

History is written by the victors: The effect of the push of the past on the fossil record

Graham E. Budd^{1,2}  and Richard P. Mann³ 

¹Department of Earth Sciences, Palaeobiology, Uppsala University, Villavägen 16, Uppsala SE 752 36, Sweden

²E-mail: graham.budd@pal.uu.se

³Department of Statistics, School of Mathematics, University of Leeds, Leeds LS2 9JT, United Kingdom

Received September 22, 2017

Accepted September 1, 2018

Survivorship biases can generate remarkable apparent rate heterogeneities through time in otherwise homogeneous birth-death models of phylogenies. They are a potential explanation for many striking patterns seen in the fossil record and molecular phylogenies. One such bias is the “push of the past”: clades that survived a substantial length of time are likely to have experienced a high rate of early diversification. This creates the illusion of a secular rate slow-down through time that is, rather, a reversion to the mean. An extra effect increasing early rates of lineage generation is also seen in large clades. These biases are important but relatively neglected influences on many aspects of diversification patterns in the fossil record and elsewhere, such as diversification spikes after mass extinctions and at the origins of clades; they also influence rates of fossilization, changes in rates of phenotypic evolution and even molecular clocks. These inevitable features of surviving and/or large clades should thus not be generalized to the diversification process as a whole without additional study of small and extinct clades, and raise questions about many of the traditional explanations of the patterns seen in the fossil record.

KEY WORDS: Crown groups, diversification rates, mass extinctions, molecular clocks, push of the past, survivorship bias.

The patterns of diversity through time have been of continuous interest ever since they were broadly recognized in the 19th century (e.g., Phillips 1840). In particular, both major radiations (such as the origin of animals (Budd and Jensen 2000) or angiosperms (Sanderson and Donoghue 1994)) and the great mass extinctions (e.g., the end-Permian (Erwin 1993) or end-Cretaceous (Friedman 2010; Hull et al. 2011)) have attracted much attention, with an emphasis on trying to understand the causal mechanisms behind these very striking patterns. For example, the “Cambrian Explosion” and “Great Ordovician Biodiversification Event” have both been discussed at great length, with mechanisms as diverse as the cooling of the Earth to bombardment with cosmic rays or secular changes in developmental mechanisms being suggested (Smith and Harper 2013). However, in the midst of this search, the effects of survival biases on creating the patterns under consideration have hardly been considered. The last decades have also seen a great deal of interest and work on mathematical approaches to diversification and extinction (e.g., Stadler et al. 2014; Wang et al.

2013; Lieberman 2001; Ezard et al. 2012; Etienne et al. 2012), including some that touch on the topics considered by this article (e.g., see especially Mooers et al. 2011; Stadler and Steel 2012; Ricklefs 2007; and Stadler 2013), but there is hardly any literature on the dynamics of clade origins from the perspective of survival biases and their effect on the fossil record. In this article, then, we wish to explore the basis for such biases and then consider how it is exported to various important aspects of the observed large-scale patterns of evolution, with particular (but not exclusive) focus on the sort of data that can be extracted from the fossil record. In the following analyses, we calculate diversification over an interval of 0.1 Myrs and plot graphs with an interval of 2 Myrs. We term the (often unvarying) rate of a particular type of event (extinction or speciation) in a model as the “background” rate; and the rate of such events measured during a particular interval of time as the “observed” rate. For example, the background rate of rolling a six using an unbiased 6-sided die is 0.17; but if in seven trials five sixes happened to be rolled, the observed rate would be 0.71.

Our mathematical models are implemented in R (R Core Team 2017) and the code for this article is available in the Supporting Information.

The “push of the past”

Nee and colleagues (Nee et al. 1994b; Harvey et al. 1994) summarized the general mathematics of stochastic birth-death models as applied to phylogenetic diversification (see also especially Ricklefs 2007). In such models, each lineage has a certain chance of either disappearing (“death,” the rate of which is usually labeled “ μ ”) or splitting into two (“birth,” rate “ λ ”). Many models that consider diversity in this way have used constant birth and death rates that have revealed much of interest about phylogenetic processes (Nee 2001; Nee 2006). As Nee et al. (1994b) pointed out, conditioning clades on survival to the present generates two biases in the rate of diversification through time: the “push of the past” (POTPa) and the “pull of the present” (POTPr).

The POTPa emerges as a feature of diversification by the fact that all modern clades (tautologically) survived until the present day. This singles them out from the total population of clades that could be generated from any particular pair of background birth and death rates: clades that happened *by chance* to start off with higher net rates of diversification have better-than-average chances of surviving until the present day. As Nee et al. (1994a) put it, such clades “got off to a flying start”; and they accumulate species faster than one would expect. Once clades become established, they are less vulnerable to random changes in the observed diversification rate, and this value therefore tends to revert to the background rate through time. Long-lived clades thus tend to show high observed rates of diversification at their origin, which then decrease to their long-term average as the present is approached. It is important to note that such an effect is only seen in the rates of appearance of total species through time, not in the rate of appearance of lineages. A similar effect should apply to now *extinct* clades (such as trilobites) that nevertheless survived a substantial length of time. This effect is analogous to the “weak anthropic principle,” which contends that only a certain subset of possible universes, that is those with particular initial conditions, could generate universes in which humans could evolve in order to experience them. Similarly, to ask the question why living clades appear to originate with bursts of diversification that then moderate through time is to miss the point that this pattern is a *necessary* condition for (most) clades to survive until the present day.

The “pull of the present” (POTPr) is, conversely, an effect seen in the number of *lineages* through time that will eventually give rise to living species, which is effectively what is being reconstructed with molecular phylogenies. As the present is approached, the number of lineages leading to recent diversity

should increase faster than the background rate of diversification because less time is available for any particular lineage to go extinct. Thus, the POTPa affects reconstructions of *diversity* through time; and the POTPr affects the number of *ancestral lineages* through time (Fig. 1A).

Despite the theoretical considerations above, molecular phylogenies, far from showing a pronounced POTPr, typically show either no change in observed diversification rate as the present is approached, or even show a marked *slow-down* in rate—the opposite effect to what might be expected. Why this might be the case has been the subject of intensive research over the last few years, with various models being proposed, the most important of which are the “protracted speciation” model of Etienne et al. (2012), and various proposals that carrying capacities of environments lead to “diversity-dependent diversification” (DDD) (see Rabosky 2013 for review; see also Cusimano and Renner 2010 for skepticism about the reality of the effect). Diversification patterns thus potentially show two sorts of slow-down: one after the initial burst of diversification; and one as the present day is reached. This article deals with the first of these effects. When clades had themselves a recent origin, these two effects can of course be confounded, so in our discussion we largely confine ourselves to old clades whose “beginning” and “end” are clearly separated.

Although the POTPa is a known effect (although sometimes confused with the “Large Clade Effect” that we discuss below (e.g., Wahlberg et al. 2009; cf Ricklefs 2007)), and emerges naturally from conditioning clade survival to the present (see e.g., Stadler 2010; Stadler et al. 2015), it has had no penetration into the paleontological literature, where it is most important. In this article, therefore, we wish to show how influential an effect it and related effects are, and discuss its implications for general discussions about the reasons behind typical patterns of diversification seen in the fossil record and molecular phylogenies, including guidelines for how it and related effects might be detected.

Mathematical Analysis

In this section, we extend on the approach of Nee et al. (1994b) by explicitly conditioning the birth-death model on the number of extant species in the crown group, and considering the full distribution of clade abundances over time rather than just the central expectation.

ESTIMATING THE NUMBER OF EXTANT AND SURVIVING CLADES THROUGH TIME

As noted by previous studies (Nee et al. 1994b; Strathmann and Slatkin 1983), in the classic birth-death model with speciation rate per species λ and extinction rate per species μ , the number of extant species, n_t in a surviving clade at a given time t from the

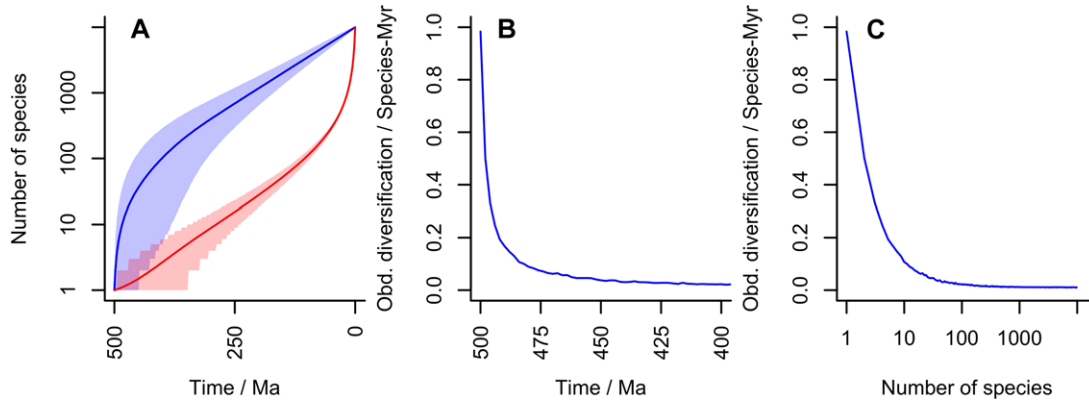


Figure 1. An example diversification with 10,000 living species, an extinction rate of 0.5 per species per million years and a diversification time of 500 Myrs. The implied speciation rate is 0.5107 per species per million years and thus the underlying diversification rate is 0.0107 per species per million years. (A) Diversity plot through time. As in all other figures, the blue line is the number of species at time t and the red line the number of species that will give rise to living species. Shading gives 95% confidence areas. Note large POTPa and POTPr. (B) Observed diversification rate at beginning of diversification (note scale of 100 Myrs). (C) Implied diversification rate correlation with diversity generated by this distribution.

