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. Because fossil calibration methods rely on the unreliable fossil record, most nodes in the tree of life are dated with poor accuracy. 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 have any practical use, I analyze the well-studied turtle clade (*Testudines*) then assess how well biogeographic dating fares compared to heavily fossil-calibrated dating results as reported in the literature. Fossil-free biogeographic dating estimated the age of the most recent common ancestor of extant turtles to be approximately 201 Ma, which is consistent with fossil-based estimates. Accuracy 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. As such, rate and time are non-identifiable under standard models of molecular substitution, so inferred amounts of evolutionary change are often reported as a compound parameter, the product of rate and time, called length. If all species' shared the same substitution rate, a phylogeny with branches measured in lengths would give relative divergence times, i.e. proportional to absolute divergence times. While it is reasonable to say species' evolution shares a basis in time, substitution rates differ between species and over macroevolutionary timescales (Wolfe et al. 1987; Martin and Palumbi 1993). Even when imposing a model of rate heterogeneity (Thorne et al. 1998; Drummond et al. 2006), only relative times may be estimated. Extrinsic information, i.e. a dated calibration point, is needed to establish an absolute time scaling, and typically takes form as a fossil occurrence or paleogeographical event.

When fossils are available, they currently provide the most accurate inroad to calibrate divergence events to geological timescales, largely because each fossil is associated with a geological occurrence time. Fossil ages 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 and constrains its time of origin (Ho and Phillips 2009; Parham et al. 2011). Node calibrations are empirical priors, not data-dependent stochastic processes, so they depend entirely 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. Since

molecular phylogenies cannot identify rate from time, then the time scaling is entirely determined by the prior, i.e. the posterior is perfectly prior-sensitive for rates and times. 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. In this case, the model of character evolution, tree prior, and fossil ages generate the distribution of clade ages, relying on the fossil ages and a morphological clock to induce time calibrations. To provide a generative process of fossilization, Heath et al. (2014) introduced 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, Gavryushkina et al. (2015) demonstrated multiple calibration techniques may be used jointly 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 too soft or too fragile of 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—namely, large vertebrates and marine invertebrates—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). For example, an ornithologist 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 origination. Using this scenario as a calibration point 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. Like fossil node calibrations, biogeographic node calibrations fundamentally rely on some prior distri-

bution of divergence times, opinions may vary from expert to expert, making results difficult to compare from a modeling perspective. 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. Ideally, all possible biogeographic and diversification scenarios would be considered, 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 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 higher when they were joined as West Gondwana (ca 120 Ma) 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 my 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 call biogeographic dating. Throughout the manuscript, I assume a rooted phylogeny, Ψ, 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, $\mathbf{Z}$, 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. Represented as a graph, A and B are vertices, and the edge (A, B) exists only during time intervals when A and B are in contact. The addition and removal of dispersal routes demarcate time intervals, or *epochs*, each corresponding to some epoch index, $k \in \{1, \ldots, K\}$. To define how dispersal rates vary with k, I use an epoch-valued continuous-time Markov chain (CTMC) (Ree et al. 2005; Ree and Smith 2008; Bielejec et al. 2014). The adjacency matrix for the k^{th} time interval's

graph is used to populate the elements of an instantaneous rate matrix for an epoch's CTMC such that the dispersal rate is equal to 1 when the indexed areas are adjacent and equals 0 otherwise. For a time-homogeneous CTMC, the transition probability matrix is typically written as $\mathbf{P}(t)$, which assumes the rate matrix, $\mathbf{Q}$, has been rescaled by some clock rate, μ, and applied to a branch of some length, t. For a time-heterogeneous CTMC, the value of the rate matrix changes as a function of the underlying time interval, $\mathbf{Q}(k)$. The transition probability matrix for the time-heterogeneous 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.

Below, I give examples of when a key divergence time is likely to precede a split event (Figure 1) or to follow a merge event (Figure 2). To simplify the argument, I assume a single change must occur on a certain branch given the topology and tip states, though the logic holds in general.

In the first scenario (Figure 1), 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 $[\mathbf{P}(s,t)]_{AB} = 0$ 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 $T2$ an $T3$ 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 bound 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

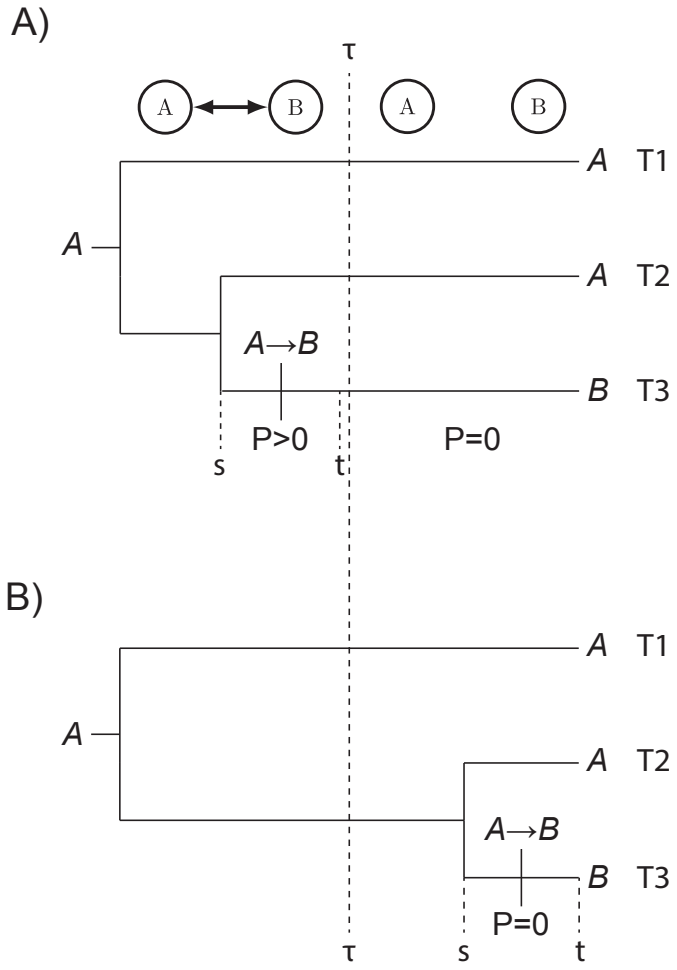


Figure 1: **Effects of a paleogeographical split on divergence times.** Area A splits from Area B at time τ. $T1$ and $T2$ have state A and the transition $A \rightarrow B$ most parsimoniously explains how $T3$ has state B. The transition probabilty for $P = [\mathbf{P}(s,t)]_{AB}$ is non-zero before 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.

tight bounds centered on τ (see Discussion).

