



# Challenges in the estimation of extinction from molecular phylogenies: A response to Beaulieu and O'Meara

Daniel L. Rabosky<sup>1,2</sup>

<sup>1</sup>*Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, Michigan 48103*

<sup>2</sup>*E-mail: drabosky@umich.edu*

Received June 9, 2015

Accepted November 13, 2015

Time-calibrated phylogenies that contain only living species have been widely used to study the dynamics of speciation and extinction. Concerns about the reliability of phylogenetic extinction estimates were raised by Rabosky (2010), where I suggested that unaccommodated heterogeneity in speciation rate could lead to positively biased extinction estimates. In a recent article, Beaulieu and O'Meara (2015a) correctly point out several technical errors in the execution of my 2010 study and concluded that phylogenetic extinction estimates are robust to speciation rate heterogeneity under a range of model parameters. I demonstrate that Beaulieu and O'Meara underestimated the magnitude of speciation rate variation in real phylogenies and consequently did not incorporate biologically meaningful levels of rate heterogeneity into their simulations. Using parameter values drawn from the recent literature, I find that modest levels of heterogeneity in speciation rate result in a consistent, positive bias in extinction estimates that are exacerbated by phylogenetic tree size. This bias, combined with the inherent lack of information about extinction in molecular phylogenies, suggests that extinction rate estimates from phylogenies of extant taxa only should be treated with caution.

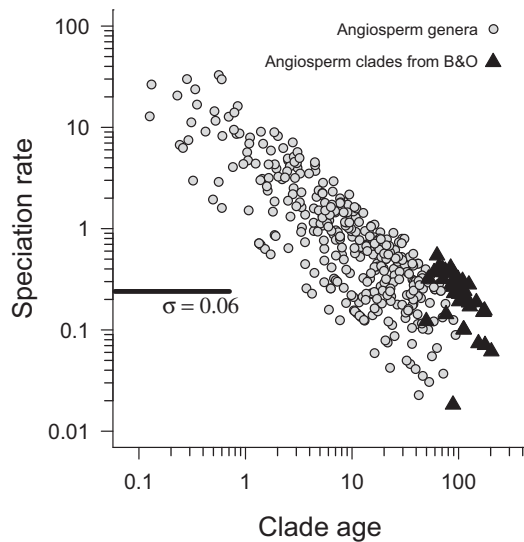
**KEY WORDS:** Birth–death model, diversification, heterotachy, model adequacy, speciation.

A seminal set of papers from the mid-1990s (Hey 1992; Nee et al. 1994a,b; Kubo and Iwasa 1995) demonstrated that branching patterns in time-calibrated phylogenetic trees contain information about speciation and extinction rates, and that these rates are potentially separable. Many previous studies have found that confidence in phylogenetic extinction estimates can be low, even when the assumptions of the inference model are perfectly satisfied (Nee et al. 1994a; Maddison et al. 2007; Davis et al. 2013). Rabosky (2010) suggested that phylogenetic extinction estimates are particularly susceptible to model inadequacy resulting from unaccommodated speciation rate heterogeneity.

In a recent paper, Beaulieu and O'Meara (2015a) suggested that the simulation model used by Rabosky (2010) was parameterized inappropriately, as it introduced biologically implausible levels of speciation rate variation. They revised the simulation protocol from Rabosky (2010) and concluded that, at least for the

specific model violations they considered, extinction rates could be estimated with some confidence from molecular phylogenies in the presence of unaccommodated speciation rate heterogeneity. They also pointed out that Rabosky's (2010) simulations led to an excess of small phylogenies and demonstrated that analyses of such small trees can lead to problematic inference. Using simple models of speciation and extinction, they found little evidence for the biased extinction estimates described by Rabosky (2010) and concluded that larger tree sizes alone are sufficient to overcome biases in extinction rate estimates generated by among-lineage variation in diversification rates.

In this article, I respond to Beaulieu and O'Meara's (2015a) critique of Rabosky (2010). Beaulieu and O'Meara (2015a) correctly point out several important issues concerning Rabosky's (2010) study, particularly with respect to the variance accumulation process that was used to introduce speciation rate



**Figure 1.** Speciation rate estimates for 361 angiosperm genera as a function of crown clade age, assuming a constant-rate birth–death process with a relative extinction rate of 0.75. Rates are inversely correlated with the timescale over which they are measured. B&O parameterized their simulations using estimates of speciation rates for major angiosperm clades (black diamonds). Under a parameterization they describe as “extreme” ( $\sigma = 0.06$ ), B&O observed a shift in mean speciation rates from a starting value of 0.196 lineages/my to 0.232 lineages/my (black line) after 50 million years of simulation time. This maximum value is much lower than speciation rates estimated for most angiosperm genera, which are approximately 80 my younger on average than the clades considered by B&O. Approximately 18.9% of genera ( $n = 68$ ) have speciation rates that are at least 10 times greater than this maximum value. Because rates are correlated with the timescale of measurement, speciation rates as estimated for ancient clades cannot be used for inference about plausible rates for young clades. A single outlier genus is not shown (*Euthamia*), which is inferred to have a crown age of 10,000 years and four extant species;  $\lambda$  for *Euthamia* was estimated as 168 lineages/my, a value similar to those inferred for some East African cichlid lineages (Coyne and Orr 2004).

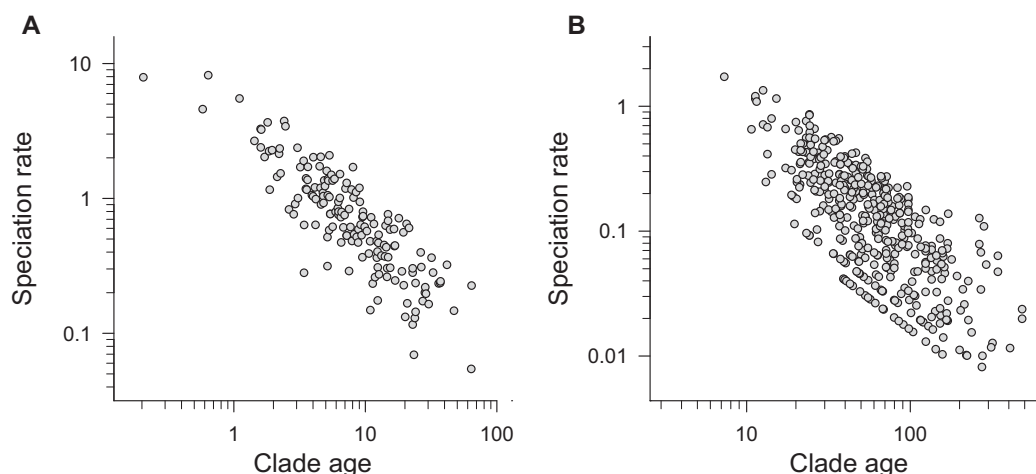
heterogeneity during the simulation of phylogenetic trees. I agree with Beaulieu and O’Meara (2015a) that Rabosky (2010) did not adequately address whether his simulations were performed in a biologically relevant area of parameter space. However, I demonstrate below that Beaulieu and O’Meara (hereafter B&O) did not adequately address this concern in their own analyses, and their study underestimated both plausible speciation rates and the magnitude of phylogenetic tree imbalance for real phylogenies. The most extreme parameterization considered plausible by B&O (“ $\sigma = 0.06$ ” Figs. 1 and 2 in their study) resulted in an 18% increase in mean speciation rate for clades after 50 million years of simulation time. This range of variation is less than that reported for most empirical studies on rate heterogeneity within species-

