Potential survival of some, but not all, diversification methods

Running title (40 characters max): Diversification method survival

Authors: Brian C. O'Meara[1*] and Jeremy M. Beaulieu[2]

Affiliations:
[1]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee, 37996-1610 USA
[2]Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas, 72701 USA

[*]**Corresponding author:** Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee, 37996-1610 USA; email: bomeara@utk.edu

Author contributions: BCO and JMB take equal responsibility for the contents of this article.

Acknowledgements: We thank members of the Beaulieu and O'Meara labs for their comments and discussions of the ideas presented here. We would also like to thank Andrew Alverson, Jim Fordyce, and Ben Fitzpatrick for their insightful comments. This work was funded by the National Science Foundation grants DEB–1916558 and DEB- 1916539**.**

Data Availability Statement: Code supporting this article is made freely available at http://flippedcoin.info/, https://github.com/bomeara/diversificationlives, and https://github.com/bomeara/CondamineEtAlExample.

Conflict of Interest: The authors have declared no conflict of interest.

Abstract

Models have long been used for understanding changing diversification patterns over time. The rediscovery that models with very different rates through time can fit a phylogeny equally well has led to great concern about the use of these models. We share and add to these concerns: even with time heterogeneous models without these issues, the distribution of the data means that estimates will be very uncertain. However, we argue that congruence issues such as this also occur in models as basic as Brownian motion and coin flipping. Taxon-heterogeneous models such as many SSE models appear not to have this particular issue.

Keywords: diversification, congruence, likelihood, identifiability

Introduction

For decades, molecular phylogenies have served as vital sources of historical information for deciphering the birth and the death dynamics of lineages. Thousands of studies of molecular phylogenies have been dedicated to investigating diversification. In theory, estimating constant birth and death rates separately is possible because each has distinguishable effects on the tree shape and branch length distributions (Nee et al. 1994). There are several extensions that expand this simple model for characterizing diversification as a function of time or diversity (e.g., Nee et al. 1992; Rabosky 2006, 2009; Bokma 2008; Rabosky and Lovette 2008; Morlon et al. 2011; Etienne et al. 2012), which are used to reconstruct lines showing speciation and extinction rates scrolling into the past, like the pen of a seismometer tracking vibrations through time. A sudden sweep up of the extinction rate arm could mean a mass extinction. A slow, downward trajectory of the speciation rate arm as time approaches the present could mean available niches have become filled up, limiting the possibilities of adding new species. And, as with constant rate birth-death models, we have been working under the assumption that even tiny changes in speciation and/or extinction through time should leave distinct signatures on the tree shape and branching structure in a molecular phylogeny.

In a recent paper by Louca and Pennell (2020), the entire enterprise of estimating diversification rates, at least from molecular phylogenies alone, has been called into question. As it turns out, for any given phylogeny there are an infinite array of congruent models each having unique functions of speciation and/or extinction rates smoothly varying through time. This is based on the property of both constant rate birth-death and time-varying models in which every lineage at any given time-point experiences the same rates, and so sampling times for either a speciation or extinction event are drawn from the same distribution (also known as a coalescent point process or CPP; see Lambert and Stadler, 2013). Under such conditions, the likelihood of a tree under a given birth-death model can be inferred simply in terms of the lineage-through-time (LTT) curve, which is a retrospective counting of the number of lineages that led to a set of species observed today, and there are always multiple qualitatively different models that can produce the same curves with the same probability. For example, one model may infer the observed diversity of Cetaceans (i.e., whales, dolphins, and relatives) is a product of dramatic changes in the rate of speciation and extinction rates over time, whereas another, *equally likely* model, may infer modern whale diversity is the product of no extinction and ever so slight changes to the speciation rate. In other words, two diametrically opposed models, particularly with regards to the role of extinction, provide *equally* valid explanations for the mode and tempo of Cetacean diversification. In some cases, such as our example above, these models will have the same number of parameters, rendering them truly indistinguishable.

It should come as no surprise, then, that one popular interpretation of these findings is that any attempt to learn anything about diversification rates from molecular phylogenies is a completely futile enterprise. A different response, which we also have seen, is the continued and uncritical use of these suspect methods sanitized with a “but see Louca and Pennell (2020)” citation. It is also worth noting that the findings of Louca and Pennell (2020) are substantially similar, though much more detailed, to the ones presented by Kubo and Iwasa (1995) a quarter century ago. These authors also described an infinite array of birth and death models fitting the data equally well, which has been effectively ignored by most later workers.

The issues raised by Louca and Pennell (2020) and Kubo and Iwasa (1995) do represent substantial methodological problems for comparative biology. However, this does not signal the end of studying diversification rates on molecular phylogenies, as some have claimed, as these problems do not extend to *all* models of diversification. Instead, they are limited to situations where the goal is to interpret diversification rates through time using what we refer to as, “time-varying, lineage homogeneous” models — again, models in which all lineages experience the same variable rates at any given point in time. These would be analogous to a non-heritable trait-dependent process (Lambert and Stadler, 2013), where changes in a trait occur the same in all species independently (e.g., global CO_2, sea-level changes, global temperature patterns). We argue that what we refer to as “lineage-specific heterogeneous” models, in which rates vary among lineages across time points, perhaps due to the inheritance of a trait (e.g., state-speciation and extinction, or SSE models; Maddison et al. 2007), should be immune to the issues of identifiability raised above. This comes with the substantial caveat that this is true if, and probably only if, the heritable rate changes are modeled as containing a single speciation and extinction rate that do not vary through time. Essentially, we will show that these models do “work” if we limit the model space to those with single rates at any time point.

We also address some of the other procedures proposed, explicitly or implicitly, by Louca and Pennell (2020): continuing with pulled diversification rate reconstruction, focusing on a point estimate only, no longer penalizing for model complexity, and how information is distributed on trees.

Overall, we make four points:

1. Model congruence can occur in areas as different as coin flipping and Brownian motion: it does not mean these models must be given up, only that certain questions are infeasible.

2. Time-varying, lineage homogeneous models that use just the information from a lineage through time curve to estimate changing speciation, extinction, diversification, turnover, or extinction fraction should be avoided due to congruence issues.
3. Pulled speciation and pulled diversification rate analyses (Louca and Pennell 2020) are identifiable, but they fail to incorporate the substantial uncertainty in reconstructions that come as a result of typically exponentially decreasing number of data points (lineages) as one approaches the root of a tree (this also plagues the methods in point 2)
4. Some SSE methods, and likely other methods that investigate heterogeneity across taxa, use information beyond that in a lineage through time curve and their utility remains intact in the face of Louca and Pennell (2020) and Kubo and Iwasa (1995).