In the second scenario (Figure 2), 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 as 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 2A). Alternatively, if $T3$ and $T4$ originate before the areas merge, then the same dispersal event

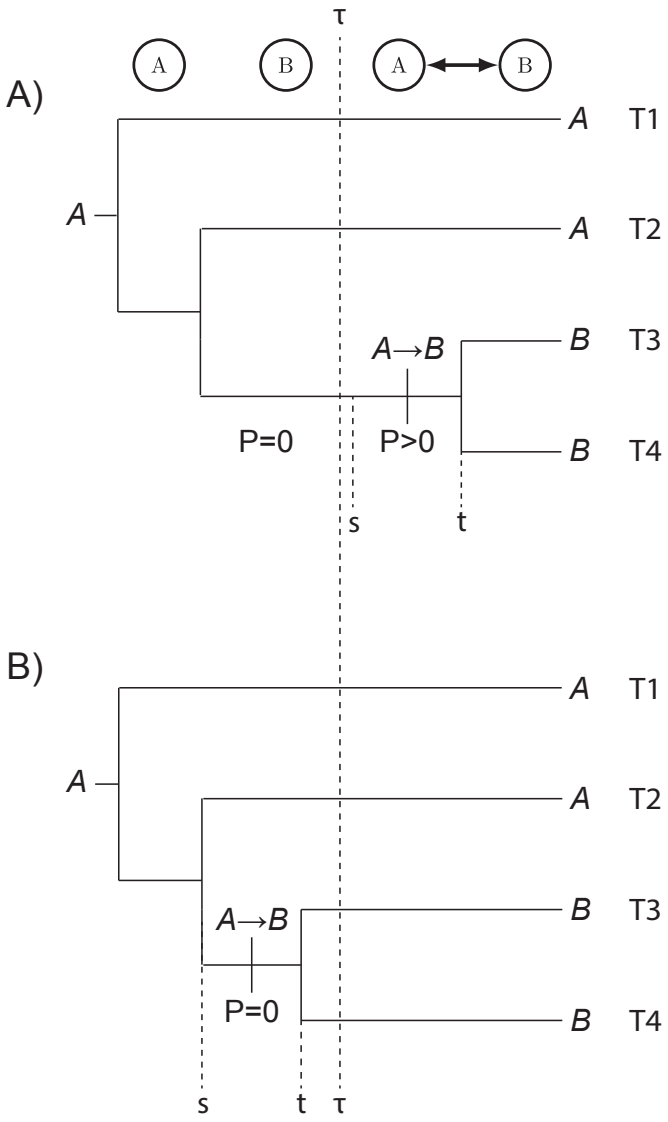


Figure 2: **Effects of paleogeographical merge on divergence times.** Area A merges with Area B at time τ. $T1$ and $T2$ have the state A and the transition $A \to B$ on the lineage leading to $(T3, T4)$ most parsimoniously explains how $T3$ and $T4$ have state B. The transition probabilty for $P = [\mathbf{P}(s,t)]_{AB}$ is zero before the paleogeographical merge event at time τ, and only non-zero afterwards. Two possible divergence and dispersal times are given: A) $T3$ and $T4$ originate after the merge when the transition $A \to B$ has non-zero probability. B) $T3$ and $T4$ originate before the merge when the transition $A \to B$ has probability zero.

on the branch ancestral to $(T3, T4)$ has probability zero (Figure 2B).

2.2 Paleogeography, graphs, and Markov chains

How biogeography may date speciation times depends critically on the assumptions of the biogeographic model. The above examples depend on the notion of *reachability*, that

two vertices (areas) are connected by some ordered set of edges (dispersal routes) of any length. In the adjacent-area dispersal model used here, one area might not be reachable from another area during some time interval, during which the corresponding transition probability is zero. That is, no path of any number of edges (series of dispersal events) may be constructed to connect the two areas. The concept of reachability may be extended to sets of partitioned areas: in graph theory, sets of vertices (areas) that are mutually reachable are called *(connected) components*. In terms of a graphically structured continuous time Markov chain, each component forms a *communicating class*: a set of states with positive transition probabilities only to other states in the set, and zero transition probabilities to other states (or communicating classes) in the state space. To avoid confusion with the "generic" biogeographical concept of components (Passalacqua 2015), and to emphasize the interaction of these partitioned states with respect to the underlying stochastic process, I hereafter refer to these sets of areas as 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 3), meaning species are free to disperse within these paleocontinents, but not between them. For example, the set of communication classses 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 3C), while at $t = -10$ there is $|S| = 1$ communicating class since a path exists between all three pairs of areas (Figure 3E).

Specifying communicating classes is partly difficult because we do not know the ease of dispersal between areas for most species throughout space and time. 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). Moreover, zero-valued rates imply that dispersal events between certain areas are not simply improbable but completely 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

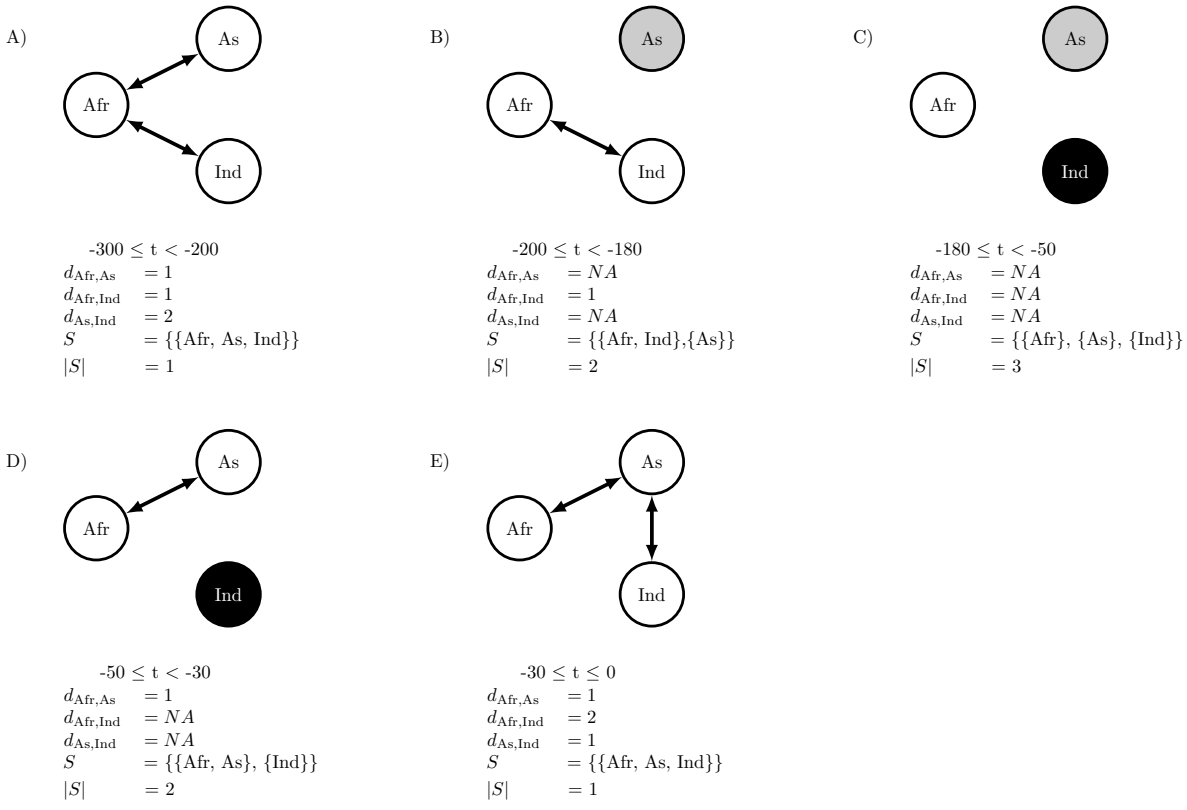


Figure 3: **Biogeographic communicating classes.** Dispersal routes shared by Africa (Afr), Asia (As), and India (Ind) are depicted for each time interval, t, over the past 300 Ma. Dispersal path lengths between areas i and j are given by $d_{i,j}$, with NA meaning there is no route between areas (areas 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.

