Systematic Biology Advance Access published May 6, 2016

Biogeographic dating of speciation times using paleogeographically informed processes

Michael J. Landis[1]

[1]*Department of Integrative Biology*
University of California, Berkeley, CA 94720-3140, U.S.A.

Michael J. Landis
University of California, Berkeley
Department of Integrative Biology
3060 VLSB #3140
Berkeley, CA 94720-3140
U.S.A.

Phone: (510) 642-1233
E-mail: mlandis@berkeley.edu

Abstract

Standard models of molecular evolution cannot estimate absolute speciation times alone, and require external calibrations to do so, such as fossils. Because fossil calibration methods rely on the incomplete fossil record, a great number of nodes in the tree of life cannot be dated precisely. However, many major paleogeographical events are dated, and since biogeographic processes depend on paleogeographical conditions, biogeographic dating may be used as an alternative or complementary method to fossil dating. I demonstrate how a time-stratified biogeographic stochastic process may be used to estimate absolute divergence times by conditioning on dated paleogeographical events. Informed by the current paleogeographical literature, I construct an empirical dispersal graph using 25 areas and 26 epochs for the past 540 Ma of Earth's history. Simulations indicate biogeographic dating performs well so long as paleogeography imposes constraint on biogeographic character evolution. To gauge whether biogeographic dating may be of practical use, I analyzed the well-studied turtle clade (*Testudines*) to assess how well biogeographic dating fares when compared to fossil-calibrated dating estimates reported in the literature. Fossil-free biogeographic dating estimated the age of the most recent common ancestor of extant turtles to be from the Late Triassic, which is consistent with fossil-based estimates. Dating precision improves further when including a root node fossil calibration. The described model, paleogeographical dispersal graph, and analysis scripts are available for use with RevBayes.

1 Introduction

Time is a simple and fundamental axis of evolution. Knowing the order and timing of evolutionary events grants us insight into how vying evolutionary processes interact. With a perfectly accurate catalog of geologically dated speciation times, many macroevolutionary questions would yield to simple interrogation, such as whether one clade exploded with diversity before or after a niche-analogous clade went extinct, or whether some number of contemporaneous biota were eradicated simultaneously by the same mass extinction event.

Only rarely does the fossil record give audience to the exact history of evolutionary events: it is infamously irregular across time, space, and species, so biologists generally resort to inference to estimate when, where, and what happened to fill those gaps. That said, we have not yet found a perfect character or model to infer dates for divergence times, so advances in dating strategies are urgently needed. A brief survey of the field reveals why.

The molecular clock hypothesis of Zuckerkandl and Pauling (1962) states that if substitutions arise (i.e. alleles fix) at a constant rate, the expected number of substitutions is the product of the substitution rate and the time the substitution process has been operating. With data from extant taxa, we only observe the outcome of the evolutionary process for an unknown rate and an unknown amount of time. In this case, rate and time are not separately identifiable, so they are estimated as their product, a compound parameter called length. If all species evolved under a single rate (a strict clock), a phylogeny with branches measured in lengths would give relative divergence times, i.e. proportional to absolute divergence times. The same is true when substitution rates vary across lineages (Wolfe et al. 1987; Martin and Palumbi 1993) and relaxed clock models are applied (Thorne et al. 1998; Drummond et al. 2006): only relative times may be estimated. Extrinsic information, e.g. a dated calibration density, is needed to establish an absolute time scale, and typically takes form as a fossil occurrence or paleogeographical event.

Fossils may be used in several ways to calibrate divergence times. The simplest method is the fossil node calibration, whereby the fossil is associated with a clade to constrain its time of origin (Ho and Phillips 2009; Parham et al. 2012). Node calibrations are empirical priors, not data-dependent stochastic processes, so they depend on experts' abilities to quantify the distribution of plausible ages for the given node. That is, node calibrations do not arise from a generative evolutionary process, so the posterior time scale is entirely determined by how the prior is specified. Rather than using prior node calibrations, fossil tip dating (Pyron 2011; Ronquist et al. 2012) treats fossil occurrences as terminal taxa with morphological characters as part of any standard phylogenetic analysis. The tree prior and model of morphological evolution are jointly applied to date speciation times from the fossils'

ages and morphological data. Introducing a generative process of fossilization, Heath et al. (2014) developed the fossilized birth-death process, by which lineages speciate, go extinct, or produce fossil observations. Using fossil tip dating with the fossilized birth-death process, Zhang et al. (2015) and Gavryushkina et al. (2015) demonstrated multiple calibration techniques may be used in tandem in a theoretically consistent framework (i.e. without introducing model violation).

Of course, fossil calibrations require fossils, but many clades leave few to no known fossils due to taphonomic processes, which filter out species with soft or fragile tissues, or with tissues that were buried in substrates that were too humid, too arid, or too rocky; or due to sampling biases, such as geographical or political biases imbalancing collection efforts (Behrensmeyer et al. 2000; Kidwell and Holland 2002). Although these biases do not prohibitively obscure the record for widespread species with robust mineralized skeletons—as is the case for many marine invertebrate and large vertebrate groups—fossil-free calibration methods are desperately needed to date the remaining majority of nodes in the tree of life.

In this direction, analogous to fossil node dating, node dates may be calibrated using paleobiogeographic scenarios (Heads 2005; Renner 2005). An ornithologist, for example, might reasonably argue that a bird known as endemic to a young island may have speciated only after the island was created, thus providing a maximum age of origin. However, using this scenario as a calibration excludes the possibility of alternative historical biogeographic explanations, e.g. the bird might have speciated off-island before the island surfaced and migrated there afterwards; see Heads (2005; 2011), Kodandaramaiah (2011), and Ho et al. (2015) for discussion on the uses and pitfalls of biogeographic node calibrations. Biogeographic node calibrations, like fossil node calibrations, fundamentally rely on some prior distribution of divergence times. This complicates model comparison since prior-encoded beliefs vary from expert to expert. Worsening matters, the time and context of biogeographic events are never directly observed, so asserting that a particular dispersal event into an island system resulted in a speciation event to calibrate a node fails to account for the uncertainty that the assumed evolutionary scenario took place at all. When estimating ancestral states, phylogenies dated

by biogeographic node calibrations should be avoided, for fear of producing falsely confident results by "double counting" the data: once when justifying the node calibrations and again when performing the biogeographic inference. Ideally, to avoid these problems, all possible biogeographic and diversification scenarios would be considered jointly, with each scenario given credence in proportion to its probability.

Inspired by advents in fossil dating models (Pyron 2011; Ronquist et al. 2012; Heath et al. 2014), which have matured from phenomenological towards mechanistic approaches (Rodrigue and Philippe 2010), I present an explicitly data-dependent and process-based biogeographic method for divergence time dating to formalize the intuition underlying biogeographic node calibrations. Analogous to fossil tip dating, the goal is to allow the observed biogeographic states at the "tips" of the tree to induce a posterior distribution of dated speciation times by way of an evolutionary process. By modeling dispersal rates between areas as subject to time-calibrated paleogeographical information, such as the merging and splitting of continental adjacencies due to tectonic drift, particular dispersal events between area-pairs are expected to occur with higher probability during certain geological time intervals than during others. For example, the dispersal rate between South America and Africa was likely to be higher when they were joined as West Gondwana (ca 120 Ma ago) than when separated as they are today. If the absolute timing of dispersal events on a phylogeny matters, then so must the absolute timing of divergence events. Unlike fossil tip dating, biogeographic dating should, in principle, be able to date speciation times only using extant taxa.

To illustrate how this is possible, I construct a toy biogeographic example to demonstrate when paleogeography may date divergence times, then follow with a more formal description of the model. By performing joint inference with molecular and biogeographic data, I demonstrate the effectiveness of biogeographic dating by applying it to simulated and empirical scenarios, showing rate and time are identifiable. While researchers have accounted for phylogenetic uncertainty in biogeographic analyses (Nylander et al. 2008; Lemey et al. 2009; Beaulieu et al. 2013), I am unaware of work demonstrating how paleogeographic calibrations may be leveraged to date divergence times via a biogeographic process. For the empirical

analysis, I date the divergence times for *Testudines* using biogeographic dating, first without any fossils, then using a fossil root node calibration. Finally, I discuss the strengths and weaknesses of the method, and how it may be improved in future work.

2 Model

2.1 The anatomy of biogeographic dating

Briefly, I will introduce an example of how time-calibrated paleogeographical events may impart information through a biogeographic process to date speciation times, then later develop the details underlying the strategy, which I refer to as *biogeographic dating*. Throughout the manuscript, I assume a rooted phylogeny with known topology but with unknown divergence times that I wish to estimate. Time is measured in geological units and as time until present, with $t = 0$ being the present, $t < 0$ being the past, and age being the negative amount of time until present. To keep the model of biogeographic evolution simple, the observed taxon occurrence matrix is assumed to be generated by a discrete-valued dispersal process where each taxon is present in only a single area at a time (Sanmartín et al. 2008). For example, taxon $T1$ might be coded to be found in Area A or Area B but not both simultaneously. Although basic, this model is sufficient to make use of paleogeographical information, suggesting more realistic models will fare better.

Consider two areas, A and B, that drift into and out of contact over time. When in contact, dispersal is possible; when not, impossible. The dispersal rate between A and B equals one when the disperal route exists, and equals zero when it does not. When A and B have a dispersal rate of zero, because the two areas are not connected by alternate dispersal routes, however circuitous, the probability of transitioning from area A to area B equals zero no matter how much time elapses. Under this construction, certain types of dispersal events are more likely to occur during certain absolute (not relative) time intervals, which potentially influences probabilities of divergence times in absolute units.

In practice, biogeographic dating requires three sources of data: molecular data to ac-

curately estimate the relative branch lengths and (if desired) topology of a rooted species tree, biogeographic data to inform the biogeographic process that in turn informs the branch lengths in absolute time, and an empirical paleogeographic model that alters the rates of biogeographic change over time. As later described in the Analysis section, these data are jointly analyzed in a Bayesian framework using RevBayes (Höhna et al. 2014), but the general concept is not wed to a particular inference methodology.

Below, I give examples of when a key divergence event is likely to precede a split event (Figures 1A,B) or to follow a merge event (Figures 1C,D). To simplify matters without compromising the general logic of the method, I assume only a single change occurs on a particular branch given the topology and tip states. In this sense, the toy examples resemble parsimony reconstructions, except they depend critically on branch length information.

[Figure 1 here.]

In the first scenario (Figures 1A,B), sister taxa $T2$ and $T3$ are found in Areas A and B, respectively. The divergence time, s, is a random variable to be inferred. At time τ, the dispersal route (A, B) is destroyed, inducing the transition probability between Areas A and B to equal zero between times τ and 0. Since $T2$ and $T3$ are found in different areas, at least one dispersal event must have occurred during an interval of non-zero dispersal probability. Then, the divergence event that gave rise to $T2$ and $T3$ must have also pre-dated τ, with at least one dispersal event occuring before the split event (Figure 1A). If $T2$ and $T3$ diverge after τ, a dispersal event from A to B is necessary to explain the observations (Figure 1B), but the model disfavors that divergence time because the required transition has probability zero. In this case, the creation of a dispersal barrier informs the latest possible divergence time, a threshold after which divergence between $T2$ and $T3$ is distinctly less probable if not impossible. It is also worth considering that a more complex process modeling vicariant speciation would provide tighter thresholds centered on τ (see Discussion).

In the second scenario (Figures 1C,D), the removal of a dispersal barrier is capable of creating a maximum divergence time threshold, pushing divergence times towards the

present. To demonstrate this, say the ingroup sister taxa $T3$ and $T4$ both inhabit Area B and the root state is Area A. Before the areas merge, the rate of dispersal between A and B is zero, and non-zero afterwards. When speciation happens after the areas merge, then the ancestor of $(T3, T4)$ may disperse from A to B, allowing $T3$ and $T4$ to inherit state B (Figure 1C). Alternatively, if $T3$ and $T4$ originate before the areas merge, then the same dispersal event on the branch ancestral to $(T3, T4)$ has probability zero (Figure 1D). By relaxing the assumptions of the toy example, the last scenario could also be explained using two independent events, one leading to $T3$ and one leading to $T4$. For most conditions, assuming the tip states and topology are given, two events will be less likely than one event.