level phylogenies. Phylogenies generated under this parameter show topological properties (tree balance) more consistent with constant-rate diversification processes than with published empirical phylogenies. The analyses in B&O relied on the assumption that speciation rates for ancient clades (e.g., orders of plants) as estimated under constant-rate diversification models provide information about maximum plausible speciation rates for much younger clades that are the focus of most lineage-level diversification analyses. This assumption is erroneous for the flowering plants they consider and for most other groups of organisms. I demonstrate below that more than half of all flowering plant genera have speciation rates that exceed the maximum value they achieve under their purportedly extreme simulation parameterization. Hence, I show that their central criticism of Rabosky (2010)—that Rabosky’s model produces unrealistically high average rates of speciation for clades—is not supported.

I then address B&O’s claim that extinction rate estimates from constant-rate birth–death models perform well even in the face of realistic levels of speciation rate heterogeneity. Using parameter values drawn from the recent empirical literature, I show that it is straightforward to construct alternative models of speciation rate variation whereby extinction rate estimates show biases that are similar to those documented by Rabosky (2010). In contrast to B&O, I find that tree size exacerbates these problems by inflating our confidence in the incorrect estimates. The conflict between my results and those of B&O is a simple consequence of the fact that B&O’s simulations incorporated only modest levels of speciation rate heterogeneity. Many phylogenies generated under their most extreme parameterizations are similar to those produced by a constant-rate birth–death process, and it is thus unsurprising that they found good statistical performance of constant-rate models when applied to those same phylogenies.

### *Beaulieu and O’Meara Underestimated Real Speciation Rates*

Rabosky (2010) and B&O studied a statistical model whereby speciation rates ( $\lambda$ ) evolve across the branches of a phylogenetic tree during the diversification process. The principal difference between their parameterizations involves the magnitude of variation in the process by which speciation rate heterogeneity arises with respect to time. In B&O, the variation introduced by this process is denoted by parameter  $\sigma$ : larger values of  $\sigma$  yield phylogenies with more variation in speciation rate. Consistent with Rabosky (2010), B&O find that among-lineage variation in speciation rate can lead to substantial biases in extinction rate estimates (Fig. 3 in B&O). Because B&O and Rabosky (2010) agree that rate heterogeneity can be problematic for extinction rate ( $\mu$ ) estimates in



**Figure 2.** Speciation rates as estimated under constant-rate birth–death models are inversely correlated with the timescale of measurement across a broad range of taxa. (A) Results for 163 species-level phylogenies of animals; data from McPeck and Brown (2007). (B) Results for 421 family-level vertebrate clades; data from Rabosky et al. (2012). Analyses in (A) and (B) used crown- and stem clade estimators of speciation rate respectively, and all analyses assumed a relative extinction rate of 0.75.

principle, the key difference between these studies involves what they view as biologically plausible variation in speciation rates.

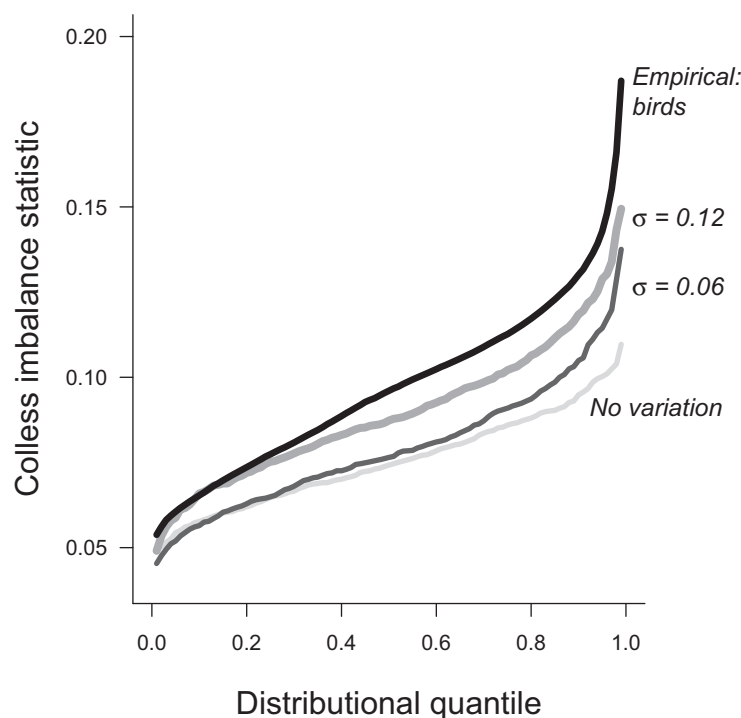
My discussion of specific parameters in B&O's model requires some background for context. B&O find that a value of  $\sigma = 0.06$  does not generate biased estimates of extinction (Fig. 1 in B&O). In B&O, Figure 3 suggests that  $\sigma = 0.12$  also generates largely unbiased estimates of relative extinction, but that higher values (e.g.,  $\sigma = 0.24$ ) are biased away from the true value. However, their study only presents results for relative extinction rates ( $\mu/\lambda$ ) equal to 0.75 (their Figure 3). I found that when true relative extinction rates are low, a value of  $\sigma = 0.12$  produces biased estimates of extinction under their model parameterization. I simulated 500 phylogenies of  $N = 100$  taxa, under  $\sigma = 0.12$  and following the simulation protocol described in B&O. The median relative extinction rate estimated for these phylogenies under a constant-rate birth–death process was 0.48 (interquartile range: 0.27–0.65). Using a likelihood ratio test, I find that 35.6% of datasets simulated in this fashion reject the true relative extinction rate (zero) with  $P < 0.001$ . Hence, the shift from “largely unbiased” to “substantially biased” relative extinction rates in B&O's model occurs somewhere between values of  $\sigma = 0.06$  and  $\sigma = 0.12$ . My discussion of this model hinges on the biological relevance of these parameter values.

B&O did not estimate among-lineage variation in speciation rate for any real phylogenies. Rather, B&O assessed the relevance of their parameters by testing whether different values of  $\sigma$  could produce an average speciation rate for simulated phylogenies that are consistent with speciation rates observed in real clades. They estimated average speciation rates for a set of large plant clades (data from Magallon and Castillo 2009), under the assumption that the clades had been diversifying under a constant-rate birth–

death process. They discussed speciation rates in terms of the average waiting times between successive speciation events that such rates would imply; this is simply the reciprocal of  $\lambda$ . Using simulations, they show that values of  $\sigma$  that would lead to problematic extinction estimates would also lead to mean speciation rates across the phylogeny that are far too high given the data on large plant clades (Fig. 3 in B&O).