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).

Fundamentally, biogeographic dating depends on how rapidly a species may disperse between two areas, and how that dispersal rate changes over time. In one extreme case, dispersals between mutually unreachable areas do not occur 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, so long as 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 3A and 3E, the minimum number of events required to disperse from India to Africa is smaller during the Triassic ($t = -250$) 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).

The concepts of adjacency, reachability, components, 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 (Landis et al. 2013) or the size of areas (Tagliacollo et al. 2015). In this study, these concepts serve the practical purpose of summarizing perhaps the most salient feature of global paleogeography—that continents were not always configured as they were 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.

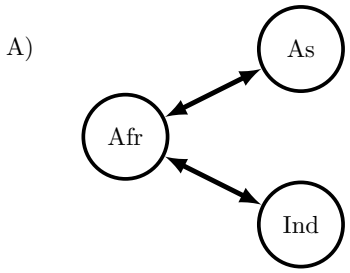
2.3 Time-heterogeneous dispersal process

Let $\mathbf{Z}$ be a vector reporting biogeographic states for $M > 2$ taxa. The objective is to construct a time-heterogeneous CTMC where transition probabilities depend on time-calibrated paleogeographical features. For simplicity, species ranges are assumed to be endemic on the continental scale, so each taxon's range may be encoded as an integer in $Z_i \in \{1, 2, \ldots, N\}$, where N is the number of areas.

The paleogeographical features that determine the dispersal process rates are assumed to be 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. 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.

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 throughout time $\tau_{k-1} \leq t < \tau_k$. As a example of an paleogeographically-informed matrix structure, take $\mathbf{G}(k)$ to be a 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 examples of $\mathbf{G} = (\mathbf{G}(1), \mathbf{G}(2), \ldots, \mathbf{G}(K))$ describing Earth's paleocontinental adjacencies are given in detail later.



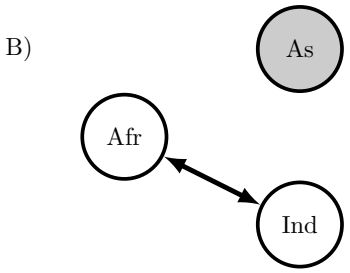
-300 ≤ t < -200

$\mu\mathbf{Q}(\text{k}) =$

	Afr	As	Ind
Afr	-	0.5	0.5
As	0.5	-	0.0
Ind	0.5	0.0	-

$\mathbf{P}(-250, -249) =$

	Afr	As	Ind
Afr	0.48	0.26	0.26
As	0.26	0.67	0.07
Ind	0.26	0.07	0.67



-200 ≤ t < -180

$\mu\mathbf{Q}(\text{k+1}) =$

	Afr	As	Ind
Afr	-	0.0	0.5
As	0.0	-	0.0
Ind	0.5	0.0	-

$\mathbf{P}(-190, -189) =$

	Afr	As	Ind
Afr	0.68	0.00	0.32
As	0.00	1.00	0.00
Ind	0.32	0.00	0.68

Figure 4: **Piecewise-constant dispersal rate matrices.** Dispersal routes shared by Africa (Afr), Asia (As), and India (Ind) are depicted for two time intervals, $-300 \leq t < -200$ and $-200 \leq t < 180$. Graphs and times correspond to those in Figure 3A,B. Transition probabilities are computed for a unit time during different epochs with a time-homogeneous biogeographic clock rate $\mu = 0.5$. A) The three areas are connected and all transitions have positive probability. B) As is unreachable from Afr and Ind, so transition probabilities into and out of As are zero.

With the paleogeographical vector $\mathbf{G}$, I define the transition rates of $\mathbf{Q}(k)$ as equal to $\mathbf{G}_z(k)$. Similar rate matrices are constructed for all K time intervals that contain possible supported root ages for the phylogeny, Ψ. Figure 4 gives a simple example for three areas, where Asia shares positive dispersal rate with Africa when they are merged and no dispersal while split.

For a piecewise-constant CTMC, the process' 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 the 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 $\tau_{k-1} \leq v_i < \tau_k$. The transition probabilty matrix over the intervals in v according to the piece-wise 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. See Bielejec et al. (2014) for an excellent review of piecewise-constant CTMCs as applied to phylogenetics.

In the above case, the times s and t are generally identifiable from μ_z so long as $\mathbf{P}_\tau(v, \mu; \tau, \mathbf{Q}) \neq \mathbf{P}_\tau(v', \mu'; \tau, \mathbf{Q})$ for any supported values of v, μ, v', and μ'. Note, I include μ as an explicit parameter in the transition probability matrix function for clarity, though they are suppressed in standard CTMC notation when t equals the product of rate and time, then the process effectively runs for the time, $\mu(t - s)$. For example, assume that $\mathbf{Q}$ is a time-homogeneous Jukes-Cantor model with no paleogeographical constraints, i.e. all transition rates are equal independent of k. The transition probability matrix for this model

is readily computed via matrix exponentiation

$$\mathbf{P}(s,t,\mu) = e^{\mu(t-s)\mathbf{Q}}.$$

Note that $\mathbf{P}(s,t,\mu) = \mathbf{P}_\tau(v,\mu;\tau,\mathbf{Q})$ when $v = (s,t)$ – i.e. the time-heterogeneous process spans no breakpoints when $m = 0$ and is equivalent to a time-homogeneous process for the interval (s,t).

For a time-homogeneous model, multiplying the rate and dividing the branch length by the same factor results in an identical transition probability matrix. In practice this means the simple model provides no information for the absolute value of μ and the tree height of Ψ, since all branch rates could likewise be multiplied by some constant while branch lengths were divided by the same constant, i.e. $\mathbf{P}(s,t,1) = \mathbf{P}(s\mu^{-1}, t\mu^{-1}, \mu)$. Similarly, since $\mathbf{P}(s,t,\mu) = \mathbf{P}(s+c, t+c, \mu)$ for $c \geq 0$, the absolute time when the process begins does not matter, only the amount of time that has elapsed. Extending a branch length by a factor of c requires modifying other local branch lengths in kind to satisfy time tree constraints, so the identifiability of the absolute time interval (s,t) depends on how "relaxed" (Drummond et al. 2006) the assumed clock and divergence time priors are with respect to the magnutude of c, which together induce some (often unanticipated) joint prior distribution on divergence times and branch rates (Heled and Drummond 2012; Warnock et al. 2015). In either case, rate and time estimates under the time-homogeneous process result from the induced prior distributions rather than by informing the process directly.

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 5; for all graphs and a link to the animation, see Supplemental Figure S1). All adjacencies were constructed visually, referencing Blakey (2006) and GPlates (Boyden et al. 2011), then corroborated using various paleogeographical sources (Table S2). The paleogeographical state per time interval is summarized

as an undirected graph, where areas are vertices and dispersal routes are edges.

