doi:10.1111/evo.136(

Hidden state models improve state-dependent diversification approaches, including biogeographical models

Daniel S. Caetano,[1,2] Brian C. O'Meara,[3] and Jeremy M. Beaulieu[1]

[1] *Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas 72701*

[2] *E-mail: dcaetano@uark.edu*

[3] *Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996-1610*

Received April 16, 2018
Accepted August 27, 2018

The state-dependent speciation and extinction (SSE) models have recently been criticized due to their high rates of "false positive" results. Many researchers have advocated avoiding SSE models in favor of other "nonparametric" or "semiparametric" approaches. The hidden Markov modeling (HMM) approach provides a partial solution to the issues of model adequacy detected with SSE models. The inclusion of "hidden states" can account for rate heterogeneity observed in empirical phylogenies and allows for reliable detection of state-dependent diversification or diversification shifts independent of the trait of interest. However, the adoption of HMM has been hampered by the interpretational challenges of what exactly a "hidden state" represents, which we clarify herein. We show that HMMs in combination with a model-averaging approach naturally account for hidden traits when examining the meaningful impact of a suspected "driver" of diversification. We also extend the HMM to the geographic state-dependent speciation and extinction (GeoSSE) model. We test the efficacy of our "GeoHiSSE" extension with both simulations and an empirical dataset. On the whole, we show that hidden states are a general framework that can distinguish heterogeneous effects of diversification attributed to a focal character.

KEY WORDS: **Biogeography, BiSSE, GeoSSE, hidden Markov, HiSSE, model averaging.**

Determining the impact that trait evolution has on patterns of lineage diversification is a fundamental and core question in evolutionary biology. The state-dependent speciation and extinction (SSE; Maddison et al. 2007) framework was developed specifically for these purposes, as it provides a means of correlating the presence or absence of a particular character state on diversification rates. Since the initial model was published, which modeled the evolution of a single binary character (i.e., BiSSE: Maddison et al. 2007; FitzJohn et al. 2009), the SSE framework has been expanded to deal with multiple state/traits (MuSSE: FitzJohn 2012), continuous traits (QuaSSE: FitzJohn 2010), to test whether change happens at speciation events or along branches (ClaSSE: Goldberg and Igic 2012; BiSSE-ness: Magnuson-Ford and Otto 2012), which also includes a nested subset of models that examines geographic range evolution (GeoSSE [geographic state-dependent speciation and extinction]: Goldberg et al. 2011), and to account for "hidden" states that may influence diversification on their own or in combination with observed states (HiSSE [hidden state speciation and extinction]: Beaulieu and O'Meara 2016).

The initial wave of interest in SSE models is quickly being replaced with widespread skepticism about their use (Rabosky and Goldberg 2015; O'Meara and Beaulieu 2016; Rabosky and Goldberg 2017). One major reason is based on the simulation study of Rabosky and Goldberg (2015). Their analyses showed that if a tree evolved under a heterogeneous branching process, completely independent from the evolution of the focal character, SSE models will almost always return high support for a model of state-dependent diversification as well as misleading parameter estimates. A similar effect was also recently demonstrated with geographic range data in the GeoSSE model (Alves et al. 2017).

There are reasons to be concerned, but, in our view, there is a deeper issue with the misinterpretation of hypothesis testing. First, rejecting the "null" model does not imply that the alternative model is the true model. It simply means that

the alternative model fits *less* badly. Misunderstanding of null hypothesis testing, and its dubious utility, has been a prominent issue for decades in other fields (i.e., Berkson 1938; Kirk 1996). Second, the apparent issue with SSE models is not a matter of high type I error rate, it is simply comparing a trivial "null" model (i.e., equal rates diversification) against a model of state-dependent diversification. Given the rich complexity of processes affecting diversification (e.g., mass extinctions, local extinctions, competition, and biogeographic changes) and trait evolution (e.g., varying population size, selection pressure, and available variation), a comparison of "one rate to rule them all" versus "something a bit more complex" will usually return the latter as a better descriptor of the data.

Aside from issues related to model rejection, and what is the appropriate "null" comparison, there are also broader issues related to any interpretation of state-dependent diversification. For instance, a common flaw in these types of analyses is the implicit assumption that a single trait is the primary driver of diversification, thereby ignoring alternate sources of rate variation (despite early warnings by Maddison et al. 2007). In reality, nearly all traits exert at least *some* effect on speciation and/or extinction rates. Even something as modest as a different base at a third codon position that leads to the same amino acid can have a tiny fitness difference (Yang and Nielsen 2008), which in theory could make a species infinitesimally closer to extinction. Moreover, even among traits we think have a greater effect, it is unlikely that *only* this one examined trait accounts for the increased diversification rates. In other words, if we are studying, say, growth habit, it is very unlikely that traits such as floral symmetry, pollination syndrome, biogeography, fruit dispersal, etc., all have exactly zero effect on diversification. It should, therefore, always be difficult to ever confidently view any one character state as the true underlying cause of changes in diversification rates.

The use of "hidden" states, and the hidden Markov modeling (HMM) approach in general, addresses multiple issues at once. These models are, however, challenging from an interpretational standpoint. For example, what does a "hidden" state mean? How does one weigh evidence for or against state-dependent and state-independent diversification in the presence of "hidden" rates? The purpose of this study is twofold. First, given the confusion over whether SSE models remain a viable means of assessing state-dependent diversification (e.g., Rabosky and Goldberg 2017), we further clarify the concept of "hidden" states as well as the misconceptions of "type I errors." Second, we demonstrate the role of hidden state models as a general framework by expanding the original GeoSSE model (Goldberg et al. 2011). We define a set of biologically meaningful models of area-independent diversification to be included in studies of area-dependent diversification that can be used in combination with the original GeoSSE formulation. Such models are especially useful given the recent clarification on the undesirable impact of cladogenetic events on the performance of dispersal-extinction-cladogenesis models (DEC: Ree and Smith 2008; DEC+J: Matzke 2014) as well as the renewed advocacy of using SSE models when examining patterns of geographic range evolution (Ree and Sanmartín 2018).

The Value of Incorporating "Hidden" States into SSE Models

As mentioned above, SSE models are routinely criticized on the grounds that when fitted to a tree evolved under a heterogeneous branching process independent from the evolution of the focal character, SSE models will almost always return high support for a state-dependent diversification model over a trivial "null" model that assumes equal diversification rates. Rabosky and Goldberg (2015) attribute these results to "type I error," however, as pointed out by Beaulieu and O'Meara (2016), this particular issue does not represent a case of type I error, but, rather, a problem of rejecting model *X*, for model *Y*, when model *Z* is true. Furthermore, rejecting model *X*, for model *Y*, does not imply that model *Y* is the true model. It simply means that model *Y* is a better *approximation* to model *Z*, than model *X*. This will generally be true if model *X* is overly simplistic (i.e., diversification rates are equal) with respect to the complexity in either model *Y* or model *Z* (i.e., diversification rates vary).

The story of the boy who cried wolf is a popular mnemonic for understanding what we mean when we refer to the difference between true type I and type II errors, which can be extended to include comparisons between complex and overly simplistic models. When the boy first cried wolf, but there was no wolf, he was making a type I error—that is, falsely rejecting the null of a wolf-free meadow. When the townspeople later ignored him when there was actually a wolf, they were making a type II error. If the sheep were instead perishing in a snowstorm, and the only options for the boy are to yell "no wolf!" or "wolf!" it is not clear what the best behavior is—"no wolf" implies no change in sheep mortality rates from when they happily gambol in a sunny meadow, even though they have begun to perish, whereas "wolf" communicates the mortality increase even though it is the wrong mechanism. It is the same here when looking at a tree coming from an unknown, but complex empirical branching process and trying to compare a constant rate model (no wolf) against a state-dependent (wolf), age-dependent (bear), or density-dependent model (snowstorm).

Another important issue is the misconception that any type of SSE model is a typical model of trait evolution like in, say, Pagel (1994) or Butler and King (2004). In these models, the phylogenetic tree certainly affects the likelihood of observing the traits, but that is the only role it plays. In contrast, other models based on the birth-death process for understanding tree growth and

shape (e.g., Nee et al. 1994; Rabosky and Lovette 2008; Morlon et al. 2011) only calculate the likelihood of the tree itself, ignoring any and all traits. An SSE model is essentially a combination of these: it computes the joint probability of the observed states at the tips and the observed tree, given the model. This is an important distinction because if a tree violates a single regime birth-death model, then even if the tip data are perfectly consistent with a simple trait evolution model, the tip data plus the tree are not. In such cases, the SSE model is very wrong in assigning rate differences to a neutral trait.

Beaulieu and O'Meara (2016) proposed a set of character-independent diversification (CID) models that are parameterized so that the evolution of each observable character state is independent of the diversification process without forcing the diversification process to be constant across the entire tree. Importantly, hidden state models are part of a more general framework that should be applied to any SSE model, regardless of a priori interest in unobserved factors driving diversification. For instance, the likelihood of a standard BiSSE model is *identical* to a HiSSE model (Beaulieu and O'Meara 2016) where the observed states each have their own unique diversification rates, with the underlying "hidden" states constrained to have the same parameter values. This is best illustrated in Figure 1. Both phylogenetic trees show variation in diversification rates, so the question is whether such rate shifts can be predicted by the observed traits or not. The state-dependent and state-independent models with respect to a focal trait both have two hidden states (*A* and *B*) and four free diversification parameters. The difference between the models in Figure 1 resides solely in the way the variation among the hidden states is partitioned. If we set the hidden rates to vary within each observed state, such that all hidden states of λ_0 and μ_0 are independent of λ_1 and μ_1 (Fig. 1, left panel), we produce a state-dependent BiSSE model. Alternatively, if the observed states share the same diversification rate, and rate variation is partitioned among hidden rate classes (Fig. 1, right panel), we produce a model independent of the focal trait. More complex state-dependent models can be created by letting multiple hidden rate categories be estimated within the observed state (see examples in Beaulieu and O'Meara 2016). The key benefit of the state-independent models (CID) is that they partition the rate variation among hidden rate categories and not among observed states.