This test is inadequate because it assumes that biologically plausible speciation rates are those that are most consistent with old, large clades. B&O claimed that their parameterizations introduce biologically relevant speciation rate variation because the long-term average speciation rate produced by their process matches that estimated for plant clades that have a median crown age of 93.9 million years and a median species richness of 2281 species. Hence, their assessment of model adequacy assumes that the diversification process is homogeneous with respect to the timescale over which it is measured. Their conclusions are only valid if the diversification parameters estimated for old, large clades are representative of those that we would observe for a sample of much younger and smaller clades.

The assumption that evolutionary rates are homogeneous across the timescale of measurement is demonstrably false for most evolutionary processes where it has been examined. For example, Gingerich (1983, 2001) demonstrated that rates of morphological evolution are inversely correlated with the interval of time over which they are measured. Rates of morphological evolution for old clades thus appear to be much slower than those for young clades. Similar results have been observed for rates of molecular evolution (Ho et al. 2011) and even for nonbiological processes, such as rates of sediment accumulation in the geological record (Sadler 1981).



**Figure 3.** Distributional quantiles for Colless's index of phylogenetic tree imbalance for (1) pure-birth phylogenies (no variation), (2) phylogenies simulated with rate heterogeneity parameter  $\sigma = 0.06$ , (3) phylogenies simulated with  $\sigma = 0.12$ , and (4) empirical phylogenies (birds). For any distributional quantile, measures of phylogenetic tree imbalance for the empirical phylogenies exceed those for phylogenies simulated under B&O's model parameterizations with  $\sigma = 0.06$  and  $\sigma = 0.12$ . A substantial fraction of phylogenies simulated under  $\sigma = 0.06$  is statistically indistinguishable from a pure-birth phylogeny with no rate variation among lineages and suggests that this parameterization does not generate levels of phylogenetic tree imbalance consistent with empirical datasets.

There is extensive evidence that speciation rates as measured over long timescales are much slower than those measured over short timescales (Ricklefs 2006; McPeck and Brown 2007; Rabosky and Adams 2012; Rosenblum et al. 2012). To demonstrate this phenomenon, I revisited the taxon considered by B&O—flowering plants—and ask whether they would have reached different conclusions about plausible speciation rates had they analyzed clades more representative of those typically studied in molecular phylogenetic studies (e.g.,  $N < 2000$ , clade ages  $< 90$  million years). I repeated the exercise from Figure 3 in B&O, but with a much younger set of clades. I used the same analytical protocol from B&O to estimate  $\lambda$  for angiosperm genera, using information on crown clade age and extant species richness and assuming a relative extinction rate of 0.75.

To obtain a working list of angiosperm genera and their species richness, I downloaded a dataset of all angiosperm species from a comprehensive database (“The Plant List” [TPL]) of all vascular plant taxa (www.theplantlist.org, Version 1.1). I restricted the dataset to “Accepted” names, giving a total of 326,711 named species. To estimate ages of genera, I used a time-calibrated phylogeny for 32,223 species of plants (Zanne et al. 2014) and identified the common set of genera from both the phylogeny

and TPL under a strict matching criterion. As additional filtering criteria, I excluded all genera that were not monophyletic in the Zanne et al. (2014) phylogeny. The Zanne et al. (2014) phylogeny contains approximately 10% of angiosperm species, and we thus expect that the crown ages of many genera are underestimated. For each genus, I used the number of taxa included in the phylogeny and the total number of known species in TPL to estimate the probability that the Zanne et al. (2014) phylogeny contained the true root node of the clade. This latter probability was derived by Sanderson (1996) for the Yule process, and I refer to it as the “crown sampling probability.” I excluded all genera for which the crown sampling probability was less than 0.80 as well as genera that did not diversify ( $N < 3$  species). The resulting quality-controlled dataset included clade age and richness data for 361 genera of angiosperms, with a median genus age of 10.8 million years and a median richness of 27 species (mean = 101.2).

I find a clear negative relationship between speciation rate and clade age for angiosperm genera (Fig. 1). Note that I display results in the more conventional units of “speciation rate,” rather than “waiting times,” as in B&O. B&O suggest that simulations with  $\sigma = 0.06$  generate speciation rates more extreme

than reality (Fig. 3B in B&O), but the mean speciation rate from simulations they performed with this parameter only reached 0.232 species/million years (Fig. 1, black horizontal line). My results suggest that fully 77% of angiosperm genera have speciation rates that are faster than this value. Some of the speciation rates in Figure 1 appear to be extreme but most (with one outlier genus; see legend of Fig. 1) are consistent with published higher end estimates for plant clades (Baldwin and Sanderson 1998; Richardson et al. 2001; Valente et al. 2010). For example, the Eurasian carnations are estimated to have speciation rates of approximately 10 lineages/million years (Valente et al. 2010; rate calculations used  $\mu/\lambda = 0.75$  for comparability to Fig. 1).

The general phenomenon of time-scaling illustrated in Figure 1 is not specific to angiosperms but is a near-universal property of diversification rates estimated under constant-rate birth–death models. Figure 2 illustrates the relationship between crown clade age and speciation rate for 163 time-calibrated animal phylogenies compiled and analyzed by McPeck and Brown (2007). Figure 2B illustrates the relationship between stem clade age and speciation rate for 421 family-level clades of vertebrates; data are taken from the Rabosky et al. (2012) compilation of clade age and species richness data for major metazoan clades.

Because their evaluation of parameter realism neglected time-scaling of rates, B&O simulated phylogenies with relatively modest variation in speciation rate. For example, consider their claim that an evolving rates model with parameter  $\sigma = 0.06$  generates speciation rates that are “more extreme than reality.” After 50 million years of simulation time (Fig. 3B in B&O), the mean speciation rate they observed for clades increased from an initial value of 0.196 to 0.232. This is a trivial (18%) increase and can hardly be described as extreme, given the timescales under consideration. Empirical estimates of changes in mean rates for heterogeneous diversification processes frequently exceed these values. The reasons for these shifts in mean diversification rates are explained by Beaulieu and O’Meara (2015a): differential net rates of species diversification associated with particular character states lead to a form of species selection for lineages with faster diversification rates. For example, Beaulieu and Donoghue (2013) estimated a shift in net diversification rates associated with fruiting type from 0.043 lineages/my (ancestral) to an average present-day rate of 0.075, for an increase in mean rate of 73.5%. Many empirical studies have reported increases in speciation rate and/or net diversification rate that are much more extreme than those simulated by B&O (e.g., Valente et al. 2010; Drummond et al. 2012; Weber et al. 2014).

Regardless of the magnitude of speciation rate estimates in Figures 1 and 2, the implications are clear: inferences about plausible speciation rates for young clades using data from ancient clades are nonsensical because speciation rates—like many other rate-associated phenomena—scale inversely with the timescale

over which they are measured. The causes of time-dependent rate scaling are debatable (Gingerich 2001; Rabosky et al. 2012) but irrelevant: B&O should not have used constant-rate speciation estimates for very old clades as a benchmark against which to compare speciation rates in young clades with several hundred or fewer species. The test used by B&O thus fails to demonstrate that Rabosky’s (2010) parameterization produces biologically irrelevant variation in speciation rate. Their analyses merely demonstrate that the process studied by Rabosky (2010) would produce unreasonable mean rates if the process were left unchecked over timescales that are much greater than most empirical studies involving species-level phylogenies.

