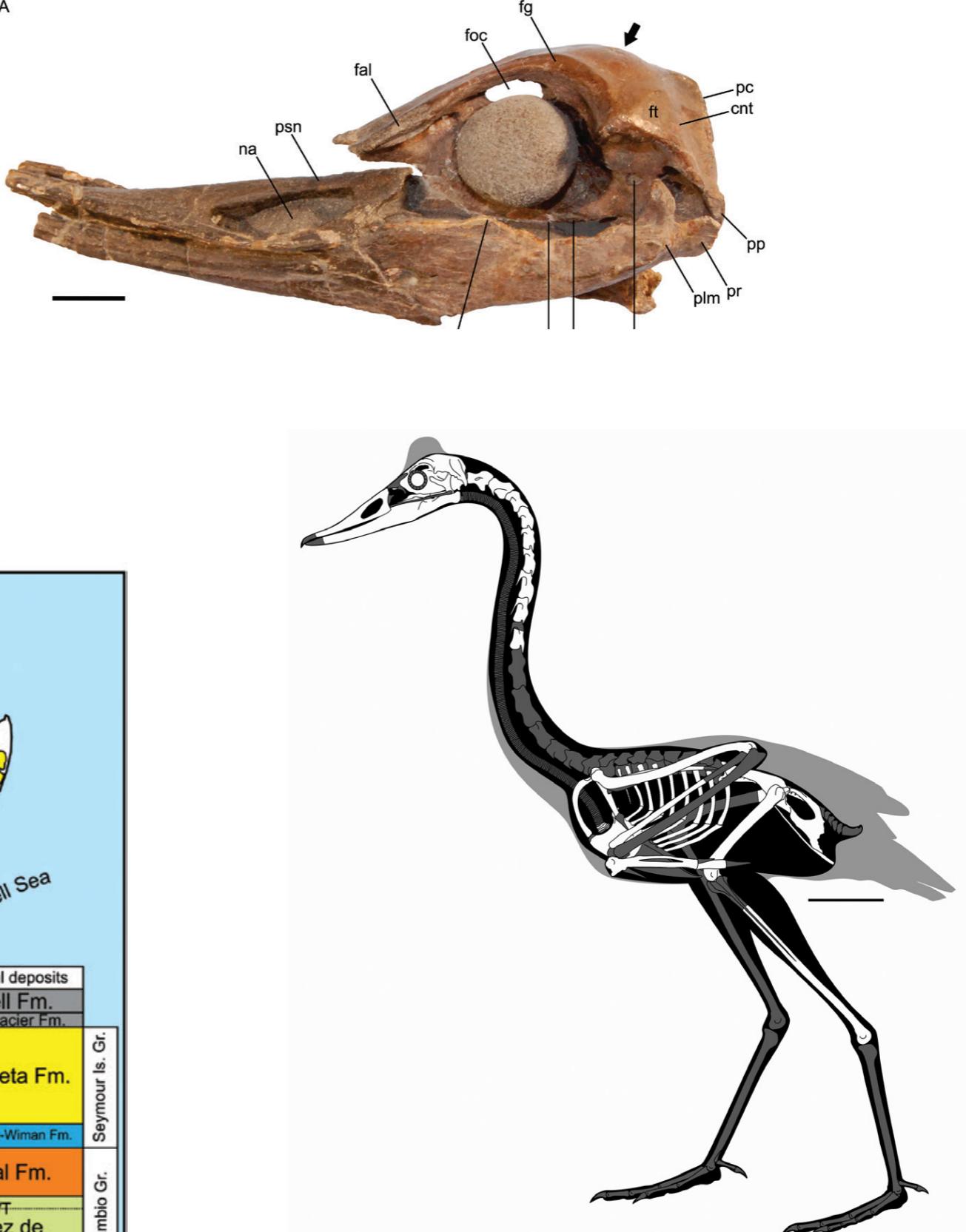
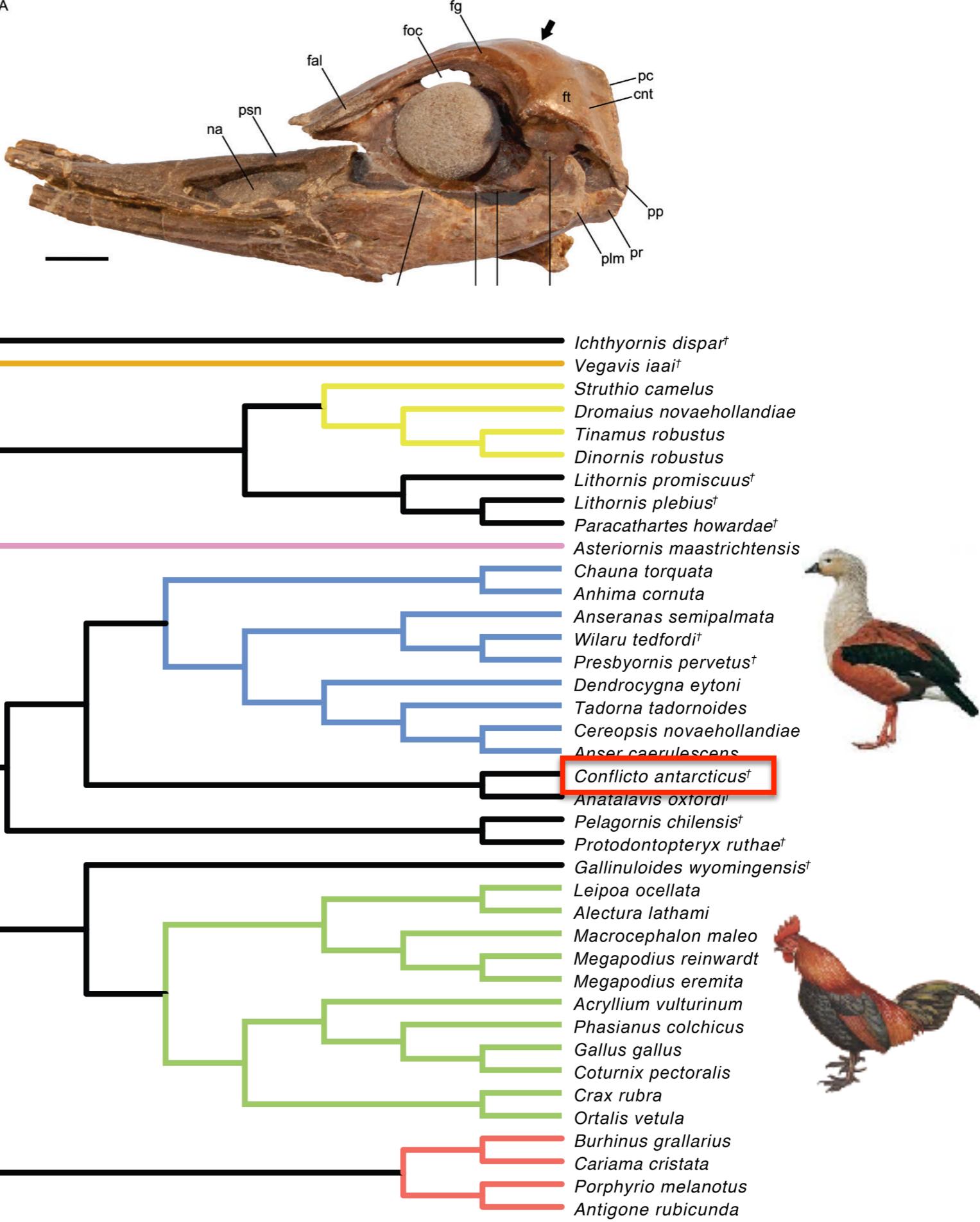
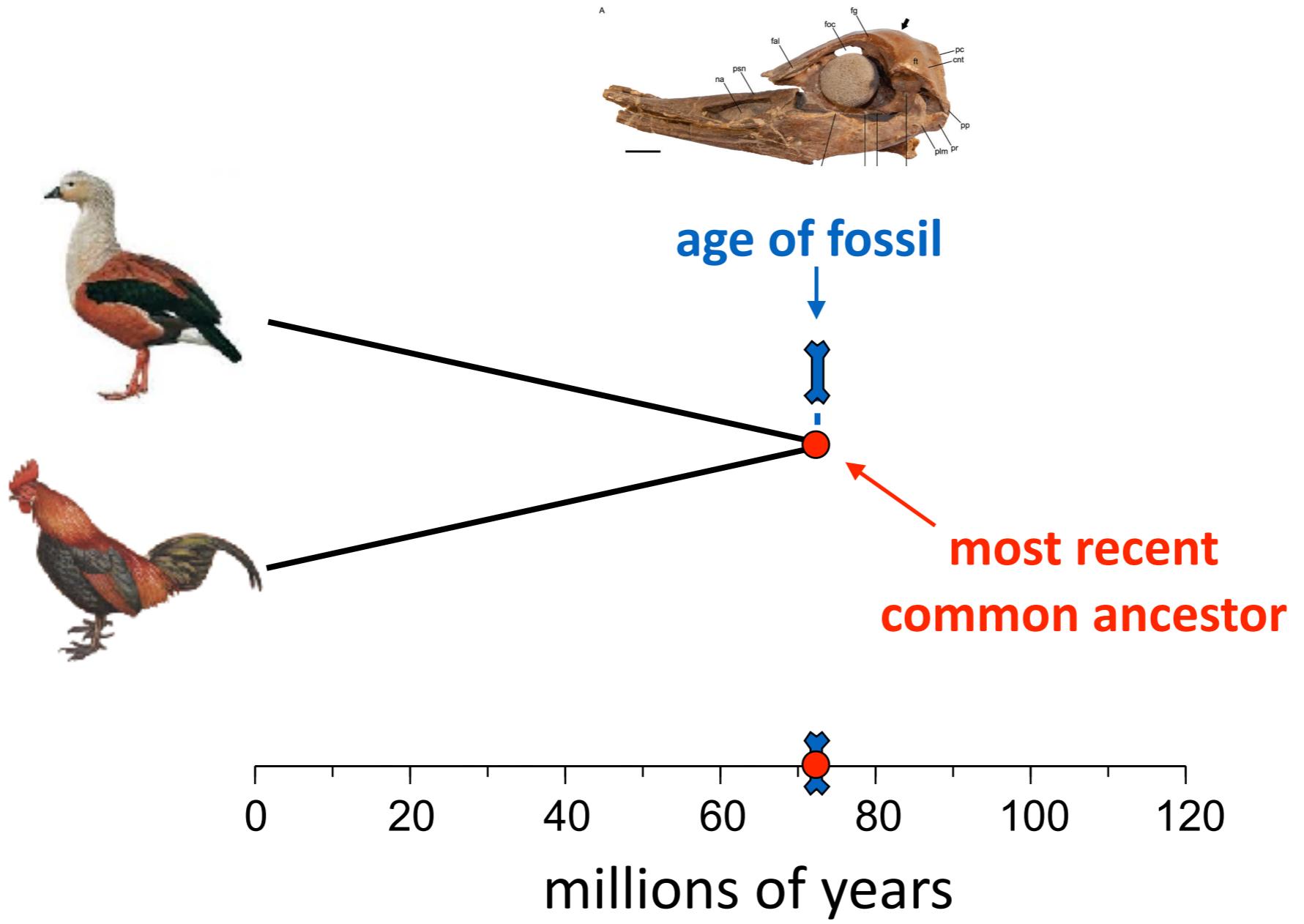


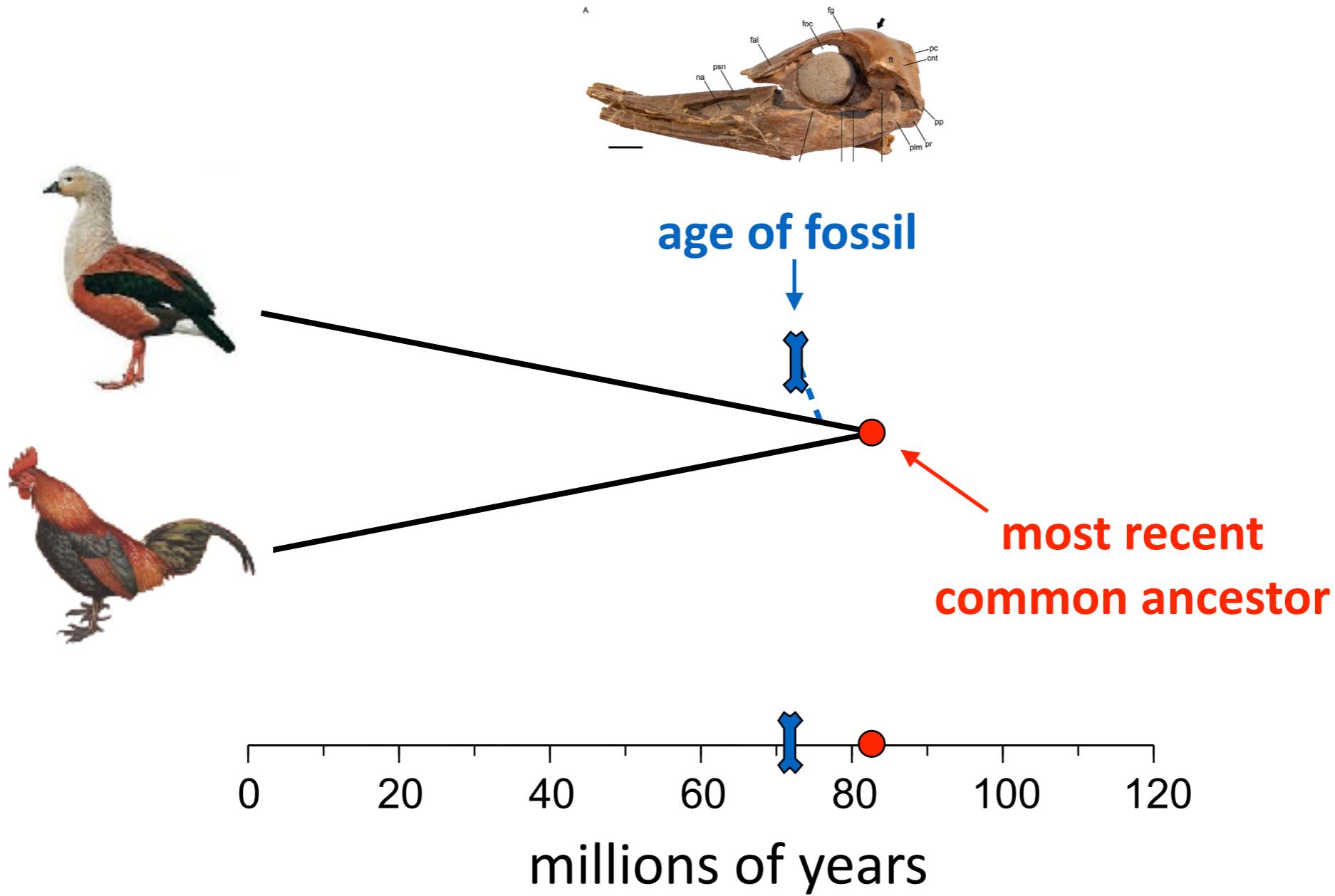
Figure 2. Skeletal anatomy of *Conflictus antarcticus* gen. et sp. nov. MLP 07-III-1-1. Bones coloured in grey are missing. Some of the bones have been mirrored (e.g. os quadratum, os coracoideum) or restored based on both homologous bones (e.g. humerus) Scale bar: 5 cm.