A natural corollary, then, is that the usefulness of the hidden state modeling approach depends entirely on the careful match of free diversification parameters between the state-dependent and state-independent diversification models (as shown in Fig. 1). These models will also return "false positives" when the proper counterbalance to a state-dependent model is not included in the set of models evaluated. This was recently demonstrated by Rabosky and Goldberg (2017). Their nonparametric approach differentiated between cases of state-dependent and

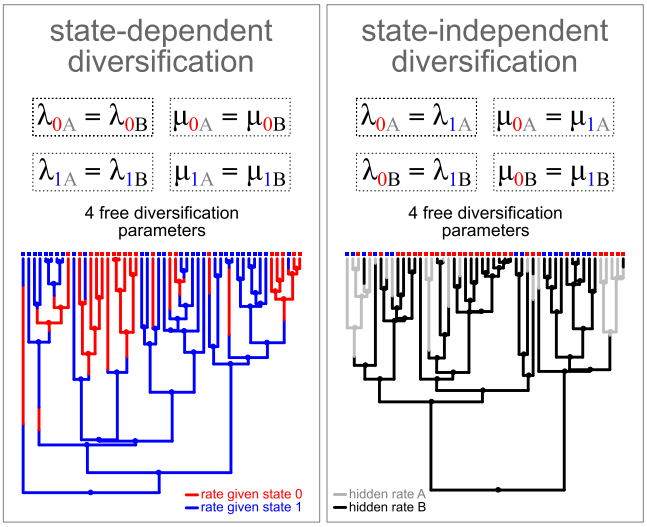

Figure 1. Hidden state models provide a framework to solve the issue with false relationships between shifts in diversification rates and traits. Left-side panel shows a case of state-dependent evolution. Here, shifts in rates of diversification in the tree are perfectly predicted by the transition between trait states 0 (red) and 1 (blue). Thus, no within-state rate variation is observed. Right-side panel shows state-independent shifts in diversification rates (hidden rate A in gray vs. hidden rate B in black) with respect to the focal trait (red and blue tips). Both models share the same number of free diversification parameters, but the variation among hidden states is partitioned differently. The state-independent model (right panel) allows for two diversification rate categories unrelated to the observed traits. A homogeneous rate null model would be inadequate for either of these trees.

state-independent diversification much better than a parametric, process-based HiSSE model under a range of benchmark scenarios (see PhyCoMB—https://github.com/phylosdd/PhyCoMB). Specifically, they found that although the inclusion of a character-independent model with two diversification rate categories (i.e., CID-2) reduced the overall "false positive" rate of the BiSSE model with no hidden state, the use of a set comprised by BiSSE, CID-2, and HiSSE models resulted in nine of 34 state-independent diversification scenarios having "false positive" rates in excess of 25%. However, we hasten to point out that this result is partly due to not including null models that are able to capture enough variation in diversification rates across the phylogenies. In fact, the model with the most distinct diversification rate categories tested by Rabosky and Goldberg (2017) assumed state-dependent diversification. A proper set of state-independent models for HiSSE methods should, necessarily, have the same number of free diversification parameters than the state-dependent models (four diversification rates in this case). When proper null models are included, the "false positive" rates dropped in all scenarios (see Fig. 2). Moreover, the improvement in performance was dramatic

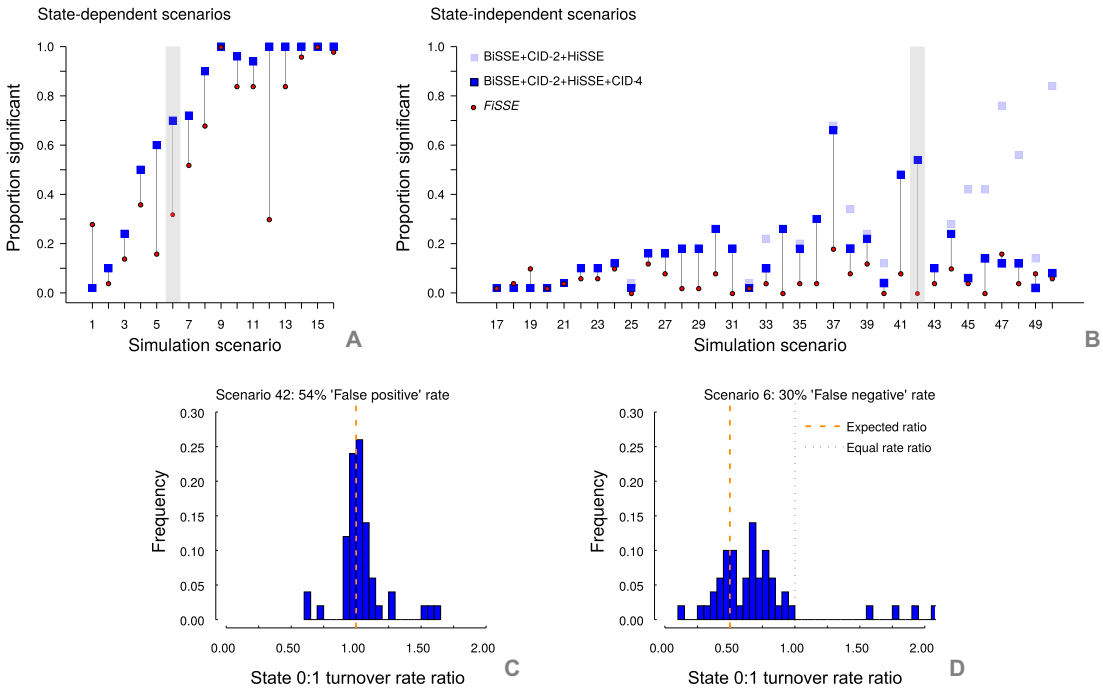

Figure 2. **A reanalysis of Rabosky and Goldberg (2017), which compared their nonparametric approach, FiSSE, against a parametric, process-based hidden state SSE model, HiSSE, under a range of benchmark simulation scenarios. Here we demonstrate the improvement of the SSE models when the four-state character-independent model, CID-4, is included in the model set (dark-blue boxes), which was not included in the original study (light-blue boxes). When this larger model set is compared against the original model set comprising just BiSSE, a two-state character-independent model, CID-2, and the character-dependent hidden state model HiSSE, the (A) power to detect the trait-dependent diversification remains unchanged. However, for the trait-independent scenarios (B), there is almost always a reduction in the "false positive" rate (as indicated by the difference in the position of the light-blue and dark-blue boxes), and in many cases the reduction is substantial. We model-averaged parameter estimates of turnover from two scenarios using our larger model set (the ones denoted in A and B by the light-gray box), a state-independent diversification scenario (C), which exhibited a 54% false positive rate (i.e., incorrectly rejected a state-independent scenario) and a true state-dependent diversification scenario (D), where the original model set showed high "false negative" rates (i.e., failed to reject a state-independent scenario 30% of the time). The dashed orange line represents the expected ratio of diversification rates between state 0 and 1 based on the simulation scenario, and the dotted black line denotes a ratio of 1, which is when rates between 0 and 1 are identical. In both cases, the center mass of the model-averaged rate ratios was centered on the "correct" generating values. In the case of the state-independent scenario, it clearly shows that despite the poor performance from a model rejection perspective, examining the model parameters indicated that, on average, there are no differences in diversification rates among observed states.**

in the same nine scenarios that previously showed high "false positive" rates (see Fig. 6 in Rabosky and Goldberg 2017). This reevaluation of results serves as a worked example of how not just any null model will suffice, but ones that are appropriately complex must also be included.

What (If Anything) is a Hidden Character?

Common difficulties that come with hidden Markov models applied to comparative analyses are "what exactly does the hidden state represent?" which leads to the most pressing question of "how should I interpret results when I find evidence for one or more hidden states influencing diversification?" Before moving forward, it is important to note a continuum across models of traits and diversification, such as the SSE models, and those that fit diversification rates to trees but ignore character information altogether (LASER: Rabosky 2006; MEDUSA: Alfaro et al. 2009; TreePar: Stadler 2011; BAMM: Rabosky 2014; among others). On one end lie models such as MEDUSA and BAMM that make no explicit hypotheses about how traits impact diversification. These models can be considered "hidden state only" models, in that shifts in diversification could be related to a single unmeasured character or, more realistically, to an evolutionary coordination among a suite of traits and trait-environment interactions. In other words, it is not controversial to assert that most characters have at least some influence on diversification, however trivial, and that even when not explicitly identifying a character focus, we are doing so implicitly. This is precisely why trait-based interpretations

often come as part of post-hoc interpretations that assert the putative causality of shifts in diversification inferred with MEDUSA or BAMM.

On the other end of the continuum, we may have a hypothesis about particular character states and their impact on diversification, which we test with any one of the many available SSE models. However, such questions now increasingly come with the added burden of mitigating factors that may erroneously produce meaningful differences in diversification among character states (Maddison and FitzJohn 2015; Beaulieu and O'Meara 2016), which ultimately leads us toward a middle ground of blending tree-only and strict SSE type models. For instance, as mentioned above, we must account for the possibility that diversification rates may actually vary independently of our special character of interest. Even when diversification rates are tied to a particular focal character state, we must also account for complicated correlations between our trait of interest and one or more unmeasured characters that can vary among clades (Beaulieu and O'Meara 2016). We should also account for additional processes, such as whether speciation events (cladogenesis) exert an effect on the character even if it may seem inconsistent with our initial hypothesis of how a particular state evolves. The important point here is that without accounting for any or all of these factors, and by explicitly staying on the strict end of the SSE spectrum, we run the risk of overstating the meaningfulness of the diversification differences among character states.

So, what, if anything, is the purpose of a hidden state model? Simply put, it is a means by which we can account for the hidden majority of factors while examining the meaningful impact of a suspected "driver" of diversification. The nature of the empirical question of whether character state *X* has a meaningfully higher rate than character state *Y* does not, and should not, change when including hidden characters. However, we emphasize that this type of question serves no purpose if character states *X* and *Y* alone are not adequate predictors for diversification rates. In this case, neither the "null" model, where diversification rates for *X* and *Y* are the same, nor the alternative model of state-dependent diversification will be adequate. If we ignore within-trait variation in diversification rates, it is really not surprising that erroneous conclusions are made. For these reasons, we advocate moving beyond just rejecting trivial null models, by making comparisons among a variety of models, looking at the weight for each, and making biological conclusions based on some summary of these models and their parameter estimates.