2.2 Time-heterogeneous dispersal process

A time-heterogeneous continuous-time Markov chain (CTMC) will be used to model the formal behavior outlined in the toy example above. The CTMC is attractive as an evolutionary model because its transition probabilities may be efficiently computed for any given instantaneous rate matrix, $\mathbf{Q}$, and the amount of time a lineage has been evolving, t. Under the CTMC, the transition probability matrix, $\mathbf{P}(t)$, reports the probability that a lineage beginning in one character state ends in any particular state after time t by considering the infinite number of possible evolutionary histories consistent with the start and end states. Structuring the dispersal rates will therefore leave its impression upon the resulting dispersal probabilities.

Here, the paleogeographical features that determine the dispersal process rates are assumed to follow a piecewise constant model, sometimes called a stratified (Ree et al. 2005; Ree and Smith 2008) or epoch model (Bielejec et al. 2014), where $K - 1$ breakpoints are dated in geological time to create K time intervals. The addition and removal of dispersal routes demarcate time intervals, or *epochs*, each corresponding to some epoch index, $k \in \{1, \ldots, K\}$. Note, these epochs are time intervals defined by the model and do not necessarily coincide with geological epochs. These breakpoint times populate the vector, $\tau = (\tau_0 = -\infty, \tau_1, \tau_2, \ldots, \tau_{K-1}, \tau_K = 0)$, with the oldest interval spanning deep into the

past, and the youngest interval spanning to the present.

For a time-homogeneous CTMC, the transition probability matrix is typically written as $\mathbf{P}(t)$, which is computed for some branch length, t, using the parameters $\mathbf{Q}$, the rate matrix, and μ, the clock rate. For a piecewise constant CTMC, the value of the rate matrix, $\mathbf{Q}(k)$, changes as a function of the underlying time interval. While a lineage exists during the k^{th} time interval, its biogeographic characters evolve according to that interval's rate matrix, $\mathbf{Q}(k)$, whose rates are informed by paleogeographical features present during the epoch spanning $\tau_{k-1} < t \leq \tau_k$. As an example of a paleogeographically informed matrix's structure, take $\mathbf{G}(k)$ to be an adjacency matrix indicating 1 when dispersal may occur between two areas and 0 otherwise, during time interval k. This adjacency matrix is equivalent to an undirected graph where areas are vertices and edges are dispersal routes. Full empirical examples of $\mathbf{G} = (\mathbf{G}(1), \mathbf{G}(2), \ldots, \mathbf{G}(K))$ describing Earth's paleocontinental adjacencies are given in detail later (Section 2.4). With the paleogeographical vector $\mathbf{G}$, I define the transition rates of $\mathbf{Q}(k)$ as equal to $\mathbf{G}(k)$, with the rate matrix of each epoch rescaled to have an average rate of one. Similar rate matrices are constructed for all K time intervals that contain possible supported root ages for the phylogeny. Supplementary Figure S1 provides an example of two epochs with differing transition probability matrices.

The transition probability matrix for the piecewise constant process, $\mathbf{P}(s, t)$, is the matrix-product of the constituent epochs' time-homogeneous transition probability matrices, and takes a value determined by the absolute time and order of paleogeographical events contained between the start time, s, and end time, t. Under this construction, certain types of dispersal events are more likely to occur during certain absolute (not relative) time intervals, which potentially influences probabilities of divergence times in absolute units.

For a piecewise constant CTMC, the process's transition probability matrix is the product of transition probability matrices spanning m breakpoints. To simplify notation, let v be the vector marking important times of events, beginning with the start time of the branch, s, followed by the m breakpoints satisfying $s < \tau_k < t$, ending with the end time of the branch, t, such that $v = (s, \tau_k, \tau_{k+1}, \ldots, \tau_{k+m-1}, t)$, and let $u(v_i, \tau)$ be a "look-up" function that gives

the index k that satisfies $\tau_{k-1} < v_i \leq \tau_k$. The transition probabilty matrix over the intervals in v according to the piecewise constant CTMC given by the vectors τ and $\mathbf{Q}$ is

$$\mathbf{P}_\tau(v, \mu; \tau, \mathbf{Q}) = \prod_{i=1}^{m+1} e^{\mu(v_{i+1} - v_i)\mathbf{Q}(u(v_i, \tau))}$$

The pruning algorithm (Felsenstein 1981) is agnostic as to how the transition probabilties are computed per branch, so introducing the piecewise constant CTMC does not prohibit the efficient computation of phylogenetic model likelihoods; Bielejec et al. (2014) provides an excellent review of piecewise constant CTMCs as applied to phylogenetics.

For a time-homogeneous model where $\tau = (-\infty, 0)$, multiplying the rate and dividing the branch length by the same factor results in an identical transition probability matrix. In practice, the time-homogeneous model provides no information for the values of the clock rate and the branch lengths in the tree, since all branch rates could likewise be multiplied by some constant while branch lengths were divided by the same constant, i.e. $\mathbf{P}_\tau((s, t), 1; \tau, \mathbf{Q}) = \mathbf{P}_\tau((\mu^{-1}s, \mu^{-1}t), \mu; \tau, \mathbf{Q})$. Similarly, assuming time homogeneity, only the elapsed amount of time—but not the absolute times the process starts and ends running—affects the transition probabilities, i.e. $\mathbf{P}_\tau((s, t), \mu; \tau, \mathbf{Q}) = \mathbf{P}_\tau((s + c, t + c), \mu; \tau, \mathbf{Q})$. In contrast, the times s and t are identifiable from μ so long as $\mathbf{P}_\tau(v, \mu; \tau, \mathbf{Q}) \neq \mathbf{P}_\tau(v', \mu'; \tau, \mathbf{Q})$ for the supported values of v, μ, v', and μ' under time-heterogeneous CTMCs. Expanding from a treeless branch to the set of branches embedded in a time tree, the node ages would be identifiable so long as no two sets of node age values share identical phylogenetic model likelihoods. Barring pathological examples, the possibility that likelihood-equivalent parameters exist will generally decrease as the structural complexity of the time-heterogeneous model increases. Proving the particular conditions for which node ages are (or are not) identifiable under a time-heterogeneous model is a challenging and open question, but may be explored numerically through simulations.

The paleogeographically structured biogeographic CTMC described in this section is one example a rate-time identifiable time-heterogeneous stochastic process. More broadly, any time-heterogeneous evolutionary process with transition probabilities influenced by dated

historical observations should be capable of inducing a time-calibrated distribution of speciation times. In this framework, the quality of the estimate would largely depend on whether those influences are modeled in a realistic, justified, and principled manner.

2.3 Paleogeography, dispersal graphs, and Markov chains

Fundamentally, biogeographic dating depends on how rapidly particular types of biogeographic events occur, and how the rates of biogeographic evolution change over time. The dispersal scenarios in Figure 1 depend on the notion of *reachability*, which I develop to build intuition for how the method works. For simplicity, I will assume the dispersal rates between pairs of areas are symmetric. One area is reachable by another area so long as they are connected by some series of dispersal routes of any length, called a *path*. A path may be thought of as any potential sequence of evolutionary events that allows a lineage to disperse from one area to another, including the use of dispersal events through intermediate areas. Under a CTMC, for any positive amount of time, the transition probability between areas is positive if the areas share at least one path, and zero if they do not. This property potentially gives rise to distinct *communicating classes*. Each communicating class contains all areas that are mutually reachable, and hence share positive transition probabilities. Conversely, the transition probabilities between areas belonging to different communicating classes equal zero, as they are unreachable. Because the area adjacencies change per epoch under a paleogeographical model, so do the corresponding transition probabilities and communicating classes. Taking terrestrial biogeography as an example, areas exclusive to Gondwana or Laurasia may each reasonably form communicating classes upon the break-up of Pangaea (Figure 2), meaning lineages are free to disperse between areas within these paleocontinents, but not to and from areas on other paleocontinents. For example, the set of communicating classes is $S = \{\{\text{Afr}\}, \{\text{As}\}, \{\text{Ind}\}\}$ at $t = -100$, i.e. there are $|S| = 3$ communicating classes because no areas share edges (Figure 2C), while at $t = -10$ there is $|S| = 1$ communicating class since a path exists between all three pairs of areas (Figure 2E).

[Figure 2 here.]

In one extreme case, dispersals between mutually unreachable areas do not occur even after infinite time, and hence have zero probability. At the other extreme, when dispersal may occur between any pair of areas with equal probability over all time intervals, then paleogeography does not favor nor disfavor dispersal events (nor divergence events, implicitly) to occur during particular time intervals. In intermediate cases, where dispersal probabilities between areas vary across time intervals, the dispersal process informs when and what dispersal (and divergence) events occur. For instance, the transition probability of going from area i to j decreases as the average path length between i and j increases. During some time intervals, the average path length between two areas might be short, thus dispersal events occur more freely than when the average path is long. Comparing Figures 2A and 2E, the minimum number of events required to disperse from India to Africa is smaller during the Triassic (e.g. at $t = -250$) than during the present ($t = 0$), and thus would have a relatively higher probability given the process operated for the same amount of time today (e.g. for a branch with the same length). Interestingly, even lineages with no biogeographic variation impart some information for when a period of biogeographic stasis might begin. It will be most probable for a lineage to remain in a given area during epochs where the area is not connected to any dispersal routes. This particular signal of stasis is difficult, if not impossible, to account for under parsimony and node-based calibration methods.

Specifying communicating classes is partly difficult because we do not know the ease of dispersal between areas for most species throughout space and time. Biologically, encoding zero-valued dispersal rates directly into the model should be avoided given the apparent prevalence of long-distance dispersal, sweepstakes dispersal, etc. across dispersal barriers (Carlquist 1966). Mathematically, zero-valued rates may imply that dispersal events between certain areas are not merely improbable but utterly impossible, creating troughs of zero likelihood in the likelihood surface for certain dated-phylogeny-character patterns (Buerki et al. 2011). In a biogeographic dating framework, this might unintentionally eliminate large numbers of speciation scenarios from the space of possible hypotheses, resulting in distorted estimates. To avoid these problems, I take the dispersal graph as the weighted average of

three distinct dispersal graphs assuming short-, medium-, or long-distance dispersal modes, each with their own set of communicating classes (see Section 2.4).

The concepts of adjacency, reachability, and communicating classes are not necessary to structure the rate matrix such that biogeographic events inform divergence times, though their simplicity is attractive. One could yield similar effects by parameterizing dispersal rates as functions of more complex area features, such as geographical distance between areas (Webb and Ree 2012; Landis et al. 2013) or the size of areas (Tagliacollo et al. 2015). In this study, the concepts introduced in this section serve the practical purpose of summarizing perhaps the most salient feature of global paleogeography—that continents were not always configured as they are today—but also illuminate how time-heterogeneous dispersal rates produce transition probabilities that depend on geological time, which in turn inform the dates of speciation times.

2.4 Adjacent-area terrestrial dispersal graph

I identified $K = 26$ times and $N = 25$ areas to capture the general features of continental drift and its effects on terrestrial dispersal (Figure 3; Supplemental Figure S8). All adjacencies were constructed visually, referencing Blakey (2006) and GPlates (Boyden et al. 2011), then corroborated using various paleogeographical sources (Supplementary Table S2). The paleogeographical state per time interval is summarized as an undirected graph, where areas are vertices and dispersal routes are edges.

[Figure 3 here.]

To proceed, I treat the paleogeographical states over time as a vector of adjacency matrices, where $\mathbf{G}_\bullet(k)_{i,j} = 1$ if areas i and j share a dispersal edge at time interval k, and $\mathbf{G}_\bullet(k)_{i,j} = 0$ otherwise. A subscript $(\bullet)$ is used to distinguish adjacency matrices generated under alternative area adjacency criteria. Temporarily, I suppress the time index, k, for the rate matrix $\mathbf{Q}(k)$, since all time intervals' rate matrices are constructed in a similar manner.