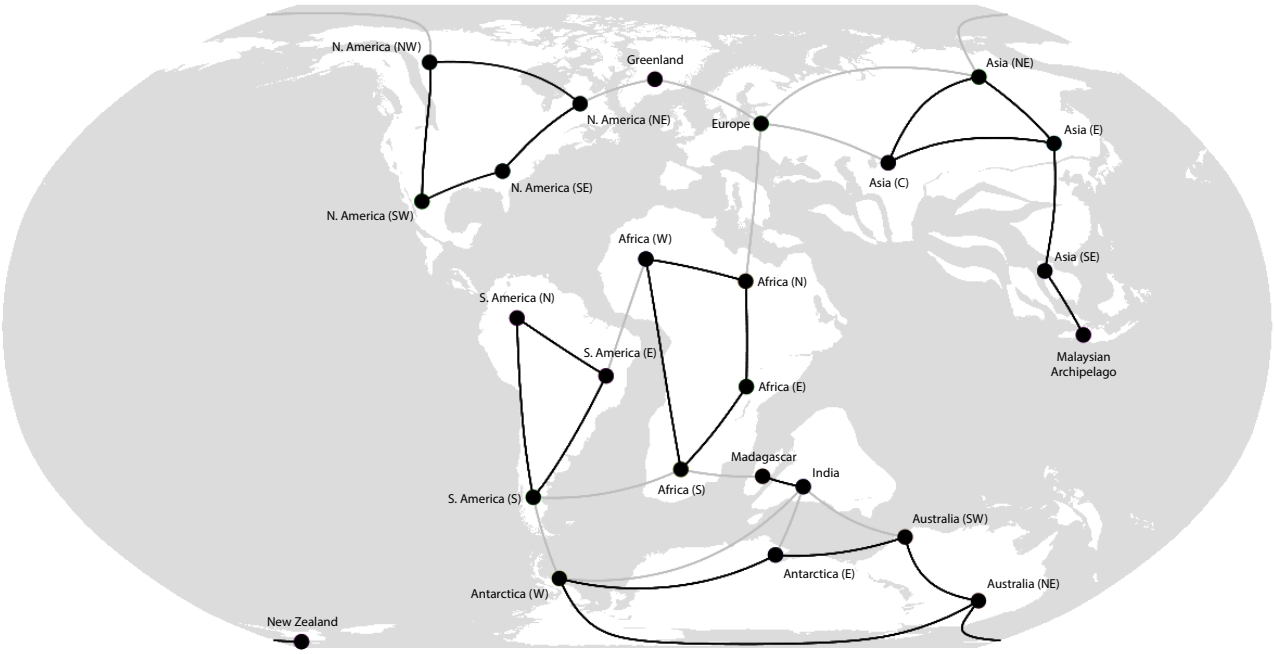


Figure 5: **Dispersal graph for Epoch 14, 110–100Ma: India and Madagascar separate from Australia and Antarctica.** A `gplates` (Gurnis et al. 2012) screenshot of Epoch 14 of 26 is displayed. Areas are marked by black vertices. Black edges indicate both short- and medium distance dispersal routes. Gray 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 dispersal graphs have 8, 1, and 1 communicating classes, respectively. India and Madagascar each have only one short distance dispersal route, which they share. 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.

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 an edge at time interval k, and $\mathbf{G}_\bullet(k)_{i,j} = 0$ otherwise. 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 geological adjacency matrices

$$\mathbf{G}_z = b_s\mathbf{G}_s + b_m\mathbf{G}_m + b_l\mathbf{G}_l \tag{1}$$

where s, m, and l correspond to short distance, medium distance, and long distance mode

parameters.

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 informative of the process' previous state or communicating class (Figure 6). 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 allowing terrestrial dispersal without travelling through intermediate areas (Figure 6A). 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 (Figure 6B). Finally, long distance dispersal graphs, $\mathbf{G}_l$, allow dispersal events to occur between any pair of areas, regardless of potential barrier (Figure 6C).

To average over the three dispersal modes, b_s, b_m, and b_l are constrained to sum to 1, causing all elements in $\mathbf{G}_z$ to take values from 0 to 1 (Eqn 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 $\mathbf{G}_z$, but are rescaled such that the average number of transitions per unit time equals 1, and 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 7.

By the argument of that continental break-up (i.e. the creation of new communicating classes; Figure 1) introduces a bound on the minimum age of divergence, and that continental joining (i.e. unifying existing communicating classes; Figure 2) introduces a bound on the maximum age of divergence, then the paleogeographical model I constructed has the greatest potential to provide both upper and lower bounds on divergence times when the number of communicating classes is large, then small, then large again. This coincides with the

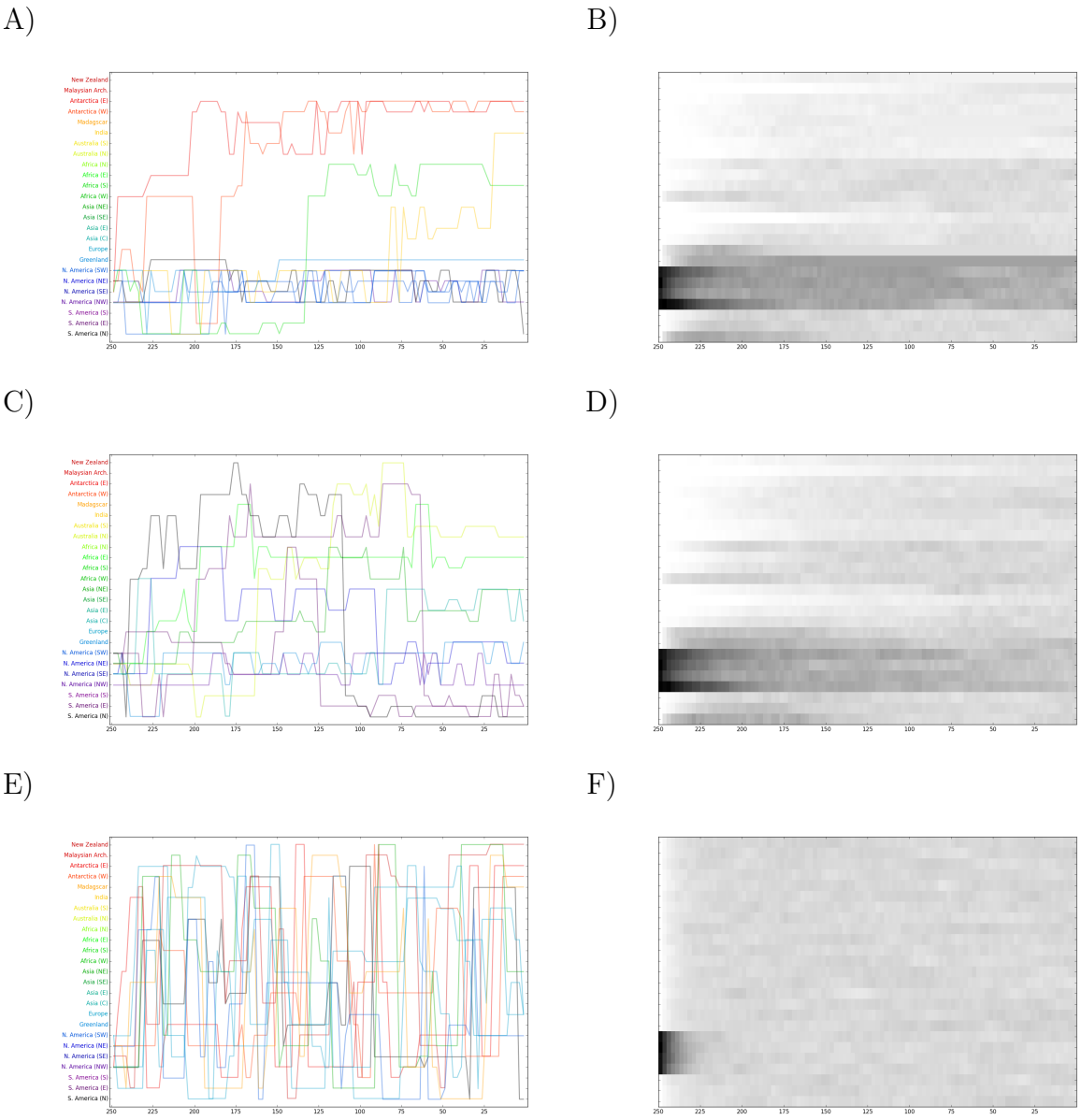
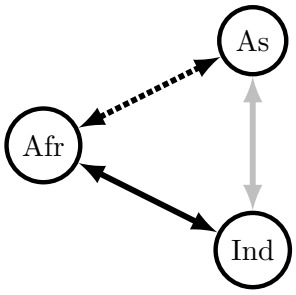