The Importance of Model Averaging

We adopt an approach that integrates estimate from every model in the set in proportion to how much the model is able to explain the pattern in the observed data using Akaike weight (w_i) (also see Silvestro et al. 2014 for an example of applying Bayesian model averaging). Model averaging largely alleviates the subjectivity of choosing thresholds to rank models, and permits the estimation of parameters taking into account the uncertainty in model fit. For instance, it is not unusual for a set of models, ranked according to their Akaike information criterion (AIC) values, or AIC weights, to suggest that three or four models fit about "equally well." The model choice framework provides no easy solution for such instances. Parameter estimates averaged across these models, on the other hand, will result in a scenario that incorporates the uncertainty associated with the fit of multiple models. For example, in the case of the datasets simulated by Rabosky and Goldberg (2017), the scenario that exhibited one of the worst "false positive" rates even after including the proper null model in the set had, on average, approximately the same model-averaged estimates of diversification rate for each of the observed states in the model (see Fig. 2). This indicated that there were no meaningful diversification rate differences among the observed states on the majority of the simulation replicates. More importantly, this is a clear example of a case in which "false positive" results are not as dire as they seem, because parameter estimates for the model show no strong effect of state-dependent diversification (see similar discussion in Cooper et al. 2016b).

We use a model-averaging procedure to summarize states and rates for every model in the set across a given tree. We begin by first computing the marginal probability of each ancestral state at internal nodes of the tree conditioned on the maximum likelihood parameter estimates using a standard marginal ancestral state reconstruction algorithm (Yang et al. 1995; Schluter et al. 1997). The marginal probability of state *i* for a focal node is the overall likelihood of the tree and data when the state of the focal node is fixed in state *i*. Note that the likeliest tip states can also be estimated, because while we observe state *1*, the underlying hidden state could either be *1A* or *1B*. For every model in the set and any given node or tip, we traverse the entire tree as many times as there are states in the model. The second step involves a weighted average of the diversification parameters for each hidden category using the marginal probabilities as the weights. For example, for focal node *n*, if the speciation rate is 0.8 for state *1A* and 0.5 for state *1B*, with marginal probabilities 0.3 and 0.7, the averaged speciation rate for state *1* at node *n* would be $0.8 \times 0.3 + 0.5 \times 0.7$. Finally, the rates and states obtained for every node and tip under each model in the set are averaged across all models using the Akaike weights. These model-averaged rates, particularly among extant species, can then be examined to determine the tendency of the diversification rates to vary among the observed character states. This procedure is implemented in the *hisse* package (Beaulieu and O'Meara 2016).

One important caveat about model averaging is making sure the models that are being averaged provide reasonable parameter estimates. Including a model with a weight of 0.00001, but a parameter estimate millions of times higher than other models' estimates for that parameter, will have a substantial effect on the model average. Even with this approach, examining results carefully, and communicating any issues and decisions to cull particular models transparently, remains important.

Similarities between BAMM and Model Averaging Using AIC Weights

Model-averaging approaches have been widely used in phylogenetics (e.g., Posada 2008; Eastman et al. 2011; Rabosky 2014; Silvestro et al. 2014), but most applications have focused on Bayesian methods. For example, BAMM (Rabosky 2014) applies a reversible-jump Markov-chain Monte Carlo (rjMCMC) sampler (Green 1995). The "reversible-jump" part means that the MCMC will visit multiple models by adding and subtracting parameters using birth and death proposal steps. This is a Bayesian method, thus a prior is used to control the weight given to different numbers of rate shifts (Rabosky 2014). The rjMCMC will adjust the complexity of the model according to the likelihood weighted by the prior (i.e., the posterior). This approach returns a posterior distribution of shifts in diversification rates and locations in the phylogenetic tree.

The SSE state-independent models using hidden states (CID: Beaulieu and O'Meara 2016) have a similar purpose to the trait-agnostic analyses of rates of diversification performed with BAMM. In both cases, we want to compute the likelihood of the tree including rate heterogeneity in the absence of any (explicit) predictor. As we discussed before, the important attribute of the CID models is that these can accommodate rate heterogeneity using hidden states. Because the true number of rate shifts in empirical trees is unknown, we need to fit multiple state-independent models by varying the number of hidden rate categories. Although BAMM has a built-in machinery to grow and shrink the number of rate regimes in the models as part of the rjMCMC, HiSSE (as well as GeoHiSSE [geographic hidden state speciation and extinction], which we describe below) requires the user to provide a set of models with an adequate number of rate categories (i.e., hidden states). However, it is possible to dredge across a set of possible models with a simple script and summarize results using model averaging (see examples on https://doi.org/10.6084/m9.figshare.c.4069580.v2).

The similarity between the application of model averaging in a likelihood framework and Bayesian model averaging is that, in both cases, the result is a collection of models weighted by some quantity proportional to a measure of fit. In the case of model averaging across a set of maximum likelihood estimates, the quantity used is the Akaike weight that is considered the weight of evidence in favor of a given model relative to all other models in the set (Burnham and Anderson 2002). In a Bayesian framework, the quantity is the frequency of a given model in the posterior distribution, which is proportional to the posterior probability of the model given the observed data (Green 1995). Both the rjMCMC used by BAMM and the model-averaging approach described here have the same ultimate goal and allow us to investigate similar questions. Given the recent popularity of rjMCMC approaches in phylogenetics, it seems natural that tests of state-dependent diversification using likelihood should focus on parameter estimates while incorporating the contribution of each model to explain the observed data and avoid the pitfalls of model rejection.

Linking "Hidden" States to Geographic Range Evolution and Diversification

The dispersal-extinction-cladogenesis models (DEC: Ree and Smith 2008; DEC+J, "jump dispersal": Matze 2014) are popular frameworks for studying geographic range evolution using a phylogenetic-based approach. They are different from SSE models in that they are standard trait evolution models—that is, the likelihood *only* reflects the evolution of ranges and the tree is considered fixed. Recently, Ree and Sanmartín (2018) raised concerns with the limitations of DEC and DEC+J models. Because the probability of the tree is not part of the DEC likelihood, cladogenetic events are independent of time, which produces odd behaviors in the parameter estimates, especially when jump dispersal events are allowed (Ree and Sanmartín 2018). A straightforward way to alleviate this issue, as mentioned by Ree and Sanmartín (2018), is to simply apply GeoSSE models (Goldberg et al. 2011) to study range evolution. Instead of optimizing ancestral areas conditioned on a fixed tree, the GeoSSE model incorporates the tree into the likelihood and has a parameter for the rate of cladogenetic speciation.

In our view, the concept of hidden variation seems the most relevant when investigating geographic range evolution. Due to the need for reducing geographical variation into coarsely defined discrete areas, parameter estimates could be strongly impacted by heterogeneous features across the landscape not captured by this categorization. A good example of the potential for hidden variation in range evolution comes from studies of diversity dynamics between tropical and temperate regions, which are often defined simply by latitude (e.g., tropical for $|\text{latitude}| < 23.5°$; Rolland et al. 2014). Such categorization necessarily overlooks the heterogeneity present in the tropics: some high elevation areas freeze,

others do not; some areas are deserts and others are lush forests, etc. However, the ability for the data to "speak" to the rate variation within a given geographic area is not currently allowed within the existing GeoSSE framework. Below, we briefly demonstrate how the HMM approach can easily be used for GeoSSE models. However, this is just one example of the utility of HMM approaches in diversification models; HMM approaches can, and we argue should, be added as options to other diversification approaches.

GEOGRAPHIC HIDDEN STATE SPECIATION AND EXTINCTION MODEL

The general form of the original GeoSSE model determines the diversification dynamics within, and transitions between, two discrete regions *0* and *1*. Under this model (see top panel in Fig. 3), a species observed at the present ($t = 0$) can be "endemic" to either area *0* or *1*, or has persistent populations in both *0* and *1*, referred to hereafter as the *01* range (i.e., "widespread"). Similar to the DEC model, range evolution occurs in two distinct modes. First, ranges can expand or contract along the branches of the phylogeny through anagenetic change. Range expansions are based on state transitions from *0* to *01* and *1* to *01*, which are parameterized in the model as the per-lineage "dispersal" rates, d_0 and d_1, respectively. Range contractions, on the other hand, describe the reverse process of transitions from *01* to *1* and *01* to *0*, which are the per-lineage rates of range contraction, x_0 and x_1, respectively (also referred as "extirpation" rates). The second mode of range evolution occurs as a product of the speciation process (i.e., cladogenesis), particularly with respect to speciation events breaking up widespread ranges into various combinations of descendant areas. The area-specific rates of "within-region" speciation are parameterized as s_0 and s_1, whereas the "between-region" rate of speciation is denoted by s_{01}. We will refrain from describing the mathematical formulation of this particular model, as these are described in detail elsewhere (Goldberg et al. 2011; Goldberg and Igic 2012). We do note, however, that our notation for the observed areas differs from Goldberg et al. (2011) to be consistent with previous work on incorporating hidden states into SSE models (see Beaulieu and O'Meara 2016).

To apply the HMM to GeoSSE for the simplest case of two hidden states, we replicate the original model across hidden states *A* and *B*. We reparameterize the model to include six distinct speciation rates, s_{0A}, s_{1A}, s_{01A}, s_{0B}, s_{1B}, and s_{01B}, and four distinct extinction rates, x_{0A}, x_{1A}, x_{0B}, and x_{1B}, allowing for three distinct net diversification rates within each observed range (Fig. 3, bottom panel). Likewise, dispersal rates from area *0* (or *1*) into *01* also show separate rates for each hidden state, parametrized as d_{0A} and d_{0B} for area *0* or d_{1A} and d_{1B} for area *1*. Shifts between the hidden states within a geographic range are modeled with the transition rate q following the same approach described by Beaulieu and O'Meara (2016), as does the inference of changes in the hidden state within each range. Herein we focus our simulation tests and empirical analyses on a spatial structure that contains only two geographical areas; however, just like the DEC model, our GeoHiSSE model can contain an arbitrarily large state space.