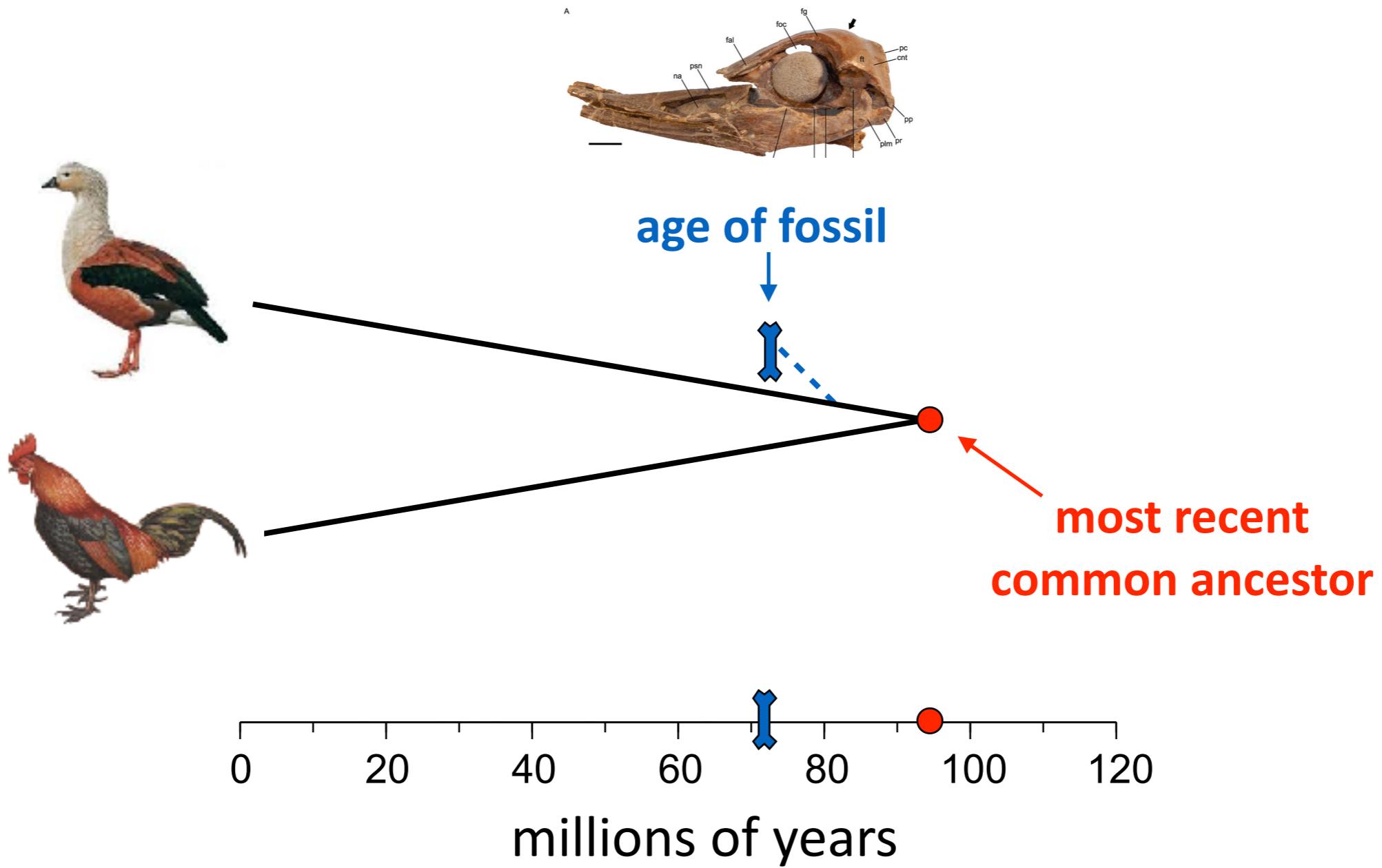
Conficto antarcticus

65.7 Ma



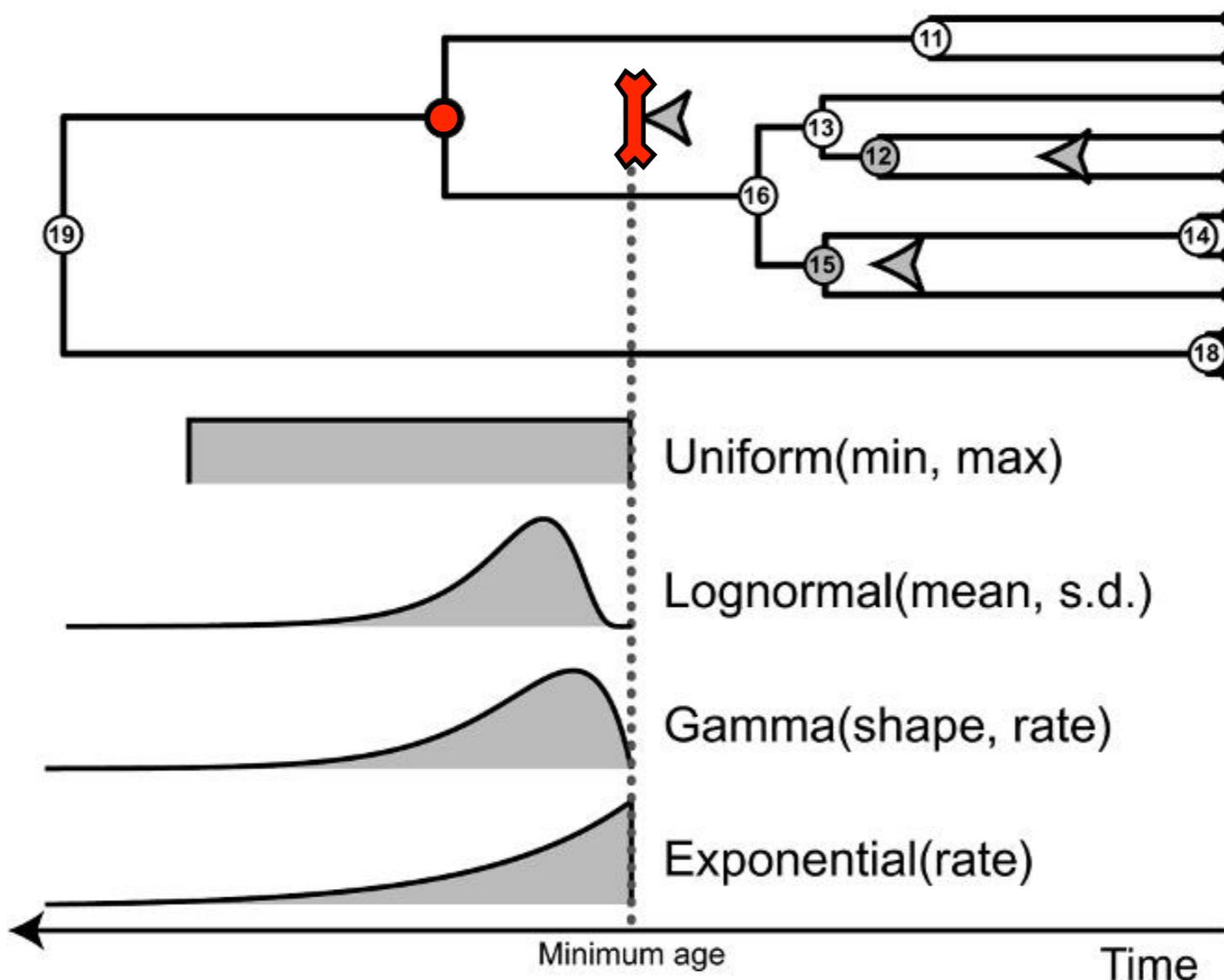




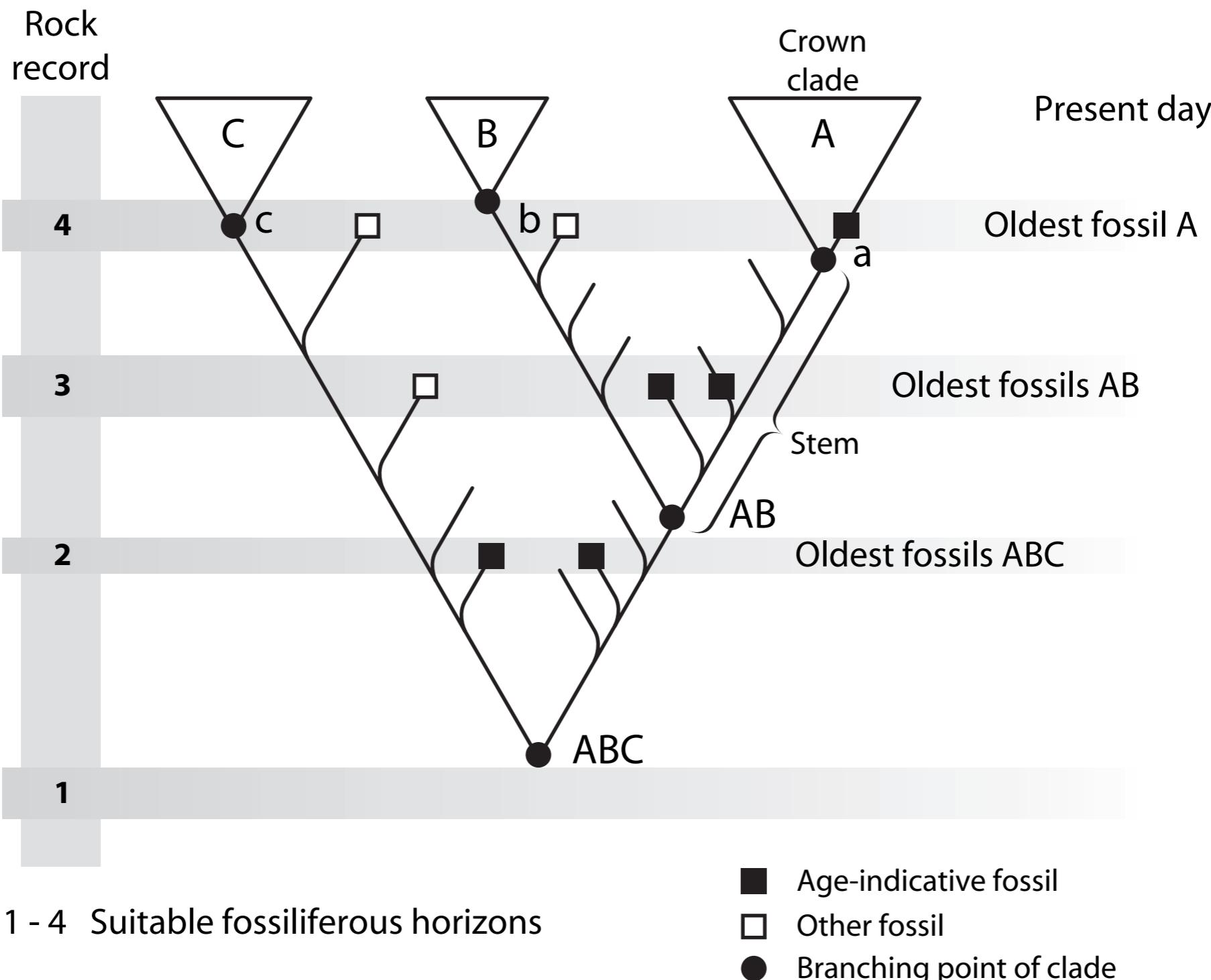


“The agony of choice”