To mitigate the effects of model misspecification, $\mathbf{Q}$ is determined by a weighted average of three geographical adjacency matrices

$$\mathbf{G} = b_s\mathbf{G}_s + b_m\mathbf{G}_m + b_l\mathbf{G}_l \quad (1)$$

where s, m, and l correspond to short-, medium-, and long-distance dispersal modes. The parameters b_s, b_m, and b_l sum to 1 and are to be estimated, thus weighing the importance of each dispersal mode according to the data.

Short-, medium-, and long-distance dispersal processes encode strong, weak, and no geographical constraint, respectively. As distance-constrained mode weights b_s and b_m increase, the dispersal process grows more prone to remaining within a particular communicating class (Supplementary Figure S2). The vector of short-distance dispersal graphs, $\mathbf{G}_s = (\mathbf{G}_s(1), \mathbf{G}_s(2), \ldots, \mathbf{G}_s(K))$, marks adjacencies for pairs of areas that allow terrestrial dispersal without traveling over bodies of water (Supplementary Figures S2A,B). Medium-distance dispersal graphs, $\mathbf{G}_m$, include all adjacencies in $\mathbf{G}_s$ in addition to adjacencies for areas separated by lesser bodies of water, such as throughout the Malay Archipelago, while excluding transoceanic adjacencies, such as between South America and Africa (Supplementary Figures S2C,D). Finally, long-distance dispersal graphs, $\mathbf{G}_l$, allow dispersal events to occur between any pair of areas, regardless of potential barrier (Supplementary Figures S2E,F).

To average over the three dispersal modes, b_s, b_m, and b_l are constrained to sum to 1, causing all elements in $\mathbf{G}$ to take values from 0 to 1 (Equation 1). Importantly, adjacencies specified by $\mathbf{G}_s$ always equal 1, since those adjacencies are also found in $\mathbf{G}_m$ and $\mathbf{G}_l$. This means $\mathbf{Q}$ is a Jukes-Cantor rate matrix only when $b_l = 1$, but becomes increasingly paleogeographically structured as $b_l \to 0$. Non-diagonal elements of $\mathbf{Q}$ equal those of $\mathbf{G}$, but are rescaled such that the average number of transitions per unit time is one, while diagonal elements of $\mathbf{Q}$ equal the negative sum of the remaining row elements. To compute transition probabilities, $\mathbf{Q}$ is later rescaled by a biogeographic clock rate, μ, prior to matrix exponentiation. The effects of the weights b_s, b_m, and b_l on dispersal rates between areas are

shown in Figure 4.

[Figure 4 here.]

By the argument that continental break-up (i.e. the creation of new communicating classes; Figures 1A,B) introduces a lower threshold on the minimum age of divergence, and that continental joining (i.e. unifying existing communicating classes; Figures 1C,D) introduces an upper threshold on the maximum age of divergence, then the paleogeographical model I constructed has the greatest potential to provide both upper and lower thresholds (soft bounds) on divergence times when the number of communicating classes is large, then small, then large again. This coincides with the formation of Pangaea, dropping from 8 to 3 communicating classes at 280 Ma, followed by the fragmentation of Pangaea, increasing from 3 to 11 communicating classes between 170 Ma and 100 Ma (Figure 5). It is important to consider this bottleneck in the number of communicating classes will be informative of root age only for fortuitous combinations of species range and species phylogeny. Just as some clades lack a fossil record, others are bound to lack a biogeographic record that is informative of origination times.

[Figure 5 here.]

3 Analysis

All posterior densities were estimated using Markov chain Monte Carlo (MCMC) as implemented in RevBayes available at `http://revbayes.com` (Höhna et al. 2014). Data and analysis scripts are available both on Dryad (`http://dx.doi.org/10.5061/dryad.dq666`) and on GitHub (`http://github.com/mlandis/biogeo_dating`). Analyses were performed on the XSEDE supercomputing cluster (Towns et al. 2014).

Simulation

Through simulation I tested whether biogeographic dating identifies rate from time. The simulation analyses were designed only to verify that the method was adequately well-behaved to merit further study. To proceed, I designed the analysis so divergence times are informed solely from the molecular and biogeographic data and their underlying processes (Table 1). As a convention, I use the subscript x to refer to molecular parameters and z to refer to biogeographic parameters. Specifically, I defined the molecular clock rate as $\mu_x = e/r$, where e gives the expected number of molecular substitutions per site and r gives the tree height. Both e and r are distributed independently by uniform priors on $(0, 1000)$. Biogeographic events occur with rate, $\mu_z = \mu_x 10^{s_z}$ where s_z has a uniform prior distribution on $(-3, 3)$. To further subdue effects from the prior on posterior parameter estimates, the tree prior assigns equal probability to all node age distributions. No node calibrations were used. Each dataset was analyzed with (+G) and without (–G) the paleogeographic-dependent dispersal process.

[Table 1 here.]

Two further assumptions were made to simplify the analyses. First, although divergence times were free to vary, the tree topology was assumed to be known. Second, molecular and biogeographic characters evolve by strict global clocks. In principle, inferring the topology or using relaxed clock models should increase the variance in posterior divergence time estimates, but not greatly distort the performance of –G relative to +G.

Phylogenies with $M = 50$ extant taxa were simulated using a birth-death process with birth rate, $\lambda = 0.25$, and death rate, $\mu = 0.15$, then rescaled so the root age equaled 250 Ma. Each dataset contained 500 nucleotides and 1 biogeographic character. Biogeographic characters were simulated under +G, where $\mathbf{G}$ is defined as piecewise constant over 25 areas and 26 time intervals in the manner described in Section 2.4. In total, I simulated 100 datasets under the parameter values given in Table 1, where these values were chosen to reflect possible empirical estimates. Each dataset was analyzed under each of two models,

then analyzed a second time to verify convergence (Gelman and Rubin 1992; Plummer et al. 2006). When summarizing posterior results, posterior mean-of-median and 95% highest posterior density (HPD95%) values were presented.

As expected, the results show the –G model extracts no information regarding the root age, so its posterior distribution equals its prior distribution, mean-of-median ≈ 499 (Figure 6A). In contrast, the +G model infers the mean-of-median root age 243 with a HPD95% interval width of 436, improving accuracy and precision in general.

[Figure 6 here.]

Estimated divergence time accuracy was assessed with the statistic

$$d = \sum_i \frac{a_i - a_i^{(\text{true})}}{a_i^{(\text{true})}} \tag{2}$$

where a is a posterior sample of the node age vector and $a^{(\text{true})}$ is the true node age vector known through simulation. When a perfectly estimates $a^{(\text{true})}$ for all node ages, $d = 0$. When estimated node ages are too young (on average), $d < 0$, and when too old, $d > 0$. Inference under +G infers an mean $d = 0.19$ with a HPD95% interval width of ≈ 1.26, while –G performs substantially worse with $d = 0.92$ and width ≈ 2.75 (Figure 6B). Posterior estimates generally favored short- over medium- and long-distance dispersal as was assumed under simulation (Figure 6C). Dispersal mode parameter estimates were $(b_s, b_m, b_l) = (0.766, 0.229, 0.003)$, respectively, summarized as median-of-medians across simulated replicates.

Empirical: *Testudines*

To assess the accuracy of the method, I performed a biogeographic dating analysis on extant turtle species (*Testudines*). Extant turtles fall into two clades, *Pleurodira*, found in the Southern hemisphere, and *Cryptodira*, found predominantly in the Northern hemisphere. Their modern distribution shadows their biogeographic history, where *Testudines* is thought

to be Gondwanan in origin with the ancestor to cryptodires dispersing into Laurasia during the Jurassic (Crawford et al. 2015). Since turtles are among the best-preserved vertebrates in the fossil record, their phylogeny and divergence times have been profitably analyzed and re-analyzed by various researchers (Joyce 2007; Hugall et al. 2007; Danilov and Parham 2008; Alfaro et al. 2009; Dornburg et al. 2011; Joyce et al. 2013; Sterli et al. 2013; Warnock et al. 2015). This makes them ideal to assess the efficacy of biogeographic dating, which makes no use of their replete fossil record: if both biogeography-based and fossil-based methods generate similar results, they co-validate each others' correctness (assuming they are not both biased in the same manner).

To proceed, I assembled a moderately sized dataset. First, I aligned cytochrome B sequences for 185 turtle species (155 cryptodires, 30 pleurodires) using MUSCLE 3.8.31 (Edgar 2004) under the default settings. Assuming the 25-area model presented in Section 2.4, I consulted GBIF (`gbif.org`) and IUCN Red List (`iucnredlist.org`) to record the area(s) in which each species was found. Species occupying multiple areas were assigned ambiguous tip states for those areas. Missing data entries were assigned to the six sea turtle species used in this study to effectively eliminate their influence on the (terrestrial) biogeographic process. To simplify the analysis, I assumed the species tree topology was fixed according to Guillon et al. (2012), which was chosen for its high species coverage. All speciation times were treated as random variables to be estimated. The tree topology and biogeographic states are shown in Supplementary Figure S4.

Like the simulation study, my aim is to show that the paleogeographically aware +G model identifies the root age in units of absolute time. To reiterate, the posterior root age should be identical to the prior root age when the model cannot inform the root age. If the prior and posterior differ, then the data under the model are informative. The root age was constrained to Uniform(0, 540), forbidding the existence of Precambrian turtles. To improve biological realism, I further relaxed assumptions about rate variability for the molecular model of evolution, both among sites (Yang 1994) and across branches (Lepage et al. 2007; Drummond et al. 2006) (Table 1).

[Figure 7 here.]

Biogeographic dating infers a posterior median root age of 219.0 with HPD95% credible interval of (134.9, 358.3) (Figure 7). This is consistent with current root age estimates informed from the fossil record (Figure 8). The posterior median of dispersal mode is $(b_s, b_m, b_l) = (0.42, 0.56, 0.02)$, with short- and medium-distance dispersal occurring at roughly equal rates and long-distance dispersal being rare by comparison. Biogeographic events occurred at a ratio of about 6:1 when compared to per-site molecular substitution events (posterior median: $\mu_x = 1.9\text{E}{-3}, \mu_z = 1.1\text{E}{-2}$). The posterior tree height mean measured in expected number of dispersal events is 2.3 with HPD95% (1.5, 3.0), i.e. as a clade-wide average, the current location of each taxon is the result of about two dispersal events.

[Figure 8 here.]

The flat prior distribution for competing dispersal modes is Dirichlet(1, 1, 1) and does not capture the intuition that short-distance dispersal should be far more common than long-distance dispersal. I encoded this intuition in the dispersal mode prior, setting the distribution to Dirichlet(100, 10, 1), which induces an expected proportion of 100:10:1 short:medium:long-distance dispersal events. After re-analyzing the data with the short-biased dispersal prior, the posterior median and HPD95% credible interval for the root age were estimated to be, respectively, 208.3 and (99.5, 290.2) (Figure 7).

Biogeographic dating is compatible with fossil dating methods, so I repeated the analysis for both flat and informative prior dispersal modes while substituting the Uniform(0, 540) prior on root age calibration for Uniform(155.7, 251.4), following Joyce et al. (2013). Although additional fossil node calibrations were available, they were not used to avoid inducing a non-uniform prior density for the root age (Warnock et al. 2015). When taking biogeography into account with the root age calibration, the model more strongly disfavors post-Pangaean origins for the clade than when biogeography is ignored, but the effect is mild (Figure 7). Posterior distributions of root age were relatively insensitive to the flat

and short-biased dispersal mode priors, with posterior median and credible intervals of 203.4 (163.9, 251.2) and 205.8 (165.0, 251.2), respectively. Because of the design of the model, it is unsurprising that the root node calibration prior effectively truncates the full, uncalibrated posterior density.

[Figure 9 here.]