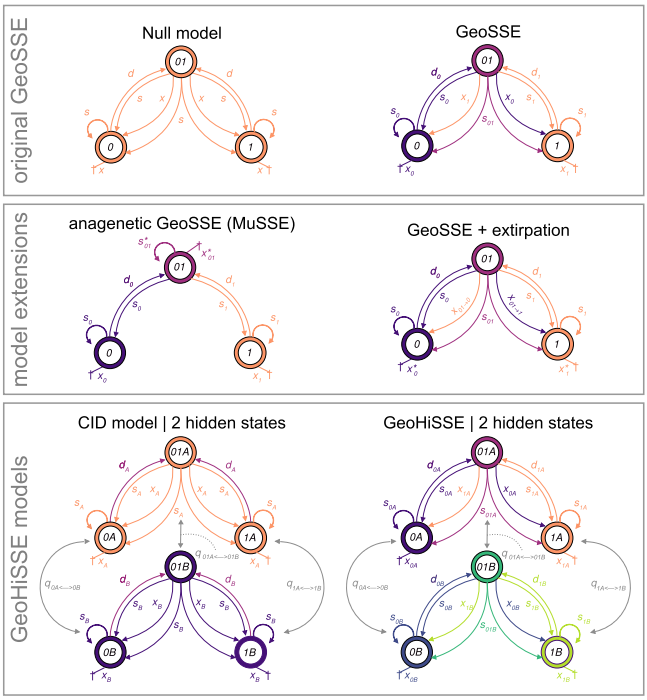

Figure 3. States and allowed transitions in the original GeoSSE model (Goldberg et al. 2011) and the models described herein. The colors denote parameters that are associated with each character combination modeled as having their own unique set of diversification rate parameters. The simple GeoSSE null model has a single rate of diversification compared with three diversification parameters in the area-dependent model. The anagenetic GeoSSE model is a special case of the MuSSE model parameterized as to emulate the transitions allowed by the original GeoSSE model. The GeoSSE+extirpation model separates rates of range reduction (e.g., $x_{01\to1}$) from the extinction of endemic lineages (e.g., x_0), but otherwise contains three unique sets of diversification parameters as in the original GeoSSE model. The area-independent GeoHiSSE model (CID) (denoted by two sets of diversification parameters shown in orange and purple) and the area-dependent GeoHiSSE model (full model) (denoted by six sets of diversification parameters shown in various colors) can have two or more hidden states. Note that all model extensions can support hidden states.

The GeoSSE model set must also be expanded to include less trivial area-independent models that incorporate rate heterogeneity from phylogenetic trees. Herein we refer to hidden state area-independent models as CID following Beaulieu and O'Meara (2016). Recently, Huang et al. (2018) implemented a CID model to obtain an area-independent model for use in model comparisons against GeoSSE. The CID model of Huang et al. (2018)

replicates the GeoSSE model across two hidden states, *A* and *B*, similar to our GeoHiSSE model described above, but constrains the between-area speciation rates $s_{01A} = s_{01B} = 0$. However, here we constrained the speciation and extinction rates to be equal in each of the hidden states such that $s_{0A} = s_{1A} = s_{01A}$, $x_{0A} = x_{1A}$, $s_{0B} = s_{1B} = s_{01B}$, and $x_{0B} = x_{1B}$ (Fig. 3, bottom panel). Note that geographic area is not completely independent from diversification due to the "effective rate of diversification" within the widespread area for a given hidden state being the sum of s_{0A}, s_{1A}, s_{01A}. There is also a global transition rate (e.g., $d_{0A\rightarrow 0B}$), which accounts for transitions among the different hidden states within each range and disallows dual transitions between areas and hidden states (i.e., $d_{0A\rightarrow 1B} = 0$). We expand the area-independent model to allow geographic ranges to be associated with as many as five different hidden states (i.e., $h \in A, B, C, D, E$). The purpose of these models is to prevent spurious assignments of diversification rate differences between observed areas in cases where diversification is affected by other traits.

FURTHER MODEL EXPANSIONS

We also relaxed and tested the behavior of two important model constraints within the GeoSSE framework. First, we wondered whether constraining range contraction and lineage extinction to be the same could be too restrictive, particularly when the ranges under consideration represent large geographical areas. Under the original GeoSSE model there are two parameters, $d_{01\rightarrow 0}$ and $d_{01\rightarrow 1}$, which denote local range contraction that are linked to extinction rates of the endemics, x_0 and x_1 (Fig. 3, top panel). However, consider a scenario where lineages have originated in a temperate region and possess a suite of traits that reduce extinction rates in this area. Movements into the tropical regions require not just getting there, but also being able to compete within this new environment. Recent attempts to disperse into the tropical zone by those lineages on the boundary separating the two areas can persist there for a time, but might eventually go locally extinct in the tropical portion. The constraint of the rate of range contraction always equaling the extinction rate of endemics prevents the detection of such dynamics. Furthermore, this will necessarily increase estimates of per-lineage extinction rates for the tropical region as a whole because of the link between range contraction and lineage extinction present in the original GeoSSE model (Goldberg et al. 2011). Herein we extended the model by separating the rate of range contraction from the process of lineage extinction (see more details in Fig. 3 and Supporting Information). More broadly, we refer to this class of models as "GeoSSE+extirpation" or "GeoHiSSE+extirpation" to represent models with and without hidden states, respectively. Removing this constraint effectively increases the number of state "transition rates" in the model. Teasing apart the effect of range reduction and extinction of endemics will likely require phylogenetic trees of large size. For instance, our simulations show parameter estimates for the GeoHiSSE+extirpation model are adequate with trees of 500 species (see Supporting Information).

Second, we included a complementary set of models that removed the cladogenetic effect from the model entirely, such that all changes occur along the branches (i.e., anagenetic change). This requires the addition of a per-lineage rate of extinction and speciation for lineages in the widespread range (x_{01} and s_{01}, respectively) as well as range contraction being distinct from the extinction of endemics. In the absence of a hidden state, this is effectively a three-state MuSSE model (FitzJohn 2012) with the transition matrix constrained such that a shift between ranges *0* and *1* has range *01* as the intermediary state. Here we also expand this particular MuSSE-type model to allow for up to five hidden states, which can be used to test hidden state, state-dependent, or state-independent diversification models (Table 1). In general, we include this particular set of models as a way of acknowledging that we really never know the "true" history of the characters or areas we observe. This is especially relevant when defined ranges are separated by limits permeable to genetic flow among populations (e.g., Goldberg et al. 2011; Bloom et al. 2014 in which noncladogenetic speciation of widespread lineages could occur, even if unlikely. Of course, these anagenetic models might provide little biological insight when applied to large geographical regions, such as continents (see empirical example below), or limits only permeable to rare dispersal events (e.g., Matzke 2014). However, this particular parameterization allows for complete area-independence due to the widespread state being treated as any other state in the model. We refer to this class of models hereafter as "anagenetic."

SIMULATIONS

We performed extensive simulations to test the behavior of the GeoHiSSE and model expansions. We evaluated models of area-dependent and area-independent diversification under a series of scenarios, including unequal frequencies between observed ranges and absence of cladogenetic events. We have relegated many of the details and tests to the Supporting Information; Section 1 tests the behavior of the original GeoSSE model under trait-independent shifts on rates and shows a clear issue with model misspecification akin to BiSSE, whereas Section 2 provides results for the simulations, which we summarize below. Briefly, we generated 100 trees containing 500 taxa for each simulation scenario. Thus, our results are relevant to trees of similar size or larger and we strongly suggest users to perform power analysis when using smaller trees. All analyses were carried out using new functions provided in the R package *hisse* (Beaulieu and O'Meara 2016). Code to replicate the simulations is also available in the data repository (https://doi.org/10.6084/m9.figshare.c.4069580.v2).

Table 1. List of 18 fitted models.

Category	Model	Description	Free parameters
CLADOGENETIC	1	CID—original GeoSSE	4
	2	Original GeoSSE, full model	7
	3	CID—GeoHiSSE, three hidden rate classes, null model	9
	4	GeoHiSSE, two rate classes, full model	15
	5	CID—GeoHiSSE, five hidden rate classes, null model	13
	6	CID—GeoHiSSE, two hidden rate classes	7
CLADOGENETIC +EXTINCTION	7	CID—GeoSSE+extirpation	6
	8	GeoSSE+extirpation, full model	9
	9	CID—GeoHiSSE+extirpation, three hidden rate classes, null model	11
	10	GeoHiSSE+extirpation, two hidden rate classes, full model	19
	11	CID—GeoHiSSE+extirpation, five hidden rate classes, null model	15
	12	CID—GeoHiSSE+extirpation, two hidden rate classes	9
ANAGENETIC	13	CID—anagenetic GeoSSE	6
	14	Anagenetic GeoSSE, full model	10
	15	CID—anagenetic GeoHiSSE, three hidden rate classes, null model	11
	16	Anagenetic GeoHiSSE, two hidden rate classes, full model	21
	17	CID—anagenetic GeoHiSSE, five hidden rate classes, null model	15
	18	CID—anagenetic GeoHiSSE, two hidden rate classes	9

***Note*: Columns show model category, model number, description of the model, and number of free parameters. Area-independent models (CID) have no relationship between diversification rates and geographic areas. "Full model" indicates that all parameters of the model are free, whereas "null model" indicates that diversification and dispersion parameters are constrained to be equal among areas in the same hidden state category. If not "full model" or "null model," the description column lists the free parameters of the model. Models indicated with "+extirpation" separate rates of range reduction from the extinction of endemic lineages.**

For the first simulation scenario (scenario A), we simulated data using a homogeneous rate GeoSSE model with the speciation rate for one of the endemic areas (area *1*) set to be two times faster than the other two possible areas (*0* and *01*). This represented the simplest case for which the original GeoSSE model is known to be adequate (Goldberg et al. 2011). It also allowed us to test whether models with multiple hidden rate classes exert undue weight when rate heterogeneity is not actually present in the data (i.e., overfitting problems). We found that even when the model set included a broad array of complex models, most of the model weight across all replicates for scenario A goes to the generating model (model 2; Fig. S8). Furthermore, even when net diversification rates between endemic ranges are averaged across all models, our estimates were congruent with the true values (Fig. 4A). When we expressed speciation and extinction rates in terms of turnover rates of an endemic area (i.e., $s_i + x_i$) and extinction fraction (i.e., x_i / s_i), the rate estimates for each node in the tree are also centered on the true parameter values, independent of tree height (Fig. S5).