Figure 6: **Sample paths for paleogeographically informed biogeographic process.** The top, middle, and bottom panels show dispersal histories simulated by the pure short (A,B), medium (C,D), and long (E,F) distance process components. All processes originate in one of the four North American areas 250 Ma. The left column shows 10 of 2000 sample paths. Color indicates the area the lineage is found in the present (A,C,E). Colors for areas match those in Figure 8. The right column heatmap reports the sample frequencies for any of the 2000 dispersal process being in that state at that time (B,D,F).

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 8). 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



$$b_s + b_m + b_l = 1.0$$

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

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

Figure 7: **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 is only by long distance with weight $b_l = 0.1$.

bound to lack a biogeographic record that is informative of origination times.

3 Analysis

All posterior densities were estimated using Markov chain Monte Carlo (MCMC) as implemented in RevBayes, available at `revbayes.com` (Höhna et al. 2014). Data and analysis scripts are available at `github.com/mlandis/biogeo_dating`. Datasets are also available on Dryad at `datadryad.org/XXX`. Analyses were performed on the XSEDE supercomputing cluster (Towns et al. 2014).

Simulation

Through simulation I tested whether biogeographic dating identifies rate from time. To do so, 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

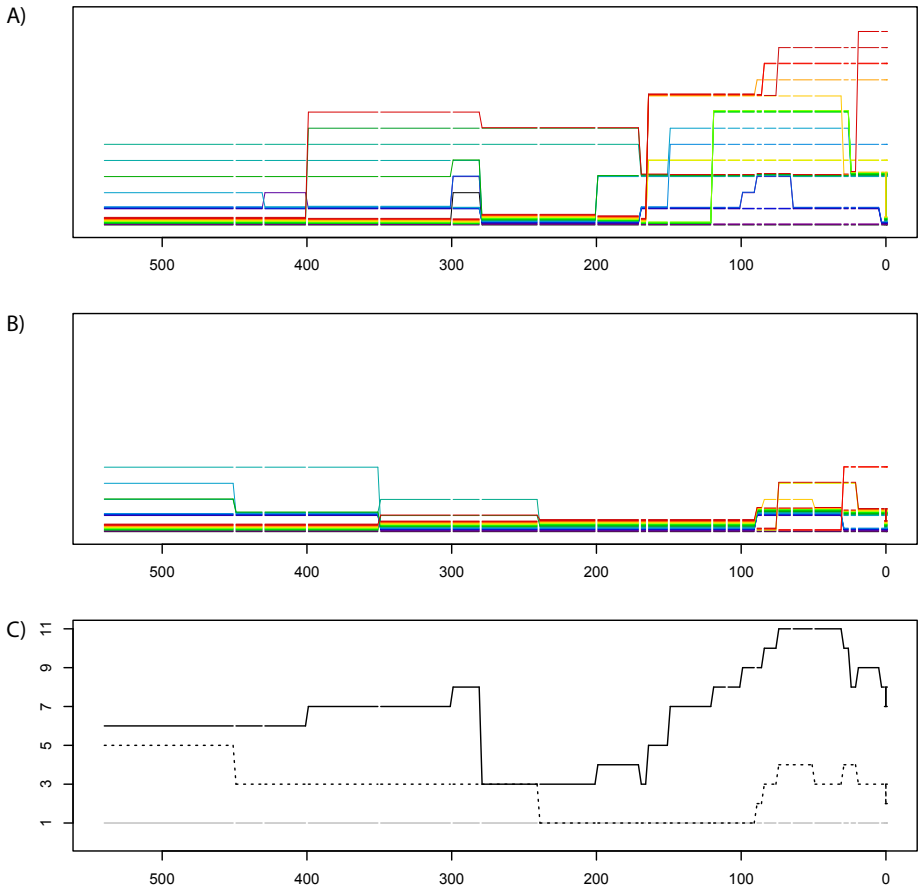


Figure 8: **Dispersal graph properties summarized over time.** communicating classes of short distance dispersal graph (A) and medium distance dispersal graph (B) are shown. Each of 25 areas is represented by one line. Colors of areas match those listed in Figure 6. Grouped lines indicate areas in one communicating class during an interval of time. Vertical lines indicate transitions of areas joining or leaving communication clases, i.e. due to paleogeographical events. When no transition event occurs for an area entering a new epoch, the line is interrupted with gap. (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.

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.

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

			Simulation		**Empirical**
	Parameter	X	$f(X)$	sim. value	$f(X)$
Tree	Root age	r	Uniform$(0, 1000)$	250	Uniform$(0, 540)$ or Uniform$(151.7, 251.4)$
	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.

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 $\mathbf{G}_z$ 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 $\approx$ 499 (Figure 9A). 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.