All posterior root state estimates favored South America for the paleogeographically informed analyses (Figure 9A). Although this is in accord with the root node calibration adopted from Joyce et al. (2013)—*Caribemys oxfordiensis*, sampled from Cuba, and the oldest described crown group testudine—the fossil is described as a marine turtle, so the accordance may simply be coincidence, especially in light of the complicated history of the Caribbean Plate. In contrast, the paleogeographically naive models support a Southeast Asian origin of *Testudines*, where, incidentally, Southeast Asia is the most frequently inhabited area among the 185 testudines. For the analysis with a flat dispersal mode prior and no root age calibration, all root states with high posterior probability appear to concur upon the posterior root age density (Figure 9B), i.e. regardless of conditioning on the northern area of South America or either of the two southern areas in North America as a root state, the posterior root age density is roughly equal. Similarly, poorly supported areas of origin contribute almost nothing to the root age density. The joint-marginal density shared by the root age and root state variables bears a strong signature of a unimodality, e.g. young and old origination times do not support conflicting areas of origin. Ancestral biogeographic state estimates for the entire clade are available as Supplementary Figures S5-S7.

4 Discussion

A major obstacle preventing the probabilistic union of paleogeographical knowledge, biogeographic inference, and divergence time estimation has been methodological, which I have attempted to redress in this manuscript. The intuition justifying prior-based fossil calibrations (Parham et al. 2012), i.e. that fossil occurrences should somehow inform divergence

times, has recently been formalized into several models (Pyron 2011; Ronquist et al. 2012; Heath et al. 2014). Here I present an analogous treatment for prior-based biogeographic calibrations, i.e. that biogeographic patterns of modern species echo time-calibrated paleobiogeographic events, by describing how epoch models (Ree et al. 2005; Ree and Smith 2008; Bielejec et al. 2014) are informative of absolute divergence times. Briefly, I accomplished this using a simple time-heterogeneous dispersal process (Sanmartín et al. 2008), where dispersal rates are piecewise constant, and determined by a graph-based paleogeographical model (Section 2.4). The paleogeographical model itself was constructed by translating various published paleogeographical reconstructions (Supplementary Table S2) into a time-calibrated vector of dispersal graphs (Figure 3 and Supplementary Figure S8). Because area adjacencies are abiotic paleogeographical features, and because the biogeographic model estimates how severely paleogeographic adjacency constrains species dispersal, the model is generally appropriate for use with other terrestrial clades.

Through simulation, I showed that biogeographic dating identifies clade age from the rates of molecular and biogeographic character change. Following that, the simulation framework could easily be extended to investigate for what phylogenetic, paleogeographic, and biogeographic conditions one is able to reliably extract information for the root age. For example, a clade with taxa invariant for some biogeographic state would contain little to no information about root age, provided the area has always existed and had a constant number of dispersal edges over time. At the other extreme, a clade with a very high dispersal rate or with a proclivity towards long-distance dispersal might provide little information due to signal saturation (Supplementary Figures S2E,F). The breadth of applicability of biogeographic dating will depend critically on such factors, but because we do not expect to see closely related species uniformly distributed about Earth nor in complete sympatry, that breadth may not be so narrow.

After validating the method through simulation, I tested whether divergence times might be estimated from extant taxa in an empirical system, *Testudines*. I assumed a flat root age calibration prior for the origin time of turtles: the posterior root age was also flat when

paleogeography was ignored, but Pangaean times of origin were strongly preferred when dispersal rates conditioned on paleogeography (Figure 7). Under the uninformative prior distributions on root age, biogeographic dating estimated turtles originated between the Late Devonian (358 Ma) and Early Cretaceous (135 Ma) epochs, with a posterior median age of 219 Ma. Under an ignorance prior where short-, medium-, and long-distance dispersal events have equal prior rates, short- and medium-distance dispersal modes are strongly favored over long-distance dispersal. Posterior estimates changed little by informing the prior to strongly prefer short-distance dispersal. Both with and without root age calibrations, and with both flat and informative dispersal mode priors, biogeographic dating placed the posterior median origin time of turtles at approximately 220–200 Ma, which is consistent with fossil-based estimates (Figure 8), albeit with less precision.

The increased uncertainty in the root age estimates under biogeographic dating may be caused by any combination of possible factors. First, in the absence of fossils, biogeographic events are historical and not observed directly. Through inference, all possible biogeographic scenarios are assigned plausibility in terms of their likelihoods, which inherently introduces uncertainty that any one history occurred. By comparison, turtle fossils can be placed with relatively high phylogenetic and temporal certainty, and thus produce more precise estimates. Second, the model and priors are designed to contrast how well biogeographic dating (+G) estimates divergence times relative to a state of complete geographical ignorance (–G). For example, using a time-homogeneous birth-death process prior will cause divergence events to follow a more clock-like pattern, which is likely to be more realistic than a uniform time tree prior, which has high node age variance by design. Additional fossil node calibrations were available for use from Joyce et al. (2013), but they induce a non-uniform root age prior distribution (Warnock et al. 2015), and would have complicated the interpretation of the results. Better behaved priors are available and recommended for rigorous empirical investigations. Third, limited by the availability of genetic data, only 185 of over 300 extant testudines were used in the analysis. Increased taxon sampling generally improves ancestral state estimation (Heath et al. 2008), which will translate to improved node age estimation for

biogeographic dating. Fourth, process-based dating methods are sensitive to the underlying model assumptions (O'Reilly et al. 2015), and, in the section to come, I explore how model inadequacies might affect biogeographic dating.

Nonetheless, biogeographic dating generates relatively uncertain root age estimates when compared to previously published fossil-based results on a per-study basis. The nine fossil-based estimates were produced using diverse techniques (Figure 8): four were estimated with node dating (Warnock et al. 2015; Alfaro et al. 2009; Joyce et al. 2013; Dornburg et al. 2011), three with maximum parsimony (Sterli et al. 2013; Joyce 2007; Danilov and Parham 2008), and two with penalized likelihood rate smoothing (Hugall et al. 2007; nucleotides and amino acids). Among these studies, the average support interval width is approximately 45 Ma while their combined supported root ages are over three times as wide, ranging from 145 Ma at the youngest to 324 Ma at the oldest (width 179 Ma). On this note, at least three of the nine fossil-based root age estimates must be incorrect due to the poor support interval overlap. For dating exercises, it is worth noting that precision should not be maximized for its own sake. Uncertainty is valuable when it is correctly measured and guards against being positively incorrect. That said, the combined fossil-based support interval (145 Ma to 324 Ma) is still 31% narrower than the supported ages reported across the biogeographic dating analyses (100 Ma to 358 Ma, width 258 Ma, excluding –G analyses).

Exercises comparing performance between fossil-based and biogeographic-based dating are left for future work. However, fossil-based and biogeography-based dating methods should not be viewed as competitive, since nothing inherently prevents them from being applied simultaneously to further improve precision. For example, under the design of the earlier biogeographic dating analysis, adding a root age prior effectively truncates the posterior density (Figure 7), thus allowing fossil-based hypotheses to tune the precision of root age estimates.

For groups with poor fossil records, biogeographic dating provides a second hope for dating divergence times. Since biogeographic dating does not rely on any fossilization process or data directly, it is readily compatible with existing fossil-based dating methods (Figure 7).

When fossils with geographic information are available, researchers have shown fossil taxa improve biogeographic inferences (Moore et al. 2008; Mao et al. 2012; Nauheimer et al. 2012; Wood et al. 2012; Meseguer et al. 2015). In principle, the processes of morphological and biogeographic evolution should guide placement of fossils on the phylogeny, and the age of the fossils should improve the certainty in estimates of ancestral biogeographic states (Slater et al. 2012), on which biogeographic dating relies. A joint tip-dated and biogeography-dated analysis under the fossilized birth-death process (Zhang et al. 2015; Gavryushkina et al. 2015) would produce improved node age estimates in a methodologically consistent framework. Joint inference of divergence times, biogeography, and fossilization stands to resolve recent paleobiogeographic conundrums that may arise when considering inferences separately (Beaulieu et al. 2013; Wilf and Escapa 2014).

Model inadequacies and future extensions

The simulated and empirical studies demonstrate biogeographic dating improves divergence time estimates, with and without fossil calibrations, but many shortcomings in the model remain to be addressed. When any model is misspecified, inference is expected to produce uncertain or, worse, spurious results (Lemmon and Moriarty 2004), and biogeographic models are not exempted. Because the biogeographic model assumed in the analysis is so very simple, a rigorous battery of simulation studies must be carried out to assess the method's robustness when faced with model violation. To this end, I discuss some of the most apparent model misspecifications below.

Anagenetic range evolution models that properly allow species to inhabit multiple areas should improve the informativeness of biogeographic data. Imagine taxa $T1$ and $T2$ inhabit areas $ABCDE$ and $FGHIJ$, respectively. Under the simple model assumed in this paper, the tip states are ambiguous with respect to their ranges, and for each ambiguous state only a single dispersal event is needed to reconcile their ranges. Under a pure anagenetic range evolution model (Ree et al. 2005), at least five dispersal events are needed for reconciliation. Additionally, some extant taxon ranges may span ancient barriers, such as a terrestrial

species found both north and south of the Isthmus of Panama. When multiple-area ranges are used, this situation almost certainly requires a dispersal event to have occurred after the isthmus was formed. For single-area species ranges coded as ambiguous states, the model effectively takes the weighted average of the lineage being only north and being only south of the isthmus, so information about the effects of the paleogeographical event on divergence times will be relatively diluted.

Any model where the diversification process and paleogeographical states (and events) are correlated or co-occuring will obviously improve divergence time estimates so long as that relationship is biogeographically realistic. Although the repertoire of cladogenetic models is expanding in terms of types of transition events, they do not yet account for geographical features, such as continental adjacency or geographical distance. Incorporating paleogeographical structure into cladogenetic models of geographically isolated speciation, such as vicariance (Ronquist 1997), allopatric speciation (Ree et al. 2005; Goldberg et al. 2011), and jump dispersal (Matzke 2014), is crucial not only to generate information for biogeographic dating analyses, but also to improve the accuracy of ancestral range estimates. Ultimately, cladogenetic events are state-dependent speciation events, so the desired process would model range evolution jointly with the birth-death process (Maddison et al. 2007; Goldberg et al. 2011), but inference under these models for large state spaces is currently infeasible. Regardless, any cladogenetic range-division event requires a widespread range, which in turn implies it was preceded by dispersal (range expansion) events. Thus, if we accept that paleogeography constrains the dispersal process, even a simple dispersal-only model will extract dating information when describing a far more complex evolutionary process.

That said, the simple paleogeographical model described herein (Section 2.4) has many shortcomings itself. It is only designed for terrestrial species originating in the last 540 Ma. The number of epochs and areas was limited by my ability to comb the literature for well-supported paleogeological events, while constrained by computational considerations (see Supplementary Information). The timing of events was assumed to be known perfectly, despite the literature reporting ranges of estimates. Rates of dispersal between areas are clas-

sified into short, medium, and long distances, but using necessarily practical and subjective criteria. Factors such as temperature, precipitation, ocean currents, ecoregion type, distances between areas, sizes of areas, carrying capacities, etc. certainly affect dispersal rates between areas in terms of magnitude and symmetry, and in terms of the stationary frequencies per epoch, but were ignored. These factors may be integrated into the existing paleogeographical model so long as the data are available. From the modeling perspective, the rate matrix equation (Equation 1) can accommodate additional layers of historical features, such as those listed above, through additional weight parameters and rate matrix vectors. Regarding how to incorporate those features, Sanmartín et al. (2008) suggest how the biogeographic rate matrix may encode various celebrated colonization models (MacArthur and Wilson 1967; Hanski 1994). Finally, these colonization rates are likely to interact with the evolving life history traits intrinsic to each lineage, such as flightedness or cold tolerance, as modeled by Sukumaran et al. (2015). All of these factors can and should be handled more rigorously in future studies by modeling these processes and factors as part of a joint Bayesian analysis (Höhna et al. 2014).