In the second and third scenarios (scenarios B and C), we introduced three and five range-independent diversification regimes, respectively. The transition between diversification rate classes followed a meristic Markov model to emulate gradual changes in diversification rates (Fig. S1). In both cases, we did not detect any meaningful differences in the net diversification rates between endemic areas (Figs. 4B, C, and S3), and the parameter estimates computed for each node and tip of the phylogeny are centered on the true values. Thus, our results show

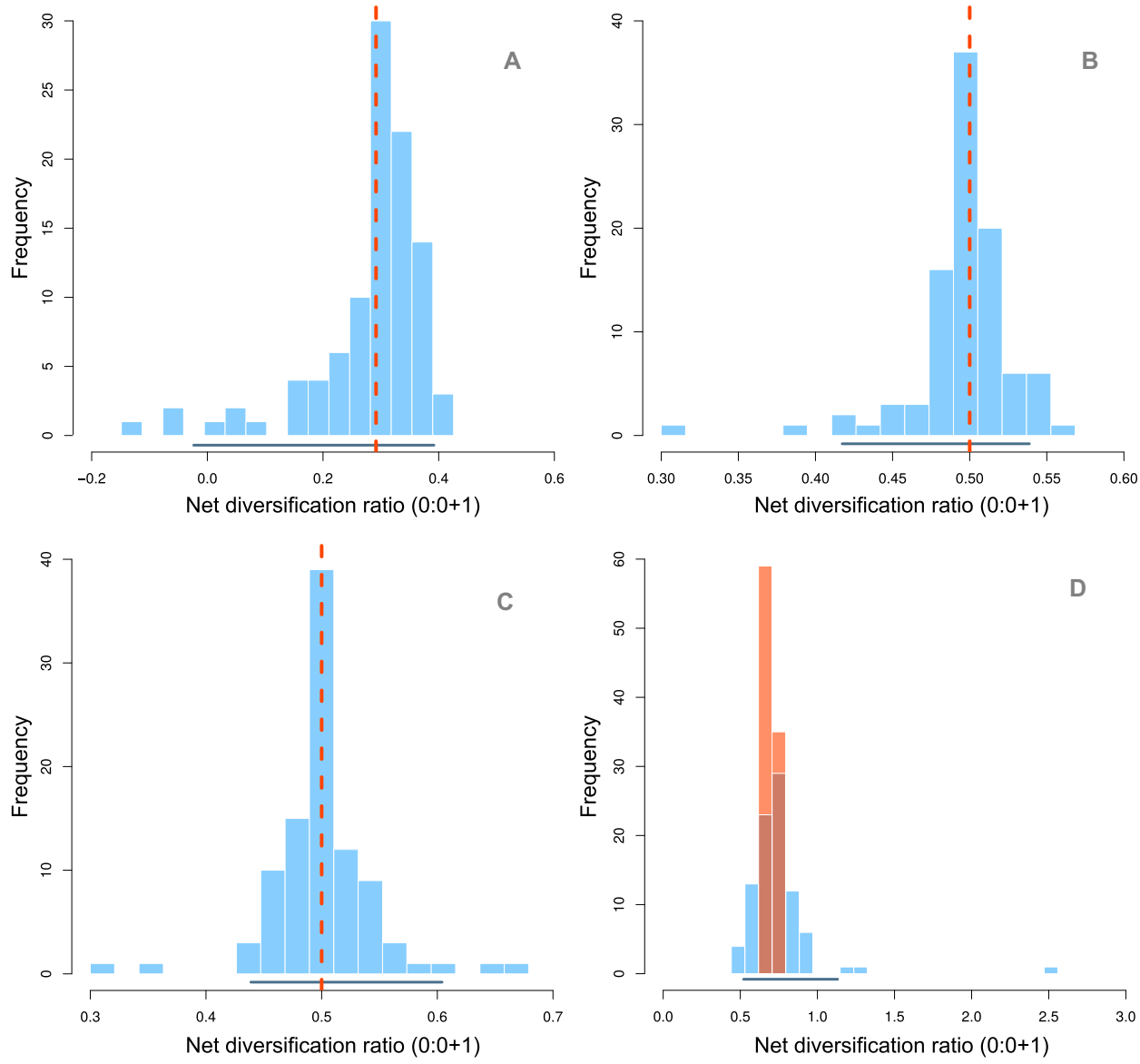

Figure 4. Effect of endemic ranges on net diversification estimated for simulation scenarios A to D. Plots show the distribution of ratios between net diversification rates for areas *0* and the sum of net diversification rates for both endemic areas (*0 + 1*) computed for each simulation replicate. Red dashed vertical lines in plots A to C represent the true value for the ratios. Horizontal lines in the bottom show the 95% density interval for each distribution of parameter estimates. Plot D shows the distribution of true ratios in orange. Estimates are the result of model averaging across 18 different models using Akaike weights. Values below 0 (plot A) or above 1 (plot D) are due to few replicates showing negative net diversification rates. See Table 1 for the list of models and Figure S3 for AIC weights.

that area-independent GeoHiSSE models can accommodate the rate heterogeneity in the simulated trees without evoking an area-dependent diversification process when none was present. We do note, however, that even when we reduced the model set and fit only simple homogeneous GeoSSE models to the same data, there is an interesting effect in that the parameter estimates averaged across all models would still lead to an area-independent diversification interpretation (see Supporting Information—Section 2).

For simulation scenario D, we represented an instance of the area-dependent model in which geography has an important effect on diversification across the phylogeny, but diversification rates vary within each range as a function of some unobserved "hidden" trait. The frequency of each hidden rate class stochastically varies across simulation replicates, and, therefore, there is no single true value for diversification rates. Nevertheless, results showed that GeoHiSSE was able to recover the correct direction in the

relationship between net diversification rates of endemic areas in the presence of heterogeneity (Figs. 4D, S5, and S6D).

We also studied two extreme cases with the objective of identifying odd behaviors when simulating datasets where (1) widespread ranges are rare or absent in extant species, and where (2) the evolution of areas are not tied to cladogenetic events. Under the original GeoSSE model, lineages need to pass through the widespread state before transitioning between endemic areas. If extant widespread lineages are rare or absent, the information to infer cladogenetic and dispersion events can become limited. To study this effect, we first simulated datasets with widespread lineages as being rarely observed at the tips (see scenario E in Table S1). Results showed that low frequency of widespread lineages does not prevent our set of models from reaching meaningful estimates using model averaging (Fig. 5E). Alternatively, we simulated the case of jump dispersal events (i.e., direct transitions between endemic distributions). For this we used a GeoSSE model to simulate the data, but we allowed lineages to disperse between endemic areas without becoming widespread first (scenario F). (Note that none of the 18 models we used to estimate parameters throughout this study allow for any instance of jump dispersal event.) Our results showed no evidence for a consistent bias in parameter estimates for both area-dependent diversification rates and speciation rates associated with cladogenetic events on widespread lineages (Fig. 5F). In summary, our approach of model averaging across a large set of candidate models does not appear sensitive to rare extant widespread areas.

Finally, we explored the extreme possibility that the widespread range was never a part of the history of the clade (scenario G). When fit to our model set, the absence of widespread areas among the extant species produces estimates of the rates of between-area speciation (s_{01}) that are highly uncertain (Fig. 5G). These estimates are orders of magnitude higher than the rates of within-area speciation. In contrast, estimates for the relative difference in net diversification rates between endemic areas did not show a strong bias (Fig. 5G). This suggests that poor estimates of between-area speciation would not strongly bias our conclusions about range-dependent diversification rates.

All previous scenarios assumed that cladogenetic events were important in the evolutionary history of the lineages. To consider the performance of the model when this is not the case, we generated datasets with transitions between areas restricted only to anagenetic dispersal events along the branches of the tree. The estimated difference in rates of within-area diversification is larger than observed in any other simulation scenario (Fig. 5H). Moreover, the absence of cladogenetic events makes estimates of between-area speciation uncertain, although raw parameter values are within the same order of magnitude of the true rates of diversification across the tree (gray lines in Fig. 5H).

On the whole, our extensive simulation study shows that parameter estimates averaged across 18 models of area-independent and area-dependent evolution are robust to a wide variety of macroevolutionary scenarios likely to be observed in empirical datasets. We also show that important violations of the expected type of data modeled with GeoSSE, such as absence of widespread lineages or cladogenetic speciation events, are not enough to significantly hinder our interpretation of the evolutionary history of the group, even when there is large ambiguity in the estimate for some parameters of the model.

Empirical Example: Hemisphere-Scale Differences in Conifer Diversification

A further question is the performance of the GeoHiSSE model, as well as our extensions to the original GeoSSE model, in an empirical setting, where we do not know the generating model. For these purposes, we focus our analyses on the evolutionary dynamics of movements between "Northern" and "Southern" conifers. There is evidence that the turnover rate—defined here as speciation + extinction rate—is generally higher for clades found exclusively in the historically Northern Hemisphere influenced floras compared to clades found almost exclusively in largely Southern Hemisphere influenced floras (Leslie et al. 2012). The falling global temperatures throughout the Cenozoic, and concomitant movements of several major landmasses northward, facilitated the emergence of colder, drier, and strongly seasonal environments within Northern Hemisphere regions (e.g., Zanazzi et al. 2007; Eldrett et al. 2009). This may have led to widespread extinction of taxa unable to survive in such environments and expansion of taxa able to thrive there (perhaps through isolated populations surviving to become new species rather than go extinct due to competition). The net effect across the clade would be an increase in speciation and extinction rates. Furthermore, the repeated cycles of range expansion and contraction due to glacial cycles would also promote isolation of populations leading both to speciation (due to allopatry), and possibly rapid extinction (due to small population size). The Southern Hemisphere, on the other hand, has maintained milder environments scattered throughout the region (e.g., Wilford and Brown 1994; McLoughlin 2001). It is important to note that these conclusions were supported by comparisons of branch length distributions and diversification models applied to various clades independently, which indicated heterogeneity among the taxonomic groups tested (see Leslie et al. 2012).