origin at time zero, obeys the following zero-truncated geometric probability distribution

$$P(n_t) = G(n_t - 1; 1 - a_t) \equiv (1 - a_t)a_t^{n_t-1} \quad (1)$$

with

$$a_t = \frac{\lambda(1 - \exp(-(\lambda - \mu)t))}{\lambda - \mu \exp(-(\lambda - \mu)t)} \quad (2)$$

It is also useful to introduce the survival probability, $s_{\Delta t}$: the probability that a lineage with one originating species will survive for a duration of time Δt ,

$$s_{\Delta t} = \frac{\lambda - \mu}{\lambda - \mu \exp(-(\lambda - \mu)\Delta t)} \quad (3)$$

For the limiting case where $\lambda - \mu \rightarrow 0$, see Strathmann and Slatkin (1983). Nee et al. (1994b) proceeded by conditioning the distribution of n_t on the tree surviving until some future time, T . Here, we also condition on there being n_T extant species at time T . By Bayes' rule:

$$P(n_t | n_T) = \frac{P(n_T | n_t)P(n_t)}{P(n_T)} \quad (4)$$

Two terms in this equation are given immediately from equation (1). We can evaluate the remaining term, $P(n_T | n_t)$, by recognizing n_T as a sum of m_t i.i.d. geometric random variables obeying equation (1) over a time period $T - t$, where m_t is the (unknown) number of species at time t that will give rise to surviving lineages. This implies that $P(n_T | m_t)$ follows a truncated negative binomial distribution (with n_T taking a minimum value of m_t):

$$P(n_T | m_t) = \text{NB}(n_T - m_t; m_t; 1 - a_{T-t})$$

$$\equiv \binom{n_T - 1}{n_T - m_t} (1 - a_{T-t})^{m_t} a_{T-t}^{n_T - m_t} \quad (5)$$

The number of lineages that survive depends on the probability, s_{T-t} for a lineage to survive from time t to time T , and follows a binomial distribution:

$$p(m_t | n_t) = \text{B}(m_t; n_t, s_{T-t}) \equiv \binom{n_t}{m_t} s_{T-t}^{m_t} (1 - s_{T-t})^{n_t - m_t} \quad (6)$$

Combining the previous equations, and summing over the unknown value of m_t , we are left with the following expression for the number of living species at time t , conditioned on the number of species in the present:

$$P(n_t | n_T) = \frac{G(n_t - 1; 1 - a_t)}{G(n_T - 1; 1 - a_T)} \sum_{m_t=1}^{n_t} \text{B}(m_t; n_t, s_{T-t}) \times \text{NB}(n_T - m_t; m_t; 1 - a_{T-t}) \quad (7)$$

where the relevant probability mass functions are as defined above. We can further evaluate the conditional probability of m_t – the number of species at time t that will have at least one descendant at time T . Here, we integrate over possible values of n_t :

$$P(m_t | n_T) = \frac{\text{NB}(n_T - m_t; m_t; 1 - a_{T-t})}{G(n_T - 1; 1 - a_T)} \sum_{n_t=1}^{\infty} \text{B}(m_t; n_t, s_{T-t}) \times G(n_t - 1; 1 - a_t) \quad (8)$$

What sort of expectations should we have about the size and duration of the POTPa? By looking at a small initial interval of time Δt from the origin and considering both the probability that the clade has diversified to two species in this interval and the

probability that it will survive to the present, we can estimate the initial observed rate of diversification for surviving clades:

$$\begin{aligned} P(n_{\Delta t} = 2 \mid \text{survive to } T) &= \frac{P(\text{survive to } T \mid n_{\Delta t} = 2)P(n_{\Delta t} = 2)}{P(\text{survive to } T)} \\ &= \frac{1 - (1 - s_{T-\Delta t})^2}{s_T} P(n_{\Delta t} = 2) \quad (9) \\ &\simeq (2 - s_T)\lambda\Delta t \end{aligned}$$

where we have assumed that $s_{T-\Delta t} \simeq s_T$ for small Δt . It follows that the initial rate of diversification, R_0 , in the POTPa can be estimated by:

$$R_0 = (2 - s_T)\lambda. \quad (10)$$

If we look back to the origins of major clades, we expect s_T to be small for geologically significant periods of time, and thus for these examples the rate can be further approximated as,

$$R_0 \simeq 2\lambda. \quad (11)$$

(we note that similar results concerning the interior branch lengths of reconstructed trees have been derived by Stadler and Steel 2012).

It is important to note that at the precise origin of a clade that will survive to the present, the observed *extinction* rate is necessarily zero, since any extinction event would terminate the clade. Thus at this point the observed speciation and diversification rates are the same.

The result above shows us that if we know λ and μ , we can immediately calculate that the expected POTPa should produce a decline in observed diversification rate from about 2λ as an initial value down to $\lambda - \mu$ as deduced from the fossil record. However, the broad confidence intervals on this value place the 95% range on this value widely: for example, in Figure 1 over the first million years, the initial rate could be as low as 0 and as high as 3 (i.e., 6λ). Thus slowdowns seen in rates of diversification that begin with a wide range of values and quickly decline (largely being over by the time of the establishment of the crown group), with reconstructed rates in the stem lineage being significantly higher than in the generated plesions (i.e., extinct branches), are attributable to the POTPa. For the case of the birds discussed below, one would expect (in the fossil record) an observed initial diversification rate of about 1.25, that is about 20 times faster than the background rate. Another example is provided by the study of diversification rates in placental mammals (Raia et al. 2013) that simulates a best fit homogeneous model with parameters $\lambda = 0.7$, $\mu = 0.6$, and $\lambda - \mu = 0.1$. A POTPa would be consistent with a decline in rates from 1.4 to 0.1 over 5–10 million years, which closely matches their reconstruction of phenotypic rates (that they show to be correlated to diversification rates in their data). Thus, the calculated POTPa corresponds closely to real-world examples.

Even though the POTPa as an *average* effect quickly declines in time, the survivorship bias that gives rise to it must persist along the surviving lineages: every clade that will survive to the present commences with one original species, which is vulnerable to extinction. As the survival rate typically remains low until close to the present, it follows that the constant renewal of the surviving stem lineages follows a quasi-fractal pattern of repetition. A further notable feature is that as the present day is approached, and the survival probability thus tends to one, the observed rate of speciation along surviving lineages declines back toward λ .

An example plot displaying the various parameters that govern this analysis is given in Figure 1A. (cf Nee et al. 1994b). This plot is for a clade that has 10,000 living species/lineages; which emerged about 500 million years ago, and which has an average lineage duration of 2 million years (i.e., $\mu = 0.5$).

The blue line gives the number of species at any particular time; and the slope of the blue line is the observed diversification rate governed by the time elapsed and total number of taxa at the Recent, n_T . Conversely, the red line gives the number of lineages that gave rise to living species/lineages. We take as the rate of speciation the maximum likelihood estimate of λ , given μ , T and n_T , which in this case is 0.5107 (rates in this article are given to a maximum of four decimal places). Thus the rate of diversification ($\lambda - \mu$) is c. 0.0107 per species per million years. As can be seen, the slopes of the two lines diverge at the beginning, representing the push of the past, and at the end, representing the pull of the present, both of which are large in this case. If there had been a *deterministic* (i.e., nonstochastic) radiation of species from the Cambrian onwards with the net diversification rate of 0.0107 species per species per million years, then instead of 10,000 there would have been only about 210 species of this taxon today. Figure 1B shows the implied large spike in the initial observed diversification rate, owing to the POTPa, with the initial observed diversification rate (~ 1) being 100 times the underlying average. Note also that this effect generates a (noncausal) correlation between diversity and diversification rate (Fig. 1C): as diversity increases, (average) rate of diversification decreases.

How realistic are the numbers in our example? The size (and thus importance) of the POTPa depends on the rate of extinction relative to the other parameters. Extinction rates have proven difficult to estimate from both molecular phylogenies (notably Rabosky 2010; but see also Beaulieu and O'Meara 2015; and Rabosky 2016) and the fossil record (see e.g., discussions in Alroy 1999; Alroy 2014; Wagner and Lyons 2010; and Hagen et al. 2017). Nevertheless, the fossil record in particular shows that extinction rates must be relatively high, as most species across a wide range of taxa only last a few million years at most in the record (e.g., Crampton et al. 2016). For example: extinction rates over all marine invertebrates have been broadly estimated at c. 0.25 per species per million years (Barnosky et al. 2011;

Raup 1991), and for Cenozoic mammals at up to 2 per species per million years (Barnosky et al. 2011; Ceballos et al. 2015). Even highly conservative (i.e., low) estimates of background extinction rates, which partly equate species with genera in the fossil record, suggest rates in excess of 0.13 (de Vos et al. 2015; cf Harnik et al. 2012; see Alroy 2014, however, for a critique of the methodology used in the latter, which produces a notable downwards bias). An example clade would be the birds, that, although may have had their crown group origin some 120 Ma (Jetz et al. 2012; but see also Prum et al. 2015 and Ksepka et al. 2017 for a more compressed view of bird evolution), nevertheless seems to have undergone a mass extinction along with the other dinosaurs 66 Ma (Longrich et al. 2011). Such a clade thus took approximately 66–70 Myrs to radiate into 10,000 species. Assuming a (probably conservative) extinction rate of 0.5 based on other land vertebrates (Loehle and Eschenbach 2012); taking 70 Ma, this would imply a speciation rate of 0.6068 and a diversification rate of 0.1068 (cf Jetz et al. 2012). Of course, all these numbers are approximate, but our aim with them is to show that the patterns we discuss in this article arise from very typical empirical values seen in analyses of extinction and diversification. Assuming that birds are a “typically” sized clade (see below), this would imply a notably enhanced rate of diversification with initial rate of 1.214 species per species per million years that would decline over about 5–10 million years (cf Ksepka et al. 2017).