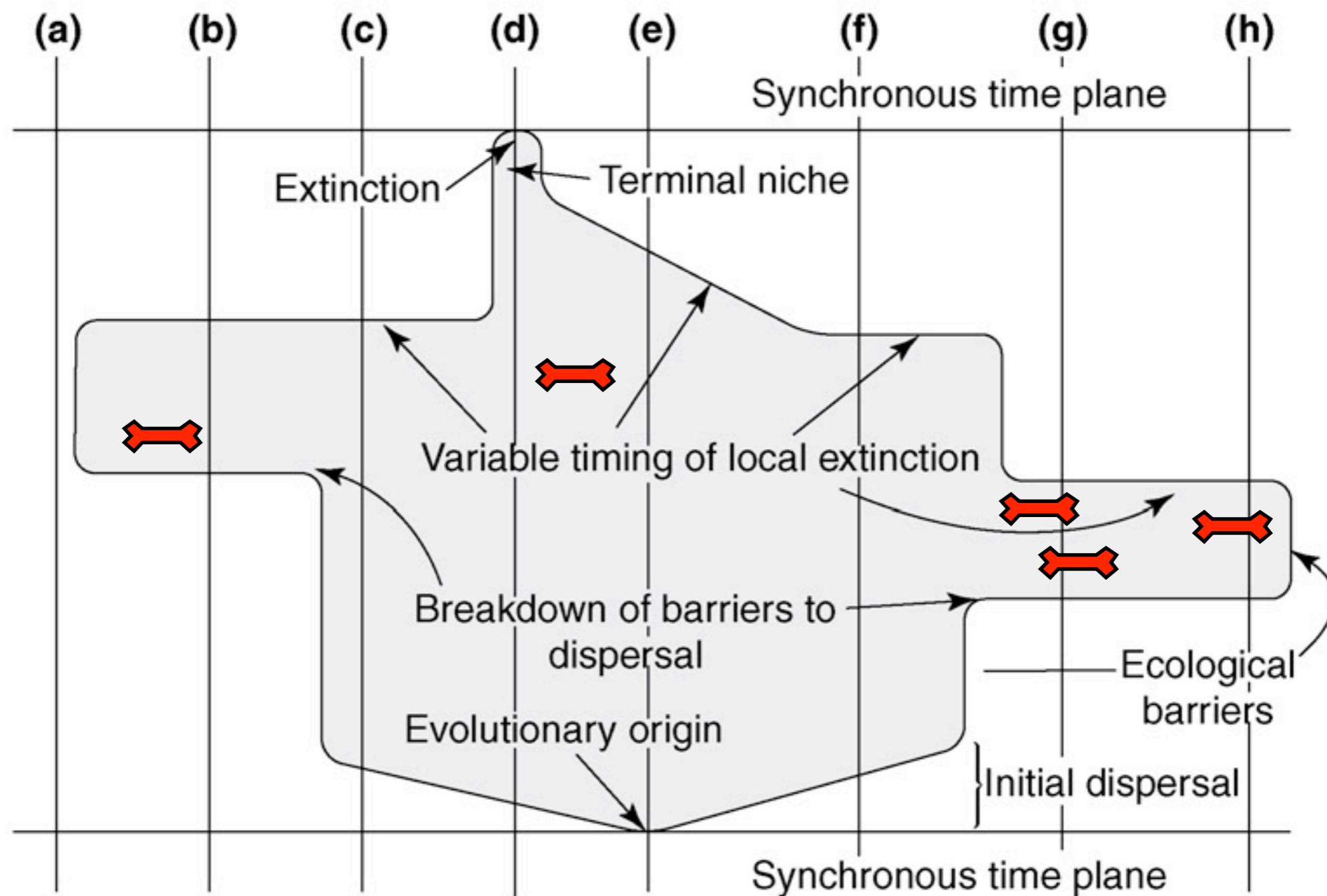
- probability functions
- parameters of the functions



How About Maximum Bounds Absence of Evidence as Evidence of Absence?



Geographic Biases



TRENDS in Ecology & Evolution

The Solution

J Mol Evol (1990) 30:400–408

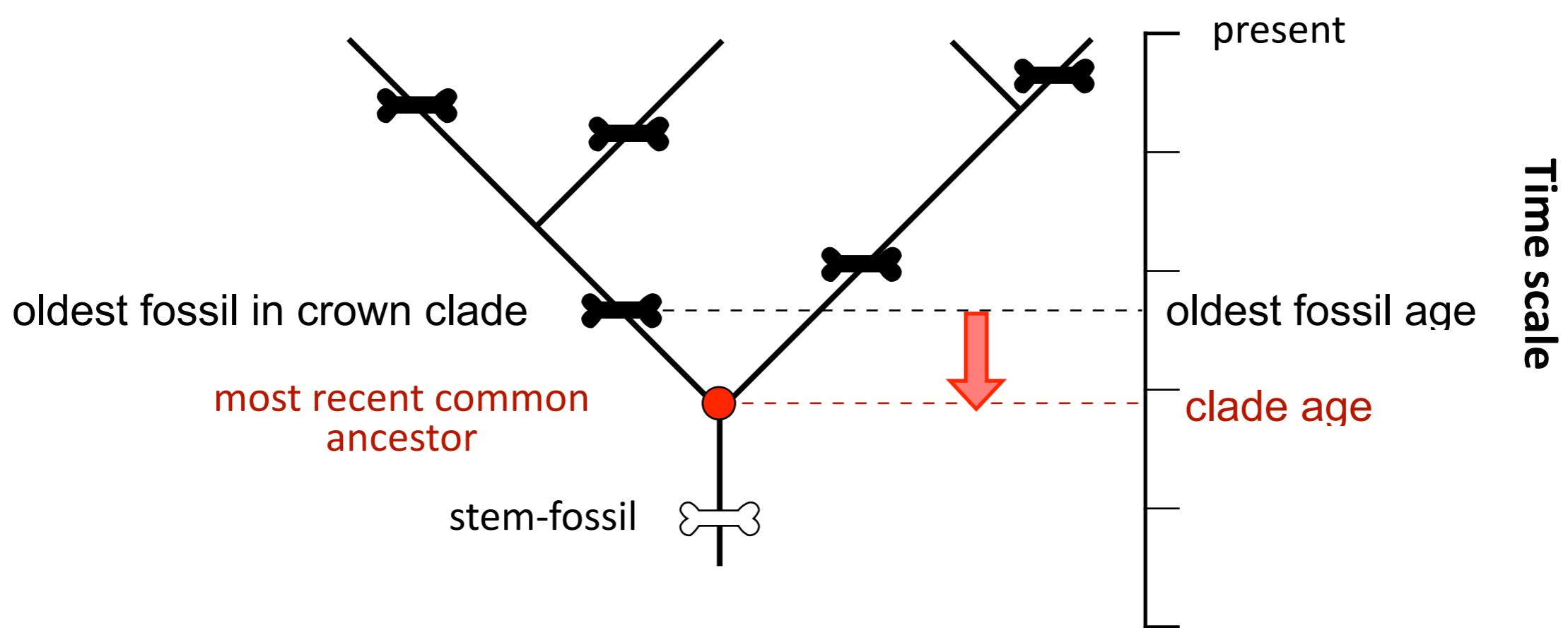
Journal of
Molecular Evolution

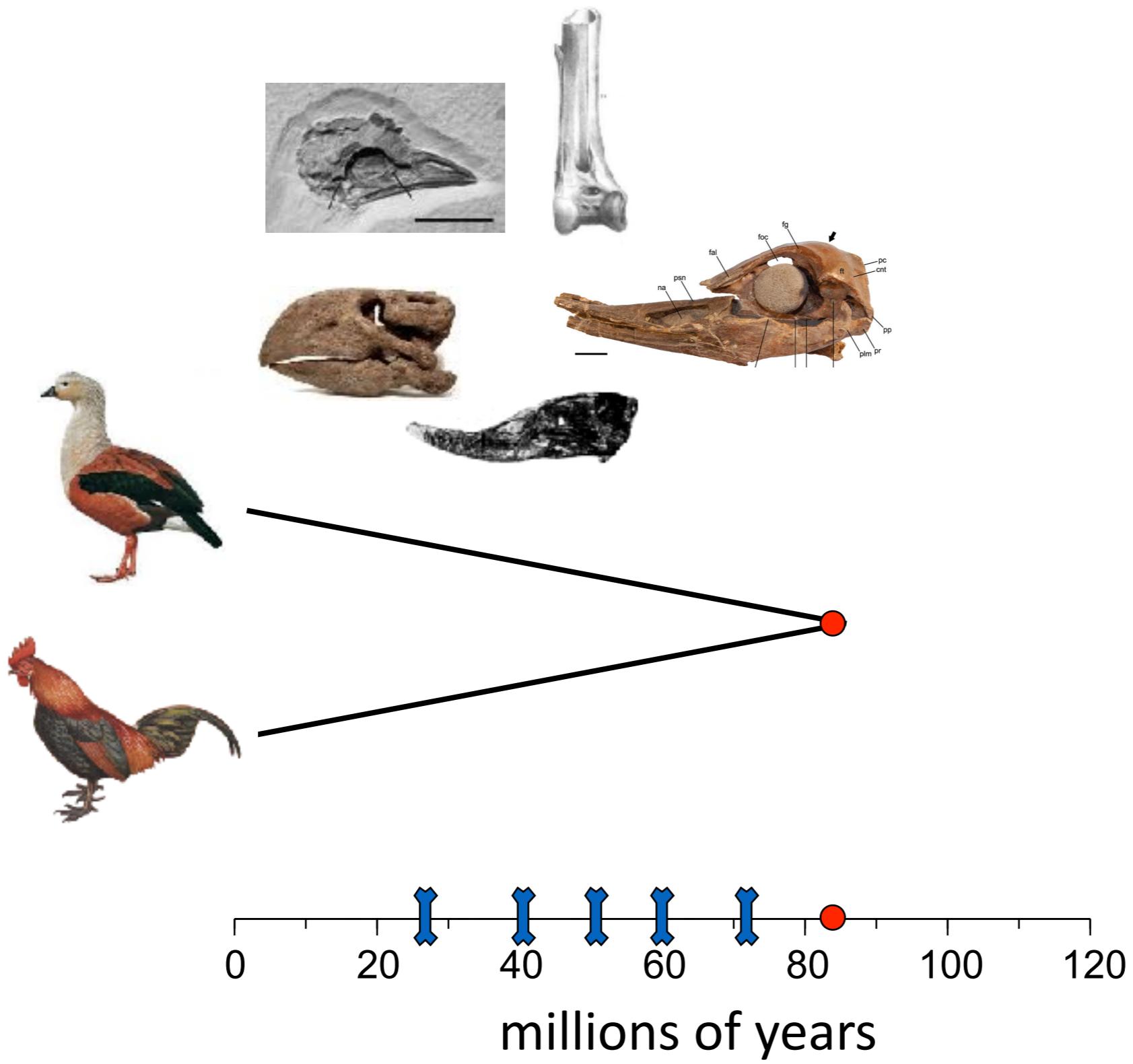
© Springer-Verlag New York Inc. 1990

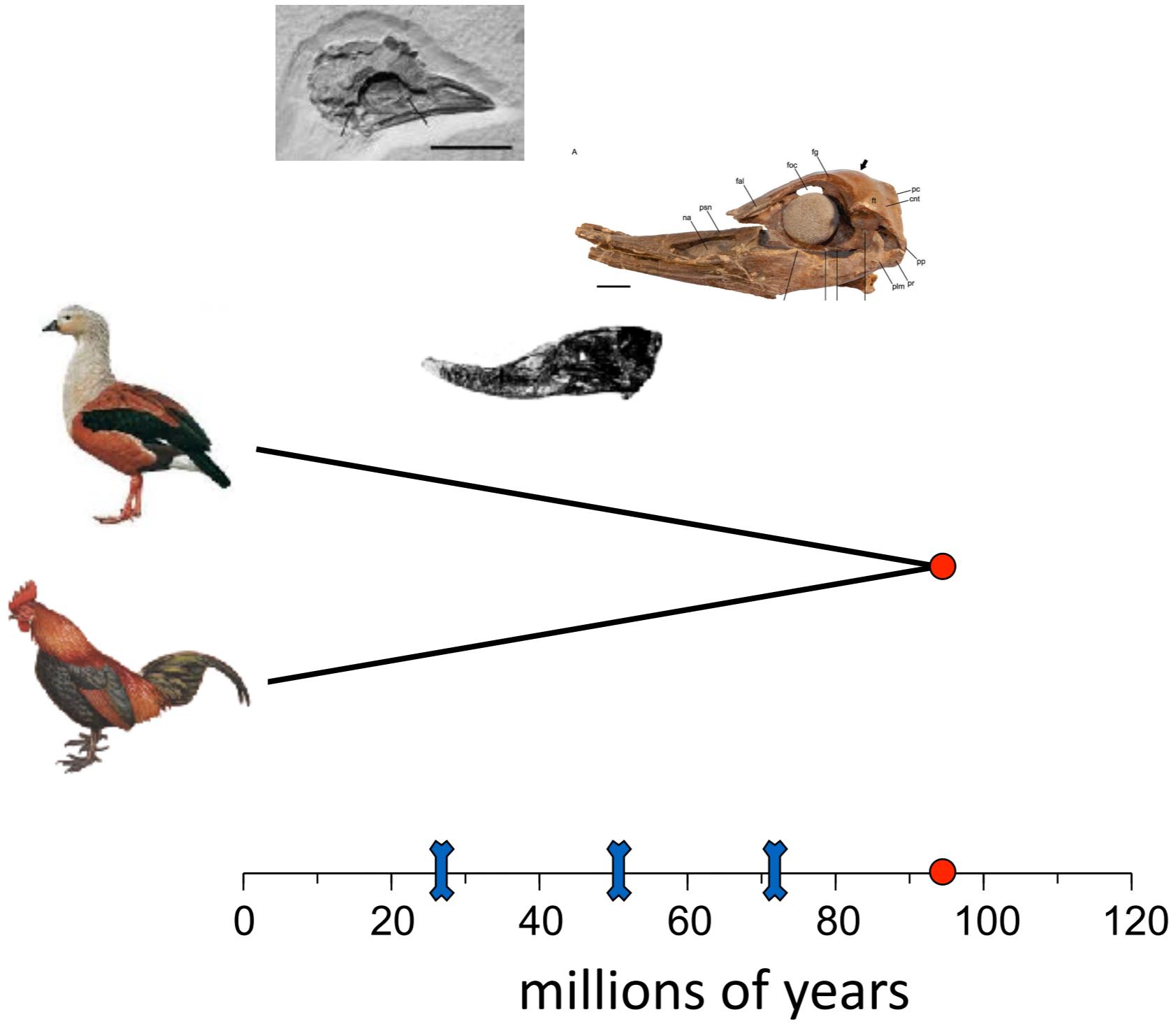
The Fossil Record and Estimating Divergence Times between Lineages: Maximum Divergence Times and the Importance of Reliable Phylogenies

Charles R. Marshall*

A fourth approach, and potentially the most powerful and widely applicable, involves quantification of the intuition that the better the fossil record, the smaller the gap between the earliest known fossil and the time of origin of the group (Shaw 1964; Paul 1982). In some cases confidence intervals on times of divergence can be calculated, and the base of these intervals may be used as lower bounds on divergence times (Marshall 1990).