Using the expanded GeoSSE framework, we reexamined the "hemisphere" diversification differences within conifers proposed by Leslie et al. (2012). We combined geographic locality information from the Global Biodiversity Information Facility

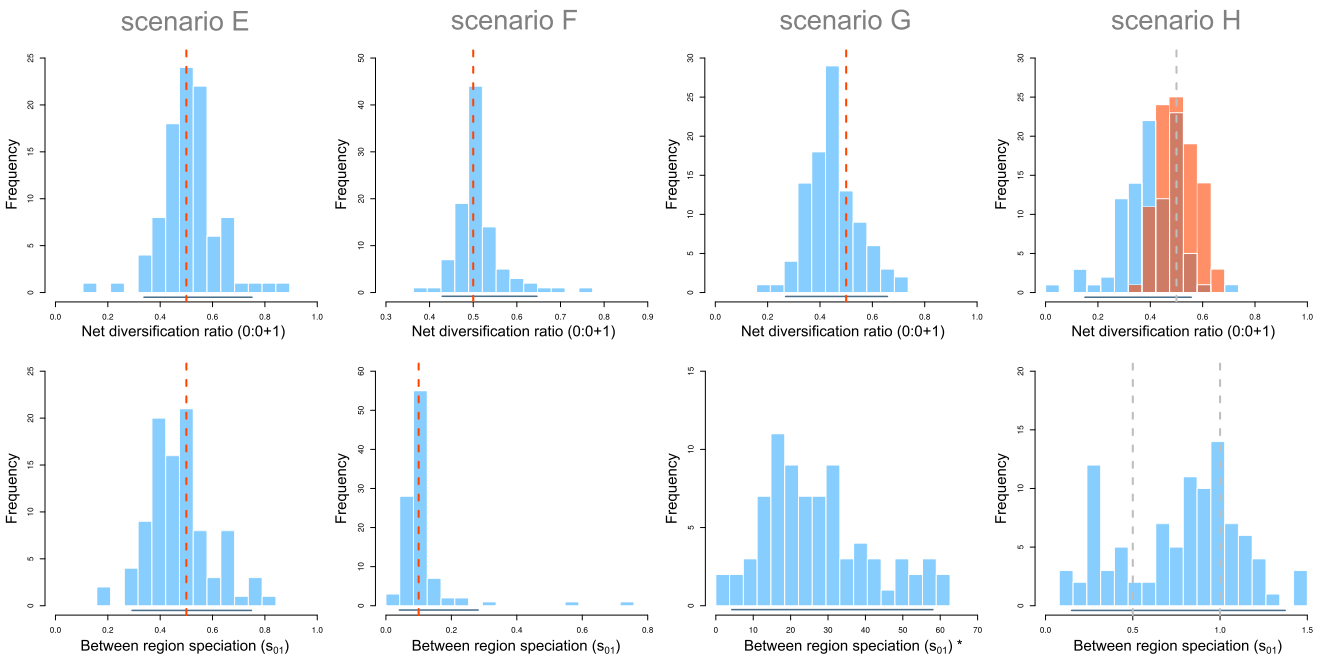

Figure 5. **Net diversification ratios between endemic areas and between region speciation rates estimated for simulation scenarios E to F. Upper row shows the distribution of ratios between net diversification rates for areas *0* and *0 + 1* computed for each simulation replicate. Lower row shows the distribution of between region speciation rates specific to the widespread area (parameter s_{01}) averaged across all tips (for E, F, and H) or nodes (for G) of the phylogeny for each simulation replicate. Red dashed vertical lines represent the true value for the parameter. Gray dashed lines mark important reference points but are not the expected value for the quantities. Upper plot H shows the distribution of true ratios in orange (see main text). Horizontal lines in the bottom show the 95% density interval for each distribution of parameter estimates. The "*" marks results based on the average across nodes instead of tips (no data available at the tips). Estimates are the result of model averaging across 18 different models using Akaike weights. See Table 1 for the list of models and Figure S4 for AIC weights.**

database (GBIF) with an updated version of the dated conifer tree from Leslie et al. (2018) that improves taxon sampling relative to the analysis of Leslie et al. (2012). This new phylogeny contains 578 species, representing around 90% of the recognized extant diversity. We note that by referring to Northern and Southern Hemispheres, we are not referring to the arbitrary 0° latitude separating the two regions. Rather, we are referring to the geographic regionalizations of the floras that contain conifer representatives. To this end, we considered any species distribution having a maximum and minimum latitude $> 10°$ as being Northern Hemisphere, $< 10°$ as Southern Hemisphere, and species distributions with a maximum latitude $> 10°$ and a minimum latitude $< 10°$ were considered "widespread" (A. Leslie, pers. comm.). Such strict thresholds in latitude used to define ranges provide the ideal scenario in which hidden states may play an important role in understanding the diversification dynamics across the clades. Finally, we pruned the Pinaceae from our analysis and focus only on movements within the Cupressophyta, which includes the Cupressales (i.e., cypresses, junipers, yews, and relatives) and the Araucales (i.e., *Araucaria*, *Agatha*, podocarps, and relatives). The decision to remove Pinaceae from our analysis was based on the uncertain relationship of Gnetales to conifers. There remains the possibility that Gnetales is sister to conifers as a whole (e.g., the "Gnetifer" hypothesis; Chaw et al. 1997; Burleigh and Mathews 2007), although most recent sequence analyses support Gnetales as sister to Pinaceae (e.g., the "Gnepine" hypothesis; Mathews 2006, 2009; Zhong et al. 2010; Ran et al. 2018). For these reasons, we focus our analyses on the Cupressophyta to ensure that our analyses reflect geographic range evolution within a monophyletic group.

Our final dataset consisted of 325 species, with 136 species designated as Northern Hemisphere, 167 designated as Southern Hemisphere, and the remaining 22 species currently persisting in both areas. Our model set included the 18 models described in Table 1 and used in our simulations, as well as an additional 17 models. Briefly, we included area-independent GeoHiSSE models that ranged from two to five hidden states rather than just those that equal the number of parameters in the area-dependent models (i.e., models consisting of three and five hidden states). We also added various MuSSE-type models that allowed and disallowed range contraction to be separate from lineage extinction and a particular set of MuSSE-type models that disallowed speciation

and extinction (i.e., the rates were set to zero) in the widespread regions to better mimic *anagenetic-only* range evolution. The entire set of models tested and their number of free parameters are described in Table S2.

The turnover rate differences, even after accounting for hidden states and the possibility of heterogeneity in area-independent diversification, supported the original findings of Leslie et al. (2012). That is, Northern Hemisphere species across Cupressophytes exhibited higher turnover rates relative to species occurring in the Southern Hemisphere (Fig. 6). The majority of the model weight is on an area-dependent GeoHiSSE model (model 4), with the remaining weight coming from one area-dependent and two area-independent GeoSSE models, all of which assumed range contraction is separate from lineage extinction (models 10–12, respectively, each described in Tables 1 and S2). The strong weight on a GeoHiSSE model with a single hidden state implied that the differences in the turnover rate for a given endemic range were similar among clades (contra Leslie et al. 2012). After incorporating uncertainty within and among models, the turnover rates within the Northern Hemisphere Cupressales (i.e., Taxaceae+Cupressaceae) and Araucariales (i.e., Araucariaceae+Podocarpaceae) have a turnover rate that was generally 2.5 times higher than Southern Hemisphere species.

Dealing with Mixed Signals among Hidden States

In cases where diversification rates (or any other rate of interest) are always higher in one particular observed state than the other for any hidden state, interpretation of the results is fairly straightforward (e.g., Fig. 6). Mapping the model-averaged diversification rates to the phylogenetic tree can also provide important, and often insightful, phylogenetic context to among-species variation in parameter estimates. However, in some instances, ignoring rate differences between combinations of observed and hidden traits could be problematic. Whether one observed state leads to higher diversification rates than the other could depend on the magnitude of the rate differences and how much time is spent in each hidden state. When the rate of diversification for $0A > 1A > 1B > 0B$, the average mean rate for *0* could be equal, higher than, or lower than the rate for state *1*. Similarly, a result with $0A = 1B$ and $0B > 1A > 0A$ could reflect area-dependent diversification on the majority of lineages if hidden states *0B* and *1A* are much more frequent than *0A* and *1B* or area-independent diversification otherwise.

As a solution to this, we recommend model averaging the "effective state proportion," which is the expected proportion of each geographic area and hidden state combination according to the equilibrium frequencies given the model parameters, tree depth, and root weights (see Appendix). This provides a useful complement to examining just rate differences among observed states. In other words, even in cases where $0A > 1A > 1B > 0B$, such that the average diversification rate among hidden states is more or less the same, we could still find that, say, 75% of species are expected to be in state *0* due to the interaction with the hidden state. This approach can create scenarios in which the majority of lineages are associated with an area-dependent diversification process, but not all. Although one might argue that hypotheses of state-dependent diversification, as tested using SSE models, predict a consistent effect of the trait across the phylogeny, in our view, it seems more realistic to consider that a small proportion of lineages are not congruent with this pattern and still accept an important effect of state-dependent diversification as a whole (see similar point in Uyeda et al. 2018).

Note that the effective state proportion may differ radically from the empirical frequency. Perhaps an unlikely series of changes early in the tree led to more taxa in state *1* than would be expected. Likewise, it is possible to have more taxa in state *1* even if the net diversification rate in state *1* and transition rate to state *1* are lower than the equivalent rates for *0*. Given this scenario, one should conclude that there is signal for a state-dependent process with observed state *0* having a positive effect on diversification, despite the higher frequency of taxa associated with state *1*. In the case of Cupressophytes, the model-averaged state proportions indicated that roughly 63.1% (±1.6%) of species are expected to occur in the Southern Hemisphere, with 31.2% (±1.3%) of conifers expected to be in the Northern Hemisphere, and 5.6% (±1.0%) of the species widespread across both regions. These expectations are largely congruent with the estimates of turnover rates between Northern and Southern Hemispheres, where the increased "boom and bust" dynamics in the Northern Hemisphere floras produces diversity within the region at much lower levels than in the floras of the Southern Hemisphere.