Although the parameters we have explored in this article thus seem to be typical of diversifications over a large range of species numbers and time, we wish to stress the important point that the clades that emerge from them that survive for long periods are rare. In our example, although the living clade survived for 500 Myrs, the median survival time of an average clade generated by these parameters is only 2 Myrs. Furthermore, only 2.1% of clades thus generated will survive for 500 Myrs. These numbers emphasize how unusual long surviving clades are, even when there is a net positive diversification rate: survival rates for other diversification scenarios are given in Figure 4. The POTPa generally has the paradoxical effect of making high extinction rates *increase* observed rates of diversification and numbers of living species—in the rare clades that managed to survive.

CROWN GROUP ORIGINS

We now turn our attention to estimating the origin time of extant crown groups. First we consider the definition of a “randomly selected” crown group used by Raup (1993): the group emerging from the common ancestor of two randomly selected extant species.

Assume that we have selected one species at random from n_T extant taxa. We now select a second. What is the probability, $W(t | n_T)$ that we select one that shares a common ancestor with the first at time t ? If there are m_t lineages at time t that will

survive to the present, then each of these must give rise to at least one extant species, leaving $n_T - m_t$ remaining extant species that do not inevitably have to join up with the other $m_t - 1$ ancestor species. Each species from this remaining set will have a $1/m_t$ probability of sharing an ancestor with the first selected, thus:

$$W(t | m_t, n_T) = \frac{n_T - m_t}{n_T m_t}. \tag{12}$$

To obtain the desired probability, $W(t | n_T)$ requires a posterior-weighted summation over the possible values of m_t :

$$W(t | n_T) = \sum_{m_t=1}^{n_T} W(t | m_t, n_T) P(m_t | n_T) \sum_{m_t=1}^{n_T} \frac{n_T - m_t}{n_T m_t} P(m_t | n_T), \tag{13}$$

where the posterior distribution on m_t is calculated as above. $W(t | n_T)$ represents a cumulative distribution function for the timing of crown group origins for randomly selected pairs of species, looking backwards in time. The corresponding probability density function, $w(t | n_T)$, is given by differentiation of $W(t | n_T)$:

$$w(t | n_T) = -\frac{dW}{dt}. \tag{14}$$

Compared to Raup’s model (which can be most closely approximated by the Yule process (see below), although he did not include a stochastic component) our model delays the average time of origin of Raupian crown groups because of the effect of the POTPr of allowing a longer period of lower early lineage diversification rates. Nevertheless, it remains true that randomly selected pairs of taxa will also tend to have early origins (Fig. 4F, I). As can be seen (Fig. 4C), the Yule process forces crown-groups defined in this way to emerge very early. Budd and Jackson (2016) simulated the origin of the *first* crown groups in clades conditioned on survival, a topic of much interest in “Cambrian Explosion” literature (e.g., Erwin 1993). In simulations that start with one lineage and go on to diversify to the Recent, the time the simulation begins can be taken as the origin of the total group, and the emergence of the crown group (for the entire clade) when m_t (the number of lineages at any time t that will give rise to living descendants) is equal to two (i.e., the basal split of the crown group is formed). Since this state can only be reached from a previous state of $m_t = 1$, the probability density $u(t)$ that the first crown emerges at time t can therefore be calculated by considering the rate of change in the probability that $m_t = 1$:

$$u(t) = -\frac{dP(m_t = 1 | n_T)}{dt}. \tag{15}$$

As summarized in Nee et al. (1994b) and Figure 1A, we can see that m_t in the early stages of diversification essentially depends on $\lambda - \mu$: $m_t \simeq \exp([\lambda - \mu]t)$. Thus a simple approximation for

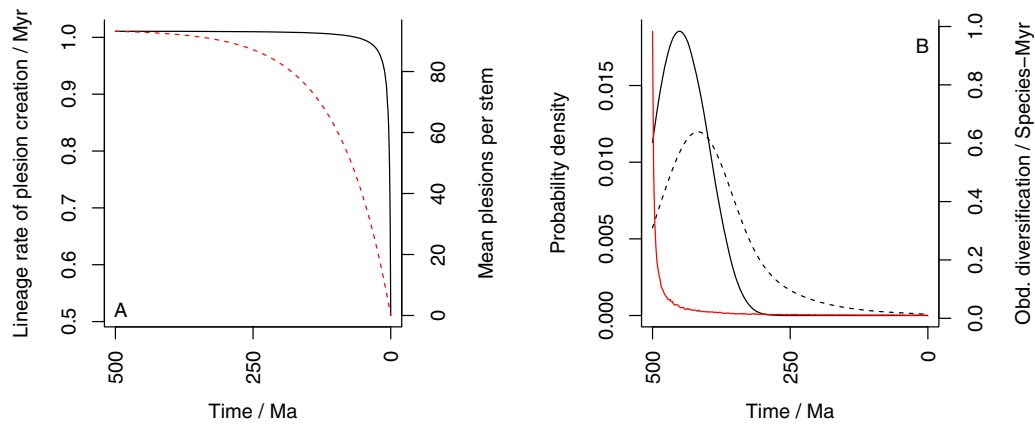


Figure 2. (A) Illustration of the rate of plesion creation along the surviving lineages (black solid line) and the mean number of plesions created along each stem group (red dashed line) through time. As rate of plesion creation is almost flat for most of the time, it follows that the decline in number of plesions per stem group depends on the stem groups decreasing in size temporally. (B) Observed diversification rate (red) and probability density functions of the first crown group (black, solid) and origin times for pairs of random living species (black, dashed) against time. All plots for a diversification over 500 Myrs in total, $n_T = 10,000$, $\lambda = 0.51$, and $\mu = 0.5$ (i.e., the same example as Fig. 1). Note the likely emergence of the first crown group as the POTPa decays.

the expected length of time it takes for the first crown-group to emerge is given by:

$$\ln(m_t) \simeq (\lambda - \mu)t. \quad (16)$$

Thus t_{cg} , the time in millions of years ago that the first crown group is expected to have emerged is simply

$$t_{cg} \simeq T - \frac{\ln 2}{\lambda - \mu}, \quad (17)$$

where T is the time elapsed since the origin of the total group. As an interesting aside, the underlying diversification rate $\lambda - \mu$ is thus approximated by:

$$\lambda - \mu \simeq \frac{\ln 2}{T - t_{cg}}. \quad (18)$$

The combination of the POTPa and the dependence of t_{cg} on $\lambda - \mu$ means that stem and crown groups exhibit different characteristics of diversification and diversity, as the first crown group tends to emerge as the effect of the POTPa fades away. An example of this is given in Figure 2B.

The Push of the Past and the Fossil Record

OVERVIEW

Within a particular total group, then, stem groups are characterized by high observed diversification rates and low diversity; and crown groups by low diversification rates and increasing diversity (cf Fig. 1C). The interaction between the crown group and the POTPa allows us to understand why it is that the crown group emerges just as the POTPa dies away: the POTPa is an effect seen

when there are few (surviving) lineages and as soon as there are two rather than one, the likelihood of the clade surviving until the present is considerably increased.

It is important to note that the high rates of observed diversification in stem groups are not *general* features, as we are applying a homogeneous model of diversification. Rather, unusually high observed diversification rates are concentrated in the stem lineage that leads to the crown group(s) (cf Stadler and Steel 2012). Stem groups should thus generate a high number of so-called “plesions” (i.e., extinct sister groups to crown groups (Budd 2001; Craske and Jefferies 1989)) that themselves will diversify and go extinct at approximately the background rate governed by $\lambda - \mu$. From equation (10) we can see that the rate of speciation along most of the stem lineages, and thus the rate of production of plesions, remains close to 2λ , although the rate slowly declines until close to the Recent, when it precipitously drops to λ . Similarly, lengths of stem-groups also decrease, over a longer timescale, as the present is reached (see Fig. 2A for graphical treatment). Such effects should be observable and inferable from the fossil record by plotting speciation and extinction rates using appropriate methods for extracting these values (e.g., Alroy 2000; Alroy 2014).

This analysis gives us a remarkable perspective on the fossil record (Fig. 3), which is after all considered on methodological grounds, as taxa in cladograms are only ever terminals, to be composed entirely of plesions (Budd 2003; but see also Gavryushkina et al. 2014) except for fossils of extant species. Average rates of speciation (and, as we shall argue below, rates of phenotypic evolution) typify the clouds of plesions that are constantly being generated (and dissipating) at a high rate; but underlying them, and hidden from view, are stem lineages that speciate at twice the normal rate. It is only briefly, at the beginnings of radiations and

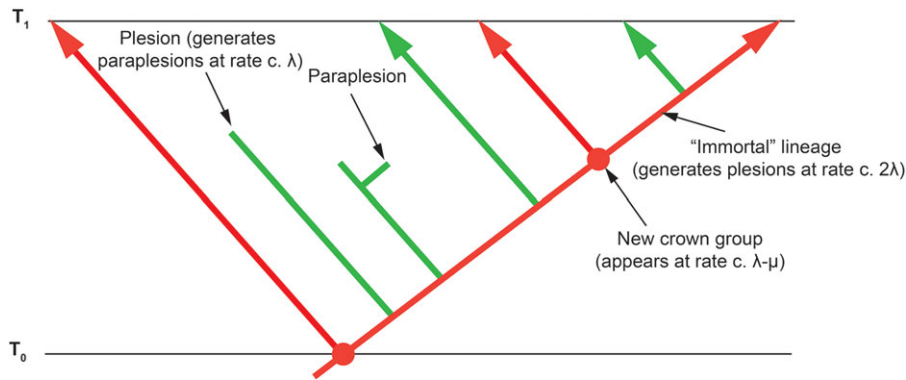


Figure 3. A small section of a tree at a time distant from the present. Red branches represent lineages that survived until the present, and where they diverge represents the birth of a new crown group. Green branches represent plesions that do not survive until the present. As per the results presented herein and in Stadler and Steel (2012), the stem lineages species generate plesions at a rate close to 2λ and the plesions themselves speciate at rate λ : the crown groups form at rate $\lambda - \mu$.

after the great mass extinctions (see below) that these obscuring clouds are stripped away, and we get to peer at the underlying hyperactive stem lineages. Once again though it must be stressed that this pattern only emerges as a result of our perspective in the Recent, which allows us to distinguish stem lineages from plesions.