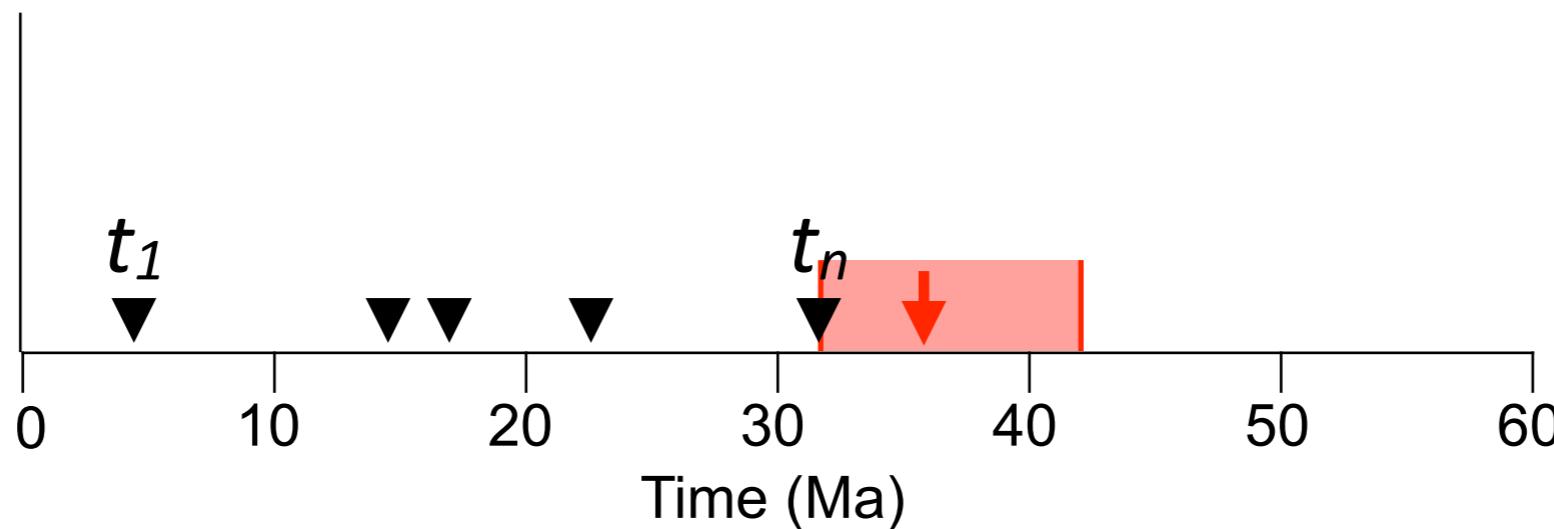


Estimating the Age of Clades

- Instead of using just the oldest fossil of the clade, we will use more fossils.
- Under the assumption that the fossils are distributed uniformly over time (the result of a uniform Poisson process) we can calculate:
 - a) unbiased point estimates and confidence intervals for the clade's age (**Strauss & Sadler 1989, Solow 2003**).

Point estimate: $t_n + t_n/n$

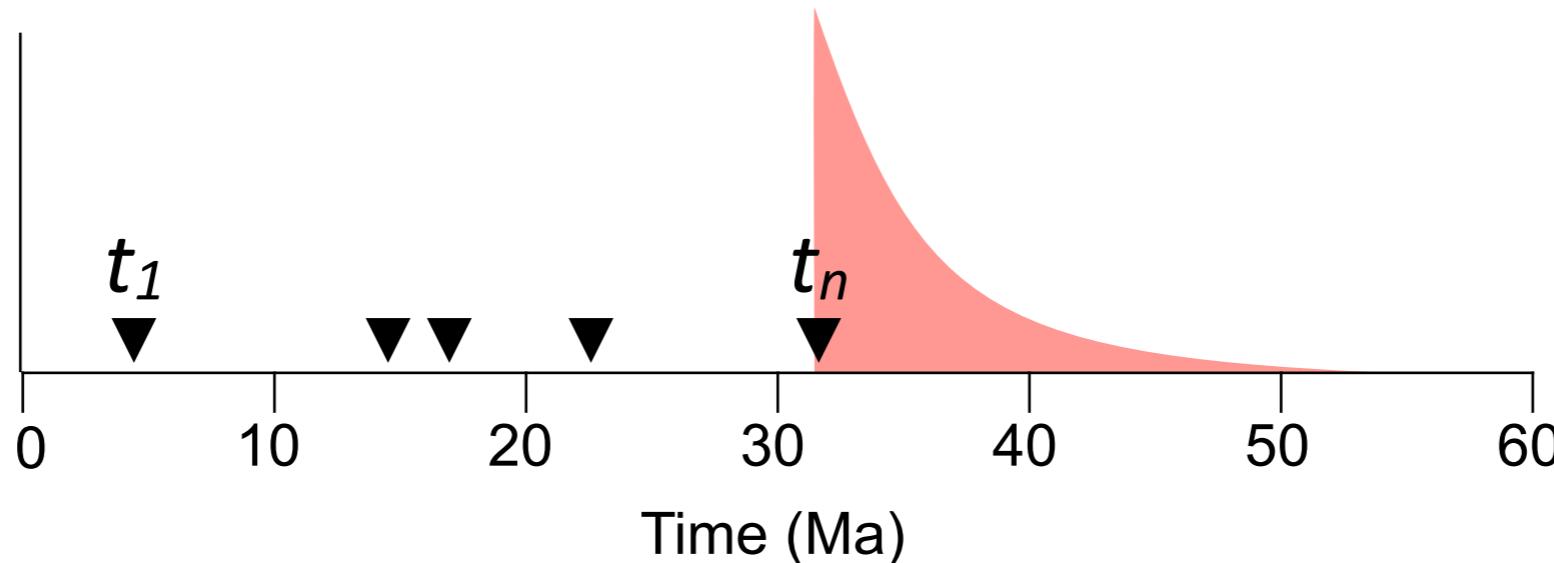
Confidence Interval: $t_n, t_n/(1-P)^{1/n}$



Estimating the Age of Clades

- Instead of using just the oldest fossil of the clade, we will use more fossils.
- Under the assumption that the fossils are distributed uniformly over time (the result of a uniform Poisson process) we can calculate:
 - a) unbiased point estimates and confidence intervals for the clade's age (Strauss & Sadler 1989, Solow 2003).
 - b) probability density for clade age ϑ :**

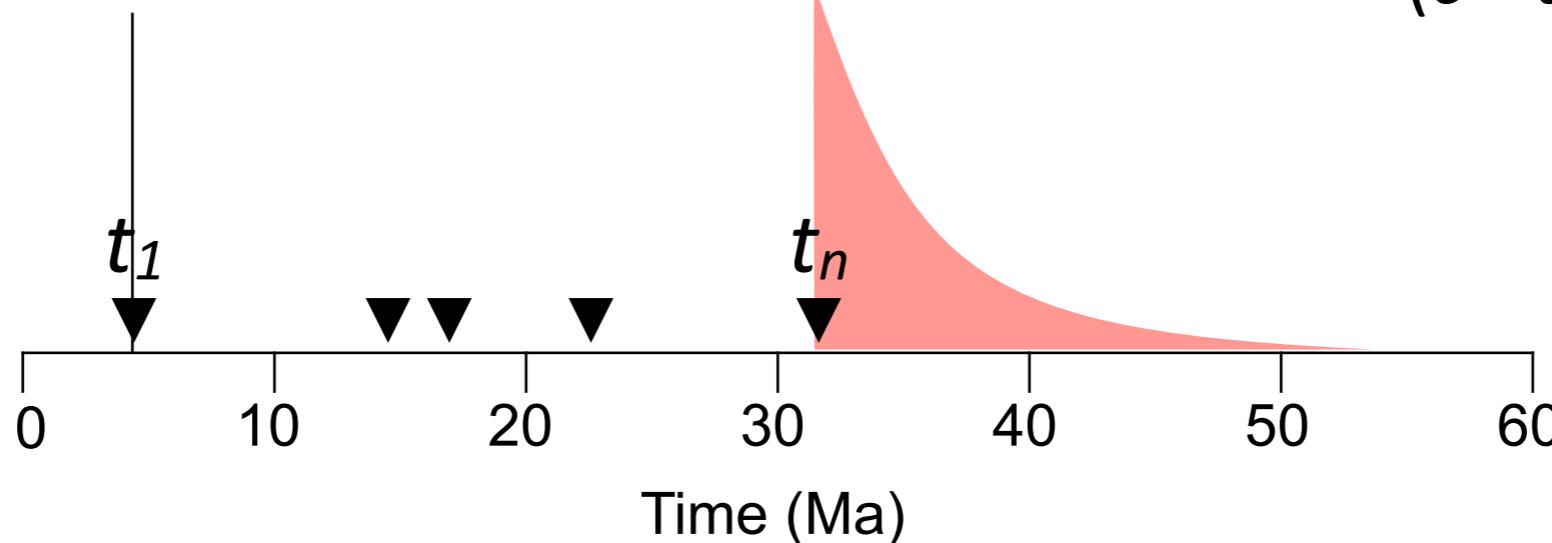
$$P(\vartheta | t_1..t_n) = \frac{1}{\vartheta^n} \text{ for } \vartheta > t_n$$



Estimating the Age of Clades

- Instead of using just the oldest fossil of the clade, we will use more fossils.
- Under the assumption that the fossils are distributed uniformly over time (the result of a uniform Poisson process) we can calculate:
 - a) unbiased point estimates and confidence intervals for the clade's age (Strauss & Sadler 1989, Solow 2003).
 - b) probability density for clade age ϑ :
 - c) **t_1 can be used as the base line**

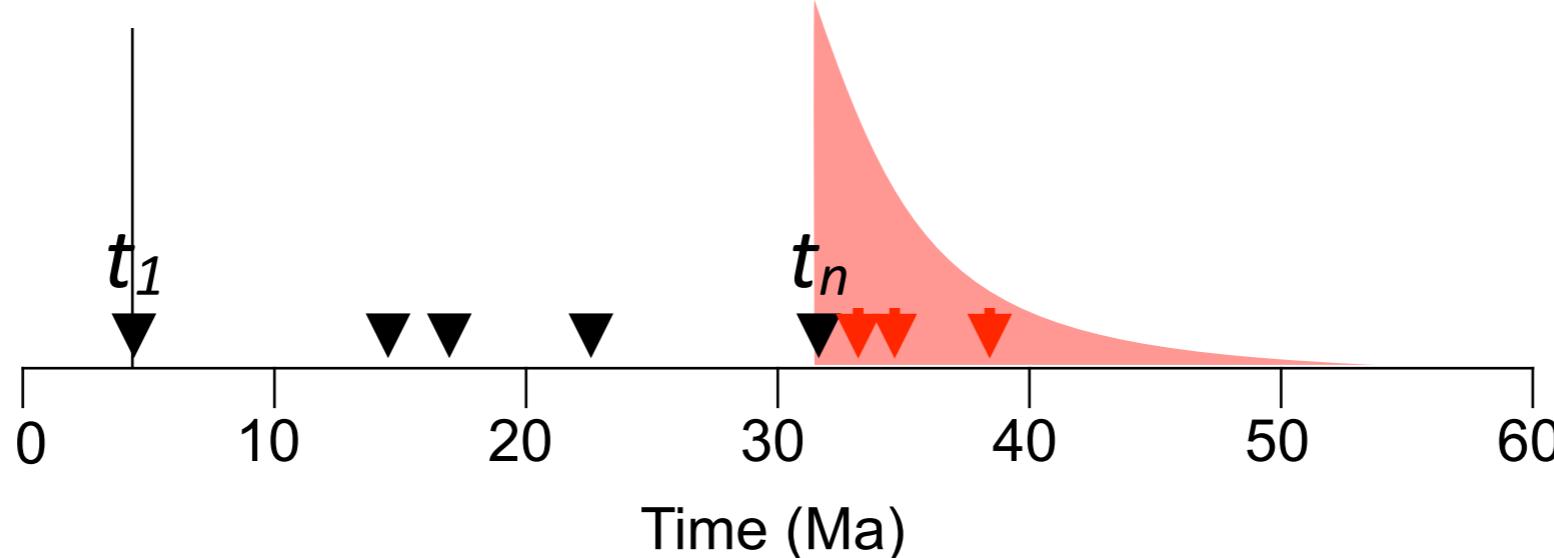
$$P(\vartheta | t_1..t_n) = \frac{1}{(\vartheta - t_1)^{n-1}} \text{ for } \vartheta > t_n$$



Estimating the Age of Clades

- Instead of using just the oldest fossil of the clade, we will use more fossils.
- Under the assumption that the fossils are distributed uniformly over time (the result of a uniform Poisson process) we can calculate:
 - a) unbiased point estimates and confidence intervals for the clade's age (Strauss & Sadler 1989, Solow 2003).
 - b) probability density for clade age ϑ :
 - c) t_1 can be used as the baseline
 - d) Quantile functions can be used for generating random numbers for simulations**