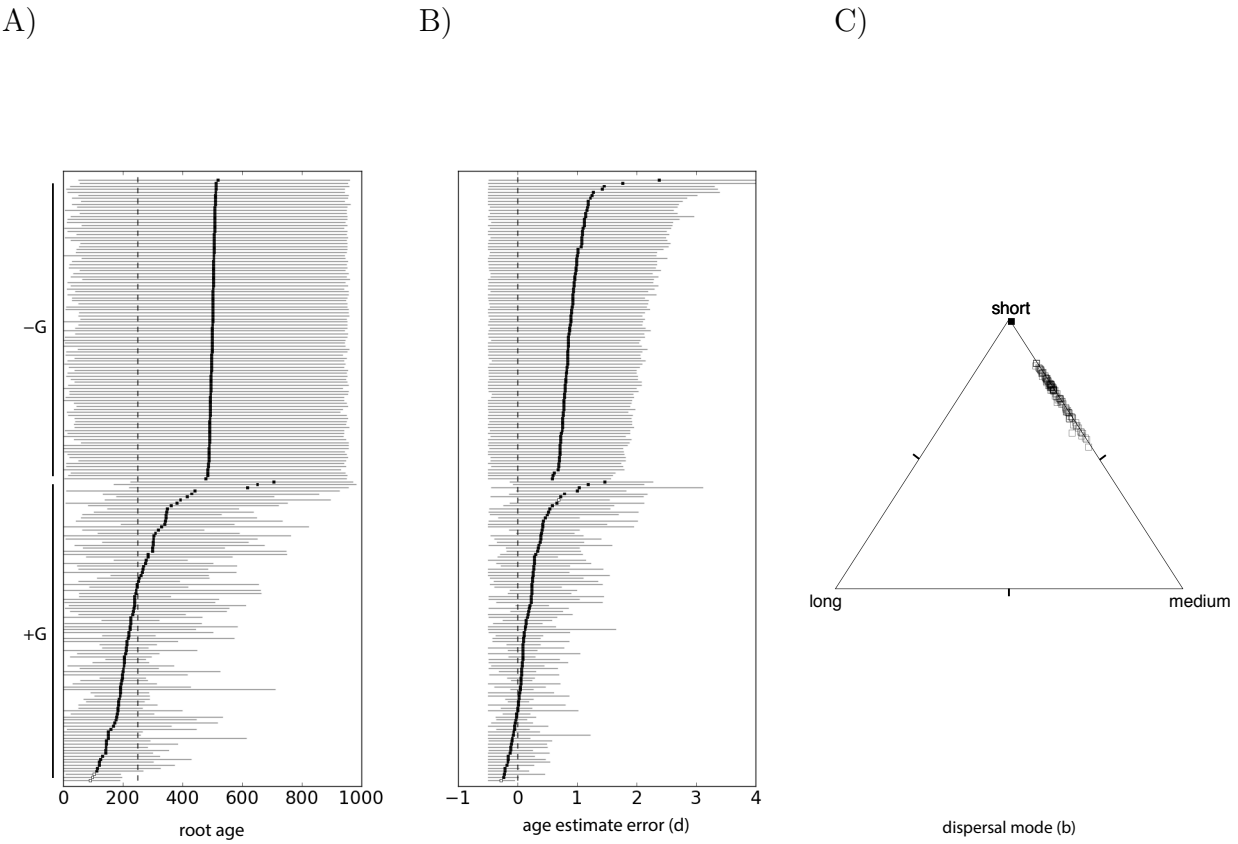


Figure 9: **Posterior estimates for simulated data.** A) Posterior estimates of root age. The true root age for all simulations is 250 Ma (dotted vertical line). B) Posterior estimates of relative node age error (Eqn 2). The true error term equals zero. Both A and B) –G analyses are on the top half, +G analyses are on the bottom. Each square marks the posterior mean root age estimate with the HPD95% credible interval. If the credible interval contains the true value, the square is filled. C) Posterior estimates of dispersal mode proportions for the +G simulations projected onto the unit 2-simplex. The filled circle gives the posterior median-of-medians, and the empty circles give posterior medians.

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_{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 9B). Posterior estimates generally favored short over medium and long distance dispersal as was as-

sumed under simulation (Figure 9C). 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* are thought to be Gondwanan in origin with the ancestor to cryptodires dispersing into Laurasia during the Jurassic (Crawford et al. 2015). Since turtles preserve so readily in the fossil record, estimates of 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 species coverage, pruning away unused taxa. All speciation times were considered random variables to be estimated. The

tree topology and biogeographic states are shown in Supplemental Figure S2. All data are recorded on `datadryad.org/XXX`.

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).

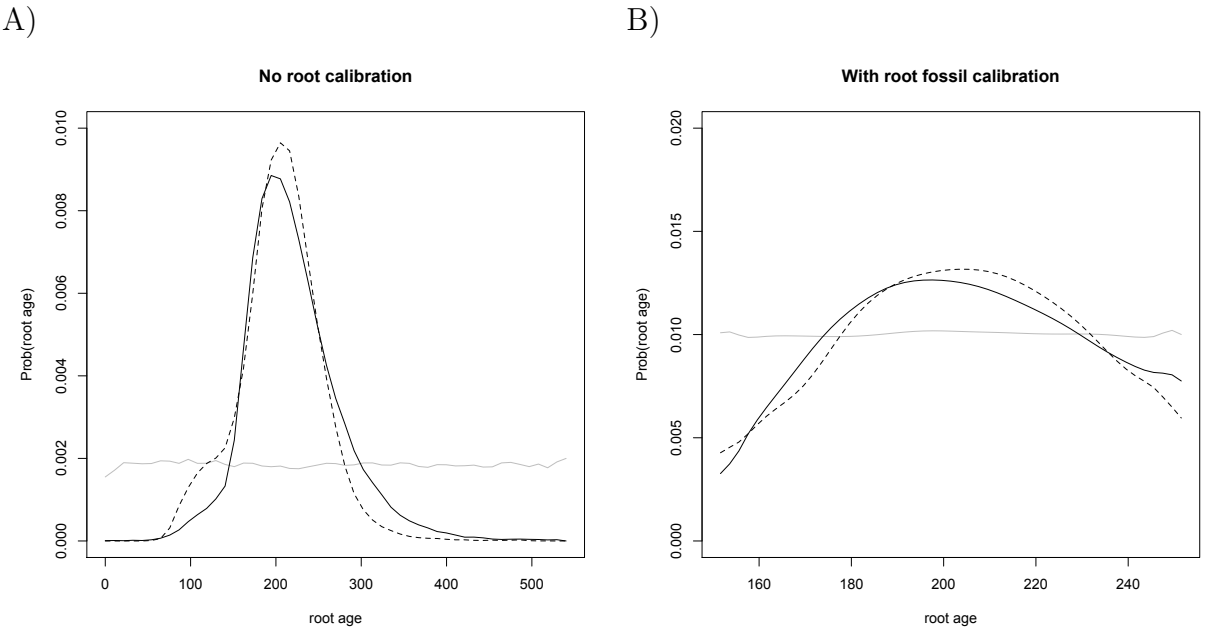


Figure 10: **Posterior root age of turtles by biogeographic dating.** Six root age posterior estimates were computed using biogeographic dating, each using variations on flat or short-biased priors for key parameters. Figure A assumes no knowledge of fossils with Uniform(0, 540) root age prior. Figure B follows Joyce et al. (2013) and assumes Uniform(151.7, 251.4) as a root node age calibration. The black solid posterior density assumes a flat prior on dispersal mode. The black dotted posterior density assumes an short-biased prior Dirichlet(100, 10, 1) on dispersal mode. The gray solid posterior density ignores paleogeography.

Biogeographic dating infers a posterior median root age of 201 with HPD95% credible interval of (115, 339) (Figure 10A). This is consistent current root age estimates informed from the fossil record (Figure 11). The posterior mode of dispersal mode is $(b_s, b_m, b_l) = (0.47, 0.51, 0.02)$, with short and medium distance dispersal occurring at approximately equal