## A Test Using Phylogenetic Tree Imbalance

I performed a second set of analyses to determine whether the purportedly extreme ( $\sigma = 0.06$  and  $\sigma = 0.12$ ) parameters considered by B&O generate phylogenies with realistic levels of phylogenetic tree imbalance. Tree imbalance is a topological measure of among-lineage variation in diversification rates (Mooers and Heard 1997): evolutionary processes with among-lineage differences in diversification should produce less-balanced phylogenies than processes where all lineages have identical rates. My analytical protocol entailed simulating phylogenies under several evolutionary models, computing a whole-tree measure of imbalance for each simulated phylogeny, and then comparing these estimates to a distribution of imbalance measures computed for an empirical set of comparably sized phylogenies.

As a measure of phylogenetic tree imbalance, I used Colless’s imbalance statistic (Colless 1982; corrected by Heard 1992). The statistic ( $I_c$ ) is computed as

$$I_c = \frac{2}{(n-1)(n-2)} \sum_{i=1}^B |R_i - L_i|,$$

where  $n$  is the number of taxa in the phylogeny,  $B$  is the number of internal nodes (typically  $n - 1$ ), and  $R_i$  and  $L_i$  are the numbers of taxa in the clades subtended by the left ( $L_i$ ) and right ( $R_i$ ) branches descended from the  $i$ th internal node. I simulated 500 phylogenies of  $N = 100$  taxa under the model described by B&O with parameters of  $\sigma = 0.06$  and  $\sigma = 0.12$ , with initial diversification rates of  $\lambda = 0.078$  and  $\mu = 0$  in each set of simulations. For comparison, I also simulated 500 phylogenies under a pure-birth process (e.g., no variation in rates among lineages).

I compared the distribution of tree imbalance statistics from simulated phylogenies to that obtained for comparably sized phylogenies of birds. I used a published distribution of time-calibrated phylogenies that is complete at the species level and which accounts for multiple sources of phylogenetic uncertainty (Jetz et al.



2012). Due to the large number of missing taxa ( $\sim 90\%$ ), I did not perform this exercise on the Zanne et al. (2014) plant phylogenies, as there is some evidence that incomplete taxon sampling can increase the apparent imbalance within phylogenetic trees when diversification rate heterogeneity is present (Heath et al. 2008). I randomly sampled 100 phylogenies from the Jetz et al. (2012) distribution (9993 species each) and extracted all subclades from these complete phylogenies that contained between 75 and 125 taxa, thus ensuring size comparability with the simulated phylogenies of  $N = 100$  taxa. The number of such clades for each complete tree in the Jetz et al. (2012) distribution ranged from 121 to 142 (median = 132), for a total of 13,174 clades. I computed  $I_c$  for all such subclades and thus obtained four distributions of imbalance statistics: (1)  $I_c$  for 500 phylogenies with no rate variation, (2)  $I_c$  for 500 phylogenies simulated with  $\sigma = 0.06$  under B&O's model of rate heterogeneity, (3)  $I_c$  for 500 phylogenies simulated with  $\sigma = 0.12$  under B&O's model, and (4)  $I_c$  for 13,174 avian clades with 75–125 taxa. Obviously, the avian clades are not statistically independent: some clades may be nested within other clades included in the dataset, and all taxa are represented in each of the 100 complete phylogenies sampled from the Jetz et al. (2012) distribution. However, this distribution enables us to ask the following question: how much imbalance should we expect for a randomly sampled avian clade with 75–125 taxa?

Phylogenies generated under B&O's model with  $\sigma = 0.06$  are only marginally more imbalanced than expected under a pure-birth model. For  $\sigma = 0.06$ , the median  $I_c$  value is 0.076 (interquartile range: 0.065–0.091), versus a median of 0.074 under a model with no rate variation (interquartile range: 0.065–0.086). For  $\sigma = 0.12$ , phylogenies are less balanced than expected under a pure-birth model (median  $I_c$ : 0.087; interquartile range: 0.075–0.10). However, the avian phylogenies are more imbalanced on average than any simulated datasets, with a median  $I_c$  of 0.098 and an interquartile range of 0.081–0.120. Fully, 76.3% of avian phylogenies have  $I_c$  values that exceed the median value estimated for simulations with  $\sigma = 0.06$ . A graphical comparison of these distributions is shown in Figure 3.

These results indicate that B&O's parameterization generates phylogenies with substantially less among-lineage rate variation than observed for comparably sized avian phylogenies sampled at random from the set of all avian phylogenies. Indeed, the only parameterization considered by B&O that yields unbiased extinction estimates ( $\sigma = 0.06$ ) results in patterns of tree imbalance that are similar to that which would be obtained from a model with no among-lineage variation in diversification rates (Fig. 3). The distribution of  $I_c$  for the  $\sigma = 0.06$  simulations is similar to the distribution obtained under a simple pure-birth process with no rate variation (Fig. 3, [1] vs. [2]). True patterns of tree imbalance for birds are likely even more severe than those reported here, because stochastic polytomy resolution (Kuhn et al. 2011) was used to

estimate the positions of 33% of taxa in the Jetz et al. (2012) phylogenies, which should equalize taxon placements across clades and reduce measured levels of imbalance.

## Rate Heterogeneity is Problematic for Extinction Estimates

Assessing the validity of parameters in the evolving rates model studied by Rabosky (2010) and B&O is difficult: the model remains effectively untested as there is yet no implementation of the model that can be fitted to real data. Here, I demonstrate that it is straightforward to formulate more realistic alternative models with rate heterogeneity and which demonstrate the statistical pathologies described by Rabosky (2010). I use a discrete character change model as the basis for introducing speciation rate heterogeneity across the branches of a diversifying clade. In contrast to the untested evolving rates model, there is at least some evidence that diversification rates are often linked to discrete character state changes (Coyne and Orr 2004; Ng and Smith 2014) or that diversification rates shift discretely on phylogenetic trees (Alfaro et al. 2009; Rabosky 2014). I simulated phylogenies under a BiSSE process (Maddison et al. 2007) with two speciation rates ( $\lambda_0, \lambda_1$ ), two extinction rates ( $\mu_0, \mu_1$ ) and unequal transition rates ( $q_{01} \neq q_{10}$ ).

I used the BiSSE process as a biologically plausible model for rate heterogeneity that describes a form of rate variation that many researchers believe exists in real data. Note that I am not explicitly analyzing trait dependent diversification: I am only using the BiSSE process to generate phylogenies that include a mixture of fast- and slow-diversifying lineages. At the end of the simulation, I simply discard information about which lineages were assigned to “fast” and “slow” diversification states. As in Rabosky (2010) and B&O, I analyzed each resulting phylogeny with a constant-rate birth–death model.