Despite these flaws, defining the paleogeographical model serves as an exercise to identify what features allow a biogeographic process to inform speciation times. Identifying significant areas and epochs remains challenging, where presumably more areas and epochs are better to approximate continuous space and time, but this is not without computational challenges (Ree and Sanmartín 2009; Webb and Ree 2012; Landis et al. 2013). Dispersal barriers are clearly clade-dependent and depend on various life history traits, e.g. benthic marine species dispersal would be poorly modeled by the terrestrial graph. Classifying dispersal edges into dispersal mode classes may be made rigorous using clustering algorithms informed by paleogeographical features, or even abandoned in favor of modeling rates directly as functions of paleogeographical features like distance. Rather than fixing epoch event times to point estimates, one might assign empirical prior distributions based on collected estimates. Ideally, paleogeographical event times and features would be estimated jointly with phylogenetic evidence, which would require interfacing phylogenetic inference with paleogeo-

graphical inference. Following Sanmartín et al. (2008), multi-clade biogeographic analyses could be used to generate the statistical power necessary to obtain reliable paleogeographical estimates. This would be a profitable, but substantial, interdisciplinary undertaking.

Conclusion

Historical biogeography is undergoing a probabilistic renaissance, owing to the abundance of georeferenced biodiversity data now hosted online and the explosion of newly published biogeographic models and methods (Ree et al. 2005; Ree and Smith 2008; Sanmartín et al. 2008; Lemmon and Lemmon 2008; Lemey et al. 2010; Goldberg et al. 2011; Webb and Ree 2012; Landis et al. 2013; Matzke 2014; Tagliacollo et al. 2015; Sukumaran et al. 2015). Making use of these advances, I have shown how patterns latent in biogeographic characters, when viewed with a paleogeographic perspective, provide information about the geological timing of speciation events. The method conditions directly on biogeographic observations to induce dated node age distributions, rather than imposing (potentially incorrect) beliefs about speciation times using node calibration densities, which are data-independent prior densities. Biogeographic dating may present new opportunities for dating phylogenies for fossil-poor clades since the technique requires no fossils. This establishes that historical biogeography has untapped practical use for statistical phylogenetic inference, and should not be considered of secondary interest, only to be analysed after the species tree is estimated.

Supplementary Material

Supplementary information is available at `http://dx.doi.org/10.5061/dryad.dq666`. Datasets and RevBayes analysis scripts are also available at `http://github.com/mlandis/biogeo_dating`.

Acknowledgements

I thank Tracy Heath, Sebastian Höhna, Josh Schraiber, Lucy Chang, Pat Holroyd, Nick Matzke, Michael Donoghue, and John Huelsenbeck for valuable feedback, support, and encouragement regarding this work. I give my special thanks to James Albert, Alexandre Antonelli, and the Society of Systematic Biologists for inviting me to present this research at Evolution 2015 in Guarujá, Brazil. Frank Anderson, Emma Goldberg, and three anonymous reviewers provided feedback that greatly improved the clarity of the manuscript.

Funding

MJL was supported by a National Evolutionary Synthesis Center Graduate Fellowship and a National Institutes of Health grant (R01-GM069801) awarded to John P. Huelsenbeck. Simulation analyses were computed using XSEDE, which is supported by National Science Foundation grant number ACI-1053575.

References

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale, and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proceedings of the National Academy of Sciences 106:13410–13414.

Beaulieu, J. M., D. C. Tank, and M. J. Donoghue. 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. BMC Evolutionary Biology 13:80.

Behrensmeyer, A. K., S. M. Kidwell, and R. A. Gastaldo. 2000. Taphonomy and paleobiology. Paleobiology 26:103–147.

Bielejec, F., P. Lemey, G. Baele, A. Rambaut, and M. A. Suchard. 2014. Inferring heterogeneous evolutionary processes through time: from sequence substitution to phylogeography. Systematic Biology 63:493–504.

Blakey, R. 2006. Global paleogeographic views of earth history: Late Precambrian to Recent. R. Blakey.

Boyden, J. A., R. D. Müller, M. Gurnis, T. H. Torsvik, J. A. Clark, M. Turner, H. Ivey-Law, R. J. Watson, and J. S. Cannon. 2011. Next-generation plate-tectonic reconstructions using GPlates. Geoinformatics: cyberinfrastructure for the solid earth sciences Pages 95–114.

Buerki, S., F. Forest, N. Alvarez, J. A. A. Nylander, N. Arrigo, and I. Sanmartín. 2011. An evaluation of new parsimony-based versus parametric inference methods in biogeography: a case study using the globally distributed plant family Sapindaceae. Journal of Biogeography 38:531–550.

Carlquist, S. 1966. The biota of long-distance dispersal. I. Principles of dispersal and evolution. Quarterly Review of Biology Pages 247–270.

Crawford, N. G., J. F. Parham, A. B. Sellas, B. C. Faircloth, T. C. Glenn, T. J. Papenfuss, J. B. Henderson, M. H. Hansen, and B. W. Simison. 2015. A phylogenomic analysis of turtles. Molecular Phylogenetics and Evolution 83:250–257.

Danilov, I. G. and J. F. Parham. 2008. A reassessment of some poorly known turtles from the Middle Jurassic of China, with comments on the antiquity of extant turtles. Journal of Vertebrate Paleontology 28:306–318.

Dornburg, A., J. M. Beaulieu, J. C. Oliver, and T. J. Near. 2011. Integrating fossil preservation biases in the selection of calibrations for molecular divergence time estimation. Systematic Biology 60:519–527.

Drummond, A. J., S. Y. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. PLoS Biology 4:e88.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. Journal of Molecular Evolution 17:368–376.

Gavryushkina, A., T. A. Heath, D. T. Ksepka, T. Stadler, D. Welch, and A. J. Drummond. 2015. Bayesian total evidence dating reveals the recent crown radiation of penguins. arXiv preprint arXiv:1506.04797 .

Gelman, A. and D. B. Rubin. 1992. Inferences from iterative simulation using multiple sequences. Statistical Science 7:457–511.

Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal effects between geographic range evolution and diversification. Systematic Biology 60:451–465.

Guillon, J.-M., L. Guéry, V. Hulin, and M. Girondot. 2012. A large phylogeny of turtles (Testudines) using molecular data. Contributions to Zoology 81:147–158.

Gurnis, M., M. Turner, S. Zahirovic, L. DiCaprio, S. Spasojevic, R. D. Müller, J. Boyden, M. Seton, V. C. Manea, and D. J. Bower. 2012. Plate tectonic reconstructions with continuously closing plates. Computers & Geosciences 38:35–42.

Hanski, I. 1994. A practical model of metapopulation dynamics. Journal of Animal Ecology Pages 151–162.

Heads, M. 2005. Dating nodes on molecular phylogenies: a critique of molecular biogeography. Cladistics 21:62–78.

Heads, M. 2011. Old taxa on young islands: a critique of the use of island age to date island-endemic clades and calibrate phylogenies. Systematic Biology 60:204–218.

Heath, T. A., S. M. Hedtke, and D. M. Hillis. 2008. Taxon sampling and the accuracy of phylogenetic analyses. J Syst Evol 46:239–257.

Heath, T. A., J. P. Huelsenbeck, and T. Stadler. 2014. The fossilized birth–death process for coherent calibration of divergence-time estimates. Proceedings of the National Academy of Sciences 111:E2957–E2966.

Ho, S. Y., K. J. Tong, C. S. Foster, A. M. Ritchie, N. Lo, and M. D. Crisp. 2015. Biogeographic calibrations for the molecular clock. Biology letters 11:20150194.

Ho, S. Y. W. and M. J. Phillips. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology 58:367–380.

Höhna, S., T. A. Heath, B. Boussau, M. J. Landis, F. Ronquist, and J. P. Huelsenbeck. 2014. Probabilistic graphical model representation in phylogenetics. Systematic Biology 63:753–771.

Hugall, A. F., R. Foster, and M. S. Y. Lee. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. Systematic Biology 56:543–563.

Joyce, W. G. 2007. Phylogenetic relationships of Mesozoic turtles. Bulletin of the Peabody Museum of Natural History 48:3–102.

Joyce, W. G., J. F. Parham, T. R. Lyson, R. C. M. Warnock, and P. C. J. Donoghue. 2013. A divergence dating analysis of turtles using fossil calibrations: an example of best practices. Journal of Paleontology 87:612–634.

Kidwell, S. M. and S. M. Holland. 2002. The quality of the fossil record: implications for evolutionary analyses. Annual Review of Ecology and Systematics Pages 561–588.

Kodandaramaiah, U. 2011. Tectonic calibrations in molecular dating. Current Zoology 57:116–124.

Landis, M. J., N. J. Matzke, B. R. Moore, and J. P. Huelsenbeck. 2013. Bayesian analysis of biogeography when the number of areas is large. Systematic Biology 62:789–804.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. Bayesian phylogeography finds its roots. PLoS Computational Biology 5:e1000520.

Lemey, P., A. Rambaut, J. J. Welch, and M. A. Suchard. 2010. Phylogeography takes a relaxed random walk in continuous space and time. Moleculate Biology and Evolution 27:1877–1885.

Lemmon, A. A. and E. M. Lemmon. 2008. A likelihood framework for estimating phylogeographic history on a continuous landscape. Systematic Biology 57:544–561.

Lemmon, A. R. and E. C. Moriarty. 2004. The importance of proper model assumption in Bayesian phylogenetics. Systematic Biology 53:265–277.

Lepage, T., D. Bryant, H. Philippe, and N. Lartillot. 2007. A general comparison of relaxed molecular clock models. Molecular Biology and Evolution 24:2669–2680.

MacArthur, R. H. and E. O. Wilson. 1967. The theory of island biogeography. Princeton University Press, New Jersey.

Maddison, W. P., P. E. Midford, and S. P. Otto. 2007. Estimating a binary character's effect on speciation and extinction. Systematic Biology 56:701–710.

Mao, K., R. I. Milne, L. Zhang, Y. Peng, J. Liu, P. Thomas, R. R. Mill, and S. S. Renner. 2012. Distribution of living Cupressaceae reflects the breakup of Pangea. Proceedings of the National Academy of Sciences 109:7793–7798.

Martin, A. P. and S. R. Palumbi. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Sciences 90:4087–4091.

Matzke, N. J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Systematic Biology Page syu056.

Meseguer, A. S., J. M. Lobo, R. Ree, D. J. Beerling, and I. Sanmartín. 2015. Integrating Fossils, Phylogenies, and Niche Models into Biogeography to Reveal Ancient Evolutionary History: The Case of Hypericum (Hypericaceae). Systematic Biology 64:215–232.

Moore, B. R., S. A. Smith, R. H. Ree, and M. J. Donoghue. 2008. Incorporating fossil data in biogeographic inference: A likelihood approach. `www.phylo.org/pdf_docs/2006_MoBot_biogeo.pdf` .

Nauheimer, L., D. Metzler, and S. S. Renner. 2012. Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. New Phytologist 195:938–950.

Nylander, J. A., U. Olsson, P. Alström, and I. Sanmartín. 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: Turdus). Systematic Biology 57:257–268.

O'Reilly, J. E., M. dos Reis, and P. C. J. Donoghue. 2015. Dating Tips for Divergence-Time Estimation. Trends in Genetics 31:637–650.

Parham, J. F., P. C. J. Donoghue, C. J. Bell, T. D. Calway, J. J. Head, P. A. Holroyd, J. G. Inoue, R. B. Irmis, W. G. Joyce, D. T. Ksepka, J. S. L. Patané, N. D. Smith, J. E. Tarver,

M. van Tuinen, Z. Yang, K. D. Angielczyk, J. M. Greenwood, C. A. Hipsley, L. Jacobs, P. J. Makovicky, J. Müller, K. T. Smith, J. M. Theodor, and R. C. M. Warnock. 2012. Best practices for justifying fossil calibrations. Systematic Biology 61:346.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence diagnosis and output analysis for MCMC. R News Pages 7–11.

Pyron, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. Systematic Biology 60:466–481.

Ree, R. H., B. R. Moore, C. O. Webb, and M. J. Donoghue. 2005. A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Evolution 59:2299–2311.

Ree, R. H. and I. Sanmartín. 2009. Prospects and challenges for parametric models in historical biogeographical inference. Journal of Biogeography 36:1211–1220.

Ree, R. H. and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57:4–14.