Caveats

The GeoHiSSE family of models are useful approaches for understanding diversification when there can be many processes at play. However, it is important that we emphasize that, like all models, they are far from perfect. First, the hidden states evolve under a continuous time Markov process, which is reasonable for heritable traits that affect diversification processes. Of course, not all heterogeneity arises in such a way. For instance, when an asteroid impact throws up a dust cloud, or causes a catastrophic fire, every lineage alive at that time is affected simultaneously. Their ability to survive may come from heritable factors, but the sudden shift in diversification caused by an exogenous event like this appears suddenly across the tree, in a manner not yet incorporated in these models. Similarly, events of clade-specific mass extinction or a secular trend affecting all species are not currently part of this

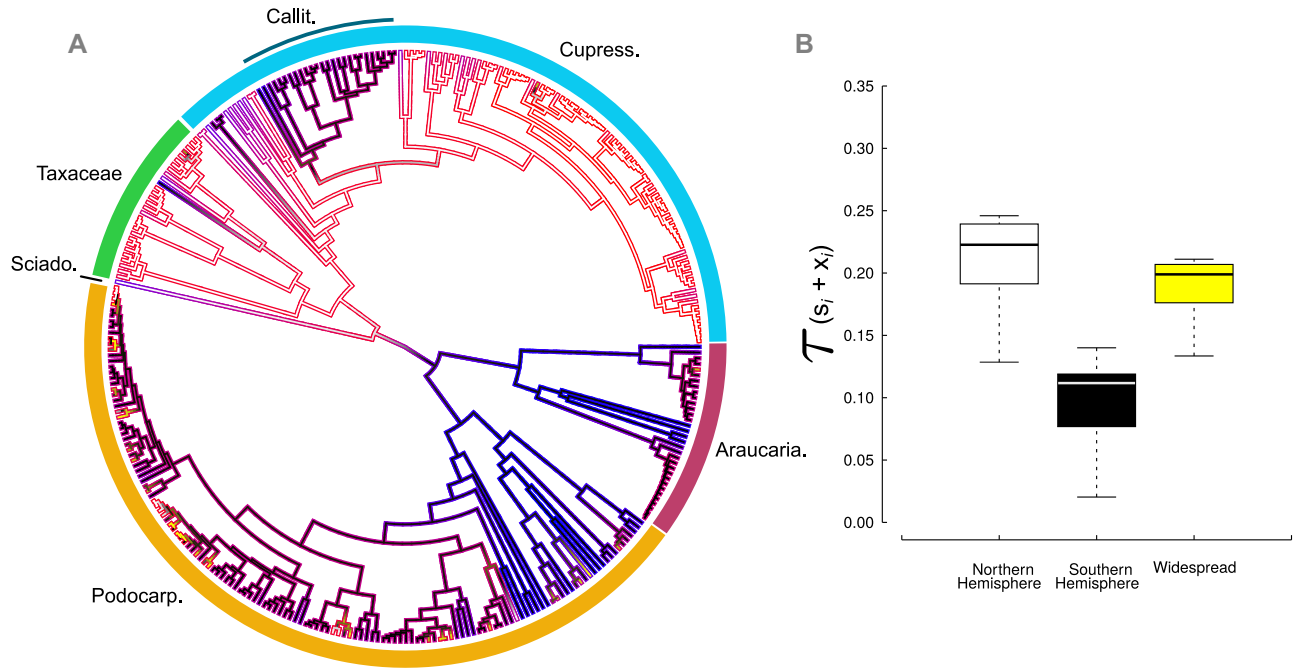

Figure 6. **(A) Geographic area reconstruction of areas and turnover rates (e.g., $\tau = s_0 + x_0$ for endemic range *0*, and $\tau = s_0 + s_1 + s_{01}$, for the widespread range, *01*) across a large set of models of Northern Hemisphere (white branches), Southern Hemisphere (black branches), and lineages in both (yellow branches) across Cupressophyta. The major clades are labeled and estimates of the most likely state and rate are based on the model-averaged marginal reconstructions inferred across a large set of models (see main text). The color gradient on a given branch ranges from the slowest turnover rates (blue) to the highest rates (red). (B) The distribution of turnover rates estimated for contemporary species currently inhabiting each of the geographic areas indicates that both Northern Hemisphere and widespread species generally experience higher turnover rates (i.e., more speciation and more extinction) relative to Southern Hemisphere species. Araucaria., Araucariaceae; Callit., Callitroids; Cupress., Cupressaceae; Podocarp., Podocarpaceae; Sciado., Sciadopityaceae.**

model. These models can require large numbers of taxa to be effective. Thus, investigating the evolution of compelling, but small clades such as the great apes, baleen whales, or the Hawaiian silverswords can be attempted with these approaches, but results are likely to be met with fairly uncertain parameter estimates. Likewise, the required number of taxa to estimate the models will increase with the number of hidden states or geographical ranges included in the model.

Conclusions

Here we show that hidden Markov models provide a general framework that has the potential to greatly improve the adequacy of any class of SSE models. The inclusion of hidden states is a means for testing hypotheses about unobserved factors influencing diversification in addition to the observed traits of interest. We emphasize, however, that they should not be treated as a separate class of SSE models, but instead viewed as complementary and should be included as part of a set of models under evaluation. They also represent a straightforward approach to incorporating different types of unobserved heterogeneity in phylogenetic trees than a simple single rate category model is able to explain. For example, our expanded suite of GeoSSE models allows accommodating heterogeneity in the diversification process as it relates to geographic areas. The GeoHiSSE models introduced here can be applied to study rates of dispersion and cladogenesis as well as to perform ancestral area reconstructions, thus being a suitable alternative to avoid the shortcomings of DEC and DEC+J models (Ree and Sanmartín 2018). Moreover, the hidden state area-independent diversification models appear adequate to explain shifts in diversification regimes unrelated to geographic ranges, and demonstrate that the area-dependent models can successfully estimate the impact of geographic areas on diversification dynamics when such a signal is present in the data.

Many phylogenetic comparative methods have been under detailed scrutiny recently (e.g., Maddison and FitzJohn 2015; Rabosky and Goldberg 2015; Cooper et al. 2016a,b; Adams and Collyer 2017; Ree and Sanmartín 2018). This is certainly a worthy endeavor given that all models will fail under certain conditions, and some models have innate flaws that render them unwise to use. One response (e.g., Rabosky and Huang 2016; Rabosky and Goldberg 2017; Harvey and Rabosky 2018) has been to move to "semi-" or "nonparametric" approaches, some of which do incorporate models internally, but with an end result that is a nonparametric test. The issue with this is that they move away from

estimating parameters and the entire exercise becomes rejecting null models that we have never considered as realistic alternatives. Such methods, in our view, provide very limited insights as they only show that the null model is wrong, with appropriate type I error, which is not the same as showing the alternate is an adequate model. A more fruitful approach may be improving upon the existing models and better communicating when methods can and cannot be applied (Cooper et al. 2016a). We should stop trying to prove that our data cannot be explained by naive, biologically blind null models and, instead, fit appropriate models such that we can learn about meaningful patterns and processes across the tree of life.

AUTHOR CONTRIBUTIONS

JMB and BCO conceived the hidden Markov approach applied to SSE models. DSC, BCO, and JMB designed simulations, derived models, implemented model averaging and R code. DSC conducted simulations. JMB conducted empirical analyses. DSC, BCO, and JMB wrote the manuscript.

ACKNOWLEDGMENTS

This work has benefited tremendously from constructive and valuable comments from C. Ané, D. Silvestro, and an anonymous reviewer. We thank members of the Beaulieu, O'Meara, and Alverson laboratories for their comments and for general discussions of the ideas presented here; we especially thank T. Nakov and J. Boyko. We would also like to specifically thank A. Alverson, E. Goldberg, and S. Smith for their insightful critiques and helpful edits on an earlier version of this manuscript. DSC would also like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES: 1093/12-6) for the opportunity to work on this project.

DATA ARCHIVING

The code used to conduct simulations and empirical analyses as well as the simulated datasets and empirical data are available on Figshare (doi: https://doi.org/10.6084/m9.figshare.c.4069580.v2). The software used to fit the models is available as an R package on github (https://github.com/thej022214/hisse).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Appendix

EFFECTIVE STATE PROPORTIONS UNDER GeoHiSSE

We can use parameters estimated under a given model to determine the expected frequency under equilibrium of each geographic area and hidden state combination across a long stretch of evolutionary time. These equilibrium frequencies are often used in a variety of ways, most notably as weights in the likelihood calculation at the root (see Goldberg et al. 2011). We rely on them to complement the examination of rate differences among the observed ranges. In the main text, we describe a situation in which the diversification rate of observed state *0* is more or less the same as state *1*, but given the interaction with the hidden characters in the model we may find that 75% of species are expected to be in area *0* over some specified length of evolutionary time.

We follow Maddison et al. (2007) and track the number of lineages in area *0*, n_0, the number of lineages in area *1*, n_1, and the number of lineages in the widespread area *01*, n_{01}, over some length of evolutionary time, t. In our case, t represents the total height of a given empirical tree, although any arbitrary unit of time can be used. The index, i, denotes the possible hidden states ($A, B, \ldots, i$) that each observed state is associated. Given all the possible events that could occur across any given interval of time, we obtain the expected number of species for each area through the following ordinary differential equations:

$$\begin{aligned}\frac{dn_{0i}}{dt} &= s_{0i}\, n_{0i} + s_{01i} n_{01i} + s_{0i} n_{01i} + d_{01i\to 0i} n_{01i} \\ &\quad + \sum_{j\neq i} d_{0j\to 0i} n_{0j} - x_{0i} n_{0i} - d_{0i\to 01i} n_{0i} \\ &\quad - \sum_{j\neq i} d_{0i\to 0j} n_{0i}, \end{aligned} \tag{1a}$$

$$\begin{aligned}\frac{dn_{1i}}{dt} &= s_{1i}\, n_{1i} + s_{01i} n_{01i} + s_{1i} n_{01i} + d_{01i\to 1i} n_{01i} \\ &\quad + \sum_{j\neq i} d_{1j\to 1i} n_{1j} - x_{1i} n_{1i} - d_{1i\to 01i} n_{1i} \\ &\quad - \sum_{j\neq i} d_{1i\to 1j} n_{1i}, \end{aligned} \tag{1b}$$

$$\begin{aligned}\frac{dn_{01i}}{dt} &= d_{0i\to 01i}\, n_{0i} + d_{1i\to 01i} n_{1i} + \sum_{j\neq i} d_{01j\to 01i} n_{01j} \\ &\quad - s_{01i} n_{01i} - d_{01i\to 0i} n_{01i} - d_{01i\to 1i} n_{01i} \\ &\quad - \sum_{j\neq i} d_{01i\to 01j} n_{01i}. \end{aligned} \tag{1c}$$

Note that we use $d_{01i\to 0i}$ and $d_{01i\to 1i}$ to denote instances in which range contraction is separate from lineage extinction, x_{0i} and x_1. However, if the model assumes range contraction is governed by the same parameter value as lineage extinction, then we simply set $d_{01i\to 0i} = x_1$ and $d_{01i\to 1i} = x_0$. We also note that when there are no hidden states, these equations reduce exactly to the equilibrium frequencies under the original GeoSSE formulation. The initial conditions are set according to the state at the root. In our case, as a means of accounting for the uncertainty in the starting state, we rely on the likelihood that each area gave rise to the data (FitzJohn et al. 2009). Once the expected number of lineages is determined after a specified T by solving the ordinary differential equation (ODE) numerically, we then sum the frequencies for each observed area across each hidden state and normalize them so that the observed area frequencies sum to 1.