Model congruence is common

It may come as a surprise that this issue of two models fitting data equally well is not new to comparative methods. Take, for instance, the inference of evolutionary trends, which, broadly defined, are identifiable patterns of trait evolution in a given direction through time. Using only extant species, can we detect horses getting bigger and with fewer digits, or increases in the mean seed size in flowering plants since the Cretaceous (e.g., Tifney 1984; Eriksson et al. 2000), or, more generally, uncover an evolutionary arms race between predator and prey (e.g., Dawkins and Krebs 1979; Abrams 1986)? It is trivial to extend a simple Brownian motion model to include a parameter that allows for the focal trait to evolve along a trend, and this is available in popular software like the R package *geiger* (Pennell et al. 2014). The likelihood for these models given the data is finite, and the simple no trend model is even nested within the trend model, so comparisons between the two are straightforward. However, as Felsenstein (1988) and Hansen and Martins (1996) have pointed out, even though trait values move in a given direction under a Brownian motion with a trend model, this does not affect the expected covariances among species trait values. That is, the expected trait differences among species is still linearly dependent on time, meaning closely related species are still expected to be more phenotypically similar than more distantly related species, which is an identical assumption under a standard Brownian motion model. Consequently, the two models have identical likelihoods when fitted to extant species only, making them indistinguishable based on their probability alone.

One might argue that in cases of clear non-identifiability any careful scientist would avoid fitting a degenerate model such as Brownian motion with a trend with just coeval terminals. However, the problem of identifiability between Brownian motion models with and without a trend is further compounded when considering the potential for dramatic effects on ancestral state reconstructions. For example, the ancestor of a clade of taxa with body sizes ranging from 10-12 kg might have a reconstructed state near 11 kg under a no trend model but could have a reconstructed state of 50 kg under a model with a trend of an incremental trait decrease through time. Such ancestral state reconstruction remains widely popular. Nevertheless, it is still a rather large leap to assert that, because these models are unidentifiable, models using Brownian motion are generally invalid for use on trees containing only modern taxa. We can still compare Brownian motion models with more complex models, such as Ornstein-Uhlenbeck models (e.g., Butler and King 2004; Beaulieu et al. 2012), Brownian models with more than one rate (e.g., O'Meara et al. 2006; Thomas et al. 2006), or models where the Brownian motion rate itself changes over time (e.g., Revell, 2021). In other words, while Brownian motion with a trend model is unidentifiable with modern taxa only, we would not, for instance, say that any model that attempts to estimate rates of evolution on such trees is uninterpretable. Some models in this space give the same likelihoods and cannot be distinguished, but many others can, which calls for care and analysis, not panic.

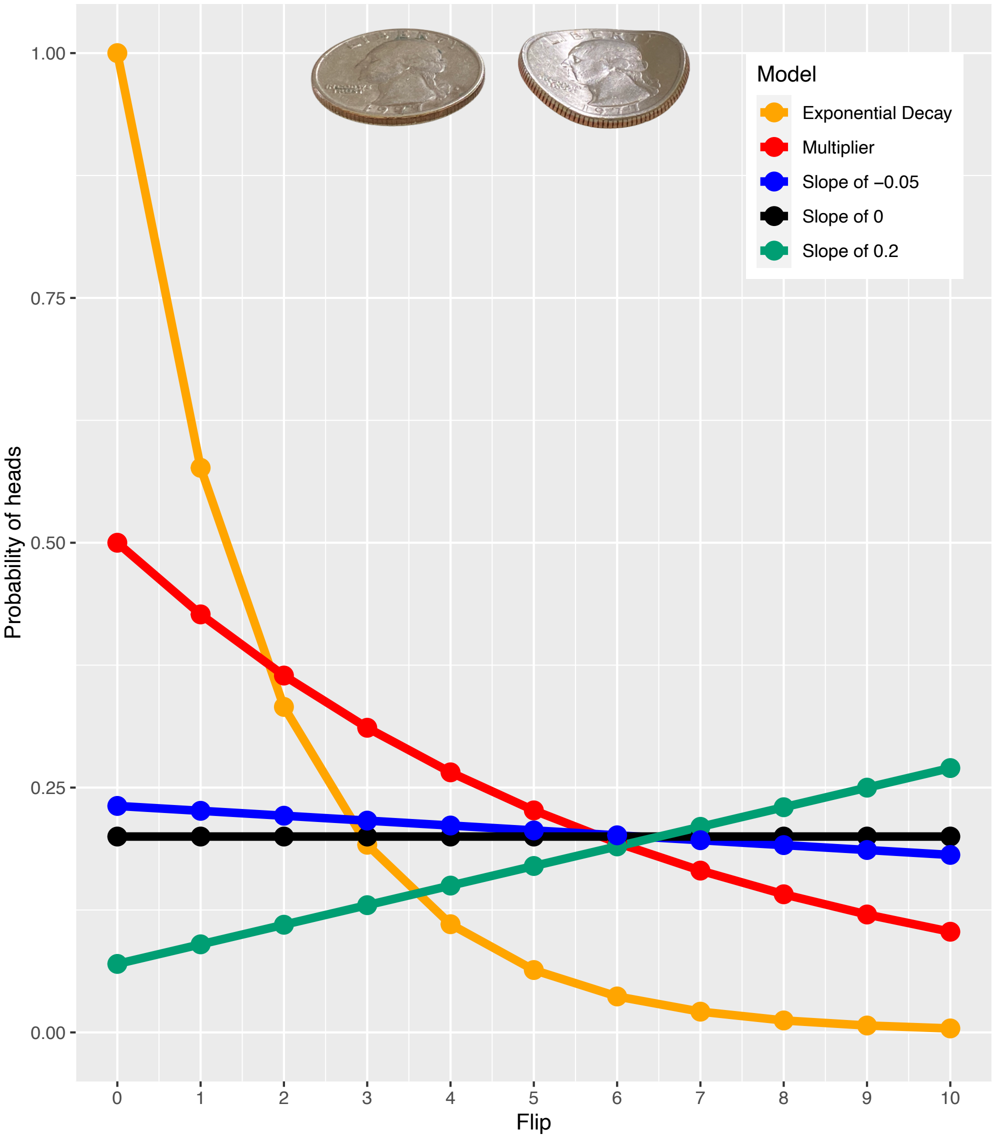

Figure 1: Probability of heads per flip on different models of coin flipping. Each of these models can fit the same dataset of two heads, eight tails with equal likelihood but make very different predictions about the next flip.