Renner, S. S. 2005. Relaxed molecular clocks for dating historical plant dispersal events. Trends in Plant Science 10:550–558.

Rodrigue, N. and H. Philippe. 2010. Mechanistic revisions of phenomenological modeling strategies in molecular evolution. Trends in Genetics 26:248–252.

Ronquist, F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Systematic Biology 46:195–203.

Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P. Rasnitsyn. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Systematic Biology 61:973–999.

Sanmartín, I., P. V. D. Mark, and F. Ronquist. 2008. Inferring dispersal: a Bayesian approach to phylogeny-based island biogeography, with special reference to the Canary Islands. Journal of Biogeography 35:428–449.

Slater, G. J., L. J. Harmon, and M. E. Alfaro. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. Evolution 66:3931–3944.

Sterli, J., D. Pol, and M. Laurin. 2013. Incorporating phylogenetic uncertainty on phylogeny-based palaeontological dating and the timing of turtle diversification. Cladistics 29:233–246.

Sukumaran, J., E. P. Economo, and L. L. Knowles. 2015. Machine learning biogeographic processes from biotic patterns: A new trait-dependent dispersal and diversification model with model choice by simulation-trained discriminant analysis. Systematic Biology Page syv121.

Tagliacollo, V. A., S. M. Duke-Sylvester, W. A. Matamoros, P. Chakrabarty, and J. S. Albert. 2015. Coordinated Dispersal and Pre-Isthmian Assembly of the Central American Ichthyofauna. Systematic Biology Page syv064.

Thorne, J., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution 15:1647–1657.

Towns, J., T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, et al. 2014. Xsede: Accelerating scientific discovery. Computing in Science & Engineering 16:62–74.

Warnock, R. C., J. F. Parham, W. G. Joyce, T. R. Lyson, and P. C. J. Donoghue. 2015. Calibration uncertainty in molecular dating analyses: there is no substitute for the prior evaluation of time priors. Proceedings of the Royal Society of London B: Biological Sciences 282:20141013.

Webb, C. O. and R. H. Ree. 2012. Historical biogeography inference in Malesia. Pages 191–215 *in* Biotic evolution and environmental change in Southeast Asia (D. Gower, K. Johnson, J. Richardson, B. Rosen, L. Ruber, and S. Williams, eds.) Cambridge University Press.

Wilf, P. and I. H. Escapa. 2014. Green Web or megabiased clock? Plant fossils from Gondwanan Patagonia speak on evolutionary radiations. New Phytologist 207:283–290.

Wolfe, K. H., W.-H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proceedings of the National Academy of Sciences 84:9054–9058.

Wood, H. M., N. J. Matzke, R. G. Gillespie, and C. E. Griswold. 2012. Treating fossils as terminal taxa in divergence time estimation reveals ancient vicariance patterns in the palpimanoid spiders. Systematic Biology 62:264–284.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. Journal of Molecular Evolution 39:306–314.

Zhang, C., T. Stadler, S. Klopfstein, T. A. Heath, and F. Ronquist. 2015. Total–evidence dating under the fossilized birth–death process. Systematic Biology Page syv080.

Zuckerkandl, E. and L. Pauling. 1962. Molecular disease, evolution, and genetic heterogeneity. Pages 189–225 *in* Horizons in Biochemistry (M. Kasha and B. Pullman, eds.) Academic Press, New York.

Tables

			Simulation		**Empirical**
	Parameter	X	$f(X)$	sim. value	$f(X)$
Tree	Root age	r	Uniform$(0, 1000)$	250	Uniform$(0, 540)$ or Uniform$(155.7, 251.4)$
	Time tree	Ψ	UniformTimeTree(r)	BD($\lambda = 0.25$,$\mu = 0.15$)	UniformTimeTree(r)
Molecular	Length	e	Uniform$(0, 1000)$	2.5	Uniform$(0, 1000)$
	Subst. rate	μ_x	e/r	determined (0.01)	e/r
	Exch. rates	r_x	Dirichlet(10)	from prior	Dirichlet(10)
	Stat. freqs	π_x	Dirichlet(10)	from prior	Dirichlet(10)
	Rate matrix	$\mathbf{Q}_x$	GTR(r_x, π_x)	determined	GTR(r_x, π_x)
	Branch rate mult.	$\rho_{x,i}$			Lognorm$(\ln\mu_x - \sigma_x^2/2, \sigma_x)$
	Branch rate var.	σ_x			Exponential(0.1)
	$+\Gamma 4$	Γ_x			Gamma(α, α)
	$+\Gamma 4$ hyperprior	α			Uniform$(0, 50)$
Biogeo.	Atlas-graph	$G(t)$	–G or +G	+G	+G
	Biogeo. rate	μ_z	$\mu_x 10^{s_z}$	determined (0.1)	$\mu_x 10^{s_z}$
	Biogeo. rate mod.	s_z	Uniform(-3, 3)	1.0	Uniform(-3, 3)
	Dispersal mode	(b_s, b_m, b_l)	Dirichlet(1)	(1000,10,1)/1011	Dirichlet$(1, 1, 1)$ or Dirichlet$(100, 10, 1)$
	Dispersal rates	$r_z(t)$	$\sum_{\{s,m,l\}} b_i G_i(t)$	determined	$\sum_{\{s,m,l\}} b_i G_i(t)$
	Stat. freqs	π_z	$(1, \ldots, 1)/25$	$(1, \ldots, 1)/25$	$(1, \ldots, 1)/25$
	Rate matrix	$\mathbf{Q}_z(t)$	GTR$(r_z(t), \pi_z)$	determined	GTR$(r_z(t), \pi_z)$

Table 1: **Model parameters.** Model parameter names and prior distributions are described in the manuscript body. All empirical priors were identical to simulated priors unless otherwise stated. Priors used for the empirical analyses but not simulated analyses are left blank. Determined means the parameter value was determined by other model parameters.

Figures

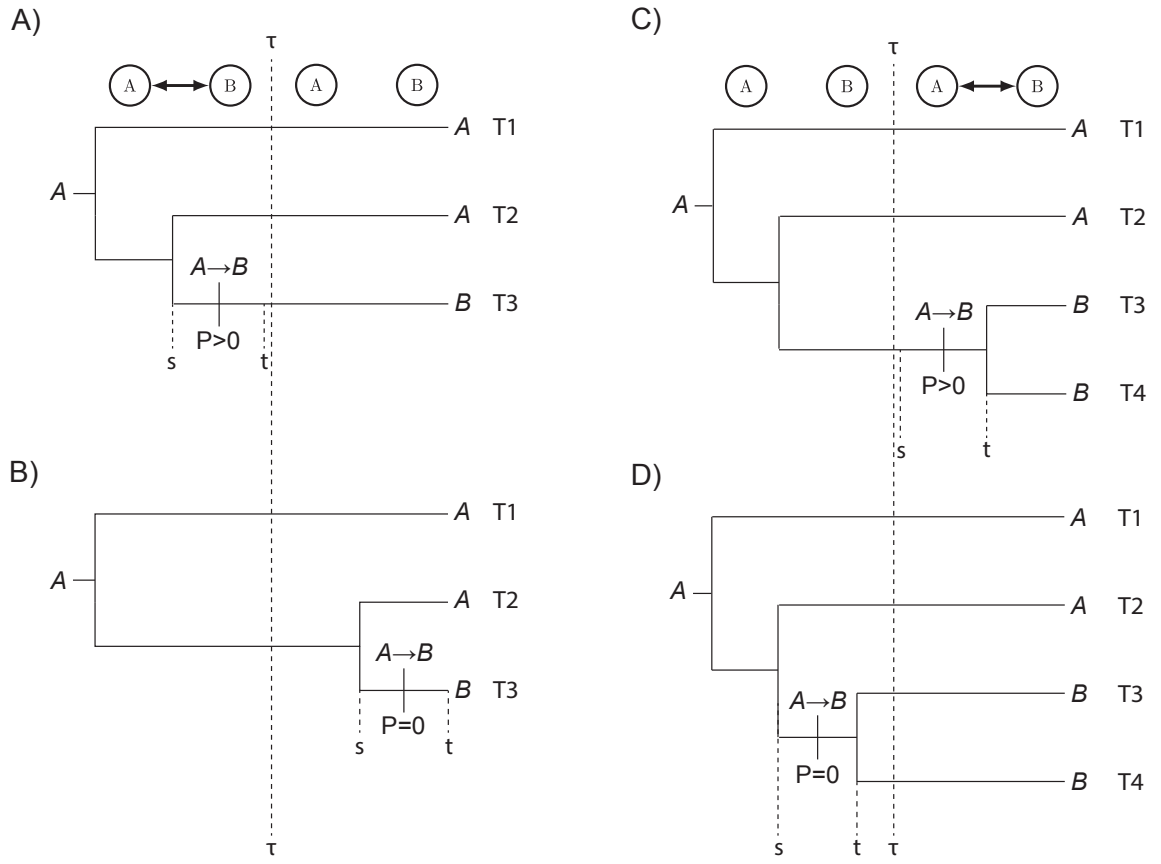


Figure 1: **Biogeographic dating toy example.** Two paleogeographic scenarios are given: a split event (A and B) and a merge event (C and D). Scenarios assume only a single dispersal event occurs for a given branch (see text). A, B) $T1$ and $T2$ have state A and the transition $A \rightarrow B$ most parsimoniously explains how $T3$ has state B. The transition probabilty for $P = [\mathbf{P}(s,t)]_{AB}$ is non-zero before the paleogeographical split event at time τ, and is zero afterwards. Two possible divergence and dispersal times are given: A) $T3$ originates before the split when the transition $A \rightarrow B$ has non-zero probability. B) $T3$ originates after the split when the transition $A \rightarrow B$ has probability zero. C, D) $T1$ and $T2$ have the state A and the transition $A \rightarrow B$ on the lineage leading to $(T3, T4)$ most parsimoniously explains how $T3$ and $T4$ have state B. The transition probabilty for $P = [\mathbf{P}(s,t)]_{AB}$ is zero before the paleogeographical merge event at time τ, and only non-zero afterwards. Two possible divergence and dispersal times are given: C) $T3$ and $T4$ originate after the merge when the transition $A \rightarrow B$ has non-zero probability. D) $T3$ and $T4$ originate before the merge when the transition $A \rightarrow B$ has probability zero.

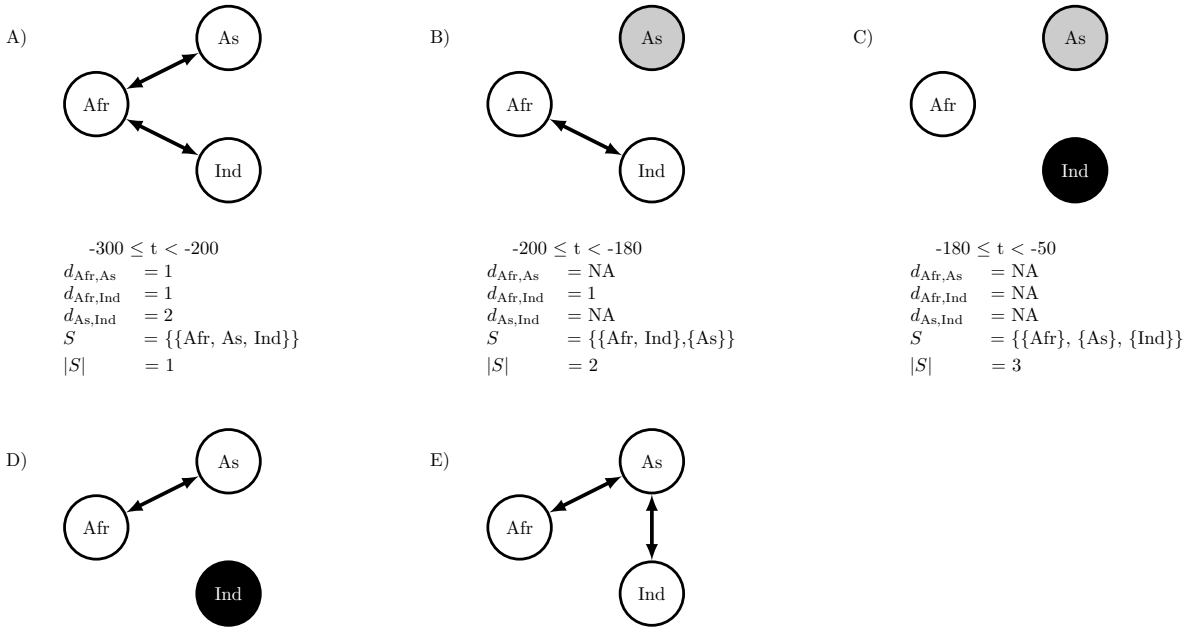


Figure 2: **Biogeographic communicating classes.** Dispersal routes shared by Africa (Afr), Asia (As), and India (Ind) are depicted for each time interval, t, over the past 300 Ma. Dispersal path lengths between areas i and j are given by $d_{i,j}$, with NA meaning there is no route between areas (areas i and j are mutually unreachable). Communicating classes per interval are given by S and by the shared coloring of areas (vertices), with $|S|$ being the number of communicating classes.