Quantile function: $t_n / (1-P)^{1/n}$



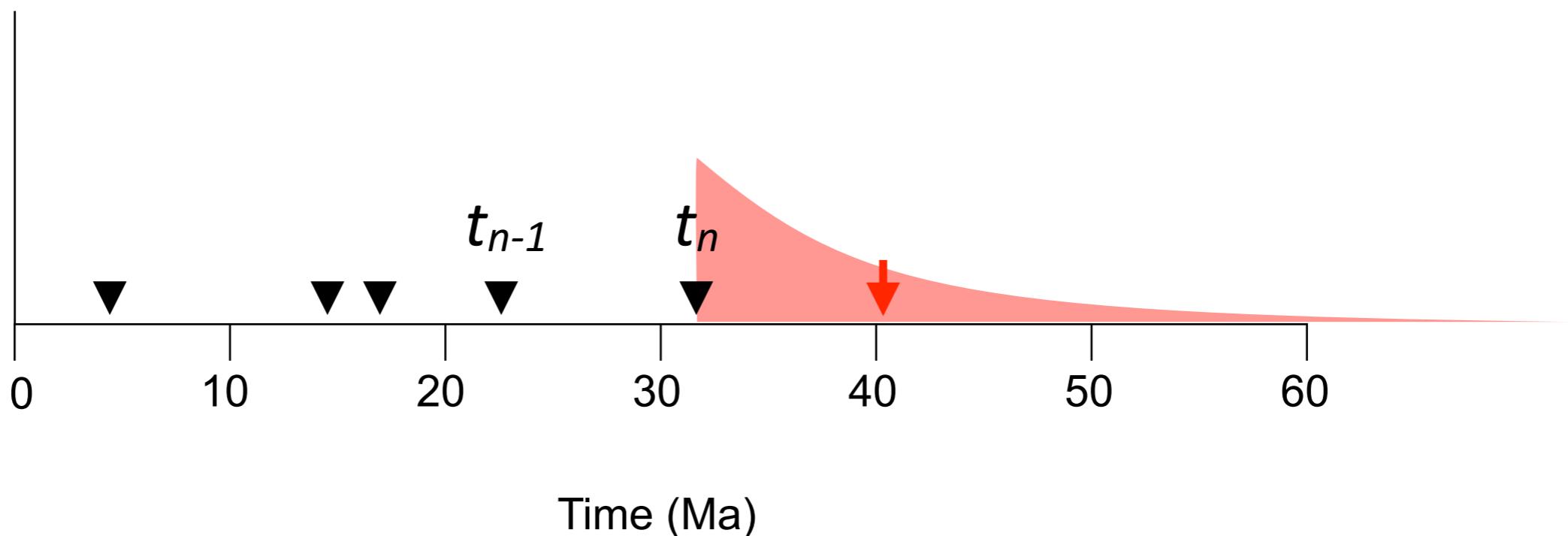
Checking Assumptions

Uniformity

Independence

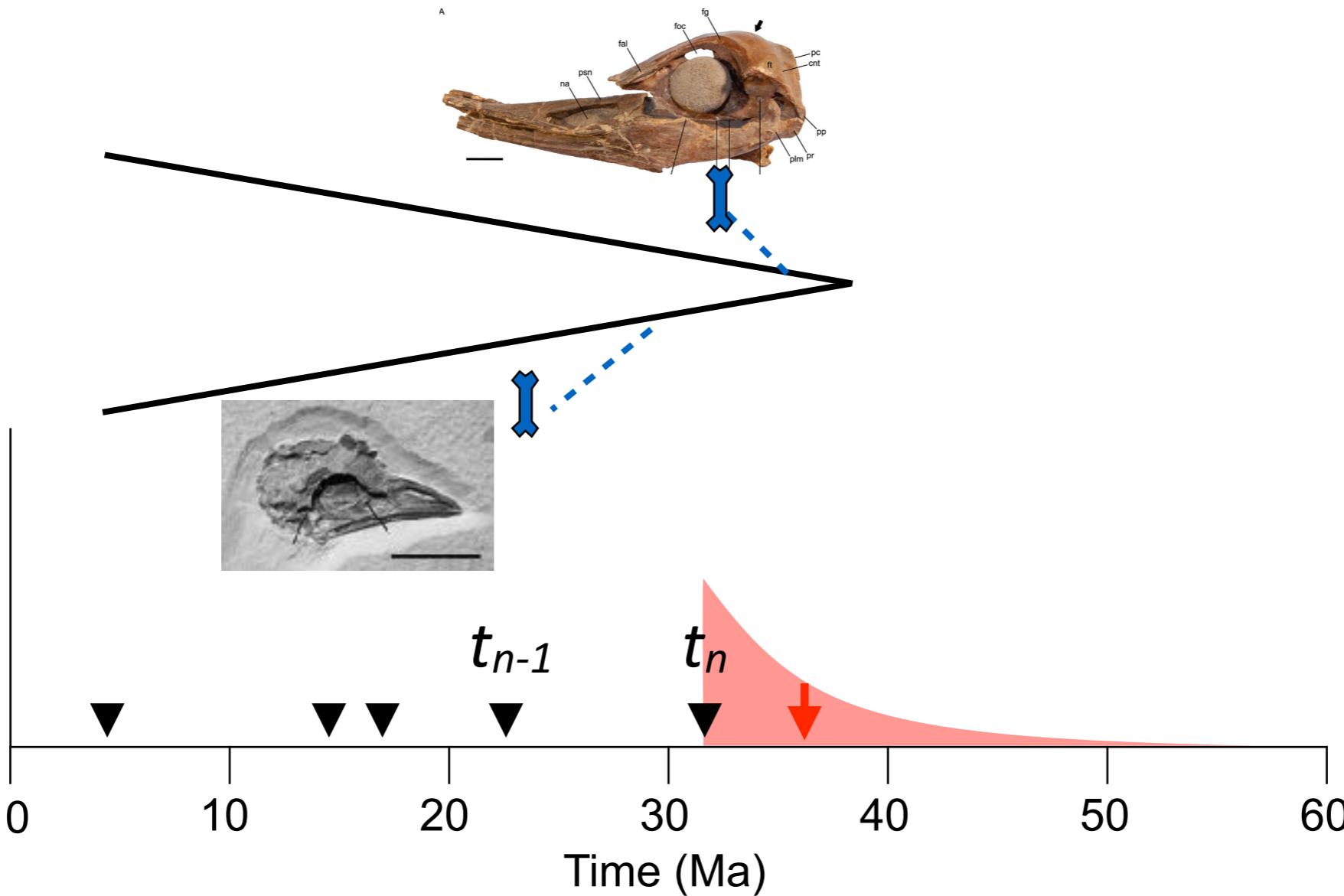
When fossil ages are not uniformly distributed

Strategy 1: use only the two oldest fossil ages (Robson & Whitlock 1964)



When fossil ages are not uniformly distributed

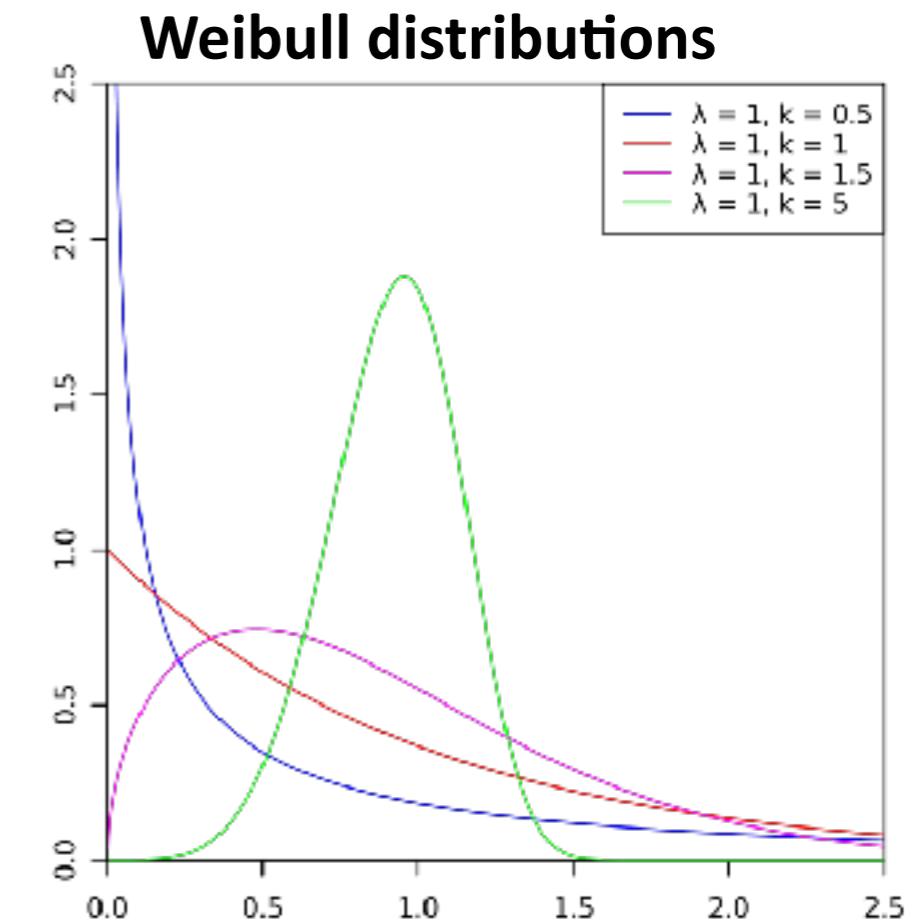
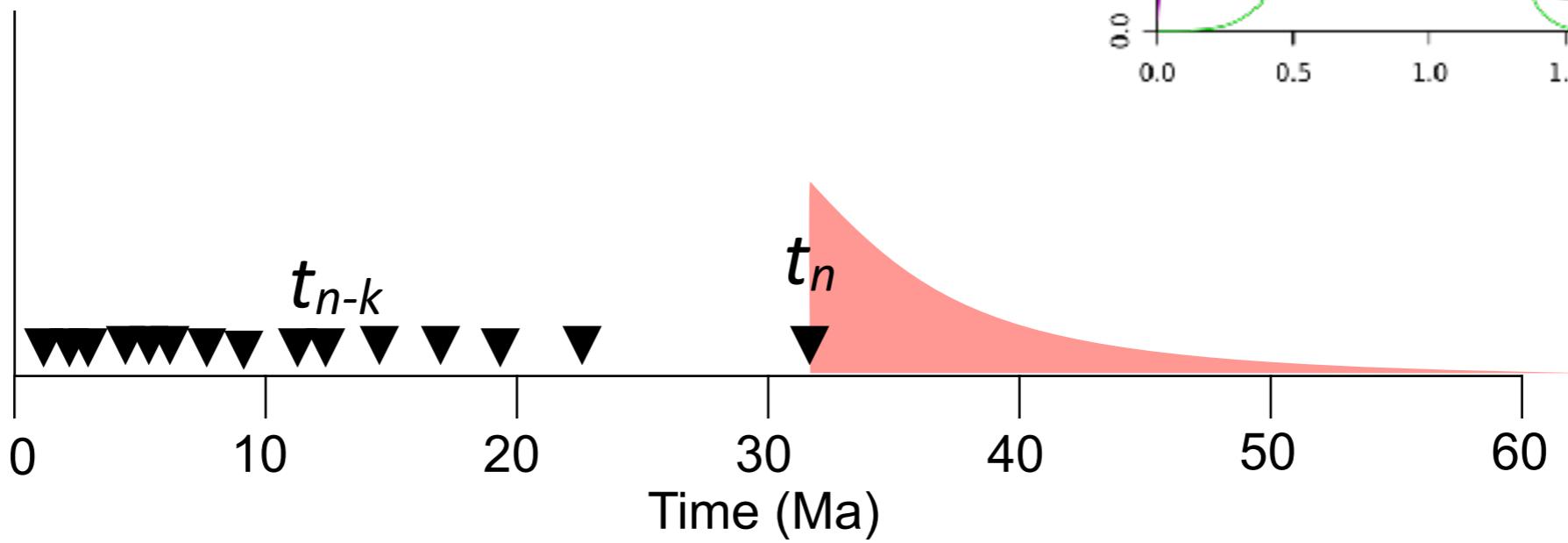
Strategy 2: If the two oldest fossils represent different lineages:
Norris Ghost-lineage method (Norris et al. 2015)



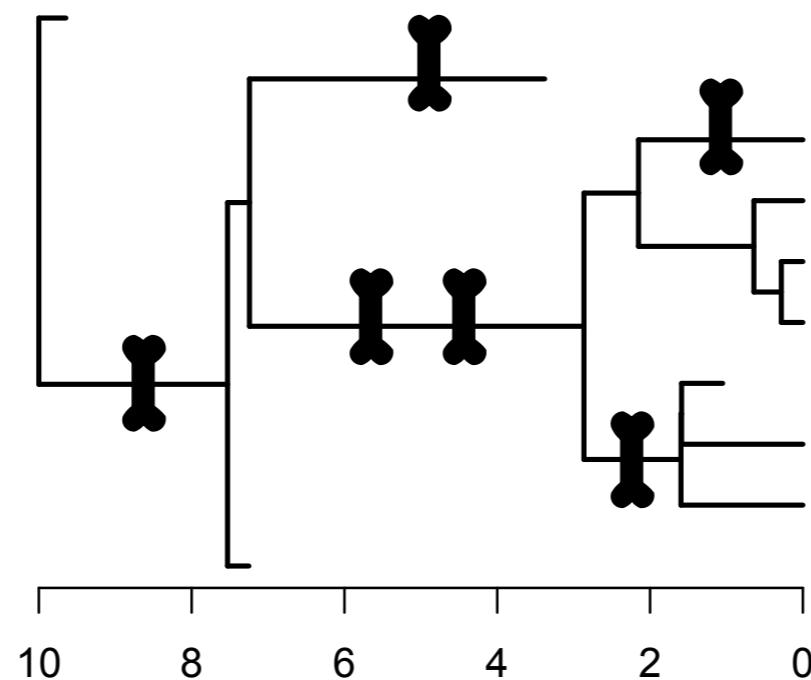
When fossil ages are not uniformly distributed