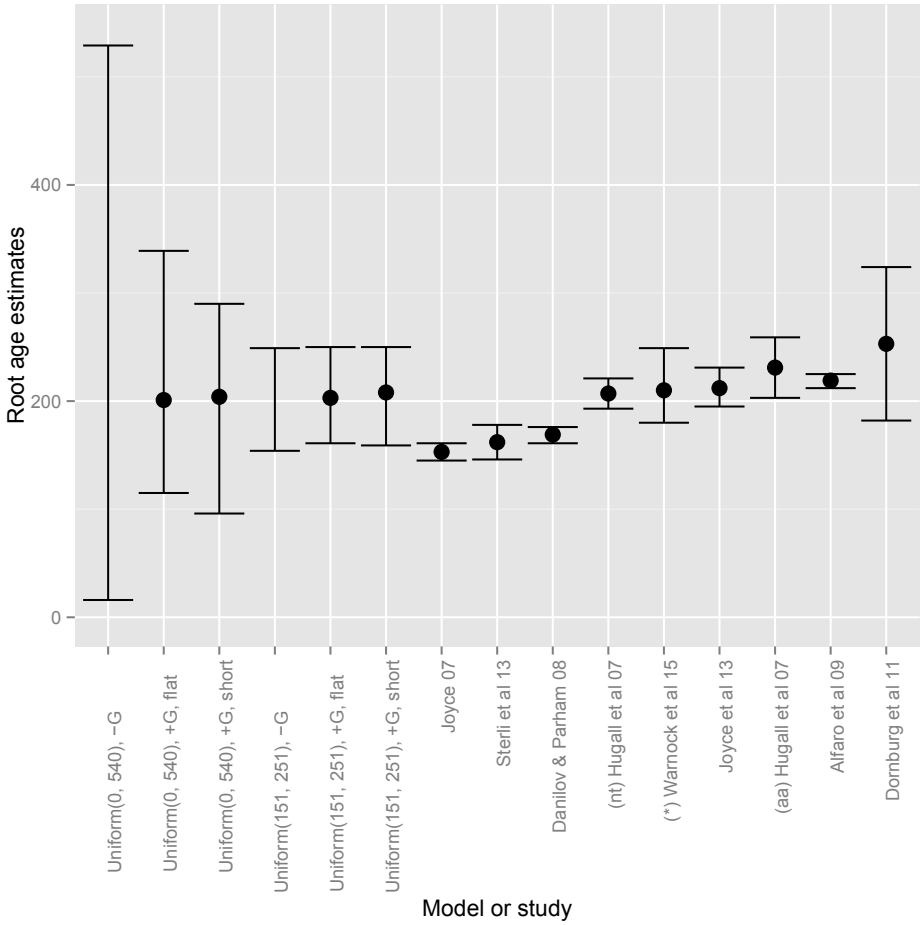


Figure 11: **Root age comparison.** Root age estimates are presented both for analysis conducted for this manuscript and as reported in existing publications. Existing estimates are as reported in Sterli et al. (2013) and supplemented recently reported results. Points and whiskers correspond to the point estimates and estimate confidence, which varies across analyses. 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.

rates and long distance dispersal being rare by comparison. Biogeographic events occurred at a ratio of about 6:1 when compared to molecular events (posterior means: $\mu_x = 1.9\text{E}{-3}, \mu_z = 1.1\text{E}{-2}$). The posterior mode tree height measured in expected number of dispersal events is 2.3 with HPD95% (1.5, 3.0), i.e. as a treewide average, the current location of each taxon is the result of about two dispersal events.

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 expected proportion of 100:10:1 short:medium:long dispersal events. After re-analyzing the data with the short-biased dispersal prior, the posterior median and HPD95% credible interval were estimated to be, respectively, 204 (96, 290) (Figure 10A).

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(151.7, 251.4) (Joyce et al. 2013). When taking biogeography into account, the model more strongly disfavors post-Pangaean origins for the clade than when biogeography is ignored, but the effect is mild. Posterior distributions of root age was relatively insensitive to the flat and short-biased dispersal mode priors, with posterior medians and credible intervals of 203 (161, 250) and 208 (159, 250), respectively.

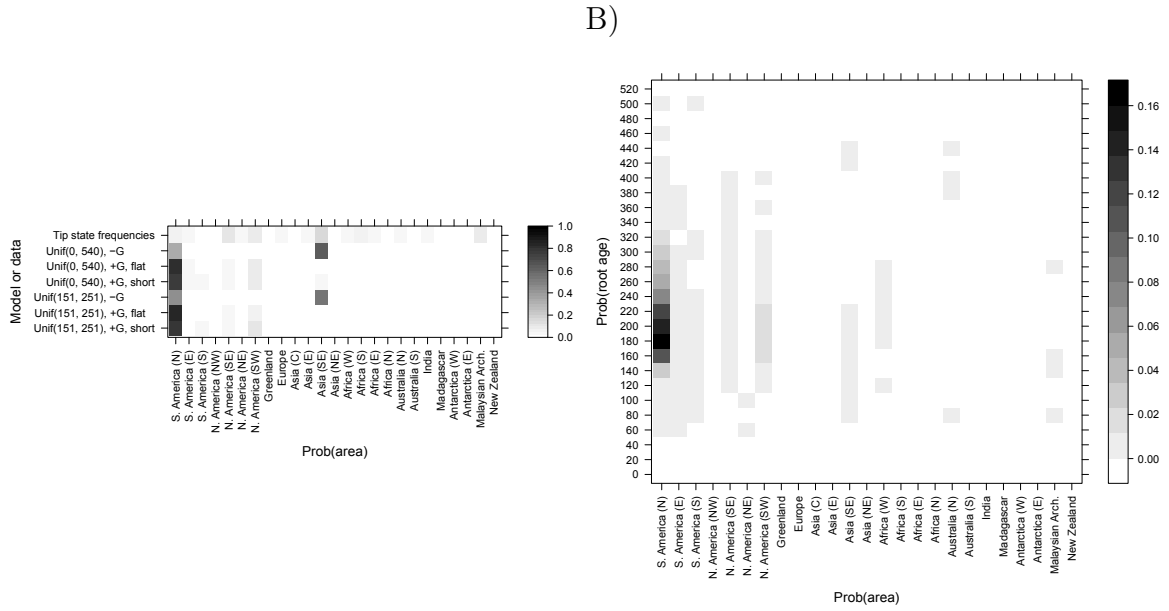


Figure 12: **Root state estimates.** A) Posterior probabilities of root state are given for the six empirical analyses. B) Joint-marginal posterior probabilities of root age and root state are given for the empirical analysis without a root calibration and with a flat dispersal mode prior. Root ages are binned into intervals of width 20.

All posterior root state estimates favored South America (N) for the paleogeographically-informed analyses (Figure 12A). Although this is in accord with the root node calibration adopted from Joyce et al. (2013)—*Caribemys oxfordiensis*, sampled from Cuba, and the oldest accepted crown group testudine—the fossil is described as a marine turtle, so the accordance may simply be coincidence. In contrast, the paleogeographically-naive models

support 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 on the posterior root age density (Figure 12B), i.e. regardless of conditioning on South America (N), North America (SE), or North America (SW) as a root state, the posterior root age density is roughly equal.

4 Discussion

The 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. 2011), 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 (Figure 5) into a time-calibrated vector of dispersal graphs.

Through simulation, I showed biogeographic dating identifies tree height from the rates of molecular and biogeographic character change. This 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 due to signal saturation (Figure 6C). 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, especially in comparison to the fossil record.

The majority of groups have poor fossil records, and 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 10B). When fossils with geographical information are available, researchers have shown fossil taxa improve biogeographical analyses (Moore et al. 2008; Wood et al. 2012). In principle, the biogeographic process 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. 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).

Because time calibration through biogeographic inferences comes primarily from the paleogeographical record, not the fossil record, divergence times may be estimated from exclusively extant taxa under certain biogeographical and phylogenetic conditions. When fossils are available, however, biogeographic dating is compatible with other fossil-based dating methods (e.g. node calibrations, fossil tip dating, fossilized birth-death). As a proof of concept, 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 10). Under the uninformative prior distributions on root age, biogeographic dating estimated turtles originated between the Mississipian (339 Ma) and Early Cretaceous (115 Ma) periods, with

a median age of 201 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 flat and biased dispersal mode priors, biogeographic dating placed the posterior mode origin time of turtles at approximately 210–200 Ma, which is consistent with fossil-based estimates (Figure 11).

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. I discuss some of the most apparent model misspecifications below.