The above equations assume explicitly that the birth-death process directly impacts range evolution through cladogenetic

events, which are not allowed if the underlying model is a MuSSE-type model. Thus, in the MuSSE-type case, we rely on the following ordinary differential equations:

$$\frac{dn_{0i}}{dt} = s_{0i}\, n_{0i} + d_{01i\to 0i} n_{01i} + \sum_{j\neq i} d_{0j\to 0i} n_{0j} - x_{0i} n_{0i} - d_{0i\to 01i} n_{0i} - \sum_{j\neq i} d_{0i\to 0j} n_{0i}, \quad (2a)$$

$$\frac{dn_{1i}}{dt} = s_{1i}\, n_{1i} + d_{01i\to 1i} n_{01i} + \sum_{j\neq i} d_{1j\to 1i} n_{1j} - x_{1i} n_{1i} - d_{1i\to 01i} n_{1i} - \sum_{j\neq i} d_{1i\to 1j} n_{1i}, \quad (2b)$$

$$\frac{dn_{01i}}{dt} = s_{01i}\, n_{01i} + d_{0i\to 01i} n_{0i} + d_{1i\to 01i} n_{1i} + \sum_{j\neq i} d_{01j\to 01i} n_{01j} - x_{01i} n_{01i} - d_{01i\to 0i} n_{01i} - d_{01i\to 1i} n_{01i} - \sum_{j\neq i} d_{01i\to 01j} n_{01i}. \quad (2c)$$

LITERATURE CITED

Adams, D. C., and M. L. Collyer. 2017. Multivariate phylogenetic comparative methods: evaluations, comparisons, and recommendations. Syst. Biol. 67:14–31.

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. Proc. Natl. Acad. Sci. USA 106:13410–13414.

Alves, D. M. C. C., J. A. F Diniz-Filho, and F. Villalobos. 2017. Geographical diversification and the effect of model and data inadequacies: the bat diversity gradient as a case study. Biol. J. Linn. Soc. 121:894–906.

Beaulieu, J. M., and B. C. O'Meara. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst. Biol. 65:583–601.

Berkson, J. 1938. Some difficulties of interpretation encountered in the application of the chi-square test. J. Am. Stat. Assoc. 33:526–536.

Bloom, D. D., Weir J. T., Piller K. R., and N. R. Lovejoy. 2014. Do freshwater fishes diversify faster than marine fishes? A test using state-dependent diversification analyses and molecular phylogenetics of New World silversides (Atherinopsidae). Evolution 67:2040–2057.

Burleigh, J. G., and S. Mathews. 2007. Assessing systematic error in the inference of seed plant phylogeny. Int. J. Plant Sci. 168:125–135.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Springer, New York.

Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. Am. Nat. 164:683–695.

Chaw, S. M., A. Zharkikh, H. M. Sung, T. C. Lau, and W. H. Li. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. Mol. Biol. Evol. 14:56–68.

Cooper, N., G. H. Thomas, and R. G. FitzJohn. 2016a. Shedding light on the 'dark side' of phylogenetic comparative methods. Methods Ecol. Evol. 7:693–699.

Cooper, N., G. H. Thomas, C. Venditti, A. Meade, and R. P. Freckleton. 2016b. A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. Biol. J. Linn. Soc. 118:64–77.

Eastman, J. M., M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. 2011. A novel comparative method for identifying shifts in the rate of character evolution on trees. Evolution 65:3578–3589.

Eldrett, J. S., D. R. Greenwood, I. C. Harding, and M. Huber. 2009. Increased seasonality through the Eocene to Oligocene transition in northern high latitudes. Nature 459:969–973.

FitzJohn, R. G. 2010. Quantitative traits and diversification. Syst. Biol. 59:619–633.

———. 2012. Diversitree: comparative phylogenetic analyses of diversification in R. Methods Ecol. Evol. 3:1084–1092.

FitzJohn, R. G., W. P. Maddison, and S. P. Otto. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Syst. Biol. 58:595–611.

Goldberg, E. E., and B. Igić. 2012. Tempo and mode in plant breeding system evolution. Evolution 66:3701–3709.

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

Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732.

Harvey, M. G., and D. L. Rabosky. 2018. Continuous traits and speciation rates: alternatives to state-dependent diversification models. Methods Ecol. Evol. 9. https://doi.org/10.1111/2041-210X.12949

Huang, D., E. E. Goldberg, L. M. Chou, and K. Roy. 2018. The origin and evolution of coral species richness in a marine biodiversity hotspot. Evolution 72:288–302.

Kirk, R. E. 1996. Practical significance: a concept whose time has come. Educ. Psychol. Meas. 56:746–759.

Leslie, A. B., J. M. Beaulieu, H. S. Rai, P. R. Crane, M. J. Donoghue, and S. Mathews. 2012. Hemisphere-scale differences in conifer evolutionary dynamics. Proc. Natl. Acad. Sci. USA 109:16217–16221.

Leslie, A. B., J. M. Beaulieu, G. Holman, C. S. Campbell, W. Mei, L. R. Raubeson, and S. Mathews. 2018. An overview of extant conifer phylogeny from the perspective of the fossil record. Am. J. Bot. 105:1531–1544.

Maddison, W. P., and R. G. FitzJohn. 2015. The unsolved challenge to phylogenetic correlation tests for categorical characters. Syst. Biol. 64:127–136.

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

Magnuson-Ford, K., and S. P. Otto. 2012. Linking the investigations of character evolution and species diversification. Am. Nat. 180:225–245.

Mathews, S. 2006. Phytochrome-mediated development in land plants: red light sensing evolves to meet the challenges of changing light environments. Mol. Ecol. 15:3483–3503.

———. 2009. Phylogenetic relationships among seed plants: persistent questions and the limits of molecular data. Am. J. Bot. 96:228–236.

Matzke, N. J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63:951–970.

McLoughlin, S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. Aust. J. Bot. 49:271–300.

Morlon, H., T. L. Parsons, and J. B. Plotkin. 2011. Reconciling molecular phylogenies with the fossil record. Proc. Natl. Acad Sci. USA 8:16327–16332.

Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. B 344:305–311.

O'Meara, B. C., and J. M. Beaulieu. 2016. Past, future, and present of state-dependent models of diversification. Am. J. Bot. 103:1–4.

Pagel, M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proc. R. Soc. B 255:37–45.

Posada, D. 2008. jModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25:1253–1256.

Rabosky, D. L. 2006. LASER: a maximum likelihood toolkit for detecting temporal shifts in diversification rates from molecular phylogenies. Evol. Bioinform. 2:247–250.

———. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One 9:e89543.

Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst. Biol. 64:340–355.

———. 2017. FiSSE: a simple nonparametric test for the effects of a binary character on lineage diversification rates. Evolution 71:1432–1442.

Rabosky, D. L., and H. Huang. 2016. A robust semi-parametric test for detecting trait-dependent diversification. Syst. Biol. 65:181–193.

Rabosky, D. L., and I. J. Lovette. 2008. Density dependent diversification in North American wood-warblers. Proc. R. Soc. B 275:2363–2371.

Ran, J.-H., T.-T. Shen, M.-M. Wang, and X.-Q. Wang. 2018. Phylogenomics resolves the deep phylogeny of seed plants and indicates partial convergent or homoplastic evolution between Gnetales and angiosperms. Proc. R. Soc. B 285. https://doi.org/10.1098/rspb.2018.1012

Ree, R. H., and I. Sanmartín. 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. J. Biogeogr. https://doi.org/10.1111/jbi.13173

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

Rolland, J., F. L. Condamine, F. Jiguet, and H. Morlon. 2014. Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. PLoS Biol. 12:e1001775.

Schluter, D., T. Price, A. Ø. Mooers, and D. Ludwig. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711.

Silvestro, D., G. Zizka, and K. Schulte. 2014. Disentangling the effects of key innovations on the diversification of Bromelioideae (Bromeliaceae). Evolution 68:163–175.

Stadler, T. 2011. Mammalian phylogeny reveals recent diversification rate shifts. Proc. Natl. Acad. Sci. USA 108:6187–6192.

Uyeda, J. C., R. Zenil-Ferguson, and M. W. Pennell. 2018. Rethinking phylogenetic comparative methods. Syst. Biol. https://doi.org/10.1093/sysbio/syy031

Wilford, G. E., and P. J. Brown. 1994. Cretaceous to Recent. Pp. 5–13 *in* R. S. Hill, ed. History of the Australian vegetation. Cambridge Univ. Press, Cambridge, U.K.

Yang, Z., and R. Nielsen. 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Mol. Biol. Evol. 25:568–579.

Yang, Z., S. Kumar, and M. Nei. 1995. A new method of inference of ancestral nucleotide and amino acid sequences. Genetics 141:1641–1650.

Zanazzi, A., M. J. Kohn, B. J. McFadden, and D. O. Terry, Jr. 2007. Large temperature drop across the Eocene-Oligocene transition in central North American. Nature 445:639–642.

Zhong, B., O. Deusch, V. V. Goremykin, D. Penny, P. J. Biggs, R. A. Atherton, S. V. Nikiforova, and P. J. Lockhart. 2010. Systematic error in seed plant phylogenomics. Genome Biol. Evol. 3:1340–1346.

Associate Editor: C. Ané
Handling Editor: Mohamed A. F. Noor

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1: Description of scenarios and parameter values used to simulate the data.
Table S2: Additional 17 models used in empirical study of conifers.
Table S3: Description of scenarios and parameter values used to simulate phylogenetic trees and range distributions under the GeoSSE+extirpation model.
Figure S1: Scheme of the transition rates between rate classes (RC0 to RC4) used for the simulation scenarios with multiple rate classes (Sims B, C, and D).
Figure S2: Proportion of widespread lineages on trees simulated under scenarios E and F.
Figure S3: Summary of model support for simulated scenarios A to D.
Figure S4: Summary of model support for simulated scenarios E to H.
Figure S5: Accuracy of turnover and extinction fraction estimates for simulations scenarios A to D.
Figure S6: Accuracy of net diversification estimates for simulations scenarios A to D.
Figure S7: Results for relative net diversification rates and Akaike model weights (AICw) for simulation scenarios B and C.
Figure S8: Distribution of Akaike weights for the model set fitted to simulation scenarios ext_A to ext_D.
Figure S9: Distribution of parameter values across 100 simulation replicates for each of the scenarios ext_A to ext_D.