Strategy 2: use Optimal Linear Estimators

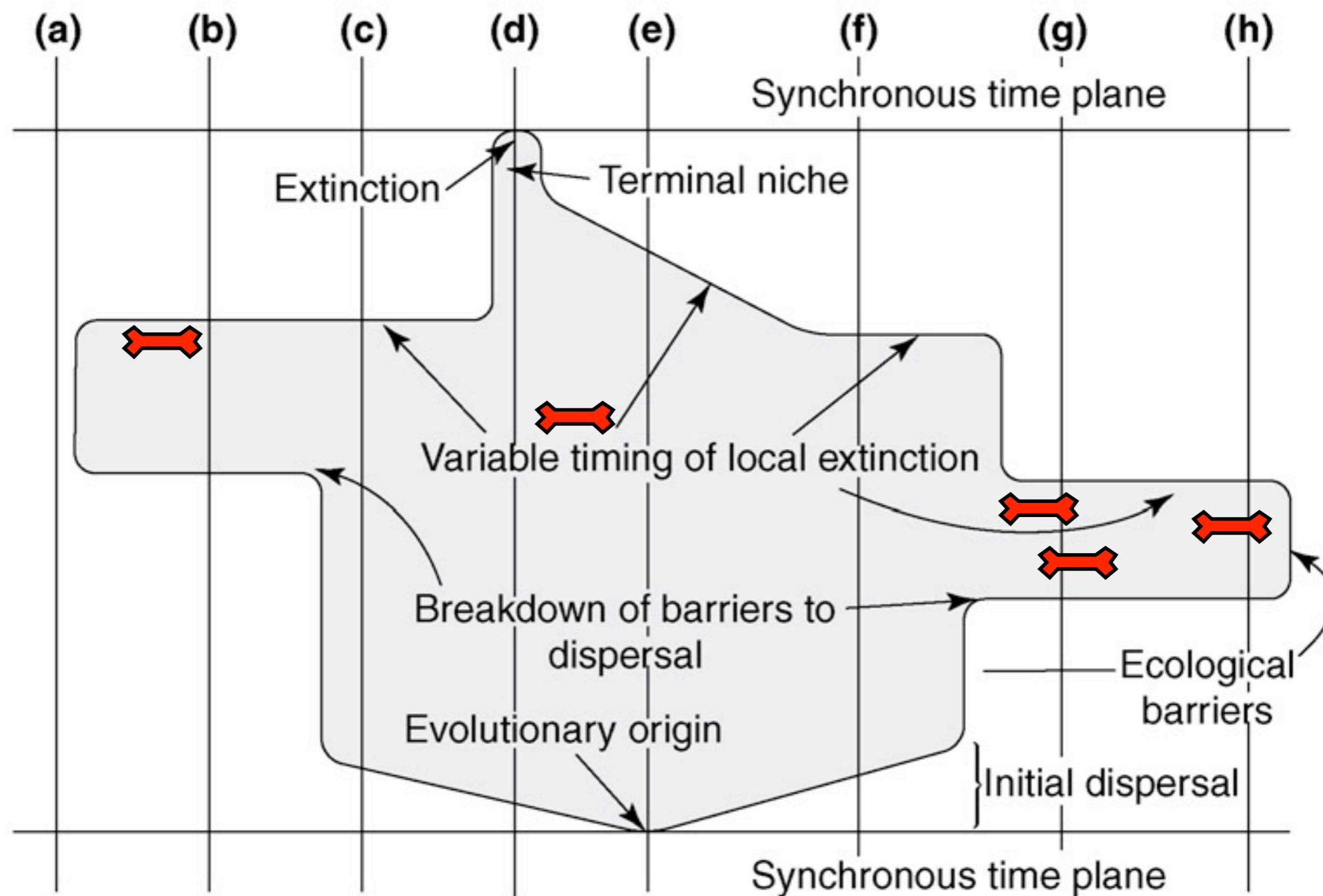
- Only the k oldest fossils are used
- Weighted sum, with weight derived from a Weibull distribution



Phylogenetic Non-independence



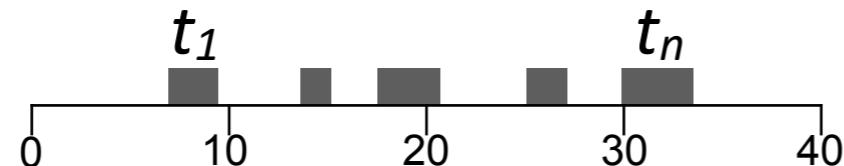
Geographic Non-independence



TRENDS in Ecology & Evolution

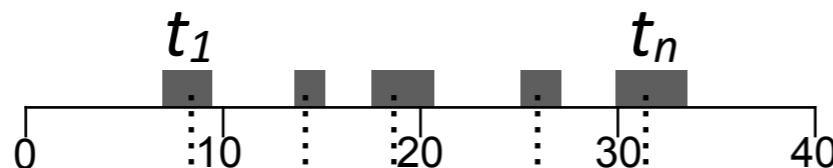
Incorporating Fossil Age Uncertainty: the clade.date algorithm

1 Collect fossil time intervals



Incorporating Fossil Age Uncertainty: the clade.date algorithm

1 Collect fossil time intervals



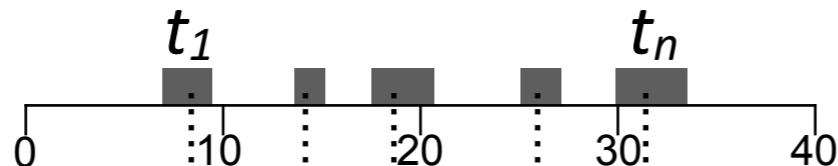
2 Monte Carlo replicates

a) sample fossil times



Incorporating Fossil Age Uncertainty: the clade.date algorithm

1 Collect fossil time intervals

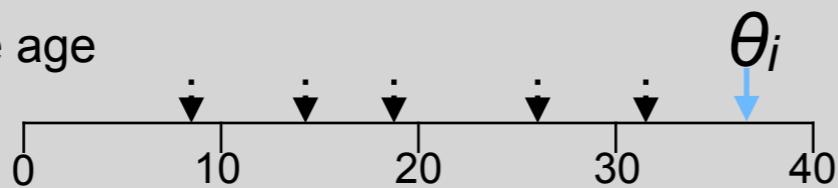


2 Monte Carlo replicates

a) sample fossil times

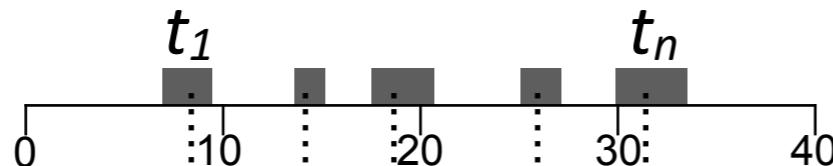


b) sample clade age



Incorporating Fossil Age Uncertainty: the clade.date algorithm

1 Collect fossil time intervals

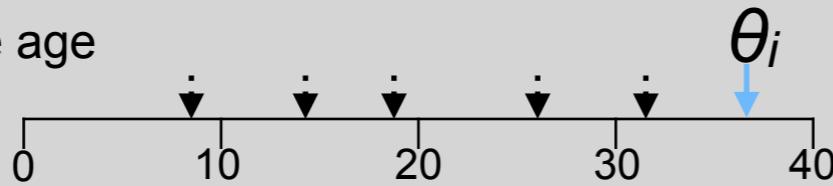


2 Monte Carlo replicates

a) sample fossil times

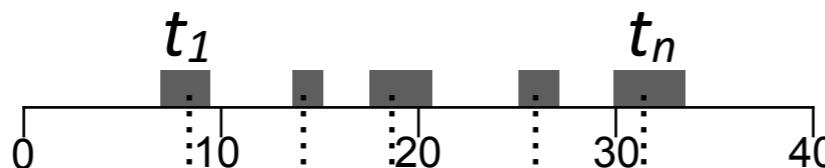


b) sample clade age



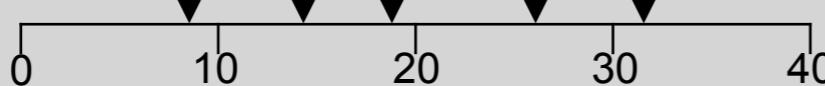
Incorporating Fossil Age Uncertainty: the clade.date algorithm

1 Collect fossil time intervals

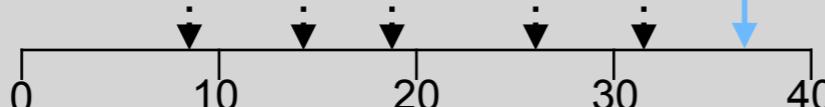


2 Monte Carlo replicates

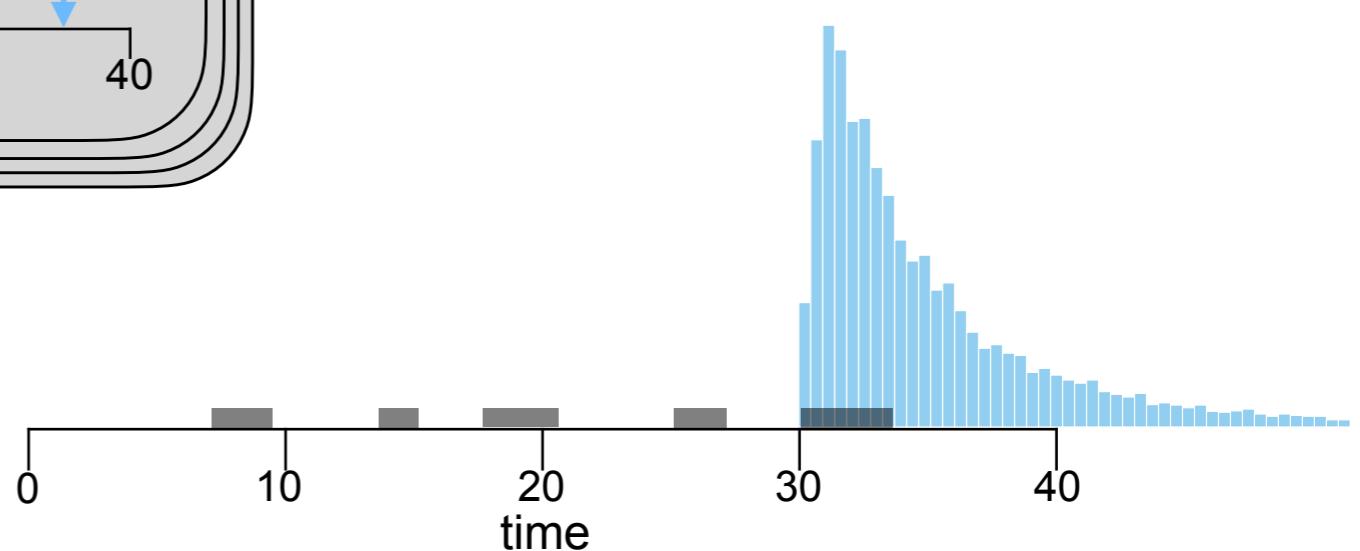
a) sample fossil times



b) sample clade age

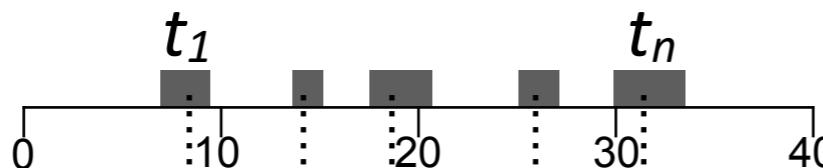


3 Collect distribution of sampled ages θ



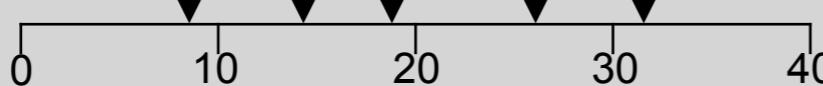
Incorporating Fossil Age Uncertainty: the clade.date algorithm