DIVERSIFICATION SCENARIOS

Armed with the mathematical analysis and example above, we are now in a position to analyze various scenarios that might play out in patterns of diversity and the fossil record. In each, we wish to examine: (i) the size of the POTPa effect; (ii) the distribution of the timing of crown group origins; and (iii) the relative proportions that the stem and crown groups take up of the total group.

THE YULE PROCESS

The Yule process (Yule 1924) governs diversification processes with no extinction, that is that $\mu = 0$. Of course this is not realistic over geologically significant time periods, but nevertheless is important to show the contrast between this and more realistic models. Furthermore, it can be used to model surviving lineages through time, that have no extinction.

Under the no-extinction model, as all species give rise to living lineages, it is clear that the blue and red lines of Figure 1 are coincident (Fig. 4A) irrespective of the error on each. There is neither a pull of the present nor a push of the past (Fig. 4B), and the slope of the line simply gives the diversification rate through the time required to lead to the observed n_T . The rate of diversification is completely constant along the mean, since the diversification rate has been selected to generate n_T (i.e., 1000 species in this case). Nonetheless, as the confidence region shows, early fluctuations in this process are possible, which we consider further later. Another feature of the no-extinction model is that

total and crown groups are nearly coincident for any particular clade, as stem-groups grow by extinction (Budd 2003). A lag at the beginning is possible though, before the first speciation event takes place.

MODELS WITH NET DIVERSIFICATION AND EXTINCTION

Figure 4D–F and G–I model two net diversification models; one with $\mu = 0.1$ and the other with $\mu = 0.5$, both with the best-fit implied λ (the maximum-likelihood value given T , n_T and the selected value of μ). As can be seen, increasing μ increases the POTPa.

If μ is set *very* low (e.g., $\mu = 0.01$ for $T = 500$ Myrs and $n_T = 1000$), then the POTPa can be much reduced. However, such models imply very implausible species longevity (a typical species would be expected to survive 100 Myrs in this model, numbers that are not realistic for the Phanerozoic. They may, however, be more appropriate to the Proterozoic (Butterfield 2007)).

Models with the same background diversification rate can have high and low turnover (e.g., $\lambda = 0.6$, $\mu = 0.5$, or $\lambda = 0.2$, $\mu = 0.1$). As models with larger diversity fluctuations in will be more vulnerable to extinction than ones with small ones, it follows that high turnover scenarios require a larger POTPa to escape the early period of vulnerability.

Mass Extinctions

We have chosen for simplicity a diversification model that is diversity-independent and has homogeneous rates of extinction both through time and for taxon-age (cf Van Valen 1973). Nevertheless, the handful of mass extinctions through time have had a large impact on diversification patterns (Benton and Emerson 2007). The most important are perhaps the end-Permian (with

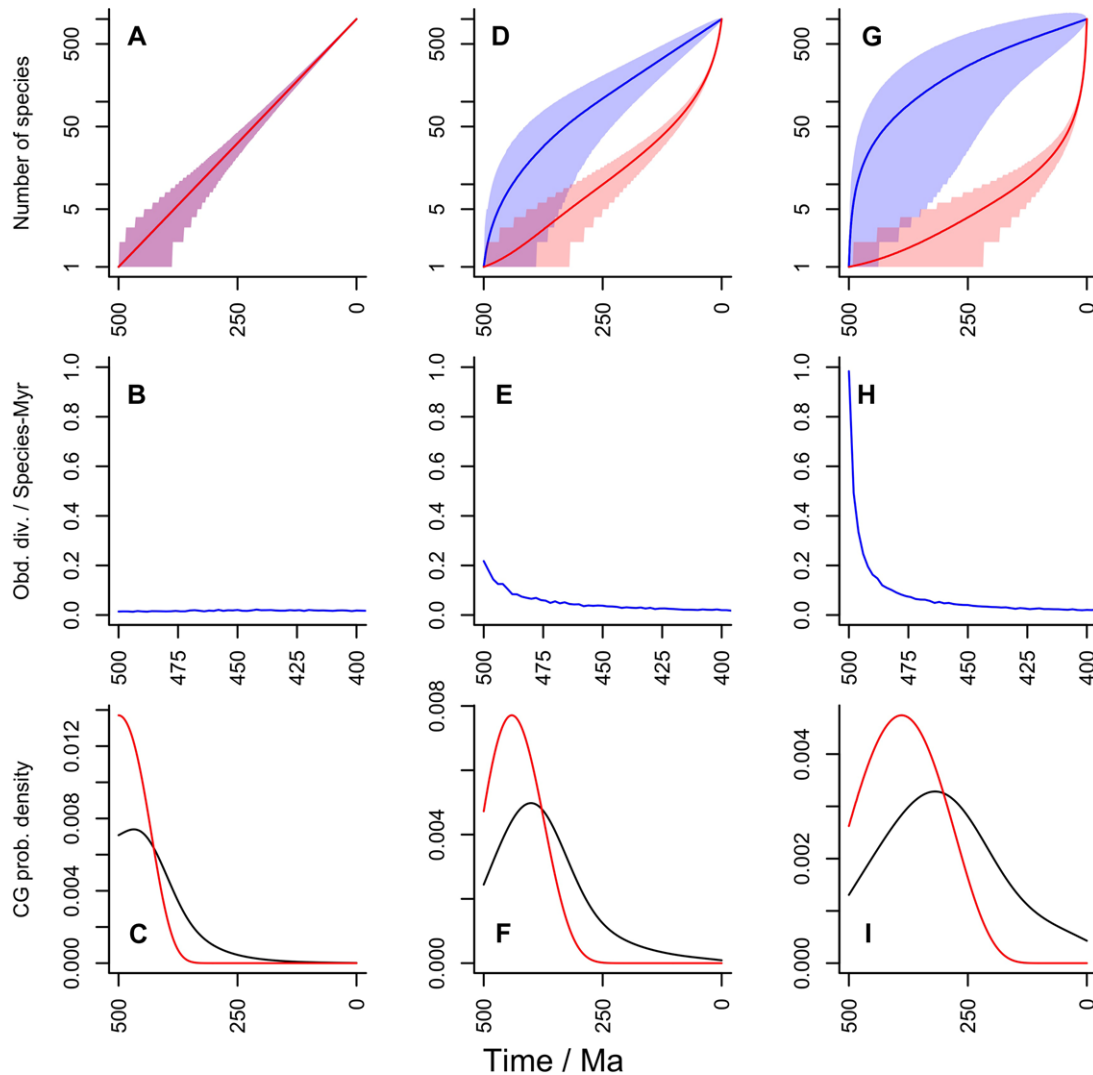


Figure 4. Patterns of diversification and stem- and crown-group formation for different (constant) diversification parameters. For each column $T = 500$ Myrs and $n_T = 1000$. All rates given per species per million years. Shading gives 95% confidence areas. Row one gives plots of diversity and diversity that gives rise to extant species through time; row two gives observed average diversification rates through time over the first 100 Myrs (i.e., the POTPa effect); row three gives probability density function plots for the appearance of the first crown group (red) and for crown groups defined by random pairs of living species. (A–C) Yule process with $\mu = 0$ and $\lambda = 0.014$. (D–F) low μ net diversification with $\mu = 0.1$ and $\lambda = 0.109$. (G–I) high μ net diversification with $\mu = 0.5$ and $\lambda = 0.505$. For the second column, median clade survival time is 10.5 Myrs and 8.2% of clades would survive 500 Myrs; for column 3, the corresponding numbers are 2 Myrs and 1%. Note different time scale on second row.

c. 80% of all species going extinct (Stanley 2016)) and the end-Cretaceous (c. 68% loss (Stanley 2016)). Such events could be considered as simply “resetting the clock”—that is if evidence exists that extinction was extremely severe in a particular clade, then T should be considered to restart at that point. Some overall patterns of diversification suggest that the only truly important mass extinction in this regard is the end-Permian one (Sepkoski 1998; Aberhan and Kiessling 2012), which divides Phanerozoic time more or less into two, with large, but largely uncommented, POTPa effects at the beginning of each. One interesting effect is that the bigger a mass extinction, the bigger the subsequent POTPa

would be, assuming *something* survives to the present. Even so, these big pushes can never make up for the lost diversity, even if they compensate for it to some extent. For example, for a diversification that started 500 Ma, and that would have generated 1000 living species without any disturbance, and with a background extinction rate of 0.5 (and implied maximum-likelihood speciation rate of 0.504), a mass extinction 250 Ma down to only one species and the subsequent POTPa and reradiation would only generate 240 living species—it is a rerun of the original radiation but in half the time. On the other hand, without any POTPa, this reradiation would be expected to generate only three living species.

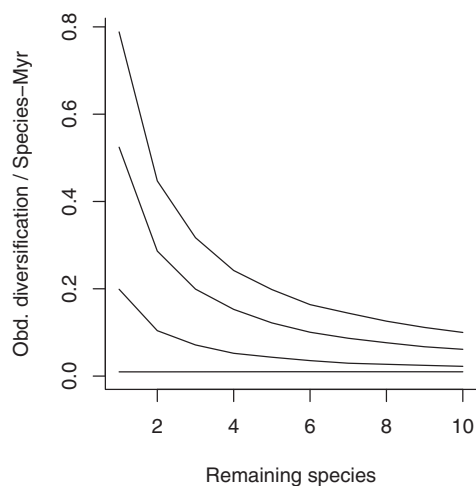


Figure 5. The impact of number of remaining species on a post-extinction POTPa on a clade that had diversified for 250 Myrs to generate 1000 species, assuming clade survival to the present, with baseline diversification $\lambda - \mu = 0.01$. The curves from bottom to top represent background extinction rates of $\mu = 0$ (Yule process), 0.1, 0.3, and 0.5 per species per million years.