This analytical procedure should not be viewed as an endorsement of the idea that simple constant-rate models should be preferred in macroevolutionary studies, and B&O agree with this point. Beaulieu and O'Meara (2015b) recently developed a “hidden-state” model that could be used to estimate extinction rates under precisely the scenario considered here. The simulations below thus use a generating model to which there is an existing analytical solution. However, B&O suggested that biologically relevant variation in speciation rates is unlikely to confound extinction estimates for moderately sized phylogenies, even when simple constant-rate models are used. My objective below is to explicitly test this assertion, as in the original Rabosky (2010) paper that was the focus of B&O's critique. Moreover, many approaches for modeling diversification rates assume rate-constancy across all or part of a phylogenetic tree, and it is thus important to

understand the extent to which unaccommodated speciation rate variation is problematic for inference.

I simulated phylogenies of  $N = 100$  and  $N = 300$  species under four relative extinction rates, such that each phylogeny was shaped by a mixture of two distinct net diversification rates. Each simulation was initiated with a “slow” diversification rate ( $r_0 = \lambda_0 - \mu_0$ ) and switched to a “fast” diversification rate ( $r_1 = \lambda_1 - \mu_1$ ) with rate  $q_{01}$ ; the reverse transition occurred with rate  $q_{10}$ . In each case, the ratio of extinction to speciation ( $\epsilon$ ) was constrained across the phylogeny. As in Rabosky (2010) and Beaulieu and O’Meara (2015a), I considered relative extinction rates of 0, 0.25, 0.5, and 0.75. For a value of  $\epsilon = 0$ , extinction rates were constrained to zero, and diversification rate heterogeneity reflects variation in the rate of speciation alone.

I obtained parameters for these simulations using two recently published studies of trait-dependent diversification. The first involved a study of fruit type and diversification in several major plant clades (Beaulieu and Donoghue 2013). Specifically, I assumed values for the best-fit BiSSE model from the Campanulidae (Table 2 in Beaulieu and Donoghue 2013), but with extinction constrained to zero ( $r_0 = 0.0431$ ,  $r_1 = 0.1108$ ,  $q_{01} = 0.002$ ,  $q_{10} = 0.0001$ ). The “fast” net diversification rate in this parameter set is 2.57 times faster than the “slow” ancestral rate. For a second set of simulations, I used net diversification rates and transition rates from a recent study of the effect of extrafloral nectaries on the diversification of *Viburnum* shrubs (Weber and Agrawal 2014). I rescaled parameters from their best-fit model reported in Table S3 by dividing all parameters by the observed value for  $\lambda_0$ . In the Weber and Agrawal (2014) study, the presence of the focal character resulted in a 4.06-fold increase in the net rate of lineage diversification. Numerous studies have reported diversification rate increases within empirical phylogenies that are as great or greater than those reported here. These include studies where derived character states have faster net diversification rates than those at the root of the tree (Hugall and Stuart-Fox 2012; Bonett et al. 2013; de Vos et al. 2014; Horn et al. 2014; Fuchs et al. 2015) as well as studies where multiprocess models have revealed high levels of among-lineage rate variation (Alfaro et al. 2009; Valente et al. 2010; Drummond et al. 2012) without explicit consideration of a character-state-dependent process. I imposed no restrictions on the frequency of “fast” and “slow” character states in each tree at the end of the simulation, requiring only that each simulated phylogeny contain at least one lineage in each rate class.

I analyzed all phylogenies with explicit consideration of the proportion of species at the end of the simulation that were in the “fast” character state, a quantity that I denote by  $F_{\text{fast}}$ . For comparison, I also generated and analyzed a set of constant-rate simulations under the four relative extinction rates. A total of 5000 phylogenies were generated for each unique set of simulation

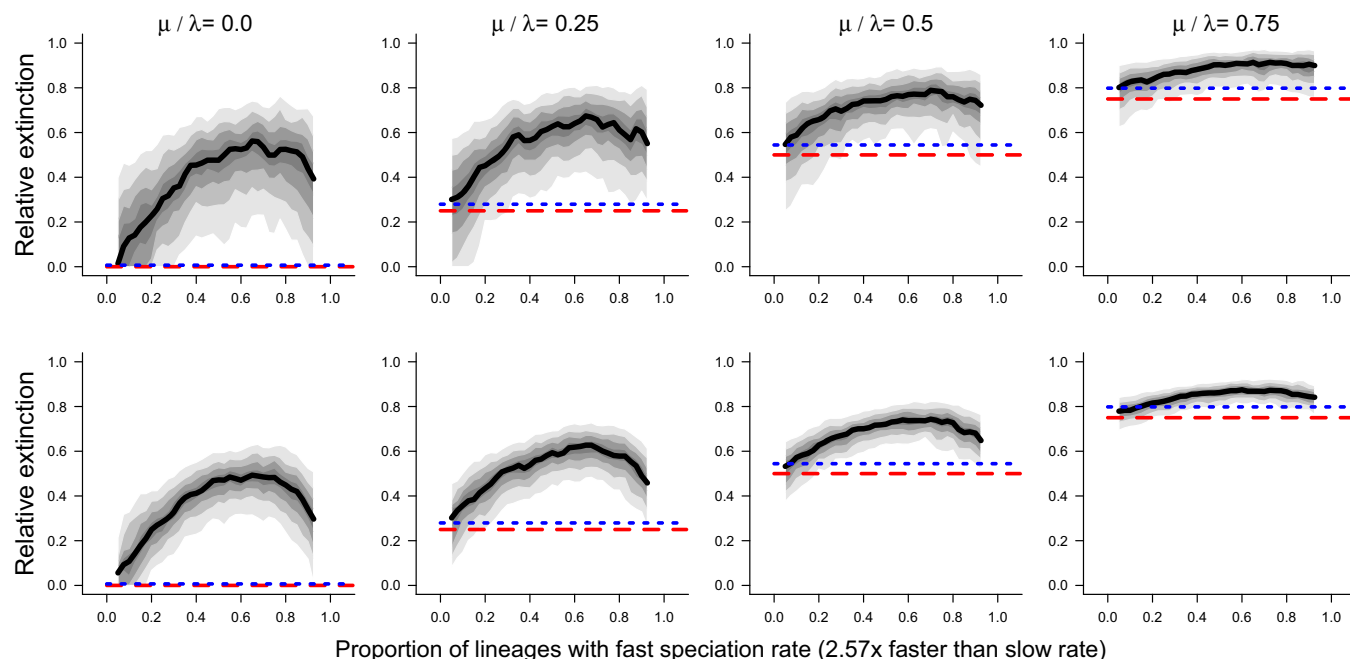
parameters considered here, and all simulation and analysis code from this study is available through the Dryad data repository ([dx.doi.org/10.5061/dryad.d4841](https://doi.org/10.5061/dryad.d4841)).