1 Collect fossil time intervals

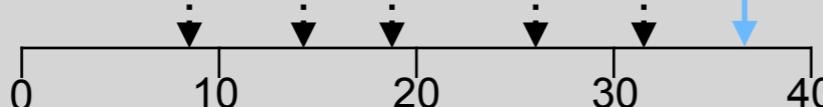


2 Monte Carlo replicates

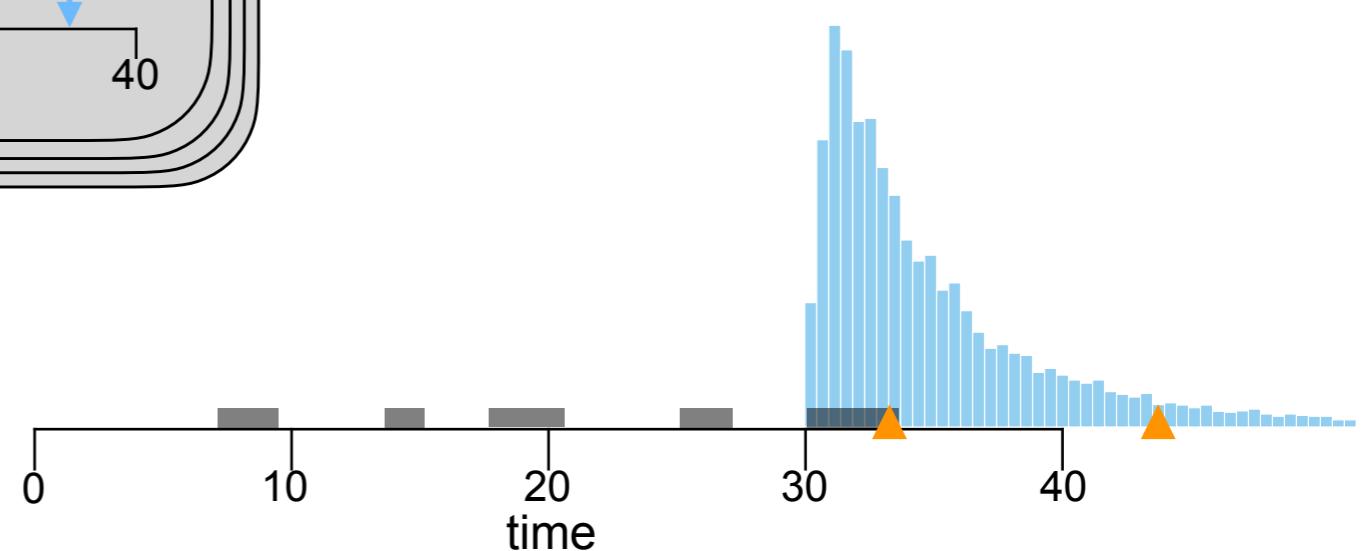
a) sample fossil times



b) sample clade age



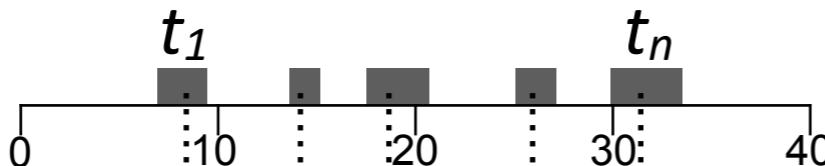
3 Collect distribution of sampled ages θ



4 Compute quantiles ▲

Incorporating Fossil Age Uncertainty: the clade.date algorithm

1 Collect fossil time intervals

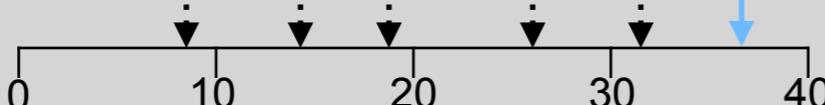


2 Monte Carlo replicates

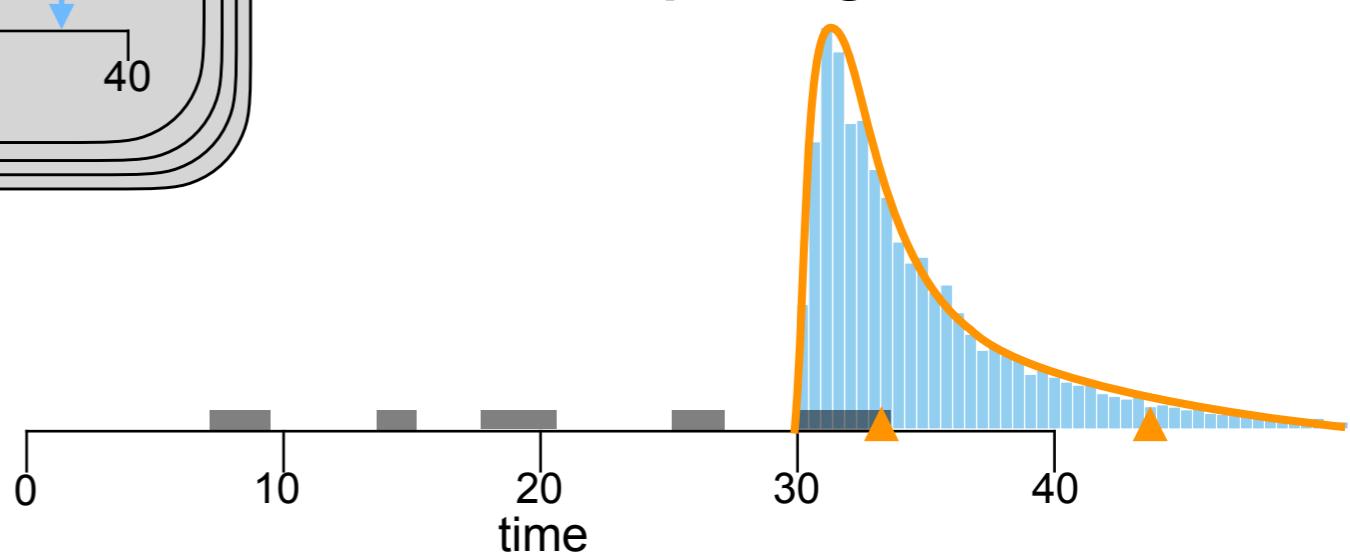
a) sample fossil times



b) sample clade age



3 Collect distribution of sampled ages θ



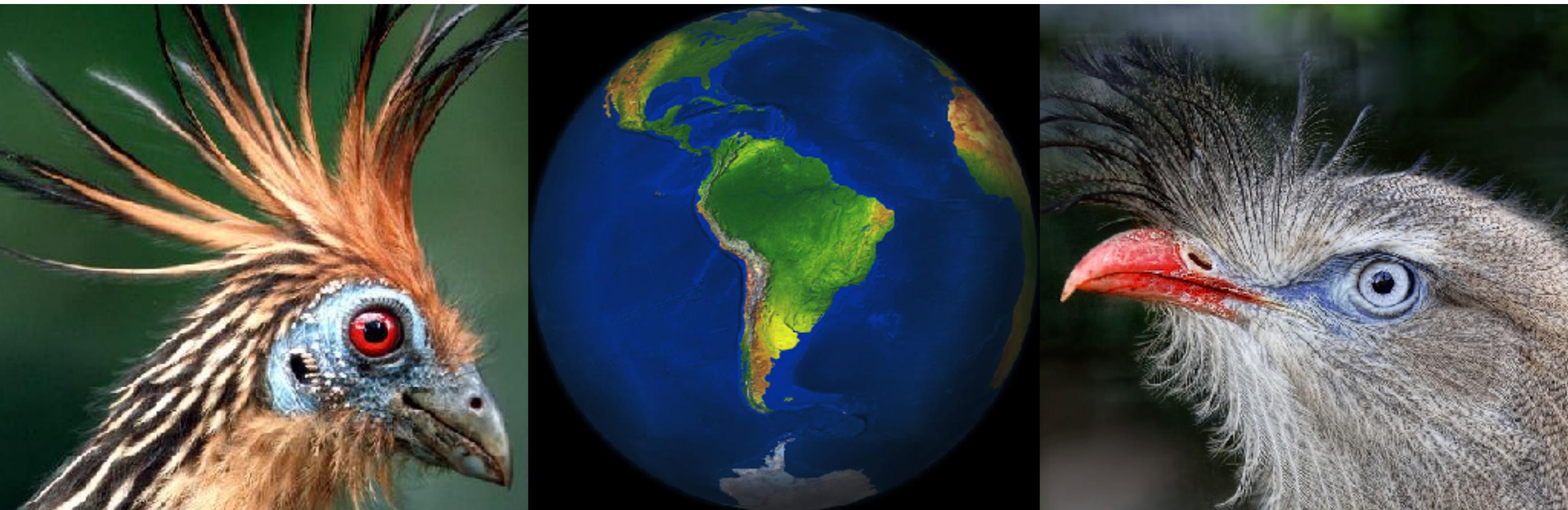
4 Compute quantiles ▲

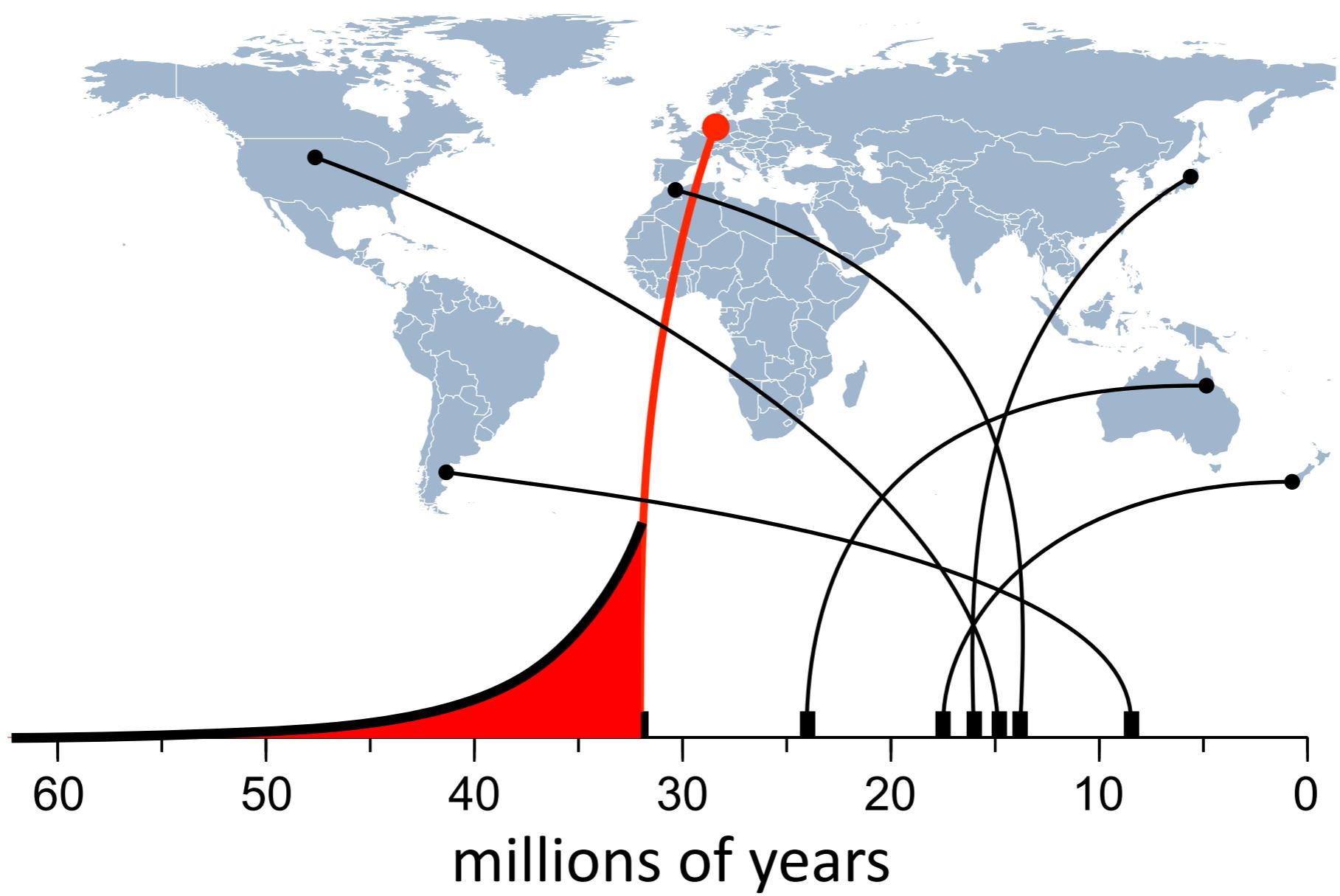
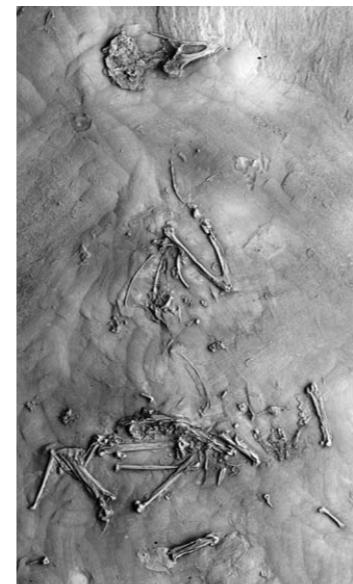
5 Fit probability density —