It is possible to model the POTPa with a standing diversity, and show how the size of the POTPa declines as surviving diversity increases. We modelled this by plotting number of survivors against immediate observed diversification rate postextinction (Fig. 5) for different rates of background extinction for a radiation that took 250 Myrs to generate 1000 species. As can be seen, extinctions can indeed generate a large POTPa, but the number of remaining species for the clade needs to be reduced to a few percent of their original numbers. Thus, really large POTPa effects after a mass extinctions are likely to be contingent on large extinctions preceding them in the clade in question. It should be noted that individual clades can be reduced to just a few surviving species, even when overall extinction rates during a mass extinction are relatively moderate (such “dead clades walking” however, are likely to be common: see for example comments in Jablonski 2002).

Another POTPa-like bias may also be having an effect, which is the effect of the POTPa on fossilization rates themselves. One of the controls on the preservation probability of a taxon is its true (as opposed to fossil record) temporal duration, and thus its extinction rate (Foote 1997; Foote and Raup 1996). When diversity drops to a low level, survivorship over the next short interval of time is compromised, with the implication that only taxa that experience unusually high rates of diversification are likely to survive—and thus enter the fossil record. Figure 6 shows there is a strong relationship between survivorship on a million or submillion year scale and diversification rates. In brief: taxa straight after a mass extinction or at the beginning of a radiation have an unusually poor chance of entering the fossil record, as their diversity is

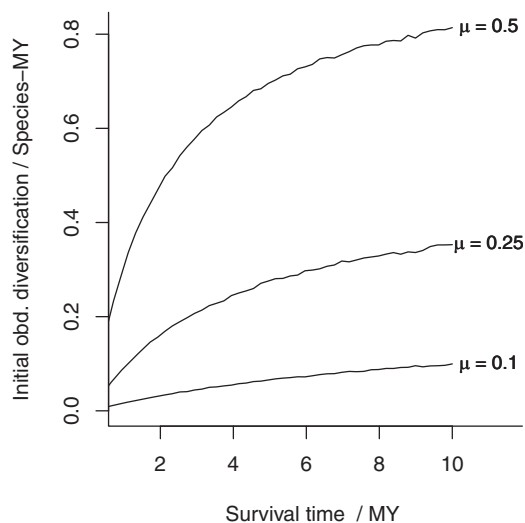


Figure 6. Expected observed initial diversification rate as a function of clade survival time for different values of underlying extinction rate, μ in a neutral model. As μ increases, a bigger and bigger POTPa is required to ensure the clade survives the first few million years.

so low and their chance of almost instant extinction is so high. However, the taxa that by chance experience high rates of early diversification are much more likely to survive long enough to generate a discoverable fossil record. Such an effect may at least partly lie behind the observation that fossilization rates seem to be depressed after mass extinctions (notably the end-Permian (Twitchett 2001)). Thus, one interesting aspect to this pattern in the record, that such “recoveries” seem to be delayed, with clades sometimes taking millions of years to show increased rates of diversification (see e.g., discussion in Sepkoski 1998), may be partly explicable by this effect too: early survivors are simply such low diversity that they tend to go extinct faster than they can enter the fossil record.

THE “COPERNICAN” NATURE OF HOMOGENEOUS BIRTH-DEATH MODELS

The various cases we have considered above show that the POTPa is in general a very important factor that cannot be neglected in trying to understand diversity patterns of the past. The most important control on the size of the POTPa is the extinction rate (compare Figs. 4E and H) although time to the Recent also has some effect. Thus, when significant time periods have passed, the POTPa is always large unless the background extinction rate is extremely low (cf Ricklefs 2007)—much lower than seems to be typical for at least Phanerozoic taxa, which typically have a life time of a few million years (Sepkoski 1998).

Because of the nature of the homogeneous model we are using, we wish to stress its “Copernican” aspect, that is that diversification is on average the same at all times. Each stem lineage

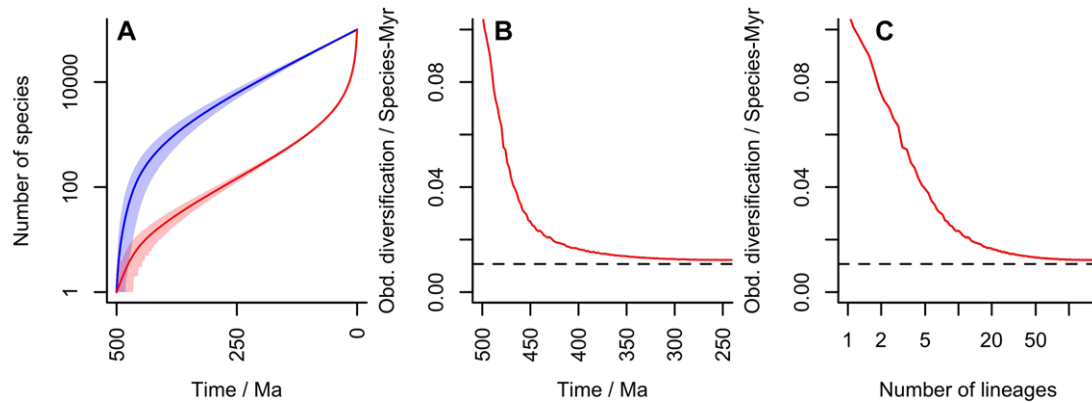


Figure 7. (A) Diversification when an exceptionally large clade is generated with the parameters of Figure 1 (here, 10 x larger than expected). An early lineage effect is introduced as this is where fluctuation in actual rates is most likely. (B) Calculated decline of lineage rate (thus in red) through time. Note that it lasts considerably longer than the POTPa although is of smaller effect (here initially c. 10 x the background rate of $\lambda - \mu$). (C) Implied lineage diversification rate correlation with lineage diversity generated by this distribution. The dashed line in (B) and (C) indicates the expected background rate (0.01) in each case.

will be characterized by a high POTPa, but as diversification continues, its distorting effect on *average* diversification rates in surviving lineages will be diminished by two factors. The first (which is small until the POTPr is reached) is that as time advances, each lineage has less time to survive until the present. The second is that as diversification proceeds, more average or even below average diversification-rate lineages will be present, and thus the overall average rate of diversification will be swamped by their diversification rates. In the first stem group, so few lineages are present that the implied POTPa on the stem lineage will have a disproportionate effect on average diversification rates. Such controls produce the characteristic decline in average actual diversification rates through time, even though an observer at any particular time would not notice any difference whatsoever.

The Large Clade Effect: An Analogy to the POTPa in Reconstructed Phylogenies

So far, we have considered the effect of survivorship biases in the blue line of Figure 1. Our exploration of the POTPa shows, however, that when conditioned on survival, it remains nearly constant at along the *surviving* lineages at close to 2λ until the Recent is approached. Hence, it largely cannot account for long-term declines in phenotypic and molecular rates along the lineages (see below) or lineage production rates themselves. However, survival alone is not the only characteristic of a clade that can lead to statistical biases. As we have shown, the birth-death model can also be conditioned on the number of extant species in the Recent, n_T . Therefore we can ask about the characteristics of outliers within the set of surviving clades, specifically those with a larger-than-

expected present diversity (Ricklefs 2007). Such outliers represent those clades that are held up as the most “successful” examples of their type and, erroneously as we shall see, are often presented as “representative” of their particular time of origination. As Pennell et al. (2012) observed through simulation analysis, larger than average clades are statistically more likely to show a slowdown relative to smaller clades. To illustrate this, we recalculated the example shown in Figure 1, but conditioned it to generate 100,000 instead of 10,000 species (Fig. 7). Under such rare conditions, more lineages need to be generated than normal, and the most likely moment to do this (as can be seen in the confidence regions of Fig. 1A) is at the beginning, when overall numbers of lineages are small and statistical fluctuations more noticeable in effect.

Under such circumstances, a *lineage* effect is produced (Phillimore and Price 2008; Pennell et al. 2012), which could be called the large clade effect (LCE). Although it is smaller than the classical POTPa (in the example of Fig. 7 the rate of speciation along the lineages increases to 2.2λ from 2λ), it has the effect of speeding up the appearance of new living lineages near the beginning, and thus makes crown groups emerge (even) earlier. The lineage through time plot thus takes on a characteristic inverted “S” shape that is often seen in plots of molecular phylogenies (e.g., Etienne et al. 2012; Harmon et al. 2003). Like the POTPa, it has a quasi-fractal organization—within a given clade, larger subclades will experience greater early diversification than smaller subclades. As we discuss below, this effect thus influences rates of evolution in large clades, and will be particularly prominent if such clades have happened to attract more than average attention, as has indeed been suggested (Ricklefs 2007). A correlation with lineage diversity is also generated Figure 7C, which in our example can be seen for about the first 20 lineages.

The initial magnitude of the LCE can be explicitly calculated in terms of the relative magnitude of the clade relative to its expected size conditioned on the background speciation and extinction rates, $\mathbb{E}(n_T | \text{survive})$. To determine the initial rate relative to the background value $R = \lambda - \mu$, we consider the probability that the new clade with one lineage diverges into two lineages within a small unit of time, both *a priori* and conditioned on the final clade size. Recalling that the distribution of n_T conditioned on m_t is negative binomial, we have:

$$\begin{aligned} \frac{R_0}{R} &= \frac{P(m_{\Delta t} | n_T)}{P(m_{\Delta t})} \\ &= \frac{P(n_T | m_{\Delta t})}{P(n_T)} \\ &= \binom{n_T - 1}{n_T - 2} \frac{(1 - a_{T-\Delta t})^2 a_{T-\Delta t}^{n_T-2}}{(1 - a_T) a_T} \quad (19) \\ &\simeq (n_T - 1) \frac{(1 - a_T)}{a_T} \\ &= \frac{n_T - 1}{\mathbb{E}(n_T | \text{survive to } T) - 1} \simeq \frac{n_T}{\mathbb{E}(n_T | \text{survive to } T)} \end{aligned}$$

Thus we can see that the expected size of the initial LCE, and thus the magnitude of the later slowdown, is proportional to the eventual clade size. It should be noted the clade containing any randomly chosen species is expected to be twice the average clade size, and thus there is a consistent bias toward this effect appearing.