As in Rabosky (2010), extinction rates estimated in the presence of diversification rate heterogeneity are substantially and predictably biased: rate heterogeneity leads to a consistent, upwards bias in the relative extinction rate  $\epsilon$  (Figs. 4 and 5). As the frequency of the “fast” character state ( $F_{\text{fast}}$ ) tends to 0 or 1, the tree becomes increasingly similar to a constant-rate phylogeny, and estimates of relative extinction become similar to the true value (Figs. 4 and 5, dashed red line). However, in the presence of even a modest mixture of fast- and slow-diversifying lineages, extinction rates are biased sharply upwards. This effect is most severe when the “fast” diversification state is found in approximately 40–70% of the lineages surviving at the end of the simulation. In all cases, relative extinction rates estimated from a true constant-rate process are similar to the generating value; median values from constant-rate simulations are shown as dotted blue lines in Figures 4 and 5. In general, across all simulation scenarios, there was little overlap between the true value of  $\epsilon$  and the distribution of estimates from simulations with rate heterogeneity, except when  $F_{\text{fast}}$  was less than 0.15.

B&O reported that extinction rate estimates are less biased for larger phylogenies. In contrast, my results suggest that tree size compounds these biases. Distributions of the estimates are narrower for larger trees ( $N = 300$ , bottom panels: Figs. 4 and 5), but the median bias is approximately the same. Hence, violations of model assumptions may be more severe for larger phylogenies because our confidence in spurious estimates is increased.

## Discussion

The results of this study support the general conclusions of Rabosky (2010): unaccommodated variation in speciation rates among lineages results in substantial yet predictable biases in the estimation of extinction from time-calibrated phylogenetic trees. The causes of this phenomenon are straightforward: increases in speciation rate, for any reason, change the distribution of waiting times in a reconstructed phylogenetic tree, relative to their expectation under a time-homogeneous pure-birth process. If speciation rates have increased—either through time, or within a particular subclade—then any analytical model that fails to account for this heterogeneity will be biased toward higher relative extinction rates. This is immediately evident from comparing the shapes of lineage through time plots with high relative extinction rates to those generated under models with increasing speciation through time (Nee et al. 1994a, Rabosky 2010): both scenarios show characteristic upturns toward the present when plotted on a semilogarithmic scale and result in similar distributions of waiting times in reconstructed phylogenies.



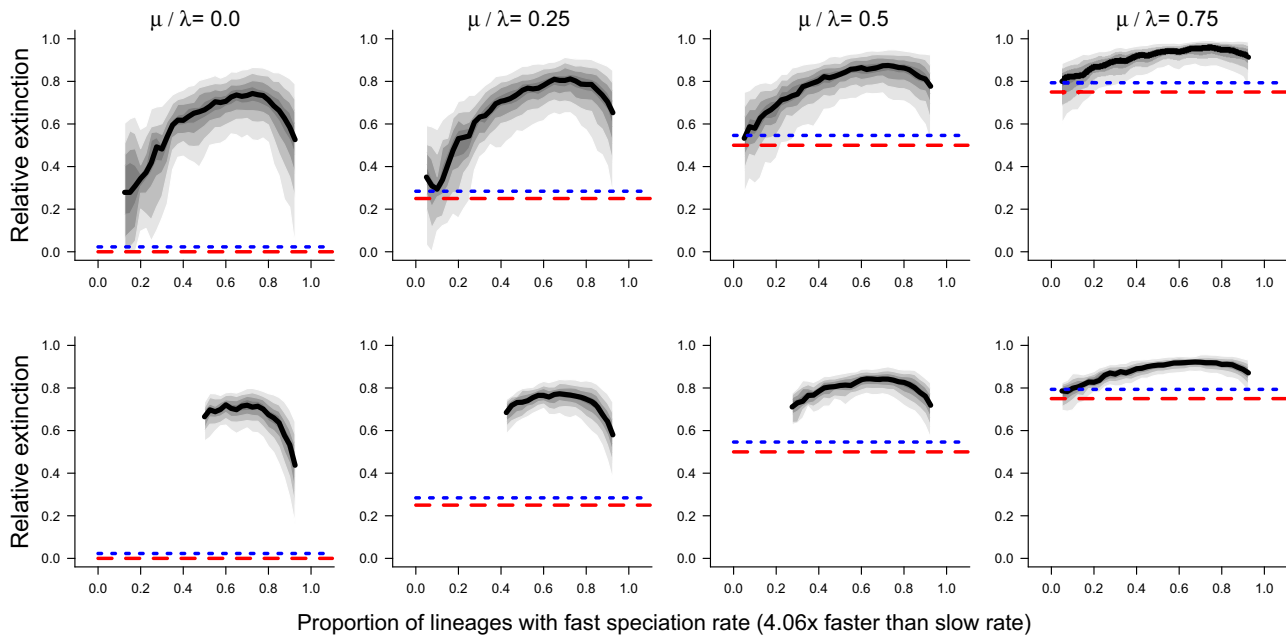
**Figure 4.** Distributions of relative extinction rate estimates for phylogenies of  $N = 100$  (top) and  $N = 300$  (bottom) taxa, simulated under a BiSSE process with parameters taken from the Campanulidae. Results are plotted as a function of the percentage of taxa in each simulated dataset possessing the “fast” character state at the end of the simulation ( $F_{\text{fast}}$ ). Red dashed line = relative extinction rate (held constant for all simulations). Solid black line = median value of relative extinction estimated by fitting a constant-rate birth–death model to datasets with speciation rate heterogeneity. Gray polygons give the 0.1–0.9 quantiles (in units of 0.1) of the distribution of estimates as a function of the “fast” state frequency at the end of the simulation ( $F_{\text{fast}}$ ). Dotted blue line = median value from 5000 simulations under the corresponding constant-rate diversification process. In each simulation scenario, the distribution of  $\varepsilon$  is biased in a positive direction. The bias diminishes as  $F_{\text{fast}}$  decreases because the phylogeny becomes more similar to a constant-rate tree. Quantiles were computed by binning simulations by their  $F_{\text{fast}}$  value in increments of 0.05; no results are shown for bins that included fewer than 20 simulated values.

Why do the results presented here differ from those obtained by B&O? It appears that the parameter space explored by B&O was unnecessarily restrictive and did not introduce biologically relevant variation in speciation rate. Admittedly, B&O at least attempted to test whether their parameters led to biologically plausible speciation rates, unlike Rabosky (2010). Beaulieu and O’Meara (2015a) point out several errors in the technical execution of Rabosky’s (2010) study. In particular, B&O convincingly demonstrated that problematic diversification inferences can result from the analysis of small phylogenetic trees (e.g.,  $N < 50$  taxa), even when the trees are generated under simple constant-rate models. As I have shown, however, B&O neglected time-scaling of evolutionary rates in their assessments of model adequacy. The consequence of this is that their simulations included a trivial degree of speciation rate variation. They implicitly assumed that distributions of speciation rates for young, species-poor clades ( $N \ll 1000$ ) could be modeled as draws from the same distribution that they estimated for very old, high-diversity plant clades. Because speciation rates for flowering plants (Fig. 1) and most other taxa (Fig. 2; Ricklefs 2006; McPeck and

Brown 2007) scale inversely with the timescale over which they are measured, B&O substantially underestimated the plausible range of speciation rate variation for real species-level phylogenies. A consequence of this is that their simulated phylogenies are characterized by less among-lineage variation in diversification rates (as quantified by phylogenetic tree balance) than real phylogenies (Fig. 3). For their simulations with  $\sigma = 0.06$ —the largest value they considered that still generates largely unbiased extinction estimates—the distribution of phylogenetic tree imbalance is generally similar to the corresponding distribution under a pure-birth process (e.g., no rate variation among lineages).