We also point out that model congruence occurs in other statistically based disciplines. Consider the classic coin-flipping example. Suppose we toss a coin 10 times, and 2 of those tosses come up heads. The most straightforward fitted binomial model indicates that the probability of observing 2 heads in 10 flips is 0.3 for a biased coin with each flip having a 20% chance of landing on heads. Now suppose that every time we touch the coin, it gets slightly dented, or a bit of metal is worn away, and it becomes less and less likely to land on one side than the other. We can devise several models that have different slopes to alter the probability of heads after a set of coin flips (Figure 1). For instance, the probability of heads can linearly increase with each flip, such that by the end the probability of heads is 10% higher than when we started flipping, and a model where the probability of heads decreases with each flip so that by the end it is 5% lower than when it started (Figure 1). Interestingly, the probability of observing 2 heads in 10 flips of the coin in each of these models is the *same* as the simple binomial model, though the linear change models infer different initial probability of heads before any flips are made as well as what the probability of the next flip being heads is. If we pre-set the 5% lower or 20% higher parameters ahead of time rather than fitting them, these have the same number of free parameters as the homogeneous binomial model.

Avoid inference of congruent diversification models

While millions of students struggling with their statistics homework might cheer the destruction of the concept of estimating the probability of heads from a set of coin flips, it is important to emphasize that even though these models are functionally congruent, each provides different predictions after a new set of coin flips are made (e.g., what is the likeliest outcome of the eleventh flip?). That is, even though they are indistinguishable from a probabilistic point of view, we can still distinguish them when new data becomes available. Of course, with comparative methods we cannot simply “flip” evolution more times to distinguish among a set of congruent models. The emphasis, then, as Morlon et al. (2020) recently pointed out, becomes what we are trying to learn about the world, given what we know about how it works. It is generally true that with coins, we have a good idea that the probability of heads does not change meaningfully over flips, so we may be willing to assume a standard binomial model and then question the fairness of a coin, perhaps as a way of extrapolating to other coins (i.e., if this Euro coin has a probability of heads of 0.502, is that true for other Euro coins?). In other words, the parameter can be of interest because the model is not really in question.

With many diversification models, the central question is about which model fits best, which is at odds with a general lack of knowledge about any system to clearly know which kind of model is appropriate ahead of time. Even with diversification models that explicitly link rates to abiotic variables such as temperature or sea level changes (e.g., Condamine et al. 2013; 2019) the goal seems more focused on which model fits best. In our view, we are not yet at the stage where we can confidently rule out a congruent model where extinction rates are driven by the position of a hypothetical dwarf star outside our solar system, which triggers periods of increased comet activity on Earth (e.g., Raup and Sepkowski 1984), over a more “sensible” model of, say, temperature clearly affecting speciation but not extinction rates. In such cases, asking questions about which of several indistinguishable models fit does not seem to us a good use of our time.

It is also important to emphasize that our argument here is not that the issues Louca and Pennell (2020) point out are trivial. In fact, there are many papers, and even entire research programs, dedicated to the development of time-varying, lineage homogeneous models of diversification, and trying to draw conclusions based on which models fit best. But, as with coin flipping or Brownian motion, knowing what conclusions can be made given the models and data and limiting our work to those areas can be important. Moreover, if even coin flipping has congruent models, there is no guarantee that even models that currently seem to avoid the congruence issue, such as pulled diversification rates recommended by Louca and Pennell (2020), do not have other congruent models with different parameters, such as models that change rates by taxa rather than solely by time.

Avoid ancestral rate reconstruction

Ancestral state reconstruction of characters remains one of the most popular and widely used approaches in phylogenetic comparative methods, despite the occasional discussion to dampen enthusiasm in them (e.g., Cunningham et al. 1998; Omland 1999; Oakley and Cunningham, 2000). Ancestral state reconstruction is useful for formulating testable hypotheses, such as the synthesis and performance evaluation of putative ancestral proteins (e.g., Thornton et al. 2003; Pillai et al. 2020), biogeographic history and movements of clades through time (e.g., Ree and Smith, 2008; Landis et al. 2020), and the order and timing of character state changes (e.g., Schluter et al. 1997; Ackerly et al. 2006). Reconstructing diversification rates through time has a similar appeal, in that they too can point to testable hypotheses about the intrinsic and extrinsic factors that drive species diversity among groups. Armed with only a phylogeny of modern taxa, we can reconstruct the seismograph tracing of how speciation rate, extinction rate, net diversification rate, or the new pulled diversification or pulled speciation rates, have changed through time. With the reconstruction of discrete or continuous characters, state information at the extant tips is generally less and less informative about states at nodes as one traverses deeper in the tree towards the root. For diversification rate models, the data are not arrayed along the tips of a tree, but rather, come from the distribution of branching events across the phylogeny.

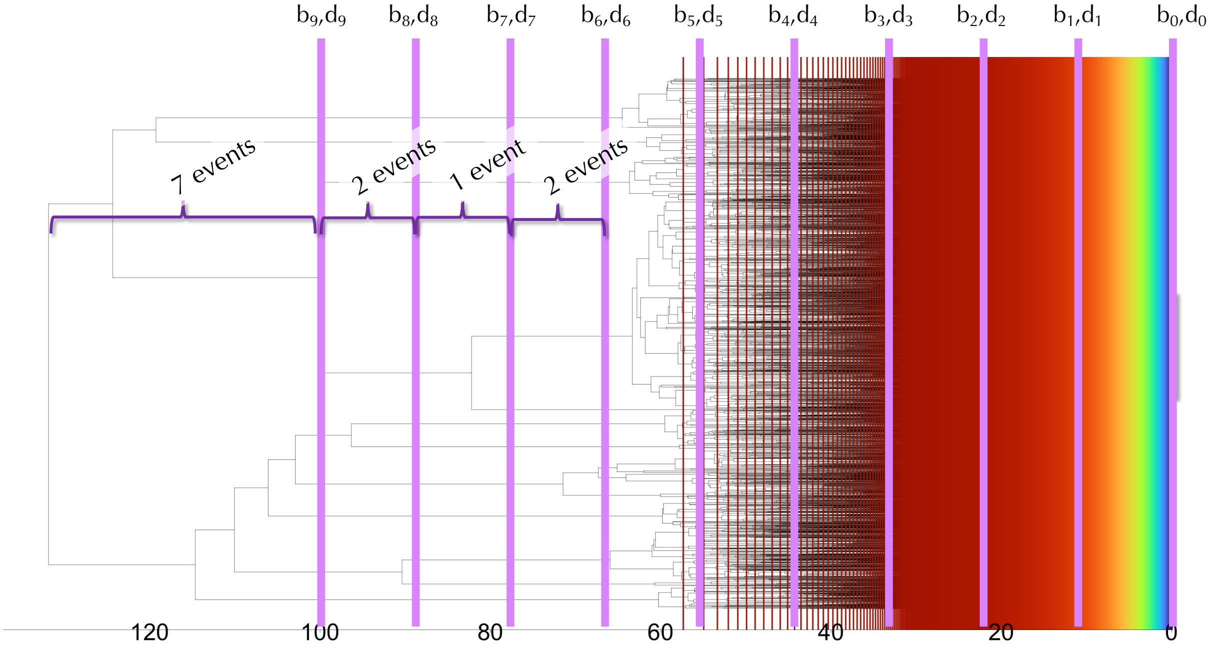

Figure 2. Million taxon tree from Louca and Pennell (2020). The purple lines separate the regimes used to estimate rates. The thin vertical lines in a rainbow distinguish regimes with 100 events within them representing equal-sized slices of data. Half the regimes are on each side of the green band, showing how much of the data are near the tips. The brackets show how many events occur in each regime. Ignoring uncertainty in branch lengths or topology, this makes a 10 Myr long edge equally informative regardless of whether it ended 3 million years ago or 300 million years ago.