Effects of the POTPa and LCE on Rates of Phenotypic and Molecular Change

The rate of phenotypic change through time is another pattern that has seen a great deal of interest (e.g., Westoll 1949; Lloyd et al. 2012; Lee et al. 2013; Ruta et al. 2006; Brusatte et al. 2010; Bronzati et al. 2015; but see also Harmon et al. 2010). A classical pattern of rates of phenotypic change is that rates are elevated at the origin of a clade and then show an exponential decline (e.g., Erwin 1993). Such a pattern looks, of course, like a POTPa effect, but this effect would seem to rely on a correlation between rates of phenotypic change and diversification. While this seems both intuitively reasonable and has much theoretical backing, this pattern has been difficult to demonstrate and indeed some studies have failed to reveal it (e.g., Adams et al. 2009; Hopkins and Smith 2015; but see also Rabosky and Adams 2012 who review the topic in general). Our model can account for such patterns by considering the fossil record to consist of plesions that are generated by a rapid rate of speciation in an underlying but unseen stem lineage. In principle at least, each of these speciation events (at least as recognized in the fossil record) should be accompanied

by a set of diagnostic synapomorphies that accumulate within the stem lineage twice as fast as they do in the plesions that arise from it. This effect does rely on some sort of correlation between phenotypic change and diversification rates. However, in at least the fossil record, where distinct taxa are recognized solely on the phenotypic differences, some such connection must exist.

As the survivorship bias that leads to the POTPa remains more or less constant along long stretches of the lineages until the present is reached, it follows that it should not generate a “slow-down” in measured rates of either phenotypic or molecular change along the lineages (i.e., measured along the red line of Fig. 1) when the present is far away. Average phenotypic rates of change should however decline through time when measured over all fossil taxa (i.e., measuring the rate of phenotypic change in the blue, rather than red, line of Fig. 1). The notable study of lungfish evolution through time by Lloyd et al. (2012) reconstructed rates of phenotypic evolution through time and indeed noted such a decline (see their Fig. 4; note that it also shows characteristic postextinction spikes; cf Raia et al. 2013 that shows a similar decline through time). However, as the authors note, the (reconstructed) stem lineage leading up to the extant lungfish retains high rates of phenotypic change much later than the initial rapid decline in overall rates, while the plesions appear to show no such pattern (the authors do not differentiate between the two in their analysis of the decline in rates). Study of bird and dinosaur phenotypic change rates also strikingly shows a similar concentration of change along the stem leading to the extant taxa (Benson et al. 2014). This pattern is exactly what the model we develop here would predict, as it confines the POTPa to the stem lineages, and suggests that their documented decline of phenotypic rates of evolution is a striking consequence of the POTPa (compare their Fig. 1 with our Fig. 3). Clearly, it would be possible to test this pattern in other groups too. The POTPa should not, however, affect rates of molecular evolution because this cannot be measured in the blue line, only in the red.

The LCE, conversely, should affect rates measured along the lineages, but in a subtle way. For a big LCE, for example the 10x larger than expected effect of our Figure 7, the rate of speciation along the lineages is initially increased only very modestly, that is in this instance from 2λ to 2.2λ . However, the initial rate of appearance of lineages increases from $\lambda - \mu$ to $10(\lambda - \mu)$. The implication of both together is that even if rates of speciation are correlated with rates of molecular and phenotypic change, neither of the latter should noticeably increase as a result of the LCE. However, the amount of that approximately constant change in large clades that is *curated* into the present along the lineages is disproportionately sourced from the early stage of the clade’s history, when the LCE is in effect, at a rate proportional to the size of the LCE. In other words, when more early species survive to become lineages, more of the total amount of phenotypic and

molecular change that took place during that interval of time is captured by those lineages and thus survives to the present, rather than being lost to the unobservable plesions.

Thus, the initial change per unit time per lineage should be increased proportionally to the size of the LCE for both molecular and phenotypic change. This can give rise to large effects, which can be seen in the study of Lee et al. (2013), where large initial rates for both phenotypic and molecular evolution can be seen as measured along the lineages. We note that the initial effect seen in Lee et al. (2013) is approximately 10× the normal rate and lasts until about 17 lineages have been created, very similar to our calculation in Figure 7C.

One implication of this finding is that in unusually large clades, one should expect a concentration of rapid molecular evolution in early lineages, and, if not corrected for, will create the effect of making molecular clocks overestimate origination times. Such an effect could in principle account for the continuing discrepancy between molecular clock estimates for the origin of the animals and the fossil record, for example Lee et al. 2013 (but only in large clades, such as the arthropods (Lee et al. 2013)—and, of course, the animals themselves). Thus, although various studies have shown that rates of molecular change may or may not be correlated with diversification rates (e.g., Barraclough and Savolinen 2001; Bromham 2003; Davies et al. 2004; Pagel et al. 2006; Lanfear et al. 2010; Goldie et al. 2011), our model suggests that it is not diversification rates per se, but rates of lineage creation, that are correlated with (curated) amounts of molecular change.

Rate Heterogeneity

A question that arises from this analysis is: “when is it appropriate to attribute heterogeneities in frequencies of events to intrinsic survivorship biases such as the POTPa and LCE rather than to adopt models where background probabilities vary exogenously through time?” To take an example from the fossil record: some extinct taxa, notably the trilobites (e.g., Bell 2013), indeed show a very rapid initial diversification, followed by a fairly drawn-out decline and final extinction. It is clear that such a decline cannot be realistically modeled by keeping the same background rate of diversification through time—it implies that the most appropriate background diversification rate has actually turned negative (cf Stanley et al. 1981). One should note here however, that given that the trilobites experienced several mass extinctions, these singular events may have successively reduced their diversity to the point where they became vulnerable to stochastic extinction, even with net positive diversification rates.

Several recent software packages (e.g., RPANDA (Morlon 2016) and TreePar (Stadler 2011b)) have been developed for detecting statistically significant rate shifts of this sort within clades

(Stadler 2013; see also Arbour and Santana 2017; Jetz et al. 2012 for examples of examination of rate shifts in different clades). How do the effects we outline here intersect with them? We have shown above the expected sizes of both the POTPa and LCE, which are themselves rate heterogeneities that arise from homogeneous models when conditioned on either/or survival and clade size. Nevertheless, there is a difference between such heterogeneities and those seen from more complex models, because the heterogeneities that emerge from survivorship bias are strictly local as opposed to global in effect. However, it remains currently unclear how they could be disambiguated from each other. Genuine DDD should improve the survivorship of early clades and thus leave fewer plesions, but this effect will resemble the LCE (see Fig. 7 for the apparent diversity dependent diversification under such conditions).

For a particular surviving clade, inferred diversification rate variation that falls within the expectations of the POTPa outlined here should not be generalized as pointing to a time-specific period of enhanced diversification, for example after mass extinctions. Furthermore, given such patterns are inevitable, they should not be taken on their own as evidence for a particular generative mechanism, even if one thinks on other grounds that such mechanisms are probable. Conversely, evidence of diversification bursts should be taken more seriously when it occurs in short-lived clades (e.g., in the fossil record) or if it occurs across a whole clade that is already well-established.

The LCE, conversely, can by definition only be measured in the lineages. We have given an expression above for its expected size, which depends on the size of a clade relative to a base-line expectation for a given background rate of diversification. We note here, however, that there are two problems with estimating clade size. The first of these is relatively straight-forward, and relates to our inability to count all living species. This has been accounted for by for example Stadler (2011a) where it is assumed that every species could be identified with a fixed probability (for sampling of higher taxa only, see Stadler and Bochma 2012). The second issue, which is more serious, concerns the nature of the species-level birth-death model we have been using. For the past, this model is an appropriate representation of the outcomes of the evolutionary process, and abundance of species is an appropriate measure of diversity. However, as we approach the present, the validity of this model arguably breaks down, as species lose their singular identity and become more accurately represented by individuals, sub-populations and nascent new species (see e.g., Etienne and Rosindell 2012; Stadler 2013). This break-down is at least partly likely to account for the apparent lack of an observed POTPr in molecular phylogenies, which has been attributed to various types of background rate heterogeneity (Stadler 2013). Although this topic is clearly an active area of research, one approach as far as detection of the LCE is concerned would be to

compare the relative *present* sizes of otherwise equivalent clades of similar ages. The LCE would predict that their early rates of lineage diversification and the size of their subsequent slowdown would be proportional to their current clade size. The general stochasticity of the whole process can of course lead to a very wide range of possible outcomes and a suitably large sample of clades would be required to reliably detect the effect. Simulation of large numbers of clades, with a range of both POTPa and LCE, and taking into account the possible disturbing effects of mass extinctions, may assist in fully understanding the range of possible outcomes of rate heterogeneity that can arise from homogeneous models.

The intersection of background rate heterogeneity and survivorship biases raises important issues about the generalizability of theories about diversification. Clades that survive until the present day are biased by the POTPa, and of those, the large clades will be further biased by the LCE. Thus, large living clades represent a very unrepresentative sample of clades in general, and their features will not be universal to the entire population of clades that have been generated by the evolutionary process. These unusual clades can of course be modeled with specific models that describe the rate shifts that must exist in them. Such models could be used to systematically generate similar large, surviving clades, but would likely fail to generalize to the more numerous, smaller and nonsurviving clades. Thus, analyzing only large, surviving clades to the exclusion of smaller and extinct clades will not demonstrate whether the properties of the studied clades results from survivorship biases, exogenous rate variation, or both. Indeed, the very simplicity of birth-death models and their powerful application suggests that the data we have are not in general sufficient to distinguish between the various ecological and evolutionary events that eventually give rise to them. As Nee remarked some years ago: “It is well known that completely different mechanisms can generate the same pattern: the distribution of parasitic worms among people is the same as the distribution of word usage in Shakespeare – the negative binomial. This means that the patterns themselves cannot inform us about mechanism and some other techniques are needed” (Nee 2002).