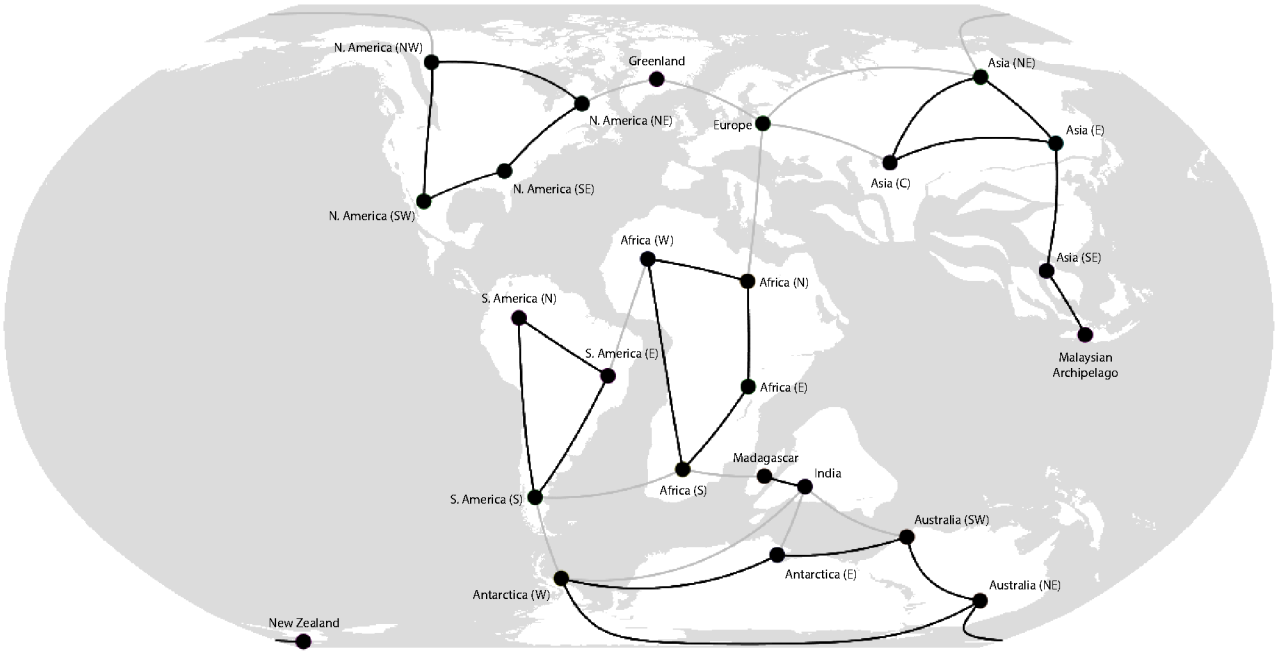
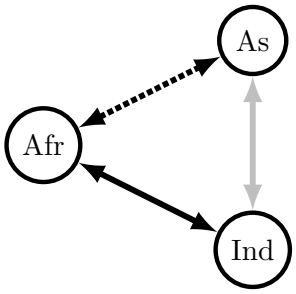


Figure 3: **Dispersal graph for Epoch 14, 110–100Ma: India and Madagascar separate from Australia and Antarctica.** A `gplates` (Gurnis et al. 2012) screenshot of Epoch 14 of 26 is displayed. Areas are marked by black vertices. Black edges indicate both short- and medium-distance dispersal routes. Gray edges indicate exclusively medium-distance dispersal routes. Long-distance dispersal routes are not shown, but are implied to exist between all area-pairs. The short-, medium-, and long-distance dispersal graphs have 8, 1, and 1 communicating classes, respectively. Only India may reach Madagascar by short-distance dispersal, and vice versa, forming one of the eight short-distance communicating classes. Both areas maintain medium-distance dispersal routes with various Gondwanan continents during this epoch. The expansion of the Tethys Sea impedes dispersal into and out of Europe. Section 2.3 explains how communicating classes are determined.



$$b_s + b_m + b_l = 1.0$$

$$\begin{aligned} q_{\text{Afr,Ind}} &= 0.7 + 0.2 + 0.1 = 1.0 \\ q_{\text{Afr,As}} &= 0.2 + 0.1 = 0.3 \\ q_{\text{As,Ind}} &= 0.1 = 0.1 \end{aligned}$$

$$\mathbf{Q} = \begin{bmatrix} - & 0.3 & 1.0 \\ 0.3 & - & 0.1 \\ 1.0 & 0.1 & - \end{bmatrix}$$

Figure 4: **Example mode-weighted dispersal matrix.** Short-, medium-, and long-distance dispersal edges are represented by solid black, dashed black, and solid gray lines, respectively. Short-, medium-, and long-distance dispersal weights are $(b_s, b_m, b_l) = (0.7, 0.2, 0.1)$. The resulting mode-weighted dispersal matrix, $\mathbf{Q}$, is computed with areas (states) ordered as (Afr, As, Ind). Afr and Ind share a short-distance dispersal edge, therefore the dispersal weight is $b_s + b_m + b_l = 1.0$. Afr and As share a medium-distance edge with dispersal weight $b_m + b_l = 0.3$. Dispersal between As and Ind occurs over a long-distance dispersal edge with weight $b_l = 0.1$.

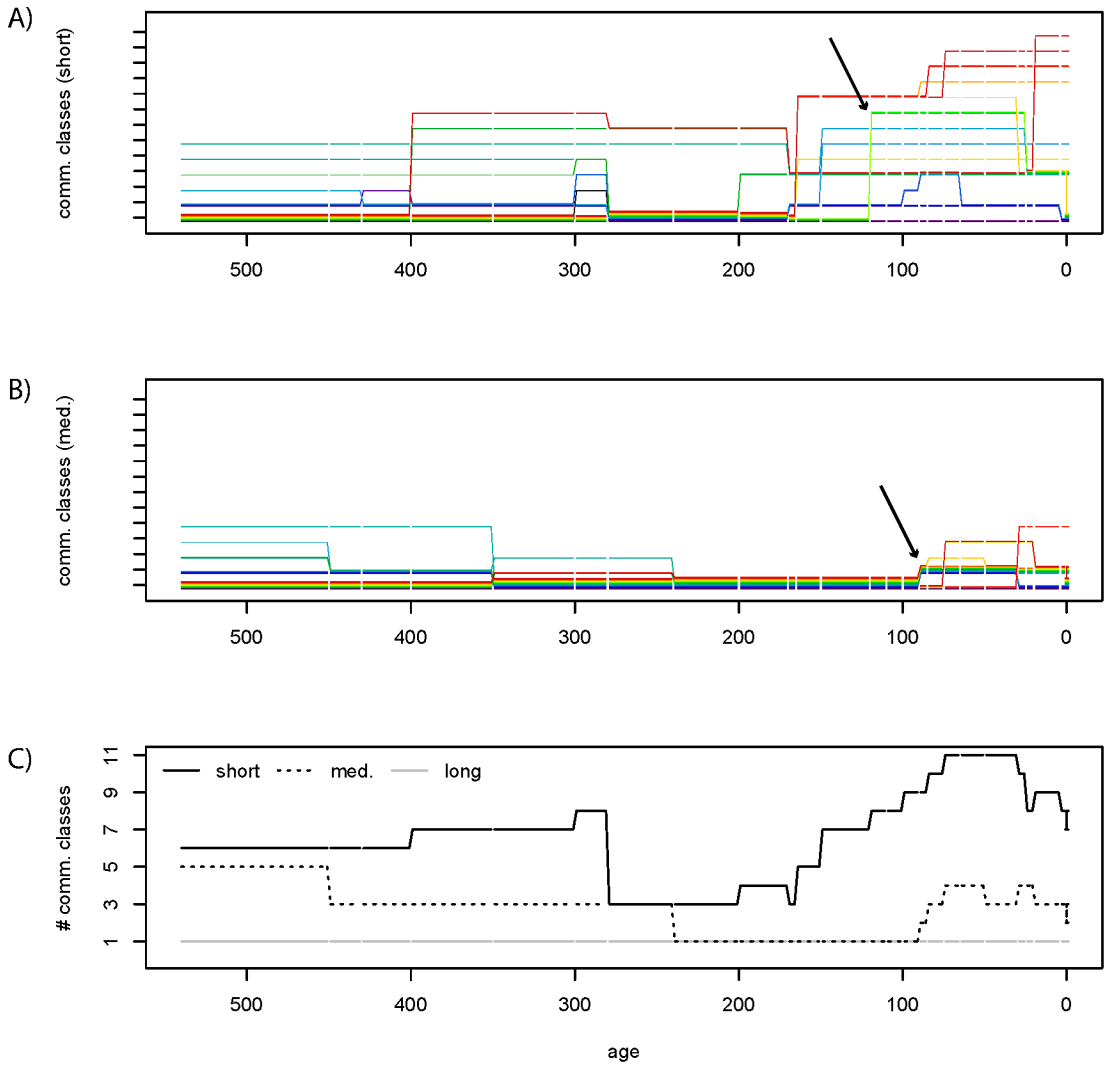


Figure 5: **Dispersal graph properties summarized over time.** Communicating classes of the short-distance dispersal graph (A) and medium-distance dispersal graph (B) are shown. Each of 25 areas is represented by one horizontal line. Colors of areas match those listed in Supplementary Figure S2 (online only). On the vertical axis, areas that share a communicating class are grouped into horizontal bands (multiple lines) whose width is proportional to the number of areas in the communicating class. Vertical lines running between horizontal bands indicate transitions of areas joining or leaving communicating classes, i.e. due to paleogeographical events. Communicating classes are arranged vertically to appear horizontally stable with respect to time, but their vertical order and relative positioning is otherwise arbitrary. When no transition event occurs for an area entering a new epoch, the line is interrupted with a gap. The two arrows mark the break-up of Gondwana, first recorded in the short-distance dispersal graph (A) at 120 Ma, then later in the medium-distance dispersal graph (B) at 90 Ma. C) Number of communicating classes: the black line corresponds to the short-distance dispersal graph (A), the dotted line corresponds to medium-distance dispersal graph (B), and the gray line corresponds to the long-distance dispersal graph, which always has one communicating class. Section 2.3 explains how communicating classes are determined.