As Maddison and FitzJohn (2015) noted, our field does not yet think in terms of the curvature of biodiversity-time, and so our expectations about the distribution of these branching events are often wrong. The number of edges on trees, under most models, increases approximately exponentially with time, although extinction complicates this, as would models with carrying capacity (e.g., Rabosky and Lovette 2008), age-dependent extinction (Alexander et al 2016), and other variations of the birth-death model. Nevertheless, lineage through time plots are commonly shown on a log scale for the number of lineages due to this nearly exponential growth. Visualizing the raw number of lineages would make the dynamics in the early parts of the plot virtually invisible due to the massive growth of the line near the present. Importantly, the midpoint of the data is the point at which half the number of lineages has accumulated, which is *not* the halfway point along the time axis.

Consider a tree split into equal-sized chunks according to some time interval, as Louca and Pennell (2020) and others have. The number of edges within a given bin naturally decreases as one moves towards the root. Now, take the extreme example from Louca and Pennell (2020) where they analyzed a tree with a million taxa (Figure 2). Even though the tree is far larger than any published study of diversification, they only estimate rates along 10-time intervals and for many of these bins there is only a trivial amount of data. For example, at the start of the 100 Myr to 90 Myr interval, there are just seven lineages, and by the end of that interval, there are only ten. The lineage through time plot, which is the data that goes into these methods, thus jumps just three times over that ten million years. This is clearly not a lot of data points for estimating speciation or extinction rates, or even a single pulled diversification rate. Each of the next several intervals have a *single* jump. That is, it goes from 10 to 11 lineages from 90 to 80 Myr, and from just 11 to 12 from 80 Myr to 90 Myr. It is no wonder that these methods perform poorly; a single event on a 12-taxon tree does not contain much information about rates, whether pulled or not. Put another way, these methods are starving for data across large portions of the tree.

A natural corollary, then, is that seismographic reconstructions of rates will contain increasing levels of uncertainty as one moves deeper in time. Nee et al. (1994) showed clearly that even rates from a constant birth-death model can carry substantial uncertainty. Yet most analyses doing the sort of work Louca and Pennell (2020) criticize, and even their examples, return a single point estimate for each parameter at a given time period. In a few cases, point estimates are summarized together across a set of trees, which is better, but still likely reflects substantially less uncertainty than what is truly present in any single estimate.

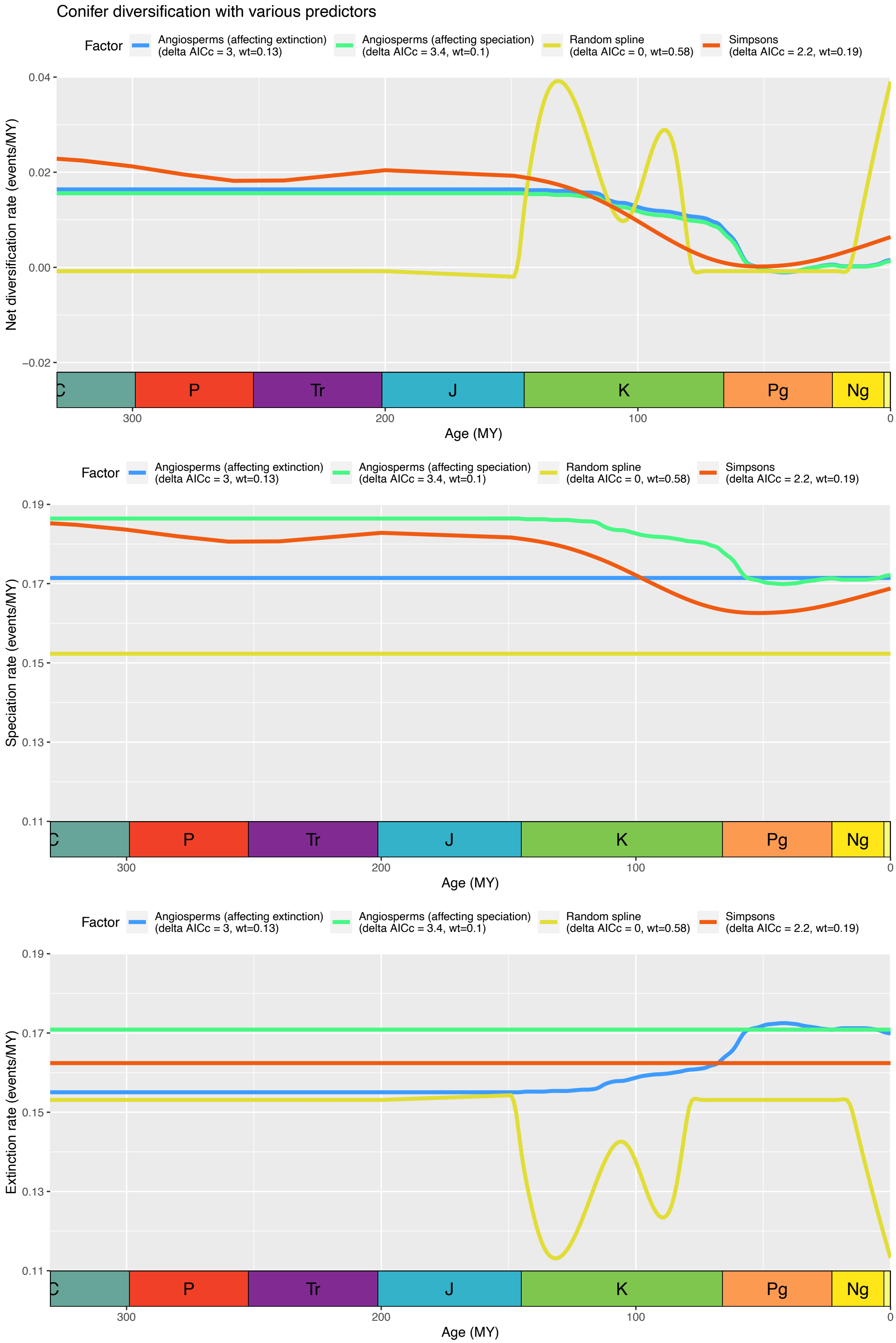
Conifer diversification with various predictors
Factor
Angiosperms (affecting extinction) (delta AICc = 3, wt=0.13)
Angiosperms (affecting speciation) (delta AICc = 3.4, wt=0.1)
Random spline (delta AICc = 0, wt=0.58)
Simpsons (delta AICc = 2.2, wt=0.19)
Net diversification rate (events/MY)
0.04
0.02
0.00
−0.02
C
P
Tr
J
K
Pg
Ng
300
200
100
0
Age (MY)
Factor
Angiosperms (affecting extinction) (delta AICc = 3, wt=0.13)
Angiosperms (affecting speciation) (delta AICc = 3.4, wt=0.1)
Random spline (delta AICc = 0, wt=0.58)
Simpsons (delta AICc = 2.2, wt=0.19)
Speciation rate (events/MY)
0.19
0.17
0.15
0.13
0.11
C
P
Tr
J
K
Pg
Ng
300
200
100
0
Age (MY)
Factor
Angiosperms (affecting extinction) (delta AICc = 3, wt=0.13)
Angiosperms (affecting speciation) (delta AICc = 3.4, wt=0.1)
Random spline (delta AICc = 0, wt=0.58)
Simpsons (delta AICc = 2.2, wt=0.19)
Extinction rate (events/MY)
0.19
0.17
0.15
0.13
0.11
C
P
Tr
J
K
Pg
Ng
300
200
100
0
Age (MY)