Summary

In this article, we have explored the patterns of diversification that can be generated by a retrospective view of a purely homogeneous model of diversification. These patterns can be substantial and highly nonhomogeneous, and it is essential to understand these “null” hypotheses before considering causal explanations for any residuals (cf Stadler and Steel 2012). Patterns of diversification through time have been much discussed in the literature (e.g., Hopkins and Smith 2015), with a common pattern being seen that diversification rates are high at the beginning of major

evolutionary radiations, in both raw diversity counts and lineages through time plots. Various mechanisms for such effects have been proposed (such as filling empty ecological niches or unusual or flexible developmental evolution). The question that the analysis above poses is, however: are such patterns inevitably generated by the push of the past and/or the large clade effect? We have shown that the POTPa is strongest when background extinction rates are high, and that in likely scenarios for the evolution of large clades, it eventually accounts for nearly all of modern diversity. Furthermore, the POTPa impacts other many aspects of diversification dynamics, including recovery from mass extinctions. Indeed, the universality of such processes extends beyond evolutionary biology, with similar patterns being observed, for instance, in the size- or age-dependent growth of companies (see e.g. Reichstein and Dahl 2004 and references therein). Even under homogeneous models, large clades can be generated at the edge of likely distributions that possess another characteristic, the “large clade effect,” which generates distinctive patterns in phenotypic and molecular evolution.

Harvey et al. (1994), when briefly describing the POTPa, commented that “If these statistical effects are not fully appreciated, it could be tempting to misinterpret such a higher early slope as evidence for lineage birth rates being higher, and/or lineage death rates being lower, at earlier times” (p. 526). Here, we have attempted to quantify both the size of, and controls on this effect, and to show just how important it is in patterns of changes of rates of evolution through time including: dependency of rates of diversification on diversity; initial bursts of diversification at the origin of clades and the effects of mass extinctions. Although it seems natural to take the history and diversification of large and ultimately successful clades such as the arthropods as proxies for evolutionary radiations as a whole (e.g., Briggs et al. 1992; Lee et al. 2013) including after mass extinctions, our analysis shows this to be particularly fraught with difficulties: the history of life was written by the victors.

AUTHOR CONTRIBUTIONS

The study and concepts within this article were jointly developed. Both authors reviewed the text and accept responsibility for the entire article.

ACKNOWLEDGMENTS

We gratefully acknowledge comments on previous versions of this manuscript by Mike Lee, Dan Rabosky, Lindell Bromham, and her group, Rachel Warnock, Joëlle Barido-Sottani, Michael Landis, and Sebastian Willman. Comments by Doug Erwin prompted our original investigation of this subject. G.E.B. acknowledges funding from the Swedish Research Council (VR).

COMPETING INTERESTS

We have no competing interests to declare.

CODE ACCESSIBILITY

The code used in this article is available as Supporting Information.

LITERATURE CITED

- Aberhan, M., and W. Kiessling. 2012. Phanerozoic marine biodiversity: a fresh look at data, methods, patterns and processes. Pp. 3–22 in J. A. Talent, ed. *Earth and life*. Springer, Berlin.
- Adams, D. C., C. M. Berns, K. H. Kozak, and J. J. Wiens. 2009. Are rates of species diversification correlated with rates of morphological evolution? *Proc. R. Soc. B* 276:2729–2738.
- Alroy, J. 1999. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Syst. Biol.* 48:107–118.
- . 2000. Understanding the dynamics of trends within evolving lineages. *Paleobiology* 26:319–329.
- . 2014. Accurate and precise estimates of origination and extinction rates. *Paleobiology* 40:374–397.
- Arbour, J. H., and S. E. Santana. 2017. A major shift in diversification rate helps explain macroevolutionary patterns in primate species diversity. *Evolution* 71:1600–1613.
- Barnosky, A. D., N. Matzke, S. Tomiya, G. O. U. Wogan, B. Swartz, T. B. Quental, C. Marshall, J. L. McGuire, E. L. Lindsey, K. C. Maguire, et al. 2011. Has the earth's sixth mass extinction already arrived? *Nature* 471:51–57.
- Barracough, T. G., and V. Savolainen. 2001. Evolutionary rates and species diversity in flowering plants. *Evolution* 55:677–683.
- Beaulieu, J. M., and B. C. O'Meara. 2015. Extinction can be estimated from moderately sized molecular phylogenies. *Evolution* 69:1036–1043.
- Bell, M. 2013. Fossil focus: trilobites. *Palaeontol. Onl.* 3:1–9.
- Benson, R. B. J., N. E. Campione, M. T. Carrano, P. D. Mannion, C. Sullivan, P. Upchurch, and D. C. Evans. 2014. Rates of dinosaur body mass evolution indicate 170 million years of sustained ecological innovation on the avian stem lineage. *PLoS Biol.* 12:e1001853.
- Benton, M. J., and B. C. Emerson. 2007. How did life become so diverse? The dynamics of diversification according to the fossil record and molecular phylogenetics. *Palaeontology* 50:23–40.
- Briggs, D. E. G., R. A. Fortey, and M. A. Wills. 1992. Morphological disparity in the Cambrian. *Science* 256:1670–1673.
- Bromham, L. 2003. Molecular clocks and explosive radiations. *J. Mol. Evol.* 57:13–20.
- Bronzati, M., F. C. Montefeltro, and M. C. Langer. 2015. Diversification events and the effects of mass extinctions on Crocodyliformes evolutionary history. *R. Soc. Open Sci.* 2:140385.
- Brusatte, S. L., M. J. Benton, G. T. Lloyd, M. Ruta, and S. C. Wang. 2010. Macroevolutionary patterns in the evolutionary radiation of archosaurs (Tetrapoda: Diapsida). *Earth Environ. Sci. Trans. R. Soc. Edinb.* 101:367–382.
- Budd, G. E. 2001. Tardigrades as 'stem-group arthropods': the evidence from the Cambrian fauna. *Zool. Anz.* 240:265–279.
- . 2003. The Cambrian fossil record and the origin of the phyla. *Integ. Comp. Biol.* 43:157–165.
- Budd, G. E., and I. S. C. Jackson. 2016. Ecological innovations in the Cambrian and the origins of the crown group phyla. *Phil. Trans. R. Soc. B* 371:20150287.
- Budd, G. E., and S. Jensen. 2000. A critical reappraisal of the fossil record of the bilaterian phyla. *Biol. Rev.* 75:253–295.
- Butterfield, N. J. 2007. Macroevolution and macroecology through deep time. *Palaeontology* 50:41–55.
- Ceballos, G., P. R. Ehrlich, A. D. Barnosky, A. Garcia, R. M. Pringle, and T. M. Palmer. 2015. Accelerated modern human-induced species losses: entering the sixth mass extinction. *Sci. Adv.* 1:e1400253.
- Crampton, J. S., R. A. Cooper, P. M. Sadler, and M. Foote. 2016. Greenhouse-icehouse transition in the late Ordovician marks a step change in extinction regime in the marine plankton. *PNAS* 113:1498–1503.
- Craske, A. J., and R. P. S. Jefferies. 1989. A new mitrate from the Upper Ordovician of Norway, and a new approach to subdividing a plesion. *Palaeontology* 32:69–99.
- Cusimano, N., and S. S. Renner. 2010. Slowdowns in diversification rates from real phylogenies may not be real. *Syst. Biol.* 59:458–464.
- Davies, T. J., V. Savolainen, M. W. Chase, J. Moat, and T. G. Barracough. 2004. Environmental energy and evolutionary rates in flowering plants. *Proc. R. Soc. B* 271:2195–2200.
- De Vos, J. M., L. N. Joppa, J. L. Gittleman, P. R. Stephens, and S. L. Pimm. 2015. Estimating the normal background rate of species extinction. *Conserv. Biol.* 29:452–462.
- Erwin, D. H. 1993. Early introduction of major morphological innovations. *Acta Pal. Pol.* 38:281–294.
- . 1993. *The great Paleozoic crisis: life and death in the Permian*. Columbia Univ. Press, New York.
- Etienne, R. S., and J. Rosindell. 2012. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Syst. Biol.* 61:204–213.
- Etienne, R. S., B. Haegeman, T. Stadler, T. Aze, P. N. Pearson, A. Pu., and A. B. Phillimore. 2012. Diversity-dependence brings molecular phylogenies closer to agreement with the fossil record. *Proc. R. Soc. B* 279:1300–1309.
- Ezard, T. H. G., P. N. Pearson, T. Aze, and A. Purvis. 2012. The meaning of birth and death (in macroevolutionary birth–death models). *Biol. Lett.* 8:139–142.
- Foote, M. 1997. Estimating taxonomic durations and preservation probability. *Paleobiology* 23:278–300.
- Foote, M., and D. M. Raup. 1996. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiology* 22:121–140.
- Friedman, M. 2010. Explosive morphological diversification of spiny-finned teleost fishes in the aftermath of the end-Cretaceous extinction. *Proc. R. Soc. B* 277:1675–1683.
- Gavryushkina, A., D. Welch, T. Stadler, and A. J. Drummond. 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. *PLoS Comp. Biol.* 10:e1003919.
- Goldie, X., R. Lanfear, and L. Bromham. 2011. Diversification and the rate of molecular evolution: no evidence of a link in mammals. *BMC Evol. Biol.* 201111:286.
- Hagen, O., T. Andermann, T. B. Quental, A. Antonelli, and D. Silvestro. 2017. Estimating age-dependent extinction: contrasting evidence from fossils and phylogenies. *Syst. Biol.* 67:458–474.
- Harmon, L. J., J. A. Schulte, A. Larson, and J. B. Losos. 2003. Tempo and mode of evolutionary radiation in iguanian lizards. *Science* 301:961–964.
- Harmon, L. J., J. B. Losos, T. J. Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, K. H. Kozak, M. A. McPeck, F. M. Roark, and T. J. Near. 2010. Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64:2385–2396.
- Harnik, P. G., H. K. Lotze, S. C. Anderson, Z. V. Finkel, S. Finnegan, D. R. Lindberg, L. H. Liow, R. Lockwood, C. R. McClain, J. L. McGuire, et al. 2012. Extinctions in ancient and modern seas. *TREE* 27:608–617.
- Harvey, P. H., R. M. May, and S. Nee. 1994. Phylogenies without fossils. *Evolution* 48:523–529.