There are four reasons to be skeptical of phylogenetic extinction estimates made in the absence of fossils. First, for a phylogeny of extant species only, there is always more information available with which to estimate speciation rates. This corollary is a simple function of the fact that at least some speciation events are actually observed: for any phylogeny of  $N$  taxa, we can be certain that at least  $N - 1$  speciation events occurred during the radiation of the clade, and we know something about the timing of those events. In the case of extinction, we have no such information because





**Figure 5.** Distributions of relative extinction rate estimates for phylogenies of  $N = 100$  (top) and  $N = 300$  (bottom) taxa, simulated under a BiSSE process with parameters taken from *Viburnum*. The fast character state entailed a 4.06-fold increase in net diversification rate relative to the root state, while holding the relative extinction rate constant. Results are plotted as a function of the percentage of taxa in each simulated dataset possessing the “fast” character state at the end of the simulation ( $F_{\text{fast}}$ ). Black solid line = median value of relative extinction estimated by fitting a constant-rate birth–death model to datasets with speciation rate heterogeneity. Gray polygons give the 0.1–0.9 quantiles of the distribution of estimates as a function of the “fast” state frequency at the end of the simulation ( $F_{\text{fast}}$ ). Regardless of  $F_{\text{fast}}$ , the distributions of estimates are strongly biased and largely fail to overlap the true value. Median values from simulations lacking rate heterogeneity (dotted blue line) are similar to the true value, indicating that rate heterogeneity has led to a marked upwards bias in  $\epsilon$ . Quantiles were computed by binning simulations by their  $F_{\text{fast}}$  value, in units of 0.05; no results are shown for bins that included fewer than 20 estimates. For  $N = 300$  taxa and low  $\epsilon$ , this parameter set generally produced few phylogenies with low values of  $F_{\text{fast}}$ , resulting in the gaps in each plot with respect to  $F_{\text{fast}}$ .

any number of events—from zero to any arbitrarily large whole number—is in principle consistent with the data. This fact underlies the observation that phylogenetic extinction estimates are characterized by substantially greater uncertainty than the corresponding speciation rate estimates, even when the assumptions of the inference model are perfectly satisfied (Maddison et al. 2007; Davis et al. 2013).

Second, lineages that have undergone declines in diversity through time typically have very little data with which it is possible to estimate extinction. We might term this phenomenon the “tuatara effect” after the enigmatic pair of reptile species that presently inhabit New Zealand. These species are members of a formerly widespread and moderately diverse line of reptiles (Rhynchocephalia), but today they are represented by a single unbroken lineage extending to the origin of squamate reptiles more than 200 million years ago. There is virtually no information that can be gleaned about the extinction dynamics of this interesting clade from a single branch on a phylogeny, other than the fact that the tuatara lineage itself has been remarkably resistant to extinction. More generally, because many groups have poor fossil records, we can never know whether a given branch on a phy-

logeny is the legacy of a historical decline in diversity. The fossil record shows extensive evidence of such declines in diversity that would be invisible on phylogenies of extant lineages only (Stanley 1979; Quental and Marshall 2013).

Third, extinction estimates for many real clades appear to be inconsistent with the reality of the fossil record. Phylogenetic extinction estimates frequently tend toward zero (Purvis 2008), which leads to an unusual paradox with respect to the results of the present study: rate heterogeneity, which is known to exist in real phylogenies, should result in extinction estimates that are positively biased, not biased toward zero. There are many reasons why phylogenetic extinction estimates might be biased toward zero, some biological in nature (Rabosky 2009; Quental and Marshall 2010, 2011) and others artifactual (Revell et al. 2005; Cusimano and Renner 2010). A recent study of diversification patterns in whales (Morlon et al. 2011) reported extinction rate estimates that were consistent with the fossil record, but other studies on the same dataset using alternative inference frameworks reported a mismatch between phylogenetic extinction estimates and those in the fossil record (Etienne et al. 2012; Rabosky 2014).

Fourth—and perhaps most worryingly—confidence in many phylogenetic extinction estimates, especially for trait-dependent models of diversification, appears to exceed what is possible if the assumptions of the inference model have been perfectly satisfied. Results of Maddison et al. (2007) and Davis et al. (2013) are sobering when considered carefully, as they point to low statistical power to detect trait-dependent extinction even with relatively large phylogenies (e.g.,  $N = 500$  tips in Maddison et al. 2007). Given what is known about the sensitivity of extinction to model inadequacy, it is possible that well-supported extinction estimates from molecular phylogenies are more likely to reflect model inadequacy than real biological results. In general, an expanded statistical toolkit is needed with which to diagnose diversification model inadequacy.

Fundamental questions regarding (molecular) phylogenetic extinction estimates remain to be addressed, including (1) why empirical extinction estimates are generally at odds with those from the fossil record, and (2) whether extinction rate parameters in birth–death models are identifiably distinct, at least in practice, from alternative parameterizations that entail speciation rate variation alone. In the absence of paleontological data, it is possible that we will never be able to obtain reliable estimates of extinction or species richness from molecular phylogenies (Quental and Marshall 2010). For these questions, the future may lie with the continued integration of paleontological and neontological data.

## ACKNOWLEDGMENTS

I thank J. Beaulieu and B. O'Meara for insightful correspondence on this topic. A. Davis Rabosky, M. Friedman, B. O'Meara, and several anonymous reviewers provided helpful comments on the manuscript. This research was supported by National Science Foundation DEB-1256330 and by the David and Lucile Packard Foundation. The author declares no conflicts of interest.

## DATA ARCHIVING

All data and R code necessary to repeat the analyses described in this article are available through the Dryad data repository: [dx.doi.org/10.5061/dryad.d4841](https://doi.org/10.5061/dryad.d4841).