Figure 3: Comparison of net diversification, speciation, and extinction rate of conifers using as a predictor the best model from Condamine et al. (2020) in blue where only extinction rate varies with angiosperm diversity, a slightly worse model from that paper (green) where speciation rate varies with angiosperm diversity, a model (yellow) that fits the data best (at least in terms of likelihood — the number of free parameters of the spline is hard to compare), and using scaled IMDB ratings of the television program the Simpsons (red) as a predictor for speciation rate (which did a better job predicting conifer diversification than angiosperm diversity did). Not shown are numerous other attempts for other predictors using other splines, linear change models, and ratings of many other television programs. Some of these also outperformed angiosperms, but many did not.

Besides unexamined uncertainty in point estimates, there is substantial uncertainty in which model fits best, even if one ignores the congruence issue. For example, Condamine et al. (2020) compared various models correlating various rates with angiosperm diversity using just a phylogenetic tree; their best model showed an exponential dependence of conifer extinction rate with the number of angiosperms. However, models nearly as good (ΔAICc < 2; see their Table S5) include an effect on speciation or both speciation and extinction (only 41% of the model weight is on variable extinction only models; 39% is on variable speciation only, and 21% on both varying). One can construct other patterns of diversification rates with very different conclusions that are better predictors. For example, in Figure 3, the yellow diversification curves predict the conifer data even better but tell a very different story of constant speciation with decreases of extinction in the Cretaceous and Neogene rather than the recovered pattern of a gradual rise of extinction in the Cretaceous onward. Even using ratings of a television show (the Simpsons, the red line) scaled for the appropriate time period predicts conifer diversification better than the postulated angiosperm mechanism. Similarly, Morlon et al. (2011) looking at a paraphyletic set of 16 cetaceans found a constant speciation but variable extinction model fit best, but there were two other models with a ΔAICc of less than 1 (including one where extinction does not vary) — this makes it hard to draw any firm conclusions from modern data alone. Careful biologists, as shown in the studies above, will limit themselves to only feasible mechanisms, but as we know from other diversification models (Rabosky and Goldberg 2015; Beaulieu and O'Meara, 2016), if presented with a very simple model and more complex alternatives only, methods using our messy, complex empirical data will leap to use the more complex predictors. That is, if the only way to incorporate the very real heterogeneity of a process is to ascribe it to some varying predictor, methods will choose that. Whether it is 16 modern taxa or a million, it is unclear what we learn from such exercises. Our energies might be better directed elsewhere.

The state of SSE models and other approaches

Louca and Pennell (2020) speculate that state-speciation and extinction models (SSE) may have similar identifiability issues. This is not an unreasonable concern. Beaulieu and O'Meara (2016) demonstrated that if a trait has no effect on speciation and/or extinction rates, the likelihood of any SSE model becomes the product of the likelihoods of the Nee et al. (1994) tree likelihood and the character model likelihood (or the sum of the log-likelihoods in log space), so the models are clearly related. One could certainly alter the SSE model to include realistic factors like mass extinctions and secular changes in rates through time, and any one of these features will undoubtedly lead to a set of models with identical likelihoods. However, in other ways, strict SSE models can be immune, because they do not split the tree into time bins. Instead, they approximately treat a tree as a series of discrete chunks — that is, a chunk in one part of the tree is in state 0, and so is impacted by the instantaneous speciation rate, λ_0, and extinction rate, μ_0, while another chunk in another part of the tree is in state 1 and so is impacted by speciation rate, λ_1, and extinction rate, μ_1 (in reality, they average over these paintings based