- Hopkins, M. J., and A. B. Smith. 2015. Dynamic evolutionary change in post-Paleozoic echinoids and the importance of scale when interpreting changes in rates of evolution. *PNAS* 112:3758–3763.
- Hull, P. M., R. D. Norris, T. J. Bralower, and J. D. Schueth. 2011. A role for chance in marine recovery from the end-Cretaceous extinction. *Nat. Geosci.* 4:856–860.
- Jablonski, D. 2002. Survival without recovery after mass extinctions. *PNAS* 99:8139–8144.
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann, and A. O. Mooers. 2012. The global diversity of birds in space and time. *Nature* 491:444–448.
- Ksepka, D. T., T. A. Stidham, and T. E. Williamson. 2017. Early Paleocene landbird supports rapid phylogenetic and morphological diversification of crown birds after the K–Pg mass extinction. *PNAS* 201700188.
- Lanfear, R., S. Y. W. Ho, D. Love, and L. Bromham. 2010. Mutation rate is linked to diversification in birds. *PNAS* 107:20423–20428.
- Lee, M. S. Y., J. Soubrier, and G. D. Edgecombe. 2013. Rates of phenotypic and genomic evolution during the Cambrian explosion. *Curr. Biol.* 23:1889–1895.
- Lieberman, B. S. 2001. Analyzing speciation rates in macroevolutionary studies. Pp 323–339 in J. M. Adrain, G. D. Edgecombe, and B. S. Lieberman, eds. *Fossils, phylogeny, and form: An analytical approach*. Springer, Berlin.
- Lloyd, G. T., S. C. Wang, and S. L. Brusatte. 2012. Identifying heterogeneity in rates of morphological evolution: discrete character change in the evolution of lungfish (Sarcopterygii; Dipnoi). *Evolution* 66:330–348.
- Loehle, C., and W. Eschenbach. 2012. Historical bird and terrestrial mammal extinction rates and causes. *Divers. Distrib.* 18:84–91.
- Longrich, N. R., T. Tokaryk, and D. J. Field. 2011. Mass extinction of birds at the Cretaceous–Paleogene (K–Pg) boundary. *PNAS* 108:15253–15257.
- Mooers, A., O. Gascuel, T. Stadler, H. Li, and M. Steel. 2011. Branch lengths on birth–death trees and the expected loss of phylogenetic diversity. *Syst. Biol.* 61:195–203.
- Morlon, H., E. Lewitus, F. L. Condamine, M. Manceau, J. Clavel, and J. Drury. 2016. Rpanda: an R package for macroevolutionary analyses on phylogenetic trees. *Meth. Ecol. Evol.* 7:589–597.
- Nee, S. 2001. Inferring speciation rates from phylogenies. *Evolution* 55:661–668.
- . 2002. Biodiversity: thinking big in ecology. *Nature* 417:229–230.
- . 2006. Birth–death models in macroevolution. *Ann. Rev. Ecol. Evol. Syst.* 37:1–17.
- Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994a. Extinction rates can be estimated from molecular phylogenies. *Phil. Trans. R Soc. B* 344:77–82.
- Nee, S., R. M. May, and P. H. Harvey. 1994b. The reconstructed evolutionary process. *Phil. Trans. R Soc. B* 344:305–311.
- Pagel, M., C. Venditti, and A. Meade. 2006. Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* 314:119–121.
- Pennell, M. W., B. A. J. Sarver, and L. J. Harmon. 2012. Trees of unusual size: biased inference of early bursts from large molecular phylogenies. *PLoS One* 7:e43348.
- Phillimore, A. B., and T. D. Price. 2008. Density-dependent cladogenesis in birds. *PLoS Biol.* 6:e71.
- Phillips, J. 1840. Palaeozoic series. *Penny Cyclopaedia* 17:153–154.
- Prum, R. O., J. S. Berv, A. Dornburg, D. J. Field, J. P. Townsend, E. M. Lemmon, and A. R. Lemmon. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526:569–573.
- R Core Team (2017). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Rabosky, D. L. 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64:1816–1824.
- . 2013. Diversity-dependence, ecological speciation, and the role of competition in macroevolution. *Ann. Rev. Ecol. Evol. Syst.* 44:481–502.
- . 2016. Challenges in the estimation of extinction from molecular phylogenies: a response to Beaulieu and O’Meara. *Evolution* 70:218–228.
- Rabosky, D. L., and D. C. Adams. 2012. Rates of morphological evolution are correlated with species richness in salamanders. *Evolution* 66:1807–1818.
- Raia, P., F. Carotenuto, F. Passaro, P. Piras, D. Fulgione, L. Werdelin, J. Saarenen, and M. Fortelius. 2013. Rapid action in the Palaeogene, the relationship between phenotypic and taxonomic diversification in Cenozoic mammals. *Proc. R Soc. B* 280:20122244.
- Raup, D. M. 1983. On the early origins of major biological groups. *Paleobiology* 9:107–115.
- . 1991. A kill curve for Phanerozoic marine species. *Paleobiology* 17:37–48.
- Reichstein, T., and M. S. Dahl. 2004. Are firm growth rates random? Analysing patterns and dependencies. *Int. Rev. App. Econ.* 18:225–246.
- Ricklefs, R. E. 2007. Estimating diversification rates from phylogenetic information. *TREE* 22:601–610.
- Ruta, M., P. J. Wagner, and M. I. Coates. 2006. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. *Proc. R Soc. B* 273:2107–2111.
- Sanderson, M. J., and M. J. Donoghue. 1994. Shifts in diversification rate with the origin of angiosperms. *Science* 264:1590–1590.
- Sepkoski, J. J. 1998. Rates of speciation in the fossil record. 1998. *Phil. Trans. R Soc. B.* 353:315–326.
- Smith, M. P., and D. A. T. Harper. 2013. Causes of the Cambrian explosion. *Science* 341:1355–1356.
- Stadler, T. 2010. Sampling-through-time in birth–death trees. *J. Theor. Biol.* 267:396–404.
- . 2011a. Simulating trees with a fixed number of extant species. *Syst. Biol.* 60:676–684.
- . 2011b. TreePar in R—estimating diversification rates in phylogenies. URL <http://cran.r-project.org/web/packages/TreePar/index.html> [accessed November 4, 2014].
- . 2013. Recovering speciation and extinction dynamics based on phylogenies. *J. Evol. Biol.* 26:1203–1219.
- Stadler, T., and F. Bokma. 2012. Estimating speciation and extinction rates for phylogenies of higher taxa. *Syst. Biol.* 62:220–230.
- Stadler, T., and M. Steel. 2012. Distribution of branch lengths and phylogenetic diversity under homogeneous speciation models. *J. Theor. Biol.* 297:33–40.
- Stadler, T., D. L. Rabosky, R. E. Ricklefs, and F. Bokma. 2014. On age and species richness of higher taxa. *Am. Nat.* 184:447–455.
- Stadler, T., T. G. Vaughan, A. Gavryushkin, S. Guindon, D. Kühnert, G. E. Leventhal, and A. J. Drummond. 2015. How well can the exponential-growth coalescent approximate constant-rate birth–death population dynamics? *Proc. R Soc. B* 282:20150420.
- Stanley, S. M. 2016. Estimates of the magnitudes of major marine mass extinctions in earth history. *PNAS* 113:6325–6334.
- Stanley, S. M., P. W. Signor, S. Lidgard, and A. F. Karr. 1981. Natural clades differ from “random” clades: simulations and analyses. *Paleobiology* 7:115–127.
- Strathmann, R. R., and M. Slatkin. 1983. The improbability of animal phyla with few species. *Paleobiology* 9:97–106.
- Twitchett, R. J. 2001. Incompleteness of the Permian–Triassic fossil record: a consequence of productivity decline? *Geol. J.* 36:341–353.
- Van Valen, L. 1973. A new evolutionary law. *Evol. Theory* 1:1–30.

- Wagner, P. J., and S. K. Lyons. 2010. Estimating extinction with the fossil record. Pp. 265–275 in A. E. Magurran and B. J. McGill, eds. *Biological diversity: frontiers in measurement and assessment*. OUP, Oxford.
- Wahlberg, N., J. Leneveu, U. Kodandaramaiah, C. Peña, S. Nylin, A. V. L. Freitas, and A. V. Z. Brower. 2009. Nymphalid butterflies diversify following near demise at the Cretaceous/Tertiary boundary. *Proc. R Soc. B* 276:4295–4302.
- Wang, S., A. Chen, J. Fang, and S. W. Pacala. 2013. Speciation rates decline through time in individual-based models of speciation and extinction. *Am. Nat.* 182:83–93.
- Westoll, T. S. 1949. On the evolution of the Dipnoi. *Genetics, paleontology and evolution*. Princeton Univ. Press, New Jersey.
- Yule, G. U. 1924. A mathematical theory of evolution, based on the conclusions of Dr. JC Willis, FRS. *Phil. Trans. R Soc. B* 213: 21–87.

Associate Editor: J. Clarke

Handling Editor: Mohamed A. F. Noor

Supporting Information

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

`run_experiments_SIR` gives the R commands necessary to produce each of the figures in the manuscript (excluding figure 3).

`fossil_trees_SIR` contains the source code necessary to perform all the mathematical analyses described in the manuscript, and should be ‘sourced’ prior to running the commands in `run_experiments_SIR`