A new time tree reveals Earth history's imprint on the evolution of modern birds

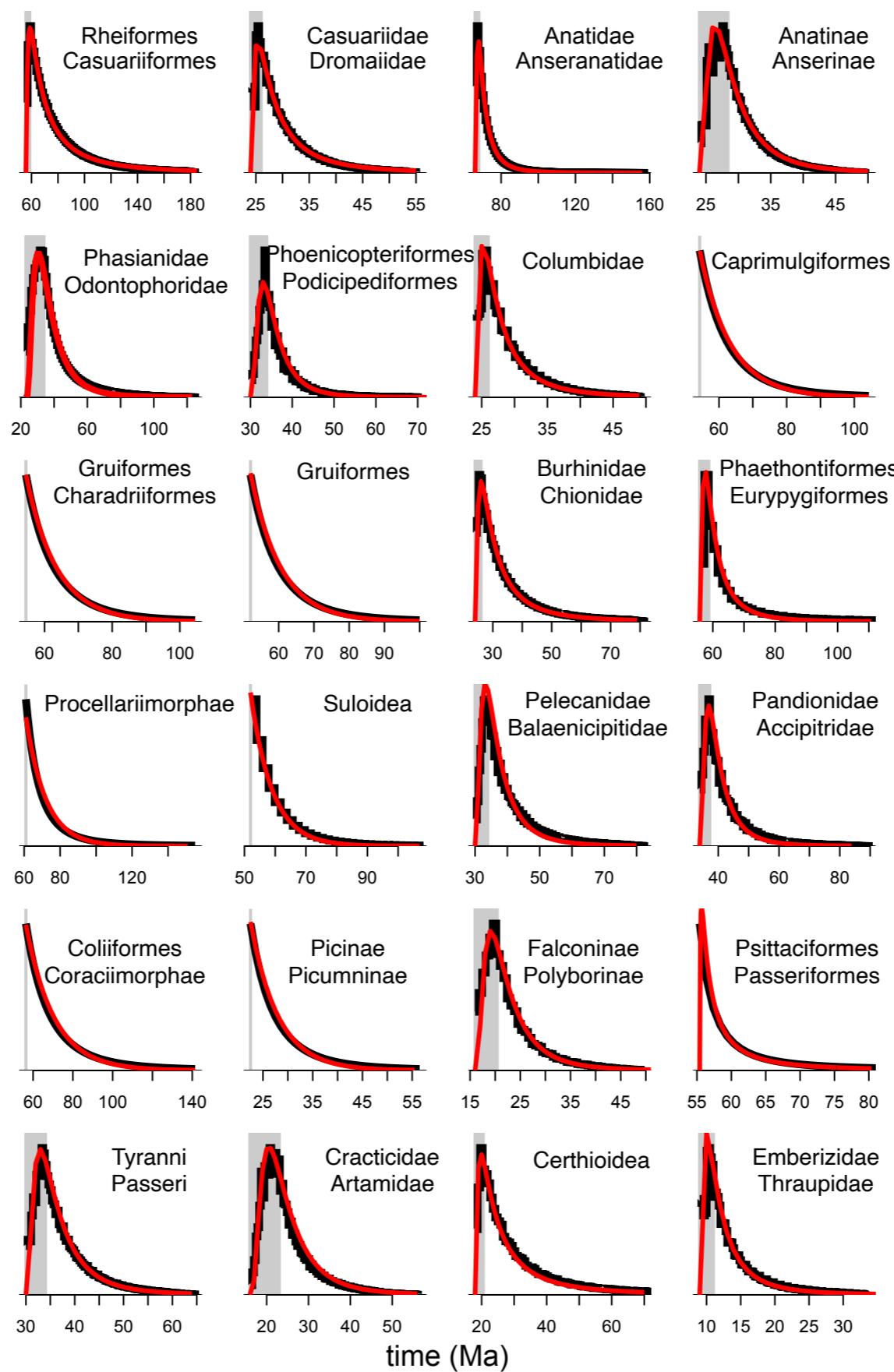
Santiago Claramunt* and Joel Cracraft*

2015 © The Authors, some rights reserved;
exclusive licensee American Association for
the Advancement of Science. Distributed
under a Creative Commons Attribution
NonCommercial License 4.0 (CC BY-NC).
[10.1126/sciadv.1501005](https://doi.org/10.1126/sciadv.1501005)





Age distributions for 24 clades using a total of 130 fossil birds



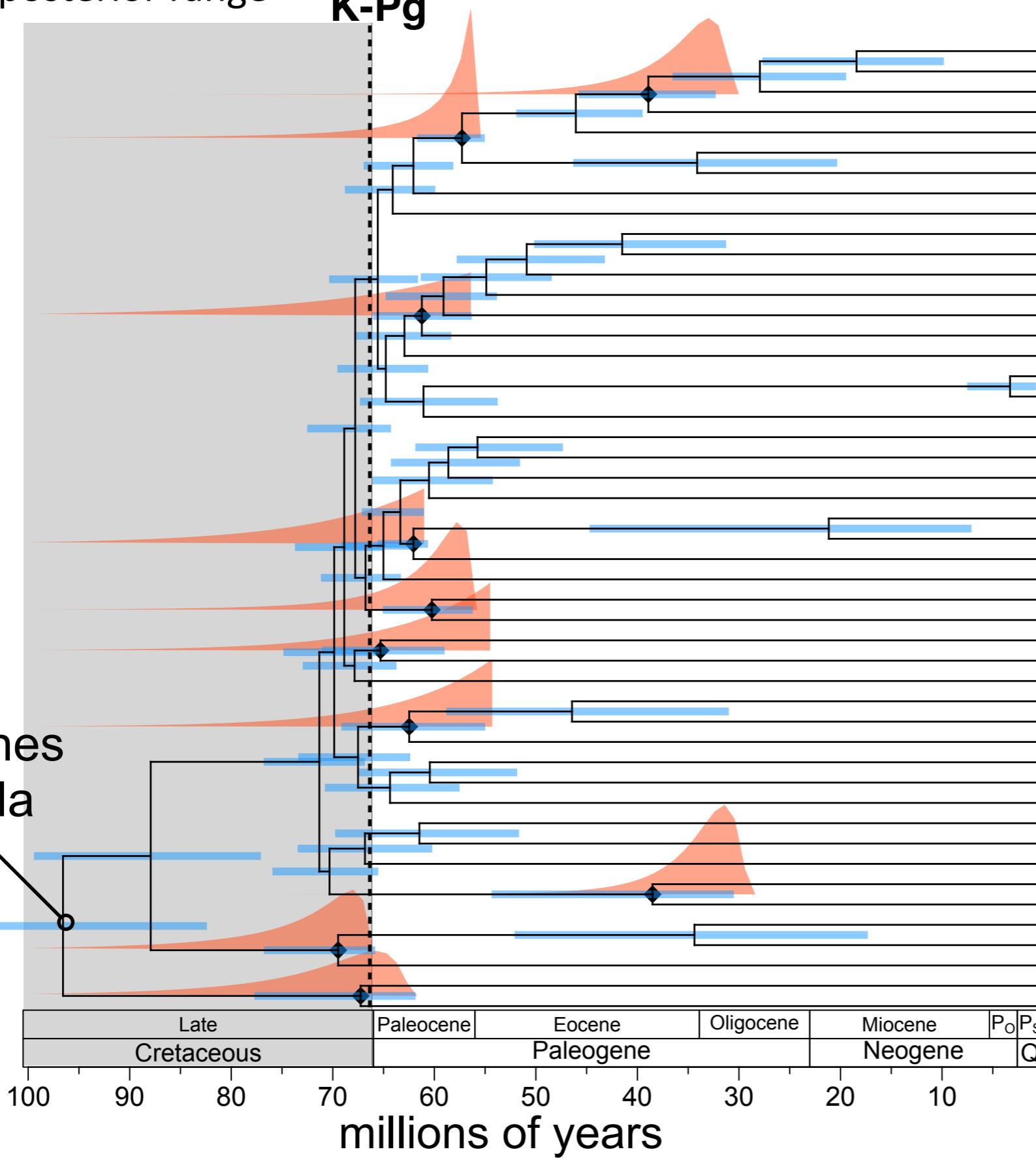
█ prior age probability
█ 95% posterior range

Genomic Dataset

K-Pg

Neornithes

96.6 Ma



Neoaves



Galloanseres

Palaeognathae



confidence intervals

K-Pg

Neornithes
91.5 Ma

Late Cretaceous Paleocene Eocene Paleogene Oligocene Miocene Neogene P_oP_s Q

100 90 80 70 60 50 40 30 20 10 0
millions of years

Neoaves

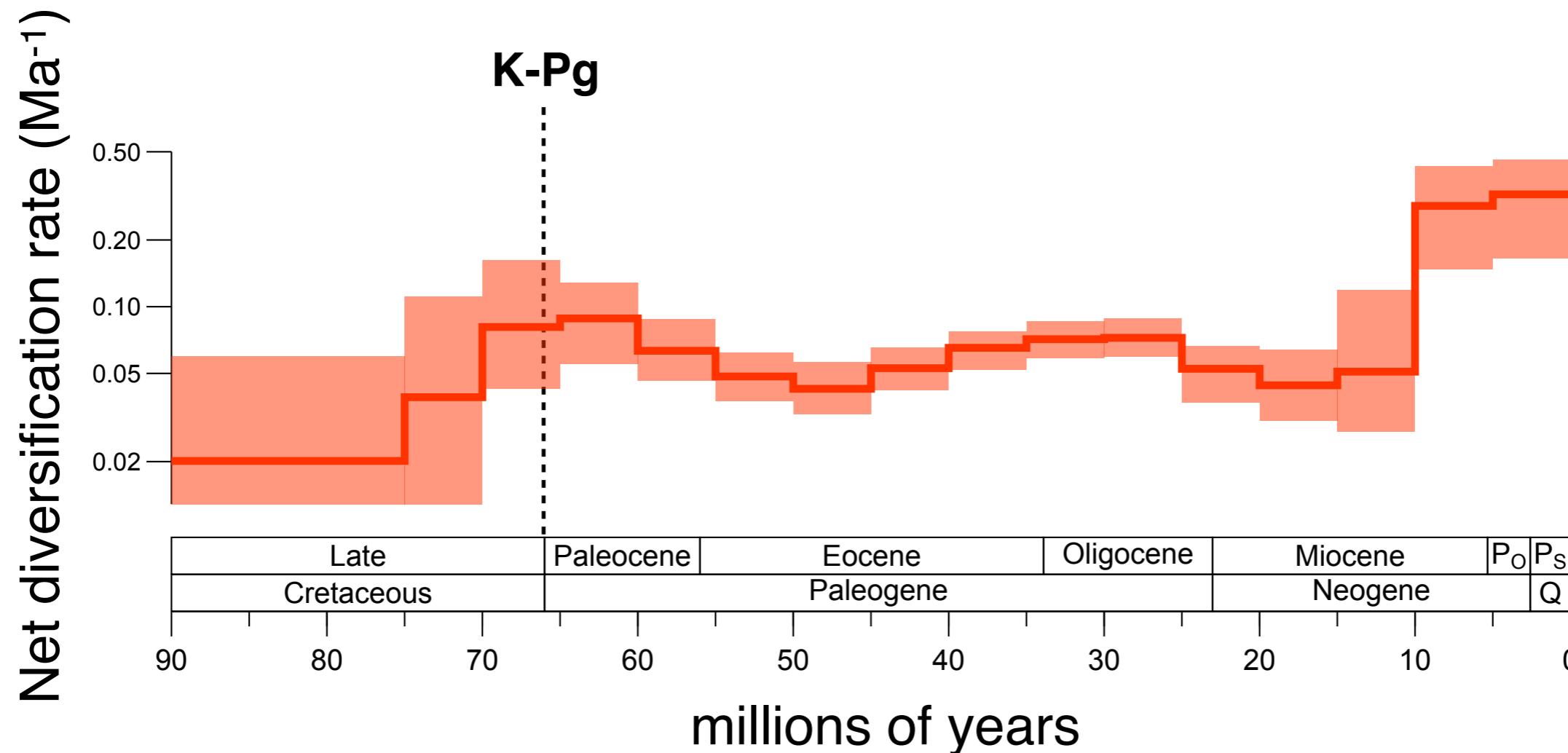


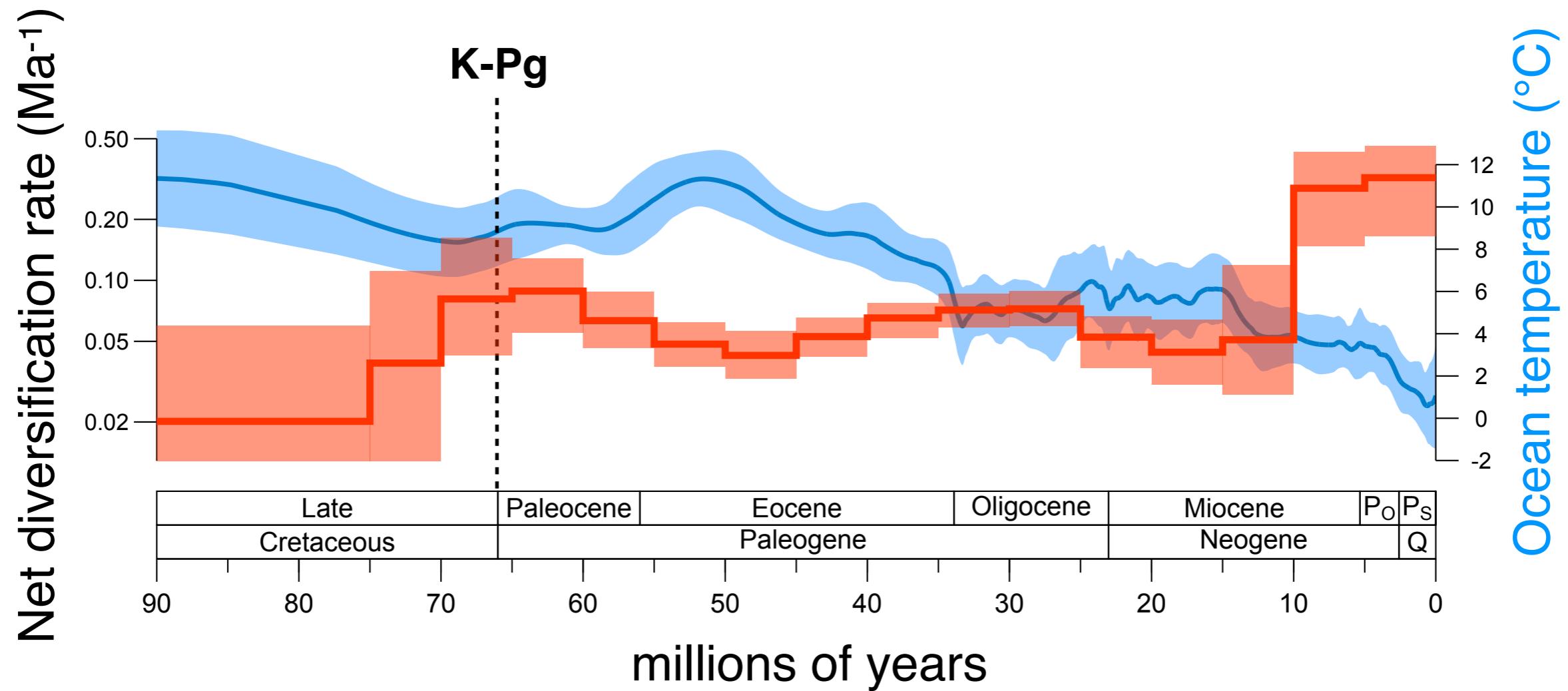
Galloanseres



Palaeognathae









2023 SSB
5TH STANDALONE MEETING
UNAM - MEXICO CITY