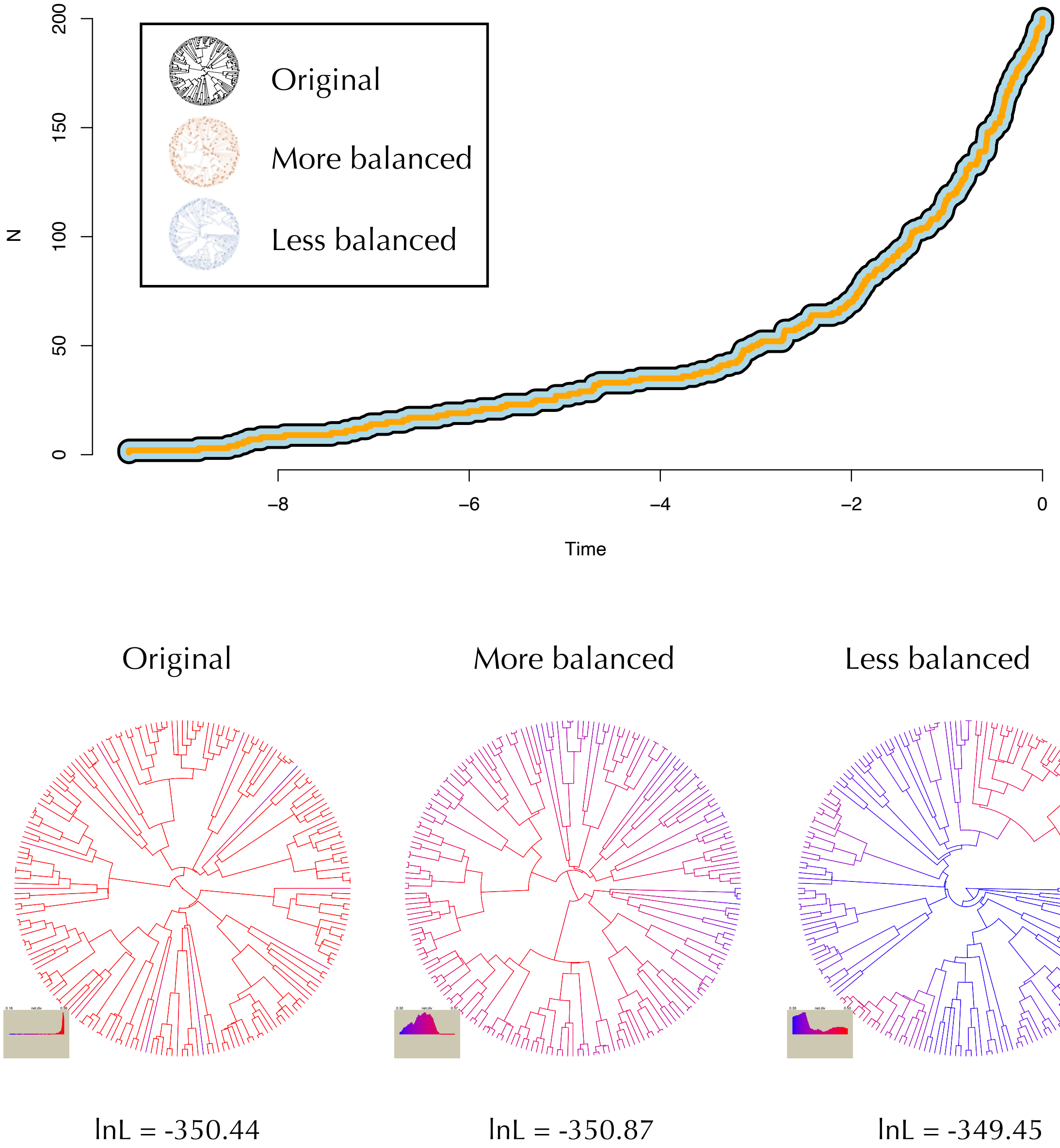

Figure 4: (A) Depicts the identical lineage through time (LTT) plots for three trees that differ in terms of tree balance. The procedure takes a simulated tree, then makes swaps across branches to either increase balance or decrease it but maintain the same lineage through time curve. (B) Depicts the log-likelihood score among the three trees under a two-rate MiSSE model. These trees produce identical log-likelihoods under taxon-homogeneous, time-heterogeneous models that use LTT data. However, this is not the case here because allowing rates to vary among clades, as our MiSSE models do, avoids the trap of having an infinite array of congruent models. Helmstetter et al. (2021) reach similarly positive conclusions about the possibility of learning about diversification from SSE models.

on their probabilities). Within each of these chunks the speciation and extinction rates are invariant, and as Nee et al. (1994) showed, constrained in this way there is a single maximum likelihood estimate of each rate. If one limits the model space to where rates are dependent on states (observed or hidden or some combination of both), then SSE models should be identifiable, though not immune to all the practical difficulties of estimating rates in the presence of extinction, finite data, errors in branch lengths and topology, and more.

We can at least empirically demonstrate that SSE models are immune to the issues of model congruence based on information in the lineage through time plot: SSE models use more information than this. In Figure 4, there are three trees with identical lineage through time curves, but different arrangements of topology. Under a constant rate Yule or birth-death model the likelihoods of these three trees are identical, as one would expect given the findings of Louca and Pennell (2020). However, if we allow for multiple rates to be inferred across the tree by fitting a hidden states only model (which we call MiSSE; see Vasconcelos et al. 2021) the three trees have different likelihood. This is because the MiSSE model uses information not accessible to LTT methods, namely, the tree topology. Other methods that fit rate heterogeneity across taxa, such as MSBD (Barido-Sottani et al. 2018) and ClaDS (Maliet et al. 2019), may also not be bound by the issues that make different LTT models congruent. Even an approach as simple as sister group comparisons (e.g., Slowinski and Guyer 1993) can detect differences in net diversification rate across pairs of clades in a way that depends on topology alone: identical lineage through time plots would have no effect on this. Taken together, this does not mean that clade-specific models of diversification could not have their own issues (even coin flipping models can have congruence, as shown above), just that the identifiability issue identified by Kubo and Iwasa (1995) and Louca and Pennell (2020) does not apply to them.

What are we really learning anyway?

Null hypothesis testing is intended to show whether an effect is significantly different from chance alone. At some point, though, comparing against chance becomes an uninteresting and dull exercise as the end point of a study. After several decades of studying diversification on molecular phylogenies and continually finding variation in rates across taxa and across time, favoring a complex model over a "dull" null hypothesis of simple constant birth-death is no longer surprising. No reasonable scientist will argue that diversification processes have remained perfectly constant through time, with no changes in extinction rates, no factors changing speciation rates, and more. We know the data comes from a heterogeneous, complex process and so any even somewhat reasonable more complex model will fit better than a simple model. As we have noted elsewhere (see Beaulieu and O'Meara 2016; Caetano et al. 2018), rejecting the "null" does not imply that the slightly more complex alternative is the true model. Like a hot gas moved from a simple bottle to a more complex bottle with greater volume, our complex data will happily expand to take the shape of the biggest container offered to it. Model rejection, model weighting, posterior probability of models are all ways of saying, "my cloud of data is more comfortable in this larger bottle than in this smaller bottle. Since the extra bulge on the larger bottle is called factor *X*, this clearly shows that factor *X* is important." However, a different bottle with the same volume but with a bulge for factor *Y* might fit as well. Good science will involve comparing different reasonable models to the data, not just comparing our slightly more complex model of interest with slightly simpler models. Much of our work on hidden rate models (e.g., Beaulieu et al. 2013; Beaulieu and O'Meara 2016; Caetano et al. 2018; Boyko and Beaulieu 2021) is motivated by this desire to give our preferred models an actual chance to lose against other models in the hope that we learn from this.

In our view, an important aspect of the work of Louca and Pennell (2020) was showing that even this limited, careful approach might not work for time-heterogeneous diversification rates: there are multiple diversification bottle shapes that fit the cloud of branching times from a tree equally well. Furthermore, approaches that seek to track the wiggles of the diversification seismograph through time tell us very little, if anything, about the past. However, we would add that instead of tracing the wiggles of a single pulled diversification rate pen on a diversification seismograph, or even take the extreme step of stopping analyses of diversification using modern phylogenies altogether, we should use the valid methods we do have to answer biological questions, in the same way we can use Brownian motion even though different parameterizations can give identical likelihoods. Focus on analyses that lead to discoveries or confirmations of biological processes that are possible given available data.