Anagenetic range evolution models that properly allow species 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 spanning both north and south of the Isthmus of Panama. This situation almost certainly requires a dispersal event to have occurred after the isthmus was formed when multiple-area ranges are used. For single-area species ranges coded as ambiguous states, the model is incapable of evaluating the likelihood that the species is found in both areas simultaneously, so additional information about the effects of the paleogeographical event on divergence times is potentially lost.

Any model where the diversification process and paleogeographical states (and events)

are correlated 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 preceeded 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. Rates of dispersal between areas are classified into short, medium, and long distances, but with subjective criteria. 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. The timing of events was assumed to be known perfectly, despite the literature reporting ranges of estimates. Certainly factors such as global temperature, precipitation, ecoregion type, etc. affect dispersal rates between areas, but were ignored. All of these factors can and should be handled more rigorously in future studies by modeling these uncertainties as part of a joint Bayesian analysis (Höhna et al. 2014).

Despite these flaws, defining the paleogeographical model serves as an excercise to identify what features allow a biogeographic process to inform speciation times. Dispersal barriers are clearly clade-dependent, e.g. benthic marine species dispersal would be poorly modeled by

the terrestrial graph. Since dispersal routes for the terrestrial graph might serve as dispersal barriers for a marine graph, there is potential for learning about mutually exclusive dispersal corridor use in a multi-clade analysis (Sanmartín et al. 2008). 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. 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). 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 paleogeographical inference. 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). 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

602 interest, only to be analysed after the species tree is estimated.

603 Acknowledgements

604 I thank Tracy Heath, Sebastian Höhna, Josh Schraiber, Lucy Chang, Pat Holroyd, Nick
605 Matzke, Michael Donoghue, and John Huelsenbeck for valuable feedback, support, and en-
606 couragement regarding this work. I also thank James Albert, Alexandre Antonelli, and the
607 Society of Systematic Biologists for inviting me to present this research at Evolution 2015
608 in Guarujá, Brazil. Simulation analyses were computed using XSEDE, which is supported
609 by National Science Foundation grant number ACI-1053575.

610 Funding

611 MJL was supported by a National Evolutionary Synthesis Center Graduate Fellowship
612 and a National Institutes of Health grant (R01-GM069801) awarded to John P. Huelsenbeck.

613

614 ## References

615 Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale,
616 and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species
617 diversity in jawed vertebrates. Proceedings of the National Academy of Sciences 106:13410–
618 13414.

619 Beaulieu, J. M., D. C. Tank, and M. J. Donoghue. 2013. A southern hemisphere origin for
620 campanulid angiosperms, with traces of the break-up of gondwana. BMC Evolutionary
621 Biology 13:80.

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

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

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

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

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

636 Carlquist, S. 1966. The biota of long-distance dispersal. i. principles of dispersal and evolu-
637 tion. Quarterly Review of Biology Pages 247–270.

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

641 Danilov, I. G. and J. F. Parham. 2008. A reassessment of some poorly known turtles from
642 the middle jurassic of china, with comments on the antiquity of extant turtles. Journal of
643 Vertebrate Paleontology 28:306–318.

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

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

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

651 Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood ap-
652 proach. Journal of Molecular Evolution 17:368–376.

653 Gavryushkina, A., T. A. Heath, D. T. Ksepka, T. Stadler, D. Welch, and A. J. Drummond.
654 2015. Bayesian total evidence dating reveals the recent crown radiation of penguins. arXiv
655 preprint arXiv:1506.04797 .

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

658 Goldberg, E. E., L. T. Lancaster, and R. H. Ree. 2011. Phylogenetic inference of reciprocal
659 effects between geographic range evolution and diversification. Systematic Biology 60:451–
660 465.

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

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

666 Heads, M. 2005. Dating nodes on molecular phylogenies: a critique of molecular biogeogra-
667 phy. Cladistics 21:62–78.

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

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

673 Heled, J. and A. J. Drummond. 2012. Calibrated tree priors for relaxed phylogenetics and
674 divergence time estimation. Systematic Biology 61:138–149.

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

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

679 Höhna, S., T. A. Heath, B. Boussau, M. J. Landis, F. Ronquist, and J. P. Huelsenbeck.
680 2014. Probabilistic graphical model representation in phylogenetics. Systematic Biology
681 63:753–771.

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

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

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

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

692 Kodandaramaiah, U. 2011. Tectonic calibrations in molecular dating. Current Zoology
693 57:116–124.

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

696 Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009. Bayesian phylogeogra-
697 phy finds its roots. PLoS Computational Biology 5:e1000520.

698 Lemey, P., A. Rambaut, J. J. Welch, and M. A. Suchard. 2010. Phylogeography takes a
699 relaxed random walk in continuous space and time. Moleculate Biology and Evolution
700 27:1877–1885.

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

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

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

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

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

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

713 Moore, B. R., S. A. Smith, R. H. Ree, and M. J. Donoghue. 2008. Incorporating fossil data
714 in biogeographic inference: A likelihood approach. Evolution .

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

718 Parham, J. F., P. C. J. Donoghue, C. J. Bell, T. D. Calway, J. J. Head, P. A. Holroyd, J. G.
719 Inoue, R. B. Irmis, W. G. Joyce, D. T. Ksepka, J. S. L. Patané, N. D. Smith, J. E. Tarver,
720 M. van Tuinen, Z. Yang, K. D. Angielczyk, J. M. Greenwood, C. A. Hipsley, L. Jacobs,
721 P. J. Makovicky, J. Müller, K. T. Smith, J. M. Theodor, and R. C. M. Warnock. 2011.
722 Best practices for justifying fossil calibrations. Systematic Biology Page syr107.

723 Passalacqua, N. G. 2015. On the definition of element, chorotype and component in biogeog-
724 raphy. Journal of Biogeography 42:611–618.

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

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

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

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

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

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

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

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

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

744 Sanmartín, I., P. V. D. Mark, and F. Ronquist. 2008. Inferring dispersal: a bayesian approach
745 to phylogeny-based island biogeography, with special reference to the canary islands. Jour-
746 nal of Biogeography 35:428–449.

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

749 Sterli, J., D. Pol, and M. Laurin. 2013. Incorporating phylogenetic uncertainty on phylogeny-
750 based palaeontological dating and the timing of turtle diversification. Cladistics 29:233–
751 246.

752 Tagliacollo, V. A., S. M. Duke-Sylvester, W. A. Matamoros, P. Chakrabarty, and J. S.
753 Albert. 2015. Coordinated dispersal and pre-isthmian assembly of the central american
754 ichthyofauna. Systematic Biology Page syv064.

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

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

760 Warnock, R. C., J. F. Parham, W. G. Joyce, T. R. Lyson, and P. C. J. Donoghue. 2015.
761 Calibration uncertainty in molecular dating analyses: there is no substitute for the prior
762 evaluation of time priors. Proceedings of the Royal Society of London B: Biological Sciences
763 282:20141013.

764 Webb, C. O. and R. H. Ree. 2012. Historical biogeography inference in Malesia. Pages 191–
765 215 *in* Biotic evolution and environmental change in Southeast Asia (D. Gower, K. John-
766 son, J. Richardson, B. Rosen, L. Ruber, and S. Williams, eds.) Cambridge University
767 Press.

768 Wilf, P. and I. H. Escapa. 2014. Green web or megabiased clock? plant fossils from gond-
769 wanan patagonia speak on evolutionary radiations. New Phytologist 207:283–290.

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

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

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

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