## LITERATURE CITED

- 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. Nat. Acad. Sci. USA* 106:13410–13414.
- Baldwin, B. G., and M. J. Sanderson. 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. USA* 95:9402–9406.
- Beaulieu, J., and M. J. Donoghue. 2013. Fruit evolution and diversification in campanulid angiosperms. *Evolution* 67:3132–3144.
- Beaulieu, J., and B. C. O'Meara. 2015a. Extinction can be estimated from moderately sized phylogenies. *Evolution* 69:1036–1043. doi:10.1111/evo.12614.
- Beaulieu, J. M., and B. C. O'Meara. 2015b. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *bioRxiv*. doi:[http://dx.doi.org/10.1101/016386](https://doi.org/10.1101/016386).
- Bonett, R. M., M. A. Steffen, S. M. Lambert, J. J. Wiens, and P. T. Chippindale. 2013. Evolution of paedomorphosis in plethodontid salamanders: ecological correlates and re-evolution of metamorphosis. *Evolution* 68:466–482.
- Colless, D. H. 1982. Phylogenetics: the theory and practice of phylogenetic systematics II (book review). *Syst. Zool.* 31:100–104.
- Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer, Cambridge.
- Cusimano, N., and S. S. Renner. 2010. Slowdowns in diversification rates from real phylogenies may not be real. *Syst. Biol.* 59:458–464.
- Davis, M. P., P. E. Midford, and W. P. Maddison. 2013. Exploring power and parameter estimation of the BiSSE method for analyzing species diversification. *BMC Evol. Biol.* 13:38.
- de Vos, J. M., C. E. Hughes, G. M. Schneeweiss, B. R. Moore, and E. Conti. 2014. Heterostyly accelerates diversification via reduced extinction in primroses. *Proc. R. Soc. Lond. B* 281:20140075.
- Drummond, C. S., R. J. Eastwood, S. T. S. Miotto, and C. E. Hughes. 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. *Syst. Biol.* 61:443–460.
- Etienne, R. S., B. Haegeman, T. Stadler, T. Aze, P. N. Pearson, A. Purvis, and A. B. Phillimore. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. Lond. B* 279:1300–1309.
- Fuchs, J., J. A. Johnson, and D. P. Mindell. 2015. Rapid diversification of falcons (Aves: Falconidae) due to expansion of open habitats in the Late Miocene. *Mol. Phy. Evol.* 82:166–182.
- Gingerich, P. D. 1983. Rates of evolution: effects of time and temporal scaling. *Science* 222:159–161.
- . 2001. Rates of evolution on the time scale of the evolutionary process. *Genetica* 112–113:127–144.
- Heard, S. B. 1992. Patterns in tree balance among cladistic, phenetic, and randomly generated phylogenetic trees. *Evolution* 46:1818–1826.
- Heath, T. A., D. J. Zwickl, J. Kim, and D. M. Hillis. 2008. Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. *Syst. Biol.* 57:160–166.
- Hey, J. 1992. Using phylogenetic trees to study speciation and extinction. *Evolution* 46:627–640.
- Ho, S. W. Y., R. Lanfear, L. Bromham, M. J. Phillips, J. Soubrier, A. G. Rodrigo, and A. Cooper. 2011. Time-dependent rates of molecular evolution. *Mol. Ecol.* 20:3087–3101.
- Horn, J. W., Z. Xi, R. Riina, J. A. Peirson, Y. Yang, Y. Dorsey, P. E. Berry, C. C. Davis, and K. J. Wurdack. 2014. Evolutionary burst in *Euphorbia* (Euphorbiaceae) are linked with photosynthetic pathway. *Evolution* 68:3485–3504.
- Hugall, A. F., and D. Stuart-Fox. 2012. Accelerated speciation in colour-polymorphic birds. *Nature* 485:631–634.
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. Mooers. 2012. The global diversity of birds in space and time. *Nature* 491:444–448.
- Kubo, T., and Y. Iwasa. 1995. Inferring the rates of branching and extinction from molecular phylogenies. *Evolution* 49:694–704.
- Kuhn, T. S., A. O. Mooers, and G. H. Thomas. 2011. A simple polytomy resolver for dated phylogenies. *Methods Ecol. Evol.* 2:427–436.
- 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.
- Magallon, S., and A. Castillo. 2009. Angiosperm diversification through time. *Am. J. Bot.* 96:349–365.
- McPeck, M. A., and J. M. Brown. 2007. Clade age and not diversification rate explains species richness among animal taxa. *Am. Nat.* 169:E97–E106.

- Mooers, A. O., and S. B. Heard. 1997. Inferring evolutionary processes from phylogenetic tree shape. *Q. Rev. Biol.* 72:31–54.
- Morlon, H., T. L. Parsons, and J. B. Plotkin. 2011. Reconciling molecular phylogenies with the fossil record. *Proc. Natl. Acad. Sci. USA* 108:16327–16332.
- Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994a. Extinction rates can be estimated from molecular phylogenies. *Proc. R. Soc. Lond. B.* 344:77–82.
- Nee, S., R. M. May, and P. H. Harvey. 1994b. The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 344:305–311.
- Ng, J., and S. D. Smith. 2014. How traits shape trees: new approaches for detecting character state-dependent lineage diversification. *J. Evol. Biol.* 27:2035–2045.
- Purvis, A. 2008. Phylogenetic approaches to the study of extinction. *Ann. Rev. Ecol. Syst.* 39:301–319.
- Quental, T. B., and C. R. Marshall. 2010. Diversity dynamics: molecular phylogenies need the fossil record. *Trends Ecol. Evol.* 25:434–441.
- . 2011. The molecular phylogenetic signature of clades in decline. *Plos One* 6:e25780.
- . 2013. How the Red Queen drives terrestrial mammals to extinction. *Science* 341:290–292.
- Rabosky, D. L. 2009. Heritability of extinction rates links diversification patterns in molecular phylogenies and fossils. *Syst. Biol.* 58:629–640.
- . 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64:1816–1824.
- . 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* 9:e89543.
- Rabosky, D. L., and D. C. Adams. 2012. Rates of morphological evolution are correlated with species richness in salamanders. *Evolution* 66:1807–1818.
- Rabosky, D. L., G. J. Slater, and M. E. Alfaro. 2012. Clade age and species richness are decoupled across the Eukaryotic tree of life. *PLoS Biol.* 10:e1001381.
- Revell, L. J., L. J. Harmon, and R. E. Glor. 2005. Underparameterized model of sequence evolution leads to bias in the estimation of diversification rates from molecular phylogenies. *Syst. Biol.* 54:973–983.
- Richardson, J. E., R. T. Pennington, T. D. Pennington, and P. M. Hollingsworth. 2001. Rapid diversification of a species-rich genus of Neotropical rain forest trees. *Science* 293:2242–2245.
- Ricklefs, R. E. 2006. Global variation in the diversification rate of passerine birds. *Ecology* 87:2468–2478.
- Rosenblum, E. R., B. A. Sarver, J. W. Brown, S. Des Roches, K. M. Hardwick, T. D. Hether, J. M. Eastman, M. W. Pennell, and L. J. Harmon. 2012. Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. *Evol. Biol.* 39:255–261.
- Sadler, P. M. 1981. Sediment accumulation rates and the completeness of stratigraphic sections. *J. Geol.* 89:569–584.
- Sanderson, M. J. 1996. How many taxa must be sampled to identify the root node of a large clade? *Syst. Biol.* 45:168–173.
- Stanley, S. M. 1979. *Macroevolution: pattern and process*. Freeman, San Francisco.
- Valente, L. M., V. Savolainen, and P. Vargas. 2010. Unparalleled rates of species diversification in Europe. *Proc. R. Soc. Lond. B* 277:1489–1496.
- Weber, M. G., and A. A. Agrawal. 2014. Defense mutualisms enhance plant diversification. *Proc. Natl. Acad. Sci. USA* 111:17442–16447.
- Zanne, A., D. C. Tank, W. K. Cornwell, J. M. Eastman, S. A. Smith, R. G. FitzJohn, D. J. McGlinn, B. C. O'Meara, A. T. Moles, P. B. Reich, et al. 2014. Three keys to the radiation of angiosperms into freezing environments. *Nature* 506:89–92.

Associate Editor: M. Friedman  
Handling Editor: M. Servedio