On the whole, it is important to recognize that *our methods are better suited for using the past to learn about the present survivors, not using the present survivors to learn about the past*. Phylogenies of extant taxa convey an enormous amount of information about species and their direct ancestors, but they also necessarily miss much of the history of a particular clade. Therefore, there will never be a clever analysis of a phylogeny of extant archosaurs (crocodilians and birds) that will result in an inference of the dynamics of the rise and fall of sauropod dinosaurs, even though they are firmly nested in that clade and must have had a huge effect on the lineages that survived while all were interacting. Yet this is exactly what we are asking of our diversification seismograph analyses of modern taxa — that is, we think we are understanding something about diversification dynamics of archosaurs in the Cretaceous from a study of their weird, few surviving lineages. However, phylogenies of extant taxa can give us information about what led to present diversity, what traits are associated with modern diversity patterns, and, perhaps, even when certain modern lineages took off. We can understand something about diversification patterns of extant birds, for example, including what traits are associated with faster diversification or turnover rates.

Perhaps the best example of procedures that illustrate where we think the field needs to reconsider are classic sister group comparisons (Mitter et al., 1988). These explicitly are about comparing modern clades and so are by their nature lineage-heterogeneous and limited to examining factors leading to modern diversity. They do not claim to allow inference about rate shifts in the past, since they attempt to control for the effect of time. There can be important corrections for even these methods (Käfer and Mousset, 2014) but they prevent scientists from spinning tales from limited information about the past. They should also be far more robust to the concerns raised by Maddison and FitzJohn (2015) than even hidden rate models. Of course, they are not without their own limitations: it can be hard to find enough comparisons; they only allow comparison of the direction of net diversification differences due to some pre-specified factor, while many of our hypotheses might relate to speciation rate, extinction rate, or, as we have advocated turnover rate (Beaulieu and O'Meara, 2016; Vasconcelos et al. 2021); they typically require only discrete characters (though see Harvey et al. 2020 and the bomeara/sisters package on github); and they require ancestral state reconstruction to find sister pairs differing by a character state. There are also questions completely inaccessible to these methods; however, accepting these limitations at the outset may have prevented years of work that relied on methods that felt scientific but gave ultimately meaningless results given the issues now understood about time-heterogeneous diversification models.

Conclusions

The reconstruction of diversification rates through time, whether of pulled or classic rates, is appealing but flawed in the same way that inference of ancestral states is appealing but also flawed. Multiple indistinguishable models give very different estimates about the past, and even for large trees, what matters is the branches and branching events at the times of interest, often when the mighty tree was a mere sapling. Moreover, this only looks at branches with modern descendants. What information it does provide is about what those lineages may have been doing, not what the clade as a whole may have been doing. Thus, approaches that seek to paint pictures about potential past diversification regimes at very incremental time periods are certainly suspect, with Louca and Pennell (2020) pointing to additional congruence issues that can affect diversification models.

Some feel that, even in the face of these congruence issues, understanding macroevolution remains an exciting and promising endeavor (Helmstetter et al. 2021). We are not nearly as optimistic. We can certainly learn about diversification processes from trees, but we need to recognize that what we can understand largely relates *only* to the surviving tips. Current SSE models and other models that infer rate heterogeneity across taxa, rather than across time, may provide additional information that lets them fit different parameters and likelihood for trees with identical lineage through time curves, avoiding the particular issue raised by Kubo and Iwasa (1995) and Louca and Pennell (2020). However, as with Brownian motion and coin flipping, congruent models can likely be found for these as well. Sister group analyses may grow in importance in future studies of diversification.

Literature Cited

Abrams, P.A. 1986. Is predator-prey coevolution an arms race? Trends in Ecology and Evolution 1:108-110.

Ackerly, D.D., D.W. Schwilk, and C.O. Webb. 2006. Niche evolution and adaptive radiation: testing the order of trait divergence. Ecology 87:S50-S61.

Alexander, H.K., A. Lambert, and T. Stadler. 2016. Quantifying age-dependent extinction from species phylogenies. Systematic Biology 65:35-50.

Barido-Sottani, J., T.G. Vaughan, and T. Stadler. 2018. Detection of HIV transmission clusters from phylogenetic trees using a multi-state birth-death model. Journal of the Royal Society Interface 15:https://doi.org/10.1098/rsif.2018.0512

Beaulieu, J.M., D.-C. Jhwueng, C. Boettiger, and B.C. O'Meara. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66: 2369-2383.

Beaulieu, J.M., B.C. O’Meara, M.J. Donoghue. 2013. Identifying hidden rate changes in the evolution of a binary morphological character: the evolution of plant habit in campanulid angiosperms. Systematic Biology 62: 725:737.

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

Boyko, J.D., and J.M. Beaulieu. 2021. Generalized hidden Markov models for phylogenetic comparative methods. Methods in Ecology and Evolution 12: 468:478.

Butler, M.A., and A.A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. The American Naturalist 164:683-695.

Bokma, F. 2008. Bayesian estimation of speciation and extinction probabilities from (in)complete phylogenies. Evolution 62:2441–2445.

Caetano, D.S., B.C. O'Meara, and J.M. Beaulieu. 2018. Hidden state models improve state-dependent diversification approaches, including biogeographic models. Evolution 72:2308-2324.

Condamine, F.L., D. Silvestro, E.B. Koppelhus, and A. Antonelli. 2020. The rise of angiosperms pushed conifers to decline during global cooling. Proceedings of the National Academy of Sciences, USA 117:28867-28875.

Condamine, F.L., J. Rolland, and H. Morlon. 2019. Assessing the causes of diversification slowdowns: temperature-dependent diversity-dependent models receive equivalent support. Ecology Letters 22:1900-1912.

Condamine, F.L., J. Rolland, and H. Morlon. 2013. Macroevolutionary perspectives to environmental change. Ecology Letters 16:72-85.

Cunningham, C.W., K.E. Omland, and T.H. Oakley. 1998. Reconstructing ancestral character states: a critical reappraisal. Trends in Ecology and Evolution 13:361-366.

Dawkins R., and J.R. Krebs. 1979. Arms races between and within species. Proceedings of the Royal Society, B 205:489-511.

Eriksson, O. E.M. Friis, and P. Lofgren. 2000. Seed size, fruit size, and dispersal systems in angiosperms from the Early Cretaceous to the Late Tertiary. The American Naturalist 156:47-58.

Etienne, R.S., B. Haegeman, T. Stadler, T. Aze, P. Pearson, A. Purvis, and A.Phillimore. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. Proceedings of the Royal Society, B 279:1300–1309.