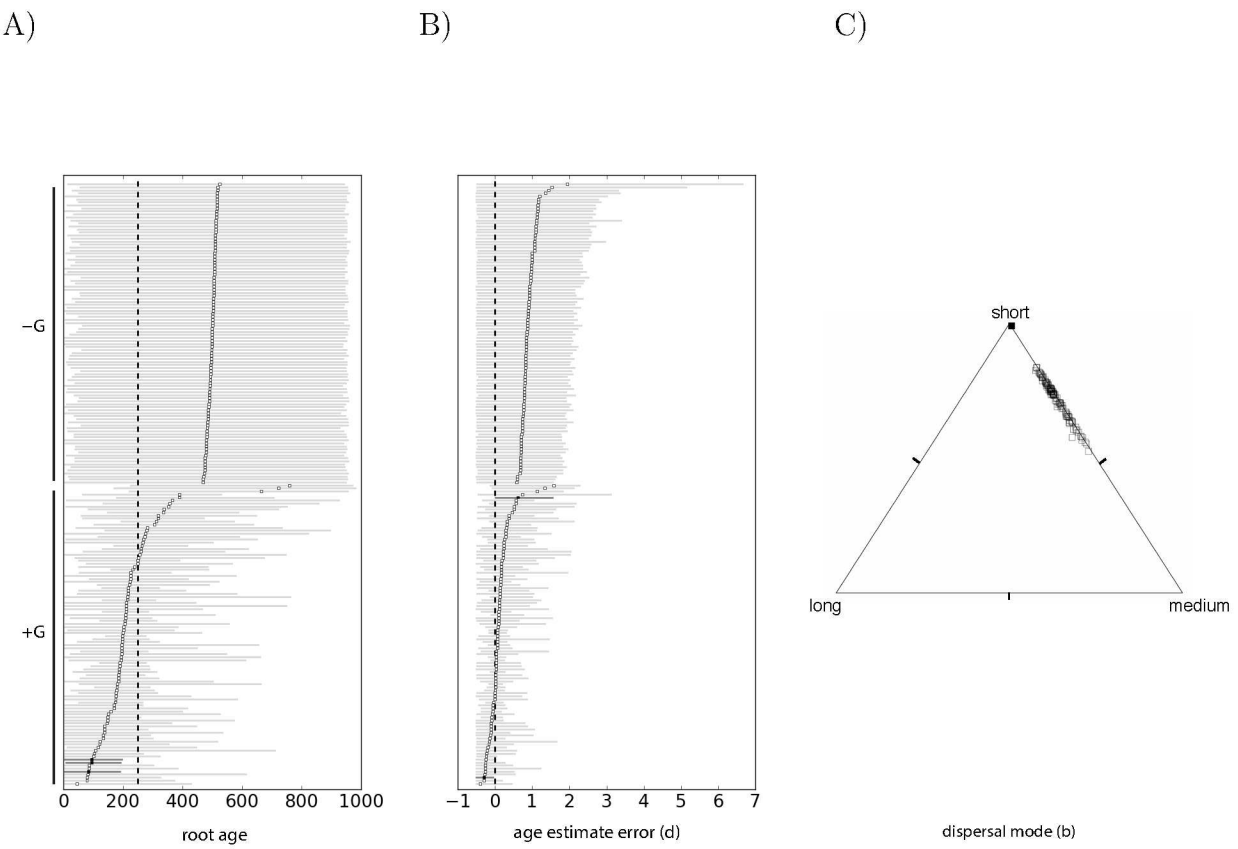


Figure 6: **Posterior estimates for simulated data.** A) Posterior estimates of root age. The true root age for all simulations is 250 Ma (dotted vertical line). B) Posterior estimates of relative node age error (Eqn 2). The true error term equals zero. Both A and B) paleogeographically non-informed (–G) analyses are on the top half, and paleogeographically informed (+G) analyses are on the bottom. Each square marks the posterior mean root age estimate with the HPD95% credible interval. Estimates whose credible interval did not contain the true value are darkened for visibility. C) Posterior estimates of dispersal mode proportions for the +G simulations projected onto the unit 2-simplex. The filled square (near "short") gives the true value, and the empty circles give posterior medians.

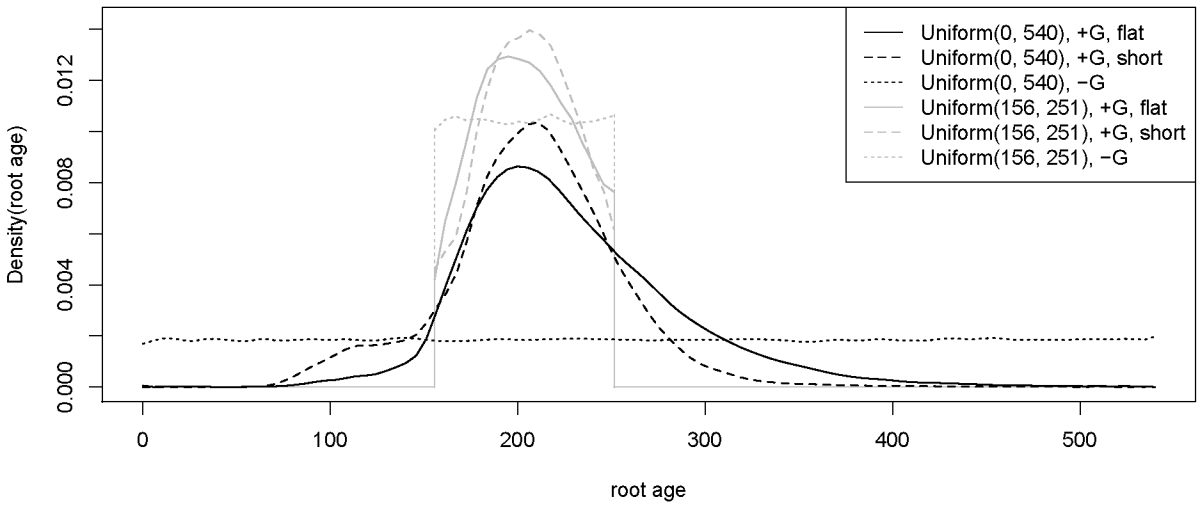


Figure 7: **Posterior root age of turtles by biogeographic dating.** Six root age posterior estimates were computed using biogeographic dating, each using variations on flat or short-biased priors for dispersal mode weight parameters. Black densities assumed no knowledge of fossils with a Uniform(0, 540) root age prior. Gray densities follow Joyce et al. (2013) and assumed Uniform(155.7, 251.4) as a root node age calibration. Line types indicate priors on dispersal mode (solid: flat prior, dashed: short-biased, dotted: paleogeography ignored).

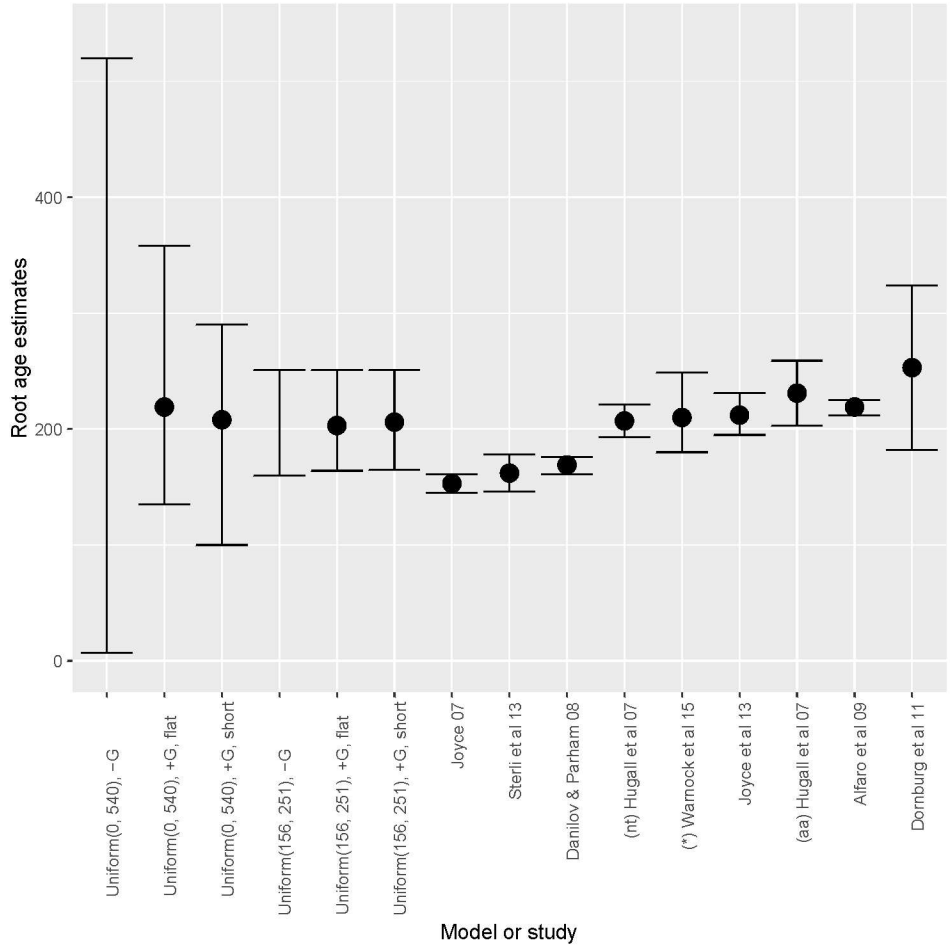


Figure 8: **Root age comparison.** Root age estimates are presented both for analysis conducted for this manuscript and as reported in the literature. Existing estimates are as reported in Sterli et al. (2013) and supplemented recently reported results. Points and whiskers correspond to the point estimates their confidence intervals. The six left estimates were computed using biogeographic dating, each using variations on flat or short-biased priors for key parameters. Two of these analyses ignore paleogeography (–G) so the posterior root age is the uniform prior root age, whose mode (not shown) equals all values supported by the prior. Hugall et al. (2007) reports ages for analyses using amino acids (aa) and nucleotides (nt). Warnock et al. (2015) reports many estimates while exploring prior sensitivity, but only uniform prior results are shown here.

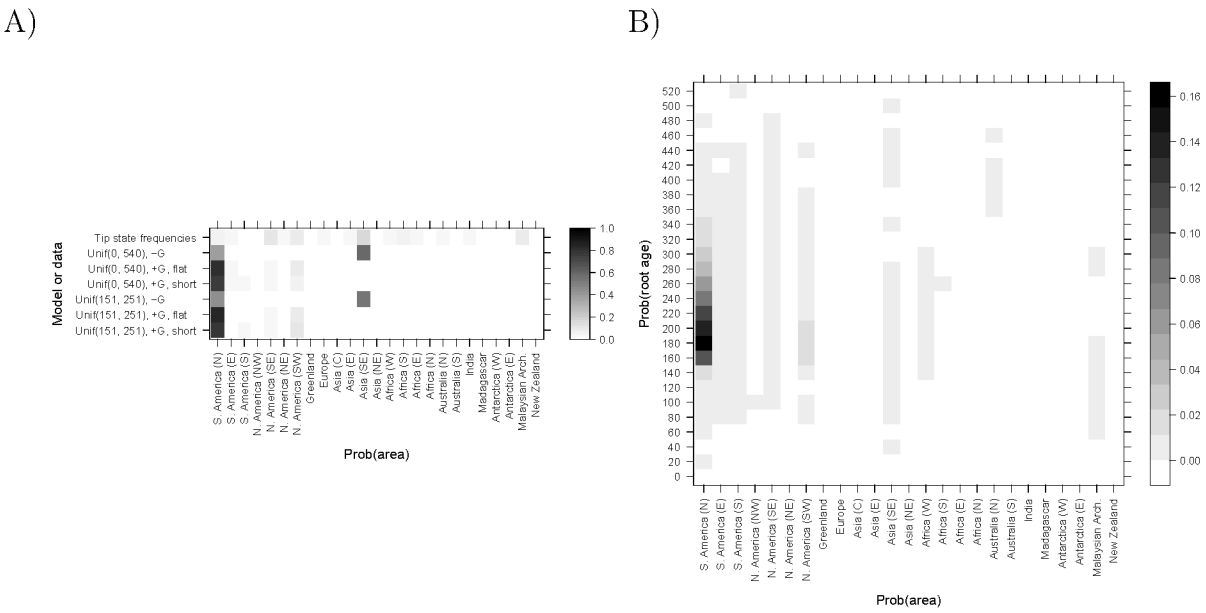


Figure 9: **Root state estimates.** A) Posterior probabilities of root state are given for the six empirical analyses. B) Joint-marginal posterior probabilities of root age and root state are given for the empirical analysis without a root calibration and with a flat dispersal mode prior. Root ages are binned into intervals of width 20.