Felsenstein J. 1988. Phylogenies and quantitative characters. Annual Review of Ecology and Systematics 19:445-471.

Hansen, T.F., and E.P. Martins. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. Evolution 50:1404-1417.

Helmstetter, A.J., S. Glemin, J. Käfer, R. Zenil-Ferguson, H. Sauquet, H. de Boer, L-P. M.J. Dagallier, N. Mazet, E.L. Reboud, T.L.P. Couvreur, F.L. Condamine. 2021. Pulled diversification rates, lineages-through-time plots and modern macroevolutionary modeling. Systematic Biology in press.

Käfer, J., and S. Mousset. 2014. Standard sister clade comparison fails when testing derived character states. Systematic Biology 63:601-609.

Kubo, T., and Y. Iwasa. 1995. Inferring the rates of branching and extinction from molecular phylogenies. Evolution 49:694-704.

Lambert, A., and T. Stadler. 2013. Birth-death models and coalescent point processes. Theoretical Population Biology 90:113-128.

Landis M.J., D.A.R Eaton, W.L. Clement, B. Park, E.L. Spriggs, P.W. Sweeney, E.J. Edwards, and M.J. Donoghue. 2020. Joint phylogenetic estimation of geographic movements and biome shifts during the global diversification of Viburnum. Systematic Biology 70:76–94.

Louca, S., and M.W. Pennell. 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature 580:502-506.

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

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

Maliet, O., F. Hartig, and H. Morlon. 2019. A model with many small shifts for estimating species-specific diversification rates. Nature Ecology and Evolution 3:1086-1092.

Mitter, C. B. Farrell, B. Wiegmann. 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? American Naturalist 132: 107-128.

Morlon, H., T.L. Parsons, and J.B. Plotkin. 2011. Reconciling molecular phylogenies with the fossil record. Proceedings of the National Academy of Sciences, USA 108:16327–16332.

Morlon, H., F. Hartig, and S. Robin. 2020. Prior hypotheses or regularization allow inference of diversification histories from extant timetrees. bioRxiv doi: https://doi.org/10.1101/2020.07.03.185074

Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994. Extinction rates can be estimated from molecular phylogenies. Philosophical Transactions of the Royal Society B344:77–82.

Nee, S., A.Ø. Mooers, and P.H. Harvey. 1992. The tempo and mode of evolution revealed from molecular phylogenies. Proceedings of the National Academy of Sciences, USA 89:8322–8326.

Oakley T.H., and C.W. Cunningham. 2000. Independent contrasts succeed where ancestor reconstruction fails in a known bacteriophage phylogeny. Evolution 54:397-405.

O'Meara, B.C., C. Ané, M.J. Sanderson, and P.C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60:922-933.

Omland, K.E. 1999. The assumptions and challenges of ancestral state reconstructions. Systematic Biology 48:604-611.

Pennell M.W., J.M. Eastman, G.J. Slater, J.W. Brown, J.C. Uyeda, R.G. FitzJohn, M.E. Alfaro, and L.J. Harmon. 2014. geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics 30:2216-2218.

Pillai, A.S., S.A. Chandler, Y. Liu, A.V. Signore, C.R. Cortez-Romero, J.L.P. Benesch, A. Laganowsky, J.F. Storz, G.K.A. Hochberg, and J.W. Thornton. 2020. Origin of complexity in haemoglobin evolution. Nature 581:480-485.

Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60:1152–1164.

Rabosky, D. L. 2009. Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. Systematic Biology 58:629–640.

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

Rabosky, D. L., and I. J. Lovette. 2008. Density-dependent diversification in North American wood warblers. Proceedings of the Royal Society, B 275:2363–2371.

Raup, D.M., and J.J. Sepkowski, Jr. 1984. Periodicity of extinction in the geological past. Proceedings of the National Academy of Sciences, USA 81:801-805.

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

Revell, L.J. 2021. A variable-rate quantitative trait evolution model using penalized likelihood. bioRxiv doi: https://doi.org/10.1101/2021.04.17.440282.

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

Slowinski, J.B., C. Guyer. 1993. Testing whether certain traits have caused amplified diversification: an improved method based on a model of random speciation and extinction. The American Naturalist 142:1019-1024.

Thomas, G.H., R.P. Freckleton, and T. Szekely. 2006. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. Proceedings of the Royal Society, B 273:1619-1624.

Thornton, J.W., E. Need, and D. Crews. 2003. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. Science 301:1714-1717.

Tifney, B.H. 1984. Seed size, dispersal syndromes, and the rise of angiosperms: evidence and hypothesis. Annals of the Missouri Botanical Garden 71:551-576.

Vasconcelos, T., B.C. O'Meara, and J.M. Beaulieu. 2021. Hidden state-only speciation and extinction models provide accurate tip-estimates of diversification rates. Submitted.

Vasconcelos, T. B.C. O’Meara, and J.M. Beaulieu. 2021. Retiring “cradles” and “museums” of biodiversity. American Naturalist in press. https://doi.org/10.1086/717412

Figure Legends

Figure 1: Probability of heads per flip on different models of coin flipping. Each of these models can fit the same dataset of two heads, eight tails with equal likelihood but make very different predictions about the next flip.

Figure 2: Million taxon tree from Louca and Pennell (2020). The purple lines separate the regimes used to estimate rates. The thin vertical lines in a rainbow distinguish regimes with 100 events within them representing equal-sized slices of data. Half the regimes are on each side of the green band, showing how much of the data are near the tips. The brackets show how many events occur in each regime.

Figure 3: Comparison of net diversification, speciation, and extinction rate of conifers using as a predictor the best model from Condamine et al. (2020) in blue where only extinction rate varies with angiosperm diversity, a slightly worse model from that paper (green) where speciation rate varies with angiosperm diversity, a model (yellow) that fits the data best (at least in terms of likelihood — the number of free parameters of the spline is hard to compare), and using scaled IMDB ratings of the television program the Simpsons (red) as a predictor for speciation rate (which did a better job predicting conifer diversification than angiosperm diversity did). Not shown are numerous other attempts for other predictors using other splines, linear change models, and ratings of many other television programs. Some of these also outperformed angiosperms, but many did not.

Figure 4: (A) Depicts the identical lineage through time (LTT) plots for three trees that differ in terms of tree balance. The procedure takes a simulated tree, then makes swaps across branches to either increase balance or decrease it but maintain the same lineage through time curve. (B) Depicts the log-likelihood score among the three trees under a two-rate MiSSE model. These trees produce identical log-likelihoods under taxon-homogeneous, time-heterogeneous models that use LTT data. However, this is not the case here because allowing rates to vary among clades, as our MiSSE models do, avoids the trap of having an infinite array of congruent models. Helmstetter et al. (2021) reach similarly positive conclusions about the possibility of learning about diversification from SSE models.